AEGANB3: An Efficient Framework with Self-attention Mechanism and Deep Convolutional Generative Adversarial Network for Breast Cancer Classification

Huong Hoang Luong[1], Hai Thanh Nguyen[2], Nguyen Thai-Nghe*[3]
FPT University, Can Tho University, Can Tho, Viet Nam[1]
Can Tho University, Can Tho, Viet Nam[2,3]

***Abstract*—Breast cancer remains a significant illness around the world, but it has become the most dangerous when faced with women. Early detection is paramount in improving prognosis and treatment. Thus, ultrasonography has appeared as a valuable diagnostic tool for breast cancer. However, the accurate interpretation of ultrasound images requires expertise. To address these challenges, recent advancements in computer vision such as using convolutional neural networks (CNN) and vision transformers (ViT) for the classification of medical images, which become popular and promise to increase the accuracy and efficiency of breast cancer detection. Specifically, transfer learning and fine-tuning techniques have been created to leverage pre-trained CNN models. With a self-attention mechanism in ViT, models can effectively feature extraction and learning from limited annotated medical images. In this study[3], the Breast Ultrasound Images Dataset (Dataset BUSI) with three classes including normal, benign, and malignant was utilized to classify breast cancer images. Additionally, Deep Convolutional Generative Adversarial Networks (DCGAN) with several techniques were applied for data augmentation and preprocessing to increase robustness and address data imbalance. The AttentiveEfficientGANB3 (AEGANB3) framework is proposed with a customized EfficientNetB3 model and self-attention mechanism, which showed an impressive result in the test accuracy of 98.01%. Finally, Gradient-weighted Class Activation Mapping (Grad-CAM) for visualizing the model decision.**

***Keywords*—Breast cancer; classification; Convolutional Neural Network (CNN); Vision Transformer (ViT); fine-tuning; transfer learning; self-attention**

I. Introduction

Breast cancer stands as one of the most prevalent and concerning malignancies affecting women globally. In addition, breast cancer poses a significant health burden and remains a leading cause of mortality among women. Breast cancer is a formidable enemy, its impact reverberating through the lives of countless individuals and families worldwide. It causes extreme physical, emotional, and socioeconomic consequences not only in women but also in men. The dangerous nature of breast cancer is its potential to metastasize. Thus, the patient needs to understand the mechanisms, risk factors, and manifestations of breast cancer for effective treatment.

Because breast cancer is one of the most common diseases in modern life, there have been many reports about the statistical indicators of this disease. Breast cancer is one of the six most common cancers in the world [1] [2] [3] and it is the leading cause of death in women [1]. In addition, there will be 1,503,694 deaths worldwide from breast cancer in 2050 (i.e., 1,481,463 women and 22,231 males) [4]. Moreover, the GLOBOCAN Cancer Tomorrow prediction tool predicts that breast cancer will rise by more than 46% in 2040 [5]. However, the incidence rates are not equal between countries around the world. For instance, developed countries are higher than developing countries at 88%, with 55.9 and 29.7 per 100,000 women, respectively. In the United States, breast cancer was a cause of death among 909,488 women between 1999 and 2020 [6]. As estimated, the US will have 310,720 new cases of female breast in 2024 [7]. In China, there were about 70,400 deaths and 303,600 new cases of breast cancer in 2015. From 2000 to 2015, the age-standardized incidence and mortality rates rose by 3.3% and 1.0% annually, respectively. It was estimated that these rates would rise by more than 11% until 2030 [8].

To resolve this problem, advancements in medical science have assisted multiple approaches aimed at tackling breast cancer from various angles. From surgery to chemotherapy and radiation therapy, treatment strategies continue to develop and help to improve patient treatment and quality of life. Among these methods, ultrasound images have come out as a valuable tool offering non-invasive and radiation-free breast cancer treatment. Addressing breast cancer requires multiple approaches integrating clinical, pathological, molecular, and imaging aspects. Thus, continual improvements in medicine and computer research are imperative to increase early detection, optimize treatment outcomes, and mitigate the impact of this formidable disease on individuals and society.

Besides, computer vision appeared as a new way for classification and segmentation of a lot of aspects of images. In a subset of computer vision, transfer learning and fine-tuning were used for extracting meaningful information from medical images. These methods have gained considerable attention for their effectiveness in adapting pre-trained convolutional neural networks (CNN) to the specific task of breast cancer analysis. Transfer learning employs knowledge from a pre-trained model on a source task and applies it to a related task with a smaller dataset [9] [10] [11]. On the other hand, fine-tuning requires further refining the parameters and layers of the pre-trained

[3]Corresponding author: Nguyen Thai-Nghe

model on the target task-specific dataset [12] [13] [14] [15]. This approach proves especially beneficial in scenarios where annotated medical image datasets are limited, which facilitates the development of robust classification models for breast cancer detection and characterization.

The advent of Vision Transformer (ViT) architectures represents a significant advancement in the field of medical imaging analysis [16] [17] [18]. Unlike traditional CNN, which relies on hierarchical feature extraction through convolutional layers. ViT introduces a self-attention mechanism that allows for direct interactions between image patches for capturing long-range dependencies within the data. This innovative approach revolutionizes breast cancer classification by enabling the network to dynamically weigh the importance of different image regions, thereby increasing its ability to discern subtle features indicative of malignancy. By using self-attention mechanisms, ViT models demonstrate superior performance in classifying breast cancer images and create more accurate diagnostic outcomes and treatment planning.

Furthermore, a combination of self-attention mechanisms and CNN architectures offers several advantages for breast cancer classification. By selectively attending to relevant image regions, these mechanisms facilitate the extraction of salient features while suppressing noise and irrelevant information. This adaptive focusing capability raises the power of CNN and enables them to effectively differentiate between benign and malignant lesions in breast cancer images. Moreover, self-attention mechanisms enable the network to capture spatial dependencies across multiple scales which allows for a more comprehensive understanding of complex structures within the breast tissue. As a result, CNN models augmented with self-attention mechanisms improve accuracy and reliability in breast cancer classification tasks.

Nowadays, computer technology gives a chance for users a lot of convenience specifically in medical treatment. This study applied several techniques for classifying breast cancer ultrasound images. Begin with applied transfer learning and fine-tuning in CNN and combine a self attention mechanism in ViT. Furthermore, DCGAN was used to augment datasets with new images that are similar to existing ones but slightly different, they can help improve the generalization of machine learning models. Moreover, Grad-CAM was used to explain the classified outcome, which helps describe the model decision.

The contributions of this paper are as follows:

- By using the capabilities of deep learning architectures, this research used DCGAN to facilitate the synthesis of realistic ultrasound images, thereby expanding limited datasets for training classification models. This augmentation process not only raises the diversity and richness of the dataset but also fosters the resilience and efficacy of machine learning algorithms in accurately discerning pathological features indicative of breast cancer.
- This study proposed a combination of CNN with self-attention mechanisms from ViT. It presents a promising approach for classifying ultrasound breast cancer images. By using the ability to extract hierarchical features from CNN and attention mechanism from ViT for capturing global dependencies. As a result, this hybrid architecture increases accuracy and other performance in breast cancer classification.
- Throughout scenarios, the proposed model demonstrated the effectiveness of augmentation techniques and a self-attention mechanism with an accuracy of 98.01%. It has an increase of 13.39% when compared with do not apply any techniques. Thus, these experiments show the AttentiveEfficientGANB3 (AE-GANB3) framework works usefully, thereby indicating its practical capabilities in medical examination and treatment.
- The utilization of Grad-CAM in classifying ultrasound breast cancer images offers insightful interpretability into the decision-making process of deep learning models in this research. By highlighting regions of interest within ultrasound images that contribute most significantly to the classification outcome, Grad-CAM aids clinicians in understanding the model's reasoning, thereby enhancing trust and facilitating informed decision-making in medical diagnostics.

The research paper includes six main parts. First, the opening section offers an introduction. Next, the subsequent section indicates an extensive review of related literature. The third part elucidates the methodology and provides explanations of the employed techniques. Following this, the fourth section delineates the experiments and details their procedures and assessments. Furthermore, the fifth section presents the results of the most important experiment and compares them with existing methods. Finally, the sixth section encapsulates essential findings and offers an analysis.

II. Related Works

CNN and ViT are two prominent methodologies employed in the realm of medical image classification. By using convolutional layers, CNNs can automatically learn relevant features from the input data, which is crucial for discerning between malignant and benign tissues in breast ultrasound scans. In [19], Sathiyabhama Balasubramaniam et al. proposed the LeNet model which applied to breast cancer data analysis and reached a high accuracy of 89.91% when classifying malignant and benign tumors. LeNet CNN is a promising technique that could be used in the future to increase the robustness and accuracy of breast cancer prediction. However, the research did not apply data augmentation to increase the training set and explanation techniques for the outcome to understand the model decision. Besides, Hua Chen et al. used ResNet50 and local binary pattern (LBP) to classify 874 breast ultrasound images (i.e. 457 benign and 417 malignant) and reached a great accuracy of 96.91% as reported in [20]. The research demonstrates that the performance of breast tumor diagnosis may be raised by integrating shallow LBP texture characteristics and multi-level depth features. According to [21]. Mohammed Alotaibi et al. employed the VGG19 model to compare three different image preprocessing procedures in dataset BUSI and gained a surprise mean accuracy of 87.8%. Thus, the study focuses on raising the predictions of deep learning models by using image preprocessing. However, the average accuracy is low which can grow by using and demonstrating the effect of data augmentation techniques.

The advancements in CNN are increasing day by day and help to create a perfect system for the classification of medical images. Clara Cruz-Ramos et al. proposed a DBFS-GMI model based on DenseNet201 and various techniques in [22]. It achieved an impressive accuracy on both datasets mini-DDSM and BUSI of 92% and 96%, respectively. Moreover, a combination of two datasets created an increase in accuracy to 97.6%. As a result, the study has developed a hybrid system that uses the CNN architecture for extracting deep learning features and several classifiers including XGBoost, AdaBoost, and MLP are applied to diagnose breast cancer. In addition, Nasim Sirjani et al. improved the InceptionV3 model and achieved an accuracy of 81% in [23]. However, these experiments run on the dataset combined on various sources which can create an imbalance in the dataset. Thus, this should be resolved by data augment techniques. In [24], Hiba Diaa Alrubaie et al. proposed a new CNN architecture which is combined by several layers such as Conv2D and MaxPooling2D to attain an accuracy of 96% in three classes classifying (i.e. benign, malignant, and normal). However, the article does not mention visual explanation techniques, which can help in the visualization of outcomes.

The versatility and adaptability of CNN make them well-suited for handling the complexities and variabilities present in ultrasound images, which facilitates robust and accurate classification of breast cancer cases. Adyasha Sahu et al proposed a model by combining the benefits of AlexNet, ResNet, and MobileNetV2 and used Laplacian of Gaussian-based modified high boosting filter (LoGMHBF) for preprocessing. As a result, the proposed model achieved the highest accuracy of 96.92% on the BUSI dataset as described in [25]. Additionally, Shao-Hua Chen et al, demonstrated that GoogLeNet and TV models have a huge effect on classifying breast cancer ultrasound images. Through various experiments, authors compare GoogLeNet, VGG16, and LeNet5 to indicate that GoogLeNet has the best accuracy of 96.37% in [26]. Next to that, four different models with VGG-Net, DenseNet, Xception, and Inception were combined to propose a fuzzy-rank-based ensemble network for classifying breast cancer on the BUSI dataset in [27]. Sagar Deep Deb et al. gained a surprising accuracy of 85.25% and they also used Grad CAM for visualization to understand the workings of the proposed model.

Besides studies on the effectiveness of CNN models on ultrasound images, other studies about breast cancer are also provided on Magnetic Resonance Imaging (MRI) or Mammograms. Quy Thanh Lu et al. illustrated the power of a customized MobileNet in classifying multiclass of breast cancer and reached impressive accuracy in four-class classify of 97.24% as reported in [28]. In addition, the study demonstrated the potential of Grad-CAM and other techniques such as data augmentation and preprocessing which increased the model performance and gave a chance to utilize MRI classification in the real world. In [29], Kiran Jabeen et al. indicated enhanced deep learning features and Equilibrium-Jaya controlled Regula Falsi and attained a surprising accuracy on two publicly available datasets CBIS-DDSM and INbreast with an average score of 95.4% and 99.7%, respectively. Thus, the proposed model demonstrated the power of classifying Mammogram images and provided a framework to improve the accuracy. Additionally, Our previous study [30] employed a fine-tuning strategy, ensemble method, and extracting inherent features to improve model reliability and classification accuracy. As a result, the model obtained an accuracy of 76.79% for binary classification.

On the other hand, Vision Transformers, a relatively novel approach, has shown promise in image classification tasks by attending to the global context of the image through self-attention mechanisms. In an experiment of [31], Ishak Pacal proposed a transformer model and compared it with other CNN architecture to see that their model outperforms other models with 88.6% accuracy. Thus, the author indicates deep learning is effective at classifying ultrasound pictures and will soon be able to be utilized in clinical trials. Besides. Behnaz Gheflati et al. proposed a ViT model to classify breast ultrasound images in the dataset BUSI and BUSI + B and achieved accuracies of 82.00% and 86.7% in [32]. In this article, the author tested the B/32 and Resnet50 models and compared the model's outcomes with the corresponding performance of the state-of-the-art. According to [33], Xiaolei Qu et al. also utilized a CNN module to extract local features and a ViT module to determine the global link between various areas to create a VGGA-ViT network. As a result, the proposed gained the highest accuracy 88.7% in dataset BUS-A and the largest accuracy 81.72% in dataset BUS-B.

Despite their architectural differences, both CNN and ViT offer valuable tools for automated diagnosis in medical imaging, contributing to enhanced efficiency and accuracy in breast cancer detection and classification. In addition, ViT is a newer approach and shows promising results in breast cancer classification tasks, albeit with slightly lower accuracies compared to CNN. Future research should focus on addressing dataset imbalances, integrating data augmentation techniques, and implementing visual explanation methods to increase model interpretability. Additionally, exploring hybrid architectures that combine CNN and ViT could further improve classification accuracy.

III. Methodology

A. The Research Implementation Procedure

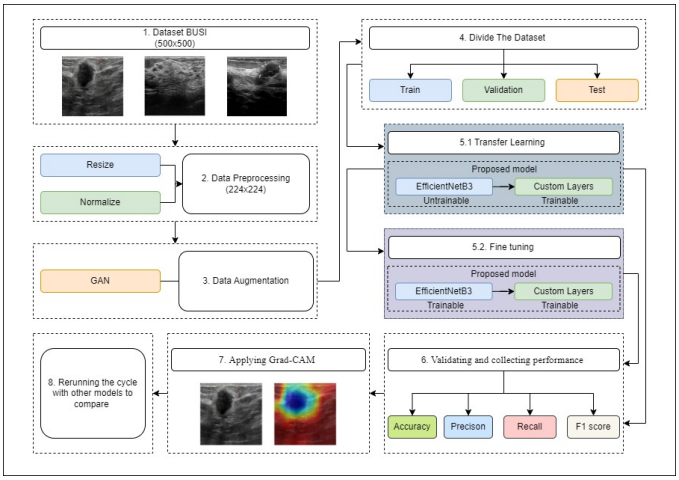

Fig. 1. The AttentiveEfficientGANB3 (AEGANB3) framework was combined with multiple steps which were numbered in detail, including applying GAN and customizing the CNN model.

This research proposed a pipeline consisting of eight steps from input to output shown in Fig. 1. The details of each step are indicated as follows:

1. Dataset BUSI: There are three classifications in the Breast Ultrasound Images Dataset (BUSI): normal, benign, and malignant. The total amount of photos is 780, with an average size of 500 × 500 pixels. Moreover, the LOGIQ E9 ultrasound system and the LOGIQ E9 Agile ultrasound system are tools utilized in the scanning procedure. Additionally, all of the photos were cropped to various proportions to eliminate unnecessary borders. Furthermore, Baheya Hospital radiologists examined and verified every picture.

2. Data Preprocessing: In this step, the technique of resizing and normalizing holds paramount importance. Resizing relates to the transformation of input data to a standardized dimension. Concurrently, normalization scales the data to a common range. Together, these preprocessing steps help to increase precision in model training.

3. Data Augmentation: This augment methodology involves training a DCGAN on existing data to generate additional samples, thereby expanding the dataset size and enhancing its diversity. The integration of GAN-based data augmentation techniques has demonstrated promising results in various domains which indicates its efficacy in raising model generalization and robustness.

4. Divide The Dataset: This scheme allocates 80% of the dataset for training, 10% for validation, and 10% for testing purposes. By following the 8-1-1 scale, this research can effectively measure the performance of breast cancer classification models ensuring reliable results in the domain of medical image analysis.

5.1 Transfer Learning: In transfer learning, a pre-trained CNN model is utilized as a feature extractor, typically trained on a large-scale dataset like ImageNet. The learned features are then used to initialize a new CNN model, which is subsequently fine-tuned on the target ultrasound breast cancer image dataset. This approach allows the model to leverage the knowledge gained from the source domain to effectively learn discriminative features for breast cancer classification.

5.2 Fine-tuning: Fine-tuning updates the parameters of the pre-trained model using backpropagation with the target dataset, thereby adapting the model to the specific characteristics of ultrasound breast cancer images. Furthermore, fine-tuning enables the optimization of model performance by adjusting the hyperparameters and architecture of the pre-trained model to better suit the target task of ultrasound breast cancer classification.

6. Validating and collecting performance: Validating and collecting the performance of models in classifying ultrasound breast cancer images requires the assessment of various metrics including accuracy (ACC), precision, recall, and F1 score. The study employs annotated datasets of ultrasound images and partitions them into training, validation, and testing subsets. Subsequently, the model is trained on the training dataset and fine-tuned using the validation set, while performance metrics such as ACC, precision, recall, and F1 score are computed using the testing set.

7. Applying Grad-CAM: Applying Grad-CAM for classifying ultrasound breast cancer images enhances interpretability and understanding of deep learning models' decision-making processes. Grad-CAM generates heatmaps highlighting regions within ultrasound images for classification decisions. By visualizing these regions, The study gives insights into which features the model prioritizes when distinguishing between benign and malignant lesions

8. Rerunning the cycle with other models to compare: In this phase, the cycle was replayed with other models to compare the performance including EfficientNetB3, DenseNet169, Xception, ViT B16, and ViT B32.

B. Dataset

The BUSI dataset serves as a valuable resource in medical imaging, specifically focusing on breast ultrasound images acquired from female individuals aged between 25 and 75 years old. In addition, this dataset was collected from 600 female patients including 780 images. These images exhibit a consistent average size of 500 by 500 pixels helping to analysis and interpretation in the area of breast cancer detection and diagnosis.

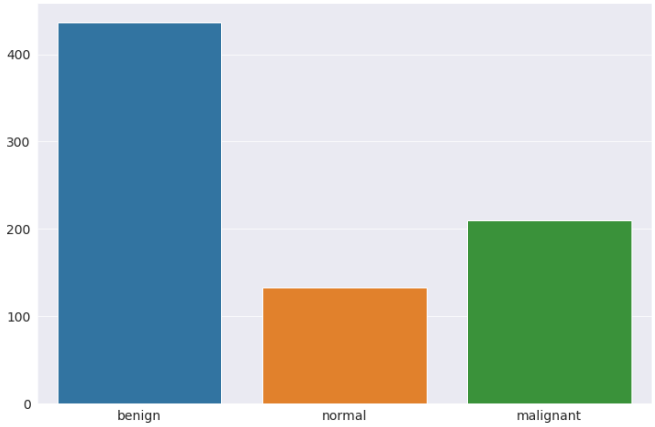

Fig. 2. The distribution between three classes including normal, benign, and malignant in the dataset BUSI.

However, a challenge in the BUSI dataset stays in its class imbalance, which could potentially skew the performance of machine learning algorithms trained on it. The distribution across the classes reveals a notable disproportion in Fig. 2, with 437 instances classified as benign, 133 as normal, and 210 as malignant. Such an imbalance poses a significant obstacle undermining their ability to accurately discern minority classes.

To mitigate this issue and increase the richness of the data set in machine learning applications. Thus, data augmentation techniques prove helpful. By augmenting the minority classes, the balance can be rectified and created equitable across all

classes. Through augmentation in Fig. 3, the instances within the benign, normal, and malignant classes can be increased to 1357, 1333, and 1330, respectively. This augmentation process not only rectifies the class imbalance but also enriches the dataset which improves performance in breast cancer detection and classification endeavors.

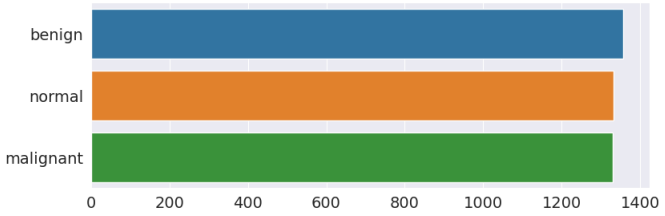

Fig. 3. The distribution between three classes in the dataset BUSI after augmentation

C. Data Preprocessing

Data preprocessing is important to ensuring the quality and efficacy of subsequent classification tasks. In the context of ultrasound images for breast cancer classification. Two fundamental preprocessing techniques resizing Eq. (1) and normalization Eq. (2) are integral steps in raising the interpretability and efficiency of classification algorithms.

The resize technique is employed to standardize the dimensions of ultrasound images. In detail, resizing from a larger dimension, such as 500x500 pixels, to a smaller dimension, like 244x224 pixels, is utilized in this study. In this resizing process, each intensity values are recalculated to fit the new dimensions while preserving the structural features essential for accurate classification. Mathematically, Let $I_{original}$ Eq. (1) denote the original ultrasound image with dimensions 500x500 pixels, and $I_{resized}$ Eq. (1) indicates the resized image with dimensions 244x224 pixels. The resizing operation can be expressed as:

$$I_{resized} = resize(I_{original}, (224, 224)) \tag{1}$$

Where $(224, 224)$ Eq. (1) illustrates the height and width of the resized image. Moreover, the pseudo-code of the resize algorithms is provided in Algorithms 1 which represents an overview of the code flow.

On the other hand, normalization assists in standardizing the pixel intensities in the ultrasound images increasing comparability and mitigating the effects of variations in illumination and contrast. By scaling the intensity values to a common range, typically between 0 and 1, normalization facilitates optimal convergence during the training phase of classification models. Mathematically, the normalization process can be represented as:

$$O(x, y) = \frac{I_{\text{resized}}\,(x, y) - \min(I_{\text{resized}}\,)}{\max(I_{\text{resized}}\,) - \min(I_{\text{resized}}\,)} \tag{2}$$

The normalization equation presented calculates the normalized pixel value $O(x, y)$ Eq. (2) at a specific position (x, y) Eq. (2) in the resized image. It involves dividing the pixel value of the resized image $I_{\text{resized}}\,(x, y)$ Eq. (2) by the range of pixel values in the resized image, which is determined by subtracting the minimum pixel value $\min(I_{\text{resized}}\,)$ Eq. (2) from the maximum pixel value $\max(I_{\text{resized}}\,)$ Eq. (2). This normalization process Eq. (2) ensures that all pixel values in the resized image fall within the range of [0, 1].

Algorithm 1 Resizing Algorithm

Require: Original Image, target_size
Ensure: Resized Image
1: Load the Original Image:
2: Define the target_size = (224,224)
3: Resize the Original Image to the target_size using the resize function:
4: $ResizedImage = \text{resize}(OriginalImage, (224, 224))$
5: **return** Resized Image

As outlined in Algorithm 2, the normalization algorithm computes the minimum and maximum pixel values present within the image. Subsequently, it iterates over each pixel in the image, normalizing its intensity value to fall within the range [0, 1]. This normalization process enhances the comparability and interpretability of images across various datasets and facilitates subsequent analysis, such as feature extraction and classification.

Algorithm 2 Normalization Algorithm

Require: Image to normalize: image
Ensure: Normalized image: normalized_image
1: min_pixel_value ← min(image)
2: max_pixel_value ← max(image)
3: **for** each pixel **in** image **do**
4: normalized_image$[x, y]$ ← $\frac{\text{image}[x,y] - \text{min\_pixel\_value}}{\text{max\_pixel\_value} - \text{min\_pixel\_value}}$
5: **end for**
6: **return** normalized_image

In conclusion, the integration of resizing and normalization techniques in the preprocessing pipeline for ultrasound images in breast cancer classification not only standardizes the data but also enhances the robustness and performance of subsequent classification algorithms. These preprocessing steps are essential for optimizing the accuracy and reliability of diagnostic systems aimed at early detection and intervention in breast cancer cases.

D. Data Augmentation with DCGAN

DCGAN have gained significant attention in recent years for their ability to generate synthetic data closely resembling real data. In medical imaging, DCGAN holds promise for tasks such as image synthesis, data augmentation, and anomaly detection. These images can then be used to augment the dataset for training a classification model, thereby improving its performance and generalization ability.

In Fig. 4, the Generator model is designed to generate synthetic ultrasound images copying real breast tissue images. The architecture comprises several layers, including dense, convolutional, and upsampling layers. The input to the Generator is a latent vector, typically drawn from a Gaussian

distribution, which is transformed into a high-dimensional representation through dense layers. Subsequently, upsampling layers increase the spatial resolution of the representation, generating images of the desired size. Batch normalization and activation functions such as ReLU ensure stable training and introduce non-linearity, respectively. The final layer produces synthetic images with pixel values normalized between 0 and 1.

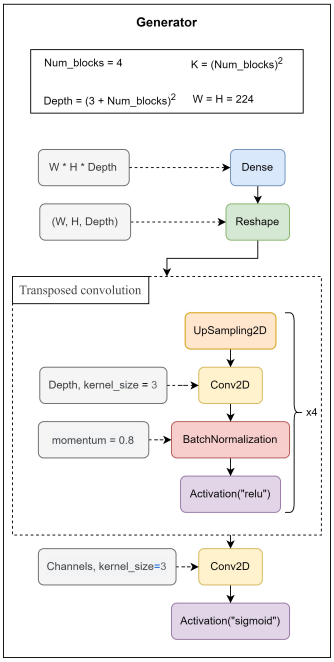

Fig. 4. The generator model of DCGAN.

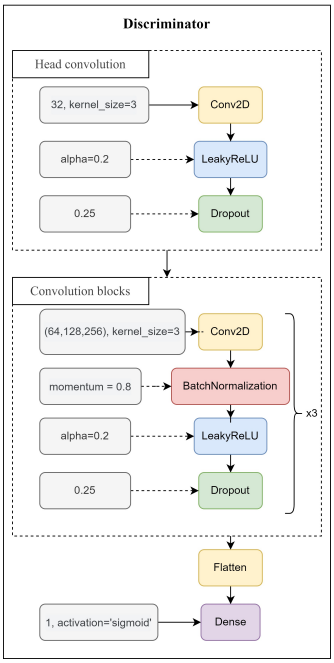

Fig. 5. The discriminator model of DCGAN.

According to the Fig. 5, the Discriminator model is responsible for distinguishing between real ultrasound images and synthetic images generated by the Generator. It consists of convolutional layers followed by batch normalization, leaky ReLU activation, and dropout layers. The architecture progressively downsamples the input images, extracting hierarchical features. The final layer performs binary classification, outputting the probability that the input image is real.

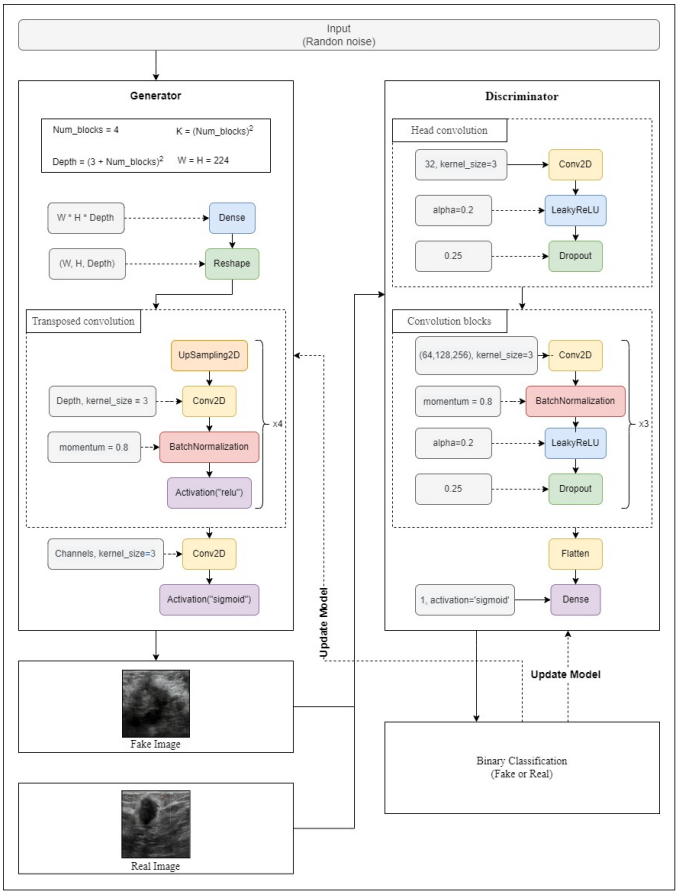

Fig. 6. The architecture of DCGAN.

During training in Fig. 6, the Generator and Discriminator are trained simultaneously in a min-max game. The Generator aims to generate images that are indistinguishable from real images, while the Discriminator aims to correctly classify between real and fake images. The two models are trained iteratively, with the Generator trying to minimize the probability of the Discriminator correctly classifying fake images, and the Discriminator trying to maximize this probability.

By iteratively updating the Generator and Discriminator models, the DCGAN learns to generate realistic ultrasound images, which can subsequently be used for tasks such as breast cancer classification. Integrating GAN-generated images into the training data can potentially improve the robustness and performance of classification models by providing additional diverse examples for learning. Moreover, future research directions include fine-tuning the DCGAN architecture, incorporating additional modalities, and expanding the dataset to improve generalization performance.

The proposed approach leverages adversarial training to

generate synthetic images that closely resemble real ultrasound images of breast tissue. Experimental results demonstrate the potential of DCGAN in enhancing the availability and diversity of medical image data for improving diagnostic accuracy in breast cancer detection.

E. Transfer Learning and Fine-tuning in AttentiveEfficient-GANB3

Transfer learning and fine-tuning are powerful techniques of deep learning, especially when dealing with tasks like image classification and segmentation. These methods allow using pre-trained models on large datasets and adapting them to new tasks with smaller datasets, thereby saving computational resources and time.

Transfer learning uses a pre-trained model which is usually trained on a large dataset like ImageNet and applying it to a new task. Instead of starting the training process from scratch, the knowledge of a model is transferred to the new task, particularly in extracting useful features from images. This is often achieved by removing the final classification layer of the pre-trained model and replacing it with a new layer suited to the specific task. On the other hand, Fine-tuning takes transfer learning a step further by not only adapting the final layers but also fine-tuning some of the earlier layers of the pre-trained model. This allows the model to adjust its learned representations to better suit the new task while still benefiting from the general features learned from the original dataset.

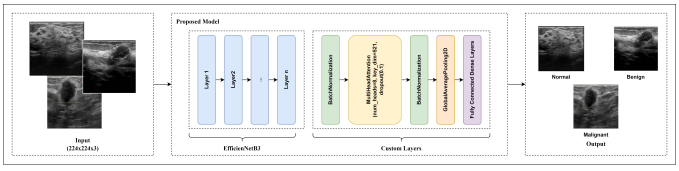

Fig. 7. The architecture of the proposed model.

In the research, transfer learning and fine-tuning can significantly improve model performance, especially when dealing with limited medical image datasets. In Fig. 7, the proposed model architecture utilizes EfficientNetB3 as the base model, which is known for its effectiveness in balancing model size and performance across various image classification tasks. Moreover, the proposed architecture integrates custom layers to further enhance its capabilities. One notable addition is the MultiHeadAttention layer, which introduces a mechanism for the model to focus on different parts of the input data independently. In the context of ultrasound images, this attention mechanism can help the model to effectively identify relevant features associated with breast cancer, thereby improving classification accuracy.

Fig. 7 includes BatchNormalization layers to stabilize and speed up the training process by normalizing the inputs to each layer. GlobalAveragePooling2D layer is used to reduce the spatial dimensions of the feature maps produced by the base model before feeding them into the final classification layers. The Dense layer serves as the final classification layer, where the model outputs predictions regarding the presence or absence of breast cancer based on the extracted features.

By using transfer learning from EfficientNetB3 and fine-tuning with custom layers such as MultiHeadAttention, the proposed model achieved strong performance in classifying breast cancer on ultrasound images, even with limited labeled and imbalanced data.

F. Visual Explanation with Gradcam

Grad-CAM is a technique used for visualizing the regions of an image that are influential in the decision-making process of a deep neural network model. It highlights the regions that the model focuses on when classifying an image. In this study, Grad-CAM can help identify the specific areas of an ultrasound image that contribute most significantly to the model's decision regarding the presence or absence of cancerous tissue. The process begins with a feedforward pass of the ultrasound image through a CNN model. This leads to create the generation of feature maps across various convolutional layers. Following this, The gradient of the score of the target class for the feature maps of the final convolutional layer is calculated. Mathematically, this can be represented as (Fig. 8):

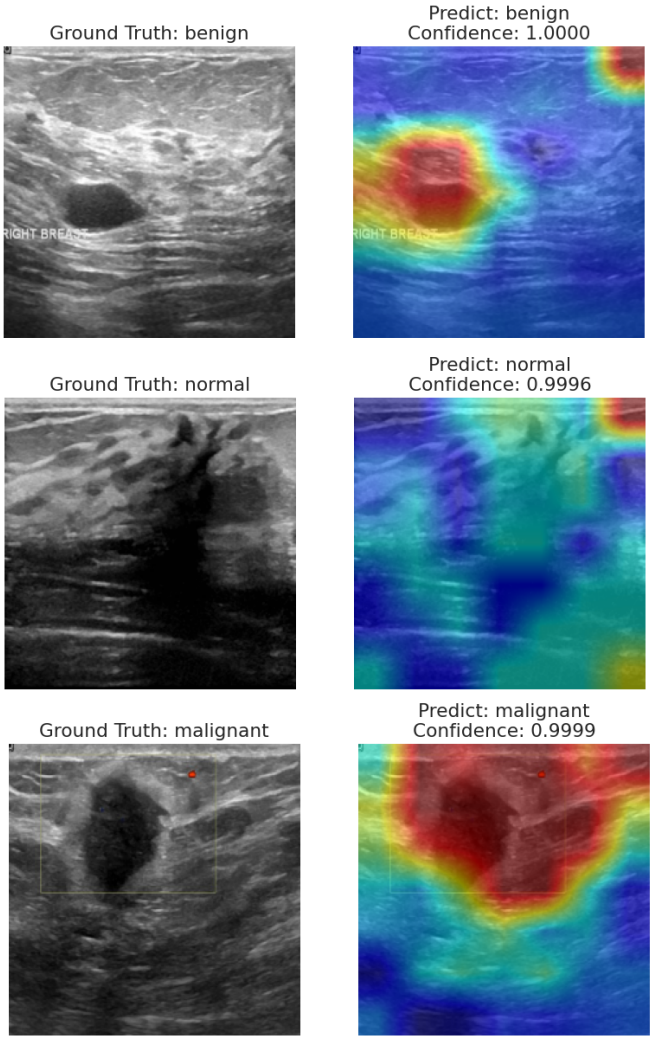

Fig. 8. The result of applying heatmap to the ultrasound image.

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial A_{ij}^k}{\partial y^c} \tag{3}$$

Where α_k^c Eq. (3) represents the importance weight associated with the $k-th$ Eq. (3) feature map for the $c-th$ Eq. (3) class. In addition, Z Eq. (3) is a normalization factor to ensure that the weights sum up to 1, preventing issues with the scale of the gradient values. Moreover, $\partial A_{ij}^k / \partial y^c$ Eq. (3) represents the partial derivative of the output score to the activation map A_{ij}^k Eq. (3). It quantifies how changes in the activation map affect the model's confidence score for class c Eq. (3). Next to that, the weighted combination step assigns the gradients of each feature map. This is achieved by weighting the gradients and applying a Rectified Linear Unit (ReLU) Eq. (4) activation function to ensure only positive influences are considered. Mathematically:

$$L_{\text{Grad-CAM}}^c = \text{ReLU}\left(\sum_k \alpha_k^c A^k\right) \tag{4}$$

Here, $L_{\text{Grad-CAM}}^c$ Eq. (4) represents the Grad-CAM heatmap for the $c-th$ Eq. (4) class. Additionally, ReLU() Eq. (4) indicates the Rectified Linear Unit activation function, which sets negative values to zero and keeps positive values unchanged. Besides, α_k^c Eq. (4) denotes the importance weight associated with the $k-th$ Eq. (4) feature map for the $c-th$ Eq. (4) class and A^k Eq. (4) signifies the $k-th$ Eq. (4) feature map from the final convolutional layer of the CNN. The equation computes a weighted sum of the feature maps A^k Eq. (4) based on their importance weights α_k^c Eq. (4) for the class c Eq. (4). Finally, This weighted sum is then passed through the ReLU activation function to generate the Grad-CAM heatmap. This heatmap effectively highlights the regions within the ultrasound image that are critical for the decision-making process. By overlaying this heatmap onto the original ultrasound image, researchers and clinicians gain valuable insights into the specific areas that contribute to the model's classification

IV. Experiments

A. Performance Metrics

In assessing the performance of breast cancer classification on ultrasound images, several metrics are commonly used: accuracy (ACC), precision, recall, and F1 score. These metrics help quantify the effectiveness of a classification model in correctly identifying cancerous and non-cancerous cases.

Accuracy Eq. (5) measures the overall correctness of the classification model and is calculated as the ratio of correctly classified instances to the total instances:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

In Eq. (5), TP (True Positives) represents the number of correctly classified cancerous cases, TN (True Negatives) is the number of correctly classified non-cancerous cases, FP (False Positives) is the number of non-cancerous cases wrongly classified as cancerous, and FN (False Negatives) is the number of cancerous cases wrongly classified as non-cancerous.

Precision Eq. (6) measures the proportion of correctly identified cancerous cases among all cases classified as cancerous. As a result, it highlights the model's ability to avoid misclassifying non-cancerous cases as cancerous:

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

Recall Eq. (7) measures the proportion of correctly identified cancerous cases among all actual cancerous cases. Thus, it indicates the model's ability to correctly detect cancerous cases:

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

The F1 score Eq. (8) is the harmonic mean of precision and recall, providing a single metric that balances between precision and recall. Hence, it gives an overall measure of the model's accuracy in identifying both cancerous and non-cancerous cases while considering the trade-off between precision and recall.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{8}$$

These metrics collectively offer a comprehensive evaluation of the performance of breast cancer classification on ultrasound images, aiding in the assessment and comparison of different classification models.

B. Scenario 1: The Performance of Classifying the Dataset without the Augmentation Method

TABLE I. The Result in Performances of Classifying Ultrasound Images without Data Augmentation Techniques

Model	Number of Parameters	Phase	Accuracy		Others metrics		
			Validation	Test	Precision	Recall	F1
DenseNet169	12.647.875	Transfer Learning	70.51%	62.82%	61.60%	62.82%	62.06%
		Fine Tuning	76.92%	73.08%	72.02%	73.08%	72.23%
Xception	20.867.627	Transfer Learning	75.64%	64.10%	63.94%	64.10%	64.00%
		Fine Tuning	80.77%	74.36%	74.34%	74.36%	73.70%
ViT B16	85.800.963	Transfer Learning	73.08%	61.54%	52.52%	61.54%	55.40%
		Fine Tuning	74.36%	66.67%	67.23%	66.67%	65.18%
ViT B32	87.457.539	Transfer Learning	69.23%	67.95%	70.22%	67.95%	64.51%
		Fine Tuning	74.36%	65.38%	66.98%	65.38%	63.72%
Proposed	**36.763.954**	**Transfer Learning**	**87.18%**	**84.62%**	**85.30%**	**84.62%**	**84.67%**
		Fine Tuning	**87.18%**	**88.46%**	**88.53%**	**88.46%**	**88.47%**

Table I presents the performance results of classifying ultrasound images without utilizing data augmentation techniques. It evaluates various models based on their accuracy during both the validation and test phases, precision, recall, and F1 score. Among the models assessed, DenseNet169, Xception, ViT B16, and ViT B32 are included. These models run over two phases: transfer learning and fine-tuning. Notably, the proposed model achieves an accuracy of 87.18% in validation and an impressive 88.46% in test of the fine-tuning phase. Despite having a larger number of parameters, ViT B16 and ViT B32 models show comparatively lower performance metrics than some other models in the table. For instance, ViT B16 has 85,800,963 parameters with a test accuracy of 67,95%, while ViT B32 has 87,457,539 parameters with a test accuracy of 65,35%, both significantly more than the proposed model

Fig. 9. The line graph illustrates about training and validation phases of accuracy in the experiment without data augmentation methods.

with 36,763,954 parameters. In addition, the performance in precision, recall, and F1 of the proposed model also achieved high scores of 85.30%, 84.62%, and 84.67%, respectively.

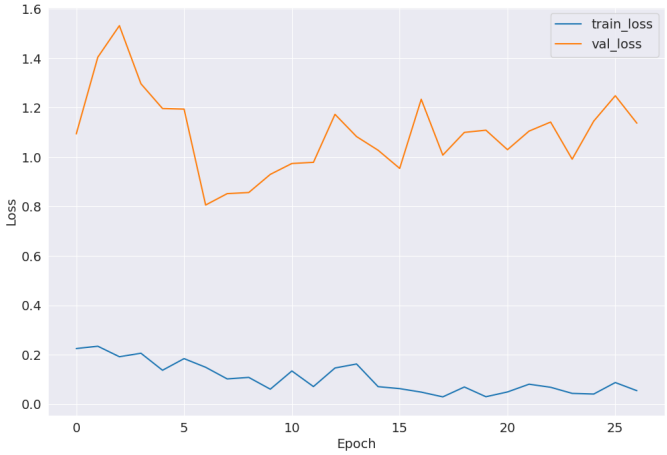

Fig. 10. The line graph illustrates about training and validation phases of loss in the experiment without data augmentation methods.

With line graphs in Fig. 9 and 10, these graphs show the trend of accuracy and loss scores during the training and validation phases. In Fig. 9, The line graph illustrating the training and validation phases of accuracy in the experiment without data augmentation methods showcases the performance of the model throughout the training process. In this specific experiment, the training accuracy reaches a high of approximately 98.08%, while the validation accuracy peaks at around 87.18%. On the other hand, Fig. 10 illustrating the training and validation phases of loss in the same experiment depicts the convergence of the model's loss function during training. In this case, the training loss reaches a low of approximately 0.0603, while the validation loss peaks at around 0.9298 during the fine-tuning phase. Besides, The confusion matrix in Fig. 11 helps evaluate the performance of breast cancer classification models by providing insight into actual and predicted percentages, enabling assessment of model accuracy and error types.

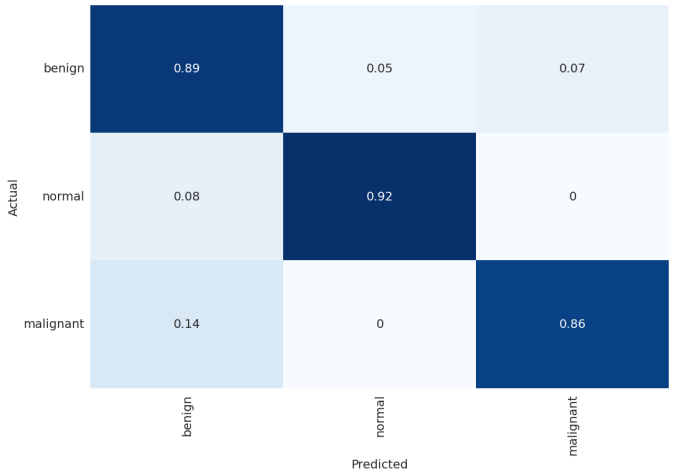

Fig. 11. The confusion matrix in the experiment without applying data augmentation methods.

C. Scenario 2: The Performance of Classifying the Dataset with Simple Augmentation Methods Such as Rotation, Flip, etc

TABLE II. THE RESULT IN PERFORMANCES OF CLASSIFYING ULTRASOUND IMAGES WITH SIMPLE DATA AUGMENTATION TECHNIQUES

Model	Number of Parameters	Phase	Accuracy		Others metrics		
			Validation	Test	Precision	Recall	F1
DenseNet169	12.647.875	Transfer Learning	70.30%	67.59%	68.15%	67.59%	67.65%
		Fine Tuning	89.10%	89.77%	89.75%	89.77%	89.73%
Xception	20.867.627	Transfer Learning	59.19%	61.62%	62.17%	61.62%	61.60%
		Fine Tuning	73.08%	74.84%	75.02%	74.84%	74.87%
ViT B16	85.800.963	Transfer Learning	59.62%	60.98%	61.71%	60.98%	61.13%
		Fine Tuning	69.02%	67.59%	69.07%	67.59%	67.36%
ViT B32	87.457.539	Transfer Learning	55.34%	55.86%	56.16%	55.86%	55.59%
		Fine Tuning	58.76%	56.29%	58.70%	56.29%	53.84%
Proposed	**36.763.954**	**Transfer Learning**	**92.31%**	**92.96%**	**92.99%**	**92.96%**	**92.93%**
		Fine Tuning	**94.66%**	**95.31%**	**95.36%**	**95.31%**	**95.29%**

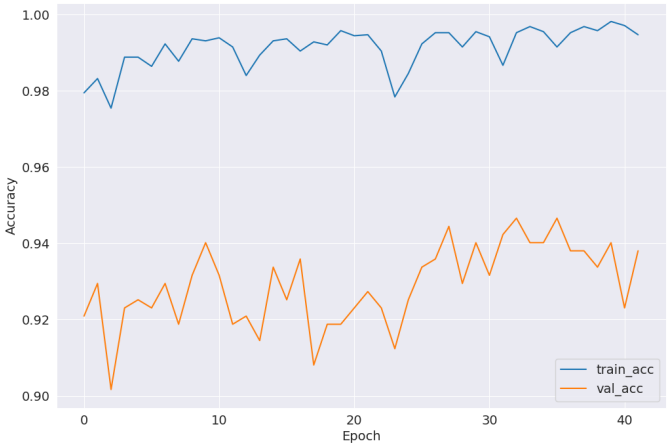

Fig. 12. The line graph illustrates about training and validation phases of accuracy in the experiment using simple data augmentation methods.

In Table II, various models are evaluated for their performance in classifying ultrasound images using simple data augmentation techniques. Notably, the proposed model stands out with the highest accuracy rates in both validation and test phases surpassing all other models. Specifically, in the fine-tuning phase, the proposed model achieved an impressive 94.66% accuracy on the validation set and 95.31% on the test set. This significant increase in accuracy suggests that the

Proposed model exhibits superior performance compared to the other models. Considering the other models, DenseNet169 has the largest growth of 22.18% between the two phases in the test set indicating that DenseNet169 is consistent with the augmentation techniques in this experiment. Besides, ViT B16 saw a slight increase when compared with Table I. On the opposite, ViT B16 fell significantly which showed that ViT did not adapt to several simple augmentation techniques.

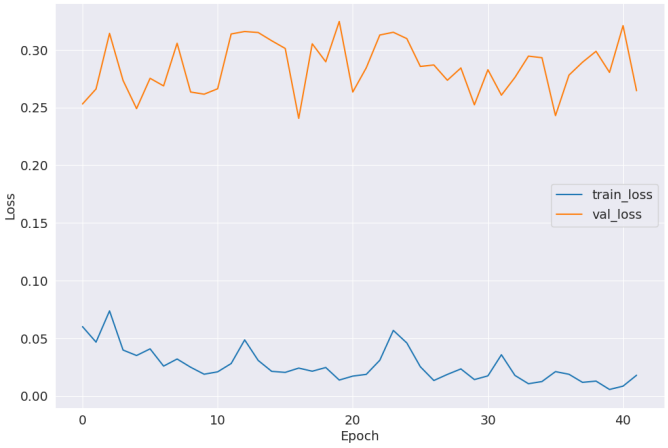

Fig. 13. The line graph illustrates about training and validation phases of loss in the experiment using simple data augmentation methods.

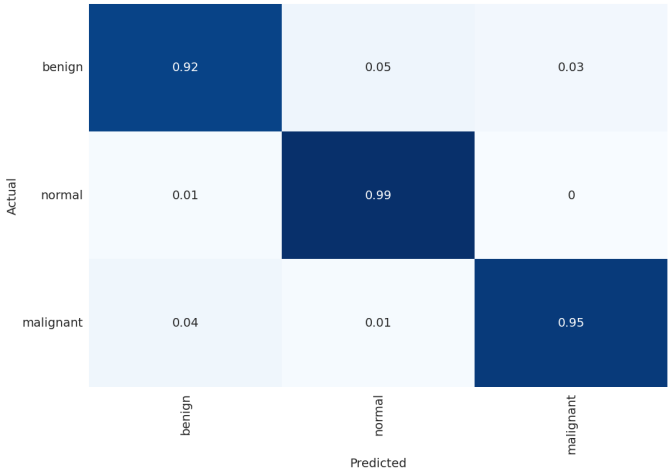

Fig. 14. The confusion matrix in the experiment using simple data augmentation methods.

The accuracy and loss scores for the two training and validation stages are shown in Fig. 12 and 13. This line chart facilitates general evaluation during the training epoch by presenting the accuracy and loss scores in an easy-to-understand and intuitive manner. Moreover, the efficacy of deep learning models for breast cancer categorization is evaluated using the confusion matrix presented in Fig. 14. Normal indicates an impressive percentage between actual and predicted of 99%. Next, benign and malignant have a huge proportion of 92% and 95%, respectively.

D. Scenario 3: The Performance of Classifying the Dataset with DCGAN Augmentation Methods

TABLE III. THE RESULT IN PERFORMANCES OF CLASSIFYING ULTRASOUND IMAGES WITH DCGAN DATA AUGMENTATION TECHNIQUE

Model	Number of Parameters	Phase	Accuracy		Others metrics		
			Validation	Test	Precision	Recall	F1
DenseNet169	12.647.875	Transfer Learning	94.78%	94.53%	94.64%	94.53%	94.52%
		Fine Tuning	97.01%	97.51%	97.51%	97.51%	97.51%
Xception	20.867.627	Transfer Learning	84.83%	83.58%	85.18%	83.58%	83.71%
		Fine Tuning	96.02%	94.78%	94.83%	94.78%	94.79%
ViT B16	85.800.963	Transfer Learning	95.27%	94.03%	94.03%	94.03%	94.03%
		Fine Tuning	95.27%	93.78%	94.24%	93.78%	93.85%
ViT B32	87.457.539	Transfer Learning	94.03%	93.03%	93.44%	93.03%	93.09%
		Fine Tuning	95.77%	94.28%	94.63%	94.28%	94.33%
Proposed	**36.763.954**	**Transfer Learning**	**97.26%**	**96.52%**	**96.52%**	**96.52%**	**96.52%**
		Fine Tuning	**97.76%**	**98.01%**	**98.01%**	**98.01%**	**98.01%**

The proposed model achieves impressive results in both transfer learning and fine-tuning phases in Table III. In transfer learning, the model achieves a validation accuracy of 97.26% and a test accuracy of 96.52%. Fine-tuning further enhances performance, with validation and test accuracies reaching 97.76% and 98.01%, respectively. Precision, recall, and F1-score metrics also demonstrate high values of 96.52% and 98.01% across both phases, indicating robust performance in classifying ultrasound images. Among other models, DenseNet169 exhibits competitive performance, especially in fine-tuning, with a test accuracy of 97.51%. Xception, although having fewer parameters compared to DenseNet169, demonstrates slightly lower accuracy in both transfer learning and fine-tuning phases. ViT B16 and ViT B32 also exhibit respectable performance, albeit with varying degrees of accuracy across transfer learning and fine-tuning. The comparative analysis highlights the efficacy of the proposed model utilizing GAN data augmentation in ultrasound image classification.

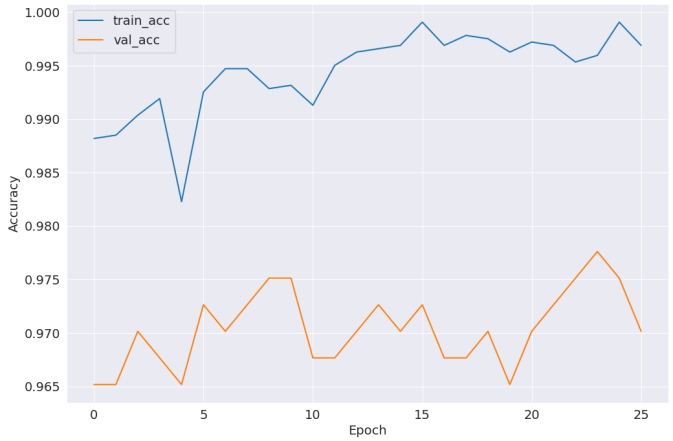

Fig. 15. The line graph illustrates about training and validation phases of accuracy in the experiment employing GAN.

Furthermore, Training and validation on both accuracy and loss scores are presented in Fig. 15 and 16. Following the figures, the evaluation performance of our model presents the balance when the dataset is changed. Moreover, Fig. 17 is provided for evaluating, optimizing, and understanding the performance of deep learning models in classifying breast cancer providing insight into actual and predicted rates that can lead to improved accuracy and reliability.

Fig. 16. The line graph illustrates about training and validation phases of loss in the experiment employing GAN.

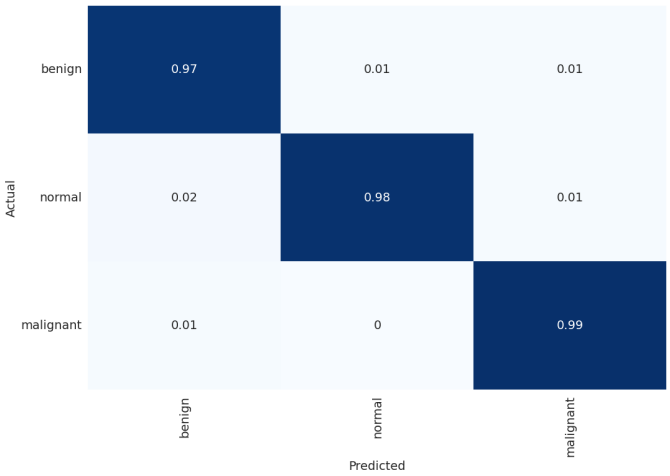

Fig. 17. The confusion matrix in the experiment employing GAN.

E. Scenario 4: The Influence of the Self-attention Mechanism on Performance over Experiments

TABLE IV. PERFORMANCE COMPARISON IN RESULTS BETWEEN WITH AND WITHOUT MULTI-HEAD ATTENTION

Data Augmentation	Model	Phase	Accuracy		Others metrics		
			Validation	Test	Precison	Recall	F1
No-Augmentation	Without Attention	Transfer learning	82.05%	82.05%	82.29%	82.05%	81.98%
		Fine tuning	82.05%	84.62%	86.78%	84.62%	85.01%
	Attention	Transfer Learning	87.18%	84.62%	85.30%	84.62%	84.67%
		Fine Tuning	87.18%	88.46%	88.53%	88.46%	88.47%
Simple Augmentation	Without Attention	Transfer Learning	83.76%	83.58%	83.78%	83.58%	83.63%
		Fine Tuning	92.95%	92.54%	92.57%	92.54%	92.54%
	Attention	Transfer Learning	92.31%	92.96%	92.99%	92.96%	92.93%
		Fine Tuning	94.66%	95.31%	95.36%	95.31%	95.29%
GAN	Without Attention	Transfer Learning	97.26%	97.01%	97.05%	97.01%	97.02%
		Fine Tuning	97.76%	97.26%	97.29%	97.26%	97.27%
	Attention (Proposed)	**Transfer Learning**	**97.26%**	**96.52%**	**96.52%**	**96.52%**	**96.52%**
		Fine Tuning	**97.76%**	**98.01%**	**98.01%**	**98.01%**	**98.01%**

Table IV provides a comprehensive comparison of model performance with and without multi-head attention across different phases and data augmentation scenarios. It primarily focuses on test accuracy and other relevant metrics like precision, recall, and F1 score.

When analyzing the results, it is clear that models with attention consistently outperform those without attention in terms of accuracy. This improvement is especially notable when data augmentation techniques are applied. For instance, in the Simple Augmentation scenario, the test accuracy increases from 83.58% to 92.96% when the attention mechanism is added to the model. The proposed attention model in the DCGAN data augmentation scenario shows superior performance compared to other configurations. In the Fine Tuning phase, the proposed attention model achieves a remarkable test accuracy of 98.01%, indicating the effectiveness of the multi-head attention mechanism.

In comparison to the first experience without applied DCGAN and Attention mechanism, the model increased by 13.39% between 98.01% and 84.62% in the fine-tuning phase of test accuracy. The observed increase in test accuracy across various experiments underscores the significance of incorporating multi-head attention mechanisms in deep learning models for enhanced performance across diverse tasks and datasets.

V. RESULTS AND COMPARISON

A. Results

After analyzing the previous scenarios, Fig. 18 was created to visualize the result in the past experiments. Specifically, the DCGAN technique demonstrated the effectiveness on dataset BUSI with an increase of 9.55% in test accuracy when compared without the augmentation technique. Moreover, the proportion is larger than 2.65% when compared with simple augmentation techniques. Other performances such as precision, recall, and f1 score also witnessed a dramatic climb with DCGAN. Besides, the result of a combination of self-attention mechanism was presented in Table IV. Thus, It indicated AE-GANB3 framework truly helped in the classification process with a surprising rise in accuracy by 13.39% from 98.01% to 84.62%. In conclusion, the proposed framework has actively contributed to the process of researching image classification using machine learning

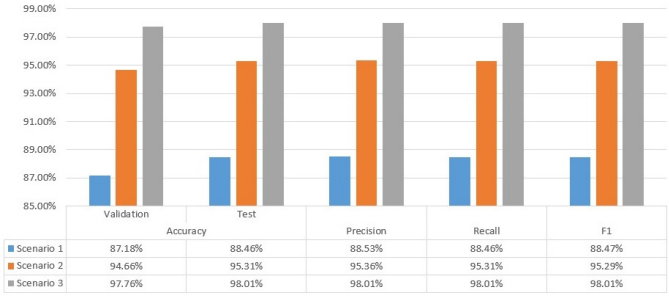

Fig. 18. The result of comparison over scenarios.

B. Comparison with others State-of-the-art Methods

Utilizing comparisons with other state-of-the-art methods is an integral aspect of research. These comparisons serve multiple purposes within the scientific community. Firstly, they establish benchmarks against which new methods can be evaluated, providing a baseline for assessing performance improvements. Secondly, such comparisons validate the effectiveness of proposed approaches, strengthening the case for their adoption. Additionally, they aid in identifying limitations or weaknesses in existing methods, offering insights for further refinement. Understanding how a new method

compares to others also provides context for its significance and relevance within the field, highlighting its innovative contributions. Moreover, comparisons can inspire new ideas for improvement by analyzing the strengths and weaknesses of existing approaches. Thus, Table V was created for comparisons rigorously, considering factors such as dataset and evaluation metrics.

TABLE V. COMPARISON WITH OTHER STATE-OF-THE-ART METHODS IN DATASET BUSI

Reference	Other methods	Year	Accuracy
Mohammed Alotaibi et al. [21]	VGG19	2023	87.8%
Clara Cruz-Ramos et al. [22]	DBFSGMI	2023	92%˜97.6%
Adyasha Sahu et al. [25]	CNN and LoGMHBF	2024	96.92%
Sagar Deep Deb et al. [27]	FRBEN	2023	85.23%
Ishak Pacal [31]	CNN +ViT	2022	88.6%
Behnaz Gheflati et al. [32]	ViT	2022	82%˜86.7%
Proposed Model			98.01%

VI. CONCLUSION

In conclusion, this research harnesses the power of deep learning architectures to address crucial challenges in medical imaging, particularly in the early detection of breast cancer. Through the utilization of DCGAN for synthesizing realistic ultrasound images and augmenting datasets, coupled with a novel hybrid CNN and ViT architecture. The study aimed to enhance the accuracy and efficacy of breast cancer classification models. The AttentiveEfficientGANB3 (AEGANB3) framework was proposed with its incorporation of augmentation techniques and self-attention mechanisms. Thus, it showed a remarkable improvement in classification accuracy, reaching an impressive 98.01% in the test set. Moreover, the integration of Grad-CAM provides valuable insights into the decision-making process of deep learning models, which enhances interpretability and fostering trust.

However, it is essential to acknowledge the limitations of this research. One such limitation is the reliance on synthetic data generated by DCGAN, which may not fully capture the variability and complexity present in real-world ultrasound images. Additionally, the interpretability provided by Grad-CAM, while insightful, may not encompass the full spectrum of factors influencing model decisions. Looking ahead, future research endeavors should aim to address these limitations and further increase the robustness and generalization capabilities of breast cancer classification models. This could involve exploring alternative data augmentation techniques, such as generative adversarial networks with more advanced architectures.

In summary, while this research represents a significant step forward in leveraging deep learning for breast cancer detection, there remain opportunities for further innovation and refinement. By addressing the identified limitations and pursuing avenues for future work, the future study can continue to advance the field of medical imaging and contribute to improved patient outcomes in the fight against breast cancer.

AVAILABILITY OF DATA, CODE, AND MATERIAL

Data for this study are published on repository link at[1] and code is at[2]

[1]https://doi.org/10.1016/j.dib.2019.104863

[2]https://github.com/lhhuong/AEGANB3

ACKNOWLEDGMENT

Luong Hoang Huong was funded by the Vingroup Innovation Foundation (VINIF) 's Master, Ph.D. Scholarship Programme, code VINIF.2023.TS.049.

We would like to extend our heartfelt gratitude to Hao Van Tran, and Phuc Tan Huynh for their invaluable contributions to this project. Their dedication, expertise, and unwavering support have been instrumental in its success.

REFERENCES

[1] B. S. Chhikara and K. Parang, "Global cancer statistics 2022: the trends projection analysis," *Chemical Biology Letters*, vol. 10, no. 1, pp. 451–451, 2023.

[2] M. R. De Miglio and C. Mello-Thoms, "Reviews in breast cancer," *Frontiers in Oncology*, vol. 13, p. 1161583, 2023.

[3] K. M. Cuthrell and N. Tzenios, "Breast cancer: Updated and deep insights," *International Research Journal of Oncology*, vol. 6, no. 1, pp. 104–118, 2023.

[4] Y. Xu, M. Gong, Y. Wang, Y. Yang, S. Liu, and Q. Zeng, "Global trends and forecasts of breast cancer incidence and deaths," *Scientific Data*, vol. 10, no. 1, p. 334, 2023.

[5] E. Heer, A. Harper, N. Escandor, H. Sung, V. McCormack, and M. M. Fidler-Benaoudia, "Global burden and trends in premenopausal and postmenopausal breast cancer: a population-based study," *The Lancet Global Health*, vol. 8, no. 8, pp. e1027–e1037, 2020.

[6] T. D. Ellington, S. J. Henley, R. J. Wilson, J. W. Miller, M. Wu, and L. C. Richardson, "Trends in breast cancer mortality by race/ethnicity, age, and us census region, united states- 1999-2020," *Cancer*, vol. 129, no. 1, pp. 32–38, 2023.

[7] R. L. Siegel, A. N. Giaquinto, and A. Jemal, "Cancer statistics, 2024," *CA: a cancer journal for clinicians*, vol. 74, no. 1, pp. 12–49, 2024.

[8] S. Lei, R. Zheng, S. Zhang, R. Chen, S. Wang, K. Sun, H. Zeng, W. Wei, and J. He, "Breast cancer incidence and mortality in women in china: temporal trends and projections to 2030," *Cancer biology & medicine*, vol. 18, no. 3, p. 900, 2021.

[9] S. Yao, Q. Kang, M. Zhou, M. J. Rawa, and A. Abusorrah, "A survey of transfer learning for machinery diagnostics and prognostics," *Artificial Intelligence Review*, vol. 56, no. 4, pp. 2871–2922, 2023.

[10] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2020.

[11] Z. Zhu, K. Lin, A. K. Jain, and J. Zhou, "Transfer learning in deep reinforcement learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[12] N. Ruiz, Y. Li, V. Jampani, Y. Pritch, M. Rubinstein, and K. Aberman, "Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation," pp. 22 500–22 510, 2023.

[13] W. Chen, Y. Liu, W. Wang, E. M. Bakker, T. Georgiou, P. Fieguth, L. Liu, and M. S. Lew, "Deep learning for instance retrieval: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[14] H. Rasheed, M. U. Khattak, M. Maaz, S. Khan, and F. S. Khan, "Fine-tuned clip models are efficient video learners," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 6545–6554.

[15] H. H. Luong, H. T. Nguyen, and N. Thai-Nghe, "A combination of active learning and deep learning for improving breast cancer prediction," in *International Conference on Advances in Information and Communication Technology*. Springer, 2023, pp. 3–10.

[16] R. Azad, A. Kazerouni, M. Heidari, E. K. Aghdam, A. Molaei, Y. Jia, A. Jose, R. Roy, and D. Merhof, "Advances in medical image analysis with vision transformers: a comprehensive review," *Medical Image Analysis*, p. 103000, 2023.

[17] F. Shamshad, S. Khan, S. W. Zamir, M. H. Khan, M. Hayat, F. S. Khan, and H. Fu, "Transformers in medical imaging: A survey," *Medical Image Analysis*, p. 102802, 2023.

[18] K. He, C. Gan, Z. Li, I. Rekik, Z. Yin, W. Ji, Y. Gao, Q. Wang, J. Zhang, and D. Shen, "Transformers in medical image analysis," *Intelligent Medicine*, vol. 3, no. 1, pp. 59–78, 2023.

[19] S. Balasubramaniam, Y. Velmurugan, D. Jaganathan, and S. Dhanasekaran, "A modified lenet cnn for breast cancer diagnosis in ultrasound images," *Diagnostics*, vol. 13, no. 17, p. 2746, 2023.

[20] H. Chen, M. Ma, G. Liu, Y. Wang, Z. Jin, and C. Liu, "Breast tumor classification in ultrasound images by fusion of deep convolutional neural network and shallow lbp feature," *Journal of digital imaging*, vol. 36, no. 3, pp. 932–946, 2023.

[21] M. Alotaibi, A. Aljouie, N. Alluhaidan, W. Qureshi, H. Almatar, R. Alduhayan, B. Alsomaie, and A. Almazroa, "Breast cancer classification based on convolutional neural network and image fusion approaches using ultrasound images," *Heliyon*, vol. 9, no. 11, 2023.

[22] C. Cruz-Ramos, O. García-Avila, J.-A. Almaraz-Damian, V. Ponomaryov, R. Reyes-Reyes, and S. Sadovnychiy, "Benign and malignant breast tumor classification in ultrasound and mammography images via fusion of deep learning and handcraft features," *Entropy*, vol. 25, no. 7, p. 991, 2023.

[23] N. Sirjani, M. G. Oghli, M. K. Tarzamni, M. Gity, A. Shabanzadeh, P. Ghaderi, I. Shiri, A. Akhavan, M. Faraji, and M. Taghipour, "A novel deep learning model for breast lesion classification using ultrasound images: A multicenter data evaluation," *Physica Medica*, vol. 107, p. 102560, 2023.

[24] H. Alrubaie, H. K. Aljobouri, Z. J. AL-Jobawi, and I. Çankaya, "Convolutional neural network deep learning model for improved ultrasound breast tumor classification," *Al-Nahrain Journal for Engineering Sciences*, vol. 26, no. 2, pp. 57–62, 2023.

[25] A. Sahu, P. K. Das, and S. Meher, "An efficient deep learning scheme to detect breast cancer using mammogram and ultrasound breast images," *Biomedical Signal Processing and Control*, vol. 87, p. 105377, 2024.

[26] S.-H. Chen, Y.-L. Wu, C.-Y. Pan, L.-Y. Lian, and Q.-C. Su, "Breast ultrasound image classification and physiological assessment based on googlenet," *Journal of Radiation Research and Applied Sciences*, vol. 16, no. 3, p. 100628, 2023.

[27] S. D. Deb and R. K. Jha, "Breast ultrasound image classification using fuzzy-rank-based ensemble network," *Biomedical Signal Processing and Control*, vol. 85, p. 104871, 2023.

[28] Q. T. Lu, T. M. Nguyen, and H. Le Lam, "Improving brain tumor mri image classification prediction based on fine-tuned mobilenet." *International Journal of Advanced Computer Science & Applications*, vol. 15, no. 1, 2024.

[29] K. Jabeen, M. A. Khan, J. Balili, M. Alhaisoni, N. A. Almujally, H. Alrashidi, U. Tariq, and J.-H. Cha, "Bc2netrf: breast cancer classification from mammogram images using enhanced deep learning features and equilibrium-jaya controlled regula falsi-based features selection," *Diagnostics*, vol. 13, no. 7, p. 1238, 2023.

[30] H. H. Luong, M. D. Vo, H. P. Phan, T. A. Dinh, L. Q. T. Nguyen, Q. T. Tran, N. Thai-Nghe, and H. T. Nguyen, "Improving breast cancer prediction via progressive ensemble and image enhancement," *Multimedia Tools and Applications*, pp. 1–28, 2024.

[31] İ. PACAL, "Deep learning approaches for classification of breast cancer in ultrasound (us) images," *Journal of the Institute of Science and Technology*, vol. 12, no. 4, pp. 1917–1927, 2022.

[32] B. Gheflati and H. Rivaz, "Vision transformers for classification of breast ultrasound images," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2022, pp. 480–483.

[33] X. Qu, H. Lu, W. Tang, S. Wang, D. Zheng, Y. Hou, and J. Jiang, "A vgg attention vision transformer network for benign and malignant classification of breast ultrasound images," *Medical Physics*, vol. 49, no. 9, pp. 5787–5798, 2022.