Journal of Robotics and Control (JRC)
Volume 6, Issue 4, 2025
ISSN: 2715-5072, DOI: 10.18196/jrc.v6i4.25075

Improving Classification Accuracy of Breast Ultrasound Images Using Wasserstein GAN for Synthetic Data Augmentation

I Gede Susrama Mas Diyasa [1*], Sayyidah Humairah [2], Eva Yulia Puspaningrum [3], Fara Disa Durry [4], Wahyu Dwi Lestari [5], Wahyu Caesarendra [6], Deshinta Arrova Dewi [7], Rangga Laksana Aryananda [8]
[1] Department of Magister Information Technology, University of Pembangunan Nasional Veteran Jawa Timur, Surabaya, Indonesia
[2, 3] Department of Informatics, University of Pembangunan Nasional Veteran Jawa Timur, Surabaya, Indonesia
[4] Faculty of Medicine, University of Pembangunan Nasional Veteran Jawa Timur, Surabaya, Indonesia
[5] Department of Mechanical Engineering, Faculty of Engineering, University of Pembangunan Nasional Veteran Jawa Timur, Surabaya, Indonesia
[6] Department of Mechanical and Mechatronics Engineering, Curtin University Malaysia, Malaysia
[7] Center for Data Science and Sustainable Technologies, INTI International University Malaysia, Malaysia
[8] Department of Data Science, University of Pembangunan Nasional Veteran Jawa Timur, Surabaya, Indonesia
Email: [1] igsusrama.if@upnjatim.ac.id, [2] sayyidahhumairah14@gmail.com, [3] evapuspaningrum.if@upnjatim.ac.id, [4] faradisa.fk@upnjatim.ac.id, [5] wahyu.dwi.tm@upnjatim.ac.id, [6] w.caesarendra@gmail.com, [7] deshinta.ad@newinti.edu.my, [8] ranggaalan90@gmail.com
*Corresponding Author

***Abstract*—Breast cancer remains one of the most prevalent cancers in Indonesia, and early detection plays a vital role in improving patient outcomes. Ultrasound imaging is a non-invasive and accessible technique used to classify breast conditions into normal, benign, or malignant categories. The advancement of deep learning, particularly Transfer Learning with Convolutional Neural Networks (CNNs), has significantly enhanced the performance of automated image classification. However, the effectiveness of CNNs heavily relies on large, balanced datasets—resources that are often limited and imbalanced in medical domains. To address this issue, this study explores the use of Wasserstein Generative Adversarial Networks (WGAN) for synthetic data augmentation. WGAN is capable of learning the underlying distribution of real ultrasound images and generating high-quality synthetic samples. The inclusion of the Wasserstein distance stabilizes training, with convergence observed around 2500–3000 epochs out of 5000. While synthetic data improves classifier performance, there remains a potential risk of overfitting, particularly when the synthetic images closely mirror the training data. Compared to traditional augmentation techniques such as rotation, flipping, and scaling, WGAN-generated data provides more diverse and realistic representations. Among the tested models, VGG16 achieved the highest accuracy of 83.33% after WGAN augmentation. Nonetheless, computational resource limitations posed challenges in training stability and duration. Furthermore, issues related to model generalizability, as well as ethical and patient privacy considerations in using synthetic medical data, must be addressed to ensure responsible deployment in real-world clinical settings.**

Keywords—Wasserstein GAN; Breast Cancer; Cancer; Ultrasound Image; Augmentation; Process Innovation.

I. INTRODUCTION

Breast cancer is one of the most common and deadly diseases affecting women worldwide [1][2][3]. According to the World Health Organization (WHO), about 2.3 million women were diagnosed for having breast cancer in 2020, and more than 685,000 of them died due to the disease. Despite improved survival rates due to early detection and better treatment, breast cancer remains a major health threat [4][5]. Early detection is crucial for improving the prognosis and survival rates of breast cancer patients [6][7]. Routine examinations and screening using medical imaging technology play a vital role in detecting cancer at an early stage when treatment is more effective [8][9]. Mammography, ultrasound, and magnetic resonance imaging (MRI) are some of the methods available for breast cancer detection [10]. Ultrasound, in particular, is a frequently used tool in breast examinations because it is non-invasive, relatively inexpensive, and easily accessible [11][12]. However, ultrasound has limitations in terms of availability and class imbalance, which can obscure important details and complicate diagnosis. Therefore, enhancing the quality and the amount of ultrasound image dataset is a significant focus in medical research [13][14][15].

Although ultrasound is a highly useful tool for breast cancer detection, it faces several significant challenges [16][17]. Ultrasound images often suffer from suboptimal quality, influenced by noise and artifacts, which can obscure critical details and make cancer detection more difficult [18][19]. This noise and these artifacts can lead to diagnostic er-rors or necessitate unnecessary additional examinations. The limited availability of medical data and the high cost of annotation often hinder the development of accurate machine learning models [20][21][22]. Collecting and annotating high-quality and big amount of medical data requires significant resources and access to medical facilities [23][24][25]. In addition, to be able to interpret ultrasound images, there is huge dependency on the expertise and

experience of radiologists, which may lead to variability between examiners [26][27][28]. This variability can result in differences in diagnosis and treatment received by patients [29]. Although several automatic detection methods have been developed, many still struggle to adequately handle the complexity and variability of medical im-ages. Existing detection algorithms may not be robust or accurate enough for wide-spread clinical application [30][31][32].

The potential utilization of ultrasound image data with the advancement of deep learning technology enables earlier detection of breast cancer without the need for invasive procedures from the outset [33][34][35]. Deep learning technology can learn features from the available data, allowing for the classification of breast conditions into several categories, namely normal, benign tumor, and cancer [36][37]. However, the limited availability of medical data and class imbalance in the data often pose obstacles in developing accurate deep learning models [38]. The collection and annotation of high-quality medical data require significant resources, access to medical facilities, and patient privacy considerations. One common approach to addressing data limitations and class imbalance, which are common issues in training CNN models is data augmentation [39][40][41].

Traditional data augmentation techniques such as geometric augmentation, color augmentation, and noise augmentation are commonly performed to improve dataset variability in medical imaging tasks [42][43][44]. These techniques encompass processes such as reflecting images, cropping and translating images, and altering the image color palette. Although these methods can help expand datasets, they have several drawbacks. The transformations applied are deterministic and often fail to generate sufficiently realistic variations, as they do not reflect the natural complexity and diversity of actual medical data [45][46]. To address these limitations, Generative Adversarial Networks (GANs) have been introduced as an advanced data augmentation approach that can generate synthetic data by learning the distribution patterns of original datasets [47]. GAN, introduced by Goodfellow, uses Jensen-Shannon (JS) divergence in the loss function [48][49]. However, the use of JS divergence is associated with the vanishing gradient problem, which causes unstable training. Additionally, GANs frequently suffer from mode collapse, where the generator fails to capture the full diversity of the dataset and repeatedly generates similar outputs [50][51][52].

To overcome the challenges in traditional GANs, a variant called Wasserstein Generative Adversarial Network (WGAN) was introduced [53][54]. WGAN replaces the JS divergence with the Wasserstein distance, a more robust metric that enables more stable training and the generation of higher-quality, more realistic synthetic images [55][56]. Despite these advancements, no prior study has quantified how GAN-based augmentation affects the robustness of classifiers on truly unseen clinical datasets, nor has it addressed the potential risks of synthetic-data-induced bias. This study aims to fill that research gap by presenting the first systematic evaluation of WGAN-augmented breast ultrasound image classification across multi-institutional datasets and proposing an ethics-aware data augmentation pipeline that safeguards patient privacy. Furthermore, we examine ethical concerns related to minority-class oversampling, which may introduce bias, and propose mitigation strategies using diversity-aware sampling to promote fairness and improve generalizability in clinical applications.

In 2021, research by Xiao et al. [57] utilized the Wasserstein GAN model for data augmentation to address class imbalance issues. This model was applied to three RNA-seq cancer patient datasets obtained from the TCGA cancer gene expression database: Breast Invasive Carcinoma (BRCA), Lung Adenocarcinoma (LUAD), and Stomach Adenocarcinoma (STAD). The datasets consisted of two classes, normal (N) and tumor (T), which were divided into testing and training data. Data augmentation with WGAN was performed only on the training data, where the number of data in the minority class was expanded in order to match the majority class as to reach the class balance. The LUAD dataset was expanded from 22 N and 110 T to 110 N and 110 T, the STAD dataset from 18 N and 223 T to 223 N and 223 T, and the BRCA dataset from 73 N and 745 T to 745 N and 745 T. Cancer condition classification was then performed using a Support Vector Machine (SVM) model. The results showed that, compared to using the original dataset alone, the SVM model exhibited significantly improved performance with the augmented dataset. The SVM model accuracy increased from 50% to 90% on the LUAD dataset, 50% to 93.33% on the STAD dataset, and 50% to 98.33% on the BRCA dataset. Building on the previous research by Xiao et al., this study discusses the augmentation of breast ultrasound image data using WGAN to generate synthetic images that can address class imbalance issues in each class.

In WGAN, the images generated by the Generator originate from the mapping of a random latent vector with dimension n. This random vector is transformed by the Generator into synthetic images that increasingly resemble the real image data. Ac-cording to the original WGAN training algorithm, the training process continues until the Generator converges [54][55]. However, in practice, the WGAN training process defines one of the hyperparameters prior to training, namely the number of epochs or training iteration steps [58][59][60]. Based on the above explanation, this study aims to conduct research on generating synthetic images using the WGAN model. It is hoped that this research will contribute to the creation of image datasets with the best possible quality to address issues of dataset availability or imbalanced datasets.

II. MATERIALS AND METHODS

This study will use an annotated breast ultrasound image dataset to train and test the WGAN model [61]. The training process of WGAN will involve two neural networks: a generator that produces synthetic ultrasound images and a discriminator that assesses the authenticity of these images [62]. The generator and discriminator will be trained iteratively until the generator is capable of producing images that closely resemble the original ones. This research focuses on the data augmentation process using WGAN, as illustrated in the block diagram in Fig. 1 and flowchart in Fig. 2

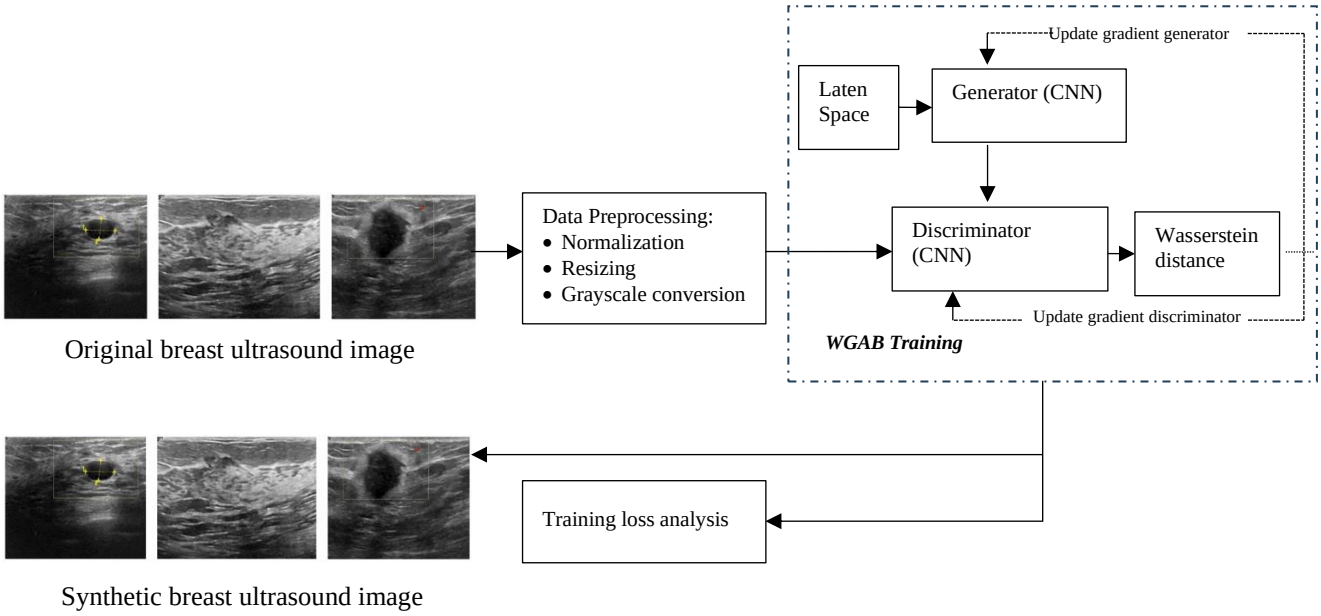

Fig. 1. Block diagram ultrasound image augmentation using WGAN

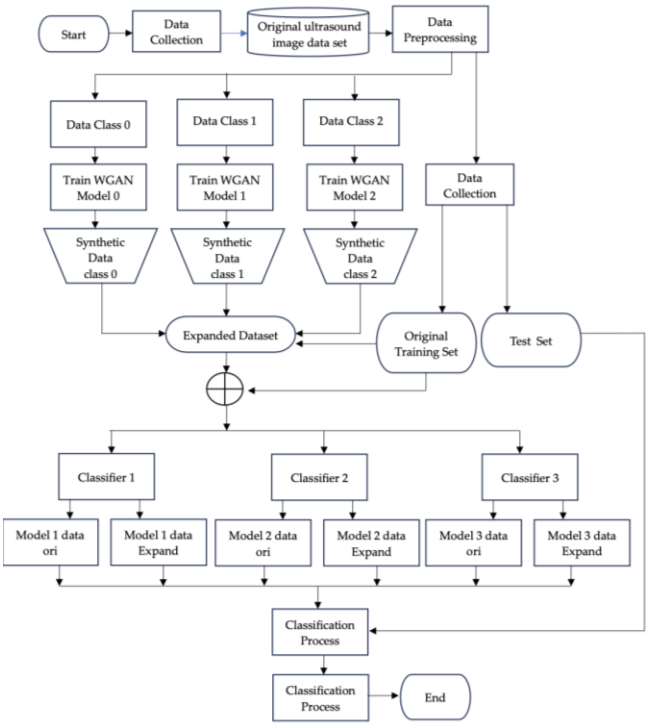

Fig. 2. Flowchart ultrasound image augmentation using WGAN

The study begins with the collection of an annotated breast ultrasound image dataset to be used for model training. Pre-processing is then conducted, which includes image data normalization, resizing, and converting the images to grayscale [63]. The training of the Wasserstein GAN, involving the generator (a neural network that produces synthetic ultrasound images) and the discriminator (a neural network that measures the distribution difference between original and synthetic ultrasound images), is performed iteratively [64][65][66]. Feedback from the discriminator is used to enhance the generator's performance. The output from the WGAN generator consists of synthetic ultrasound images that closely resemble the original images with high quality [67][68][69]. These synthetic images are then used for data augmentation to increase both the size and variability of the dataset which then will be the input data for classification process [70][71][72].

This study will use an annotated breast ultrasound image dataset to train and test the WGAN model. The training process of WGAN will involve two neural networks: a generator that produces synthetic ultrasound images and a discriminator that assesses the authenticity of these images. The generator and discriminator will be trained iteratively until the generator is capable of producing images that closely resemble the original ones. This research focuses on the data augmentation process using WGAN, as illustrated in the block diagram and flowchart in Fig. 1 and in Fig. 2.

A. Breast Ultrasound Image Data Acquisition

The dataset used in this study is derived from research conducted by Al-Dhabyani et al. (2020). This dataset contains breast ultrasound image data from several individuals with varying conditions. There are 437 images classified as Benign, 133 images categorized as Normal, and 210 images classified as Malignant. In this study, the Benign, Normal, and Malignant classes will be referred to as classes 0, 1, and 2, respectively, as shown in Fig. 3.

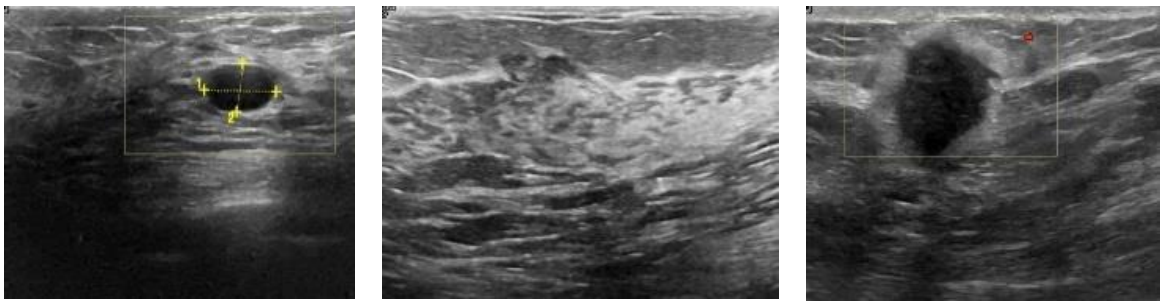

Fig. 3. Dataset sample from class 0 (a), 1 (b), and 2 (c)

Fig. 3 in this document presents sample images from the breast ultrasound dataset used in the study. The images are categorized into three distinct classes, each representing different medical conditions. On the left, there are images from Class 0, which consists of images diagnosed as Benign. These images depict lesions or changes that do not show signs of cancer, providing an overview of relatively safe conditions. In the centre, there are images from Class 1, representing Normal conditions. These images show healthy breast tissue without any detected abnormalities. This class serves as a reference for distinguishing between normal and abnormal conditions. On the right, there are images from Class 2, which depict Malignant conditions. These images indicate the presence of abnormalities that may be cancerous, making them crucial for further diagnosis and management. By presenting these three classes side by side, Fig. 3 offers a clear visualization of the characteristic differences between benign, normal, and malignant conditions. This is particularly important in the context of research, as it aids researchers and medical practitioners in understanding and developing better detection methods for various breast conditions.

B. Pre-processing

Pre-processing in the augmentation of breast cancer ultrasound images involves a series of steps to prepare the data before using it to train the model [73][73]. The two main aspects of pre-processing are normalization and image resizing. Below is the process undertaken for pre-processing the breast ultrasound image dataset using WGAN.

1. Normalization is the process of adjusting the pixel intensity range in an image to achieve a uniform distribution. This helps improve the convergence of the optimization algorithm and reduces scale differences that can affect the model's performance [74]. In this study, image normalization was performed using OpenCV version 4.5, where each pixel value was scaled using the formula equation (1), resulting in a range of [−1,1]. This normalization approach ensures that the data are centered around zero, which is beneficial for stabilizing the training process of deep neural network.

$$Pixel_{normalized} = \frac{Pixel_{original}}{127.5} - 1 \quad (1)$$

Where:

- Pixel_original refers to the pixel intensity value in the original image.
- The value 127.5 is derived from the maximum pixel intensity in the original image, which is 255, divided by 2.

2. Image resizing is the process of changing the image size to a predefined uniform size. In this study, each image is resized to 128×128 pixels. This resizing process ensures that all image data will be compatible with the WGAN Generator and Discriminator networks [75][76].
3. Conversion to grayscale is carried out to ensure that breast ultrasound images do not contain any colors originating from external factors such as ultrasound equipment, which are not part of the breast tissue ultrasound. The images should be in grayscale format.

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

C. Wasserstein GAN Training

The WGAN training process begins with initializing the parameters [77][78]. It then proceeds in a main loop that continues until the generator's parameters converge. Each iteration of the main loop involves several updates to the Critic or Discriminator. In each Discriminator iteration, a batch of real data is first sampled from the original data distribution. Then, a batch of data is sampled from random noise or the latent space. The Discriminator's gradient is computed to update its parameters, followed by a process of weight clipping. After several Discriminator updates, the Generator's parameters are updated. A batch of data from the latent space is sampled again, and the generator's gradient is computed to update the Generator's parameters. While the original algorithm repeats this process continuously until the generator's parameters converge, in this study, the iterations in the main loop are limited by the number of epochs or steps.

The training configuration used in this work includes a learning rate of 0.00005, a batch size of 128, five Discriminator updates for each Generator update (n_critic = 5), and a weight clipping value of 0.01. To identify the optimal settings, hyperparameter tuning was performed using a grid search over several learning rate and clipping value combinations. To mitigate the common problem of mode

collapse where the Generator produces repetitive outputs with limited variability, we applied early stopping during training. In addition, preliminary experiments were conducted using gradient penalty as an alternative to weight clipping, aiming to enhance training stability and encourage the generation of more diverse and realistic synthetic images.

In WGAN, the Wasserstein distance is implemented in the Discriminator's loss function as shown in equation (2), calculated by averaging the scores for real and fake images [79][80][81]. The difference between the average scores of fake and real images is used as the loss, which the discriminator tries to maximize to effectively distinguish between real and fake images. For the Generator, the loss function is calculated by taking the negative of the average score the Discriminator assigns to the fake images. This encourages the generator to create images that receive high scores from the discriminator, indicating that the images appear more realistic.

$$W(P_r, P_r) = \inf_{\gamma \in \prod(P_r, P_g)} E_{(x,y)\sim\gamma}[\|x - y\|] \quad (2)$$

Table I shows the details of the algorithm of WGAN training according to the original WGAN research by Arjovsky.

TABLE I. WASSERSTEIN GAN ALGORITHM

WGAN Algorithm
Parameters: α = learning rate, c = parameter clipping, m = batch size, $n$critic = Critic iteration for each Generator iteration

While θ has not converged do
For t = 0, …, $n$critic do
Sample $\{x^{(i)}\}_{i=1}^{m} \sim P_r$ a batch from the real data
Sample $\{z^{(i)}\}_{i=1}^{m} \sim p(z)$ a batch of prior samples
$g_w \leftarrow \nabla_w \left[\frac{1}{m}\sum_{i=1}^{m} f_w(x^{(i)}) - \frac{1}{m}\sum_{i=1}^{m} f_w\left(g_\theta(z^{(i)})\right)\right]$
$w \leftarrow w + \alpha \cdot RMSProp(w, g_w)$
$w \leftarrow \text{clip}(w, -c, c)$
End for
Sample $\{z^{(i)}\}_{i=1}^{m} \sim p(z)$ a batch of prior samples.
$g_\theta \leftarrow -\nabla_\theta \frac{1}{m}\sum_{i=1}^{m} f_w\left(g_\theta(z^{(i)})\right)$
$\theta \leftarrow \theta - \alpha \cdot RMSProp(\theta, g_\theta)$
End while

D. Evaluating the Effectiveness of WGAN-based Augmentation

The original preprocessed image dataset will be going through two different processes. Firstly, the complete preprocessed dataset will be used for WGAN training. On the other hand, the preprocessed dataset is split into training and test set. The training set which will be combined with the synthetic images generated by WGAN generator will be the expanded dataset. The original training set and expanded dataset then will be the input of CNN Classifiers during the training. Then, the trained classifiers will be tested by the test set as to measure the performance difference of the classifiers with different datasets. The performance is measured by four metrics, namely accuracy, precision, recall, and F1-score. In this work, we are using transfer learning classifiers, which are VGG16, ResNet50, MobileNetV2, and YOLOv8.

III. RESULTS AND DISCUSSION

A. Results of WGAN Training

Each of the generator and discriminator models of the WGAN in this study is constructed using convolutional neural networks with layer architectures as shown in Fig. 4. For the training process, several hyperparameters are defined by the researchers to adapt the WGAN model to the existing dataset, ensuring that the WGAN model can generate synthetic data of good quality and consistent with the dataset used. Table II displays the hyperparameters involved in the WGAN training process and their values.

The Generator network (Fig. 4, left) starts by transforming a 128-dimensional latent vector through a dense layer into a 3D tensor, then progressively upsamples it using Conv2DTranspose layers with batch normalization and LeakyReLU activations to produce realistic 128×128 grayscale ultrasound images. The Discriminator network (Fig. 4, right) processes 128×128 input images through successive Conv2D layers with LeakyReLU and Dropout, reducing spatial dimensions while increasing feature depth, before flattening and outputting a single value estimating the Wasserstein distance. Together, these architectures enable smooth image generation and effective discrimination by capturing both structural and semantic features.

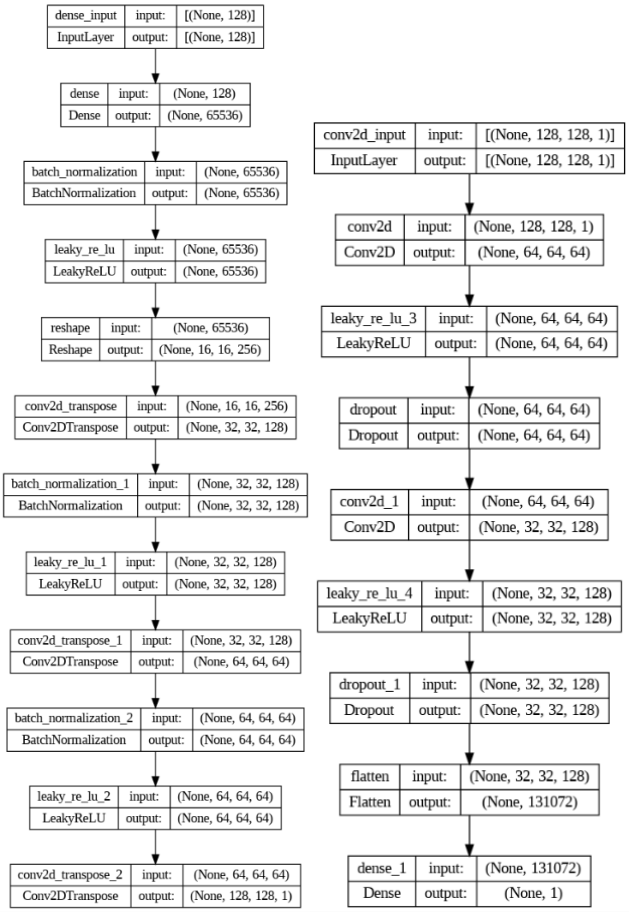

Fig. 4. WGAN Architecture; Generator (left) and Critic (right)

TABLE II. HYPERPARAMETERS

Hyperparameters	***Value***
Epoch	5000
Weight clipping value	0.01
Optimizer	RMSProp
Learning rate optimizer	5e-5
Batch size	128
Number of discriminator iteration for each generator iteration	5

During the training process over 5000 epochs for each dataset class, the loss of the generator and discriminator is recorded. Fig. 5, Fig. 6, and Fig. 7 present the loss graphs for each WGAN training process. Fig. 5, Fig. 6, and Fig. 7 display the loss of each Generator and Discriminator model during the training process using image data from classes 0, 1, and 2. The loss of the Wasserstein GAN (WGAN) training process using three different datasets reveals several important aspects regarding model convergence and stability. Unlike traditional GANs, WGAN's Discriminator or Critic does not perform an evaluation by classifying input data as real or fake but rather computes the Wasserstein distance between two distributions, namely the real data distribution and the synthetic data distribution generated by the Generator.

An evaluation of the loss graphs in Fig. 5, Fig. 6, and Fig. 7 indicates that the stabilization of the discriminator loss occurs more rapidly compared to the generator loss. The graphs show that in the early epochs of training, there are marked fluctuations in the loss values. This phenomenon indicates that the WGAN model makes significant initial adjustments in model weights in response to different data distributions. The synthetic images produced by the Generator in the early stages of training are still poor, resulting in a significant difference between the distributions of real image data and the synthetic images generated by the Generator. As the training progresses, the WGAN Generator model improves in producing synthetic images.

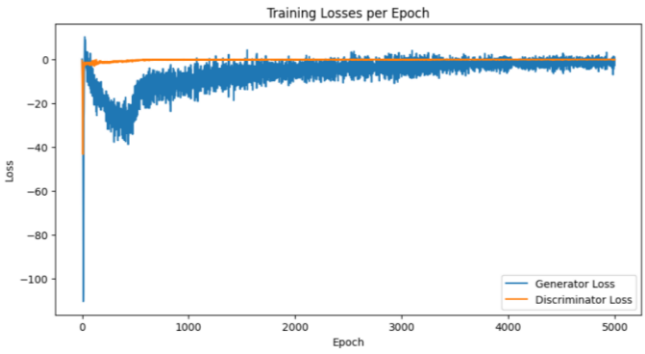

Fig. 5. Training loss WGAN class 0

Based on the analysis of the loss patterns, the model for class 0 (Fig. 5) begins to show stability after approximately 3000 epochs, with minimal fluctuations thereafter. If training continues without an epoch limit, the model will likely remain stable with slight improvements in the quality of synthetic images. The model for class 1 (Fig. 6) shows stabilization after 2500 epochs, but with significant variation still present. The loss pattern indicates that the model for class 1 requires more epochs to achieve the level of stability reached by the model for class 0. Compared to the model for class 1, the model for class 2 (Fig. 7) shows a loss pattern similar to the model for class 0 and achieves stability more quickly, around 3000 epochs. Adding more epochs to the training process can further minimize loss and enhance stability, though the improvement in synthetic data quality may not be very significant.

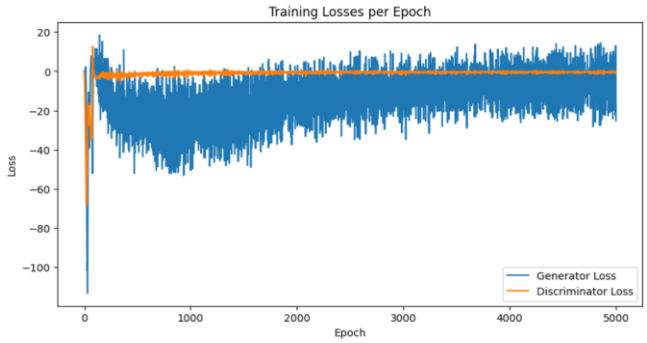

Fig. 6. Training loss WGAN class 1

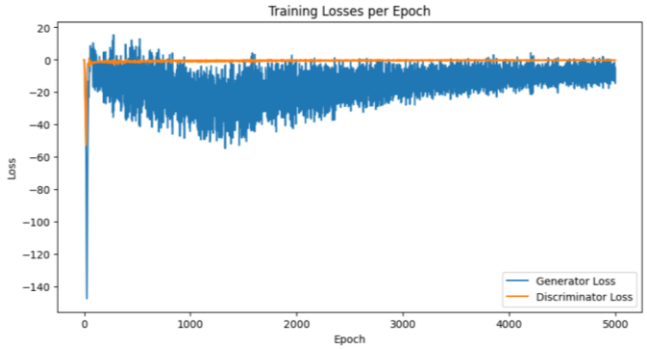

Fig. 7. Training loss WGAN class 2

Thus, the stability evaluation of the WGAN training shows that the models for classes 0 and 2 achieve stability faster than class 1, which requires more epochs to reach stability. The increasingly stable loss patterns at the end of training indicate that the models have successfully approximated the real data distribution, signifying good convergence. The differing behaviors of each model relate to the complexity and variability of the data in each class. The training dataset size for class 1 is the smallest, which, based on its loss graph, requires more time to reach the stability level achieved by the models for classes 0 and 2. This evaluation provides a clearer picture of the stability and convergence of the WGAN models used and shows how the models can be further improved with additional training if necessary. The implementation of Wasserstein distance in the discriminator loss provides an indicator of training progress, enabling researchers to monitor and adjust the model as needed. With achieved stabilization, the WGAN model can reliably generate high-quality synthetic images, which is crucial in addressing data limitations in the medical field, especially in breast ultrasound image data.

B. Image Synthetic Augmented by WGAN

The size of the expanded ultrasound image dataset, as shown in Table III, compares the sizes of the original image dataset and the dataset after the augmentation process.

TABLE III. SIZE OF DATASET

Dataset/Metrics	Accuracy	Precision	Recall	F1-score
Original data	0.8030	0.8000	0.8121	0.8005
Expanded data	0.8133	0.8150	0.8012	0.8175

Fig. 8, Fig. 9, and Fig. 10 display 5 synthetic image samples for each class generated by the WGAN Generator during the second augmentation process, which was previously trained using image data from the corresponding classes. Fig. 8 shows synthetic image samples for class 0 generated by the generator trained on class 0 data. Fig. 9 shows synthetic image samples for class 1 (Normal category) generated by the WGAN model generator trained on class 1 data. Fig. 10 shows synthetic image samples for class 2 (Malignant/Cancer category) generated by the WGAN model generator trained on class 2 image data.

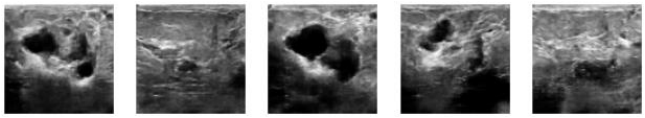

Fig. 8. Synthetic image class 0

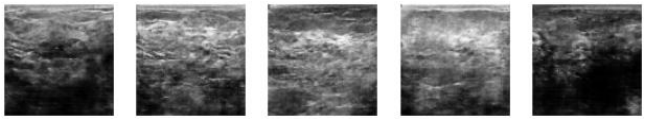

Fig. 9. Synthetic image class 1

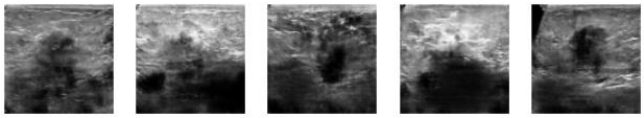

Fig. 10. Synthetic image class 2

C. Prediction Using CNN Classifiers

After expanding the original dataset, the classifiers performance utilizing different datasets is examined based on evaluation metrics. The performance of the classifiers with each dataset is presented in Table IV, Table V, Table VI and Table VII.

TABLE IV. EVALUATION METRICS OF VGG16

Dataset/Metrics	*Accuracy*	*Precision*	*Recall*	*F1-score*
Original data	0.7628	0.7380	0.7496	0.7433
Expanded data	0.8333	0.8490	0.8021	0.8219

TABLE V. EVALUATION METRICS OF RESNET50

Dataset/Metrics	*Accuracy*	*Precision*	*Recall*	*F1-score*
Original data	0.6795	0.7000	0.6400	0.6395
Expanded data	0.7308	0.7100	0.6976	0.6945

TABLE VI. EVALUATION METRICS OF MOBILENETV2

Dataset/Class	Class		
	0 - Benign	1 - Normal	2 - Malignant
Original	437	133	210
Expand	437 + 50	133 + 354 = 487	210 + 277 = 487

TABLE VII. EVALUATION METRICS OF YOLOV8

Dataset/Metrics	*Accuracy*	*Precision*	*Recall*	*F1-score*
Original data	0.8077	0.7929	0.7951	0.7939
Expanded data	0.8205	0.8243	0.8151	0.8094

We compared multiple pretrained models that have been known best for the classification, including VGG16, ResNet50, and MobileNetV2. In addition, we also incorporated YOLOv8, a pretrained model known for object detection but can also be utilized for classification task to our experiment. Each model was evaluated based on accuracy, precision, recall, and F1-score. Based on the results from Table IV, Table V, Table VI and Table VII, all models exhibit similar performance behaviors in their prediction capabilities according to the evaluation metrics scores. All evaluation metrics indicate that the three models improve in prediction accuracy when the dataset is augmented compared to the original dataset.

VGG16 consistently achieved the highest results, with an accuracy of 83.33%, outperforming the other models in all metrics. ResNet50 and MobileNetV2 also exhibited notable improvements, particularly when augmented data generated by WGAN was used. The results achieved by YOLOv8 only differs by 2 percent compared to that of VGG16. Even though YOLOv8's strong performance in object detection tasks, its evaluation metrics in this classification context did not surpass those of VGG16. It is indicating that its architecture, while powerful for object detection, might not be as well-suited for direct image classification tasks compared to models like VGG16. This suggests that while YOLOv8 may be effective for other types of image-based tasks, more specialized models like VGG16 may offer better results for ultrasound image classification.

Overall, the best accuracy, precision, and F1-score were achieved by the VGG16 model, with scores of 83.33%, 84.90%, and 82.19%, respectively. In contrast, the best recall was obtained by the MobileNetV2 model, with a score of 81.51%. The combination of the VGG16 model with the expanded dataset proved to be the most effective, showing the most significant improvements across all evaluation metrics, including accuracy, precision, recall, and F1-score. This improvement suggests that the model can leverage additional data to learn more complex and accurate features, which is crucial in medical applications such as breast cancer detection. The increase in precision and F1-score particularly indicates that the model becomes more reliable in correctly predicting positive cases, thereby reducing misclassification errors, which can have serious consequences in a clinical context.

To further evaluate the classifier's discriminative performance across different breast condition categories, the VGG16 model was assessed using the Receiver Operating Characteristic (ROC) curve, as illustrated in Fig. 11. The ROC curve provides a visual representation of the trade-off between sensitivity (true positive rate) and false positive rate for each class. Based on the simulation results, Class 2 (Malignant) achieved the highest AUC value of 0.79, suggesting that the model is highly effective in distinguishing malignant cases, which is essential in early and accurate clinical diagnosis. Class 0 (Benign) followed with an AUC of 0.75, while Class 1 (Normal) had the lowest AUC of 0.63-likely due to the limited number of normal class samples in the original dataset prior to augmentation. These findings reinforce the importance of balanced and representative data

in optimizing classifier performance across all diagnostic categories.

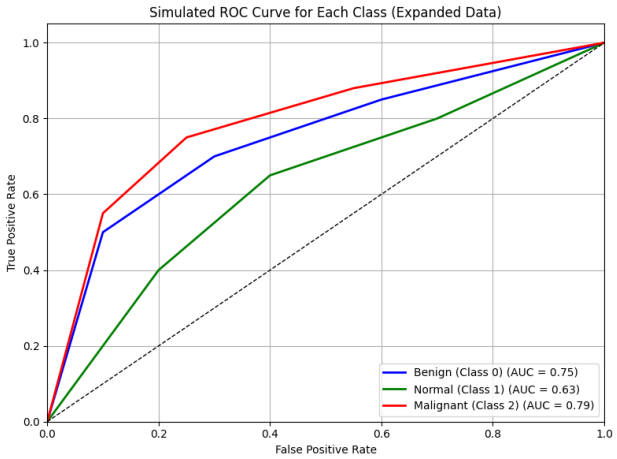

Fig. 11. ROC curve of VGG16

IV. CONCLUSIONS

The potential of WGAN for data augmentation in the medical imaging field is very promising. This study demonstrates how the implementation of Wasserstein GAN on limited breast ultrasound medical data can be conducted with stable training processes, resulting in synthetic image data that closely resemble original breast ultrasound images. The stability of the WGAN training process is related to the implementation of the Wasserstein distance as the loss function in the Discriminator. This implementation also facilitates monitoring and interpreting model performance by researchers. The differences in stability among the WGAN models for each class are influenced by the dataset size of each class. Overall, the issue of limited medical data, which often hinders research, can be addressed through data augmentation by WGAN. The effectiveness of utilizing WGAN in data augmentation is apparent in the classifier's performances, where all of the evaluation metrics of each classifier is increasing, with the best accuracy score is achieved by VGG16 model with 83.33% accuracy. This comprehensive analysis underscores the importance of data augmentation in enhancing model performance, especially in critical domains where accuracy and reliability are paramount.

From a practical standpoint, the augmented dataset produced using WGAN has the potential to support the development of computer-aided diagnostic (CAD) tools in clinical settings, thereby assisting radiologists and clinicians in making more accurate and timely diagnoses. However, this study also acknowledges certain limitations, such as the potential lack of diversity in the generated images and risks of overfitting if synthetic data dominate the training process. These aspects warrant careful consideration in future implementations. Researchers recommend that in future studies, each model be trained with a different number of epochs based on the needs of each model. Additionally, the application of weight clipping in WGAN in this study could be replaced with an alternative such as the implementation of gradient penalty, providing smoother weight constraints in the model and resulting in more stable training performance and better synthetic images.

REFERENCES

[1] S. Hyuna *et al*., "B. GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," *Global Cancer Statistics 2020*, vol. 71, no. 3, pp. 209-249, 2021.

[2] S. Sari and M. Naghavi, "Global, regional, and national burden of breast cancer and its attributable risk factors among women, 1990–2017," *Research Square*, Oct. 2020, doi: 10.21203/rs.3.rs-97176/v1.

[3] R. A. Salman, "Prevalence of women breast cancer," *Cell. Mol. Biomed. Rep*., vol. 3, no. 4, pp. 185–196, 2023, doi: 10.55705/cmbr.2023.384467.1095.

[4] T. I. Mamun, S. Younus, and M. H. Rahman, "Gastric cancer Epidemiology, modifiable and non-modifiable risk factors, challenges and opportunities: An updated review," *Cancer Treat. Res. Commun.*, vol. 41, 2024, doi: 10.1016/j.ctarc.2024.100845.

[5] M. Arnold *et al*., "Current and future burden of breast cancer: Global statistics for 2020 and 2040," *The Breast*, vol. 66, pp. 15–23, 2022, doi: 10.1016/j.breast.2022.08.010.

[6] I. U. Haq, H. Ali, H. Y. Wang, L. Cui, and J. Feng, "BTS-GAN: Computer-aided segmentation system for breast tumor using MRI and conditional adversarial networks," *Engineering Science and Technology, an International Journal*, vol. 36, pp. 1-10, 2022.

[7] O. Ginsburg *et al*., "Breast cancer early detection: a phased approach to implementation," *Cancer*, vol. 126, pp. 2379–2393, May 2020, doi: 10.1002/cncr.32887.

[8] J. Loud and J. Murphy, "Cancer screening and early detection in the 21st century," *Seminars in Oncology Nursing*, vol. 33, no. 2, pp. 121–128, May 2017, doi: 10.1016/j.soncn.2017.02.002.

[9] N. Chacko and R. Ankri, "Non-invasive early-stage cancer detection: current methods and future perspectives," *Clinical and Experimental Medicine*, vol. 25, p. 17, 2025, doi: 10.1007/s10238-024-01513-x.

[10] L. Jiawei *et al*., "Non-invasive biomarkers for early detection of breast cancer," *Cancer*, vol. 12, no. 2767, pp. 1-28, 2020.

[11] G. Ghulam *et al*., "Digital image processing for ultrasound images: A comprehensive review," *Int. J. Innov. Creat. Change*, vol. 15, no. 3, pp. 1335–1354, 2021.

[12] N. A. Alam, M. M. U. Khandaker, R. Mahbubur, M. M. R. Manu, and K. N. Mostofa, "A novel automated system to detect breast cancer from ultrasound images using deep fused features with super resolution," *Intell.-Based Med.*, vol. 10, p. 100149, pp. 1–15, 2024.

[13] T. Rebeca, M. David, I. I. Carlos, A. G. Rodrigo, A. V. Fernando, and L. H. Joaquin, "Recent advances in artificial intelligence-assisted ultrasound scanning," *Appl. Sci.*, vol. 13, p. 3693, pp. 1–17, 2023.

[14] W. Yu, G. Xinke, M. He, Q. Shouliang, Z. Guanjing, and Y. Yudong, "Deep learning in medical ultrasound image analysis: A review," *IEEE Access*, vol. 4, pp. 1–15, 2016.

[15] R. Alyaa, "Effects of artifacts on the diagnosis of ultrasound image," *Medico-Legal Update*, vol. 21, no. 4, pp. 327–336, 2021.

[16] R. Iacob *et al*., "Evaluating the role of breast ultrasound in early detection of breast cancer in low- and middle-income countries: A comprehensive narrative review," *Bioengineering*, vol. 11, no. 3, p. 262, Mar. 2024, doi: 10.3390/bioengineering11030262.

[17] S. O. Aboagye, J. A. Hunt, G. Ball, and Y. Wei, "Portable noninvasive technologies for early breast cancer detection: A systematic review," *Comput. Biol. Med.*, vol. 182, p. 109219, Nov. 2024, doi: 10.1016/j.compbiomed.2024.109219.

[18] O. M. Zeyad and A. E. E. Latif, "Breast cancer detection in ultrasound imaging," *World J. Adv. Res. Rev.*, vol. 12, no. 1, pp. 308–314, 2021.

[19] J. E. Randall, M. S. Ryan, and L. Alfonso, "Twelve key challenges in medical machine learning and solutions," *Intell.-Based Med.*, vol. 6, p. 100068, pp. 1–9, 2022.

[20] A. Imane, I. Ali, and C. Ikram, "Cost-sensitive learning for imbalanced medical data: A review," *Artif. Intell. Rev.*, vol. 57, no. 80, pp. 1–72, 2024.

[21] C. Chiranjib, B. Manojit, P. Soumen, and S. Lee, "From machine learning to deep learning: Advances of the recent data-driven paradigm shift in medicine and healthcare," *Curr. Res. Biotechnol.*, vol. 7, p. 100164, 2024.

[22] Y. Yang *et al*., "Expert recommendation on collection, storage, annotation, and management of data related to medical artificial intelligence," *Intell. Med.*, vol. 3, pp. 144–149, 2023.

[23] B. Z. Hussain, I. Andleeb, M. S. Ansari, A. M. Joshi, and N. Kanwal, "Wasserstein GAN based chest X-ray dataset augmentation for deep learning models: COVID-19 detection use-case," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBS)*, pp. 2058–2061, 2022.

[24] Y. Yang *et al*., "Standardization of collection, storage, annotation, and management of data related to medical artificial intelligence," *Intell. Med.*, vol. 3, no. 2, Dec. 2021, doi: 10.1016/j.imed.2021.11.002.

[25] Y. P. Sari Min, W. Y. Christiansi, D. Mau-Buti, and A. Prambudi, "Implementing electronic medical records through big data in healthcare facilities," *J. Sci. Res. Educ. Technol. (JSRET)*, vol. 3, no. 1, 2024.

[26] Y. H. Kim, "Artificial intelligence in medical ultrasonography: Driving on an unpaved road," *Ultrasonography*, vol. 40, no. 3, pp. 313–317, May 2021, doi: 10.14366/usg.21031.

[27] M. P. Caserta *et al*., "Ultrasound practice redesign to improve image quality: Implementation of a quality control sonographer," *J. Am. Coll. Radiol.*, vol. 17, no. 12, Aug. 2020, doi: 10.1016/j.jacr.2020.07.015.

[28] R. Tenajas, D. Miraut, C. I. Illana, R. Alonso-Gonzalez, F. Arias-Valcayo, and J. L. Herraiz, "Recent advances in artificial intelligence-assisted ultrasound scanning," *Appl. Sci.*, vol. 13, no. 6, p. 3693, Mar. 2023, doi: 10.3390/app13063693.

[29] D. Z. Haq and C. Fatichah, "Ultrasound image synthetic generating using deep convolution generative adversarial network for breast cancer identification," *IPTEK J. Technol. Sci.*, vol. 34, no. 1, pp. 13–25, 2021.

[30] A. Carriero, L. Groenhoff, E. Vologina, P. Basile, and M. Albera, "Deep learning in breast cancer imaging: State of the art and recent advancements in early 2024," *Diagnostics*, vol. 14, p. 848, 2024.

[31] T. Sahar *et al*., "Anomaly detection in laser powder bed fusion using machine learning: A review," *Results in Engineering*, vol. 17, p. 100803, 2023.

[32] C. S. Stafie, I.-G. Sufaru, C. M. Ghiciuc, I.-I. Stafie, E.-C. Sufaru, S. M. Solomon, and M. Hancianu, "Exploring the intersection of artificial intelligence and clinical healthcare: A multidisciplinary review," *Diagnostics*, vol. 13, no. 12, p. 1995, Jun. 2023, doi: 10.3390/diagnostics13121995.

[33] T. L. Nikmah, R. M. Syafei, and D. N. Anisa, "Inception ResNet v2 for early detection of breast cancer in ultrasound images," *J. Inf. Syst. Explor. Res.*, vol. 2, no. 2, pp. 93–102, 2024.

[34] I. S. M. D. Gede, Y. P. Eva, M. Hatta, and S. Ariono, "New method for classification of spermatozoa morphology abnormalities based on macroscopic video of human semen," in *Proc. 2019 Int. Seminar Appl. Technol. Inf. Commun. (iSemantic)*, pp. 133–140, 2019.

[35] L. Balkenende, J. Teuwen, and R. M. Mann, "Application of deep learning in breast cancer imaging," *Seminars in Nuclear Medicine*, vol. 52, no. 5, pp. 584-596, 2022.

[36] A. Heikal, E. G. Amir, E. Samir, and M. Z. Rashad, "Fine tuning deep learning models for breast tumor classification," *Sci. Rep.*, vol. 14, p. 10753, pp. 1–26, 2024.

[37] Z. Rguibi, A. Hajami, D. Zitouni, Y. Maleh, and A. Elqaraoui, "Medical variational autoencoder and generative adversarial network for medical imaging," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 32, no. 1, pp. 494–505, 2023.

[38] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[39] F. J. Moreno-Barea, J. M. Jerez, and L. Franco, "Improving classification accuracy using data augmentation on small data sets," *Expert Systems with Applications*, vol. 161, p. 113696, 2020.

[40] M. Li, Y. Jiang, Y. Zhang, and H. Zhu, "Medical image analysis using deep learning algorithms," *Front. Public Health*, vol. 11, p. 1273253, Nov. 2023, doi: 10.3389/fpubh.2023.1273253.

[41] J. M. Górriz *et al*., "Computational approaches to explainable artificial intelligence: Advances in theory, applications and trends," *Information Fusion*, vol. 100, p. 101945, Dec. 2023.

[42] Z. Shangguan, Y. Zhao, W. Fan, and Z. Cao, "Dog image generation using deep convolutional generative adversarial networks," in *Proc. Int. Conf. Universal Village (UV)*, pp. 1–6, 2020.

[43] Y. Chandola, V. Uniyal, Y. Bachheti, and R. Rawat, "Data augmentation techniques applied to medical images," *Int. J. Res. Publ. Rev.*, vol. 5, no. 7, pp. 483–501, Jul. 2024.

[44] T. Islam, M. S. Hafiz, J. R. Jim, M. M. Kabir, and M. F. Mridha, "A systematic review of deep learning data augmentation in medical imaging: Recent advances and future research directions," *Healthc. Anal.*, vol. 5, p. 100340, Jun. 2024.

[45] C. Chadebec and S. Allassonnière, "Data augmentation with variational autoencoders and manifold sampling," in *Deep Generative Models, and Data Augmentation, Labelling, and Imperfections*, pp. 184–192, 2021.

[46] E. Strelcenia and S. Prakoonwit, "Improving cancer detection classification performance using GANs in breast cancer data," *IEEE Access*, vol. 11, pp. 1–24, 2017.

[47] L. Wei, X. Linchuan, L. Zhixuan, W. Senzhang, C. Jiannong, M. Chao, and C. Xiaohui, "Sketch-then-edit generative adversarial network," *Knowl.-Based Syst.*, vol. 203, p. 106102, 2020.

[48] L. Cai, Y. Chen, N. Cai, W. Cheng, and H. Wang, "Utilizing Amari-alpha divergence to stabilize the training of generative adversarial networks," *Entropy*, vol. 22, p. 410, 2020.

[49] P. Zhaoqing, Y. Weijie, W. Bosi, X. Haoran, V. S. Sheng, L. Jianjun, and K. Sam, "Loss functions of generative adversarial networks (GANs): Opportunities and challenges," *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 4, no. 4, pp. 500–522, 2020.

[50] P. Sharma, M. Kumar, H. K. Sharma, and S. M. Biju, "Generative adversarial networks (GANs): Introduction, taxonomy, variants, limitations, and applications," *Multimed. Tools Appl.*, vol. 83, no. 41, pp. 88811-88858, 2024.

[51] Y. Kossale, M. Airaj, and A. Darouichi, "Mode collapse in generative adversarial networks: An overview," in *Proc. 2022 8th Int. Conf. Optimization Appl. (ICOA)*, pp. 1-6, Oct. 2022, doi: 10.1109/ICOA55659.2022.9934291.

[52] Z. Ding, S. Jiang, and J. Zhao, "Take a close look at mode collapse and vanishing gradient in GAN," in *Proc. 2022 IEEE 2nd Int. Conf. Electron. Technol., Commun. Inf. (ICETCI)*, pp. 597-602, May 2022, doi: 10.1109/ICETCI55101.2022.9832406.

[53] A. Martin, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Mach. Learn. Res.*, vol. 70, pp. 214–223, 2017.

[54] X. Guo, K. W. Seea, Y. Li, B. Bilal, Z. Liang, and Y. Wang, "A time-series Wasserstein GAN method for state-of-charge estimation of lithium-ion batteries," *J. Power Sources*, vol. 581, p. 233472, 2023.

[55] E. Strelcenia and S. Prakoonwit, "Improving cancer detection classification performance using GANs in breast cancer data," *IEEE Access*, vol. 1, pp. 71594–71615, 2023.

[56] Y. Xiao, J. Wu, and Z. Lin, "Cancer diagnosis using generative adversarial networks based on deep learning from imbalanced data," *Comput. Biol. Med.*, vol. 135, p. 104540, 2021.

[57] A. Kunapinun, M. N. Dailey, D. Songsaeng, M. Parnichkun, C. Keatmanee, and M. Ekpanyapong, "Improving GAN learning dynamics for thyroid nodule segmentation," *Ultrasound Med. Biol.*, vol. 49, no. 2, pp. 416–430, 2023.

[58] R. N. Abirami, P. M. D. R. Vincent, K. Srinivasan, U. Tariq, and C. Y. Chang, "Deep CNN and deep GAN in computational visual perception-driven image analysis," *Complexity*, vol. 2021, p. 5541134, 2021.

[59] W. Liu, L. Duan, Y. Tang, and J. Yang, "Data augmentation method for fault diagnosis of mechanical equipment based on improved Wasserstein GAN," in *Proc. Int. Conf. Prognostics Syst. Health Manag. (PHM-Jinan)*, pp. 103–111, 2020.

[60] Y. Jiménez-Gaona, D. Carrión-Figueroa, and M. J. Rodríguez-Álvarez, "GAN-based data augmentation to improve breast ultrasound and mammography mass classification," *Biomed. Signal Process. Control*, vol. 94, p. 106255, 2024.

[61] I. Showrov, A. M. D. Tarek, R. N. Hadiur, R. J. Jamin, M. F. Mridha, K. Mohsin, N. Arai, and J. Song, "Generative adversarial networks (GANs) in medical imaging: Advancements, applications, and challenges," *IEEE Access*, vol. 12, pp. 35728–35753, 2024.

[62] M. Alruily, W. Said, A. M. Mostafa, M. Ezz, and M. Elmezain, "Breast ultrasound images augmentation and segmentation using GAN with identity block and modified U-Net 3+," *Sensors*, vol. 23, no. 20, p. 8599, Oct. 2023.

[63] M. R. Islam, M. M. Rahman, M. S. Ali, A. A. N. Nafi, M. S. Alam, T. K. Godder, M. S. Miah, and M. K. Islam, "Enhancing breast cancer segmentation and classification: An ensemble deep convolutional

neural network and U-Net approach on ultrasound images," *Machine Learn. Appl.*, vol. 16, p. 100555, 2024.

[64] R. Remtulla, A. Samet, M. Kulbay, A. Akdag, A. Hocini, A. Volniansky, S. K. Ali, and C. X. Qian, "A future picture: A review of current generative adversarial neural networks in vitreoretinal pathologies and their future potentials," *Biomedicines*, vol. 13, no. 2, p. 284, Jan. 2025, doi: 10.3390/biomedicines13020284.

[65] A. Aggarwal, M. Mittal, and G. Battineni, "Generative adversarial network: An overview of theory and applications," *Int. J. Inf. Manage. Data Insights*, vol. 1, no. 1, p. 100004, Apr. 2021, doi: 10.1016/j.jjimei.2020.100004.

[66] M. Ibrahim, Y. Al Khalil, S. Amirrajab, C. Sun, M. Breeuwer, J. Pluim, B. Elen, G. Ertaylan, and M. Dumontier, "Generative AI for synthetic data across multiple medical modalities: A systematic review of recent developments and challenges," *Comput. Biol. Med.*, vol. 189, p. 109834, May 2025.

[67] N. Fatima, F. Mento, S. Afrakhteh, T. Perrone, A. Smargiassi, R. Inchingolo, and L. Demi, "Synthetic lung ultrasound data generation using autoencoder with generative adversarial network," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 72, no. 5, May 2025.

[68] T. Pang, J. H. D. Wong, W. L. Ng, and C. S. Chan, "Semi-supervised GAN-based radiomics model for data augmentation in breast ultrasound mass classification," *Computer Methods and Programs in Biomedicine*, vol. 203, p. 106018, 2021.

[69] M. R. Islam *et al*., "Enhancing breast cancer segmentation and classification: An Ensemble Deep Convolutional Neural Network and U-net approach on ultrasound images," *Machine Learning with Applications*, vol. 16, p. 100555, 2024.

[70] D. Kumar, M. A. Mehta, and I. Chatterjee, "Empirical analysis of deep convolutional generative adversarial network for ultrasound image synthesis," *Open Biomed. Eng. J.*, vol. 15, pp. 71–77, 2021.

[71] T. Kumar, R. Brennan, A. Mileo, and M. Bendechache, "Image Data Augmentation Approaches: A Comprehensive Survey and Future Directions," in *IEEE Access*, vol. 12, pp. 187536-187571, 2024.

[72] F. Rahat, M. S. Hossain, M. R. Ahmed, S. K. Jha, and R. Ewetz, "Data Augmentation for Image Classification Using Generative AI," *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 4173-4182, 2025.

[73] P. Sangeeta and B. Debnath, "An enhanced multi-scale deep convolutional orchard capsule neural network for multi-modal breast cancer detection," *Healthc. Anal.*, vol. 5, p. 100298, 2024.

[74] M. Alhassan and M. Fuseini, "Data augmentation: A comprehensive survey of modern approaches," *Array*, vol. 16, p. 100258, 2022.

[75] B. Kevin, M. Anna, M. Angel, and R. José, "Automatic generation of artificial images of leukocytes and leukemic cells using generative adversarial networks (SyntheticCellGAN)," *Comput. Methods Programs Biomed.*, vol. 229, p. 107314, 2023.

[76] C. Wang, B. Jin, Y. Wang, N. X. Neal, and Z. Sheng, "WGAN-E: A generative adversarial networks for facial feature security," *Electronics*, vol. 9, p. 486, 2020.

[77] H. Lee, J. Kim, E. K. Kim, and S. Kim, "Wasserstein generative adversarial networks based data augmentation for radar data analysis," *Appl. Sci.*, vol. 10, p. 1449, 2020.

[78] M. Han and Y. Cho, "De-blurring using perceptual similarity Wasserstein generative adversarial network based de-blurring using perceptual similarity," *Appl. Sci.*, vol. 9, p. 2358, 2019.

[79] T. Luca, M. A. Ali, and A. H. Hashim, "Novel hybrid integrated Pix2Pix and WGAN model with gradient penalty for binary images denoising," *Syst. Soft Comput.*, vol. 6, p. 200122, 2024.

[80] A. Ponraj, P. Nagaraj, D. Balakrishnan, P. N. Srinivasu, J. Shafi, W. Kim, and M. F. Ijaz, "A multi-patch-based deep learning model with VGG19 for breast cancer classifications in the pathology images," *Digital Health*, vol. 11, pp. 1–21, 2025, doi: 10.1177/20552076241313161.

[81] D. Diana, T. B. Kurniawan, D. A. Dewi, M. K. Alqudah, M. Z. Zakari, and E. F. B. E. Fuad, "Convolutional neural network based deep learning model for accurate classification of durian types," *J. Appl. Data Sci.*, vol. 6, no. 1, pp. 101–114, Jan. 2025.