

Article

Not peer-reviewed version

Data Augmentation of Breast Ultrasound Images Using Wasserstein Generative Adversarial Networks

[I Gede Susrama Mas Diyasa](#)*, Sayyidah Humairah, Eva Yulia Puspaningrum, Fara Disa Durry, [Wahyu Dwi Lestari](#), [Wahyu Caesarendra](#)

Posted Date: 28 October 2024

doi: 10.20944/preprints202410.2205.v1

Keywords: Wasserstein GAN; Breast Cancer; Ultrasound Image; Augmentation



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Data Augmentation of Breast Ultrasound Images Using Wasserstein Generative Adversarial Networks

I Gede Susrama Mas Diyasa ^{1,*}, Sayyidah Humairah ², Eva Yulia Puspaningrum ², Fara Disa Durry ³, Wahyu Dwi Lestari ⁴ and Wahyu Caesarendra ⁵

¹ Department of Magister Information Technology, University of Pembangunan Nasional Veteran Jawa Timur, Surabaya, Indonesia

^{2a} Department of Informatics, University of Pembangunan Nasional Veteran Jawa Timur, Surabaya, Indonesia

³ Faculty of Medicine, University of Pembangunan Nasional Veteran Jawa Timur, Surabaya, Indonesia

⁴ Department of Mechanical Engineering, Faculty of Engineering, University of Pembangunan Nasional Veteran Jawa Timur, Surabaya, Indonesia

⁵ Faculty of Mechanical Engineering, Opole University of Technology, Opole, Poland

* Correspondence: igsusrama.if@upnjatim.ac.id; Tel.: +062 81999471017

Abstract: Breast cancer is among the most prevalent cancers in Indonesia. One of the methods for early detection of breast cancer is utilizing ultrasound images to classify breast conditions into normal, benign tumors, or cancer. The advent of deep learning technology facilitates the image analysis process, such as the use of Transfer Learning Convolutional Neural Networks (CNNs). Generally, CNN models require large and balanced datasets to perform well in classification tasks. However, medical datasets like breast ultrasound images tend to be limited and imbalanced. The Wasserstein GAN (WGAN) is a generative data augmentation method capable of producing synthetic images by learning the pattern of the distribution of real image data. The implementation of the Wasserstein distance results in the training process of WGAN demonstrating stability from epochs 3000, 2500, and 3000 out of a total of 5000 epochs. The quality of the synthetic images improves with an increasing number of training iterations. By using WGAN, all of the evaluation metrics of each classifier are increasing, with the best accuracy score is achieved by VGG16 model with 83.33% accuracy

Keywords: wasserstein GAN; breast cancer; ultrasound image; augmentation

1. Introduction

Breast cancer is one of the most common and deadly diseases affecting women worldwide [1]. 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 [2]. Early detection is crucial for improving the prognosis and survival rates of breast cancer patients [3]. Routine examinations and screening using medical imaging technology play a vital role in detecting cancer at an early stage when treatment is more effective. Mammography, ultrasound, and magnetic resonance imaging (MRI) are some of the methods available for breast cancer detection [4]. Ultrasound, in particular, is a frequently used tool in breast examinations because it is non-invasive, relatively inexpensive, and easily accessible [3,5]. 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 [6–8].

Although ultrasound is a highly useful tool for breast cancer detection, it faces several significant challenges. Ultrasound images often suffer from suboptimal quality, influenced by noise and artifacts, which can obscure critical details and make cancer detection more difficult [9,10]. This noise and these artifacts can lead to diagnostic errors 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 [11–13]. Collecting and annotating high-quality and big amount of medical data requires significant resources and access to medical facilities [14]. 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. This variability can result in differences in diagnosis and treatment received by patients [15]. Although several automatic detection methods have been developed, many still struggle to adequately handle the complexity and variability of medical images. Existing detection algorithms may not be robust or accurate enough for widespread clinical application [16].

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 [17–19]. 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 [20,21]. However, the limited availability of medical data and class imbalance in the data often pose obstacles in developing accurate deep learning models [22]. 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 [23].

The commonly performed process is data augmentation using traditional techniques, which include geometric augmentation, color augmentation, and noise augmentation. These techniques encompass processes such as reflecting images, cropping and translating images, and altering the image color palette [24]. However, conventional augmentation methods have several drawbacks. Although these methods can increase data variability, they tend to be limited in generating sufficiently realistic variations. This is because the transformations performed are deterministic and often do not reflect the natural diversity of the actual data [25,26]. As an alternative, GAN (Generative Adversarial Network) is a generative data augmentation method that can produce synthetic data by learning the distribution patterns of original data [27]. GAN, introduced by Goodfellow, uses Jensen-Shannon (JS) divergence in the calculation of the loss function [28,29]. The implementation of JS divergence has a drawback of being identical to the occurrence of vanishing gradients, leading to unstable GAN training. Besides this issue, GAN often experiences mode collapse, where the GAN model fails to capture the diversity of the entire data distribution and is fixated on generating data with certain patterns [30]. To address the limitations of traditional GAN, a variant known as the Wasserstein Generative Adversarial Network (WGAN) was introduced [31,32]. WGAN uses the Wasserstein distance as a metric to measure the difference between the original data distribution and the generated data distribution, enabling more stable training and more realistic results [33,34].

In 2021, research by Xiao et al. [35] 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. According to the original WGAN training algorithm, the training process continues until the Generator converges [33,34]. However, in practice, the WGAN training process defines one of the hyperparameters prior to training, namely the number of epochs or training iteration steps [36–38]. 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

2. Materials and Methods

This study will use an annotated breast ultrasound image dataset to train and test the WGAN model [39]. 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 [40]. 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 Figure 1 and flowchart in Figure 2.

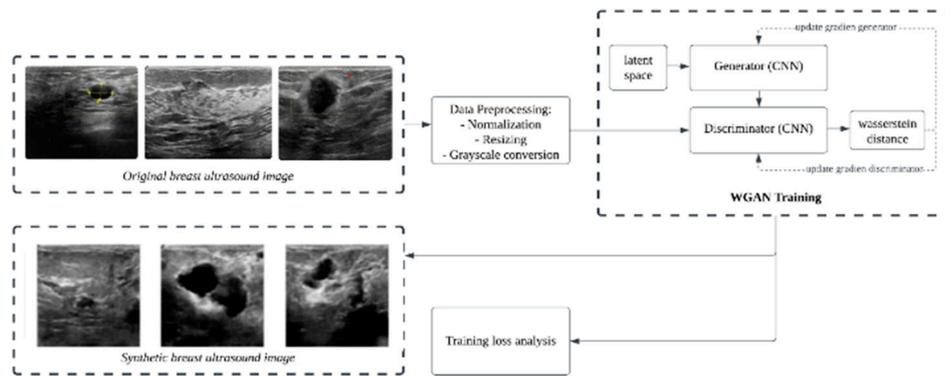


Figure 1. Block Diagram Ultrasound Image Augmentation Using WGAN.

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 [41]. 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. 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. 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.

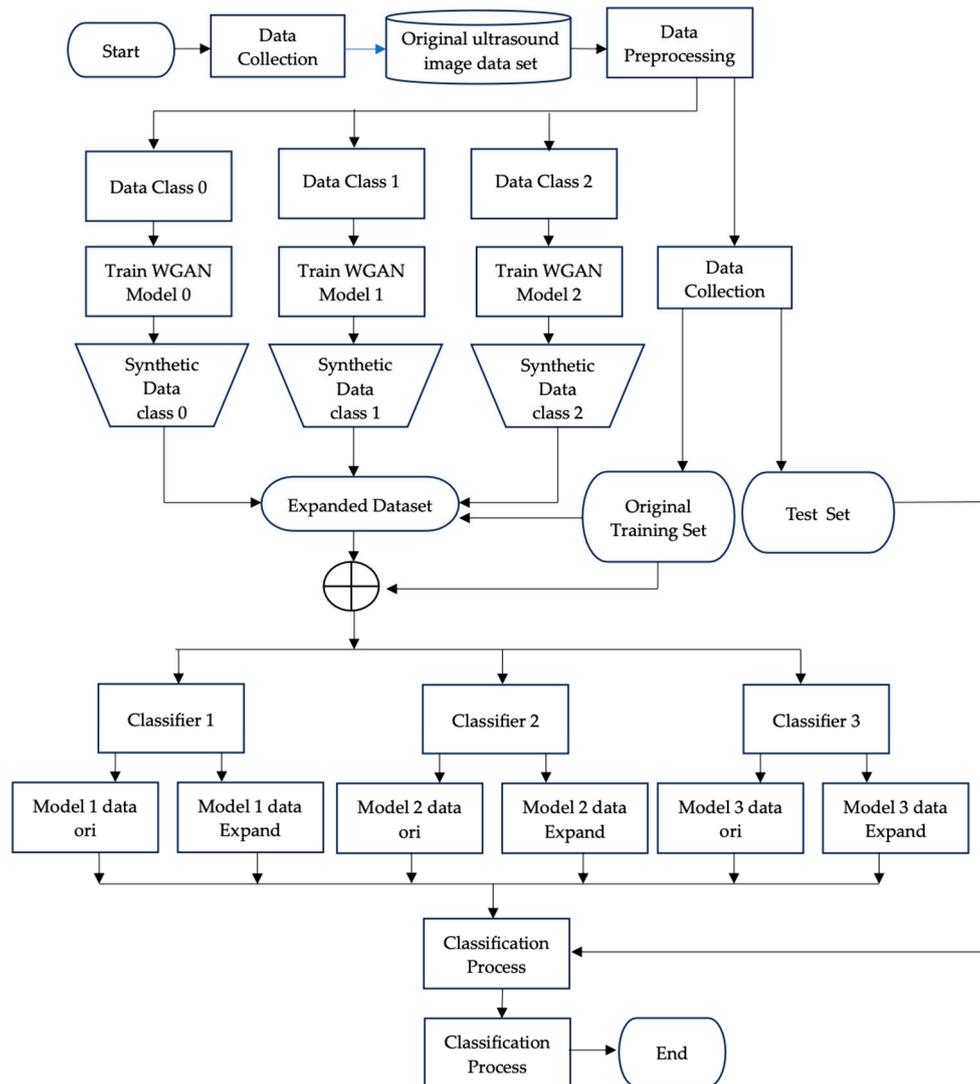


Figure 2. Flowchart Ultrasound Image Augmentation Using WGAN.

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 Figures 1 and 2.

2.1. 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 Figure 3.

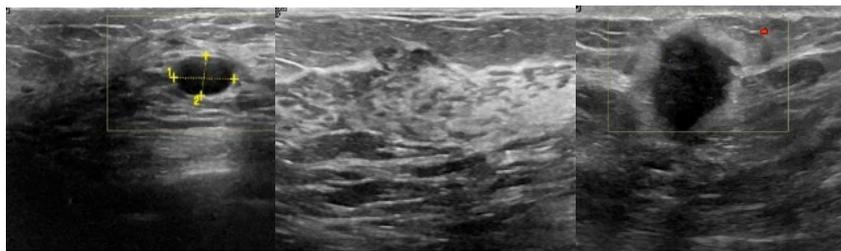


Figure 3. Dataset Sample from Class 0 (left), 1 (middle), 2 (right).

Figure 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 center, 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, Figure 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.

2.2. 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 [42,43]. 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 [44]. The normalization formula is shown in Equation (1).

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

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.
 - Pixel_normalized represents the normalized pixel intensity value.
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 x 128 pixels. This resizing process ensures that all image data will be compatible with the WGAN Generator and Discriminator networks [45,46].
 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.

2.3. Wasserstein GAN Training

The WGAN training process begins with initializing the parameters [47,48]. 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.

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 [49,50]. 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.

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

Table 1 shows the details of the algorithm of WGAN training according to the original WGAN research by Arjovsky.

Table 1. Wasserstein GAN Algorithm.

<i>WGAN Algorithm</i>
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 $t = 0, \dots, n_{critic}$ 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 $g_w \leftarrow \nabla_w \left[\frac{1}{m} \sum_{i=1}^m f_w(x^{(i)}) - \frac{1}{m} \sum_{i=1}^m f_w(g_\theta(z^{(i)})) \right]$ $w \leftarrow w + \alpha \cdot RMSProp(w, g_w)$ $w \leftarrow \text{clip}(w, -c, c)$ End for Sample $\{z^{(i)}\}_{i=1}^m \sim p(z)$ a batch of prior samples. $g_\theta \leftarrow -\nabla_\theta \frac{1}{m} \sum_{i=1}^m f_w(g_\theta(z^{(i)}))$ $\theta \leftarrow \theta - \alpha \cdot RMSProp(\theta, g_\theta)$ End while

2.3. 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.

3. Results and Discussion

3.1. 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 Figure 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 2 displays the hyperparameters involved in the WGAN training process and their values.

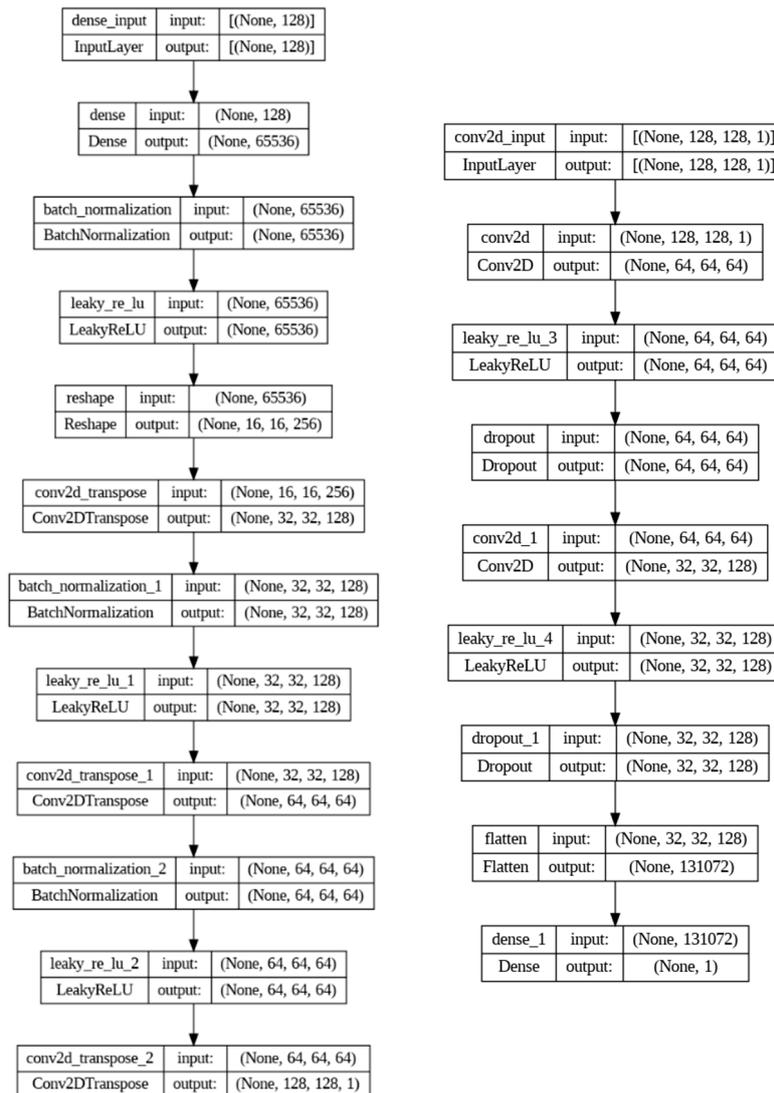


Figure 3. WGAN Architecture; Generator (left) and Critic (right).

Table 2. Hyperparameters

Hyperparameters	Value
Epoch	5000
Weight clipping value	0.01
Optimizer	RMSProp
Learning rate optimizer	5e-5
Batch size	128

<i>Hyperparameters</i>	<i>Value</i>
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. Figures 5–7 present the loss graphs for each WGAN training process. Figures 5, 6, and 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 Figures 5–7 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.

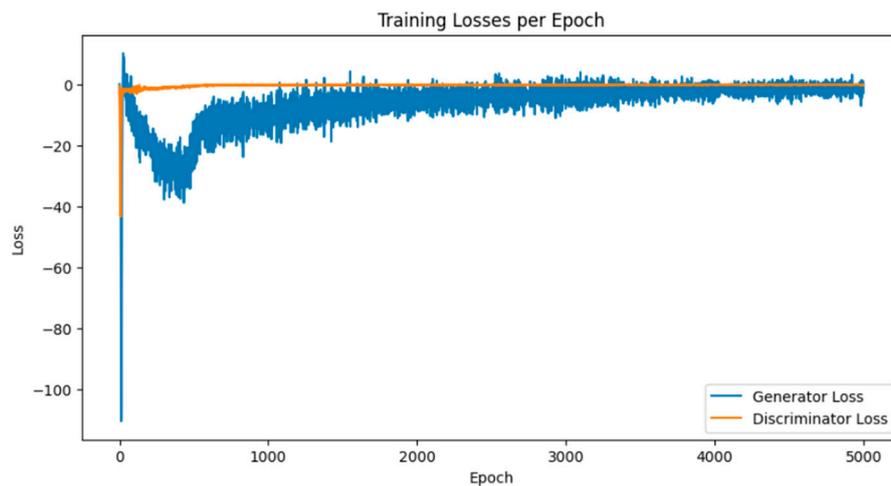


Figure 5. Training Loss WGAN Class 0.

Based on the analysis of the loss patterns, the model for class 0 (Figure 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 (Figure 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 (Figure 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.

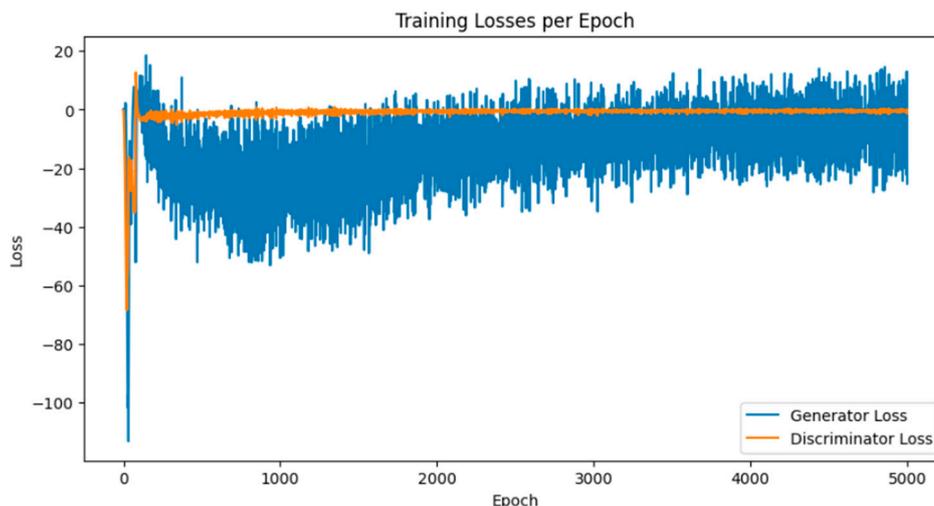


Figure 6. Training Loss WGAN Class 1

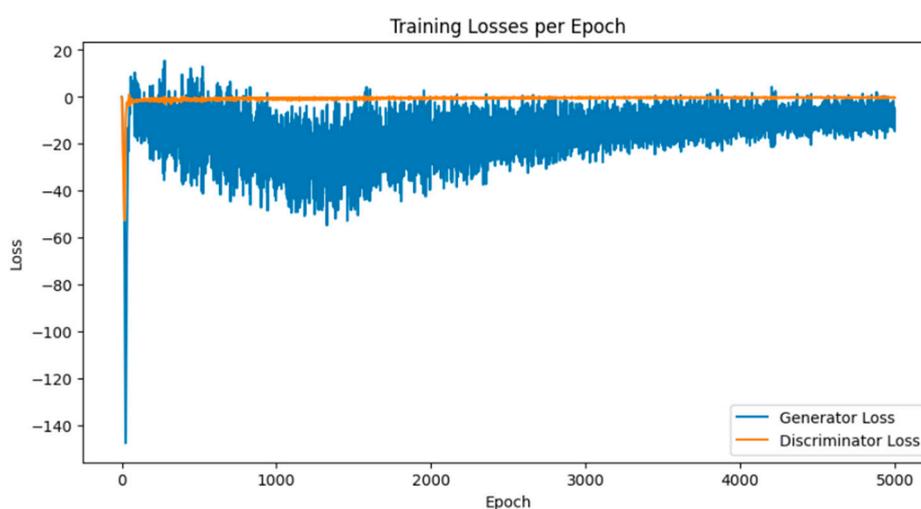


Figure 7. Training Loss WGAN Class 2

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.

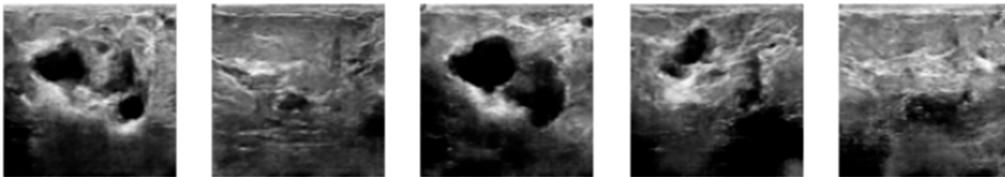
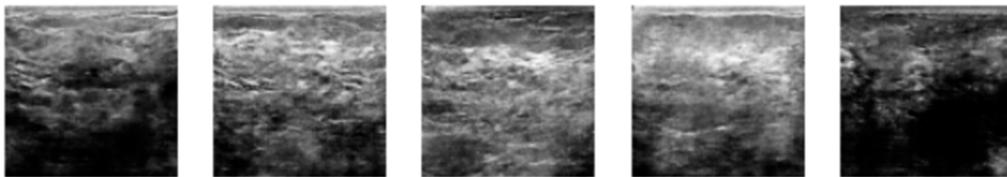
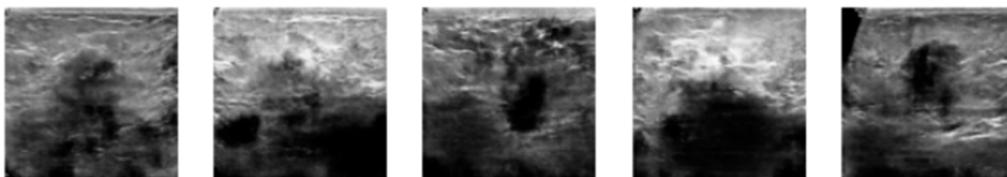
3.2. Image Synthetic Augmented by WGAN

The size of the expanded ultrasound image dataset, as shown in Table 3, compares the sizes of the original image dataset and the dataset after the augmentation process.

Table 3. Size of Dataset.

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

Figures 8–10 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. Figure 7 shows synthetic image samples for class 0 generated by the generator trained on class 0 data. Figure 8 shows synthetic image samples for class 1 (Normal category) generated by the WGAN model generator trained on class 1 data. Figure 9 shows synthetic image samples for class 2 (Malignant/Cancer category) generated by the WGAN model generator trained on class 2 image data.

**Figure 8.** Synthetic Image Class 0**Figure 9.** Synthetic Image Class 1**Figure 10.** Synthetic Image Class 2

3.3. 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 Tables 4–7.

Table 4. Evaluation Metrics of VGG16

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

Table 5. Evaluation Metrics of ResNet50

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

Table 6. Evaluation Metrics of MobileNetV2

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

Table 7. Evaluation Metrics of YOLOv8

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

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 Tables 4–7, 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.

4. 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.

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.

Supplementary Materials: The following supporting information can be downloaded at preprints.org, Figure S1: title; Table S1: title; Video S1: title.

Author Contributions: The writing of this paper involved significant contributions from several researchers, each playing a vital role in various aspects of the study. I Gede Susrama Mas Diyasa served as the lead author, guiding the overall manuscript preparation, formulation of the research idea, and result analysis. Sayyidah Humairah was responsible for conducting the coding and testing of the algorithms used in this research, ensuring that all technical procedures were executed effectively and yielded accurate data. Eva Yulia Puspaningrum focused on the machine learning analysis, evaluating the performance of the applied models and interpreting the results from a scientific data perspective. To ensure the accuracy of the findings in a medical context, Fara Disa Durry, an experienced medical doctor, handled the validation of the results, confirming that the research outcomes were applicable and relevant in the medical field. In terms of overall supervision, Caesarendra acted as the supervisor, providing guidance and direction throughout the research development process and ensuring that each phase adhered to academic standards. Finally, Wahyu Dwi Lestari played a crucial role in funding acquisition, securing the resources necessary to support this research. This solid collaboration among the researchers enabled the study to proceed smoothly and produced valuable findings that contribute to the advancement of science.

Funding: Funded by LPPM Universitas Pembangunan Nasional Veteran Jawa Timur through Penelitian Kolaborasi scheme with grant number 366/UN.63/LPPM/2024

Data Availability Statement: The data supporting the findings of the paper titled "Data Augmentation of Breast Ultrasound Images Using Wasserstein Generative Adversarial Networks" are not publicly available due to restrictions on the confidentiality of medical images. The original breast ultrasound images used in this study are derived from clinical sources and contain sensitive patient information. However, the augmented datasets generated using the Wasserstein Generative Adversarial Networks (WGANs) as part of this research are available from the corresponding author upon reasonable request, provided that appropriate ethical approvals are in place and confidentiality agreements are maintained. The implementation code for the data augmentation model, as well as detailed instructions for reproducing the experiments, can be accessed via a public repository, [https://s.id/Breast_MDPI], which is freely available for academic and non-commercial use.

Acknowledgments: Deepest gratitude is extended to the Data Science Laboratory for providing computing resources essential for conducting the computational tasks in this research. The technical support and provision of equipment played a crucial role in completing the data analysis and model testing. Sincere thanks are also directed to the leadership and staff of the Research and Community Service Institute (LPPM) at UPN Veteran Jawa Timur for their administrative support throughout the research process. Their assistance and coordination greatly facilitated the smooth progression of this study, ensuring its successful completion.

Conflicts of Interest: The authors declare no conflict of interest regarding this research. All findings are presented objectively and are not influenced by any personal interests that could affect the representation or interpretation of the data. The funding sponsor, LPPM UPN Veteran Jawa Timur, had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results. All research stages were conducted independently by the research team to ensure the integrity and validity of the reported outcomes.

References

- Hyuna, S.; Jacques, F.; Rebecca, L. S.; Mathieu, L.; Isabelle, S.; Ahmedin, J.; and Freddie, B. GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *Global Cancer Statistics* 2020, 2021, vol. 71 (3), pp. 209-249.
- Melina, A.; Eileen, M.; Harriet, R.; Allini, M.; Deependra, S.; Mathieu, L.; Jerome, V.; Julie, R. G.; Fatima, C.; Sabine, S.; and Isabelle, S. Current and future burden of breast cancer: Global statistics for 2020 and 2040. *The Breast*, 2022, vol. 66, pp.15-23.
- Haq, I. U.; Ali, H.; Wang, H. Y.; L. Cui, and Feng, J. BTS-GAN: Computer-aided segmentation system for breast tumor using MRI and conditional adversarial networks. *Engineering Science and Technology, an International Journal*, 2022, Vol. 36, 101154, pp. 1-10.
- Jiawei, L.; Xin, G.; Zhimin, F.; Lai-Ming, C.; Yan, L.; Xiaojia, W.; Wen-Ming, C. and Dong-Xu, L. Non-invasive biomarkers for early detection of breast cancer. *Cancer MDPI*, 2020, Vol. 12 (2767), pp. 1-28.
- Ghulam, G.; Anum, S.; Syeda, N. B.; Aqsa, K.; Hina, S.; Sadia, P.; Akkasha, L.; Sajid, M. and Saeed, M. Digital Image Processing for Ultrasound Images: A Comprehensive Review. *International Journal of Innovation, Creativity and Change*, 2021, Vol. 15 (3), pp. 1335-1354.
- Alam, N. A.; Khandaker, M. M. U.; Mahbubur, R.; Manu, M. M. R.; and Mostofa, K. N. A Novel Automated System to Detect Breast Cancer From Ultrasound Images Using Deep Fused Features With Super Resolution. *intelligence-based Medicine*, 2024, Vol. 10 (100149), pp. 1-15.
- Rebeca, T.; David, M.; Carlos, I. I.; Rodrigo, A. G.; Fernando, A. V.; and Joaquin, L. H. Recent Advances in Artificial Intelligence-Assisted Ultrasound Scanning. *Applied Sciences*, 2023, Vol. 13 (3693), pp. 1-17.
- Yu, W.; Xinke, G.; He, M.; Shouliang, Q.; Guanqing Z.; and Yudong, Y. Deep Learning in Medical Ultrasound Image Analysis: A Review. *IEEE Access*, 2016, Vol.4, pp. 1-15.
- Alyaa, R. Effects of Artifacts on the Diagnosis of Ultrasound Image. *Medico-legal Update*, 2021, vol. 21 (4), pp. 327-336.
- Zeyad, O. M.; and Latif, A. E. E. Breast Cancer Detection in Ultrasound Imaging. *World Journal of Advanced Research and Reviews*, 2021, vol 12 (01), pp. 308-314.
- Randall, J. E.; Ryan, M. S.; and Alfonso, L. Twelve key challenges in medical machine learning and solutions. *Intelligence-Based Medicine*, 2022, vol. 6 (100068), pp.1-9.
- Imane, A.; Ali, I.; and Ikram, C. Cost-sensitive Learning for Imbalanced Medical Data: A Review. *Artificial Intelligence Review*, 2024, vol. 57 (80), pp. 1-72.
- Chiranjib, C.; Manojit, B.; Soumen, P.; and Sang-soo, L. From machine learning to deep learning: Advances of the recent data-driven paradigm shift in medicine and healthcare. *Current Research in Biotechnology*, 2024, vol. 7 (100164), pp. 1-20.
- Yang, Y.; Li, R.; Xiang, Y.; Lin, D.; Yan, A.; Chen, W.; Li, Z.; Lai, W.; Wu, X.; Wan, C.; Bai, W.; Huang, X.; Li, Q.; Deng, W.; Liu, X.; Lin, Y.; Yan, P.; and Lin, H. Expert Recommendation on Collection, Storage, Annotation, And Management Of Data Related To Medical Artificial Intelligence. *Intelligence Medicine*, 2023, vol. 3, pp. 144-149.
- Hussain, B. Z.; Andleeb, I.; Ansari, M. S.; Joshi, A. M. and Kanwal, N. Wasserstein GAN based Chest X-Ray Dataset Augmentation for Deep Learning Models: COVID-19 Detection Use-Case. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 2058–2061
- Haq, D. Z. and Fatichah, C. Ultrasound Image Synthetic Generating Using Deep Convolution Generative Adversarial Network for Breast Cancer Identification. *IPTEK The Journal of Technology and Science*, 2021. Vol. 34, (1), pp. 13-25.
- Carrero, A.; Groenhoff, L.; Vologina, E.; Basile, P.; and Albera, M. Deep Learning in Breast Cancer Imaging: State of the Art and Recent Advancements in Early 2024. *Diagnostics*. 2024, vol. 14 (848), pp. 1-36.
- Nikmah, T. L.; Syaifei, R. M.; and Anisa, D. N. Inception ResNet v2 for Early Detection of Breast Cancer in Ultrasound Images. *Journal of Information System Exploration and Research*, 2024, vol. 2 (2), pp. 93-102
- Gede, I. S. M. D; Eva, Y. P.; Hatta, M.; and Ariono, S. New Method For Classification Of Spermatozoa Morphology Abnormalities Based On Macroscopic Video Of Human Semen. *2019 International Seminar on Application for Technology of Information and Communication (iSemantic)*, 2019, pp. 133-140.
- Alessandro, C.; Léon, G.; Elizaveta, V.; Paola, B. and Marco A. Deep Learning in Breast Cancer Imaging: State of the Art and Recent Advancements in Early 2024. *Diagnostics Journal*, 2024, Vol. 14(848), pp. 1-36
- Heikal, A.; Amir, E. G.; Samir, E.; and Rashad, M. Z. Fine Tuning Deep Learning Models For Breast Tumor Classification. *Scientific Reports*, 2024, vol. 14 (10753), pp. 1-26.
- Rguibi, Z.; Hajami, A.; Zitouni, D.; Maleh, Y. and Elqaraoui, A. Medical variational autoencoder and generative adversarial network for medical imaging. *Indonesian Journal of Electrical Engineering and Computer Science*, 2023. Vol. 32 (1), pp. 494–505.
- Krizhevsky, A.; Sutskever, I. and Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*. 2017. Vol. 60 (6), pp. 84-90

24. Perez, L. and Wang, L. The Effectiveness of Data Augmentation in Image Classification using Deep Learning, 2017. Pp. 1-8
25. Shangguan, Z.; Zhao, Y.; Fan, W. and Cao, Z. Dog Image Generation using Deep Convolutional Generative Adversarial Networks. *International Conference on Universal Village, UV 2020, IEEE Xplor*, Institute of Electrical and Electronics Engineers Inc. 2020, pp. 1-6
26. Chadebec, C. and Allasonnière, S. Data Augmentation with Variational Autoencoders and Manifold Sampling. *Deep Generative Models, and Data Augmentation, Labelling, and Imperfections*, Springer International Publishing, 2021, pp. 184–192.
27. Strelcenia, E and Prakoonwit, S. Improving Cancer Detection Classification Performance Using GANs in Breast Cancer Data. *IEEE Access*, 2017. Vol. 11, pp. 1-24
28. Wei, L.; Linchuan, X.; Zhixuan, L.; Senzhang, W.; Jiannong, C.; Chao, M.; and Xiaohui, C. Sketch-then-Edit Generative Adversarial Network. *Knowledge-Based Systems*, 2020, Vol. 203 (106102), pp. 1-11
29. Cai, L.; Chen, Y.; Cai, N.; Cheng, W.; and Wang, H. Utilizing Amari-Alpha Divergence to Stabilize the Training of Generative Adversarial Networks. *Entropy*, vol. 22 (410), pp. 1-19.
30. Zhaoqing, P.; Weijie, Y.; Bosi, W.; Haoran, X.; Victor, S. S.; Jianjun, L.; and Sam, K. Loss functions of generative adversarial networks (gans): Opportunities and challenges," *IEEE Transactions Emerging Topics in Computational Intelligence*, 2020, Vol. 4 (4), pp. 500–522.
31. Sharma, P.; Kumar, M.; Sharma, H. K.; and Biju, S. M. Generative adversarial networks (GANs): Introduction, Taxonomy, Variants, Limitations, and Applications. *Multimedia Tools and Applications*, 2024, pp. 1-49.
32. Martin, A.; Soumith, C.; and Leon, B. Wasserstein Generative Adversarial Networks, *Proceedings of Machine Learning Research*, 2017, vol. 70, pp. 214-223.
33. Xinyu, G.; Seea, K.W.; Yabin, L.; Bilal, B; Liang, Z.; and Yunpeng, W. A time-series Wasserstein GAN method for state-of-charge estimation of lithium-ion batteries, *Journal of Power Sources*, 2023, vol. 581(233472), pp. 1-15.
34. Emilija, S.; and Simant, P. Improving Cancer Detection Classification Performance Using GANs in Breast Cancer Data, *IEEE Access*, 2023, Vol. 1, pp. 71594-71615.
35. Xiao, Y.; Wu, J. and Lin, Z. Cancer diagnosis using generative adversarial networks based on deep learning from imbalanced data. *Computer in Biologi and Medicine*, 2021. Vol. 135 (104540). pp. 1-10.
36. Kunapinun, A.; Dailey, M. N.; Songsaeng, D.; Parnichkun, M.; Keatmanee, C. and Ekpanyapong, M. Improving GAN Learning Dynamics for Thyroid Nodule Segmentation. *Ultrasound in Medicine & Biology*, 2023. Vol. 49 (2), pp. 416–430,
37. Abirami, R.N.; Durai Raj Vincent, P. M.; Srinivasan, K., Tariq, U. and Chang, C. Y. Deep CNN and Deep GAN in Computational Visual Perception-Driven Image Analysis. *Complexity, Hindawi*, 2021. Vol. 2021 (5541134), pp. 1-30.
38. Liu, W.; Duan, I.; Tang, Y.; and Yang, J. Data Augmentation Method for Fault Diagnosis of Mechanical Equipment Based on Improved Wasserstein GAN. *Proceedings International Conference on Prognostics and System Health Management, PHM-Jinan 2020, IEEE-Xplor*, 2020, pp. 103–111.
39. Jiménez-Gaona, Y.; Carrión-Figueroa, D.; and Rodríguez-Álvarez, M. J. Gan-based data augmentation to improve breast ultrasound and mammography mass classification. *Biomedical Signal Processing and Control*, 2024, vol. 94 (106255), pp. 1-18.
40. Showrov, I.; Tarek, A. MD.; Hadiur, R. N.; Jamin, R. J.; Mridha, M. F.; Mohsin, K.; , Nobuyoshi, A.; and Jungpil, S.; Generative Adversarial Networks (GANs) in Medical Imaging: Advancements, Applications, and Challenges. *IEEE Access*.2024. Vol. 12, pp. 35728-35753
41. Alruily, M.; Said, W.; Mostafa, A. M.; Ezz, M.; and Elmezain, M. Breast Ultrasound Images Augmentation and Segmentation Using GANwithIdentity Block and Modified U-Net 3+. *Sensors*, 2023, col. 23 (8599), pp. 1-16.
42. Islam, M. R.; Rahman, M. M.; Ali, M. S.; Nafi, A. A. N.; Alam, M. S.; Godder, T. K.; Miah, M. S.; and Islam, M. K. Enhancing Breast Cancer Segmentation and Classification: An Ensemble Deep Convolutional Neural Network And U-Net Approach on Ultrasound Images. *Machine Learning with Application*, 2024, vol. 16 (100555), pp. 1-12.
43. Sangeeta, P.; and Debnath, B. An Enhanced Multi-Scale Deep Convolutional Orchard Capsule Neural Network For Multi-Modal Breast Cancer Detection. *Healthcare Analytics*, 2024, vol. 5 (100298), pp. 1-17.
44. Alhassan, M.; Fuseini, M. Data augmentation: A comprehensive survey of modern approaches. *Array*. 2022. Vol. 16 (100258), pp. 1-29.
45. Kevin, B.; Anna, M.; Angel, M.; and José, R. Automatic generation of artificial images of leukocytes and leukemic cells using generative adversarial networks (syntheticcellgan). *Computer Methods and Programs in Biomedicine*, 2023, vol. 229 (107314), pp. 1-16.
46. Chunxue, W.; Bobo, J.; Yan, W.; Neal, N. X.; and Sheng, Z. WGAN-E: A Generative Adversarial Networks for Facial Feature Security. *Electronics*, 2020, vol. 9, 486, pp. 1-20.

47. Hansoo L.; Jonggeun, K.; Eun Kyeong, K.; and Sungshin, K. Wasserstein Generative Adversarial Networks Based Data Augmentation for Radar Data Analysis. *applied sciences*, 2020, vol. 10 (1449), pp. 2-16.
48. Minsoo, H.; and Yoonsik, C.; De-blurring using Perceptual Similarity Wasserstein Generative Adversarial Network Based De-Blurring Using Perceptual Similarity. *Applied Sciences*, 2019, vol. 9 (2358), pp. 1-27.
49. Luca, T.; Ali, M. A.; and Hashim A. H. Novel hybrid integrated Pix2Pix and WGAN model with Gradient Penalty for binary images denoising. *Systems and Soft Computing*, 2024, vol. 6 (200122), pp. 1-11.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.