Indonesian Journal of Electronics, Electromedical Engineering, and Medical Informatics e-ISSN: 2656-8624
Homepage: https://ijeeemi.org/; Vol. 7, No. 3, pp. 526-538, August 2025

RESEARCH PAPER

OPEN ACCES

Implementation of the Wasserstein Generative Adversarial Network with Gradient Penalty (WGAN-GP) Method to Address Class Imbalance in Alzheimer's Disease Magnetic Resonance Imaging (MRI) Datasets

Muhammad Faiq Alamudin[1], Muhammad Itqan Mazdadi[2], Radityo Adi Nugroho[3], Triando Hamonangan Saragih[4], Muliadi[5] , and Vijay Anant Athavale

[1] Department of Computer Science, Faculty of Mathematics and Natural Science, Lambung Mangkurat University, Banjarbaru, Indonesia
[2] Walchand Institute of Technology, Solapur, India

ABSTRACT

Class imbalance in medical imaging datasets often leads to biased machine learning models, particularly in Alzheimer's disease (AD) diagnosis using MRI. This study proposes the use of Wasserstein Generative Adversarial Networks with Gradient Penalty (WGAN-GP) to mitigate class imbalance in AD MRI datasets. Realistic MRI images were synthesized for underrepresented AD stages, and the quality of the generated data was quantitatively validatedusing the Fréchet Inception Distance (FID), with the lowest FID score recorded at 31.84, indicating a high degree of realism and diversity. The synthetic images were used to augment a dataset of 6,400 T1-weighted scans for training four Convolutional Neural Network (CNN) architectures: ResNet-50, AlexNet, VGG-16, and VGG-19. Results demonstrated statistically significant improvements in balanced accuracy across all models ($p < 0.01$ for all comparisons). The AlexNet + WGAN-GP combination achieved the highest accuracy of 98.54%, representing a mean improvement of 4.76% (95% CI: 2.45% to 6.98%) over its baseline. Significant gains were also observed for ResNet-50, VGG-16, and VGG-19. These enhancements were consistent across multiple evaluation metrics, including precision, recall, F1-score, and AUC. These findings confirm that WGAN-GP is a highly effective and statistically validated strategy for boosting the diagnostic accuracy of CNN models in Alzheimer's disease classification.

PAPER HISTORY
Received March 2, 2025
Revised June 20, 2025
Accepted August 2, 2025
Published August 6, 2025

KEYWORDS
WGAN-GP;
Alzheimer's Disease;
CNN;
Synthetic Data Generation;
Class Imbalance

AUTHOR EMAIL
fqalamudin@gmail.com
mazdadi@ulm.ac.id
radityo.adi@ulm.ac.id
triando.saragih@ulm.ac.id
muliadi@ulm.ac.id
vijay.athavale@gmail.com

1. INTRODUCTION

Alzheimer's disease is a progressive and incurable neurodegenerative disorder characterized by the destruction of brain cells, leading to severe memory impairment. Early detection of Alzheimer's disease can mitigate further neuronal damage, thereby preventing irreversible memory loss and preserving cognitive function in patients [1]. In the early detection of Alzheimer's disease, neuroimaging stands as the most promising approaches, as brain degeneration can be non-invasively identified through significant cerebral atrophy using magnetic resonance imaging (MRI) [2]. The insidious nature of Alzheimer's disease, distinguished by its gradual onset of symptoms, makes accurate and early diagnosis difficult [3]. In this context, deep learning approaches such as Convolutional Neural Networks (CNNs) offer innovative solutions by automatically processing unstructured data such as MRI images without relying on manual feature extraction [4]. However, the imbalances in medical datasets, particularly those containing rare cases, can lead to performance losses and misleading results during model training. The Alzheimer's MRI Preprocessing Dataset sourced from Kaggle exhibits significant class imbalance [3]. It comprises four classes: Mild Demented (896 images), Moderate Demented (64 images), Very Mild Demented (2,240 images), and Non Demented (3,200 images). This dataset is categorized as unstructured data, consisting entirely of MRI images. Classes with relatively fewer samples compared to others may result in prediction errors and adversely affect model outcomes. To address this imbalance issue, data augmentation techniques can be employed to increase the number of images in underrepresented classes while mitigating class imbalance [5].

In this study, we use the variation of GAN to handle the class imbalance issue. The Generative Adversarial Networks (GAN) model consists of two neural networks, the discriminator and the generator, designed to generate synthetic data for unstructured data types such as images [6]. Generative Adversarial Networks (GANs) are grounded in an unsupervised deep learning framework. The *generator* (G) receives random noise vectors as input and synthesizes data samples that approximate real data

Corresponding author: Muhammad Itqan Mazdadi, mazdadi@ulm.ac.id, Department of Computer Science, Lambung Mangkurat University, Jl. A. Yani Km 36, Banjarbaru 70714, Indonesia. **DOI:** https://doi.org/10.35882/ijeeemi.v7i3.109
Copyright © 2025 by the authors. Published by Jurusan Teknik Elektromedik, Politeknik Kesehatan Kemenkes Surabaya Indonesia. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

distribution. Concurrently, the *discriminator* (D) evaluates both authentic and synthetic data instances, aiming to distinguish between them. The adversarial training process involves alternating iterative updates: G is optimized to generate realistic samples capable of deceiving D, while D is trained to maximize its discriminative accuracy. This min-max competition drives both networks toward equilibrium, where synthetic data becomes nearly indistinguishable from real samples [7]. The WGAN model was introduced to address the vanishing gradient problem in GANs, which arises from the use of divergence measure such as Jensen-Shannon (JS) or Kullback-Leibler (KL) to measure the distance between real and model-generated data distribution. Instead, WGAN employs the Wasserstein Distance, enabling more stable and meaningful training updates [8]. However, Wasserstein Generative Adversarial Networks (WGANs) can still suffer from unstable gradient values, which may become either vanishingly small or excessively large. To address this issue, the WGAN with gradient penalty method (WGAN-GP) is employed. This approach constrains gradient norms, ensuring Lipschitz continuity in the critic network. Compared to the original WGAN, WGAN-GP achieves faster convergence, more stable training dynamics, and generates higher-quality synthetic images [9].

Various strategies have been explored to address class imbalance in AD classification, with common approaches including oversampling techniques and algorithmic adjustments. A comparative study, for instance, evaluated several of these methods on an Alzheimer's MRI dataset. It was found that a widely used oversampling method like SMOTE could paradoxically decrease overall model accuracy compared to the baseline. While the ADASYN method offered some improvement, an algorithmic approach known as Weight Balancing proved to be the most effective among the compared methods, demonstrating superior performance across key evaluation metrics [3].

The classification of Alzheimer's disease stages from MRI scans using basic CNN, VGG-16, and VGG-19 models achieved a maximum accuracy of 77.66% (VGG-19), indicating considerable room for improvement [10]. Prior studies [11] utilizing the same Alzheimer's MRI benchmark dataset from Kaggle reported suboptimal performance, with baseline CNN architecture employing VGG-16 and ResNet-50 achieving only 85.07% and 75.25% accuracy, respectively . However, this suboptimal accuracy highlights the challenges posed by class imbalance and underscores the need for advanced techniques such as synthetic data augmentation to improve model generalizability. Therefore, integrating WGAN-GP offers a robust solution. Unlike traditional methods, WGAN-GP mitigates class imbalance by synthesizing high-fidelity MRI images for underrepresented Alzheimer's stages (Mild, Moderate, and Very Mild) through its gradient penalty mechanism, which enforces Lipschitz continuity without relying on restrictive weight clipping. This approach ensures stable training dynamics and generates realistic data augmentation, enabling CNNs to learn discriminative patterns directly from raw pixel data without manual feature engineering. Consequently, this hybrid approach enhances model generalizability by balancing class distributions and capturing intricate pathological features critical for accurate Alzheimer's classification.

Based on the previous description, this research aims to improve the early detection of Alzheimer's disease (AD) stages with higher accuracy using CNN architectures (AlexNet, ResNet-50, VGG-16, and VGG-19), while reducing dependency on large, perfectly balanced datasets through synthetic data augmentation via Wasserstein Generative Adversarial Networks with Gradient Penalty (WGAN-GP). The main contributions of this study are as follows, 1) this research highlights the importance of applying Wasserstein Generative Adversarial Networks with Gradient Penalty (WGAN-GP) in the classification of Alzheimer's disease stages, 2) it gives a comparative analysis of CNN architectures in the classification of Alzheimer's disease stages, 3) it proposes a reproducible framework for synthetic data augmentation in medical imaging using WGAN-GP to address class imbalance, which can be generalized to other neurodegenerative disease classification tasks. In addition, this study contributes to the development of more effective diagnostic tools for Alzheimer's disease (AD) stage classification

The structure of this paper is as follows: Section II discusses the dataset used, the proposed methods, and the proposed training and testing schemes. Section III displays the results of baseline CNN architectures accuracy and the use of WGAN-GP. Section IV discusses the interpretation of the findings, and the comparison with other studies and limitations of the approach. Finally, Section V, concludes the study by restating the objectives, summarizing the main findings, and outlining directions for future works.

2. MATERIALS AND METHOD

A. Dataset

The dataset used in this study is an image dataset obtained from Kaggle, titled the Alzheimer's Dataset, consisting of MRI images, categorized into four distinct classes representing the progression of Alzheimer's disease based on MRI scans: Mild Demented (896 images), Moderate Demented (64 images), Non-Demented (3,200 images), and Very Mild Demented (2,240 images). These classes reflect varying stages of cognitive impairment, which are critical for understanding disease progression [12]. Table 1 displays the statictical distribution of thedataset collected from Kaggle before employing WGAN-GP. The dataset had been previously preprocessed by the original author on the Kaggle platform, including cropping all images to a standardized size of 128x128 pixels. Previous research from [5], [13], [14] has also used the same dataset, which used deep learning especially CNN models to classify Alzheimer's disease. Fig. 2. presents samples image from each class in the dataset. The dataset was partitioned using a standard hold-out method using a one-fold approach, meaning the data was divided into a single, fixed training and testing set for all experiments. Specifically, 80% of the data was allocated for training, and the remaining 20%

Corresponding author: Muhammad Itqan Mazdadi, mazdadi@ulm.ac.id, Department of Computer Science, Lambung Mangkurat University, Jl. A. Yani Km 36, Banjarbaru 70714, Indonesia. **DOI:** https://doi.org/10.35882/ijeeemi.v7i3.109
Copyright © 2025 by the authors. Published by Jurusan Teknik Elektromedik, Politeknik Kesehatan Kemenkes Surabaya Indonesia. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

was reserved as a test set for final evaluation. The test set was strictly separated from the training and validation data to provide an unbiased evaluation of the final model's performance on unseen data.

Table 1. Statistical distribution of the dataset prior to WGAN-GP implementation

Class	Training	Testing	Total	Percentage
Mild	716	180	896	14%
Moderate	51	13	64	1%
Non	2,560	640	3,200	50%
Very Mild	1,792	448	2,240	35%

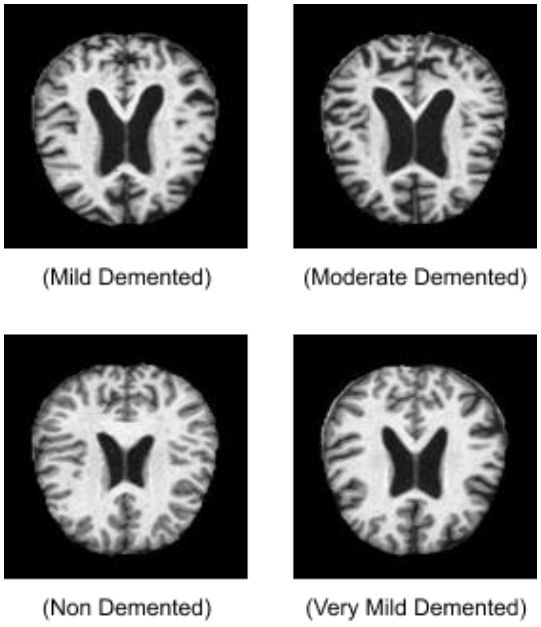

Fig. 2. Sample Images from each class in the dataset provided for this research

B. Data Augmentation

When a minority class has notably fewer samples than the majority class, it results in class imbalance, which poses challenges for classifier models and gives poor generalization performance. Data augmentation mitigates data collection challenges in resource-limited scenarios [15]. Current approaches to addressing the data imbalance problem are generally categorized into two main classes of solutions. The first class is characterized by the utilization of data-level preprocessing techniques, including oversampling and undersampling, to achieve a more balanced data distribution [16]. Various studies have employed a diverse range of methods to manage the issue of imbalanced datasets, including ADASYN [17], [18], SMOTE [3], [19], ROS [18] and WGAN-GP [7], [16], [20], [21].

This research utilizes WGAN-GP, a variation of GANs to address the inherent class imbalance within the employed dataset. To achieve this, the model was trained using pre-cropped 128x128 pixel images from the Kaggle dataset. These images were then further processed by normalizing their pixel values to a [-1, 1] scale and augmenting them with horizontal flipping for increased variability.

GANs employ generative models to produce entirely new data points that closely resemble the real data distribution. Approaches of this generative nature allow for the construction of artificial data points that are not only realistic but also proficient in capturing key distributional attributes of the training data [22].

The original GANs architercture is considered highly successful, especially in its capability to generate realistic, high-resolution images. What distinguishes GANs from other generative modeling techniques is their foundation in game theory, as opposed to traditional optimization. More specifically, GANs generate data by leveraging a

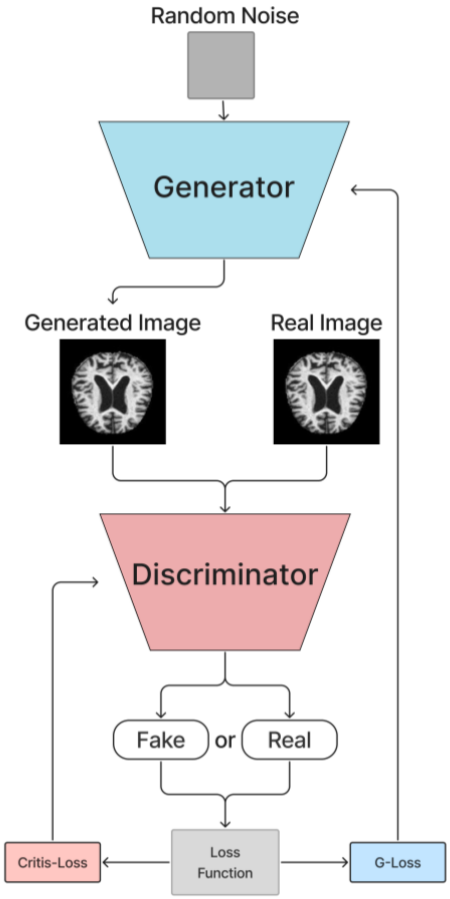

Fig. 1. A schematic of the generator and discriminator networks constituting a GANs

competitive dynamic between two neural networks. The *generator*, this network implicitly defines the model's probability distribution. Its main job is to create samples from this distribution. It takes a random noise and learns a function to transform this noise into realistic-looking samples. The *discriminator*, this network looks at a sample. It estimates whether it's a real sample (from the original training data) or fake (produced by the generator). In essence, the *discriminator* tries to tell real data apart from generated data [23]. A common analogy used to explain GANs involves a counterfeiter and a police detective. The *generator* acts as the counterfeiter who

Corresponding author: Muhammad Itqan Mazdadi, mazdadi@ulm.ac.id, Department of Computer Science, Lambung Mangkurat University, Jl. A. Yani Km 36, Banjarbaru 70714, Indonesia. **DOI:** https://doi.org/10.35882/ijeeemi.v7i3.109
Copyright © 2025 by the authors. Published by Jurusan Teknik Elektromedik, Politeknik Kesehatan Kemenkes Surabaya Indonesia. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

makes fake money, learning to make better fakes to avoid detection. The *discriminator* is the detective, learning to improve at finding flaws in the counterfeits. This ongoing contest drives both the counterfeits and the detective to progressively enhance their respective capabilities [24]. Fig. 1. illustrates how GANs work with their generator and discriminator.

The Wasserstein GAN (WGAN), introduced by [8], utilizes the Wasserstein distance to generate helpful gradients for updating the generator component. Although WGAN provides more stable training compared to standard GANs, its original weight clipping technique can prevent convergence. Therefore [9] proposed an improved version of WGAN replaces weight clipping with a gradient penalty. The gradient penalty further enhances training stability, encouraging the *discriminator* to behave as an actual Wasserstein distance estimator [22]. In essence, WGAN-GP provides enhanced training stability, essential for generating high-quality medical images, by resolving common GAN issues like mode collapse during training. This is achieved through the use of a Wasserstein loss function combined with a gradient penalty that enforces the Lipschitz constraint, resulting in smoother and more reliable gradients [25].

More sophisticated approaches for image generation have been developed in recent times. StyleGAN and improved StyleGAN [26], [27]. The style-based design and progressive training approach enabled high-resolution image generation with fine-grained attribute control. However, while these methods produce highly realistic natural images, these model rely on massive datasets and intensive computational power restricts its use in medical imaging. WGAN-GP, conversely, maintains consistent performance using smaller data samples and less complex architectures [25].

BigGAN [28] elevates image quality through larger model architectures, broader datasets, class-conditional generation, and orthogonal regularization; however, its substantial computational and data requirements limit medical imaging applications. WGAN-GP provides a resource-efficient and scalable alternative for synthetic medical image generation, without compromising output fidelity.

By coordinating a mixture of expert GANs, MEGAN [29] generates images conditioned on multimodal inputs including text, visual data, and categorical labels. This strategy yields richer diversity and higher-fidelity outputs in multimodal synthesis. Such capabilities hold promise for medical imaging, where combining scans with clinical annotations or reports could enhance synthetic image utility. Yet, the architecture's inherent complexity and intensive training requirements pose challenges for medical applications with limited data availability.

Generative Adversarial MultiTask Learning (GAMTL) [30] synthesizes images while executing complementary tasks like facial recognition or sketch generation. This joint training leverages task relationships to enhance output quality. For medical imaging, integrating diagnostic objectives could ensure clinically viable synthetic data for downstream tools. However, coordinating balanced learning across multiple tasks increases training complexity.

Despite newer approaches, researchers [25] point out that WGAN-GP continues to be a highly effective and practical option. Its combination of training stability, computational efficiency, and reliable performance on smaller datasets makes it especially well suited for synthesizing medical images such as MRIs. This is crucial for tasks demanding the faithful reproduction of pathological features and robust, feasible computation.

Table 2. The generator is designed to transform a 128-dimensional latent noise vector into a 128x128 pixel image

Layer	Layer Type	Filter, Kernel, Stride	Activation Function
1	Dense	8192 units	-
2	Reshape	-	-
3	Conv2D-Transpose	512, (5,5), (2,2)	LeakyReLU (α=0.2)
4	Conv2D-Transpose	256, (5,5), (2,2)	LeakyReLU (α=0.2)
5	Conv2D-Transpose	128, (5,5), (2,2)	LeakyReLU (α=0.2)
6	Conv2D-Transpose	64, (5,5), (2,2)	LeakyReLU (α=0.2)
7	Conv2D-Transpose	32, (5,5), (2,2)	LeakyReLU (α=0.2)
8	Conv2D (output)	3, (5,5), (1,1)	tanh

Table 3. Discriminator (critic) takes a 128x128 pixel image and outputs a single scalar value representing its perceived "realness"

Layer	Layer Type	Filter, Kernel, Stride	Activation Function
1	Conv2D	32, (5,5), (2,2)	LeakyReLU (α=0.2)
2	Conv2D	64, (5,5), (2,2)	LeakyReLU (α=0.2)
3	Conv2D	128, (5,5), (2,2)	LeakyReLU (α=0.2)
4	Conv2D	256, (5,5), (2,2)	LeakyReLU (α=0.2)
5	Conv2D	512, (5,5), (2,2)	LeakyReLU (α=0.2)
6	Flatten	64, (5,5), (2,2)	-
7	Dropout	0.5	-
8	Dense (output)	3, (5,5), (1,1)	Linear

Table 4. Summary of key hyperparameters and configuration settings for model training

Parameter	Value
Optimizer	Adam
Learning Rate	0.0002
Beta_1	0.5
Beta_2	0.999
Noise Dimension	128

Corresponding author: Muhammad Itqan Mazdadi, mazdadi@ulm.ac.id, Department of Computer Science, Lambung Mangkurat University, Jl. A. Yani Km 36, Banjarbaru 70714, Indonesia. **DOI:** https://doi.org/10.35882/ijeeemi.v7i3.109
Copyright © 2025 by the authors. Published by Jurusan Teknik Elektromedik, Politeknik Kesehatan Kemenkes Surabaya Indonesia. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

Batch Size	128
GP Weight	15
Discriminator Step	5
Loss Function	Wasserstein Loss

The generator, as shown in Table 2, converts a 128-dimensional latent noise vector into a 128x128 pixel image using several Conv2DTranspose layers and LeakyReLU activations. The discriminator, detailed in Table 3, outputs a scalar value representing the "realness" of the image after processing through Conv2D layers and a final linear activation. Table 4 summarizes the model's training hyperparameters, including an Adam optimizer with a learning rate of 0.0002 and a batch size of 128. The loss function used is Wasserstein loss, with additional settings for training stability and performance

C. Classification

This study utilizes a Convolutional Neural Network (CNN) architecture, which has the ability to process structured grid data including MRI scans. Its hierarchical design captures multi-scale spatial features, while convolutional filters enable the extraction of richer, more abstract representations than conventional classification models [14]. Convolutional Neural Networks offer a diverse range of architectural designs. This study utilizes four of these configurations:

1. AlexNet

AlexNet was introduced by Krizhevesky in 2012 [31]. The AlexNet model is structured with eight layers, comprising five convolutional layers and three fully connected layers, which are succeeded by an output layer. A key distinction of AlexNet is in its implementation of the ReLu activation function, a departure from the sigmoid function, which enabled faster training and better performance, and also reduced the impact of the vanishing gradient problem [32].

2. ResNet-50

As a leading architecture in image classification, ResNet [33] models demonstrate exceptional performance across visual recognition tasks. The ResNet-50 variant incorporates 50 computational layers structured as: an initial convolutional layer, followed by maxpooling, and 16 residual blocks. Each residual block contains one 3x3 convolutional layer, two 1x1 convolutional layers, and identity-mapping skip connections between input and output tensors. The network's 32x spatial downsampling ratio reduces feature map dimensions to 1/32nd of the original input resolution before reaching the final fully connected layer. For MRI-based diagnostics, ResNet-50 models serve dual purposes: they offer clinically viable vertebral fracture and malignancy assessment and excel in multi-class recognition tasks [34].

3. VGG-16

VGG-16 integrates convolutional, pooling, and fully connected layers. The architecture comprises 16 convolutional layers, of which 13 are used for feature extraction and 3 for filtering, and each layer has a ReLU layer [10]. The VGG16 architecture demonstrates exceptional performance in medical image classification tasks owing to its capacity to capture fine-grained visual patterns and textural nuances [35]. Characterized by progressive depth expansion, VGG-16 processes standardized 224x224 pixel inputs, and a filter size of 3x3. Classification outputs are generated through a Softmax-activated final layer that computes class probability distributions [36].

4. VGG-19

The VGG-19 model comprises 19 layers in total, specifically featuring 16 convolutional layers and three fully connected layers. It is pretrained on more than one million images from the ImageNet database. The VGG-19 has 224x224 pixels for image input. The architecture utilizes a 3x3 pixel filter size, a stride of two, and incorporates maximum pooling layers [37]. In its entirety, the VGG-19 model encompasses roughly 138 million computational parameters [38].

Table 5. Summary of the training hyperparameters applied to the four CNN models (AlexNet, ResNet50, VGG16, and VGG19)

Parameter	Value
Optimizer	SGD (AlexNet & ResNet50) Adam(VGG16 & VGG19)
Learning Rate	0.001
Epoch	50
Batch Size	128
Loss Function	Categorical Crossentropy

To ensure a fair and direct comparison, the training process for all four CNN classifiers was standardized using a consistent set of core hyperparameters, as summarized in Table 5. All models were trained for a maximum of 50 epochs, using a batch size of 128, a learning rate of 0.001, and the Categorical Crossentropy loss function. To prevent overfitting and ensure optimal training duration, an Early Stopping callback was implemented; this monitored the validation loss and halted training if no improvement was observed for 10 consecutive epochs, restoring the weights from the best-performing epoch.

For the VGG16, VGG19, and ResNet50 models, a transfer learning approach was adopted, utilizing weights pre-trained on ImageNet. This strategy involved training only a new classifier head, where a Dropout layer with a rate of 0.5 was applied for regularization. The VGG16 and VGG19 models were specifically optimized using the Adam optimizer, while ResNet50 used SGD with a momentum 0.95. In contrast, the AlexNet architecture was trained from scratch. Consequently, a more extensive regularization strategy was employed, incorporating Batch Normalization layers after its convolutional layers and Dropout layers (0.5) within its dense layers. This model was optimized using SGD with a momentum of 0.9. a uniform data protocol was applied across all experiments: class labels were converted into a one-hot categorical format to ensure compatibility with the categorical crossentropy loss function.

Corresponding author: Muhammad Itqan Mazdadi, mazdadi@ulm.ac.id, Department of Computer Science, Lambung Mangkurat University, Jl. A. Yani Km 36, Banjarbaru 70714, Indonesia. **DOI:** https://doi.org/10.35882/ijeeemi.v7i3.109
Copyright © 2025 by the authors. Published by Jurusan Teknik Elektromedik, Politeknik Kesehatan Kemenkes Surabaya Indonesia. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

D. Classification

A confusion matrix is a metric used to evaluate classification performance by quantifying prediction outcomes into four categories: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). True Positives (TP) represent instances where positive cases are correctly identified, while True Negatives (TN) indicate accurate identification of negative cases. Conversely, False Positives (FP) occur when negative cases are incorrectly classified as positive, and False Negatives (FN) arise when positive cases are mistakenly predicted as negative. Fig. 3. presents the confusion matrix used in this study. Utilizing the constituent elements of the confusion matrix allows for the measurement of key classifier performance indicators

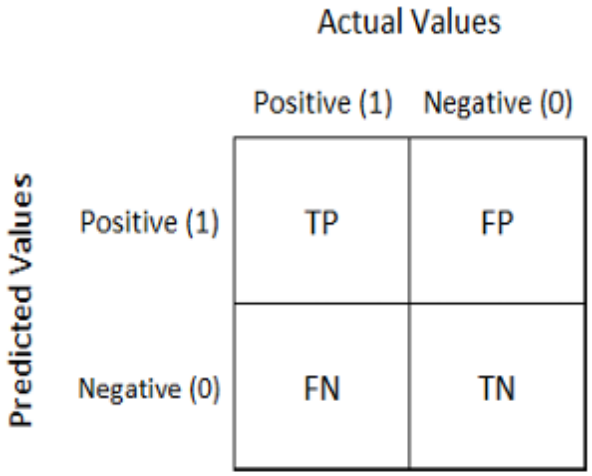

Fig. 3. Confusion matrix table to evaluate classifier performance to calculate Accuracy, Precision, Recall and F1-Score

such as accuracy, precision, recall, and F1-score [39].

Accuracy, as defined in Eq. 1., represents the count of samples correctly classified across all classes.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+FN} \quad (1)$$

Precision in Eq. 2. corresponds to the proportion of samples correctly predicted as positive (true positives) out of all samples the classifier assigned to the positive class.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Sensitivity in Eq. 3. measures the percentage of actual positives correctly identified for a given class among all samples originally from that class.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

As specified in Eq. 4., the F1-Score quantifies overall classifier effectiveness by combining precision and recall into a single harmonic mean.

$$F1 - Score = 2 \times \left(\frac{Precision\ x\ Recall}{Precision+Recall}\right) \quad (4)$$

Based on the preceding materials and method explanation, Fig. 4. shows step by step from the dataset, data augmentation, classification and evaluation. All images in the dataset were already in consistent pixel dimensions. The training phase (labeled “Imbalance Data Training” in the flowchart), involved implementing CNN architectures under two conditions, with and without WGAN-GP. Both configurations were evaluated using a confusion matrix.

To quantitatively evaluate the performance of the images generated by the WGAN-GP model, the Fréchet Inception Distance (FID) was employed. This metric assesses the quality of synthetic images by measuring the similarity between the feature distributions of the generated and real datasets. Using a pre-trained Inception v3 model, the FID score effectively captures both the fidelity and diversity of the generated samples, where a lower score indicates a higher quality result [40].

Given the stochastic nature of neural network training influenced by factors such as random weight initialization and data shuffling, it is essential to validate that such gains are statistically meaningful and not merely the result of random variation. To ensure the robustness of the findings, each experimental configuration was executed indepentently 10 times, providing a sample results suitable for formal statistical analysis.

The inclusion of p-values is necessary to formally test the null hypothesis and quantify the probability that the observed improvements could have occurred by chance, In statistical hypothesis testing, a result is considered statistically significant when the obtained P-value is lower than a predetermined significance level (commonly α = 0.05). A P-value < 0.05 indicates that the probability of obtaining the observed data (or more extreme results) is very low if the null hypothesis (H_0) were true. Therefore, there is sufficient evidence to reject the null hypothesis and conclude that there is a real effect or difference. P values are better written as inequalities, such as P < 0.01 when P = 0.009 [41]. However, to provide a more complete picture beyond a simple statement of significance, 95% confidence intervals (CI) are also presented [42]. The CI offers a plausible range for the true magnitude of the performance increase, thereby providing insight into the practical importance of the augmentation method. This dual approach allows for a more rigorous and transparent assessment of the WGAN-GP's effectiveness.

E. Computational Environment

All computational experiments were conducted on the Kaggle platform, utilizing an NVIDIA P100 GPU for all tasks, including the training of the WGAN-GP model and the four CNN architectures. The implementation was developed in Python (version 3.11.13), utilizing the TensorFlow framework (version 2.18) along with its integrated Keras API for model development.

Corresponding author: Muhammad Itqan Mazdadi, mazdadi@ulm.ac.id, Department of Computer Science, Lambung Mangkurat University, Jl. A. Yani Km 36, Banjarbaru 70714, Indonesia. **DOI:** https://doi.org/10.35882/ijeeemi.v7i3.109
Copyright © 2025 by the authors. Published by Jurusan Teknik Elektromedik, Politeknik Kesehatan Kemenkes Surabaya Indonesia. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

3. RESULTS

This section presents the empirical findings from the implementation of WGAN-GP for data augmentation and the evaluation of various CNN architectures for Alzheimer's Disease classification. This study utilized MRI scans initially characterized by a significant class imbalance across the four diagnostic categories: Mild Demented, Moderate Demented, Non Demented and Very Mild Demented. The original distribution is already shown in . Following the application of WGAN-GP, data augmentation was performed to generate synthetic MRI scans to balance the training data.

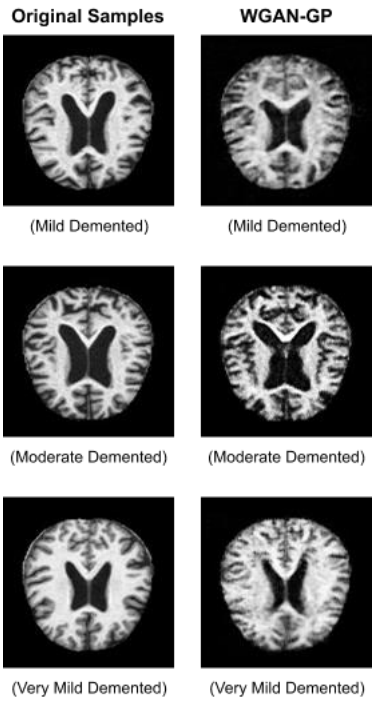

Fig. 5. Representative examples demonstrating WGAN-GP's capability to create realistic synthetic MRI scans (right column) compared to actual patient MRIs (left column)

Table 6. Class distribution after applying WGAN-GP, each training class achieved a balanced 2560 image count.

Class	Training	Testing	Total	Percentage
Mild	2,560	180	2,740	23.78%
Moderate	2,560	13	2,573	22.33%
Non	2,560	640	3,200	27.77%
Very Mild	2,560	448	3,008	26.11%

Table 6 shows the class distribution after WGAN-GP implementation. A balanced class distribution within the training dataset was achieved by supplementing the Mild Demented class with 1,844 images, the Moderate Demented class with 2,509 images, and the Very Mild Demented class with 768 images.

WGAN-GP generates these synthetic images, with samples displayed in Fig. 5. To quantitatively evaluate the quality of the synthetic images generated by the WGAN-GP, model checkpoints were saved and the Fréchet Inception Distance (FID) score was calculated at several key epochs. A lower FID score signifies a higher similarity between the distributions of the real and generated images, indicating superior quality and diversity.

The decision to select 1,000 epochs as the optimal

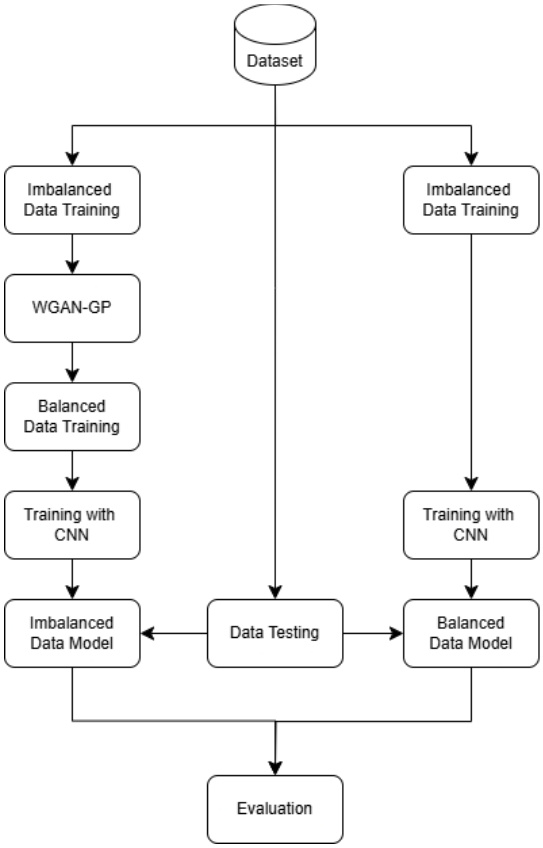

Fig. 4. Based on the preceding Materials and Methods description, these flowcharts illustrate the research methodology structured around the proposed approach.

training duration was based on a comparative analysis of these checkpoint results, as presented in Table 7. At the 1,000 epoch checkpoint, the FID scores reached their lowest values. Since further training to 1,500 epochs did not yield additional improvements and instead showed a slight degradation in scores, this model was determined to provide the best balance between image quality and training efficiency.

Table 7. Comparison of Fréchet Inception Distance (FID) scores at various epoch checkpoints for each class

Class	Epoch 100	Epoch 500	Epoch 1000	Epoch 1500
Mild	123.40	47.50	**34.94**	38.43
Moderate	369.70	204.45	**89.77**	94.65
Very Mild	63.03	32.96	**31.84**	32.84

Corresponding author: Muhammad Itqan Mazdadi, mazdadi@ulm.ac.id, Department of Computer Science, Lambung Mangkurat University, Jl. A. Yani Km 36, Banjarbaru 70714, Indonesia. **DOI:** https://doi.org/10.35882/ijeeemi.v7i3.109
Copyright © 2025 by the authors. Published by Jurusan Teknik Elektromedik, Politeknik Kesehatan Kemenkes Surabaya Indonesia. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

Table 8. Results of the comparative statistical analysis summarizing the performance of CNN architectures with and without WGAN-GP data augmentation. The statistics shown were derived from 10 independent runs for each model configuration.

Model	Metric	Without WGAN-GP (Mean ± Stdev)	With WGAN-GP (Mean ± Stdev)	95% CI	p-value (inequality)
ResNet50	Accuracy	94.43% (± 0.61)	98.17% (± 0.33)	+3.26% to +4.25%	p < 0.001
	Precision	94.41% (± 0.61)	98.18% (± 0.31)	+3.26% to +4.25%	p < 0.001
	Recall	94.41% (± 0.64)	98.17% (± 0.33)	+3.26% to +4.24%	p < 0.001
	F-1 Score	94.36% (± 0.61)	98.17% (± 0.32)	+3.31% to +4.30%	p < 0.001
	AUC	92.36% (± 0.55)	97.73% (± 0.69)	+4.54% to +5.99%	p < 0.001
AlexNet	Accuracy	93.78% (± 3.19)	98.54% (± 0.26)	+2.45% to +6.98%	p < 0.01
	Precision	95.10% (± 2.42)	98.56% (± 0.31)	+1.56% to +5.37%	p < 0.01
	Recall	93.78% (± 3.19)	98.54% (± 0.26)	+2.45% to +6.98%	p < 0.01
	F-1 Score	93.78% (± 3.21)	98.54% (± 0.26)	+2.45% to +7.06%	p < 0.01
	AUC	94.03% (± 3.13)	98.88% (± 0.53)	+2.86% to +6.85%	p < 0.01
VGG16	Accuracy	91.49% (± 0.47)	94.72% (± 0.44)	+2.76% to +3.70%	p < 0.001
	Precision	91.54% (± 0.44)	94.74% (± 0.44)	+2.74% to +3.66%	p < 0.001
	Recall	91.49% (± 0.47)	94.72% (± 0.44)	+2.76% to +3.70%	p < 0.001
	F-1 Score	91.50% (± 0.47)	94.72% (± 0.44)	+2.75% to +3.69%	p < 0.001
	AUC	93.48% (± 0.72)	95.64% (± 0.65)	+1.42% to +3.04%	p < 0.001
VGG19	Accuracy	90.46% (± 0.46)	95.43% (± 0.41)	+4.74% to +5.58%	p < 0.001
	Precision	90.25% (± 0.48)	95.46% (± 0.43)	+4.56% to +5.39%	p < 0.001
	Recall	90.46% (± 0.46)	95.43% (± 0.41)	+4.74% to +5.58%	p < 0.001
	F-1 Score	90.33% (± 0.46)	95.43% (± 0.41)	+4.69% to +5.49%	p < 0.001
	AUC	93.52% (± 0.58)	96.32% (± 0.65)	+2.24% to +3.32%	p < 0.001

After balancing the training dataset, eight experimental configurations were applied: four CNN architectures, each tested with and without WGAN-GP augmentation; each configuration was executed 10 independent times. The evaluation of four CNN architectures ResNet-50, AlexNet, VGG-16, and VGG-19 demonstrated a consistent improvement in classification performance when trained with the WGAN-GP augmented dataset compared to the original imbalanced dataset.

Table 8. shows that the application of Wasserstein GAN with Gradient Penalty (WGAN-GP) for data augmentation resulted in statistically significant performance imptovements across all tested Convolutional Neural Network (CNN) architectures. The evaluation, spanning metrics of Accuracy, Precision, Recall, F-1 Score, and AUC, consistently demonstrated that augmenting the training dataset with synthetic images yielded superior outcomes compared to training on the original data alone. The statistical significance of these improvements ($p < 0.05$ for all comparisons) underscores the efficacy and robustness of WGAN-GP as a technique to overcome data scarcity and improve the generalization capabilities of the models in this context. While all architectures benefited from the augmentation, a comparative analysis identifies the AlexNet model paired with WGAN-GP as the definitive top performer. This configuration achieved the highest mean scores across every evaluation metric: Accuracy (98.54%), Precision (98.56%), Recall (98.54%), F-1 Score (98.54%), and AUC (98.88%). This superior performance suggests a strong synergy between the feature extraction capabilities of the AlexNet architecture and the nature of the synthetic data produced by the WGAN-GP for this specific classification task.

Table 9. Aggregated confusion matrix from 10 independent runs of the AlexNet + WGAN-GP model on the test set

True Label	Predicted Mild	Predicted Moderate	Predicted Non	Predicted VeryMild
Mild	**1760**	0	13	27
Moderate	0	**128**	1	1
Non	13	0	**6317**	70
VeryMild	10	0	52	**4418**

Beyond elevating average performance, WGAN-GP augmentation also markedly improved model stability, most notably in the AlexNet architecture. The standard deviation of AlexNet's accuracy was drastically reduced from ±3.19 in the baseline model to a highly consistent ±0.26 with augmentation. This demonstrates that the technique not only boosts predictive accuracy but also enhances the reliability and reproducibility of the model's performance across independent runs, a critical attribute for deployment in real-world applications.

Corresponding author: Muhammad Itqan Mazdadi, mazdadi@ulm.ac.id, Department of Computer Science, Lambung Mangkurat University, Jl. A. Yani Km 36, Banjarbaru 70714, Indonesia. **DOI:** https://doi.org/10.35882/ijeeemi.v7i3.109
Copyright © 2025 by the authors. Published by Jurusan Teknik Elektromedik, Politeknik Kesehatan Kemenkes Surabaya Indonesia. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

Table 10. Performance metrics of the AlexNet + WGAN-GP model, aggregated from 10 independent runs

AlexNet+ WGAN-GP	Precision	Recall	F1-Score
Mild	98.71%	97.78%	98.24%
Moderate	100%	98.46%	99.22%
Non	98.97%	98.70%	98.83%
VeryMild	97.83%	98.62%	98.22%
Accuracy		98,54%	
Macro Avg	98.88%	98.39%	98.63%
Weighted Avg	98.54%	98.54%	98.54%

The aggregated results from 10 runs show that the AlexNet + WGAN-GP model achieves very high and stable performance, with an overall accuracy of 98.54% as summarized in Table 10. This strong performance is also reflected in the Macro Avg F1-Score of 98.63%, which indicates a balanced classification capability across all classes, including the minority 'Moderate' class, which achieved perfect precision. Table 9. provides further detail through the aggregated confusion matrix shows that the vast majority of predictions fall on the main diagonal. The few errors that occurred are concentrated in classes that are clinically difficult to distinguish, such as the confusion between the 'Non' and 'Very Mild' classes. This combination of strong quantitative metrics with an explainable error pattern confirms the reliability and robustness of the proposed model.

4. DISCUSSION

The findings of this study establish the Wasserstein GAN with Gradient Penalty (WGAN-GP) not merely as a data augmentation tool, but as a powerful feature space refiner for Alzheimer's disease classification. By learning the underlying data manifold of brain imaging, WGAN-GP generates novel, high-fidelity synthetic images that represent plausible variations of neurodegenerative pathology, rather than simple interpolations. This process of creating authentic, on-manifold data proved transformative, elevating performance across all tested architectures. More significantly, this study unveiled a critical insight: diagnostic accuracy is ultimately determined by the synergy between model architecture and data quality. This was most evident in the AlexNet architecture, which, when guided by this enriched synthetic data, achieved a near-perfect classification benchmark, establishing that optimal model capacity paired with superior data can surpass even deeper, more complex networks.

In the context of prior research, as summarized in Table 11., our WGAN-GP approach demonstrates a significant performance edge. Previous studies have achieved commendable results using simpler oversampling techniques; for instance, a notable benchmark was set at 97.90% accuracy using ADASYN [3]. While this highlights the effectiveness of addressing data imbalance, our method achieves comparable, and in the case of our augmented AlexNet, superior results. We attribute this to WGAN-GP's ability to generate more diverse and authentic data compared to the interpolative nature of SMOTE or ADASYN. Furthermore, our results surpass those of architectures using standard augmentation, such as the Lightweight CNN which reached 95.93% [5], and even a modified AlexNet without generative augmentation that scored 95.70% [11]. This comparison underscores that the key differentiator was not the base architecture alone, but the transformative quality of the WGAN-GP-generated data. Finally, it is crucial to note that not all generative methods yield similarly strong results. Another study using a standard GAN on the ADNI dataset reported only 68% accuracy [43]. This starkly contrasts with our findings and suggests that the specific choice of the WGAN-GP architecture, known for its training stability and high-fidelity output, was critical to achieving a new state-of-the-art performance in Alzheimer's classification.

These significant performance gains come at a pronounced computational cost, as evidenced by training time increases of 55-118% across architectures, a trade-off most acutely manifested in computationally intensive networks. This efficiency-performance dichotomy highlights WGAN-GP's role as a force multiplier for

Table 11. Benchmarking Results for Various Alzheimer's Disease Classification Models from Prior Studies

Researcher	Dataset	Classification Model	Oversampling Model	Result
[3]	Alzheimer's Disease (Kaggle)	CNN	SMOTE ADASYN	94.17% 97.90%
[5]	Alzheimer's Disease (Kaggle)	CNN	Image Data Generator	Lightweight CNN: 95.93%
[11]	Alzheimer's Disease (Kaggle)	CNN	-	VGG-16: 85.07% ResNet-50: 75.25% Mod AlexNet: 95.70%
[44]	Alzheimer's Disease (Kaggle)	CNN	SMOTE	90.78%
[43]	Alzheimer's Disease (ADNI)	InceptionV3	GAN	68%

Corresponding author: Muhammad Itqan Mazdadi, mazdadi@ulm.ac.id, Department of Computer Science, Lambung Mangkurat University, Jl. A. Yani Km 36, Banjarbaru 70714, Indonesia. **DOI:** https://doi.org/10.35882/ijeeemi.v7i3.109
Copyright © 2025 by the authors. Published by Jurusan Teknik Elektromedik, Politeknik Kesehatan Kemenkes Surabaya Indonesia. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

diagnostic capability rather than an optimization tool, prioritizing clinical accuracy over operational expediency while establishing clear architecture-specific implementation thresholds for real-world deployment. This is a crucial point regarding the real-world applicability of our method. We acknowledge that the training phase involving WGAN-GP is computationally intensive. However, it is essential to distinguish between the one-time training cost and the ongoing deployment (inference) cost. The significant computational overhead is an upfront investment incurred only during the model development stage. This intensive process aims to create a highly robust and accurate classifier by overcoming the limitations of data scarcity and imbalance. The final product intended for a clinical setting is the trained CNN classifier, not the GAN itself.

The clinical utility of this final, highly accurate classifier is profound. Its enhanced accuracy and reliability can directly support earlier diagnosis and reduce the risk of misdiagnosis, which are critical factors in managing neurodegenerative diseases. Furthermore, the model serves as a robust decision support tool for clinicians, providing an objective assessment to complement their expertise. This ultimately contributes to greater diagnostic efficiency and improved patient management pathways.

5. CONCLUSION

The results conclusively demonstrate that data augmentation using a WGAN-GP provides a substantial and statistically significant enhancement to the performance of all evaluated CNN architectures for this classification task. Among the tested configurations, the AlexNet model augmented with WGAN-GP emerged as the top performer, achieving the highest scores across all metrics, including an accuracy of 98.54% (±0.26). This approach not only elevates the models' predictive capabilities but also markedly improves their stability, as evidenced by the significant reduction in standard deviation. Therefore, this study validates WGAN-GP as a highly effective and reliable technique for generating synthetic data to overcome dataset limitations and boost model performance in this domain.

For future work, several promising avenues can be explored. First, investigating other advanced GAN variants may yield synthetic images with even greater realism and diversity, potentially leading to further performance gains. Second, exploring a hybrid approach that combines WGAN-GP augmentation with other techniques such as advanced transfer learning or other forms of data augmentation could uncover synergistic effects. Finally, testing the generalizability of this method on different medical imaging modalities (e.g., PET, CT scans) or larger, multi-centric datasets would be a valuable step to validate its robustness and applicability in broader clinical contexts.

REFERENCES

[1] A. Nawaz, S. M. Anwar, R. Liaqat, J. Iqbal, U. Bagci, and M. Majid, “Deep Convolutional Neural Network based Classification of Alzheimer’s Disease using MRI Data,” in *Proceedings - 2020 23rd IEEE International Multi-Topic Conference, INMIC 2020*, Institute of Electrical and Electronics Engineers Inc., Nov. 2020. doi: 10.1109/INMIC50486.2020.9318172.

[2] N. R. Thota and D. Vasumathi, “WASSERSTEIN GAN-GRADIENT PENALTY WITH DEEP TRANSFER LEARNING FOR 3D MRI ALZHEIMER DISEASE CLASSIFICATION,” 2022. [Online]. Available: https://adni.loni.usc.edu/

[3] E. Öter, A. Hastalığı, S. Için, V. Dengeleme, Y. Karşılaştırmalı, and B. Çalışması, “A Comparative Study on Data Balancing Methods for Alzheimer’s Disease Classification,” 2024.

[4] Q. Li and M. Q. Yang, “Comparison of machine learning approaches for enhancing Alzheimer’s disease classification,” *PeerJ*, vol. 9, Feb. 2021, doi: 10.7717/peerj.10549.

[5] A. A. A. El-Latif, S. A. Chelloug, M. Alabdulhafith, and M. Hammad, “Accurate Detection of Alzheimer’s Disease Using Lightweight Deep Learning Model on MRI Data,” *MDPI Diagnostics*, vol. 13, no. 7, Apr. 2023, doi: 10.3390/diagnostics13071216.

[6] I. J. Goodfellow *et al.*, “Generative Adversarial Nets,” 2014. [Online]. Available: http://www.github.com/goodfeli/adversarial

[7] Y. Zhang *et al.*, “GAN-based one dimensional medical data augmentation,” *Soft comput*, vol. 27, no. 15, pp. 10481–10491, Aug. 2023, doi: 10.1007/s00500-023-08345-z.

[8] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein Generative Adversarial Networks,” 2017.

[9] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, “Improved Training of Wasserstein GANs,” Mar. 2017, [Online]. Available: http://arxiv.org/abs/1704.00028

[10] S. A. Ajagbe, K. A. Amuda, M. A. Oladipupo, O. F. AFE, and K. I. Okesola, “Multi-classification of alzheimer disease on magnetic resonance images (MRI) using deep convolutional neural network (DCNN) approaches,” *International Journal of Advanced Computer Research*, vol. 11, no. 53, pp. 51–60, Mar. 2021, doi: 10.19101/ijacr.2021.1152001.

[11] H. Acharya, R. Mehta, and D. Kumar Singh, “Alzheimer Disease Classification Using Transfer Learning,” in *Proceedings - 5th International Conference on Computing Methodologies and Communication, ICCMC 2021*, Institute of Electrical and Electronics Engineers Inc., Apr. 2021, pp. 1503–1508. doi: 10.1109/ICCMC51019.2021.9418294.

[12] M. Yildirim and A. Cinar, “Classification of Alzheimer’s disease MRI images with CNN based

Corresponding author: Muhammad Itqan Mazdadi, mazdadi@ulm.ac.id, Department of Computer Science, Lambung Mangkurat University, Jl. A. Yani Km 36, Banjarbaru 70714, Indonesia. **DOI:** https://doi.org/10.35882/ijeeemi.v7i3.109
Copyright © 2025 by the authors. Published by Jurusan Teknik Elektromedik, Politeknik Kesehatan Kemenkes Surabaya Indonesia. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

hybrid method," *Ingenierie des Systemes d'Information*, vol. 25, no. 4, pp. 413–418, Aug. 2020, doi: 10.18280/isi.250402.

[13] S. Sharma, K. Guleria, S. Tiwari, and S. Kumar, "A deep learning based convolutional neural network model with VGG16 feature extractor for the detection of Alzheimer Disease using MRI scans," *Measurement: Sensors*, vol. 24, Dec. 2022, doi: 10.1016/j.measen.2022.100506.

[14] T. Morris, Z. Liu, L. Liu, and X. Zhao, "Using a Convolutional Neural Network and Explanaiable AI to Diagnose Dementia Based on MRI Scans," *Alzheimer's and Dementia*, vol. 19, no. 4, pp. 1598–1695, Apr. 2023, doi: 10.1002/alz.13016.

[15] A. Mumuni and F. Mumuni, "Data augmentation: A comprehensive survey of modern approaches," Dec. 01, 2022, *Elsevier B.V.* doi: 10.1016/j.array.2022.100258.

[16] H. Zhang, R. Wang, R. Pan, and H. Pan, "Imbalanced fault diagnosis of rolling bearing using enhanced generative adversarial networks," *IEEE Access*, vol. 8, pp. 185950–185963, 2020, doi: 10.1109/ACCESS.2020.3030058.

[17] S. Al-Otaibi, M. Mujahid, A. R. Khan, H. Nobanee, J. Alyami, and T. Saba, "Dual Attention Convolutional AutoEncoder for Diagnosis of Alzheimer's Disorder in Patients Using Neuroimaging and MRI Features," *IEEE Access*, vol. 12, pp. 58722–58739, 2024, doi: 10.1109/ACCESS.2024.3390186.

[18] A. Bruno, V. Dos Santos, and D. R. Carvalho, "Generating synthetic 2019-nCoV samples with WGAN to increase the precision of an Ensemble Classifier," *Iberoamerican Journal of Applied Computing*, pp. 1–13, 2020.

[19] P. Gayathri *et al.*, "Deep Learning Augmented with SMOTE for Timely Alzheimer's Disease Detection in MRI Images," 2024. [Online]. Available: www.ijacsa.thesai.org

[20] A. Yeafi, M. Islam, S. K. Mondal, K. M. I. H. Nashad, and M. S. U. Yusuf, "A Semi-supervised Approach For Brain Tumor Classification Using Wasserstein Generative Adversarial Network with Gradient Penalty," in *2023 6th International Conference on Electrical Information and Communication Technology, EICT 2023*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/EICT61409.2023.10427898.

[21] L. Bi and G. Hu, "Improving Image-Based Plant Disease Classification With Generative Adversarial Network Under Limited Training Set," *Front Plant Sci*, vol. 11, Dec. 2020, doi: 10.3389/fpls.2020.583438.

[22] L. Abou-Abbas, K. Henni, I. Jemal, and N. Mezghani, "Generative AI with WGAN-GP for boosting seizure detection accuracy," *Front Artif Intell*, vol. 7, 2024, doi: 10.3389/frai.2024.1437315.

[23] I. Goodfellow *et al.*, "Generative adversarial networks," *Commun ACM*, vol. 63, no. 11, pp. 139–144, Oct. 2020, doi: 10.1145/3422622.

[24] L. Goetschalckx, A. Andonian, and J. Wagemans, "Generative adversarial networks unlock new methods for cognitive science," Sep. 01, 2021, *Elsevier Ltd*. doi: 10.1016/j.tics.2021.06.006.

[25] H. Anaya-Sánchez, L. Altamirano-Robles, R. Díaz-Hernández, and S. Zapotecas-Martínez, "WGAN-GP for Synthetic Retinal Image Generation: Enhancing Sensor-Based Medical Imaging for Classification Models," *Sensors*, vol. 25, no. 1, Jan. 2025, doi: 10.3390/s25010167.

[26] T. Karras, S. Laine, and T. Aila, "A Style-Based Generator Architecture for Generative Adversarial Networks," *IEEE Trans Pattern Anal Mach Intell*, vol. 43, no. 12, pp. 4217–4228, 2021, doi: 10.1109/TPAMI.2020.2970919.

[27] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and Improving the Image Quality of StyleGAN," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 8107–8116. doi: 10.1109/CVPR42600.2020.00813.

[28] A. Brock, J. Donahue, and K. Simonyan, "Large Scale GAN Training for High Fidelity Natural Image Synthesis," Sep. 2018, [Online]. Available: http://arxiv.org/abs/1809.11096

[29] D. K. Park, S. Yoo, H. Bahng, J. Choo, and N. Park, "MEGAN: Mixture of Experts of Generative Adversarial Networks for Multimodal Image Generation," May 2018, [Online]. Available: http://arxiv.org/abs/1805.02481

[30] W. Wan and H. J. Lee, "Generative Adversarial Multi-Task Learning for Face Sketch Synthesis and Recognition," in *2019 IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 4065–4069. doi: 10.1109/ICIP.2019.8803617.

[31] M. Z. Alom *et al.*, "The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches," Sep. 2018, doi: https://doi.org/10.48550/arXiv.1803.01164.

[32] S. Joshi, R. Shah, Y. Chandola, and V. Uniyal, "Classification of Brain MRI Images using End-to-End Trained AlexNet & End-to-End Pre-Trained MobileNet," 2024. [Online]. Available: www.ijrpr.com

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, Dec. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

Corresponding author: Muhammad Itqan Mazdadi, mazdadi@ulm.ac.id, Department of Computer Science, Lambung Mangkurat University, Jl. A. Yani Km 36, Banjarbaru 70714, Indonesia. **DOI:** https://doi.org/10.35882/ijeeemi.v7i3.109
Copyright © 2025 by the authors. Published by Jurusan Teknik Elektromedik, Politeknik Kesehatan Kemenkes Surabaya Indonesia. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

[34] K. Liu *et al.*, "Prediction of Primary Tumor Sites in Spinal Metastases Using a ResNet-50 Convolutional Neural Network Based on MRI," *Cancers (Basel)*, vol. 15, no. 11, Jun. 2023, doi: 10.3390/cancers15112974.

[35] A. Ritahani Ismail, Syed Qamrun Nisa, S. A. Shaharuddin, S. I. Masni, and S. A. Suharudin Amin, "Utilising VGG-16 of Convolutional Neural Network for Medical Image Classification," *International Journal on Perceptive and Cognitive Computing*, vol. 10, no. 1, pp. 113–118, Jan. 2024, doi: 10.31436/ijpcc.v10i1.460.

[36] M. Toğaçar, B. Ergen, Z. Cömert, and F. Özyurt, "A Deep Feature Learning Model for Pneumonia Detection Applying a Combination of mRMR Feature Selection and Machine Learning Models," *IRBM*, vol. 41, no. 4, pp. 212–222, Aug. 2020, doi: 10.1016/j.irbm.2019.10.006.

[37] A. ŞENER and B. ERGEN, "Enhancing Brain Tumor Detection on MRI Images Using an Innovative VGG-19 Model-Based Approach," *Sakarya University Journal of Science*, vol. 27, no. 5, pp. 1128–1140, Oct. 2023, doi: 10.16984/saufenbilder.1302803.

[38] M. Mateen, J. Wen, Nasrullah, S. Song, and Z. Huang, "Fundus image classification using VGG-19 architecture with PCA and SVD," *Symmetry (Basel)*, vol. 11, no. 1, Jan. 2019, doi: 10.3390/sym11010001.

[39] F. M. J. M. Shamrat *et al.*, "AlzheimerNet: An Effective Deep Learning Based Proposition for Alzheimer's Disease Stages Classification From Functional Brain Changes in Magnetic Resonance Images," *IEEE Access*, vol. 11, pp. 16376–16395, 2023, doi: 10.1109/ACCESS.2023.3244952.

[40] R. Zhao, "Deep Generative Models in Brain MRI Synthesis for Alzheimer's Disease Research," 2024.

[41] C. Andrade, "The P Value and Statistical Significance: Misunderstandings, Explanations, Challenges, and Alternatives Website: Quick Response Code," 2019.

[42] A. Hazra, "Using the confidence interval confidently," *J Thorac Dis*, vol. 9, no. 10, pp. 4125–4130, Oct. 2017, doi: 10.21037/jtd.2017.09.14.

[43] H. Tufail, A. Ahad, I. Puspitasari, I. Shayea, P. J. Coelho, and I. M. Pires, "Deep Learning in Smart Healthcare: A GAN-based Approach for Imbalanced Alzheimer's Disease Classification," *Procedia Comput Sci*, vol. 241, pp. 146–153, 2024, doi: 10.1016/j.procs.2024.08.021.

[44] Z. Li, Y. Wang, Z. Jiang, Z. Luo, J. Wu, and T. T. Toe, "Ensemble of CNN Models for Identifying Stages of Alzheimer's Disease: An Approach Using MRI Scans and SMOTE Algorithm," in *2023 3rd International Symposium on Computer Technology and Information Science, ISCTIS 2023*, Institute of Electrical and Electronics Engineers Inc., 2023, pp. 496–500. doi: 10.1109/ISCTIS58954.2023.10213182.

AUTHOR BIOGRAPHY

Muhammad Faiq Alamudin is a computer science student at the Faculty of Mathematics and Natural Sciences, Universitas Lambung Mangkurat, having commenced his studies in 2020. His research interests lie in the field of Alzheimer's Disease Classification, with a particular focus on data augmentation. His final project will entail a comparative analysis between with and without a data augmentation in the classification of alzheimer's disease.

Muhammad Itqan Mazdadi He is a lecturer in the Department of Computer Science at Lambung Mangkurat University, with research interests primarily in Data Science and Computer Networking. His academic journey began at Lambung Mangkurat University, where he earned his undergraduate degree in Computer Science in 2013. To further deepen his expertise, he pursued and successfully completed a master's degree in Informatics from the Islamic University of Indonesia, Yogyakarta. Currently, he holds the position of Secretary in the Computer Science Department at Lambung Mangkurat University, contributing to both academic and administrative responsibilities. His research focuses on exploring advancements in data-driven technologies and network systems. Through his role as a lecturer and researcher, he actively engages in academic development, fostering innovation in his field. His dedication to education and research enables him to support students and collaborate on projects that enhance technological progress. Email: mazdadi@ulm.ac.id Orcid ID: 0000-0002-8710-4616.

Radityo Adi Nugroho received his bachelor's degree in Informatics from the Islamic University of Indonesia and a master's degree in Computer Science from Gadjah Mada University. Currently, he is an assistant professor in the Department of Computer Science at Lambung Mangkurat University. His research interests include software defect prediction and computer vision. He can be contacted at email: radityo.adi@ulm.ac.id. Orcid ID: 0000-0002-7326-7668.

Corresponding author: Muhammad Itqan Mazdadi, mazdadi@ulm.ac.id, Department of Computer Science, Lambung Mangkurat University, Jl. A. Yani Km 36, Banjarbaru 70714, Indonesia. **DOI:** https://doi.org/10.35882/ijeeemi.v7i3.109
Copyright © 2025 by the authors. Published by Jurusan Teknik Elektromedik, Politeknik Kesehatan Kemenkes Surabaya Indonesia. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).

Triando Hamonangan Saragih is currently a lecturer at the Department of Computer Science, Lambung Mangkurat University. He has a deep involvement in the academic world with a main focus on various aspects of Data Science. His academic journey began with completing his undergraduate studies in Informatics at Brawijaya University, Malang, which he successfully completed in 2016. To deepen her understanding in the field of Computer Science, she then continued her education to the master's level at the same university and successfully earned a master's degree in 2018. Her research interests center on Data Science, where she continues to develop her expertise in data analysis, predictive modeling, and the application of data-driven technologies. As an academic, he is active in scientific research and collaboration to develop innovative solutions that can be applied in various fields. He also seeks to contribute to the development of science through publications and academic discussions. Email: triando.saragih@ulm.ac.id Orcid ID: 0000-0003-4346-3323

Muliadi is a lecturer in the Department of Computer Science at Lambung Mangkurat University, specializing in Artificial Intelligence, Decision Support Systems, and Data Science. His academic journey commenced with earning a bachelor's degree in Informatics Engineering from STMIK Akakomin in 2004. To further enhance his knowledge, he pursued and successfully obtained a master's degree in Computer Science from Gadjah Mada University in 2009. With a strong foundation in Data Science, he possesses extensive expertise in analyzing and interpreting complex datasets. Additionally, he has valuable skills in Start-up Business Development, Digital Entrepreneurship, and Data Management. His experience allows him to contribute significantly to both academic and practical applications of technology. Beyond his role as a lecturer, he actively engages in research and collaborative projects aimed at advancing technological innovation. Through his expertise, he strives to bridge the gap between academia and industry, fostering solutions that drive digital transformation and business growth. Email: muliadi@ulm.ac.id Orcid ID: 0000-0003-2871-9482.

Professor Dr **Vijay Anant Athavale** was awarded PhD in Computer Science 2003 from Barkatullah University, Bhopal, India. He has a rich teaching, research and administrative experience of more than 34 years. He has served in Government and Private Universities and Colleges in India and abroad. He has served academic assignments at Canada, the Republic of Yemen and the Republic of Sudan. He has served as an expatriate expert for the United Nations Development Programme to Ethiopia. He has been instrumental in shaping of several Institutions of repute offering Technical and Management Education in India. He spearheaded many innovative initiatives including the industry academic integration and internationalization. As a thought leader and a subject-matter expert, he has presented numerous sessions at Conferences, FDPs, and Workshops and his work has been cited in over 690 publications

Corresponding author: Muhammad Itqan Mazdadi, mazdadi@ulm.ac.id, Department of Computer Science, Lambung Mangkurat University, Jl. A. Yani Km 36, Banjarbaru 70714, Indonesia. **DOI:** https://doi.org/10.35882/ijeeemi.v7i3.109
Copyright © 2025 by the authors. Published by Jurusan Teknik Elektromedik, Politeknik Kesehatan Kemenkes Surabaya Indonesia. This work is an open-access article and licensed under a Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).