Generative Fabrication of Medical Images for Machine Learning Training

Andres G. Calzada-Jasso[1], Andrei Tchernykh[1,2],
Ixchel D. Avendaño-Pacheco[1]
[1]*CICESE Research Center*
[2]*Institute for System Programming, RAS, Moscow, Russia*
Ensenada, Baja California, Mexico
andres.calzada@cicese.edu.mx, chernykh@cicese.mx,
ixchel@cicese.edu.mx

Jorge M. Cortés-Mendoza[3], Bernardo Pulido-Gaytan[3]
[3]*National College of Ireland*
Dublin, Ireland
jorgemario.cortesmendoza@ncirl.ie,
luisbernardo.pulidogaytan@ncirl.ie

Mikhail Babenko[4]
[4]*North-Caucasus Federal University*
Stavropol, Russia
mgbabenko@ncfu.ru

Alfredo Goldman[5]
[5]*University of São Paulo*
São Paulo, Brazil
gold@ime.usp.br

Horacio González-Vélez[3]
[3]*National College of Ireland*
Dublin, Ireland
horacio@ncirl.ie

***Abstract*—Training in supervised machine learning is based on the availability of datasets; however, medical datasets must comply with stringent privacy regulations. Generative Adversarial Networks (GANs) are a relevant alternative to solve the limitation of small medical datasets due to their ability to generate additional data with desired features. A significant drawback of these models is that they may produce unrealistic, blurred, or insufficiently diverse images. This paper proposes a data augmentation technique using GANs to create synthetic Magnetic Resonance Imaging (MRI) of four stages of Alzheimer's Disease (AD): non-demented, very mild demented, mild demented, and moderate demented. We designed a GAN based on the Pix2Pix model, which learns the features of each AD stage. Generated images are evaluated by multistage Convolutional Neural Network (CNN) models, greyscale histograms of the distribution of pixel intensities, and brain mass measurements on binarized images. The results indicate that AD synthetic MRI effectively captures disease patterns, demonstrating the potential of GANs to improve training and diagnosis of neurodegenerative diseases.**

Keywords—Alzheimer's Disease, Binarization, Convolutional Neural Network, Data Augmentation, Generative Adversarial Networks, Histograms

I. Introduction

Recent technological advances have significantly transformed medical imaging, offering innovative solutions for diagnosis and treatment planning. Magnetic Resonance Imaging (MRI), Computed Tomography (CT), and Positron Emission Tomography (PET) have helped improve diagnostic precision. Segmentation, classification, and personalised medicine address the challenges posed by neurodegenerative diseases, where tools have proven essential for accuracy and efficiency in clinical practice [1].

Generative Adversarial Networks (GANs) have revolutionised medical imaging by enabling the synthesis of realistic medical data. They can provide critical support in situations where real datasets are limited. Some studies demonstrated the ability of GANs to generate high-quality synthetic images in domains such as cardiac MRI, liver CT, and retinal imaging, making them valuable for training Deep Learning (DL) models [2]. Synthetic images may not always capture the full complexity of real datasets, but they remain a powerful tool for data augmentation. They facilitate the development of robust Machine Learning (ML) models for segmentation and disease diagnosis.

Alzheimer's Disease (AD) is a slowly progressive neurodegenerative disease that leads to dementia. It is characterized by neurotic plaques—unexpected aggregations of altered glial cells and swollen cellular processes of nerve cells—and neurofibrillary tangles, resulting from accumulating a protein called amyloid beta peptide (Aβ) in affected brain areas, such as the medial temporal lobe and neocortical structures [3]. The first record of this disease showed a massive loss of neurons in the patient's brain, who suffered from memory loss and personality changes. This condition was described as a serious disease of the cerebral cortex [4].

Worldwide, 47 million people live with dementia, and by 2050, the number is expected to increase to 131 million [3]. It has a significant impact as a global health concern because the condition continues to grow, with millions of new cases diagnosed annually. AD is the most widespread of all types of dementia, comprising roughly 70% of all diagnoses. Its early identification and treatment are crucial to managing the disease with pharmacological treatments and adequate care [5].

Computational tools have emerged as significant support for the medical field in the diagnosis of AD. In recent years, the use of DL algorithms in various areas of medicine has grown significantly [6]. These tools can be effectively applied in clinical practice, e.g., in cancer diagnosis [7], cardiac rhythm analysis and heart failure prediction [8], as well as in radiography and scans to detect strokes, fractures, and malignant lesions, among others.

DL faces significant issues related to privacy and data sharing of patient information [9], [10]. One of the main challenges in DL is the large amount of data required to train these models. A possible solution to increase data and preserve information privacy is the use of GANs [11], these models have emerged as an important technique in image processing to generate artificial data [12].

GAN models have two main components: A Generator (G) produces synthetic images, and a Discriminator (D) distinguishes between real and synthetic images. This approach can produce indistinguishable images between the generated and real images. GANs are efficient methods for medical diagnosis, cost reduction, and improving accuracy. The solid theoretical foundations of GANs enable their application across a wide range of domains, including computer vision, image segmentation [13], time series synthesis, image editing [14], text-to-image translation [15], and natural language processing [16]. A probability function represents the input dataset, and D, based on a Neural Network (NN) model, evaluates this probability to determine if a sample belongs to the population of real instances.

GANs can generate high-quality synthetic data to improve the DL training process in diagnosing diseases such as AD. They are flexible models that can improve the resolution of MRI by improving the quality [17]. These models offer greater consistency and comparison between clinical studies performed by medical specialists [18]. Furthermore, GANs are important tools for medical image segmentation because they help to create more efficient models to identify brain structures affected by AD. More accurate DL models are crucial for developing treatments and monitoring the progression of the disease. The state-of-the-art highlights the use of these models for segmentation to improve accuracy and reduce the time required to process large volumes of medical image data [19].

This paper aims to generate synthetic images similar to real ones that facilitate the training of DL in the context of AD, where datasets have limitations due to the low number of instances, and obtaining images is complicated due to ethical and privacy restrictions [20]. Our main contributions are:

- a GAN model based on Pix2Pix for MRI data augmentation to address the lack of data in the available medical datasets of AD; and
- evaluation of the efficiency of the GAN model using three verification methods: CNN models, histograms, and binarization of the images.

The content of the paper is structured as follows. The next section presents related work in the field. Section III describes the methodology of our research. Section IV summarises the results and main contributions. Finally, Section V presents the conclusions and future work.

II. Related Work

GANs are effective tools for a variety of applications in different areas, e.g., image synthesis, improved image resolution, and synthetic data creation [21]. This section presents relevant information on the domain of GANs and their application in medicine.

A. Generative adversarial network models

GANs are widely used for synthesising realistic images. Although general-purpose GANs are capable of generating diverse and artistic visuals, they often underperform in specialised domains because their training process considers images of many objects. In contrast, task-specific GANs are trained on specific-purpose datasets, enabling them to generate more accurate and relevant outputs.

The first GAN model, introduced by Ian Goodfellow, established the foundation for extensive research into generative modelling [12]. In recent years, numerous variants and enhancements of the original architecture have been proposed, each addressing specific limitations and improving performance across different application areas. The most recent models improve the quality of images generated by the original GAN, some of them perform better for specific tasks, and others have been adapted to work with medical images. GANs introduce a new technique confronting two CNN models, G and D.

Figure 1 shows the GAN general scheme, where G introduces random noise to produce synthetic images resembling real ones, while D evaluates their authenticity. This process is the basis for the creation of images, which allows the number of instances of existing datasets to increase [22].

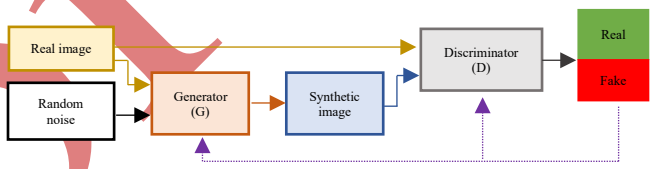

Fig. 1. The basic scheme of a GAN model.

GANs are also used in image synthesis and translation between image modalities. These models allow images of one modality to be estimated from another, facilitating more efficient clinical workflows. For example, pseudo-CT images can be generated from MRI images, which is applied in MRI-only radiation therapy planning, reducing radiation exposure and the costs of CT or PET scans. GANs also enable the generation of PET images from MRI, the creation of high-resolution MRI images from low-field images, and the synthesis of missing MRI contrasts [23].

B. GAN models in medicine

GANs use different approaches to generate more realistic synthetic images. Deep Convolutional GAN (DCGAN) [24] introduces convolutional layers to better control the features generated by realistic images. Its architecture improves the stability and quality of synthetic images. Pix2Pix [25] facilitates accurate image-to-image translation in structural shifting tasks through a U-Net and PatchGAN architecture, which simplifies the creation of fine details. CycleGAN [26] generates translation between domains using a cycle loss function without pairs aligned to ensure consistency, which allows solving a wider range of tasks and domains.

Super-Resolution GAN (SRGAN) [27] improves the resolution of generated images with a residual block network and perceptual loss, increasing the fidelity in image production.

Wasserstein GAN (WGAN) [28] uses the Wasserstein distance that solves stability problems in training and provides a more reliable measurement of the differences between the generated and real data. Finally, Progressive Growing of GAN (Pro-GAN) [29] employs a progressive resolution increase technique that optimises image quality and reduces stability problems.

Table 1 summarises the key characteristics of the most widely used GAN models in the medical domain, highlighting the specific contributions of each model to image generation. These advances have significantly enhanced the generative capabilities of GANs and broadened their applicability in medical imaging.

C. *Application of GAN models in medicine*

The capability of GANs to generate realistic synthetic data from a limited dataset has been helpful in the medical field, where datasets do not have the required number of instances to train ML models, and their combination is restricted due to the protection of private data. GANs provide improvements in the diagnosis and treatment of diseases, such as data augmentation, improved image resolution, and segmentation of anatomical images.

Table 2 summarises the main characteristics of various medical applications of GAN, focusing on improving AD diagnosis by MRI. It includes details on the application of GAN, dataset, disease, organ of images, classification task (if performed), number of images extracted from the dataset, modality of MRI images, and evaluation metrics. The authors introduce techniques to improve classification accuracy, such as attention and transformation of modes. In addition, they work with MRI and PET images or a combination of both.

The AD Neuroimaging Initiative (ADNI) and the Open Access Series of Imaging Studies (OASIS) are widely used datasets in the field. However, not all related works focus on improving the classification of disease stages, and the number of classes used in the different investigations varies. These GAN models outperform the results of various models in the literature. In addition, their flexibility allows for the definition of different evaluation criteria to measure the performance of the models. Finally, CNNs are a standard method for validating the generated image using GANs.

In this paper, we address the generation of images with AD using a GAN, where the model learns image features to create new images with the disease of the four stages. The validation of synthetic images considers CNNs, greyscale histograms, and brain mass on binarized images to assess GAN efficiency in generating images with AD. Properly validating synthetic images can demonstrate the potential of GANs to create quality images for the training and diagnosis of neurodegenerative diseases.

D. *Ethical implications*

Although synthetic data addresses data scarcity and privacy issues, generating realistic medical images remains a challenge. Moreover, no current regulations guarantee the safe and ethical use of synthetic images, so more rigorous oversight laws are necessary, where interdisciplinary bodies integrated by physicians, AI developers, lawyers, among others, collaborate to develop, maintain, and continually refine best practices for synthetic data [30].

Generative AI has been integrated into healthcare to improve diagnostic accuracy or reduce errors by accessing knowledge during clinical services [31]. This integration has enhanced the people's awareness of potential misuse of these tools, including heightened concerns around security, privacy, and manipulation, where manipulation of human likeness and falsification of evidence underlie the most prevalent tactics in real-world cases of misuse [32]. Hence, generative AI tools must only help physicians validate diagnostics instead of being a method to define diagnostics.

TABLE I. GAN MODELS IN MEDICINE

Model	Description	Loss function	Advantages	Ref
DCGAN	Convolutional.	Cross-entropy	Simple, efficient, and easy to train	[24]
Pix2Pix	U-Net and PatchGAN	L1, Adversarial	Accurate in image-to-image translation	[25]
CycleGAN	Cyclic G-D	Cycle Consistency, Adversarial	Without paired data, style transfer, or domain change	[26]
SRGAN	Deep residual convolutions	Perceptual, Adversarial	High-resolution, visual detail, and quality	[27]
WGAN	Wasserstein distance	Wasserstein Function	Training stability, low collapse issues, and quality improvement	[28]
Pro-GAN	Progressive resolution G	Adversarial	High-quality and detailed images	[29]

TABLE II. GAN MODELS APPLIED TO MEDICINE

Application	Purpose	Image type	Dataset	Classes	GAN type	Evaluation*	No. images	Classification	Ref.
Tumour	Anonymisation	MRI	ADNI, BRATS	5	Pix2Pix	CNN	3,416 and 264	x	[33]
Histopathology	Data augmentation	Tissue	Dartmouth-Hitchcock Medical Centre	3	Cycle-GAN	CNN	427	x	[34]
Alzheimer	MRI-PET mode change	MRI	ADNI	4	GANDALF	CNN	1,525	✓	[35]
Alzheimer	Anomaly detection	MRI	OASIS-3	2	MADGAN	AUC	1,133	x	[36]
Alzheimer	Data augmentation	PET	ADNI	3	DCGAN	PSNR, SSIM	-	x	[37]
Age estimation	Data augmentation	MRI	ABIDE I-II	2	WGAN-GP	CNN	862	x	[38]
Alzheimer	Data augmentation	MRI	ADNI	2	BrainNetGAN	CNN	220	✓	[39]
Hand	Sequence transformation	MRI	One patient	1	Pix2Pix and CycleGAN	PSNR, SSIM	790	x	[40]
Alzheimer	Image harmonisation	MRI	ADNI, OASIS, AIBL	2	AG-GAN with attention	CNN	1,016, 628, and 235	✓	[41]
Alzheimer	Data augmentation	MRI	OASIS	3	Standard GAN	CNN	3,000	✓	[42]
Alzheimer	Data augmentation	MRI	OASIS	3	Standard GAN with attention	CNN	3,000	✓	[43]

*Area Under Curve (AUC), Peak Signal Noise Ratio (PSNR), Structural Similarity Index (SSIM).

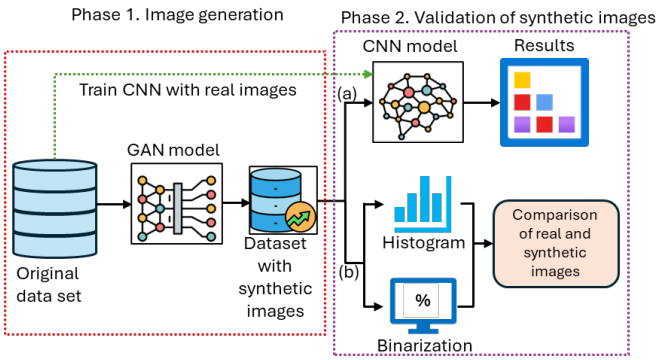

Fig. 2. Two-phase methodology for the training and evaluation of the GAN.

III. Methodology

The proposed methodology involves two phases: The first phase, or the image generation phase, consists of training a GAN model with a real-image dataset of the four stages of AD to generate a new set of synthetic images. The second phase, or validation of synthetic images, involves verifying the synthetic set of images by training a CNN with the dataset of phase one. Once the CNN is trained, the images generated by the GAN will be evaluated by the CNN to show the model's accuracy.

In addition, the second phase considers two methods to validate the images: histogram and binarization. A histogram line of the image can describe the percentage of healthy and damaged brain mass. So, we can use the comparison of histograms between real and synthetic images to estimate the level of dementia in patients. Similarly, image binarization can represent an estimate of the percentage of healthy brain mass.

Figure 2 shows the two-phase methodology for training and evaluating the GAN, where the red and blue boxes mark the steps of the first and second phases, respectively. Also, phase two describes two verification method routes. Route (a) shows the model evaluation using a CNN. Meanwhile, route (b) presents the histogram and binarization processes for the same purpose.

A. Generative Adversarial Networks

GAN models are unsupervised DL models [12], commonly used for data augmentation. We propose using Pix2Pix GAN to generate synthetic images [25]. This GAN is an image-to-image translation model that uses a conditional-type architecture based on U-Net for G and a PatchGAN for D. Pix2Pix generates images very similar to the originals because it identifies and preserves the most important information in the image-to-image translation in as much detail as possible.

Figure 3 shows the Pix2Pix GAN for the generation of synthetic images. It shows the structure of the encoder-decoder U-Net and PatchGAN. The U-Net generator features encoder and decoder layers connected by skip connections, allowing for preserving spatial details during image reconstruction. PatchGAN *D* evaluates small regions of the generated image, rather than the entire image, to distinguish between real and fake images, improving accuracy by capturing local patterns. PatchGAN evaluates small regions of the generated image, rather than the entire image, to distinguish between real and fake images, improving accuracy by capturing local patterns.

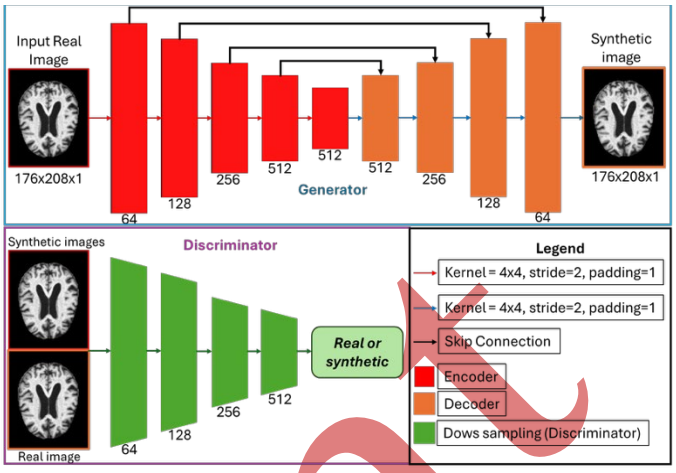

Fig. 3. Pix2Pix GAN model

The validation phase involves evaluating the synthetic images generated by Pix2Pix GAN using three methods. First, a CNN model is used. Second, histogram analysis is used to compare pixel intensity distributions. Third, a binarization technique estimates the proportion of healthy brain mass. The following sections describe each of these validation methods in detail.

B. CNN to evaluate GAN efficiency

The Visual Geometry Group Net (VGGNet) is an object recognition method developed by Oxford University [44]. Its VGG-16 version can outperform several state-of-the-art CNNs for the detection of objects. It has a uniform architecture of 3x3 filters across the network with 13 convolutional layers, five max-pooling layers, and three dense layers. VGG-16 has been extensively explored for classifying images in the medical domain due to the simplicity of its structure.

C. Histogram method to evaluate GAN efficiency

Comparison of histogram curves of real and synthetic images that consider the characteristics of the image is an alternative method for detecting AD [45]. This method measures the grey-to-white matter ratio, identifies the maximum number of pixels based on intensity, and applies a Gaussian filter to smooth the total number of pixels per intensity. Figure 4 shows examples of smooth histogram curves for the four stages of real images in the data set.

For this analysis, values close to zero correspond to the white part of the images, and values close to 255 correspond to black. The healthy brain image has more pixels with an intensity between 0 and 50 (red line). This number decreases according to the severity of the disease: very mild (blue line), mild (green line), and moderate (purple line). This pattern describes the progression of the disease with higher white areas for images with healthy brains and higher black areas for damaged brains. In addition, a lower presence of pixels with an intensity between 200 and 255 is a characteristic of healthy brains (red line); this number increases according to the severity of the disease. The number of pixels with an intensity between 50 and 200 does not reflect a pattern or characteristics of the AD stages.

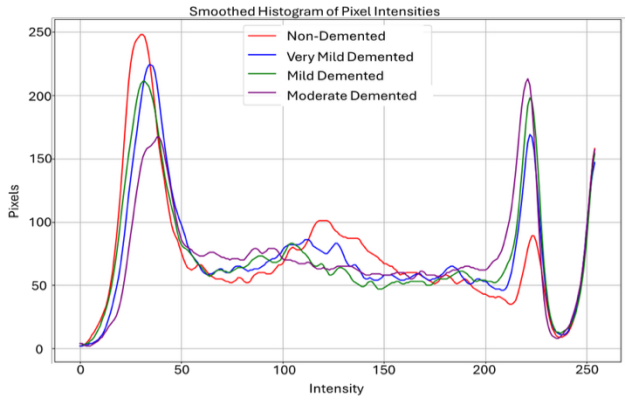

Fig. 4. Examples of smooth histogram curves of the four stages of AD.

D. Binarization to evaluate GAN efficiency

A binarization process estimates the percentage of healthy brain mass described in the image with the white colour. Hence, the complement of brain mass is considered the damaged part of the brain. The binarization of the images defines a mask to obtain the values corresponding to the brain and ignore the contour accompanying images, which do not provide relevant information. Figure 5 shows the binarization process.

Initially, the original image is smoothed, and the Canny algorithm is applied to detect the edges of the brain with the closing technique to define the contour of the brain mass. Later, the largest contour is defined and drawn on the mask so that it can be filled with one's values and finally obtain the mask corresponding to the brain. Finally, the mask consists of zero values that point out the black area without relevant information and one value that points out the white area with the necessary information about the image.

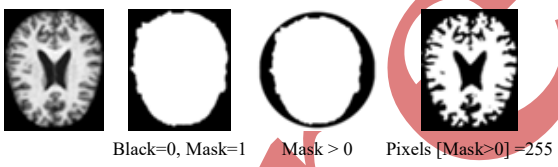

Fig. 5. Binarization of images for the estimation of brain mass.

A threshold defines the binarization of the images. If a pixel has a value lower than or equal to the threshold, then the pixel is considered white. In other cases, the pixel is set to black. The binarization process is carried out using the following equation

$$\text{percentage} = \left(\frac{white_{pixels}}{total_{pixels}}\right) * 100 \quad (1)$$

where the percentage of the white area is calculated by counting the pixels of the white colour and dividing them by the total number of pixels within the mask.

Figure 6(a) shows an example of real images to define the percentage values using the binarization process. Table 3 presents the percentage of brain mass for the four stages. Healthy brains have a greater amount of brain mass, and this value decreases as the disease progresses in stages.

TABLE III. PERCENTAGE OF BRAIN MASS OF REAL IMAGES.

AD stages	Brain mass (%)
None	66.49
Very Mild	60.21
Mild	56.53
Moderate	52.86

IV. EXPERIMENTAL RESULTS

We evaluated the performance of GAN for generating AD images. The implementation based on Python 3.12.14 and the PyTorch 2.4.0 library is performed on a server with Debian 6 OS, 2 processors Intel Xeon Gold 5420+, 56 cores, 100 GB RAM, and a NVIDIA A100-PCIE-40GB GPU.

A. Dataset

Our implementation uses MRI images from the Kaggle Alzheimer's Dataset [46] to train the GAN model and a CNN to diagnose AD. The dataset has images of a single pre-selected slice, ready to be processed without a pre-processing stage. Figure 6 shows four examples of the images in the dataset for three stages of the demented and healthy brain. Table 4 presents the distribution of the images in the dataset for the four stages. All images have a pixel resolution of 176 x 208 pixels, axial slice, and a JPG format.

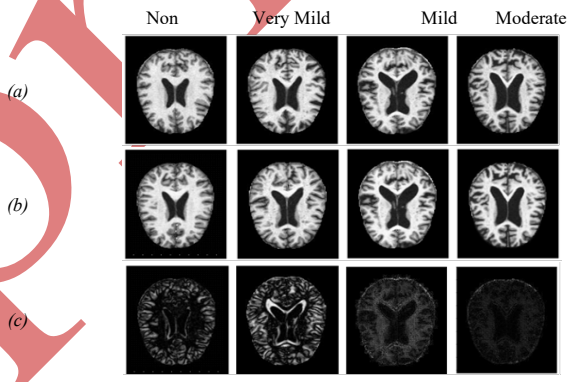

Fig. 6. Difference (c) between (a) real and (b) synthetic images.

TABLE IV. DATASET CHARACTERISTICS

Non-demented	Very Mild demented	Mild demented	Moderate demented
3,200	2,240	896	64

B. Image generation using GAN

The configuration of the hyperparameters for the GAN is fundamental because they affect the model's performance. We use a set of hyperparameters reported in the literature for training and evaluation [39], [42], [43]. The GAN was trained using the following hyperparameters: 100 epochs, batch size of 16, learning rate of 0.0002, and Adam optimiser. The training time of the GAN model with 5,757 images is 44.83 minutes, and the generation time of 100 images is 0.0193 minutes, considering this configuration.

The loss function combines Binary Cross Entropy (BCE) with L1 regularisation. BCE is defined by:

$$BCE = -\frac{1}{N}\sum_{i=1}^{N}(y_i log(p_i) + (1 - y_i)\log(1 - p_i)) \quad (2)$$

where N is the number of samples, y_i is the target value (1 for real samples and 0 for generated samples), p_i is the prediction of D, probability that a sample is a real image.

The L1-norm is

$$L1 = \frac{1}{n}\sum_{i=1}^{n}|\bar{y}_i - \hat{y}_i|$$

where n is the number of pixels in the image, $\bar{y}_i$ is the pixel value in the real image and $\hat{y}_i$ is the pixel value in the generated image.

Figure 6 shows a comparison of real and synthetic images from the four AD stages. The columns define the AD stages, and the rows present the type of image. Row (c) shows the difference between real images (a) and synthetic images (b) generated by the GAN based on the real images of row (a). Row (c) highlights the variation between the two images.

Figures 7-10 present loss function values for G and D during the GAN training with four AD stages. In Figure 7, the values of G tend to be higher than the values D, which suggests that G faces difficulties in creating sufficiently realistic images. Furthermore, the abrupt changes in both losses indicate a problem of model convergence. In Figure 8, the model significantly reduces the difference between the generated and real images during the first epochs, indicating good initial progress. However, some fluctuations appear during training. These results suggest that the model is learning correctly, although fine-tuning can help to achieve more stable convergence.

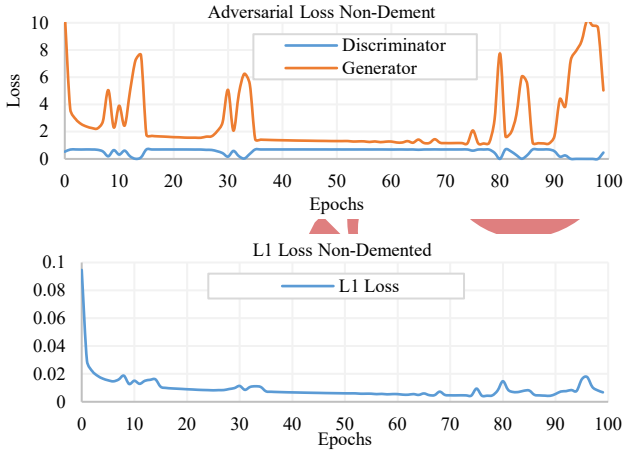

Fig. 7. Loss functions of the GAN for the Non-Demented stage.

Figure 9 shows the behaviour of the loss function during the training of GAN with Mild Demented images. Loss values of G show fluctuations after epoch 50, similar to the training of the GAN model for Very Mild Demented.

Figure 10 shows the values of the loss function for the training of the GAN with the Moderate Demented stage. The value of G decreases rapidly and stabilizes at low values, indicating that G creates realistic images that are difficult to differentiate for D. The values of D remain close to zero, which suggests that D remains effective in identifying real images, achieving a balance with G. Likewise, the L1 values decrease at the beginning and stabilise, implying that the images G maintain a structural similarity with the real ones. Overall, L1 indicates that the model shows good convergence by generating images that resemble real ones.

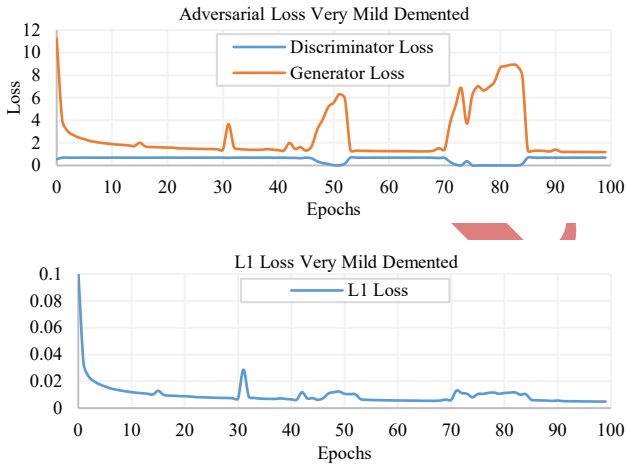

Fig. 8. Loss functions of the GAN for the Very Mild dementia stage.

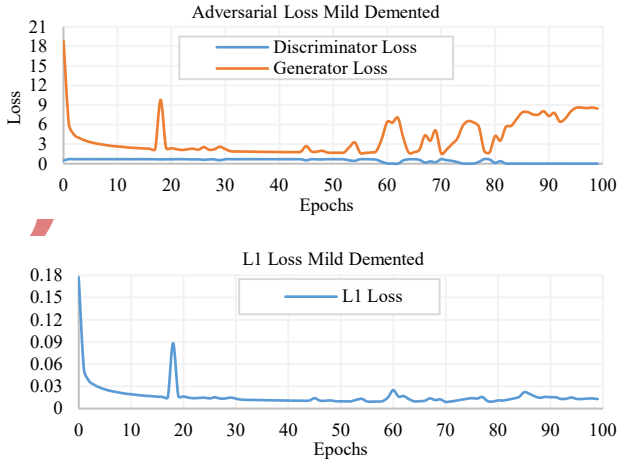

Fig. 9. Loss functions of the GAN for the Mild Demented stage.

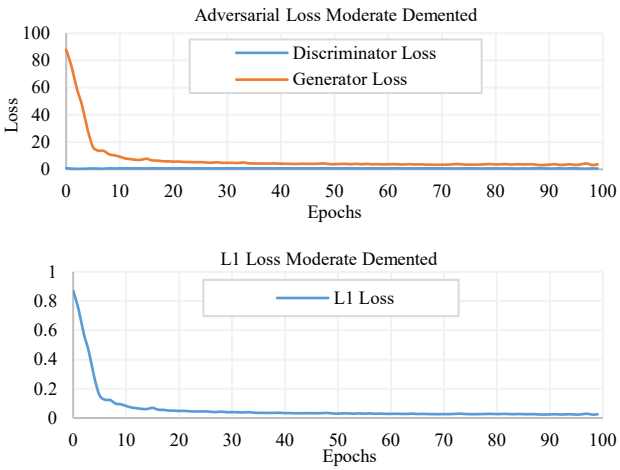

Fig. 10. Loss functions of the GAN for the Moderate Demented stage.

A total of 25 images are generated using the trained Pix2Pix GAN model per stage, totalling 100. We used this set of images to validate the efficiency of GAN before producing a more significant number of images to balance the dataset.

Figures 1A to 4A show examples of the images generated using the GAN model for the four AD stages; see the Appendix.

C. *Evaluation of synthetic images using histograms*

The first evaluation method uses histograms to test the efficiency of GAN to create magnetic resonance imaging for the stages of AD. The histogram patterns of the real and synthetic images show their similarities. Hence, we can use them to classify the synthetic images generated using GANs.

Figure 11 shows the comparison of histograms between synthetic and real images. The distribution of values in the histogram indicates the similarity between the images generated by the GAN and the real images. It suggests the efficiency of the model in replicating the features of real images.

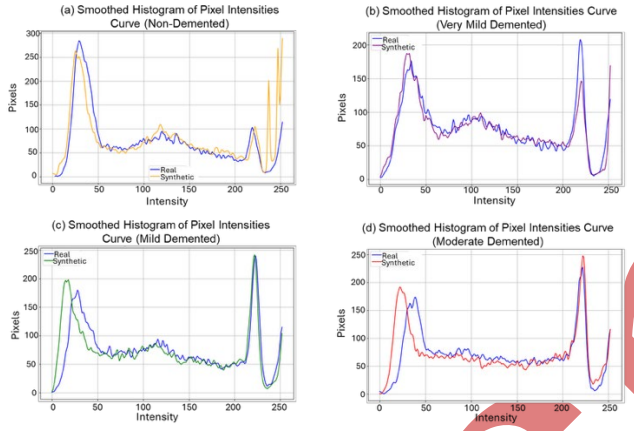

Fig. 11. Comparison of histogram curves of real and synthetic images.

D. *Evaluation of synthetic images using binarization*

The binarization method is the second method used to evaluate the quality of the synthetic images generated by the GAN. Four synthetic images from different stages are randomly selected for evaluation. The threshold for the binarization process is set at 128, which provides better control over the process, contrary to the adaptive thresholds in the literature [45], [47], [48].

Figure 6(b) shows the images for evaluation with their respective real images; see 6(a). Table 5 shows the percentage of brain mass in synthetic images using the binarization process with a static threshold. Synthetic images show percentages similar to those observed in the real images: the percentage of healthy brain mass decreases as the disease progresses. The binarization method provides a measure that can be used to classify the AD stages of synthetic images.

TABLE V. PERCENTAGE OF BRAIN MASS OF SYNTHETIC IMAGES

Dementia Stages	Brain mass (%)
None	64.07
Very Mild	62.07
Mild	59.45
Moderate	55.15

E. *Evaluation of synthetic images using VGG-16*

The VGG-16 model is the third method used to evaluate the quality of the synthetic images generated by the GAN. The VGG-16 model was trained with the following hyperparameters: 100 epochs, a batch size of 64, a learning rate of 0.001, the Adam optimiser, and Binary Cross Entropy as a loss function. We use the values reported in the literature to set the hyperparameters [49], [50], [51], [52]. Additionally, we apply a simple split of 80-20 for VGG-16 training with the dataset described in Table 4.

Figure 12 shows the values of the loss function for the training and validation of the CNN model. Both values are stable after epoch 40, suggesting a good fit and consistent performance. The small difference between training and validation loss values reflects that the model generalises well, achieving balanced and accurate learning.

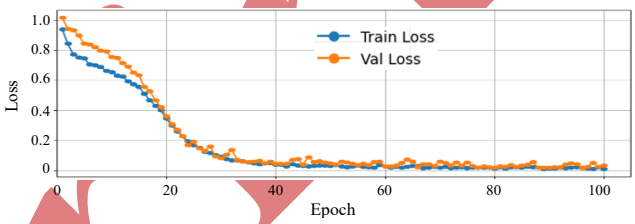

Fig. 12. The loss function of VGG-16 for the classification of AD stages.

Figure 13 shows the confusion matrix for the evaluation of VGG-16 with a set of 100 synthetic images, 25 per stage. The diagonal values show that all the synthetic images are correctly classified, so the model accuracy is 100%.

Stages	Non	Very Mild	Mild	Moderate
Non	**25**	**0**	**0**	**0**
Very Mild	**0**	**25**	**0**	**0**
Mild	**0**	**0**	**25**	**0**
Moderate	**0**	**0**	**0**	**25**

Fig. 13. Confusion matrix of VGG-16 for the GAN-generated image classification.

To further assess the quality and consistency of the MRI generated by the GAN—previously validated through the VGG-16 classification (see Figure 13)—we performed a Principal Component Analysis (PCA). This analysis evaluates the similarity between real and synthetic images based on their extracted feature representations. As a case study, we consider the MD stage, for which images were generated in pairs. Feature extraction was applied to both real and synthetic images prior to PCA, following the methodology described in [48]. Figure 14 presents the resulting distribution of features in the PCA space, where the orange points denote synthetic images and the blue points correspond to real images.

Figure 14 shows that the synthetic images exhibit a feature distribution closely aligned with that of the real images. Most of the synthetic feature points are tightly clustered around those of the real samples, with minimal dispersion. This compact distribution indicates that the GAN can generate synthetic

images that effectively capture the characteristics of the target stage of dementia. Consequently, these synthetic samples are reliable for augmenting training datasets and enhancing the development of balanced and robust classification models.

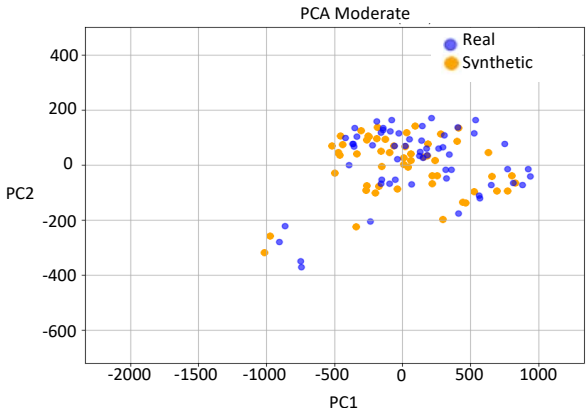

Fig. 14. Characteristics of real and synthesised images generated by the GAN of the MD stage.

V. Conclusion And Future Work

Data augmentation techniques are fundamental in machine learning, especially when data cannot be shared due to privacy restrictions. In recent years, Generative Adversarial Networks (GANs) have been viewed as an important and prominent alternative to solve the limitation of small datasets. Unfortunately, these models exhibit limitations for several specific domains, such as medicine, which are important current research topics due to the benefits for the field.

In this paper, we use three techniques to verify the efficiency of a GAN in generating magnetic resonance imaging images of the four stages of Alzheimer's disease. We train a Pix2Pix GAN model to generate synthetic images of brains without demented, very mild demented, mild demented, and moderate demented. The techniques to evaluate the efficiency of the GAN are based on Convolutional Neural Networks (CNN), Histograms, and image binarization.

The results show that histograms, image binarization, and CNNs are important alternatives to evaluate the efficiency of the GAN in generating images for AD stages. The histograms show patterns similar to the original images. Meanwhile, binarization of synthetic images provides similar percentages of brain mass to those generated with real images. The verification with the CNN based on VGG-16 demonstrates the model's efficiency in generating images of the four stages. The precision of the evaluation is 100% for all the stages of the disease.

More studies are required to assess the actual performance and effectiveness of the approach in a more complex domain. Some topics to be addressed are a large number of synthetic images, an exhaustive search of hyperparameters, adaptive thresholds for the binarization of the images, deep analysis of the histograms, additional metrics for image evaluation, verifying the biologically implausible patterns in the synthetic images, and a comparison with the state-of-the-art models in the literature.

Acknowledgements

The research conducted in this publication was jointly funded by the European Commission under grants 101186291 (SMARCO: SMARt COmmunities Skills Development in Europe) and 101084013 (DIGITAL4Business), the Irish Research Council under grant number GOIPD/2023/1341, and the Ministry of Science and Higher Education of the Russian Federation under grant number 075-15-2022-294.

References

[1] L. Pinto-Coelho, "How artificial intelligence is shaping medical imaging technology: A survey of innovations and applications," *Bioengineering*, vol. 10, no. 12, p. 1435, Dec. 2023, doi: 10.3390/bioengineering10121435.

[2] Y. Skandarani, P.-M. Jodoin, and A. Lalande, "GANs for medical image synthesis: An empirical study," *J. Imaging*, vol. 9, no. 3, p. 69, Mar. 2023, doi: 10.3390/jimaging9030069.

[3] V. J. De-Paula, M. Radanovic, B. S. Diniz, and O. V. Forlenza, "Alzheimer's Disease," in *Protein Aggregation and Fibrillogenesis in Cerebral and Systemic Amyloid Disease*, vol. 65, J. R. Harris, Ed., in Subcellular Biochemistry, vol. 65. , Dordrecht: Springer Netherlands, 2012, pp. 329–352. doi: 10.1007/978-94-007-5416-4_14.

[4] G. Cipriani, C. Dolciotti, L. Picchi, and U. Bonuccelli, "Alzheimer and his disease: a brief history," *Neurol Sci*, vol. 32, no. 2, pp. 275–279, Apr. 2011, doi: 10.1007/s10072-010-0454-7.

[5] N. Huynh and G. Deshpande, "A review of the applications of generative adversarial networks to structural and functional MRI based diagnostic classification of brain disorders," *Front. Neurosci.*, vol. 18, p. 1333712, Apr. 2024, doi: 10.3389/fnins.2024.1333712.

[6] R. Najjar, "Redefining radiology: a review of artificial intelligence integration in medical imaging," *Diagnostics*, vol. 13, no. 17, p. 2760, Aug. 2023, doi: 10.3390/diagnostics13172760.

[7] X. Jiang, Z. Hu, S. Wang, and Y. Zhang, "Deep learning for medical image-based cancer diagnosis," *Cancers*, vol. 15, no. 14, p. 3608, Jul. 2023, doi: 10.3390/cancers15143608.

[8] F. M. Asch *et al.*, "Human versus artificial intelligence–based echocardiographic analysis as a predictor of outcomes: An analysis from the world alliance societies of echocardiography COVID study," *Journal of the American Society of Echocardiography*, vol. 35, no. 12, pp. 1226-1237.e7, Dec. 2022, doi: 10.1016/j.echo.2022.07.004.

[9] B. Pulido-Gaytan *et al.*, "Privacy-preserving neural networks with Homomorphic encryption: Challenges and opportunities," *Peer-to-Peer Netw. Appl.*, vol. 14, no. 3, pp. 1666–1691, May 2021, doi: 10.1007/s12083-021-01076-8.

[10] B. Pulido-Gaytan and A. Tchernykh, "Self-learning activation functions to increase accuracy of privacy-preserving Convolutional Neural Networks with homomorphic encryption," *PLoS ONE*, vol. 19, no. 7, p. e0306420, Jul. 2024, doi: 10.1371/journal.pone.0306420.

[11] N. K. Singh and K. Raza, "Medical image generation using generative adversarial networks," 2020, *arXiv*. doi: 10.48550/ARXIV.2005.10687.

[12] I. J. Goodfellow *et al.*, "Generative adversarial networks," 2014, *arXiv*. doi: 10.48550/ARXIV.1406.2661.

[13] F. Yuan, Z. Zhang, and Z. Fang, "An effective CNN and Transformer complementary network for medical image segmentation," *Pattern Recognition*, vol. 136, p. 109228, Apr. 2023, doi: 10.1016/j.patcog.2022.109228.

[14] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," Mar. 29, 2019, *arXiv*: arXiv:1812.04948. Accessed: Aug. 12, 2024. [Online]. Available: http://arxiv.org/abs/1812.04948

[15] A. Razavi, A. van den Oord, and O. Vinyals, "Generating diverse high-fidelity images with VQ-VAE-2," Jun. 02, 2019, *arXiv*: arXiv:1906.00446. Accessed: Aug. 12, 2024. [Online]. Available: http://arxiv.org/abs/1906.00446

[16] L. Alzubaidi *et al.*, "A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications," *J Big Data*, vol. 10, no. 1, p. 46, Apr. 2023, doi: 10.1186/s40537-023-00727-2.

[17] P. Welander, S. Karlsson, and A. Eklund, "Generative adversarial networks for image-to-image translation on multi-contrast MR images - a comparison of CycleGAN and Unit," 2018, *arXiv*. doi: 10.48550/ARXIV.1806.07777.

[18] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: A review," 2018, doi: 10.48550/ARXIV.1809.07294.

[19] T. Neff, C. Payer, D. Stern, and M. Urschler, "Generative adversarial network based synthesis for supervised medical image segmentation," May 2017, *Verlag der Technischen Universität Graz*. doi: http://dx.doi.org/10.3217/978-3-85125-524-9-30.

[20] T. Neff, C. Payer, D. Štern, and M. Urschler, "Generative adversarial networks to synthetically augment data for deep learning based image segmentation," May 2018, *Verlag der Technischen Universität Graz*. doi: http://dx.doi.org/10.3217/978-3-85125-603-1-07.

[21] A. Beers *et al.*, "High-resolution medical image synthesis using progressively grown generative adversarial networks," 2018, *arXiv*. doi: 10.48550/ARXIV.1805.03144.

[22] Y. Chen, A. G. Christodoulou, Z. Zhou, F. Shi, Y. Xie, and D. Li, "MRI super-resolution with GAN and 3D multi-level DenseNet: smaller, faster, and better," 2020, *arXiv*. doi: 10.48550/ARXIV.2003.01217.

[23] S. Dayarathna, K. T. Islam, S. Uribe, G. Yang, M. Hayat, and Z. Chen, "Deep learning based synthesis of MRI, CT and PET: Review and analysis," *Medical Image Analysis*, vol. 92, p. 103046, Feb. 2024, doi: 10.1016/j.media.2023.103046.

[24] A. Radford, L. Metz, and S. Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," 2015, *arXiv*. doi: 10.48550/ARXIV.1511.06434.

[25] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," 2016, *arXiv*. doi: 10.48550/ARXIV.1611.07004.

[26] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," 2017, *arXiv*. doi: 10.48550/ARXIV.1703.10593.

[27] C. Ledig *et al.*, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," 2016, *arXiv*. doi: 10.48550/ARXIV.1609.04802.

[28] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017, *arXiv*. doi: 10.48550/ARXIV.1701.07875.

[29] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," 2017, *arXiv*. doi: 10.48550/ARXIV.1710.10196.

[30] L. R. Koetzier *et al.*, "Generating Synthetic Data for Medical Imaging," *Radiology*, vol. 312, no. 3, p. e232471, Sep. 2024, doi: 10.1148/radiol.232471.

[31] D. Yim, J. Khuntia, V. Parameswaran, and A. Meyers, "Preliminary Evidence of the Use of Generative AI in Health Care Clinical Services: Systematic Narrative Review," *JMIR Med Inform*, vol. 12, p. e52073, Mar. 2024, doi: 10.2196/52073.

[32] N. Marchal, R. Xu, R. Elasmar, I. Gabriel, B. Goldberg, and W. Isaac, "Generative AI Misuse: A Taxonomy of Tactics and Insights from Real-World Data," 2024, *arXiv*. doi: 10.48550/ARXIV.2406.13843.

[33] H.-C. Shin *et al.*, "Medical Image Synthesis for Data Augmentation and Anonymization Using Generative Adversarial Networks," in *Simulation and Synthesis in Medical Imaging*, vol. 11037, A. Gooya, O. Goksel, I. Oguz, and N. Burgos, Eds., in Lecture Notes in Computer Science, vol. 11037. , Cham: Springer International Publishing, 2018, pp. 1–11. doi: 10.1007/978-3-030-00536-8_1.

[34] J. Wei *et al.*, "Generative Image Translation for Data Augmentation in Colorectal Histopathology Images," 2019, *arXiv*. doi: 10.48550/ARXIV.1910.05827.

[35] H.-C. Shin *et al.*, "GANDALF: Generative Adversarial Networks with Discriminator-Adaptive Loss Fine-tuning for Alzheimer's Disease Diagnosis from MRI," 2020, *arXiv*. doi: 10.48550/ARXIV.2008.04396.

[36] C. Han *et al.*, "MADGAN: unsupervised Medical Anomaly Detection GAN using multiple adjacent brain MRI slice reconstruction," 2020, *arXiv*. doi: 10.48550/ARXIV.2007.13559.

[37] J. Islam and Y. Zhang, "GAN-based synthetic brain PET image generation," *Brain Inf.*, vol. 7, no. 1, p. 3, Dec. 2020, doi: 10.1186/s40708-020-00104-2.

[38] R. Li, M. Bastiani, D. Auer, C. Wagner, and X. Chen, "Image Augmentation Using a Task Guided Generative Adversarial Network for Age Estimation on Brain MRI," 2021, *arXiv*. doi: 10.48550/ARXIV.2108.01659.

[39] C. Li, Y. Wei, X. Chen, and C.-B. Schonlieb, "BrainNetGAN: Data augmentation of brain connectivity using generative adversarial network for dementia classification," 2021, *arXiv*. doi: 10.48550/ARXIV.2103.08494.

[40] J. A. Grant-Jacob, C. Everitt, R. W. Eason, L. J. King, and B. Mills, "Exploring sequence transformation in magnetic resonance imaging via deep learning using data from a single asymptomatic patient," *J. Phys. Commun.*, vol. 5, no. 9, p. 095015, Sep. 2021, doi: 10.1088/2399-6528/ac24d8.

[41] S. Sinha, S. I. Thomopoulos, P. Lam, A. Muir, and P. M. Thompson, "Alzheimer's disease classification accuracy is Improved by MRI harmonization based on attention-guided generative adversarial networks," in *17th International Symposium on Medical Information Processing and Analysis*, A. Walker, L. Rittner, E. Romero Castro, N. Lepore, J. Brieva, and M. G. Linguraru, Eds., Campinas, Brazil: SPIE, Dec. 2021, p. 24. doi: 10.1117/12.2606155.

[42] Wong Pui Ching, Shahrum Shah Abdullah, Mohd Ibrahim Shapiai, and A. K. M. Muzahidul Islam, "Performance Enhancement of Alzheimer's Disease Diagnosis Using Generative Adversarial Network," *ARASET*, vol. 45, no. 2, pp. 191–201, May 2024, doi: 10.37934/araset.45.2.191201.

[43] P. C. Wong, S. S. Abdullah, and M. I. Shapiai, "Double-enhanced convolutional neural network for multi-stage classification of Alzheimer's disease," *Adv. sci. technol. eng. syst. j.*, vol. 9, no. 2, pp. 9–16, Mar. 2024, doi: 10.25046/aj090202.

[44] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," 2014, *arXiv*. doi: 10.48550/ARXIV.1409.1556.

[45] N. Pasnoori, T. Flores-Garcia, and B. D. Barkana, "Histogram-based features track Alzheimer's progression in brain MRI," *Sci Rep*, vol. 14, no. 1, Jan. 2024, doi: 10.1038/s41598-023-50631-1.

[46] S. Dubey, "Alzheimer's Dataset (4 class of Images)." Accessed: Aug. 12, 2024. [Online]. Available: https://www.kaggle.com/datasets/tourist55/alzheimers-dataset-4-class-of-images

[47] S. Roy, D. Bhattacharyya, S. K. Bandyopadhyay, and T.-H. Kim, "An improved brain MR image binarization method as a preprocessing for abnormality detection and features extraction," *Front. Comput. Sci.*, vol. 11, no. 4, pp. 717–727, Aug. 2017, doi: 10.1007/s11704-016-5129-y.

[48] S. Roy, "A New Efficient Binarization Method for MRI of Brain Image," *SIPIJ*, vol. 3, no. 6, pp. 35–51, Dec. 2012, doi: 10.5121/sipij.2012.3604.

[49] D. A. Arafa, H. E.-D. Moustafa, H. A. Ali, A. M. T. Ali-Eldin, and S. F. Saraya, "A deep learning framework for early diagnosis of Alzheimer's disease on MRI images," *Multimed Tools Appl*, vol. 83, no. 2, pp. 3767–3799, Jan. 2024, doi: 10.1007/s11042-023-15738-7.

[50] R. Jain, N. Jain, A. Aggarwal, and D. J. Hemanth, "Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images," *Cognitive Systems Research*, vol. 57, pp. 147–159, Oct. 2019, doi: 10.1016/j.cogsys.2018.12.015.

[51] 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," *IJACR*, vol. 11, no. 53, pp. 51–60, Mar. 2021, doi: 10.19101/IJACR.2021.1152001.

[52] R. A. Hazarika, D. Kandar, and A. K. Maji, "An experimental analysis of different deep learning based models for Alzheimer's disease classification using brain magnetic resonance images," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 10, pp. 8576–8598, Nov. 2022, doi: 10.1016/j.jksuci.2021.09.003.

APPENDIX

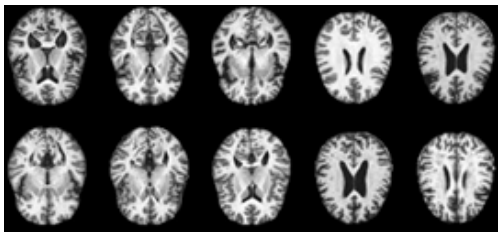

Fig 1A. Examples of synthetic images with Non-Demented

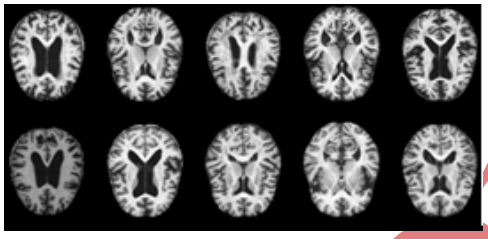

Fig 2A. Examples of synthetic images with Very Mild Demented

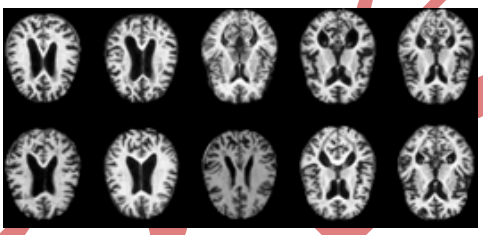

Fig 3A. Examples of synthetic images with Mild Demented

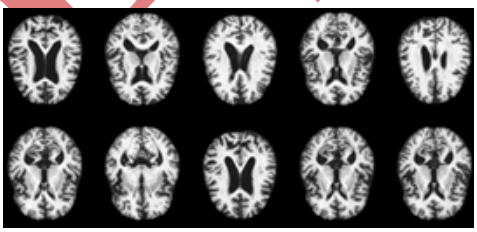

Fig 4A. Examples of synthetic images with Moderate Demented