Svedberg Open
DISSEMINATION OF KNOWLEDGE

International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: https://www.svedbergopen.com/

Research Paper Open Access

Magnetic Resonance Image Enhancement For Early Brain Tumor Diagnosis Using Wasserstein Generative Adversarial Network With Gradient Penalty

Shahla AbdulWahhab AbdulQader[1], Omaima Nazar A. AlAllaf AlMustawi[2*]

[1]Assistant Professor, Department of Network Technologies and Intelligent Systems Engineering, Mosul Polytechnic College, Northern Technical University, Mosul, Iraq, ORCID ID: https://orcid.org/0000-0002-8951-8513, Email: shahla_wa1971@ntu.edu.iq
[2*]Associate Professor, Department of Information Systems, College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, 11432, Saudi Arabia, ORCID ID: https://orcid.org/0000-0002-6724-0573, Email: onalmustawi@imamu.edu.sa
[2*]Correspondence: omaimaalallaf@gmail.com

Abstract— Early detection of brain tumours is really important since symptoms are often vague, which can delay diagnosis. MRI scans can be affected by patient movement, bodily changes, and the need for expert interpretation. Generative Adversarial Networks have become a powerful tool for improving the quality of medical images for brain cancer diagnosis and filling in missing data. So, in this work, we aim to improve the quality of MRI images of brain tumours using an advanced deep learning model based on GAN, specifically the Wasserstein GAN with Gradient Penalty (WGAN-GP) with a generator G and Dual Discriminator D (Dual D). This method is used to make grayscale images from the Br35H (2020) dataset, which includes medical brain tumour images, clearer and more accurate. We conducted many experiments with different network architectures and learning parameters in this research. Our results show an improvement in generated MRI Brain Tumor images with an accuracy of 99.634% and a PSNR of 42.42 dB. This research is important because it helps doctors and researchers get clearer images, making it easier to diagnose Brain Tumors accurately. It also helps improve medical treatment plans and outcomes since original images often have noise or distortions that can affect accuracy.

Keywords—Deep Learning (DL); Generative Adversarial Network (GAN); Wasserstein GAN with Gradient Penalty (WGAN-GP)

I. INTRODUCTION

AI has become a promising field and offers potential solutions for creating and improving medical images. Brain tumours are a serious illness, and medical experts often face a big shortage of imaging data because of patient privacy concerns [1]. Brain tumour segmentation aims to tell the difference between healthy tissue and tumour-affected areas. This step is crucial for diagnosing brain tumours and planning treatment to boost the chances of successful therapy. MRI provides detailed insights that are key for accurate brain tumour diagnosis and can help replace manual detection that relies on medical professionals' expertise. Also, traditional methods for diagnosing brain tumours using MRI scans are expensive, need high doses of radiation, take a long time to gather data, and can be inconsistent. So, accurately spotting and detecting brain tumours is really important for early diagnosis and creating the right treatment plans for patients.

Lately, Deep Neural Networks (DNNs) have shown great results in image processing, from high-level recognition and segmentation to low-level tasks like denoising, super-resolution, and reconstructing compressed images. Even with their progress, DNNs still have challenges, like improving performance and dealing with different levels of image damage. The field of medical image enhancement using DNNs has made big strides since the mid-2010s. Deep Conventional Neural Networks (CNN) were introduced in [2]..[7]. This uses autoencoders with symmetric skip connections to improve image restoration. It tackles detail loss by passing info from early layers to later ones in a balanced way, helping to better reconstruct medical images. Next came the stage of developing Generative Adversarial Network (GAN) Deep Learning (DL) methods.

GAN methods can be used for image-to-image translation, turning low-resolution images into high-resolution ones. For example, generating MRI images from CT scans can help create multimodal datasets from just one type. Research in Medical Image-to-Image Translation has advanced with models like

CycleGAN and Pix2Pix. This lets you convert an image from one medical imaging style to another while keeping all the details, helping to simulate or enhance medical images without needing perfectly matched datasets. While GANs can create large datasets, their limited diversity and quality have recently been boosted by denoising diffusion probabilistic models, which perform better in generating natural images.

The progress of GAN and Wasserstein GAN with Gradient Penalty (WGAN-GP) in improving medical brain tumour images is discussed, including the technical developments and challenges tackled up to 2025. Wasserstein GAN (WGAN) was introduced by Arjovsky et al. [3], swapping the standard Jensen-Shannon divergence for the Wasserstein distance in its loss function, which helps stabilise training and reduces mode collapse. Still, WGAN struggled to stick to the Lipschitz constraint, which led to an improvement conducted by Ishaan Gulrajani [4], where WGAN-GP was suggested by adding a Gradient Penalty to better enforce the rule and make training more stable. This tweak opened the door for more uses, like improving medical images.

Recent reviews by Showrov et al. [8] and Jabbar Hussain et al. [9] have summed up the big advances GANs have made in medical imaging, looking at challenges like keeping training stable, not having enough data, and needing to create high-quality images with accurate medical details. These studies highlighted the need to use other GAN methods like Attention, Multi-scale, and Perceptual Loss, and to combine generative models with other DL models to boost performance. In 2025, comparisons of different GANs setups for Brain MRI was conducted by Chethan et al. [10], pointing out the differences between attention-based and traditional models. The strengths and weaknesses of each setup were covered in the same study, with tips on choosing the right model based on the medical application and the required accuracy and speed.

The evolution from basic autoencoders to advanced GANs with multi-level attention, multi-scale perceptual loss, gradient penalty, semi-supervised generation, and diffusion techniques has really improved Brain Tumor image enhancement. These models also boost diagnostic accuracy, provide reliable training data, and improve the quality and precision of Brain Tumor images to help doctors with diagnosis and treatment. The main challenges are still GAN training stability, limited labelled data, and privacy concerns. So, we need smart generation and discrimination techniques to enhance performance while keeping models interpretable, efficient, and generalisable for Brain Tumor diagnosis. This is the main trend in research in this area.

Therefore, our work is focused on improving the accuracy and quality of MRI Brain Tumor images using the DL Wasserstein GAN with Gradient Penalty model (WGAN-GP) with a Generator and Dual D. The grayscale Brain Tumor images were taken from the Br35H Dataset. The importance of our work is in helping medical professionals get clearer MRI images, which help in more accurately diagnosing brain tumours. The rest of this research is organised like this: Section II goes over the related studies; Section III talks about the different GAN approaches; Section IV explains the research method; Section V shows the experimental results and discussion; and finally, Section VI wraps up the work with ideas for future research.

II. RELATED STUDIES

Lots of practical guides for effectively training GAN and WGAN-GP have been put together by many researchers. This makes it easier for researchers to use these models and broaden their applications, especially in areas that need high quality like medicine. GAN multi-model for medical data was introduced in 2017 by Agisilaos Chartsias et al. [11] to boost diagnosis by combining info from different images like MRI and CT. This helps improve the accuracy of tumour detection and analysis. Low-dose CT (LDCT) scans can lower the amount of radiation a patient gets. A Progressive Wasserstein GAN with a weighted structurally-sensitive hybrid loss function (PWGAN-WSHL) was suggested by Wang and Xueli Hu in 2021 [12] to help remove noise from LDCT images and make CT denoising more effective. The hybrid loss function cut down on artifacts while keeping more detail in the denoised results. Following this, there's been a fast push to develop and improve medical image generation models to enhance original MRI images [13]. In 2018, Shin et al. [13] conducted GANs for generating images and boosting data while keeping things private by removing personal info without messing with patient confidential.

In 2024, specific applications in the medical field started to appear, like Wasserstein GAN with Gradient Penalty (WGAN-GP) and nearest-neighbour interpolation used by Hector et al. [14] to create high-quality synthetic images for diabetic retinopathy classification. Meanwhile, in 2021, a Brain Tumour detection system was suggested by Anuja Arora et al. [15] for segmenting Gliomas in pre-operative MRI scans using a U-Net-based DL model. The first step is to transform the input image data, which then goes through various techniques like subset division, narrow object region, Brain category slicing, and feature scaling. These steps are done before feeding the data into the U-Net DL model. This model is used for pixel label segmentation on the tumour regions and experiments were based on the BraTS 2018 dataset.

Attention techniques have started making their way into medical GANs. In 2021, Krithika and Suganth [16] introduced different GAN setups for analysing medical images. In 2023, a multi-scale attention GAN (MAGAN) was suggested by Guojin Zhong et al. [17] to boost medical images from unpaired ones. This model was trained using the adversarial interaction of two generators and two discriminators. The idea is to combine multi-scale info when extracting features by building a feature pyramid and filtering out irrelevant stuff, highlighting important areas based on attention distribution, which helps imaging and improves the quality of enhanced images. In 2024, Qihong Yang et al. [18] suggested a multimodal Brain Tumor MR segmentation approach using domain adaptation. First, they used clustering methods in the pre-training stage for each modality to pick target domain images that complement the source domain best, then generated pseudo labels. Second, they added feature adapters to improve feature alignment and made a network that responds to both source and target domain images to use the multimodal image info. Meanwhile, in 2023, Asiri et al. [19] used a GAN DL method where two NNs compete to get better at creating realistic MRI images. The GAN network mainly has two parts: the generator and the discriminator. The generator is a convolutional NN and the discriminator is a deconvolutional NN. They collected the Contrast-Enhanced MRI (CE-MRI) dataset from 2005 to 2020 from various hospitals in China, which has four classes. Their results reached an accuracy of 96%.

In recent years, quite a few studies [20], [21], [22], [23], [24] and [25] have come up that mix deep competitive networks (GAN) and U-Net models to boost accuracy in tumor segmentation. They focused on using Attention GANs and Self-Attention to balance accuracy and speed, leading to big improvements in generation and segmentation quality compared to the usual models. Also, in 2021, Qian Wang et al. [26], researchers suggested a GAN algorithm to get more image features by adding multi-level GAN convolutions. They did this by adding multiple residual blocks and global residuals to extract and learn features from noisy images to avoid losing image details. Their network used a weighted sum of feature loss and perceptual loss to keep detailed image information while removing noise. Their experiments showed that their algorithm can effectively remove image noise and make the images look better. Meanwhile, in 2025, Hugo et al. [27], researchers looked into how to fix overfitting by carefully testing different image data augmentation techniques. They added different image samples to make the dataset bigger and more diverse. They showed results checking how these techniques affect image classification and semantic segmentation.

Interest in MR-only treatment planning using synthetic CTs (synCTs) has grown quickly in radiation therapy. One big challenge is figuring out techniques that work with medical images showing unusual anatomy. In 2020, a spatial attention-guided GAN model was suggested by Hajar [28] to create accurate synCTs from T1-weighted MRI images for typical anatomy. Tests on fifteen brain cancer patients showed that attention-GAN did better than current synCT models, with average MAEs of 85.223±12.08, 232.41±60.86, and 246.38±42.67 Hounsfield units between synCT and CT-SIM across the whole head, bone, and air areas, respectively.

The use of GANs has become more popular in LDCT denoising research recently because of their big performance benefits. Differences in LDCT images' noise and artifact levels can make GANs less robust and effective at denoising. In 2024, Yanling et al. [29] suggested a scale-sensitive GAN to make GANs more robust in LDCT denoising. Their network has an error feedback pyramid generator, which boosts the network's ability to spot noise and artifacts in LDCT images by doing feature extraction and noise reduction at different scales. They also added a multi-scale residual discriminator with better feature extraction to make it easier to tell real samples from artificial ones. Cross-modal generation has become an important way to tackle the problem of missing modalities in medical imaging.

Current methods rely on CNNs or vision transformers (ViTs) and their variants as the main models. But this comes with issues like limited receptive fields and high computational costs. To deal with this, in 2025, Xinmiao and Yang [30] came up with a latent feature space-based multi-scale residual ViT generative adversarial model (LMRT-NET). Their model combines the ViTs global sensitivity with the CNNs local accuracy while cutting down on computational costs. The LMRT-NET generator was based on encoder-decoder architecture. This is done to improve performance and reduce computational expenses by employing multi-scale dynamic aggregation residual ViT (DART) blocks in the latent feature space. Their module was consisted of two layers of residual convolutional blocks and transformer blocks of varying scales. The transformer blocks help the convolutional blocks capture contextual features and the lower-level blocks support higher-level blocks in learning high-dimensional global information. In 2025, Idriss et al. [31] proposed an automated classification model based on CNNs to improve diagnostic consistency and speed. They used a dataset of 3,060 MRI images, and their model employed the Grad-CAM technique to visualize key regions influencing its decisions. They conducted rigorous testing and evaluated performance using accuracy, precision and recall metrics. Their results demonstrated that the CNN-based model provides superior classification performance and increased transparency compared to conventional methods.

In 2023, Gustav et al. [32] came up with Medfusion: a conditional latent DDPM for medical image generation, and tested how it stacked up against GANs. Medfusion was trained and compared with StyleGAN-3 using fundoscopy images from the AIROGS dataset, radiographs from the CheXpert dataset, and histopathology images from the CRCDX dataset. Based on earlier studies, Progressively Growing GAN (ProGAN) and Conditional GAN (cGAN) were used as extra benchmarks on the CheXpert and CRCDX datasets, respectively. Medfusion showed equal or better fidelity across all three datasets. Also, in 2023, Ashfak et al. [33] created semi-supervised GAN models that let you use less labelled data while still nailing Brain Tumour classification, which helps tackle the lack of labelled data in the medical field.

More advanced model called UTAD-Net was introduced in 2024 by Chuan et al. [34], which focused on unsupervised translation between different brain tumour imaging types and showed it could accurately identify and translate tumour areas. The model has two parts: a teacher network and a student network. The teacher network first learns an end-to-end mapping from the source to the target modality using unpaired images and their tumour masks. Then, the translation knowledge is passed on to the student network, allowing it to create more realistic tumour areas and complete images without needing masks. In 2024, Abid et al. [35] suggested a DL GAN approach that uses transfer learning and auto-encoder techniques to improve Brain Tumor segmentation. The GAN has a generator and a discriminator to get better quality segmentation results. They used down-sampling and up-sampling in the generator for Tumor segmentation. The encoder was designed to keep as much information as possible, and the decoder rebuilds the image from these encodings. Transfer learning was applied at the bottleneck with the DenseNet model. This method improved the accuracy of Brain Tumor segmentation, especially for small tumors.

Hybrid structures like Hybrid Dual-GAN (Ultra-Resolution GAN) (WGAN-GP with Real-ESRGAN) have been proposed to create high-quality brain images, mixing the realistic generation skills of WGAN-GP with the detailed enhancement and accuracy from Real-ESRGAN. The main models in image generation right now, focusing on GANs and diffusion models, were covered in the work. It pointed out that GANs are great at quickly producing diverse and high-quality images but often run into problems like mode collapse and training instability. On the other hand, diffusion models are known for creating detailed images, though they need a lot of computing power, making them tricky if resources are limited. The study also talked about the current challenges these models face, like biases and inefficiencies, and suggested possible solutions and future research directions [36][37]. Finally, in 2026, M. Yaseen et al. [38], researchers ran a survey to look at the current state of diffusion models for medical image segmentation. They went through the theory, new methods, and ways to make computations more efficient. They checked out recent improvements in latent diffusion frameworks, architectures, and segmentation models, while also talking about the real-world challenges of using these models in clinics. Their review covers uses across different medical imaging types like MRI, ultrasound, and X-ray. They showed that diffusion models have made great strides in tackling core issues like limited data, differences between observers, and measuring uncertainty.

III. WGAN-GP APPROACHES

Traditional GAN has a generator and a discriminator that compete while the network is training. The generator tries to create data that looks like the real thing, while the discriminator tries to tell the difference between the real and generated data. Training can get tricky because of issues like vanishing gradients, instability, no guarantees for the loss function to converge, and mode collapse, where the generator just spits out the same patterns. To fix these problems, the network was turned into WGAN (Wasserstein GAN), which handles some of these issues using the Wasserstein Distance metric. Eq. (1) shows how to calculate the Earth-Mover (EM) distance or Wasserstein Distance.

$$W(P_r, P_g) = inf_{\gamma \in \Pi(P_r, P_g)} E_{(x,y)\sim\gamma}[\|x - y\|] \quad (1)$$

Where Π (Pr, Pg) represents the set of all joint distributions γ(x, y) whose marginal are Pr and Pg respectively.

And γ(x, y) indicates how much “mass" must be transported from x to y, this is to transform from Pr to Pg distributions. The EM distance here represent the “cost" of the optimal transport plan [3].

Whereas, Eq. (2) represents the fact that $W(P_r, P_0)$ might have nicer properties when optimized than $JS(P_r, P_0)$. In addition, the infimum in Eq.1 is highly intractable. This is representing the Kantorovich-Rubinstein duality [3].

$$W(P_r, P_\vartheta) = sup_{\|f\|_L \leq 1} \; E_{x\sim P_r}[f(x)] - E_{x\sim P_\vartheta}[f(x)] \quad (2)$$

The supremum is over all the 1-Lipschitz functions f: X→R. And, if we replace $\|f\|_L \leq 1$ for $\|f\|_L \leq K$ (consider K-Lipschitz for some constant K), then we end up with $K.W(P_r, P_g)$.

WGAN tackles instability and mode collapse problems in regular GANs. The main difference is that WGAN uses Wasserstein distance (Earth Mover’s distance, EMD) for a smoother loss and better stability, while

GAN uses Jensen-Shannon (JS) divergence or Kullback-Leibler (KL) divergence. Also, WGAN uses a Critic instead of a typical Discriminator, and the Critic doesn't label samples as real or fake with probability; it gives a numerical score showing how close the generated distribution is to the real one using the Wasserstein distance. This make training more stable and improves quality of the generated samples.
The critic network D is used to estimate the function f, and the critic D has to be a Lipschitz-1 function, which limits how fast the gradients can change so they don't go over 1; so the aim is Eq. (3). The WGAN value function is built using the Kantorovich-Rubinstein duality (Eq. (3)) [39].

$$min_G \, max_{D \in D} \; E_{x \sim P_r} \; [D(x)] - \; E_{\tilde{x} \sim P_g} \; [D(\tilde{x})] \qquad (3)$$

Where D is the set of 1-Lipschitz functions and Pg is once again the model distribution implicitly defined by ˜x = G(z), z ∼ p(z). Then under an optimal discriminator case (called a critic, since it's not trained to classify), minimizing the value function with respect to the generator parameters minimizes W(Pr, Pg).

A. Gradient Penalty (WGAN-GP Loss Equation)

In 2017, Ishaan Gulrajani [4] proposed an alternative way to enforce the Lipschitz constraint. The differentiable function is 1-Lipschitz if and only if its gradients have norm at most 1 everywhere. Therefore, they considered directly limiting the norm of the gradient of the critic's output with respect to its input. They applied simple version of the constraint by imposing a penalty on the gradient norm for random samples $\hat{x} \sim \mathbb{P}_{\hat{x}}$ to avoid tractability problems. This is to compute Eq. (4) [4].

$$L = E_{\tilde{x} \sim P_g} [D(\tilde{x})] - E_{x \sim P_r} [D(x)] + \lambda E_{\tilde{x} \sim P_{\tilde{x}}} [(\|\nabla_{\tilde{x}} D(\tilde{x})\|_2 - 1)^2] \qquad (4)$$

Where x∼P_r represent the real sample and $\tilde{x} \sim \mathbb{P}_g$ represent the generated sample. And ˆx represent the random sample for straight line between x and ˆx. As a Sampling Distribution, they implicitly define Pˆx by uniformly sampling points along straight lines connecting pairs of points drawn from the data distribution Pr and the generator distribution Pg. This is motivated by the fact that the optimal critic has straight lines with gradient norm 1 connecting paired points from Pr and Pg in Eq. (1). Since enforcing the unit gradient norm constraint everywhere is impractical, enforcing it only along these straight lines seems sufficient and has been shown experimentally to give good results.
As a Penalty coefficient, all experiments in their work use ƛ = 10, which they found to work well across a variety of architectures and datasets ranging from toy tasks to large ImageNet CNNs [4]. The WGAN-GP network was seen as better because the Gradient Penalty gives a smoother limit on the critic's function instead of a strict weight limit. This helps make training more stable and improves the quality of the generated samples, which cuts down problems like vanishing gradients and mode collapse. The Generator's weights are updated to minimise LG as per Eq. (5) [4].

$$£_G \; = -E_{g^x} \; [D(x_g)] \qquad (5)$$

B. Reconstruction Loss (L1 Loss)

Reconstruction loss encourages the model to make images really similar in pixels to the original. This is measured between the high-resolution output and the real image. It's often used with medical images to keep their structure. The discriminator still does its normal job, but now the generator has to not only fool the discriminator but also stay close to the real output in L2 distance. We're also thinking about using L1 distance instead of L2, Eq. (6), since L1 usually helps reduce blurriness [7].

$$£_{L1}(G) = \; E_{x,y,z} \; [\|y - G(x,y)\|_1] \qquad (6)$$

C. Perceptual Loss (Multi-Scale)

Perceptual Loss is also known as Perceptual Loss (Multi-Scale VGG19 Features) [40][41]. It's used to measure the difference in high-level features between the real and generated image using VGG19. It's also used to calculate a high-resolution 128x128 image using VGG19 locally to improve fine details in images as shown in Eq. (7).

$$£\, perc = \sum_{L \in £} \; || \; (G(x)) \varphi l - \varphi l(y) \; ||1 \qquad (7)$$

Where φl represents a feature from layer l in VGG19 (conv1_2, conv2_2…).

D. LPIPS Loss (Learned Perceptual Image Patch Similarity)

The LPIPS Loss is an AI metric for checking image quality. It measures how similar images are from a human perspective. Lower values mean the generated image is closer in quality to the original. Values are usually near zero, so smaller numbers show the images look more alike to humans. A value of 0 means a

perfect match. Eq. (8) is used to measure similarity between images based on how humans perceive them and is often used as a metric rather than the main training loss [42].

$£_{LPIPS}$ = LPIPS (G (x), y) (8)

E. Feature Matching Loss

The Feature Matching Loss encourages the Generator to match the intermediate features of the distinguished examples rather than just fooling them. Equation 9 measures this as a distortion in the layers of the general and local distinguishable features, which helps speed up learning and improves the gradients for the Generator [43][44].

$$£_{FM} = \sum \left|\left|(G(x))D(i) - D(i)(y)\right|\right| 1 \quad (9)$$

Where D(i) is the output from layer i in the discriminator.

F. Structural Consistency Loss (MSSSIM)

It is used to maintain the structural framework between the generated and original image. Eq. (10) is calculated on the high-resolution image [45].

$£_{struct}$ = 1 - MSSSIM (G(x), y) (10)

G. Total Generator Loss (our Proposed)

We conducted a set of loss functions as shown in Eq. (11) to be used in our proposed WGAN-GP approach.

$£_{total} = ƛ_1 . £_{WGAN} + ƛ_2 . £_{perc} + ƛ_3 . £_{L1} + ƛ_4 . £_{FM} + ƛ_5 . £_{struct}$ (11)

IV. RESEARCH METHODOLOGY

This section gives more details about the main steps of our suggested improved WGAN-GP approach for enhancing MRI brain images to help detect brain tumors.

A. Data Preparation and Pre-processing

The input image samples for the proposed model are Grayscale Brain Tumor images from the Br35H Dataset [46]. The Br35H Dataset has 4 types of Brain images (“Yes”, “No”, “Pred”, and “Br35H-Mask-RCNN”) stored in 4 folders. The “Yes” folder has MRI Brain images with Tumor infection and is often used to classify or detect Tumors. These are in the "Positive" class and we used them as a training set to train a GAN model to generate pathological images. The “No” folder has MRI images of a normal Brain without a Tumor ("Negative Class") and can be used for GAN training if we want to create images of a healthy Brain.

The “Pred” or Prediction class (Predicted Masks) usually contains images generated by a previously trained model, like processed images or ones with Tumors identified (segmented). These aren’t typically used for direct training, but for evaluating results. We chose to use 1500 image samples from the “Yes” folder in the Br35H Dataset, using 80% for training and 20% for testing to make sure the images are randomly distributed. The image size was changed to 128×128 pixels to balance detail and accuracy, while being good for memory and processing time, which works well with convolution layers. We did a few transformations to get the images ready for training, including:

- **Normalizing**: this is key in GAN networks, changing image values from [0-255] to [0-1] for the network. The image values should be within [-1,1].
- **Data Augmentation** was used to boost data variety, help the model generalize better, and cut down overfitting. We used several data augmentation methods to make the model more robust and reduce overfitting, like random rotation, flipping, intensity scaling, and adding noise. The DataLoader handled batches of data during training. All these data changes create a realistic simulation of brain images and make the model's training more stable. Table I shows the different data augmentation techniques.

TABLE I DIFFERENT TECHNIQUES OF DATA AUGMENTATION

Data Augmentation	Description

Rando Rotation	Simulating different angles from ±15° to 30±°
Flipping	Random Horizontal and Vertical Flipping. Compensating for different brain states
Intensity Scaling	Adjusting the brightness. Simulating different camera devices. Changing the contrast and brightness.
Gaussian Noise Injection	Add noise to strengthen the model against interference.
Elastic Deformation	ElasticGrid: Flexible distortion through simulating tissue transformations of Tumor.
Crop & Resize	Improving the positioning of model. Crop the image center, then resize.

B. GAN Network Architecture

The proposed model has two parts: a Generator (G) and a Dual-Discriminator. The Generator Network uses an Attention U-Net with Residual and Self-Attention Blocks. The Attention U-Net architecture includes a Residual Block, Self-Attention, and Multi-Scale Output. Fig.1 shows the overall setup of the proposed method.

1) The Generator (G)

The Generator is used to make high-resolution images with good structure and fine details. Its architecture is based on a modified U-Net.

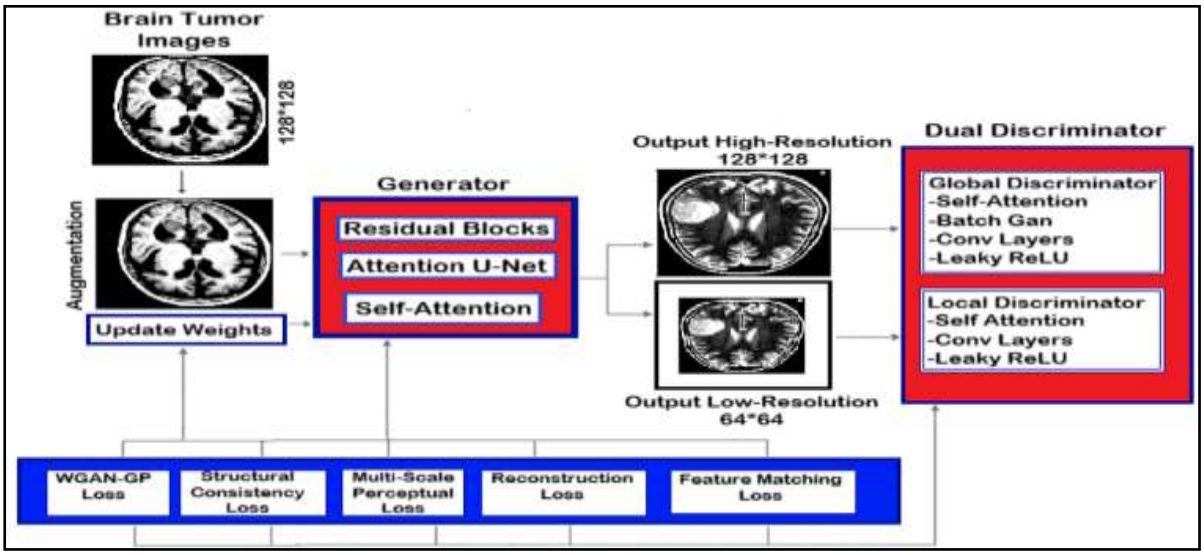

Fig. 1 General architecture of the proposed approach for improving Brain images to detect Brain Tumors.

a) The Encoder

The encoder has a bunch of convolutional layers (Conv) with down sampling (MaxPooling), and it includes Residual Blocks (ResNet) to boost deep learning (deepen representation power) and stop gradient vanishing.

b) Self-Attention layers

Self-Attention layers are built into the Encoder and Decoder to capture long-range spatial dependencies. It helps the model focus on important areas like the tumour. Self-Attention lets the model concentrate on key regions in the image across all locations (global dependencies) and strengthens connections between distant regions Tumor boundaries). The Decoder: UpSampling. Self-Attention includes:

- Attention Gates smartly pick info from the encoder and only use the useful stuff. In U-Net, they stop unnecessary info from getting through from the Encoder and send the important features to the Decoder. They sit just before merging the upstream and downstream layers (Skip Connections).
- Residual connections and self-attention to make the output better. The self-attention is used in two spots in the generator: in the middle of the encoder after the second layer, and in the decoder before the last layer.

c) Skip Connections

Skip Connections pass spatial features between levels. Each level in the Encoder is connected to its match in the Decoder. They also keep the spatial details of the image. Here's what the layers do: The Decoder uses UpSampling, Convolution, and Attention to rebuild the output.

d) The Decoder architecture

The Decoder includes: Enhanced Attention U-Net with Residual Blocks; and Multi-Scale Outputs. It takes a 128×128 image and outputs two images.

- **High-resolution output (fake_high):** The 128×128 (the main one).
  fake_high = Conv(64→3, kernel=3, padding=1) → Tanh(): 128×128×3
- **Auxiliary low-resolution output (fake_low):** An 64x64 auxiliary version used to calculate the loss at a smaller scale.
  fake_low = (fake_high, size=64×64): 64×64×3

The decoder input is a noisy or low-quality brain image while the decoder output is a perceptually improved medical image. ResBlock is used in both the Encoder and Decoder. Each ResBlock includes:
Input→Conv→BN→ReLU→Conv→BN→Add(input)→ReLU
This let info flow straight from previous layers, makes training more stable, and help pick up deep features.

2) Dual-Discriminator (Local and Global)

Dual-Discriminator uses PatchGAN with Self-Attention to tell real images from fake ones and ensures the images it generates are good. Images are evaluated using local classification maps with both local and global accuracy. A 64×64 crop from the centre is used to check fine details like tumour texture, edges, or tissue boundaries, providing tissue details (Local D). There are two types of Dual-Discriminator, Global and Local, and both use PatchGAN with Self-Attention. Using BatchNorm and LeakyReLU helps speed up and deepen learning. Fig.2 shows the main steps for the Generator and the Dual Global and Local Discriminator in this approach.

a) Global Discriminator (D)

The Global Discriminator takes a 28×128 image and checks it based on its overall structure. It uses:

- PatchGAN: Labels each patch as real or fake.
- Self-Attention: Makes sure the model focuses on tumour structure or brain symmetry.
- Conv layers + LeakyReLU + InstanceNorm: outputs a probability matrix for each patch.

Global D: H=W=128 → PatchMap = 16×16
Local D: H=W=64 → PatchMap = 7×7

b) Local Discriminator (D_local)

The Local Discriminator checks out the center 64×64 patch (local structure). It does this by putting a center crop into the 64x64 image and looking at the local parts. It judges how real the image patches look instead of the whole image. This helps focus on the important details. Plus, it's designed the same as D, just smaller.

c) Self-Attention layers in Discriminator

Self-Attention layers in the Discriminator help it learn overall consistency and structure. They give patch-by-patch probabilities showing if something is real or generated. Usually used after the second or third layer, they focus on small local details. This is handy for checking structural relationships and makes D and D_local better at noticing complex geometric details in the image.

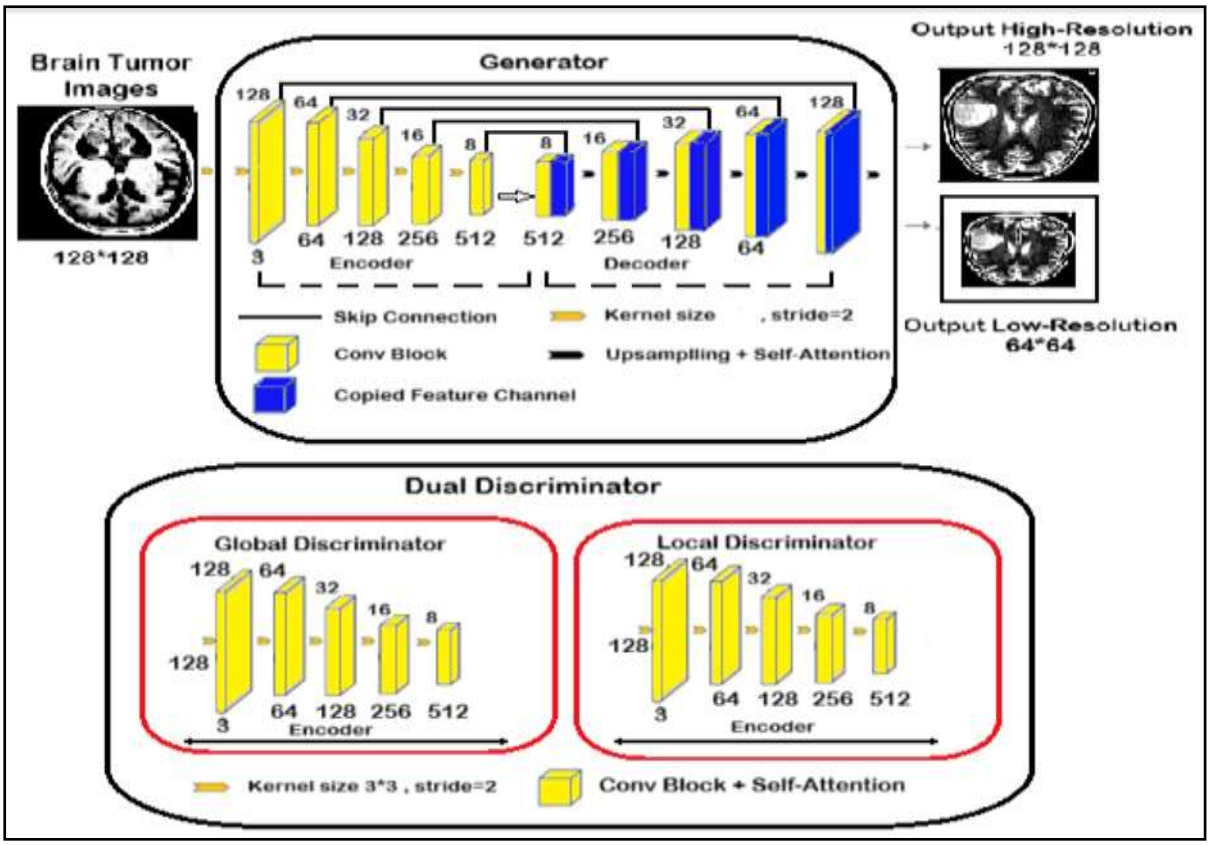

Fig. 2 Generator, Dual Global and Local Discriminator

3) Loss Functions

We tried out different loss functions in our suggested WGAN-GP approach. Table II shows the usual values of the parameters for the functions we used.

- Wasserstein Loss with Gradient Penalty (WGAN-GP) for both discriminators to make training more stable (2-Discriminator Loss, 1-Generator Loss).
- Multi-Scale Perceptual Loss using VGG19 features from different layers.
- Reconstruction Loss (L1).
- Feature Matching Loss (using intermediate features of D and D_local).
- LPIPS Loss: optional perceptual distance for fine textures.
- Structural Consistency Loss: via MSSSIM between generated and real images.

C)Training Procedure Details

This section lays out all the key stuff needed to make sure the improved WGAN-GP model trains stably, efficiently, and with high quality for enhancing MRI Brain Tumor images.

- **Batch Size**: A small, fixed batch size is used for training to make the most of memory and keep adversarial training steady. Usually, smaller batch sizes are more stable for WGAN-GPU and allow more frequent updates.
- **Single D**: Batch size depends on memory limits.
- **Dual D**: Batch size can be 4, 8, or 16 on the GPU. Number of Epochs: The model is trained across multiple Epochs (many options available).
- **Number of Epochs**: The model is trained over several Epochs with various options.
- **Discriminator Update Frequency:** In line with WGAN-GP recommendations, the Discriminator (Critic) is updated more frequently than the Generator (approximal 5 times). This allows the Discriminator to provide more stable gradients before the Generator is updated.
- **Optimizers:** Both Generator and Discriminator are trained using the Adam optimizer with WGAN-GP-specific parameters: Discriminator: Updated 5 times per generator step; Optimizer: Adam with $\beta1=0.5$, $\beta2=0.999$, LR=0.0002; and Learning Rate: Initially 0.0002, reduced using CosineAnnealingLR scheduler.

- **Learning Rate Scheduler:** To gradually reduce the learning rate and to stabilize long training and avoid overshooting. CosineAnnealingLR is used: Scheduler = 100. It reduces the learning rate over time in a cosine pattern to allow smooth convergence.
- **Stability Enhancements:** Label Smoothing is used for training stability. Rather than using labels of 1.0 for real images, a smoothed value is used: Real Label= 0.9. This helps prevent the Discriminator from becoming too confident. Real image labels are set to 0.9 instead of 1.0 to reduce overfitting. Label Noise: Also used for training stability in some cases, randomly flipping a small percentage of labels is used to inject noise and improve generalization.
- **Data Augmentation:** many augmentation techniques were applied to increase robustness and prevent overfitting: Random rotations (±15–30°); Horizontal and vertical flips; Random brightness/contrast changes; Elastic deformation; Cropping and resizing; and Gaussian noise injection.
- **Evaluation Metrics:** Performance is measured every 10 epochs using: PSNR (Peak Signal-to-Noise Ratio) [47]; LPIPS (Learned Perceptual Image Patch Similarity); FID (Frechet Inception Distance) [48]; and SSIM (Structural Similarity Index) [49].
- **Output Image Saving:** generated images per epoch are saved for qualitative assessment.
- **Best Model Saving:** The Generator is monitored and saved based on PSNR performance:
- **Update weight:** Update Dual D for reduced D^L; and update G for reduced G^L.

TABLE II TYPICAL VALUES OF THE USED FUNCTIONS PARAMETERS

Parameters	Typical value
Wasserstein	1.0
Gradient Penalty Loos to control training stability of WGAN-GP	10
L1 (Reconstruction Loss)	1
Perceptual Loss (Lperc) that depend on quality of the images required.	0.1 – 1
Feature Matching Loos (LFM)	10
Structural Loos (Lstruct)	1.0
Image size (real image of the human Brain) and it is depended on image size.	128 x128
Number of layers of VGG19 used. This is used to extract multi-level features.	conv1_2, conv2_2, conv3_4
Learning. Common settings for training a GAN	Adam Optimizer, lr = 1e-4
Batch size	4, 8,16
Epoch	Manu option
Execution environment	GPU
No. of Layers(G)	200
No. of Layers(D)	D=100.D Local=100

V. RESULTS AND DISCUSSION

Many experiments were conducted in this research. All of them were run on a PC with an Intel(R) Core (TM) i5-10210U CPU, 64 GB of RAM, and 10 GB NVIDIA RTX 3080 graphics card. The model was built using Python 3.313 and PyTorch version 2.1.0, with CUDA 11.8 for GPU-accelerated training. Training was done locally using custom setup that included all needed dependencies and files. Different metrics were used to check this model (FID, SSIM, LPIPS, and PSNR). The FID (Frechet Inception Distance) is a solid metric for checking the quality of medical images made by GAN models compared to real ones, as shown in Eq. (12).

It looks at how much the statistical distribution of generated image features matches that of real image features. It checks similarity just at pixel level [48]. Steps to get FID:

- Feature extraction: Get features from real and generated images.
- Distribution calculation: Work out the mean and covariance of the image features.
- Distance measurement: FID work out distance between two feature distributions using Frechet distance.

To check the results, if the values are under 50, they're seen as very good in some cases. But for some, values under 10 are considered excellent. When the value gets close to 0, the generated images are seen as very realistic. On the other hand, if the value is high, it means the generated images are quite different from real ones and their quality is poor [48].

$$FID = \|m - m_w\|_2^2 + T_r\ \left(C + C_w - 2(CC_w)^{1/2}\right) \qquad (12)$$

The SSIM (Structural Similarity Index) analysis is important because it measures quality from a perceptual perspective. When SSIM value close to 1, this is indicating strong structural similarity between the original and generated images. If the SSIM is steadily increasing, reaching values of 0.90 or higher, for example, it indicates excellent performance.

This metric compares images based on structure and composition, not just the numerical value of pixels. It focuses on the similarity of shapes and details in images. Its value ranged from 0 to 1.

A value of 1 indicates a perfect match between the two images. The closer the value is to 1, the higher the structural similarity—that is, the generated images resemble the originals in terms of detail and [49].

Many experiment results were tested on the proposed WGAN-GP model using various optimisations in a GPU environment and compared with each other. It's also known that the original WGAN-GP network uses UNET, which is common in medical image processing, especially for optimization and segmentation. It has an encoder and a decoder path, with skip connections between matching layers, helping preserve fine details when reconstructing the image. During the experiments, UNet was swapped with Attention U-Net, which focuses on the most important parts of the image rather than processing all the information, reducing noise.

Attention units are used in skip connections to filter out important information. They also boost performance by capturing fine details, especially in tricky or complex areas, which are key for spotting tumor structures. The model development process began to enhance medical images of brain tumor areas using the WGAN-GP approach. This is done with a network that has one generator and one discriminator, using between 30 and 300 images and choosing the number of iterations between 100 and 300.

We ran experiments with several models (M1 to M10) using a batch size of 8, and the M9 model came out as the best with an average PSNR of 27.95 and SSIM of 0.5624. This shows that this model can restore images fairly well, but the visual quality is still limited. The images also have some noise and lose fine structural details, as seen in Fig.3 and Fig.4 for model M8 and Fig.5 and Fig.6 for model M9.

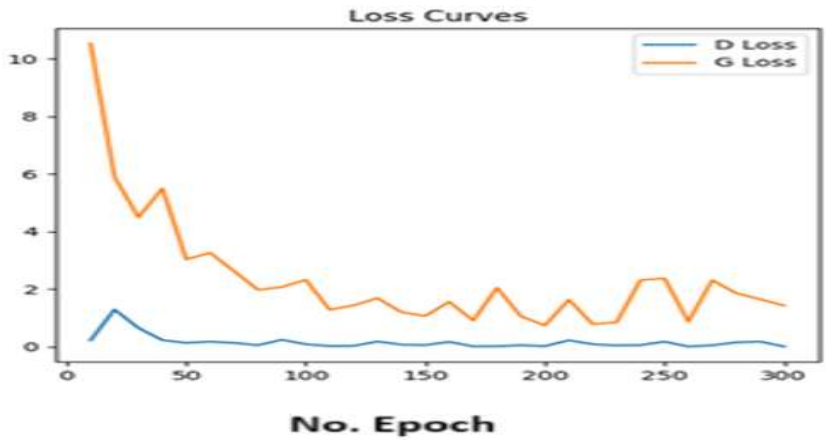

Fig. 3 Loss Curves for Model (M8) of WGAN-GP training

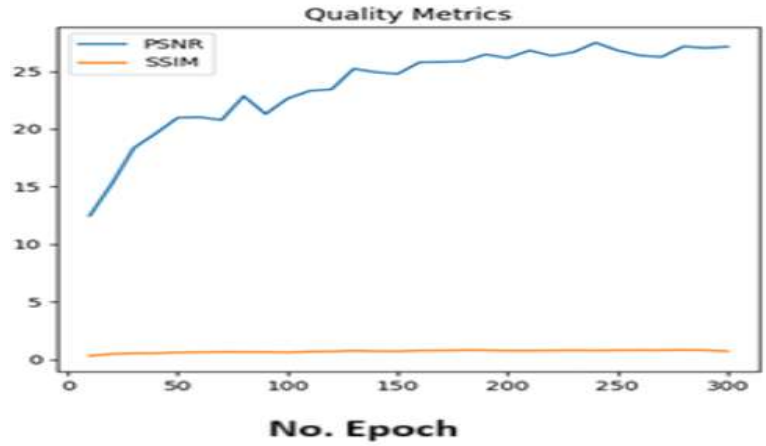

Fig. 4 Quality Metrics for Model (M8) of WGAN-GP training

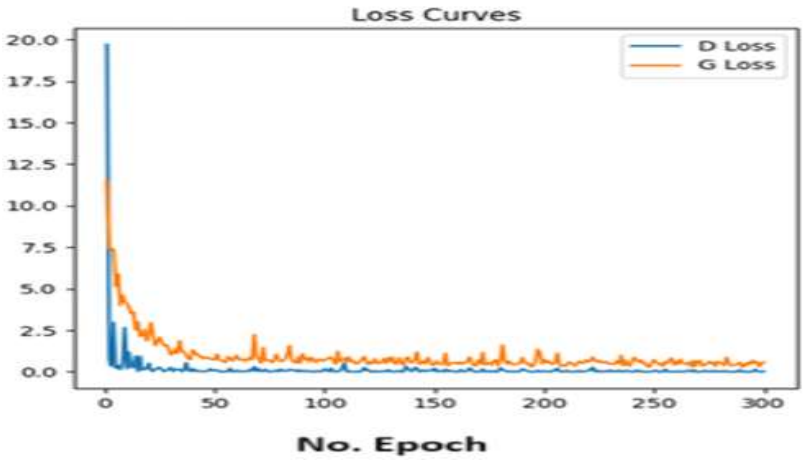

Fig. 5 Loss Curves for Model (M9) of WGAN-GP training

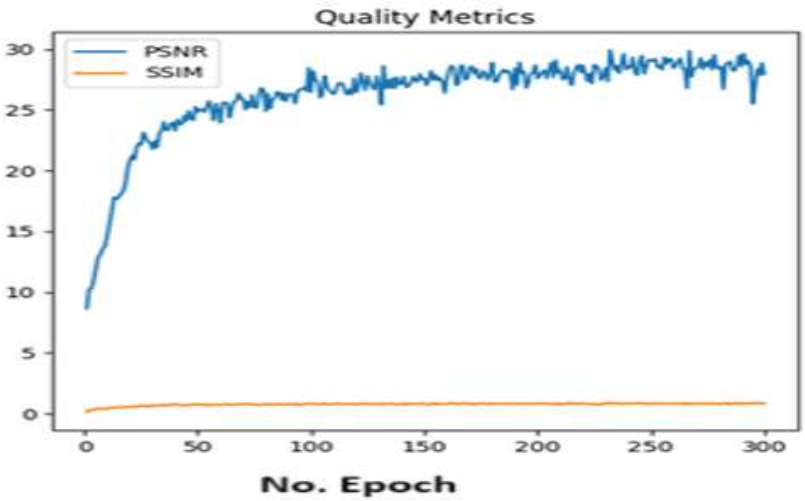

Fig. 6 Quality metrics for Model (M9) of WGAN-GP training

During this process, we noticed some issues, like oscillations in the training curves, poor edge quality, and fine details of the tumor region in the generated images, plus instability in the model's output. These problems were down to the network architecture, not enough used losses, and relying on just one discriminator for global image evaluation. So, we changed the network architecture to make it better and show its effect on image enhancement. We did this by swapping the generator with a Self-Attention U-Net Network. Then we added Attention Gates layers with Residual Blocks in the U-Net structure. This led to a big improvement in the structural details of the tumour regions, with the PSNR going up to 36.76 dB and the SSIM up to 0.9716. On the other hand, adding self-attention to the generator and discriminator helped the model get spatial relationships better. This makes the improved images look more visually coherent, as shown in Fig.7 and Fig.8 for the M10 model.

To make the generated images better and more accurate, a Global Dual Discriminator was used to look at the whole image, while a Local Dual Discriminator focused on the centre of the image (64×64), with a batch size of 4 and the same CPU setup. This helped boost detail resolution in sensitive parts of the brain, giving a big jump in image accuracy. Using Multi-Scale Generator Outputs helped the generator create both a high-resolution image (128×128) and a low-resolution one (64×64). This helps with Multi-Layer Learning and keeps an eye on image quality at different levels and details. This approach balanced losses across multiple scales, reaching a PSNR of 42.22, as shown in Fig.9 and Fig.10 for the M11 model. Table III shows the experimental results of WGAN-GP models for enhancing tumour images. We also used advanced and multiple losses in the multi-level model to boost the quality of the generated image: multi-scale perceptual loss using VGG19 (from multiple layers); feature matching loss using feature layers; and structural consistency loss using MSSSIM.

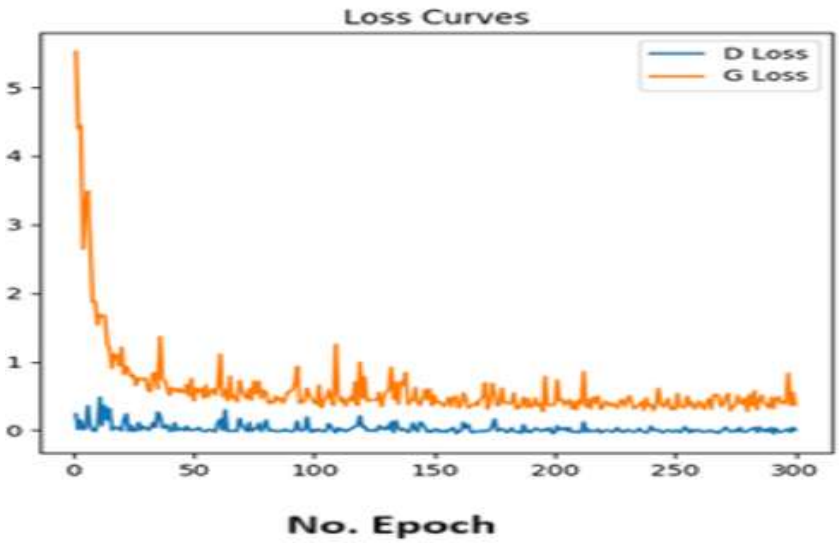

Fig. 7 Loss Curves for Model (M10) of WGAN-GP training

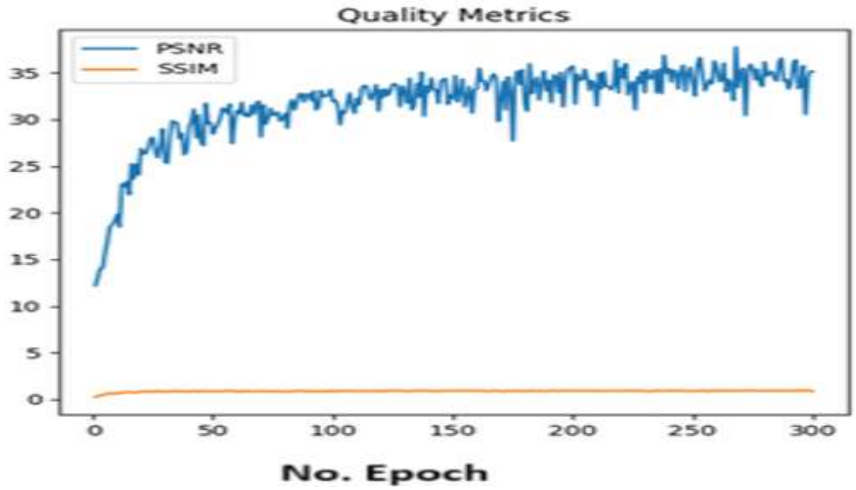

Fig. 8 Quality metrics for Model (M10) of WGAN-GP training

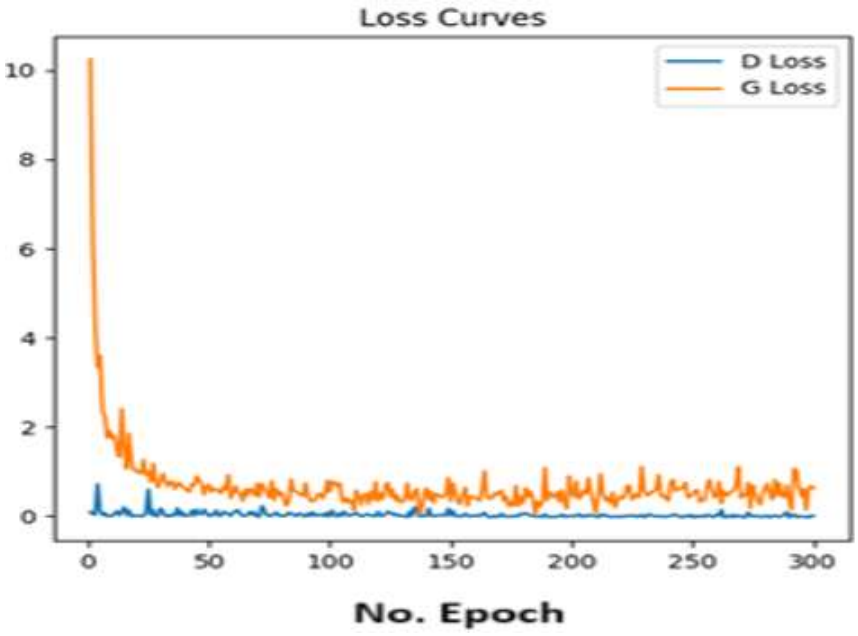

Fig.9 Loss Curves for Model (M11) of WGAN-GP training

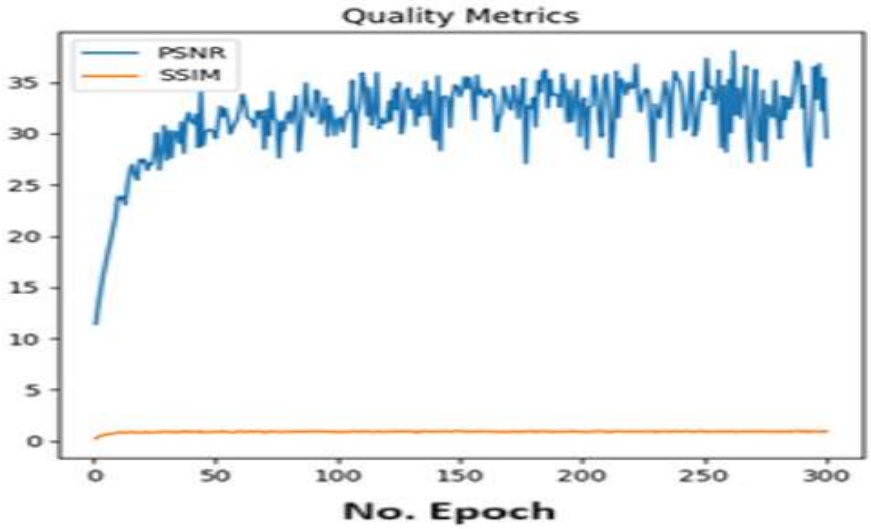

Fig. 10 Quality metrics for Model (M11) of WGAN-GP training

TABLE III RESULTS OF WGAN-GP MODELS FOR IMAGE ENHANCEMENT

Model	No. of images	Epoch	Type D	Batch-size	G-Loos	D-Loos	PSNR	SSIM	FID	LPIPS
M1	30	100	Single D	8	0.534	0.5023	18.8105	0.6121	55.21	0.08543
M2	30	200	=	8	0.5416	0.5182	18.1231	0.5122	53,66	0.08765
M3	30	300	=	8	0.6419	0.6043	17.7018	0.6419	56.54	0.129861
M4	100	100	=	8	0.6182	0.5208	19.2341	0.6132	51.91	0.092754
M5	200	100	=	8	0.6982	0.6541	20.1276	0.6871	52.43	0.074376
M6	200	200	=	8	0.4012	0.1261	23.2871	0.6987	50.98	0.043215
M7	300	300	=	8	0.4325	0.1245	23.8741	0.6731	50.65	0.305432
M8	100	300	=	8	1.4232	0.0006	27.14	0.7187	33.32	0.025435
M9	300	300	=	8	0.0528	0.0212	27.95	0.5624	20.99	0.010873
M10	300	300	=	8	0.5137	0.0382	36.76	0.9716	13.42	0.009088
M11	300	300	Dual D	4	0.4321	0.0016	42.22	0.9255	9.2	0.007032
M12	300	300	=	8	0.4315	0,0023	42.54	0.9213	9.0102	0.007321
M13	1000	300	=	16	0.5432	0.0027	42.42	0.9463	9.0.101	0.006326

On the other hand, the traditional GAN network works by making the generator trick the discriminator into producing images that look real. But the nice thing about using a feature matching loss is that it gives the generator an extra guide, helping it match the deep feature representations from the discriminator with real and generated images. Inside the discriminator, there are several layers that create feature maps

showing different aspects of the image (like edges, shapes, textures, complex structures, fine details), which the discriminator picks up when checking out the real image.
The generator tries to make images so that their feature maps look like the feature maps of real images at different levels in the discriminator. This has led to a bunch of improvements in model M11.
• Better training stability: it helped the generator learn image features so they look 'real',
• Finer details: the generated images not only copied the outer look but also the inner structural features in the discriminator,
• Less Mode Collapse: the generator used to make very similar images, but Feature Matching Loss encouraged more variety and realism, making unnatural images less likely.

When all these improvements were combined in the M11 model, the training began with relatively high PSNR values from the first iterations, reaching a PSNR value of 34 at the beginning of training, that is, in the first 50 epochs. It continued to improve steadily until the PSNR values reached 42.22 and SSIM = 0.9952, which are results that clearly outperformed all previous versions of the proposed models.

On the other hand, when developing the models M1 to M10 using one feature with a batch size of 8 to feed images for training, in the case of model M10 and a training sample size of 300 images, this would pass 37 images per batch. This led to a steady improvement in the PSNR value and made training with the WGAN-GP model more stable. Since the WGAN-GP algorithm relies on computing the GP, which involves calculations for every batch, a small batch size strikes a good balance between training stability and speed. But when we tried using 16 and 32 for the Batch Size, it caused gradient instability during training, so it's better to keep the size small. On the other hand, a Batch Size of 8 improved generalization by updating the weights more often and in a more varied way, helping the model generalize better as it learned from different images in each batch. “Fig.11” shows the M10 model with a Batch Size of 8 for the epochs (300, 290, 170, 150, 110, 80), with gradual improvement from the start of training on a GPU.

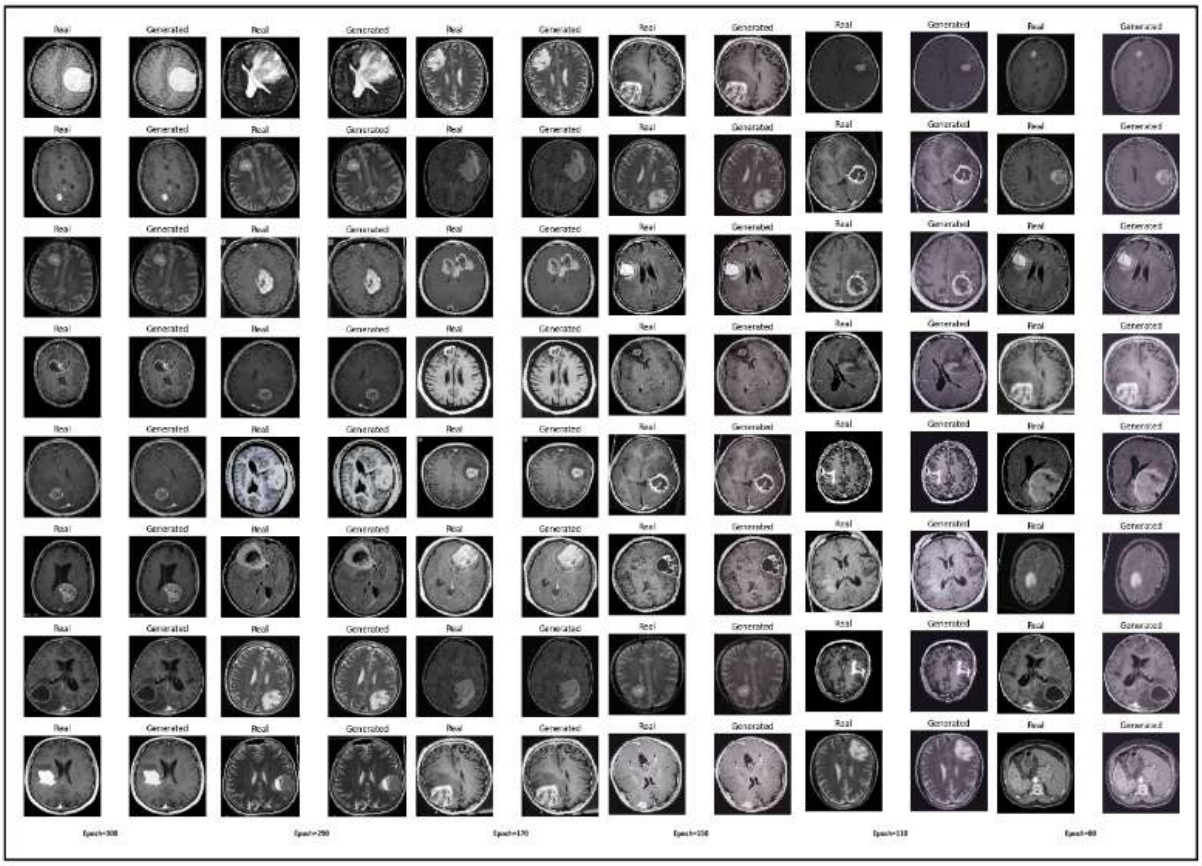

Fig.11 The M10 model with a Batch Size=8 for the epochs (300, 290, 170, 150, 110, 80)

However, when making improvements, we developed model M11 and ramped up computations by using dual discriminators: a global one to check entire images at 128×128 resolution and a local one to check focused areas at 64×64 resolution. This made each training iteration a lot more demanding computationally. So, the batch size was cut down to 4, which improved memory use since adding more discriminators means each image is checked twice, doubling computations and upping memory use,

especially with extra losses like LPIPS, Perceptual Loss and structural consistency loss. This made training better and more stable. Also, smaller batch sizes help make gradient updates for weights smoother, lowering the chance of value spikes or crashes from differences in evaluation by the two discriminators.
On the other hand, this hit a good balance between quality and resources. Even though the batch size was reduced, performance didn't take a hit; in fact, the experiments showed a big boost in metrics like PSNR, which went over 40, showing that this change really helped improve the model's results. So, dropping the batch size to 4 after adding the double discriminators was a smart move that balanced training stability, quality, computational efficiency, and image accuracy. Fig.12 shows model M11 with a batch size of 4 for Epochs=(300, 290, 170, 150, 110, 80) with high PSNR and big improvements from the start of training on a GPU. There was also a clear visual and perceptual upgrade when checking FID.

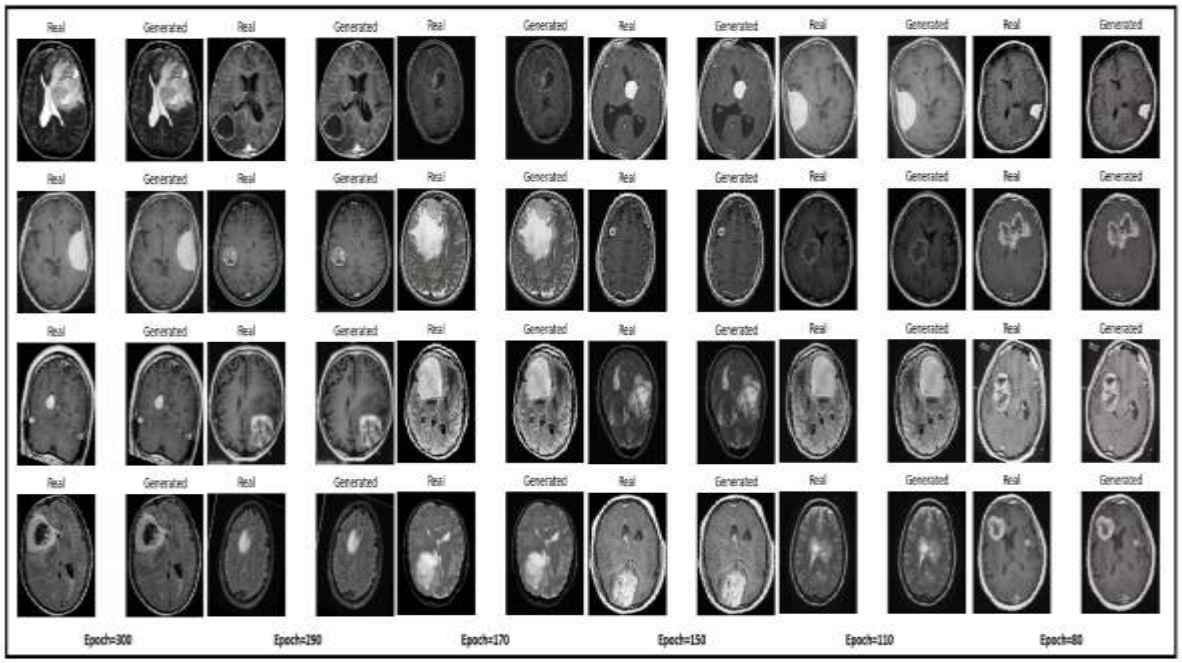

Fig. 12 The M11 model with a Batch Size of 4 for the iterations Epoch=(300,290,170,150,110,80)

When using a batch size of 8 for the M10 model and picking 480 real images and 480 generated images from all iterations, this sample was taken randomly every 10 epochs. Using such a large number of images in the calculation shows that the result is more stable and accurate, reaching FID=13.42. Also, it shows the model's quality across a wide dataset, with the generated images looking very close to the real ones in terms of details and variety. There might be tiny differences that could be improved, but they don't really affect the overall look. Still, these details were better in the M11 model for the same samples. The FID value hit 9.2, showing that the generated images are really high quality and match the original images closely. This shows the model is strong and does a great job at making realistic images, and it can even be used to generate images for diagnosing brain tumors in medical applications.
To check out the M11 model, it was retrained using the same settings as the M12 and M13 models, with batch sizes of 8 and 16. Changing the batch size sped up operations in the WGAN-GP model and made it possible to use features that need heavy computation like Attention Blocks, Perceptual Loss with VGG19, and Dual Discriminator. This didn't really mess with training results or image accuracy. Basically, training was super quick, especially when increasing the training samples from 300 to 1000 images for the same M13 model parameters. "Fig.13" shows the M12 model with a batch size of 8 for epochs (300, 290, 270, 200, 150, 110) with a high PSNR value and big improvement from the start of training on a GPU.

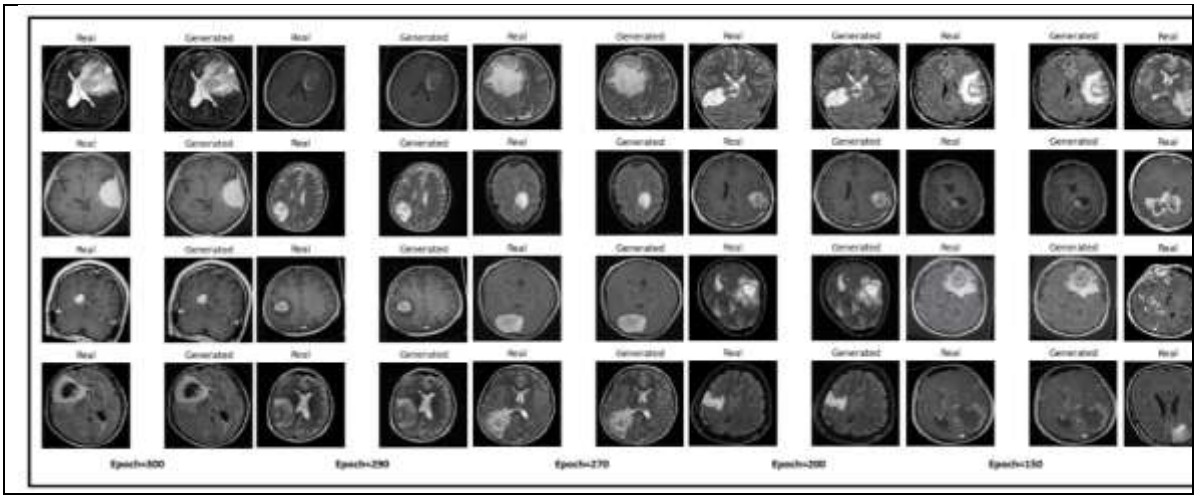

Fig.13 Results of M12 model with a Batch Size of 8 for epochs (300, 290, 270, 200, 150, 110)

To check the results, we tested models M11, M12 and M13 using 100 images that weren't part of the training. They all gave great results, but it's safe to say that model M13 is the best for quickly generating real images with a GPU and BatchSize=16, as shown in Fig.14. This helps speed up diagnosis and treatment since speed is key for this type of disease, with a high PSNR and low SSIM shown in Fig.15 for all 100 test images.

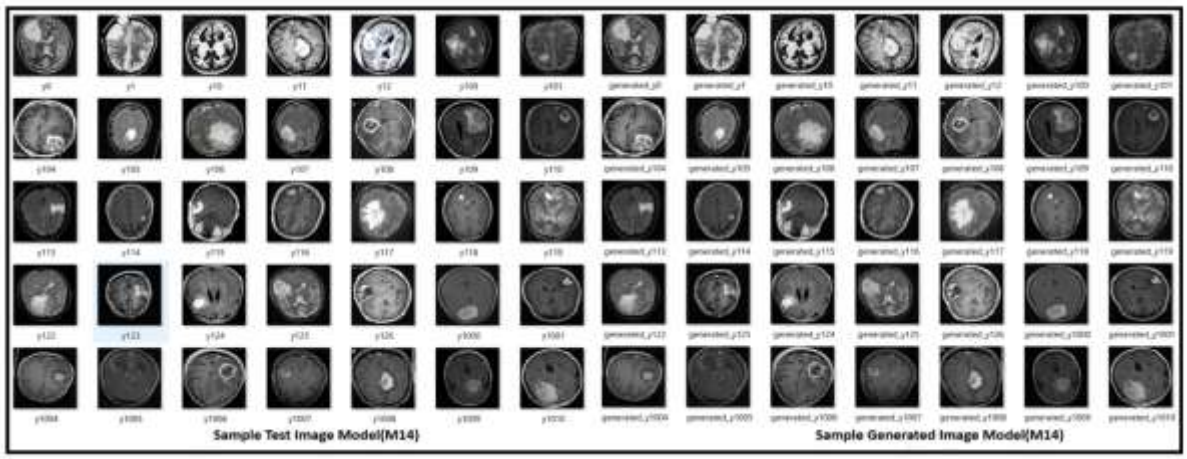

Fig. 14 Testing Images Samples from Model (M13) with a Batch Size of 16

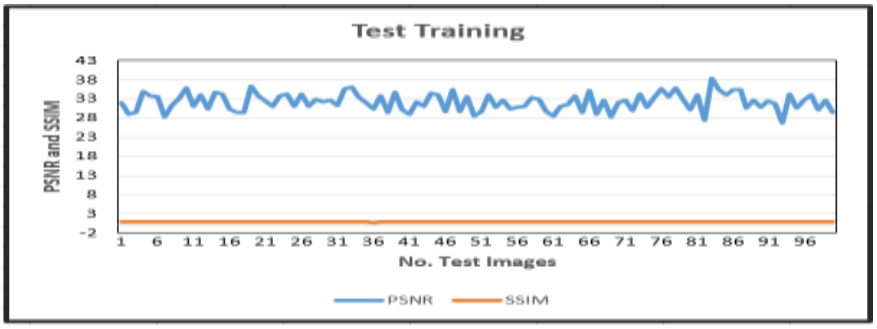

Fig.15 PSNR and SSIM for 100 Testing Images Samples

The SSIM values range from 0 to 1, which means SSIM keeps increasing with training. This shows that the model is getting better at the details and shapes of images, getting closer to the quality of real images visually. At the same time, the LPIPS values decrease as training goes on, meaning the model generates images that look more like what a human would see in a real image. Fig.15 shows the SSIM and LPIPS values over the epochs. Fig.16 shows two test image samples individually to clarify model M13. Also, "Fig.17" shows three test image samples individually for model M13. And "Fig.18" shows three test image samples individually for model M13.

As a result, we can say that our suggested Wasserstein GAN with Gradient Penalty approach can be used to achieve high-strength MRI enhancement for early Brain Tumor diagnosis. Table IV shows the model type, main specifications, used data base, strengths and limitations related to different related studies and our suggested approach.

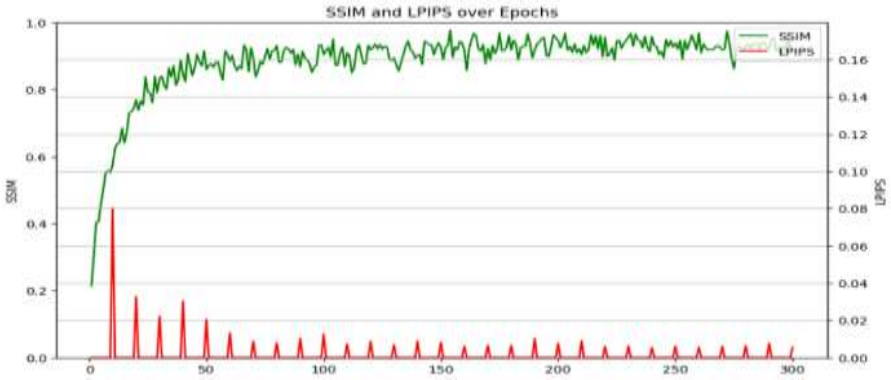

Fig.16 Values of SSIM and LPIPS over the Epochs.

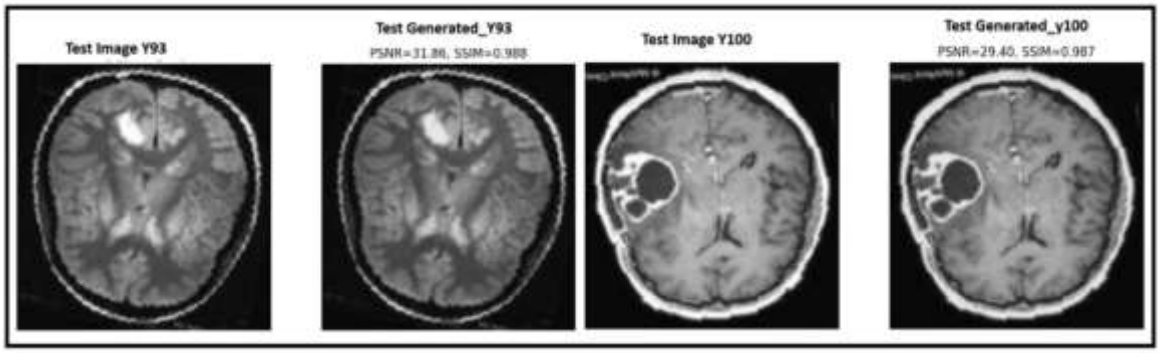

Fig.17 Two Testing Image Samples of are individual for clarification of Model (M13).

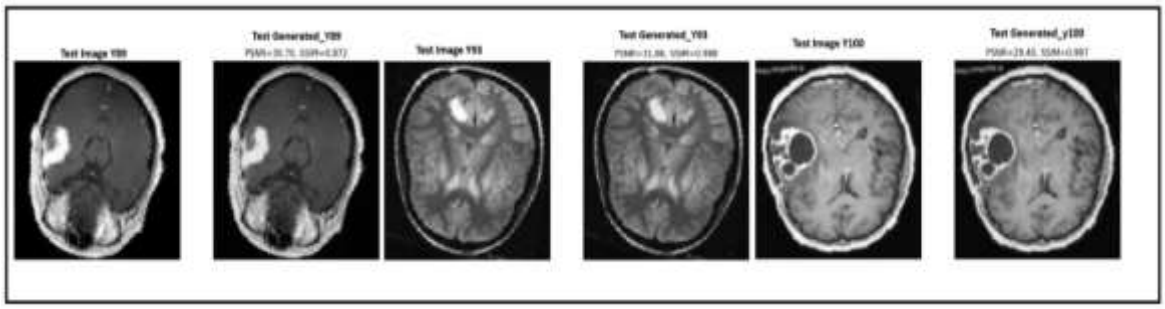

Fig.18 Three Testing Image Samples of are individual for clarification of Model (M13).

TABLE IV DIFFERENCES BETWEEN REFERENCES AND OUR APPROACH

Reference	Model Type	Main Idea	Used Database	Results	Limitations	Used Attention
[11]	CycleGAN (unpaired image-to-image translation).	Synthesize one modality (MR) from another (CT) using unpaired data to expand training sets for	Cardiac MR and CT images from different patients (unpaired).	-15% accuracy improvement in segmentation when training with real and synthetic data.	-They cannot directly evaluate synthetic images due to lack of ground truth. -Limited to cardiac datasets.	Not used Attention

Reference	Model Type	Main Idea	Used Database	Results	Limitations	Used Attention
		segmentation tasks.		-Synthetic images alone achieve accuracy with 95% of real data model performance.	-Require post-processing for segmentation refinement.	
[13]	GAN for image synthesis (MRI brain tumors).	Generate synthetic MRI images for data augmentation and anonymization to overcome class imbalance.	BRATS (2015) and ADNI Two public brain MRI datasets.	-Improve tumour segmentation performance using synthetic images. -Synthetic data can replace real data with minimal loss in accuracy. - Accuracy=96 %.	-Synthetic image realism may vary. -Limited testing on other imaging modalities. -Potential domain gap between real and synthetic data.	Not used Attention
[19]	GAN	Create artificial realistic data for MRI brain tumor images. Data Augmentation.	Contrast-Enhanced MRI (CE-MRI), Four classes collected from Chinese hospitals (2005-2020).	-The model achieved accuracy that surpasses previous methods in generating improved brain tumour images.	-They didn't assess image quality using metrics (SSIM, FID, PSNR). -They didn't test the impact of generated images on classification or segmentation.	Not used Attention
[20]	Improved 2D U-Net (CNN)	Improving brain tumor segmentation using 2D U-Net with removal of irrelevant background to reduce execution time.	BraTS 2017, BraTS 2018, BraTS 2019, BraTS 2020.	-Good segmenting brain tumors from MRI images. -Remove irrelevant background to reduce computation time, and improve extraction of different tumour areas. -Good accuracy.	-They focused on 2D images and didn't take advantage of the 3D spatial relationships between slices.	Not used Attention
[21]	Dual CNN	Classifying three types of brain tumors	Dataset of 3064 images with three	-The model achieved accuracy	-The model is limited to classification	Not used

Reference	Model Type	Main Idea	Used Database	Results	Limitations	Used Attention
	(VGG-16 and Custom CNN).	using a DCTN model based on merging VGG-16 with a customized CNN.	types of brain tumors (Glioma, meningioma, and pituitary).	during training and testing after about 22 experiments using different architectures.	with no comprehensive evaluation metrics and no checking against external databases.	Attention
[24]	Cascade CNN (C-CNN) with Distance -Wise Attention.	Segment brain tumors from multi-model MRI efficiently using attention mechanism focusing on tumor regions.	BRATS (2018) multi-model MRI dataset (T1, T1c, T2, FLAIR).	-Dice scores, whole Tumour = 0.9203. - Enhance Tumour (0.9113). -Tumour Core (0.8726). - Accuracy=98 %.	-The model requires multi-model MRI input. -Limited to BRATS 2018 dataset. -Performance depends on preprocessing accuracy.	Distance -Wise Attention (DWA)
[30]	Hybrid GAN (CNN + Vision Transformer)	LMRT-NET model to generate multi-modal medical images using latent feature space with multi-scale transformers and multi-level information fusion.	Spinal Disease, BraTS2020 (Multi-Contrast Brain Tumor MRI Dataset), Multi-Model Pelvic MRI-CT Dataset (MalePelvis).	-The model achieved higher quality in image generation and better generalization compared to the latest methods. -PSNR=30.62. -SSIM=0.634.	-The model needs multi-model data. -The structure is complex and relies on ViTs, which increases training requirements.	Multi-scale Residual Vision Transformer – DART Blocks
[31]	CNN	Developing an automated framework to classify brain tumors from MRI images using CNN with decision interpretation through Grad-CAM.	3060 MRI Brain Images.	-The model achieved better accuracy and classification with increased transparency using Grad-CAM compared to traditional methods.	-The model relies on a single database and it is limited to classification without generation or segmentation.	Not used Attention. Grad-CAM is used for interpretation, not Attention inside the model.
[34]	GAN and Teacher – Student Network	Unsupervised translation of multimodal brain image focusing on	BRATS 2020 dataset of brain tumor images of four	-Improve quality of generated images,	-The model relies on tumour masks during network training.	Not used Attention.

Reference	Model Type	Main Idea	Used Database	Results	Limitations	Used Attention
		tumor areas using knowledge distillation.	modalities: Flair, T1, T1ce and T2 Unpaired Multi-modal MRI.	-Get competitive results. -Enhance segmentation performance.	-The training is more complicated because there are two networks.	Tumor-aware mechanism and not Attention Module
[35]	GAN, Autoencoder, Transfer Learning (DenseNet)	Improving brain tumor segmentation using GAN with Autoencoder and DenseNet.	BraTS 2021 for MRI brain Tumor image (1251 MR images).	-Improve tumour segmentation accuracy compared to traditional CNN-based methods.	-Instability of GAN training. -The model needs large amount of data. -The model affected by differences in tumour characteristics.	Not used Attention
[36]	Hybrid GAN + CNN	Using GAN to generate synthetic MRI images and improve CNN training for brain tumor classification.	Kaggle Brain MRI Dataset (2022), 100 images (75 tumors, 25 healthy).	-Results showed that the hybrid model has better classification accuracy compared to traditional methods.	-Small database (no standard DB like BraTS) -No segmentation mechanisms.	Not used Attention
Our Paper	WGAN-GP with Self-Attention One Generator, Two discriminators to generate and enhance MRI	Enhancement of generated MRI Brain Tumor.	Br35H (2020) MRI datasets (1000 sample).	-Enhancement of generated MRI Brain Tumor. -PSNR = 42.42 dB -Accuracy= 99.634%.	-Lack of real patient images. So, we conducted our study on Br35H (2020) dataset. -The model require laptops with high performance specifications.	Self-Attention in D and G

VI. CONCLUSION

Catching brain tumors early is really important because the symptoms can be pretty vague, which can delay diagnosis. MRI scans can get messed up by patient movement, changes in the body, and need experts to interpret them. Generative Adversarial Networks (GANs) have become a great way to boost the quality of medical images for brain cancer diagnosis and fill in missing data. So, in this work, we're looking to improve the quality of MRI images of brain tumors using a fancy deep learning model based on GAN, specifically the Wasserstein GAN with Gradient Penalty (WGAN-GP) with a generator G and Dual Discriminator D (Dual D). This method is used to make grayscale images from the Br35H (2020) dataset, which has medical brain tumor images, clearer and more accurate. We ran loads of experiments with different network architectures and learning settings in this study.

The cumulative improvements in GAN architecture, losses, dual features, and output resolution variety have really boosted the quality of generated medical images. The big jump in PSNR from the first iterations after the latest updates shows that the M13 model has started learning better and more efficiently right from the start. You could say this improved model is a strong and promising setup for making high-quality medical images and can be trusted for clinical diagnostics or to back up other models.

After adding all the upgrades, we noticed that the performance curves kicked off with high values right from the start, unlike the baseline model which took dozens of iterations to hit average results. This shows that the improved model's architecture and its losses were set up strongly from the get-go, showing in great early results. These results show that the step-by-step improvement approach based on solid architecture and picking advanced loss functions has really boosted the model's performance.

The final model creates top-notch images with a perceptual and anatomical accuracy of 99.634% and a PSNR of 42.42 dB. This makes it great for medical image analysis, keeping things precise and safe. In future work, we'll use Deep Learning with upgraded versions of GANs that are better at generalising features to boost resolution and enhance coloured MRI images for Brain Tumour diagnosis.

REFERENCES

[1] Wei K, Kong W, Liu L, Wang J, Li B, Zhao B, Li Z, Zhu J, and Yu G. CT, "CT Synthesis from MR Images Using Frequency Attention Conditional Generative Adversarial Network," Computers in Biology and Medicine, vol.170, 107983, March 2024, PMID: 38286104, DOI: 10.1016/j.compbiomed.2024.107983

[2] Xiao-Jiao Maoy, Chunhua Shen, and Yu-Bin Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in Proc. 30th Conference on Neural Information Processing Systems (NIPS), pp. 2810 – 2818, 5th December 2016, Barcelona, Spain, DOI: 10.5555/3157382.3157412

[3] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein Generative Adversarial Networks," in Proc. 34th International Conference in Machine Learning (ICML), vol. 70, pp. 214–223, 6th August 2017, DOI: 10.5555/3305381.3305404

[4] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron Courville, "Improved Training of Wasserstein GANs," in Proc. 31th International Conference on Neural Information Processing Systems (NIPS'17), pp. 5769 – 5779, 25 December 2017, DOI:10.48550/arXiv.1704.00028

[5] Yuansan Liu, Jeygopi Panisilvam, Peter Dower, Sejeong Kim, and James Bailey, "Bidirectional adversarial autoencoders for high precision eta surface design," APL Machine Learning, Vol.3, 036110, August 2025, DOI: 10.1063/5.0284338

[6] Christopher X. Ren, Amanda Ziemann, James Theiler, and Alice M.S. Durieux, "Cycle-Consistent Adversarial Networks for Realistic Pervasive Change Generation in Remote Sensing Imagery," in Proc IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI), 29-31 March 2020, Albuquerque, NM, USA, DOI: 10.1109/SSIAI49293.2020.9094603

[7] Phillip Isola, Jun-Yan Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21-26 July 2017, IEEE Xplore: 09 Nov 2017, ISSN: 1063-6919, DOI: 10.1109/CVPR.2017.632

[8] Showrov Islam, Md. Tarek Aziz, Hadiur Rahman Nabil, Jamin Rahman Jim, M. F. Mridha, Md. Mohsin Kabir, Nobuyoshi Asai, and Jungpil Shin, "Generative Adversarial Networks (GANs) in Medical Imaging: Advancements, Applications, and Challenges," IEEE Access, Vol.12, pp. 35728 – 35753, 26 February 2024, DOI: 10.1109/ACCESS.2024.3370848

[9] Jabbar Hussain, M. Båth, and J. Ivarsson, "Generative adversarial networks in medical image reconstruction: A systematic literature review," Computers in Biology and Medicine, Vol.191, June 2025, 110094, DOI: 10.1016/j.compbiomed.2025.110094

[10] Chethan B K, Sumukh Gowda B U, A Yashwantha, Rakesh R, and K Krupank, "A Comparative Analysis of GAN Architectures for Brain MRI Applications," International Journal for Multidisciplinary Research (IJFMR), Vol.7, Issue 2, March-April 2025, E-ISSN: 2582-2160, DOI: 10.36948/ijfmr.2025.v07i02.38144

[11] Agisilaos Chartsias, T. Joyce, R. Dharmakumar, and S. A. Tsaftaris, Adversarial Image Synthesis for Unpaired Multi-modal Cardiac Data: Simulation and Synthesis in Medical Imaging, Lecture Notes in Computer Science, Springer, Cham, 2017, vol.10557, DOI: https://doi.org/10.1007/978-3-319-68127-6_1

[12] Guan Wang and Xueli Hu, "Low-dose CT denoising using a Progressive Wasserstein generative adversarial network," Computers in Biology and Medicine, Vol.135, August 2021, 104625, DOI: 10.1016/j.compbiomed.2021.104625

[13] Hoo-Chang Shin, et al, "Medical image synthesis for data augmentation and anonymization using generative adversarial networks," in Proc. Simulation and Synthesis in Medical Imaging: 3rd International Workshop (SASHIMI 2018), Granada, Spain, 16 September 2018, pp.1 – 11, DOI: 10.1007/978-3-030-00536-8_1

[14] Hector Anaya-Sanchez, L. Altamirano-Robles, R. Diaz-Hernandez, and S. Zapotecas-Martinez, "WGAN-GP for Synthetic Retinal Image Generation: Enhancing Sensor-Based Medical Imaging for Classification Models", Sensors, vol.25, no. 167, pp.1-20, December 2024, DOI: 10.3390/s25010167

[15] Anuja Arora, A. Jayal, M. Gupta, P. Mittal, and S. C. Satapathy, "Brain Tumor Segmentation of MRI Images Using Processed Image Driven U-Net Architecture," Computers, Vol.10, no.11, 139, 2021, DOI: 10.3390/computers10110139

[16] M. Krithika alias Anbu Devi, and K. Suganth, "Review of Medical Image Synthesis using GAN Techniques," ITM Web of Conferences 37, 01005, 2021, DOI:10.1051/itmconf/20213701005 ICITSD-2021

[17] Guojin Zhong, Weiping Ding, Long Chen, Yingxu Wang, and Yu-Feng Yu, "Multi-Scale Attention Generative Adversarial Network for Medical Image Enhancement," IEEE Transactions on Emerging Topics in Computational Intelligence, Vol.7, Issue.4, pp: 1113 –1125, August 2023, 03 March 2023, DOI: 10.1109/TETCI.2023.3243920

[18] Qihong Yang, Ruijun Jing, and Jiliang Mu, "Multi-Modal MR Image Segmentation Strategy for Brain Tumors Based on Domain Adaptation," Computers, Vol.13, no.12, 347, 2024, DOI: 10.3390/computers13120347

[19] Abdullah A. Asiri, et al, "Multi-Level Deep Generative Adversarial Networks for Brain Tumor Classification on Magnetic Resonance Images," Intelligent Automation & Soft Computing, vol.36, no.1, pp.127-143, 2023, DOI:10.32604/iasc.2023.032391

[20] MD Abdullah Al Nasim, Abdullah Al Munem, and Maksuda Islam, "Brain Tumor Segmentation using Enhanced U-Net Model with Empirical Analysis," in Proc. 25th International Conference on Computer and Information Technology (ICCIT), December 2022, DOI: 10.1109/ICCIT57492.2022.10054934

[21] Aya M. Al-Zoghby, Esraa M.K. Al-Awadly, Ahmad Moawad, Noura Yehia, and Ahmed I. Ebada, "Dual Deep CNN for Tumor Brain Classification," Diagnostics, Vol.13, 2023, 2050. DOI:10.3390/diagnostics13122050

[22] Yuan Cao and Yinglei Song, "New Approach for Brain Tumor Segmentation Based on Gabor Convolution and Attention Mechanism," Applied Sciences, Vol.14, No.11, 4919, 2024, DOI: 10.3390/app14114919

[23] Zhenmou Yuan, et. al, "SARA-GAN: Self-Attention and Relative Average Discriminator Based Generative Adversarial Networks for Fast Compressed Sensing MRI Reconstruction," Frontiers in Neuroinformatics, Vol.14, 1 Nov 2020, Article 611666, DOI: 10.3389/fninf.2020.611666

[24] Ramin Ranjbarzadeh, et. al, "Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images," Scientific Reports, Vol.11, 10930, 2021, DOI: 10.1038/s41598-021-90428-8

[25] Y Skandarani, P Jodoin and A Lalande, "GANs for Medical Image Synthesis: An Empirical Study," Journal of Imaging, Vol.9, no.69, 2023, DOI: 10.3390/jimaging9030069

[26] Qian Wang, Han Liu, Guo Xie, and Youmin Zhang, "Image Denoising Using an Improved Generative Adversarial Network with Wasserstein Distance," in Proc. 40th Chinese Control Conference, 26-28th July 2021, Shanghai, China, DOI: 10.23919/CCC52363.2021.9550033

[27] Hugo Carlesso, et. al, "GeMix: Conditional GAN-Based Mixup for Improved Medical Image Augmentation," arXivLabs, July 2025, DOI: 10.48550/arXiv.2507.15577

[28] Hajar Emami, Ming Dong and Carri Glide-Hurst, "Attention-Guided Generative Adversarial Network to Address Atypical Anatomy in Synthetic CT Generation," in Proc. IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI), August 2020, DOI: 10.1109/IRI49571.2020.00034

[29] Yanling Wang, et. al, "Scale-Sensitive Generative Adversarial Network for Low-Dose CT Image Denoising," IEEE Engineering in Medicine and Biology Society Section, IEEE Access, Vol.12, pp. 98693- 98706, 24 July 2024, DOI: 10.1109/ACCESS.2024.3425606

[30] Xinmiao Zhu and Yang Li, "A Latent Multi-Scale Residual Transformer Approach for Cross-Modal Medical Image Synthesis," IEEE Access, Vol.13, pp. 58351-58363, 28 March 2025, DOI: 10.1109/ACCESS.2025.3555879

[31] Idriss Teledjieu, Irzum Shafique, and Muhammad Khan, "Improving Classification Performance in Brain Tumor Based on Convolutional Neural Networks," International Journal of Computer Sciences and Engineering, Vol.13, Issue.7, pp.10-17, July 2025, DOI: 10.26438/ijcse/v13i7.1017

[32] Gustav Müller-Franzes, et al. "A multimodal comparison of latent denoising diffusion probabilistic models and generative adversarial networks for medical image synthesis," Scientific Reports, Vol. 13, no.12098, 26 July 2023, DOI: 10.1038/s41598-023-39278-0.
[33] Ashfak Yeafi, M. Islam, S. Mondal, K. Nashad, and Md. Yusuf, "A Semi-supervised Approach For Brain Tumor Classification Using Wasserstein Generative Adversarial Network with Gradient Penalty," in Proc. 6th Int. Conf on Electrical Information and Communication Technology (EICT), 07-09 December 2023, Khulna, Bangladesh, IEEE Xplore: 13 Feb 2024, DOI: 10.1109/EICT61409.2023.10427898
[34] Chuan Huang, Jia Wei, and Rui Li, "Unsupervised Tumor-Aware Distillation for Multi-Modal Brain Image Translation," in Proc. International Joint Conference on Neural Networks (IJCNN), 30 June 2024 - 05 July 2024, Yokohama, Japan, IEEE Xplore, 09 September 2024, DOI: 10.1109/IJCNN60899.2024.10651523
[35] Abid Ali, et al, "Brain Tumor Segmentation Using Generative Adversarial Networks," IEEE Access, vol. 12, pp. 183525-183541, 27th August 2024, DOI: 10.1109/ACCESS.2024.3450593.
[36] Monia A. A. Al-Hobishi and Muhammed F. Abdullah, "Hybrid GAN-CNN Model for Brain Tumor Detecting and Classifying Diseases Based on MRI Images," Journal of Science and Technology, Vol.30, no.8, 2025, DOI: 10.20428/jst.v30i8.2995
[37] Haotian Wang, "Comparative Analysis of GANs and Diffusion Models in Image Generation," Highlights in Science Engineering and Technology, Vol. 120, 2024, in Proc. 5th International Conference on Automation, Mechanical Control and Computational Engineering (AMCCE), pp.59-66, December 2024, DOI: 10.54097/9gba9v27
[38] Muhammad Yaseen, Maisam Ali, Sikandar Ali, and Hee-Cheol Kim, "Diffusion-Based Approaches for Medical Image Segmentation: An In-Depth Review," Electronics, Vol.15, no. 7, 1400, 27 March 2026, DOI: 10.3390/electronics15071400
[39] C. Villani, Optimal transport: old and new, Vol.338, Springer Science & Business Media, 2008.
[40] Justin Johnson, Alexandre Alahi, Li Fei-Fei, "Perceptual Losses for Real-Time Style Transfer and Super-Resolution," in Proc. European Conference on Computer Vision, Lecture Notes in Computer Science 9906:694-711, October 2016, DOI:10.1007/978-3-319-46475-6_43
[41] Wang Xintao, et al, "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks," in Proc. Computer Vision (ECCV 2018) Workshops, Lecture Notes in Computer Science (2019), vol 11133. Springer, Cham. DOI: 10.1007/978-3-030-11021-5_5
[42] Richard Zhang, et al, "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric," in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18-23 June 2018, pp.586-595, IEEE Xplore: 16 Dec 2018, DOI: 10.1109/CVPR.2018.00068.
[43] Ting-Chun Wang, et al, "High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs," in Proc. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 30 Nov 2017, DOI:10.1109/CVPR.2018.00917
[44] Tim Saliman, et, al. "Improved Techniques for Training GANs," in Proc. 30th International Conference on Neural Information Processing Systems (NIPS'16), pp.2234 – 2242, 5-10 Dec 2016, Barcelona, Spain, DOI: 10.5555/3157096.3157346
[45] Hang Zhao, O. Gallo, I. Frosio and J. Kautz, "Loss Functions for Image Restoration with Neural Networks," IEEE Transactions on Computational Imaging, Vol. 3, no1, pp. 47-57, March 2017, DOI: 10.1109/TCI.2016.2644865.
[46] Ahmed Hamada, "Br35H :: Brain Tumor Detection 2020," IEEE DataPort, 2020. [Online]. DOI: 10.21227/t5k6-5r54
[47] Alan Conrad Bovik and Jerry D. Gibson, Handbook of Image and Video Processing, Academic Press, The University of Texas at Austin, Academic Press, Inc., 6277 Sea Harbor Drive Orlando, FL, United States, ISBN:978-0-12-119790-2, 1 March 2000, DOI: doi/book/10.5555/556230
[48] Martin Heusel, et, al. "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium," in Proc. 31th International Conference on Neural Information Processing Systems (NIPS'17), pp.6629 – 6640, Long Beach California, USA, 4-9 Dec 2017, DOI: 10.48550/arXiv.1706.08500
[49] Zhou Wang, et. al, "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing, vol.13, no.4, pp.600-612, April 2004, DOI: 10.1109/TIP.2003.819861