Generating MRI images using style transfer learning

MSc Research Project
Master of Science in Data Analytics

Dnyaneshwari Mahajan
Student ID: X20191138

School of Computing
National College of Ireland

Supervisor: Qurrat UI Ain

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Dnyaneshwari Mahajan
Student ID:	X20191138
Programme:	Master of Science in Data Analytics
Year:	2022-23
Module:	MSc Research Project
Supervisor:	Qurrat UI Ain
Submission Due Date:	15/12/2023
Project Title:	Generating MRI images using style transfer learning
Word Count:	6379
Page Count:	20

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Dnyaneshwari Mahajan
Date:	24th January 2023

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Generating MRI images using style transfer learning

Dnyaneshwari Mahajan
X20191138

Abstract

In the medical field, the availability of medical imaging for practitioners or radiologists is a huge challenge because of the cost of medical imaging and patient data privacy laws. Deep learning has emerged as an important technology in computer vision applications due to technologies like GAN compared to traditional augmentation techniques. State-of-art style transfer (Generative Adversarial Networks) GANs play a crucial role in generating images in the medical field. The aim of this research project is to build a model for generating MRI (Magnetic resonance imaging) images with different contrast levels using a novel approach of combing the CycleGAN framework and customized U-Net segmentation(CycleGAN+UNET).

This research conducted four experiments on 2 publicly available unpaired datasets of T1-styled images and T2-styled images. At the training phase, the Cyclic consistency loss for the final model is 3.75 which is comparatively low than other models. The generator loss and discriminator loss are 0.93 and 0.79 respectively, the values of the generator and discriminator are very close to each other which means they are balanced. The identity loss is 0.20 which is comparatively low than other models and indicates that the reconstructed image is similar to the original image. The result shows that model performance is better when generating same-styled images T1 to T1 and T2 to T2 with maximum SSIM (Structural similarity) scores of 0.87, 0.94 and PSNR (Peak-signal-to-noise-ratio) scores of 33.09db, and 30.72db respectively. Improvement is needed to generating images with different styles images T1 to T2 and T2 to T1 because of low SSIM scores of 0.35 and 0.55 respectively.

Keywords— GAN; CycleGAN; MRI; SSIM; PSNR; U-Net; Deep learning

1 Introduction

MRI is a widely used imaging technique to understand a significant amount of information about diseases in the human body part. Multiple MRI images help doctors to diagnose the patient accurately but due to the cost of data and less availability of data, it can cause a problem. What if only one kind of MRI is necessary and another can be created automatically? Would it not simultaneously improve the treatment plan and saves the patients money and time?

For various issues, several types of MRIs are necessary. It might not be possible to diagnose the issue with one kind of MRI. An additional type of MRI, could improve diagnostic accuracy and improve the treatment of patients(Zhang et al.; 2021).

In this research project, a new approach is proposed to solve the problem of inadequate amount of training data for practitioners and doctors to study different cases of abnormalities in brain MRI images. The deep learning-based model is developed to generate MRI images with different contrast levels using less amount of MRI images. a novel customized tuned model is developed using CycleGAN style transfer framework with Unet-based image segmentation, Adam optimizer, and ReLu activation function. This model generates T1-styled MRI images to T2-styled MRI images and vice-versa. T1 and T2 are two different datasets with brain MRI images with different contrast levels. The attributes of images in both datasets are described in Table 1.

Table 1: MRI Image Attributes.

Type	**T1 style**	**T2 style**
Water	Dark	Very Bright
Fat	Very Bright	Dark
Bone	Dark	Dark
Muscle	Intermediate	Dark
Tumours	Intermediate	Bright

1.1 Background and Motivation

Medical misdiagnosis is a serious problem, but it is also unfortunately common to occur. MRI diagnosis is one of the medical field's most complicated and challenging tasks. Diagnosing MRI images and predicting diseases is not a straightforward binary process such as normal and abnormal. To interpret the scan radiologist and practitioners may often need to use different imaging variations which can significantly improve diagnosis accuracy by providing practitioners with comprehensive knowledge. However, due to patient data privacy laws, it is complicated and costly to access a big dataset of MRI images. This problem motivates this research to build a model for generating MRI images using deep learning techniques to solve the issue of misdiagnosis in the medical domain. Deep learning approaches required a large amount of data to train the model to provide accurate results but in the medical domain due to storage of data and data privacy rules, it is complicated to access datasets. The common approach is to take a small but labelled dataset and augment it to improve the efficiency of a model.

Generative Adversarial Network is a very popular and powerful technology in training deep neural networks(Chong and Ho; 2021). Deep learning is an important technology in the computer field. This technique is used to build medical-related applications. It is hard to achieve great accuracy with a low amount of data when building medical applications based on MR images. Basic data augmentation techniques such as affine, elastic transformation or flipping, and stretching provide limited and unrealistic performance(R and S; 2021). Since medical images are standard and follow strict regulations, this research used a model based on the style transfer learning-based GAN approach.

1.2 Research Question

How well does style transfer learning apply to generate artificial MRI images to provide a solution for misdiagnosis in the medical field?

1.3 Research Objective

The main objective of this research project is to build an optimized model using CycleGAN Generative adversarial network and U-net-based segmentation for generating artificial MRI images. Using the style transfer learning model generates a T2-weighted MRI image from the T1-weighted MRI image and vice-versa. T1-styled and T2-styled MRI images contain different values of attributes such as type of water, bone, fat, muscle, and tumours of brain tissues. A novel approach will be applied to GAN with U-Net segmentation to produce MRI images of different contrast levels to strengthen the quality of disease diagnosis in medical applications. The model is evaluated using SSIM and PSNR scores.

1.4 Report Structure

Continuing further, a remaining document is in the following sections. In section 2, related work on GAN-based techniques, its conclusion and evaluation techniques are discussed. Section 3 comprises the detailed methodology for model development. Section 4 provides the complete design architecture of the model. Section 5 contains the implementation of the final model. Section 6 provides detailed results of an experiment using evaluation techniques. The last section 7 contains the conclusions and future work of the research project.

2 Related Work

In the medical imagining field, the Process of Generating MRI images is very challenging and important. less availability of datasets is a problem for radiologists or practitioners which can cause misdiagnosis of diseases. To provide an effective solution to this problem various research have been proposed in past few years. The brief literature review of related research is discussed below to support the idea of building a model by assessing and improving it. This section helped a research project to build a model for generating MRI images by understanding and evaluating the previous research. The related work section is divided into 4 subsections mentioned as follows.

2.1 CycleGAN is the best method for Data Augmentation of MRI images

Over the last decade, most research is carried out on generating MRI images using various deep-learning techniques.(R and S; 2021),(Mok and Chung; 2019) compared deep convolutional GAN with traditional augmentation techniques such as affine image transformation, and elastic image transformation. The authors (Singh and Raza; 2021) concluded that GAN-based augmentation is adopted well to overcome the disadvantages of the aforementioned techniques because the traditional techniques fail to a diversity of medical data. In the medical field, high-level information such as the size and shape of the tumor in the brain, fat, and water this kind of sensitive contextual information could be affected by using basic augmentation techniques which is very risky.

To generate MRI images different GAN-based frameworks such as DCGAN (Deep convolutional GAN), WGAN (Wasserstein GAN), CGAN(Conditional GAN) and CycleGAN are used(Han et al.; 2018). A framework like CycleGAN is outperformed by others because it contains the additional component in loss function which is cyclic-consistency loss which ensures the backward and forward cyclic consistencies (Tmenova et al.; 2019). (Zhang et al.; 2021) compared CGAN and CycleGAN framework to generate T1-styled MRI images to T2-styled MRI images whether they found that CycleGAN performed better than CGAN because the image generated by CGAN was blurred.

The authors (Zhang et al.; 2021) proposed a model using CycleGAN deep neural network for generating artificial data of angiograms. The two data set are used to transfer style from 1 dataset to another and vice-versa to generate the data with different contrast level. The authors built a model and trained for 200 epochs which shows a good result with a 0.906 SSIM score and 31.303 PSNR score. This paper did not apply any optimization function which could help to increase the performance of the model.

The authors (Xu et al.; 2019) proposed a model using semi-supervised guided CycleGAN with Res-Net18 segmentation. The generator of the module is trained by unsupervised adversarial loss and supervised pixel loss. The result is compared with the baseline model and the authors concluded that semi-supervised works great to generate realistic tumour images from normal images and vice-versa with boosting 9.78 % sensitivity.

This section helped this research project choose the CycleGAN framework for generating MRI images over other aforementioned data augmentation techniques. CycleGAN performs well than other data augmentation techniques because it contains cyclic-consistency loss which enforces cyclic consistency and prevents the generator to generate fake images (Singh and Raza; 2021). Also to improve the performance of the model use of optimization functions, the image segmentation model and other factors are important which are discussed in the next section.

2.2 The need for the best Image segmentation and Optimization Technique

The U-Net is a widely used technique for medical image segmentation. The structure of the U-net is an auto-encoder where the encoder extracts the features of an image and the decoder creates a segmentation map using those features.(Chen et al.; 2021) proposed a model using U-Net image segmentation to generate medical images. U-Net shows better performance for generating data with a small dataset. The authors used the Adam optimizer to increase the learning rate and to produce improved output. The approach of using U-net segmentation (Zhu et al.; 2019) and adam optimizer produces good results but the authors used the basic GAN framework which motivates this research to use this approach with the CycleGAN framework to enhance model accuracy.

The authors (Bui et al.; 2020) proposed a model with U-Net based generator with instance normalization and a patch-based neural network as a discriminator. These two segmentation networks are joined and trained together with a fixed number of weights. After applying this segmentation model accuracy increased from 84.67% to 90.01% compared with manual segmentation CycleGAN. The author also applied the Adam optimizer to the model for better results. The Adam optimizer is used while training the model because it is recent and a popular algorithm since it is effective and requires less memory(R and S; 2021).

The authors (Mukherkjee et al.; 2022) proposed a new approach of combining two DCGANs and one WGAN to select the best 2 images with great SSIM scores for aggregation. After the aggregation, the VGG-19 model is applied to generate style-transferred images. They also applied the Adam optimizer to increase the learning rate of the model.

This approach to generating the image is different but limited to just combining two images and applying a pre-trained VGG 19 deep learning model. This research needs further improvement by using a customised deep learning model to increase performance.

The authors (Xu et al.; 2020) compared VGG19, ResNet, DenseNet and U-Net segmentation techniques with general style transfer model with GAN where they conclude that the model framework can improve segmentation accuracy by 2 to 4 % if U-Net segmentation is applied to create a generator than other techniques although authors state that VGG19 is the best choice for feature extraction.

The authors (Quan et al.; 2018) compared singleGAN, ReconGAN, and RefineGAN models for generating MRI images whether they conclude that RefinGAN produces fewer errors and outperformed with less running time, and good image quality, thus it will work great in medical applications. RefineGAN is based on U-Net, CNN and GAN. The authors used the U-Net architecture of convolutional autoencoder and deep residual network with skip connections to RefineGAN. It is based on cyclic consistency loss with fully residual CNN. This architecture is similar to CyleGAN but it follows just forward propagation. This paper motivated the research project to combine U-Net segmentation with CycleGAN to improve the model efficiency.

This section helped the research project, select the U-net segmentation technique for image segmentation because it helps to increase the accuracy of a model than manual segmentation and the Adam optimizer as an optimization function. The literature review of this section did not combine CycleGAN with U-net segmentation which motivate this research to use this approach. According to previous research papers(Kim et al.; 2023), Adam Optimizer helps to increase the learning rate and decrease processing time(R and S; 2021). So, the Adam optimizer is applied to a research project model.

2.3 The need for Instance Normalization and choosing the best techniques for evaluation

The instance normalization layer enables a model to understand style information to improve the performance speed at the training phase.(Xu et al.; 2020) compared the performance of a model without normalization and after performing shifting, scaling-up image size, matching style channel-wise, and style transfer of the original image. Authors state that after performing instance normalization model training time is reduced after performing mathematical operations to convert it into a single vector as a single unit which is independent of contexts(Huang and Belongie; 2017).

(Liu et al.; 2021) Adaptive instance normalization is used to insert the style of an image into a generator and batch normalization to successfully generate images and to enhance the learning of image features.(Chong and Ho; 2021) chosen instance normalization than batch normalization for the discriminator to learn instance-oriented mean and to overcome hardware restrictions on training batch size and covariance shift. It helps to increase the performance and stability of the training model.

SSIM and PSNR are the most frequently used evaluation methods for checking whether generated images are real or not (Oh et al.; 2022). PSNR value provides the quality of the generated image and the SSIM score is used to check similarities between two images to check generated image is similar to the original image or not (Horé and Ziou; 2018),(Wang et al.; 2022).

This section helped the research project by indicating the importance of instance normalization to improve the performance and stability of the model through a literature review of the previous study. Also helped to choose appropriate evaluation techniques such as SSIM and PSNR.

3 Methodology

To Generate MRI images with different contrast levels, this research used a U-Net segmentation model for creating a generator and CycleGAN architecture. The research consists set of steps which follow KDD (Knowledge Discovery in Database) process methodology. KDD approach followed to continue the research project work as it covers everything required to accomplish the work. The methodology consists of several 5 steps explained as follows.

3.1 Data Selection

The data gathering was a major task, for this research because of patient privacy laws the availability of the medical imagining dataset is very less. This research required data with 2 separate datasets of MRI images with different contrast levels or different styles. The image data is taken from the git hub library[1]open-source platform. The dataset consists of 2 separate folders one with 42 T1-styled brain MRI images and the second with 46 T2-styled brain MRI images. The attributes of these T1 and T2-styled images are shown in Table 1.

3.2 Data Preprocessing

After the data section and loading of the image dataset into the image array are done, data preprocessing needs to be carried out to achieve the desired results by improving the image features. In this phase, image resizing, image reshaping, image shuffling and image batching are done to capture better image information. It is observed that brain images naturally have more height than the width so, the rectangular format is resized to a square format of 256×256 dimension. The image batching is done with batch size = 2. Image normalization is a process of the normalizing intensity of an image pixel. Image normalization is important to convert input images into a normal range of pixel values (Depeursinge et al.; 2017). The formula for image normalization is as follows:

$$\text{Inorm} = \frac{I - \text{Min}(I) * 255}{\text{Max}(I) - \text{Min}(I)} \tag{1}$$

Here, ‘I’ is the grayscale value of the original image, ’Inorm’ is the grayscale value of the normalized output image, ’Min(I)’ is the minimum grayscale of the original image, ’Max(I)’ is the maximum grayscale of the original image[2]. So, the final formulas for the normalization of the T1 images and T2 images dataset after putting values in equation 1 are,

$$T1norm = (T1images/127.5) - 1.0 \tag{2}$$

$$T2norm = (T2images/127.5) - 1.0 \tag{3}$$

[1]https://github.com/hackassin/Brain-MRI-Style-Transfer-With-GAN/blob/main/MRI

[2]https://www.baeldung.com/cs/instance-vs-batch-normalization

Here, ‘T1norm’, and’ T2norm’ are grayscale values of normalized output images, ‘T1images’, and ‘T2images, are grayscale values of original images and Min(I)=127.5 and Max(I)=255 because images in dataset are grayscale whose range is in between [0,255].

3.3 Data Transformation

Input image data is converted into the format of a TensorFlow dataset using tf.data.dataset() method. In this phase, (Sandfort et al.; 2019) instance normalization is performed on data to normalize it into one sample or instance so that training of the model on normalized data will reduce training time. The data after preprocessing and transformation is shown in Fig.1. The following Fig.1 indicates that data after normalization capture better information than data before normalization.

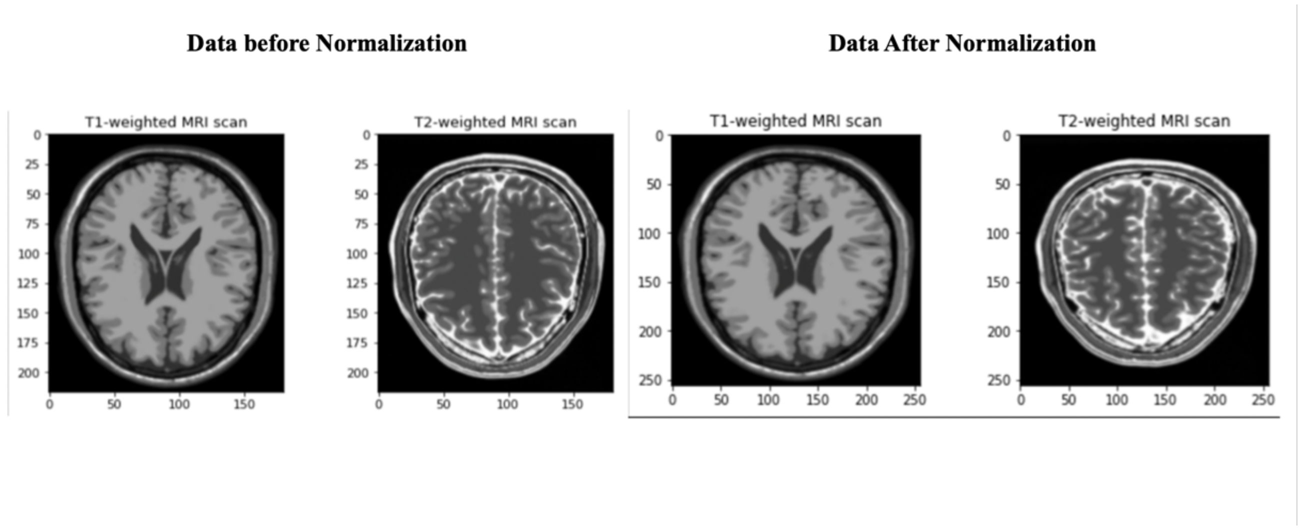

Figure 1: Comparison of data before and after normalization

3.4 Modelling

In this phase, CycleGAN architecture is used to build a model for generating synthetic MRI images as mentioned in section 2.1. CycleGAN architecture contains 2 generators and 2 discriminators. As mentioned in section 2.2, the U-Net segmentation method is used to build the generator which is a combination of convolution and transposed convolution neural network. A discriminator is a classifier which is based on a convolution neural network.

CycleGAN architecture as shown in Fig.2 is a combination of 2 GANs. It contains 2 generators and 2 discriminators. 1^{st} generator transforms the image in the domain ‘a’ into the image in domain ‘b’. Once this fake image is formed 2^{nd} generator transforms the fake image in domain ‘b’ into an image in domain ‘a’ that is the original image. During the training phase of the model, the discriminator checks whether the image generated by generators is real or fake.

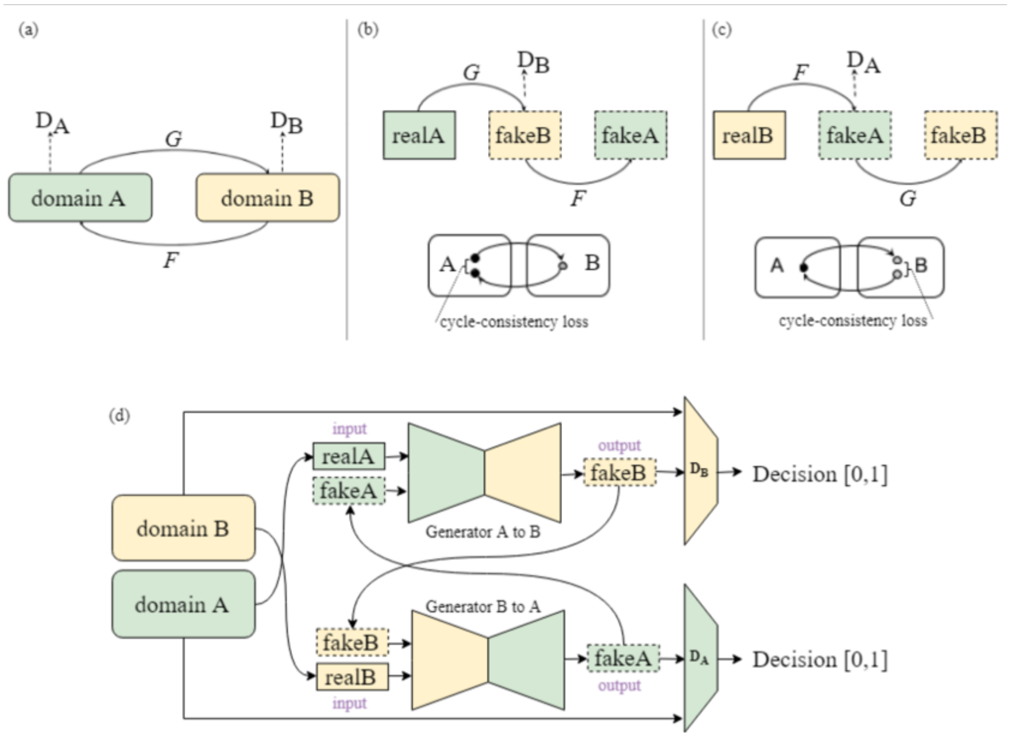

Figure 2: CycleGAN Architecture:Unpaired Image-to-Image Translation (Tmenova et al.; 2019)

In CycleGAN architecture as shown in Fig.2, (Sandfort et al.; 2019) the generator receives feedback from another generator which ensure that the image generated by the generator is cyclic consistent or not. Cyclic consistency, meaning that applying both generators in a sequence should provide an image that is similar. To evaluate the quality of a generated image during training model is done using cyclic consistency loss which helps to ensure forward and backward cyclic consistency. All the losses to evaluate generated image quality during the training phase are briefly described in the next subsection **??**.

The Adam optimizer is also applied to both generator and discriminator with a 2e-4 learning rate for reducing the training time of the model. The Leaky ReLU activation function is also applied to the model before the training phase to speed up the training process2.3.

3.5 Evaluation

The model is evaluated using Cyclic consistency loss, identity loss, generator loss, and discriminator loss during the training phase to choose the best working model.

The cyclic consistency loss is an L1 distance between the reconstructed image and the original image. It indicates the stability of the network. The generator loss and discriminator loss have an adversarial relationship. If the generator loss is up than the discriminator loss then the generator is working well and vice-versa. The losses must not be 0 because if the losses are 0 that means the model is overfitting. The identity loss indicates the similarity between the reconstructed image and the original image(Zhang et al.; 2021).

There are some parameters are important to understand before deriving equations such as G=Generator, D=discriminator, x=Original Data, and z=input noise. Equation 4 indicates generator loss and Equation 5 indicates discriminator loss.

$$\mathbf{L}(\mathbf{G}) = \min[\log(\mathbf{1} - \mathbf{D}(\mathbf{G}(\mathbf{z})))] \tag{4}$$

$$\mathbf{L}(\mathbf{D}) = \max[\log(\mathbf{D}(\mathbf{x})) + \log(\mathbf{1} - \mathbf{D}(\mathbf{G}(\mathbf{z})))] \tag{5}$$

$$L_{cyc}\left(G, D_A, B, A\right) = E_{a \sim p_{\text{data}}\ (a)}\left[|F(G(a)) - a|_1\right] + E_{b \sim p_{\text{data}}\ (b)}\left[|G(F(b)) - b|_1\right] \tag{6}$$

Equation 6 indicates cyclic consistency losses. The parameter for this equation is based on fig.2 where CycleGAN contains 2 mapping functions G:A→B, F:B→A(Zhu et al.; 2017).

After choosing the best working model, the image generated by a model on the test dataset is evaluated using SSIM and PSNR scores. A higher the value of SSIM score for reconstructed images shows that model working well. PSNR score represents the quality of a generated image(Sandfort et al.; 2019).

3.5.1 SSIM:

$$\operatorname{SSIM}(\mathbf{x}, \mathbf{y}) = \frac{\left(2\mu_x\mu_y + C_1\right)\left(2\sigma_{xy} + C_2\right)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \tag{7}$$

The SSIM (Horé and Ziou; 2018) is calculated as shown in the equation.7 where, μ_x is the mean of pixel sample x, μ_y is the mean of pixel sample y, σ_x^2 is a variance of x, σ_y^2 is a variance of y, $\sigma_x y$ is the covariance of x and y, $c1 = (k1L)^2$ and $c1 = (k1L)^2$ are variables to stabilize denominator part, L is 2 bit per pixel value,k1=0.01 and k2=0.03 by default.The SSIM score range between 0 to 1, and 1 represents a perfect match for generated image with the original image.

3.5.2 PSNR:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \tag{8}$$

$$\mathrm{PSNR} = 20 \cdot \log_{10}\left(MAX_I\right) - 10 \cdot \log_{10}(MSE) \tag{9}$$

For calculating PSNR (Horé and Ziou; 2018) first MSE is calculated as shown in equation.8 where MSE is the mean squared error for m*n images 'I' with 'K' as a noise. Then PSNR is calculated as shown in equation.9 where, MAX_I is the maximum pixel value of an image. Here, the image is 8 bits per sample so, the value of MAX_I is 255. The greater the value of PSNR indicates less amount of noise which means generated image has closer likeness to the original image.

4 Design Specification

The model design of the U-Net-based Generator for synthesizing MRI images is shown in fig.3.

In this research project, the CycleGAN model is implemented comprised of 2 generators and 2 discriminators as shown in fig.2. The generator_g and generator_f both generators are built using the U-Net segmentation model. The generator_g translate the domain A (T1-weighted images) to domain B (T2-weighted images) and generator_f translates the domain B (T2-weighted images) to domain A (T1-weighted images) as shown in fig.2.

The design of fig.3 shows U-shaped architecture when it rotates 90 degrees. U-Net generator is a combination of convolutional and transposed convolutional neural networks. The model has a simple U-shaped structure(Schonfeld et al.; 2020). Downsampling of sequential layers that works as an encoder and upsampling of sequential layers that work as a decoder with skip connections as shown in fig.3.

The Upsampling is performed using a transposed convolutional neural network to increase image dimensions and reduce the number of channels. The hyperparameter tuning is done at the time of defining an Upsample() function by passing values such as stride= 2(Filter will move by 2 pixels), padding=" same"(Output size of image is same as input size), keeping the dropout value constant to 0.5 and applying ReLU activation function to result.

The Downsampling is performed using a convolutional neural network to reduce image dimensions and increase the number of channels (Zhu et al.; 2019). The hyperparameter tuning is done at the time of defining a Downsample() function by passing values such as stride= 2, padding=" same", keeping the dropout value constant to 0.5 and applying the ReLU activation function to the result.

The image of 256*256 pixels is passed as input to the U-Net generator so the shape of an input image is (256,256,1) where channel=1 for an input image. The model contains 8 sequential Downsampling layers as shown in fig.3. In each sequential hidden layer, the image dimensions reduce to half because the dropout value is 0.5 and channel numbers are increased. Sequential layer 1 takes the input image of shape (128,128,64) from the output of the previous layer and provides the output image of shape (64,64,128) as input to the next layer.

After Downsampling is done Upsampling is performed by passing the output of sequential layer 8 as input to sequential layer 9. In the upsampling, image dimensions increased by twice and channel numbers decreased. As shown in fig.3 sequential layer 9 takes the input image of shape (2,2,1024) and provides the output image of shape (4,4,512). This process continues until sequential layer 14 and at the end, convolutional transpose is taken to generate an image with the same dimensions as an input image.

The Downsample and Upsample layers are connected by bypass skip connection. Skip connection ensures the sequence of the downsampling layer is the same as the upsampling layer (Zhu et al.; 2019).

This is the architecture design of the generator model which is the same for both generators 'generator_g' and 'generator_f'. The model-building process is done before the training phase. At the time of the training phase, the training performed on the train datasets by integrating these 2 modules in a single model by passing some parameters such as losses, gradient to an optimizer and passing the number of epochs.

There are 3 models developed by passing different parameters and the best model is

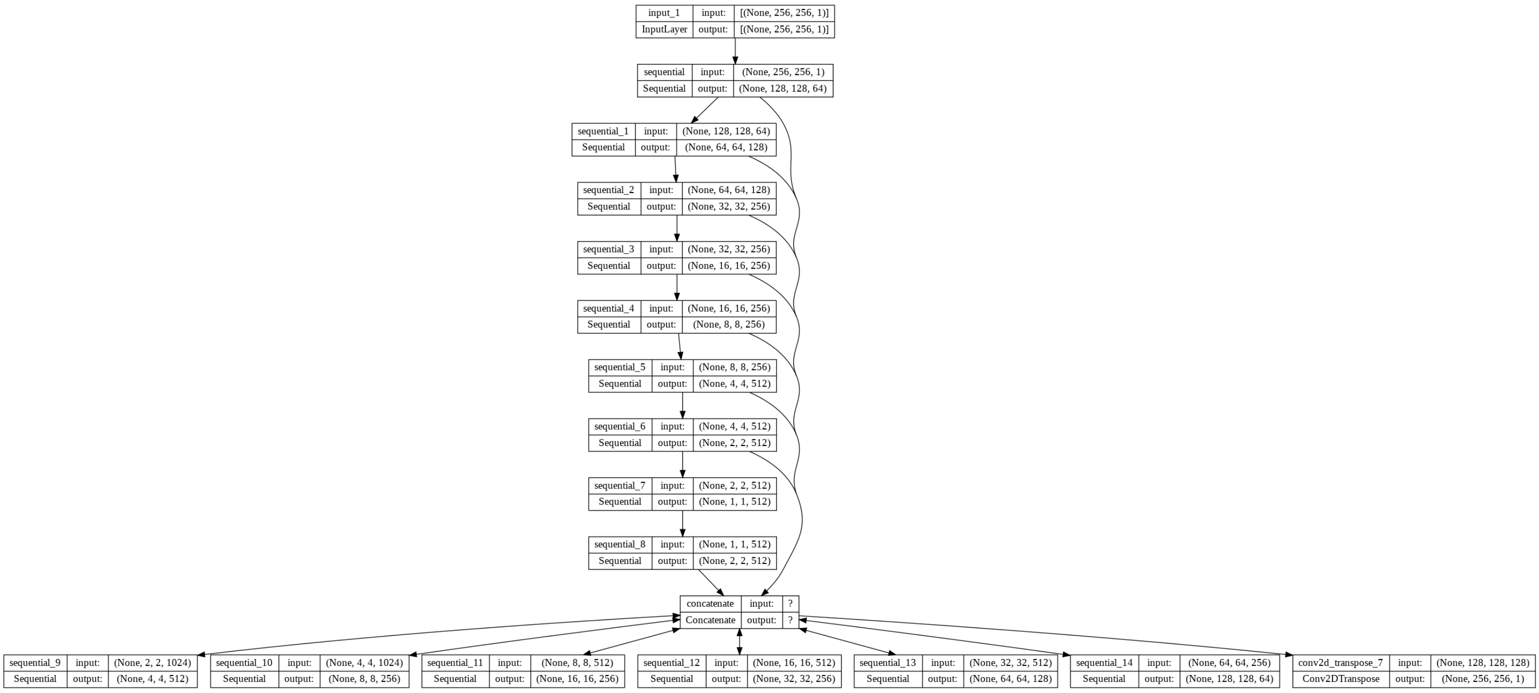

Figure 3: The Architecture Design of the generator model based on U-Net segmentation: Network Structure of Generator with Upsampling and Downsampling Layers

selected for training. A brief discussion of these models is provided in section.5.

5 Implementation

In this research project, 4 experiments are performed for finalizing the best model such as V1, V2, V3, and V4. These 4 experiments represent 4 different models with different parameters. Figure.4 contain all the models for generating MRI images with important hyperparameters. By changing some parameters of the model, the output produced by this model is changed. After assessing the losses of each model the final model is selected as shown in Table.3.

The Losses shown in Table.2 are losses at the last epoch of respective models.

Table 2: Losses of Experiments

Experiments	**Cyclic consistency Loss**	**Generator loss**	**Discriminator loss**	**Identity loss**
V1	5.1	2.3	1.9	1.6
V2	4.8	2.1	1.8	1.3
V3	3.85	1.2	0.96	0.39
V4	3.75	0.93	0.79	0.20

Experiment versions	V1	V2	V3	V4
Image resize dimension (resize_dim_1, resize_dim_2)	128x128	256x256	256x256	256x256
Batch size (batch_size)	40	5	2	2
Instance normalization: random normal initializer range	0., 0.02	0., 0.02	0., 0.02	0., 0.02
Downsampling, Upsampling: filter strides	2	2	2	2
Downsampling, Upsampling filter padding	same	same	same	same
Upsampling: dropout value	0.5	0.5	0.5	0.5
activation function	Leaky ReLU	Leaky ReLU	Leaky ReLU	Leaky ReLU
Unet generator model: model stack: (downsampling layer, upsampling layer)	down1, down2, down3, down4, bottleneck, up1, up2, up3, up4, last	down1, down2, down3, down4, bottleneck, up1, up2, up3, up4, last	down1, down2, down3, down4, down5, down6, down7, bottleneck, up1, up2, up3, up4, up5, up6, up7, last	down1, down2, down3, down4, down5, down6, down7, bottleneck, up1, up2, up3, up4, up5, up6, up7, last
Discriminator model: model stack: (downsampling layer, zero padding layer)	down1(input), down2(down1), zero_pad1(down2), convolution, instance normalization, activation, zero_pad2(activation), last	down1(input), down2(down1), zero_pad1(down2), convolution, instance normalization, activation, zero_pad2(activation), last	down1(input), down2(down1), down3(down2), zero_pad1(down3), convolution, instance normalization, activation, zero_pad2(activation), last	down1(input), down2(down1), down3(down2), zero_pad1(down3), convolution, instance normalization, activation, zero_pad2(activation), last
Loss: Weight for Discriminator loss	0.5	0.5	0.5	0.5
Loss: Weight for Cycle loss	10	10	10	10
Loss: Weight for Identity loss	0.5	0.5	0.5	0.5
Optimizer: optimizer type	Adam	Adam	Adam	Adam
Optimizer: learning rate	2.00E-04	2.00E-04	2.00E-04	2.00E-04
Model Training: Number of epochs	300	300	300	100

Figure 4: List of Experiments with Hyperparameter Tuning: V1, V2, V3, V4

5.1 Experiment 1 (V1)

In this experiment, the model for generating an MRI image is created by passing the hyperparameters mentioned in Fig.4. The resized image of shape(128x128) is passed as input to the model. At the time of training the model, it is observed that the model is not learning enough due to less number of neural network layers. As shown in fig.3, there are only 4 upsampling layers and 4 downsampling layers present in the model stack which are not enough to train a model. The dataset contains less number of an image so, batch size needs to be reduced.

As shown in Table.2, the Cyclic consistency loss 3.5 for V1 is 5.1 which indicates that the L1 distance between real and Recycled images or reconstructed images is comparatively high than in other models. The generator loss and discriminator loss are 2.3 and 1.9 respectively. The generator and discriminator have an adversarial relationship so, the generator performed well for this model because generator loss is greater than discriminator loss. Although the values of the generator and discriminator are very close to each other that means they are balanced. The identity loss is 1.6 which is comparatively high than other models.

5.2 Experiment 2 (V2)

The difference between models V1 and V2 is the shape of the image is changed from (128x128) to (256x256) for capturing more information and the batch size is changed from 40 to 5 because a larger batch size takes more processing time and reduces the performance of the model. All other parameters are the same for both models. As per Table.2, the Cyclic consistency loss for V2 is 4.8 which is other models but lower than V1. The generator loss and discriminator loss are 2.1 and 1.8 respectively. The generator and discriminator have an adversarial relationship so, the generator performed well for this model because generator loss is greater than discriminator loss. Although the values of the generator and discriminators are very close to each other which means they are balanced. The identity loss is 1.3 which is comparatively high than other models. At the time of training the model, it is observed that the model is working well than V1. The dataset is small hence, to reduce overfitting and provide more parameters for training needs to be increased in neural network layers.

5.3 Experiment 3 (V3)

The difference between V2 and V3 models is batch size is changed from 5 to 2 to check if it will work the same or different when reduced batch size is to 2. The number of upsampling, and downsampling layers are increased to 7 each to increase the number of parameters at the training phase to provide better output. All other parameters are the same for both models. As per Table.2, the Cyclic consistency loss for V3is 3.85 which is comparatively low than V1 and V2. The generator loss and discriminator loss are 1.2 and 0.96 respectively. The generator and discriminator have an adversarial relationship so, the generator performed well for this model because generator loss is greater than discriminator loss. Although the values of the generator and discriminator are very close to each other which means they are balanced. The identity loss is 0.39 which is comparatively low than V1 and V2. At the time of training the model, it is observed that the model is working well than V1 and V2 but this model trains for 300 epochs which causes low memory error and takes a long training time. Also after the 100th epoch model shows a repetition of the same output. And to avoid overfitting the model.

5.4 Experiment 4 (V4)

The only difference between model V3 and V4 is the number of epochs reduced from 300 to 100 to avoid overfitting because results produced by the V3 model show repetition of output after the 100th epoch. All other parameters are the same for both models including batch size because a generated image is slightly realistic when it is changed to 2. As per Table.2, the Cyclic consistency loss for V4 is 3.75 which is comparatively low than other models. The generator loss and discriminator loss are 0.93 and 0.79 respectively. The generator and discriminator have an adversarial relationship so, the generator performed well for this model because generator loss is greater than discriminator loss. Although the values of the generator and discriminator are very close to each other that means they are balanced. The identity loss is 0.20 which is comparatively low than other models and indicates that the reconstructed image is similar to the original image. At the time of training the model, it is observed that the model is working well than other models. The V4 model is selected as the final model for this research project after assessing and comparing all the models.

The V4 model is selected as a final model for this research project after assessing and comparing all the models.

6 Evaluation

After the implementation of four CycleGAN-based models, V4 model outperformed all other models because of fewer loss values and based on observation at the training phase. This section illustrates the evaluation results of a final selected model 5.4 using SSIM and PSNR scores.

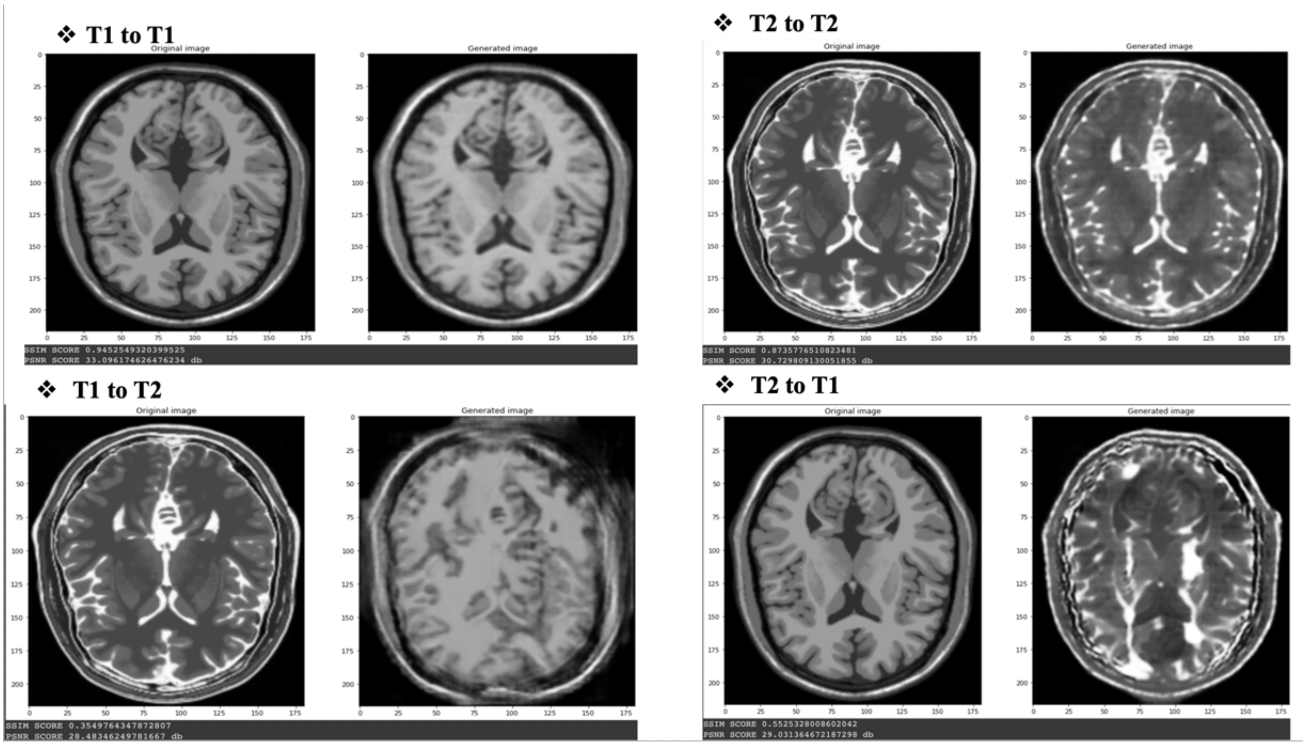

Figure 5: The V4 model Generation results of T1-weighted images to T1-weighted images (T1→T1), T2-weighted images to T2-weighted images (T2→T2), T1-weighted images to T2-weighted images (T1→T2), T2-weighted images to T1-weighted images (T2→T1) in MRI.

Table.3 and Fig.5 contain all the results for the final model. The V4 model is applied to the test dataset to check the final results. The V4 model generates MRI images at four different conditions as shown in Table.3. The following subsections explain the 4 different results generated by the V4 model.

Table 3: Evaluation Metrics for Generated images

Image Generation Style	SSIM Score	PSNR Score
T1→T1	0.94	33.09 db
T2→T2	0.87	30.72 db
T1→T2	0.35	28.48 db
T2→T1	0.55	29.03 db

6.1 Result 1: Generating T1-styled MRI image when T1-styled MRI image is provided

The model reconstructs the image when the T1-styled image from test dataset and generator_f are provided as shown in fig.5. The SSIM score is 0.94. According to 5 and the SSIM score, the reconstructed image is a closer resemblance to the original image. which indicates that the model is working well. The PSNR score is 33.09 dB.The PSNR score indicate the quality of generated image which is better than earlier similar works (Mukherkjee et al.; 2022) (Tmenova et al.; 2019).

6.2 Result 2: Generating T2-styled MRI image when T2-styled MRI image is provided

T2-styled image and generator_g are provided to the model to reconstruct the T2-styled image as shown in Fig.5. The SSIM score is 0.87. According to Fig.5 and the SSIM score, the reconstructed image is a closer resemblance to the original image. which indicates that the model is working well. The PSNR score is 30.72 dB which is better than earlier similar works (Mukherkjee et al.; 2022) (Tmenova et al.; 2019).

6.3 Result 3: Generating T1-styled MRI image when T2-styled MRI image is provided

T1-styled image and generator_g are provided to the model to reconstruct the T2-styled image as shown in Fig.5. The SSIM score is 0.35. According to Fig.5 and the SSIM score, the generated image is not similar to the original image structure. This indicates that the model needs improvement for generating T2-styled contrast images. The PSNR score is 28.48 dB.

6.4 Result 4: Generating T2-styled MRI image when T1-styled MRI image is provided

T2-styled image and generator_f are provided to the model to reconstruct the T2-styled image as shown in Fig.5. The SSIM score is 0.55. According to Fig.5 and the SSIM score, the reconstructed image is a little bit closer to the original image structure. which indicates that the model is working better for generating T1-styled contrast images than Result 3 but still needs to be improved. The PSNR score is 29.03 dB which is better than earlier similar works (Mukherkjee et al.; 2022).

6.5 Discussion

The main objective of this research is a build a model using CycleGAN and U-Net segmentation to generate MRI image data for practitioners and doctors to avoid misdiagnosis. This objective is achieved by this research project. It was challenging to carry out this research from the start, especially when gathering contrast-styled MRI images. In this research work, customized CycleGAN-based models are created by hyperparameter tuning. Four of them are mentioned in section.5.The hyperparameter tuning is performed based on output produced by initial model V1. In Model V2 batch size is changed from 40 to 5 because a larger batch size takes more processing time and reduces the performance of the model and the image is changed from (128x128) to (256x256) for capturing more information. In model V3 batch size is changed from 5 to 2 to check if it will work the same or different when the batch size is reduced to 2. The number of upsampling, and downsampling layers are increased to 7 each to increase the number of parameters at the training phase to provide better output. In model V4 the number of epochs reduced from 300 to 100 to avoid overfitting because results produced by the V3 model show repetition of output after the 100th epoch all other parameters are the same as V3. After assessing the working of a model during the training phase the best model is selected for generating MRI images with different contrast. The model selection is done based on Cyclic consistency loss, the relationship between generator and discriminator loss and identity loss. Model V4 is mentioned in section.5.4 is outperformed. The final model V4 is used to generate MRI images on the test dataset. The results are evaluated using SSIM and PSNR scores provided in section.6.

The final model generates the artificial MRI image using style transfer learning CycleGAN framework with U-NET segmentation. A greater SSIM score indicates high similarity exists between two images. The research question is aimed to generate realistic artificial MRI images so, the high SSIM score between the original and reconstructed image indicates that the model is working well in generating realistic MRI images. The model generating MRI images with different styles works well if the SSIM score is high. The PSNR score indicates the quality of a generated image. A typical value for a PSNR score is between 30 to 50 db for 8-bit grey-scale images. The final model V4 generates T1 to T1 styled MRI images with 0.94 SSIM score and 33.09 db PSNR score, and T2 to T2 styled MRI images with 0.87 SSIM score and 30.72 db PSNR score. This research model is working well for generating the same styled images. While generating differently styled images that is T1 to T2 with 0.35 SSIM score and 28.48 db PSNR score, T2 to T1 with 0.55 SSIM score 29.03 db PSNR score. That means improvement is needed for generating differently styled images which is the main objective of this research project. The higher the score of SSIM between generated and original image the model is working accurately.

Similar work is mentioned with the same kind of datasets in Table.4, (Mukherkjee et al.; 2022) where SSIM scores are 0.77 and 0.80 for a customized model for Generating T1-styled images from T2 and T2-styled images from T1 respectively. These scores are better than research model V4. This indicates that the model needs improvement to

Table 4: Evaluation of Final model(CycleGAN+UNET) with other GAN models:Comparision between final model and similar previous work models

Reserch	Method	SSIM Score		PSNR Score	
		T1-T2	T2-T1	T1-T2	T2-T1
(Mukherkjee et al.; 2022)	AGGrGAN	0.77	0.80	29.14	29.63
(Tmenova et al.; 2019)	CycleGAN	0.96	0.73	31.3	21.96
(Zhang et al.; 2021)	VGG16	0.90	0.93	28.10	28.65
(Deng et al.; 2021)	StarGAN	0.93	0.92	25.5	24.97
(Deng et al.; 2021)	pGAN	0.93	0.92	26.52	25.52
(Deng et al.; 2021)	LGAN	0.94	0.93	27.99	27.24
Our Final Model	CycleGAN+UNET	0.35	0.55	28.48	29.03

increase the SSIM score.

PSNR scores are quite close to the earlier research work (Mukherkjee et al.; 2022),(Tmenova et al.; 2019),(Zhang et al.; 2021) which means the quality of generated images is good compared to previous literature work but could be improved by reducing noise from images.

After comparing the results with previous work4 that performed better than model V4, most of the research papers used pre-trained segmentation models such as VGG-19 (Zhang et al.; 2021), Res-Net50 (Mukherkjee et al.; 2022), 3D-DenseNet (Bui et al.; 2020), MobileNetV2 (Zhang et al.; 2021),LGAN ,PGAN ,StarGAN (Deng et al.; 2021). These pre-trained models could improve the performance of the model instead of U-Net segmentation.

The model performs well while generating an image of the same style that is from T1 to T1 and T2 to T2. But the model needs improvement while generating T2 to T1 and T1 to T2 styled images.

The objective of this research project is to solve the problem of the data availability issue that occurred in the medical field because radiologists and practitioners need huge datasets for diagnosing diseases accurately. At this stage, this model works well while generating the same style of images but needs improvement for generating different-style MRI images. The medical domain is a very sensitive area, so the model should generate accurate results because a false diagnosis can cause a serious issue with patients' health. The final model is not providing enough accurate output so, these models should not be deployed in the medical domain for practitioners and doctors at this point. In future, this research will be extended to improve the model performance.

7 Conclusion and Future Work

Generating MRI images with different contrast levels is one of the important areas in the medical domain. In the medical domain, doctors required a different type of imagining to analyse and diagnose the disease but this is a very costly process. This research will help patients by reducing the cost required for different medical imaging and also help a practitioner to diagnose diseases with additional images. The novel approach of building a customized model with tuned hyperparameters using the CycleGAN framework and U-Net segmentation is tested in this research. The dataset used in this research is quite small than previous research. The four models are created in section.5 from which the best model is selected based on losses. Then the final model is evaluated using SSIM and PSNR scores for four results. As shown in section 6. The model performance is better when generating same-styled images but improvement is needed for generating images with different styles or contrast.

The medical domain is a very sensitive area because any false diagnosis can affect patients' life badly. At this stage, this model will not be recommended to practitioners and doctors because it needs improvement when generating different style MRI images.

In future work, research will be more focused on gathering high-resolution colour image data or 3D image datasets to build a model for generating high-quality MRI images. In future, this research will be extended to find the best approach by comparing the pretrained segmentation model with the customized U-Net model.

References

Bui, T. D., Wang, L., Lin, W., Li, G. and Shen, D. (2020). 6-month infant brain mri segmentation guided by 24-month data using cycle-consistent adversarial networks, IEEE, pp. 359–362.

Chen, X., Li, Y., Yao, L., Adeli, E. and Zhang, Y. (2021). Generative adversarial u-net for domain-free medical image augmentation.
URL: *http://arxiv.org/abs/2101.04793*

Chong, C. K. and Ho, E. T. W. (2021). Synthesis of 3d mri brain images with shape and texture generative adversarial deep neural networks, *IEEE Access* **9**: 64747–64760.

Deng, C., Chen, Z., Wang, R., Su, W. and Qu, Y. (2021). Modality-shared mri image translation based on conditional gan, IEEE, pp. 1286–1290.

Depeursinge, A., Fageot, J. and Al-Kadi, O. S. (2017). Chapter 1 - fundamentals of texture processing for biomedical image analysis: A general definition and problem formulation, *in* A. Depeursinge, O. S. Al-Kadi and J. Mitchell (eds), *Biomedical Texture Analysis*, The Elsevier and MICCAI Society Book Series, Academic Press, pp. 1–27.
URL: *https://www.sciencedirect.com/science/article/pii/B9780128121337000016*

Han, C., Hayashi, H., Rundo, L., Araki, R., Shimoda, W., Muramatsu, S., Furukawa, Y., Mauri, G. and Nakayama, H. (2018). Gan-based synthetic brain mr image generation, IEEE, pp. 734–738.

Horé, A. and Ziou, D. (2018). Image quality metrics: Psnr vs. ssim, *2018 20th International Conference on Pattern Recognition*, pp. 2366–2369.

Huang, X. and Belongie, S. (2017). Arbitrary style transfer in real-time with adaptive instance normalization, *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.

Kim, E., Cho, H.-H., Kwon, J., Oh, Y.-T., Ko, E. S. and Park, H. (2023). Tumor-attentive segmentation-guided gan for synthesizing breast contrast-enhanced mri without contrast agents, *IEEE Journal of Translational Engineering in Health and Medicine* **11**: 32–43.

Liu, M., Maiti, P., Thomopoulos, S., Zhu, A., Chai, Y., Kim, H. and Jahanshad, N. (2021). Style transfer using generative adversarial networks for multi-site mri harmonization.

Mok, T. C. W. and Chung, A. C. S. (2019). Learning data augmentation for brain tumor segmentation with coarse-to-fine generative adversarial networks, *in* A. Crimi, S. Bakas, H. Kuijf, F. Keyvan, M. Reyes and T. van Walsum (eds), *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, Springer International Publishing, Cham, pp. 70–80.

Mukherkjee, D., Saha, P., Kaplun, D., Sinitca, A. and Sarkar, R. (2022). Brain tumor image generation using an aggregation of gan models with style transfer, *Scientific Reports* **12**: 9141.

Oh, Y.-T., Ko, E. and Park, H. (2022). Tdm-stargan: Stargan using time difference map to generate dynamic contrast-enhanced mri from ultrafast dynamic contrast-enhanced mri, IEEE, pp. 1–5.

Quan, T. M., Nguyen-Duc, T. and Jeong, W.-K. (2018). Compressed sensing mri reconstruction using a generative adversarial network with a cyclic loss, *IEEE Transactions on Medical Imaging* **37**: 1488–1497.

R, R. T. and S, V. K. K. (2021). Artificial mri image generation using deep convolutional gan and its comparison with other augmentation methods, IEEE, pp. 1–6.

Sandfort, V., Yan, K., Pickhardt, P. J. and Summers, R. M. (2019). Data augmentation using generative adversarial networks (cyclegan) to improve generalizability in ct segmentation tasks, *Scientific Reports* **9**: 16884.

Schonfeld, E., Schiele, B. and Khoreva, A. (2020). A u-net based discriminator for generative adversarial networks, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Singh, N. K. and Raza, K. (2021). Medical image generation using generative adversarial networks: A review.

Tmenova, O., Martin, R. and Duong, L. (2019). Cyclegan for style transfer in x-ray angiography, *International Journal of Computer Assisted Radiology and Surgery* **14**: 1785–1794.

Wang, J., Qiao, L., Lv, H. and Lv, Z. (2022). Deep transfer learning-based multi-modal digital twins for enhancement and diagnostic analysis of brain mri image, *IEEE/ACM Transactions on Computational Biology and Bioinformatics* pp. 1–1.

Xu, Y., Li, Y. and Shin, B.-S. (2020). Medical image processing with contextual style transfer, *Human-centric Computing and Information Sciences* **10**: 46.

Xu, Z., Qi, C. and Xu, G. (2019). Semi-supervised attention-guided cyclegan for data augmentation on medical images, IEEE, pp. 563–568.

Zhang, J., Liu, L., Zhang, H., Gao, S. and Ma, Y. (2021). Image-to-image bidirectional generation net for mr images, ACM, pp. 85–91.

Zhu, J.-Y., Park, T., Isola, P. and Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks, *Proceedings of the IEEE international conference on computer vision*, pp. 2223–2232.

Zhu, Y., Zhou, Z., Liao, G., Yang, Q. and Yuan, K. (2019). Effects of differential geometry parameters on grid generation and segmentation of mri brain image, *IEEE Access* **7**: 68529–68539.