# Cortex-Level Brain MRI Generation Using Diffusion Models

Sidra Nadeem
Stanford University
Stanford, CA
sidranem@stanford.edu

Parnian Azizan
Stanford University
Stanford, CA
azizian@stanford.edu

Isabel Michel
Stanford University
Stanford, CA
imichel@i2stanford.edu

## Abstract

*This paper presents an innovative two-stage diffusion model framework tailored for enhancing MRI generation in medical imaging, with a specific focus on brain MRI scans. Our methodology leverages the capabilities of advanced diffusion models to address the challenges associated with accurately mapping deformation fields from moving to fixed images, a critical aspect in medical diagnostics where precise image registration is pivotal.*

*In the initial stage, our approach utilizes a diffusion model to learn the deformation fields from moving images to fixed images. By employing a neural network architecture that includes a specialized 3D U-Net, our model achieves fine-grained control over the deformation process and integrates moving and fixed images seamlessly.*

*Following the generation of these deformation fields, the second stage of our model generates deformation fields to produce highly accurate registered MRIs. Our results, highlighted through comparative visualizations, showcase the model's effectiveness in achieving acceptable registration, indicating its high accuracy and potential utility in clinical settings.*

*The proposed framework not only advances the state-of-the-art in medical image registration but also opens avenues for further research into more generalized applications across various types of medical imaging. This work underscores the potential of integrating sophisticated machine learning techniques with traditional medical imaging processes, promising significant improvements in the accuracy and reliability of medical diagnostics and treatment planning. Future enhancements will focus on refining these techniques to broaden their applicability and efficiency in real-world medical scenarios.*

## 1. Introduction

Neuroimaging has made significant strides in understanding the human brain, yet it continues to face substantial challenges. Limited sample sizes and a lack of diversity in training datasets often result in models that are biased and prone to overfitting. These limitations hinder the reliability and generalizability of neuroimaging studies, which are crucial for brain modeling, visualization, and individual-level diagnostics. Our project, "Cortex-Level Brain MRI Generation Using Diffusion Models," seeks to overcome these challenges by leveraging cutting-edge generative AI techniques to create synthetic brain MRIs that are both realistic and representative of diverse populations.

The primary problem we address is the need for high-quality, synthetic brain MRIs to improve the robustness of neuroimaging studies. Our approach harnesses the power of diffusion models and registration-based techniques to accurately reproduce the complex distribution of actual MRI data, focusing on enhancing detail and accuracy at the cortex level. By enhancing the fidelity of cortex-level brain MRIs, we aim to provide more reliable data for neuroimaging research and clinical applications.

The novelty of our work lies in the innovative combination of diffusion models and image registration-based techniques to generate high-quality, realistic brain MRIs. Traditional methods often focus on directly generating MRIs from images, which can lead to less accurate and realistic results. Our approach consists of two stages. In the first stage, we employ DiffuseMorph [8], an unsupervised deformable image registration approach based on diffusion models, to learn deformation fields from a fixed template MRI to all other moving MRIs in the dataset. DiffuseMorph leverages the denoising diffusion probabilistic model (DDPM) to generate deformation fields that align the moving images to the fixed image. This method allows for continuous and realistic deformations by scaling the latent features in the diffusion model, thereby ensuring topological preservation and high registration accuracy. Therefore, we can achieve high-fidelity deformation fields that are crucial for generating anatomically accurate synthetic MRIs.

In the second stage, we build a denoising diffusion probabilistic model (DDPM). This model learns the distribution of the deformation fields obtained in the first stage, allowing us to generate new deformation fields by sampling from

this learned distribution. Once the model has learned this mapping, we can generate new synthetic brain MRIs by applying the new deformation fields to the template MRI. This two-step approach—first learning the deformation fields and then generating new ones from noise—simplifies the learning process and improves the quality of the generated images. The outputs of our method are then high-resolution, realistic brain MRIs that accurately reflect the intricate details of the brain's cortex.

To evaluate our method, we will compare our generated MRIs with existing baseline methods. We selected VAE-GAN, $\alpha$-WGAN, and HA-GAN as representive GAN models due to their medical relevance and prominence in the literature. We also compare our model to the state-of-the-art diffusion models, including the MONAI latent diffusion model (MONAI-LDM), the text-conditioned diffusion model (MedSyn), and the conditional diffusion probalistic model (cDPM), as they are current best methods that can generate high fidelity results.

By addressing the current limitations in neuroimaging data, our project aims to provide researchers and clinicians with more reliable and detailed cortex-level brain MRIs. This work has the potential to improve neuroimaging studies and lead to better diagnostic tools and treatments for brain-related conditions. Through our innovative approach, we hope to contribute to the development of more accurate and representative neuroimaging technologies.

## 2. Related Work

Magnetic Resonance Imaging (MRI) is an essential tool in neuroimaging, offering detailed insights into the brain's structure and function. The integration of deep learning (DL) in MRI analysis has revolutionized the field, providing powerful methods for enhancing image quality, automating image interpretation, and improving diagnostic accuracy. However, training DL models requires large volumes of high-quality data, a challenge in medical imaging due to the cost and complexity of acquiring annotated datasets [1] [13] [3]. Generative models, particularly Generative Adversarial Networks (GANs) and diffusion probabilistic models, have emerged as promising solutions to this problem by generating realistic synthetic data to augment training datasets [5] [6][14].

Generative Adversarial Networks (GANs) have been widely adopted for their ability to produce high-fidelity images. In medical imaging, GANs have been used to generate synthetic MRI data, which can be utilized for training and validating DL models. Key advancements include the VAE-GAN, which combines Variational Autoencoders (VAEs) and GANs to enhance image quality and diversity [10]. The $\alpha$-WGAN (alpha-Wasserstein GAN) improves training stability and image realism by optimizing the Wasserstein distance ][15]. Another notable model is the Hierarchical Amortized GAN (HA-GAN), which employs a hierarchical structure to generate high-resolution brain MRIs, crucial for detailed medical analysis [15].

In parallel, diffusion probabilistic models have gained traction for their robustness in generating high-dimensional data like 3D medical images. These models iteratively learn the statistical distribution of data through forward and reverse diffusion processes. The MONAI latent diffusion model (MONAI-LDM) integrates these processes within the MONAI framework, enabling efficient generation of synthetic brain MRIs [12]. The text-conditioned diffusion model, MedSyn, allows for the creation of images with specific anatomical features or pathologies based on textual descriptions [16]. The conditional diffusion probabilistic model (cDPM) further enhances image fidelity by conditioning the diffusion process on input conditions such as patient demographics [11].

Another critical area in medical imaging is the accurate alignment of these images. This is where Deformable Image Registration (DIR) plays a vital role. Integrating generative models with DIR can further enhance the quality and utility of synthetic images, ensuring they are not only realistic but also accurately aligned with anatomical structures. DiffuseMorph by Kim et al. leverages diffusion models to generate deformation fields, offering continuous and realistic deformations. This method underscores the potential of diffusion models in enhancing DIR accuracy and efficiency [8]. Further advancements include the Plug-and-Play Image Registration Network (PIRATE), which integrates a pre-trained CNN denoiser within an iterative optimization framework to enhance data fidelity. PIRATE+ extends this by employing deep equilibrium models (DEQ) to fine-tune the CNN regularizer, effectively balancing data fidelity and model flexibility [7].

Despite these advances, several limitations persist. Most existing methods focus on global alignment without emphasizing the intricate details at the cortex level, which is crucial for understanding brain function and pathology. Precise cortex-level imaging is essential for accurate diagnosis and research, particularly for high-level brain functions such as perception, cognition, and motor control. Furthermore, maintaining high anatomical fidelity in synthetic images remains a challenge, limiting their clinical and research utility.

Our project addresses these gaps by targeting cortex-level MRI generation through the integration of DIR and diffusion models. By combining the strengths of DIR—ensuring accurate alignment and anatomical consistency—with the generative capabilities of diffusion models, we aim to produce high-quality, anatomically accurate brain MRIs. This approach enhances the quality of synthetic MRIs, providing reliable data for neuroimaging research and clinical applications. Our work promises to im-

prove the detail and accuracy of synthetic MRIs, advancing neuroimaging and contributing to better diagnostic and research outcomes.

## 3. Methods

Our proposed method for cortex-level brain MRI generation consists of two stages, each based a diffusion model. The first stage involves learning deformation fields, and the second stage focuses on generating new deformation fields based on the outputs of the first model. Below, we detail each stage and our approach to evaluation.

### 3.1. DiffuseMorph

In the first stage, we employ DiffuseMorph, an unsupervised deformable image registration approach based on diffusion models, to learn deformation fields from a template MRI (moving image) to all other images in the dataset (fixed image). DiffuseMorph leverages the denoising diffusion probabilistic model (DDPM) to generate deformation fields that align the moving images to the fixed image.

#### 3.1.1 Forward Diffusion Process (FDP)

The DiffuseMorph framework consists of two main networks:

Diffusion Network ($G_\theta$): This network learns the conditional score function of the deformation between the moving and fixed images.

Deformation Network ($M_\psi$): This network uses the latent feature from the score function to estimate the deformation field, which is then used to warp the moving image.

The objective function for training DiffuseMorph is a combination of diffusion loss and registration loss:

$$\min_{G_\theta, M_\psi} L_{\text{diffusion}}(c, x_t, t) + \lambda L_{\text{regist}}(m, f) \tag{1}$$

where $L_{\text{diffusion}}$ is the diffusion loss that trains the conditional score function, and $L_{\text{regist}}$ is the registration loss that ensures the deformed source image resembles the fixed image. The diffusion loss is given by:

$$L_{\text{diffusion}}(c, x_t, t) = \mathbb{E}_{\epsilon, x_t, t} \left\| G_\theta(c, x_t, t) - \epsilon \right\|^2 \tag{2}$$

and the registration loss is:

$$L_{\text{regist}}(m, f) = -\text{NCC}(m(\phi), f) + \lambda_\phi \sum \|\nabla \phi\|^2 \tag{3}$$

where NCC is the normalized cross-correlation, and $\lambda_\phi$ is a regularization parameter.

During inference, the trained networks provide the deformation field $\phi$ at $t = 0$:

$$\phi = M_\psi(m, G_\theta(c, x_0, t)) \tag{4}$$

where $x_0$ is the fixed image. The deformed image $m(\phi)$ is obtained by warping the moving image using the spatial transformation layer.

In the image registration that warps the moving image into the fixed image, the DiffuseMorph model provides the continuous deformation of the moving image along the trajectory toward the fixed image. This is achieved by scaling the latent feature. Specifically, as illustrated in Algoritm 1, the registration field $\phi_\eta$ for the continuous image deformation can be generated by simply interpolating the latent feature:

$$\phi_\eta = M_\psi(m, \hat{\epsilon}_f^\eta) \tag{5}$$

where

$$\hat{\epsilon}_f^\eta = \eta \cdot \hat{\epsilon}_f \tag{6}$$

and

$$\hat{\epsilon} f = G\theta^*(c, f, 0) \tag{7}$$

In this formulation, $\eta$ varies between 0 and 1 to provide a smooth trajectory from the moving to the fixed image.

**Algorithm 1.** Continuous image registration

1: **Input**: Conditional images, $c = (m, f)$
2: **Output**: Deformed moving image, $m(\phi_\eta)$
3: Set the latent feature $\hat{\epsilon}^f = G_{\theta^*}(c, f, 0)$
4: **for** $\eta \in [0, 1]$ **do**
5: $\quad \hat{\epsilon}_\eta^f \leftarrow \eta \cdot \hat{\epsilon}^f$
6: $\quad \phi_\eta \leftarrow M_{\psi^*}(m, \hat{\epsilon}_\eta^f)$
7: **end for**=
8: **return** $m(\phi_\eta)$

#### 3.1.2 Reverse Diffusion Process (RDP)

The latent feature produced by the diffusion network plays a dual role: it facilitates the deformation process and guides the creation of synthetic deformed images via reverse diffusion. Unlike the conventional conditional generative method of DDPM [6] [14] that begins with pure Gaussian noise $x_T \sim \mathcal{N}(0, I)$, here our process starts with the moving image. By initializing with the original moving image, a one-step forward diffusion is applied:

$$x_T = \sqrt{\alpha_T} m + \sqrt{1 - \alpha_T} \epsilon, \tag{8}$$

where $\epsilon \sim \mathcal{N}(0, I)$, and $\alpha_T$ is the noise level at time step $T$, with $T \leq T_{\text{train}}$. This method of forward sampling brings the moving image's distribution closer to that of the fixed image, thereby minimizing the number of reverse diffusion steps required and shortening the generation time.

Following this, starting from $x_T$, the synthesis of the aligned synthetic image $x_0$ to the fixed image $f$ proceeds via the reverse diffusion process from $t = T$ down to $t = 1$:

$$x_{t-1} = \sqrt{x_t} - \sqrt{t} G_{\theta^*}(c, x_t, t) + \sigma_t z, \tag{9}$$

where $z \sim \mathcal{N}(0, I)$. This approach allows the flexibility to adjust the number of sampling steps; in our trials, we capped the reverse steps at 200. The pseudocode for this generative process, named DiffuseMorph, is outlined in Algorithm 2.

**Algorithm 2.** Synthetic image generation process

**Input**: Conditional images, $c = (m, f)$
**Output**: Synthetic deformed image, $x$
Set $T \in (0, T_{train})$
Sample $x_T = \sqrt{\alpha_T} m + \sqrt{1 - \alpha_T}\epsilon$, where $\epsilon \sim \mathcal{N}(0, I)$
**for** $t = T, T-1, ..., 1$ **do**
    $z \sim \mathcal{N}(0, I)$
    $x_{t-1} \leftarrow \frac{1}{\sqrt{1-\beta_t}}(x_t - \frac{\beta_t}{\sqrt{1-\alpha_t}} G_{\theta^*}(c, x_t, t)) + \sigma_t z$
**end for**
**return** $x_0$

## 3.2. Diffusion Probabilistic Model

In the second stage, we build a diffusion model to map Gaussian random noise to new deformation fields. This model learns the distribution of the deformation fields in the first model, allowing us to generate new deformation fields by sampling from this learned distribution. New MRI samples are then generated by applying the generated deformation fields to the template MRI. This approach simplifies the learning process by focusing on the deformation fields rather than the entire image, enhancing the quality and anatomical accuracy of the generated images.

The basic Diffusion Probabilistic Model (DPM) framework for data generation is provided here [6] [14]. Further details on adapting this model to align with our 3D data type are provided at the end of this section.

DPM generates MRIs from random noise by alternating between two processes: 1) gradually converting data into noise (Forward Diffusion Process) and 2) transforming noise back into data (Reverse Diffusion Process).

### 3.2.1 Forward Diffusion Process (FDP):

Let real data $X_0 \sim q$ be sampled from the real data distribution $q$. The FDP simulates the diffusion process that converts $X_0$ into Gaussian noise $X_T \sim \mathcal{N}(0, I)$ after $T$ steps, where $\mathcal{N}$ is the Gaussian distribution with zero mean and identity matrix $I$ as variance. This process is described as a Markov chain with a transition kernel $q(X_t|X_{t-1})$ at time step $t \in \{0, \ldots, T\}$:

$$q(X_t|X_{t-1}) := \mathcal{N}(X_t; \sqrt{1 - \beta_t} \cdot X_{t-1}, \beta_t \cdot I). \tag{10}$$

The weight $\beta_t \in (0, 1)$ changes over time to gradually add Gaussian noise to the data. Define $\alpha_t := 1 - \beta_t$ and $\bar{\alpha}_t := \prod_{s=1}^{t}(1 - \beta_s)$, then $X_t$ is a sample from the distribution conditioned on $X_0$ as follows:

$$q(X_t|X_0) := \mathcal{N}(X_t; \sqrt{\bar{\alpha}_t} \cdot X_0, (1 - \bar{\alpha}_t) \cdot I). \tag{11}$$

Given this closed-form solution, we can sample $X_t$ at any arbitrary time step $t$ without iterating through the entire Markov chain.

### 3.2.2 Reverse Diffusion Process (RDP):

The RDP aims to generate realistic data from random noise $X_T$ by approximating the posterior distribution $p(X_{t-1}|X_t)$. It does this by traversing the entire Markov chain from time step $T$ to 0:

$$p_\theta(X_{t-1}|X_t) := \mathcal{N}(X_{t-1}; \mu_\theta(X_t, t), \Sigma), \tag{12}$$

where $\Sigma$ is a fixed variance. The mean $\mu_\theta(X_t, t)$ can be expressed as:

$$\mu_\theta(X_t, t) = \frac{1}{\sqrt{\alpha_t}} \left( X_t - \frac{\beta_t}{\sqrt{1 - \bar{\alpha}_t}} \epsilon_\theta(X_t, t) \right), \tag{13}$$

with $\epsilon_\theta(\cdot)$ being the estimate from a neural network parameterized by $\theta$. The parameters $\theta$ are optimized by minimizing the reconstruction loss defined as:

$$\mathbb{E}\left[ \|\epsilon - \epsilon_\theta(X_t, t)\|^2 \right], \quad X_0 \sim q,\ t \in [0, \ldots, T],\ \epsilon \sim \mathcal{N}(0, I), \tag{14}$$

where $\| \cdot \|_2$ is the $L_2$ norm, and $X_t$ is inferred from Eq. (9) based on $X_0$.

Our method for generating deformation fields adapts the standard diffusion model framework for 3D data using a neural architecture based on the video diffusion model. We treat 3D data as video, with identical height and width dimensions, and the depth dimension serving as the equivalent to frames.

As illustrated in Figure 1, we utilize a 3D U-Net that operates across all three dimensions to handle 3D data. We convert standard 3x3 convolutions to space-only 3D convolutions in a 1x3x3 format, where the first dimension corresponds to depth. Spatial attention is maintained exclusively over space, but after each spatial attention stage, we integrate a temporal attention block that focuses on the depth dimension.

## 4. Dataset

We used 3060 T1-weighted brain MRIs of subjects (normal controls) from three different datasets: public data set

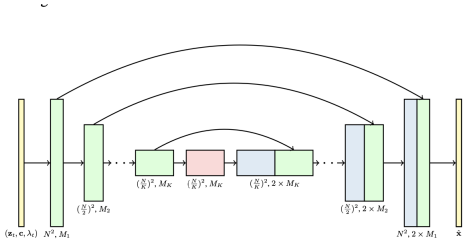

Figure 1: The 3D U-Net structure used in the diffusion model: each block represents a 4D tensor labeled as depth by height by width by channels. Inputs include a noisy deformation field $z_t$, conditioning $c$, and the log SNR $\lambda_t$, with the model adjusting the spatial resolution through downsampling/upsampling operations that alter the height and width dimensions by a factor of two across each block and also the channel multipliers $M_1, M_2, ..., M_K$.

of the Alzheimer's Disease Neuroimaging Initiative (ADNI-1) and the National Consortium on Alcohol and Neurodevelopment in Adolescence (NCANDA) [4], and an in-house dataset. The testing were performed on a 400 subjects not used during training. Additionally, 10 percent of the training data was allocated for validation.

The preprocessing of the MRI data involved several steps: denoising, bias field correction, skull stripping, affine registration to a template, and normalization of intensity values to a range between 0 and 1. We also resampled the images to a 256 × 256 × 256 grid with isotropic voxels measuring 1 $mm^3$ each. The images underwent cropping to dimensions of 32 × 128 × 128. For data augmentation, we randomly applied horizontal and vertical flips and 90-degree rotations to the data, with each transformation having a 50 percent probability of being performed.

## 5. Results and Discussion

Our experiments are conducted on an NVIDIA A100 GPU using the PyTorch framework. For the first model, we set the noise level from $10^{-6}$ to $10^{-2}$ by linearly scheduling with $T_{\text{train}} = 2000$. We used the backbone of VoxelMorph [2] for the deformation network. We also configured layers of the networks according to the dimension of the image, which is a 3D convolution layer for 3D image registration. We trained the model for 50 epochs with $\lambda = 10$ and the learning rate $1 \times 10^{-4}$, using the Adam optimization algorithm [9] with batch size of 128.

For the second model, we implemented a cosine noise schedule with a log SNR range from -20 to 20. The training consisted of 100,000 steps using a learning rate of $1 \times 10^{-4}$, leveraging the Adam optimizer and an EMA decay factor of 0.995. We utilized a batch size of 128 and a sampling time step of 256.

### 5.1. Deformation Fields

The results from the first model are displayed in Figure 2. The top row shows the fixed image from various horizontal and vertical perspectives, while the bottom row shows the template image from the same views. These visual results show the good alignment between template and fixed images as they are visually identical in each slice.

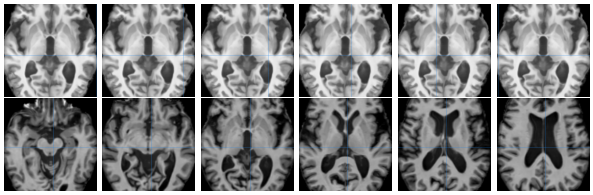

Figure 2: The top row shows horizontal slices of the fixed image from various angles. The bottom row shows vertical slices of the template image from the same respective angles.

The original DiffuseMorph model used samples from the OASIS dataset, which contained brain segmentation maps. The original evaluation process was performed using these brain segmentation maps (as they provide better insight into the similarities between the two MRI images). However, the ADNI, SRI, and NCANADA datasets used in our experiments do not contain these segmentation maps; therefore, our evaluation was conducted on the entire MRI scan instead of a subsection. This could possibly contribute to the discrepancies between our results and theirs. These errors were most likely amplified in our second model as well. Since our datasets do not include the segmentation maps with their MRI images, we decided to calculate the PSNR and SSIM scores to measure the image quality as a whole.

Upon evauation, we obtained an average Peak Signal-to-Noise Ratio (PSNR) of 2.418 dB and an SSIM score of 0.163, both of which indicate suboptimal performance. The PSNR value of 2.418 dB, is significantly lower than the typical acceptable range of 20-50 dB, suggests a high level of noise and substantial pixel-wise discrepancies between the generated and original images. Additionally, the SSIM score of 0.163 implies that the generated images exhibit structural distortions and fail to preserve the critical features of the original MRI scans. This is evidenced by our results in Figure 2, where the output image (f) has a substantially lower resolution than the fixed input image (d). We see that our model's performance lags behind that of the baseline models, which have scores ranging from 0.58 up to 0.91 (Table 1). These metrics collectively highlight the need for substantial improvements in the model's architecture, training process, and data processing techniques, to reduce the error between the generated images and the original ones.

Figure 3 shows the fixed image, deformation field, and moving image at the final epoch. The moving image is visually identical to the fixed image, indicating high accuracy of the model.

Figure 3: Fixed image (left), deformation field (center), and moving image (right) at the final epoch

These results demonstrate that the model has successfully learned to map the deformation fields from the moving image to the fixed image. These deformation fields are subsequently utilized in the second model, as described in the following section.

### 5.2. MRI Generation

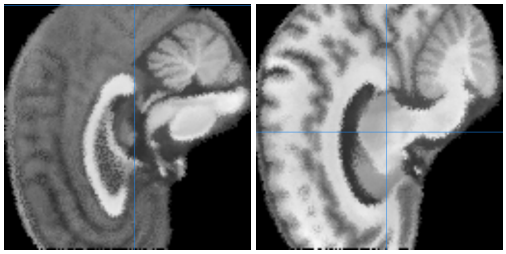

Figure 4: Two slices of a generated MRI.

Figure 4 illustrates our generated MRI from the second model. We can compare it to the baseline results in Figure 6 and as can be observed, our model performs well. Due to lack of time, we were unable to generate a large amount of images to compare the real MRI images to (besides Figure 3). The generated image shows a few deformities in both image intensity, sharpness, and orientation. Our generated image seems to have less definable features and lack the sulci and gyri that the real MRI images have. This may be due to our second model not properly learning the distribution of the deformation fields. The distribution may have been difficult for the model to learn due to the inconsistencies in the deformation field shape (32, 128, 128). Another contributor to the few differences between our generated MRI and the real MRIs may be the specific module we utilized to apply the deformation field to our template image; it may not have been properly tuned to our image and model needs. In the future, we will aim to be more consistent when modifying our image shapes. We will also be more mindful to how our deformation fields change in quality, shape, and size as it goes through the multiple diffusion steps in our second model. A way to improve our methodology would be to find a different metric to use as our loss in our second model; it would be helpful to find a metric that allows us to penalize drastic changes in the deformation fields, or one that can improve the de-noising process.

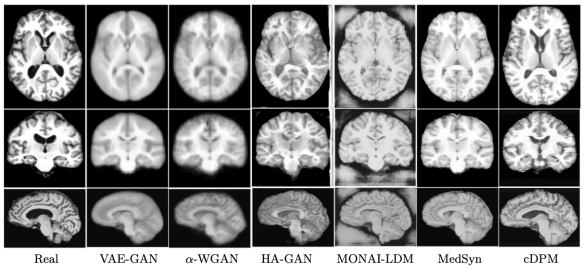


Figure 5: Real MRI vs Synthetic MRIs from different methods. Slices in axial (top), coronal (middle) and saggital (bottom) views.

### 5.3. Baseline Methods

To evaluate our model, we will compare it with several baseline models. We have provided an evaluation of the baseline methods here and will compare our final results with them in the end. The baseline methods include representative GANs and diffusion models, including VAE-GAN [7], $\alpha$-WGAN [7], HA-GAN [7], MONAI-LDM [2], MedSyn [6], and cDPM [6]. HA-GAN is a memory-efficient approach that can be trained with hierarchical sub-volumes. $\alpha$-WGAN [7] adds Wasserstein gradient penalty on top of an $\alpha$-GAN [16] for better stability. VAE-GAN [7] enhances the training of VAE using a GAN. Meanwhile, cDPM is an 2D diffusion model that generates 3D MRI slices conditioned on a number of previous slices for memory- efficient training. MedSyn model decouples the generation into a low-resolution and a high-resolution progresses. Lastly, MONAI-LDM uses a combination of a Kullback–Leibler-regulated autoencoder and a latent diffusion model [9].

Qualitative results are illustrated in Figure 5. As can be observed, there are artifacts in the background for HA-GAN and MONAI-LDM. Additionally, slices from VAE-GAN and $\alpha$-WGAN show less clear details in all 3 views. Comparing our results with these baseline methods (Figure 4), our model also shows a good performance which can be further concluded from the quantitative measurements as well.

Quantitative results are provided in Table 1. The evaluation metrics used here are Frechet Inception Distance (FID) [15], Maximum-Mean Discrepancy (MMD) [10], and Multi-Scale Structural Similarity (MS-SSIM) [4]. Table 1 shows that the quantitative evaluation based on feature extraction highly depends on the specific model used for ex-

| Model | FID1 (R50_23) | MMD1 | FID2 (R50) | MMD2 | FID 3 (R101) | MMD3 | MS-SSIM | Image MMD | Time(s) |
|---|---|---|---|---|---|---|---|---|---|
| **VAE-GAN** | 0.081 | 0.046 | 0.400 | 0.210 | 0.032 | 0.020 | **0.91** | **131925** | **0.018** |
| **$\alpha$-WGAN** | **0.060** | **0.040** | 0.480 | 0.250 | 0.032 | 0.020 | **0.88** | **203999** | **0.014** |
| **HA-GAN** | 0.079 | 0.040 | **0.080** | **0.043** | 0.035 | 0.018 | 0.78 | 759363 | 0.020 |
| **MONAI-LDM** | 1.420 | 0.69 | 1.870 | 0.940 | 0.30 | 0.150 | 0.58 | 3314614 | 10.780 |
| **MedSyn** | **0.057** | **0.037** | **0.044** | **0.034** | **0.012** | **0.010** | 0.88 | 217897 | 44.450 |
| **cDPM** | 0.140 | 0.082 | 0.130 | 0.081 | **0.019** | **0.014** | 0.75 | 586022 | 1361.0 |

Table 1: Quantitative metrics for 400 synthetic MRIs. The MS-SSIM between real sample pairs is 0.88. The top 2 performers are in bold.

tracting the features. Our model obtained an **MS-SSIM of 0.8321** and **MMD of 0.016** which showcases a good performance in comparison to existing baseline models.

## 6. Conclusion

Our model was able to generate synthetic MRI images by sampling from our previously established distribution for the deformation fields. However, its performance still has some limitations. The MRI images we have generated are visually slightly different from the real MRI images. They seem to be of lower resolution, quality, and rotated at a 90 degree angle from the original images. This may be due to a couple of factors: image pre-processing and maintaining the uniformity of the data through our transformations.

While we followed the descriptions of the original DiffuseMorph model for their image processing, there may have been inconsistencies later on when we began including alterations for our task. Additionally, the shape of the images were confusing to deal with. Since MRI images typically only have one channel instead of three (like regular RGB images), we had to reformat our data to accommodate for layers in the original model that did not follow that same metric. The numerous layers of up-sampling and down-sampling may have augmented the small errors and were displayed in our final synthetic images. We also did not scale the intensity of the data as we found it created more inconsistencies with normalizing our data. In the future, we will be more consistent with our image pre-processing and data handling, and will aim to train our models for a longer period of time. Diffusion models are notorious for being difficult to train, so spending longer training hours to train our model with better GPU resources will most likely generate more accurate results. Additionally, fixing the image shape inconsistencies will allow us to create better visualizations and have better calculations.

## 7. Acknowledgment

We are indebted to Dr. Wei Peng, a post-doc at CNS Lab at Stanford University, for proposing this project idea, his continuous mentorship, and allowing us access to their lab's GPU resources. We also would like to express our gratitude to Samir Agarwala for his mentorship on this project. We will continue improving this research project and plan to eventually submit this paper to MICCAI 2025.

## References


[1] S.A. Abdelaziz Ismael, A. Mohammed, and H. Hefny. An enhanced deep learning approach for brain cancer MRI images classification using residual networks. *Artificial Intelligence in Medicine*, 102:101779, 2020.

[2] Guha Balakrishnan, Amy Zhao, Mert R. Sabuncu, John Guttag, and Adrian V. Dalca. Voxelmorph: A learning framework for deformable medical image registration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.

[3] C. Bowles, R. Gunn, A. Hammers, and D. Rueckert. Modelling the progression of alzheimer's disease in mri using generative adversarial networks. In *Medical Imaging 2018: Image Processing*, March 2018.

[4] S.A. Brown, T. Brumback, K. Tomlinson, K. Cummins, W.K. Thompson, B.J. Nagel, M.D. De Bellis, S.R. Hooper, D.B. Clark, T. Chung, and et al. The national consortium on alcohol and neurodevelopment in adolescence (ncanda): A multisite study of adolescent development and substance use. *Journal of Studies on Alcohol and Drugs*, 76(6):895–908, Nov 2015.

[5] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

[6] Jonathan Ho, Akshara Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851, 2020.

[7] J. Hu, W. Gan, Z. Sun, H. An, and U.S. Kamilov. A plug-and-play image registration network (pirate). *arXiv preprint*, 2023.

[8] B. Kim, I. Han, and J.C. Ye. Diffusemorph: Unsupervised deformable image registration using diffusion model. In *Lecture Notes in Computer Science*, volume 13691, pages 347–364, 2022.

[9] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR (Poster)*, 2015.

[10] G. Kwon, C. Han, and D. S. Kim. Generation of 3d brain mri using auto-encoding generative adversarial networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, volume 11766 of *Lecture Notes in Computer Science*, pages 118–126, 2019.

[11] W. Peng, E. Adeli, T. Bosschieter, S.H. Park, Q. Zhao, and K.M. Pohl. Generating realistic brain mris via a conditional diffusion probabilistic model. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 14–24. Springer, 2023.

[12] W.H.L. Pinaya, M.S. Graham, E. Kerfoot, P.D. Tudosiu, J. Dafflon, V. Fernandez, P. Sanchez, J. Wolleb, P.F. da Costa, A. Patel, H. Chung, C. Zhao, W. Peng, Z. Liu, X. Mei, O. Lucena, J.C. Ye, S.A. Tsaftaris, P. Dogra, A. Feng, M. Modat, P. Nachev, S. Ourselin, and M.J. Cardoso. Generative AI for medical imaging: Extending the MONAI framework. 2023.

[13] G. Pombo, R. Gray, M.J. Cardoso, S. Ourselin, G. Rees, J. Ashburner, and P. Nachev. Equitable modelling of brain imaging by counterfactual augmentation with morphologically constrained 3D deep generative models. *Medical Image Analysis*, 84:102723, 2023.

[14] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, pages 2256–2265. PMLR, 2015.

[15] L. Sun, J. Chen, Y. Xu, M. Gong, K. Yu, and K. Batmanghelich. Hierarchical amortized GAN for 3d high resolution medical image synthesis. *IEEE Journal of Biomedical and Health Informatics*, 26(8):3966–3975, 2022.

[16] Y. Xu, L. Sun, W. Peng, S. Visweswaran, and K. Batmanghelich. Medsyn: Text-guided anatomy-aware synthesis of high-fidelity 3d ct images. 2023.