Gene-to-Image: Decoding Brain Images from Genetics via Latent Diffusion Models

Sooyeon Jeon, Yujee Song, and Won Hwa Kim(✉)

Pohang University of Science and Technology, Pohang, South Korea
{jsuyeon,yujees,wonhwa}@postech.ac.kr

Abstract. Imaging genetics, discovering associations between imaging and genetic variations, has emerged as a promising avenue to advance the understanding of neurological disorders. However, the majority of existing studies focus on selecting disease-related features to improve prediction accuracy using statistical analysis or learning-based methods. Despite the data-intensive nature in medical imaging, understanding of how genetics affect brain structures in image generation remains largely unexplored. In this paper, we propose a novel approach that generates brain images from genetics leveraging latent diffusion models. Specifically, attention-based diffusion models are conditioned on genetic information, which allows us to enhance the quality and relevance of the generated images in the context of Alzheimer's diagnosis (AD). We validated our model on T1 MRI and single nucleotide polymorphism (SNP) in the Alzheimer's Disease Neuroimaging Initiative (ADNI). Our model yields real-like synthetic images demonstrating AD-specific variation that helps to increase accuracy in a downstream classification of AD. Overall, our study highlights the potential of diffusion models in imaging genetics to facilitate accurate diagnosis and understanding of AD.

Keywords: Imaging genetics · Diffusion Models · Alzheimer's disease

1 Introduction

Imaging genetics, the study of associations between genetic variations such as single nucleotide polymorphisms (SNPs) and imaging phenotypes, explores how genetic factors contribute to variations in brain structure, function and connectivity. This interdisciplinary field offers insights into the pathology and progression of neurological disorders such as Alzheimer's disease (AD), as understanding the genetic underpinnings is crucial for developing effective diagnostic and therapeutic strategies considering the complex interplay of multiple genetic variants implicated in the disease. Since AD is a progressive disorder where early diagnosis is paramount, the pursuit of imaging genetics research holds particular significance in helping early detection and intervention strategies [3,16]. Most AD studies are primarily focused on identifying AD-specific symptoms and

I. Rekik et al. (Eds.): PRIME 2024, LNCS 15155, pp. 48–60, 2025.
https://doi.org/10.1007/978-3-031-74561-4_5

biomarkers through statistical analysis [20] or extracting features to enhance disease prediction with deep learning methods [14].

While such studies have significantly advanced understanding of pathology and progression, utilizing the relationships between sequence (i.e., gene) and structure (i.e., image) for data synthesis is heavily underexplored. In recent years, deep learning-based generative models have shown remarkable achievement in generating realistic samples [4], however, there remains a significant gap in the literature concerning integrating genetics into the medical image generation despite the rich source of information from genetic data. Generative models learn the underlying probability distribution that approximates the data distribution of a training set as accurately as possible [8]. In particular, image generation under various conditions (e.g., labels and types), highlights the potential of integrating various data modalities to guide the image generation process [13,15,27,30]. One notable example is Latent Diffusion Models (LDMs) [18], which represent images as a sequence of diffused representations, where each step in the sequence gradually transforms a latent noise vector into the desired image. LDMs demonstrated their efficiency and performance in both unconditional and conditional tasks such as text-to-image translation for natural images, conditioning textual descriptions on the generation process [29].

The example above presents an opportunity for novel research avenues in imaging genetics to leverage genetic data to enhance the quality and relevance of generated images to AD-specific patterns and the role of genetics in articulating brain structure. This research is particularly significant because it involves generating medical images, a task that is inherently challenging due to the typically limited number of available medical data. Having a limited amount of data makes the problem ill-posed, particularly when considering the high dimensionality of medical imaging such as MRI scans. To effectively learn the underlying probability distribution of this high-dimensional data, a substantial amount of training data is required. When the sample size is small, it is challenging to capture the complexity and variability inherent in the data. Incorporating additional information, such as genetic data, can mitigate this problem in that it better constrains the solution space and improves the reliability of the generated images. Further, image generation holds promise as an effective data augmentation for data-intensive medical image analysis. As a small sample size is a common problem in medical imaging studies, accurate synthetic data offer an efficient means to deal with data scarcity and address data imbalance problems and privacy issues [28].

Therefore, in this study, we propose a novel approach for generating MRI images from SNP data by harnessing the LDMs, offering a new perspective on bridging the genotype-phenotype gap. By integrating genetic information (i.e., a sequence) into LDMs via an attention mechanism, we aim to produce high-fidelity MRI that improves the accuracy and clinical relevance of imaging-based AD diagnosis. To our best knowledge, it is the first work in image generation from genetics.

The contributions of our work are three-fold: **1)** We propose a method to incorporate genome data into latent diffusion models, showcasing the use of genetics for image generation. **2)** We emphasize the effectiveness of the attention

mechanism in LDMs to capture genetic influence, enhancing the relevance of generated images to conditions. **3)** We extensively validate our model on the Alzheimer's Disease Neuroimaging Initiative (ADNI) study, which yields real-like synthetic images that potentially help training downstream prediction model.

2 Related Work

Deep Learning-Based Imaging Genetics. Several studies have employed deep learning architectures to extract features from imaging and genetic data, identifying disease-related biomarkers and developing predictive models for disease diagnosis. [11,23,24,31] proposed deep learning frameworks with deep neural networks or convolutional neural networks to extract features of each modality and integrate them as an input for the classifiers. [7] utilize a hierarchical graph convolutional network to embed genetic information and combine the latent embedding and image encoder for disease prediction.

Multivariate Statistical Technique. To maximize the correlation between genetic data and images, several studies employed Canonical correlation analysis (CCA) is typically employed to investigate relationships between genetics and images. Du et al. introduced a regularization to bring structural sparsity [6] and proposed a multi-task sparse CCA to study the relationship between the longitudinal imaging and SNPs [5]. Similarly, [12] proposed multi-task-based structured sparse CCA, integrating complementary multi-modal imaging data.

Medical Image Generation. While there has been no research on generating medical images from genomic data, significant progress has been made in generating medical images using generative models. Recent studies have employed various generative adversarial networks (GANs) [8] and diffusion models to create realistic medical images. [21] demonstrated the effectiveness of GANs in generating synthetic abnormal MRI images for data augmentation, improving the performance of lesion detection algorithms. [1] applied CycleGANs for multimodal image-to-image translation, significantly enhancing the quality of synthesized PET or MRI data. More recently, [17] explored brain imaging generation with latent diffusion models, showing the potential of diffusion models in producing high-fidelity medical images.

3 Method

We first explain latent diffusion models (LDMs) based on Denoising Diffusion Probabilistic Models (DDPMs) [10] (Sect. 3.1), then we introduce our model for generating MRI image slices from genetic data with LDM (Sect. 3.2).

3.1 Latent Diffusion Models

Diffusion Models, inspired by non-equilibrium statistical physics, slowly destroy the original data by successively adding random noise, converting data

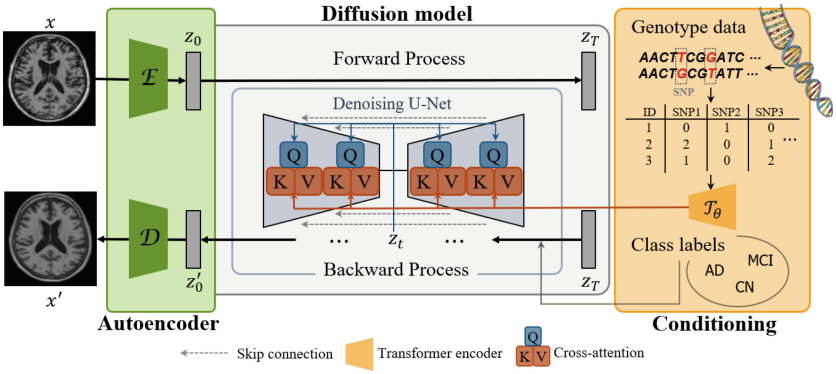

Fig. 1. The overall architecture of our model. In the autoencoder module, the encoder $\mathcal{E}$ compresses the input image into a low-dimensional latent representation, and the decoder $\mathcal{D}$ reconstructs the latent representation generated through the diffusion process back into image space. The diffusion model takes the latent transformed by the autoencoder's encoder as input, conditioning on genotype data and diagnostic labels to generate new latent representations through a denoising U-Net. Genotype data is embedded using a transformer encoder and fed into a cross-attention module, while diagnostic labels are concatenated with the input of the diffusion process.

distribution into a simple known distribution such as a Gaussian distribution (i.e., forward process) [22]. Then, the models learn to reverse this process to reconstruct data distribution, which allows us to learn a tractable and flexible distribution (i.e., reverse process). This iterative process allows the model to generate high-quality images by progressively restoring information from the noise space to the data space. For both processes, it is explicitly formulated as a Markov chain, in which each step only depends on the previous step.

For a forward process, suppose a data distribution $x_0 \sim q(x_0)$ is gradually destroyed into a tractable distribution from time $t = 0$ to $t = T$. By the Markov property, the joint distribution conditioned on x_0, denoted as $q(x_1, ..., x_T|x_0)$, can be factorized into the transition kernels.

$$q(x_1, ..., x_T|x_0) := \prod_{t=1}^{T} q(x_t|x_{t-1}) \tag{1}$$

When the noise added to data is pure Gaussian noise, the transition kernel $q(x_t|x_{t-1})$ at time step t is formulated as follows:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t\mathbf{I}), \tag{2}$$

where β_t is a variance schedule. Using Eq. (2), one can represent closed form of $q(x_t|x_0)$ at an arbitrary time t. With $\alpha_t := 1-\beta_t$, $\bar{\alpha}_t := \prod_{s=1}^{t} \alpha_s$ and a Gaussian vector $\boldsymbol{\epsilon}$, we have

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)\mathbf{I}), \quad x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}. \tag{3}$$

For the backward process, $q(x_{t-1}|x_t)$ is also Gaussian distribution as in the forward process if β_t is sufficiently small [22]. In practice, as it is computationally intractable, we approximate $q(x_{t-1}|x_t)$ by estimating the mean and variance through a parameterized model p_θ.

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \mathbf{\Sigma}_\theta(x_t, t)). \tag{4}$$

To train the parameter θ so that p_θ closely approximates the target distribution q, the learning objective L is set to minimize the Kullback-Leibler (KL) divergence between two distributions. Simplifying the learning objective to be computable, it is expressed in closed form as

$$L(\theta) \sim D_{KL}(q(x_{t-1}|x_t, x_0)\|p_\theta(x_{t-1}|x_t)). \tag{5}$$

Using Bayes' rule, one can compute the mean and variance of $q(x_{t-1}|x_t, x_0)$,

$$q(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}; \tilde{\mu}_t(x_t, x_0), \tilde{\beta}_t \mathbf{I}), \tag{6}$$

$$\text{where} \quad \tilde{\mu}_t(x_t, x_0) = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_t) \quad \text{and} \quad \tilde{\beta}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t. \tag{7}$$

If $\mathbf{\Sigma}_\theta$ is not trained, Eq. (5) is transformed to minimizing the difference of the mean, $\tilde{\mu}_t$ and μ_θ. Further, by Eq. (7), this is converted to the objective with noise matching and can be simplified to

$$L(\theta) \sim \mathbb{E}[\|\epsilon - \epsilon_\theta(x_t, t)\|_2^2]. \tag{8}$$

Latent Diffusion Models are probabilistic generative models that leverage the concept of diffusion processes to efficiently generate high-quality images. They operate diffusion processes not in the image space but on the encoded latent representation. Therefore, the LDMs consist of an autoencoder for image compression and retrieval and a diffusion model for generation.

First, the encoder of the autoencoder, $\mathcal{E}$, compresses input data x to a low dimensional latent variable $z = \mathcal{E}(x)$, where latent representations are trained to follow a Gaussian distribution. With the noise sample computed by the mean and variance resulting from encoding, the decoder, $\mathcal{D}$, produces new data similar to the input. With the trained autoencoder, the compressed latent z is fed into a diffusion model, which is trained to learn the distribution of z. The objective Eq. (8) is modified accordingly as

$$L(\theta) \sim \mathbb{E}[\|\epsilon - \epsilon_\theta(z_t, t)\|_2^2]. \tag{9}$$

At last, the diffusion model generates samples from the latent distribution by the reverse process, and the trained decoder transforms the latent samples back into the image space and reconstructs the desired output image.

3.2 Gene-To-Image Translation via Attention

Figure 1 presents an overview of our methods. MRI slices are transformed into latent representations through an autoencoder, compressing them into low-dimensional features. The latent representations undergo a diffusion process to

generate new representations, conditioned on diagnostic labels and genotype data. Diagnostic labels are concatenated with the latent representations, while genotype data is encoded by a transformer encoder and incorporated through a cross-attention module. Finally, the generated representations are reconstructed back into MR images.

The essence of our work is to incorporate genetic data into the diffusion model so that the model effectively reflects genetic factors governing the generative process on the generated images. To condition LDMs on genetic data, we preprocess genetic data with a transformer encoder and apply a cross-attention mechanism in U-Net [19]. Given genetic data $y \in \mathbb{R}^{F_g}$, where F_g denotes the dimension of genetic features (e.g. SNP features), the transformer encoder converts y to an intermediate representation $\mathcal{T}_\theta(y) \in \mathbb{R}^{F_g \times F_e}$ with the embedding dimension F_e. It is then fed into the intermediate layers of the U-Net through a cross-attention layer, implementing $Attention\ (Q,\ K,\ V) = softmax\ (\frac{QK^T}{\sqrt{d}}) \cdot V$, where

$$Q = W_Q \cdot \mathcal{I}(z_t), \quad K = W_K \cdot \mathcal{T}_\theta(y), \quad V = W_V \cdot \mathcal{T}_\theta(y). \tag{10}$$

$\mathcal{I}(z_t)$ is an intermediate output of the latent representation of the image from the U-Net and W_V, W_Q, and W_K are learnable weight matrix. Finally, we define the overall learning objective for the diffusion model with the conditions as follows:

$$L(\theta) \sim \mathbb{E}[\|\epsilon - \epsilon_\theta(z_t, t, \mathcal{T}_\theta(y))\|_2^2]. \tag{11}$$

From Eq. (11), the model jointly learns both $\mathcal{T}_\theta$ and ϵ_θ to estimate the pure noise ϵ, which allows us to estimate the mean of the backward trajectory (i.e., Eq. (4)), ultimately generating synthetic images conditioned on genetic data.

4 Experiments

4.1 Dataset and Preprocessing

Dataset. We evaluated the performance of the proposed model on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. We selected 911 samples from the ADNI-1 and ADNIGO/2 phases, and each sample contains MRI images, genotype data, and corresponding diagnostic labels. The dataset consists of 291 control (CN), 438 mild cognitive impairment (MCI), and 182 AD subjects.

Preprocessing. For the image data, the baseline T1-weighted MRI scans were extracted and we used the MNI ICBM 152 template to perform spatial normalization and correct intensity non-uniformity, following the CIVET pipeline. Then, we obtained slices of size 217 × 181 in the axial view of the images and resized them to 256 × 256 in the experiments. For genotype data, ADNI provides the SNP genotype data using the Illumina method with PLINK format, which contains 620,901 SNPs for ADNI-1 and 730,525 SNPs for ADNIGO/2. To select SNP markers statistically significant and relevant to AD, two procedures were conducted. Quality control (QC) was initially performed using PLINK software to filter out samples or SNPs based on the summary statistic measures. The

SNP data was screened out with the following conditions: missing genotype rate $> 5\%$, minor allele frequency $< 1\%$, and Hardy-Weinberg exact test $< 10^{-5}$. After the QC, 553,789 and 672,256 SNPs were retained for each phase. To further identify SNPs relevant to AD, we used SNPs belonging to the AD gene candidates listed in the AlzGene database [2] and finally selected 49 SNP markers. Each SNP value was encoded in three categories: normal (0), heterogeneous variant (1), and homogeneous variant (2). On the other hand, the $\epsilon 4$ allele of APOE is the strongest known genetic factor for AD, but it is not included in the ADNI sequencing data. Instead, ADNI provides the genotyped data of the two SNPs (rs429368, rs7412) that define $\epsilon 2$, 3, and 4 of APOE, and we extracted information only for $\epsilon 4$. It is also encoded into three categories, 0, 1, and 2.

4.2 Experimental Settings

Baselines. We validate the performance of generating MRI images with our proposed genotype-conditioned LDM, conditional generative adversarial network (CGAN) [15], two text-to-image GANs (StackGAN and AttnGAN) [27,30], and label-conditioned LDM on the ADNI dataset. In detail, we employed CGAN and label-conditioned LDM with MRI images and diagnostic labels, where the labels are concatenated to images. StackGAN and AttnGAN are text-conditioned two-stage and multi-stage GAN models based on the given text, which we replaced with genotype data. Our model includes three LDM-based models: one that uses only genotype data (G2I), one that uses both genotype data and labels without attention (G2I-l), and one that uses both genotype data and labels with attention applied (G2I-l+Att.).

Evaluation. To evaluate the quality and validity of the synthesized images, we use various metrics to 1) measure sample quality and 2) perform a downstream 3-way classification for AD diagnosis. On the image quality evaluation, we compare the Peak Signal-to-Noise Ratio (PSNR), the Fréchet Inception Distance (FID) [9], the Structural Similarity Metric (SSIM) [26], and the Multiscale Structural Similarity Metric (MS-SSIM) [25]. PSNR measures the reconstruction quality of the generated image, FID measures the distance between real and generated data distributions using features extracted from Inception-V3, and MS-SSIM and SSIM measure the similarity between samples with luminance, contrast and structure. For the classification, we used a simple 2-layered Convolution Neural Network. We use 85 % of the dataset for training and select 27 images of the remaining data as a test set for fair evaluation for each class. We augment our training dataset by combining the original images with generated images to ensure a total of 500 images per class. Additionally, we augment our training dataset by combining the original images with generated images to ensure a total of 500 images per class. We compare all models with accuracy, Macro-precision, Macro-recall, and Macro-F1-score.

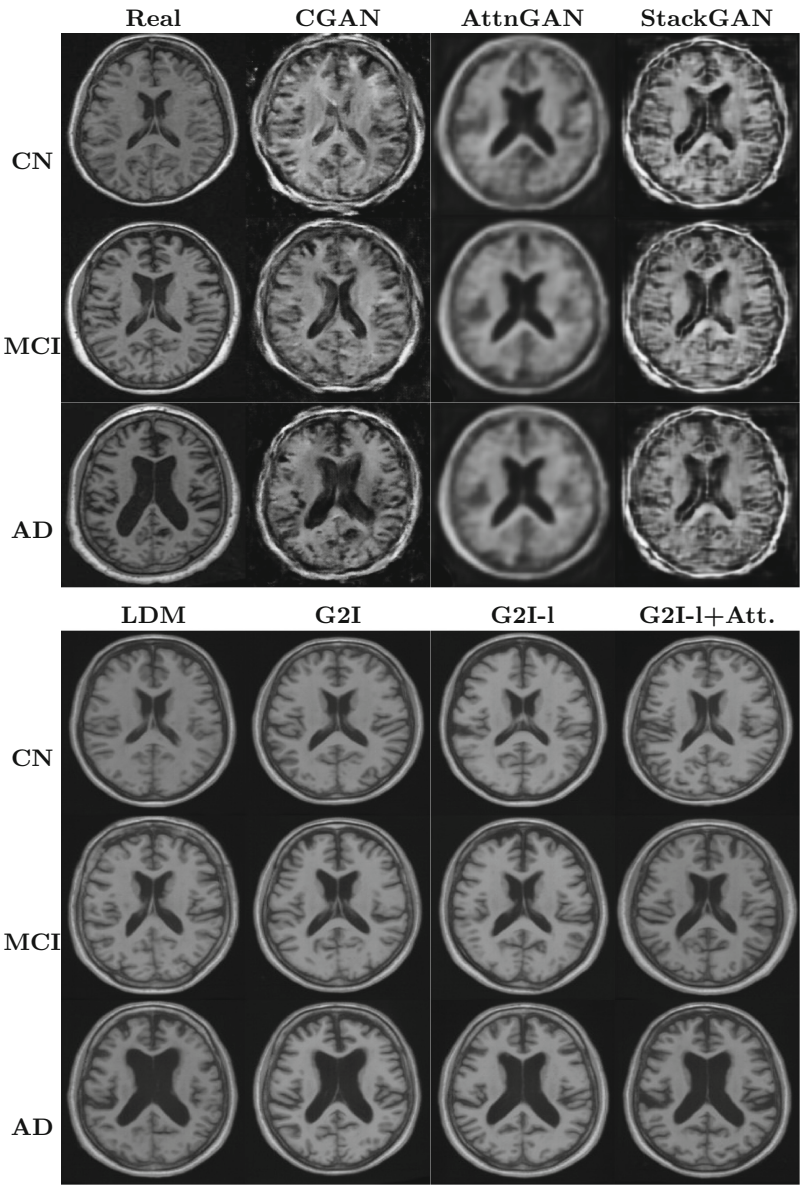

Fig. 2. Real images and the samples of brain MRI generated from baselines and our models for three diagnostic groups.

5 Experimental Results

Sample Quality Evaluation. Figure 2 shows images generated from our models alongside real images and other baseline models across three different groups. Notably, only our genotype-conditioned LDM and label-conditioned LDM

Table 1. Quantitative evaluations of samples from our models and baseline models. This is the average calculated over a total of 3960 samples.

Model	PSNR ↑	FID ↓	MS-SSIM ↓	SSIM ↓
CGAN [15]	14.900	141.300	**0.506**	**0.407**
AttnGAN [27]	**18.227**	161.976	0.873	0.887
StackGAN [30]	15.707	192.704	0.835	0.804
LDM [18]	16.491	73.260	0.654	0.608
G2I	16.889	71.551	0.704	0.648
G2I-l w/o Att.	16.968	80.711	0.733	0.675
G2I-l w/ Att.	16.825	**71.237**	0.688	0.645

produced complete images, showcasing the ability of LDM to generate visually sound MRI. The LDM-based models (LDM and G2I models) generally reflected the increased ventricle and cortical atrophy as AD progressed, generating images that closely resemble real ones. In contrast, text-to-GAN models (AttnGAN and StackGAN) exhibited consistent results across groups, with no significant difference observed. CGAN yielded images that varied by group but highlighted the challenges in generating realistic MRI images.

As seen in Table 1, Our models achieved a high PSNR score after AttnGAN, G2I-l with attention exhibited the lowest FID score, and CGAN exhibited the lowest MS-SSIM and SSIM scores. Although AttnGAN achieved the highest PSNR value, this metric can be misleading due to the blurriness in its generated images as depicted in Fig. 2. For FID scores, both G2I and G2I-l with attention exhibited low values. Specifically, G2I-l with attention showed an improvement in FID compared to using only one of the modalities. However, the FID deteriorated when the attention mechanism was not applied. Regarding MS-SSIM and SSIM, CGAN showed notably low scores due to its ability to generate different images based on diagnostic stages but inability to produce complete images, resulting in significant differences between samples (Fig. 2). Conversely, text-to-image GANs consistently displayed high SSIM and FID scores indicating that they failed to capture the underlying data distribution and generate similar images across different conditions, i.e., mode-collapse. In our models, incorporating genetic data with attention (G2I and G2I-l with attention) resulted in high PSNR and low FID. Although they exhibited a slightly higher MS-SSIM and SSIM compared to LDM, the difference was not significant. Thus, considering multiple metrics collectively, utilizing genetic data with attention effectively produces high-quality images, aligning well with the given conditions.

The effectiveness of the G2I-l with attention model in generating images that accurately reflect group-specific conditions is also shown in Fig. 3. This figure shows the group differences in mean images generated by LDM-based models. The LDM model using only labels produces significant differences within the ventricle due to generated images that appear blurred or disconnected. In contrast, the G2I-l with attention model shows differences only at the ventricle boundaries. This indicates that incorporating genetic data with an attention mechanism produces images that better represent diagnostic states.

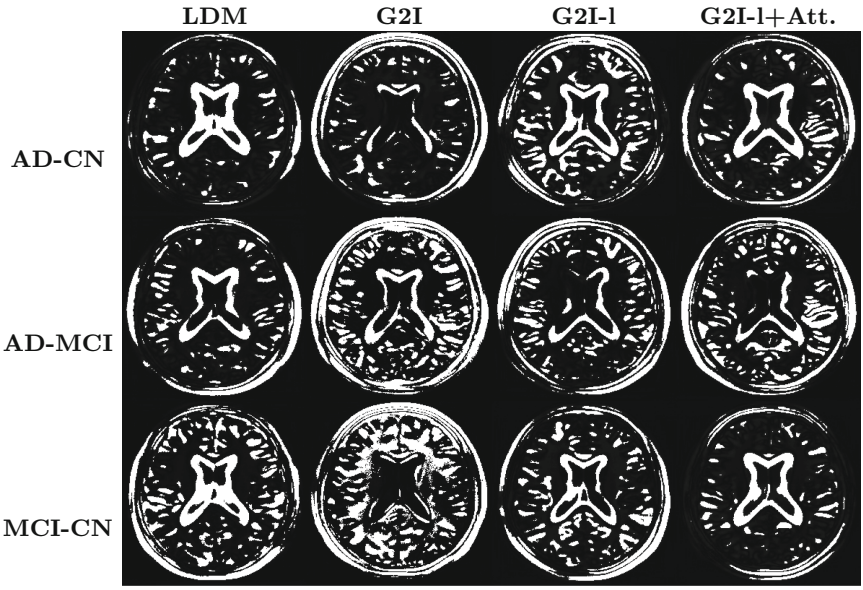

Fig. 3. Group differences in the mean images generated by LDM-based models.

Downstream Classification with Augmentation. Table 2 shows the classification performance between our models and baseline models. As shown in the table, G2I-l with attention consistently outperformed all other models across all metrics. Using only labels (LDM) to generate images resulted in an approximate 4%p improvement in accuracy while using only genetic data (G2I) did not show any difference. When both types of information were utilized with attention, accuracy improved, nearly 8%p higher compared to the original images. However, without the attention mechanism, the accuracy was lower compared to using labels alone. This underscores the effectiveness of leveraging genetic data,

Table 2. AD-stage classification performance comparisons.

Model	Accuracy (%)	Precision	Recall	F1-score
Original images	53.13	45.59	54.79	41.96
CGAN [15]	56.25	50.46	55.31	40.03
AttnGAN [27]	57.29	52.38	58.60	42.49
StackGAN [30]	55.21	44.31	54.92	37.73
LDM [18]	57.29	48.81	58.38	39.71
G2I	53.13	50.97	52.75	40.28
G2I-l w/o Att.	55.21	52.71	56.71	41.57
G2I-l w/ Att.	**61.46**	**53.45**	**63.25**	**44.28**

especially when combined with attention mechanisms for image generation, in image-based AD diagnosis. Also, the observed performance improvement, compared to using only original images, suggests that employing our model for image generation is an effective image augmentation technique, resolving class imbalances caused by the small number of AD patients and enhancing diagnostic accuracy.

6 Conclusion

We introduced a novel approach, the Gene-to-Image (G2I) model, leveraging LDMs to generate MRI slices from genotype data. G2I incorporates genotype into the diffusion model with an attention mechanism, enhancing image fidelity and relevance to Alzheimer's disease (AD) conditions. Through comprehensive validation, including assessment of image quality and AD classification accuracy, our model consistently demonstrated superior performance, affirming its effectiveness and potential use for imaging genetics.

Acknowledgement. This research was supported by NRF-2022R1A2C2092336 (70%), RS-2022-KH12 7855 (10%), RS-2022-KH128705 (10%), and RS-2019-II191906 (AI Graduate Program at POSTECH) (10%) from South Korea.

References

1. Armanious, K., et al.: MedGAN: medical image translation using GANs. Comput. Med. Imaging Graph. **79**, 101684 (2020)
2. Bertram, L., et al.: Systematic meta-analyses of Alzheimer disease genetic association studies: the Alzgene database. Nat. Genet. **39**(1), 17–23 (2007)
3. Bigos, K.L., Weinberger, D.R.: Imaging genetics-days of future past. Neuroimage **53**(3), 804–809 (2010)
4. Cao, H., et al.: A survey on generative diffusion models. IEEE Trans. Knowl. Data Eng. **36**, 2414–2830 (2024)
5. Du, L., et al.: Identifying progressive imaging genetic patterns via multi-task sparse canonical correlation analysis: a longitudinal study of the ADNI cohort. Bioinformatics **35**(14), i474–i483 (2019)
6. Du, L., et al.: Detecting genetic associations with brain imaging phenotypes in Alzheimer's disease via a novel structured SCCA approach. Med. Image Anal. **61**, 101656 (2020)
7. Ghosal, S., et al.: A biologically interpretable graph convolutional network to link genetic risk pathways and imaging phenotypes of disease. In: International Conference on Learning Representations (2021)
8. Goodfellow, I., et al.: Generative adversarial networks. Commun. ACM **63**(11), 139–144 (2020)
9. Heusel, M., et al.: GANs trained by a two time-scale update rule converge to a local NASH equilibrium. In: Advances in Neural Information Processing Systems, vol. 30 (2017)
10. Ho, J., et al.: Denoising diffusion probabilistic models. Adv. Neural. Inf. Process. Syst. **33**, 6840–6851 (2020)

11. Kanyal, A., et al.: Multi-modal deep learning on imaging genetics for schizophrenia classification. In: 2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW), pp. 1–5 (2023)
12. Kim, M., et al.: Multi-task learning based structured sparse canonical correlation analysis for brain imaging genetics. Med. Image Anal. **76**, 102297 (2022)
13. Li, B., et al.: Controllable text-to-image generation. In: Advances in Neural Information Processing Systems, vol. 32 (2019)
14. Lin, E., et al.: Deep learning with neuroimaging and genomics in Alzheimer's disease. Int. J. Mol. Sci. **22**(15), 7911 (2021)
15. Mirza, M., et al.: Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784 (2014)
16. Peng, J., et al.: Structured sparse kernel learning for imaging genetics based Alzheimer's disease diagnosis. In: Medical Image Computing and Computer-Assisted Intervention– MICCAI 2016, pp. 70–78 (2016)
17. Pinaya, W.H., et al.: Brain imaging generation with latent diffusion models. In: Mukhopadhyay, A., Oksuz, I., Engelhardt, S., Zhu, D., Yuan, Y. (eds.) MICCAI Workshop on Deep Generative Models, vol. 13609, pp. 117–126. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-18576-2_12
18. Rombach, R., et al.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10684–10695 (2022)
19. Ronneberger, O., et al.: U-Net: convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015, pp. 234–241 (2015)
20. Sheng, J., et al.: Predictive classification of Alzheimer's disease using brain imaging and genetic data. Sci. Rep. **12**(1), 2405 (2022)
21. Shin, H.-C., et al.: Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In: Gooya, A., Goksel, O., Oguz, I., Burgos, N. (eds.) SASHIMI 2018. LNCS, vol. 11037, pp. 1–11. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00536-8_1
22. Sohl-Dickstein, J., et al.: Deep unsupervised learning using nonequilibrium thermodynamics. In: International Conference on Machine Learning, pp. 2256–2265. PMLR (2015)
23. Venugopalan, J., et al.: Multimodal deep learning models for early detection of Alzheimer's disease stage. Sci. Rep. **11**(1), 3254 (2021)
24. Wang, J.X., et al.: Alzheimer's disease classification through imaging genetic data with IGnet. Front. Neurosci. **16**, 846638 (2022)
25. Wang, Z., et al.: Multiscale structural similarity for image quality assessment. In: The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003, vol. 2, pp. 1398–1402 (2003)
26. Wang, Z., et al.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
27. Xu, T., et al.: AttnGAN: fine-grained text to image generation with attentional generative adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1316–1324 (2018)
28. Yu, M., et al.: How good are synthetic medical images? An empirical study with lung ultrasound. In: International Workshop on Simulation and Synthesis in Medical Imaging, pp. 75–85 (2023)
29. Zhang, C., et al.: Text-to-image diffusion model in generative AI: a survey. arXiv preprint arXiv:2303.07909 (2023)

30. Zhang, H., et al.: StackGAN: text to photo-realistic image synthesis with stacked generative adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 5907–5915 (2017)
31. Zhou, T., et al.: Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis. Hum. Brain Mapp. **40**(3), 1001–1016 (2019)