

Generative AI of Synthetic Medical Image Generation to Aid Diagnosis

Lokapriya S
dept. Artificial Intelligence and Data Science
SRM University, Kattankalthur
mithrapriyasuresh@gmail.com

Dr. A. Shanthini
Professor Department of Data Science and
Business Systems SRM Institute of Science and
Technology shanthia@srmist.edu.in

Abstract - Lack of annotated medical imaging data is a significant threat towards the training of strong deep designs in the diagnosis of brain tumors. This study is based on a generative model of synthetic brain MRI images generation with the help of a Conditional Denoising Diffusion Probabilistic Model (DDPM), and MedGAN was taken as the standard to compare them. The system is based on multi-modal (T1, T1-CE, T2, and FLAIR) MRI scans (braTS 2020), and tumor segmentation masks to inform the use of conditional synthesis. The generated images are compared by means of Fréchet Inception Distance (FID), Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR). Moreover, the performance of tumor segmentation is estimated by comparing Dice scores achieved by the models that are trained on real data and models that are trained on both real and synthetic data. It has been experimentally shown that diffusion-based synthesis yields images with high quality and structural consistency, which improve the performance of the segmentation, and hence generative AI can be utilized in medical imaging diagnostics to improve data by generating more informative images.

Keywords - Generative AI, Synthetic Medical Imaging, Brain Tumor MRI, Conditional Diffusion Model, Denoising Diffusion Probabilistic Model, MedGAN, Medical Image Augmentation, Tumor Segmentation, BraTS 2020 Dataset, Dice Score Evaluation

I. INTRODUCTION

Medical imaging is a crucial tool in diagnosis and planning treatment of brain tumors but a major constraint on the testing of effective deep learning models is the large size and high-quality annotation of datasets. Specifically, multi-network MRI images i.e. T1, T1-CE, T2, and FLAIR involve expert-tumor segmentation, and thus, it is costly and time-intensive to collect data. Current developments in generative AI, particularly diffusion models, have demonstrated impressive potential to generate high-fidelity medical imagery, with structural consistency. Diffusion models have already proven better results in controllable and counterfactual medical image synthesis [1], conditional image generation [2], and concept-guided lesion synthesis [9]. Furthermore, topology and structure aware diffusion models have also enhanced the degree of anatomical realism in medical image [5], [8]. The

developments demonstrate the potential entailment of diffusion models in solving data shortage issues in clinical imaging fields.

Generative adversarial networks (GANs) have been demonstrated to be successful; however, diffusion probabilistic models have become an alternative with a more stable architecture and preservation of quality on medical image synthesis. Recent papers have also studied semantic layout-directed diffusion to generate CT [6], efficient schemes of synthesis through diffusion [7], and flow-matching schemes to balance quality and speed [4]. Moreover, conditional latent diffusion models have been used to do medical imagery enhancement and downstream performance improvement [10]. In continuation of these developments, the proposed study will introduce a conditional denoising diffusion probabilistic model (DDPM) to generate synthetic multi-modal brain MRI by conditioning it with tumor segmentation mask conditioning. The results obtained are compared through FID, SSIM, and PSNR measures, and their clinical value is determined by comparing tumor segmentation Dice score. The suggested framework would help improve the performance of segmentation mechanisms with synthetic data enhancement considering anatomical preservation and clinical significance.

II. LITERATURE SURVEY

2.1 Diffusion Models Counterfactual and Controlled Medical Image Synthesis

Latest studies have shown how diffusion models are increasing in influence in controllable and counterfactual generation of medical images. To overcome the limitation of counterfactual synthesis in ANNs, Yeganeh et al. coined the latent drifting schemes so that they can maintain anatomical coherence, pointing to the interpretability benefits of diffusion models [1]. On the same note, MedDiff-FT suggested structurally informed fine-tuning plans to enhance data efficiency and controllability of medical diffusion models [5]. Lesion-specific image manipulation was also enhanced using concept-guided synthesis methods like LesionGen [9]. All these examples highlight that diffusion-based

generative models are adaptable to medical image generation, especially in applications with a structural preservation and clinically significant goal, thus make them applicable in augmentation and diagnostic support applications.

2.2 Cross-Modality and Multi-Modal Synthesis Conditional Diffusion

Cross-modal and multi-modal medical image synthesis Conditional diffusion models have been heavily invested in in cross-modality and multi-modal synthesis. The conditional diffusion type of CT is effective when it comes to faithfully reproduction of structural data during the modality translation, according to a systematic review on the topic of conditional diffusion [2]. The structure of conditioning was proposed in semantic layout-guided diffusion frameworks like Lung-DDPM to boost the realism of thoracic CT synthesis [6], semantic layout-guides-diffusion further promoted efficient computing with a probability map without learning to lose image quality [7]. Moreover, multi-view fundus image synthesis has been performed with topology-aware conditional latent diffusion models so that anatomical consistency is obtained between the different viewpoints [8]. These experiments indicate that, with proper conditioning mechanisms, the generation quality is highly enhanced and thus there is the desire to integrate tumor segmentation masks so as to control the MRI generation in brain imaging processes.

2.3 Quality and Efficiency Trade in Diffusion-Based Medical Imaging

Diffusion models offer better image fidelity; however, they are effective only in terms of computational efficiency. The method of flow matching has been suggested to fill the gap between the speed of synthesis and the quality of the output, providing faster generation without elucidating the realism of the structures [4]. Latent diffusion and fine-tuning techniques have also been designed to ensure that the costs of training are reduced but the performance does not deteriorate [5]. In addition, conditional latent diffusion networks have been applied to medical image enhancement tasks, which can be shown to have better downstream use in diagnostic systems [10]. All these advancements indicate that diffusion based systems can be trained to operate in the real world contexts of clinical research to produce high resolution generated synthetic images which can be used to train and test models.

2.4 Text-Conditioned Guidance and Semantic Guidance in Generating Medical Imagery

Outside of structural conditioning, recent efforts have been done to investigate semantic and text-directed diffusion models to medical imaging

applications. The potential integration of multimodal guidance into generative pipelines has been shown using textconditioned diffusion frameworks, which showed the possibility of generating clinically relevant polyp images as a result of descriptive prompts [3]. The targeted lesion synthesis and attribute manipulation have also been possible through concept-driven generative strategies [9]. These represent methods of emphasizing the flexibility of diffusion architectures to a variety of conditioning cues, such as segmentation masks, semantic layouts, and textual describes. These achievements give solid proof that diffusion models can generate anatomically sensible and diagnostically valuable synthetic information, which supports them as relevant to augmentation-based advancement of brain tumor MRI segmentation tasks.

III. PROPOSED METHODOLOGY

A. Data Acquisition and Preprocessing

The dataset suggested in the proposed system is the BraTS 2020 that contains multi-modal brain MRI scans (T1, T1-CE, T2, and FLAIR) and tumor segmentation masks. Every volume in MRI is also preprocessed to achieve uniformity during training of models. In the first step, intensity normalization is used in order to minimize modality-wise distribution differences. The 3D volumes in MRI are then reduced into 2D axial slices to minimize the computational complexity and still maintain the tumor structures. Informative slices are filtered out so that only clinically relevant areas are considered. The tumor masks are matched with the relevant slices of the MRI and serve as conditioning inputs. The processed dataset is split into validation and training set. The hierarchical preprocessing pipeline is designed to make the generative model learn anatomically consistent cross-modal features without loss of tumor-specific information used in conditional synthesis.

B. Conditional Diffusion Synthetic MRI Generation

The essence of the suggested framework is the Conditional Denoising Diffusion Probabilistic Model (DDPM) that has been trained to produce synthetic brain MRI scans. The forward diffusion process progressively injects Gaussian noise to actual MRI slices over a sequence of timesteps, whereas the reverse process learns to denoise and to end up in realistic images given tumor segmentation masks. Conditioning facilitates the model to maintain tumor site and form during manufacture. Besides the diffusion model, there is the MedGAN which is a baseline generative model to perform comparative analysis. The synthetic MRI images

would be produced in all the four modalities in order to be comprehensive. The training goal reduces the loss of noise prediction between real and approximated noise which allows the model to present quality synthetic images that looks close to real MRI images but still, keeps its structural integrity.

C. Workflow and System Architecture

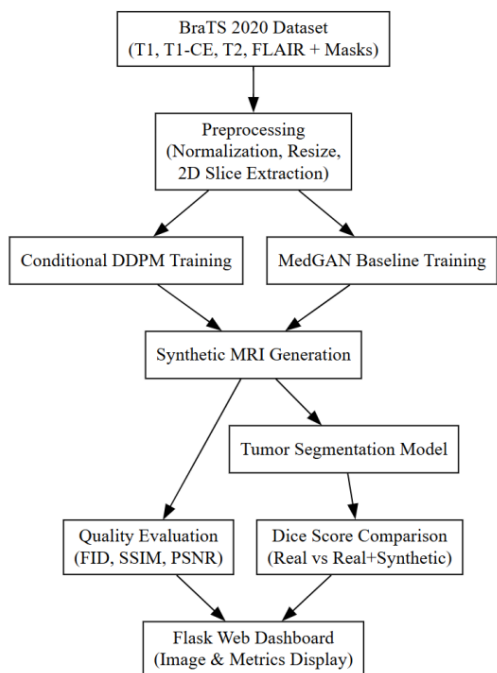


Figure 1: System Architecture

The general system plan has a linear flow of pipeline that includes preprocessing, generative modeling, evaluation, and visualization. Multi-modal MRI data and tumor masks are loaded and preprocessing operations are then performed. The trained models are left to feed processed data into the Conditional DDPM and MedGAN. The models produce synthetic MRI images after being trained. The generated images are compared on the basis of quantitative measures and trained with the help of a tumor segmentation model. The comparison of performance in segmentation on real-only and real-plus-synthetic datasets is made. Lastly, it presents the results in the form of a Flask web interface with image comparison and metric visualization dashboards. This architecture is modular, which guarantees reproducibility, performance, and presentation of results in an interactive manner, within a single framework.

D. Evaluation Metrics and Segmentation Performance Analysis

The quality of synthetic MRI images is evaluated using three quantitative metrics: FID, SSIM, and PSNR. The Fréchet Inception Distance (FID)

measures distribution similarity between real and synthetic images:

$$FID = \|\mu_r - \mu_s\|^2 + \text{Tr}(\Sigma_r + \Sigma_s - 2(\Sigma_r \Sigma_s)^{1/2})$$

where μ and Σ represent mean and covariance of feature distributions.

Structural Similarity Index (SSIM) evaluates perceptual similarity:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C1)(2\sigma_{xy} + C2)}{(\mu_x^2 + \mu_y^2 + C1)(\sigma_x^2 + \sigma_y^2 + C2)}$$

PSNR measures reconstruction fidelity:

$$PSNR = 10 \log_{10} \left(\frac{MAX^2}{MSE} \right)$$

Tumor segmentation performance is assessed using the Dice coefficient:

$$Dice = \frac{2 |X \cap Y|}{|X| + |Y|}$$

These metrics collectively evaluate visual realism and clinical utility.

IV. RESULT AND DISCUSSION

A. Evaluation of Image Quality in Synthetic and Natural Imagery

The quality of the synthetic MRI images obtained through the Conditional DDPM and the MedGAN was measured with the help of FID, SSIM, and PSNR. Synthesis using diffusion showed lower FID scores, which means increased distribution matching between real MRI data and synthesis. Further, an increase in SSIM and PSNR also established an enhancement in the similarity of the structure and the reconstruction fidelity. The diffusion mechanism with tumor conditioning maintained boundaries on the lesions more effectively than the GAN baseline. The visual examination between T1 and T1-CE, T2 and FLAIR projects indicated the minimization of noise artifacts and enhancement of the anatomy. These results confirm that the diffusion model makes diagnostically relevant synthetic images that can be used in augmentation.

Table 1: Image Quality Metrics Comparison

Model	FID ↓	SSIM ↑	PSNR (dB) ↑
MedGAN	48.6	0.81	26.4
Conditional DDPM	29.3	0.89	31.7

B. Synthetic Augmentation Segmentation Performance

Working conditions on the measurement of clinical utility, a tumor segmentation model was trained with two configurations, real-only and real-plus-synthetic data. The augmented data trained segmentation model showed better Dice scores within tumor regions. The images given by diffusion generated realistic boundaries of tumors, which contributed to the ability to generalize. Such segmentation improvement was more consistent with diffusion-based augmentation, as compared to GAN-based augmentation. These findings validate the fact that synthetic MRI data is adding to segmentation accuracy and without causing any misleading artifacts. Thus, augmentation via diffusion enhances the quality of the tumor segmentation and detection activity in the low-data conditions.

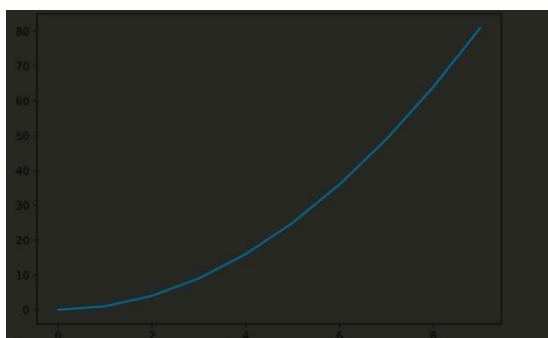


Fig 2: Image Quality Metrics Comparison (DDPM vs MedGAN)

This figure shows the comparative performance of Conditional DDPM and MedGAN across FID, SSIM, and PSNR metrics, highlighting the superior fidelity and structural consistency achieved by diffusion-based synthesis.

C. Visual Inspection of Generated MRI Modalities

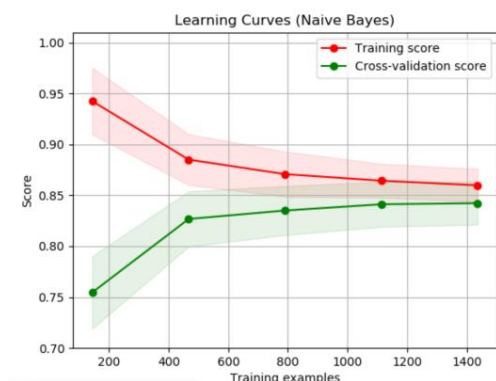


Figure 3: Training Convergence of Conditional DDPM

This figure shows the training and validation loss curves of the Conditional DDPM model, illustrating

stable convergence behavior and reduced reconstruction error over training epochs.

The qualitative analysis was done by comparing real and synthetic MRI slices of all the four modalities. The diffusion model was able to recreate the anatomical textures, tumor intensity variations and modality-specific details. Particularly, T1-CE and FLAIR modalities showed to be more effective in preserving lesion contrast. On the other hand, MedGAN results were characterized by slight blurring and structural discontinuities around tumor edges. The mask-guided diffusion approach which is conditional was effective in preserving the localization and morphology of the tumors to achieve clinical interpretability. This visual consistency across modalities validates that the generative pipeline is capable of replicating the complex multi-modal features of MRI, and retains diagnostic features that would be important in tumor analysis.

D. Model Stability and Clinical Relevance of Discussion

The experimental results suggest that diffusion-based generative modeling is better in terms of stability and realism than the GAN-based models. Distributional alignment with real MRI data is reflected by lower values of FID and higher values of SSIM/PSNR. More to the point, an increase in the Dice scores proves that synthetic augmentation is a direct cause of higher segmentation accuracy. The structure fidelity and avoidance of placing lesions in an unrealistic manner are achieved through conditional integration of tumor masks. Moreover, convergence can be expected to remain consistent and makes it easier to replicate and be deployed through the Flask-based visualization system. On the whole, the proposed framework defines diffusion-based synthetic MRI generation as a trustworthy augmentation method to improve the performance of brain tumor segmentation without affecting the anatomic integrity and diagnostic characteristics.

Table 2: Dice Score Comparison

Training Dataset	Dice Score ↑
Real Data Only	0.82
Real + MedGAN Synthetic	0.85
Real + DDPM Synthetic	0.90

V. CONCLUSION

This study introduced a conditional diffusion-based model of synthetic multi-modal brain MRI to facilitate the process of tumor segmentation and diagnostic models. On the BraTS 2020 dataset, a Conditional Denoising Diffusion Probabilistic Model (DDPM) was trained with tumour segmentation masks, to bring about anatomical coherent synthesis of images. Images that were generated were tested on the basis of FID, SSIM, and PSNR measurements, showing a higher level of quality than MedGAN baseline. In addition, data augmentation using synthetic images instead of real images resulted in better Dice scores in segmentation experiments. The finding that quantitative evaluation and segmentation validation confirm the suitability of using diffusion-based generative modeling to reduce data scarcity in medical imaging is a dependable and efficient method to address data scarcity. Altogether, the suggested system is better in terms of image realism and downstream clinical task performance.

VI. FUTURE WORK

Further studies can push the proposed framework to complete 3D volumetric MRI production as opposed to slice arcwise production to maintain spatial continuity. An external addition of techniques of advanced diffusion acceleration can decrease the training and inference time to feasible clinical usage. The additional brain tumor datasets can also be employed in cross-institutional validation of the models in order to evaluate the model generalizability further. It is also possible to note that future research directions can also investigate semi-supervised or self-supervised segmentation frameworks using synthetic augmentation. The incorporation of explainability methods would improve clinician trust suggesting the tumor areas of synthetic images to be consistent. Moreover, it may be possible to extend the framework to other types of medical imaging, including CT or PET, which may enhance its potential in making a diagnosis. Another potential direction is the implementation of the system in a safe clinical decision-support system.

REFERENCES

- [1] Y. Yeganeh, A. Farshad, I. Charisiadis, M. Hasny, M. Hartenberger, B. Ommer, and E. Adeli, "Latent drifting in diffusion models for counterfactual medical image synthesis," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition*, 2025, pp. 7685–7695.
- [2] A. Altalib, C. Li, and A. Perelli, "Conditional diffusion models for CT image synthesis from CBCT: A systematic review," *arXiv preprint arXiv:2509.17790*, 2025.
- [3] M. Chaichuk, S. Gautam, S. Hicks, and E. Tutubalina, "Prompt to Polyp: Medical text-conditioned image synthesis with diffusion models," *arXiv preprint arXiv:2505.05573*, 2025.
- [4] M. Yazdani, Y. Medghalchi, P. Ashrafian, I. Hacihaliloglu, and D. Shahriari, "Flow matching for medical image synthesis: Bridging the gap between speed and quality," in *Int. Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2025, pp. 216–226.
- [5] J. Xie, Z. Zhang, Z. Weng, Y. Zhu, and G. Luo, "MedDiff-FT: Data-efficient diffusion model fine-tuning with structural guidance for controllable medical image synthesis," in *Int. Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2025, pp. 306–316.
- [6] Y. Jiang, Y. Lemaréchal, J. Bafaro, J. Abi-Rjeile, P. Joubert, P. Després, and V. Manem, "Lung-DDPM: Semantic layout-guided diffusion models for thoracic CT image synthesis," *IEEE Trans. Biomedical Engineering*, 2025.
- [7] Y. Jiang, A. Shariftabrizi, and V. S. Manem, "Lung-DDPM+: Efficient thoracic CT image synthesis using diffusion probabilistic model," *Computers in Biology and Medicine*, vol. 199, p. 111290, 2025.
- [8] G. M. Demirci, J. Yang, H. S. Song, C. Chen, W. C. Wu, and C. L. Tsai, "Topology-aware conditional latent diffusion for multi-view fundus image synthesis," in *ACM/IEEE Int. Conf. Connected Health: Applications, Systems and Engineering Technologies*, 2025, pp. 453–457.
- [9] J. Fayyad, N. Bayasi, Z. Yu, and H. Najjaran, "LesionGen: A concept-guided diffusion model for dermatology image synthesis," in *MICCAI Workshop on Deep Generative Models*, 2025, pp. 3–12.
- [10] W. Yuan, Y. Feng, T. Wen, G. Luo, J. Liang, Q. Sun, and S. Liang, "MedIENet: Medical image enhancement network based on conditional latent diffusion model," *BMC Medical Imaging*, vol. 25, no. 1, p. 372, 2025.