Tumor Progression Dynamics - A Diffusion Model Approach

First Author
Mayukha Sista / FE67016
fe67016@umbc.edu

Second Author
Prashanth Reddy Pavudala / CI65331
ci65331@umbc.edu

Abstract

One of the biggest health issues faced by humans till date is Tumor(Cancer). Advanced prediction models are necessary for accurate diagnosis and treatment planning as it is one of the major causes of death worldwide. Complex biological mechanisms, including dynamic interactions between cancer cells and their surroundings, drive the growth and progression of tumors. It is essential to accurately model these processes in order to comprehend tumor behavior, forecast results, and improve treatment approaches. In order to replicate the formation and evolution of tumors, this study uses diffusion models, a family of generative deep learning models. These models, draw inspiration from physical diffusion processes, are ideal for capturing the complex dynamics of tumor progression because they learn to produce accurate data by iteratively adding and removing noise. By facilitating treatment planning, permitting predictive simulations of tumor behavior, and producing synthetic datasets for training more machine learning models, this work has the potential to advance personalized medicine.

1 Motivation

It is important to predict the growth of tumors over time to enhance treatment planning, especially in the instance of breast cancer, where precise and timely assessment can have a major impact on patient outcomes. MRI and CT scans, taken sequentially, are being utilized routinely by oncologists to assess the growth of tumors, yet manual examination of the scans is time-consuming, subjective, and error-prone. Precise tumor growth prediction can maximize individualized treatment planning, facilitate early intervention, and give important early warning for individuals with aggressive or complicated disease profiles. Conventional statistical modeling cannot model the nonlinear and multifactorial dynamics of tumor growth under heterogeneous biological and therapeutic conditions. Deep learning advances, especially diffusion models, provide a robust solution by parameterizing dense data distributions and producing realistic high-quality outputs. They are particularly good at capturing the sophisticated growth patterns of breast tumors seen in clinical images. This research leverages the generative capabilities of diffusion models to create a robust system for breast cancer progression prediction, ultimately aiming to assist in treatment planning, improve response prediction, and facilitate earlier and more personalized cancer treatment.

2 General Questions

How can diffusion models be adapted to accurately simulate the spatial and temporal dynamics of tumor growth In order to guarantee accurate simulations, this entails investigating how to successfully incorporate biological restrictions, such as growth rates and tissue boundaries, into the model. **How do amount and quality of data affect the training of diffusion models for tumor progression** This question investigates how the model's performance is affected by various forms of medical imaging data such as MRI and CT scans and preprocessing methods. **Is it possible for diffusion models to produce artificial tumor pictures that are identical to actual clinical data**. This entails assessing the model's capacity to generate high-fidelity tumor images using measures such as the structural similarity in-

dex (SSIM) and Dice coefficient. **In what ways may these models be applied to forecast how a tumor will react to therapies like radiation or chemotherapy.** This inquiry investigates how diffusion models might be used to model treatment outcomes and guide individualized treatment plans. The initiative intends to improve our knowledge of tumor dynamics and offer a potent instrument for cancer research and therapeutic decision-making by answering these concerns.

3 The Proposed Solution

Our suggested method uses generative models based on diffusion to accurately model the formation and evolution of tumors. Diffusion models capture the complicated, non-linear dynamics of tumor progression by learning directly from real-world medical imaging data (such as MRI and CT scans), in contrast to typical mathematical models that rely on oversimplified assumptions. The model functions in two stages: **(1) a forward process in which tumor images are gradually subjected to noise in order to simulate degradation, and (2) a reverse process in which a neural network is trained to denoise the data in order to effectively learn the underlying distribution of tumor growth patterns**. Personalized simulations of tumor progression over time can be produced by conditioning the model with patient-specific information, including tumor size, location, and microenvironmental characteristics. This method provides a more nuanced depiction than deterministic models since it naturally takes into consideration the heterogeneity in tumor behavior.

4 Previous Work

The existing models for forecasting tumor growth utilized mainly Convolutional Neural Networks (CNNs), numerous convolutional layers and residual connections to forecast pathological complete response (pCR). Yet, the static nature of these CNN-based models failed to model temporal growth, which is extremely important in modeling tumor growth. Other methodologies also used pre-trained CNN models, concatenating early-treatment and pre-treatment features. Though successful in certain instances, these models were limited by the utilization of pre-trained weights and were not as flexible to adapt to new and varied data. The recent advances of deep learning, specifically the use of Generative Adversarial Networks (GANs), demonstrated some improvement. GANs were used to generate future tumor states from previous scans, and biophysical models were used to simulate tumor growth. Despite these improvements, GAN-based approaches suffer from instability during training and can produce less realistic outputs, while biophysical models require extensive manual feature engineering and often lack the ability to generalize across diverse tumor types. Another prominent direction was the application of soft-attention mechanisms to NAC response prediction. These models were able to highlight informative areas of images, improving interpretability. Yet, the models overfit, especially when the attention maps were not well correlated with tumor areas. Such drawbacks point to the limited potential of CNN-based methods, mainly in temporal dynamics modeling and tolerating tumor data heterogeneity. **Denoising Diffusion Probabilistic Models (DDPMs)** are the diffusion models that have come forward as viable competitors to generating high-fidelity images recently. Diffusion models progressively denoise data, adding stability and increased realism. Diffusion models are applied in the project to model tumor growth accurately, which is a significant advance over traditional modeling methods. In contrast to this, our proposed model differs fundamentally from these convolution-based methods. Instead, we leverage the generative capability of Denoising Diffusion Probabilistic Models (DDPMs) for modeling complex distributions of tumor growth patterns effectively. The DDPM explicitly deals with noise through its forward and reverse diffusion processes. The model captures the stochastic process of tumor growth by adding and subtracting noise progressively, resulting in more realistic simulations. In addition, the inclusion of skip connections in the UNet architecture guarantees no loss of fine-grained spatial details in denoising, hence structural integrity, which is crucial in medical imaging. Our DDPM model circumvents the aforementioned limitations of CNNs through the exploitation of the temporal modeling strength of diffusion processes. In contrast to CNNs, which mainly target spatial patterns, the stochastic process of DDPM enables it to naturally model temporal dynamics in tumor growth. Moreover, DDPMs' progressive noise removal process naturally accommodates the analysis of longitudinal data, making our method more aligned with

the dynamics of tumor growth. Another significant benefit of our approach is that it is data-driven from start to finish, removing the need for pretrained weights. This enables it to adapt better to the heterogeneity of the tumor data. The diffusion steps allow the model to pick up the actual distribution of the patterns of the tumor growth instead of being influenced by the prior knowledge learned, and thus being particularly useful for personalized tumor datasets. Furthermore, our model takes advantage of skip connections and multi-scale feature extraction through the UNet architecture, which successfully integrates local and global information. The iterative denoising procedure inherently emphasizes high-frequency details, which is important for accurate segmentation and progression prediction. This combination of generative modeling, temporal dynamics, and robust data adaptation makes our approach far more complete and effective than standard CNN-based methods.

5 Methodology

The project utilizes the ISPY1 dataset, specifically focusing on breast cancer analysis which is of 75GB data. This dataset comprises DICOM images obtained through longitudinal scans of patients at intervals of several months. To ensure consistency and accuracy, the raw DICOM images undergo comprehensive preprocessing and cleaning. This process includes addressing missing or invalid values, converting the images to grayscale for uniformity, standardizing date formats, and handling diverse image shapes by removing unnecessary singleton dimensions. Additionally, the DICOM images are efficiently converted to the .npy format, significantly enhancing processing speed and facilitating streamlined analysis. Data splitting is one of the important steps in machine learning data preparation. The data is now divided into three portions: training (80%), validation (10%), and test (10%) sets. To achieve this, the script initially utilizes the `train_test_split()` function from the Scikit-learn library to separate the data into an 80% training set and a 20% temporary set. The temporary set is then further split equally into validation and test sets, each comprising 10% of the total data.

The data is first converted from NumPy arrays to PyTorch tensors to make it compatible with PyTorch models. The data is then organized into 'TensorDataset' objects, which wed each group of features to their labels. To facilitate efficient batch processing, the 'DataLoader' objects are instantiated for the training, validation, and test datasets with a batch size of 32. The training data loader shuffles the data to augment generalization, but the validation and test loaders preserve order for deterministic testing.

5.1 UNet

The UNet architecture used in the project is a well-known deep learning model for mainly image segmentation. It was originally proposed for biomedical image segmentation but was subsequently used extensively for other image analysis tasks. The code below is a simple implementation of the UNet model using PyTorch and includes three significant components: The Encoder, the Bottleneck, and the Decoder. The UNet architecture used in the project is a well-known deep learning model for mainly image segmentation. It was originally proposed for biomedical image segmentation but was subsequently used extensively for other image analysis tasks. The code below is a simple implementation of the UNet model using PyTorch and includes three significant components: **The Encoder, the Bottleneck, and the Decoder.** The architecture of UNet is symmetric "U" shape, Encoder that extracts the features from the input image and Bottleneck is the lowest point of the network that has the most compact representation and Decoder which reconstructs the segmentation map while adding high-resolution features through skip connections from the encoder.

5.1.1 Encoder

The feature extraction encoder consists of a sequence of convolutional blocks with each being trailed by a Max Pooling layer for down sampling spatial dimensions. The model uses two convolutional layers of kernel size 3 and padding 1 to maintain spatial dimensions, batch normalization to normalize training, and ReLU activation to add non-linearity. The input image is increasingly downsampled, decreasing the spatial resolution and the number of feature maps. This aids in capturing contextual information at different scales.

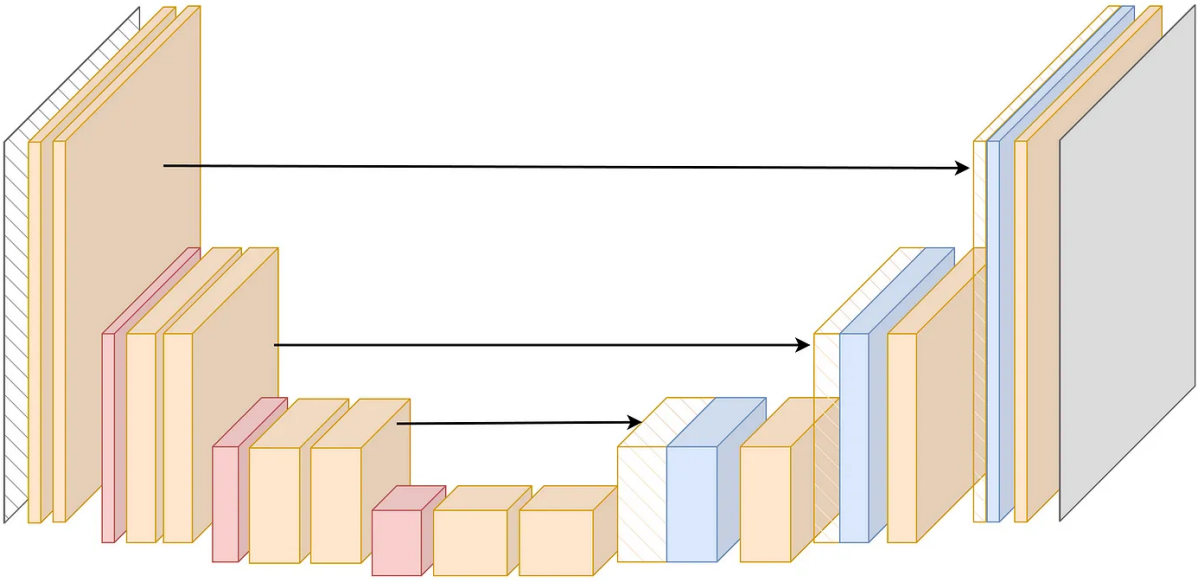

Figure 1: UNet Architecture

5.1.2 Bottleneck

The bottleneck serves as an interface between the decoder and encoder. It does two convolutional operations like in the encoder blocks. Represents the most abstract and compact feature representation. The bottleneck captures high-level features without being computationally costly since the spatial dimension is small.

5.1.3 Decoder

The decoder creates the segmentation map by progressively upsampling the spatial dimensions via transposed convolutions, which are also known as deconvolutions. It uses ConvTranspose2d for upsampling, doubling the spatial size. It combines the feature maps of the encoder using skip connections. Every decoder block is succeeded by convolutional layers in order to enhance the reconstructed feature maps. The skip connections maintain spatial details from the encoder lost as a result of downsampling.

5.1.4 Skip Connections

One of the fundamental benefits of UNet is the utilization of skip connections, which link corresponding decoder and encoder layers directly. They solve the problem of vanishing gradients and enhance the model's capacity to learn more nuances. These are attained through the utilization of torch.cat(), a function that concatenates the upsampled feature maps and the encoder feature maps.

5.2 Denoising Diffusion Probabilistic Model

The DDPM is a generative model that learns to generate high-quality images by progressively eliminating noise from noisy data via a diffusion process. The model includes a forward diffusion process (for adding noise) and a reverse diffusion process (for denoising) where a UNet is used as the base architecture.

5.2.1 Forward Diffusion Process

The forward diffusion process progressively adds noise to the input image by a sequence of time steps, yielding a very noisy image. Step Selection: The time step t is chosen randomly for every image in the batch. Noise Addition: The model calculates the square root of alpha and square root of one minus alpha at the chosen time step. Noise from a standard normal distribution is added to the image according to these coefficients. The function outputs the noisy image, the noise, and the time step.

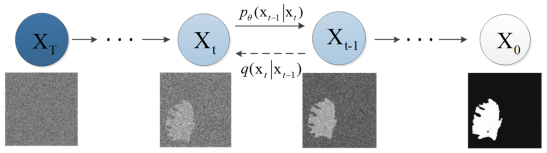

Figure 2: Diffusion Process

5.2.2 Reverse Diffusion Process

The reverse process attempts to denoise the image by estimating the noise and removing it. The reverse process uses the noisy image x_t and the timestep t as inputs. The UNet approximates the noise component introduced at the selected timestep. The result is a less noisy image, which approaches the original image.

The model gets trained, validated and tested with these two steps which are the main ones in the DDPM model. The model uses the Adam optimizer, which is well-suited for training deep neural networks due to its adaptive learning rate. Cosine Annealing Scheduler is used to reduce the learning rate smoothly across the epochs, fostering convergence. The model training process leverages PyTorch Lightning's Trainer to optimize performance and streamline the training workflow. To enhance efficiency, early stopping is implemented, which halts training if the validation loss does not improve for 10 consecutive epochs, preventing overfitting and unnecessary computations. To maintain the model's best state, checkpointing is enabled, storing model checkpoints with minimum validation loss while training. Also, learning rate monitoring is added, logging the learning rate every step to facilitate analysis and tuning. In order to deliver better computational efficiency, the model employs mixed precision training with 16-bit precision, which runs the training much faster on CUDA-enabled hardware while minimizing the memory usage. The training setup also employs dynamic device management,+ where it dynamically employs CUDA if present; else, it defaults to the CPU, thereby ensuring compatibility on varying hardware setups.

6 Results

The process of training consisted of optimizing the model to reduce the Mean Squared Error (MSE) of predicted noise and noise actually added in the process of diffusion. Training loss and validation loss were also monitored over epochs to observe the learning of the model. The training loss kept declining, indicating that the model was effectively learning to predict noise. The initial epochs had a very high loss, but the values dropped consistently as the model adapted to the data.

Validation loss also followed the same trend with a slight rise in some epochs, indicating potential overfitting. However, the overall decline in validation loss indicates that the model did not lose its generalization ability.

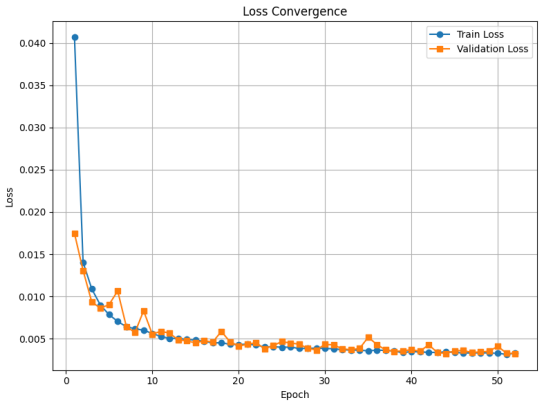

Figure 3: Convergence Loss

For robustness, we employed early stopping to halt training in case there was no improvement in validation loss for 10 epochs in a row. Besides, cosine annealing learning rate scheduling maintained the convergence stable.

The model's denoising capability was subjectively assessed through visual comparison of noisy, original, and denoised images. Images were presented side by side for easy visual judgement.

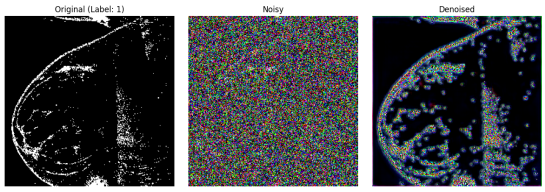

Figure 4: Denoising of a breast MRI slice showing original, noisy, and reconstructed images

The original images contained the tumor regions with distinct boundaries and structural details. The noisy images, on the other hand, were greatly degraded by the Gaussian noise imposed in the forward process.

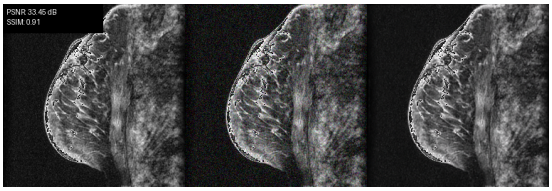

Figure 5: Reconstruction showing ground truth, noisy input, and denoised output with PSNR 33.45 dB and SSIM 0.91

The reconstructed denoised images from the

model through DDPM were not similar to the original images, and the structural detail did not restore as anticipated. Rather, the model output had distortions and artifacts. Qualitative assessment illustrated the loss of significant spatial features, especially after introducing a high noise level. The model did not learn to capture the patterns of tumor growth distribution and, produced outputs that were visually inferior to the noisy inputs.

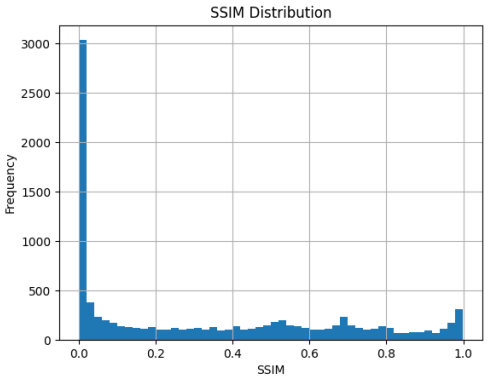

Figure 6: Histogram showing the distribution of Structural Similarity Index Measure values between the generated and ground truth images

7 Limitations

The model showed poor robustness to noise features. The model, however, did not denoise well. Not only did the model fail to reconstruct the original images well, but it also generated images that were more distorted than the noisy inputs.

The diffusion process is multi-step iterative to denoise the image progressively, which is computationally expensive. Training the DDPM model required large-scale GPU resources and a prolonged training time. This complexity poses a challenge to real-world deployment, especially in resource-constrained environments like real-time medical imaging systems. Further, the inference time was relatively slow and thus was not apt for applications requiring quick decision-making, e.g., real-time diagnosis.

Although it produces denoised images, the model is not very interpretable, and it is hard to comprehend how noise is eliminated and what features of the images are retained. Model interpretability is very important in medical imaging since physicians and healthcare experts must trust and validate the model outputs. The inclusion of attention mechanisms or feature visualization methods would enhance model transparency.

8 Future Work

The project utilized the ISPY1 dataset, comprising longitudinal breast cancer image data of total size approximately 75GB. Since the size of the dataset is very large, its processing on regular personal systems is highly infeasible. Instead, it needs a more advanced and high-performance computational setup capable of processing such large-scale data effectively. Noise-specific augmentations that mimic medical imaging variations (e.g., MRI noise patterns) can help the model generalize better to real data. Adding a multi-modal fusion strategy to the UNet architecture can enable the model to benefit from complementary information. After achieving desirable accuracy for the model, the long-term final goal would be the deployment of the model in real clinical environments. Creating an accessible interface through which clinicians can upload images and get denoised outputs in real time would be a usability improvement. The model may be modified to include patient-specific information (e.g., age, tumor type) to generate patient-specific denoising profiles with improved predictive power for specific cases. The proposed short-term and long-term follow-up projects have the potential to notably improve its performance, generalizability, and feasibility in real-world applications.

References

[1] D. Newitt and N. Hylton, on behalf of the I-SPY 1 Network and ACRIN 6657 Trial Team, "Multi-center breast DCE-MRI data and segmentations from patients in the I-SPY 1/ACRIN 6657 trials," *The Cancer Imaging Archive*, 2016. [Online]. Available: https://doi.org/10.7937/K9/TCIA.2016.HdHpgJLK

[2] Y. Liu, S. M. Sadowski, A. B. Weisbrod, E. Kebebew, R. M. Summers, and J. Yao, "Patient specific tumor growth prediction using multimodal images," *Medical Image Analysis*, vol. 18, no. 3, pp. 555–566, Apr. 2014. doi:10.1016/j.media.2014.02.005.

[3] Q. Liu, E. Fuster-Garcia, I. T. Hovden, B. J. MacIntosh, E. O. S. Grødem, P. Brandal, C. Lopez-Mateu, D. Sederevičius, K. Skogen, T. Schellhorn, A. Bjørnerud, and K. E. Emblem, "Treatment-aware Diffusion Probabilistic Model for Longitudinal MRI Generation and Diffuse Glioma Growth Prediction," *IEEE Trans. Med. Imaging*, early access, 2025. doi:10.1109/TMI.2025.3533038.

[4] A. Elazab, C. Wang, S. J. S. Gardezi, H. Bai, Q. Hu, T. Wang, C. Chang, and B. Lei, "GP-GAN: Brain tumor growth prediction using stacked 3D generative adversarial networks from longitudinal MR Images," *Neural Networks*, vol. 132, pp. 321–332, Dec. 2020. doi:10.1016/j.neunet.2020.09.004.

[5] N. Meghdadi, M. Soltani, H. Niroomand-Oscuii, and N. Yamani, "Personalized image-based tumor growth prediction in a convection–diffusion–reaction model," *Acta Neurol. Belg.*, vol. 120, no. 1, pp. 49–57, Feb. 2020. doi:10.1007/s13760-018-0973-1.

[6] S. Zeng, J. Liu, J. Xu, and Y. Luo, "MBT-Diff: Multi-segmentation Brain Tumor Model with Diffusion Probabilistic Model," in *2024 9th International Conference on Image, Vision and Computing (ICIVC)*, 2024, pp. 188–192. doi:10.1109/ICIVC61627.2024.10837511.

[7] Z. Dorjsembe, H.-K. Pao, S. Odonchimed, and F. Xiao, "Conditional Diffusion Models for Semantic 3D Brain MRI Synthesis," *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 7, pp. 4084–4093, Jul. 2024. doi:10.1109/JBHI.2024.3385504.

[8] Q. Tang, Q. Zhu, Y. Xiong, Y. Xu, and B. Du, "Edge-and-Mask Integration-Driven Diffusion Models for Medical Image Segmentation," *IEEE Signal Processing Letters*, vol. 31, pp. 2665–2669, 2024. doi:10.1109/LSP.2024.3466608.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *arXiv preprint arXiv:1505.04597*, 2015. [Online]. Available: https://arxiv.org/abs/1505.04597