CIS 5220 – Final Project – Technical Report

Team Name

April 2023

Team Members:

- Chakradhar Paladi; paladi; Email: paladi@seas.upenn.edu
- Jayadev Chevireddi; jayadev; Email: jayadev@seas.upenn.edu
- Sahachar Reddy Tippana; sahachar; Email: sahachar@seas.upenn.edu

---

Abstract

The main objective of our project is to enhance the performance of deep learning models for brain MRI tumor detection, which is a critical problem in the field of medical imaging. The existing datasets are limited in terms of quality and diversity, leading to overfitting and reduced generalizability of models. To address this issue, we propose a novel approach using Stable Diffusion models for data augmentation. Our approach generates realistic and diverse tumor images that can augment the existing dataset and ultimately improve the performance of CNNs. We have compared our approach with existing methods such as flips and rotates, Variational Autoencoders (VAEs), and Generative Adeversarial Networks(GANs), and have found that Stable Diffusion yield better results in our approach. Our contribution to the field lies in presenting a new and effective way of augmenting datasets for brain MRI tumor detection, which can have far-reaching implications in improving the accuracy of medical imaging models.

1 Introduction

The motivation for our project is to address the limitations of existing datasets for medical imaging, especially in the context of brain MRI tumor detection for rare types of cancer. The dataset sizes are often limited and lack diversity, leading to overfitting and reduced performance of deep learning models. Moreover, obtaining additional data can be challenging due to factors such as patient privacy, the cost and time required for acquiring physical copies of medical tests, and in our case: difficult to find rare cases of cancer. By generating more images, we can improve the detection of rare forms of cancer and ultimately improve their detection.

Our contributions to the field lie in using Diffusion Models to generate new images that can augment the existing dataset. We have explored the use of Stable Diffusion methods and found that they generate more realistic and diverse images compared to other methods. Our project focuses on generating specific case of brain tumor images (owing to the nature of the dataset). We have generated the said images and are to analyze the effects of augmenting datasets with them.

2 Related Work

In recent years, deep learning models have shown remarkable performance in the field of medical image analysis, especially in the detection and classification of brain tumors from magnetic resonance images (MRI). The success of these models can be attributed to their ability to learn complex features from the input images, which is essential for accurate tumor classification.

One of the most popular deep learning architectures for this task is the ResNet-18, introduced by He et al. in 2016 [He+16]. ResNet-18 has been used in several studies for brain tumor classification from MRI. For instance, Isin and Polat (2017) used ResNet-18 for the detection of brain tumors from MRI images, achieving an accuracy of 93.25% [Nas+19]. Kamnitsas et al. (2017) proposed an efficient multi-scale 3D CNN architecture with fully connected CRF for accurate brain lesion segmentation, which was based on the ResNet-18 [Kam+17].

Overall, ResNet-18 has shown promising results in brain tumor classification from MRI images and has become a popular choice for deep learning-based medical image analysis.

Classical random augmentation techniques, such as rotation, scaling, and flipping, have been widely used to enhance the performance of deep learning models on image classification tasks. These techniques help to increase the diversity of the training data and improve the generalization ability of the model. Augmentation techniques have been successfully applied in various medical image analysis tasks, including the classification of brain tumors from MRI images. For instance, in a study by Tajbakhsh et al. (2020) [Taj+16], augmentation techniques were used to improve the performance of a deep learning model for brain tumor segmentation.

Recent studies have also shown the potential benefits of using VAE-based augmentation techniques in medical imaging analysis. In a recent study by Wang et al. (2022), VAE-based augmentation was applied to a brain tumor MRI dataset to enhance the classification performance of a deep learning model [Ahm+22]. The authors found that VAE-based augmentation improved the classification accuracy of the model compared to traditional data augmentation methods.

Recently, generative adversarial networks (GANs) have been applied to generate realistic medical images for data augmentation. In the context of brain tumor classification, GAN-based augmentation has been shown to be effective in improving the performance of ResNet18 models. A study by Ahmed et al. (2022) improved the classification average accuracy from 72.63% to 96.25%, which indicates the effectiveness of GAN-based augmentation in improving the performance of ResNet models for brain tumor classification [Ahm+22]. Another study by Sarraf et al. (2020) also used GAN-based augmentation to generate additional brain MRI images for the classification of Alzheimer's disease, and demonstrated improved performance compared to traditional data augmentation techniques [ST16].

Recently, a new data augmentation method called Denoising Diffusion Data Augmentation (DDPM) has been proposed, which generates realistic and diverse images by modeling the underlying data distribution. In this study, we aim to evaluate the effectiveness of DDPM-based augmentation on a brain tumor MRI image dataset to improve the classification performance of a ResNet18 model.

Several studies have demonstrated the effectiveness of Diffusion-based Probabilistic Models (DDPM) for data augmentation in image classification tasks. For instance, the authors of the paper "Diffusion Models Beat GANs on Image Synthesis" showed that DDPM-based augmentation improves the performance of image classification tasks when compared to GAN-based augmentation methods [DN21]. Therefore, we hypothesize that DDPM-based augmentation can also enhance the performance of the ResNet-18 model in detecting and classifying brain tumors.

3 Dataset and Features

Although brain MRI images with tumors are a rarity, internet age has made it possible to collect it. We are leveraging the Brain Tumor MRI dataset from Kaggle [1] for this project. This dataset contains 7023 images of human brain MRI images which are classified into 4 classes: glioma, meningioma, no tumor and pituitary. The data has been split into train and test set and the classes are almost equally distributed throughout our training dataset as shown in the below figure 1. The split distribution is as shown in the figure 2.

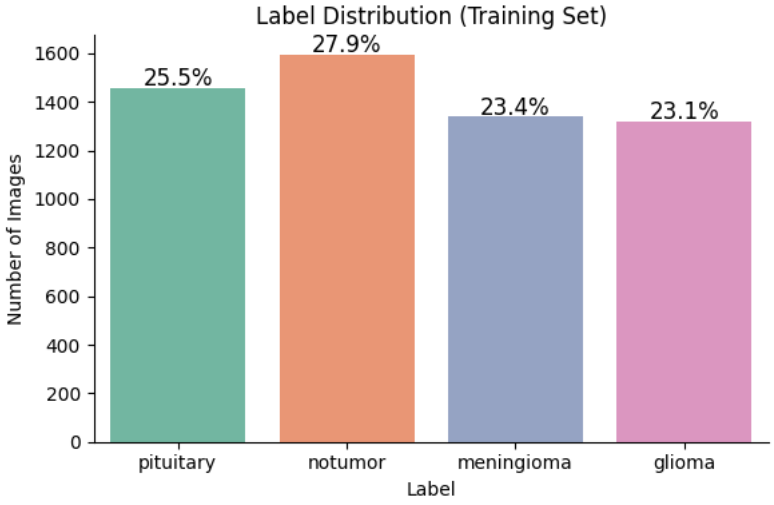

Figure 1: Training data distribution from Brain Tumor MRI dataset

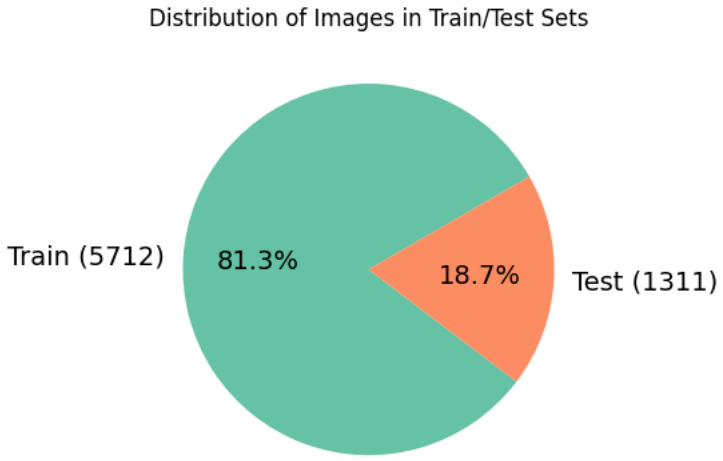

Figure 2: Training data distribution from Brain Tumor MRI dataset

3.1 Data Pre-Processing

Due to the rare nature of the data in the dataset, usually its difficult to collect this kind of data from a single source and hence, the images in our dataset are not consistent. We pre-processed our images to have a uniform size of 128x128, converted the raw data to gray scale and also mapped the class labels to integer values. The below figures 3 and 4 visualizes the raw data and pre-processed data to take a note in the difference.

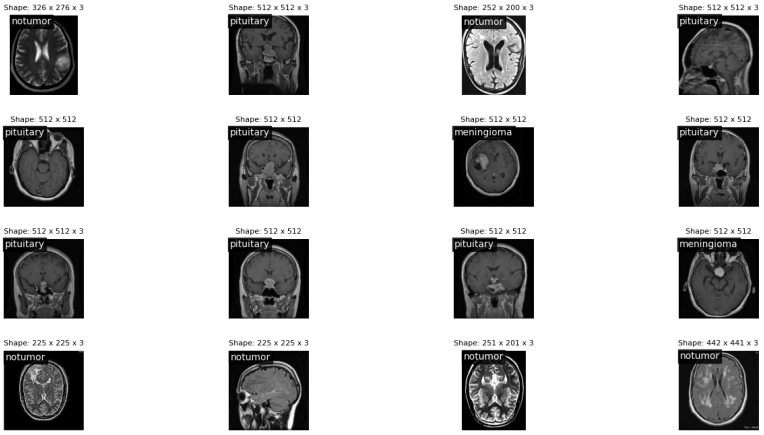

Figure 3: Raw data, before pre-processing

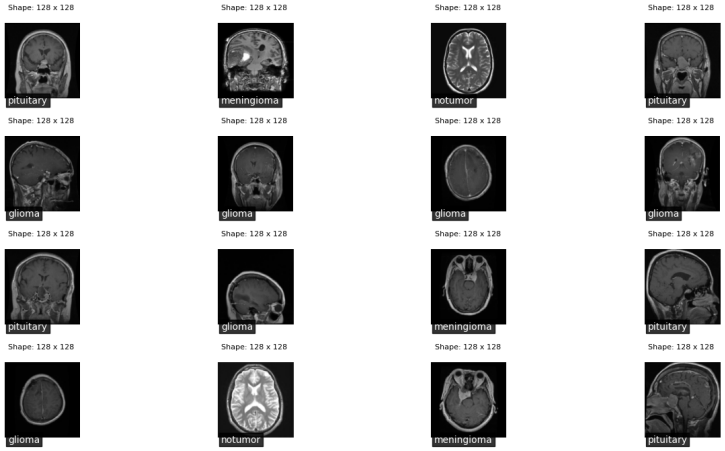

Figure 4: Pre-processed data

4 Methodology

For our task of increasing thea data instances of brain tumors, we employed classical random augmentation techniques with 20% and 50%, and deep learning based image generation techniques such as VAEs, GANs and Diffusion modeling. The generated outputs are then used to train and validate our baseline classification model. The classifiers trained are compared with respect to their loss curves, training and validation accuracies achieved. The figure 5 illustrates the flow chart of the process followed.

4.1 Classification using modified ResNet-18: Baseline

ResNet-18 is an 18-layer deep convolutional neural network that uses residual connections to overcome the vanishing gradient problem, which is a common issue in very deep neural networks. The architecture of ResNet-18 allows the gradient to flow directly from the input to the output, thus enabling the network to learn more complex features.

We are using a pretrained ResNet-18 backbone model for classifying the brain MRI image data into glioma, meningioma, pituitary and no tumor classes. The ResNet-18 model is retrained on our datasets while freezing all the layers but the last fully connected layer so that it can be trained to predict our 4 classes. The below figure 6 shows the architecture of the ResNet-18 used in this study.

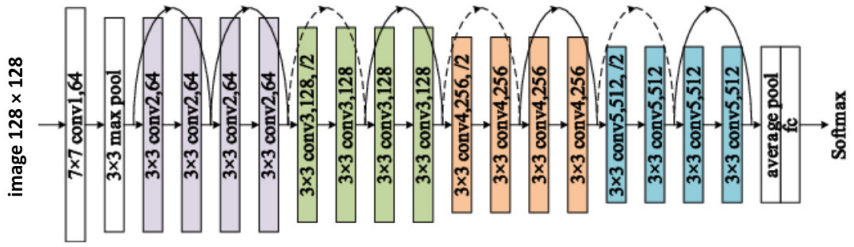

Figure 5: Classification model - ResNet-18 architecture

4.2 Augmentation

4.2.1 Classical Random Augmentation - 20% and 50%

Accurately classifying brain tumor MRI images can be a challenging task, especially in the presence of noise and variability in the imaging data. One way to improve the performance of classification models on these images is through the use of random augmentation techniques. In this context, classical random augmentation techniques refer to common image transformations such as rotations, flips, and scaling that are applied to the original images to create a larger and more diverse training set.

By training a ResNet18 model on this augmented dataset, the hope is that the model will learn to generalize better to new, unseen images, ultimately improving the accuracy of brain tumor classification. In this article, we will explore the use of classical random augmentation techniques like Random Horizontal Flip, Random Rotation, and Random Resized Crop on the brain tumor MRI image dataset to enhance the performance of a ResNet18 model for classification.

We will also see the effects of percentage of random augmentation on the dataset fow which we performed 20% and 50% augmentation on the original dataset and trained the ResNet-18 using the three datasets individually.

4.2.2 Variational Autoencoders: Advanced Architecture-1

VAE-based augmentation is a generative data augmentation approach to synthesize new data samples that are similar but not identical to the original images. VAEs learn the distribution of the input data and can generate new samples from the learned distribution. This approach has shown promising results in various computer vision tasks, including medical image analysis. In this article, we will explore the use of VAE-based augmentation on our brain tumor MRI image dataset to enhance the classification performance of a ResNet-18 model.

The VAE architecture as shown in the figure 6 is a type of neural network that uses a probabilistic approach to learn a low-dimensional representation of the input image data, known as the latent space. Our implementation of VAE consists of an encoder, a decoder, and a reparameterization module that connects the encoder and decoder. Our encoder is defined as a sequential series of four 2D convolutional layers with ReLU activation functions that reduce the input image to a latent representation. The output of the encoder is then fed to two fully connected layers with linear activation functions that compute the mean and log variance of the latent distribution.

The reparameterization module uses the mean and log variance values obtained from the encoder to sample a latent vector from a Gaussian distribution using the reparameterization trick. The trick involves generating a random noise vector and multiplying it with the standard deviation of the Gaussian distribution before adding the mean value.

Our decoder is defined as a sequential series of four transposed convolutional layers with ReLU activation functions that convert the latent representation back to the original image. The final layer uses a sigmoid activation function to ensure that the output is within the range of 0 to 1.

The model is trained using a loss function which is the sum of reconstruction loss and KL Divergence to generate random samples from the latent space distribution. The reconstruction loss is the mean squared error (MSE) between the output generated by the VAE and the target input image. The KL divergence loss measures the divergence between the distribution of the latent variables and a normal distribution. The KL divergence loss encourages the VAE to learn a compact and continuous representation of the input data in the latent space.

The VAE loss is defined by the following equation:

$$L_{VAE} = \frac{1}{N} \sum_{i=1}^{N} [MSE(x_i, \hat{x}i) + \beta D_{KL}(q(z_i|x_i)||p(z))]$$

where MSE is the mean squared error between the input image x_i and the output image $\hat{x_i}$, D_{KL} is the KL divergence between the distribution $q(z_i|x_i)$ and the prior distribution $p(z)$, β is a hyperparameter that balances the reconstruction loss and KL divergence loss, and N is the batch size.

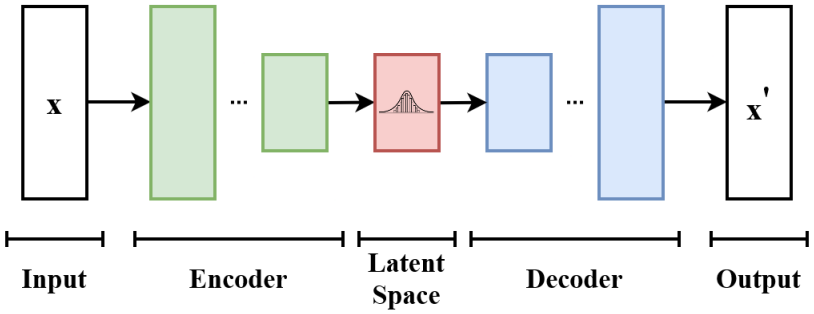

Figure 6: The basic scheme of a variational autoencoder architecture []

4.2.3 Generative Adversarial Networks: Advanced Architecture-2

GANs are composed of two neural networks, a generator and a discriminator, which learn to generate synthetic images that are indistinguishable from real ones. GAN-based augmentation has shown promising results in improving the performance of deep learning models for medical image analysis tasks such as segmentation and classification. In this article, we will explore the use of GAN-based augmentation on our brain tumor MRI image dataset to enhance the classification performance of our ResNet-18 model.

Our GAN architecture is designed for generating 128x128x1 MRI images. The generator consists of two transposed convolutional layers followed by a fully connected layer. The activation function used in the generator is ReLU, except for the last layer which uses the tanh activation function.

The discriminator is composed of a fully connected layer followed by two convolutional layers. The activation function used in the discriminator is Leaky ReLU with a slope of 0.2, except for the last layer which does not have an activation function. The loss function used in this GAN architecture is the mean squared error (MSE) loss.

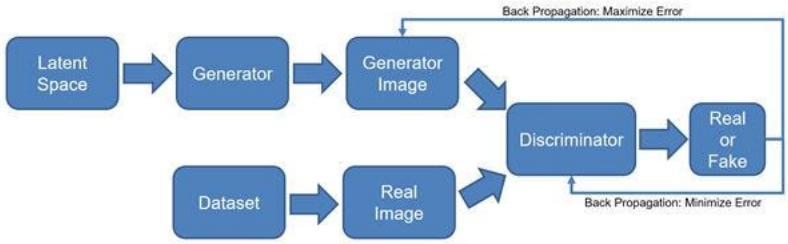

Figure 7: The basic work flow of a GAN architecture []

Overall, the GAN architecture generates images by training the generator and discriminator in an adversarial way as shown in the figure 7. The generator learns to generate images that can fool the discriminator into believing they are real, while the discriminator learns to distinguish between real and fake images. The MSE loss is used to calculate the difference between the generated and real images, which is then used to update the parameters of both the generator and discriminator. Through this iterative process of training the generator and discriminator, the GAN architecture generates high-quality images that resemble the real MRI images. We trained our GAN model for 200 epochs and the generated images were to better than that generated by the VAE.

4.2.4 Denoising Diffusion Probabilistic Model: Adv Architecture-3

The Denoising Diffusion Probabilistic Model (DDPM) is a generative model used for image generation, which is designed to model image generation as a diffusion process. The DDPM model uses a noise scheduler to control the amount of noise added to an image as it is transformed over a fixed number of timesteps.

In our particular architecture, the input to the model is a 128x128x3 MRI image. The generator model uses a UNET 2D architecture consisting of four Resnet down-sampling blocks to halve the image size, two spatial self-attention blocks, and four Resnet up-sampling blocks. There are skip connections that link the features on the downsample path to the corresponding layers in the upsample path. The activation function used throughout the model is ReLU. The loss function used for training is the mean squared error (MSE) loss.

The optimizer used is AdamW, which is a variant of the Adam optimizer with additional weight decay. The learning rate scheduler used is the cosine learning rate scheduler, which gradually decreases the learning rate over time.

The DDPM scheduler is set to 1000 timesteps, meaning that the diffusion process is carried out over 1000 steps. At each step, noise is added to the image, and the image is transformed using the generator model. The output of the generator is then compared to the original image using the MSE loss function.

The training loop also uses Accelerate for easy TensorBoard logging, gradient accumulation, mixed precision training, and multi-GPUs training.

Overall, our architecture aims to generate high-quality MRI images by modeling the diffusion process and using a UNET-based generator model with attention blocks for capturing spatial dependencies. The optimizer and learning rate scheduler are used to improve the training process and achieve better results. Below is a basic Unet2d architeture in the figure 8.

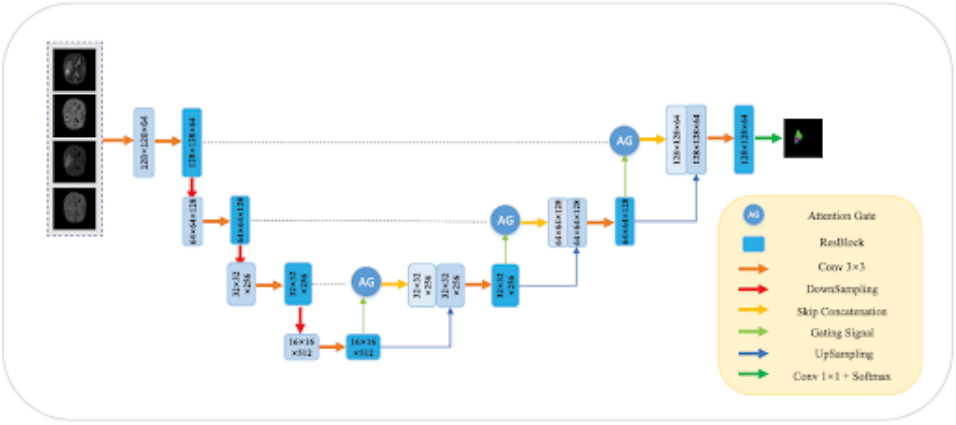

Figure 8: The basic work flow of a UNet-2D architecture

5 Results

Start with a sentence or two on what you want to show. Describe how you showed that and what you learned. This section should include your results, i.e. the performance of your models with regards to your chosen performance metric, as well as tables/visualizations of these results, the training process, confusion matrices for classification projects... (whatever works for your project, make a sensible choice here!). Please report the loss function that you've minimized and additional measures of performance/quality you looked at, too.

In this study, we aim to demonstrate the effectiveness of using Denoising Diffusion Probabilistic Modeling (DDPM) for generating synthetic brain MRI images with tumors and evaluate its performance as a data augmentation technique. We compare the performance of models trained on various augmented datasets.

To evaluate the performance of the models, we chose a suitable loss function and additional performance metrics, such as accuracy. We trained models on 20% and 50% augmented datasets using different augmentation techniques and monitored the training process, capturing loss curves and accuracy images for each case.

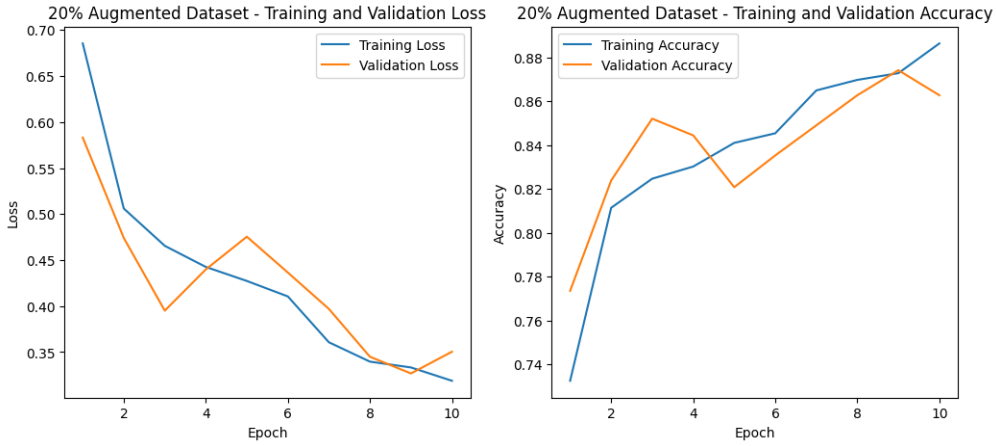

Figure 9: Loss and Accuracy Curves of 20% Augmented Dataset

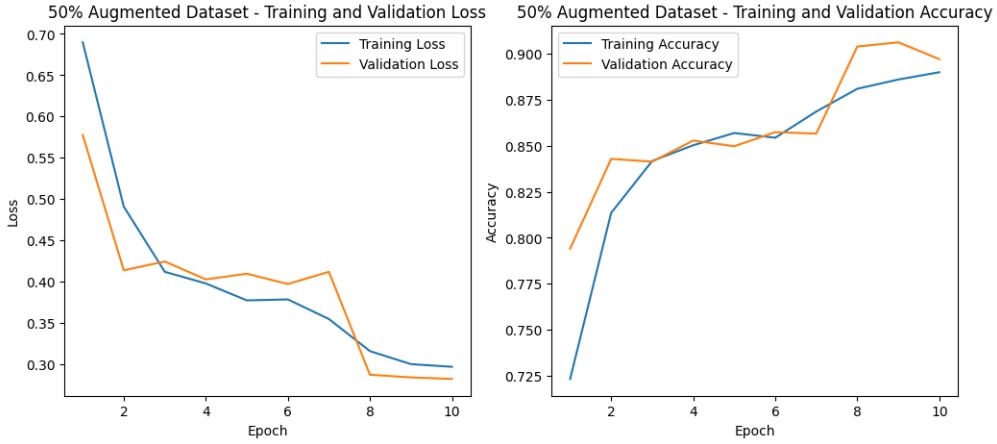

Figure 10: Loss and Accuracy Curves of 50% Augmented Dataset

Images generated by Variational Autoencoders (VAEs) and their loss curves after 30 epochs of training time.

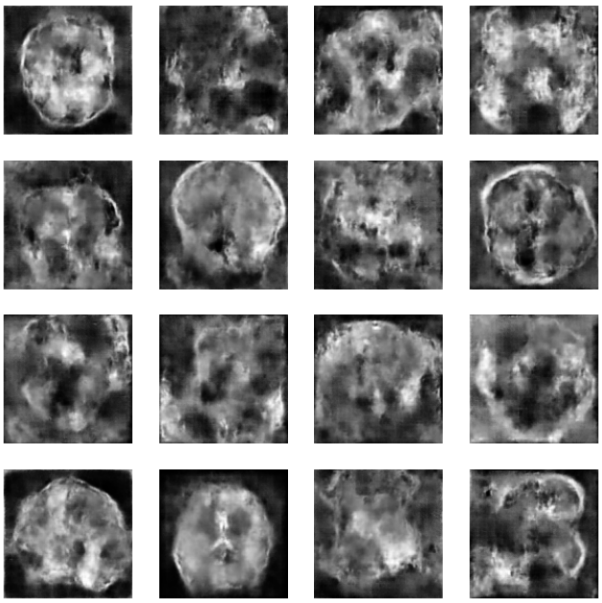

Figure 11: Images Generated by VAE

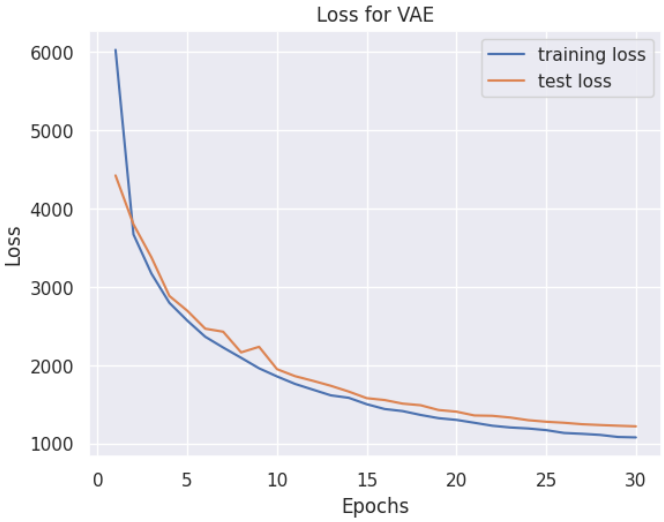

Figure 12: Loss Curves of VAE

Images generated by Generative Adversarial Networks (VAEs) and their loss curves after 200 epochs of training time.

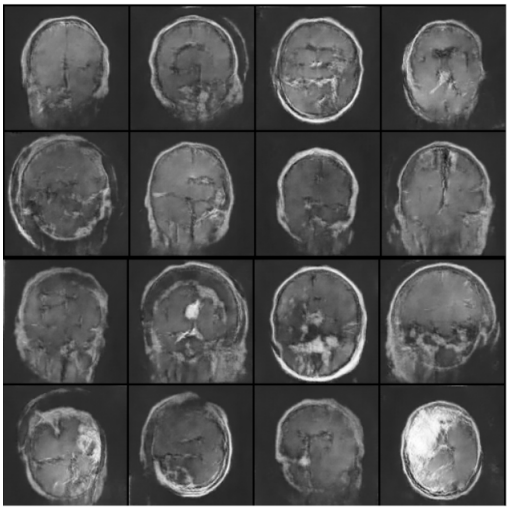

Figure 13: Images Generated by GAN

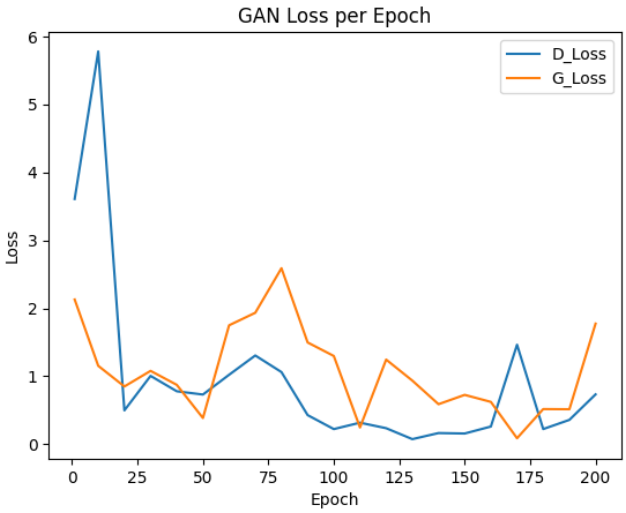

Figure 14: Loss Curves of GAN

6 Discussion

Approach	Loss function	Optimizer	Findings
20% Aug - ResNet	CrossEntropy	Adam	86.27% Val_acc
50% Aug - ResNet	CrossEntropy	Adam	89.5% Val_acc
VAE	MSE	Adam	Indistinct Images
GAN	MSE	Adam	Stability issues
DDPM/UNet	MSE	Adam	Optimal Images*

* *The images have a hint of noise, solved using denoising techniques.*

6.1 Findings

Our project comprised of three stages, starting with non-deep learning techniques (flip, rotation, and crop), advancing to generative models (VAE and GAN), and finally to our primary focus, Stable Diffusion Models.

In the initial phase, non-deep learning data augmentation methods (20% and 50% augmentation through flip, rotation, and crop) yielded a validation accuracy of 86.27% and 89.5% respectively. These basic techniques, while simple, provided a useful benchmark for subsequent, more complex models.

Next, we experimented with generative models, specifically Variational Autoencoders (VAE) and Generative Adversarial Networks (GAN). The VAE model, trained for 30 epochs, produced images of lower quality, and the loss reached a plateau, indicating limitations in its ability to generate diverse and realistic medical images. The GAN model, trained for 200 epochs, generated significantly better images than the VAE. However, after prolonged training, it began to overfit, resulting in less accurate representations of brain MRI structures.

The core of our project was the application of Stable Diffusion Models, specifically the Denoising Diffusion Probabilistic Model (DDPM) with a DDPM noise scheduler and a U-Net 2D architecture. We developed four separate models, each generating images of glioma, pituitary, meningioma, and no tumor. We generated 200 images for each tumor type. The resulting images were remarkably realistic and diverse, contributing to a validation accuracy of 86% when incorporated into the original dataset. However, it's important to note that about 30% of the generated images exhibited high levels of noise, which potentially limited the achievable validation accuracy.

Our work offers a substantial contribution towards addressing the challenges of diagnosing rare brain cancers. The scarcity and lack of diversity in available data are significant roadblocks in this field, and our project tackles these issues by generating high-quality, diverse synthetic images.

By incorporating these synthetic images into the original dataset, we enhanced

its size by over 13%, improving its diversity and potentially the robustness and generalizability of deep learning models trained on it. This approach is especially valuable in scenarios where data acquisition is costly, time-consuming, or simply not feasible due to the rarity of the condition.

Compared to existing research, our results emphasize the potential of Stable Diffusion Models, particularly the Denoising Diffusion Probabilistic Model, for generating realistic and diverse medical images. While our results were promising, the issue of image noise in a significant percentage of generated images indicates that further optimization is required to fully leverage the power of these models. Despite this, our findings contribute valuable insights to the growing body of knowledge on data augmentation techniques for medical imaging and provide new avenues for future research in this area.
Elaborate on how the performances of your non-DL, your base-DL, and your advanced models compare to each other and why one might have worked better than the others.

6.2 Limitations and Ethical Considerations

Technical Limitations :
a. ***Computational complexity***:
DDPM is a computationally intensive model that requires significant computational resources and time to train. This can be a barrier for researchers with limited access to high-performance computing facilities or tight deadlines.

b. ***Model robustness***:
The quality of the generated images depends heavily on the quality of the training data and the model's hyperparameters. Incorrectly tuned hyperparameters or insufficiently diverse training data can result in poor-quality images, leading to suboptimal model performance.

c. ***Limited generalizability***:
The performance of DDPM-generated images for data augmentation may not necessarily transfer to other medical imaging tasks or different MRI modalities. This limitation highlights the need for thorough validation across multiple tasks and modalities before widespread adoption.

Ethical Considerations:
a. ***Privacy concerns***:
Generating realistic brain MRI images using DDPM can potentially lead to the inadvertent disclosure of sensitive patient information. Ensuring that the generated images do not retain any identifiable information is essential to protect patient privacy.

b. ***Misrepresentation of data***:
The use of artificial images in medical research raises concerns about the potential misrepresentation of data. It is crucial to transparently report the use of DDPM-generated images in any research findings and acknowledge the limitations associated with their use.

c. ***Bias in model training***:
The generated images may inherit and even amplify biases present in the original dataset, leading to biased model predictions. Researchers should be cautious of such biases and, where possible, employ techniques to mitigate their impact.

6.3 Future Research Directions

Several future research directions can be pursued to improve the use of DDPM in data augmentation for brain MRI datasets and to extend its applicability in the medical imaging domain.

Higher Resolution Images : One potential direction for future work is to focus on generating higher-resolution images using DDPM. Higher-resolution images can provide more detailed and accurate information, which could lead to improved performance in downstream medical imaging tasks. To achieve this, researchers can explore novel architectures and training techniques specifically tailored for high-resolution image synthesis.

Model Scaling: Another area of interest is model scaling, which involves studying the relationship between the size of the model, the amount of training data, and the quality of the generated images. By understanding how these factors interact, researchers can optimize the trade-offs between computational resources, training time, and image quality, leading to more efficient and effective DDPM models.

Generalizing to Other Datasets: Extending the applicability of DDPM-generated images to other medical imaging datasets and modalities is a crucial direction for future research. This can involve validating the performance of DDPM-generated images on a diverse range of medical imaging tasks, such as different MRI modalities, CT scans, or X-rays, and exploring techniques to adapt the DDPM model to new datasets effectively.

Semantically Segmented Datasets: An additional direction for future work is to semantically segment the original dataset and use it to generate synthetic data with corresponding semantic information. This can provide richer information for downstream tasks, such as segmentation or classification, and help improve the overall performance of medical imaging models trained on DDPM-generated data.

7 Conclusions

In this study, brain MRI images with tumors have been generated using Denoising Diffusion Probabilistic Modeling (DDPM). This advanced generative model has shown promising results in creating realistic and diverse images for data augmentation purposes. Compared to other generative models such as Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs), DDPM has demonstrated a significant improvement in the quality of synthetic images generated.

However, it is important to note that the synthetic images produced by DDPM still contain some level of noise. While this noise may not necessarily hinder the model's performance in certain tasks, it is essential to explore methods for reducing noise and further improving the quality of the generated images. This could involve optimizing the DDPM model's architecture, training techniques, or employing post-processing techniques to denoise the synthetic images.

In terms of data augmentation, the performance of our DDPM-generated images is a little bit less to that of classical augmentation techniques, such as flipping, rotating, and resizing. This demonstrates the potential of DDPM as a viable alternative for data augmentation in the context of brain MRI datasets. However, there is still room for improvement, and future research should focus on enhancing the performance of DDPM-generated images in data augmentation.

In summary, the use of Denoising Diffusion Probabilistic Modeling in generating synthetic brain MRI images with tumors has shown promising results, offering an alternative data augmentation approach. While there are some limitations, such as the presence of noise in the generated images, the performance is a little bit less than classical augmentation techniques. By focusing on further improvements and investigating the model's generalizability to other tasks, DDPM has the potential to become a valuable tool in medical imaging research and the development of more accurate and robust machine learning models.

References

[Ahm+22] Bilal Ahmad et al. "Brain Tumor Classification Using a Combination of Variational Autoencoders and Generative Adversarial Networks". In: *Biomedicines* 10.2 (2022). ISSN: 2227-9059. DOI: 10.3390/biomedicines10020223. URL: https://www.mdpi.com/2227-9059/10/2/223.

[DN21] Prafulla Dhariwal and Alex Nichol. *Diffusion Models Beat GANs on Image Synthesis*. 2021. arXiv: 2105.05233 [cs.LG].

[] *GAN workflow*. https://medium.com/@Packt_Pub/inside-the-generative-adversarial-networks-gan-architecture-2435afbd6b3b. [Online].

[He+16] Kaiming He et al. "Deep Residual Learning for Image Recognition". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.

[Kam+17] Konstantinos Kamnitsas et al. "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation". In: *Medical Image Analysis* 36 (2017), pp. 61–78. ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2016.10.004. URL: https://www.sciencedirect.com/science/article/pii/S1361841516301839.

[Nas+19] Md Abdullah Nasim et al. "Brain Tumor Detection using Convolutional Neural Network". PhD thesis. June 2019. DOI: 10.13140/RG.2.2.15562.52163.

[ST16] Saman Sarraf and Ghassem Tofighi. "Deep Learning-based Pipeline to Recognize Alzheimer's Disease using fMRI Data". In: *bioRxiv* (2016). DOI: 10.1101/066910. eprint: https://www.biorxiv.org/content/early/2016/07/31/066910.full.pdf. URL: https://www.biorxiv.org/content/early/2016/07/31/066910.

[Taj+16] Nima Tajbakhsh et al. "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?" In: *IEEE Transactions on Medical Imaging* 35.5 (2016), pp. 1299–1312. DOI: 10.1109/TMI.2016.2535302.

[] *VAE basic*. https://en.wikipedia.org/wiki/Variational_autoencoder. [Online].