Synthetic Medical Image Generation with Diffusion Model

ECE-GY 7123: Deep Learning Final Project
Anubha Singh[1], **Kavya Gupta**[2], **Khushi Sharma**[3]

New York University
[1]as18806@nyu.edu [2]kg3373@nyu.edu [3]ks7406@nyu.edu

Problem Statement

In the realm of medical diagnostics, particularly dermatology, the creation and utilization of extensive image datasets are paramount for the training and refinement of Artificial Intelligence (AI) driven diagnostic systems. Such systems are designed to assist healthcare professionals in quickly identifying a range of skin diseases, including malignant cancers and at the same time, simultaneously mitigating privacy concerns associated with the use of real patient data. However, the development of these datasets is frequently hindered by numerous challenges, such as significant data imbalance, variability in image quality, and stringent privacy regulations. To address these issues, our project proposes the innovative application of fine-tuning a pre-trained diffusion model to generate synthetic images. This approach not only promises to augment the available data but also to enhance the overall training process of machine learning models aimed at improving the diagnosis and treatment of skin cancer.

Literature Review

In the field of synthetic data generation for medical imaging, text-to-image stable diffusion models are emerging as a pivotal technology due to their robust training dynamics and exceptional mode coverage. The literature reveals a growing interest in leveraging these models for medical applications, as evidenced by recent studies. Chambon et al. (2022) have explored adapting pretrained vision-language foundational models to medical imaging domains, underscoring the potential of these models to seamlessly transition from general image synthesis to specific medical use cases. Their work suggests that fine-tuning these models on medical imagery can enhance their applicability and accuracy in clinical settings.

Further contributing to the body of knowledge, Farooq et al. (2024) have demonstrated how stable diffusion models can be employed to generate synthetic skin lesion data, thereby enhancing the performance of skin disease classifiers that use advanced machine learning architectures like Vision Transformers (ViT) and Convolutional Neural Networks (CNNs). This study highlights the efficacy of synthetic images in improving disease classification accuracy while adhering to privacy standards, thus offering a dual benefit. Additionally, insights from Corvi et al. (2023) into the intriguing properties of synthetic images from generative models like GANs and diffusion models detail how these technologies can create visually accurate and diverse datasets, which are critical for training robust diagnostic tools. These contributions collectively emphasize the significant impact of stable diffusion models in the medical imaging sector, marking a promising direction for future research and application.

Dataset

The ISIC 2020 Skin Cancer Dataset, sourced from the International Skin Imaging Collaboration (ISIC) Archive, serves as our dataset for this project. As one of the most extensive collections of dermatology-related image datasets available, it provides a rich repository of skin lesion images indispensable for the development of accurate and effective diagnostic models. The dataset includes diverse imagery of skin conditions such as melanoma, nevus, and basal cell carcinoma, which are among the most common and clinically significant types of skin cancer. Comprising 33,126 metadata entries, the dataset encapsulates not only the images but also accompanying patient demographics and clinical details, including patient ID, sex, age, and the general anatomic site of the lesion. This comprehensive metadata is vital for facilitating nuanced analyses and enabling researchers to consider factors such as age-related variability in skin lesions or gender differences in skin cancer prevalence.

Technical Details

Model Details

Stable Diffusion, a neural network model, generates images through four main stages: first, an Image Encoder converts input images into numerical vectors in latent space. Then, a Text Encoder turns text descriptions into high-dimensional vectors. Next, a Diffusion Model uses these vectors to create new images in the latent space, guided by text. Finally, an Image Decoder reconstructs these latent images into detailed, pixel-based visuals. This model not only generates images from text but also performs inpainting, outpainting, and image-to-image translations.

In the **Vision Transformer (ViT)** model for image generation, the architecture employs an encoder-decoder setup where latent embeddings serve as input. These embeddings, enhanced with positional information, pass through the transformer's encoder. The decoder then iteratively constructs the image by generating and assembling patches, utilizing self-attention mechanisms to ensure spatial coherence and detail. This model is then trained through techniques that predict subsequent patches based on previous ones, effectively leveraging the transformer's capacity to manage complex dependencies for creative image synthesis.

We used the following hyperparameters for our ViT model:
Learning Rate: 2e-5
Batch Size: 32
Weight Decay: 0.01
Optimizer: AdamW
Scheduler: get_linear_schedule_with_warmup()

Methodology

Fine-Tuning

We attempted to fine-tune a diffusion model (runwayml/stable-diffusion-v1-5) using the ISIC 2020 dataset, employing DreamBooth, a targeted training methodology designed to update diffusion models by training on a minimal set of images representing specific subjects or styles. This approach involves associating a unique prompt with representative images.

For the fine-tuning process, our dataset included approximately 25 images from each type of benign skin cancer mole as specific instances. Additionally, we incorporated around 250 images encompassing both benign and malignant skin cancer moles as broader classes. The images were organized in a structured dictionary with the following components:

- **instance_prompt**: A prompt describing a distinct instance of a skin lesion image.
- **class_prompt**: A prompt encapsulating the overall class of skin lesion images.
- **instance_data_dir**: The directory path where images of the specified instance are stored.
- **class_data_dir**: The directory path where images representing the general class are stored.

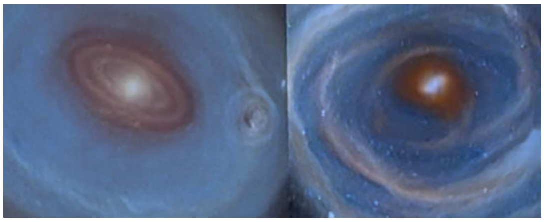

Figure 1: Synthetically Generated Skin Lesion

Despite the rigor of our methodology, we faced substantial challenges inherent to the complexities of stable diffusion during the fine-tuning process.

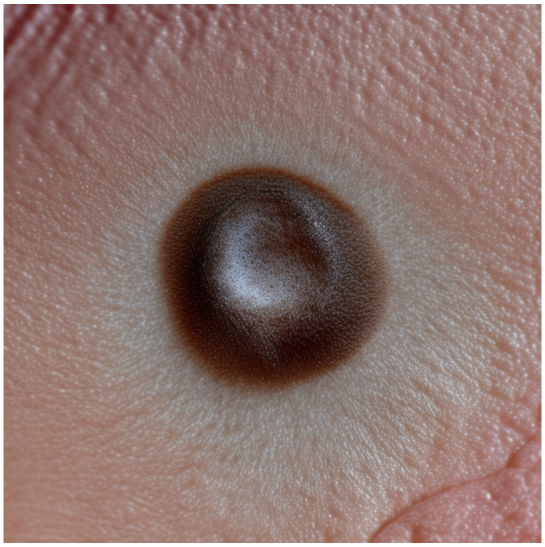

Figure 2: Synthetically Generated Skin Lesion where Instance Prompt was given as 'Malignant Skin Lesion' and Class Prompt was given as 'Skin Lesion'

When we utilized fewer epochs, the resulting images, as depicted in Figure 1, were of poor quality. However, upon extending the training duration, we obtained improved images resembling those shown in Figure 2, which exhibited a notable enhancement in quality. Despite these promising results, our efforts were hindered by limitations such as constrained GPU resources and prolonged training duration. Consequently, we faced challenges in generating a large number of high-quality images. These constraints significantly impeded our ability to achieve satisfactory outcomes and effectively utilize the fine-tuning technique to enhance model performance as desired.

Our next step was to look into a paper by Muhammad Ali Farooq et al., called "Derm-T2IM". In this paper, they developed a stable diffusion model for generating synthetic images of skin cancer by using a similar methodology as we attempted. Their model fine-tuning employed preprocessing of the ISIC dataset to center lessons and also advanced methods like removal of hair from images.

We attempted to run their fine-tuned model to generate images but unfortunately due to less documentation and significant computational resources required for the task, we hit a roadblock. The resources that we had access to lacked the processing power and memory capacity needed to handle the computations involved in training the diffusion model. Despite our best efforts, the complexity of the model and the limitations of our setup made it exceedingly difficult to successfully deploy the model presented in the paper.

Thus, for our next step, we decided to focus on validation of synthetic images using a skin lesion classifier based on Vision Transformer.

Synthetic Data Validation

For the validation of synthetic data, we engaged in a comprehensive evaluation of the generated images sourced from the ”Derm-T2IM” study by Muhammad Ali Farooq et al. Our approach involved employing a skin lesion classifier based on the Vision Transformer (ViT) architecture to assess the authenticity and clinical relevance of these synthetic images. We began by fine-tuning a pre-trained ViT model exclusively with a dataset comprising **real** cancer images to establish a reliable baseline. This baseline provided a critical reference point for assessing the impact of synthetic data integration. Subsequently, we fine-tuned this model again using a hybrid dataset that consisted of an equal mix, i.e. a 50-50 mix of synthetic (or generated) and real cancer images. This methodological adaptation led to an enhancement in model performance, with the baseline accuracy of 81.73% improving to 84.06%, thereby indicating a substantial increase of 2.3%.

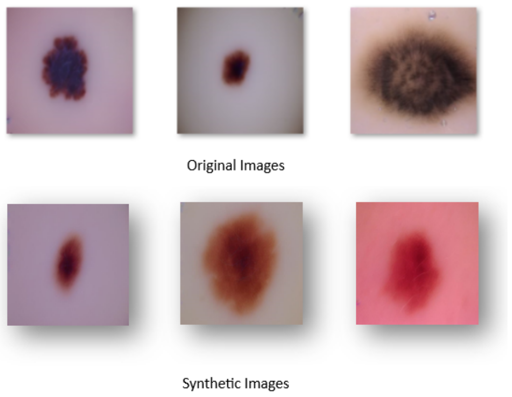

Figure 3: Original VS Synthetic Images.

Results

The results of our validation efforts clearly demonstrate the efficacy of incorporating synthetic images into the training regimen of diagnostic models. We saw a notable increase of 2.33% in the final Test Accuracy, i.e. it increased from a baseline accuracy of 81.73% to 84.06%. This increase not only highlights the qualitative improvements in model robustness due to the synthetic data but also reinforces the potential utility of synthetic imagery in improving the diagnostic accuracy of medical image analysis models.

Table 1: Model Performance Comparison

Model	Test Accuracy (%)
Baseline	81.73
Hybrid	84.06

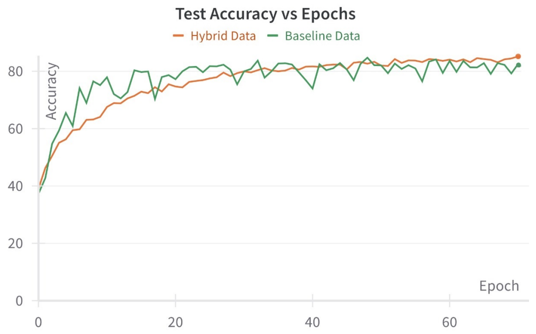

Figure 4: Test Accuracy vs Epoch.

Conclusion

This study has demonstrated the substantial potential of employing stable diffusion models for the synthesis of dermatological images, particularly in the context of enhancing the training and performance of AI-driven diagnostic systems for skin cancer. By fine-tuning a pre-trained diffusion model with the ISIC 2020 dataset and utilizing DreamBooth for targeted training, we aimed to overcome significant challenges such as data imbalance and privacy concerns inherent in medical imaging. Despite encountering limitations related to computational resources and memory capacities, our refined approach using a Vision Transformer-based classifier to validate synthetic images resulted in a noticeable improvement in diagnostic accuracy. The final test accuracy increased, thus affirming the effectiveness of integrating synthetic data into diagnostic models. This advancement not only underscores the practical benefits of synthetic images but also opens avenues for future research to further optimize these models for clinical applications, ensuring they meet the high standards required for medical diagnostics.

References

Chambon, P.; Bluethgen, C.; Langlotz, C. P.; and Chaudhari, A. 2022. Adapting Pretrained Vision-Language Foundational Models to Medical Imaging Domains. arXiv:2210.04133.

deca.ai. 2024. Stable Diffusion. https://deci.ai/deep-learning-glossary/stable-diffusion/.

Farooq, M. A.; Yao, W.; Schukat, M.; Little, M. A.; and Corcoran, P. 2024. Derm-T2IM: Harnessing Synthetic Skin Lesion Data via Stable Diffusion Models for Enhanced Skin Disease Classification using ViT and CNN. arXiv:2401.05159.

Hugging Face. 2024. DreamBooth.

MICCAI 2023 PRIME Workshop. 2023. DermoSegDiff: A Boundary-aware Segmentation Diffusion Model for Skin Lesion Delineation.

R. Corvi, G. P. K. N., D. Cozzolino; and Verdoliva, L. 2023. Intriguing properties of synthetic images: from generative adversarial networks to diffusion models.

GitHub Link

The link to our Public GitHub Repository is: Synthetic-Image-Through-Diffusion