GENERETIVE AI

Synthetic Data Generation for Healthcare

REVIEW - 2

Group Members :
Komal Katare (202201040142)
Palak Yerawar (202201040195)
Surabhi Kharkate (202201040215)
Nilambari Mahajan (202201040118)

Project Guide : Mrs. Savita Mane
Assistant Professor
Computer Department

CONTENT

- Problem Statement
- Objectives
- Dataset Selection
- Literature review
- Preprocessing
- Implementation on design
- System architecture / design
- Mathematical formulas
- Evaluation Metrices
- Comparative Analysis
- Deployment with Ethical Considerations
- Conclusion

PROBLEM STATEMENT

Skin disease classification faces challenges of data scarcity, imbalance, and privacy concerns. This project generates synthetic dermoscopy images using VAE, GAN, and Diffusion model. The enhanced dataset improves Vision Transformer (ViT) performance while ensuring data privacy.

OBJECTIVES

1. To generate high-quality synthetic dermoscopic images using VAE, GAN, and Diffusion Models.

2. To classify skin lesions using Vision Transformers (ViTs).

3. To evaluate synthetic data quality with Fréchet Inception Distance (FID).

4. To assess diagnostic performance using accuracy, precision, recall, and F1-score.

5. To address dataset imbalance and improve generalization through synthetic data augmentation.

DATASET SELECTION

The dataset used in this project is the **HAM10000 Skin Lesion Dataset**, part of the ISIC archive. It contains **10,015 dermoscopic images** of pigmented skin lesions in JPEG format, with metadata such as age, sex, and lesion site. The dataset covers seven diagnostic categories, including melanoma, nevus, basal cell carcinoma, and others, making it well-suited for skin cancer classification. It is widely used in medical AI research to tackle data scarcity, class imbalance, and standardization challenges.

Dataset link: https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000

LITERATURE REVIEW

Sr.No.	Title	Year	Description	Ideas
1.	The Role of Vision Transformers in Dermoscopic Skin Disease Diagnosis	Year: 2025	A comprehensive review analyzing how Vision Transformers (ViTs) are applied for skin disease diagnosis using dermoscopic images (HAM10000 dataset).	Summarizes existing ViT-based methods, highlights challenges like data scarcity and lesion variability, and proposes future research directions.
2.	LesionGen: A Concept-Guided Diffusion Model forDermatology Image	Year: 2025	Introduces a diffusion-based model that generates realistic dermatology images guided by clinical concepts.	Uses text-to-image diffusion guided by medical concepts to create diverse synthetic lesion images for improving diagnostic model training.
3.	DermViT: Diagnosis-Guided Vision Transformer	Year: 2024	Proposes a diagnosis-guided Vision Transformer architecture tailored for robust and efficient skin lesion classification.	Enhances ViT with domain-specific modules (context pyramid, hierarchical attention, feature gating) to better mimic dermatologist reasoning and boost accuracy.

Sr. No.	Title	Year	Description	Ideas
4.	Derm-T2IM: Stable Diffusion for Skin Lesion Synthesis	Year : 2024	Presents a Stable Diffusion–based text-to-image model customized for realistic dermoscopic skin lesion generation.	Fine-tunes diffusion models with dermatology-specific prompts to create diverse, high-quality synthetic lesion images for training data augmentation.
5.	Fine-tuning GLIDE for Dermoscopic Image Synthesis — Frontiers in Medicine	Year : 2023	Adapts OpenAI's GLIDE text-to-image diffusion model for dermoscopic image generation.	Uses prompt-based fine-tuning on medical image datasets to synthesize realistic skin lesion images, improving model generalization under limited data.
6.	Boosting Lesion Segmentation via Diffusion + Prompts	Year : 2023	Proposes a diffusion-driven framework to enhance lesion segmentation performance using semantic prompts.	Combines diffusion models with text/image prompts to guide segmentation networks, improving accuracy and boundary precision in dermoscopic tasks.

PREPROCESSING

- **Image Resizing** to 64×64 pixels

Why: Reduces computational load and ensures all images have a uniform input size for the model.

- **Normalization** to pixel range [-1, 1]

Why: Stabilizes and accelerates model training, especially with activation in the decoder.

- **Data Augmentation** (flips, rotations)

Why: Doubles dataset from 10k to 20k samples to reduce overfitting and improve the model's ability to generalize.

MODEL SELECTION AND DESCRIPTION

1. **Variational Autoencoder (VAE) :**

- **Encoder** compresses skin lesion images into a latent space (compact representation).
- **Latent space** sampling allows generating new variations of lesions.
- **Decoder** (with deconvolution layers) reconstructs images from latent vectors.
- **Training Setup:** latent_dim=128, trained for 120 epochs using Adam optimizer (lr=1e-3), batch_size=64, loss = MSE + KL divergence, samples generated using 1000 latent vectors for FID evaluation.
- Enables data augmentation by generating realistic lesion images.
- Ensures diversity while preserving important medical features.

Applications in this Project :

- Balances the dataset by generating more samples of rare lesion classes.
- Provides high-quality synthetic skin lesion images for better training of classifiers.

MODEL SELECTION AND DESCRIPTION

2. DC Generative Adversarial Network (GAN) :

- **Generator** creates synthetic skin lesion images from random noise/latent space.
- **Discriminator** checks whether images are real (from dataset) or fake (from generator).
- Continuous competition improves image realism and quality.
- Generates rare lesion types, balancing dataset distribution.
- Enhances classifier performance by providing diverse training samples.
- **Training Setup:** Trained at 64×64 resolution for 120 epochs using DCGAN architecture, Adam optimizer (learning rate = 0.0002), batch size = 64, and Binary Cross-Entropy loss for adversarial training.

Application in Project:

- Balancing class distribution (handling imbalanced datasets).
- Style transfer (e.g., adapting lesions across imaging devices).

MODEL SELECTION AND DESCRIPTION

3. Diffusion Models :

- Gradually adds noise to lesion images and then learns to reverse the process.
- Denoising steps reconstruct high-quality, detailed synthetic images.
- Produces more realistic and sharper lesions than GANs/VAEs.
- Useful for generating large-scale, high-resolution datasets.
- Improves generalization of diagnostic models in clinical settings.
- **Training Setup:** Trained on images at 64×64 resolution for 120 epochs, Adam optimizer, batch_size=128, Learning rate=2e-4 and MSE loss to predict noise across timesteps.

Application in Project:

- Producing clinically realistic data for training AI.
- Reducing overfitting by creating unlimited samples.

MODEL SELECTION AND DESCRIPTION

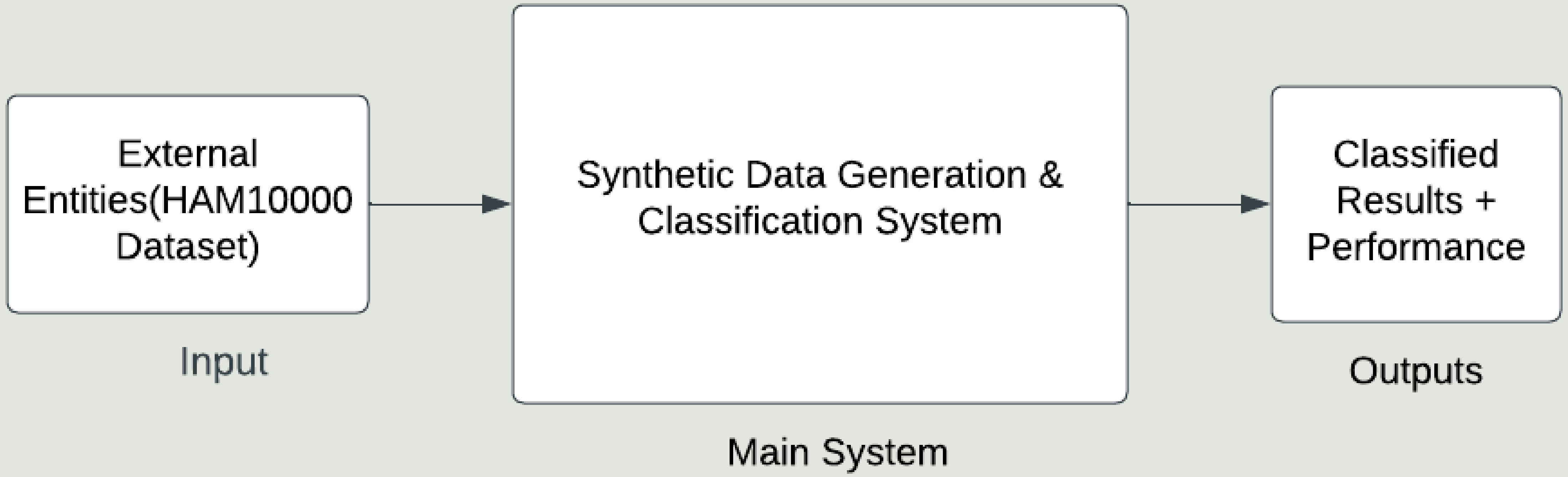
4. Vision Transformer (ViT) :

- Splits lesion images into fixed-size patches (like small tiles).
- Each patch is **encoded** into embeddings and processed by transformer layers.
- **Self-attention** mechanism captures global relationships across patches.
- Learns fine-grained patterns for accurate **disease classification**.
- Achieves higher accuracy compared to traditional CNNs on large datasets.
- **Training Setup:** ViT model fine-tuned for 100 epochs on HAM10000 images using AdamW optimizer (learning rate = 5e-5), batch size = 32, cross-entropy loss, early stopping, and pre-trained ImageNet weights.

Application in Project:

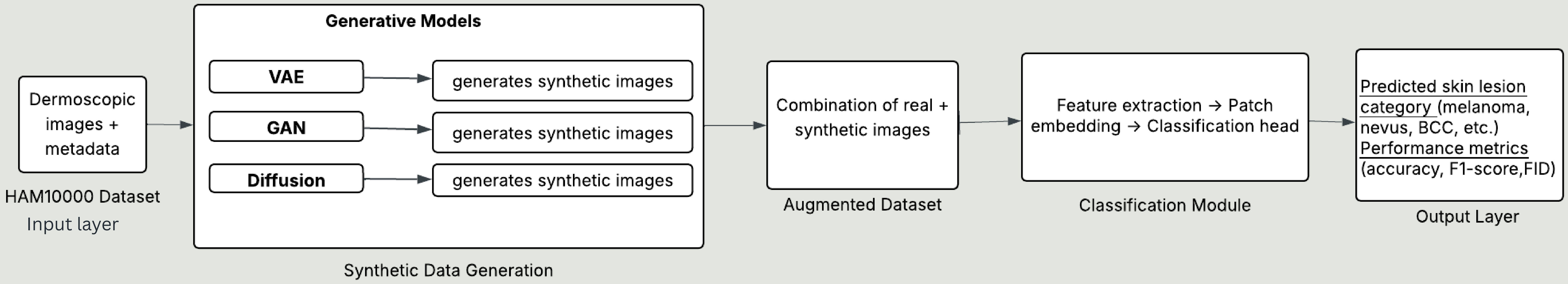
- Robust feature extraction using self-attention.
- Better handling of large-scale datasets than CNNs.
- Useful in multi-class diagnosis and decision support systems.

SYSTEM ARCHITECTURE / DESIGN



LEVEL - 0

SYSTEM ARCHITECTURE / DESIGN



LEVEL - 1

MATHEMATICAL FORMULAS

1. Variational Autoencoder (VAE)

VAE loss Function : Captures both images reconstruction and latent regularization

L_VAE = L_reconstruction + KL(q(z|x) || p(z))

MATHEMATICAL FORMULAS

2. Generative Adversarial Network (GAN)

Adversarial training between Generator (G) and Discriminator (D):

$$L_D = - [E_{x \sim pdata} [\log D(x)] + E_{z \sim p(z)} [\log(1 - D(G(z)))]]$$

$$L_G = - E_{z \sim p(z)} [\log D(G(z))]$$

MATHEMATICAL FORMULAS

3. Diffusion Model

Progressive denoising objective :

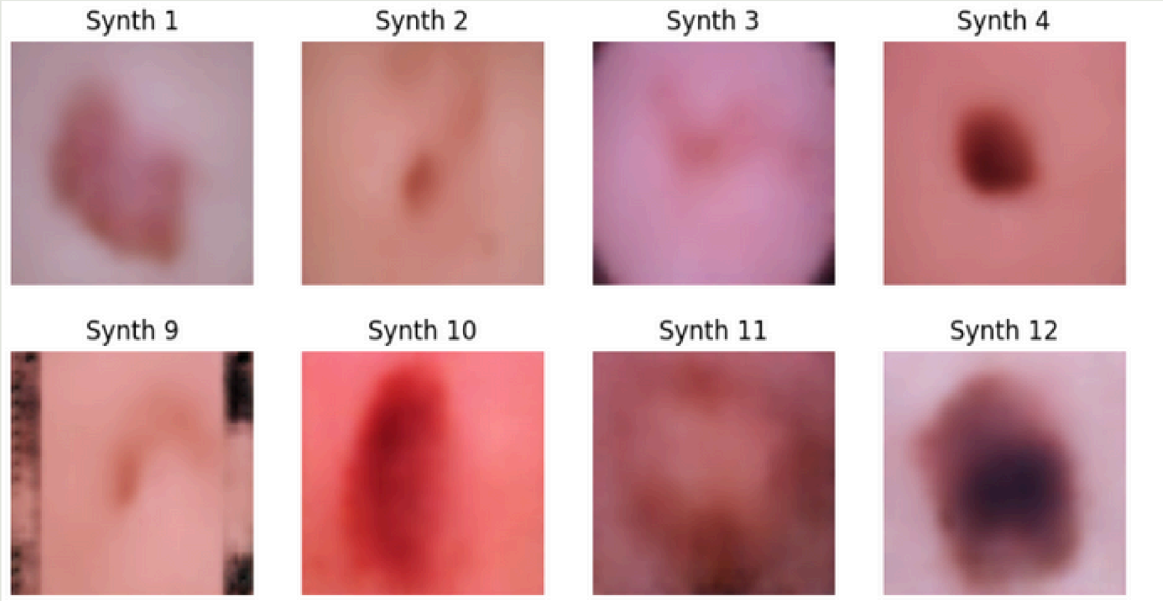
$$L_{diffusion} = E_{x, \varepsilon, t} \left[|| \varepsilon - \varepsilon_\theta(x_t, t) ||^2 \right]$$

MATHEMATICAL FORMULAS

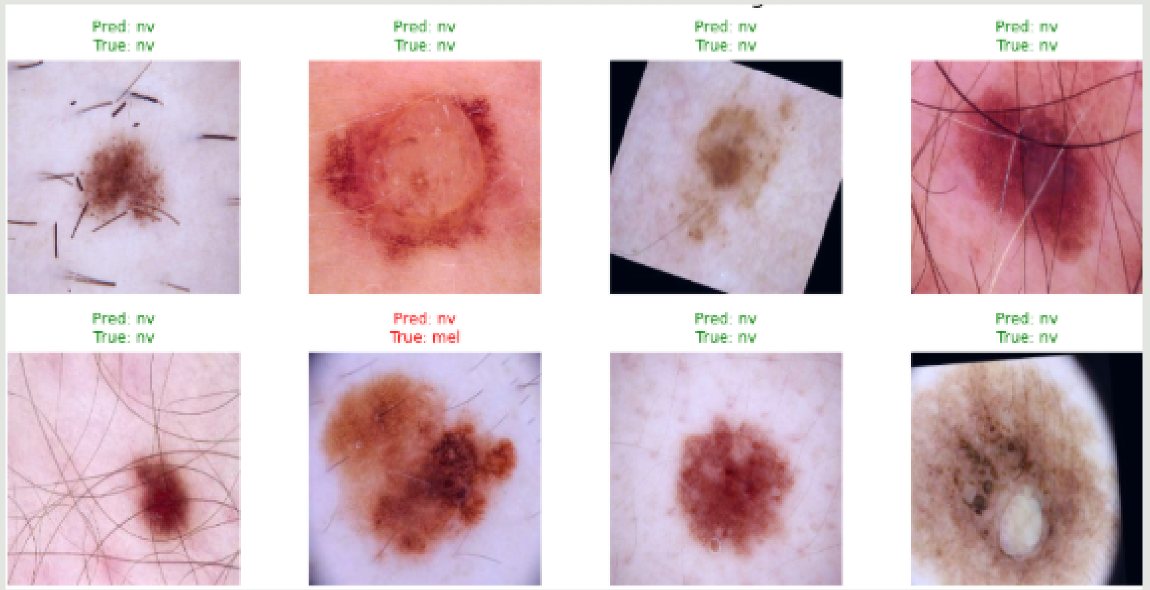
4. Vision Transformer (ViT)

Standard cross-entropy for multi-class diagnosis :

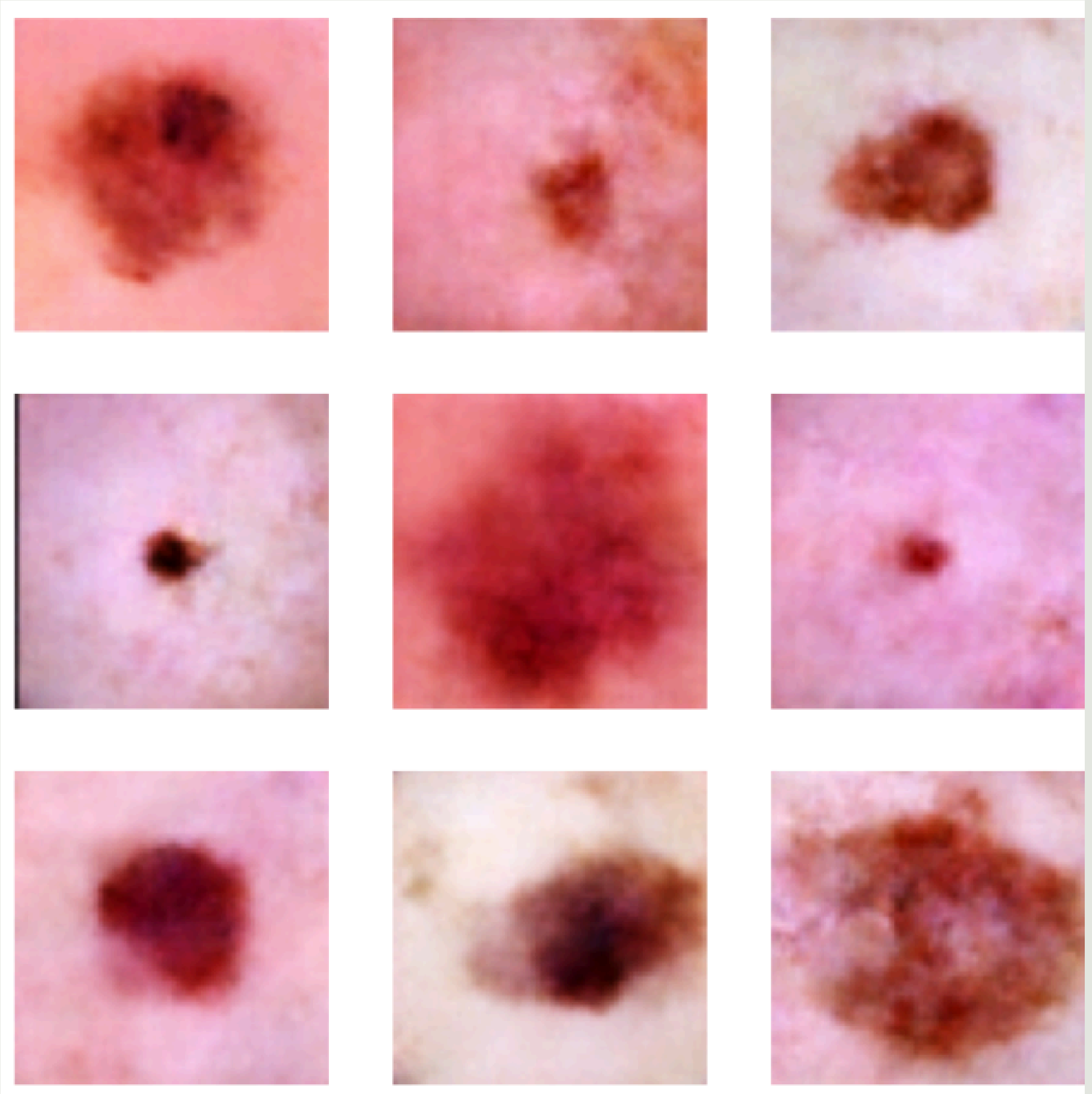
$$L_{CE} = - \Sigma_i \, y_i \log(\hat{y}_i)$$



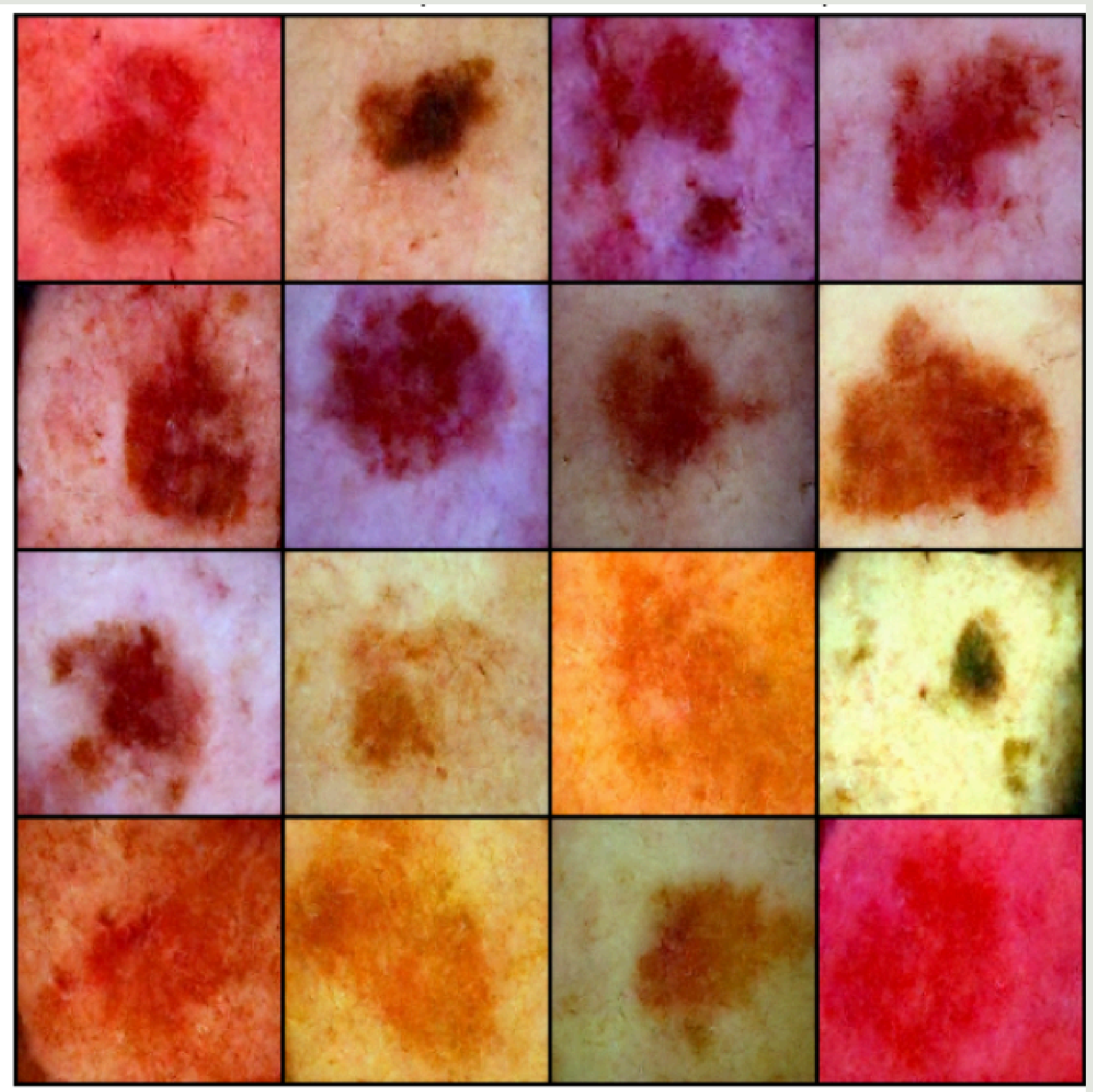
VAE



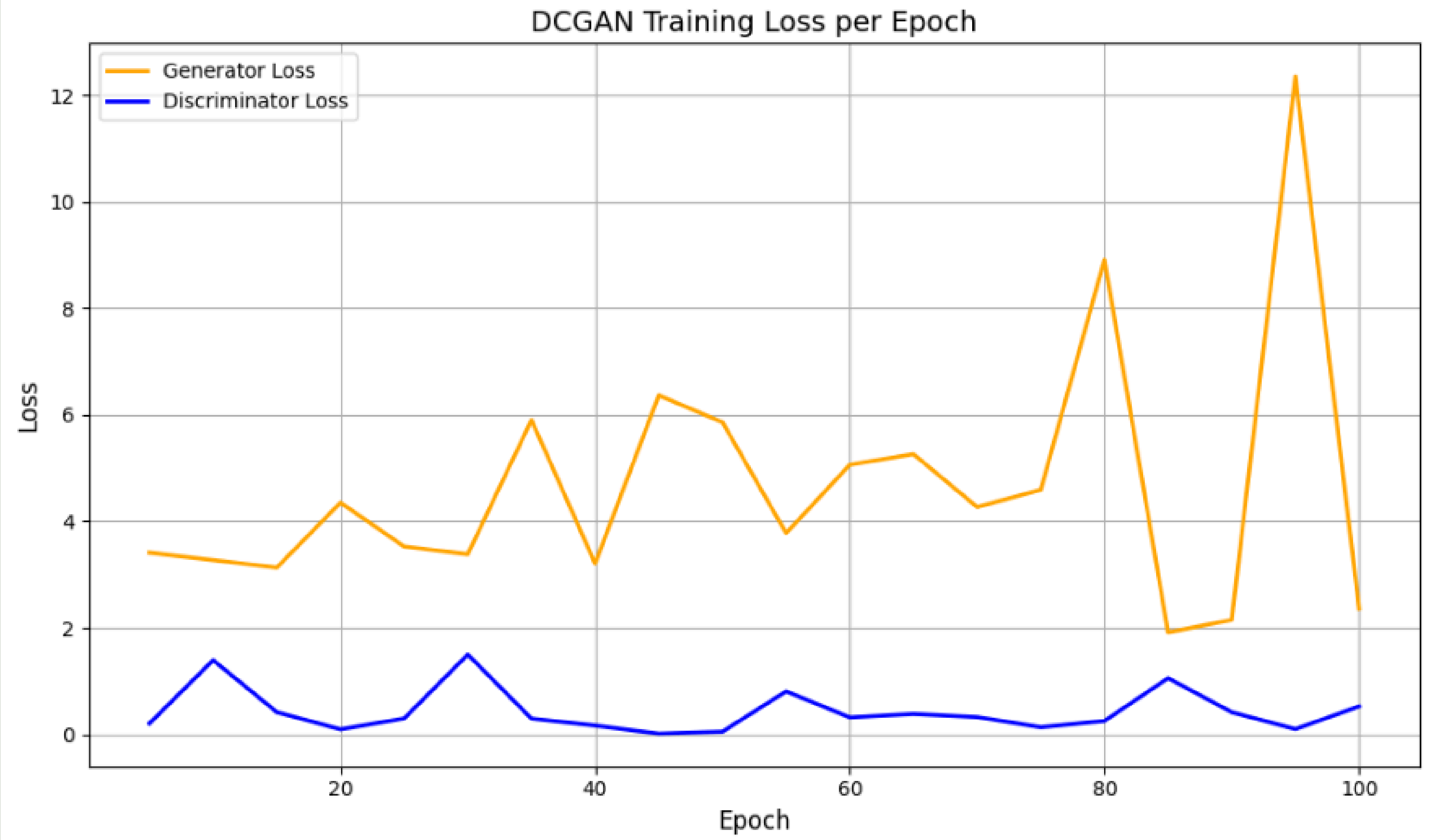
VIT

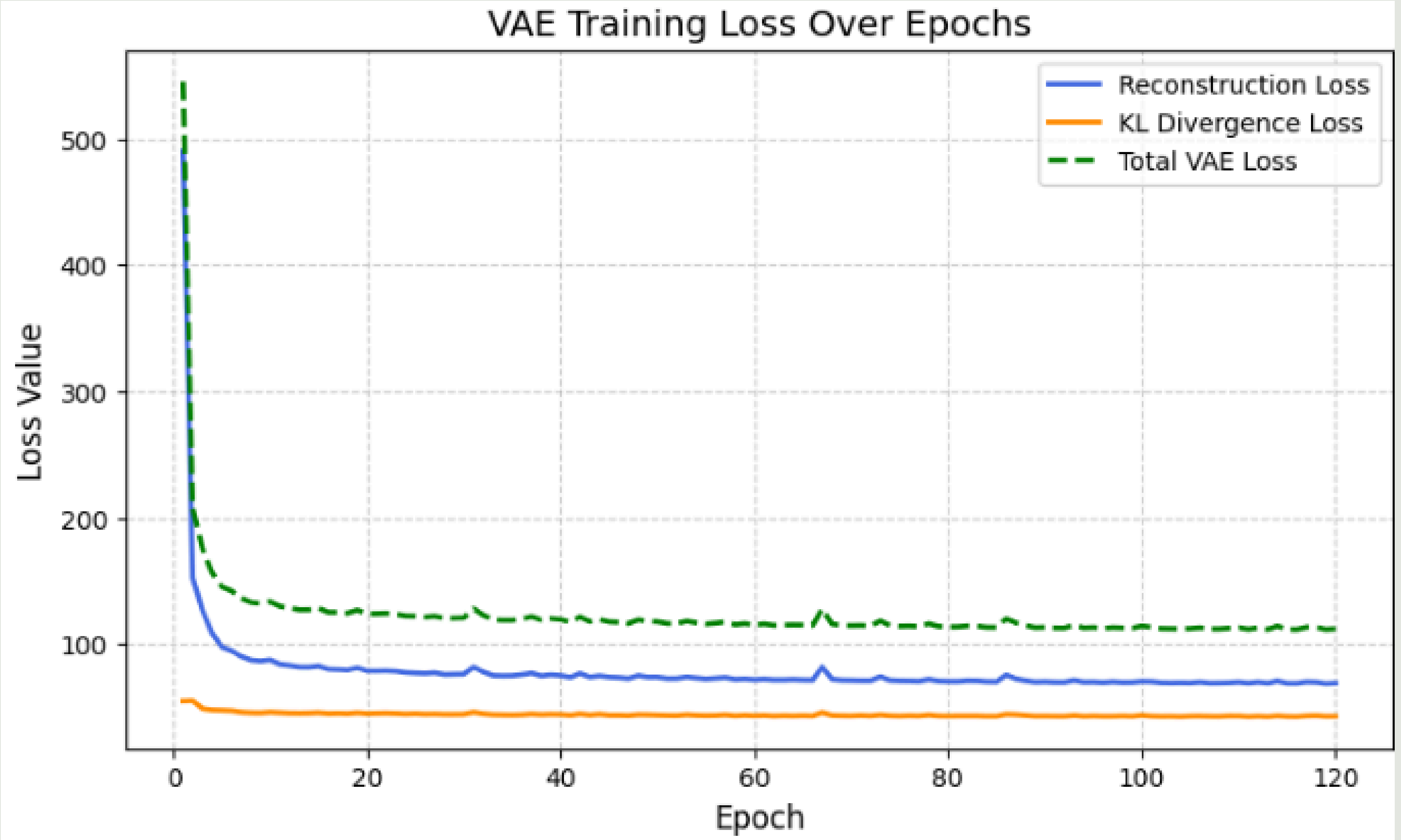


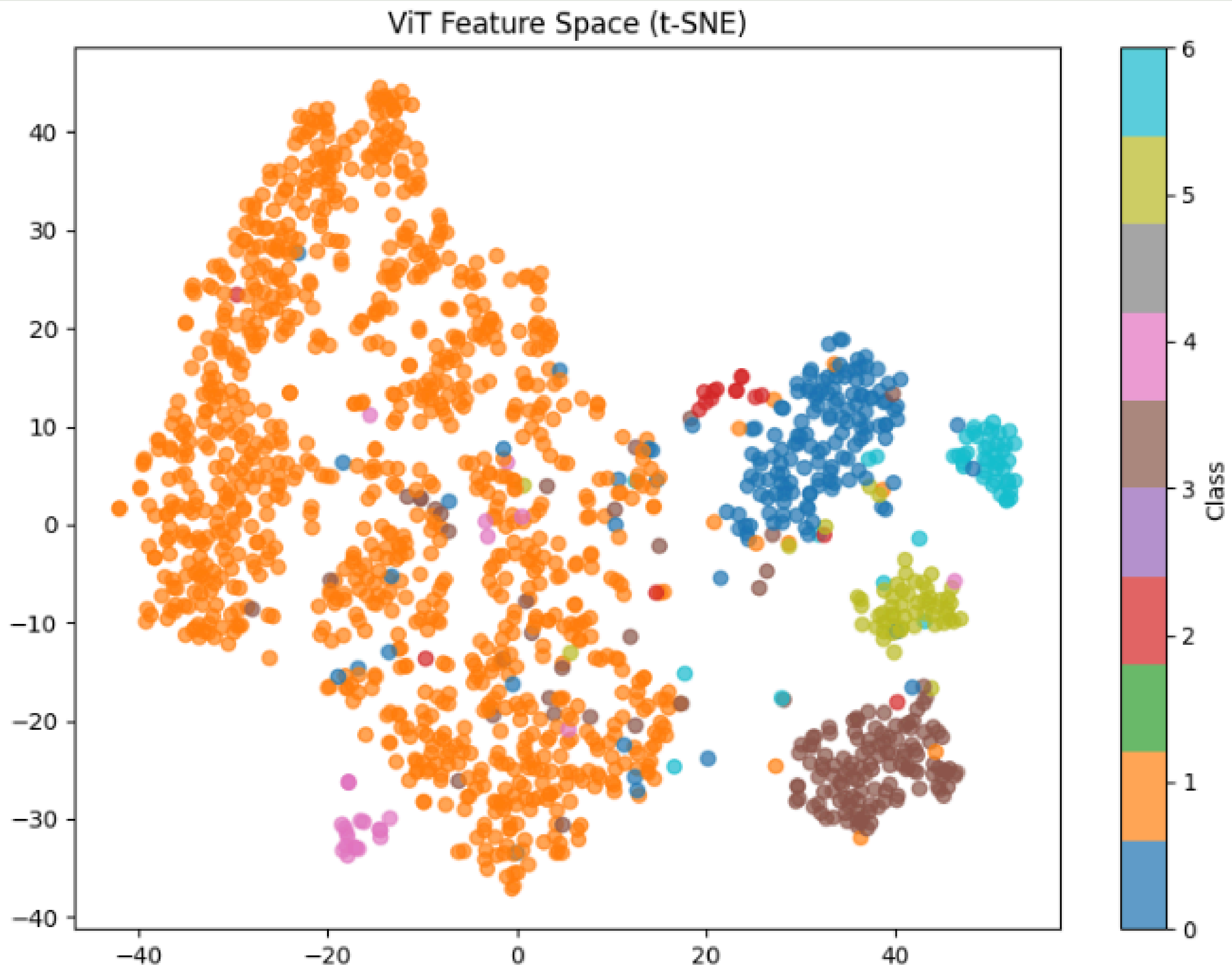
GAN

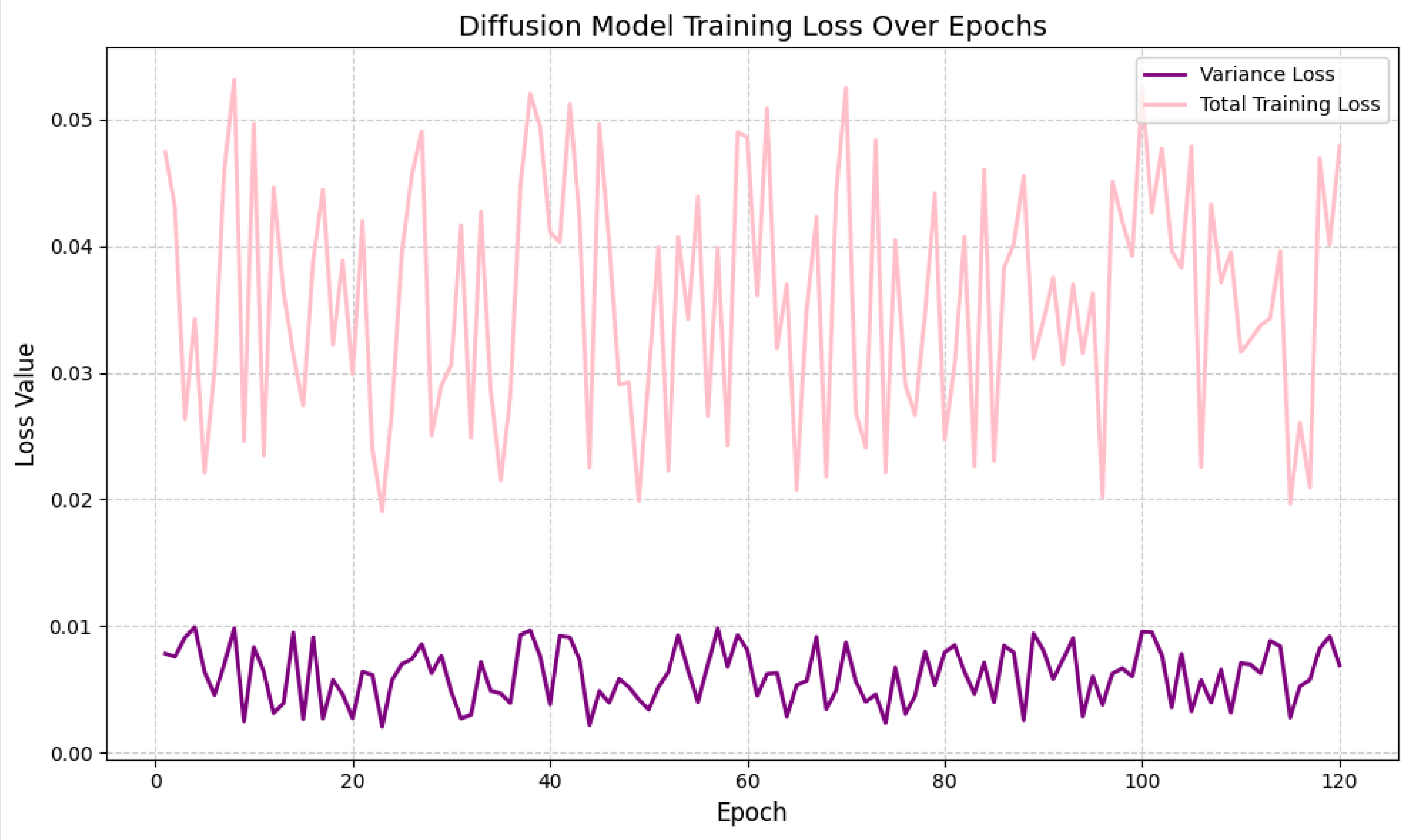


DIFFUSION


DCGAN Training Loss per Epoch
Generator Loss
Discriminator Loss
Loss
Epoch


VAE Training Loss Over Epochs
Reconstruction Loss
KL Divergence Loss
Total VAE Loss
Loss Value
Epoch


ViT Feature Space (t-SNE)
Class


Diffusion Model Training Loss Over Epochs
Variance Loss
Total Training Loss
Loss Value
Epoch

COMPARATIVE ANALYSIS

WITH EVALUATION METRICS

Metric	VAE	GAN	Diffusion	ViT
FID Score	250.98	156.54	203.44	—
SSIM	0.8099	0.6585	0.4537	—
LPIPS	0.1838	0.6247	0.7434	—
Inception Score	3.30 ± 0.08	2.82 ± 0.07	3.92 ± 0.38	—
Weighted F1-score	—	—	—	0.6996
Avg. Intra-class Cos Similarity	—	—	—	0.5148
Classifier Accuracy	—	—	—	0.7367

Parameter	VAE	GAN	Diffusion	ViT
Epochs	120	120	120	100
Optimizer	Adam	Adam	Adam	AdamW
Learning Rate (LR)	1e-3	0.0002	2e-4	5e-5
Batch Size	64	64	128	32
Loss Function	MSE + KL Divergence	Binary Cross-Entropy (adversarial)	MSE (noise prediction)	Cross-Entropy
Training Metrics	Recon Loss: 69.1633, KL Loss: 42.9225	D Loss: 0.2028, G Loss: 2.1655	Best Model @ Epoch 90, Final Loss: 0.0177	Train Acc: 1.0000, Val Loss: 3.6208, Val Acc: 0.7367
Complexity	Low (simple encoder-decoder)	Medium (two-network training)	High (iterative denoising steps)	Very High (transformer-based)
Inference Speed	Fast	Medium	Slow	Fast
Method Type	Probabilistic latent variable	Adversarial generation	Denoising-based generation	Transformer-based classification
Activation Function	ReLU, Sigmoid	LeakyReLU, Tanh	SiLU (Swish)	GELU
Output Nature	Smooth latent reconstructions	Sharp realistic images	High-quality, noise-free samples	Class probabilities
Training Objective	Minimize reconstruction + KL divergence	Generator fools discriminator	Predict and remove noise progressively	Minimize classification cross-entropy

ETHICAL CONSIDERATIONS

- **Data Privacy (HIPAA):** Ensure all medical images are anonymized and compliant with global data protection laws.
- **Informed Consent:** Use datasets only with proper patient consent and ethical approval.
- **Bias & Fairness:** Prevent demographic or skin-type bias in generated synthetic data and model predictions.
- **Transparency & Accountability:** Clearly document model behavior, data sources, and synthetic generation methods.
- **Secure Use & Storage:** Protect datasets from misuse through encryption, limited access, and research-only usage.

CONCLUSION

This project demonstrates that synthetic data generation using VAE, GAN, and Diffusion models can effectively address challenges of data scarcity, imbalance, and privacy in healthcare datasets. By leveraging the HAM10000 skin lesion dataset and applying advanced generative models, we created diverse, high-quality synthetic samples that enhance training for diagnostic systems. Among these, the Diffusion model performed best, producing the most realistic and high-quality images with the lowest loss. The integration of Transformer-based ViT further improved classification performance, highlighting the potential of combining synthetic data with state-of-the-art models for robust and reliable medical AI solutions.

THANK YOU!!!