Generating Blood Cells with Machine Learning

Débora Sofia Valério Taborda de Seiça

University of Coimbra, Portugal
uc2018278107@student.uc.pt

1 Introduction

Generative machine learning models have recieved significant attention in recent years due to their capability to synthesize realistic and diverse data samples. Such models learn the probability distributions of real-world datasets and generate new instances. These techniques have numerous applications, including data augmentation, anomaly detection, image restoration, and even medical diagnostics [6].

In the context of medical imaging, generative approaches have the potential to mitigate common challenges such as limited data availability, imbalanced class distribution, and the high costs associated with obtaining labeled medical data. Specifically, generative models like Generative Adversarial Networks [1] (GANs), Variational Autoencoders [4] (VAEs), and Diffusion Models [3] have shown great promise in generating realistic medical images.

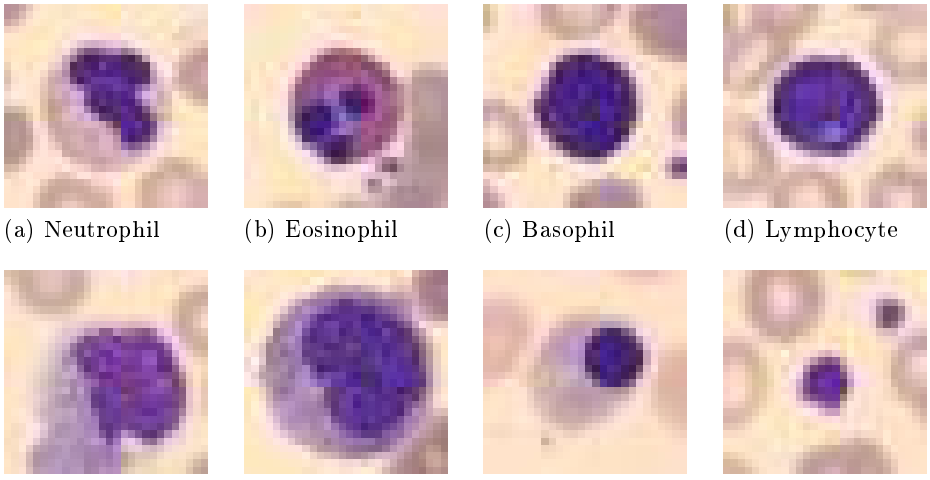
In this project, I address the task of synthesizing realistic blood cell images from the BloodMNIST dataset [5]. The BloodMNIST dataset comprises RGB images of peripheral blood smear cells, categorized into eight distinct cell types: neutrophils, eosinophils, basophils, lymphocytes, monocytes, immature granulocytes, erythroblasts, and platelets (Fig. 1).

2 Problem Statement

The primary objective of this study is to implement, train, and comparatively evaluate three prominent generative model classes: GANs, VAEs, and Diffusion Models. We assess their effectiveness quantitatively by computing the Fréchet Inception Distance (FID) [2], a robust metric to measure the similarity between generated and real image distributions. Additionally, we discuss qualitative results to highlight each model's strengths and limitations. The ultimate goal is to identify an optimal generative framework capable of producing realistic and diverse synthetic blood cell images.

3 Approach

We implemented and evaluated three classes of generative models. Each model was adapted to generate 28×28 RGB images that resemble the blood cell samples in the BloodMNIST dataset.



(a) Neutrophil (b) Eosinophil (c) Basophil (d) Lymphocyte

(e) Monocyte (f) I. Granulocyte (g) Erythroblast (h) Platelet

Fig. 1. Examples of each blood cell class from the BloodMNIST dataset.

For each approach, the model was trained on the entire dataset, combining the original training, validation, and test splits. All images were normalized to the $[-1, 1]$ range. We maintained a batch size of 128 and used the Adam optimizer for all models. Each training was repeated with five random seeds for later evaluation stability.

VAE: The Variational Autoencoder consists of a convolutional encoder that maps the input image into a latent Gaussian distribution, and a decoder that reconstructs the image from sampled latent vectors. The model is trained using a reconstruction loss combined with a Kullback-Leibler divergence term to regularize the latent space [4].

GAN: We implemented a Deep Convolutional GAN (DCGAN), with a generator that learns to synthesize realistic images from latent noise vectors, and a discriminator that learns to distinguish between real and generated samples [1]. Both networks use strided convolutions and LeakyReLU activations. The generator is trained to fool the discriminator, while the discriminator is trained to correctly classify real and fake inputs.

Diffusion: The DDPM is a generative model based on a Markovian process that gradually adds noise to images and then learns to reverse this corruption step by step [3]. The training objective is to predict the added noise at each timestep, enabling the model to sample clean images from pure noise.

To ensure fairness, all models were trained using the same hardware environment and epochs, and evaluated using the Fréchet Inception Distance (FID) as a common metric.

4 Experimental Setup

All code and experiments are documented in the accompanying Jupyter Notebook, developed with Python 3.10.12 and PyTorch 2.1.0, running under CUDA 11.8 on a machine equipped with an NVIDIA GeForce GTX 970 GPU. The BloodMNIST dataset was loaded via the MedMNIST v2 framework, and all images were resized to 28×28 with three color channels. To simplify the setup, the training, validation, and test splits were merged into a single training set of 17,092 images.

Each generative model was trained for 20 epochs using a batch size of 128. The Adam optimizer was employed with default parameters ($\beta_1 = 0.9$, $\beta_2 = 0.999$), and a learning rate of 10^{-3} for the VAE and Diffusion models, and 2×10^{-4} for the GAN.

During evaluation, each model generated 10,000 synthetic images from latent noise vectors (or denoised noise in the case of the diffusion model). These samples were compared against 10,000 randomly selected real images using FID. To ensure statistical robustness, this evaluation was repeated across five independent runs with different different seeds. The final score for each model is reported as the mean μ and standard deviation σ of the FID values across these runs.

Let FID_i denote the score obtained in run i, for $i = 1, \ldots, 5$. The reported result is

$$\text{FID}_{\text{final}} = \mu \pm \sigma \tag{1}$$

where

$$\mu = \frac{1}{5} \sum_{i=1}^{5} \text{FID}_i \tag{2}$$

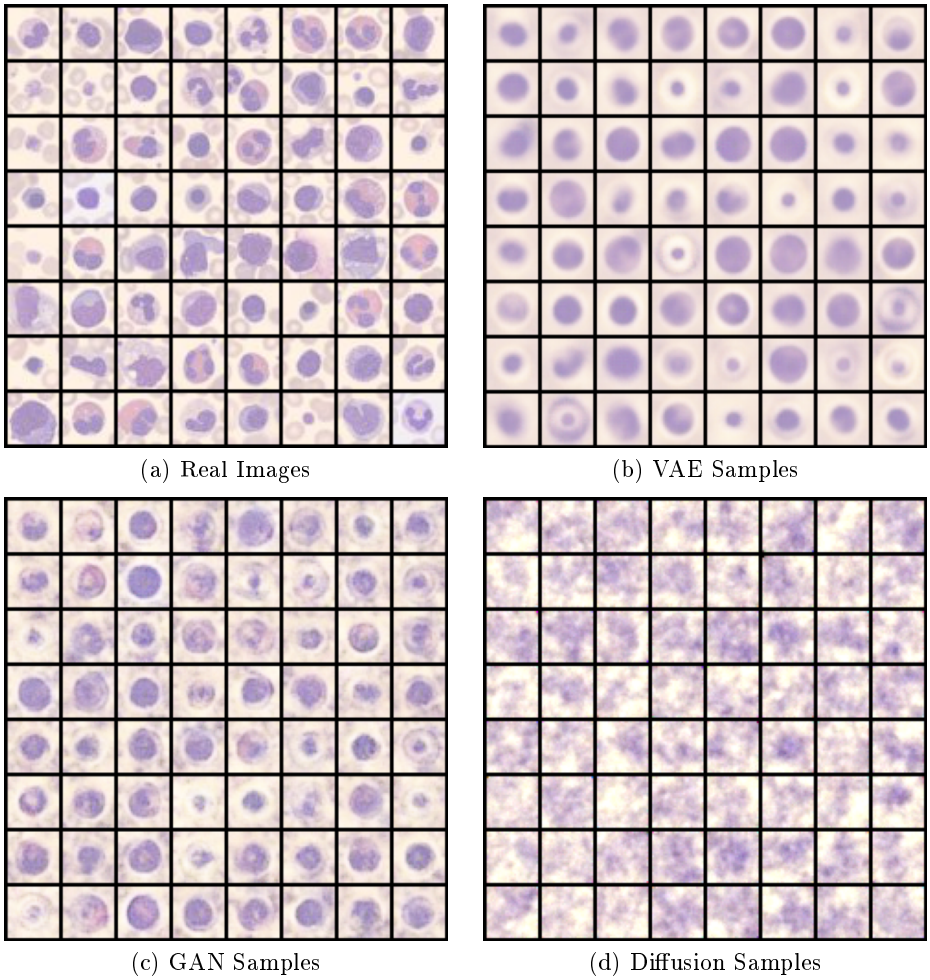
$$\sigma = \sqrt{\frac{1}{5} \sum_{i=1}^{5} (\text{FID}_i - \mu)^2} \tag{3}$$

5 Results and Analysis

Table 1 summarizes the mean and standard deviation of the FID scores obtained for each model across five independent runs.

Table 1. Fréchet Inception Distance (FID) results over 5 runs. Lower is better.

Model	Mean	Std. Dev.
VAE	201.73	0.23
GAN	180.57	0.71
Diffusion	423.70	0.64



(a) Real Images (b) VAE Samples

(c) GAN Samples (d) Diffusion Samples

Fig. 2. Qualitative comparison of real and generated blood cell images.

Figure 2 shows both the real images and the visual samples generated by each model. The results reveal clear qualitative differences. The VAE samples exhibit smooth but often blurry outputs due to its probabilistic reconstruction objective. Diffusion models produce a lump of blobs, barely resembling a blood cell. GANs generate highly detailed and diverse samples, closely resembling real blood cells.

Overall, GANs achieve the best quantitative and qualitative performance, indicating superior capability in modeling complex visual structures present in biomedical data.

6 Conclusion

In this study, we implemented and evaluated three prominent classes of generative models, Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), and Denoising Diffusion Probabilistic Models (DDPMs), on the BloodMNIST dataset. Each model was trained under consistent conditions and assessed using both qualitative visualizations and the Fréchet Inception Distance (FID) metric.

The results revealed clear differences in performance. The VAE produced smooth but blurry reconstructions due to its probabilistic decoder. Diffusion models failed to produce visually coherent images, resulting in noisy outputs that barely resemble blood cells. In contrast, GANs achieved the best visual quality, generating diverse and realistic samples that closely match real data.

These observations align with the FID scores reported in Table 1, where GANs obtained the lowest value (180.57), outperforming both VAEs (201.73) and diffusion models (423.70). This demonstrates the effectiveness of adversarial training for modeling complex biomedical structures.

Overall, GANs offer the best balance of fidelity and diversity for blood cell image synthesis in this context. Future work may explore hybrid or conditional architectures to further improve control over generated samples and class specificity.

References

1. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. Advances in Neural Information Processing Systems **27**, 2672–2680 (2014)
2. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. Advances in Neural Information Processing Systems **30**, 6626–6637 (2017)
3. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems **33**, 6840–6851 (2020)
4. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 (2013)

5. Yang, J., Shi, R., Wei, D., Liu, Z., Zhao, L., Ke, B., Pfister, H., Ni, B.: Medmnist v2: A large-scale lightweight benchmark for 2d and 3d biomedical image classification. Scientific Data **10**, 41 (2023)
6. Yi, X., Walia, E., Babyn, P.: Generative adversarial network in medical imaging: A review. Medical Image Analysis **58**, 101552 (2019)