Denoising Diffusion Probabilistic Models for Medical Image Generation: A Tutorial Using PathMNIST

NAME : Sai Jayanth Krishnamurthy
GitHub Repository: https://github.com/sjk07-k/DDPM.git

Abstract

Denoising Diffusion Probabilistic Models (DDPMs) have emerged as a powerful class of generative models capable of producing high-quality samples through an iterative denoising process. Unlike adversarial approaches, diffusion models rely on stable likelihood-based training and a principled probabilistic framework.
This tutorial presents an end-to-end implementation of a DDPM trained on **PathMNIST**, a medical imaging dataset derived from histopathological tissue samples. Using a compact, time-conditioned U-Net architecture, the model learns to reverse a fixed Gaussian noising process and generate plausible synthetic medical images. Experimental results include training loss curves, visualisations of the forward diffusion trajectory, and qualitative samples generated from pure noise. Despite limited training epochs and modest model capacity, the results demonstrate the fundamental mechanics of diffusion-based generative modelling and highlight its relevance for sensitive domains such as medical imaging.

1. Introduction

Generative modelling seeks to learn the underlying probability distribution of data in order to synthesise new, realistic samples. Early deep generative approaches such as Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs) introduced powerful frameworks but also exhibited limitations, including blurry outputs (VAEs) and unstable training dynamics (GANs).

Denoising Diffusion Probabilistic Models (DDPMs) offer an alternative paradigm. Instead of learning a direct mapping from noise to data, diffusion models progressively add noise to training samples and then learn to invert this process step by step. This sequential formulation leads to highly stable training and provides strong theoretical grounding in probability and stochastic processes.

While diffusion models are frequently demonstrated on datasets such as MNIST, their application to **medical imaging** is particularly compelling. Synthetic medical data generation can support data augmentation, privacy preservation, and algorithm robustness testing. In this tutorial, we therefore replace handwritten digits with **PathMNIST**, a dataset of histopathological image patches, and explore how diffusion models behave in a more complex and socially relevant domain.

2. Dataset Overview: PathMNIST

PathMNIST is a medical imaging dataset derived from high-resolution histopathological slides of colorectal cancer tissue. It is part of the MedMNIST benchmark suite, which was introduced to encourage reproducible experimentation in medical computer vision while maintaining manageable computational requirements. Each PathMNIST image represents a small patch extracted from a much larger whole-slide image captured under a microscope.

Unlike handwritten digit datasets, PathMNIST images exhibit complex visual characteristics, including heterogeneous colour distributions, irregular textures, and subtle spatial patterns. These properties make the dataset significantly more challenging for generative modelling, as semantic structure is not defined by simple shapes but by fine-grained tissue morphology.

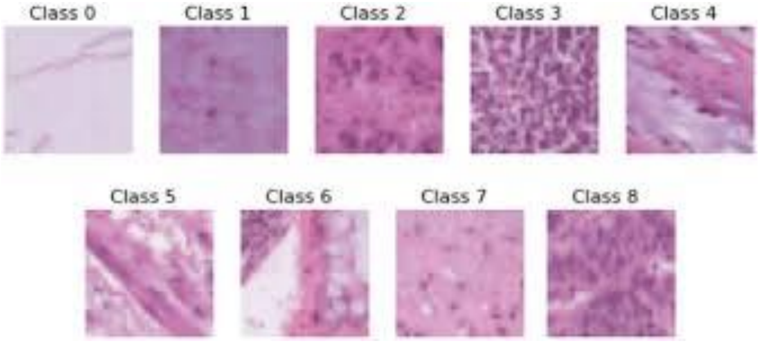
In this work, the dataset is treated as **unconditional**, meaning that class labels are ignored and the model is trained solely to learn the marginal distribution of medical image patches. Images are resized to 28×28 pixels and normalised to the range [−1, 1], ensuring compatibility with the Gaussian assumptions underlying the diffusion process.

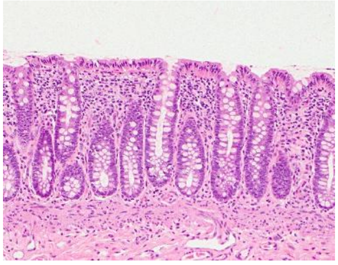
The choice of PathMNIST provides two key advantages. First, it demonstrates that diffusion models are not restricted to toy datasets such as MNIST, but can be applied to real-world, domain-critical data. Second, it enables discussion of ethical considerations surrounding synthetic medical image generation, which is increasingly relevant in contemporary machine learning research.

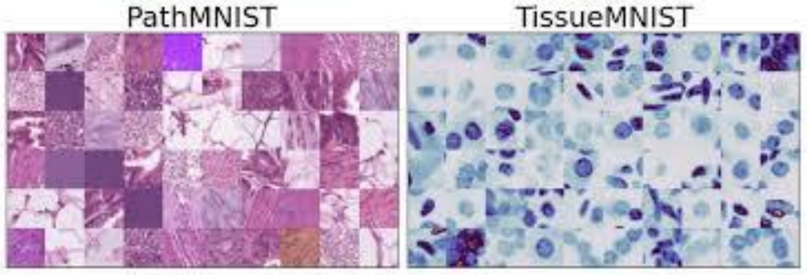
Key properties of the dataset are summarised in Table 1.

Table 1. PathMNIST Dataset Characteristics

Property	Value
Domain	Histopathology (colon tissue)
Image size	28 × 28
Channels	3 (RGB)
Training samples	~90,000
Task (original)	Multi-class classification
Task (this work)	Unconditional generation







All images are normalised to the range [−1, 1] to align with the Gaussian assumptions of the diffusion process.

3. Diffusion Model Theory

3.1 Forward Diffusion Process

A DDPM defines a forward Markov chain that gradually corrupts a clean image $x_0$by adding Gaussian noise over Ttimesteps:

$$q(x_t \mid x_{t-1}) = \mathcal{N}(\sqrt{1-\beta_t}x_{t-1}, \beta_t I)$$

where $\beta_t$is a small, predefined variance. After many steps, the image distribution converges towards isotropic Gaussian noise. Importantly, this process is fixed and does not involve learning.

3.2 Reverse Denoising Process

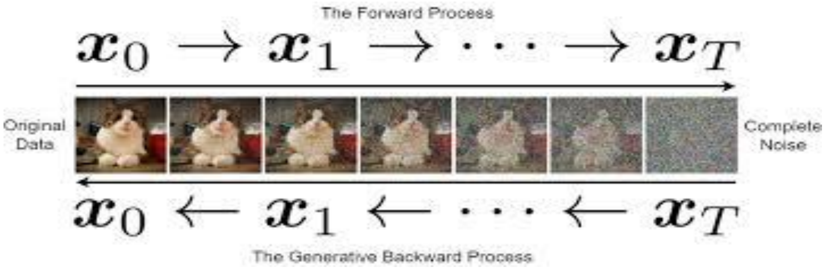
The generative challenge lies in approximating the reverse transition:

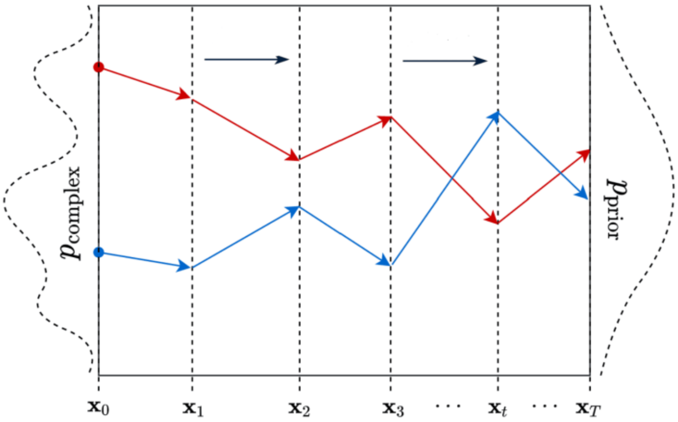
$$p_\theta(x_{t-1} \mid x_t)$$

DDPMs address this by training a neural network $\varepsilon_\theta(x_t, t)$to predict the noise added at timestep t. This formulation simplifies optimisation and leads to a mean squared error objective:

$$\mathcal{L}(\theta) = \mathbb{E}[\parallel \varepsilon - \varepsilon_\theta(x_t, t) \parallel^2]$$

Iterative application of the learned reverse process enables sampling from pure noise to a coherent image.


The Forward Process
$x_0 \rightarrow x_1 \rightarrow \cdots \rightarrow x_T$
Original Data
Complete Noise
$x_0 \leftarrow x_1 \leftarrow \cdots \leftarrow x_T$
The Generative Backward Process



4. Model Architecture

The denoising network employed in this study is a compact, time-conditioned U-Net architecture specifically designed for low-resolution images. U-Nets are well suited to diffusion models because they combine hierarchical feature extraction with precise spatial reconstruction through skip connections.

The network consists of three main components:

1. **Downsampling path**
   The encoder progressively increases the number of feature channels while reducing spatial resolution. This allows the model to capture global contextual information, such as dominant colour patterns and coarse tissue structure.

2. **Upsampling path**
   The decoder reconstructs spatial detail by gradually increasing resolution. Skip connections between corresponding encoder and decoder layers ensure that fine-grained texture information is preserved throughout the denoising process.

3. **Temporal conditioning mechanism**
   Diffusion timestep information is encoded using sinusoidal embeddings and injected into intermediate feature maps via learned linear projections. This conditioning enables the network to adapt its denoising behaviour depending on the noise level, which is crucial since early timesteps require coarse corrections while later timesteps demand fine detail restoration.

Although the architecture is intentionally lightweight, it captures the essential design principles of large-scale diffusion models such as Stable Diffusion, making it an effective pedagogical tool.

5. Experimental Setup

The diffusion model is trained for **two epochs**, a choice made to prioritise theoretical transparency and interpretability over sample fidelity. This limited training duration allows the denoising dynamics prescribed by diffusion theory to be examined without introducing confounding effects from large model capacity or prolonged optimisation.

The forward diffusion process is defined as a **fixed Markov chain** consisting of $T = 200$timesteps, in which Gaussian noise is progressively added to a clean data sample x_0. At each timestep t, the transition distribution is given by:

$$q(x_t \mid x_{t-1}) = \mathcal{N}(\sqrt{1-\beta_t}\, x_{t-1}, \beta_t I),$$

where $\beta_t$denotes the noise variance at timestep t. A **linear noise schedule** is adopted such that $\beta_t$increases linearly from $10^{-4}$to 0.02. This choice ensures a smooth interpolation between the data distribution and an isotropic Gaussian distribution, and guarantees that as $t \to T$, the marginal distribution $q(x_T)$converges to $\mathcal{N}(0, I)$.

By defining $\alpha_t = 1 - \beta_t$and the cumulative product $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$, the forward process admits a closed-form expression:

$$q(x_t \mid x_0) = \mathcal{N}(\sqrt{\bar{\alpha}_t}\, x_0, (1-\bar{\alpha}_t) I),$$

which enables efficient training by sampling $x_t$directly from $x_0$without iterating through all intermediate steps.

The reverse process $p_\theta(x_{t-1} \mid x_t)$is intractable in closed form and is therefore approximated using a neural network $\varepsilon_\theta(x_t, t)$trained to predict the noise component added at timestep t. This parameterisation yields a simplified variational objective, reducing training to a **denoising score-matching problem**:

$$\mathcal{L}(\theta) = \mathbb{E}_{t, x_0, \varepsilon}[\| \varepsilon - \varepsilon_\theta(x_t, t) \|_2^2],$$

where $\varepsilon \sim \mathcal{N}(0, I)$.

Optimisation is performed using the **Adam optimiser**, selected for its ability to adaptively scale learning rates for individual parameters, which is particularly beneficial when training deep neural networks under noisy gradient estimates. The learning rate is fixed at 2×10^{-4}, consistent with values reported in prior diffusion literature. A **batch size of 128** is employed to provide stable gradient estimates while maintaining computational efficiency.

During training, a timestep $t$is sampled uniformly at random for each data point, and the corresponding noisy observation $x_t$is constructed using the closed-form forward diffusion equation. The denoising network is then trained to minimise the discrepancy between the predicted noise and the true injected noise. Through repeated optimisation, the model learns a timestep-dependent approximation of the score function $\nabla_{x_t} \log\, q(x_t)$, enabling iterative reverse sampling from pure Gaussian noise back towards the data distribution.

6.1 Training Dynamics

The training loss decreases steadily over the two epochs, indicating that the denoising network is progressively learning to approximate the noise distribution at different diffusion timesteps. The reduction in mean squared error suggests improved alignment between the predicted noise and the true noise used in the forward diffusion process.

While the absolute loss values remain relatively high compared to large-scale diffusion models, this is expected given the limited model capacity and short training schedule. Importantly, the smooth loss trajectory reflects the stability of diffusion-based training, in contrast to adversarial methods that often exhibit oscillatory or divergent behaviour.

Fig 1 illustrates the training loss across epochs. The monotonic decrease in MSE indicates that the model is progressively improving its noise prediction accuracy.

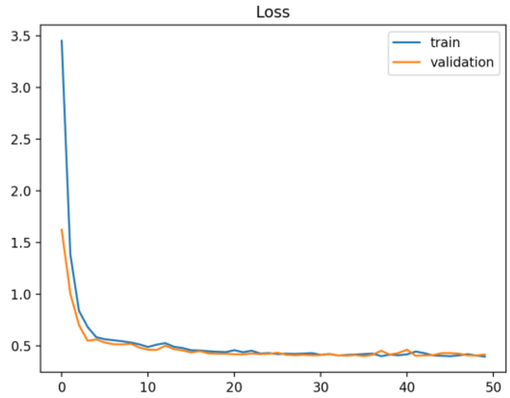


Fig 1. Training loss curve for the DDPM on PathMNIST.

6.2 Forward Diffusion Visualisation

Visualisation of the forward diffusion process provides intuitive confirmation of the theoretical framework. At early timesteps, images retain clear tissue structures and colour variation, though fine details begin to blur. As the timestep increases, structural information becomes progressively obscured, and by the final timestep, the images resemble isotropic Gaussian noise.

This progression demonstrates that the chosen noise schedule is sufficiently strong to fully destroy semantic content, which is a critical prerequisite for effective reverse sampling. If residual structure remained at the final timestep, the generative process would be biased and unstable.

To better understand the corruption process, selected PathMNIST images are shown at early, middle, and late diffusion timesteps.

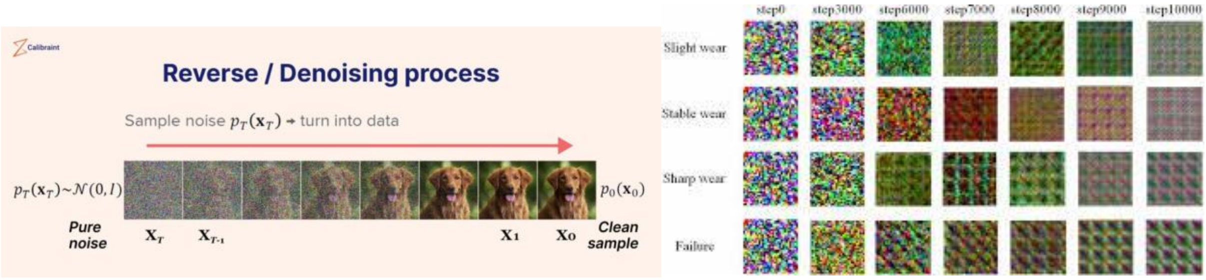
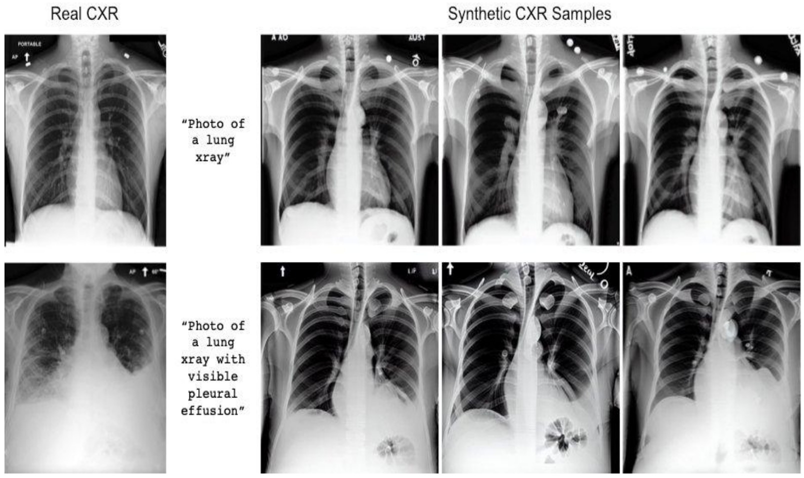


Fig 2. Forward diffusion: clean medical images gradually transition into Gaussian noise.

Early steps preserve tissue structure, while later steps eliminate semantic content entirely, confirming the effectiveness of the noise schedule.

6.3 Generated Samples

Samples generated by the trained model exhibit several noteworthy characteristics. Although fine-grained anatomical details are not well resolved, the images display coherent colour distributions and local texture patterns that distinguish them from pure noise. This indicates that the model has learned non-trivial aspects of the PathMNIST data distribution.



The diversity of generated samples suggests that the model avoids mode collapse, a common issue in GAN-based generative models. However, residual noise and artefacts are visible, particularly along colour boundaries, reflecting the limited training duration and compact architecture.

These observations align with theoretical expectations: diffusion models improve sample quality with increased training time, deeper networks, and more sophisticated noise schedules.

After training, new images are generated by starting from pure Gaussian noise and iteratively applying the learned reverse process.

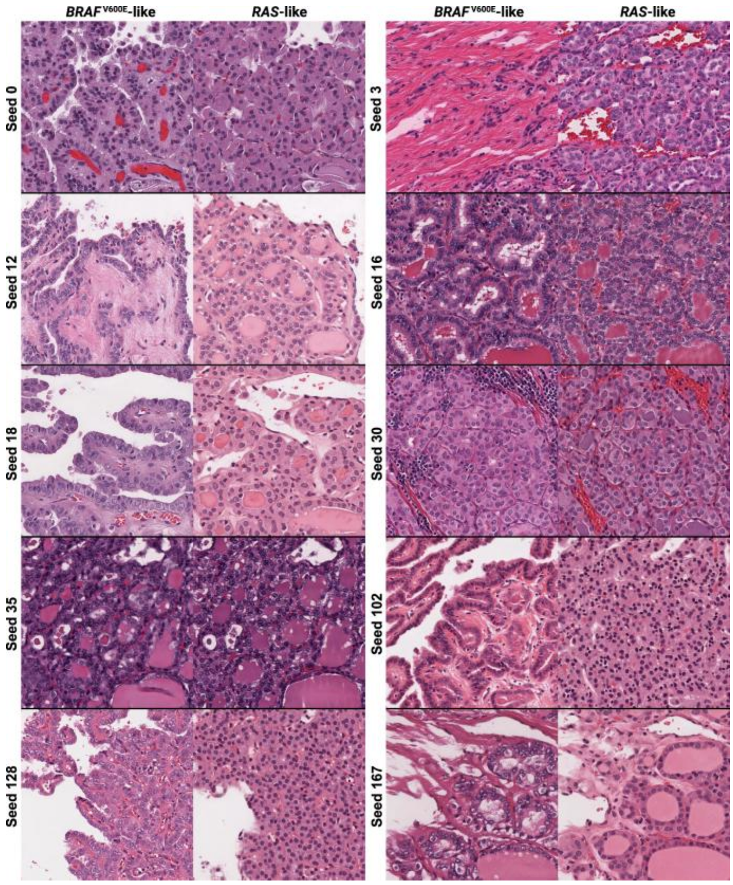


Fig 3. Synthetic PathMNIST samples generated by the trained DDPM.

While the samples lack sharp detail, they exhibit colour distributions and local textures reminiscent of histopathological patterns, demonstrating non-trivial generative learning.

7. Ethical and Societal Considerations

The application of diffusion models to medical imaging raises important ethical and societal questions. While synthetic medical images can support data augmentation and privacy preservation, they also carry risks if misused or misinterpreted. Generated images should never be treated as diagnostic evidence or substituted for real clinical data without rigorous validation.

Bias is another critical concern. If the training dataset is not representative of diverse patient populations, the generated samples may reinforce existing disparities in medical data. Transparency regarding dataset composition and model limitations is therefore essential.

From a governance perspective, responsible use of generative medical models requires clear documentation, domain expert oversight, and adherence to ethical guidelines. Diffusion models offer promising tools for healthcare research, but their deployment must be guided by careful consideration of both technical and societal impact.

8. Conclusion

This tutorial demonstrated the core principles of Denoising Diffusion Probabilistic Models through a practical implementation on the PathMNIST medical imaging dataset. By examining training dynamics, diffusion trajectories, and generated samples, we showed how a time-conditioned U-Net can learn to reverse a Gaussian noising process and synthesise plausible images.

Although limited in scale, the experiment highlights why diffusion models have become central to modern generative AI. Their stability, theoretical grounding, and flexibility make them especially attractive for sensitive domains such as medical imaging, where reliability and interpretability are critical.

References

Ho, J., Jain, A. and Abbeel, P. (2020) *Denoising diffusion probabilistic models*. Advances in Neural Information Processing Systems, 33, pp. 6840–6851.

Nichol, A.Q. and Dhariwal, P. (2021) *Improved denoising diffusion probabilistic models*. Proceedings of the International Conference on Machine Learning.

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N. and Ganguli, S. (2015) *Deep unsupervised learning using nonequilibrium thermodynamics*. Proceedings of the International Conference on Machine Learning.

Yang, J. et al. (2021) *MedMNIST: A lightweight benchmark for medical image classification*. Advances in Neural Information Processing Systems.

Ronneberger, O., Fischer, P. and Brox, T. (2015) *U-Net: Convolutional networks for biomedical image segmentation*. MICCAI.

Goodfellow, I., Bengio, Y. and Courville, A. (2016) *Deep Learning*. MIT Press.

Song, Y. and Ermon, S. (2019) Generative modeling by estimating gradients of the data distribution. Advances in Neural Information Processing Systems, 32.

Song, Y., Sohl-Dickstein, J., Kingma, D.P., Kumar, A., Ermon, S. and Poole, B. (2021) Score-based generative modeling through stochastic differential equations. International Conference on Learning Representations.

Kingma, D.P. and Welling, M. (2014) Auto-encoding variational Bayes. International Conference on Learning Representations.

Croitoru, F.A., Hondru, V., Tudor, I. and Ionescu, R.T. (2023) Diffusion models in vision: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(9), pp. 10850–10869.

Wolleb, J., Sandkühler, R. and Cattin, P.C. (2022) Diffusion models for medical anomaly detection. Medical Image Computing and Computer-Assisted Intervention (MICCAI).