

QUANTUM DIFFUSION MODELS FOR MEDICAL IMAGING

Francesco Aldo Venturelli^{1,2}, Stefano Martina^{3,4}, Marco Parigi³, Filippo Caruso^{3,4}, Alba Cervera-Lierta²
and Miguel A. González Ballester^{1,2,5}.

¹*BCN Medtech, Universitat Pompeu Fabra, Barcelona, Spain,*

²*Barcelona Supercomputing Center, Barcelona, Spain,*

³*Department of Physics and Astronomy, University of Florence, Sesto Fiorentino, Florence, Italy,*

⁴*LENS - European Laboratory for Non-Linear Spectroscopy, University of Florence, Sesto Fiorentino, Florence, Italy,*

⁵*ICREA, Barcelona, Spain.*

INTRODUCTION

Diffusion Models (DMs) are a leading class of physics-inspired generative AI (GenAI) models. DMs obtained state-of-the-art results with Denoising Diffusion Probabilistic Models (DDPMs) [1], which progressively add noise to data and then reverse the process by learning the noise added at each forward time step (denoising) [2]. Despite their benefits, they require intensive computational resources, long training procedures and many diffusion steps to permit the model to learn how to map the features from pure noisy samples to original ones. In parallel, due to the rise of quantum computing and its potential to represent exponentially large systems with a polynomial number of qubits [3], there has been a growing interest in quantum GenAI that has yielded to the introduction of quantum diffusion models (QDMs) [4]. The majority of the works in the context of QDMs propose to implement the backward part with a parametrized quantum circuit that is iteratively trained to minimize an appropriate loss function [5, 6, 7]. Our formulation goes in the opposite direction: we execute the forward dynamics on a real noisy intermediate-scale quantum (NISQ) device [8] of IBM to take advantage of the intrinsic quantum noise of the machine, and use a classical learning model for data generation.

SUMMARY OF THE CONTRIBUTIONS

The state-of-the-art of hybrid and quantum algorithms applied to real-world data often faces challenges due to the impossibility for current NISQ devices to deal with large and resource-intensive data such as high-resolution images and volumes. Our work contributes in different ways to circumvent this issue: we demonstrate that currently available NISQ devices can be used to model the forward process of a DM without the need of error correcting codes, directly exploiting intrinsic quantum noise of the machine. Second, we demonstrate that our hybrid quantum forward dynamics can handle complex data samples such as RGB colored images, large-scale grayscale images, and 3D volumes, such as medical imaging datasets. Our main contributions are:

- The generation of real-world complex medical data of independent size, in particular grayscale, RGB 2D and 3D images (BloodMNIST [9], BraTS2020 [10, 11, 12] and FractureMNIST [9]);
- The introduction of a classical U-Net model with an appropriate time embedding and skip connections;
- The introduction of the optimization level, a native error mitigating scheme to reduce the aggressiveness of the noise of the machine;
- The evaluation of the generated samples through three distinct metrics widely used in the context of GenAI for medical imaging such as Kullback-Leibler divergence (KL-div.) [13], Frechét Inception Distance (FID) [14] and Structural Similarity Index Measure (SSIM) [15] to produce a fair comparison with the classical scenarios;
- The introduction of a weighted loss function that incorporates the KL-div. for accurate probability distribution generation;
- A comparison between two versions of the corresponding classical Markov chain with variable time scheduler function proposed in [16, 17].

This formulation extends the capabilities of previous attempts in [18]. A detailed description of our method can be found in our publication [19].

RESULTS

We execute multiple model trainings on the three datasets: a single class of blood cells RGB images of size 64×64 composed of 32 color intensity states belonging to BloodMNIST dataset, MRI grayscale tumoral brain images of BraTS2020 dataset of dimension 190×190 with 64 color intensity states and 3D grayscale volumes of 64 color intensity states of dimension $28 \times 28 \times 28$ extracted from FractureMNIST dataset.

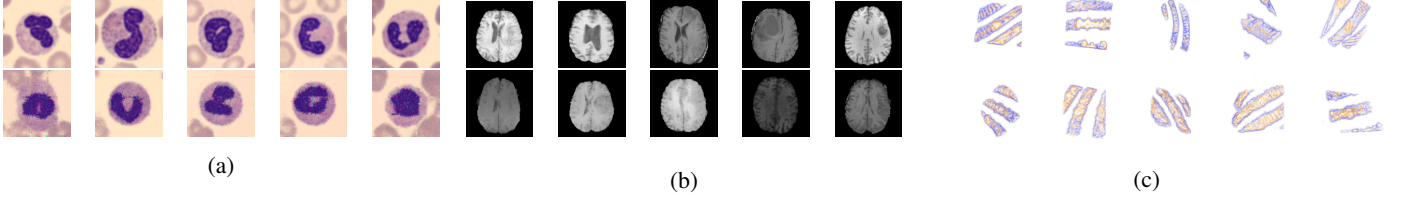


Figure 1. Qualitative comparison between original and generated samples for three medical imaging datasets. (a) BloodMNIST images, (b) BraTS2020 MRI scans, and (c) FractureMNIST3D volume renderings. The top rows correspond to five random original images and volumes, while the bottom rows are the generated ones.

We obtain comparable results with the classical case, achieving $FID \approx 243, 297, 156$ and $SSIM \approx 0.377, 0.676, 0.602$ for BloodMNIST, BraTS2020 and FractureMNIST respectively.

METHODS

Classical diffusion models in discrete state space

Classical DDPMs are represented by a forward Markov chain that progressively transforms an initial data sample into pure noise, and a reverse procedure that learns to remove the noise at each time step. In classical discrete-state space DMs [16], the data $\mathbf{x}$ are scalar random variables of $K \in \mathbb{N}$ categories. The forward dynamics is given by

$$q(\mathbf{x}_t | \mathbf{x}_{t-1}) = \text{Cat}(\mathbf{x}_t; p = \mathbf{x}_{t-1} \mathbf{Q}_t), \quad (1)$$

where $\text{Cat}(\mathbf{x}; p)$ is a categorical distribution over the one-hot row vector $\mathbf{x}$ with probabilities p , and the transition matrix $[\mathbf{Q}_t]_{ij} = q(x_t = j | x_{t-1} = i)$ tells us how likely a walker in the state $x_{t-1} = i$ has the chance to move to the site $x_t = j$ after a time-step. The prior $p(x_T)$ is a K -discrete Uniform distribution. In the backward part, deep learning models estimate

$$p_\theta(\mathbf{x}_{0:T}) = p(\mathbf{x}_T) \prod_{t=1}^T p_\theta(\mathbf{x}_{t-1} | \mathbf{x}_t). \quad (2)$$

Hybrid quantum diffusion models on NISQ device

In our case, we use a native quantum algorithm particularly suitable for being executed on IBM NISQ devices and that fits in the domain of the discrete state space DMs [16, 17]: the discrete-time quantum walk (DTQW) [20]. We use N_q qubits to embed $V = 2^{N_q}$ classical states into which the walker can jump to, and an extra qubit to reproduce the coin operator for a total number of $N_q + 1$ qubits. The topology on which the DTQW operates is represented by nearest-neighbor connections of degree 2 and 3, forming a ring-chain network. In this way, we converge to the Uniform distribution by exploiting the quantum noise of the machine and the periodic boundary conditions of the topology. Moreover, the DTQW needs few time steps since efficient implementations exists for currently available IBM quantum devices [21]. Mathematically, the DTQW state is represented as

$$\rho = |\Psi\rangle\langle\Psi| = |s_C\rangle\langle s_C| \otimes |\psi_P\rangle\langle\psi_P|, \quad (3)$$

the tensor product of the quantum state of the coin operator $|s\rangle_C \in \mathcal{H}_C$ and the one associated to the shift operator acting on the positions $V |\psi\rangle_P \in \mathcal{H}_P$, written in terms of the density operator to represent general quantum state

mixtures. To construct the dynamics of the walker, we apply a sequence of unitary operators that perform a single (forward/backward) step as formalized

$$\rho_{t+1} = \mathcal{E}(U\rho_t U^\dagger), \quad (4)$$

where $\mathcal{E}$ corresponds to a noisy set of channels. In our case, given a quantum walker initially at the position $|i_0\rangle$, which corresponds to the intensity value (and intensity value graph node) i_0 of the color state space, the new value after t evolution steps is obtained by

$$i_t \sim \text{Cat}(\mathbf{p} = \text{diag}(\rho_t)). \quad (5)$$

The sampling operation from the categorical distribution is analogous to a quantum measurement of the state, since it allows to decouple the quantum evolution of a generic-pixel DTQW from its application to the real images by estimating $\text{diag}(\rho_t)$ with a finite number of shots (8 192 in our case).

Classical backward model

We employ a U-Net with variable structure, depending on the data used to train it. For instance, given a batch of 2D data samples of discrete values

$$\mathbf{x} \in \{0, \dots, L-1\}^{C \times W \times H}, \quad (6)$$

where $L \in \{0, \dots, 2^{N_q} - 1\}$ is the number of discrete intensity values corresponding to the positional qubits N_q , C is the number of channels, W and H are the image dimensions; the model maps it into a higher dimensional vector that comprises a time dimension. We use a time embedding to condition the model on the discrete temporal index $t \in [1, T]$ to represent different distributions at different time steps. The structure of the U-Net comprises a sequence of 5 downsampling encoders consisting of a variable number of convolutional layers controlled by a tunable parameter h_L . The other part of the model contains 4 upsampling decoders that use transpose convolutions for upsampling and are connected to the same encoder index l via *skip connections*. At each decoder, the upsampled feature map is concatenated with the corresponding encoded feature map.

Metrics and loss

The used metrics are the KL div [13] divergence $D_{\text{KL}}(P \parallel Q)$

$$D_{\text{KL}}(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}, \quad (7)$$

which computes the divergence between the original pixel-intensity distribution of the entire dataset and the generated distribution. The FID [14]

$$\text{FID} = \|\mu_r - \mu_g\|_2^2 + \text{tr} \left(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{1/2} \right) \quad (8)$$

that is a commonly used statistical metric to assess the quality of images generated by models. μ_r and μ_g represent the mean feature vectors of real and generated samples extracted using a pre-trained *Inception-v3* network, while Σ_r and Σ_g denote their corresponding covariance matrices. Finally we calculate the SSIM [15] between the generated image and the real reference image to assess perceptual quality

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \quad (9)$$

where μ_x and μ_y denote the mean pixel intensities of the two image distributions, σ_x and σ_y their standard deviations, and σ_{xy} the covariance. The constants $c_1 = (k_1 L)^2$ and $c_2 = (k_2 L)^2$ stabilize the division, where L is the dynamic range of the pixel values (typically $2^{\text{bits per pixel}} - 1$), and the default parameters are $k_1 = 0.01$ and $k_2 = 0.03$. We decided to introduce a different kind of loss to induce a more accurate pixel-value distribution for a specific image class:

$$\mathcal{L} = \sum_{t=1}^T \left(-\alpha_t \sum_{x \in \mathcal{X}} y_x \log(p_x) - (1 - \alpha_t) \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)} \right). \quad (10)$$

The first term is the cross-entropy, also used for the training of the model with the other two datasets, and the second one is the KL-divergence between the original and target pixel-value distributions. $\alpha_t \in [0, 1]$ is a time-dependent factor that is proportionally scaled with t so that for higher values of t the model learns mainly to match the pixel-value distribution and for lower values the model is trained to reconstruct the image $\mathbf{x}_t$.

References

- [1] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pages 2256–2265. pmlr, 2015.
- [2] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851, 2020.
- [3] Andreas Bärttschi, Francesco Caravelli, Carleton Coffrin, Jonhas Colina, Stephan Eidenbenz, Abhijith Jayakumar, Scott Lawrence, Minseong Lee, Andrey Y Lokhov, Avnish Mishra, et al. Potential applications of quantum computing at los alamos national laboratory. *arXiv preprint arXiv:2406.06625*, 2024.
- [4] Marco Parigi, Stefano Martina, and Filippo Caruso. Quantum-noise-driven generative diffusion models. *Advanced Quantum Technologies*, 8(12):2300401, 2025.
- [5] Chuangtao Chen, Qinglin Zhao, MengChu Zhou, Zhimin He, Zhili Sun, and Haozhen Situ. Quantum generative diffusion model: A fully quantum-mechanical model for generating quantum state ensemble, 2024.
- [6] Francesca DeFalco, Andrea Ceschini, Alessandro Sebastianelli, Bertrand LeSaux, and Massimo Panella. Quantum latent diffusion models. *Quantum Machine Intelligence*, 6(2):85, Nov 2024.
- [7] Kübra Yeter-Aydeniz, Nora M Bauer, Pranay Jain, and Max Masnick. Hybrid quantum-classical latent diffusion models for medical image generation. *arXiv preprint arXiv:2508.09903*, 2025.
- [8] Kishor Bharti, Alba Cervera-Lierta, Thi Ha Kyaw, Tobias Haug, Sumner Alperin-Lea, Abhinav Anand, Matthias Degroote, Hermann Heimonen, Jakob S. Kottmann, Tim Menke, Wai-Keong Mok, Sukin Sim, Leong-Chuan Kwek, and Alán Aspuru-Guzik. Noisy intermediate-scale quantum algorithms. *Rev. Mod. Phys.*, 94:015004, Feb 2022.
- [9] Jiancheng Yang, Rui Shi, Donglai Wei, Zequan Liu, Lin Zhao, Bilian Ke, Hanspeter Pfister, and Bingbing Ni. Medmnist v2-a large-scale lightweight benchmark for 2d and 3d biomedical image classification. *Scientific Data*, 10(1):41, 2023.
- [10] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging*, 34(10):1993–2024, 2015.
- [11] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Maciej Rozycki, Justin S. Kirby, et al. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific Data*, 4:170117, 2017.
- [12] Spyridon Bakas, Mauricio Reyes, Andras Jakab, Stefan Bauer, Markus Rempfler, Alessandro Crimi, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. *arXiv preprint arXiv:1811.02629*, 2018.
- [13] Solomon Kullback and Richard A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22(1):79–86, 1951.
- [14] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems*, volume 30, 2017.
- [15] Zhou Wang, Eero P. Simoncelli, and Alan C. Bovik. Multiscale structural similarity for image quality assessment. In *Proceedings of the Thirty-Seventh Asilomar Conference on Signals, Systems & Computers*, volume 2, pages 1398–1402, November 2003.
- [16] Jacob Austin, Daniel D Johnson, Jonathan Ho, Daniel Tarlow, and Rianne Van Den Berg. Structured denoising diffusion models in discrete state-spaces. *Advances in neural information processing systems*, 34:17981–17993, 2021.
- [17] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International conference on machine learning*, pages 8162–8171. PMLR, 2021.
- [18] Marco Parigi, Stefano Martina, Francesco Aldo Venturelli, and Filippo Caruso. Physics-inspired generative ai models via real hardware-based noisy quantum diffusion. *arXiv preprint arXiv:2505.22193*, 2025.
- [19] Francesco Aldo Venturelli, Stefano Martina, Marco Parigi, Filippo Causo, Alba Cervera-Lierta, and Miguel Ángel González Ballester. Quantum diffusion models for medical imaging. *In preparation*.
- [20] Julia Kempe. Quantum random walks: an introductory overview. *Contemporary Physics*, 50(1):339–359, 2009.
- [21] Luca Razzoli, Gabriele Cenedese, Maria Bondani, and Giuliano Benenti. Efficient implementation of discrete-time quantum walks on quantum computers. *Entropy*, 26(4):313, 2024.