Gonçalo
Rodrigues Silva

Diagnóstico de Doenças de Pele com Treino Adversarial e Modelos de Difusão

Skin Disease Diagnosis with Adversarial Training and Diffusion Models

PROPOSTA DE TESE

palavras-chave

Classificação de Lesões de Pele, Modelos Generativos, Modelos de Difusão, Treino Adversarial e Aumento de Dados

resumo

Este documento apresenta o trabalho realizado como parte da unidade curricular "Preparação da Dissertação", com o objetivo de estabelecer as bases para o meu projeto de dissertação de mestrado. O estudo proposto visa investigar a aplicação do treino adversarial no contexto de modelos de difusão para melhorar a generalização de classificadores de aprendizagem profunda no diagnóstico de lesões de pele. Dada a importância de capturar detalhes subtis em imagens dermatoscópicas, o controle preciso proporcionado pelo processo generativo dos modelos de difusão oferece uma abordagem promissora para a aumento de dados. A primeira parte deste documento fornece uma breve introdução à área de pesquisa, destacando a motivação e os objetivos. A segunda parte apresenta uma revisão da literatura, explorando os fundamentos teóricos dos modelos de difusão aplicados à imagem médica. Isso inclui os Modelos Probabilísticos de Difusão com Remoção de Ruído (DDPMs), Modelagem Generativa Baseada em Pontuações e sua conexão com Equações Diferenciais Estocásticas. Com base na revisão da literatura, a parte final apresenta o trabalho preliminar realizado e o plano de trabalho proposto. O trabalho preliminar envolveu o desenvolvimento de um modelo de difusão para a geração de dígitos numéricos com base no conjunto de dados MNIST, descrevendo o pré-processamento dos dados, o treino do modelo e a avaliação de desempenho. O plano de trabalho delineia as principais tarefas a serem realizadas para alcançar os objetivos da pesquisa, incluindo o desenvolvimento e a implementação de uma estrutura de treino adversarial no contexto da classificação de lesões de pele.

keywords

Skin Lesion Classification, Generative Models, Diffusion Models, Adversarial Training, Data Augmentation

abstract

This document presents the work conducted as part of the 'Dissertation Preparation' curricular unit, aimed at establishing the basis for my master's thesis project. The proposed study aims to investigate the application of adversarial training within the diffusion model framework to enhance the generalizability of deep learning classifiers for skin lesion diagnosis. Given the importance of capturing subtle details in dermoscopic images, the precise control enabled by the generative process of diffusion models offers a promising approach for data augmentation. The first part of this document provides a brief introduction to the research area, outlining the motivation and objectives. The second part provides a literature review, exploring the theoretical foundations of diffusion models applied in medical imaging. This includes Denoising Diffusion Probabilistic Models (DDPMs), Score-Based Generative Modeling, and their connection to Stochastic Differential Equations. Based on the literature review, the final part presents the preliminary work carried out and the proposed work plan. The preliminary work involved the development of a diffusion model for genereting numerical digits based on the MNIST dataset, describing the data preprocessing, model training, and performance evaluation. The work plan outlines the key tasks to be undertaken to achieve the research objectives, including the development and implementation of an adversarial training framework within the context of skin lesion classification.

Contents

Contents i

List of Figures iii

List of Tables iv

Acronyms v

1 Introduction 1

1.1 Motivation . . . 1

1.2 Objectives . . . 2

1.3 Dissertation Outline . . . 2

2 State-of-the-Art 3

2.1 Medical Imaging and Deep learning . . . 3

2.2 Skin Lesion Data Scarcity and Data Augmentation . . . 4

2.2.1 The Problem of Data Scarcity . . . 4

2.2.2 The Role of Data Augmentation and Generative Models . . . 6

2.3 Diffusion Models Introduction . . . 9

2.3.1 Theoretical Foundation . . . 9

2.3.2 Modern Advancements . . . 10

2.4 Denoising Diffusion Probabilistic Models (DDPM) . . . 10

2.4.1 Forward Diffusion Process . . . 11

2.4.2 Reverse Denoising Process . . . 11

2.4.3 Model Simplifications and Training Objective . . . 12

2.5 Noise-Conditioned Score Networks (NCSN) . . . 12

2.5.1 Theory and Methodology . . . 12

2.5.2 Limitations and Challenges . . . 14

2.5.3 Relationship to Diffusion Models . . . 14

2.6 Stochastic Differential Equations (SDE) . . . 15

2.6.1 Forward and Reverse Processes . . . 15
2.6.2 Numerical Solutions and Sampling . . . 16
2.6.3 Advantages and Challenges . . . 17
2.7 The Importance of Iterative Processes in Diffusion Models . . . 17
2.8 Performance Metrics for Diffusion Models . . . 18
2.8.1 Inception Score (IS) . . . 19
2.8.2 Fréchet Inception Distance (FID) . . . 21
2.9 Summary . . . 22

3 Methodology and Work Plan **23**
3.1 Work Done . . . 23
3.2 Work Plan . . . 26

References **28**

A Additional Content **31**
A.1 Diffusion Models in Medical Imaging . . . 31
A.2 Comparison of DDPMs, NCSNs, and SDEs . . . 32

List of Figures

1.1 Examples of synthetic melanoma images generated by a StyleGAN2-Ada model with unrealistic details 1

2.1 Evolution of deep learning algorithm types in medical imaging (2012-2022) 4
2.2 ISIC Challenge - Phase 1: Lesion Segmentation 5
2.3 ISIC Challenge - Phase 2: Dermoscopic Feature Detection 6
2.4 ISIC Challenge - Phase 3: Disease Classification 6
2.5 Overview of the Generative Adversarial Networks (GAN) architecture 7
2.6 Graphical model of the Denoising Diffusion Probabilistic Model 10
2.7 Annealed Langevin Dynamics intermediate Samples 14
2.8 Forward and Reverse Processes in Stochastic Differential Equations 16

3.1 MNIST dataset images at different stages of the diffusion process. 25
3.2 Gantt chart showing the work done and work plan for the following months. 27

List of Tables

2.1 Summary of GANs used for data augmentation in medical imaging. 8

A.1 Summary of studies on diffusion models in medical imaging, highlighting key tasks such as image synthesis, classification, and data augmentation. 31

A.2 Comparison of DDPMs, NCSNs, and SDEs 32

Acronyms

AI	Artificial Intelligence
CNN	Convolutional Neural Network
DL	Deep Learning
GAN	Generative Adversarial Networks
ISIC	International Skin Imaging Collaboration
DDPM	Denoising Diffusion Probabilistic Models
NCSN	Noise Conditional Score Networks
SDE	Stochastic Differential Equation
FID	Fréchet Inception Distance
IS	Inception Score
KL	Kullback-Leibler
MCMC	Markov Chain Monte Carlo
ODE	Ordinary Differential Equation
EM	Euler-Maruyama
PC	Predictor-Corrector
CIFAR	Canadian Institute For Advanced Research
ALD	Annealed Langevin Dynamics
MSE	Mean Squared Error
VLB	Variational Lower Bound

CHAPTER 1

Introduction

1.1 Motivation

Skin lesions are a growing health concern requiring early and accurate diagnosis for successful treatment. Dermatologists rely on visual examination and biopsies, but subjectivity in visual assessment and limited access to specialists can create difficulties for early detection. Deep learning has achieved remarkable success in medical image analysis, with Convolutional Neural Network (CNN) demonstrating promising results in skin lesion classification [1], [2]. However, a significant challenge in training these models lies in the limited availability of high-quality, labeled medical data. Generative models can address this problem by creating synthetic images that augment the real dataset, enhancing the model's ability to generalize predictions. While GAN has been a popular choice for generating images [3], diffusion models [4] offer advantages in the context of skin lesion diagnosis. First, by learning the underlying distribution of real data, these models often provide more realistic and diverse images compared to GANs, who suffer from artifacts and inconsistencies (see Figure 1.1). Second, GANs involve a complex training process where a generator and a discriminator compete with each other, making them unstable and prone to collapse. Conversely, diffusion models follow a well-defined training procedure with a clear objective, leading to more stable and reliable training.

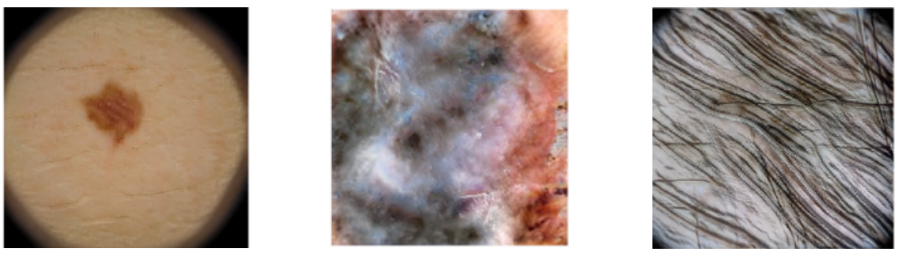


Figure 1.1: Examples of synthetic melanoma images generated by a StyleGAN2-Ada model with unrealistic details [3]: (a) checkerboard pattern, (b) mode collapse and (c) lack of skin lesion.

1.2 Objectives

This dissertation proposes a structured approach to investigate the potential of adversarial training with diffusion models in order to address data scarcity and improve the generalizability of deep learning classifiers. In the critical task of skin lesion classification, where subtle details can be crucial, the higher fidelity of diffusion-generated images can be advantageous [5], [6]. Additionally, diffusion models allow for more control over the generation process by guiding the model toward specific image features. This can be particularly beneficial in skin lesion diagnosis, where focusing on generating synthetic images that mimic challenging or rare lesion types can improve the classifier's ability to handle those cases.

Bearing this in mind, the research will focus on the following specific objectives:

1. Train a deep CNN classifier for skin lesion classification using a dataset of dermatoscopy images.
2. Develop a diffusion model to generate synthetic skin lesion images that are similar to real data
3. Implement an adversarial (iterative) training process where the diffusion model is guided to generate synthetic data that targets challenging cases identified on a held-out test set.
4. Evaluate the impact of the generated synthetic data on the classifier's performance, comparing its accuracy, sensitivity, and specificity on the test set before and after adversarial training.

1.3 Dissertation Outline

The rest of this dissertation is organized as follows:

- **Chapter 2** discusses the key concepts and the related work in the field of diffusion models that are relevant to the objectives of the dissertation.
- **Chapter 3** outlines the planned work for the dissertation, detailing the methodologies to be employed for their achievement.

CHAPTER 2

State-of-the-Art

This chapter introduces related work, mainly focusing on the research contributions in the area of Diffusion models and their applications in the field of medical imaging. It will introduce Diffusion models, how they work, the relevance for the theme, and finally, will evaluate metrics for Diffusion models.

2.1 Medical Imaging and Deep learning

In recent years, thanks to technology improvements, the healthcare industry has experienced a considerable transition. As noted in Choy *et al.* [1], the exponential growth of medical data combined with rising computing power has been a major force behind this transformation. This combination of circumstances has made it possible to transform medical imaging through the use of Artificial Intelligence (AI) techniques, particularly Deep Learning (DL).

DL, inspired by the neural networks of the human brain, has become a potent instrument for analyzing complex medical images. DL models outperform conventional machine learning methods by using multi-layered neural networks to extract intricate patterns and features from massive datasets.

An important turning point in DL history was the release of AlexNet [7] in 2012, which showed how these models could perform at the cutting edge of image recognition tasks [1]. Since then, CNN have become increasingly dominant in medical imaging applications, as demonstrated by a comprehensive systematic review of 13,857 references by Choy *et al.* [1]. Their analysis revealed that by 2021, CNNs represented approximately 85% of all deep learning applications in the field. (Figure 2.1).

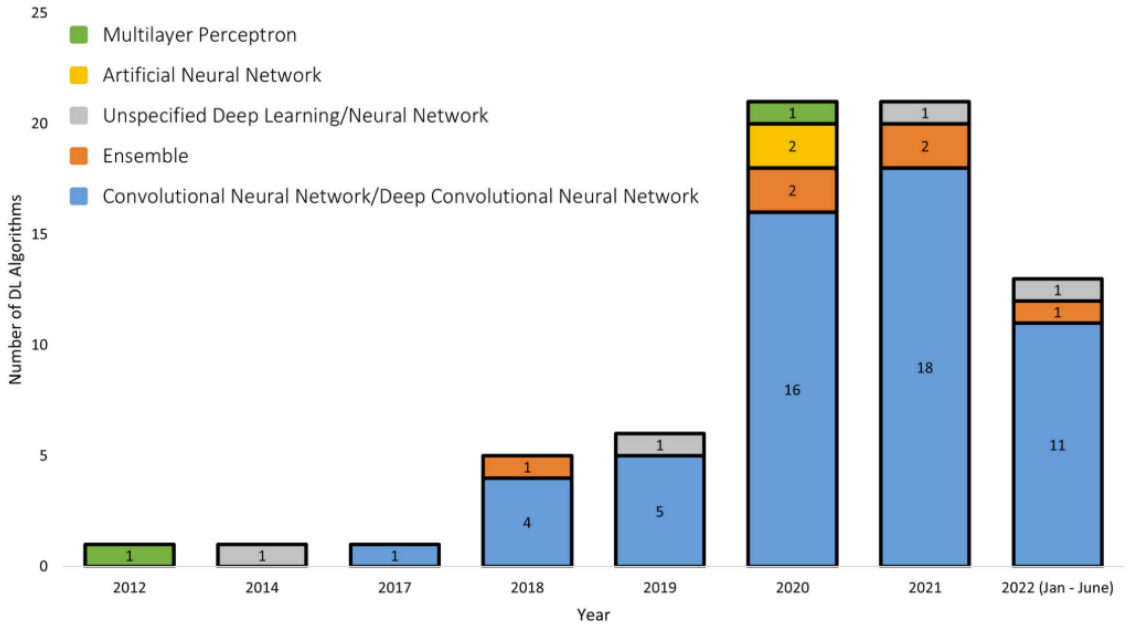


Figure 2.1: Evolution of deep learning algorithm types in medical imaging (2012-2022), showing the increasing dominance of CNNs. Taken from Kazerouni *et al.* [4]

.

In visually demanding medical fields including radiology, ophthalmology, pathology, and dermatology, the effects of DL are more noticeable. These domains are perfect for DL-based solutions because of the volume of image data and the urgent requirement for precise interpretation. DL has been shown in numerous studies to have the ability to improve patient outcomes by increasing diagnosis accuracy and streamlining clinical operations.

DL technology has the potential to revolutionize medical imaging as it develops further. We may foresee the creation of increasingly complex diagnostic tools by utilizing AI, which will eventually result in earlier disease identification, more accurate treatment planning, and improved patient care.

2.2 Skin Lesion Data Scarcity and Data Augmentation

2.2.1 The Problem of Data Scarcity

The diagnosis and monitoring of skin lesions present unique challenges in medical imaging, particularly when addressing the issue of data scarcity. As highlighted in the systematic review by Choy *et al.* [1] and the findings of Codella *et al.* [8], the lack of diverse, high-quality, and well-annotated datasets is a significant barrier to progress in skin lesion analysis.

Deep learning models are inherently data-hungry, requiring extensive datasets to combat overfitting and achieve robust generalization. The scarcity of rich, representative datasets directly impacts the performance and reliability of these models, especially in real-world applications. This challenge is particularly pressing in dermatology, where datasets must cover a wide variety of skin types, ages, and ethnicities to ensure fair and unbiased diagnostic capabilities. However, current datasets often exhibit significant demographic biases, with

limited representation of underrepresented groups. For instance, only 19% of dermatology studies utilizing deep learning explicitly mention participant ethnicity and/or Fitzpatrick skin type.

The factors contributing to data scarcity in dermatology include strict privacy regulations, the high cost and time required for expert annotation, and the natural imbalance in the prevalence of certain skin disorders. Without datasets that are both large enough to meet the needs of deep learning models and representative enough to generalize effectively, the development of robust diagnostic systems remains a significant challenge.

To try to mitigate this problem, there is an organization that has been in the front of addressing this data scarcity challenges, The International Skin Imaging Collaboration (ISIC). ISIC has combined data from top clinical centers across the globe, since its founding, to build one of the biggest publicly available archives of dermoscopy images. By 2017, this organization had retrieved over 20,000 images from various devices used at different clinical centers, as reported by Codella *et al.* [8].

ISIC's dedication to advancing automated skin lesion diagnosis has led to the creation of benchmark challenges, the first of which appeared in 2016, dividing the problem into three phases: **lesion segmentation** (Figure 2.2), **dermoscopic feature detection** (Figure 2.3), and **disease classification** (Figure 2.4). A similar challenge was held in 2017, with the same phases, but in comparison to the previous year, the scale of the challenge enlarged significantly, with a total of 593 registrations and 46 finalized submissions.

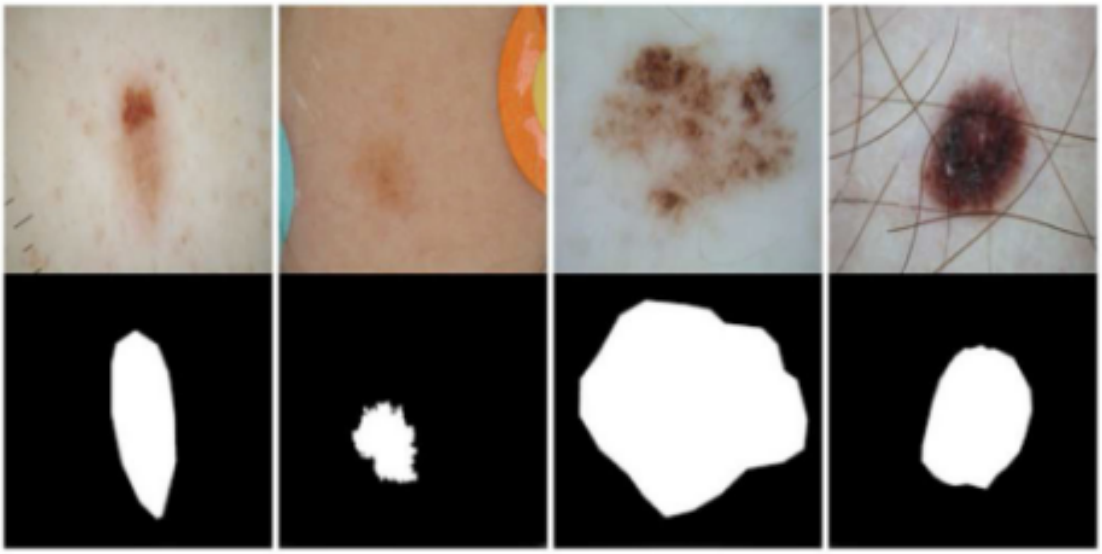


Figure 2.2: "Phase 1: Lesion Segmentation", Top: Original Images, Bottom: Segmentation Masks. Taken from Codella *et al.* [8].

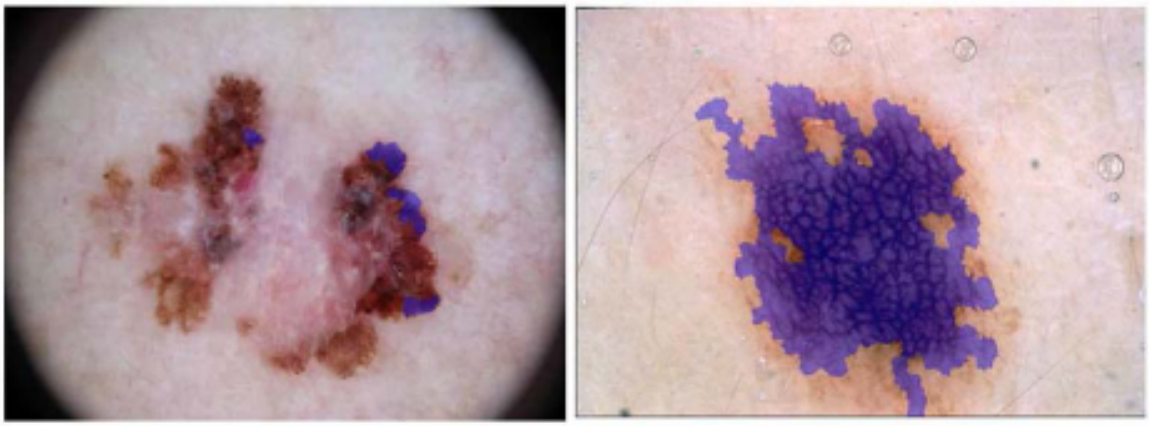


Figure 2.3: "Phase 2: Dermoscopic Feature Detection", Ground truth labels highlighted in purple. Left: Streaks. Right: Pigment Network. Taken from Codella *et al.* [8].

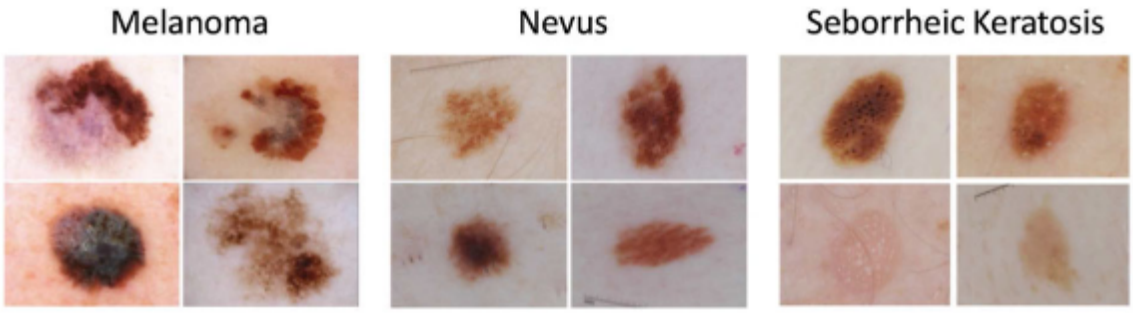


Figure 2.4: "Phase 3: Disease Classification", Ground truth labels written above. Taken from Codella *et al.* [8].

Currently, ISIC keeps organizing challenges and competitions yearly to further grow its archive, playing a crucial role in advancing the field by providing standardized datasets and evaluation metrics that allow researchers to compare their approaches effectively. Collaboration can be called the basis of the development and validation of automated skin lesion analysis algorithms.

Despite these efforts, the scarcity of high-quality, diverse, and well-annotated datasets remains a significant challenge, so, through the years multiple techniques of data augmentation and different types of generative models have been improved and discovered to help mitigate the problem even more.

2.2.2 The Role of Data Augmentation and Generative Models

Data augmentation is a widely used technique in deep learning to artificially expand the size and diversity of training datasets without explicitly collecting new data. Traditional data augmentation methods rely on applying various transformations to the original images, such as geometric transformations, color and intensity adjustments, and adding noise or filters [9], [10].

Geometric transformations that are commonly used include cutting certain parts of the image, flipping it horizontally or vertically, rotating it by a specific angle or scaling it up or down. These are used to improve the model's performance while dealing with invariance in the data. Color transitions can mimic changes in lighting and stains by varying brightness, contrast, saturation, and hue. Applying smoothing or sharpening filters and adding random noise can make the model more resilient to variations in image quality that are frequently seen in real-world data [11].

While traditional augmentation techniques are simple to implement and computationally inexpensive, they have some limitations. The transformations are hand-engineered based on domain knowledge and may not always capture the full spectrum of possible variations. Some transformations may also produce unrealistic images that do not preserve the key characteristics of the original data, potentially misleading the model. Despite these drawbacks, traditional augmentation remains a powerful tool to improve model generalization, especially when training data is scarce [9], [11].

Generative models, on the other hand, offer a more sophisticated approach to data augmentation by learning the underlying distribution of the training data and generating new samples that are indistinguishable from the real data. These models can generate high-quality images that capture the complexity and diversity of the original data, providing a more effective way to expand the training dataset. One of the most popular generative models, presented by Goodfellow *et al.* [12], is the GAN, which can be used to create images from a training dataset.

A GAN consists of two neural networks, a generator and a discriminator, that are trained simultaneously in a two-player adversarial game. The generator attempts to produce synthetic images that are indistinguishable from real images, while the discriminator tries to correctly classify images as real (from the training data) or fake (from the generator), as seen in Figure 2.5. Through this competitive optimization, the generator learns to capture the underlying data distribution and can be used to generate novel realistic images [3], [13].

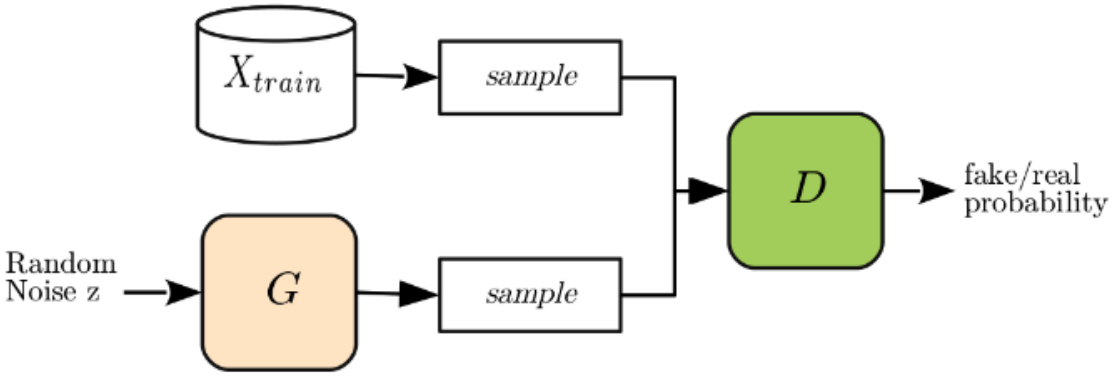


Figure 2.5: Overview of the GAN architecture. Taken from Hayes *et al.* [14].

Various types of GANs have been explored for data augmentation in medical imaging (See

Table 2.1), such as Deep Convolutional GANs (DCGANs) for generating skin lesion images [15], Cycle-Consistent GANs (CycleGANs) for cross-modality synthesis in MRI and CT [16], and Progressive Growing GANs (PGGANs) for high-resolution mammogram synthesis [17]. Compared to traditional augmentation, GANs can learn more complex transformations and produce highly realistic images that preserve anatomical and pathological characteristics. GAN-based augmentation has shown success in improving model performance for various medical image analysis tasks, including skin lesion classification [18], brain tumor segmentation [19], and lung nodule detection [20].

Table 2.1: Summary of GANs used for data augmentation in medical imaging.

GAN Type	**Application**	**Reference**
Deep Convolutional GANs (DCGANs)	Generating skin lesion images	[15]
Cycle-Consistent GANs (CycleGANs)	Cross-modality synthesis in MRI and CT	[16]
Progressive Growing GANs (PGGANs)	High-resolution mammogram synthesis	[17]
Applications of GAN-based Augmentation		
Improving model performance for skin lesion classification		[18]
Improving model performance for brain tumor segmentation		[19]
Improving model performance for lung nodule detection		[20]

Impact on Model Performance

Data augmentation improves deep learning models in medical imaging, being utilized to increase the size and diversity of training datasets, diversity being important to make the model generalize better and reduce overfitting [9]. This is crucial for medical datasets, which are often small due to the difficulty of obtaining labeled data.

Augmentation also addresses class imbalance. Rare conditions or underrepresented groups can cause uneven datasets. Oversampling through augmentation balances the data and helps the model learn unbiased features [21]. It enhances training stability, too. Diverse data helps the model avoid poor optimization paths, leading to faster and better convergence. It is effective not only during training but also when applied to test data. The combination of geometric and color transformations often yields superior results, even surpassing benchmarks without requiring additional data. Further optimization, such as hyperparameter tuning, can provide additional improvements. Performance metrics also improve with augmentation. Geometric and color transformations boosted skin lesion classification accuracy in multiple CNN networks when compared to using the original data alone [11].

Current Limitations of Existing Methods

Data augmentation techniques show promise but, like almost everything, have its limitations. The main problem with the use of GANs is its unstable training. The generator and discriminator can be sensitive to hyperparameters and may experience mode collapse, where

only limited samples are produced [22]. This reduces augmentation effectiveness and needs careful tuning.

Another challenge is the quality control of the generated images. Even with well-trained GANs, there is a risk of producing unrealistic or artifactual samples that may not adequately capture the characteristics of the original data [23].

Validating augmentation methods is also complex and expensive. It involves training and testing multiple models with different settings, which uses significant resources, especially for large datasets and deep models. These challenges call for better and more efficient methods. New approaches like diffusion models show potential for improving stability, quality, and validation in medical imaging.

2.3 Diffusion Models Introduction

Diffusion models are a type of generative models that have been gaining popularity in the field of deep learning because of their ability to generate high-quality images and have a more stable training, unlike their GAN counterparts.

Modern diffusion are built around the foundation established by Sohl-Dickstein *et al.* [24], who introduced an approach of probabilistic modeling inspired by non-equilibrium thermodynamics. This paper was groundbreaking because it was responsible for the theoretical foundation of what would later allow the development of powerful diffusion-based generative models. The central innovation of this paper was a forward diffusion process, or noising process, that gradually adds noise to destroy a data structure, paired with a learned reverse process that removes the noise to reconstruct the data. This approach balanced flexibility and tractability, enabling expressive models with exact sampling capabilities.

2.3.1 Theoretical Foundation

Adding to what was already said, the idea behind the diffusion process is to start with a sample from a simple Gaussian noise distribution and iteratively refine it to match the target data distribution. This is achieved by learning a series of denoising steps that reverse a predefined noising process.

The noising process gradually corrupts the data by adding Gaussian noise over a series of discrete steps, indexed by a "timestamp" or time step t. This timestamp serves as an index that indicates the noise level at a particular step in the sequence, ranging from the maximum noise (at $t = T$, corresponding to pure Gaussian noise) to no noise (at $t = 0$, corresponding to clean data). During training, the model learns to predict and remove the noise present at each timestamp, conditioned on the noisy input and the corresponding t.

By applying the denoising steps in reverse order—starting from pure noise and moving back to $t = 0$—the model generates samples that resemble the training data. The timestamp thus plays a critical role in controlling the progression of the denoising process and guiding the model at each step toward reconstructing realistic samples. The mathematical formulation of the diffusion process applied to the diffusion models is further explained in section 2.4.

2.3.2 Modern Advancements

Building on this theoretical foundation, later research has significantly enhanced diffusion models. Current advancements have optimized their efficiency, scalability and practical use cases.

Some of these improvements can be grouped into 3 main approaches, these being discussed in the following sections:

- **Denoising Diffusion Probabilistic Models (DDPM)**: Refining the reverse diffusion process for high-quality image synthesis while optimizing computational requirements.
- **Noise Conditional Score Networks (NCSN)**: Utilizing score-matching techniques to directly model the gradient of the data distribution, enabling precise control over generated samples.
- **Stochastic Differential Equation (SDE)**: Generalizing the diffusion framework with continuous-time formulations for more flexible and scalable modeling.

In Table A.1, we summarize several studies on diffusion models and their applications in the medical imaging domain, providing a concise overview of the advancements in diffusion-based approaches, highlighting their contributions to specific tasks such as image synthesis, classification, and data augmentation. This summary underscores the versatility and potential of diffusion models in addressing challenges within medical imaging.

2.4 Denoising Diffusion Probabilistic Models (DDPM)

By introducing specific optimizations into the 2-way pipeline, Ho *et al.* [25] was able to achieve state-of-the-art performance on many tasks, creating the so called Denoising Diffusion Probabilistic Models (DDPM)S.This research was particularly important because it revived diffusion models, otherwise an abandoned idea, which were a type of model that was theoretically appealing but had been largely ignored due to practical limitations, but this new approach transformed it into a state-of-the-art generative model.

During the explanation of the Forward and Reverse processes, it will be necessary to have a general understanding of the process as a whole and as such, having the Figure 2.6 in mind during the next steps is crucial, as it creates a visual intuition on what is going to be explained and facilitates further understanding.

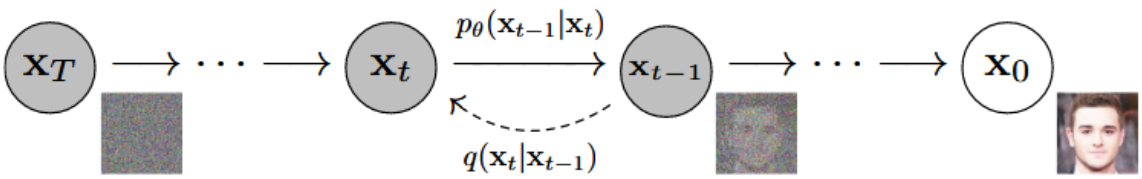


Figure 2.6: Graphical model of the Denoising Diffusion Probabilistic Model. Taken from [25]

.

The figure above shows both the forward and reverse processes of the diffusion model. The forward process, here represented as the function q, starts with a clean data point x_0

$(t = 0)$, while the reverse process, here represented as the function p, starts with a noisy data point x_T $(t = T)$.

Having now presented the basis of the process and taking into account the theoretical foundation, we will now recap and dive into the Forward Diffusion process, and later on the Reverse Denoising process.

2.4.1 Forward Diffusion Process

The forward diffusion process systematically corrupts training data by introducing Gaussian noise across multiple timesteps, gradually transforming a data point x_0 into approximately pure Gaussian noise after T timesteps. This process follows a Markov chain with Gaussian transitions, where at each step t, a controlled amount of noise is added according to a variance schedule β_t. This schedule, typically starting with small values and increasing over time, determines how much noise is added at each step.

The process can be mathematically expressed as:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t\mathbf{I}) \tag{2.1}$$

Here, x_t represents the data at timestep t, following a Gaussian distribution with mean and variance determined by the cumulative noise added to the original data. The mean of the distribution is determined by scaling the data from the previous timestep, x_{t-1}, by $\sqrt{1-\beta_t}$, while the variance is given by $\beta_t\mathbf{I}$, with $\mathbf{I}$ denoting the identity matrix.

Through the study of Kingma and Welling [26] on Auto-Encoding Variational Bayes it was learned that the variances β_t can be learned by reparameterization. Using this reparameterization, It's possible to sample x_t directly for any t without the need to sample from the previous time step x_{t-1}.The following equation shows how the forward diffusion process can be expressed:

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)\mathbf{I}) \tag{2.2}$$

where $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$. This provides a direct relationship between the data at any timestep t and the original data x_0, enabling efficient computation and modeling in the diffusion process.

2.4.2 Reverse Denoising Process

The reverse denoising process aims to iteratively remove the noise introduced in the forward diffusion process, generating clean samples from Gaussian noise. This is modeled as a learned Markov chain with Gaussian transitions:

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)) \tag{2.3}$$

Here, $\mu_\theta(x_t, t)$ and $\Sigma_\theta(x_t, t)$ represent the mean and covariance matrix, respectively, both parameterized by a neural network θ and conditioned on the timestep t. The generation process starts with Gaussian noise at timestep T and progressively refines it through this learned reverse process until a clean sample emerges at $t = 0$.

2.4.3 Model Simplifications and Training Objective

To enhance computational efficiency and stability, several practical simplifications have been applied to the reverse process. A notable simplification involves fixing $\Sigma_\theta(x_t, t)$ as a time-dependent constant $\sigma_t^2 I$, rather than learning it as a variable parameter. Additionally, two parameterizations for $\mu_\theta(x_t, t)$ have been explored:

$$\mu_\theta(x_t, t) = \begin{cases} \text{Direct prediction of } x_{t-1} \\ \text{Prediction of noise component } \epsilon \end{cases} \tag{2.4}$$

Empirical evidence suggests that predicting the noise component ϵ instead of directly predicting x_{t-1} leads to superior performance. This approach allows the model to focus specifically on learning the noise characteristics of the diffusion process.

The training objective for diffusion models typically involves minimizing the difference between the predicted and actual noise components across all timesteps. This is often formulated as a Mean Squared Error (MSE) loss:

$$L_{\text{simple}} = \mathbb{E}_{t,x_0,\epsilon}\left[\|\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, t)\|^2\right] \tag{2.5}$$

This simplified loss function provides several advantages:

- Stable training through prediction of noise component ϵ,
- Simplified implementation without multiple loss terms
- Better empirical performance compared to alternative parameterizations

Ho *et al.* [25] demonstrated, as a result of these simplifications and optimizations, that the use of diffusion models can generate, as proposed, high quality samples.

2.5 Noise-Conditioned Score Networks (NCSN)

Noise-Conditioned Score Networks (NCSN), introduced by Song and Ermon [27], are an important advancement in score-based generative modeling. The paper concluded that the use of NCSNs can generate high-quality samples when compared to other generative models, such as GANs, without the need of adversarial training. However, the authors also noted that the current training methods were very unstable under some settings and were also limited to low resolution images. So, the authors proposed state-of-the-art methods and solution on their follow-up paper «Improved Techniques for Training Score-Based Generative Models» [28]. In the following subsection 2.5.1 we will present the main concepts and improvements made by the authors.

2.5.1 Theory and Methodology

These models aim to estimate the gradients of the logarithmic data density, referred to as the 'score', while applying varying levels of noise to the data. By learning this information across different noise levels, NCSNs can generate high-quality samples from complex data distributions. [27] [28].

Why Noise Levels are used?

In real-world data, information normally resides on low-dimensional structures (or manifolds) embedded in high-dimensional spaces. In these situations scores can be poorly defined, principally in areas where data points are limited. However, by perturbing the data with Gaussian noise at varying levels it spreads the data across the space, making it easier to estimate the score function [28].

For example, if we take a clean data point and add Gaussian noise to it with standard deviation σ, we get the perturbed version of the data. This relationship can be expressed as:

$$x' \sim \mathcal{N}(x, \sigma^2 I) \tag{2.6}$$

where x is the clean data point, x' is the perturbed data point, and I is the identity matrix.

Score Function and Training Objective

The score function of a probability density $p(x)$ is defined as:

$$\nabla_x \log p(x), \tag{2.7}$$

which points in the direction of the steepest increase in the log-probability density. In NCSNs, a neural network $s_\theta(x, \sigma)$ is trained to approximate this score for data perturbed with Gaussian noise of standard deviation σ. The training objective is based on *denoising score matching*, which minimizes the discrepancy between the predicted score and the true score of the perturbed data distribution:

$$\mathcal{L}(\theta) = \frac{1}{2L} \sum_{i=1}^{L} \mathbb{E}_{p(x)} \mathbb{E}_{p_{\sigma_i}(\tilde{x}|x)} \Big[\big\| s_\theta(\tilde{x}, \sigma_i) + \frac{\tilde{x} - x}{\sigma_i^2} \big\|^2 \Big], \tag{2.8}$$

where L is the number of noise levels, and $p_{\sigma_i}(\tilde{x}|x)$ is the distribution of data x perturbed by Gaussian noise with variance σ_i^2 [29].

The primary innovation of NCSNs is the ability to estimate scores across various noise scales by training a single scoring network conditioned on the noise level σ. This eliminates the need for separate models for each noise level, significantly simplifying the framework.

Annealed Langevin Dynamics

Once the model is trained, the sample generation in NCSNs is done using Annealed Langevin Dynamics (ALD), an iterative process rooted in Markov Chain Monte Carlo (MCMC) methods [30]. Starting from an initial random sample, the algorithm iteratively refines the sample using the estimated score:

$$\tilde{x}_{t+1} = \tilde{x}_t + \alpha s_\theta(\tilde{x}_t, \sigma_i) + \sqrt{2\alpha}\epsilon_t, \tag{2.9}$$

where α is the step size, $\epsilon_t \sim \mathcal{N}(0, I)$ represents Gaussian noise, and σ_i is the noise level at step i. The noise levels are gradually reduced (annealed) across iterations, allowing the

process to refine the sample progressively and converge to the target distribution. The final step often includes a denoising operation to remove residual noise, further enhancing sample quality [27] [28]. In Figure 2.7, we can see a visual representation of the Annealed Langevin Dynamics intermediate steps.

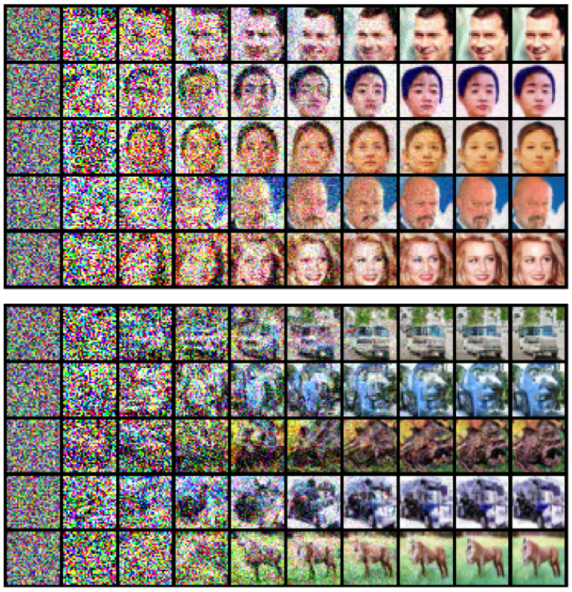


Figure 2.7: Annealed Langevin Dynamics intermediate Samples. Taken from [28].

2.5.2 Limitations and Challenges

Despite the great results achieved in [27] and [28], NCSNs face several limitations or challenges:

- **Computational Complexity:** The iterative sampling process via Langevin dynamics is computationally expensive, especially for high-resolution data [28].
- **Hyperparameter Sensitivity:** The performance of NCSNs is sensitive to the choice of noise levels, step sizes, and other hyperparameters.
- **Convergence Issues:** Ensuring convergence during sampling can be challenging, particularly in high-dimensional spaces.

2.5.3 Relationship to Diffusion Models

Conceptually, diffusion models—especially score-based diffusion models—and NCSNs are comparable. In both methods, data is perturbed with noise and then iteratively denoised to produce samples. The primary distinction is that, whereas diffusion models generally characterize the process as forward and reverse diffusion through time, NCSNs specifically highlight learning score functions at diverse noise levels. These methods have been brought together by the creation of Stochastic Differential Equations (SDE) frameworks, which emphasize their common mathematical fundamentals and bring us to the following section.

2.6 Stochastic Differential Equations (SDE)

Stochastic Differential Equations (SDEs) provide a continuous-time framework for generative modeling, representing a significant advancement in the field of diffusion models. Introduced by Song et al. [31], SDEs unify and extend earlier approaches, such as Denoising Diffusion Probabilistic Models (DDPM) [24], [25] and Noise-Conditioned Score Networks (NCSN) [27], into a broader mathematical framework.

The core idea behind SDEs in generative modeling is to describe the forward diffusion process (data $\rightarrow$ noise) and reverse process (noise $\rightarrow$ data) as solutions to stochastic differential equations. This framework enables the use of continuous-time dynamics to manipulate probability distributions in a principled and flexible manner.

2.6.1 Forward and Reverse Processes

The forward SDE progressively adds noise to data points x according to the following equation:

$$\mathrm{d}x = f(x,t)\mathrm{d}t + g(t)\mathrm{d}w, \tag{2.10}$$

where:

- $f(x,t)$: Drift coefficient that determines the deterministic component of the dynamics.
- $g(t)$: Diffusion coefficient controlling the stochastic component.
- $\mathrm{d}w$: Wiener process (standard Brownian motion).

This forward process transforms the data distribution $p_0(x)$ into a simple prior distribution $p_T(x)$, often a Gaussian, over a continuous time interval $t \in [0,T]$. The reverse process, derived using results from Anderson [32], is given by:

$$\mathrm{d}x = [f(x,t) - g(t)^2 \nabla_x \log p_t(x)]\mathrm{d}t + g(t)\mathrm{d}\bar{w}, \tag{2.11}$$

where $\mathrm{d}\bar{w}$ is a Wiener process [33] under time reversal. The term $\nabla_x \log p_t(x)$, referred to as the score function (see subsection 2.5.1), guides the reverse process by removing noise and reconstructing data.

All the previous steps can be visualized in Figure 2.8.

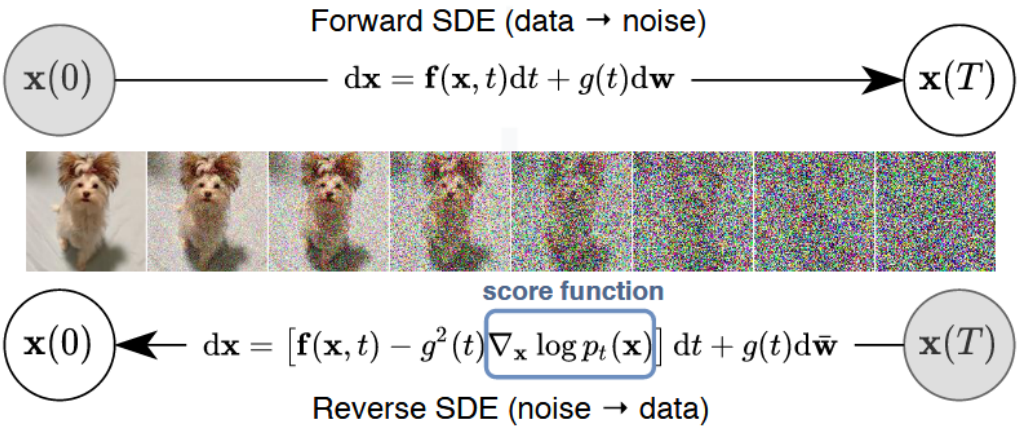


Figure 2.8: Forward and Reverse Processes in Stochastic Differential Equations. Taken from [31].

2.6.2 Numerical Solutions and Sampling

Normally, generating samples from an SDE involves simulating the reverse process, these samples being generated from a target distribution. This simulation can be done using numerical techniques, the most common sampling techniques being:

- Euler-Maruyama (EM) Method
- Predictor-Corrector (PC) Sampling
- Probability Flow Ordinary Differential Equation (ODE)

These methods ensure flexibility, accuracy, and efficiency in solving reverse-time dynamics, and will be briefly covered in the following sections without going into too much detail.

Euler-Maruyama (EM) Method

The Euler-Maruyama method is a numerical scheme that approximates the solution of a stochastic differential equation by discretizing time. Instead of solving the continuous-time reverse SDE directly, the method computes an approximate solution iteratively over small timesteps Δt. For the reverse-time SDE, this is expressed as:

$$\tilde{x}_{t+1} = \tilde{x}_t + \Delta t \cdot [f(\tilde{x}_t, t) - g(t)^2 \nabla_x \log p_t(\tilde{x}_t)] + g(t)\sqrt{\Delta t} \cdot \epsilon, \tag{2.12}$$

where $\epsilon \sim \mathcal{N}(0, I)$ is standard Gaussian noise.

While the method is straightforward and computationally efficient, its accuracy depends on the choice of Δt. Smaller time steps improve precision but increase computational cost, but making Δt too large may lead to accumulation of errors [31], [34].

Predictor-Corrector (PC) Sampling

The Predictor-Corrector framework, first presented in the context of score-based generative models in [31], enhances the accuracy of sampling by combining deterministic and stochastic updates:

1. **Prediction:** A deterministic step is performed using the Euler-Maruyama update to move the sample closer to the target distribution.
2. **Correction:** A stochastic step, such as Langevin dynamics, refines the sample by ensuring alignment with the score function $\nabla_x \log p_t(x)$.

This hybrid approach balances the trade-off between computational cost and sample quality [31].

Probability Flow ODE

An alternative to stochastic sampling involves transforming the reverse-time SDE into a deterministic Ordinary Differential Equation (ODE):

$$\mathrm{d}x = \left[f(x,t) - \frac{1}{2} g(t)^2 \nabla_x \log p_t(x) \right] \mathrm{d}t. \tag{2.13}$$

This ODE, known as the Probability Flow ODE, produces the same marginal distributions as the original SDE but eliminates stochasticity. It offers advantages such as exact likelihood computation and efficient sampling via adaptive solvers. However, its deterministic nature may limit sample diversity. [31]

2.6.3 Advantages and Challenges

Stochastic Differential Equations provide a robust and unified framework for generative modeling, offering numerous advantages while presenting some challenges. One key benefit is the flexibility of the framework, allowing the modeling of complex data distributions using both stochastic and deterministic sampling methods. Techniques such as Predictor-Corrector sampling can enhance accuracy while maintaining computational viability. Moreover, Probability Flow ODEs enable exact likelihood computation. Lastly, as mentioned previously, SDEs can generate high-quality samples without adversarial training.

However, the framework's computational demands can be substantial, particularly for iterative sampling methods like Euler-Maruyama, which require small time steps for precision. Additionally, the performance of these methods is sensitive to hyperparameter choices, including step sizes, noise schedules, drift and diffusion coefficients.

For an easier understanding of the differences and similarities between the approaches discussed in section 2.4, section 2.5 and section 2.6, a comparison is presented in Table A.2.

2.7 The Importance of Iterative Processes in Diffusion Models

As we already concluded, a key characteristic of diffusion models is their step-by-step approach to generating data. This process is different from the one-step generation mechanism used by GANs, allowing for a constant improvement in the quality and fidelity of the generated samples since diffusion models iteratively refine samples through a series of steps.

This iterative process not only improves performance but also provides significant flexibility and control over the generation process, making diffusion models particularly beneficial in tasks requiring high precision, such as the one we are focusing on.

One of the primary benefits of iterative processes is the model's ability to observe and modify intermediate steps in the generation process. By examining these intermediate states, we can gain valuable insights into the model's behavior, identify areas of improvement, and intervene when necessary [35]. This level of clarity is absent in GANs since their generation process occurs in a single step, it can not provide transparency into the underlying processes.

Furthermore, iterativity plays a crucial role in avoiding mode collapse, a common issue in GANs where the generator produces limited and repetitive samples. By using an iterative refinement process, diffusion models maintain diversity in the generated data, better capturing the underlying data distribution [28].

The iterative nature of diffusion models also facilitates the control over the characteristics of the generated samples. Parameters such as the noise schedule β_t, or its derived counterparts α_t and $\bar{\alpha}_t$, can be fine-tuned to adjust the amount of noise added at each step of the forward diffusion process [25]. This allows researchers to influence specific aspects of the generation, such as the resolution, texture, or even the appearance of certain features in the images, since the way the noise is added affects how the model learns to reverse the process. For example, by modifying these parameters, it becomes possible to focus on generating images with certain characteristics, such as lesion boundaries, color distribution, or texture patterns, which are crucial in medical imaging applications [4].

Similar to Conditional GANs, which use labels to guide the generation process, diffusion models can incorporate conditioning variables to produce images aligned with specific categories or characteristics [36]. However, diffusion models go further by being capable of more granular control over features such as the size, shape, or texture of the generated structures. For instance, in the context of skin lesion image generation, these models can be tuned to generate lesions with varying sizes, distinct borders, or specific patterns associated with particular diseases [5].

In practical applications like skin lesion classification, the iterative process can also enable the generation of synthetic images that mimic real-world variability, addressing challenges such as data scarcity and class imbalance. By refining details progressively, diffusion models ensure that both global patterns and critical fine-grained features are accurately represented, making them a powerful tool for augmenting datasets in complex domains like medical diagnostics.

Despite its advantages, iterativity introduces challenges, mainly in terms of computational cost. The step-by-step refinement process can cause a major increase in the time and resources needed for training and sampling. However, advancements such as optimized noise schedules, accelerated sampling techniques, and hybrid approaches, all presented throughout this chapter, have shown promise in mitigating these computational challenges while maintaining high-quality outputs.

2.8 Performance Metrics for Diffusion Models

In this section will be presented the main evaluation metrics used to assess the performance of diffusion models, focusing on the most common ones used in the literature, those being presented in the following diagram:

$$\text{Evaluation Metrics} \begin{cases} \text{Sample Quality Metrics} \begin{cases} \text{Inception Score (IS)} \\ \text{Fréchet Inception Distance (FID)} \end{cases} \end{cases}$$

Like previous generative models, the evaluation of diffusion models have an important role in understanding how advantageous they are across different areas. However, due to the particular characteristics of diffusion models, such as their iterative noise-to-signal generation process and probabilistic nature, the evaluation process requires a more complex approach to quantify their generative capabilities. This meaning that evaluation metrics like accuracy or loss convergence, which are typically used in traditional machine learning models, are not enough to evaluate diffusion models, or at least not as useful.

2.8.1 Inception Score (IS)

With the constant evolution of generative models, principally GANs, that were generating increasingly realistic images, it lead to a challenge in thFSue field: how to automatically evaluate the quality of these images?

Human evaluation was used in multiple instances, but it was not ideal when we are dealing with very large datasets or when quick iterative improvements were needed during model development. So, the need for the use of new evaluation metrics became important. The Inception Score (IS) was one of the first metrics to be adopted, introduced by Salimans *et al.* [22] in their work on improving GANs.

How it Works

The Inception Score works by utilizing a pre-trained neural network called Inception v3 [37], which was originally designed to classify images from the ImageNet dataset [38].

When evaluating a generative model, the process begins by generating a large batch of images. Each of these images is then passed through the Inception network, which in return will output a set of probabilities for each image, representing how likely the image belongs to a certain class, from 0 to 1 for each class.

When analyzing each image, the network is trying to determine how good the generated image is by looking at the clarity and diversity of the sample. When an image is clear and realistic, the network assigns a high probability to one specific class. This means, for example, if the image clearly shows a dog, the "dog" class should get a higher probability than the class "cat" or "car", as well as all other classes. On the other hand, when looking for diversity, this metric examines the distribution of classifications across the entire batch of generated images. A good generator should produce images that cover many classes of the dataset rather than focusing on just a few types of images.

Calculation Method

The factors name above are measured using the Kullback-Leibler (KL) divergence [39], which is a measure of how similar two probability distributions are. In this case, the KL divergence is used to measure the similarity between two probability distributions: the

conditional class distribution for each image and the marginal distribution across all generated images, the latter being the sum of all the probability distribution outputted by the Inception network for each image. Mathematically, the Inception Score (IS) can be expressed as:

$$IS(G) = \exp(\mathbb{E}_{x \sim p_g} D_{KL}(p(y|x) \| p(y))) \tag{2.14}$$

where $p(y|x)$ represents the conditional class distribution obtained from the Inception model for a generated image x, and $p(y)$ represents the marginal class distribution across all generated images. The score can be interpreted as the exponential of the KL divergence between these distributions, for better readability, higher scores meaning quality images. Barratt and Sharma [40] showed in their work that the Inception model can be improved, as well as the way the IS is calculated.

Limitations

However, research compiled about Inception Score by Barratt and Sharma [40], has revealed several significant limitations of the Inception Score:

- **Network Sensitivity**: The score varies significantly depending on the specific weights and implementation of the Inception network used, even when classification accuracy remains similar
- **Dataset Mismatch**: Using the score for datasets other than ImageNet (for which the Inception network was trained) can lead to misleading results, as the class predictions may not align with the actual dataset classes
- **Vulnerability to Exploitation**: The score can be artificially inflated by generating adversarial examples that maximize class confidence without actually improving image quality
- **Insufficient Overfitting Detection**: The metric alone cannot determine whether a model is simply memorizing training data

Transition to Better Metrics

The Inception Score (IS) was initially a valuable starting point for evaluating generative models, but its limitations restrict its utility, especially for models like diffusion models. IS measures the quality and confidence of generated images based on their fit into distinct object categories, relying on a pre-trained classifier (typically Inception v3). While useful for datasets with well-defined classes like ImageNet, IS does not directly compare generated images to real ones, making it less reliable for measuring fidelity to the true data distribution. This shortcoming is particularly evident with diffusion models, which often generate more complex or abstract data that may not align with predefined categories.

These limitations showed the need for a stronger evaluation metric. The Fréchet Inception Distance (Fréchet Inception Distance (FID)) appeared as a good alternative, addressing several shortcomings of IS. Unlike IS, FID compares the generated images with the real data distribution.

2.8.2 Fréchet Inception Distance (FID)

As it was mentioned in the previous section, the Fréchet Inception Distance (FID) came to address the limitations of the Inception Score (IS), enabling the comparison of the generated images with the real data distribution, thus providing a more robust analysis of the similarity of the two distributions. This metric, which name came from the Inception model and the Fréchet distance, was first introduced by Heusel *et al.* [41] in their work on evaluating the performance of GANs, becoming a commonly used metric for evaluating generative models, especially in image synthesis tasks.

How it Works

To compute the FID, both real and generated images are analyzed using a pre-trained model (e.g., Inception-v3), which extracts key features from the images. These features are summarized as two sets of statistics: the average values (means, μ) and the spread or variation (covariance, Σ).

FID calculates the similarity between these statistics using the Fréchet Distance, a measure originally developed to compare curves but extended here to higher-dimensional distributions [3], [41]. A lower FID score indicates that the generated images are more similar to the real ones in terms of quality and diversity.

FID is particularly effective at identifying issues like mode collapse, where the generative model fails to capture the full diversity of the real data distribution. Additionally, FID is robust to adversarial noise and artifacts, making it a more reliable metric than some likelihood-based methods.

Calculation Method

The computation of FID involves three primary steps. First, real and generated images are passed through the Inception-v3 model to extract feature embeddings. Next, the mean (μ) and covariance (Σ) of these embeddings are calculated for both real and generated image distributions. Finally, the Fréchet Distance formula is used, with p referring to the real image distribution and q to the generated image distribution, adapted from [3], [41]:

$$FID = ||\mu_p - \mu_q||_2^2 + \mathrm{Tr}(\Sigma_p + \Sigma_q - 2(\Sigma_p\Sigma_q)^{1/2}), \tag{2.15}$$

where μ_p and Σ_p are the mean and covariance matrix of the real images, and μ_q and Σ_q are the corresponding statistics for the generated images.

In this formula:

- $||\mu_p - \mu_q||_2^2$ represents the squared Euclidean norm, which measures the distance between the means of the real and generated distributions.
- $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix, which is the sum of its diagonal elements. It is used here to quantify the differences between the covariances of the two distributions.

This formula accounts for both differences in means (global distribution shifts) and covariances (diversity within distributions).

Limitations

Even though FID appeared as an improvement over IS, it's not without its own limitations. One of the main issues is that assumes that the extracted features follow a Gaussian distribution, a nice assumption but not always true, which can make FID less reliable in more complex scenarios. Additionally, this metric also depends heavily on the pre-trained Inception-v3 model, with no guarantees that the extracted features capture all relevant aspects of the data.

Furthermore, FID considers only the first two moments of the distributions (mean, μ, and covariance, Σ), which may miss higher-order statistics that could be important for certain applications.

It's also worth mentioning that FID can be quite demanding in terms of sample size, needing a lot of samples to achieve stable and meaningful results, which can make computation resource-intensive. Because of these limitations, it's a good idea to pair FID with other metrics to get a fuller picture of how well a generative model is performing.

2.9 Summary

This chapter focused on making an overview of key advancements and challenges in generative modeling and their relevance to medical imaging. It began by discussing the intersection of deep learning and medical imaging, emphasizing the problem of data scarcity and the critical role of data augmentation and generative models.

A detailed introduction to diffusion models followed, with a focus on Denoising Diffusion Probabilistic Models (DDPM), their forward diffusion and reverse denoising processes, and the simplifications that make them effective. Noise Conditional Score Networks (NCSN) were also explored, highlighting their theoretical underpinnings, limitations, and connections to diffusion models.

Stochastic Differential Equation (SDE) were introduced as a unifying framework, with emphasis on forward and reverse processes, numerical solutions for sampling, and their strengths and weaknesses. To consolidate understanding, a comparative analysis of DDPMs, NCSN, and SDEs was provided, focusing on their objectives, noise handling, mathematical frameworks, and extensions. For more details refer to Table A.2.

Finally, performance metrics such as Inception Score (IS) and Fréchet Inception Distance (FID) were reviewed, offering a basis for evaluating the generative quality of these models. Collectively, this chapter establishes the foundation for understanding the state-of-the-art in diffusion-based generative modeling and its applications in medical imaging.

CHAPTER 3

Methodology and Work Plan

After a thorough review of the key topics, background, concepts, and technologies, we are ready to focus on the main goal of this thesis: improving a CNN model of skin disease diagnosis with adversarial training and diffusion models. This chapter explains the methodology used in this project, including what has been done so far and the plan for the next steps.

3.1 Work Done

The initial phase of my thesis focused on conducting a comprehensive literature review to guarantee a solid understanding of the foundation for the study. This review covered several key areas like, applications of deep learning in skin lesion diagnosis, relevant performance metrics, data augmentation techniques, synthetic data generation methodologies, and multiples frameworks of diffusion models. This analysis was very helpful to identify the current gaps and challenges in the field, guiding the direction of the research.

In addition, multiple public datasets were explored, mainly the ISIC archive, but also datasets like the HAM10000 [42] are being carefully considered for the implementation.

In parallel, I implemented a foundational version of DDPMs. This implementation was inspired by several sources, including Vandegar [43], detailed guides by Erdem [44], and the implementation of Pulfer [45]. These resources were very helpful to help me grasp the concepts and some practical aspects of DDPMs.

The implementation, done either by me or the other sources, were informed by important paper in the field, notably *Denoising Diffusion Implicit Models* by Song *et al.* [46] and *Improved Denoising Diffusion Probabilistic Models* by Nichol and Dhariwal [35]. These papers offered the necessary mathematical formulations and theoretical advancements to construct the model effectively, in addition to the papers covered in the literature review.

The implementation was done using Python as the main programming language along with some machine learning and data processing libraries. PyTorch was used as the basis for defining and training the neural networks, handling tensors, and implementing the forward and reverse diffusion processes. Numpy was used for array manipulation and other preprocessing

tasks, while Matplotlib was used for data visualization. Additionally, to facilitate the loading and normalization of the chosen dataset (MNIST), was used the torchvision library.

As mentioned, the MNIST dataset[47] was used. The MNIST dataset, consisting of 70,000 grayscale images of handwritten digits, serves as a benchmark for testing generative models because of its simplicity and adoption in the machine learning community. Each image in the dataset is a 28x28 pixel array, where each pixel value ranges from 0 to 255.

The implementation began by visualizing the original dataset (Figure 3.1a). This provided a baseline for understanding the transformations applied to the data during the diffusion process.

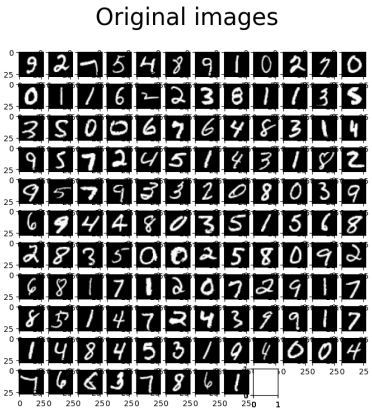
As it was mentioned multiple times, the core idea of DDPMs involves a two-step process: the forward diffusion process and the reverse diffusion process.

To test the forward diffusion process, progressively increasing levels of Gaussian noise were applied to the MNIST images. This can be viewed in the Figures 3.1b, 3.1c, 3.1d, and 3.1e, which correspond to noise levels of 25%, 50%, 75%, and 100%, respectively. These figures demonstrate how the images gradually lose their structure and become indistinguishable from pure noise as the noise level increases.

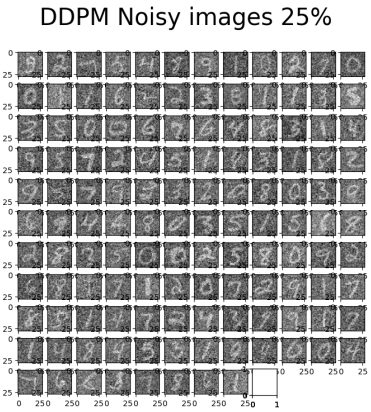
Before training, initial images generated by the model were also visualized to test the model's ability to generate images from random noise. These images were also used to see the starting point of the reverse diffusion process. The generated images before training were of low quality and lacked any discernible structure, as shown in Figure 3.1f.

After training, the generated images demonstrated significant improvement in quality, as shown in Figure 3.1g. Since the model was not trained for a long time, the generated images were not perfect, but they already resembled the original MNIST digits, showing that the model was able to understand the underlying data distribution.

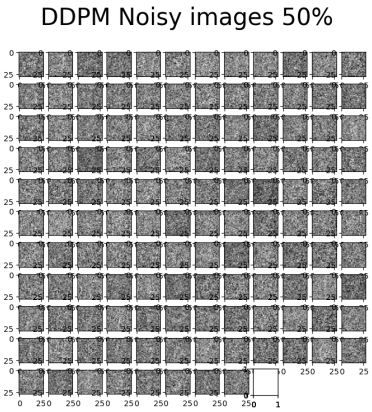
This result also highlights the importance of iterative refinement during the reverse diffusion process, where the model gradually removes noise to reconstruct the original data.



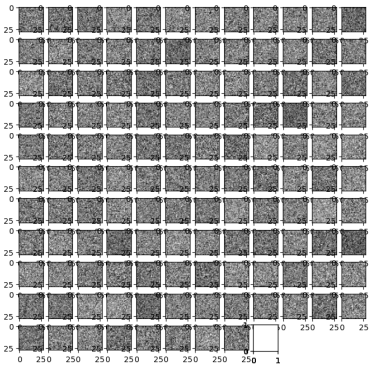
(a) Original Images



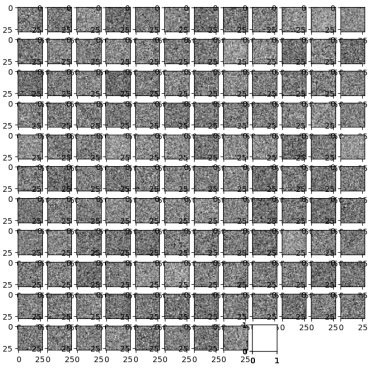
(b) 25% Noise



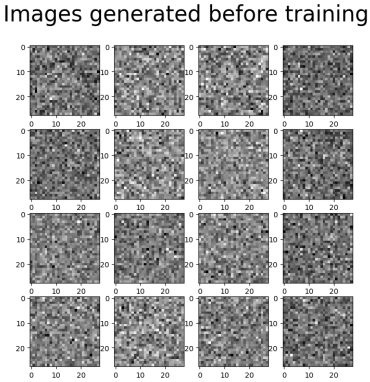
(c) 50% Noise



(d) 75% Noise



(e) 100% Noise



(f) Generated Images Before Training

(g) Generated Images After Training

Figure 3.1: MNIST dataset images at different stages of the diffusion process.

3.2 Work Plan

For the following months, the project is going to be divided into four main stages:

1. Deep Classifier Deployment
2. Diffusion Model
3. Adversarial Training
4. Dissertation and Documentation

The first task is to deploy a deep classifier. This includes selecting a pre-trained CNN architecture that is appropriate for image classification and training the classifier on a real dataset of skin lesion images. Following this, the classifier's performance will be evaluated on a separate test set using metrics such as accuracy, sensitivity, and specificity.

The second task is continuing the research on diffusion models and implement simple models to gain more knowledge about the topic, and then be able to choose an adequate architecture. Additionally, hyperparameters of the diffusion model will be fine-tuned to achieve optimal performance, and the quality and realism of the generated synthetic images will be analyzed.

Since my work is intended to be compared to the thesis of Silva [3], I will likely use the same datasets as the prior research to ensure consistency and facilitate meaningful comparisons. However, the final selection of datasets is still under consideration. In the initial stages of diffusion model testing, I plan to work with the CIFAR-10 dataset, a widely used benchmark for image generation tasks created by Canadian Institute For Advanced Research (CIFAR), before I start using medical images. This will allow for an initial testing and optimization of the diffusion model in a more controlled and understood environment.

The third task is centered on adversarial training. This involves developing an iterative training loop that includes:

1. Identifying misclassified images from the test set.
2. Using these challenging cases to guide the diffusion model to generate synthetic data that replicates their features.
3. Re-training the classifier on the real and newly generated synthetic data.

The performance metrics will be compared before and after adversarial training to evaluate the effectiveness of the synthetic data in improving classification, particularly for challenging cases.

Finally, the last stage, that is in reality a continuous process throughout the semester, involves preparing the dissertation and supporting documentation. This includes writing the master dissertation, creating comprehensive code repositories, and developing reports to ensure reproducibility and support future research efforts.

The following Gantt chart, Figure 3.2, shows the work proposal for the following months, as well as the work already done. The timeline is divided into two semesters: the first highlighting the tasks completed during the initial semester, while the second details the planned activities for the second semester.

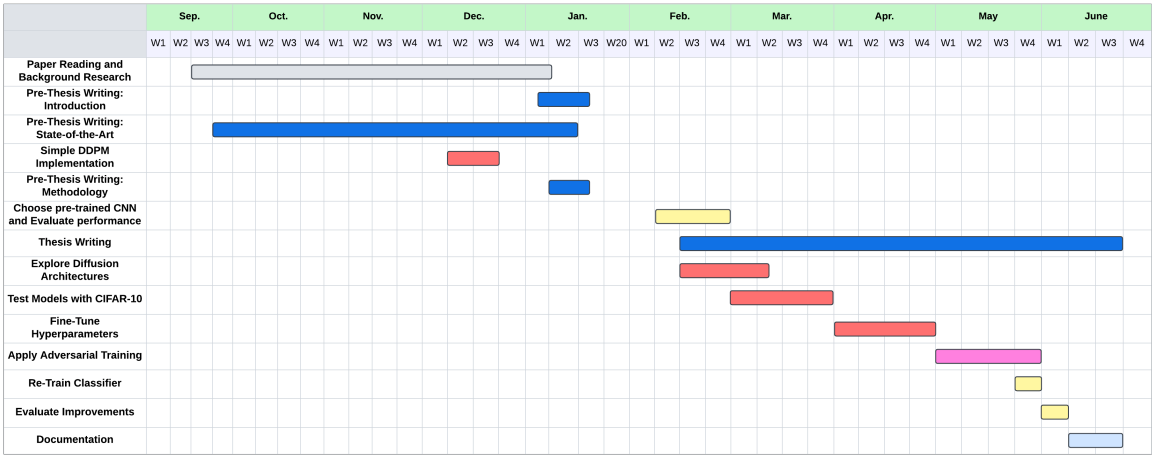


Figure 3.2: Gantt chart showing the work done and work plan for the following months.

References

[1] S. P. Choy, B. J. Kim, A. Paolino, *et al.*, «Systematic review of deep learning image analyses for the diagnosis and monitoring of skin disease», *npj Digital Medicine*, vol. 6, no. 1, ISSN: 2398-6352. DOI: 10.1038/s41746-023-00914-8. [Online]. Available: https://www.nature.com/articles/s41746-023-00914-8 (visited on 11/19/2024).

[2] K. P. Venkatesh, M. M. Raza, G. Nickel, S. Wang, and J. C. Kvedar, «Deep learning models across the range of skin disease», *npj Digital Medicine*, vol. 7, no. 1, 32, s41746–024–01033–8, Feb. 10, 2024, ISSN: 2398-6352. DOI: 10.1038/s41746-024-01033-8. [Online]. Available: https://www.nature.com/articles/s41746-024-01033-8 (visited on 11/19/2024).

[3] G. A. S. Silva, «Data augmentation and deep classification with generative adversarial networks», Accepted: 2021-09-30T15:00:50Z Journal Abbreviation: Aumento de dados e classificação profunda com redes adversárias generativas, masterThesis, Jul. 28, 2021. [Online]. Available: https://ria.ua.pt/handle/10773/32283 (visited on 11/19/2024).

[4] A. Kazerouni, E. K. Aghdam, M. Heidari, *et al.*, «Diffusion models in medical imaging: A comprehensive survey», *Medical Image Analysis*, vol. 88, p. 102 846, Aug. 2023, ISSN: 13618415. DOI: 10.1016/j.media.2023.102846. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1361841523001068 (visited on 11/19/2024).

[5] M. Akrout, B. Gyepesi, P. Holló, *et al.*, *Diffusion-based data augmentation for skin disease classification: Impact across original medical datasets to fully synthetic images*, Jan. 12, 2023. DOI: https://doi.org/10.48550/arXiv.2301.04802. arXiv: 2301.04802[cs]. [Online]. Available: http://arxiv.org/abs/2301.04802 (visited on 11/19/2024).

[6] L. W. Sagers, J. A. Diao, M. Groh, P. Rajpurkar, A. S. Adamson, and A. K. Manrai, *Improving dermatology classifiers across populations using images generated by large diffusion models*, Nov. 23, 2022. arXiv: 2211.13352[eess]. [Online]. Available: http://arxiv.org/abs/2211.13352 (visited on 11/19/2024).

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, «ImageNet classification with deep convolutional neural networks», in *Advances in Neural Information Processing Systems*, vol. 25, Curran Associates, Inc., 2012. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html (visited on 11/20/2024).

[8] N. C. F. Codella, D. Gutman, M. E. Celebi, *et al.*, «Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)», in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, Washington, DC: IEEE, Apr. 2018, pp. 168–172, ISBN: 978-1-5386-3636-7. DOI: 10.1109/ISBI.2018.8363547. [Online]. Available: https://ieeexplore.ieee.org/document/8363547/ (visited on 11/19/2024).

[9] C. Shorten and T. M. Khoshgoftaar, «A survey on image data augmentation for deep learning», *Journal of Big Data*, vol. 6, no. 1, p. 60, Jul. 6, 2019, ISSN: 2196-1115. DOI: 10.1186/s40537-019-0197-0. [Online]. Available: https://doi.org/10.1186/s40537-019-0197-0 (visited on 11/25/2024).

[10] L. Perez and J. Wang, *The effectiveness of data augmentation in image classification using deep learning*, Dec. 13, 2017. DOI: 10.48550/arXiv.1712.04621. arXiv: 1712.04621. [Online]. Available: http://arxiv.org/abs/1712.04621 (visited on 11/20/2024).

[11] F. Perez, C. Vasconcelos, S. Avila, and E. Valle, «Data augmentation for skin lesion analysis», in *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based*

Procedures, and Skin Image Analysis, D. Stoyanov, Z. Taylor, D. Sarikaya, *et al.*, Eds., vol. 11041, Series Title: Lecture Notes in Computer Science, Cham: Springer International Publishing, 2018, pp. 303–311, ISBN: 978-3-030-01200-7 978-3-030-01201-4. DOI: 10.1007/978-3-030-01201-4_33. [Online]. Available: https://link.springer.com/10.1007/978-3-030-01201-4_33 (visited on 11/19/2024).

[12] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, *et al.*, *Generative adversarial networks*, Jun. 10, 2014. DOI: 10.48550/arXiv.1406.2661. arXiv: 1406.2661. [Online]. Available: http://arxiv.org/abs/1406.2661 (visited on 11/25/2024).

[13] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, *Generative adversarial networks: An overview*, Oct. 19, 2017. DOI: 10.48550/arXiv.1710.07035. arXiv: 1710.07035. [Online]. Available: http://arxiv.org/abs/1710.07035 (visited on 11/25/2024).

[14] J. Hayes, L. Melis, G. Danezis, and E. D. Cristofaro, «LOGAN: Evaluating privacy leakage of generative models using generative adversarial networks», *ArXiv*, May 22, 2017. [Online]. Available: https://www.semanticscholar.org/paper/LOGAN%3A-Evaluating-Privacy-Leakage-of-Generative-Hayes-Melis/7c4f52328c2869bdff8034d2867baa5b67d0ce27 (visited on 11/28/2024).

[15] S. Kazeminia, C. Baur, A. Kuijper, *et al.*, «GANs for medical image analysis», *Artificial Intelligence in Medicine*, vol. 109, p. 101 938, Sep. 1, 2020, ISSN: 0933-3657. DOI: 10.1016/j.artmed.2020.101938. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0933365719311510 (visited on 11/25/2024).

[16] J. Jiang, Y.-C. Hu, N. Tyagi, *et al.*, «Tumor-aware, adversarial domain adaptation from CT to MRI for lung cancer segmentation», *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, vol. 11071, pp. 777–785, Sep. 2018. DOI: 10.1007/978-3-030-00934-2_86. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6169798/ (visited on 11/25/2024).

[17] D. Korkinof, T. Rijken, M. O'Neill, J. Yearsley, H. Harvey, and B. Glocker, *High-resolution mammogram synthesis using progressive generative adversarial networks*, Sep. 3, 2019. DOI: 10.48550/arXiv.1807.03401. arXiv: 1807.03401. [Online]. Available: http://arxiv.org/abs/1807.03401 (visited on 11/25/2024).

[18] A. Bissoto, F. Perez, E. Valle, and S. Avila, *Skin lesion synthesis with generative adversarial networks*, Feb. 8, 2019. DOI: 10.48550/arXiv.1902.03253. arXiv: 1902.03253. [Online]. Available: http://arxiv.org/abs/1902.03253 (visited on 11/25/2024).

[19] C. Han, K. Murao, T. Noguchi, *et al.*, *Learning more with less: Conditional PGGAN-based data augmentation for brain metastases detection using highly-rough annotation on MR images*, Aug. 22, 2019. DOI: 10.48550/arXiv.1902.09856. arXiv: 1902.09856. [Online]. Available: http://arxiv.org/abs/1902.09856 (visited on 11/25/2024).

[20] M. J. M. Chuquicusma, S. Hussein, J. Burt, and U. Bagci, *How to fool radiologists with generative adversarial networks? a visual turing test for lung cancer diagnosis*, Jan. 9, 2018. DOI: 10.48550/arXiv.1710.09762. arXiv: 1710.09762. [Online]. Available: http://arxiv.org/abs/1710.09762 (visited on 11/25/2024).

[21] M. Buda, A. Maki, and M. A. Mazurowski, «A systematic study of the class imbalance problem in convolutional neural networks», *Neural Networks*, vol. 106, pp. 249–259, Oct. 1, 2018, ISSN: 0893-6080. DOI: 10.1016/j.neunet.2018.07.011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608018302107 (visited on 12/02/2024).

[22] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, *Improved techniques for training GANs*, Jun. 10, 2016. DOI: 10.48550/arXiv.1606.03498. arXiv: 1606.03498. [Online]. Available: http://arxiv.org/abs/1606.03498 (visited on 12/02/2024).

[23] J. P. Cohen, M. Luck, and S. Honari, *Distribution matching losses can hallucinate features in medical image translation*, Oct. 3, 2018. DOI: 10.48550/arXiv.1805.08841. arXiv: 1805.08841. [Online]. Available: http://arxiv.org/abs/1805.08841 (visited on 12/02/2024).

[24] J. Sohl-Dickstein, E. A. Weiss, N. Maheswaranathan, and S. Ganguli, *Deep unsupervised learning using nonequilibrium thermodynamics*, Nov. 18, 2015. DOI: 10.5555/3045118.3045358. arXiv: 1503.03585[cs]. [Online]. Available: http://arxiv.org/abs/1503.03585 (visited on 11/19/2024).

[25] J. Ho, A. Jain, and P. Abbeel, *Denoising diffusion probabilistic models*, Dec. 16, 2020. arXiv: 2006.11239[cs]. [Online]. Available: http://arxiv.org/abs/2006.11239 (visited on 11/19/2024).

[26] D. P. Kingma and M. Welling, *Auto-encoding variational bayes*, Dec. 10, 2022. DOI: 10.48550/arXiv.1312.6114. arXiv: 1312.6114[stat]. [Online]. Available: http://arxiv.org/abs/1312.6114 (visited on 12/31/2024).

[27] Y. Song and S. Ermon, *Generative modeling by estimating gradients of the data distribution*, Oct. 10, 2020. arXiv: 1907.05600[cs]. [Online]. Available: http://arxiv.org/abs/1907.05600 (visited on 11/19/2024).

[28] Y. Song and S. Ermon, «Improved techniques for training score-based generative models», in *Advances in Neural Information Processing Systems*, vol. 33, Curran Associates, Inc., 2020, pp. 12 438–12 448. [Online]. Available: https://papers.neurips.cc/paper_files/paper/2020/hash/92c3b916311a5517d9290576e3ea37ad-Abstract.html (visited on 11/20/2024).

[29] P. Vincent, «A connection between score matching and denoising autoencoders», *Neural Computation*, vol. 23, no. 7, pp. 1661–1674, Jul. 2011, Conference Name: Neural Computation, ISSN: 0899-7667. DOI: 10.1162/NECO_a_00142. [Online]. Available: https://ieeexplore.ieee.org/document/6795935 (visited on 01/13/2025).

[30] R. M. Neal, «Probabilistic inference using markov chain monte carlo methods», 1993.

[31] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, *Score-based generative modeling through stochastic differential equations*, Feb. 10, 2021. arXiv: 2011.13456[cs]. [Online]. Available: http://arxiv.org/abs/2011.13456 (visited on 11/19/2024).

[32] B. D. Anderson, «Reverse-time diffusion equation models», *Stochastic Processes and their Applications*, vol. 12, no. 3, pp. 313–326, May 1982, ISSN: 03044149. DOI: 10.1016/0304-4149(82)90051-5. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/0304414982900515 (visited on 01/13/2025).

[33] N. Wiener, «Differential-space», *Journal of Mathematics and Physics*, vol. 2, no. 1, pp. 131–174, 1923, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sapm192321131, ISSN: 1467-9590. DOI: 10.1002/sapm192321131. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/sapm192321131 (visited on 01/14/2025).

[34] P. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations* (Stochastic Modelling and Applied Probability). Springer Berlin Heidelberg, 2011, ISBN: 9783540540625. [Online]. Available: https://books.google.pt/books?id=BCvtssom1CMC.

[35] A. Nichol and P. Dhariwal, *Improved denoising diffusion probabilistic models*, Feb. 18, 2021. arXiv: 2102.09672[cs]. [Online]. Available: http://arxiv.org/abs/2102.09672 (visited on 11/19/2024).

[36] P. Dhariwal and A. Nichol, *Diffusion models beat GANs on image synthesis*, Jun. 1, 2021. DOI: 10.48550/arXiv.2105.05233. arXiv: 2105.05233[cs]. [Online]. Available: http://arxiv.org/abs/2105.05233 (visited on 01/18/2025).

[37] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, *Rethinking the inception architecture for computer vision*, Dec. 11, 2015. DOI: 10.48550/arXiv.1512.00567. arXiv: 1512.00567[cs]. [Online]. Available: http://arxiv.org/abs/1512.00567 (visited on 01/05/2025).

[38] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, «ImageNet: A large-scale hierarchical image database», in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, ISSN: 1063-6919, Jun. 2009, pp. 248–255. DOI: 10.1109/CVPR.2009.5206848. [Online]. Available: https://ieeexplore.ieee.org/document/5206848 (visited on 01/05/2025).

[39] S. Kullback and R. A. Leibler, «On information and sufficiency», *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 79–86, Mar. 1951, Publisher: Institute of Mathematical Statistics, ISSN: 0003-4851, 2168-8990. DOI: 10.1214/aoms/1177729694. [Online]. Available: https://projecteuclid.org/journals/annals-of-mathematical-statistics/volume-22/issue-1/On-Information-and-Sufficiency/10.1214/aoms/1177729694.full (visited on 01/05/2025).

[40] S. Barratt and R. Sharma, *A note on the inception score*, Jun. 21, 2018. DOI: 10.48550/arXiv.1801.01973. arXiv: 1801.01973[stat]. [Online]. Available: http://arxiv.org/abs/1801.01973 (visited on 01/04/2025).

[41] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, *GANs trained by a two time-scale update rule converge to a local nash equilibrium*, Jan. 12, 2018. DOI: 10.48550/arXiv.1706.08500. arXiv: 1706.08500[cs]. [Online]. Available: http://arxiv.org/abs/1706.08500 (visited on 01/04/2025).

[42] P. Tschandl, C. Rosendahl, and H. Kittler, «The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions», *Scientific Data*, vol. 5, p. 180 161, 2018. DOI: 10.1038/sdata.2018.161. [Online]. Available: https://doi.org/10.1038/sdata.2018.161.

[43] M. Vandegar, *Papers in 100 lines of code - denoising diffusion probabilistic models*, https://github.com/MaximeVandegar/Papers-in-100-Lines-of-Code/tree/main/Denoising_Diffusion_Probabilistic_Models, Available at https://github.com/MaximeVandegar/Papers-in-100-Lines-of-Code/tree/main/Denoising_Diffusion_Probabilistic_Models, 2023.

[44] K. Erdem, *Step-by-step visual introduction to diffusion models*, Available at https://medium.com/@kemalpiro/step-by-step-visual-introduction-to-diffusion-models-235942d2f15c, 2020.

[45] B. Pulfer, *Denoising diffusion probabilistic models implementation*, Available at https://github.com/BrianPulfer, 2023.

[46] J. Song, C. Meng, and S. Ermon, *Denoising diffusion implicit models*, Oct. 5, 2022. arXiv: 2010.02502[cs]. [Online]. Available: http://arxiv.org/abs/2010.02502 (visited on 11/19/2024).

[47] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, «Gradient-based learning applied to document recognition», *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998, ISSN: 00189219. DOI: 10.1109/5.726791. [Online]. Available: http://ieeexplore.ieee.org/document/726791/ (visited on 01/15/2025).

[48] G. Müller-Franzes, J. M. Niehues, F. Khader, *et al.*, «A multimodal comparison of latent denoising diffusion probabilistic models and generative adversarial networks for medical image synthesis», *Scientific Reports*, vol. 13, no. 1, p. 12 098, Jul. 26, 2023, Publisher: Nature Publishing Group, ISSN: 2045-2322. DOI: 10.1038/s41598-023-39278-0. [Online]. Available: https://www.nature.com/articles/s41598-023-39278-0 (visited on 01/18/2025).

[49] S. Mukhopadhyay, M. Gwilliam, V. Agarwal, *et al.*, *Diffusion models beat GANs on image classification*, Jul. 17, 2023. DOI: 10.48550/arXiv.2307.08702. arXiv: 2307.08702[cs]. [Online]. Available: http://arxiv.org/abs/2307.08702 (visited on 01/18/2025).

[50] P. Patcharapimpisut and P. Khanarsa, «Generating synthetic images using stable diffusion model for skin lesion classification», in *2024 16th International Conference on Knowledge and Smart Technology (KST)*, ISSN: 2473-764X, Feb. 2024, pp. 184–189. DOI: 10.1109/KST61284.2024.10499667. [Online]. Available: https://ieeexplore.ieee.org/document/10499667/?arnumber=10499667 (visited on 01/18/2025).

[51] M. A. Farooq, W. Yao, M. Schukat, M. A. Little, and P. Corcoran, «Derm-t2im: Harnessing synthetic skin lesion data via stable diffusion models for enhanced skin disease classification using ViT and CNN», in *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jul. 15, 2024, pp. 1–5. DOI: 10.1109/EMBC53108.2024.10781852. arXiv: 2401.05159[cs]. [Online]. Available: http://arxiv.org/abs/2401.05159 (visited on 01/18/2025).

[52] R. Zhang, Y. Yao, Z. Tan, *et al.*, *FairSkin: Fair diffusion for skin disease image generation*, Oct. 31, 2024. DOI: 10.48550/arXiv.2410.22551. arXiv: 2410.22551[cs]. [Online]. Available: http://arxiv.org/abs/2410.22551 (visited on 01/18/2025).

[53] J. Wang, Y. Chung, Z. Ding, and J. Hamm, *From majority to minority: A diffusion-based augmentation for underrepresented groups in skin lesion analysis*, Jul. 30, 2024. DOI: 10.48550/arXiv.2406.18375. arXiv: 2406.18375[cs]. [Online]. Available: http://arxiv.org/abs/2406.18375 (visited on 01/18/2025).

APPENDIX A

Additional Content

A.1 Diffusion Models in Medical Imaging

Table A.1: Summary of studies on diffusion models in medical imaging, highlighting key tasks such as image synthesis, classification, and data augmentation.

Reference	**Topic**	**Explanation**
Dhariwal and Nichol (2021, [36])	Image synthesis	Showed that diffusion models outperform GANs for image synthesis, achieving state-of-the-art FID scores with classifier guidance for diversity-fidelity trade-offs.
Müller-Franzes *et al.* (2023, [48])	Medical image synthesis	Proposed Medfusion, a diffusion model, outperforming GANs in diversity and fidelity for medical imaging datasets.
Mukhopadhyay *et al.* (2023, [49])	Image classification	Demonstrated that diffusion models excel in generative and classification tasks, surpassing BigBiGAN in ImageNet classification.
Patcharapimpisut and Khanarsa (2024, [50])	Skin lesion classification	Used diffusion models for generating synthetic skin lesion images, improving recall by 4-5% and accuracy by 1-2%.
Farooq *et al.* (2024, [51])	Skin lesion classification	Developed Derm-T2IM for generating synthetic dermatoscopic images, boosting robustness in ViT and CNN models.
Zhang *et al.* (2024, [52])	Medical Image generation	Introduced FairSkin, addressing biases in image quality and feature learning across skin tones for equitable diagnosis.

Akrout *et al.* (2023, [5])	Data augmentation	Leveraged diffusion models to augment skin disease datasets, maintaining classification accuracy with synthetic data.
Wang *et al.* (2024, [53])	Minority skin lesion analysis	Proposed diffusion-based augmentation to improve diagnostics for underrepresented skin types using majority group data.

A.2 Comparison of DDPMs, NCSNs, and SDEs

Table A.2: Comparison of DDPMs, NCSNs, and SDEs

Aspect	**Denoising Diffusion Probabilistic Models (DDPMs)**	**Noise-Conditioned Score Networks (NCSNs)**	**Stochastic Differential Equations (SDEs)**
Objective	Optimizes a Variational Lower Bound (VLB) or directly predicts noise for denoising.	Learns the score function $\nabla_x \log p(x)$ using score matching.	Models forward and reverse processes as solutions to SDEs, providing a continuous-time framework.
Noise Handling	Models a continuous noise corruption and denoising process over time steps.	Conditions explicitly on discrete noise levels σ_i.	Uses continuous noise levels described by drift and diffusion coefficients.
Training Method	Employs a reweighted MSE loss to learn the reverse denoising process.	Uses denoising score matching to estimate gradients of the log-density.	Uses score-based training on SDE trajectories, often combining score matching and likelihood maximization.
Sampling Process	Reverse diffusion guided by a trained neural network.	Annealed Langevin Dynamics with iterative score-based updates.	Numerical solvers like Euler-Maruyama, Predictor-Corrector, or Probability Flow ODEs.
Noise Schedule	Continuous or discretized noise schedules governing the diffusion and reverse processes.	Predefined discrete noise levels.	Governed by the drift and diffusion terms of the SDE, offering higher flexibility.
Mathematical Framework	Based on discrete-time Markov chains and Gaussian noise transitions.	Grounded in score matching and Langevin dynamics.	Grounded in stochastic calculus, unifying score-based methods and diffusion processes.

Table A.2 (continued)

Aspect	**Denoising Diffusion Probabilistic Models (DDPMs)**	**Noise-Conditioned Score Networks (NCSNs)**	**Stochastic Differential Equations (SDEs)**
Generative Quality	High quality with optimized versions offering faster sampling.	High quality but slower sampling due to ALD.	Flexible and robust, enabling both high-quality and efficient sample generation.
Extensions	Enhanced with improved sampling methods, conditional generation, and hybrid architectures.	Extended to continuous noise levels, bridging with SDE frameworks.	General framework encompassing both DDPMs and NCSNs, enabling extensions to complex data distributions.