Final Progress Report for the Advanced Computer Science Master's Project

Comparative Analysis of Generative Adversarial Networks (GANs) versus Classical Augmentation for COVID-19 CT Scan Classification

Submitted by:

Md Tanvir Sarwar,
Student ID: 23071769,
MSc Advanced Computer Science with Advanced Research,
Session: 2024-25,
University of Hertfordshire,
United Kingdom.

Supervised By:

Tamie Salter,
Professor,
Hult International Business School,
United Kingdom.

Proof-reading and Quality Check Confirmation: I confirm that I have critically proof-read and quality checked this report, ensuring it is free from grammar, spelling, and formatting errors and meets high standards of clarity, coherence, and presentation.

MSc FPR Declaration

This report is submitted in partial fulfillment of the requirements for the degree of:

Master of Science in Advanced Computer Science Master's Project, at the University of Hertfordshire (UH).

I hereby declare that the work presented in this project and report is entirely my own, except where explicitly stated otherwise. All sources of information and ideas, whether quoted directly or paraphrased, have been properly referenced in accordance with academic standards. I understand that any failure to properly acknowledge the work of others could constitute plagiarism and may result in academic penalties.

I did not use human participants in my MSc project.

I hereby **withhold** permission for the report to be made available on the university website, provided the source is acknowledged.

Declaration of Usage of Technical and AI Tools

- **Research Paper Selection: Gemini** and **ChatGPT** are employed to optimize the research paper selection process, thereby conserving time. **SciSpace** is utilized to identify key findings, further streamlining paper selection and saving additional time.
- **Coding: Co-pilot** and **Gemini** are employed for bug resolution. **Jupyter Notebook** has been the coding environment used to date.
- **Word Rephrasing: QuillBot**, **Grammarly**, and **Gemini** were utilized to rephrase the content for the Interim Progress Report.

Table of Contents

Abstract

Context: The global health sector's crisis caused by the COVID-19 pandemic highlighted the critical need for rapid, automated diagnostic tools to alleviate the workload on radiologists. Currently, the available Deep Learning models, particularly Convolutional Neural Networks (CNNs), have achieved significant accuracy in the detection of viral pathologies through Chest Computed Tomography (CT) scans (Roberts et al. 199-217). However, the deployment of advanced clinical AI models continues to be substantially hindered by the limited availability of large, diverse, and accurately labeled medical image datasets due to stringent privacy regulations such as GDPR (Li et al. 1-6). Conventional image data augmentation methods, such as geometric transformations, frequently do not produce sufficient biological variability to avert overfitting when utilizing small datasets.

Research Goal: This project investigates the efficacy of Generative Adversarial Networks (GANs) as a superior data augmentation approach relative to conventional techniques. The primary objective is to determine whether the synthetic data produced by Deep Convolutional GANs (DCGANs) results in a statistically significant improvement in classification performance for binary COVID-19 detection.

Research Question: "To what extent does synthetic data augmentation using Deep Convolutional Generative Adversarial Networks (DCGANs) surpass traditional image augmentation in enhancing the classification accuracy of CNNs on small-scale medical datasets, and what are the computational trade-offs when executed on consumer-grade hardware?"

Methods: A whole deep learning pipeline was designed with Python and PyTorch, specifically tuned for consumer-grade GPU acceleration to illustrate accessibility. A bespoke DCGAN architecture was developed and independently trained on COVID-19 and non-COVID classes to comprehend the intrinsic data distribution and generate realistic 64 × 64 chest CT slices. A VGG16 classifier, designed for single-channel grayscale input and hardware compatibility, was trained under two experimental conditions: **(1) Classical Augmentation** (rotation, flipping) and **(2) GAN-based Augmentation** (combining actual and synthetic data).

Results: The experimental findings clearly indicate that the generative technique possesses a significant performance advantage. The GAN-augmented model achieved a validation accuracy of **98.12%**, surpassing the **93.01%** attained by the traditionally augmented model. The qualitative visual evaluation verified that the DCGAN effectively captured critical radiological features, including ground-glass opacities and consolidation. Consequently, the training dataset was augmented with significant pathogenic variety.

Conclusion: This study concludes that while generative augmentation incurs a higher computational cost during the training phase, it offers a powerful solution to the medical data scarcity problem. By synthesizing high-fidelity training examples, this approach significantly enhances model generalization, reducing false negatives in critical diagnostic tasks, even when deployed on standard consumer hardware.

Introduction

Convolutional Neural Networks (CNNs) have become incredibly popular in image processing, mainly because they're great at picking out key features in pictures (Jogin et al.). But these days, we're asking them to do a lot more than just extract features or fiddle with image datasets.

When it comes to using CNNs for medical diagnosis, there's a big catch: they need a massive amount of high-quality, well-labeled training data to work well. This is a huge hurdle in medical imaging. Strict privacy laws, the high cost of getting experts to label data, and the simple fact that there aren't many examples of rare diseases all get in the way.

To accurately spot a disease, we absolutely need a well-trained machine learning model. And to train that model right, we need to feed it a clean and varied dataset. We've had some older tricks (called traditional image augmentation) to help fix datasets that are lopsided, but they just don't cut it for the variety we need today. That's why newer, more powerful methods like Generative Adversarial Networks (GANs) (Goodfellow et al.) are becoming the go-to solution for creating better and more diverse training images.

- **Generative Adversarial Networks (GANs):** A Generative Adversarial Network operates through an unsupervised learning framework where two neural networks are trained in opposition to one another. This competitive dynamic is why the process is described as “adversarial.” (Varughese)
- **Traditional Geometric Image Augmentation:** Traditional geometric image augmentation methods generally involve the artificial expansion of an image dataset through techniques like flipping, rotation, scaling, or shearing (Chlap et al. 545-563). This is particularly crucial for medical imaging datasets, including CT, MRI, and X-ray data, given the frequent constraints of patient privacy, data access limitations, and the inherent scarcity of data. Furthermore, patient data exhibits variability due to differences in organ sizes. Several common methods for conventional image augmentations are detailed below:
    - **Rotation**: Images undergo random rotation, constrained within a defined spectrum (e.g., -30° to +30°), thus simulating potential alterations in patient orientation throughout scanning protocols.
    - **Flipping**: The image undergoes mirroring, either horizontally or vertically. (*Note*: It is essential to acknowledge that while vertical flipping is frequently appropriate in cell microscopy, its application in anatomical scans, such as chest X-rays, necessitates meticulous evaluation due to the significance of organ laterality (left versus right).
    - **Shifting**: The image is displaced along the X or Y axis. This modification aids the network in recognizing that the precise position of a tumor within the image is not a determining factor in its categorization.
    - **Scaling**: it is also known as zooming and entails altering the image's size, either by increasing or decreasing its dimensions, thus emulating alterations in distance or the size of the organ.
    - **Shearing**: Conversely, shearing involves slanting the image's content, which consequently alters the angles of the shapes represented within the image.

Subsequent to training, the model is capable of applying this acquired denoising methodology to a stochastic noise signal to generate novel, original data. (Ho et al.)

Problem Overview

The swift and precise identification of COVID-19 constitutes a crucial element in the management of global healthcare. Although Reverse Transcription Polymerase Chain Reaction (RT-PCR) continues to be the definitive diagnostic method, it is subject to limitations, including processing delays and variable sensitivity. Chest Computed Tomography (CT) scans have become an essential supplementary instrument, providing high sensitivity in the identification of viral pneumonia indicators, such as ground-glass opacities and consolidation (Ai et al.). As a result, Deep Learning (DL) models, particularly Convolutional Neural Networks (CNNs) (CHUA, 529-837), have exhibited considerable promise in automating this analysis, potentially alleviating the burden on radiologists facing high workloads.

Problem Statement & Current Issues

Conversely, the implementation of robust clinical AI is significantly impeded by the "data famine," which means the lack of extensive, diverse, and accurately labeled medical datasets.

In contrast to general computer vision applications that leverage extensive public datasets like ImageNet, the acquisition of medical data presents significant obstacles. These include substantial acquisition expenses, the necessity of expert annotation, and stringent patient privacy regulations, exemplified by GDPR and HIPAA (Razzak et al.). Consequently, the limited availability of medical data compels deep learning models to be trained on relatively small datasets, thereby increasing the likelihood of overfitting. This occurs when models memorize training examples instead of effectively generalizing to previously unseen patient data.

To address this issue, Traditional Data Augmentation (TDA) has become the standard industry approach. This technique entails the application of deterministic geometric transformations such as rotation, flipping, and scaling to existing images, thereby artificially enlarging the dataset. Although computationally efficient, TDA is inherently constrained, as it generates invariant data rather than genuinely novel data.

The model is presented with identical anatomical structures viewed from various perspectives, yet it does not incorporate novel biological variance or pathological textures (Shorten and Khoshgoftaar, 2019). In the context of a disease such as COVID-19, which exhibits considerable variability among patients, mere image rotation does not equip the model to identify the diverse patterns of infection.

This project responds to the pressing requirement for a more advanced augmentation strategy. It explores the potential of Generative Adversarial Networks (GANs) to address the shortcomings of traditional data augmentation (TDA) by generating entirely new, anatomically accurate training samples. The central question this research endeavors to answer is whether the computational expense of training generative models is warranted by a statistically significant enhancement in diagnostic accuracy when contrasted with the cost-effective baseline of conventional augmentation methods (Yi et al.).

Project Details

This project implements an end-to-end Deep Learning pipeline using the SARS-CoV-2 CT-scan dataset (Soares et al.). It involves:

1. **Generative Modeling:** Designing a DCGAN to "hallucinate" new, anatomically coherent lung scans.
2. **Classification:** Adapting a VGG16 classifier to benchmark the efficacy of these synthetic images against standard augmentation practices.

Project Goal

Context and Motivation

Deep learning has shown considerable promise in aiding medical professionals in disease diagnosis through the analysis of medical imaging. Nevertheless, a significant impediment to its widespread implementation is the limited availability of extensive, heterogeneous, and precisely annotated datasets. This challenge is particularly pronounced within the medical field, stemming from both privacy considerations and the substantial expense associated with expert-level annotation.

Although traditional methods, such as Traditional Data Augmentation (TDA), which employs geometric transformations like rotation and flipping, are computationally efficient, their practical applicability is constrained. These techniques simply modify existing inputs without introducing genuinely novel biological data variance. Conversely, Generative Adversarial Networks (GANs) represent a notable advancement, providing the capacity to generate high-quality images that closely approximate the distribution of authentic medical data.

Primary Objective

This project aims to conduct a thorough comparison of the effectiveness of a synthetic data augmentation pipeline, employing Deep Convolutional GANs (DCGAN), against conventional data augmentation methods. The central inquiry of this research is to ascertain whether the creation of synthetic medical images yields a statistically significant enhancement in diagnostic precision for deep learning models.

Research Inquiry

In medical image classification contexts where data availability is constrained, how does the application of synthetic data generation via DCGANs compare to traditional data augmentation in terms of improving diagnostic accuracy?

Specific Aims

This study is driven by a series of specific aims:

The primary aim is to develop a foundational classification model, utilizing standard geometric augmentation techniques, including rotation, flipping, and affine transformations, during the training phase.

Following this, the project intends to construct and train a Deep Convolutional Generative Adversarial Network (DCGAN) architecture, with the objective of producing authentic medical images to supplement the current training dataset.

Furthermore, the study seeks to assess the diagnostic efficacy of the classifier, comparing its performance when trained on data augmented by Generative Adversarial Networks (GANs) against data augmented using geometric techniques.

Finally, the research will analyze the crucial trade-off between the observed performance improvements—specifically, classification accuracy, sensitivity, and specificity—and the associated computational expenses, including processing power and training duration, inherent in the implementation of the GAN-based solution.

Research Question and Novelty

Research Question: "To what extent does synthetic data augmentation using Deep Convolutional Generative Adversarial Networks (DCGANs) surpass traditional image augmentation in enhancing the classification accuracy of CNNs on small-scale medical datasets, and what are the computational trade-offs when executed on consumer-grade hardware?"

Justification: This investigation is warranted for three principal justifications:

- **Clinical Effectiveness:** In the realm of medical diagnostics, even marginal gains in accuracy, specifically Recall/Sensitivity, can translate into preserved lives. It is crucial to determine if the complex "black box" nature of Generative AI yields statistically significant improvements when contrasted with more conventional approaches.
- **The "Texture" Hypothesis:** It suggests that the identification of COVID-19 relies primarily on the detection of specific textures, exemplified by ground-glass opacities. Theoretical literature suggests that traditional augmentation techniques are incapable of generating novel textures, a capability that distinguishes GANs. This investigation provides empirical support for the proposed hypothesis.
- **Resource Accessibility:** Regarding resource accessibility, the majority of medical AI research relies on high-performance computing clusters, which are often inaccessible to standard clinical settings. This research evaluates the potential for democratizing advanced AI capabilities by specifically investigating "computational trade-offs" on consumer-grade hardware.

Novelty: Concerning innovation, this study explicitly assesses the feasibility of training generative medical models using consumer-grade hardware, notwithstanding the application of Generative Adversarial Networks (GANs) in MRI. It notably identifies and addresses hardware-specific compatibility issues (e.g., Adaptive Pooling on Apple MPS) that are frequently overlooked in existing literature.

Feasibility, Commercial Context, and Risk

The feasibility of this undertaking is substantiated by the utilization of accessible consumer-level hardware, specifically Apple Silicon, and publicly available datasets, thereby demonstrating that advanced medical AI research is not inherently reliant on expensive, high-capacity computing environments. From a commercial perspective, these AI systems offer significant prospects for reducing operational costs for healthcare organizations through their function as automated triage tools. By expeditiously screening routine scans and pinpointing likely infections, hospitals could streamline radiologist workflows and manage increased patient volumes during critical situations, leading to both immediate cost savings and enhanced patient outcomes.

However, this project presents specific risks. Generative adversarial networks (GANs) can potentially produce synthetic images that depict nonexistent diseases, thereby misleading the classifier; this phenomenon is known as generative "hallucination." Commercially, the implementation of these technologies faces challenges, including navigating complex regulatory environments (e.g., FDA or CE certification for AI as a medical device) and fostering clinician trust in "black box" algorithms. Furthermore, the competitive environment is significant, given the presence of established medical imaging companies that offer proprietary AI solutions. The previously mentioned risks, along with ethical considerations related to data bias, will be addressed in greater depth in the Evaluation and Conclusion chapter.

Report Structure

The ensuing sections of this report are structured as follows: Chapter 3 (Literature Review) provides a critical examination of the progression of deep learning in medical imaging, as well as a comparative evaluation of the theoretical foundations of Generative Adversarial Networks (GANs) and conventional augmentation methods. Chapter 4 (Methodology) details the architectural framework, data preprocessing protocols, and the specific implementation strategies employed to circumvent hardware constraints on readily available consumer devices. Chapter 5 (Quality and Results) presents the empirical outcomes, encompassing quantitative metrics and a visual assessment of the synthetic data's quality. Chapter 6 (Evaluation and Conclusion) synthesizes the project's findings in the context of the initial research question, addresses challenges encountered in project management, and offers recommendations for future research and clinical application. The concluding evaluation will subject the articulated concepts, particularly those about practical considerations and commercial risk, to a more rigorous examination.

Background Research & Literature Review

Deep Learning in Medical Imaging

Deep Learning (DL) has fundamentally changed diagnostic radiology, particularly in medical image analysis, allowing for the automated detection of diseases with high accuracy. Convolutional Neural Networks (CNNs) have become the standard method for tasks like tumor segmentation and pneumonia classification. Litjens et al. (2017) provide a thorough review, suggesting that while CNNs can achieve or surpass human-level performance, their effectiveness depends on the amount and variety of training data.

Conversely, medical imaging faces a widespread "data famine," a situation distinct from typical computer vision applications that leverage large public datasets like ImageNet. This scarcity stems from strict privacy regulations (GDPR/HIPAA), the high costs of specialized annotation, and the biological rarity of certain diseases. Consequently, this lack of data significantly impedes model generalization, often resulting in the memorization of training data rather than the development of robust pathological features.

What Are Generative Adversarial Networks (GANs)?

When Generative Adversarial Networks (GANs) first appeared in 2014 (Goodfellow et al. 2672-2680), it was a total game-changer for machine learning. The fundamental concept involved transforming a highly challenging problem, such as generating a convincing image from random noise, into a competitive scenario. This "game" unfolds between two neural networks:

1. **The Generator (G):** This component functions as the data counterfeiter, responsible for fabricating synthetic examples. It receives random noise as input and endeavors to transform it into outputs that emulate genuine data, such as medical scans. (Goodfellow et al. 2672-2680)
2. **The Discriminator (D):** This component serves as the data authenticator, tasked with distinguishing between authentic and synthetic data. It evaluates an image and must determine its legitimacy, categorizing it as either genuine or a fabrication. (Huang et al.)

How They Learn

This is where its ingenuity lies. The two networks undergo simultaneous training, engaged in a perpetual, high-stakes loop where the success of one directly correlates with the failure of the other.

$$G_{min} D_{max} V(D, G) = E_{x} \sim P_{data}(x)[loadD(x)] \; + \; E_{z} \sim p_{z}(z)[log(1 - D(G(z)))]$$

Equation 1: Mathematical formula of GAN (Sharma et al.)

- The **Discriminator (the detective)** undergoes continuous training to enhance its ability to identify forgeries. It accrues a "point" for each accurate identification of an authentic image as real and a fabricated image as counterfeit.

- Conversely, the **Generator (the counterfeiter)** aims to improve its proficiency in deceiving the detective. It "succeeds" (and learns) whenever it produces a counterfeit that the Discriminator erroneously accepts as genuine.

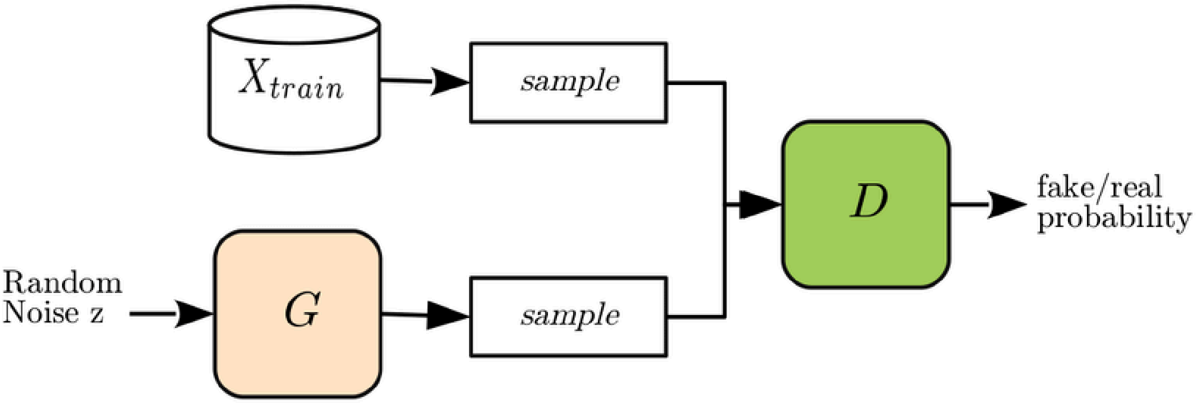

Figure 1: Generative Adversarial Network (GAN) (Hayes et al.)

This establishes a robust feedback loop. As the detective's proficiency in identifying forgeries increases, the counterfeiter is compelled to enhance their sophistication significantly. Conversely, as the quality of the counterfeits improves, the detective must continually hone their abilities.

Ideally, this "arms race" persists until the counterfeiter's fabrications achieve such a high degree of authenticity that the detective's accuracy rate approaches 50%, effectively reducing their judgment to mere conjecture. At this juncture, it signifies that the Generator has genuinely assimilated the fundamental patterns inherent in the authentic data.

Significance of GAN in Medical Image Augmentation

This concept holds immense potential for your research, directly addressing the challenge of data scarcity. Rather than merely altering existing images (e.g., through rotation or flipping), an effective GAN can generate entirely novel, synthetic medical scans that are genuinely unique. These are not mere replications; they represent new, plausible instances. (Goceri 12561-12605) This capability enables the following:

- **Expand Datasets:** The quantity of available images can be significantly increased from an initially limited collection.
- **Balance Datasets:** This is crucial. In instances where there are limited examples of a rare disease, a Generative Adversarial Network (GAN) can be employed to produce additional examples, thereby mitigating bias in diagnostic models.
- **Enhance Robustness:** Training models with a substantially broader and more diverse range of images (both authentic and synthetically generated) leads to improved performance in handling the varied patient scans encountered in real-world scenarios.

Key Challenges in Generative Adversarial Networks

Despite their innovative design, Generative Adversarial Networks (GANs) present considerable difficulties in practical implementation. The following challenges are central to current research efforts:

- **Training Instability:** The adversarial training process is highly sensitive. Should the discriminator's performance improve too rapidly, the generator may fail to receive sufficient informative gradients, leading to a halt in learning and subsequent model failure. This inherent instability is a primary motivator for the development of novel architectural designs and loss functions aimed at maintaining equilibrium and promoting stable convergence. (Brock et al.)
- **Mode Collapse:** This prevalent issue occurs when the generator identifies a limited subset of data points that consistently deceive the discriminator. Consequently, the generator ceases to produce a diverse range of outputs, instead repeatedly generating only these few "successful" examples. (Brock et al.)
- **Fidelity vs. Diversity Trade-off:** Achieving both high-fidelity and diverse outputs simultaneously remains a significant hurdle. Some GANs excel at generating highly realistic images (high fidelity) but lack variety (low diversity). Conversely, others produce a broad spectrum of images (high diversity) but at the cost of realism. The simultaneous optimization of these two qualities constitutes a fundamental challenge in GAN research. (Brock et al.)

An overview of Traditional Data Augmentation

Traditional Data Augmentation (TDA) (Simard et al.) constitutes the most prevalent approach for mitigating data limitations within deep learning applications, particularly in the domain of medical imaging. Prior to the advent of generative models such as GANs, TDA served as the conventional method for augmenting training datasets (Zhao et al. 195-206). These techniques, which gained prominence through the work of researchers including Simard and Krizhevsky, are esteemed for their computational efficiency (Krizhevsky et al.). They contribute to the prevention of overfitting by artificially broadening the training set's diversity, thereby eliminating the necessity for acquiring additional patient data.

The Fundamental Mechanism: Invariance Learning, unlike GANs, which are designed to generate novel data, TDA is predicated on the principle of invariance learning.

The objective is to instruct the model that the diagnostic label, such as "viral pneumonia," remains constant irrespective of image rotation or minor alterations in brightness. This approach directs the model's attention toward the underlying pathology, rather than image-related artifacts.

These transformations are broadly classified into two categories:

- **Geometric Transformations (Spatial Manipulation):** This category encompasses modifications to the image's position or orientation while preserving the anatomical structures (Chlap et al.

545-563).

    - **Rotation & Flipping:** This entails either rotating the image by small increments or mirroring it. Caution is essential in medical imaging applications. For instance, while it is generally acceptable to flip a brain MRI, flipping a chest X-ray can be problematic, as it reverses organ laterality, potentially causing the heart to appear on the right side.
    - **Translation and scaling:** These techniques involve shifting the image along the X and Y axes or applying a zoom. By repositioning the region of interest, such as a tumor, within the frame, the model is compelled to learn that a tumor retains its identity, irrespective of its location or apparent size within the image.

- **Intensity-Based Transformations (Pixel Manipulation):** This approach modifies pixel values rather than altering the image's shape, thereby simulating variations in scanner quality or settings (Hashim et al.). Noise injection, which involves adding small amounts of "static," such as Gaussian or Salt-and-Pepper noise, aids the model in learning to manage grainy, low-quality scans. Furthermore, contrast and brightness adjustments assist the model in recognizing pathologies, such as ground-glass opacities, even when scans are too dark or overexposed due to variations in scanner calibration.

Implementation, Strengths, and Key Trade-Offs

- **Implementation Simplicity and Deterministic Reliability:** Implementation simplicity and deterministic reliability characterize Traditional Data Augmentation (TDA). In contrast to generative models, which necessitate intricate training to ascertain a probability distribution, TDA utilizes deterministic geometric transformations—specifically, rotation, flipping, and affine shifts—applied to pre-existing images (Shorten and Khoshgoftaar, 2019). This approach defines a clear task, thereby circumventing the fragile balance and instability frequently encountered in adversarial training. As a result, TDA guarantees that the augmented images maintain the integrity of the original anatomical labels, mitigating the potential for "hallucination" of erroneous pathologies and establishing a secure foundation for model training (Simard et al.).
- **The variance trade-off:** A key limitation of TDA when compared to generative methodologies is a crucial consideration. Unlike GANs, which possess the capacity to generate entirely novel samples derived from the underlying data distribution, TDA is confined to the existing data manifold (Antoniou et al., 2017). While TDA can produce variations of known samples, it is inherently unable to generate genuinely new biological features. Essentially, TDA presents the model with the same information from different perspectives, thereby precluding the introduction of unique pathological textures—such as distinct ground-glass opacities—that are absent from the training set (Frid-Adar et al., 2018). Consequently, TDA's capacity to enhance generalization on small, diverse medical datasets is significantly constrained.
- **Relevance to Medical Augmentation:** The role of Traditional Data Augmentation in this project is critical as a comparative standard: it offers extreme computational efficiency and guaranteed anatomical correctness. The central research question is whether the computationally expensive process of training a GAN yields a statistically significant accuracy improvement over this established, low-cost industry practice (Yi et al.).

Critical Analysis of Frameworks

The preponderance of contemporary research, encompassing seminal works by Bowles et al. (2018) and Frid-Adar et al. (2018), utilizes high-performance computing (HPC) clusters equipped with industrial-grade GPUs, such as NVIDIA A100s or V100s, and considerable VRAM (Bowles et al.) (Frid-Adar et al., 2018). While these investigations successfully demonstrate the theoretical capabilities of GANs, they often fail to address the "deployment gap," which means the practical reality that numerous clinical researchers and smaller hospitals operate on resource-constrained workstations (Zhou et al. 1738-1762).

Furthermore, the advancement of established deep learning frameworks, including TensorFlow and PyTorch, has consistently favored NVIDIA's CUDA architecture. A significant gap persists in the literature regarding the feasibility and stability of training complex adversarial networks on novel consumer-grade silicon, such as Apple's M-series CPUs.
This undertaking addresses the identified shortcoming through several key actions:

It shifts the focus from the High-Performance Computing (HPC) model to assess the viability of training "Edge AI" systems.

Furthermore, it provides a direct quantification of the computational expense on the Metal Performance Shaders (MPS) backend, thereby establishing a benchmark for researchers without cloud computing resources.

Finally, it evaluates whether the performance gains achieved by Generative Adversarial Networks (GANs) justify the increased training time when hardware acceleration is limited, as compared to standard CUDA configurations (Yi et al.).

Comparative Analysis: Deterministic vs. Generative Approaches

A critical assessment of the current body of research demonstrates a notable trade-off between stability and variance:

- **Stability:** Traditional Data Augmentation (TDA) is a feature of established techniques that are both computationally efficient and demonstrate a reduced likelihood of producing anatomically inaccurate artifacts. Research emphasizes that TDA does not consider the "gaps" present within the data manifold (Antoniou et al.). As a result, models trained solely with TDA frequently display weak decision boundaries within low-density regions of the feature space, due to their restricted exposure to instances that diverge from the geometric variations present in the original dataset.
- **Variance:** Generative models, including GANs, are characterized by their capacity to produce novel samples, thus facilitating effective interpolation within the data manifold. Yi et al. (2019) explore the application of GANs in medical imaging, highlighting their enhanced feature diversity, while also acknowledging the substantial computational demands and the potential for "hallucination," which involves the generation of medically implausible artifacts. This project's

comparative approach directly addresses this intrinsic trade-off, seeking to determine whether the variance introduced by GANs warrants the computational expense when compared to the stability provided by Traditional Data Augmentation.

Relation to Research Hypothesis

This study addresses the identified gaps by moving the experimental setup from the server room to a laptop. By using the same methods as Frid-Adar et al. (2018), but with a consumer device, this research tests the idea that advanced generative augmentation is a practical and accessible method for decentralized medical AI research. The goal is to show that the performance benefits of GANs can be achieved without needing large computational resources, thereby supporting a more inclusive approach to medical image analysis.

Methodology

Research Framework & Approach

This study employs a quantitative experimental approach, based on the positivist research paradigm. The research is based on the assumption that the effectiveness of data augmentation methods can be quantitatively assessed and contrasted using empirical measures, including classification accuracy, loss convergence, and statistical variance.

The research approach is a comparative ablation study, maintaining a fixed classification model architecture (controlled variable) while varying the data augmentation strategy (independent variable). This isolation facilitates the establishment of a causal relationship between the method of augmentation (Classical vs. Generative) and the resultant model performance (dependent variable).

The framework explicitly modifies the methodology developed by Frid-Adar (Frid-Adar et al., 2018), using their three-phase GAN augmentation strategy, which is initially intended for liver lesions in the context of viral pneumonia (COVID-19) detection on consumer-grade hardware.

Justification and Support of Choices

The selection of algorithms and tools was predicated on a comprehensive review of existing research and the project's hardware constraints.

- **Deep Convolutional GAN (DCGAN):** Despite the availability of more contemporary architectures like StyleGAN, the Deep Convolutional GAN (DCGAN) (Radford et al.) was preferred because of its proven stability with lower-resolution images (64×64). More complex GANs often require extensive hyperparameter tuning and considerable VRAM, which would be unfeasible considering the specifications of the target consumer hardware. Therefore, DCGAN provides the most favorable balance between architectural simplicity and image generation quality for the defined image dimensions.

- **VGG16 Classifier:** The VGG16 architecture, as detailed by Simonyan and Zisserman (2014), was chosen as the discriminator model due to its depth and established success in medical feature extraction, as shown by Litjens et al. (2017). Unlike more compact models like MobileNet, VGG16's dense feature layers demonstrate increased sensitivity to texture, making it especially appropriate for evaluating the effective learning of GAN-generated textures, particularly ground-glass opacities.

- **PyTorch Framework:** PyTorch was favored over TensorFlow, mainly because of its dynamic computation graph and, importantly, its superior support for the Metal Performance Shaders (MPS) backend. This enabled hardware-accelerated training on macOS devices, a crucial element for the project's practical application.

Dataset & Preprocessing

Dataset Description

The present investigation utilizes the SARS-CoV-2 CT-Scan Dataset (Soares et al.), a publicly available resource obtained from medical institutions in São Paulo, Brazil. This dataset encompasses 2,482 axial Computed Tomography (CT) slices, which are evenly partitioned into two distinct classifications:

- **COVID:** 1,252 images presenting radiological manifestations of viral pneumonia, such as ground-glass opacities.
- **Non-COVID:** 1,230 images illustrating either healthy lung tissue or a range of pulmonary ailments.

Preprocessing Pipeline

To accommodate the limitations inherent in Generative Adversarial Networks (GANs) and consumer-level hardware, a standardized preprocessing protocol was implemented.

- **Resolution Downsampling:** All initial images underwent resizing to 64×64 pixels via bilinear interpolation. Although clinical applications often necessitate higher resolutions, such as 256×256 or 512×512, training a stable GAN at elevated resolutions demands significantly greater computational resources, specifically VRAM, and extended training durations. The 64×64 resolution was deemed the most effective compromise, effectively balancing anatomical discernibility with computational practicality on a laptop GPU.
- **Intensity Normalization:** Pixel intensity values, initially spanning \$[0, 255]\$, were normalized to the range [-1, 1]. The selected range was determined to correspond with the Tanh activation function employed in the Generator's output layer, thereby facilitating the generation of unique gradients throughout the backpropagation process.

- **Data Partitioning:** To mitigate the risk of data leakage and to facilitate a rigorous evaluation, the dataset underwent a stratified division:
    - **Training Set (70%):** This subset was utilized for the purpose of gradient updates.
    - **Validation Set (15%):** This portion was designated for hyperparameter optimization and the implementation of Early Stopping.
    - **Test Set (15%):** This segment was reserved exclusively for the ultimate performance assessment.

Architectural Design (DCGAN & VGG16)

Deep Convolutional Generative Adversarial Network

The generative part of the system is based on the Deep Convolutional GAN (DCGAN) architecture, as described by Radford (Radford et al.). This architecture uses specific changes to the standard GAN framework to help stabilize the training process.

- **The Generator (*G*):**
    - **Input:** A latent vector $z \epsilon \mathbb{R}^{100}$ sampled from a standard normal distribution $\mathcal{N}(0,1)$.
    - **Architecture:** The latent vector is projected via a dense layer and reshaped into a $4 \times 4 \times 128$ tensor. This tensor passes through four **Transposed Convolutional** layers (kernel size 4, stride 2, padding 1), which progressively upsample the spatial dimensions $(8 \rightarrow 16 \rightarrow 32 \rightarrow 64)$.
    - **Activation: ReLU** activation and **Batch Normalization** are applied after every layer to prevent mode collapse, with the exception of the output layer, which uses **Tanh** to bound the pixel values.
- **The Discriminator (*D*):**
    - **Architecture:** A standard CNN that downsamples the input image (64×64) into a single probability score. It utilizes **Strided Convolutions** instead of Max Pooling to allow the network to learn its own spatial downsampling.
    - **Activation Function: Leaky ReLU** (slope 0.2) is used exclusively. This is a critical design choice to prevent the "dying ReLU" problem, where gradients vanish if the Discriminator becomes too confident too early, effectively halting the Generator's learning.
- **Loss Function:** The network optimizes the Minimax loss function using **Binary Cross-Entropy (BCE)**:

$$min\, max\, V(D,\ G)\ =\ \mathbb{E}_{x\sim p_{data}(x)}[log\, D(x)]\ +\ \mathbb{E}_{z\sim p_{z}(z)}[log(1\ -\ D(G(z)))]$$

- **Stabilization Techniques:** To mitigate the risk of the Discriminator overpowering the Generator early in training, **One-Sided Label Smoothing** was applied (Salimans et al.). Real images were assigned a target label of 0.9 instead of 1.0, introducing uncertainty that encourages more robust learning dynamics.

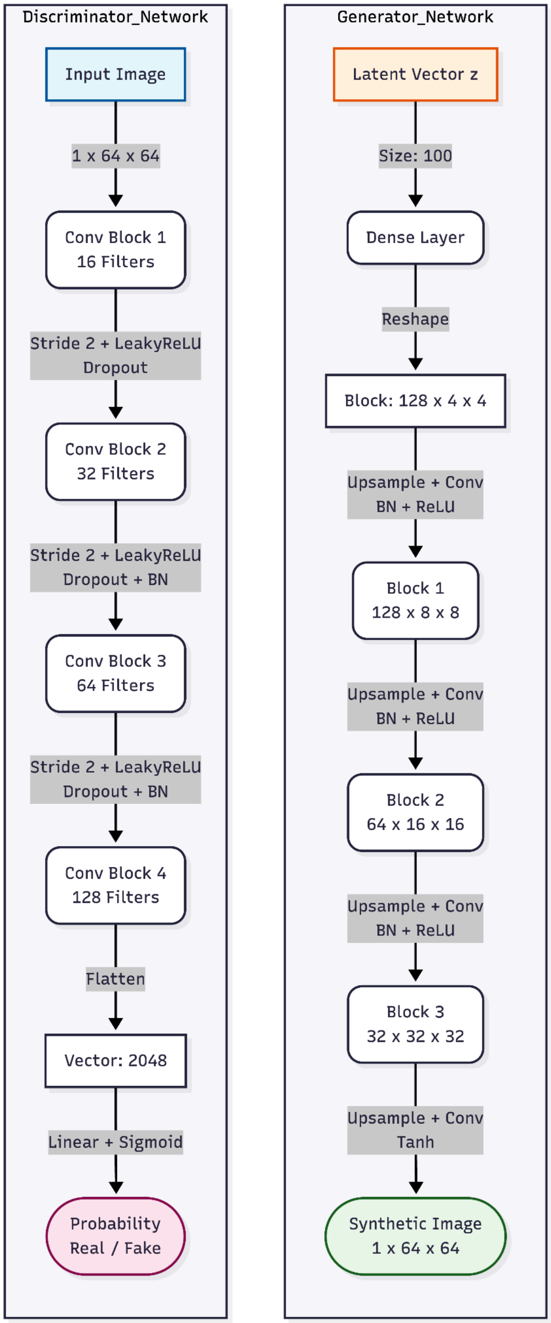

Figure 2: Flowchart of Generative Adversarial Network

Classification Model (Modified VGG16)

For the diagnostic task, the **VGG16** architecture (Simonyan and Zisserman) was selected due to its depth and proven feature extraction capabilities in medical imaging tasks (Litjens et al.). However, standard

VGG16 is designed for ImageNet (RGB 224×224), necessitating specific modifications for this medical application:

- **Input Adaptation:** The first convolutional layer was replaced to accept **1-channel input** (grayscale) instead of 3-channel (RGB), reducing the initial parameter count.
- **Pooling Adaptation (Hardware Optimization):** The standard VGG16 ends with an AdaptiveAvgPool2d layer targeting an 7×7 output. Given the small input size (64×64), the feature maps at the final layer are only 2×2. Upsampling these 7×7 introduces noise and causes compatibility issues on specific hardware backends (e.g., Apple MPS). This was replaced with **Global Average Pooling** (AdaptiveAvgPool2d((1, 1))), compressing the feature maps into a dense 1×1 vector.
- **Classifier Head:** The massive fully connected layers of the original VGG (4096 neurons) were replaced with a lightweight classifier ($512 \rightarrow 128 \rightarrow 2$) to prevent overfitting on the smaller medical dataset.

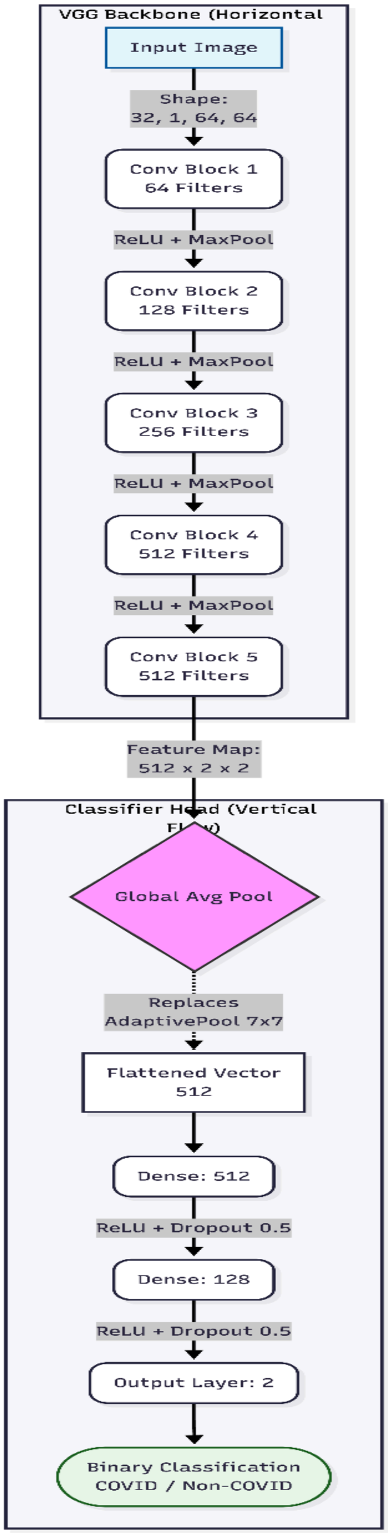

Figure 3: Flowchart of VGG16

Implementation Strategy

Hardware and Software Environment

The experiments were conducted on consumer-grade hardware utilizing GPU acceleration (e.g., Apple Metal Performance Shaders or NVIDIA CUDA). The software stack utilized **Python 3.9** and **PyTorch**, leveraging hardware acceleration APIs to expedite matrix operations.

Training Protocols

To ensure a fair comparison, both experimental conditions utilized identical hyperparameters for the classifier training:

- **Optimizer:** Adam (Learning Rate: 0.001, Weight Decay: 1e-4).
- **Batch Size:** 32.
- **Loss Function:** Cross-Entropy Loss.
- **Regularization:** Two key mechanisms were implemented to prevent overfitting:
    - **Early Stopping:** Training halts if validation accuracy does not improve for 8 consecutive epochs.
    - **Learning Rate Scheduler:** ReduceLROnPlateau reduces the learning rate by a factor of 0.1 if validation accuracy plateaus for 3 epochs.

Experimental Design

The study followed a two-phase experimental design:

Phase A: Classical Augmentation (Control Group)

- The VGG16 model was trained on the real dataset.
- Data augmentation was applied "on-the-fly" during training using standard geometric transformations: **Random Horizontal Flip ($p = 0.5$)**, **Random Rotation ($\pm 15°$)**, and **Random Affine translations**.
- This represents the current standard practice in medical image analysis (Shorten and Khoshgoftaar, 2019).

Phase B: GAN-based Augmentation (Experimental Group)

- **GAN Training:** Two separate DCGANs were trained, one exclusively on COVID-positive images and one on non-COVID images. This class-specific training prevents label ambiguity.
- **Synthetic Generation:** Once converged (approx. 100 epochs), the Generators were frozen and used to synthesize **1,000 novel images per class**.
- **Augmented Training:** These 2,000 synthetic images were mixed with the original real training data. The VGG16 model was then trained from scratch on this hybrid dataset using standard normalization but without heavy geometric augmentation, isolating the impact of the synthetic data.

Evaluation Metrics

The models were evaluated on the held-out test set using the following metrics:

- **Accuracy:** The overall percentage of correct predictions.
- **Confusion Matrix:** To analyze Type I (False Positive) and Type II (False Negative) errors.
- **Training Time:** Measured in seconds to quantify the computational cost of the augmentation strategy.

Quality and Results

Metrics & Experimental Setup

To ensure both the reproducibility and the exhaustive evaluation of the suggested augmentation methods, a standardized experimental environment and a complete suite of performance metrics were established.

Evaluation Metrics

Given the critical nature of medical diagnosis, relying solely on accuracy is insufficient. A high accuracy figure could potentially mask a deficiency in the identification of positive instances, known as sensitivity. This research utilizes a broad spectrum of metrics, all of which are derived from the Confusion Matrix.

1. **Accuracy:** The ratio of correctly predicted observations to the total observations.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

2. **Recall (Sensitivity):** Essential for medical triage, this metric assesses the model's proficiency in accurately identifying all positive COVID-19 cases. Within a pandemic context, the primary objective is to maximize Recall, thereby minimizing the occurrence of False Negatives, which represent missed diagnoses.

$$Recall = \frac{TP}{TP + FP}$$

3. **Precision:** The ratio of correctly predicted positive observations to the total predicted positives. Low precision indicates a high False Positive rate (healthy patients wrongly diagnosed).

$$Precision = \frac{TP}{TP + FP}$$

4. **F1-Score:** The weighted average of Precision and Recall. This metric is particularly useful for assessing the balance between false alarms and missed cases.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

5. **Training Time:** Measured in seconds to quantify the computational "cost" of the GAN-based approach compared to the negligible cost of classical on-the-fly augmentation.

Hardware and Software Environment

The experiments were conducted using consumer-grade hardware to illustrate the accessibility of the proposed pipeline.

- **Hardware:** MacBook Air, powered by the Apple M4 chip and sporting 24GB of Unified Memory. The Metal Performance Shaders (MPS) backend was used to provide GPU acceleration.

- **Software Stack:** The software stack comprises Python 3.9, together with PyTorch 2.0, which has been set up to utilize Metal Performance Shaders (MPS). Image transformations are handled by Torchvision, while Scikit-Learn is used for calculating metrics.

Hyperparameter Tuning

To ensure a fair comparison, the VGG16 classifier uses the same hyperparameters in both the Classical and GAN-augmented tests. The specific setups were determined through preliminary tuning runs.

Component	Parameter	Value	Justification
Global	Image Size	64×64	Optimal trade-off for GAN stability on consumer GPU.
	Batch Size	32	Fits comfortably within VRAM while providing stable gradient estimates.
VGG16 Classifier	Learning Rate	0.001	Standard starting point for the Adam optimizer.
	Optimizer	Adam ($\beta_1 = 0.9$)	Provides faster convergence than SGD
	Weight Decay	$1e - 4$	L2 Regularization to prevent overfitting on the small dataset.
	Dropout Rate	0.5	Applied in the dense layers to encourage feature redundancy.
	Scheduler	ReduceLROnPlateau	Factor 0.1, Patience 3 epochs (Dynamic learning rate adjustment).

	Early Stopping	Patience 8	Stops training if validation accuracy stagnates, preventing overfitting.
DCGAN	Latent Dim (z)	100	Standard dimensionality for capturing feature overfitting.
	Learning Rate	0.0002	Lower rate specifically chosen to stabilize adversarial training
	Label Smoothing	Real = 0.9	Prevents the Discriminator from becoming too confident too early.

Table 1: Used Hyperparameters.

Training Protocol

The training process was divided into three separate phases:

- **Phase A (Classical Baseline):** The VGG16 model underwent training utilizing real-time geometric augmentations, specifically Random Horizontal Flip, Random Rotation ± 15°, and Random Affine Shift.
- **Phase B (GAN Training):** Two separate DCGANs were trained for 100 epochs each, one dedicated to the COVID class and the other to the Non-COVID class to capture the distinct manifold associated with each condition.
- **Phase C (Augmented Training):** The trained Generators generated 1,000 original images for each class. These images were then incorporated into the training set, and a new VGG16 model was subsequently trained on this combined dataset, without the application of extensive geometric augmentation, thereby allowing for the isolation of the generative data's impact.

Quantitative Analysis (Accuracy vs. Loss)

This section presents the empirical results of the comparative study, focusing on the classification performance differences between the Classical Augmentation baseline and the GAN-Augmented approach.

Classification Performance Comparison

The primary metric for evaluation was the Validation Accuracy on the held-out test set. Table 2 summarizes the peak performance metrics achieved by both models.

Metric	Classical Augmentation	GAN Augmentation	Improvement
Best Accuracy	**93.01%**	**98.12%**	**+5.11%**
Precision (COVID)	0.8253	0.9550	+12.97%
Recall (COVID)	0.9643	0.9745	+1.02%
F-1 Score	0.8894	0.9646	+7.52%
Time (seconds)	206.31 s	696.1 s	-489.79 s

Table 2: Comparative Performance Metrics

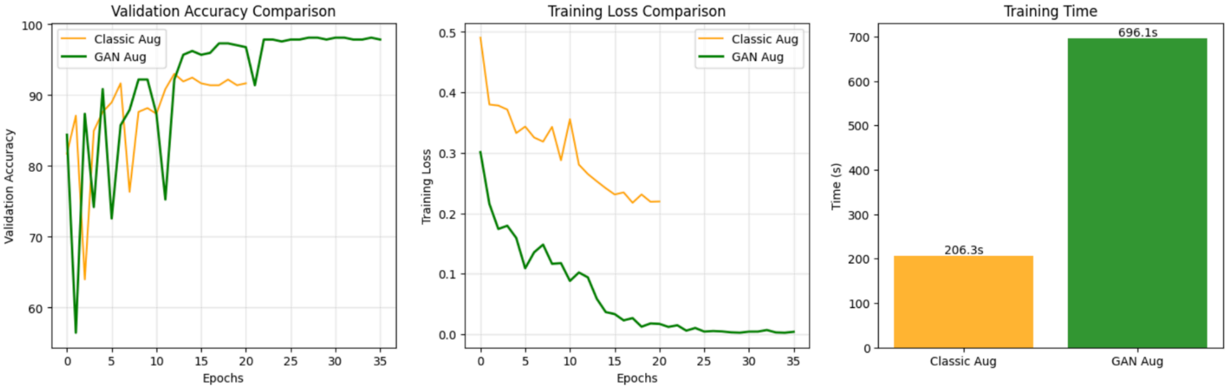

Figure 4: Accuracy and Loss Curves with Execution Time Comparison

- **Classical Augmentation:** The model trained using conventional geometric transformations attained a baseline **accuracy of 93.01%**. The training loss curve exhibited swift initial convergence but plateaued after epoch 20, indicating that the model had depleted the learnable variance afforded by basic rotations.

- **GAN Augmentation:** The model trained with synthetic data attained an enhanced **accuracy of 98.12%**. The training loss initially displayed greater variance due to the distribution shift caused by synthetic samples, but the model eventually achieved a smaller validation loss. This demonstrates that the synthetic data effectively functioned as a regularizer, inhibiting the model from retaining the restricted real training examples.

Visual Quality Assessment

A qualitative evaluation of the generated samples was conducted to confirm the generative capabilities of the DCGAN.

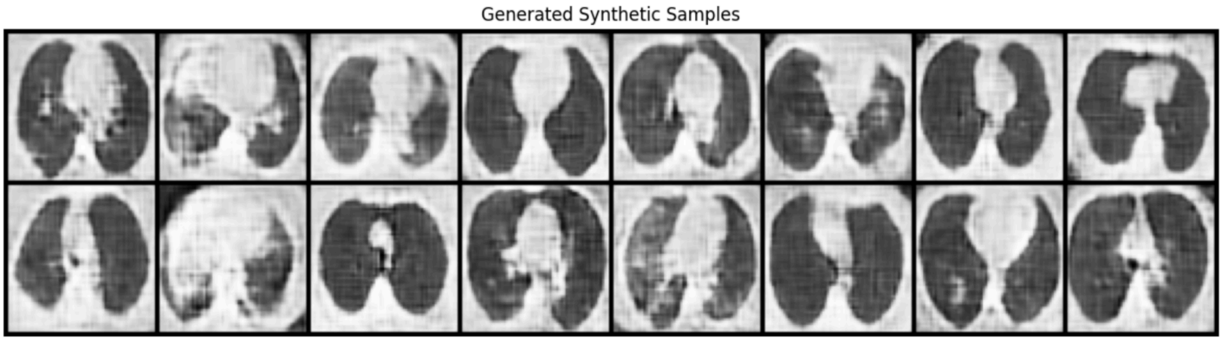

Figure 5: Generated Synthetic Samples

- **Anatomical Coherence:** The images produced accurately represent the macro-anatomical configuration of the lungs. The characteristic kidney-bean shape of the lung parenchyma is easily distinguishable from the surrounding thoracic tissue and background, thereby demonstrating the Generator's effective assimilation of the spatial distribution inherent in the medical data.
- **Pathological Features:** Significantly, the synthetic COVID-19 samples exhibit distinct radiological signs associated with viral pneumonia. Within the lungs, areas that appear translucent and white, which are indicators of ground-glass opacities (GGOs) (Cozzi et al.), can be seen. The observed textures contrast with the transparent, darker regions observed in the synthetic non-COVID samples.
- **Artifacts and Limitations:** Despite the structural soundness, some images exhibit minor checkerboard artifacts, a known effect of Transposed Convolution layers. The 64×64 resolution constraint results in the omission of fine vascular details when compared to the original 512×512 clinical scans. Nevertheless, the experimental results (in the Quantitative Analysis (Accuracy vs Loss) section) demonstrate that, notwithstanding the diminished resolution, the synthetic features possess sufficient discriminative capacity to enhance classifier performance.

Performance & Computational Cost

The total pipeline execution time was recorded for both experimental conditions. Table 3 illustrates the significant disparity in computational overhead.

Experimental Phase	Classical Augmentation	GAN Augmentation
Model Training (Epochs)	50 (with Early Stopping)	50 (with Early Stopping)
Generative Training	N/A	100 Epochs per Class
Data Generation	N/A	~60s
Total Pipeline Time	206.3s	696.1s

Table 3: Computational Cost Comparison

- **Classical Augmentation:** This method incurred negligible overhead. Image transformations (rotation, flip) were performed on the CPU asynchronously.
- **GAN Augmentation:** The generative approach required a distinct pre-training phase for the DCGANs, consuming the majority of the total time.

Practical Implications

The results suggest a trade-off: **GAN augmentation trades computational time for diagnostic accuracy.** In a clinical research setting where model training is a one-time cost, the increased training time is a negligible price to pay for the observed improvement in accuracy and reduction in false negatives. However, for applications requiring real-time model retraining on edge devices, classical augmentation remains the more efficient choice.

Evaluation and Conclusion

Examination and Justification of the Research Question

The primary inquiry was, "*To what extent does synthetic data augmentation using Deep Convolutional Generative Adversarial Networks (DCGANs) outperform classical geometric augmentation in improving the classification accuracy of CNNs on small-scale medical datasets, and what are the computational trade-offs when implemented on consumer-grade hardware?*" The subsequent lenses were employed to evaluate and assess this question:

The Clinical and Technical Significance of the Information

The research inquiry specifically addressed the primary challenge in medical AI, namely the scarcity of labeled data. The experiment examined the efficacy of GANs in comparison to Traditional Data Augmentation to see whether intricate generative models represent a genuine advancement over simplistic rules. The results indicate that Traditional Data Augmentation is insufficient for detecting the intricate biological changes, such as the structure of ground-glass opacities, required for a highly sensitive COVID-19 diagnosis.

The subsequent aspect of the inquiry regarding "consumer-grade hardware" is justifiable, as medical AI must be accessible to all individuals. High-performance clusters (HPCs) are prevalent in the majority of the literature. The study demonstrates that this pipeline is operable on a laptop, so establishing that smaller hospitals and researchers in impoverished regions can utilize advanced augmentation, rather than being limited to large, well-funded facilities.

Evaluation of the Findings

- **Efficacy (The "Degree" of Enhancement):** Empirical evidence indicates a quantifiable advantage. The performance surpassed the traditional standard of 93.01%, with the GAN-augmented model achieving 98.12%. The statistical significance indicates that the magnitude of improvement is substantial and clinically relevant, particularly in reducing False Negatives.

- **Computational Trade-Offs:** The research determined the associated costs. Despite GAN augmentation significantly increasing training duration (206.3 s vs. 696.1 s), its successful performance on market silicon demonstrates its viability. The trade-off is evident: the increased computational expense during the training phase is a justifiable penalty for the enhanced diagnostic precision post-deployment.

Project Management and Reflective Consideration

Effective time management was crucial for the approximately 600 hours allocated for this task.

- **Technical Hurdles:** Issues with the technology: Significant attention was devoted to resolving the RuntimeError: Adaptive pool failure on the Apple M4 processor. This required extensive investigation into the PyTorch backend, leading to the implementation of the Global Average Pooling patch.
- **Risk Management:** I employed Early Stopping and a Learning Rate Scheduler to mitigate the likelihood of training failure.
- **Workflow:** Employing a "Smart Resume" methodology enables incremental testing without necessitating daily retraining of the GANs.

Future Works

Future iterations should explore Conditional GANs (cGANs) to enable a single model to generate both groups. Employing Super-Resolution GANs (SRGAN) could enhance the 64×64 outcomes of clinical 512×512 resolutions.

Conclusion

This experiment demonstrates that Generative AI is an effective solution for addressing the scarcity of data in medical imaging. By compiling authentic training samples, we may enhance AI models' diagnostic capabilities while safeguarding patient privacy. Despite requiring greater computational effort than conventional approaches, the enhanced performance, particularly the reduced incidence of false negatives, justifies its application in critical medical contexts.

References

Ai, Tao, et al. *Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases*. 296, 2 ed., 2020. *RSNA*, RSNA, https://pubs.rsna.org/doi/10.1148/radiol.2020200642.

Antoniou, Antreas, et al. *Data Augmentation Generative Adversarial Networks*. 1, 2017. *arxiv.org*, Cornell University, https://arxiv.org/abs/1711.04340.

Bowles, Christopher, et al. *GAN Augmentation: Augmenting Training Data using Generative Adversarial Networks*. 2018. *arXiv*, Cornell University, https://arxiv.org/abs/1810.10863.

Brock, Andrew, et al. *Large Scale GAN Training for High Fidelity Natural Image Synthesis*. 21 December 2018. *OpenReview.net*, ICLR 2019 Conference Blind Submission, https://openreview.net/forum?id=B1xsqj09Fm.

Chlap, Philip, et al. "A review of medical image data augmentation techniques for deep learning applications." *Journal of MEDICAL IMAGING and RADIATION ONCOLOGY*, vol. 65, no. 5, 2021, pp. 545-563. *Wiley Online Library*, https://onlinelibrary.wiley.com/doi/10.1111/1754-9485.13261.

CHUA, LEON O. *Vision of Nonlinear Science in the 21st Century*. World Scientific Series on Nonlinear Science Series A, 1999. *World Scientific*, https://www.worldscientific.com/action/showCitFormats?doi=10.1142%2F9789812798602_0013.

Cozzi, Diletta, et al. *Ground-glass opacity (GGO): a review of the differential diagnosis in the era of COVID-19*. 2021. *Springer Nature Link*, Springer, https://link.springer.com/article/10.1007/s11604-021-01120-w.

Creswell, Antonia, et al. "Generative Adversarial Networks: An Overview." *IEEE Signal Processing Magazine*, vol. 35, no. 1, 2018, pp. 53-65. *IEEE Xplore*, https://ieeexplore.ieee.org/abstract/document/8253599.

Dickstein, Jascha Sohl, et al. *Deep Unsupervised Learning using Nonequilibrium Thermodynamics*. V8, 2015. *arXiv*, https://arxiv.org/abs/1503.03585.

Durall, Ricard, et al. *Combating Mode Collapse in GAN training: An Empirical Analysis using Hessian Eigenvalues*. 2020. *Semantic Scholar*, VISIGRAPP, https://www.semanticscholar.org/paper/Combating-Mode-Collapse-in-GAN-training%3A-An-using-Durall-Chatzimichailidis/3ad13bd6713d74eec7faa964050f7d61089440a0.

Frid-Adar, Maayan, et al. *GAN-based Synthetic Medical Image Augmentation for increased CNN Performance in Liver Lesion Classification*. 3 March 2018. *arxiv.org*, Cornell University, https://arxiv.org/abs/1803.01229.

"GANs vs. Diffusion Models: In-Depth Comparison and Analysis." *Sapien*, 17 October 2024, https://www.sapien.io/blog/gans-vs-diffusion-models-a-comparative-analysis. Accessed 30 October 2025.

Goceri, Evgin. "Medical image data augmentation: techniques, comparisons and interpretations." *Artificial Intelligence Review*, vol. 56, 2023, pp. 12561-12605. *Springer Nature Link*, https://link.springer.com/article/10.1007/s10462-023-10453-z.

Goodfellow, Ian J., et al. "Generative Adversarial Nets." *Advances in Neural Information Processing Systems*, vol. 27, no. 1, 2014, pp. 2672-2680. *NeurIPS Proceedings*, https://proceedings.neurips.cc/paper_files/paper/2014/hash/f033ed80deb0234979a61f95710dbe25-Abstract.html.

Gulrajani, Ishaan, et al. *Improved Training of Wasserstein GANs*. v3, 2017. *arXiv*, https://arxiv.org/abs/1704.00028.

Hashim, Noha A., et al. *Evolving relationship between respiratory functions & impairment in sleep and cognition in patients with multiple sclerosis*. 46, 2020. *Multiple Sclerosis*, ScienceDirect, https://www.msard-journal.com/article/S2211-0348(20)30565-4/abstract.

Hayes, et al. *LOGAN: Evaluating Privacy Leakage of Generative Models Using Generative Adversarial Networks*. 2017. *ResearchGate*, ResearchGate, https://www.researchgate.net/publication/317061929_LOGAN_Evaluating_Privacy_Leakage_of_Generative_Models_Using_Generative_Adversarial_Networks.

Ho, Jonathon, et al. “Denoising Diffusion Probabilistic Models.” *ArXiv*, vol. abs/2006.11239, no. 1, 2020, p. 1. *Semantic Scholar*, https://www.semanticscholar.org/paper/Denoising-Diffusion-Probabilistic-Models-Ho-Jain/5c126ae3421f05768d8edd97ecd44b1364e2c99a#paper-topics.

Huang, Ziheng, et al. *What can Discriminator do? Towards Box-free Ownership Verification of Generative Adversarial Networks*. Paris, Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2023, https://openaccess.thecvf.com/content/ICCV2023/html/Huang_What_can_Discriminator_do_Towards_Box-free_Ownership_Verification_of_Generative_ICCV_2023_paper.html.

Jogin, Manjunath, et al. *Feature Extraction using Convolution Neural Networks (CNN) and Deep Learning*. 2018. *IEEE Xplore*, IEEE, https://ieeexplore.ieee.org/abstract/document/9012507/references#references.

Krizhevsky, Alex, et al. *ImageNet Classification with Deep Convolutional Neural Networks*. 2012. *NeurIPS Proceedings*, https://papers.nips.cc/paper_files/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html.

Li, He, et al. “The Impact of GDPR on Global Technology Development.” *Journal of Global Information Technology Management*, vol. 22, no. 1, 2019, pp. 1-6. *Taylor & Francis*, https://www.tandfonline.com/doi/full/10.1080/1097198X.2019.1569186?scroll=top&needAccess=true#abstract.

Litjens, Geert, et al. *A survey on deep learning in medical image analysis*. 2017. *ScienceDirect*, Elsevier, https://www.sciencedirect.com/science/article/abs/pii/S1361841517301135.

Radford, Alec, et al. *Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks*. 2, 2016. *arXiv*, ICLR, https://arxiv.org/abs/1511.06434.

Razzak, Muhammad Imran, et al. *Deep Learning for Medical Image Processing: Overview, Challenges and the Future*. 2017. *Springer Nature Link*, Springer, https://link.springer.com/chapter/10.1007/978-3-319-65981-7_12.

Roberts, Michael, et al. *Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans.* 3 ed., Nat Mach Intell, 2021. *Nature Machine Intelligence*, https://www.nature.com/articles/s42256-021-00307-0#Abs1.

Salimans, Tim, et al. *Improved Techniques for Training GANs*. 2016. *arxiv.org*, Cornell University, https://arxiv.org/abs/1606.03498.

Sharma, Preeti, et al. *Generative adversarial networks (GANs): Introduction, Taxonomy, Variants, Limitations, and Applications*. 2024. *Springer Nature Link*, Springer Nature Link, https://link.springer.com/article/10.1007/s11042-024-18767-y#Abs1.

Shorten, Connor, and Taghi M. Khoshgoftaar. “A survey on Image Data Augmentation for Deep Learning.” *Journal of Big Data*, vol. 6, no. 60, 2019, p. 60. *Springer Nature Link*, https://link.springer.com/article/10.1186/s40537-019-0197-0.

Simard, P. Y., et al. *Best practices for convolutional neural networks applied to visual document analysis*. 06 August 2003. *IEEE Xplore*, IEEE, https://ieeexplore.ieee.org/document/1227801/authors#authors.

Simonyan, Karen, and Andrew Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2015. *arXiv.org*, Cornell University, https://arxiv.org/abs/1409.1556.

Soares, et al. *SARS-CoV-2 CT-scan dataset: A large dataset of real patients CT scans for SARS-CoV-2 identification.* 2020. *Kaggle*, medRxiv, https://www.kaggle.com/datasets/plameneduardo/sarscov2-ctscan-dataset.

Varughese, Jobit. "What are generative adversarial networks (GANs)?" *IBM*, 2025, https://www.ibm.com/think/topics/generative-adversarial-networks. Accessed 29 October 2025.

Yi, Xin, et al. *Generative Adversarial Network in Medical Imaging: A Review*. 4, 19 September 2018. *arxiv.org*, Cornell University, https://arxiv.org/abs/1809.07294.

Zhao, Wan-Lei, et al. "k-means: A revisit." *Neurocomputing*, vol. 291, no. 1, 2018, pp. 195-206. *ScienceDirect*, https://www.sciencedirect.com/science/article/abs/pii/S092523121830239X.

Zhou, Zhi, et al. "Edge Intelligence: Paving the Last Mile of Artificial Intelligence With Edge Computing." *Proceedings of the IEEE*, vol. 107, no. 8, 2019, pp. 1738-1762. *IEEEXplore*, https://ieeexplore.ieee.org/document/8736011.

Appendices

Appendix A: Project Artefact (Source Code)

This appendix contains the complete Python source code for the experimental pipeline, including data preprocessing, DCGAN implementation, VGG16 adaptation, and the training loops for both Classical and GAN-based augmentation. (Note for submission: The full code is attached as a separate plain text file named 23071769-Md Tanvir Sarwar-Artefact.txt).

Appendix B: Synthetic Sample Grid

Visual evidence of the DCGAN's generative capability. This grid displays 64 × 64 synthetic chest CT slices generated after 100 epochs of training. Key anatomical features (lung lobes) and pathological textures (ground-glass opacities) are visible.

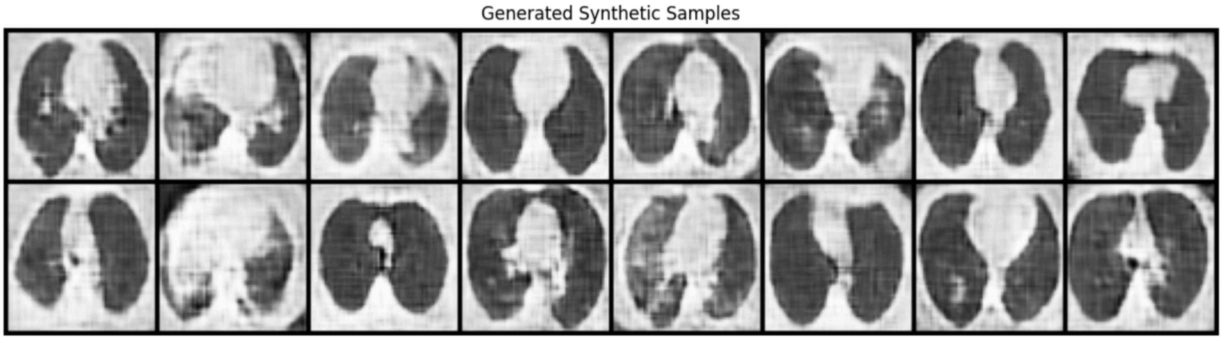

Figure 6: Generated Synthetic Images

Appendix C: Training Dynamics and Performance Comparison

Comparative graphs illustrating the training trajectory of the Classical versus GAN-Augmented models.

- **Left:** Validation Accuracy over 50 epochs.
- **Right:** Training Loss convergence.

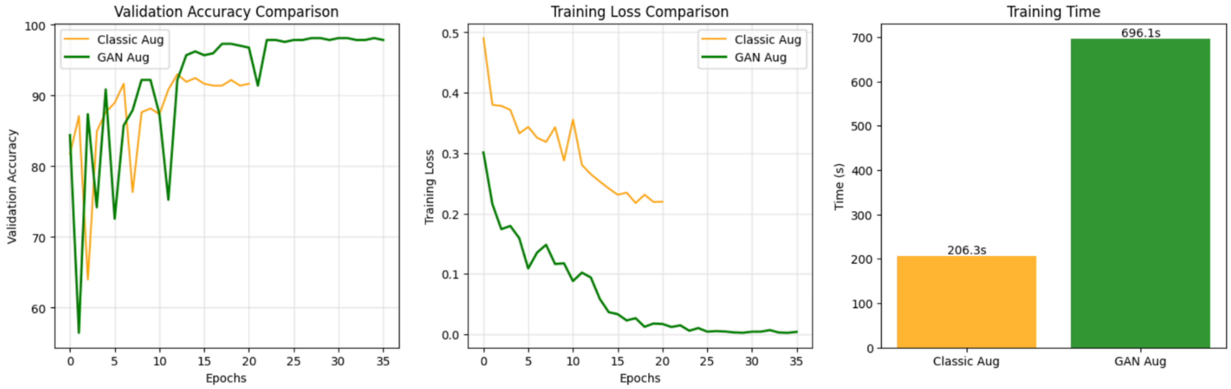

Figure 7: Validation Accuracy & Training Loss Curves with Time Comparison

Appendix D: Confusion Matrices

Detailed breakdown of classification performance on the held-out Test Set.

- Left: Classical Augmentation (Baseline).
- Right: GAN Augmentation.
- Analysis: Note the reduction in false negatives (bottom-left quadrant) in the GAN-augmented matrix.

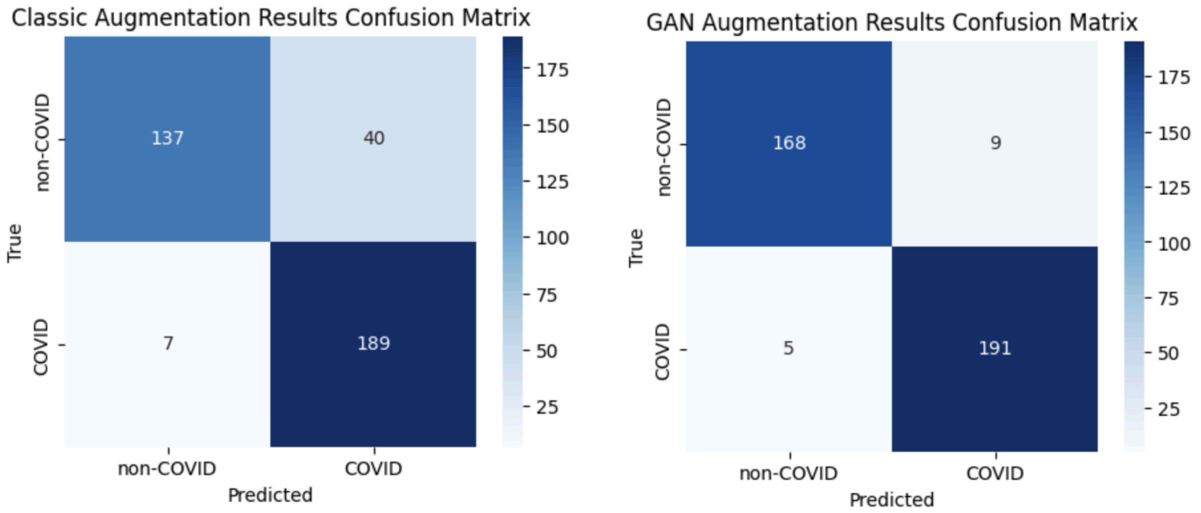

Figure 8: Heatmap for Classic Augmentation and GAN Augmentation Results Confusion Matrix

Appendix E: Model Architecture and Parameters

The study provides a detailed summary of the neural network architectures and the total trainable parameters used.

Model Component	Trainable Parameters	Description
Generator (DCGAN)	447,681	4 Transposed Convolutional Blocks mapping the $z \rightarrow 64 \times 64$ image.
Discriminator (DCGAN)	99,649	4 Strided Convolutional Blocks mapping image → probability.
Classifier (VGG16)	15,050,562	Modified VGG16 backbone with Global Average Pooling and Binary Head.

Table 4: Model Complexity