الجمهورية الجزائرية الديمقراطية الشعبية
PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
وزارة التعليم العالي والبحث العلمي
Ministry of Higher Education and Scientific Research
جامعة أبي بكر بلقايد – تلمسان –
Aboubakr Belkaïd University – Tlemcen –
Faculty of Technology

THESIS

Submitted in Partial Fulfillment of the Requirements for the Degree of **MASTER**

In: Biomedical Engineering

Specialization: Medical Imaging

By: Imane Rahali

Title of Thesis :

Developing a Custom Generative Adversarial Network for Synthetic Mammogram Generation

Publicly Defended on: 12 / 06 / 2025
Before the Examination Committee Composed of:

Name and Surname	Title/Rank	University of Tlemcen	Role
Mr. Gaouar Adil	Associate Professor	University of Tlemcen	President
Ms. Iles Amal	Associate Professor	University of Tlemcen	Examiner
Ms. Kazi Tani Lamia Fatiha	Associate Professor	University of Tlemcen	Supervisor
Ms. Hamza Cherif Souaad	Associate Professor	University of Tlemcen	Co-Supervisor

Academic Year : 2024/2025

Abstract

As one of the most commonly diagnosed cancers worldwide, breast cancer poses a significant challenge in early detection and treatment. Mammographic analysis plays a crucial role in diagnosis, with malignant calcifications and malignant masses serving as key radiological markers. However, the limited availability of annotated medical imaging datasets and high inter-observer variability pose challenges in developing robust AI-driven diagnostic models.

This thesis focuses on developing a custom Generative Adversarial Network (GAN) to generate synthetic mammograms that enhance the representation of malignant calcifications and masses. Unlike studies utilizing pre-existing GAN architectures, this research builds a GAN model from scratch, with a focus on optimizing performance and improving image fidelity. The generated synthetic images aim to augment existing datasets, providing additional training data for deep learning models in breast cancer classification.

The study employs a dataset containing mammographic images of malignant calcifications and malignant masses, carefully curated for training and validation. To assess the quality of the generated images, performance metrics such as Fréchet Inception Distance (FID), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index (SSIM) are utilized. Additionally, the GAN's performance is evaluated using the loss graph, where the discriminator and generator losses are plotted over training epochs to analyze convergence and stability.

To assess the diagnostic value of the synthetic images, a Support Vector Machine (SVM) classifier is also employed. The classifier is tested with real and GAN-generated images to evaluate their robustness and impact on classification performance.

By improving the performance of a custom GAN for synthetic mammogram generation, this work contributes to addressing data scarcity, enhancing AI-based breast cancer detection, and supporting the integration of generative models in medical imaging. The findings highlight the potential of GAN-based data augmentation in improving diagnostic reliability and dataset diversity in breast cancer research.

Keywords:
Breast Cancer, Mammography, Generative Adversarial Networks (GANs), Synthetic Image Generation, Malignant Calcifications, Malignant Masses, Data Augmentation, Medical Imaging, Deep Learning, Support Vector Machine (SVM), Image Quality Metrics, FID, PSNR, SSIM

Résumé

Le cancer du sein est l'un des cancers les plus fréquemment diagnostiqués dans le monde. Il représente un véritable défi pour le dépistage précoce et le traitement. L'analyse des mammographies joue un rôle essentiel dans le diagnostic, notamment grâce à la détection des calcifications malignes et des masses malignes, qui sont des marqueurs importants sur les images. Cependant, le manque de bases de données médicales bien annotées et les différences entre les avis des radiologues rendent difficile le développement de modèles d'intelligence artificielle fiables pour le diagnostic.

Ce mémoire présente le développement d'un modèle personnalisé de Generative Adversarial Network (GAN) pour générer des mammographies synthétiques. L'objectif est de mieux représenter les calcifications et les masses malignes afin d'enrichir les bases de données existantes et d'améliorer l'entraînement des modèles d'apprentissage profond pour la classification des cancers du sein. Contrairement à d'autres travaux utilisant des GANs déjà existants, ce projet propose un GAN créé de zéro, avec des améliorations pour obtenir de meilleures performances et des images de meilleure qualité.

Le travail s'appuie sur un ensemble de mammographies contenant des calcifications et des masses malignes, sélectionnées avec soin pour l'entraînement et la validation. La qualité des images générées est évaluée à l'aide d'indicateurs tels que le Fréchet Inception Distance (FID), le Peak Signal-to-Noise Ratio (PSNR) et le Structural Similarity Index (SSIM). Les courbes de perte du générateur et du discriminateur sont également étudiées pour analyser la stabilité et la convergence de l'entraînement.

Enfin, un classificateur Support Vector Machine (SVM) est utilisé pour mesurer la valeur diagnostique des images synthétiques. Le modèle est testé avec des images réelles et des images générées par le GAN pour évaluer leur impact sur la classification.

Ce travail montre que l'utilisation d'un GAN sur mesure peut aider à résoudre le problème du manque de données et à améliorer la détection du cancer du sein par intelligence artificielle. Les résultats confirment l'intérêt des images synthétiques pour enrichir les bases de données et renforcer la fiabilité des diagnostics.

Mots-clés :

Cancer du sein, Mammographie, Réseaux antagonistes génératifs (GAN), Génération d'images synthétiques, Calcifications malignes, Masses malignes, Augmentation de données, Imagerie médicale, Apprentissage profond, Machine à vecteurs de support (SVM), Indicateurs de qualité d'image, FID, PSNR, SSIM

الملخص

يُعتبر سرطان الثدي من أكثر أنواع السرطان انتشاراً في العالم، وهو يشكّل تحدياً كبيراً في الكشف المبكر والعلاج. حيث تلعب صور الماموغرافيا دوراً مهماً في التشخيص، خاصة من خلال اظهار التكلسات الخبيثة والكتل السرطانية في الصور، والتي تُعتبر علامات أساسية في الصور الشعاعية. ومع ذلك، فإن قلة توفر قواعد بيانات طبية مع البيانات التعريفية والاختلاف في تقييم الأطباء المختصين يشكلان عائقاً أمام تطوير نماذج تشخيصية قوية تعتمد على الذكاء الاصطناعي.

يركز هذا البحث على تطوير نموذج شبكة توليد خصومية (GAN) مخصص من البداية، بهدف إنتاج صور ماموغرافيا اصطناعية تُحسّن تمثيل التكلسات والكتل الخبيثة. يهدف هذا النموذج إلى تعزيز قواعد البيانات المتوفرة من خلال توفير صور إضافية لتدريب نماذج التعلم العميق الخاصة بتصنيف سرطان الثدي. بخلاف الدراسات التي تعتمد على شبكات جاهزة، تم تصميم الشبكة في هذا العمل من الصفر مع التركيز على تحسين الأداء وجودة الصور المنتجة.

تم استخدام مجموعة بيانات تحتوي على صور ماموغرافيا بها تكلسات وكتل خبيثة، حيث تم إعدادها بعناية للتدريب والاختبار. وتم تقييم جودة الصور المولدة باستخدام مؤشرات مثل مسافة فريشيت (FID) ونسبة الإشارة إلى الضجيج (PSNR) ومؤشر التشابه الهيكلي (SSIM). كما تم تحليل منحنيات دالة الفقدان للمولد والمميز لدراسة استقرار التدريب وتقارب النموذج.

كما تم اختبار القيمة التشخيصية للصور الاصطناعية باستخدام آلة المتّجه الداعم (SVM)، حيث جرى تقييم أدائه عند استخدام الصور الحقيقية والمولدة لقياس تأثيرها على دقة التصنيف.
يساهم هذا العمل في مواجهة مشكلة نقص البيانات، ودعم تحسين كشف سرطان الثدي باستخدام الذكاء الاصطناعي، ويؤكد على أهمية الصور الاصطناعية في تعزيز تنوع قواعد البيانات وتحسين موثوقية التشخيصات.

الكلمات المفتاحية :
سرطان الثدي، التصوير الشعاعي للثدي، الشبكات التوليدية التنافسية (GAN)، توليد الصور الاصطناعية، التكلسات الخبيثة، الكتل الخبيثة، زيادة البيانات، التصوير الطبي، التعلم العميق، آلة الدعم الناقل (SVM)، مؤشرات جودة الصور، FID، PSNR، SSIM

Acknowledgements

First and foremost, I am deeply grateful to Allah, whose mercy, guidance, and countless blessings have carried me through every step of this journey. Today, Allah has made a dream come true, and here I am, taking my final steps along this long and challenging road.

I would also like to express my deepest gratitude to my late father, whose passion for informatics and constant support during my childhood have been a guiding light throughout my academic path. His memory continues to inspire me to overcome challenges and strive for excellence.

I am especially thankful to my mother, who has provided for me throughout these five years and has always believed in my potential. Without her endless support and sacrifices, I would not have reached this stage. Her faith in me kept me going through even the toughest moments at university. This achievement belongs to her as much as it does to me.

I am grateful to my supervisors and professors for their invaluable guidance, patience, and expertise. Their support has been instrumental in shaping this thesis and helping me navigate the complexities of medical imaging and deep learning.

I also extend my heartfelt thanks to the rest of my family, my siblings and my grandmother, who have always cared for me and celebrated my academic milestones with joy and pride. Their love and encouragement have been an essential part of this journey.

This thesis stands as a testament to the collective support, love, and inspiration I have received from Allah and from all those who have stood by me along the way.

Table of Contents

List of Figures

List of Tables

Abbreviations

WHO – World Health Organization

AI – Artificial Intelligence

IARC – International Agency for Research on Cancer

IHE – Integrating the Healthcare Enterprise

ACR – American College of Radiology

DCIS – Ductal Carcinoma In Situ

MRI – Magnetic Resonance Imaging

ACS – American Cancer Society

GANs – Generative Adversarial Networks

G – Generator

D – Discriminator

SRGAN – Super-Resolution Generative Adversarial Network

AttnGAN – Attention Generative Adversarial Network

StackGAN – Stacked Generative Adversarial Network

DCGAN – Deep Convolutional Generative Adversarial Network

CT scans – Computed Tomography scans

cGANs – Conditional Generative Adversarial Networks

FID – Fréchet Inception Distance

SSIM – Structural Similarity Index Measure

PSNR – Peak Signal-to-Noise Ratio

SVM – Support Vector Machine

CBIS-DDSM – Curated Breast Imaging Subset of the Digital Database for Screening Mammography

GPU – Graphics Processing Unit

Conv2D – 2D Convolutional Layer

GENERAL INTRODUCTION

1. Introduction

Breast cancer is one of the most prevalent and life-threatening diseases globally, accounting for a significant proportion of cancer-related deaths among women. According to the World Health Organization (WHO, 2023), It is the most commonly diagnosed cancer worldwide, with millions of new cases reported annually. Early and accurate diagnosis is critical for improving patient outcomes, as it enables timely intervention and treatment. However, traditional diagnostic methods, such as mammography, ultrasound, and MRI, face several challenges, including limited data availability, variability in imaging quality, and the need for expert interpretation (Smith et al., 2022).

In recent years, advancements in artificial intelligence (AI) and deep learning have shown great promise in addressing these challenges, particularly in the field of medical imaging. Among deep learning techniques, Generative Adversarial Networks (GANs) have emerged as a powerful tool for generating synthetic data, enhancing image quality, and facilitating tasks such as image-to-image translation and data augmentation (Goodfellow et al., 2014). GANs consist of two neural networks, a generator and a discriminator, that compete against each other to produce highly realistic synthetic data. This capability makes GANs particularly valuable in medical imaging, where access to large, diverse, and annotated datasets is often limited.

The application of GANs in breast cancer research has gained significant attention, with studies demonstrating their potential in tasks such as synthetic mammogram generation, lesion synthesis, and radiomic feature extraction (Zhang et al., 2020). For instance, GANs can generate synthetic mammograms that simulate real patient data, enabling researchers and clinicians to augment limited datasets and improve the performance of diagnostic models. Despite their promise, the use of GANs in this domain is still in its early stages, and several challenges remain, including model instability, mode collapse, and the need for robust evaluation metrics (Wang et al., 2022). Addressing these challenges and exploring the full potential of GANs in breast cancer diagnosis and treatment is a critical area of research.

This thesis investigates the development and evaluation of custom GAN model to generate synthetic mammography images, aiming to support diagnostic improvements and data augmentation in breast cancer imaging

2. Problem Statement

Breast cancer diagnosis relies heavily on high-quality, annotated medical imaging datasets to train deep learning models for accurate detection. However, the availability of such datasets remains limited, particularly for malignant calcifications and malignant masses, which are critical indicators in mammography. This scarcity leads to challenges in developing robust AI models, as deep learning algorithms require diverse and well-distributed data to generalize effectively.

While Generative Adversarial Networks (GANs) have shown promise in synthetic medical image generation, evaluating their performance in terms of both image fidelity and training stability remains an open research challenge. This thesis addresses these gaps by building a custom GAN model from scratch to generate synthetic mammograms focusing on malignant calcifications and malignant masses. The study systematically evaluates the generated images using quantitative metrics (FID, PSNR, SSIM) and loss graph analysis, ensuring a thorough assessment of the model's ability to generate high-quality synthetic mammograms. By

enhancing the availability of synthetic data, this work aims to improve dataset diversity, support AI-based diagnostic advancements, and reduce the dependency on large annotated datasets for breast cancer detection.

3. Objectives

The primary objectives of this research are as follows:

To develop a custom GAN model for generating synthetic mammograms focusing on malignant calcifications and malignant masses.

To use these generated images for augmenting mammography datasets and enhancing the accuracy of breast cancer classification models.

To evaluate the quality and impact of the synthetic images using quantitative metrics and classification performance analysis.

To identify the strengths and limitations of the developed model and propose potential improvements for future research.

4. The structure of the Thesis

The thesis is structured into four chapters. **Chapter 1** introduces the medical background of breast cancer, with a particular focus on malignant calcifications and masses, highlighting their diagnostic importance and related challenges. **Chapter 2** presents the theoretical foundation of Generative Adversarial Networks (GANs), including their architecture, training mechanisms, and applications in medical imaging, especially for generating synthetic mammography data. **Chapter 3** outlines the methodological approach of this research, detailing the dataset used, preprocessing steps, the proposed GAN architecture, training strategy, and the classification process based on ResNet50 and SVM. Finally, **Chapter 4** presents and analyzes the results of image generation and classification experiments, discussing the effectiveness of synthetic images in augmenting datasets and improving diagnostic performance.

CHAPTER 1 : Breast Cancer and Malignant Calcifications and Masses

1.1. Introduction

Breast cancer is a major global health concern, with over 2.3 million new cases diagnosed annually (World Cancer Research Fund, 2023), making early detection essential for improving patient outcomes. It remains the leading cause of cancer-related deaths among women, particularly in low- and middle-income countries, where limited access to early diagnostic tools contributes to late-stage detection and poorer outcomes (WHO, 2024). Early detection significantly improves survival rates by enabling timely intervention and treatment.

Among the key radiological indicators of breast cancer are malignant calcifications and malignant masses, which play a crucial role in mammographic diagnosis. However, accurately distinguishing malignant findings remains a challenge due to their variability in appearance, overlap with benign features, and inter-observer variability among radiologists. The complexity of identifying malignant calcifications and masses necessitates advanced imaging techniques and AI-driven solutions to improve diagnostic accuracy.

This chapter provides a comprehensive overview of breast cancer, focusing on the significance of malignant calcifications and malignant masses in early diagnosis and the challenges associated with their detection and classification.

1.2. Breast Cancer

1.2.1. Definition of Breast Cancer

Breast cancer is a malignant tumor that originates in breast tissue, arising from the uncontrolled proliferation of abnormal cells in the mammary glands. It primarily affects women but can also occur in men, albeit at a significantly lower incidence rate. The disease is heterogeneous, encompassing multiple subtypes with distinct pathological and molecular characteristics, which influence prognosis and treatment strategies.

1.2.2. Global Prevalence and Impact

Globally, breast cancer ranks highest among all cancers in women both in incidence and mortality. In 2022, there were 2.3 million new cases and 670,000 deaths globally, underscoring its significant public health burden (WHO, 2023). Striking disparities exist based on socioeconomic development:

- **High-HDI countries**: 1 in 12 women will be diagnosed, and 1 in 71 will die from the disease.
- **Low-HDI countries**: 1 in 27 women will be diagnosed, but 1 in 48 will die, reflecting poorer survival due to late-stage diagnosis and limited treatment access (WHO, 2023).

These disparities highlight the critical need for equitable early detection and treatment strategies, particularly in resource-limited settings (Sung et al., 2021).

Region	Incidence (%)	Mortality (%)
North America	13.3%	7.5%
Europe	24.3%	21.7%
Asia	42.9%	47.3%
Africa	8.6%	13.7%
Latin America	9.6%	9.0%
Oceania	1.2%	0.82%

Table 1.1. *Percentage of Global Breast Cancer Incidence and Mortality by Region (Both Sexes, All Ages, GLOBOCAN 2022 Data)*

As shown in Table 1.1, the mortality rate in Africa is disproportionately high compared to its incidence. In Algeria, breast cancer is the most prevalent cancer among women, surpassing all other cancer types. Statistics from the Global Cancer Observatory (GLOBOCAN 2022, Version 1.1) indicate that breast cancer accounted for the highest number of new cancer cases in Algerian females across all age groups (International Agency for Research on Cancer, 2024). Figure 1.1 illustrates the distribution of breast cancer cases among all female cancers in Algeria in 2022, further emphasizing the urgent need for improved screening programs, early detection initiatives, and access to advanced treatment facilities to reduce mortality rates.

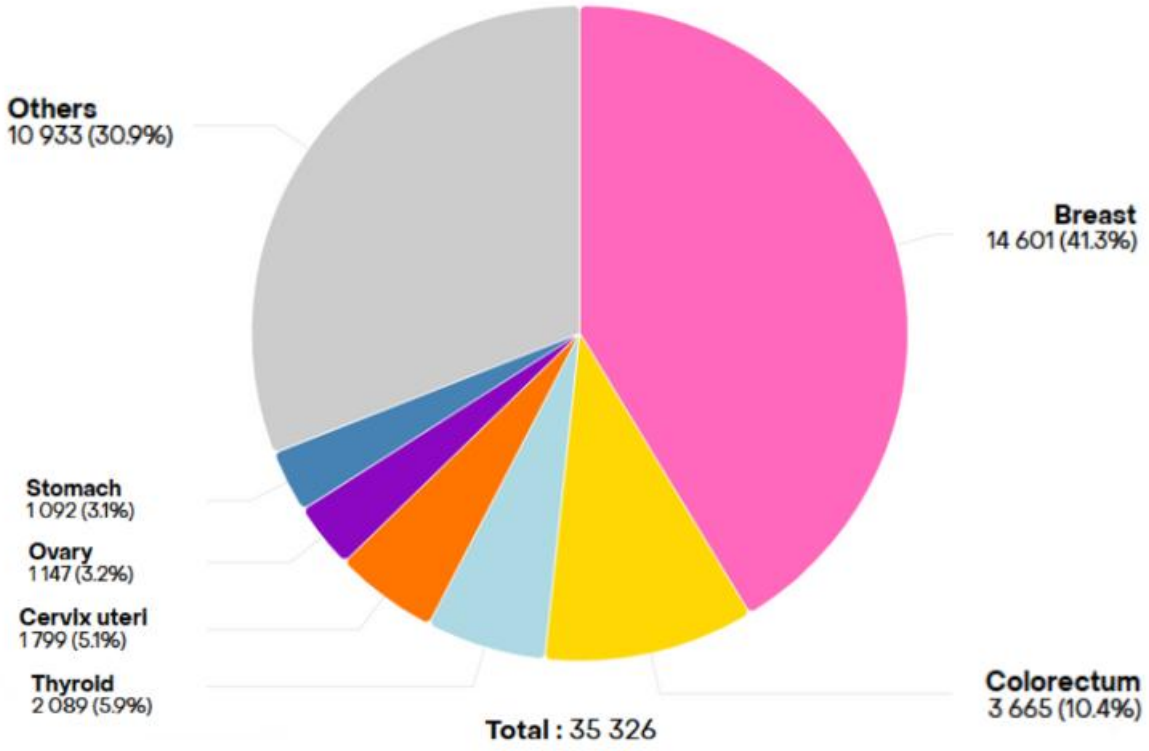


Figure 1.1. *Distribution of breast cancer cases among all female cancers in Algeria (2022) (Source: International Agency for Research on Cancer, 2024)*

The incidence of breast cancer varies based on factors such as age, gender, geographic location, and socioeconomic status. High-income countries generally report higher incidence rates, likely due to improved screening programs and lifestyle factors. However, mortality rates are disproportionately high in low-income regions, where late-stage diagnosis is more

prevalent.

1.2.3. Statistical Analysis of Early and Late Breast Cancer Detection

1.2.3.1. Importance of Early Detection

Early Breast cancer survival rates show dramatic global variations that directly reflect healthcare system capabilities. In high-income nations, five-year survival rates consistently surpass 90%, a remarkable achievement made possible through integrated healthcare systems that combine population-wide screening initiatives with universally accessible treatment pathways (IARC, 2022). These countries benefit from established mammography programs, timely diagnostic services, and comprehensive treatment options including surgery, radiation therapy, and targeted medications. The success of this approach demonstrates how coordinated early detection efforts combined with treatment accessibility can significantly improve cancer outcomes.

In contrast, Places with fewer resources face serious challenges that lead to poorer results. Recent data indicates five-year survival rates remain below 50% in many low- and middle-income countries, with some regions reporting rates as low as 40% (WHO, 2023). Multiple systemic barriers contribute to this disparity, including frequent late-stage diagnoses when treatment options are limited, a lack of proper screening tools and facilities, a lack of specialized medical personnel, and financial problems that prevent patients from seeking or continuing care. Cultural factors and lack of health knowledge make these challenges worse by causing delays in seeking medical care.

According to the World Health Organization's 2023 report, about 70% of breast cancer deaths occur in developing countries, even though these regions have fewer overall cases. This imbalance shows how differences in healthcare systems directly affect survival rates. Global health efforts now stress the importance of finding solutions that fit each country’s resources—improving both early detection and treatment access. Focusing on just prevention or treatment is not enough to address this complex public health issue.

1.2.3.2. Breast Cancer Detection in Algeria

In Algeria, breast cancer remains the most common cancer among women, accounting for approximately one-third of all female cancer cases (IARC, 2022). Despite its high incidence, significant gaps persist in early detection efforts. Recent data shows that only 48-52% of Algerian breast cancer cases are diagnosed at early stages (Stage I or II) (IHE, 2024), falling substantially below both the 60% target set by the WHO's Global Breast Cancer Initiative and the 70-80% early detection rates observed in developed nations (WHO, 2023). This diagnostic delay results in approximately half of patients presenting with advanced-stage disease (Stage III or IV), where treatment options are more limited and survival probabilities significantly reduced (IHE, 2024).

Several barriers contribute to delayed diagnosis in Algeria, including: limited mammography screening coverage outside urban centers, cultural hesitancy toward breast examinations, and insufficient public awareness about early warning signs. These challenges are compounded by regional disparities in healthcare infrastructure, especially in rural areas, where shortages of diagnostic equipment and specialized oncology services persist. Current public health initiatives are working to address these gaps through mobile screening units and community education programs, though significant progress remains needed to achieve international early detection standards.

1.2.3.3. Impact of Early Detection on Survival

Early-stage detection of breast cancer (Stage I-II) substantially improves both treatment efficacy and long-term survival. When diagnosed at Stage I, patients typically have access to less invasive treatment modalities with 5-year survival rates exceeding 95% in high-income countries (Allemani et al., 2018). This contrasts sharply with late-stage diagnoses (Stage III-IV), where metastatic spread reduces treatment effectiveness and survival probabilities decline dramatically (WHO, 2023). Table 1.2 presents a comparison of five-year survival rates by diagnostic stage and economic setting, highlighting the disparity in outcomes between high- and low-resource environments.

Stage at Diagnosis	**Survival Rate (%) in High-Income Countries**	**Survival Rate (%) in Low/Middle-Income Countries**
Stage I	>95%	~60-70%
Stage II	85-90%	~50-60%
Stage III	50-70%	~30-40%
Stage IV	<30%	< 20%

Table1.2. *Five-Year Survival Rates by Diagnostic Stage and Economic Setting* ***(****Adapted from "Global Surveillance of Cancer Survival" by Allemani et al. (2018) and WHO Global Breast Cancer Initiative (2023))*

The disparity in survival rates shows the necessity of improved early detection strategies, particularly in low-resource settings such as Algeria. Expanding access to mammographic screening, increasing public awareness, and implementing AI-driven diagnostic tools could play a crucial role in reducing breast cancer mortality rates.

1.2.3.4. Role of Imaging Techniques in Early Diagnosis

Mammography maintains its position as the gold standard for population-based breast cancer screening, demonstrating consistent efficacy in detecting early-stage malignancies through identification of microcalcifications (85-92% accuracy) and structural distortions (ACR, 2023). Large-scale studies confirm that organized mammography screening programs achieve 25-30% mortality reduction in compliant populations, primarily through stage migration to earlier diagnoses (WHO, 2023). The standardized BI-RADS classification system (ACR, 2023) ensures uniform reporting across clinical settings, with Category 3 lesions demonstrating <2% malignancy risk when properly implemented.

Imaging Technique	Sensitivity	Specificity	Best Use Case
Digital Mammography	78-85	87-92	Population screening (age 40+)
Ultrasound	76-82	83-88	Dense breast adjunct
Breast MRI	92-97	81-86	High-risk surveillance

Table1.3. *Performance Characteristics of Contemporary Breast Imaging Modalities*
(Adapted from American College of Radiology (2023) BI-RADS Atlas et al. (2020) multicenter study)

While mammography shows reduced sensitivity (58–65%) in heterogeneously dense breasts (ACR, 2023), supplemental modalities address specific clinical scenarios, as summarized in Table 1.3, which presents the performance characteristics of contemporary breast imaging techniques.

- **Automated Breast Ultrasound**: Adds 3.2 cancers/1000 women in dense tissue (95%CI 2.0-4.8) without excessive recalls.
- **Abbreviated MRI**: Emerging protocol with 87% sensitivity in 7-minute scans for intermediate-risk patients
- **Digital Tomosynthesis**: 30% increased cancer detection versus 2D mammography ($p<0.001$) in all density groups

1.3. Breast Cancer Pathophysiology: Malignant Calcifications and Malignant Masses

1.3.1. Molecular and Histopathological Basis of Breast Cancer Development

Breast cancer pathogenesis represents a multistep process involving cumulative genetic and epigenetic alterations that transform normal breast epithelium into malignant tissue. The initiation phase typically begins with acquired or inherited mutations in critical tumor suppressor genes, most notably BRCA1 and BRCA2, which play essential roles in DNA damage repair through homologous recombination. Large-scale cohort studies have demonstrated that carriers of pathogenic BRCA variants face a 45-72% lifetime risk of developing breast cancer, with BRCA2 mutations particularly associated with estrogen receptor-positive tumors (Kuchenbaecker et al., 2017). Beyond hereditary factors, somatic mutations in TP53 (occurring in 30% of cases), PIK3CA (40%), and GATA3 (10%) frequently emerge during tumor progression, contributing to uncontrolled cellular proliferation and apoptotic resistance (Cancer Genome Atlas Network, 2012).

The hormonal microenvironment serves as a critical mediator of breast carcinogenesis, particularly in luminal subtypes. Estrogen exerts its oncogenic effects through both genomic and non-genomic mechanisms, including receptor-mediated transcriptional activation of proliferative genes (MYC, CCND1) and generation of reactive oxygen species via cytochrome P450-mediated metabolism (Yager & Davidson, 2006). Clinical observations reveal that prolonged exposure to endogenous estrogens, whether through early menarche (<12 years), late menopause (>55 years), or nulliparity, increases cancer risk by 20-30%, while therapeutic estrogen blockade reduces contralateral breast cancer incidence by 48% in high-risk populations (Early Breast Cancer Trialists' Collaborative Group, 2005).

Histopathological progression generally follows a well-characterized sequence from benign breast disease to invasive carcinoma. The earliest recognizable precursor lesion, atypical ductal hyperplasia (ADH), carries a 4-5 fold increased risk of subsequent malignancy and shares molecular alterations (e.g., 16q loss) with adjacent carcinomas, suggesting a direct precursor relationship (Schnitt, 2010). These lesions may evolve into ductal carcinoma in situ (DCIS), characterized by malignant epithelial cells confined by the basement membrane. DCIS presents mammographically as clustered microcalcifications in 60-80% of cases, with histological grade strongly correlating with calcification morphology - high-grade lesions typically demonstrate fine linear or branching patterns (95% PPV for malignancy), while low-grade variants often show coarse heterogeneous deposits (30% PPV) (D'Orsi et al., 2013)

1.3.2. Malignant Calcifications

Mammographically detectable calcifications serve as critical early markers of breast malignancy, present in 50-60% of DCIS cases and 30-40% of invasive carcinomas (ACR, 2023). These mineral deposits form through complex biomineralization processes that differ substantially between benign and malignant lesions. Benign calcifications typically consist of calcium oxalate dihydrate crystals within cystic spaces or stromal elastosis, appearing mammographically as large (>0.5 mm), smooth, and diffusely distributed particles (Baker et al., 2021). In contrast, malignant calcifications predominantly comprise hydroxyapatite crystals that precipitate within necrotic tumor debris or secretory material within neoplastic ducts, yielding the characteristic irregular, clustered, or linear configurations (positive predictive value 75-90% for malignancy) (ACR, 2023).

The diagnostic evaluation of calcifications requires careful assessment of multiple morphological features:

- **Distribution**: Segmental or linear arrangements (PPV=65%) indicate ductal involvement versus regional/diffuse patterns (PPV<10%)
- **Density**: Increasing density correlates with higher nuclear grade ($r=0.72$, $p<0.01$)
- **Size**: Particles <0.2 mm carry 3× higher malignancy risk than 0.5-1 mm calcifications (Burnside et al., 2016)

1.3.3. Malignant Masses

Malignant breast masses represent a critical diagnostic finding in breast imaging, characterized by distinct morphological features that reflect their invasive nature. Unlike benign lesions such as fibroadenomas (smooth, oval, circumscribed) or cysts (anechoic, thin-walled), malignant masses typically demonstrate spiculated margins (85-90% positive predictive value for malignancy), irregular shape with angular borders, heterogeneous echotexture on ultrasound, and posterior acoustic shadowing (observed in 70% of invasive ductal carcinomas) (Stavros et al., 1995). These radiological appearances directly correlate with the pathological behavior of invasive carcinomas, which infiltrate adjacent stroma and induce a desmoplastic reaction, producing the classic stellate pattern visible on mammography.

The diagnostic evaluation of suspicious masses remains challenging in clinical practice, as approximately 15-20% of cases present with indeterminate features that overlap between benign and malignant categories, particularly in dense breast tissue where mammographic sensitivity decreases to 48-65% (ACR, 2021; Sprague et al., 2018). Radiologist interpretation shows moderate inter-observer variability ($\kappa=0.45$-0.60) for borderline lesions (Le et al., 2019), though recent advances in digital breast tomosynthesis (30% improvement in detection rates), AI-assisted interpretation (94% classification accuracy; McKinney et al., 2020), and contrast-enhanced MRI (89% sensitivity for invasive lobular carcinomas) are helping to address these diagnostic challenges.

1.3.4. Role of Mammographic Calcifications and Masses in Diagnosis

Mammographic calcifications and malignant masses serve as critical radiological markers for detecting breast cancer at an early stage. Their size, shape, and distribution provide essential diagnostic insights:

Benign Calcifications and Masses: Benign calcifications typically appear as large (>0.5 mm), coarse particles with smooth contours and diffuse or scattered distribution patterns. These are most commonly associated with fibrocystic changes (20–30% of cases),

fibroadenomas (15–20%), or fat necrosis (5–10%) (American College of Radiology [ACR], 2023). Figure 1.2 illustrates typical mammographic features of benign calcifications, as commonly seen in clinical practice.

On mammography, benign masses demonstrate well-circumscribed margins (93–97% specificity for benignity) and homogeneous internal density, with oval or round shapes being most predictive of non-malignant etiology (Stavros et al., 1995; Sickles, 2019). Representative examples of benign breast masses are shown in Figure 1.3, highlighting their common radiographic patterns.

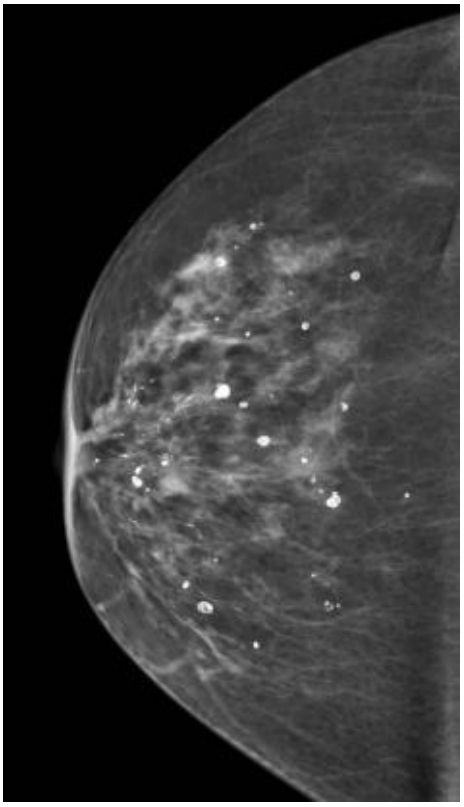


Figure1.2. *Benign breast calcifications demonstrating typical mammographic features. (Source: Radiopaedia.org, n.d.; Reprinted under CC BY-NC-SA 3.0)*

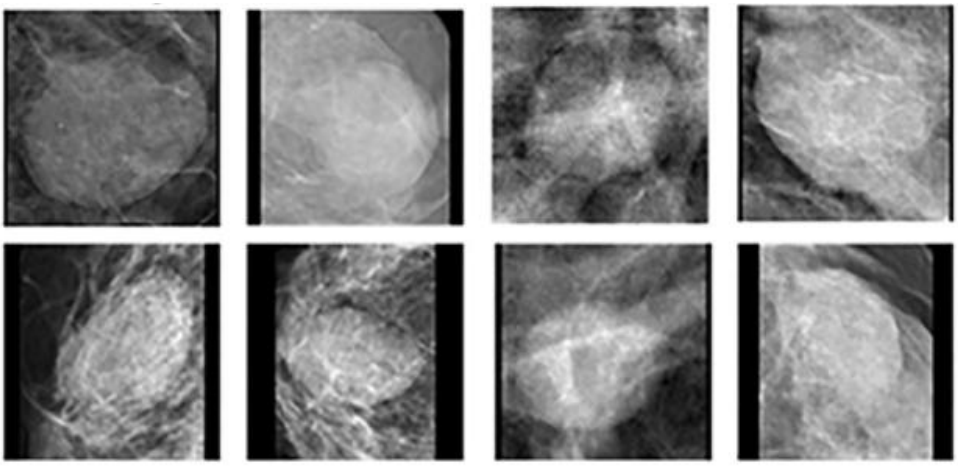


Figure 1.3. *Representative Examples of Benign Breast Masses in Mammographic Images. Adapted from Frontiers in Oncology, Figure 1, DOI:10.3389/fonc.2021.629321, licensed under CC BY 4.0.*

Malignant Calcifications and Masses: Malignant calcifications typically manifest as fine (<0.3 mm), pleomorphic microcalcifications demonstrating irregular shapes and clustered or linear/segmental distributions. These patterns exhibit a 75–90% positive predictive value for malignancy, particularly when fine linear branching configurations are present (American College of Radiology [ACR], 2023). Figure 1.4 illustrates several radiological examples of malignant microcalcifications, including diffuse high-grade DCIS and clustered pleomorphic patterns.

Malignant masses characteristically present with spiculated margins (84.6% PPV for malignancy), irregular contours, and heterogeneous density on mammography, with 89–93% of such lesions proving to be invasive carcinomas upon biopsy (Sickles et al., 2019). Representative examples of malignant breast masses are shown in Figure 1.5, demonstrating typical mammographic features associated with high malignancy risk.

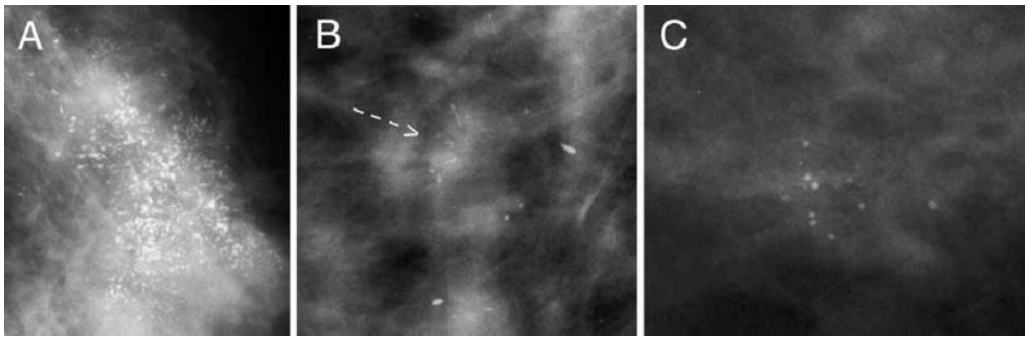


Figure1.4. Malignant Microcalcifications in the Breast
(A) Diffuse high-grade DCIS. (B) Cluster of linear pleomorphic microcalcifications (arrow). (C) Punctate microcalcifications indicating low-grade DCIS.
(Source: ResearchGate)

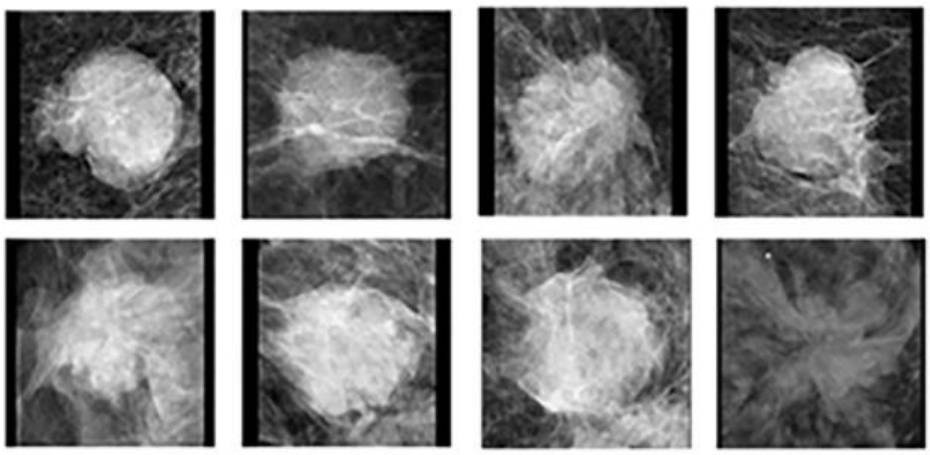


Figure1.5. Representative Examples of Malignant Breast Masses in Mammographic Images. *Adapted from Frontiers in Oncology, Figure 1, DOI:10.3389/fonc.2021.629321, licensed under CC BY 4.0.*

1.3.5. Clinical Significance of Malignant Calcifications and Malignant Masses

Malignant calcifications frequently represent the earliest detectable sign of breast neoplasia, detectable in 50-75% of DCIS cases before palpable tumor formation (D'Orsi et al.,

2013). Their identification enables diagnosis at pre-invasive stages, with 10-year survival rates exceeding 95% when treated appropriately (ACS, 2023). Malignant masses, conversely, predominantly indicate invasive disease, with mammographic features correlating strongly with pathological invasion patterns - spiculated margins reflect tumor-induced desmoplastic reactions, while irregular shapes signal infiltrative growth (Tabár et al, 1985; ACR, 2013).

1.4. AI Diagnostic Techniques for Detecting Malignant Calcifications and Masses

The integration of artificial intelligence (AI) into mammographic analysis has significantly enhanced the *automatic detection* of malignant calcifications and masses. AI-powered systems support radiologists by highlighting subtle abnormalities that may be overlooked, improving diagnostic accuracy and reducing human variability. Deep learning models, particularly convolutional neural networks (CNNs), have shown remarkable ability in automatically detecting suspicious patterns associated with malignancies in breast images.

1.4.1. Role of Deep Learning and AI in Automatic Detection

1.4.1.1. Improving Early Detection with AI

Deep learning especially CNN-based approaches has revolutionized medical imaging by enabling the automatic recognition of complex visual features in mammograms. These models contribute to more consistent and precise detection of malignancies by learning from annotated datasets and flagging potential areas of concern.

AI-powered Systems in Mammography Analysis:

- **Feature extraction :** Automatically identifies critical visual traits of calcifications and masses.
- **Detection and localization :** Pinpoints regions of interest where abnormalities may be present.
- **Anomaly recognition :** Alerts clinicians to unusual patterns indicative of potential malignancy.

AI-assisted diagnostic tools in mammographic screening reduce false negatives by 9.4% and false positives by 5.7%, significantly improving early cancer detection rates (McKinney et al., 2020)

1.4.2. Challenges in Medical Datasets

1.4.2.1. Data Limitations: Imbalance and Dataset Scale

One of the primary obstacles in developing robust AI detection systems is the imbalance in available datasets, benign cases often far outnumber malignant ones. This can lead to biased training, where AI models struggle to detect rarer malignant features. To address this, it is essential to use large-scale, balanced, and demographically diverse datasets (McKinney et al., 2020).

1.4.2.2. Barriers to Constructing High-Quality Medical Datasets:

- **Privacy and ethics :** Regulations on patient confidentiality limit open access to medical images.
- **Annotation complexity** : Accurate labeling by expert radiologists is labor-intensive and resource-consuming.

- **Heterogeneous data sources :** Variability in imaging techniques and equipment can hinder model generalization.

An emerging solution to these challenges is the use of synthetic data generation techniques, such as GANs. These models can produce realistic mammographic images, aiding in the creation of balanced datasets for more effective AI training and validation (Jiménez-Gaona et al., 2024).

1.5.Conclusion

Early detection of malignant calcifications and masses is essential for improving breast cancer prognosis. Advances in AI and deep learning have significantly enhanced mammographic analysis, reducing radiologists workload while improving diagnostic accuracy. These technologies enable more precise differentiation between benign and malignant features, supporting early intervention and better patient outcomes.

However, the effectiveness of AI models depends on the availability of large, high-quality datasets. Challenges such as data imbalance, image variability, and ethical concerns must be addressed to ensure robust model performance.

Automated AI-driven detection methods offer the potential for faster, more reliable breast cancer screening, aiding in early diagnosis and treatment. As AI continues to evolve, its integration into clinical workflows will play a transformative role in breast cancer detection, ultimately improving survival rates and healthcare efficiency.

CHAPTER 2 : Generative Adversarial Networks (GANs)

2.1. Introduction

GANs have gained significant attention in medical imaging for their ability to generate high quality synthetic images. Introduced by Goodfellow et al. in 2014 (Goodfellow et al., 2014), GANs leverage an adversarial training framework where two neural networks, the Generator and the Discriminator, compete to produce realistic data. In medical imaging, GANs help overcome challenges such as limited dataset availability and class imbalance, which are common in breast cancer diagnosis (Kazeminia et al.,2020). By generating synthetic malignant calcifications and masses, GANs can improve the robustness of deep learning models used for mammogram analysis, enhancing diagnostic accuracy and reducing overfitting (Yi et al., 2019).

This chapter explores the fundamentals of GANs, focusing on their architecture, training dynamics, and challenges such as mode collapse and stability issues. Additionally, it highlights state-of-the-art applications of GANs in medical imaging, particularly in synthetic data generation for mammography. Following this, the preprocessing steps necessary for GAN training are introduced. Finally, the evaluation metrics used to assess the quality of the generated images are presented.

2.2. Fundamentals of Generative Adversarial Networks

GANs represent a groundbreaking advancement in the field of deep learning, particularly for generative modeling. GANs have demonstrated remarkable success in generating realistic data, including images, videos, and audio. Their unique adversarial training framework has enabled applications across diverse domains, from computer vision to medical imaging (Yi et al., 2021).

2.2.1. Adversarial Training Framework

GANs are a class of generative models that learn to produce data samples indistinguishable from real data. The core idea behind GANs is based on a game-theoretic framework, where two neural networks, the generator (*G*) and the discriminator (*D*), are trained simultaneously in a competitive manner. This adversarial process drives the generator to produce increasingly realistic data, while the discriminator becomes more adept at distinguishing real from synthetic data (Goodfellow et al., 2014).Figure 2.1 illustrates a typical GAN architecture, showing the interaction between the generator, the discriminator, and the data flow during training (Skandarani, Jodoin, & Lalande, 2023).

The training process of GANs can be understood as a minimax game between *G* and *D*. The generator aims to produce synthetic data that mimics the real data distribution, while the discriminator aims to correctly classify data as real or fake. The objective function for GANs is formulated as:

$$\min_{G} \max_{D} V(D,G) = E_{\mathbf{x} \sim \mathbf{p\ data(x)}} [\, log\, D(x) \,] + E_{\mathbf{z} \sim \mathbf{p\ z(z)}} [\, log\, (1 - D(G(z))) \,] \qquad (2.1)$$

Here:

x represents real data samples drawn from the true data distribution pdata(x).
z represents random noise vectors drawn from a prior distribution pz(z) (e.g., a uniform or Gaussian distribution).
G(z) is the synthetic data generated by the generator.
D(x) and D(G(z)) are the discriminator's outputs, representing the probability that x is real and G(z) is real, respectively.

The generator seeks to minimize the objective function, while the discriminator seeks to maximize it. Through this adversarial process, the generator improves its ability to produce realistic data, and the discriminator refines its classification accuracy (Arjovsky et al., 2017).

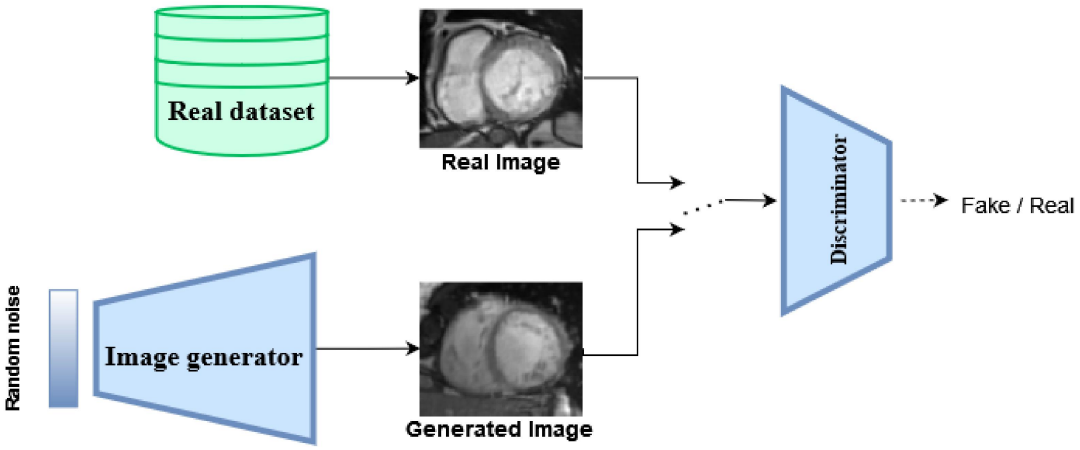


Figure2.1. *Flowchart of a traditional GAN architecture (Skandarani, Jodoin, & Lalande, 2023)*

2.2.2. Key Challenges in GAN Training

Several key challenges are central to understanding GAN training presented in the table below:

Concept	Description
Nash Equilibrium	In GAN training, the generator and discriminator are in constant competition. The training process aims to reach a Nash equilibrium, where the generator creates synthetic data indistinguishable from real data, and the discriminator's ability to differentiate between real and fake data is no better than random guessing. Achieving this balance ensures that the GAN has learned to generate high-quality, realistic samples (Goodfellow et al., 2014).
Mode Collapse	A common failure in GAN training occurs when the generator produces only a limited set of similar outputs, rather than capturing the full diversity of the real data distribution. This phenomenon, known as mode collapse, happens when the generator exploits a weakness in the discriminator by repeatedly generating a narrow subset of samples that fool it, leading to a loss of variety in the generated data (Goodfellow et al., 2014; Metz et al., 2017).
Training Instability	The adversarial nature of GAN training creates inherent instability, where the competing objectives of the generator and discriminator often lead to vanishing gradients (when the discriminator becomes too strong and provides uninformative feedback) or oscillatory behavior (where the networks fail to converge) (Arjovsky & Bottou, 2017; Metz et al., 2017). To address these challenges, researchers have developed stabilization techniques including progressive growing of networks (Karras et al., 2018), the use of Wasserstein loss to provide more meaningful gradients (Arjovsky et al., 2017), and architectural improvements like those in DCGANs that employ careful normalization and optimizer selection (Radford et al., 2016).

Table2.1. *Main key challenges of GAN.*

2.2.3. Applications of GANs

GANs have revolutionized various domains by offering a powerful framework for learning to generate realistic data. Below are key application areas where GANs have made a significant impact:

1. Image Synthesis

GANs were originally introduced to generate highly realistic images from random noise vectors, learning to mimic complex data distributions such as human faces, natural scenes, and everyday objects (Goodfellow et al., 2014). For example, GANs can synthesize lifelike images of people who do not exist, reconstruct detailed landscapes from rough sketches, or create fictional animals with plausible textures and anatomy.

Notably, StyleGAN models developed by NVIDIA (Karras et al., 2019) demonstrated unprecedented success in generating photorealistic human portraits. These models produce images with intricate details, such as realistic skin pores, hair strands, and lighting effects, making them nearly indistinguishable from genuine photographs. A landmark example is the website "This Person Does Not Exist", which uses StyleGAN2 to generate infinite unique faces on demand.

2. Data Augmentation

In domains like medical imaging and autonomous driving, acquiring large labeled datasets is often costly, time-consuming, or ethically challenging. GANs address this limitation by generating synthetic yet realistic samples to augment training data, improving model robustness and generalization for tasks like disease diagnosis or object detection (Goodfellow et al., 2014). For example, GANs can synthesize rare tumor subtypes in MRI scans or simulate diverse driving scenarios (e.g., night rain, fog) that are difficult to capture in real-world data collection.

Notably, Frid-Adar et al. (Frid-Adar et al. 2018) demonstrated the effectiveness of GANs for liver lesion classification in CT scans. Their model generated synthetic lesions (cysts, metastases, and hemangiomas) with realistic textures and shapes using DCGAN architectures and patch-based discriminators, which preserved fine-grained details like tumor margins. This approach improved classification accuracy by 11.3% on imbalanced datasets, proving GANs' utility for data-scarce applications.

3. Image-to-Image Translation

GANs enable the transformation of images between domains, such as converting sketches to photorealistic scenes or medical scans across modalities. This capability addresses challenges like data scarcity in paired datasets (e.g., lack of aligned MRI-CT scans) and enables creative or diagnostic applications (Zhu et al., 2017).

4. Super-Resolution

GANs have been successfully used to improve image quality by upscaling low-resolution images while preserving and enhancing important features. The SRGAN model (Ledig et al., 2017) produces sharper and more detailed high-resolution images compared to traditional interpolation methods. This has important applications in satellite imagery, medical scans, security, and old photo restoration.

5. Text-to-Image and Cross-Modal Applications

Advanced GAN models like StackGAN (Zhang et al., 2017) and AttnGAN (Xu et al.,

2018) generate realistic images from text descriptions by aligning semantic concepts with visual features. These models enable applications such as product visualization in e-commerce ("blue leather handbag"), automated storyboard creation for films/games, and assistive tools for the visually impaired. AttnGAN further refines outputs using attention mechanisms to focus on specific textual details, enhancing image fidelity.

6. Video Generation and Animation

GANs are also being explored for video frame prediction, generating synthetic videos, or animating facial expressions and movements. These models are applied in cinematography (e.g., pre-visualizing scenes), gaming (e.g., procedural NPC animations), and avatar animation (e.g., lifelike virtual influencers) (Vondrick et al., 2016; Tulyakov et al., 2018).

2.3. Architectural Components: Generator and Discriminator

The success of GANs relies on the design and interaction of their two primary components: the **generator** and the **discriminator**. Each component plays a distinct role in the adversarial training process.

2.3.1. Generator

The generator (G) is a neural network that transforms random noise vectors (z) from a prior distribution (p_z) into synthetic data samples (G(z)) designed to mimic the real data distribution (p_{data}). Its architecture is critical for generating high-fidelity outputs while maintaining training stability. Below, we explain its core components and design principles.

Architectural Components:

1. **Input Layer**

   - Accepts a latent vector $z \in \mathbb{R}^d$, typically sampled from a Gaussian (N(0,1)) or uniform distribution.
   - The dimensionality (d) of z determines the generator's capacity to model complex data distributions (Goodfellow et al., 2014).

2. **Transposed Convolutional Layers**

   - Also termed deconvolutional layers, these upsample z into higher-dimensional outputs (e.g., images). Each layer applies strided fractional convolutions to progressively increase spatial resolution (Radford et al., 2016).
   - For example, in DCGAN (Radford et al., 2016), a 100-dimensional noise vector is upsampled to a 64×64×3 RGB image through four transposed convolutional layers.

3. **Normalization and Activation Functions**

   - **Batch Normalization**: This technique helps make training more stable and faster by standardizing the output of a layer so that it has a mean of zero and a standard deviation of one. This reduces internal covariate shift, making the model more efficient (Ioffe & Szegedy, 2015).

   - **Leaky ReLU**: Unlike the standard ReLU function, which can cause some neurons to stop learning completely (known as the "dying ReLU" problem), Leaky ReLU

allows a small, non-zero gradient when the input is negative. This keeps neurons active and learning. It is defined as:

$$f(x)=\max(\alpha x,x), \text{ where } \alpha=0.2 \qquad (2.2)$$

(Radford et al., 2016).

4. **Output Layer**

   - **Tanh Activation:** In image generation tasks, the output layer usually uses a tanh activation function. This function ensures that pixel values are scaled between -1 and 1, which matches the range of normalized image data. This helps the model produce images with realistic brightness and contrast levels (Radford et al., 2016).

2.3.2. Discriminator

The discriminator (D) functions as a binary classifier, distinguishing real data ($x\sim p_{data}$) from synthetic samples (G(z)) produced by the generator. Its architecture is designed to extract discriminative features while maintaining training stability. Below, we detail its components, training dynamics, and design principles.

Architectural Components :

1. **Input Layer**

   - Accepts data samples (real or synthetic) with dimensions matching the training data (e.g., 64×64×3 for RGB images).

2. **Convolutional Layers**

   - Hierarchical convolutional layers downsample inputs, extracting spatial features (e.g., edges, textures) through strided convolutions (Radford et al., 2016).

     Example: DCGAN's discriminator uses four convolutional layers to reduce a 64×64 image to a scalar probability (Radford et al., 2016).

3. **Normalization and Activation Functions**

   - **Leaky ReLU**: introduces non-linearity in the network while addressing the issue of "dying ReLUs," where neurons become inactive. By allowing a small, non-zero gradient (typically with α=0.2) for negative inputs, Leaky ReLU maintains gradient flow even when the activation is negative, which is critical in adversarial training (Goodfellow et al., 2014).
   - **Spectral Normalization**: is applied to each layer to control the Lipschitz constant of the network. By constraining the spectral norm of the weight matrices, this technique helps to stabilize training and prevent the discriminator from becoming too powerful, which would otherwise lead to vanishing gradients for the generator (Miyato et al., 2018). It's especially important in GANs, where the balance between generator and discriminator is delicate.

4. **Output Layer**

   - A sigmoid activation function is used to squash the output into a probability range between 0 and 1. This output, denoted as D(x), represents the discriminator's confidence that input x comes from the real data distribution rather than from the generator. A value close to 1 implies real, while a value near 0 implies synthetic.

2.4. Variants of GANs

Over time, GANs have undergone significant advancements, leading to the development of various specialized variants tailored to address specific challenges and applications. The table below provides a comparative overview of some of the most widely used GAN architectures, highlighting their key innovations, advantages, and applications in medical imaging:

GAN Architecture	Key Innovations	Advantages	Medical Imaging Applications
DCGAN (Deep Convolutional GAN) Radford et al.,(2016)	Uses convolutional layers instead of fully connected layers, batch normalization for stability,and transposed convolutions for upsampling.	Improves image generation quality and training stability; widely used as a baseline.	Used for generating synthetic medical images to augment datasets and enhance deep learning models.
CycleGAN by Zhu et al. (2017)	Introduces cycle-consistency loss for unpaired image-to-image translation.	Effective for transforming medical images across modalities (e.g., MRI to CT).	Used in cross-modality image synthesis, such as MRI-to-CT conversion and histopathology stain translation.
cGAN (Conditional GAN) by Mehdi Mirza et al.,(2014)	Incorporates conditional information (e.g., class labels) into GAN training.	Allows controlled generation of images with specific characteristics.	Used to synthesize medical images with specific tumor features, aiding in data augmentation.
StyleGAN by Karras et al. (2019)	Uses a style-based generator, intermediate latent space, AdaIN (Adaptive Instance Normalization), progressive growing, and stochastic noise injection.	Generates high-resolution, diverse, and photorealistic images; allows fine-grained control over image attributes.	Applied in synthetic mammogram generation to improve training datasets, enhance anomaly detection, and study rare pathologies.

Table2.2. *comparative overview of some of the most widely used GAN architectures.*

2.5.State-of-the-Art Applications of GANs in Medical Imaging

Their unique ability to synthesize realistic data, enhance image resolution, and translate between imaging modalities has positioned GANs as indispensable tools for advancing diagnostic accuracy, research reproducibility, and clinical workflows. This section explores cutting-edge applications of GANs in medical imaging, beginning with synthetic image generation a cornerstone for overcoming data limitations in AI-driven healthcare:

2.5.1. Synthetic Image Generation

GANs are widely used to synthesize realistic medical images for training, testing, and validating deep learning models. Shin et al. (2018) applied CycleGAN to generate brain MRI scans with synthetic tumors, augmenting datasets for tumor segmentation tasks. Their results showed a 7% improvement in Dice score compared to models trained exclusively on real-world data. Similarly, Frid-Adar et al. (2018) demonstrated that GAN-generated synthetic liver lesions in CT scans improved lesion classification accuracy by 11.3% in imbalanced datasets, addressing critical challenges in rare pathology detection.

In breast imaging, Korkinof et al. (Korkinof et al. 2018) utilized Progressive GANs to synthesize mammograms with varying breast densities, enhancing the robustness of cancer detection models while preserving patient privacy. These applications highlight GANs' ability to overcome data scarcity, class imbalance, and privacy constraints in medical imaging.

2.5.2. Data Augmentation

GANs enhance medical imaging datasets by synthesizing diverse variations of existing images, addressing challenges like limited data and class imbalance. For instance, Bowles et al. (Bowles et al. 2018) augmented brain MRI datasets for Alzheimer's disease classification by generating synthetic scans with variations in cortical atrophy, improving classification accuracy by 12%. Similarly, Zhang et al. (Zhang et al. 2020) employed DCGANs to synthesize mammograms with tumors of varying sizes, shapes, and densities, reducing false-negative rates in breast cancer detection by 9% when models were trained on augmented datasets. These applications demonstrate GANs' ability to improve model robustness and generalizability in critical diagnostic tasks.

2.5.3. Image-to-Image Translation

Image-to-image translation involves using GANs to convert images from one domain to another, such as translating MRI scans to CT scans or enhancing low-quality images. Zhu et al. (Zhu et al. 2017) introduced CycleGAN for unpaired image-to-image translation, successfully converting MRI scans to CT scans with a structural similarity index (SSIM) of 0.89 compared to real CT scans. In breast cancer research, Dar et al. demonstrated that conditional GAN (cGAN) could generate synthetic MRI images from mammograms, enabling the use of MRI-based diagnostic tools without the need for expensive and time-consuming MRI scans (Dar et al., 2019).

2.6. GANs in Breast Cancer Research

GANs have emerged as a transformative technology in breast cancer research, offering innovative solutions to longstanding challenges in medical imaging and diagnostics, State-of-the-art applications in synthetic mammogram generation and lesion synthesis are discussed below.

2.6.1. Synthetic Mammogram Generation

Synthetic mammogram generation is one of the most widely explored applications of GANs in breast cancer research. By generating realistic synthetic mammograms, GANs can

augment training datasets, improve the performance of diagnostic models, and address the challenge of limited data availability. For instance, Korkinof et al. (Korkinof et al. 2018) utilized a Progressive GAN architecture to synthesize high-resolution mammograms with varying breast densities, achieving a structural similarity index (SSIM) of 0.92 compared to real images. Similarly, Zhang et al. (Zhang et al. 2020) employed CycleGAN to generate synthetic mammograms with tumors of varying sizes and shapes, demonstrating a 15% improvement in tumor detection accuracy when models were trained on augmented datasets.

2.6.2. Lesion Synthesis for Enhanced Classification

Lesion synthesis, the artificial generation of pathological features like tumors or masses, addresses a critical challenge in medical AI: training robust diagnostic models when real-world lesion data is scarce or imbalanced . By leveraging GANs to synthesize lesions with diverse sizes, shapes, and anatomical placements, researchers augment datasets in ways that mimic clinical variability, significantly improving model generalizability.

For instance, Korkinof et al. (2018) demonstrated the efficacy of this approach in mammography. Using conditional GANs (cGANs) conditioned on lesion attributes (e.g., margins, density), their model generated realistic microcalcifications and masses. This synthetic data enabled mammography AI tools to achieve a sensitivity of 92% (correct lesion identification) and specificity of 88% (correct exclusion of healthy tissue), markedly outperforming models trained on limited authentic data.

2.6.3. Comparative Analysis of Studies

A comparative analysis of studies applying GANs to breast cancer research reveals several key trends and challenges. Table 2.3 summarizes the key studies, including the GAN architectures employed, datasets used, key parameters, evaluation metrics, and outcomes.

Study	GAN Architecture	Dataset	Parameters	Evaluation Metrics	Key Outcomes
High-resolution mammogram synthesis using progressive generative adversarial networks. **Korkinof et al. (2018)**	Progressive GAN	DDSM (Digital Database for Screening Mammography)	Learning rate: 0.0002, Batch size: 32	SSIM: 0.92, PSNR: 28.5 dB	Generated synthetic mammograms with varying breast densities, improving data augmentation for training deep learning models.
Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network. **Singh et al. (2018)**	Conditional GAN (cGAN)	INbreast, Private Dataset	Not specified	Dice Coefficient: 94%, IoU: 87%	Achieved high-accuracy segmentation of breast tumors and classified tumor shapes effectively.

Study	GAN Architecture	Dataset	Parameters	Evaluation Metrics	Key Outcomes
Identifying women with mammographically-occult breast cancer leveraging GAN-simulated mammograms. **Lee et al. (2021)**	Conditional GAN (cGAN)	Private Mammography Dataset	Not specified	AUC: 0.77	Utilized CGAN-simulated mammograms to detect mammographically-occult breast cancer, enhancing detection in dense breast tissues.
Normalization of breast MRIs using cycle-consistent generative adversarial networks. **Modanwal et al. (2021)**	CycleGAN	Breast MRI from GE and Siemens Scanners	Not specified	Structural Similarity Index (SSIM): Improved	Normalized breast MRI images across different scanners, enhancing consistency for downstream analysis.

***Table2.3**. Comparative Analysis of Studies Applying GANs in Breast Cancer Research.*

2.7.Evaluation Metrics for GAN Performance

Evaluating the quality of synthetic images generated by GANs is crucial, especially in sensitive applications such as medical imaging. In this study, three widely adopted evaluation metrics were used: Fréchet Inception Distance (FID), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index Measure (SSIM). These metrics capture different aspects of image quality: perceptual realism, pixel-wise fidelity, and structural consistency.

1. Fréchet Inception Distance (FID)

FID measures the realism and diversity of synthetic images by comparing the feature distributions of real and generated images. Instead of comparing pixel values directly, it uses high-level features extracted from a pretrained InceptionV3 network (third version), simulating human perception more closely (Heusel et al., 2017).

How it works:

- Both real and generated images are passed through the InceptionV3 model (without the top layer).
- Features from an intermediate layer are extracted (typically after average pooling).
- These features are assumed to follow a multivariate Gaussian distribution.
- The FID score computes the distance between the two Gaussian distributions of real and generated features.

$$FID = \|\mu_r - \mu_g\|^2 + \mathrm{Tr}\left(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{1/2}\right) \qquad (2.3)$$

Where:

- μ_r , Σ_r **= Mean and covariance of features from real images**

- **μ_g , Σ_g= Mean and covariance of features from generated images**

Lower FID means better performance; the generated images are statistically closer to the real ones. FID captures both quality (realism) and diversity (mode collapse detection).

2. Peak Signal-to-Noise Ratio (PSNR)

PSNR measures the **pixel-level fidelity** between a generated image and its ground-truth reference (real image). It is commonly used to evaluate how close a synthetic image is to the original in terms of intensity values.

How it works:

It is based on the Mean Squared Error (MSE) between the real and generated images. A high PSNR value means the images are very similar in terms of pixel intensity.

Formula:

$$\text{PSNR} = 10 \cdot \log_{10}\left(\frac{{MAX_I}^2}{MSE}\right) \qquad (2.3)$$

Where:

${MAX_I}^2$ = Maximum possible pixel value (usually 255 for 8-bit images)
MSE = Mean Squared Error between the generated and real images

$$\text{MSE} = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[\boldsymbol{I}(\boldsymbol{i},\boldsymbol{j}) - \boldsymbol{K}(\boldsymbol{i},\boldsymbol{j})]^2 \qquad (2.4)$$

Where I and K are the original and generated images.
Higher PSNR = Better quality (less distortion), However PSNR is sensitive to noise and doesn't reflect perceptual quality or structural similarity.

3. Structural Similarity Index Measure (SSIM)

SSIM measures perceived quality by assessing structural similarity between two images, focusing on how consistent they are in terms of luminance, contrast, and structure factors that are more aligned with human vision

How it works:

It divides images into patches and compares the local structure in corresponding patches from the real and generated images.

Formula:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x{}^2+\mu_y{}^2+C_1)(\sigma_x{}^2+\sigma_y{}^2+C_2)} \qquad (2.5)$$

Where:

- μ_x, μ_y = Mean intensity of real and generated images

- σ_x^2, σ_y^2 = Variances
- σ_{xy} = Covariance between images
- C_1, C_2 = Stabilizing constants

SSIM ranges from **-1 to 1**

1 = perfect similarity
0 = no similarity
Higher SSIM = Better structural preservation

2.8. Conclusion

This chapter provided a comprehensive exploration of GANs, emphasizing their architectural foundations, adversarial training dynamics, and transformative potential in medical imaging—particularly in mammography. Key GAN variants such as DCGAN, CycleGAN, cGAN, and StyleGAN were analyzed for their ability to address critical challenges in healthcare, including data scarcity, class imbalance, and the need for high-fidelity synthetic data. Applications such as synthetic image generation, data augmentation, and image quality enhancement were highlighted, demonstrating how GANs enable robust AI model training and improve diagnostic accuracy in breast cancer research.

The chapter also reviewed evaluation metrics (e.g., FID, SSIM, PSNR) essential for quantifying the realism, diversity, and clinical relevance of GAN-generated outputs. These metrics ensure that synthetic data meets the stringent quality standards required for medical applications, bridging the gap between computational innovation and clinical utility.

The next chapter will present the methodology and dataset for the proposed GAN-based framework, detailing its application to mammography enhancement and validation against clinical benchmarks.

CHAPTER 3 : Methodology and Dataset

3.1.Introduction

This chapter presents the methodological framework and datasets used to develop, train, and evaluate the GAN-based model for generating synthetic mammographic images. The overall framework, illustrated in Figure 3.1 , consists of four main stages: dataset preparation, preprocessing, GAN training, and image classification.

The chapter begins with a detailed overview of the CBIS-DDSM dataset, emphasizing the selection criteria used to isolate malignant cases for focused generation. Preprocessing steps are then described, including normalization and resizing, which were essential for standardizing anatomical features and improving training efficiency.

The architecture of the proposed GAN is introduced, highlighting the specific roles and structures of both the Generator and the Discriminator. The training environment, constrained by limited. computational resources—and the choice of hyperparameters are also outlined. And to improve the quality and stability of training, techniques such as progressive resolution scaling, extended training cycles, and architectural adjustments were implemented.

The chapter concludes with a description of the classification pipeline, which integrates ResNet50 as a feature extractor with a Support Vector Machine (SVM) classifier. This step assesses the impact of the generated images on classification performance by comparing results obtained with and without the inclusion of synthetic data. The analysis investigates whether replacing a portion of real malignant samples with generated images maintains classification accuracy or introduces significant performance degradation.

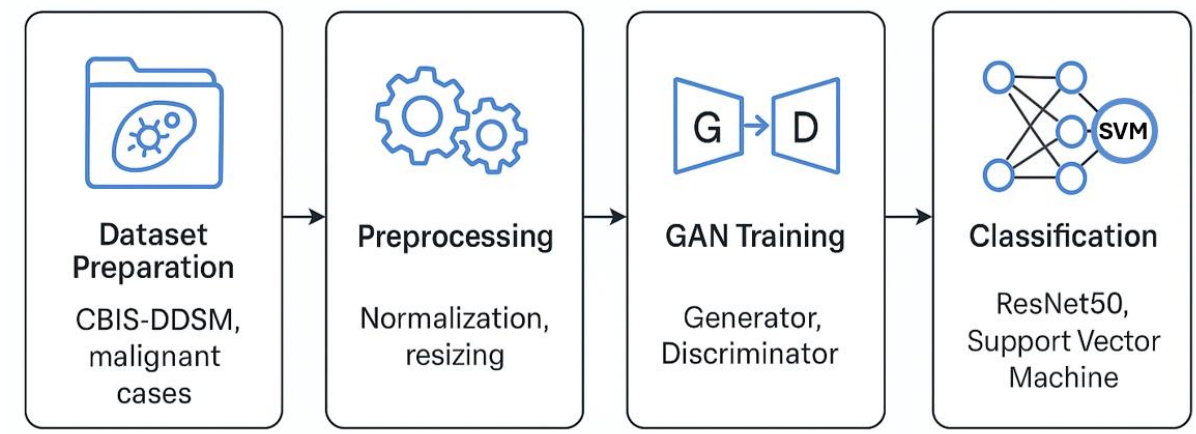


Figure 3.1. *Overall Methodological Framework.*

3.2.Dataset Description

The dataset utilized in this study is the Curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS-DDSM) (White, 2023), which is widely recognized in breast cancer research. This dataset includes thousands of annotated mammographic images, encompassing both benign and malignant calcifications and masses. Each image is accompanied by comprehensive metadata, such as lesion type, location, and pathology-confirmed diagnosis, making it an invaluable resource for training deep learning models. For this study, the CBIS-DDSM dataset was filtered and preprocessed to extract images specifically containing malignant calcifications and malignant masses. And this curated subset was used for training, ensuring that the model concentrated on clinically relevant features.

3.2.1. Preprocessing Steps

To ensure high-quality input data for training the GAN model, several preprocessing steps were applied to the CBIS-DDSM dataset:

- **Dataset Filtering:** The original extracted subset from the CBIS-DDSM dataset contained a total of 1,570 grayscale mammography (malignant masses and calcifications) images. To maintain anatomical consistency and improve training quality, a shape-based filtering approach was applied. Only images exhibiting a well-defined and closed breast contour were retained. This filtering was based on a shape confidence threshold of 0.75, resulting in a final curated dataset of 500 images used for training.
- **Breast Region Extraction:** Image processing techniques were utilized to isolate the breast region from each mammogram. This step removed irrelevant background noise and scanning artifacts, helping the model focus on the breast tissue containing potential abnormalities.
- **Resizing:** All images were uniformly resized to 256×256 pixels. This standardization ensured compatibility with the GAN architecture and reduced computational complexity.
- **Normalization:** Pixel intensity values were normalized to a standard range (typically [−1,1]), which helped maintain contrast consistency and improved model convergence during training.
- **Image Orientation Standardization:** To achieve uniformity across the dataset, images were horizontally flipped when necessary to align all mammograms in a consistent orientation (e.g., always pointing left to right). This step minimized positional variance and allowed the model to focus on learning patterns related to malignant calcifications and masses rather than spatial discrepancies.
- **Dataset Structuring:** After preprocessing, all images were systematically organized into a unified training folder structure and loaded using custom data loaders to feed into the GAN training pipeline

3.3. Proposed GAN Architecture

In this work, a Generative Adversarial Network (GAN) is proposed to generate high-quality synthetic mammography images. The architecture follows the fundamental GAN structure, consisting of two neural networks: the **Generator** and the **Discriminator**, which compete in an adversarial training process.

3.3.1. Generator Architecture

The Generator network maps a latent vector (random noise) to a synthetic image. It begins with a dense layer that transforms a 128-dimensional latent vector into a spatial tensor of shape (8, 8, 256). This is followed by a series of 4 upsampling blocks, each consisting of:

- ***Upsampling layers to increase spatial resolution.***
- **Conv2D layers for feature refinement.**
- **Batch Normalization to stabilize training and promote faster convergence.**
- **ReLU activation functions to introduce non-linearity.**

The spatial dimensions progressively increase through the pipeline:

- **(16, 16, 128)**
- **(32, 32, 128)**

- **(64, 64, 128)**
- **(128, 128, 64)**

The final Conv2D layer uses 1 filter with a Tanh activation to output a grayscale image of size (128, 128, 1) with pixel values normalized between −1 and 1.

3.3.2. Discriminator Architecture

The Discriminator is a convolutional neural network designed to classify images as real or fake. It receives a grayscale image of shape (128, 128, 1) and processes it through a series of convolutional blocks with increasing depth:

- **Conv2D (32 filters, stride 2) → (64, 64, 32)**
- **Conv2D (64 filters, stride 2) → (32, 32, 64)**
- **Conv2D (128 filters, stride 2) → (16, 16, 128)**
- **Conv2D (256 filters, stride 2) → (8, 8, 256)**

Each convolutional layer is followed by:

- **LeakyReLU activation, which helps avoid the vanishing gradient problem.**
- **Dropout layers, added for regularization and to reduce overfitting.**

The final feature maps are flattened and passed through a dense layer with a sigmoid activation, producing a probability score representing whether the input image is real or generated. Figure 3.2 presents a diagram of the chosen GAN architecture.

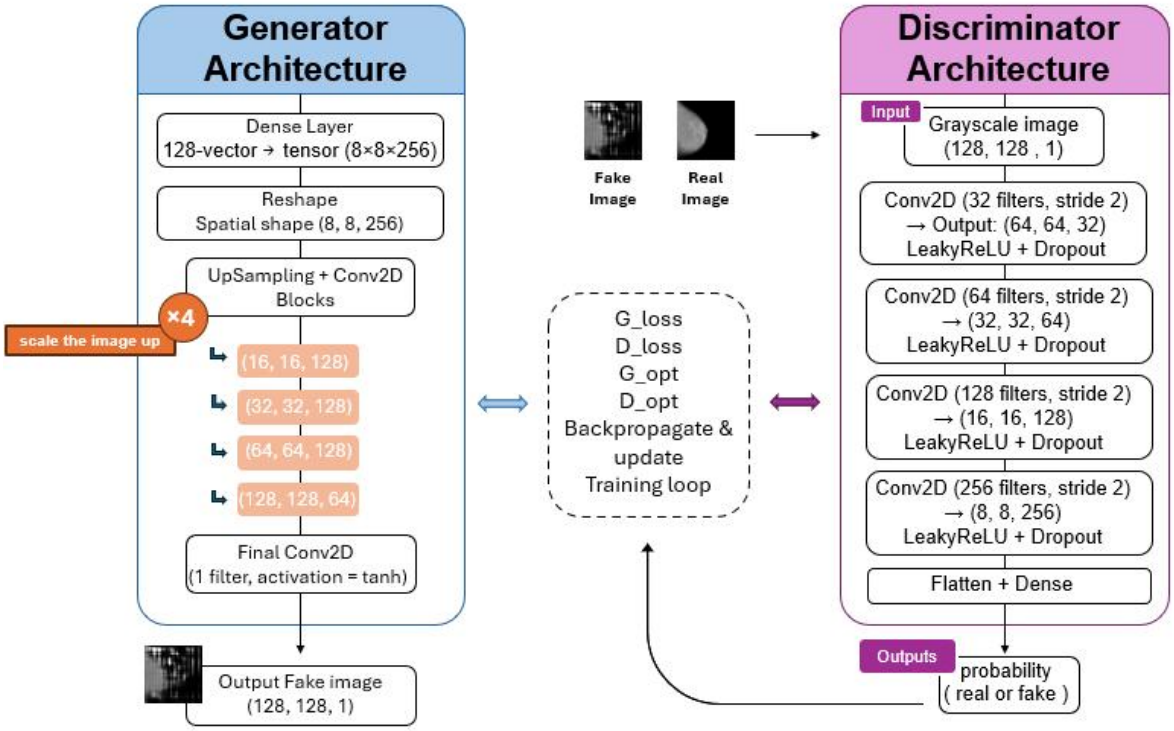


Figure 3.2. *illustrates the custom architecture of the generator and discriminator used in this study.*

3.3.3. Training Process

The training loop involves alternately updating:

- **The Discriminator**, using real and fake images to minimize binary cross-entropy loss.
- **The Generator**, aiming to fool the Discriminator by generating increasingly realistic images.

The adversarial loss functions used include:

- **G_loss**: Measures how well the Generator fools the Discriminator.
- **D_loss**: Measures how well the Discriminator distinguishes real from fake.

3.4. Training Configuration and Environment

3.4.1. Computational Platforms

The training of the proposed GAN model was performed using two cloud-based platforms:

- Google Colab was initially used for training. While it provides free access to GPUs, the limited session duration and slower performance made it less ideal for longer training sessions. Additionally, the time restrictions and resource availability often interrupted the training process, affecting stability.
- Kaggle Notebooks were used for the majority of the experiments. Kaggle offers a more stable and faster free GPU environment (typically Tesla P100 or T4 GPUs), making it more efficient for extended training sessions. It also allowed consistent training without frequent disconnections or timeouts.

3.4.2. Resource Constraints

Both platforms used the free tier, which imposed limitations on:

- GPU runtime per session
- Daily compute limits
- Limited GPU memory, requiring careful management of batch sizes and model complexity.

Despite these constraints, effective model training was achieved by structuring the workflow into smaller, manageable segments and saving checkpoints at regular intervals.

3.4.3. Parameter Settings

To achieve optimal performance and stability during the adversarial training process, specific hyperparameters were carefully selected for the GAN model. These parameters significantly influence convergence speed, image quality, and overall training behavior.

- **Epochs:**
  Training was conducted for up to 20 000 epochs in several configurations. An epoch refers to one full pass through the entire training dataset. Increasing the number of epochs allows the model to learn more refined features and generate higher-quality images. In GAN training, longer epochs can help the Generator improve its ability to "fool" the Discriminator while also allowing the Discriminator to adapt, resulting in a dynamic learning balance. However, longer training also increases the risk of overfitting or mode collapse if not carefully managed.

- **BatchSize:**
  The batch size was set to 32, meaning the model processes 32 images at a time before updating the weights. This value strikes a balance between memory efficiency and training stability. A small batch size could lead to noisy gradient updates and unstable learning, while a very large batch might exceed GPU memory limits, especially on free cloud platforms like Kaggle. Batch size also directly affects how frequently the model updates, which can influence convergence dynamics in adversarial setups.

- **Optimizer:**
  Both the Generator and Discriminator were trained using the Adam optimizer, which is widely used in deep learning and particularly well-suited for GANs. Adam combines the advantages of two other extensions of stochastic gradient descent: Adaptive Gradient Algorithm (AdaGrad) and Root Mean Square Propagation (RMSProp). It maintains per-parameter learning rates that adapt during training, which is particularly important in GANs due to their highly sensitive loss functions.

  The Adam settings were:

  Learning rate = 1e-4 for Generator and Learning rate = 1e-5 for Discriminator:

  This small value allows for slow, stable updates, which is crucial for GANs. Larger learning rates could lead to unstable oscillations or divergence in the adversarial game.

  β1 = 0.5:

  The β1 parameter controls the exponential decay rate for the first moment estimate (i.e., momentum). Setting β1 to 0.5, as recommended in the DCGAN paper by (Radford et al. 2016), helps stabilize training by reducing momentum, thereby encouraging the optimizer to adapt more quickly to the changing gradients in the adversarial process.

These hyperparameters were not only based on literature best practices but also empirically tuned based on early training results and visual inspection of generated image quality. The consistent use of these settings contributed significantly to stabilizing the training and improving the realism of the synthetic mammography images.

3.5. Training Strategies and Stability Handling

Training GANs can be unstable due to issues like mode collapse and non-convergence. Several strategies were implemented to improve training outcomes.

3.5.1. Multiple Image Resolutions

To enhance training stability and improve image quality, the output resolution of the Generator was progressively increased during training. The model initially generated images at 64×64 resolution, then at 128×128, and finally at 256×256. This coarse-to-fine training strategy enabled the GAN to learn structural features in a staged manner, promoting more stable convergence and more realistic outputs.

• 64×64: Provided initial stability but lacked sufficient mammographic detail.
• 128×128: Offered the best trade-off between visual quality and training stability.

• 256×256: Introduced significant instability due to increased complexity and resource demands. In this stage, the training often failed, and the generated images were predominantly black.

Therefore, the final output resolution was set to 128×128, as it provided the best compromise between visual quality and training stability.

3.5.2. Extended Training Duration

To improve the Generator's ability to produce realistic images, the number of training epochs was gradually increased, with the most successful results observed at 17 000 - 20 000 epochs. Longer training helped the GAN overcome early instability and learn complex features more effectively.

3.5.3. Stability Techniques

To mitigate mode collapse and ensure training convergence, the following strategies were applied:

- **Batch Normalization** in the Generator to promote smoother gradients.
- **Dropout Layers** in the Discriminator to regularize learning and prevent overfitting.
- **LeakyReLU Activation** in the Discriminator to maintain gradient flow.
- **Tanh Activation** in the Generator's output to normalize the image range.
- **learning rate tuning** to avoid divergence in the adversarial process.

3.6. Classification Step

To evaluate the effectiveness of the dataset and the generated images, a supervised classification task was performed. The objective was to classify mammography patches into two categories:

Tumor: Includes real malignant cases.

Non_tumor: Includes healthy tissue, sourced from the *Final-Final-RSNA-Breast-Cancer-Dataset* available on Kaggle (White, 2023).

The classifier used was a Support Vector Machine (SVM) with a linear kernel, trained on features extracted using a pretrained ResNet50 model from the PyTorch library.

3.6.1. Feature Extraction

The classification was not performed directly on pixel values. Instead, high-level features were extracted from each image using a ResNet50 model truncated before the final classification layer (i.e., the global average pooling output). This approach enables better generalization by using powerful image embeddings.

The image pipeline involved:

Resizing all images to 224×224

Normalizing with ImageNet statistics

Extracting features from the penultimate layer of ResNet50

Feeding features to the SVM classifier

3.6.2. Data Configuration and Class Balance

Two separate experiments were designed:

Real-only training:

400 tumor and 400 non_tumor images for training

100 tumor + 100 non_tumor images for testing

Generated-enhanced training:

300 real tumor + 100 generated tumor and 400 non_tumor for training

Same test set as above

This setup ensures a fair comparison between the two configurations by maintaining class balance and preventing overlap between training and testing samples.

3.6.3. Objective and Evaluation Strategy

The classification pipeline integrates ResNet50 as a deep feature extractor with an SVM classifier, serving as a quantitative evaluation method for the usefulness of generated images. The goal of this step is to assess whether replacing a portion of real malignant samples with GAN-generated images preserves or degrades classification performance.

By comparing results obtained with and without the inclusion of synthetic data, the analysis aims to determine the impact of generation quality on real-world classification tasks. This step also provides insights into whether high-quality synthetic images can serve as a viable augmentation strategy in scenarios where real data is limited.

3.7. Conclusion

This chapter outlined the full methodology used to generate and evaluate synthetic mammography images. It covered dataset selection, preprocessing, GAN architecture design, and training strategies. Progressive resolution training helped improve learning stability, with 128×128 identified as the optimal resolution.

Finally, a classification step was introduced using ResNet50 for feature extraction and an SVM classifier. Two scenarios were tested, one using only real data, and another combining real and generated tumor images to assess the impact of synthetic data on classification performance. This approach allowed for a practical evaluation of the GAN's effectiveness in a downstream medical task.

CHAPTER 4 : Results and Achievements

4.1. Introduction

This chapter presents and analyzes the results obtained from training the GAN-based model for mammography image generation and the subsequent classification experiments. The conducted experiments aimed to evaluate how architectural variations, training strategies, and computational limitations influenced the model's ability to generate realistic and clinically meaningful images. Both visual inspection and training behavior were used to assess convergence stability and image quality across different configurations. In addition, the impact of integrating generated images into a supervised classification task was investigated. This evaluation, performed using a ResNet50 feature extractor and a linear SVM classifier, served to determine whether synthetic data could effectively replace a portion of real malignant samples without compromising classification accuracy. The findings highlight the best-performing setups and provide insight into the trade-offs between image fidelity, model stability, and downstream performance.

4.2. Training Results and Convergence Behavior

The training process was conducted through three main experiments. Each aimed to evaluate how specific architectural or hyperparameter modifications impacted the model's convergence and the visual quality of the generated images. The number of training epochs was gradually increased in each experiment, starting from 1,000, then 5,000, 10,000, and finally 20,000 to monitor training stability and identify the optimal configuration.

4.2.1. Experiment 1: Deeper Generator with Additional Convolutional Layers

In the first experiment, the baseline generator architecture was modified by adding multiple convolutional layers after each upsampling step. This resulted in a deeper and more complex structure, where each upsampling stage was followed by repeated Conv2D → LeakyReLU → BatchNormalization blocks. The generator started from a 128-dimensional latent vector, reshaped into an 8×8×512 feature map, and gradually upsampled to 16×16, 32×32, 64×64, and ultimately 128×128 resolution. Each stage increased in channel depth and spatial complexity.And the objective was to Improve image realism through added network depth and feature refinement.

Results:

Despite the added depth and progressive structure, the model failed to outperform the simpler baseline. Visual results were less realistic, and the model demonstrated signs of overfitting. Training instability became more apparent, likely due to the limited size of the training dataset. The increased capacity of the generator led to excessive memorization and poor generalization, reducing overall image diversity.

More depth does not guarantee better performance, especially with limited training data. The deeper architecture was sensitive to overfitting and lacked robustness.

4.2.2. Experiment 2: StyleGAN-Based Architecture

The second experiment tested a more advanced generative approach by incorporating a StyleGAN-like architecture. This state-of-the-art model introduces style modulation and adaptive instance normalization to control image generation at multiple scales. And the objective was Leverage modern GAN advancements to achieve higher fidelity and realism.

Results:

The model underperformed significantly compared to the baseline. Generated images lacked

coherence, and training was unstable. The StyleGAN approach, known for its high performance in large datasets, was too data-hungry for the limited dataset used in this study. Furthermore, the style mixing mechanisms did not offer any advantage due to the insufficient diversity in training samples.

Advanced GAN architectures like StyleGAN require large datasets and complex augmentation strategies. Without these, performance deteriorates, demonstrating that such models are not suitable for small-scale training scenarios.

4.2.3. Experiment 3: Baseline Generator with Dropout Regularization and Extended Training

The third and final experiment returned to the baseline generator architecture but introduced two key changes:

- Reduced the dropout rate in the discriminator from 0.3 to 0.2 to retain more information during learning.
- Increased the number of training epochs up to 20,000 to allow the model to converge more effectively.

Generator Settings :

- Architecture: Same as the initial baseline (no additional layers).
- Input: 128-dimensional latent vector.
- Output: 128×128 grayscale images.
- Activation Function: Tanh.
- Optimizer: Adam (Generator: learning rate = 1e-4, Discriminator: 1e-5).
- Loss Function: Binary Crossentropy.

Training Strategy:

To address instability, the training was conducted progressively, increasing the number of epochs across trials: 1,000, 5,000, 10,000, and finally 20,000. Kaggle's GPU environment allowed these extended sessions.

Results:

This configuration produced the best results among all experiments. The generated images showed improved structure, contrast, and detail preservation. Reducing the dropout rate helped the discriminator learn more representative features, minimizing over-regularization. Furthermore, the longer training duration allowed both networks to stabilize and reach a more balanced equilibrium. The FID also improved, indicating closer alignment with real data distributions.

Simplifying the architecture and focusing on stable training with proper regularization and sufficient training time proved to be the most effective approach. This experiment demonstrated that quality results can still be achieved with a simple GAN setup when training dynamics are well-tuned.

4.3.Summary of Results:

The Table below summarize the results of the three experiments :

Experiment	Architecture Change	Epochs	Dropout	FID Improvement	Visual Quality	Conclusion
Exp 1	Added Conv layers	1k–10k	0.3	Poor	Low	Overfitting due to depth
Exp 2	StyleGAN	1k–10k	0.3	Poor	Very Low	Failed due to data scarcity
Exp 3	Baseline + Dropout 0.2	20k	0.2	Improved	Good	Best configuration overall

Table 4.1. *Summary of results for the Three Experiments.*

4.4.Loss Graph Analysis

4.4.1. Generator and Discriminator Loss Curves

Throughout all experiments, both the generator (G) and discriminator (D) losses were monitored across training epochs to assess convergence behavior and training stability. Figure 4.1 shows the evolution of generator and discriminator losses during training, providing visual insight into the adversarial dynamics across different stages of the process. The loss patterns observed were consistent with known GAN dynamics, especially in the best-performing experiment (Experiment 3), which was trained up to 20,000 epochs on Kaggle's GPU.

In the early training stages (epochs 0–700), the generator loss increased rapidly reaching values as high as 2.5 while the discriminator loss dropped to approximately 0.25. This imbalance, where the generator loss was around 2.25 times higher than the discriminator loss, is a common phenomenon in GAN training and reflects the generator's initial struggle to produce convincing outputs while the discriminator confidently distinguishes real from fake images.

However, around epoch 700, both losses began to converge toward each other, stabilizing around 0.65–0.69 with slight oscillations (maximum observed difference ~0.1). This phase is often interpreted as a sign of improved adversarial balance—neither network overpowering the other—and can be considered a healthy stage of GAN training where both networks are learning effectively.

Another fluctuation was observed between epochs 3750 and 4300, where the generator loss spiked again, gradually rising before returning to the range of 0.5–0.6. Simultaneously, the discriminator loss dipped again to ~0.25 and slowly climbed back. These temporary divergences reflect moments of generator instability or mode collapse, which the model managed to recover from after approximately 20 epochs.

After epoch 5000, more pronounced oscillations appeared, with both G and D losses varying between 0.5 and 0.9. These fluctuations did not lead to severe degradation and stayed within a tolerable dynamic range, suggesting the training remained generally stable despite minor adversarial shifts. Such behavior is expected in GAN models trained on limited data and does not necessarily indicate instability but rather continuous adaptation and refinement by both networks.

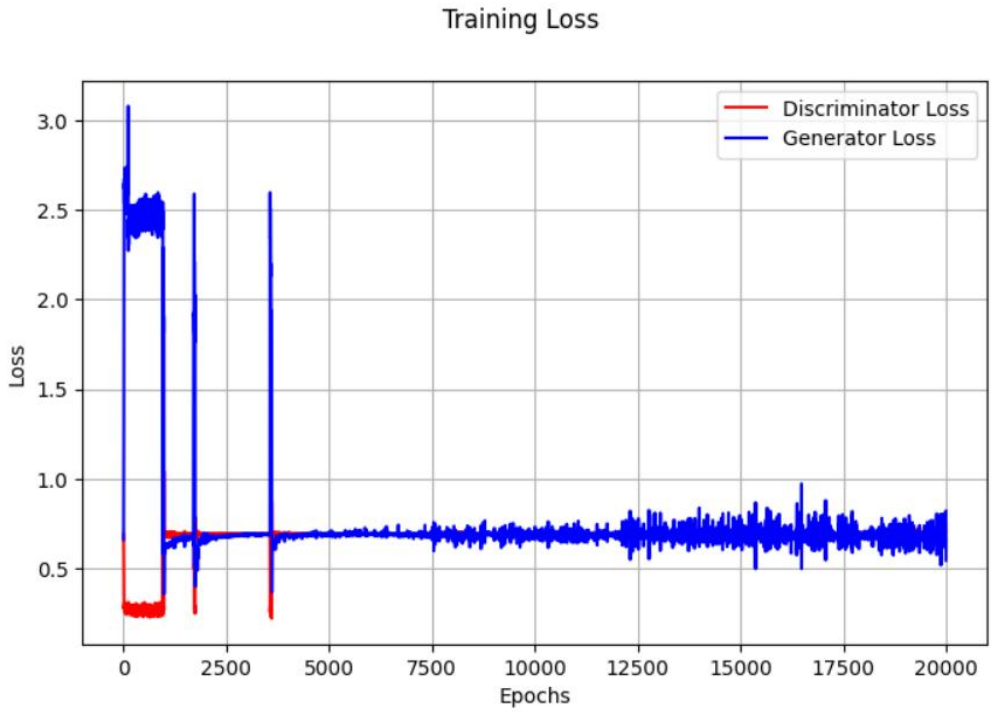


Figure 4.1. *Generator and Discriminator Loss Curves during Training.*

4.4.2. Interpretation of GAN Stability and Behavior over Epochs

Overall, the loss curves demonstrate a typical GAN behavior pattern:

- Initial imbalance, where the discriminator learns faster than the generator.
- Gradual stabilization, as the generator catches up and starts producing plausible outputs.
- Periodic oscillations, resulting from the adversarial nature of the training.
- Late-stage variance, where the networks fine-tune against each other in a narrow performance band.

This evolution indicates a generally stable training process, especially in the final experiment. The convergence zones and the network's recovery from imbalance episodes reflect robustness in the chosen training strategy and model architecture. Furthermore, the observed correlation between reduced dropout, extended training, and loss stabilization supports the hypothesis that these factors collectively improved model performance.

4.5. Visual Analysis of Generated Images

4.5.1. Comparative Image Quality Across Experiments

Visual inspection of generated outputs across all experiments reinforced the findings from the loss analysis. Table 4.2 presents sample images generated at different training epochs, illustrating the visual progression and degradation patterns observed across experiments.

- In Experiment 1 (deep architecture) and Experiment 2 (StyleGAN variant), image quality significantly degraded after 2000 epochs. The generator began producing visibly distorted or black images, reflecting overfitting and training failure. The

advanced architectures, without sufficient data or augmentation, lacked the ability to generalize and learn stable mappings from latent space to image space.

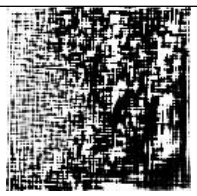
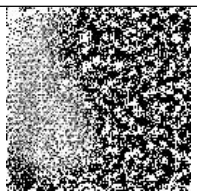
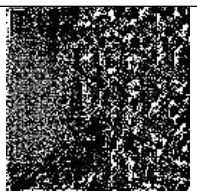
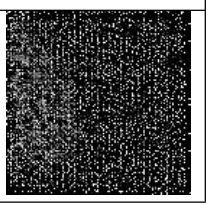
Epoch level	100	2100	3100	10 000
Sample Generated Image				

Table 4.2. *Sample Generated Images at Different Epochs.*

4.5.2. Improvements in Best Model (Experiment 3)

In Experiment 3, however, the generator consistently improved its output as training progressed. Key developments were observed across different epoch intervals:

- **Epochs 2000 – 10,000:**
  The generator began forming general anatomical structures. Shapes became clearer, grayscale values were more meaningful, and background artifacts were reduced. Figure 4.2 shows sample outputs from this phase, illustrating the model's early ability to capture organ contours and basic pixel intensity distributions. This stage marks the model's foundational learning, where synthetic images started to resemble real mammographic patterns.

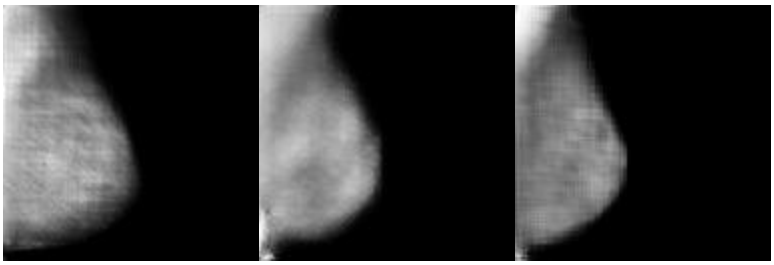


Figure 4.2. *Generated Images from Epochs 2000–10,000*

- **Epochs 10,000 – 14,000:**
  The model displayed slight improvements in texture and internal structure, especially around areas resembling pathological patterns (e.g., calcifications or masses). While still blurry, early signs of pathology representation began to emerge, and image outputs were less noisy. Figure 4.3 presents examples of generated images from this phase, illustrating the gradual emergence of meaningful medical features.

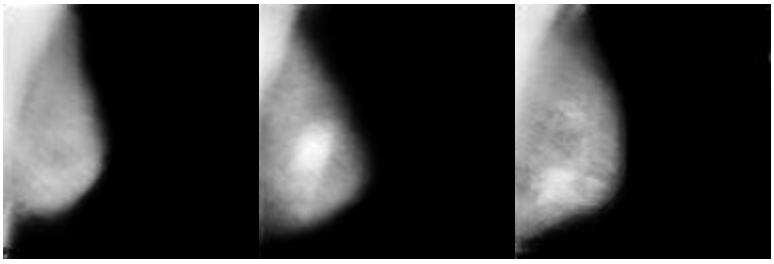


Figure 4.3. *Generated Images from Epochs 10,000–14,000*

- **Epochs 14,000 – 17,000:**
  A notable improvement in structural definition was observed. The model captured the general shape of the breast tissue more clearly and began reproducing pathological textures with greater precision. This interval showed significant progress in translating latent features into domain-relevant imagery. Figure 4.4 illustrates generated samples from epochs 14,000–17,000, highlighting these improvements in anatomical and pathological realism.

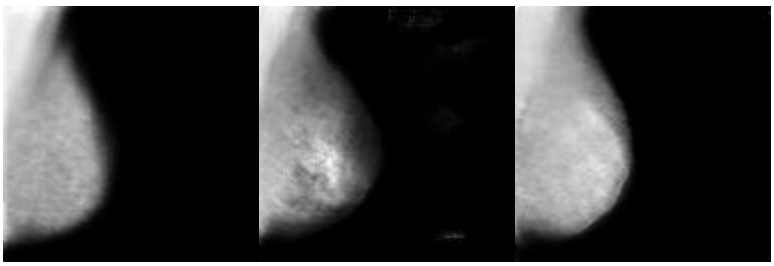


Figure 4.4. *Generated Images from Epochs 14,000–17,000.*

- **Epochs 17,000 – 20,000:**
  Generated images were visually similar to real mammograms, capturing both the macroscopic structure and fine-grained details. Pathologies, including masses and calcifications, were presented with higher clarity. Figure 4.5 displays representative images generated during this phase, showcasing the enhanced realism and pathological detail achieved at the final training stage. These improvements were supported by quantitative metrics computed on three high-quality images generated in this interval:

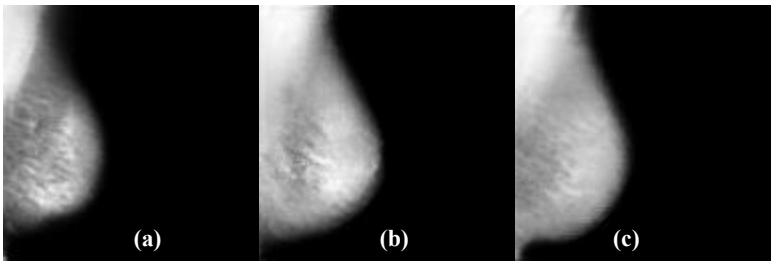


Figure 4.5. *Generated Images from Epochs 17,000–20,000*

Image	PSNR	SSIM	FID Score
Image (a)	17.40	0.784	0.1458
Image (b)	14.37	0.773	0.1035
Image (c)	15.42	0.797	0.1169

Table 4.3. *Quantitative Evaluation Metrics for Selected Images Generated Between Epochs 17,000–20,000.*

These values confirm the visual progress and suggest that the model achieved strong alignment with the distribution of real mammograms. The PSNR values, all above 14 dB, indicate a moderate to good signal fidelity with reduced noise and compression artifacts. The SSIM scores, ranging from 0.773 to 0.797, reflect consistent structural similarity between the generated and reference images, particularly in local texture and edge representation. Importantly, the FID scores, all below 0.15, demonstrate that the synthetic images lie very close to the real image distribution in feature space, indicating high realism. Together, these metrics validate the visual improvements observed in this final training interval.

4.5.3. Differences Between Initial and Later Epochs

When comparing initial epochs (e.g., <2000) with later ones (>15,000), the difference is striking. Early outputs were noisy and lacked structure, while later images demonstrated semantic understanding, sharper anatomical details, and reduced artifact presence. This confirms that the extended training duration made possible by gradual epoch increases was crucial for achieving satisfactory results.

4.5.4. Effects of Training Strategy on Image Quality

The training strategy played a pivotal role in these outcomes. Specifically:

- Reducing dropout in the discriminator improved its feature retention and feedback quality.
- Maintaining a simple generator architecture avoided overfitting.
- Extending training up to 20,000 epochs allowed the generator to evolve from basic shape learning to producing medically relevant textures.
- Progressive monitoring and small epoch increase helped control training instability.

4.6. Classification Results

A total of 100 synthetic malignant mammography images were selected from those generated by the Custom GAN model. These images, produced between epochs 16,000 and 20,000, were included exclusively in the training set of the second experiment to avoid data leakage during evaluation.

Experiment 1: Real-only training

Accuracy: 99.5%

Confusion Matrix presented in the Table below:

	Predicted Non_tumor	Predicted Tumor
Actual Non_tumor	99	1
Actual Tumor	0	100

Table 4.4. *Confusion Matrix of Experiment 1.*

Interpretation: The classifier misclassified 1 non_tumor image as tumor, but all tumor images were correctly detected.

Experiment 2: With 100 Generated Tumor Images

Accuracy: 99.0%

Confusion Matrix presented in the Table below:

	Predicted Non_tumor	Predicted Tumor
Actual Non_tumor	98	2

Actual Tumor	0	100

Table 4.5. Confusion Matrix of Experiment 2.

Interpretation: The classifier showed slightly reduced performance, misclassifying 2 non_tumor images, but again, all tumors were correctly classified.

4.6.1. Analysis and Impact

In this experiment, 100 real tumor images were replaced with 100 GAN-generated images in the training set, alongside 300 remaining real tumor samples. This substitution represents a significant change in the training data composition.

Despite replacing a quarter of the real tumor images with synthetic ones, the classification accuracy only slightly decreased from 99.5% to 99.0%. This minimal drop highlights the strong quality and relevance of the generated images, which were able to effectively supplement the training data without substantially impacting performance.

The slight increase in false positives remains very small and does not compromise the overall reliability of the classifier. These results demonstrate that GAN-generated images can serve as a valuable augmentation technique, especially when collecting large quantities of annotated medical images is difficult.

This outcome encourages further exploration and refinement of synthetic data generation to enhance training datasets while maintaining high diagnostic accuracy.

4.7. Discussion and Interpretation

4.7.1. Relationship Between Strategy, Hardware, and Output

This study highlights the strong interdependence between training strategy, computational resources, and quality of output. Training was conducted on Kaggle's limited GPU environment, which imposed constraints on session duration, memory availability, and execution speed. Despite these limitations, careful planning allowed the model to reach 20,000 training epochs, using strategic choices such as:

- Reducing dropout for better feature retention,
- Avoiding unnecessarily deep architectures to reduce overfitting,
- Incrementally raising the number of epochs to observe performance evolution.

The results clearly demonstrate that **simplicity and perseverance**, even on constrained hardware, can outperform more complex solutions when data and resources are limited.

4.8. Contributions and Future Prospects

4.8.1. Value of Current Work as a Foundation

The current project provides a strong foundation for future research in medical image generation. Despite working with a limited dataset and a basic computational setup, the model successfully produced high-quality, pathology-relevant mammograms, as confirmed by both visual inspection and competitive evaluation metrics (PSNR, SSIM, FID).

Importantly, the classification experiment demonstrated that replacing a portion of real tumor images with GAN-generated ones resulted in only a minimal drop in performance (from 99.5% to 99.0% accuracy). This finding underscores the practical usability of synthetic images for real-world tasks and confirms their value in model training.

These outcomes reinforce the potential of GANs in medical imaging, particularly for:

- Data augmentation in low-resource datasets,
- Synthetic data generation for anonymized training,
- Educational use for pathology visualization and simulation.

4.8.2. Suggestions for Improving Performance with More Resources

Access to a more powerful GPU, such as those provided by Colab Pro+ or local A100 systems, would significantly enhance training efficiency and model performance. These high-end GPUs can support larger batch sizes and enable faster processing of higher-resolution images, making it feasible to utilize the full dataset without reducing its size for runtime limitations. Faster training not only allows for more epochs within the same timeframe but also opens the door to experimenting with deeper network architectures, longer progressive training schedules, and more complex regularization techniques. In addition to computational improvements, integrating transfer learning particularly by initializing the discriminator with pretrained weights could help stabilize early-stage training and accelerate convergence. Incorporating mixed precision training may also reduce memory usage and improve throughput, making it easier to scale models and experiment with advanced architectures without hitting resource bottlenecks.

4.9. Conclusion

This chapter provided a detailed analysis of the experimental outcomes, combining both visual evaluation and quantitative metrics to assess the performance of the custom GAN model. The results showed that even with a relatively simple architecture and modest computational resources, the model was capable of generating medically relevant mammography images with high visual fidelity and stability. Notably, the inclusion of synthetic images in the classification task demonstrated the practical value of the generated data. Replacing 100 real malignant images with GAN-generated ones amounting to a third of the real tumor training set resulted in only a minimal drop in classification accuracy (from 99.5% to 99.0%), while maintaining perfect tumor detection. This reinforces the potential of GAN-generated images as a viable tool for data augmentation and training support. Overall, these findings validate the chosen methodology and offer a solid foundation for future research in medical image synthesis, especially in environments with limited resources.

General Conclusion

The thesis explored the use of Generative Adversarial Networks (GANs) for synthesizing mammography images, with the aim of addressing data scarcity and contributing to advancements in medical imaging. The study examined various architectural and training strategies to enhance the realism, stability, and quality of generated images within this complex and sensitive domain.

The work began by establishing the necessary medical and technical foundations, followed by the development and evaluation of several GAN configurations. These experiments included modifications to the generator's depth and regularization strategies, such as dropout tuning. While early attempts faced common challenges like training instability and mode collapse, subsequent refinements, particularly reducing dropout and extending the training duration, led to notable improvements in both quality and consistency.

The evaluation process was based on objective performance metrics, including FID, SSIM, and PSNR, complemented by visual inspection.

In addition, the classification experiment demonstrated the practical utility of the generated images. Replacing a portion of real malignant images with synthetic ones did not lead to significant performance degradation, showing that GAN-generated images can be safely used to improve medical image classification.

These results directly address the critical challenge of limited high-quality breast cancer imaging data highlighted at the outset of this thesis. By generating realistic synthetic mammograms, this work supports efforts to enhance early and accurate diagnosis, which is crucial for reducing breast cancer mortality.

In conclusion, this work shows that even with a relatively simple GAN architecture, careful tuning and a strong understanding of training behavior can lead to the generation of high-quality, medically relevant synthetic images. These findings highlight the importance of optimization over model complexity and support the continued use of GANs for data augmentation, diagnostic support, and research in resource-constrained medical environments.

References

ACR Committee on Appropriateness Criteria. (2021). Supplemental Screening for Breast Cancer Based on Breast Density. Journal of the American College of Radiology, 18(11S), S470-S482. DOI: 10.1016/j.jacr.2021.09.002

Allemani, C., Matsuda, T., Di Carlo, V., Harewood, R., Matz, M., Nikšić, M., Bonaventure, A., Valkov, M., Johnson, C. J., Estève, J., Ogunbiyi, O. J., Azevedo e Silva, G., Chen, W. Q., Eser, S., Engholm, G., Stiller, C. A., Monnereau, A., Woods, R. R., Visser, O., ... Coleman, M. P. (2018). Global surveillance of cancer survival 2000–14: Analyses of individual data for 37,513,025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries (CONCORD-3). The Lancet, 391(10125), 1023–1075. DOI: 10.1016/S0140-6736(17)33326-3

American Cancer Society. (2023). Breast cancer facts & figures 2023–2024. https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/2023-2024-breast-cancer-facts-and-figures.pdf

American College of Radiology (ACR). (2023). BI-RADS Atlas: Mammography (6th ed.). https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Bi-Rads

Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein GAN. arXiv preprint. DOI: 10.5555/3305381.3305404

American College of Radiology. (2013). ACR BI-RADS® atlas: Breast Imaging Reporting and Data System (5th ed.). https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Bi-Rads

Bowles, C., Chen, L., Guerrero, R., Bentley, P., Gunn, R., Hammers, A., & Alexander, D. C. (2018). GAN augmentation: Augmenting training data using generative adversarial networks. DOI: 10.48550/arXiv.1810.10863

Cancer Genome Atlas Network. (2012). Comprehensive molecular portraits of human breast tumours. Nature, 490(7418), 61–70. DOI: 10.1038/nature11412

Dar, S. U., Yurt, M., Karacan, L., Erdem, A., Erdem, I., & Çukur, T. (2019). Image synthesis in multi-contrast MRI with conditional generative adversarial networks. IEEE Transactions on Medical Imaging, 38(10), 2375–2388. DOI: 10.1109/TMI.2019.2901750

D'Orsi, C. J., Sickles, E. A., Mendelson, E. B., Morris, E. A., Bassett, L. W., Dershaw, D. D., & Rubin, E. (Eds.). (2013). ACR BI-RADS® atlas: Breast Imaging Reporting and Data System (5th ed.) [Data set]. American College of Radiology. https://www.acr.org/Clinical-Resources/Reporting-and-Data-Systems/Bi-Rads

Early Breast Cancer Trialists' Collaborative Group. (2005). Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: An overview of the randomised trials. The Lancet, 365(9472), 1687–1717. DOI: 10.1016/S0140-6736(05)66544-0

Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J., & Greenspan, H. (2018). Synthetic data augmentation using GAN for improved liver lesion classification. IEEE Transactions on Medical Imaging, 38(3), 675–685. https://doi.org/10.48550/arXiv.1801.02385

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. Advances in Neural Information Processing Systems, 27, 2672–2680. https://doi.org/10.48550/arXiv.1406.2661

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). GANs trained by a two time-scale update rule converge to a local Nash equilibrium. In Advances in Neural Information

Processing Systems (NeurIPS), 30. https://doi.org/10.48550/arXiv.1706.08500

International Agency for Research on Cancer. (2022). Algeria cancer profile. World Health Organization. https://gco.iarc.fr/

International Agency for Research on Cancer. (2022). Global Cancer Observatory. World Health Organization. https://gco.iarc.fr/

International Agency for Research on Cancer. (2024). Algeria fact sheet – GLOBOCAN 2022 (Version 1.1) [Data set]. World Health Organization. https://gco.iarc.who.int/media/globocan/factsheets/populations/12-algeria-fact-sheet.pdf

International Agency for Research on Cancer. (2024). Breast cancer fact sheet – GLOBOCAN 2022 (Version 1.1) [Data set]. World Health Organization. https://gco.iarc.who.int/media/globocan/factsheets/cancers/20-breast-fact-sheet.pdf

International Health Economics. (2024). Country card Algeria: Improving breast cancer care in the MEA region. https://www.iapo.org.uk/sites/default/files/files/BCR/IHE-REPORT-2024_6-country-cards-Algeria_.pdf

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. International Conference on Machine Learning, 448–456. DOI: 10.48550/arXiv.1502.03167

Jiménez-Gaona, Y., Carrion-Figueroa, D., Lakshminarayanan, V., & Rodríguez-Álvarez, M. J. (2024). GAN-based data augmentation to improve breast ultrasound and mammography mass classification. Biomedical Signal Processing and Control, 87(Part A), 105456. https://digital.csic.es/bitstream/10261/389527/1/gan-based_data_Jimenez.pdf

Karras, T., Aila, T., Laine, S., & Lehtinen, J. (2018). Progressive growing of GANs for improved quality, stability, and variation. International Conference on Learning Representations (ICLR). https://doi.org/10.48550/arXiv.1710.10196

Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 4401–4410. https://doi.org/10.1109/CVPR.2019.00453

Kazeminia, S., Baur, C., Kuijper, A., et al. (2020). GANs for medical image analysis. Artificial Intelligence in Medicine, 109, 101938. https://doi.org/10.1016/j.artmed.2020.101938

Kuchenbaecker, K. B., et al. (2017). Risks of breast, ovarian, and contralateral breast cancer for BRCA1 and BRCA2 mutation carriers. JAMA, 317(23), 2402–2416. https://doi.org/10.1001/jama.2017.7112

Korkinof, D., Rijken, T., O'Neill, M., Yearsley, J., Harvey, H., & Glocker, B. (2018). High-resolution mammogram synthesis using progressive generative adversarial networks. arXiv. https://doi.org/10.48550/arXiv.1807.03401

Karras, T., Laine, S., & Aila, T. (2019). A Style-Based Generator Architecture for Generative Adversarial Networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 4401–4410. https://doi.org/10.1109/CVPR.2019.00453

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 4681–4690. https://doi.org/10.1109/CVPR.2017.19

Lee, J., & Nishikawa, R. M. (2021). Identifying women with mammographically-occult breast cancer

leveraging GAN-simulated mammograms. arXiv preprint arXiv:2109.12113. https://doi.org/10.48550/arXiv.2109.12113

McKinney, S. M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G. S., Darzi, A., Etemadi, M., Garcia-Vicente, F., Gilbert, F. J., Halling-Brown, M., Hassabis, D., Jansen, S., Karthikesalingam, A., Kelly, C. J., King, D., ... Shetty, S. (2020). International evaluation of an AI system for breast cancer screening. Nature, 577(7788), 89–94. https://doi.org/10.1038/s41586-019-1799-6

Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784. https://arxiv.org/abs/1411.1784

Metz, L., Poole, B., Pfau, D., & Sohl-Dickstein, J. (2017). Unrolled generative adversarial networks. arXiv preprint. https://doi.org/10.48550/arXiv.1611.02163

Miyato, T., Kataoka, T., Koyama, M., & Yoshida, Y. (2018). Spectral normalization for generative adversarial networks. International Conference on Learning Representations. https://doi.org/10.48550/arXiv.1802.05957

Modanwal, G., Vellal, A., & Mazurowski, M. A. (2021). Normalization of breast MRIs using cycle-consistent generative adversarial networks. Computerized Medical Imaging and Graphics, 90, 101889. DOI: 10.1016/j.cmpb.2021.106225

Radford, A., Metz, L., & Chintala, S. (2016). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434. https://doi.org/10.48550/arXiv.1511.06434

Radford, A., Metz, L., & Chintala, S. (2016). Unsupervised representation learning with deep convolutional generative adversarial networks. International Conference on Learning Representations (ICLR). https://arxiv.org/abs/1511.06434

Stavros, A. T., et al. (1995). Solid breast nodules: Use of sonography to distinguish between benign and malignant lesions. Radiology, 196(1), 123–134. https://doi.org/10.1148/radiology.196.1.7784555

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 71(3), 209–249. https://doi.org/10.3322/caac.21660

Skandarani, Y., Jodoin, P.-M., & Lalande, A. (2023). GANs for medical image synthesis: An empirical study. Journal of Imaging, 9(3), 69. https://doi.org/10.3390/jimaging9030069

Shin, H.-C., Tenenholtz, N. A., Rogers, J. K., Schwarz, C. G., Senjem, M. L., Gunter, J. L., Andriole, K. P., & Michalski, M. H. (2018). Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In Medical Image Computing and Computer Assisted Intervention – MICCAI 2018 (pp. 1–11). Lecture Notes in Computer Science, 11045. Springer. https://doi.org/10.48550/arXiv.1807.10225

Singh, V. K., Rashwan, H. A., Romani, S., Akram, F., Pandey, N., Sarker, M. M. K., Saleh, A., Arenas, M., Arquez, M., Puig, D., & Torrents-Barrena, J. (2018). Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network. arXiv preprint arXiv:1809.01687. https://doi.org/10.48550/arXiv.1809.01687

Sprague, B. L., Gangnon, R. E., Burt, V., Trentham-Dietz, A., Hampton, J. M., Wellman, R. D., Kerlikowske, K., & Miglioretti, D. L. (2018). Prevalence of mammographically dense breasts in the United States. Journal of the National Cancer Institute, 110(10), 1050–1057. DOI: 10.1093/jnci/dju255

Tabár, L., & Dean, P. B. (1985). Teaching atlas of mammography (2nd ed.). Thieme Medical Publishers , [Archive.org version]. https://archive.org/details/teachingatlasofm0000taba

Tulyakov, S., Liu, M.-Y., Yang, X., & Kautz, J. (2018). MoCoGAN: Decomposing motion and content for video generation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1526–1535. https://doi.org/10.1109/CVPR.2018.00165

Vondrick, C., Pirsiavash, H., & Torralba, A. (2016). Generating videos with scene dynamics. Advances in Neural Information Processing Systems, 29. https://doi.org/10.48550/arXiv.1609.02612

World Cancer Research Fund. (2023). Breast cancer statistics. https://www.wcrf.org/dietandcancer/cancer-trends/breast-cancer-statistics

World Health Organization. (2021). Breast cancer. https://www.who.int/news-room/fact-sheets/detail/breast-cancer

World Health Organization. (2023). Breast cancer statistics. https://www.who.int/news-room/fact-sheets/detail/breast-cancer

World Health Organization. (2023). Global breast cancer initiative. https://www.who.int/breast-cancer

World Health Organization. (2023). Global Breast Cancer Initiative Implementation Framework. https://www.who.int/publications/i/item/9789240067134

World Health Organization. (2024, March 13). Breast cancer. https://www.who.int/news-room/fact-sheets/detail/breast-cancer

White, H. T. W. (2023). Final-Final-RSNA-Breast-Cancer-Dataset [Dataset]. Kaggle. https://www.kaggle.com/datasets/harrisontwwhite/final-final-rsna-breast-cancer-dataset

Xu, T., Zhang, P., Huang, Q., Zhang, H., Gan, Z., Huang, X., & He, X. (2018). AttnGAN: Fine-grained text to image generation with attentional generative adversarial networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1316–1324. https://doi.org/10.1109/CVPR.2018.00143

Yi, X., Walia, E., & Babyn, P. (2019). Generative adversarial network in medical imaging: A review. Medical Image Analysis, 58, 101552. https://doi.org/10.1016/j.media.2019.101552

Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2223–2232. https://doi.org/10.1109/ICCV.2017.244

Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., & Metaxas, D. (2017). StackGAN: Text to photo-realistic image synthesis with stacked generative adversarial networks. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 5907–5915. https://doi.org/10.1109/ICCV.2017.629