UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Programa de Doctorado en Tecnologías de la Salud y Bienestar

Breast medical images classification through the application of deep learning processing technologies

Author: Yuliana Jiménez Gaona

Supervisors:

PhD. María José Rodríguez Álvarez
Instituto de Instrumentación para la Imagen Molecular I3M
Universitat Politècnica de València, Valencia, Spain

PhD. Vasudevan Lakshminarayanan
Theoretical and Experimental Epistemology Lab, School of Optometry and Vision Science
University of Waterloo, Waterloo, Canada

July, 2024

The Doctoral Thesis is presented as a **compendium** of the following publications:

- Deep-Learning-Based Computer-Aided Systems for Breast Cancer Imaging: A Critical Review.
- Breast Mass Regions Classification from Mammograms using Convolutional Neural Networks and Transfer Learning.
- Ultrasound Breast images denoising using Generative Adversarial Networks (GANs).
- GAN-based data augmentation to improve breast Ultrasound and Mammography Mass Classification
- BraNet: A mobil Application for Breast image classification based on Deep Learning algorithms.

The study was carried out between January 2020 and June 2024 at respected institutions such as the Institute of Instrumentation for Molecular Imaging (I3M) of the Universitat Politècnica de València (Valencia, Spain), the Theoretical and Experimental Epistemology Lab, School of Optometry and Vision Science University of Waterloo (Waterloo, Canada), and the Private Technical University of Loja (Loja, Ecuador). It is worth noting that this research project was co-funded by the Spanish Government Grant PID2019-107790RB-C22, which aimed to develop software for a continuous PET crystal system to be applied in breast cancer treatment.

Acknowledgements

I would like to express my special appreciation and thanks to my supervisor, PhD.María José Rodríguez and PhD.Vasudevan Lakshminarayanan, thank you for supporting me and advising me throughout these years. Their encouragement and teachings have shaped my research from different scientific perspectives, making me the researcher that I am today. Without you, the competition of this doctoral program would not have been possible.

I would also like to thank the Institute for Molecular Imaging Instrumentation (I3M) for their support during my stay in Valencia as a student and as a researcher. I would also like to thank the Vice-Rectorate for Research at UTPL for its contributions and support with a grant, especially for the PhD. Silvia Gonzalez for her support. Special thanks to my friend and college Darwin, who motivated me to study this doctorate, and to all of you who shared this time with me over the years.

I would like to thank my husband Oscar, my mother Fanny, my father Mario, to all my brothers Mario, Dario, Marco and friends for their unconditional support throughout of these years. Finally, this work is dedicated with special love to my three children, Emiliano, Elías and Sarita, who have been the main driving force behind my studies.

CONTENTS

RESUMEN

El cáncer de mama es una de las principales causas de muerte en mujeres de todo el mundo. Supone el 18.2% de las muertes por cáncer en la mujer y la primera causa de muerte en mujeres entre 40 y 55 años según la Sociedad Española de Senología y Patología Mamaria (SESPM). Una forma eficiente de disminuir este porcentaje es diagnosticarlo de forma temprana mediante exámenes de rayos x (Mamografía, Tomografía por emisión de positrones, Imagen de resonancia magnética, Tomografía computarizada), Ultrasonido, Tomosíntesis, Histopatología y Termografía. En la actualidad dentro del campo de la radiómica estos datos clínicos están siendo procesados con el uso de algoritmos de inteligencia artificial, especialmente para el preprocesamiento, segmentación y clasificación de lesiones malignas o benignas presentes en las imágenes médicas. Además, el desarrollo de estos sistemas computacionales asistidos para diagnóstico y detección temprana de anomalías presentes en la mama, ayudan al médico con una segunda opinión al diagnóstico manual tradicional. En consecuencia, el objetivo de este estudio es construir modelos de aprendizaje profundo y automático para la detección, segmentación y clasificación de lesiones mamarias en imágenes de mamografía y ultrasonido. Los hallazgos de este estudio brindan diversas herramientas de aumento de datos, super resolución, segmentación y clasificación automática de imágenes de mama para mejorar la precisión en los algoritmos de clasificación de lesiones mamarias.

ABSTRACT

Breast cancer is one of the most common causes of death in women worldwide. It accounts for 18.2% of cancer deaths in women and is the leading cause of death in women between 40 and 55 years of age, according to the Spanish Society of Senology and Breast Pathology (SESPM). An effective way to reduce this rate is through early diagnosis using radiological imaging (mammography, positron emission tomography, magnetic resonance imaging, computed tomography), Ultrasound,

Tomosynthesis, Histopathology and Thermography. Currently, the field of radiomics is processing these clinical data using artificial intelligence algorithms, for pre-processing, segmentation, and classification of malignant or benign lesions present in medical images. In addition, the development of these computer-aided systems for diagnosis and early detection of breast abnormalities helps the radiologists with a second opinion to the traditional manual diagnosis. Therefore, the aim of this study is to build deep and machine learning models for the detection, segmentation, and classification of breast lesions in mammography and ultrasound images. The results of this study provide several tools for data augmentation, super-resolution, segmentation, and automatic classification of breast images to improve the accuracy of breast lesion classification algorithms.

RESUMEN VALENCIANO

El càncer de mama és una de les principals causes de mort en dones de tot el món. La mortalitat relacionada amb esta mena de càncer és més alta en comparación amb altres tipus de càncer. Una forma eficient de disminuir este percentatge és diagnosticar-lo de manera primerenca mitjançant exàmens de raigs x (Mamografia, Tomografía per emissió de positrons, Imatge de ressonància magnètica, Tomografia computada), Ultrasò, Tomosíntesi, Histopatologia i Termografia. En la actualidad dins del camp de la radiómica estes dades clíniques estan sent processados amb l'ús d'algorismes d'intel·ligència artificial, especialment per al preprocesamiento, segmentació i classificació de lesions malignes o benignes presents en les imatges mèdiques. A més, el desenvolupament d'estos sistemes computacionals asistidos per a diagnòstic i detecció precoç d'anomalies presents en la mama, ajuden al metge amb una segona opinió al diagnòstic manual tradicional. En conseqüència, l'objectiu d'este estudi és construir models d'aprenentatge profundo i automàtic per a la detecció, segmentació i classificació de lesions mamàries en imatges de mamografia i ultrasò. Les troballes d'este estudi brinden vaig donar-verses ferramentes d'augment de dades, super resolució, segmentació i classificación automàtica d'imatges de mama per a millorar la precisió en els algorismes de classificació de lesions mamàries.

Director's Authorization

Dra. María José Rodríguez Álvarez,
Profesora titular en el Departamento de Matemática Aplicada
Universidad Politécnica de Valencia, DNI 76125754D

HAGO CONSTAR:

Que como director de la tesis doctoral de Yuliana del Cisne Jiménez Gaona con DNI 1103898019, autorizo a presentar la tesis doctoral "*Breast medical images classification through the application of deep learning processing technologies*", mediante la modalidad de compendio de artículos al disponer de los siguientes artículos publicados:

- Deep-Learning-Based Computer-Aided Systems for Breast Cancer Imaging: A Critical Review.
- Breast Mass Regions Classification from Mammograms using Convolutional Neural Networks and Transfer Learning.
- Ultrasound Breast images denoising using Generative Adversarial Networks (GANs).
- GAN-based data augmentation to improve breast Ultra-sound and Mammography Mass Classification.
- BraNet: A mobil Application for Breast image classification.

Por todo ello fimo esta carta de autorización

Fdo.: María José Rodríguez Álvarez
Valencia, Aphril 2024
Departamento de Matemática Aplicada
Universitat Politècnica de València
Camino de Vera, s/n 46022
Valencia-España

Codirector's Authorization

Dr. Vasudevan Lakshminarayanan

As co-director of the doctoral thesis of Yuliana del Cisne Jiménez Gaona with DNI 1103898019, I authorize to present the doctoral thesis "Breast medical images classification through the application of deep learning processing technologies" through the modality of compendium of articles by having the following articles published: Deep-Learning-Based Computer-Aided Systems for Breast Cancer Imaging: A Critical Review.

- Breast Mass Regions Classification from Mammograms using Convolutional Neural Networks and Transfer Learning.
- Ultrasound Breast images denoising using Generative Adversarial Networks (GANs).
- GAN-based data augmentation to improve breast Ultra-sound and Mammography Mass Classification.
- BraNet: A mobil Application for Breast image classification.

At this moment, sign this letter of authorization.

Fdo.: Vasudevan Lakshminarayanan

V Lakshminarayanan

Canada, Aphril 2024
Theoretical and Experimental Epistemology Lab,
School of Optometry and Vision Science
University of Waterloo,
Waterloo ON N2L3G1, Canada

Thesis Outline

This thesis report is structured as a compendium of works previously published or accepted for publication. The contents of each section in this PhD Thesis report are structured as follows:

The first section is devoted to the Thesis Director's Authorization for the doctoral student to present the work, which includes all the necessary documentation to present the thesis in this format (Section 1).

The Introduction chapter (Chapter 1) offers a background of the main concepts appearing in this dissertation. Also, the research's Motivation, Objectives, and Related Work are presented in these Sections (1.1, 1.2, and 1.3).

Material chapter (Chapters 2, 3, 4, 5, and 6) includes the full references of the articles that make up the body of the thesis, which are presented as follows: (Chapter 2) Deep-Learning-Based Computer-Aided Systems for Breast Cancer Imaging: A Critical Review**,** (Chapter 3) Breast Mass Regions Classification from Mammograms using Convolutional Neural Networks and Transfer Learning, and (Chapter 4) Ultrasound Breast images denoising using Generative Adversarial Networks (GANs), (Chapter 5) GAN-based data augmentation to improve breast Ultrasound and Mammography Mass Classification and (Chapter 6) BraNet: A mobile Application for Breast image classification based on Deep Learning algorithms.

Overall Discussion is presented in Chapter 7; in this section, we discuss the main results of this research and further work that could be performed. Thus, Chapter 8 summarizes all the conclusions obtained from this research.

The Thesis Presentation Application, Compendium of published articles, and Acceptance of the co-authors for the doctoral student details are presented in Appendix I.

Chapter 1
Introduction

1.1 Motivation

Breast cancer occurs in every country of the world mainly in women, and her lives are lost because of this type of cancer more than any other type. In 2020, there were 2.3 million women diagnosed with breast cancer and 685 000 deaths globally. As of the end of 2020, there were 7.8 million women alive who were diagnosed with breast cancer in the past 5 years, making it the world's most prevalent cancer.

Certain factors increase the risk of breast cancer including increasing age, obesity, harmful use of alcohol, family history of breast cancer, gene mutations in the genes BRCA1, BRCA2 and PALB-2, history of radiation exposure, reproductive history, tobacco use and postmenopausal hormone therapy.

However, breast cancer treatment can be highly effective, especially when the disease is identified early state, to reduce the mortality rate be. The image techniques that have been most used for breast screening and cancer detection are Digital Mammography and Ultrasound, since provide efficient control and while exposing the patient to minimal radiation. Mammography and Ultrasound are indeed powerful tools in the field of breast cancer diagnosis, screening and medical imaging. They serve as complementary image modality in detecting and characterizing breast abnormalities, and their combined use enhances the accuracy of breast cancer diagnosis:

Digital Mammography: This technique is widely used for breast cancer screening in asymptomatic women and for diagnostic purposes in women with breast symptoms or abnormalities. Digital mammography is a radiographic imaging technique that uses low-dose x-rays to create detailed images of the breast tissue. It is particularly effective in detecting small tumors or abnormalities that may not be palpable during a physical examination.

Ultrasound: It is often used as a complementary imaging modality alongside mammography, and it is particularly helpful for evaluating breast abnormalities identified on mammograms or for further characterizing breast lumps or cysts. It can differentiate between fluid-filled cysts and solid masses and provide additional information about the nature of a breast lesion. Ultrasound uses high-frequency sound waves to produce real-time images of the breast tissue.

The combination of mammography and ultrasound allows healthcare providers to gather comprehensive information about breast abnormalities. In some cases, other imaging techniques like 3D mammography (tomosynthesis) is an advanced form of mammography that captures multiple images from different angles, Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI); Thermography Images may also be used for further evaluation when needed.

The focus on anatomic details in medical imaging, particularly in breast imaging, is critical for early cancer detection and accurate diagnosis. These imaging techniques help identify the size, location, and characteristics of breast lesions, aiding in the formulation of appropriate treatment plans. Additionally, ongoing advancements in technology and image interpretation techniques continue to improve the accuracy and reliability of breast cancer diagnosis through mammography and ultrasound.

These breast imaging studies can be interpreted with a section of the latest version of the Breast Imaging Reporting and Data System (BIRADS) , but traditional manual diagnosis requires an intense workload on the part of expert pathologists,. Thus, with the advent of the artificial intelligence and computer-aided diagnosis systems based on deep learning model, provides fast and powerful image analysis assistance to pathologists in their diagnostic tasks, looking for a second opinion to early detection and reduce global breast cancer mortality.

Consequently, this thesis focuses on the evaluation of different experimental deep learning-based algorithms, mainly used for 1. Breast image Pre-processing: (i) Denoising, (ii) Super-resolution, and (iii) Data augmentation. 2. Segmentation and 3. Classification. Hence, the central hypothesis of this work is that automatic segmentation and classification algorithms based on deep learning help with fast and effective early diagnosis and breast anomaly classification, which are evidenced in the articles submitted.

1.2 Objectives

The main objective of this thesis is to implement deep learning-based algorithms for mammography and ultrasound 2D breast image processing, as an alternative to traditional machine learning algorithms used in breast image classification.

Three specific objectives were defined as follows:

1) Conduct a literature review about deep learning-CAD architectures used for breast tumor diagnosis/detection in breast imaging classification compared with the traditional CAD system. (article 1 Deep-Learning-Based Computer-Aided Systems for Breast Cancer Imaging: A Critical Review)

2) Enhance breast medical images by implementing deep learning-based algorithms for breast imaging super-resolution and denoising through convolution and generative pretraining models. (article 2 Breast Mass Regions Classification from Mammograms using Convolutional Neural Networks and Transfer Learning and article 3 Ultrasound Breast images denoising using Generative Adversarial Networks (GANs))

3) Implement deep learning-based algorithms for breast region of interest (ROIs) for data augmentation, segmentation, and automatic classification by distinguishing benign from malignant regions. (article 4 GAN-based data augmentation to improve breast Ultra-sound and Mammography Mass Classification)

4) Implement a on-line phase using a mobile graphical interface with the best deep learning models trained in off-line phases, such as breast region segmentation and classification. (article 5 BraNet: A mobil Application for Breast image classification)

5) Evaluate the performance and accuracy of deep learning algorithms using several statistical metrics.

1.3 Related Work

Relate work is structured in four parts, which constitute a framework for the work developed in this Thesis: breast cancer, screening, deep learning algorithms, and a short revision of the state of the art, which constitutes an introduction to the publications (included in the chapters 3-6) that represent the body of the Thesis.

1.3.1 Breast Cancer

Nowadays, cancer is one of the leading causes of human deaths in worldwide. Breast cancer has elevated morbidity and mortality in women and is higher than other cancers [1-4]. The incidence rate of breast cancer has been reported to range from 19.3 per 100,000 women in East Africa to 89.7 per 100,000 women in Western Europe [5,6]. In Latin America an estimated 114,900 women are diagnosed and an estimated 37,000 women die of breast cancer every year in this region. In addition, both incidence and mortality are increasing [7]. In general, about 12% in USA and European countries suffer from this disease during their life. It has also been reported that the number of deaths will continue to grow in next years, and this number is expected to rise to 74,000 every year in 2030 [8].

Because of the human body anatomy, women are more vulnerable to breast cancer than men. The breast's anatomy involves a combination lobes, ducts, nipples and glandular tissue, fat, connective tissue, and supporting structures like Cooper's ligaments, all of which contribute to the mechanical properties and localization of breast tissue.

Modeling breast mechanics requires a comprehensive understanding of these components and their interactions. Epithelial tumors usually grow within the lobes as well as in the ducts and later form a lump generating breast cancer [9-11]. Breast cancer classification refers to the categorization of breast cancer based on various characteristics and features of the disease. Accurate classification is essential for determining the appropriate treatment plan and predicting the prognosis for a patient. There are several aspects to consider when classifying breast cancer:

(i) *Histological Classification* (Ductal Carcinoma in Situ (DCIS), Invasive Ductal Carcinoma (IDC), Invasive Lobular Carcinoma (ILC) and Mixed Histology) [12].

(ii) *Molecular Classification* (Hormone Receptor Status, HER2 (human epidermal growth factor receptor 2). Tumors are classified as HER2-positive (HER2+) or HER2-negative (HER2-).Triple-Negative Breast Cancer (TNBC): This subtype is negative for ER, PR, and HER2 and is often associated with a more aggressive course) [13].(iii) *Staging:* Breast cancer is staged based on the size of the tumor, lymph node involvement, and the presence of distant metastasis. Grade: The grade of breast cancer is determined by the appearance of cancer cells under a microscope. It helps predict how fast the cancer may grow and spread.

(iv) *Clinical Stage:* This considers the overall clinical assessment, including imaging studies, to determine the extent and severity of the disease, and there are different methods of clinical detection, non-invasive: Mammography [14], Ultrasound [15], Positron Emission Tomography (PET) [16], Magnetic Resonance Imaging (MRI) [17], Digital Brest Tomosynthesis (DBT) [18]; Thermography Images [19] and invasive (histology and biopsy) [20,21]. The biopsy image is the only microscopic tissue procedure that can definitively determine whether the suspicious area is cancerous.

1.3.2 Breast Imaging Screening

Breast screening technologies including mammography and ultrasound and are currently the most widely used imaging methods for detecting breast cancer early, since this method involves relatively low cost, wide availability and lower cost compared to some other imaging techniques, are effective at detecting early-stage breast cancer [8]. Ultrasound (also known as ultrasonography or sonography) is indeed a complementary imaging test to mammography in the evaluation of breast health, is no radiation exposure, useful for women with dense breast tissue and can provide additional information about breast lesions. While mammography remains the primary screening tool for breast cancer detection of very small lesions, even masses and microcalcifications, ultrasound plays a valuable role in certain situations that can be benign or malignant [22]. In practice, these two imaging modalities are often used together to provide a more comprehensive evaluation of breast health. However, the thermogram is more proper screening and has lower cost than other types of screening methods like the mammogram, ultrasound, and magnetic resonance imaging [23,24]. Other advanced imaging techniques, such as breast MRI, may also be used in specific cases, such as for high-risk individuals or to further evaluate suspicious findings from mammography and ultrasound. The choice of screening method depends on factors like the individual's age, breast density, personal and family history, and the presence of specific risk factors for breast cancer. These breast resonance studies can be interpreted by a traditional manual diagnosis with the BIRADS system version [25], however it requires an intense workload, a strong professional skills and rich experience on the part of expert pathologists [26], who are prone to misdiagnosis [27].

1.3.3 Computer Aides Diagnosis/Detection systems (CAD)

For this reason, artificial intelligence algorithms are being developed for medical image processing such as machine learning and deep learning-based image diagnostic models [28-31], usually called Computer-Aided Diagnosis/Detection system (CAD). CAD are solving the main implications in Biomedicine and Radiomics, looking for reducing the false positives and improving the automatic diagnosis and efficiency in the location and monitoring of tumor processes. The CAD workflow starts with (i) Public databases collection. (ii) Regions of interest (RoIs) cropping and mask extraction, (iii) Image preprocessing, (iv) Segmentation, (v) Classification and (vi) Performance evaluation metrics [9].

The first step involves gathering a large and diverse dataset of medical images. These databases often contain images of various medical conditions, including those related to tumors, and serve as the foundation for training and testing machine learning and deep learning models. During the *Database collection* some public (CBIS-DDSM, Inbreast, Mini-MIAS, UDIAT, BUSI) or private clinical datasets are being used as input for deep learning models training. Then, in medical images some *Regions of Interest (RoIs) and Mask are Cropping and* Extracted to precisely locate these regions**,** because not all parts of the image are relevant for diagnosis. CAD systems identify and extract RoIs, which are specific areas in the image that contain potential abnormalities. The images are cropping and resizing, conserving only the RoIs and their mask during segmentation process (looking for better computational performance during the training).

After that the *Image Preprocessing,* is essential to enhance the quality of the medical images and make them suitable for analysis. Common preprocessing steps include pectoral removal, noise reduction, image enhancement, image super resolution, data augmentation, contrast adjustment, normalization and resizing. *Segmentation* is a critical step where the CAD system precisely outlines and identifies the boundaries of lesions or areas of interest within the breast tissue. This process can help localize and isolate potential abnormalities like tumors. Its process is necessary to achieve better accuracy and precision during segmentation and classification steps. The *classification* problem is a fundamental cognitive task in computer vision, which is accomplished by the identification of certain anatomical or pathological features that can discriminate one anatomical structure or tissue in benignant or potentially malignant. CAD system employs machine learning or deep learning algorithms to classify the identified RoIs.

These algorithms learn from the features and patterns within the RoIs to make diagnostic predictions. Among the most popular deep learning algorithms used for solving these task are Convolutional Neural Networks (CNN) [15-18], Generative Adversarial Networks (GAN)[32-34], Autoencoders (AE)[35,36] and Recurrent Neural Networks (RNN)[37], which have achieved great success in large-scale medical image recognition [27,31].

CNNs become the backbone of many medical image analysis tasks due to their ability to automatically learn hierarchical features from images. CNNs are widely used for tasks like image classification, object detection, and image segmentation.

GANs consist of two neural networks, a generator, and a discriminator, that are trained together in a competitive manner. GANs have been applied in medical image generation, data augmentation, and image-to-image translation tasks. They can help in creating synthetic medical images that can be used to augment datasets and improve model generalization.

Autoencoders are neural networks designed for unsupervised learning. They are used for tasks like image denoising, dimensionality reduction, and anomaly detection. In medical image analysis, AEs can be used for feature extraction and representation learning. RNNs are designed to work with sequential data, and they are often used in tasks involving sequences of medical images. In some cases, RNNs are applied to analyze temporal aspects of medical data, such as tracking changes in medical images over time.

Finally, *Performance Evaluation Metrics* to assess the CAD system's accuracy and effectiveness, various metrics may include sensitivity, specificity, accuracy, and area under the receiver operating characteristic curve (AUC-ROC). They help measure how well the CAD system can detect and classify breast lesions.

To conclude, several studies [38-60] indicate that DL methods, especially Autoencoders, GAN [61-63] and CNN [64-65] networks can be a promising methodology in medical image analysis. In preprocessing, GAN are specially used in data augmentation, Autoencoders in Denoising and CNNs in segmentation and classification.

Consequently, this thesis focuses on the evaluation of different experimental deep learning-based algorithms, resulting in the following articles:

(i) Deep-learning-based computer-aided systems for breast cancer imaging: a critical review.
(ii) Breast mass regions classification from mammograms using convolutional neural networks and transfer learning.
(iii) Ultrasound Breast images denoising using Generative Adversarial Networks (GANs).
(iv) GAN-based data augmentation to improve breast Ultrasound and Mammography Mass Classification.
(v) BraNet: A mobil Application for Breast image classification based on Deep Learning algorithms.

1.4 Contribution to Knowledge

This thesis offers five novel contributions for the assessment of patients with breast cancer.

The first contribution provides a critical review of the literature on deep learning applications in breast tumor diagnosis using ultrasound and mammography images. It also summarizes recent advances in CAD systems, which make use of new deep learning methods to automatically recognize breast images and improve the accuracy of diagnoses made by radiologists. The results demonstrate that in most cases, the deep learning architectures outperformed traditional methodologies.

The second contribution introduces a novel approach to enhance the quality of digital mammography images through pre-processing techniques, improving breast cancer detection accuracy. Two convolutional networks (EDSR and RDN) were implemented as image super-resolution techniques. The PSNR and SSIM evaluation metrics indicate that the EDSR-based on Unet model outperforms the RDN-based on Resnet model in image-enhanced super-resolution, thus leading to more precise breast tissue segmentation and subsequent promising classification results implementing the Resnet model.

The third contribution provides a new deep learning model named Generative Adversarial Networks (GANs) as speckle noise reduction in breast ultrasound (US) images. Speckle noise degrades visual radiological interpretation and generally causes several difficulties in identifying malignant and benign regions in US images, this work presents two GANs models (Conditional GAN and Wasserstein GAN) for speckle denoising of breast US regions of interest (ROIs), looking for reduce speckle noise while preserving features and details in the ROIs.

The fourth contribution trained GAN models to generate synthetic breast image data and explore the possibility of improving the Resnet classification model's performance. The accuracy of the classification process was compared using real and synthetic data combined with GANs and CNN models. Various GAN models, including Wasserstein GAN with Gradient Penalty (WGAN-GP), Cycle GAN, Conditional GAN, and Spectral Normalization GAN (SNGAN), were tested for data augmentation in breast regions of interest (ROIs) using mammography and US databases. The quality and diversity of the synthetic data were assessed using feature-based, nonreference-based, and reference-based metrics. Moreover, the Resnet performance model was evaluated using the accuracy, F1 score, precision, and recall average values. Classification results showed high accuracy without data augmentation in both breast image types. Indicating not only is data augmentation needed to improve the image classification process, but also previous preprocessing and characterizing ROIs by abnormality type is crucial to generate diverse synthetic data and improve accuracy using combined GANs and CNN models.

The final contribution presents an open-source preclinical mobile application named "BraNet" for mammography (DM) and ultrasound (US) breast imaging segmentation and classification, using a client-server architecture implemented in Python for iOS and Android devices, providing radiologists with second opinions to reduce false diagnosis.

Chapter 2

DEEP-LEARNING-BASED COMPUTER-AIDED SYSTEMS FOR BREAST CANCER IMAGING: A CRITICAL REVIEW

Jiménez-Gaona, Y., Rodríguez-Álvarez, M. J., & Lakshminarayanan, V.

Applied Sciences, 2020, vol. 10, no 22, p. 8298. https://doi.org/10.3390/app10228298

Abstract

Purpose: This paper provides a critical review of the literature on deep learning applications in breast tumor diagnosis using ultrasound and mammography images. **Aim:** To summaries recent advances in computer-aided diagnosis/detection (CAD) systems, which make use of new deep learning methods to automatically recognize breast images and improve the accuracy of diagnoses made by radiologists. **Methodology:** This review is based upon published literature in the past decade (January 2010–January 2020), where we obtained around 250 research articles, and after an eligibility process, 59 articles were presented in more detail. **Results:** The main findings in the classification process revealed that new DL-CAD methods are useful and effective screening tools for breast cancer, thus reducing the need for manual feature extraction. The breast tumor research community can utilize this survey as a basis for their current and future studies.

Key words: breast cancer; computer-aided diagnosis; convolutional neural networks; deep learning; mammography; ultrasound

1.Introduction

Due to the anatomy of the human body, women are more vulnerable to breast cancer than men. Breast cancer is one of the leading causes of death for women globally [1–4] and is a significant public health problem. It occurs due to the uncontrolled growth of breast cells. These cells usually form tumours that can be seen from the breast area via different imaging modalities.

To understand breast cancer, some basic knowledge about the normal structure of the breast is important. Women's breasts are constructed of lobules, ducts, nipples, and fatty tissues (Figure 1) [5]. Normally, epithelial tumours grow inside the lobes, as well as in the ducts, and later form a lump [6], generating breast cancer.

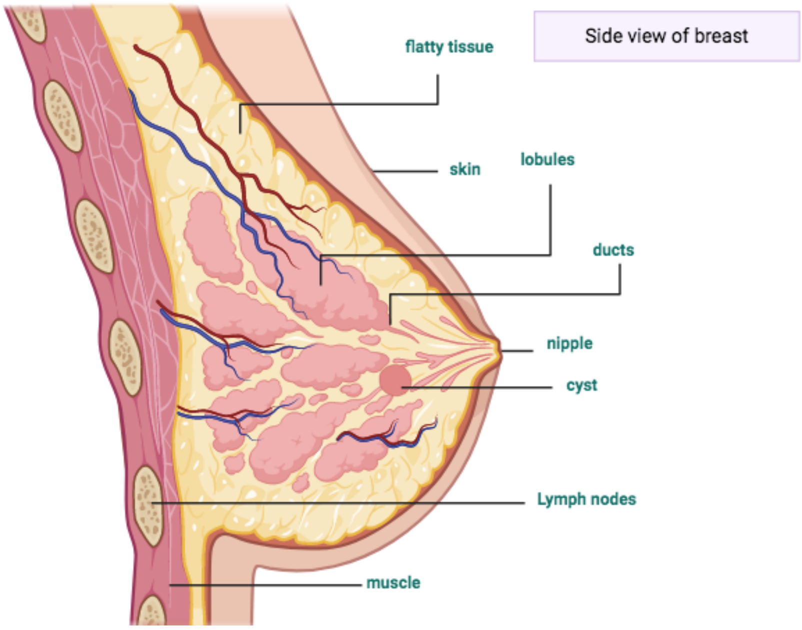


Figure 1. This scheme represents the anatomy of a woman's breast. Inside the lobes are the zones where the epithelial tumours or cyst grow. Source: Biorender (2020). Retrieved from https://app.biorender.com/biorender-templates

Breast abnormalities that can indicate breast cancer are masses and calcifications [7]. Masses are benign or malignant lumps and can be described in terms of their shape (round, lobular, oval, and irregular) or their margin (obscured, indistinct, and spiculated) characteristics. The spiculated masses are the kind of masses that have a high probability of malignancy. A spiculated mass is a lump of tissue with spikes or points on the surface. It is suggestive but not a confirmation of malignancy. It is a common mammography finding in breast carcinoma [8].

On the other hand, microcalcifications are small granular deposits of calcium and may reveal themselves as clusters or patterns (like circles or lines) and appear as bright spots in a mammogram. Benign calcifications are usually larger and coarser with round and smooth contours. Malignant calcifications tend to be numerous, clustered, small, varying in size and shape, angular, and are irregularly shaped [7,9].

Breast cancer screening aims to detect benign or malignant tumours before the symptoms appear, and hence reduce mortality through early intervention [2]. Currently, there are different screening methods, such as mammography [10], magnetic resonance imaging (MRI) [11], ultrasound (US) [12], and computed tomography (CT) [13]. These methods help to visualize hidden diagnostic features. Out of these modalities, ultrasound and mammograms are the most common screening methods for detecting tumours before they become palpable and invasive [2,14–16]. Furthermore, they may be utilized effectively to reduce unnecessary biopsies [17]. These two are the modalities that are reviewed in this article.

A drawback in mammography is that the results depend upon the lesion type, the age of the patient, and the breast density [18–24]. Dense breasts that are "radiographically" hard to see exhibit a low contrast between the cancerous lesions and the background [25,26].

Digital mammography (DM) has some limitations, such as low sensitivity, especially in dense breasts, and therefore other modalities, such as US, are used [12]. US is a non-invasive, non-radioactive, real-time imaging technique that provides high-resolution images [27]. However, all these techniques require manual interpretation by an expert radiologist. Normally, the radiologists try to do a manual interpretation of the medical image via a double mammogram reading to enhance the accuracy of the results [28]. However, this is time-consuming and is highly prone to mistakes [3,29]. Because of these limitations, different artificial intelligence algorithms are gaining attention due to their excellent performance in image recognition tasks.

Different breast image classification methods have been used to assist doctors in reading and interpreting medical images, such as traditional computer-aided diagnosis/detection (CAD) systems [8,30–32] based on machine learning (ML) [33–35], or based on modern CAD-deep learning (DL) system [36–42].

The goal of CAD is to increase the accuracy and sensitivity rates to support radiologists in their diagnosis decisions [43,44]. Recently, Gao et al. [45] developed a CAD system for screening mammography readings that demonstrated an approximately 92% accuracy in the classification. Likewise, other studies [46,47] used several CNNs for mass detection in mammography's and ultrasounds [48–50].

In general, DL-CAD systems focus on CNNs, which is the most popular model used for intelligent image analysis and for detecting cancer with good performance [51,52]. With CNNs, it is possible to automate the feature extraction process as an internal part of the network, thus minimizing human interference. DL-CAD systems have added broader meaning with this approach, distinguishing it from traditional CAD methods.

The next-generation technologies based on the DL-CAD system solve problems that are hard to solve with traditional CAD [12,33]. These problems include learning from complex data [53-54], image recognition [55], medical diagnosis [56,57], and image enhancement [58]. In using such techniques, the image analysis includes pre-processing, segmentation (selection of a region of interest—ROI), feature extraction/selection, and classification.

In this review, we summarize recent upgrades and improvements in new DL-CAD systems for breast cancer detection/diagnosis using mammograms and ultrasound imaging and then describe the principal findings in the classification process. The following research questions were used as the guidelines for this article:

- How the new DL-CAD systems provide breast imaging classification in comparison with the traditional CAD system?
- Which artificial neural networks implemented in DL-CAD systems give better performance regarding breast tumour classification?
- Which is the main DL-CAD architectures used for breast tumour diagnosis/detection?
- What are the performance metrics used for evaluating DL-CAD systems?

2. Materials and Methods

2.1 Flowchart of the Review

The research process is shown in Figure 2, which was in accordance with the PRISMA (Preferred reporting items for systematic reviews and meta-analyses) flow diagram and protocol [59].

Furthermore, the systematic review process follows the flow diagram and protocol (Figure 3) given in [60].

We identified appropriate studies in PubMed, Medline, Google Scholar, and Web of Science databases, as well as conference proceedings from IEEE (Institute of Electrical and Electronics Engineers), MICCAI (Medical Image Computing and Computer Assisted Intervention), and SPIE (Society of Photographic Instrumentation Engineers), published between January 2010 and January 2020. The search was designed to identify all studies in which DM and US were evaluated as a primary detection modality for breast cancer and were both used for screening and diagnosis. A comprehensive search strategy including free text and MeSH terms was utilized, including terms such as: “breast cancer,” “breast tumor,” “breast ultrasound,” “breast diagnostic,” “diagnostic imaging,” “deep learning,” “CAD system,” “convolutional neural network,” “computer-aided detection,” “computer-aided diagnoses,” “digital databases,” “mammography,” “mammary ultrasound,” “radiology information,” and “screening.”

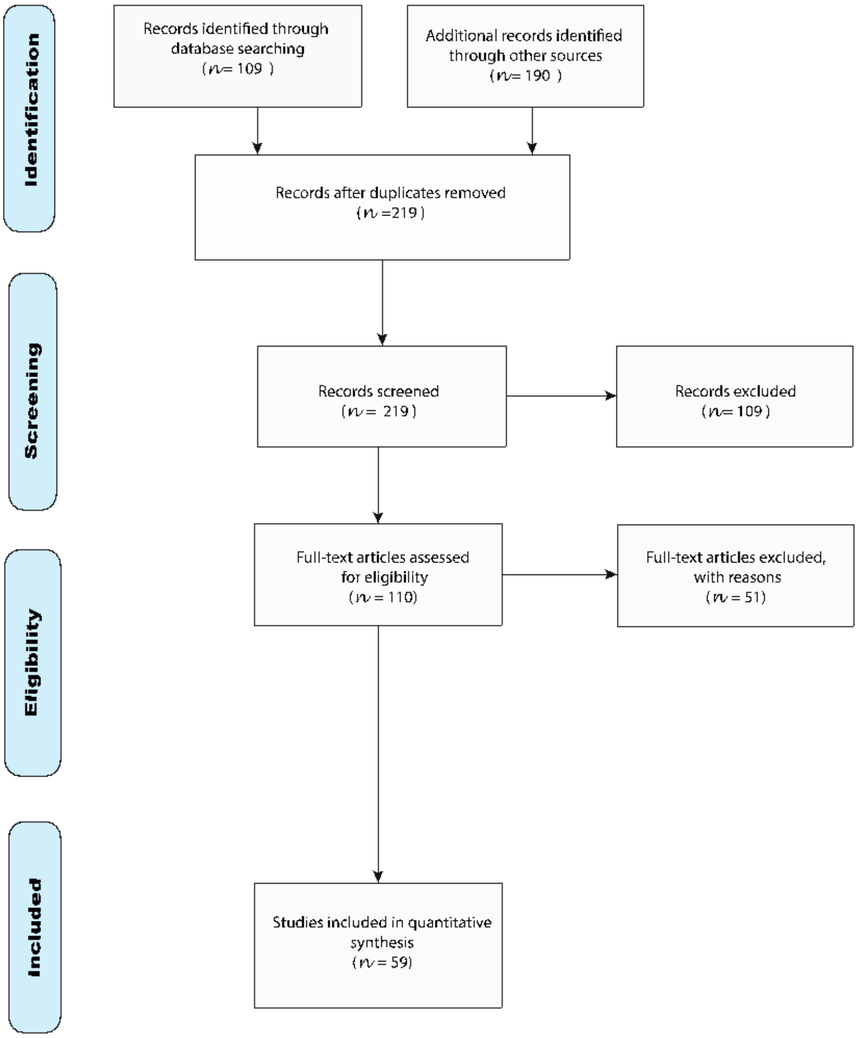


Figure 2. PRISMA flow diagram. Source: own preparation.

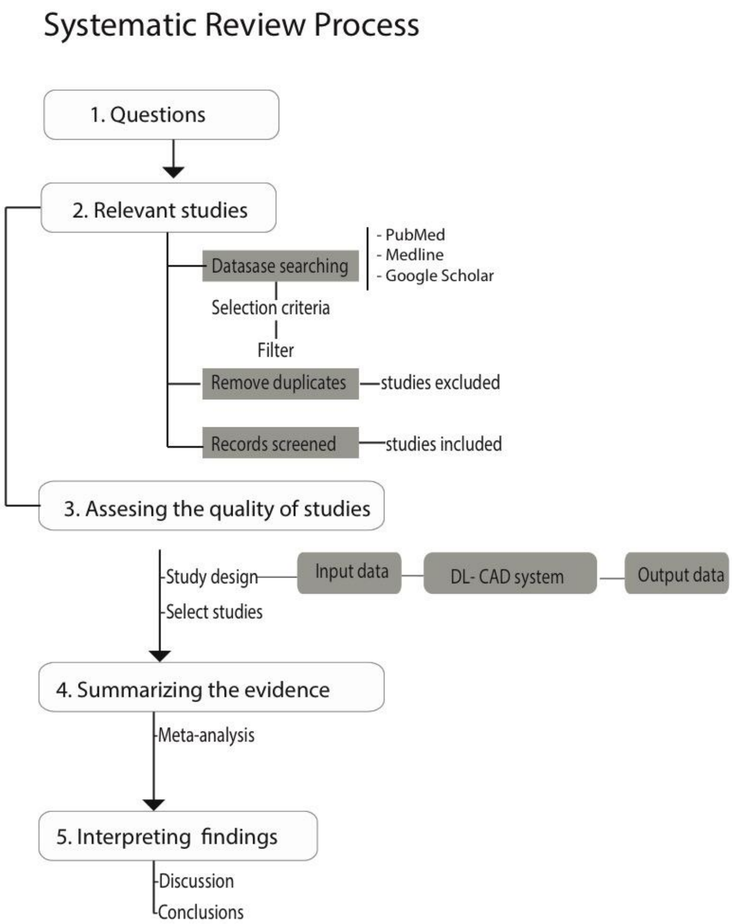


Figure 3. This flowchart diagram represents the review process of articles in this paper. DL-CAD: deep learning computer-aided diagnosis/detection. Source: own preparation.

Inclusion Criteria

Articles were included if they assessed computer-aided diagnosis (CADx) and/or computer-aided detection (CADe) for breast cancer, DL in breast imaging, deep CNN, DL in mass segmentation and classification in both DM and US, deep neural network architecture, transfer learning, and feature-based methods regarding automated DM breast density measurements. From a review of the abstracts, we manually selected the relevant papers.

Exclusion Criteria

Articles were excluded if the study population included other screening methods, such as MRI, CT, PET (positron emission tomography), or if other machine learning techniques were used.

Study Design

The general modern DL-CAD design was divided into four sections (Figure 4). First, different mammography and ultrasound public digital databases were analyzed as input data for the DL-CAD system. The second section includes the preprocessing and postprocessing in the next-generation DL-CAD.

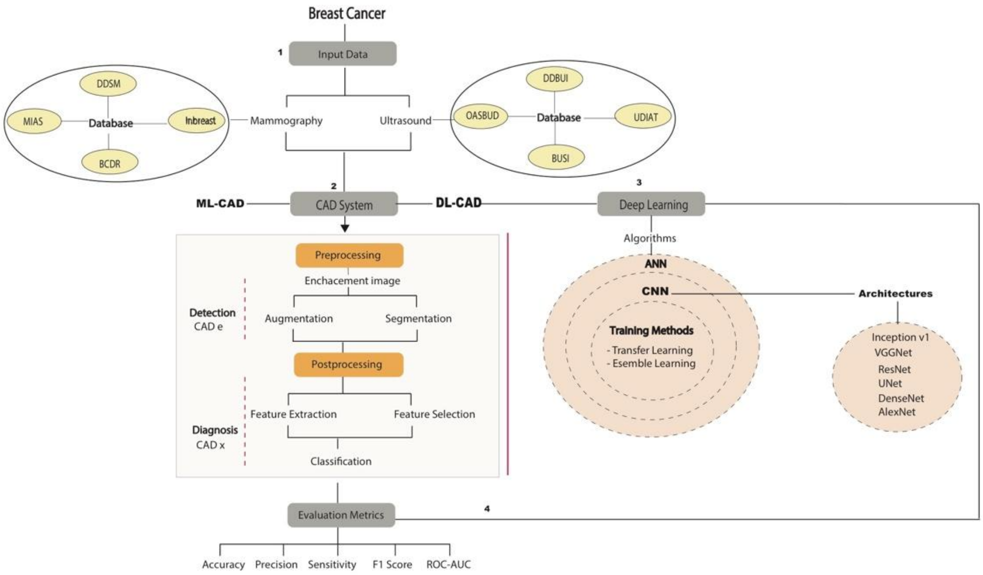


Figure 4. The general diagram is a flowchart that describes how a modern CAD system process can be used with DM and US images from public and private databases. Normally, the CAD system consists of several stages, such as segmentation, feature extraction/selection, and classification. However, DL-CAD systems are based on CNN models and architectures for automatic feature extraction/selection and classification with convolutional and fully connected layers through self-learning. Finally, CAD systems are validated by different metrics. ANN: artificial neural network, BCDR: Breast Cancer Digital Repository, BUSI: Breast Ultrasound Image Dataset, CADe: computer-aided detection, CADx: computer-aided diagnosis, DDBUI: Digital Database for Breast Ultrasound Images, DDSM: Digital Database for Screening Mammography, MIAS: Mammographic Image Analysis Society Digital Mammogram Database, OASBUD: Open Access Series of Breast Ultrasonic Data, ROC–AUC: receiver operating

characteristic curve–area under the curve, UDIAT: Ultrasound Diagnostic Ultrasound Centre of the Parc Tauli, VGGNet: Visual Geometry Group. Source: own preparation.

In the third part, full articles were analyzed to compile the successful CNNs used in DL architectures. Furthermore, the best evaluation metrics were analyzed to measure the accuracy of these algorithms. Finally, a discussion and conclusions about these classifiers are presented.

Public Databases

Normally, DL models are tested using private clinical images or publicly available digital databases that are used by researchers in the breast cancer area. The amount of public medical images is increasing because most of the DL-CAD systems require a large amount of data. Thus, DL algorithms are applied to available digitized mammograms, such as those from MIAS (Mammographic Image Analysis Society Digital Mammogram Database) [61], DDSM (Digital Database for Screening Mammography), IRMA (Image Retrieval in Medical Application) [62,63], INbreast [64], and BCDR (Breast Cancer Digital Repository) [45,65], as well as public US databases, such as BUSI (Breast Ultrasound Image Dataset), DDBUI (Digital Database for Breast Ultrasound Images), and OASBUD (Open Access Series of Breast Ultrasonic Data) from the Oncology Institute in Warsaw, Poland, and the private US collected datasets, such as SNUH (Seoul National University Hospital, Korea) [48], Dataset A (collected in 2001 from a professional didactic media file for breast imaging specialists) [66], and Dataset B collected from the UDIAT(Ultrasound Diagnostic Ultrasound Centre of the Parc Tauli) Corporation, Sabadell, Spain. These widely used datasets are listed in Table 1.

Table 1. Summary of the most used public breast cancer databases in the literature.

Type	Database	Annotations	Link	Author
Mammograms	DDSM	2620 patients including mediolateral oblique (MLO) and craniocaudal (CC) images for classification.	http://www.eng.usf.edu/cvprg/Mammography/Database.html	Jiao et al. [67]
	BCDR	736 biopsies prove lesion of 344 patients, including CC and MLO images for classification.	https://bcdr.eu/	Arevalo et al. [68]
	INbreast	419 cases, including CC and MLO images of 115 patients, for detection and diagnosis.	http://medicalresearch.inescporto.pt/breastresearch/index.php/Get_INbreast_Database	IMoreira et al. [64]
	Mini-MIAS	322 digitized MLO images of 161 patients for segmentation, detection, and classification.	http://peipa.essex.ac.uk/info/mias.html	Peng et al. [69]
Ultrasound	BUSI	The dataset consists of 600 female patients. The 780 images include 133 normal images without masses, 437 images with cancer masses, and 210 images with benign masses. This set is utilized for classification, detection, and segmentation.	https://scholar.cu.edu.eg/?q=afahmy/pages/dataset	Dhabyani et al. [70]
	DDBUI	285 cases and 1132 images in total for classification.	https://www.atlantis-press.com/proceedings/jcis2008/1735	Tian et al. [71]
	Dataset A	Private dataset with 306 (60 malignant and 246 benign) images, which are utilized for detection.	goo.gl/SJmoti	Yap et al. [48]
	Dataset B	Private dataset with 163 (53 malignant and 110 benign) images.		Byra et al. [66]
	SNUH	Private dataset with a total of 1225 patients with 1687 tumors. This study includes biopsy diagnosis.		Moon et al. [49]
	OASBUD	52 malignant and 48 benign masses, which are utilized in image processing algorithms.	http://bluebox.ippt.gov.pl/~hpiotrzk	Piotrzkowska et al. [72]

	Image Net	882 US images (678 benign and 204 malignant lesions), which are utilized in object recognition, image classification, and automatic object clustering.	http://www.image-net.org/	Deng et al. [73]

SNUH: Seoul National University Hospital.

CAD Focused on DM and US

The CAD systems are divided into two categories. One is the traditional CAD system and the other is the DL-CAD system (Figure 5). In the traditional CAD system, the radiologist or clinician defines features in the image, where there can be problems regarding recognizing the shape and density information of the cancerous area. A DL-CAD system, on the other hand, creates such features by itself through the learning process [74].

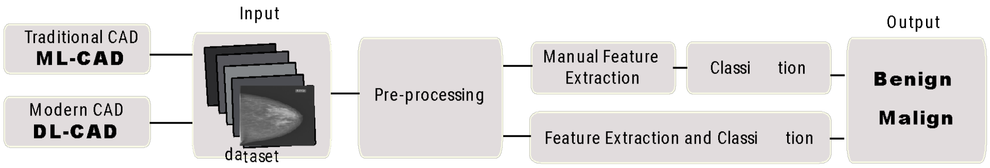


Figure 5. The scheme describes the main difference between the traditional machine learning (ML)-CAD system and the DL-CAD system. Source: own preparation.

Furthermore, CAD systems can be broken down into two main groups: CADe and CADx. The main difference between CADe and CADx is that the first refers to a software tool that assists in ROI segmentation within an image [75], identifying possible abnormalities and leaving the interpretation to the radiologist [8]. On the other hand, CADx serves as a decision aid for radiologists to characterize findings from a CADe system. Several significant CAD works are described in Table 2.

Table 2. The traditional CAD system summary with DM and US breast cancer images. It covers four stages: (i) image processing, (ii) segmentation, (iii) feature extraction and selection, and (iv) classification.

Reference	Models	Description	Application
[76,77]	Pixel-based, which is based on the curvature of the edge and clustering [3,78]: ventional (CLAHE), region-based, feature-based (wavelet), and fuzzy.	Pectoral removal techniques are not sufficient to provide accurate results. Thereby, intensity-based methods, line detection, statistical techniques, wavelet methods, and the active contour technique have also been tried for segmenting this area. Its accuracy varies from 84% to 99%, where the active contour technique	Preprocessing

		reached the highest value of 99%, followed by the wavelet method with 93% [79]. Enchancement techniques are divided into three categories: spatial, frequency domain, and a mixture of these two. These categories can be classified into four models. The region-based method requires a seed point and it is time-consuming.	
[80,81]	Local thresholding and region-growing [82]; edge detection, template matching [12,83], and a multiscale technique [84]; NN [85].	The thresholding method shows greater stability but is dependent on the parameter selection. Furthermore, is not sufficient for segmenting fatty tissue in a DM because its images contain noise and have low contrast and intensity. The region-growing method is well-known in micro-calcification detection and uses pixel properties for segmenting fatty tissue. Edge detection utilizes the wavelet transform in a multiscale structure to represent signals and variations in US images. Template matching requires a comparison with a given image (ROI) with a template image to measure the similarity between both. Finally, an NN utilizes a multi-layered perceptron with a hidden layer for extracting the contours of tumors automatically; nevertheless, training an NN is time-consuming.	Segmentation
[86]	PCA [87], LDA [88], and GLCM [89].	ture selection methods: wrapper and filter (chi-square [90]). The most well-known feature extraction techniques are PCA, LDA, GLCM, gain ratio, recursive feature [91], RF, WPT [92,93], Fourier power spectrum [94], Gaussian [95] and DBT [29]. PCA feature extraction techniques are better at reducing the high-dimensional correlated features into low dimensional features [87].	ture Selection and extraction
[12,33]	SVM [96,97] and ANN [98,99].	SVM is useful in DM classification because these are highly overlapping and nonlinear in their feature space. It minimizes the generalization error during the process of testing data and is much more accurate and computational efficient because of the reduced number of parameters. ANN: Backpropagation, SOM, and hierarchical ANN. The performance of back-propagation is better than that of linear classifiers. However, the training process is stochastic and unrepeatable, even with the same data and initial conditions. Prone to overfitting due to the complexity of the model structure.	Classification

Finally, advantages and disadvantages from other classifiers have been previously discussed in several studies: KNN [100], BDT [101], simple logistic classifier [102], and DBN [103]

Preprocessing

It is known that the database characteristics can significantly affect the performance of a CAD scheme, or even a particular processing technique. Furthermore, it can develop a scheme that yields erroneous or confusing results [104] since radiological images contain noise, artefacts, and other factors that can affect medical and computer interpretations. Thus, the first step in preprocessing is to improve the image quality, contrast, removal noise, and pectoral muscle [105].

Image Enhancement

The main purpose of image preprocessing is to enhance the image and suppress noise while preserving important diagnostic features [106,107]. Preprocessing for breast cancer diagnosis also consists of delineation of the tumors from the background, breast border extraction, and pectoral muscle removal. The pectoral muscle segmentation is a challenge in mammogram image analysis because the density and texture information is similar to that of the breast tissues. Furthermore, it depends on the standard view used during mammography. Generally, mediolateral oblique (MLO) and craniocaudal (CC) views are used [78].

As noted, DM includes many sources of noise, which are classified as a high-intensity, low-intensity, or tape artefacts. The principal noise models observed in mammography are salt and pepper, Gaussian, speckle, and Poisson noise.

In the same way, US images suffer from noise, such as intensity inhomogeneity, a low signal-to-noise ratio, high speckle noise [108,109], blurry boundaries, shadow, attenuation, speckle interference, and low contrast. Speckle noise reduction techniques are categorized in filtering, wavelet, and compound methods [12].

Thus, many traditional filters can be applied for noise removal, including a wavelet transform, median filter, mean filter, adaptive median filter, Gaussian filter, and adaptive Wiener filter [3,110–113]. Furthermore, different traditional methods, such as histogram equalization (HE) [114,115], adaptive histogram equalization (AHE) [116], and contrast-limited adaptive histogram equalization (CLAHE) [117], can be used to enhance the image.

On the other hand, deep CNNs [118] are gaining attention for improving super-resolution [119] images (SR) based on a CNN, namely, (i) multi-image super-resolution and (ii) single-image super-resolution [120,121]. Among the most used algorithms for generating high-resolution (HR) imaging [122,123] are nearest-neighbor interpolation [124], bilinear interpolation [125], and bicubic interpolation [126].

Image Augmentation

Deep CNN depends on large datasets to avoid overfitting and is necessary for good DL model performance [127]. Thus, limited datasets are a major challenge in medical image processing [128] and it is necessary to implement data augmentation techniques. There are two common techniques for increasing the data in DL, namely, data augmentation and transfer learning/fine-tuning [129,130]. Examples of DL models that have been trained with data augmentation are Imagenet [74] and transfer learning [47].

The image augmentation algorithms include basic image manipulations (flipping, rotation, transformation, feature space augmentation, kernel, mixing images, and random erasing [131]) and DL manipulations (generative adversarial networks (GANs)) [132], along with a neural style transfer [133] and meta-learning [128]). These techniques increase the amount of data by preprocessing input image data via operations such as contrast enhancement and noise addition, which have been implemented in many studies [134–140].

Image Segmentation

This processing step plays an important role in image classification. Segmentation is the separation of ROIs (lesions, masses, and microcalcifications) from the background of the image.

In traditional CAD systems, the tasks of specifying the ROI, such as an initial boundary or lesions, are accomplished with the expertise of radiologists. The traditional segmentation task in DM can be divided into four main classes: (i) threshold-based segmentation, (ii) region-based segmentation, (iii) pixel-based segmentation, and (iv) model-based segmentation [3,78]. Furthermore, US image segmentation includes several techniques: threshold-based, region-based, edge-based, water-based, active-contour-based, and neural-network-learning-based techniques [141,142].

The accuracy of the segmentation affects the results of CAD systems because numerous features are used for distinguishing malignant and benign tumors (texture, contour, and shape of lesions). Thus, the features may only be effectively extracted if the segmentation of tumors is performed with great accuracy [106,142]. This is why researchers are using DL methods, especially CNNs, because these methods have shown excellent results on segmentation tasks. Furthermore, DL-CAD systems are independent of human involvement and are capable of autonomously modeling breast US and DM knowledge using constraints. Two strategies have been utilized for full image sizes for training CNNs for DM and US instead of ROIs: (1) high-resolution [143] and (2) patch-level [144] images. For example, recent network architectures that have been used to produce segmented regions are YOLO [145], SegNet [146,147], UNet [148], GAN [149], and ERFNet [150].

Postprocessing

Image Feature Extraction and Selection

After the segmentation, feature extraction and selection are the next steps to remove the irrelevant and redundant information of the data being processed. Features are characteristics of the ROI taken from the shape and margin of lesions, masses, and calcifications. These features can be categorized into texture and morphologic features [12,86], descriptors, and model-based features [52], which help to discriminate between benign and malignant lesions. Most of the texture features are calculated from the entire image or ROIs using the gray-level value and the morphological features.

There are some traditional techniques used for feature selection, such as searching algorithms, the chi-square test, random forest, gain ratio, and recursive feature elimination [91]. In addition, other traditional techniques used for the feature extraction include principal component analysis (PCA), wavelet packet transform (WPT) [92,93], grey-level co-occurrence matrix (GLCM) [91], Fourier power spectrum (FPS) [94], Gaussian derivative kernels [95], and decision boundary features (DBT) [151].

However, in some classification processes, such as an ANN or support vector machine (SVM), the dimension of the vectors affects both the computational time and the performance [152] because this depends on the number of features extracted. Thus, feature selection techniques reduce the size of the feature space, improving the accuracy and computation time by eliminating redundant features [153]. DL models produce a set of image features from the data [154], whose main advantage is that they extract features and perform classifications directly. Providing good extraction and selection of the features is a crucial task for DL-CAD systems; for example, some CNNs that are capable of extracting features have been presented by different authors [155,156].

Classification

During the classification process, the dimension of feature vectors is important because these affect the performance of the classifier. The features of breast US images can be divided into four types: texture, morphological, model-based, and descriptor features [86]. After the features have been extracted and selected, they are input into a classifier to categorize the ROI into malignant and benign classes. The commonly used classifiers include linear, ANN, Bayesian neural networks, decision tree, SVM, template matching [106], and CNNs.

Recently, deep CNNs, which are hierarchical architectures trained on large-scale datasets, have shown stunning performances regarding object recognition and detection [157], which suggests that these could also improve breast lesion detection in both US and DM methods. Some researchers are interested in lesion [158], microcalcification [159,160], and mass [161,162] classification in DM and US [15–154,163–165] images based on CNN models.

Deep Learning Models

DL in medical imaging is mostly represented by a basic structure called a CNN [57,75]. There are different DL techniques, such as GANs, deep autoencoders (DANs), restricted Boltzmann machine (RBM), stacked autoencoders (SAEs), convolutional autoencoders (CAEs), recurrent neural networks (RNNs), long short-term memory (LSTM), multiscale convolutional neural network (M-CNN), and multi-instance learning convolutional neural network (MIL-CNN) [3]. DL techniques have been implemented to train neural networks for breast lesion detection, including ensemble [75] and transfer learning [129,157,166] methods. The ensemble method combines several basic models to get an optimal model [167], and transfer learning is an effective DL method to pre-train models to deal with small datasets, as in the case of medical images.

ANNs are composed of an input and output layer, plus one or more hidden layers, as shown in Figure 6. In the field of breast cancer, three types of ANN are frequently used: backpropagation, SOM, and hierarchical ANNs. To train an ANN with a backpropagation algorithm, the error function is given to calculate the gradient descent. This error propagates in the backward direction and the weights are adjusted for error reduction. This processing is repeated until the error becomes zero or is a minimum [3].

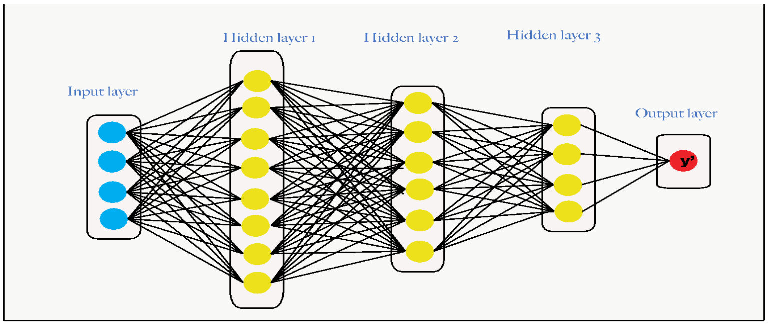


Figure 6. An ANN learns by processing images, where each of which contains an input, hidden, and result layer. Source: own preparation.

Convolutional Neural Networks

CNNs are the most widely used Neural Networks when it comes to DL and medical image analysis. The CNN structure has three types of layers: (i) convolution, (ii) pooling, and (iii) full-connection layers, which are stacked in multiple layers [74]. Thus, a CNN's structure is determined by different parameters, such as the number of hidden layers, the learning rate, the activation function (rectified linear unit (ReLU)), pooling layer for feature map extraction, loss function (softmax), and the fully connected layers for classification, as shown in Figure 7.

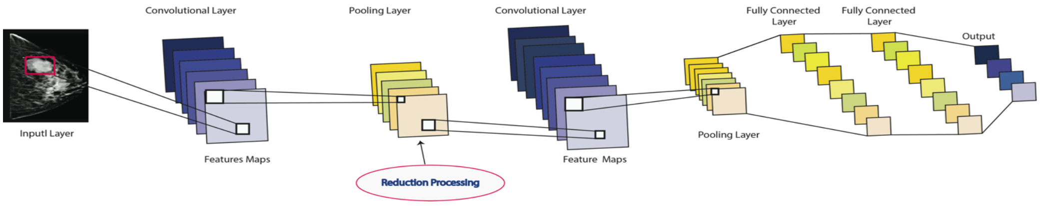


Figure 7. A feed-forward CNN network, where the convolutional layers are the main components, followed by a nonlinear layer (rectified linear unit (ReLU)), pooling layer for feature map extraction, loss function (softmax), and the fully connected layers for classification. The output can be either benign or malignant classes.

Furthermore, there are several methods for improving a CNN's performance, such as dropout and batch normalization. Dropout is a regularization method that is used to prevent a CNN model from overfitting. A batch normalization layer speeds up the training of CNNs and reduces the sensitivity to network initialization.

Evaluation Metrics

Different quantitative metrics are used to evaluate the classifier performance of a DL-CAD system. These include accuracy (Acc), Sensitivity (Sen), Specificity (Spe), area under the curve (AUC), F1 score, and a confusion matrix. The statistical equations are shown in Tables 3 and 4.

Table 3. Confusion matrix for a binary classifier that is used to distinguish between two classes, namely, benign, and malignant. TP: true positive; FN: false negative, FP: false positive, TN: true negative, TPR: true positive rate, FPR: false positive rate.

Classes	**Predicted Classes**		**Equation**
	C_1	**C_2**	
C_1 (Benign)	TP	FN	$TPR = \left(\frac{TP}{TP+FN}\right)$
C_2 (Malignant)	FP	TN	$FPR = \left(\frac{FP}{FP+TN}\right)$

Table 4. Validation assessment measures.

Model	**Equation**
Accuracy	$\mathrm{Acc} = \left(\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}}\right)$

Sensitivity	TPR
Specificity	$TNR = \left(\frac{TN}{TN + FN}\right)$
Precision	$Precision = \left(\frac{TP}{TP + FP}\right)$
F1 Score	$F1\ score = 2x\left(\frac{precision \times recall}{precision + recall}\right)$
MCC	$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$

The receiver operating characteristic curve (ROC) is a graph for plotting the true positive rate (TPR) versus a false positive rate (FPR) and is derived from the AUC. The TPR and the FPR are also called sensitivity (recall) and specificity, respectively, as shown in Figure 8.

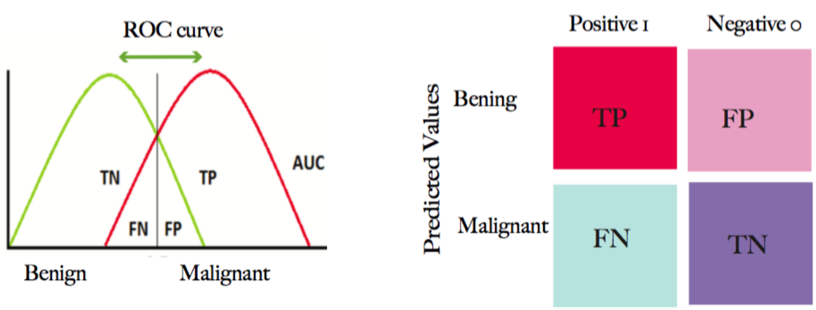


Figure 8. The confusion matrix for the ROC. The number of images correctly predicted by the classifier is located on the diagonal. The ROC curve utilizes the TPR on the *y*-axis and the FPR on the *x*-axis. Source: own preparation.

The AUC provides the area under the ROC curve and a perfect score has a range from 0.5 to 1. A 100% correct classified version will have an AUC value of 1 and it will be 0 if there is a 100% wrong classification [168].

Cross-validation is a statistical technique that is used to evaluate predictive models by partitioning the original samples into training, validation, and testing sets. There are three types of validation: (1) hold-out splits (training 80% and testing 20%), (2) three-way data split (training 60%, validation 20%, and testing 20%), and (3) K-fold cross-validation (from 3 to 5 k-fold for a large data set, 10 k-fold for a small dataset), where the data are split into k different subsets depending on their size [65].

3. Results

3.1 CNN Architectures

A model's performance depends on the architecture and the size of the data. In this sense, there are different CNN architectures that have been proposed: AlexNet [169], VGG-16 [170], ResNet [171], Inception (GoogleNet) [172], and DenseNet [173]. These networks have shown promising performance in recent works for image detection and classification. Table 5 shows a brief description of these networks.

Table 5. Summary of CNN architecture information for breast imaging processing.

Reference	Model	Description	Training Method	Application
Krizhevsky et al. [169]	AlexNet	A deep CNN evaluated using the Imagenet [65] LSVRC-2010 dataset [173], with top-1 and top-5 error rates of 37.5% and 17.0%, respectively. This achieved a top-5 test error rate of 15.3% compared to 26.2% (ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) 2012).	Dropout model	Classification
Samala et al. [174]	DL-CNN	CAD system for masses in DBT volume, which is trained using transfer learning. The best AUC obtained was 0.933 and the improvement was statistically significant ($p < 0.05$).	CNN architecture	Detection tomosynthesis from DM
Simoyan et al. [170]	VGG-VD	The very deep (VD)-CNN models (VGG-VD16 and VGG-VD19 [158]) were evaluated in ILSVRC 2014 (ImageNet).	Deep ConvNet architecture	Classification
He et al. [171]	ResNet	An ensemble of these residual nets achieved a 3.57% error on the ImageNet (ILSVRC 2015) test set.	ResNet with a depth of up to 152 layers 8× deeper	Classification
Huang et al. [172]	DenseNet	DenseNet was proposed to reduce the vanishing gradient problem, to reduce the number of parameters, and to strengthen the feature propagation.	ImageNet with a CNN	Object recognition
Szegedy et al. [27]	Inception v5	A deep CNN was evaluated in ILSVRC 2014.	Deep-CNN	Classification and detection
Das et al. [175]	VGGNet	BreakHist dataset with 58 malignant and 24 benign cases was evaluated with a deep CNN. The best accuracy percentage was reached with 100× (89.06%).	MIL architecture	Histopathology

Cao et al. [152]	Deep CNN	Private dataset that contains 577 benign and 464 malignant cases.	Detection: Fast R-CNN, Faster R-CNN, YOLOv3, and SSD; Classification: AlexNet, VGG, ResNet, GoogleNet, ZFNet, and Densenet	US lesion detection and classification
Chiao et al. [153]	Deep CNN	Private US imaging dataset that contains 307 images with 107 benign and 129 malignant cases.	Mask R-CNN with ROI alignment; based on a faster R-CNN using an RPN to extract features	Sonogram lesion detection and classification
Yap et al. [48]	LeNet, UNet, deep CNN	This work studies the performance of CNNs in breast US detection using two private datasets A and B.	LeNet [163], U-Net [148], and transfer learning [176]	US breast lesion detection
Geras, K. et al. [176]	Multi-view DL-CNN	INbreast [77] and DDSM [58] databases were used; the model achieved an AUC of 0.68%.	The CNN is jointly trained using stochastic gradient descent with backpropagation [175,176] and data augmentation via random cropping [168]	High-resolution, augmentation, and DM classification
Han et al. [62]	GoogleNet with ensemble learning	Dataset contains a total of 7408 US breast images, with 657 used as the training set and 829 as the test set. The accuracy reached was 90.21%.	The CNN was trained with 10-fold cross-validation. Data augmentation was carrying out with the Caffe method	Data augmentation, detection, and classification of breast lesions in US

Dhung el et al. [178]	LeNet for CNN models in cascade R-CNN	INbreast dataset was used, with 115 cases and 410 images from MLO and CC views. The results showed that the DL-CAD system is able to detect 90% of masses, with a segmentation accuracy of 85% and the classification reached a sensitivity of 0.98 and a specificity of 0.7.	DL detection: Fast R-CNN, multiscale-DBN, and random forest; DL segmentation: CRF; DL classification: regression method.	Detection, segmentation, and classification of masses in DM
Singh et al [165]	GAN	The Mendeley database [179] was used, which contains 150 malignant and 100 benign tumors. The performance metrics achieved scores of dice = 93.76% and IoU = 88.82%.	Segmentation with GAN learning.	Segmentation and classification of US images
Cheng, J. Z. [37]	SDAE based CADx	The method was carried out on a private database, with 520 breast sonograms (275 benign and 245 malignant lesions). The AUC performance reached 0.80%.	An SDAE (OverFeat) model was used to classify with the ensemble method.	Breast lesion/nodules diagnosis and classification of US images

3.2 Performance Metrics

Furthermore, brief reviews of the DL architectures based on DM and US breast images, along with their evaluation metrics, are presented in Tables 6 and 7 [50,180].

Table 6. The quantitative indicators that were used to evaluate the performance between different CNN architectures in DM datasets.

Reference	Database	Deep CNN Model	Acc (%)	Sen (%)	Spec (%)	Precision (%)	F1 Score (%)	AUC (%)
Al-Masni et al. [145]	DDSM with 600 DM.	CNN YOLO5: Fold cross-validation in both datasets; mass classification	99	93.20	78	-	-	87.74
	DDSM augmentation with 2.400	Mass detection	97	100	94	-	-	96.45
Ragab et al. [168]	DDSM with 2620 cases	Deep-CNN-based linear SVM using ROI manually	79	76.3	82.2	85	80	88
	CBIS- DDSM with 1644 cases	ROI threshold	80.5	77.4	84.2	86	81.5	88
		SVM-based medium Gaussian	87.2	86.2	87.7	88	87.1	94

Duggento et al. [180]	CBIS-DDSM	Deep CNN	71	84.4	62.4	-	-	77
Chougrad et al. [181]	BCDR	Inceptionv3	96.67	-	-	-	-	96
	DDSM		97.35	-	-	-	-	98
	INbreast		95.50	-	-	-	-	97
	MIAS		98.23	-	-	--		99

Table 7. The quantitative indicators that were used to evaluate different CNN architectures' performances on US datasets.

Reference	Database	Deep CNN Model	Acc (%)	Sen (%)	Spec (%)	Precision (%)	F1 Score (%)	AUC (%)
Moon et al. [49]	BUSI SNUH	VGGNet-like	84.57	73.65	93.12	89.34	80.74	91.98
		VGGNet 16	84.57	73.64	93.12	89.34	80.74	93.22
		ResNet 18	81.60	86.49	77.77	75.29	80.50	91.85
		ResNet 50	81.60	75.68	86.24	81.16	78.32	88.83
		ResNet 101	84.57	75,00	92.06	88.10	81.02	91.04
		DenseNet 40	85.46	79.05	90.48	86.67	82.69	93.52
		DenseNet 12	86.35	77.70	93.12	89.84	83.33	92.48
		DenseNet 161	83.09	69.59	93.65	9.57	78.33	89.18
Byra et al. [66,182]	ImageNet	VGG19	88.7	0.848	0.897	-	-	93.6
	UDIAT		84	0.851	0.834	-	-	89.3
	OASBUD [150]		83	0.807	0.854	-	-	88.1
Cao et al. [152]	Private dataset consisting of 579 benign and 464 malignant cases	Single Shot Detector (SSD)300 + ZFNet	96.89	67.23	-	-	79.38	-
		YOLO	96.81	65.83	-	-	78.37	-
		SSD300 + VGG16	96.42	66.70	-	-	78.85	-
Han et al. [62]	Private database with a total of 7408 US images with 4254 benign and 3154 malignant lesions	CNN-based GoogleNet	91.23	84.29	96.07		-	91
Shan et al. [35]	Private database containing	ANN	78.1	78	78.2	-	-	82.3

283 breast US images (133 cases are benign and 150 cases are malignant)

Furthermore, Table 8 gives a brief overview of the new DL-CAD systems' approaches and the traditional ML-CAD systems.

Table 8. DL-CAD systems vs. traditional ML-CAD systems.

Reference	Application	Method	Dataset	Acc (%)	Sen (%)	Spec (%)	AUC (%)	Error (%)
Dheeba [183]	DM classification	ML wavelet neural network	Private database consisting of 216 multiview CC and MLO images.	93.67	94.16	92.10	96.85	96.85
Trivizakis et al. [184]	DM classification	ML with transfer learning and features based on ImageNet and CNN architecture	Mini MIAS and	79.3	-	-	84.2	-
			DDSM	74.8	-	-	78.00	-
Samala et al. [185]	DM classification	Multitask transfer learning by a Deep CNN	ImageNet	90	-	-	82	-
Jadoon et al. [186]	DM extraction and classification	CNN + wavelet	IRMA, DDSM, and MIAS	81.83	-	-	83.1	15.43
		CNN + SVM		83.74	-	-	83.9	17.46
Debelee et al. [42]	DM extraction	CNN + SVM	MIAS	97.46	96.26	100	-	-
			DDSM	99	99.48	98.16	-	-
			MIAS	87.64	96.65	75.73	-	-
		MLP	DDSM	97	97.40	96.26	-	-
			MIAS	91.11	86.66	100	-	-
		KNN + SVM	DDSM	97.18	100	95.65	-	-
Ahmed et al. [187]	DM detection	Deep CNN with five–fold cross-validation	INbreast	80.10	80	-	78	-
Xu et al. [51]	US image segmentation	Deep CNN	Private 3D breast US	90.13	88.88	-	-	-
Shan et al. [35]	US image	ML decision tree	Private breast US consisting of 283 images, where 133 cases are benign and	77.7	74.0	82.0	80	-
		ANN		78.1	78.0	78.2	82	-
		Random forest		78.5	75.3	82	82	-

	segmentation	SVM	150 cases are malignant	77.7	77.3	78.2	84	-
Gu et al. [188]	3D US image segmentation	Preprocessing: morphological reconstruction; segmentation: region-based approach	Private database with 21 cases, with masses prior to biopsy	85.7	-	-	-	-
Zhang et al. [36]	US image feature extraction and classification	DL architecture	The private dataset consisting of 227 elastography images, with 135 benign tumors and 92 malignant tumors	93.4	88.6	97.1	94.7	-
Almajalid et al. [147]	US image segmentation	DL-CNN architecture U-net	The private dataset containing 221 BUS images	82.52	78.66	18.59	-	-
Singh et al. [189]	US image classification	ML fuzzy c-means and backpropagation ANN	178 breast US containing 88 benign and 90 malignant cases	95.86	95.14	96.58	95.85	-
Cheng et al. [37]	US (sonogram) classification	DL-SDAE	520 breast US (275 benign and 245 malignant lesions)	82.4	78.7	85.7	89.6	_
Shi, et al. [190]	US image classification	Deep polynomial network	A total of 200 pathology-proven breast US images	92.40	92.67	91.36	-	-

4. Discussion and Conclusions

Considering that breast tumor screening using DM has some consequences and limitations because a higher number of unnecessary biopsies and ionizing radiation exposure endangers the patient's health [12], along with low specificity and high FP results, which imply higher, recall rates and higher FN results [191]. This is why US is used as the second choice for DM. Thus, US imaging is one of the most effective tools in breast cancer detection because it has been shown to achieve high accuracy in mass detection, classification [38], and diagnosis of abnormalities in dense breasts [192].

For the abovementioned reasons, we have addressed using both kinds (DM and US) of images in this review, focusing on different ML and DL architectures applied in breast tumor processing, and offering a general overview of databases and CNNs, including their relation and efficacy in performing segmentation, feature extraction, selection, and classification tasks [192].

Thus, according to the research shown in Table 1, the most utilized databases for DM images are MIAS and DDSM, and for US image classification, the public databases BUSI, DDBUI, and OASBUD are most used. The DM images contributed to 110 and 168 published conference papers for the DDSM and MIAS databases, respectively [5]. However, the databases report some limitations and advantages; for example, the MIAS database contains a limited number of images, strong noise, and low-resolution images. In contrast, the DDSM contains a big dataset. Likewise, INbreast contains high-resolution images but has a small data size. BCDR, in comparison with DDSM, has been used in a few studies. Some details about the others strengths and limitations of these databases are discussed in Abdelhafiz [65].

Thereby, Table 2 shows a summary of traditional ML-CAD systems that use public and private databases of DM and US breast images. It covers (i) image preprocessing and (ii) postprocessing steps. This is in contrast with Table 5, which shows a brief summary of DL-CAD systems based on CNN architectures in both types of digital breast images. Thus, in Table 5, various DL architectures and their training strategies for detection and classification tasks are discussed. Based on the most popular datasets, CNN seems to perform rather well, as demonstrated by Chiao et al., Yap et al., and Samala et al. [48,153,174]. Furthermore, [169,173] used several preprocessing and postprocessing techniques for high-resolution [58] data augmentation, segmentation, and classification. The most commonly CNNs used are AlexNet, VGG, ResNet, DenseNet, Inception (GoogleNet), LeNet, and UNet, which employ recent Python libraries for implementing CNNs, such as Tensorflow, Caffe, and Keras, with different hyper-parameters to training the network [55].

Most of these DL architectures use a large data set; thus, it is required to apply an augmentation technique to avoid overfiting and to have better performance during classification. In this sense, the researchers mentioned in Tables 6 [145,168,180,181] and 7 [35,49,62,66,152,182] the authors used transfer learning and ensemble methods, such as data augmentation, to improve the performance of the CNN network, reaching an 89.86% accuracy and 0.9578% AUC in DM, and an AUC of 0.68% on US images. Furthermore, Singh et al. [165] showed that the results obtained with a GAN for breast tumor segmentation outperformed the UNet model, and the SegNet and ERFNet models yielded the worst segmentation results on US images.

In addition, according to Cheng et al. [37], DL techniques could potentially change the design paradigm of CADx systems due to their several advantages over the traditional CAD systems. These are as follows: First, DL can directly extract features from the training data. Second, the feature selection process will be significantly simplified. Third, the three steps of feature extraction, selection, and classification can be realized within the same deep architecture. Thus, SDAE architecture can potentially address the issues of high variation in either the shape or appearance of lesions/tumors. Furthermore, various studies [41,55,39,40] prove that those CNN methods that compare images from CC and MLO views improve the accuracy of detection and reduce the FPR.

Furthermore, different evaluation metrics are described in Tables 3 and 4 as corroboration of the performance of these techniques. The results in Tables 6 and 7 describe different research where their authors have used a variety of datasets (Table 1), approaches, and performance metrics to evaluate CNN techniques in DM and US imaging. For example, better results were achieved in DM analysis by Al-Masni [145] with YOLO5 using DDSM data augmentation, while Chougrad et al. [181] used a deep CNN (Inception V3) with DDSM and MIAS datasets. On the other hand, Moon et al. [49] introduced a DenseNet model to analyze private (BUSI and SNUH) US datasets. Byra et al. [66] achieved high accuracy with the VGG19 deep CNN model using the ImageNet database. Similarly, Cao et al. [152] attained an accuracy of 96.89% with SSD + ZFNet and Han et al. [62] reached 91.23% using a private dataset with GoogleNet.

Likewise, Table 8 contains a literature review for the comparison of the evaluation metrics between DL-CAD systems and traditional ML-CAD systems. Even though Table 8 shows that Deheeba et al. [183] presented a good traditional wavelet neural network CAD system with high accuracy (93.67%) and AUC of 96.85%, Debelee et al. [42] exceeded this percentage using a CNN + SVM DL-CAD system with DDSM (99%) and MIAS (97.18%) DM datasets. In US images Zhang et al. [36] and Shi et al. [190] proved that a DL-CAD based on CNN and a deep polynomial network achieved better results in terms of accuracy (93.4 and 92.40%) and AUC (94.7%), respectively. In the same way, DL-CAD reached higher values than ML-CAD when used on private US images. For example, Shan et al. [35] and Singh et al. [41] showed ML based on an ANN for segmentation and classification that reached accuracies of 78.5 and 95.86% and an AUC of 82%, respectively. These works demonstrate that in most cases, the DL architectures outperformed traditional methodologies.

To conclude, the use of DL could be a promising new technique to obtain the main features for automatic breast tumor classification, especially in dense breasts. Furthermore, in medical image analysis, using DL has proven to be better for researchers compared to a conventional ML approach [41,42]. It appears as though DL provides a mechanism to extract features automatically through a self-learning network, thus boosting the classification accuracy. However, there is a continuing need for better architectures, more extensive datasets that overcome class imbalance problems, and better optimization methods.

Finally, the main limitation in this work is that several algorithms and results are not available in the open literature because of proprietary intellectual property issues.

Author Contributions: Conceptualization, Y.J.-G. and V.L.; methodology, Y.J.-G.; formal analysis, Y.J.-G., M.J.R.-Á., and V.L.; investigation, Y.J.-G.; resources, Y.J.-G.; writing—original draft preparation, Y.J.-G.; writing—review and editing, Y.J.-G., M.J.R.-Á., and V.L.; V.L. visualization, Y.J.-G.; supervision, M.J.R.-Á. and V.L.; project administration, M.J.R.-Á. and V.L.; funding acquisition, V.L. All authors have read and agreed to the published version of the manuscript.

Abbreviations:

ANN: artificial neural network
CADx: computer-aided diagnosis
CADe: computer-aided detection
CNN: convolutional neural network
DM: digital mammography
DL: deep learning

DNN: deep neural network
DL-CAD: deep learning CAD system
CC: craniocaudal
MC: microcalcifications
ML: machine learning
MLO: mediolateral oblique
ROI: region of interest
US: ultrasound
MLP: Muli-layer perceptron
DBT: digital breast tomosynthesis
MIL: multiple instances learning
CRF: conditional random forest
RPN: region proposal network
GAN: generative adversarial network
IoU: intersection over union
SDAE: stacked denoising auto-encoder
CBIS: Curated Breast Imaging Subset
YOLO:You Only Look Once
ERFNet: Efficient Residual Factorized Network
CLAHE: contrast-limited adaptive histogram equalization
PCA: principal component analysis
LDA: linear discriminant analysis
GLCM: grey-level co-occurrence matrix
RF: random forest
DBT: decision boundary features
SVM: support vector machine
NN: neural network
SOM: self-organizing map
KNN: K-nearest neighbor
BDT: binary decision tree
DBN: deep belief networks
WPT: wavelet packet transform

References

1. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R. L.; Torre, L. A.; Jemal, A. Global cancer statistics. *CA Cancer J. Clin.* **2011**, *68*, 394–424, doi:10.3322/caac.20107.
2. Gao, F.; Chia, K.-S.; Ng, F.-C.; Ng, E.-H.; Machin, D. Interval cancers following breast cancer screening in Singaporean women. *Int. J. Cancer* **2002**, *101*, 475–479, doi:10.1002/ijc.10636.
3. Munir, K.; Elahi, H.; Ayub, A.; Frezza, F.; Rizzi, A. Cancer Diagnosis Using Deep Learning: A Bibliographic Review. *Cancers* **2019**, *11*, 1235, doi:10.3390/cancers11091235.

4. American Cancer Society. Breast Cancer Facts and Figures 2019; American Cancer Society: Atlanta, GA, USA, 2019.
5. Nahid, A.-A.; Kong, Y. Involvement of Machine Learning for Breast Cancer Image Classification: A Survey. *Comput. Math. Methods Med.* **2017**, *2017*, 1–29, doi:10.1155/2017/3781951.
6. Skandalakis, J.E. Embryology and anatomy of the breast. In *Breast Augmentation*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 3–24.
7. Dheeba, J.; Singh, N.A. Computer aided intelligent breast cancer detection: Second opinion for radiologists—A prospective study. In *Computational Intelligence Applications in Modeling and Control*; Springer: Cham, Switzerland, 2015; pp. 397–430.
8. Ramadan, S.Z. Methods Used in Computer-Aided Diagnosis for Breast Cancer Detection Using Mammograms: A Review. *J. Heal. Eng.* **2020**, *2020*, 1–21, doi:10.1155/2020/9162464.
9. Chan, H.-P.; Doi, K.; Vybrony, C.J.; Schmidt, R.A.; Metz, C.E.; Lam, K.L.; Ogura, T.; Wu, Y.; MacMahon, H. Improvement in Radiologists?? Detection of Clustered Microcalcifications on Mammograms. *Investig. Radiol.* **1990**, *25*, 1102–1110, doi:10.1097/00004424-199010000-00006.
10. Olsen, O.; Gøtzsche, P.C. Cochrane review on screening for breast cancer with mammography. *Lancet* **2001**, *358*, 1340–1342, doi:10.1016/s0140-6736(01)06449-2.
11. Mann, R.M.; Kuhl, C.K.; Kinkel, K.; Boetes, C. Breast MRI: guidelines from the European Society of Breast Imaging. *Eur. Radiol.* **2008**, *18*, 1307–1318, doi:10.1007/s00330-008-0863-7.
12. Jalalian, A.; Mashohor, S.B.; Mahmud, H.R.; Saripan, M.I.B.; Ramli, A.R.B.; Karasfi, B. Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review. *Clin. Imaging* **2013**, *37*, 420–426, doi:10.1016/j.clinimag.2012.09.024.
13. Sarno, A.; Mettivier, G.; Russo, P. Dedicated breast computed tomography: Basic aspects. *Med. Phys.* **2015**, *42*, 2786–2804, doi:10.1118/1.4919441.
14. Njor, S.; Nyström, L.; Moss, S.; Paci, E.; Broeders, M.; Segnan, N.; Lynge, E. Breast Cancer Mortality in Mammographic Screening in Europe: A Review of Incidence-Based Mortality Studies. *J. Med. Screen.* **2012**, *19*, 33–41, doi:10.1258/jms.2012.012080.
15. Morrell, S.; Taylor, R.; Roder, D.; Dobson, A. Mammography screening and breast cancer mortality in Australia: an aggregate cohort study. *J. Med. Screen.* **2012**, *19*, 26–34, doi:10.1258/jms.2012.011127.
16. Marmot, M.; Screening, T.I.U.P.O.B.C.; Altman, D.G.; Cameron, D.A.; Dewar, J.A.; Thompson, S.G.; Wilcox, M. The benefits and harms of breast cancer screening: an independent review. *Br. J. Cancer* **2013**, *108*, 2205–2240, doi:10.1038/bjc.2013.177.
17. Liu, C.-Y.; Hsu, C.-Y.; Chou, Y.-H.; Chen, C.-M. A multi-scale tumor detection algorithm in whole breast sonography incorporating breast anatomy and

tissue morphological information. In Proceedings of the 2014 IEEE Healthcare Innovation Conference (HIC), Seattle, WA, USA, 8–10 October 2014; pp. 193–196.
18. Pisano, E.D.; Gatsonis, C.; Hendrick, E.; Yaffe, M.; Baum, J.K.; Acharyya, S.; Conant, E.F.; Fajardo, L.L.; Bassett, L.; D'Orsi, C.; et al. Diagnostic Performance of Digital versus Film Mammography for Breast-Cancer Screening. *New Engl. J. Med.* **2005**, *353*, 1773–1783, doi:10.1056/nejmoa052911.
19. Carney, P.A.; Miglioretti, D.L.; Yankaskas, B.C.; Kerlikowske, K.; Rosenberg, R.; Rutter, C.M.; Geller, B.M.; Abraham, L.A.; Taplin, S.H.; Dignan, M.; et al. Individual and Combined Effects of Age, Breast Density, and Hormone Replacement Therapy Use on the Accuracy of Screening Mammography. *Ann. Intern. Med.* **2003**, *138*, 168–175, doi:10.7326/0003-4819-138-3-200302040-00008.
20. Woodard, D.B.; Gelfand, A.E.; Barlow, W.E.; Elmore, J.G. Performance assessment for radiologists interpreting screening mammography. *Stat. Med.* **2007**, *26*, 1532–1551, doi:10.1002/sim.2633.
21. Cole, E.; Pisano, E.D.; Kistner, E.O.; Muller, K.E.; Brown, M.E.; Feig, S.A.; Jong, R.A.; Maidment, A.D.A.; Staiger, M.J.; Kuzmiak, C.M.; et al. Diagnostic Accuracy of Digital Mammography in Patients with Dense Breasts Who Underwent Problem-solving Mammography: Effects of Image Processing and Lesion Type1. *Radiology.* **2003**, *226*, 153–160, doi:10.1148/radiol.2261012024.
22. Boyd, N.F.; Guo, H.; Martin, L.J.; Sun, L.; Stone, J.; Fishell, E.; Jong, R.A.; Hislop, G.; Chiarelli, A.; Minkin, S.; et al. Mammographic Density and the Risk and Detection of Breast Cancer. *New Engl. J. Med.* **2007**, *356*, 227–236, doi:10.1056/nejmoa062790.
23. E Bird, R.; Wallace, T.W.; Yankaskas, B.C. Analysis of cancers missed at screening mammography. *Radiol.* **1992**, *184*, 613–617, doi:10.1148/radiology.184.3.1509041.
24. Kerlikowske, K.; Carney, P.A.; Geller, B.; Mandelson, M.T.; Taplin, S.; Malvin, K.; Ernster, V.; Urban, N.; Cutter, G.; Rosenberg, R.; et al. Performance of screening mammography among women with and without a first-degree relative with breast cancer. *Ann. Intern. Med.* **2000**, *133*, 855–863, doi:10.7326/0003-4819-133-11-200012050-00009.
25. Ertosun, M.G.; Rubin, D.L. Probabilistic visual search for masses within mammography images using deep learning. In Proceedings of the 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Washington, DC, USA, 9–12 November 2015; pp. 1310–1315.
26. Nunes, F.L.S.; Schiabel, H.; Goes, C.E. Contrast Enhancement in Dense Breast Images to Aid Clustered Microcalcifications Detection. *J. Digit. Imaging* **2006**, *20*, 53–66, doi:10.1007/s10278-005-6976-5.
27. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the Inception Architecture for Computer Vision. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.

28. Dinnes, J.; Moss, S.; Melia, J.; Blanks, R.; Song, F.; Kleijnen, J. Effectiveness and cost-effectiveness of double reading of mammograms in breast cancer screening: findings of a systematic review. *Breast* **2001**, *10*, 455–463, doi:10.1054/brst.2001.0350.
29. Robinson, P.J. Radiology's Achilles' heel: Error and variation in the interpretation of the Röntgen image. *Br. J. Radiol.* 1997, 70, 1085–1098.
30. Rangayyan, R.M.; Ayres, F.J.; Desautels, J.L. A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs. *J. Frankl. Inst.* **2007**, *344*, 312–348, doi:10.1016/j.jfranklin.2006.09.003.
31. Jalalian, A.; Mashohor, S.; Mahmud, R.; Karasfi, B.; Saripan, M.I.B.; Ramli, A.R.B. Foundation and methodologies in computer-aided diagnosis systems for breast cancer detection. *EXCLI J.* 2017, *16*, 113–137.
32. Vyborny, C.J.; Giger, M.L.; Nishikawa, R.M. COMPUTER-AIDED DETECTION AND DIAGNOSIS OF BREAST CANCER. *Radiol. Clin. North. Am.* **2000**, *38*, 725–740, doi:10.1016/s0033-8389(05)70197-4.
33. Giger, M.L. Machine Learning in Medical Imaging. *J. Am. Coll. Radiol.* **2018**, *15*, 512–520, doi:10.1016/j.jacr.2017.12.028.
34. Xu, Y.; Wang, Y.; Yuan, J.; Cheng, Q.; Wang, X.; Carson, P.L. Medical breast ultrasound image segmentation by machine learning. *Ultrasonics.* **2019**, *91*, 1–9, doi:10.1016/j.ultras.2018.07.006.
35. Shan, J.; Alam, S.K.; Garra, B.; Zhang, Y.; Ahmed, T. Computer-Aided Diagnosis for Breast Ultrasound Using Computerized BI-RADS Features and Machine Learning Methods. *Ultrasound Med. Biol.* **2016**, *42*, 980–988, doi:10.1016/j.ultrasmedbio.2015.11.016.
36. Zhang, Q.; Xiao, Y.; Dai, W.; Suo, J.; Wang, C.; Shi, J.; Zheng, H. Deep learning based classification of breast tumors with shear-wave elastography. *Ultrason.* **2016**, *72*, 150–157, doi:10.1016/j.ultras.2016.08.004.
37. Cheng, J.-Z.; Ni, D.; Chou, Y.-H.; Qin, J.; Tiu, C.-M.; Chang, Y.-C.; Huang, C.-S.; Shen, D.; Chen, C.-M. Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans. *Sci. Rep.* **2016**, *6*, 24454, doi:10.1038/srep24454.
38. Shin, S.Y.; Lee, S.; Yun, I.D.; Kim, S.M.; Lee, K.M. Joint Weakly and Semi-Supervised Deep Learning for Localization and Classification of Masses in Breast Ultrasound Images. *IEEE Trans. Med. Imaging* **2018**, *38*, 762–774, doi:10.1109/tmi.2018.2872031.
39. Wang, J.; Ding, H.; Bidgoli, F.A.; Zhou, B.; Iribarren, C.; Molloi, S.; Baldi, P. Detecting Cardiovascular Disease from Mammograms With Deep Learning. *IEEE Trans. Med. Imaging* **2017**, *36*, 1172–1181, doi:10.1109/tmi.2017.2655486.
40. Dhungel, N.; Carneiro, G.; Bradley, A.P. Fully automated classification of mammograms using deep residual neural networks. In Proceedings of the 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), Melbourne, Australia, 18–21 April 2017; pp. 310–314.

41. Kooi, T.; Litjens, G.J.S.; Van Ginneken, B.; Gubern-Mérida, A.; Sánchez, C.I.; Mann, R.; Heeten, A.D.; Karssemeijer, N. Large scale deep learning for computer aided detection of mammographic lesions. *Med. Image Anal.* **2017**, *35*, 303–312, doi:10.1016/j.media.2016.07.007.
42. Debelee, T.G.; Schwenker, F.; Ibenthal, A.; Yohannes, D. Survey of deep learning in breast cancer image analysis. *Evol. Syst.* **2020**, *11*, 143–163, doi:10.1007/s12530-019-09297-2.
43. Keen, J.D.; Keen, J.M.; Keen, J.E. Utilization of computer-aided detection for digital screening mammography in the United States, 2008 to 2016. *J. Am. Coll. Radiol.* **2018**, *15*, 44–48.
44. Henriksen, E.L.; Carlsen, J.F.; Vejborg, I.; Nielsen, M.B.; A Lauridsen, C. The efficacy of using computer-aided detection (CAD) for detection of breast cancer in mammography screening: a systematic review. *Acta Radiol.* **2018**, *60*, 13–18, doi:10.1177/0284185118770917.
45. Gao, Y.; Geras, K.J.; Lewin, A.A.; Moy, L. New Frontiers: An Update on Computer-Aided Diagnosis for Breast Imaging in the Age of Artificial Intelligence. *Am. J. Roentgenol.* **2019**, *212*, 300–307, doi:10.2214/ajr.18.20392.
46. Pacilè, S., Lopez, J., Chone, P., Bertinotti, T., Grouin, J. M., & Fillard, P. (2020). Improving Breast Cancer Detection Accuracy of Mammography with the Concurrent Use of an Artificial Intelligence Tool. *Radiology: Artificial Intelligence*, *2*(6), e190208.
47. Huynh, B.Q.; Li, H.; Giger, M.L. Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. *J. Med. Imaging* **2016**, *3*, 034501, doi:10.1117/1.jmi.3.3.034501.
48. Yap, M.H.; Pons, G.; Marti, J.; Ganau, S.; Sentis, M.; Zwiggelaar, R.; Davison, A.K.; Marti, R. Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks. *IEEE J. Biomed. Heal. Informatics* **2018**, *22*, 1218–1226, doi:10.1109/jbhi.2017.2731873.
49. Moon, W.K.; Lee, Y.-W.; Ke, H.-H.; Lee, S.H.; Huang, C.-S.; Chang, R.-F. Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks. *Comput. Methods Programs Biomed.* **2020**, *190*, 105361, doi:10.1016/j.cmpb.2020.105361.
50. Yap, M.H.; Goyal, M.; Osman, F.; Ahmad, E.; Martí, R.; Denton, E.; Juette, A.; Zwiggelaar, R. End-to-end breast ultrasound lesions recognition with a deep learning approach. In *Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging*; International Society for Optics and Photonics: Bellingham, WA, USA, 2018; Volume 10578, p. 1057819.
51. Xu, B.; Wang, N.; Chen, T.; Li, M. Empirical evaluation of rectified activations in convolutional network. *arXiv* **2015**, arXiv:1505.00853.
52. Swiderski, B.; Kurek, J.; Osowski, S.; Kruk, M.; Barhoumi, W. Deep learning and non-negative matrix factorization in recognition of mammograms. In Proceedings of the Eighth International Conference on Graphic and Image Processing (ICGIP 2016), Tokyo, Japan, 29–31 October 2016; International

Society for Optics and Photonics: Bellingham, WA, USA, 2017; Volume 10225, p. 102250B.
53. Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444.
54. Miotto, R.; Wang, F.; Wang, S.; Jiang, X.; Dudley, J.T. Deep learning for healthcare: review, opportunities and challenges. *Briefings Bioinform.* **2018**, *19*, 1236–1246, doi:10.1093/bib/bbx044.
55. Shin, H.-C.; Roth, H.R.; Gao, M.; Lu, L.; Xu, Z.; Nogues, I.; Yao, J.; Mollura, D.; Summers, R.M. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Trans. Med. Imaging* **2016**, *35*, 1285–1298, doi:10.1109/tmi.2016.2528162.
56. Lee, J.-G.; Jun, S.; Cho, Y.-W.; Lee, H.; Kim, G.B.; Seo, J.B.; Kim, N. Deep Learning in Medical Imaging: General Overview. *Korean J. Radiol.* **2017**, *18*, 570–584, doi:10.3348/kjr.2017.18.4.570.
57. Suzuki, K. Overview of deep learning in medical imaging. *Radiol. Phys. Technol.* **2017**, *10*, 257–273, doi:10.1007/s12194-017-0406-5.
58. Moradmand, H., Setayeshi, S., Karimian, A. R., Sirous, M., & Akbari, M. E. (2012). Comparing the performance of image enhancement methods to detect microcalcification clusters in digital mammography. Iranian journal of cancer prevention, 5(2), 61.
59. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Int. J. Surg.* **2010**, *8*, 336–341, doi:10.1016/j.ijsu.2010.02.007.
60. Khan, K.S.; Kunz, R.; Kleijnen, J.; Antes, G. Five steps to conducting a systematic review. *J. R. Soc. Med*. **2003**, *96*, 118–121.
61. Suckling J, P. (1994). The mammographic image analysis society digital mammogram database. *Digit. Mammo*. **1994**, *1069*, 375–386.
62. Han, S.; Kang, H.-K.; Jeong, J.-Y.; Park, M.-H.; Kim, W.; Bang, W.-C.; Seong, Y.-K. A deep learning framework for supporting the classification of breast lesions in ultrasound images. *Phys. Med. Biol.* **2017**, *62*, 7714–7728, doi:10.1088/1361-6560/aa82ec.
63. De Oliveira, J. E., Deserno, T. M., & Araújo, A. D. A. (2008, October). Breast Lesions Classification applied to a reference database. In 2nd International Conference, Hammanet, Tunisia 29-31 October, 2008: E-Medical Systems (pp. 29–31).
64. Moreira, I.C.; Amaral, I.; Domingues, I.; Cardoso, A.; Cardoso, M.J.; Cardoso, J.S. Inbreast: Toward a full-field digital mammographic database. *Acad. Radiol.* **2012**, *19*, 236–248.
65. Abdelhafiz, D.; Yang, C.; Ammar, R.; Nabavi, S. Deep convolutional neural networks for mammography: advances, challenges and applications. *BMC Bioinform.* **2019**, *20*, 281, doi:10.1186/s12859-019-2823-4.
66. Byra, M.; Jarosik, P.; Szubert, A.; Galperin, M.; Ojeda-Fournier, H.; Olson, L.; O'Boyle, M.; Comstock, C.; Andre, M. Breast mass segmentation in ultrasound

with selective kernel U-Net convolutional neural network. *Biomed. Signal. Process. Control.* **2020**, *61*, 102027, doi:10.1016/j.bspc.2020.102027.
67. Jiao, Z.; Gao, X.; Wang, Y.; Li, J. A deep feature based framework for breast masses classification. *Neurocomputing* **2016**, *197*, 221–231, doi:10.1016/j.neucom.2016.02.060.
68. Arevalo, J.; González, F.A.; Ramos-Pollán, R.; Oliveira, J.L.; Lopez, M.A.G. Representation learning for mammography mass lesion classification with convolutional neural networks. *Comput. Methods Programs Biomed.* **2016**, *127*, 248–257, doi:10.1016/j.cmpb.2015.12.014.
69. Peng, W.; Mayorga, R.; Hussein, E. An automated confirmatory system for analysis of mammograms. *Comput. Methods Programs Biomed.* **2016**, *125*, 134–144, doi:10.1016/j.cmpb.2015.09.019.
70. Dhabyani, W.; Gomaa, M.; Khaled, H.; Fahmy, A. Dataset of breast ultrasound images. *Data Brief.* **2020**, *28*, 104863, doi:10.1016/j.dib.2019.104863.
71. Tian, J.-W.; Wang, Y.; Huang, J.-H.; Ning, C.-P.; Wang, H.-M.; Liu, Y.; Tang, X.-L. The Digital Database for Breast Ultrasound Image. In Proceedings of the Proceedings of the 11th Joint Conference on Information Sciences (JCIS); Atlantis Press, 2008.
72. Piotrzkowska-Wróblewska, H.; Dobruch-Sobczak, K.; Byra, M.; Nowicki, A. Open access database of raw ultrasonic signals acquired from malignant and benign breast lesions. *Med. Phys.* **2017**, *44*, 6105–6109, doi:10.1002/mp.12538.
73. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition. Miami, FL, USA, 20–25 June 2009, pp. 248-255.
74. Fujita, H. AI-based computer-aided diagnosis (AI-CAD): the latest review to read first. *Radiol. Phys. Technol.* **2020**, *13*, 6–19, doi:10.1007/s12194-019-00552-4.
75. Sengupta, S.; Singh, A.; Leopold, H.A.; Gulati, T.; Lakshminarayanan, V. Ophthalmic diagnosis using deep learning with fundus images – A critical review. *Artif. Intell. Med.* **2020**, *102*, 101758, doi:10.1016/j.artmed.2019.101758.
76. Shih, F.Y. *Image Processing and Pattern Recognition*: Fundamentals and Techniques; John Wiley & Sons: Hoboken, NJ, USA, 2010;.
77. Biltawi, M., Al-Najdawi, N. I. J. A. D., & Tedmori, S. A. R. A. Mammogram enhancement and segmentation methods: classification, analysis, and evaluation. In Proceeding of The 13th international Arab conference on information technology. Jordan, 10–13 December 2012
78. Dabass, J.; Arora, S.; Vig, R.; Hanmandlu, M. Segmentation techniques for breast cancer imaging modalities—A review. In Proceedings of the 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 10–11 January 2019; pp. 658–663.
79. Ganesan, K.; Acharya, U.R.; Chua, K.C.; Min, L.C.; Abraham, K.T. Pectoral muscle segmentation: A review. *Comput. Methods Programs Biomed.* **2013**, *110*, 48–57, doi:10.1016/j.cmpb.2012.10.020.

80. Huang, Q.; Luo, Y.; Zhang, Q. Breast ultrasound image segmentation: a survey. *Int. J. Comput. Assist. Radiol. Surg.* **2017**, *12*, 493–507, doi:10.1007/s11548-016-1513-1.
81. Noble, J.A.; Boukerroui, D. Ultrasound image segmentation: a survey. *IEEE Trans. Med. Imaging* **2006**, *25*, 987–1010, doi:10.1109/tmi.2006.877092.
82. Kallergi, M.; Woods, K.; Clarke, L.P.; Qian, W.; Clark, R.A. Image segmentation in digital mammography: Comparison of local thresholding and region growing algorithms. *Comput. Med. Imaging Graph.* **1992**, *16*, 323–331, doi:10.1016/0895-6111(92)90145-y.
83. Tsantis, S.; Dimitropoulos, N.; Cavouras, D.; Nikiforidis, G. A hybrid multi-scale model for thyroid nodule boundary detection on ultrasound images. *Comput. Methods Programs Biomed.* **2006**, *84*, 86–98, doi:10.1016/j.cmpb.2006.09.006.
84. Ilesanmi, A.E.; Idowu, O.P.; Makhanov, S.S. Multiscale superpixel method for segmentation of breast ultrasound. *Comput. Biol. Med.* **2020**, *125*, 103879, doi:10.1016/j.compbiomed.2020.103879.
85. Chen, D.; Chang, R.-F.; Kuo, W.-J.; Chen, M.-C.; Huang, Y.-L. Diagnosis of breast tumors with sonographic texture analysis using wavelet transform and neural networks. *Ultrasound Med. Biol.* **2002**, *28*, 1301–1310, doi:10.1016/s0301-5629(02)00620-8.
86. Cheng, H.D.; Shan, J.; Ju, W.; Guo, Y.; Zhang, L. Automated breast cancer detection and classification using ultrasound images: A survey. *Pattern Recognit.* **2010**, *43*, 299–317, doi:10.1016/j.patcog.2009.05.012.
87. Hasan, H.; Tahir, N.M. Feature selection of breast cancer based on principal component analysis. In Proceedings of the 2010 6th International Colloquium on Signal Processing & Its Applications, Mallaca City, Malaysia, 21–23 May 2010; pp. 1–4.
88. Chan, H.-P.; Wei, D.; A Helvie, M.; Sahiner, B.; Adler, D.D.; Goodsitt, M.M.; Petrick, N. Computer-aided classification of mammographic masses and normal tissue: linear discriminant analysis in texture feature space. *Phys. Med. Biol.* **1995**, *40*, 857–876, doi:10.1088/0031-9155/40/5/010.
89. Maglogiannis, I.; Zafiropoulos, E.; Kyranoudis, C. Intelligent segmentation and classification of pigmented skin lesions in dermatological images. In *Hellenic Conference on Artificial Intelligence*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 214–223.
90. Jin, X.; Xu, A.; Bie, R.; Guo, P. Machine learning techniques and chi-square feature selection for cancer classification using SAGE gene expression profiles. In International Workshop on Data Mining for Biomedical Applications; Springer: Berlin/Heidelberg, Germany, 2006; pp. 106–115.
91. Verma, K.; Singh, B.K.; Tripathi, P.; Thoke, A.S. Review of feature selection algorithms for breast cancer ultrasound image. In New Trends in Intelligent Information and Database Systems; Springer: Cham, Switzerland, 2015; pp. 23-32.

92. Sikorski, J. (2004). Identification of malignant melanoma by wavelet analysis. Proceedings of Student/Faculty Research Day, CSIS, Pace University New York, NY, USA, 7 May 2004.
93. Chiem, A.; Al-Jumaily, A.; Khushaba, R.N. A novel hybrid system for skin lesion detection. In Proceedings of the 2007 3rd International Conference on Intelligent Sensors, Sensor Networks and Information, Melbourne, Australia, 3–6 December 2007; pp. 567–572.
94. Tanaka, T.; Torii, S.; Kabuta, I.; Shimizu, K.; Tanaka, M. Pattern Classification of Nevus with Texture Analysis. *IEEJ Trans. Electr. Electron. Eng.* **2007**, *3*, 143–150, doi:10.1002/tee.20246.
95. Zhou, H.; Chen, M.; Rehg, J.M. Dermoscopic interest point detector and descriptor. In Proceedings of the 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Boston, MA, USA, 28 June–1 July 2009; pp. 1318–1321.
96. Singh, B.; Jain, V.K.; Singh, S. Mammogram Mass Classification Using Support Vector Machine with Texture, Shape Features and Hierarchical Centroid Method. *J. Med. Imaging Heal. Informatics* **2014**, *4*, 687–696, doi:10.1166/jmihi.2014.1312.
97. Sonar, P.; Bhosle, U.; Choudhury, C. Mammography classification using modified hybrid SVM-KNN. In Proceedings of the 2017 International Conference on Signal Processing and Communication (ICSPC), Coimbatore, India, 28–29 July 2017; pp. 305–311.
98. Pal, N.R.; Bhowmick, B.; Patel, S.K.; Pal, S.; Das, J. A multi-stage neural network aided system for detection of microcalcifications in digitized mammograms. *Neurocomputing* **2008**, *71*, 2625–2634, doi:10.1016/j.neucom.2007.06.015.
99. Ayer, T.; Chen, Q.; Burnside, E.S. Artificial Neural Networks in Mammography Interpretation and Diagnostic Decision Making. *Comput. Math. Methods Med.* **2013**, *2013*, 1–10, doi:10.1155/2013/832509.
100. Al-Hadidi, M.R.; Alarabeyyat, A.; Alhanahnah, M. Breast cancer detection using k-nearest neighbor machine learning algorithm. In Proceedings of the 2016 9th International Conference on Developments in eSystems Engineering (DeSE), Liverpool, UK, 31 August–2 September 2016; pp. 35–39.
101. Sumbaly, R.; Vishnusri, N.; Jeyalatha, S. Diagnosis of Breast Cancer using Decision Tree Data Mining Technique. *Int. J. Comput. Appl.* **2014**, *98*, 16–24, doi:10.5120/17219-7456.
102. Landwehr, N.; Hall, M.A.; Frank, E. Logistic Model Trees. *Mach. Learn.* **2005**, *59*, 161–205, doi:10.1007/s10994-005-0466-3.
103. Abdel-Zaher, A.M.; Eldeib, A.M. Breast cancer classification using deep belief networks. *Expert Syst. Appl.* **2016**, *46*, 139–144, doi:10.1016/j.eswa.2015.10.015.
104. Nishikawa, R.M.; Giger, M.L.; Doi, K.; Metz, C.E.; Yin, F.-F.; Vyborny, C.J.; Schmidt, R.A. Effect of case selection on the performance of computer-aided detection schemes. *Med. Phys.* **1994**, *21*, 265–269, doi:10.1118/1.597287.

105. Rodríguez-Ruiz, A.; Teuwen, J.; Chung, K.; Karssemeijer, N.; Chevalier, M.; Gubern-Merida, A.; Sechopoulos, I. Pectoral muscle segmentation in breast tomosynthesis with deep learning. In *Medical Imaging 2018: Computer-Aided Diagnosis*; International Society for Optics and Photonics: Bellingham, WA, USA, 2018; Volume 10575, p. 105752J.
106. Guo, R.; Lu, G.; Qin, B.; Fei, B. Ultrasound Imaging Technologies for Breast Cancer Detection and Management: A Review. *Ultrasound Med. Biol.* **2018**, *44*, 37–70, doi:10.1016/j.ultrasmedbio.2017.09.012.
107. Cadena, L.; Castillo, D.; Zotin, A.; Diaz, P.; Cadena, F.; Cadena, G.; Jimenez, Y. (2019) Processing MRI Brain Image using OpenMP and Fast Filters for Noise Reduction. In Proceedings of the World Congress on Engineering and Computer Science 2019, San Francisco, CA, USA October 22–24, 2019.
108. Kang, C.-C.; Wang, W.-J.; Kang, C.-H. Image segmentation with complicated background by using seeded region growing. *AEU - Int. J. Electron. Commun.* **2012**, *66*, 767–771, doi:10.1016/j.aeue.2012.01.011.
109. Prabusankarlal, K.M.; Thirumoorthy, P.; Manavalan, R. Computer Aided Breast Cancer Diagnosis Techniques in Ultrasound: A Survey. *J. Med. Imaging Heal. Informatics* **2014**, *4*, 331–349, doi:10.1166/jmihi.2014.1269.
110. Abdallah, Y.M.Y.; Elgak, S.; Zain, H.; Rafiq, M.; Ebaid, E.A.; Elnaema, A.A. Breast cancer detection using image enhancement and segmentation algorithms. *Biomed. Res.* **2018**, *29*, 3732–3736, doi:10.4066/biomedicalresearch.29-18-1106.
111. K.U., S.; S., G.R. Objective Quality Assessment of Image Enhancement Methods in Digital Mammography - A Comparative Study. *Signal. Image Process. : Int. J.* **2016**, *7*, 1–13, doi:10.5121/sipij.2016.7401.
112. George, M.J.; Sankar, S.P. Efficient preprocessing filters and mass segmentation techniques for mammogram images. In Proceedings of the 2017 IEEE International Conference on Circuits and Systems (ICCS), Thiruvananthapuram, India, 20–21 December 2017; pp. 408–413.
113. Talha, M.; Sulong, G.B.; Jaffar, A. Preprocessing digital breast mammograms using adaptive weighted frost filter. *Biomed. Res.* **2016**, *27*, 1407–1412.
114. Thitivirut, M.; Leekitviwat, J.; Pathomsathit, C.; Phasukkit, P. Image Enhancement by using Triple Filter and Histogram Equalization for Organ Segmentation. In Proceedings of the 2019 12th Biomedical Engineering International Conference (BMEiCON), Ubon Ratchathani, Thailand, 19–22 November 2019; pp. 1–5.
115. Gandhi, K.R.; Karnan, M. Mammogram image enhancement and segmentation. In Proceedings of the IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Coimbatore, India, 28–29 December 2010; pp. 1–4.
116. Pizer, S.M.; Amburn, E.P.; Austin, J.D.; Cromartie, R.; Geselowitz, A.; Greer, T.; Romeny, B.M.T.H.; Zimmerman, J.B.; Zuiderveld, K.K. Adaptive histogram

equalization and its variations. *Comput. Vision, Graph. Image Process.* **1987**, *39*, 355–368, doi:10.1016/s0734-189x(87)80186-x.
117. Pisano, E.D.; Zong, S.; Hemminger, B.M.; DeLuca, M.; Johnston, R.E.; Muller, K.; Braeuning, M.P.; Pizer, S.M. Contrast Limited Adaptive Histogram Equalization image processing to improve the detection of simulated spiculations in dense mammograms. *J. Digit. Imaging* **1998**, *11*, 193–200, doi:10.1007/bf03178082.
118. Wan, J.; Yin, H.; Chong, A.-X.; Liu, Z.-H. Progressive residual networks for image super-resolution. *Appl. Intell.* **2020**, *50*, 1620–1632, doi:10.1007/s10489-019-01548-8.
119. Umehara, K.; Ota, J.; Ishida, T. Super-Resolution Imaging of Mammograms Based on the Super-Resolution Convolutional Neural Network. *Open J. Med. Imaging* **2017**, *7*, 180–195, doi:10.4236/ojmi.2017.74018.
120. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 295–307, doi:10.1109/tpami.2015.2439281.
121. Kim, J.; Lee, J.K.; Lee, K.M. Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1646–1654.
122. Jiang, Y.; Li, J. Generative Adversarial Network for Image Super-Resolution Combining Texture Loss. *Appl. Sci.* **2020**, *10*, 1729, doi:10.3390/app10051729.
123. Glasner, D.; Bagon, S.; Irani, M. Super-resolution from a single image. In Proceedings of the 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 29 September–2 October 2009; pp. 349–356.
124. Schultz, R.; Stevenson, R.; R.R., S.; R.L., S. A Bayesian approach to image expansion for improved definition. *IEEE Trans. Image Process.* **1994**, *3*, 233–242, doi:10.1109/83.287017.
125. Gribbon, K.T.; Bailey, D.G. A novel approach to real-time bilinear interpolation. In Proceedings of the DELTA 2004 Second IEEE International Workshop on Electronic Design, Test and Applications, Perth, Australia, 28–30 January 2004; pp. 126–131.
126. Zhang, L.; Wu, X. An edge-guided image interpolation algorithm via directional filtering and data fusion. *IEEE Trans. Image Process.* **2006**, *15*, 2226–2238, doi:10.1109/tip.2006.877407.
127. Sun, C.; Shrivastava, A.; Singh, S.; Gupta, A. Revisiting unreasonable effectiveness of data in deep learning era. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 843–852.
128. Shorten, C.; Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60, doi:10.1186/s40537-019-0197-0.
129. Weiss, K.; Khoshgoftaar, T.M.; Wang, D. A survey of transfer learning. *J. Big Data* **2016**, *3*, 1, doi:10.1186/s40537-016-0043-6.

130. Shao, L.; Zhu, F.; Li, X. Transfer Learning for Visual Categorization: A Survey. *IEEE Trans. Neural Networks Learn. Syst.* **2015**, *26*, 1019–1034, doi:10.1109/tnnls.2014.2330900.
131. Zhong, Z.; Zheng, L.; Kang, G.; Li, S.; Yang, Y. Random Erasing Data Augmentation. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, New York, USA, 7–12 February 2020; pp. 13001–13008.
132. Mirza, M.; Osindero, S. Conditional Generative Adversarial Nets Mehdi. *arXiv* **2014**, arxiv:1411.1784.
133. Gatys, L.; Ecker, A.; Bethge, M. A Neural Algorithm of Artistic Style. *J. Vis.* **2016**, *16*, 326, doi:10.1167/16.12.326.
134. Mordang, J.-J.; Janssen, T.; Bria, A.; Kooi, T.; Gubern-Mérida, A.; Karssemeijer, N. Automatic Microcalcification Detection in Multi-vendor Mammography Using Convolutional Neural Networks. In *Public-Key Cryptography – PKC 2018*; Springer Science and Business Media LLC, 2016; pp. 35–42.
135. Akselrod-Ballin, A.; Karlinsky, L.; Alpert, S.; Hasoul, S.; Ben-Ari, R.; Barkan, E. A region based convolutional network for tumor detection and classification in breast mammography. In *Deep Learning and Data Labeling for Medical Applications*, Springer: Cham, Switzerland, 2016; pp. 197–205.
136. Zhu, W.; Xie, X. Adversarial Deep Structural Networks for Mammographic Mass Segmentation 2016, 095786.
137. Lévy, D.; Jain, A. Breast mass classification from mammograms using deep convolutional neural networks. *arXiv* **2016**, arXiv:1612.00542.
138. Sert, E.; Ertekin, S.; Halici, U. Ensemble of convolutional neural networks for classification of breast microcalcification from mammograms. In Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Seogwipo, Korea, 11–15 July 2017; pp. 689–692.
139. Dhungel, N.; Carneiro, G.; Bradley, A.P. The automated learning of deep features for breast mass classification from mammograms. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2016; pp. 106–114.
140. Saeed, J.N. A SURVEY OF ULTRASONOGRAPHY BREAST CANCER IMAGE SEGMENTATION TECHNIQUES. *Acad. J. Nawroz Univ.* **2020**, *9*, 1–14, doi:10.25007/ajnu.v9n1a523.
141. Gardezi, S.J.S.; ElAzab, A.; Lei, B.; Wang, T. Breast Cancer Detection and Diagnosis Using Mammographic Data: Systematic Review. *J. Med. Internet Res.* **2019**, *21*, e14464, doi:10.2196/14464.
142. Gomez, W.; Rodriguez, A.; Pereira, W.C.A.; Infantosi, A.F.C. Feature selection and classifier performance in computer-aided diagnosis for breast ultrasound. In Proceedings of the 2013 10th International Conference and Expo on Emerging Technologies for a Smarter World (CEWIT), Melville, NY, USA 21–22 October 2013; pp. 1–5.

143. Dong, C.; Loy, C.C.; He, K.; Tang, X. Learning a deep convolutional network for image super-resolution. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2014; pp. 184–199.
144. Lotter, W.; Sorensen, G.; Cox, D. A multi-scale CNN and curriculum learning strategy for mammogram classification. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Springer: Cham, Switzerland, 2017; pp. 169–177.
145. Al-Masni, M.A.; Al-Antari, M.A.; Park, J.-M.; Gi, G.; Kim, T.-S.; Rivera, P.; Valarezo, E.; Choi, M.-T.; Han, S.-M. Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system. *Comput. Methods Programs Biomed.* **2018**, *157*, 85–94, doi:10.1016/j.cmpb.2018.01.017.
146. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495, doi:10.1109/tpami.2016.2644615.
147. Almajalid, R.; Shan, J.; Du, Y.; Zhang, M. Development of a deep-learning-based method for breast ultrasound image segmentation. In Proceedings of the 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 17–20 December 2018; pp. 1103–108.
148. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Proceedings of the Lecture Notes in Computer Science*; (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Berlin/Heidelberg, Germany, 2015.
149. Kim, T.; Cha, M.; Kim, H.; Lee, J.K.; Kim, J. Learning to discover cross-domain relations with generative adversarial networks. *arXiv* **2017**, arXiv:1703.05192.
150. Romera, E.; Alvarez, J.M.; Bergasa, L.M.; Arroyo, R. ERFNet: Efficient Residual Factorized ConvNet for Real-Time Semantic Segmentation. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 263–272, doi:10.1109/tits.2017.2750080.
151. Lee, C.; Landgrebe, D. Feature extraction based on decision boundaries. *IEEE Trans. Pattern Anal. Mach. Intell.* **1993**, *15*, 388–400, doi:10.1109/34.206958.
152. Cao, Z.; Duan, L.; Yang, G.; Yue, T.; Chen, Q. An experimental study on breast lesion detection and classification from ultrasound images using deep learning architectures. *BMC Med. Imaging* **2019**, *19*, 1–9, doi:10.1186/s12880-019-0349-x.
153. Chiao, J.-Y.; Chen, K.-Y.; Liao, K.Y.-K.; Hsieh, P.-H.; Zhang, G.; Huang, T.-C. Detection and classification the breast tumors using mask R-CNN on sonograms. *Med.* **2019**, *98*, e15200, doi:10.1097/md.0000000000015200.
154. Dhungel, N.; Carneiro, G.; Bradley, A.P. Automated Mass Detection in Mammograms Using Cascaded Deep Learning and Random Forests. *2015 International Conference on Digital Image Computing: Techniques and Applications (DICTA)* 2015, 1–8.

155. Hu, F.; Xia, G.-S.; Hu, J.; Zhang, L. Transferring Deep Convolutional Neural Networks for the Scene Classification of High-Resolution Remote Sensing Imagery. *Remote. Sens.* **2015**, *7*, 14680–14707, doi:10.3390/rs71114680.
156. Zhu, W.; Lou, Q.; Vang, Y.S.; Xie, X. Deep multi-instance networks with sparse label assignment for whole mammogram classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2017; pp. 603–611.
157. Kooi, T.; Gubern-Merida, A.; Mordang, J.-J.; Mann, R.; Pijnappel, R.; Schuur, K.; Heeten, A.D.; Karssemeijer, N. A comparison between a deep convolutional neural network and radiologists for classifying regions of interest in mammography. In *International Workshop on Breast Imaging*; Springer: Cham, Switzerland, 2016; pp. 51–56.
158. Kooi, T.; Van Ginneken, B.; Karssemeijer, N.; Heeten, A.D. Discriminating solitary cysts from soft tissue lesions in mammography using a pretrained deep convolutional neural network. *Med. Phys.* **2017**, *44*, 1017–1027, doi:10.1002/mp.12110.
159. Chan, H.-P.; Lo, S.-C.B.; Sahiner, B.; Lam, K.L.; Helvie, M.A. Computer-aided detection of mammographic microcalcifications: Pattern recognition with an artificial neural network. *Med. Phys.* **1995**, *22*, 1555–1567, doi:10.1118/1.597428.
160. Valvano, G.; Della Latta, D.; Martini, N.; Santini, G.; Gori, A.; Iacconi, C.; Ripoli, A.; Landini, L.; Chiappino, D.; Iacconi, C. Evaluation of a Deep Convolutional Neural Network method for the segmentation of breast microcalcifications in Mammography Imaging. In *Precision Medicine Powered by pHealth and Connected Health*; Springer Science and Business Media LLC, 2017; Vol. 65, pp. 438–441.
161. Carneiro, G.; Nascimento, J.; Bradley, A.P. Unregistered multiview mammogram analysis with pre-trained deep learning models. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Cham, Switzerland, 2015; pp. 652–660.
162. Carneiro, G.; Nascimento, J.; Bradley, A.P. Automated Analysis of Unregistered Multi-View Mammograms With Deep Learning. *IEEE Trans. Med. Imaging* **2017**, *36*, 2355–2365, doi:10.1109/tmi.2017.2751523.
163. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324.
164. Huynh, B.; Drukker, K.; Giger, M. MO-DE-207B-06: Computer-Aided Diagnosis of Breast Ultrasound Images Using Transfer Learning From Deep Convolutional Neural Networks. *Med. Phys.* **2016**, *43*, 3705, doi:10.1118/1.4957255.
165. Singh, V.K.; Rashwan, H.A.; Abdel-Nasser, M.; Sarker, M.; Kamal, M.; Akram, F.; Pandey, N.; Romani, S.; Puig, D. An Efficient Solution for Breast Tumor Segmentation and Classification in Ultrasound Images using Deep Adversarial Learning. *arXiv* **2019**, arXiv:1907.00887.

166. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
167. Liu, Y.; Yao, X. Ensemble learning via negative correlation. *Neural Networks* **1999**, *12*, 1399–1404, doi:10.1016/s0893-6080(99)00073-8.
168. Ragab, D.A.; Sharkas, M.; Marshall, S.; Ren, J. Breast cancer detection using deep convolutional neural networks and support vector machines. *PeerJ* **2019**, *7*, e6201, doi:10.7717/peerj.6201.
169. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
170. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
171. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv* **2015** arXiv:1512.03385.
172. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; Volume 2017, pp. 2261–2269.
173. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252, doi:10.1007/s11263-015-0816-y.
174. Samala, R.K.; Chan, H.-P.; Hadjiiski, L.; Helvie, M.A.; Wei, J.; Cha, K. Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography. *Med. Phys.* **2016**, *43*, 6654–6666, doi:10.1118/1.4967345.
175. Das, K.; Conjeti, S.; Roy, A.G.; Chatterjee, J.; Sheet, D. Multiple instance learning of deep convolutional neural networks for breast histopathology whole slide classification. In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; pp. 578–581.
176. Geras, K.J.; Wolfson, S.; Shen, Y.; Wu, N.; Kim, S.; Kim, E.; Heacock, L.; Parikh, U.; Moy, L.; Cho, K. High-resolution breast cancer screening with multi-view deep convolutional neural networks. *arXiv* **2017**, arXiv:1703.07047.
177. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning Internal Representations by Error Propagation. *Learning Internal Representations by Error Propagation* 1985.
178. Dhungel, N.; Carneiro, G.; Bradley, A.P. A deep learning approach for the analysis of masses in mammograms with minimal user intervention. *Med. Image Anal.* **2017**, *37*, 114–128, doi:10.1016/j.media.2017.01.009.

179. Rodrigues, P. S. (2017). Breast Ultrasound Image. Mendeley Data, 1. http://dx.doi.org/10.17632/wmy84gzngw.1
180. Duggento, A.; Aiello, M.; Cavaliere, C.; Cascella, G.L.; Cascella, D.; Conte, G.; Guerrisi, M.; Toschi, N. An Ad Hoc Random Initialization Deep Neural Network Architecture for Discriminating Malignant Breast Cancer Lesions in Mammographic Images. *Contrast Media Mol. Imaging* **2019**, *2019*, 5982834–9, doi:10.1155/2019/5982834.
181. Chougrad, H.; Zouaki, H.; Alheyane, O. Deep Convolutional Neural Networks for breast cancer screening. *Comput. Methods Programs Biomed.* **2018**, *157*, 19–30, doi:10.1016/j.cmpb.2018.01.011.
182. Byra, M.; Sznajder, T.; Korzinek, D.; Piotrzkowska-Wroblewska, H.; Dobruch-Sobczak, K.; Nowicki, A.; Marasek, K. Impact of Ultrasound Image Reconstruction Method on Breast Lesion Classification with Deep Learning. *Public-Key Cryptography – PKC 2018* 2019, 41–52.
183. Dheeba, J.; Singh, N.A.; Selvi, S.T. Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach. *J. Biomed. Informatics* **2014**, *49*, 45–52, doi:10.1016/j.jbi.2014.01.010.
184. Trivizakis, E.; Ioannidis, G.S.; Melissianos, V.D.; Papadakis, G.Z.; Tsatsakis, A.; Spandidos, D.A.; Marias, K. A novel deep learning architecture outperforming 'off-the-shelf' transfer learning and feature-based methods in the automated assessment of mammographic breast density. *Oncol. Rep.* **2019**, *42*, 2009–2015, doi:10.3892/or.2019.7312.
185. Samala, R.K.; Chan, H.-P.; Hadjiiski, L.M.; Helvie, M.A.; Richter, C.; Cha, K.H. Cross-domain and multi-task transfer learning of deep convolutional neural network for breast cancer diagnosis in digital breast tomosynthesis. In *Medical Imaging 2018: Computer-Aided Diagnosis*; International Society for Optics and Photonics: Bellingham, WA, USA, 2018; Volume 10575, p. 105750Q.
186. Jadoon, M.M.; Zhang, Q.; Haq, I.U.; Butt, S.; Jadoon, A. Three-Class Mammogram Classification Based on Descriptive CNN Features. *BioMed Res. Int.* **2017**, *2017*, 1–11, doi:10.1155/2017/3640901.
187. Ahmed, A.H.; Salem, M.A.-M. Mammogram-Based cancer detection using deep convolutional neural networks. In Proceedings of the 2018 13th International Conference on Computer Engineering and Systems (ICCES), Cairo, Egypt, 18–19 December 2018; pp. 694–699.
188. Gu, P.; Lee, W.-M.; Roubidoux, M.A.; Yuan, J.; Wang, X.; Carson, P.L. Automated 3D ultrasound image segmentation to aid breast cancer image interpretation. *Ultrason.* **2016**, *65*, 51–58, doi:10.1016/j.ultras.2015.10.023.
189. Singh, B.K.; Verma, K.; Thoke, A.S. Fuzzy cluster based neural network classifier for classifying breast tumors in ultrasound images. *Expert Syst. Appl.* **2016**, *66*, 114–123, doi:10.1016/j.eswa.2016.09.006.
190. Shi, J.; Zhou, S.; Liu, X.; Zhang, Q.; Lu, M.; Wang, T. Stacked deep polynomial network based representation learning for tumor classification with small

ultrasound image dataset. *Neurocomputing* **2016**, *194*, 87–94, doi:10.1016/j.neucom.2016.01.074.
191. Liu, J.; Zarshenas, A.; Wei, Z.; Yang, L.; Fajardo, L.; Suzuki, K.; Qadir, A. Radiation dose reduction in digital breast tomosynthesis (DBT) by means of deep-learning-based supervised image processing. In *Medical Imaging 2018: Image Processing*; International Society for Optics and Photonics: Bellingham, WA, USA, 2018; Volume 10574, p. 105740F.
192. Zou, L.; Yu, S.; Meng, T.; Zhang, Z.; Liang, X.; Xie, Y. A Technical Review of Convolutional Neural Network-Based Mammographic Breast Cancer Diagnosis. *Comput. Math. Methods Med.* **2019**, *2019*, 1–16, doi:10.1155/2019/6509357.

Chapter 3

BREAST MASS REGIONS CLASSIFICATION FORM MAMMOGRAMS USING CONVOLUTIONAL NEURAL NETWORKS AND TRANSFER LEARNING

Yuliana Jiménez-Gaona, María José Rodríguez-Álvarez, Diana Carrión-Figueroa, Darwin Castillo-Malla, and Vasudevan lakshminarayanan

Journal of Modern Optics, Volume 70, 2023 - Issue 10, Pages 645-660, https://doi.org/10.1080/09500340.2024.2313724

Abstract

Purpose: This study introduces a novel approach aimed at enhancing the quality of digital mammography images through pre-processing techniques, to improve breast cancer detection accuracy. **Aim:** The primary objective is to enhance image resolution, thus leading to more precise breast tissue segmentation and subsequent classification utilizing convolutional neural networks (CNNs). **Methodology:** Three recognized public mammography databases: CBIS-DDSM, Mini-MIAS, and Inbreast were used as pre-processing data. **Results:** Our statistical findings revealed that the EDSR method (PSNR = 39.05 dB/ SSIM = 0.90) consistently outperformed the visual quality of images when compared to SR-RDN (PSNR = 32.68 dB/SSIM = 0.82). Similarly, UNet demonstrated superior performance over SegNet, boasting an average Intersection over Union (IoU) of 0.862, an average Dice coefficient of 0.991, and an accuracy rate of 0.947 in Region of Interest (RoI) segmentation results. **Conclusion:** the ResNet model contributed to enhanced accuracy compared to conventional machine learning algorithms. However, it did not surpass state-of-the-art deep CNN-based classifiers, achieving an accuracy rate of 75%.

Key words: Breast cancer, classification, convolutional neural network, mammography, segmentation, super resolution, image processing.

1.Introduction

The early detection of breast lesions remains a significant challenge in the field of medical research [1]. Various screening methods [2,3] and less invasive approaches to breast cancer detection [4-8], including x-ray radiographic techniques, have been developed to address this issue. However, digital mammography (DM) stands out as a superior diagnostic modality for the early detection of breast lesions [9]. DM offers precise control and data acquisition while minimizing radiation exposure to patients, making it a critical tool in the fight against breast cancer [10].

Deep learning-based computer-assisted diagnostic (CAD) systems have emerged as a promising technology for medical image processing [11-13], playing a significant role in aiding radiologists in both screening and acting as a second reader to improve diagnostic accuracy. One significant challenge with deep learning (DL) training models is their reliance on extensive datasets for training. However, the limited dataset sizes are often due to privacy and data protection concerns, among other reasons [14-16].

Another critical aspect is that the accuracy of lesion detection heavily depends on image quality. DM images frequently exhibit various types of noise, including Salt and Pepper, Gaussian, Speckle, and Poisson noise [17]. These issues often stem from factors like image transfer, blurring, compression, or general image degradation, which lead to the production of low-resolution (LR) images. As a result, image super-resolution becomes a pivotal element in computer vision.

Multiple super-resolution techniques have been proposed to enhance the quality of medical images, thus improving the accuracy of segmentation and classification processes. These processes are crucial in the accurate diagnosis of cancer. Convolutional neural networks (CNNs) have been adapted for enhancing image resolution, segmentation, and classification tasks [17-22]. These techniques have consistently demonstrated exceptional performance in image reconstruction, employing single image super-resolution (SISR) [23,24] and Multi-Image Super-Resolution (MISR) algorithms, with SISR being widely adopted due to its remarkable efficiency [25,26].

1.1 Related work-state of the art

Data Augmentation

Original image data augmentation based on basic transformations including: spatial translation, rotation, horizontal flipping, random cropping [11] and oversampling. Other augmentation methods are also available and include [27]: geometric transformations, colour space transformations, kernel filters, image blending, random erasure, feature space augmentation, adversarial training, GAN-based augmentation [28-30], neural style transfer, and meta-learning schemes. These include spatial translation, rotation, horizontal flipping, random cropping [11] and oversampling.

Other augmentation strategies extend to more advanced methods [27], including geometric transformations, colour space transformations, kernel filters, image blending, random erasure, feature space augmentation, adversarial training, GAN-based A noteworthy observation comes from Yu et al. [15], who have substantiated that deep convolutional neural networks (CNNs) can experience substantial enhancements in performance when trained on augmented data as opposed to non-augmented data.

However, it is essential to bear in mind the insights provided by Yadav et al. [31], who have explored the impacts of both simple and complex data augmentation techniques. Their findings suggest that highly intricate transformations may not consistently outperform simpler ones, and in some cases, overly complex augmentations may introduce additional noise into the feature set, potentially detrimental to the learning process. augmentation [28-30], neural style transfer, and meta-learning schemes.

Hence, a judicious balance between diversity and noise in training data is recommended when selecting data augmentation methods.

Single Image Super Resolution

The root causes of this degradation are multifaceted, originating from the real-world clinical settings where medical imaging data is acquired. Factors like equipment conditions, patient movements, and technical constraints influence to compromise image quality. Therefore, the necessity arises to bolster the resolution of these images before they undergo the rigors of segmentation and classification.

Image SR emerges as the pivotal technique, elevating images from a low-resolution (LR) state to high resolution (HR) [32,35]. This process takes center stage in ameliorating the screening process, especially when addressing challenges like macrocalcifications or dense breast tissue. The enhancement profoundly impacts the precision of subsequent classification and segmentation processes.

SISR algorithms are categorize into four distinct types: 1) prediction models, 2) edge-based methods, 3) image statistics and 4) example-based or patch-based [18,33].

Traditional SR techniques encompass nearest-neighbor interpolation, such as bilinear interpolation, bicubic interpolation, and learning-based methods. Despite their simplicity and efficiency, they often grapple with reduced accuracy [36].

To surmount these limitations, Convolutional Neural Networks (CNNs) have been harnessed to generate HR images, using techniques like Convolutional SR-CNN [37-40] and Generative Adversarial Networks (GAN) SR-GAN [41-44]. A case in point is the model introduced by Jiang et al. [41], known as "TSGAN," which combines texture loss and encourages local information matching with a gradient penalty. This model achieved commendable metrics with an average PSNR of 27.99 dB and an SSIM of 0.778, metrics used to assess image quality and signal reconstruction.

An array of CNN and GAN-based methods has been devised, encompassing Multi-scale deep super-resolution systems (MDSR), Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN)[45], Residual Dense Block (RDB) [46], Efficient sub-pixel convolutional neural network (ESPCN) [47], Very Deep Network for SR (VDSR)[48], SR-ResNet, Sparse Coding-based Network (SCN)[49], Deep Recursive Convolutional Network (DRCN)[50], and Deep Recursive Residual Network (DRRN) [51]. Nevertheless, these methods at times fail to exploit the full spectrum of information residing in each convolutional layer and the hierarchical features crucial for reconstruction, thereby limiting their architectural optimality [45].

Current research endeavors revolve around the adaptation of additional CNN techniques for super-resolution and segmentation [38,52-58]. For instance, Tong et al. [59] devised a dense skip connection to circumvent the vanishing gradient problem plaguing very deep networks. Abbass et al. [46] and Zhang et al. [37] introduced a Residual Dense Network (RDN) grounded in the DenseNet architecture, characterized by a higher growth rate and the adept utilization of all hierarchical features from the LR image, yielding a marked enhancement in overall performance.

Worth noting is the application of SRCNN to mammography images, showcasing its superiority over conventional interpolation methods when enhancing digital mammography images of dense breasts. Dong et al. [38,39] propounded an SRCNN image reconstruction technique based on end-to-end (E2E) mapping, eclipsing traditional methods like sparse coding, kernel regression, and random forest [60,61]. Evidently, conventional SR methods, reliant on mapping functions from dictionaries, pale in comparison to modern Dense Neural Networks (DNN) employing E2E mapping approaches [62].

Additionally, the deployment of the Enhanced Deep Residual Network (EDSR), predicated on multiple ResNet architectures, has garnered attention [32]. While experimentation involving different scaling factors and optimizers yielded no single superior optimization technique [62], the caveat is that an increased number of layers may engender a surge in parameters, potentially bottlenecking image detail. Lim et al. [45] introduced single and multi-scale SR networks premised on SRResNet with a deeper residual design [48]. Residual learning techniques have manifestly led to improved performance by eliminating superfluous modules compared to antecedent methods. Their results showcase the achievement of higher PSNR values, specifically 1.57 dB for SRResNet and 2.14 dB for EDSR, respectively.

Image Quality Assessment

When it comes to evaluating the quality of super-resolution images, a range of established metrics is commonly employed. These metrics include the Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Metric (SSIM), Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Signal-to-Noise Ratio (SNR), Multi-Scale Structural Similarity (MS-SSIM), Task-Specific Similarity Assessment (TSSA), and Mean Opinion Score (MOS) [63].

However, within the domain of medical imaging super-resolution, the predominant metrics of choice are PSNR and SSIM. These metrics take center stage when quantifying the quality of the generated image in comparison to the original image [64]. Figure 1 illustrates the scale values for PSNR and SSIM.

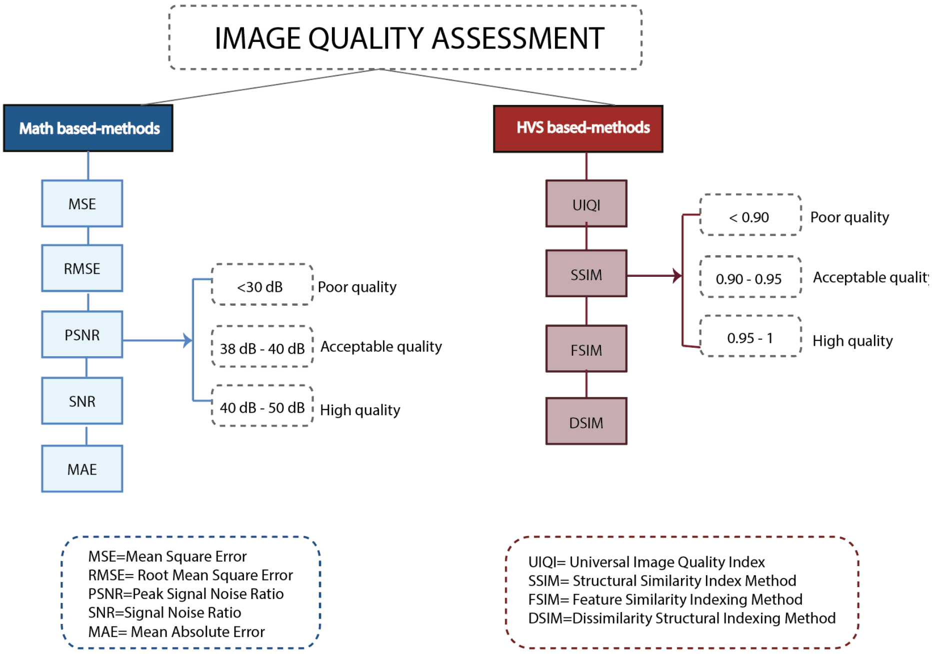


Fig. 1. Super resolution quality metrics, divided into (i) Math based methods and (ii) Human visual system-based methods [65-66].

For medical applications, the significance of PSNR and SSIM cannot be overstated. These metrics provide a valuable quantitative means of assessing the degree to which the enhanced images faithfully represent the original data. In the context of medical diagnostics, this fidelity is of utmost importance.

Segmentation

Segmentation is a critical task in medical image analysis, involving the separation of the region of interest (ROI) from the background in an image. Accurate tumour segmentation in medical images is particularly challenging due to the presence of various image artefacts and complexities. To address these challenges, researchers [67-70] have increasingly turned to deep learning (DL) methods, with a particular focus on CNNs. Various network architectures such as Visual geometric group (VGG-16) [71], ResNet [72], UNet [9,73-74], SegNet [75], ERFNet [76], have been applied to image segmentation.

Feature extraction and classification

Deep learning has established itself as a dominant approach for medical image classification. DL models have the capability to automatically extract relevant image features based on the shape of the ROI. Once these features are extracted and selected, they are utilized as input for a classifier, enabling the categorization of ROI samples into malignant or benign classes.

Several pre-trained networks have demonstrated high accuracy in medical image classification, including VGGNet, ResNet, DenseNet, and Inception [78-81]. Each of these CNN architectures has made significant contributions to the field of deep learning for computer vision tasks. The choice of architecture depends on the specific problem and involves trade-offs between factors such as model size, computational efficiency, and accuracy.

For example, DenseNet incorporates shorter skip connections between layers in a feed-forward architecture, resulting in enhanced accuracy, reduced susceptibility to overfitting, and efficient training through a cross-layer connection structure. VGGNet follows the classic CNN network structure, comprising a stack of convolutional, max-pooling, activation layers, and fully connected classification layers. ResNet offers flexibility with its fundamental shortcut connections that are task-dependent.

In contrast, Inception networks employ convolution kernels of various sizes and pooling operations within a single layer [81]. Taking these factors into account, the proposed method aims to enhance the resolution of breast images, ultimately improving the accuracy of segmentation and classification. This approach leverages convolutional neural networks and transfer learning models.

Specifically, two novel CNN-based algorithms, EDSR (Enhanced Deep Residual Network) and RDN (Residual Dense Network), are introduced to address super-resolution challenges. For image segmentation, two primary pre-trained models, UNet and SegNet, are employed, while a CNN model (ResNet50) is used for image classification. The work is organized into four main sections: "Related Work," "Materials and Methods," "Results and Discussion," and "Conclusion and Future Work."

2. Methodology

2.1 Datasets

When evaluating computer-aided diagnosis (CAD) systems for breast cancer in mammography, several challenges arise. These include the absence of a standardized evaluation dataset and the necessity to adhere to ethical, regulatory, patient privacy, and data security considerations. Consequently, many CAD systems are assessed using private datasets or unspecified subsets of public databases. For this study, we harnessed three open-access datasets:

(i) CBIS-DDSM (Curated Breast Imaging Subset –Digital Database for Screening Mammography) [82]. This database consists of 2620 cases, each containing two different views: mediolateral oblique (MLO) and craniocaudal (CC). The images are stored in DICOM format and dimensions of approximately 3784 x 5912 pixels.

(ii) mini-MIAS (Mammographic Image Analysis Society) [83,84]. This dataset offers 322 MLO mammograms from 161 patients. The images have a resolution of 1024x1024 pixels and are categorized into 208 normal, 63 benign, and 51 malignant images.

(iii) Inbreast [85], this dataset comprises a total of 115 cases, with 90 of them having two views (MLO and CC). The image matrix size varies, featuring dimensions of 3328 x 4084 or 256 x 3328 pixels. Images in this dataset are stored in the DICOM format, see table 1.

The datasets were selected based on predefined inclusion criteria that encompass demographic characteristics (age ≥ 40 years and female gender) and clinical characteristics. The selected datasets adhere to specific criteria, including normal/cancer/benign cases with verified pathology information, breast density, and abnormality descriptions. Additionally, the datasets include images with two different views, MLO and CC, while excluding normal cases.

The research methodology for breast lesion classification using deep networks is structured into five key steps, as visualized in Figure 2: 1) Manual mask RoIs extraction, 2) RoIs cropping and data augmentation. 3) Super-resolution using EDSR and SR-RDSN algorithms. 4) RoIs segmentation using UNet and SegNet algorithms. 5) RoIs classification using ResNet-50 and finally, 6) Image quality evaluation using statistical metrics.

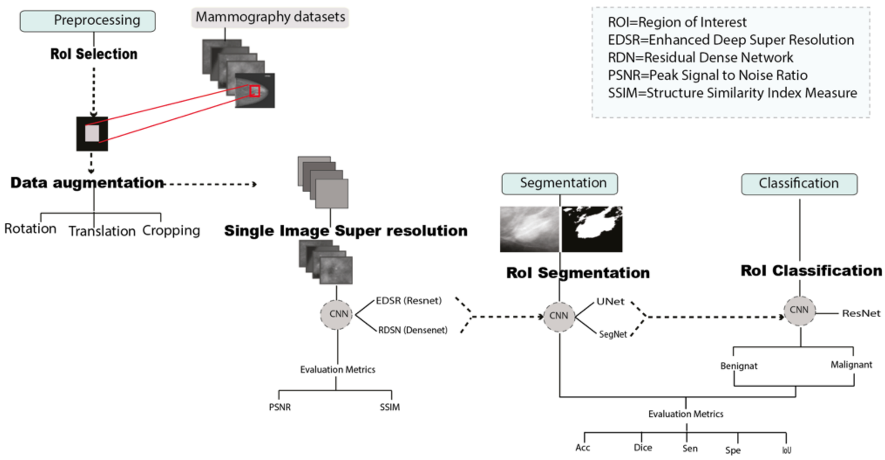


Fig. 2. The diagram describes the flowchart to breast lesion classification.

This comprehensive methodology encompasses data sources, image processing steps, and deep learning techniques employed for breast lesion classification and image quality enhancement. It provides a systematic approach to address the complexities of breast cancer diagnosis in mammography.

Manual Selection of RoI masks

The initial phase of our methodology involved the meticulous manual selection of a total of 784 binary Regions of Interest (RoIs) and their corresponding binary masks. This selection process is visually demonstrated in Figures 3a and 3b. The manual selection was carried out with the assistance of ImageJ software, which is accessible at [https://imagej.net/ij/docs/intro.html](see table 1).

Table 1. The distribution of benign and malignant cases per dataset.

Database	**Benignant**	**Malignant**	**Total**
CBIS-DDSM [82]	305	318	623
Mini-MIAS [83,84]	17	25	42
Inbreast [85]	70	49	119
Total	392	392	784

Our dataset encompasses 392 benign and 392 malignant images. To streamline the computational performance and facilitate the subsequent training of network models, all images were consistently resized to a dimension of 128 x 128 pixels.

This methodical data preparation is of paramount importance, as it forms the foundation for the subsequent phases in our methodology. It ensures that the network models can effectively learn and extract features from the RoIs, thereby contributing to the overall efficiency of the breast lesion classification system.

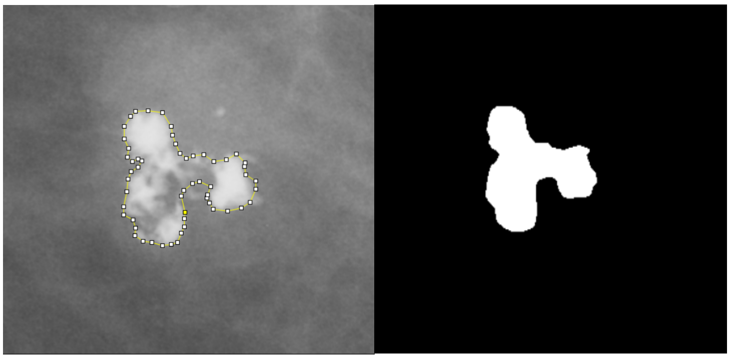


Fig. 3. (a)Manually RoI selection and (b) Binary mask.

Figures 3a and 3b provide a visual representation of the diligently selected RoIs and their corresponding binary masks, underscoring the crucial role of this data curation process in our methodology.

Data augmentation

Traditional data augmentation techniques are widely adopted in medical image analysis to mitigate the issue of limited data for training deep learning models. These techniques involve applying various transformations to the existing data, enhancing the model's generalization capabilities, and reducing the risk of overfitting.

In our study, the original dataset images underwent expansion through the application of basic geometric transformation operations, including: Blurring (Blur_1.5); Flipped (fliph, flipv); Translation (trans_20_20), rotation (rot_90, rot_180) and scaling [26] to generate new RoIs images from the selected databases. This data augmentation process yielded an additional 4704 RoI images. When combined with the original set of 784 RoIs, the dataset's size was extended to a total of 5486 RoIs.

Cross-validation

Cross-validation is a statistical technique employed to effectively partition the augmented dataset into subsets for model evaluation. In our case, the dataset was divided into three subsets: a training set, a validation set, and a testing set. The proportions for this division were determined as follows: Training Set: 4380 images (80%), Testing Set: 546 images (10%) and Validation Set: 560 images (10%).

This random yet systematic division of the dataset, as outlined in Table 2, ensures robust evaluation of the CNN models. Cross-validation is a key component in validating the performance and reliability of the breast lesion classification system.

Table 2. Data split into three sets: training, validation, and test.

Datasets	**Benignant**	**Malignant**	**Total**
Training	2190	2190	4380
Validation	280	280	560
Testing	273	273	546
Total	2743	2743	5486

Table 2 offers a concise summary of the dataset division based on the cross-validation technique, underscoring its integral role in our methodology.

Image Quality Assessment Using PSNR and SSIM

To assess the quality of our processed images, we employed well-known metrics such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Metric (SSIM). These metrics require a reference image (ground truth) for comparison. In our case, we calculated these indexes by comparing the images generated by our model with high-resolution images (ground truth) from the CBIS-DDSM breast images. Importantly, the CBIS-DDSM dataset used for this evaluation was distinct from the data used for training the model, ensuring that the assessment was conducted on unseen data. The results of PSNR and SSIM are presented in Table 4.

Image Super Resolution Using Transfer Learning

In our study, we employed transfer learning to tackle the challenge of image Super-Resolution. We utilized two distinct models for this task: EDSR (Enhanced Deep Super-Resolution) based on the ResNet architecture and SR-RDN (Super-Resolution Residual Dense Network) based on the DenseNet architecture.

The training process involved the careful selection and optimization of hyperparameters, which play a vital role in enhancing the accuracy of the Convolutional Neural Networks (CNNs) [86]. The optimization was conducted through a systematic exploration of different hyperparameter combinations, with Python's GridSearchCV (ParameterGrid) using scikit-learn. This approach enabled us to identify the combination of hyperparameters that minimized the margin of error, resulting in the most effective models.

Key hyperparameters details are described below:

EDSR hyperparameters: Scaling factor: 0.1, number of epochs: 20, Loss function:L1, number of blocks:50, optimization algorithm: ADAM/SGD/RMsProp, ResBlocks: 32, Number of filters:256, Upsampling factor x3.

RDN hyperparameters: Convolutional layer size: 3×3, Kernel size for local and global feature fusion: 1×1. Kernel size for convolutional layer: 3×3 with zero-padding, Local and global feature fusion layers: 64 filters with a Kernel size of 1x1, Upsampling factor: x4, Number of epochs: 30, Number of blocks: 16, Number of layers: 8, Batch size: 50, Patch size: 10, Activation function: ReLU.

Segmentation

UNet architecture

The UNet architecture is a fundamental component of our segmentation process. This model involves applying a series of convolutional operations to the input image, effectively compressing information, and detecting essential features. Subsequently, a new image is generated using the learned features acquired during the contraction process.

The hyperparameter details of the UNet architecture are outlined in Table 3, and you can also refer to Figure 4 for a visual representation of this architecture.

Tabla 3. Hyperparameters for U-net and SegNet training architecture.

Hiperparameter	Unet/ Segnet
Number of epochs	40
Batch size	4
Steps	125
Optimizer	Adadelta
Learning rate	0.001
Loss function	Binary-crossentropy
Activation function	ReLu, Sigmoid

Table 3 presents a comprehensive overview of the hyperparameters associated with the UNet model, while Figure 4 provides a graphical representation of the architecture's structure. This UNet architecture plays a crucial role in the segmentation of breast lesions, contributing to the accuracy of our analysis.

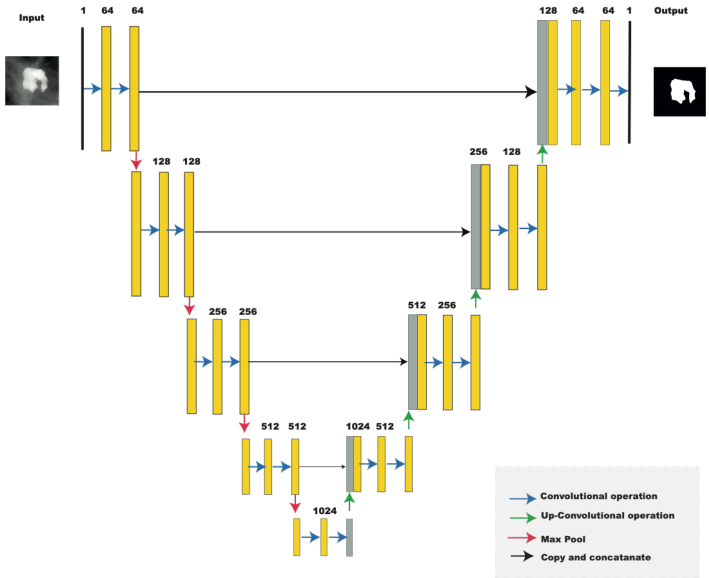


Fig. 4. The U-net architecture consists of an expansive path on the right side (upsampling) and a contracting path on the left side (downsampling). The contracting path follows the typical CNN architecture, which each yellow box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The grey boxes represent duplicated feature maps. The arrows denote the different operations.

SegNet architecture

SegNet consists of an encoder-decoder network (see Figure 5) followed by a pixel-wise classification layer. It lacks fully connected layers, relying solely on convolutional layers. The decoder upsamples its input using pooling indices, while the encoder produces a feature map, and subsequently performs convolution to densify the feature map. The final decoder output feature maps are then fed to a Softmax classifier for pixel-wise classification.

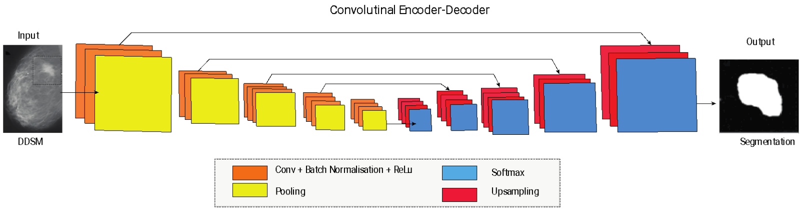


Fig. 5. SegNet architecture flowchart, adapted from Badrinarayanan, V. [75]

We used the same hyperparameters values for training U-net and SegNet networks (see Table 3).

Classification

The Deep Residual Network (ResNet) is one of the pre-trained models used in transfer learning, particularly in computer vision. It has been introduced for automatic feature extraction and classification, to address the problem of vanishing gradients providing good performance with less training time and fewer data samples compared to training a deep network from scratch.

In this research, we chose to train the ResNet-50 model, which is widely used in medical image classification. The model skips one or more layers and manages the gradient vanishing problem, in addition to its ease optimization. Model accuracy can be improved by increasing its depth. Therefore, two or three layers of the ResNet-50 model are directly connected to each layer (not to the adjacent layer), employing the ReLu non-linear activation function (Figure 6). The hyperparameters details are described below:

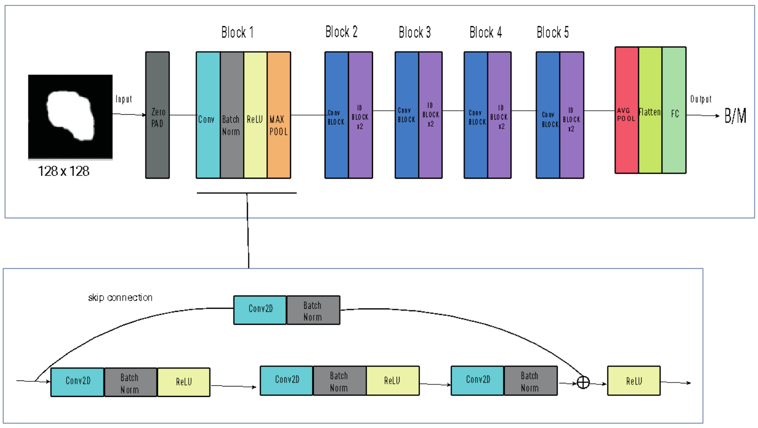


Fig. 6. The ResNet-50 architecture consists of 5 blocks, each containing 3 convolutional identity blocks and 3 convolutional blocks with skip connections.

Resnet hyperparameters: We trained the model with data augmentation and without data augmentation using standard benchmark data sets with the next hyperparameter values: Adadelta optimizer $\varepsilon=1e^{-07}$, $\delta= 0.95$, a learning rate = 0.001, batch size = 2, number of epochs = 60 with 2190 steps, categorical-cross entropy loss function over 2,190 iterations. We experimented with two activation functions: *ReLU* was used during the training of convolutional layers, while the *Sigmoid* function was used for binary class prediction.

All CNN models were training using a cloud service based on Jupyter Notebooks on Google Colab Pro GPU (model V100) and python libraries such as Keras, Matplolib and TensorFlow.

Statistical Measures

Likewise, the most frequently used statistical metrics for assessing image restoration and HR image quality are PSNR and SSIM index. A higher PSNR value indicates higher image quality, while a small value implies high numerical differences between images [87]. Typical PSNR values range from 30 dB to 50 dB, and SSIM values between -1 and 1, where 1 indicates perfect similarity, 0 indicates no similarity, and -1 indicates perfect anti-correlation (see Figure 1). For segmentation performance evaluation, we use Dice (F1 score) and Intersection over union (IoU or Jaccard index). Finally, for classification performance evaluation, we consider accuracy (Acc), precision (Prec), sensitivity (Sen), specificity (Spec) and Area Under the Curve (AUC) [11].

3. Results

This section discusses the most important results, assessment metrics, and graphs obtained from the network training.

3.1 Image Quality comparison

Table 4 presents the average PSNR and SSIM values after the image quality evaluation. These metrics offer valuable insights into the quality of the processed images when compared to high-resolution ground truth images from the CBIS-DDSM dataset. The EDSR model showed a significant improvement compared to the other model.

Tabla 4. The most relevant PSNR/SSIM values for x3 and x4 factor scaling.

ID	SR-RDN		ID	EDSR	
	PSNR (dB)	SSIM		PSNR (dB)	SSIM

DDSM_043__blur1	40.76	0.97	**DDSM_043__blur1**	46.41	0.97
DDSM_0607	40.46	0.95	**DDSM_0381__blur1**	46.23	0.97
DDSM_0120	40.33	0.95	**DDSM_0275__blur1**	45.95	0.97
DDSM_0468	40.22	0.94	**DDSM_0228__blur1**	45.76	0.97
DDSM_0466	40.20	0.96	**DDSM_0120**	45.63	0.91
DDSM_0168__blur 1	40.18	0.94	**DDSM_0467**	45.57	0.97
DDSM_0188	40.14	0.94	**DDSM_0466**	45.51	0.97
DDSM_0275__blur 1	40.14	0.97	**DDSM_0212__blur1**	45.38	0.97
DDSM_0369	40.05	0.95	**DDSM_079__blur1**	45.35	0.95
DDSM_0538	39.98	0.90	**DDSM_0368**	45.2	0.92
DDSM_0212__blur 1	35.87	0.98	**DDSM_021__blur1**	39.5	1
DDSM_0228__blur 1	37.00	0.97	**DDSM_0295__blur1**	38.17	1
DDSM_0275__blur 1	40.14	0.97	**DDSM_0106__blur1**	41.14	0.97
DDSM_043__blur1	40.76	0.97	**DDSM_0212__blur1**	45.38	0.97
DDSM_0189__blur 1	35.12	0.96	**DDSM_0228__blur1**	45.76	0.97
DDSM_0256__blur 1	37.91	0.96	**DDSM_0275__blur1**	45.95	0.97
DDSM_0381__blur 1	39.11	0.96	**DDSM_0381__blur1**	46.23	0.97
DDSM_0466	40.20	0.96	**DDSM_043__blur1**	46.41	0.97
DDSM_0467	39.95	0.96	**DDSM_0466**	45.51	0.97
DDSM_0549	38.83	0.96	**DDSM_0467**	45.57	0.97

3.2 Data dispersion

Figures 7a-d present the statistical results, where a and b are the dispersion data obtained from SR-RDN. The blue data in Figure 7a represent the PSNR metric, ranging from 30 to 40 dB, while the red points in Figure 7b represent the SSIM metric, ranging from -1 to 1. PSNR is used to measure the quality of the restored image when it is affected by noise and blur. Similarly, SSIM is defined as a function of luminance comparison.

The linear dependence factor is computed using the correlation coefficient in SSIM index and can be find broad applications in mammographic diagnosis and cancer detection fields [63].

In figure 7c, the blue data show more signal with a higher quality rate and better-quality image using EDSR algorithm. Figure 7d presents better SSIM statistical results using EDSR in comparison to 7b using SR-RDN algorithm. It indicates better luminance (ranging from 0.90 to 0.95), contrast, and structural information in restructured EDSR images.

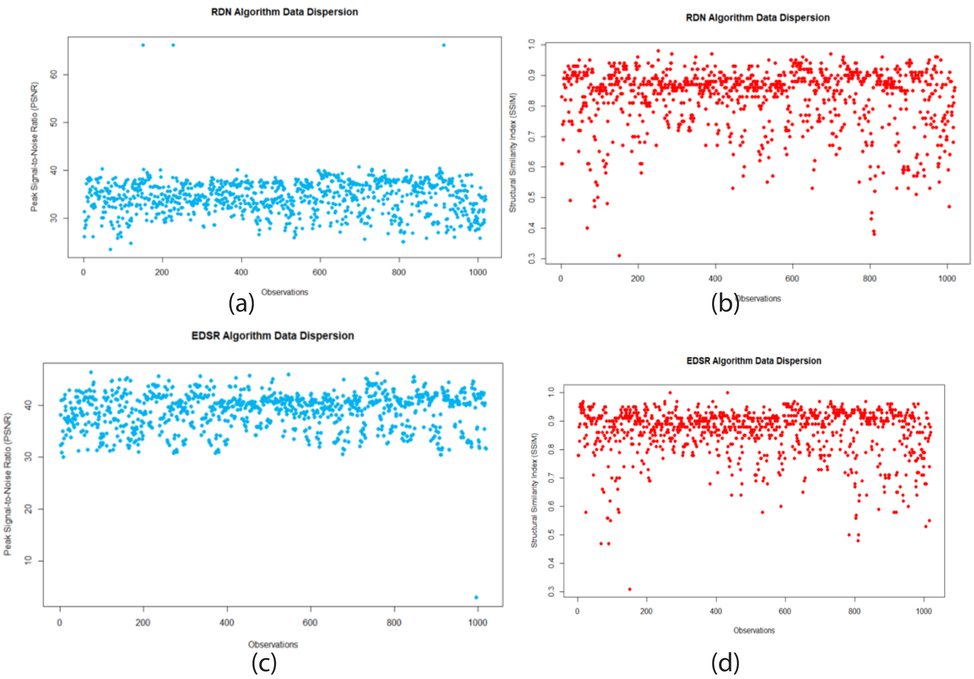


Fig. 7. Dispersion values in both super-resolution algorithms SRRDN and EDSR. (a) PSNR vs. observations in RDN algorithm. (b) SSIM vs. observations in RDN algorithm. (c) PSNR vs. observations in EDSR algorithm. (d) SSIM vs. observations in EDSR algorithm.

3.3 SegNet and Unet comparison

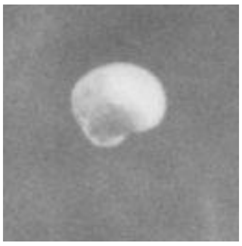
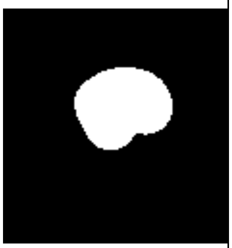
The experiments (see table 6) demonstrate that UNet achieved better RoI segmentation performance (IoU=0.862, Dice=0.991 and Acc=0.947) than the SegNet model, across all mammogram datasets used in this study.

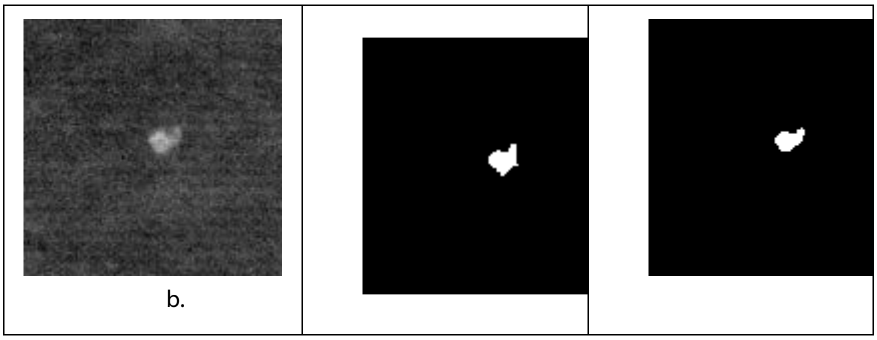
Tabla 6. Average values compared from segmentation models results.

Method	Segmentation metrics					
	Acc	**Prec**	**Sen**	**Spec**	**Dice**	**IoU**
Unet	0.947	0.930	0.925	0.956	0.991	0.862
SegNet	0.889	0.848	0.836	0.950	0.810	0.709

Table 7 displays the RoI input image with their manual segmentation and automatic segmentation using U-net and SegNet models.

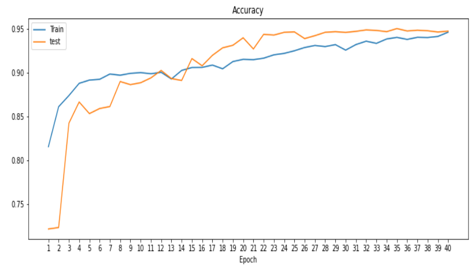
Table 7. Comparison between manual segmentation and automatic RoI segmentation (Unet/Segnet) from original RoI images: a) DDSM_0504), b) DDSM_0526.

Original RoI	**Manual Segmentation**	**Unet/Segnet**
 a		

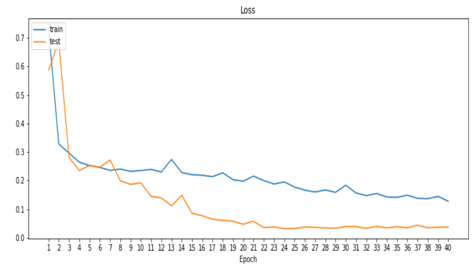


b.

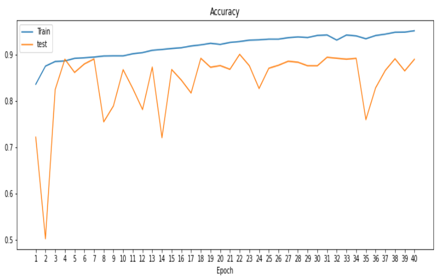
Therefore, it is important to monitor their evolution and performance of the models during training and validation. Figure 8 presents plots of the indices obtained for each epoch during the testing of different models. Figures 8a and 8b display the loss value and accuracy by each epoch in the Unet, while Figures 8c and 8d show the loss value and accuracy by each epoch in Segnet. The results indicate that the Unet model is more stable, with a consistent learning rate, and does not exhibit overfitting.



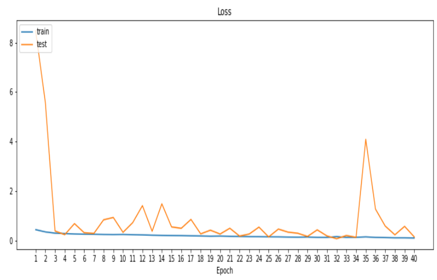
a.



b.



c.



d.

Fig. 8. Accuracy and loss values in a. Unet training dataset, b. Unet validation dataset c. Segnet training dataset and d. Segnet validation dataset.

3.4 Classification using ResNet-50

The results of the classification experiments are described in Table 8, where the values show a high accuracy using the hyperparameters with data augmentation.

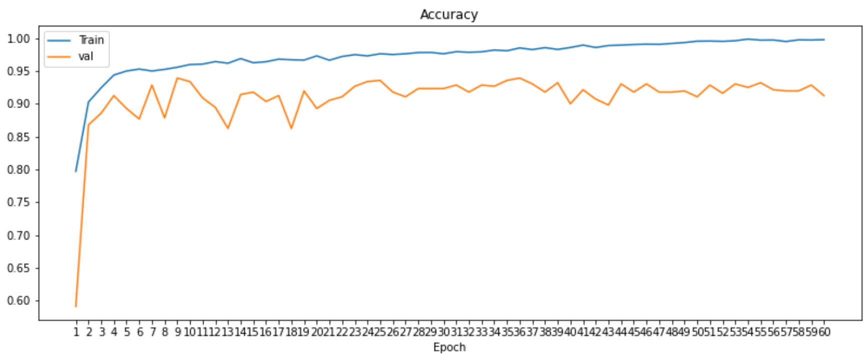
Tabla 8. ResNet-50 evaluation metrics on two datasets.

ResNet-50 model	Average classification metrics					
	Image number	**Acc**	**Prec**	**Sen**	**Spec**	**F1 score**
Data augmentation	5486	0.75	0.68	0.55	0.82	0.64
Without Data augmentation	784	0.68	0.65	0.77	0.59	0.71

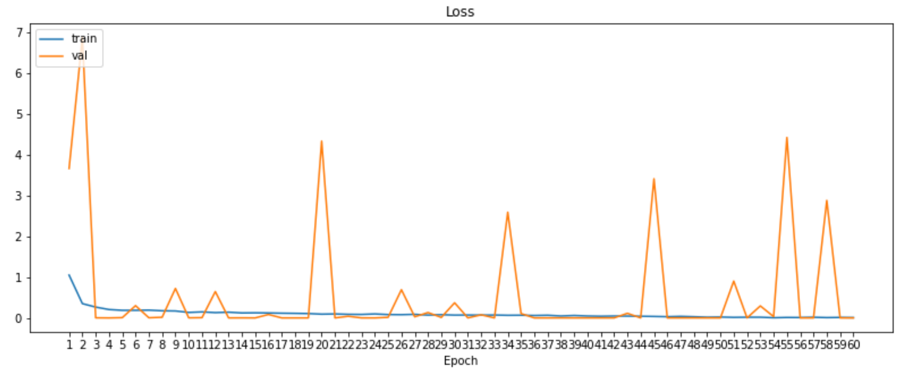
The ResNet model improved the accuracy to 75%, surpassing other traditional machine learning algorithms when using enhanced RoI data augmentation images by affine transformation.

The ResNet model utilizes a loss function to indicate it proximity to making correct predictions. The risk during the training process is the potential for the model to overfit to the training set, meaning it might learn an overly specific function that performs well on the training data but fails to generalize to unseen images.

Figures 9a and 9b illustrate that the model learns effectively from the training and validation data sets.



a.



b.

Fig. 9. (a) Line plots of model accuracy on the training (blue) and validation (orange) datasets. (b) Loss values on the training (blue) and validation (orange) datasets for each epoch.

This indicates that the model is likely to perform well on new images. However, it's important to note that while these factors can contribute to a model's ability to perform well on new images, there are no guarantees. The real-world performance of a machine learning model often requires ongoing monitoring, refinement, and adaptation to changing data distributions and conditions.

4. Discussion

DM images are often acquired at lower resolutions to minimize radiation exposure while maintaining adequate diagnostic quality. However, low-resolution images can compromise the ability to detect subtle features or abnormalities, such as microcalcifications in breast cancer lesions. The use of CNNs for SR image enhancement and segmentation can indeed be valuable and significantly improve the visibility of fine details, making it easier for radiologists to identify and classify breast cancer lesions, especially in low-resource settings where DM images may have poor resolution.

After enhancing the image resolution, the next step is to locate and delineate regions of interest, such as potential breast cancer lesions, within the mammogram. Once the image is enhanced and the lesions are segmented, CNNs can be used for the actual classification of breast cancer lesions into benign and malignant.

In our work, to test the proposed method, three public mammography databases were selected: CBIS-DDSM, Mini-MIAS and Inbreast. In the SR task the EDSR results provide high enhancement in image quality, with the PSNR and SSIM index (39.05 dB and 0.90) exceeding those of SR-RDN (32.68 dB and 0.82).

The average EDSR index values indicated that successfully reconstructed of detailed textures and edges in the RoIs and exhibited better quality output, in comparison with other results in the literature [45-47]. Lim et al. [45] proposed the EDSR CNN-based algorithm to improve super resolution in natural image (DIV2K), by removing the batch normalization layer, accelerating the training process, and achieving better performance (PSNR of 32.62 dB and SSIM of 0.8984) compared to other methods such as ESPCN (30.90 dB/-), VDSR (31.35 dB/0.8838), DRCN (31.53 dB/0.8838), SRResNet (32.05 dB/0.9019), RDN (32.61 dB/0.9003), MDSR (32.60 dB/0.8982), and DBPN (32.47 dB/0.898).

Additionally, another custom CNN (Unet) was used to perform image segmentation on the high-resolution RoIs generated by EDSR, achieving an average Intersection over Union (IoU) of 0.862, an average Dice similarity coefficient of 0.991, and an accuracy of 0.947 in segmentation results, surpassing the results of the SegNet model.

Similar research has proposed UNet as a segmentation network. Almajalid et al. [73] used UNet for breast tumour segmentation using ultrasound images and achieved a F1 score of 0.994 with the training set and 0.8252 with the testing set. Likewise, Zhou et al. [88] improved UNet by using skip connections and achieved an average IoU gain of 3.9 over the standard U-Net.

Our results align with Vianna et al. [88], who compared U-Net and SegNet for the breast lesions segmentation in ultrasonography, were U-Net demonstrated better Dice results (86.35%) compared to Segnet-Dice of 81.1%.

By analysing the segmented regions, ResNet50 provides a classification or likelihood score for the presence of breast lesions. Table 8 shows enhanced model results for breast lesion image classification using transfer learning with data augmentation through affine transformation, in comparison with training model without data augmentation.

However, the average classification results show that synthetic data cannot fully substitute real images for training CNN classifiers. The absence of real images in the training set can lead to overfitting and lower model accuracy (75%) compared to other state-of-the-art deep CNN-based classifiers, such as Dense Convolutional Network (Densenet) [80,90-91], VGGNet [92], and Inception [93]. Chen et al. [94] introduced data augmentation and ResNet transfer learning for the automatic extraction of features and classification of mammography images, achieving good performance metrics (Acc= 93.15%, Spe=92.17%, Sen=93.83%, AUC=0.95, and loss=0.15). Similarly, Wu et al. [79] presented a deep CNN-ResNet method for breast cancer classification, achieving an AUC of 0.895 in predicting the presence of breast cancer.

These results may be attributed to the theory explained by Lan et al. [95] regarding generated images by traditional augmentation methods. Such images tend to share a similar distribution with the original ones and may not be suitable for processing medical images. Guan et al. [96] demonstrated that RoIs generated by GANs are more similar to real RoIs than affine-transformed RoIs in terms of mean, standard deviation, skewness, and entropy.

However, we acknowledge that breast tumour classification using DM has limitations when using traditional data augmentation, as the model did not significantly improve accuracy (75%) compared to other state-of-the-art deep CNN-based classifiers. This limitation could potentially be overcome by incorporating GAN models for the generation of synthetic data.

One major limitation of our work is the limited number of SR studies based on CNN models using 2D breast images such as mammography and ultrasound, as most of the literature primarily focuses on urban and natural images. In our study, we developed a deep CNN approach for mammography SR, segmentation, and classification of RoIs, resulting in good indices and quality values.
In summary, employing CNNs for SR image enhancement, segmentation, and breast cancer lesion classification can significantly enhance diagnostic capabilities in DM, particularly in settings where resource constraints may limit traditional diagnostic approaches. This approach has the potential to improve early detection and enhance patient outcomes in breast cancer diagnosis.

Conclusions

This article has presented a novel Computer-Aided Diagnosis (CAD) system framework based on deep learning for breast mammography super-resolution, segmentation, and classification, utilizing the concept of transfer learning. We implemented data augmentation with affine transformations for RoIs to enhance the performance of CNN networks. The synthetic RoIs data served as input to two different SR-CNN algorithms, EDSR and SR-RDN, resulting in improved image quality with enhanced resolution and precision for the subsequent segmentation and classification tasks.

We found that EDSR outperformed SR-RDN in the super-resolution task, as evidenced by higher PSNR and SSIM indices. Additionally, the U-Net model was selected as the preferred RoI segmentation technique due to its more reliable results, as demonstrated by Dice, Intersection over Union (IoU), and accuracy metrics.

However, while the ResNet-50 architecture improved accuracy over traditional machine learning algorithms when using generated images with affine transformations, it could not achieve the same accuracy (75%) as other state-of-the-art deep classifiers. This limitation can be attributed to the inability of traditional data augmentation to accurately simulate the real distribution of medical images, as opposed to generative models.

In summary, the importance of SR in lesion segmentation depends on the specific context and characteristics of the medical images in question. Clinical validation and evaluation should guide the selection of image enhancement methods to ensure they enhance diagnostic accuracy without introducing unintended effects.

Future research may involve comparative studies to assess the impact of SR on segmentation accuracy. Additionally, data augmentation techniques based on Generative Adversarial Network (GAN) models will be explored for the generation of synthetic mammography data. This synthetic data could be used as a training dataset for alternative breast mass classifiers based on convolutional networks (e.g., DenseNet, NasNet, VGGNet) with the aim of improving breast lesion classification accuracy and reducing overfitting.

Author Contributions: Conceptualization, Y.J.G., V.L.; methodology, Y.J.G,D.C, A.V, D.C.M; formal analysis, Y.J.G., M.J.R.-Á., and V.L.; investigation, Y.J.G.; resources, Y.J.G.; writing—original draft preparation, Y.J.G.; writing—review and editing, Y.J.G., M.J.R.-Á., and V.L.; visualization, Y.J.G.; supervision, M.J.R.-Á. and V.L.; project administration, M.J.R.-Á. and V.L.; funding acquisition M.J.R.-Á. All authors have read and agreed to the published version of the manuscript.

Data availability. The data that support the findings of this study are openly available inCode 1 https://opticapublishing.figshare.com/s/d3fb2f7112f41c58fa07 (Ref. [96]).

Abbreviations:

AUC	Area under curve
CAD	Computer aided system
CC	Cranio caudal
CNN	Convolutional neural network
DNN	Deep neural network
DDSM	Digital Database for Screening Mammography
DM	Digital mammography
DL	Deep learning
EDSR	Enhanced Deep Residual Network
E2E	End to End
ESRGAN	Enhanced Super-Resolution Generative Adversarial Networks
ESPCN	Efficient sub-pixel convolutional neural network
GAN	Generative adversarial network
HR	High resolution
IoU	Intersection over Union
LR	Low resolution

MDSR Multi-scale deep super-resolution
MLO Mediolateral Oblique
PSNR Peak signal to Noise Ratio
RoI Region of interest
RDNResidua Dense Network
RDB Residual Dense Block
RNNRecurrent Neural Network
ReLU Rectified Linear Unit
SR-GAN Super-Resolution Using a Generative Adversarial Network
SSIM Structural Similarity Index Metric
SISR Single image super resolution
SegNet Segmentation Network
TP True positive
TN True negative
FP False positive
FN False negative
VGG Visual geometric group
VDSR Very Deep Network for SR

4.1 References

[1]. American Cancer Society. 2018. Global Cancer: Facts & Figures, 4th edition URL:http://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/global-cancer-facts-and-figures/global-cancer-facts-and-figures-4th-edition. Pdf

[2]. Oeffinger, K. C., Fontham, E. T., Etzioni, R., et al.(2015). Breast cancer screening for women at average risk: 2015 guideline update from the American Cancer Society. Jama, 314(15), 1599-1614.

[3]. Mainiero, M. B., Lourenco, A., Mahoney, M. C., Newell, M. S., et al. (2016). ACR appropriateness criteria breast cancer screening. Journal of the American College of Radiology, 13(11), R45-R49.

[4]. Mann, R. M., Kuhl, C. K., Kinkel, K., & Boetes, C. (2008). Breast MRI: guidelines from the European society of breast imaging. European radiology, 18(7), 1307-1318.

[5]. Jalalian, A., Mashohor, S. B., Mahmud, H. R., et al. (2013). Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review. Clinical imaging, 37(3), 420-426.

[6]. Sarno, A., Mettivier, G., & Russo, P. (2015). Dedicated breast computed tomography: basic aspects. Medical physics, 42(6Part1), 2786-2804.

[7]. Ponraj, D. N., Jenifer, M. E., Poongodi, P., & Manoharan, J. S. (2011). A survey on the preprocessing techniques of mammogram for the detection of breast cancer. Journal of Emerging Trends in Computing and Information Sciences, 2(12), 656-664.

[8]. Kalles, V., Zografos, G. C., Provatopoulou, X., Koulocheri, D., & Gounaris, A. (2013). The current status of positron emission mammography in breast cancer diagnosis. Breast Cancer, 20(2), 123-130.

[9]. Abdallah, Y. M., Elgak, S., Zain, H., Rafiq, M., Ebaid, E. A., & Elnaema, A. A. (2018). Breast cancer detection using image enhancement and segmentation algorithms.

[10]. Hendrick, R. E. (2010). Radiation doses and cancer risks from breast imaging studies. Radiology, 257(1), 246-253.

[11]. Jiménez-Gaona, Y., Rodríguez-Álvarez, M. J., & Lakshminarayanan, V. (2020). Deep-Learning-Based Computer-Aided Systems for Breast Cancer Imaging: A Critical Review. Applied Sciences, 10(22), 8298.

[12]. Chiao, J. Y., Chen, K. Y., Liao, K. Y. K., Hsieh, P. H., Zhang, G., & Huang, T. C. (2019). Detection and classification the breast tumors using mask R-CNN on sonograms. Medicine, 98(19).

[13]. Chougrad, H., Zouaki, H., & Alheyane, O. (2018). Deep convolutional neural networks for breast cancer screening. Computer methods and programs in biomedicine, 157, 19-30.

[14]. Hussain, Z., Gimenez, F., Yi, D., & Rubin, D. (2018). Differential data augmentation techniques for medical imaging classification tasks. [American Medical Informatics Association.]. AMIA ... Annual Symposium Proceedings - AMIA Symposium. AMIA Symposium, 2017, 979–984.

[15]. Yu, X., Wu, X., Luo, C., & Ren, P. (2017). Deep learning in remote sensing scene classification: A data augmentation enhanced convolutional neural network framework. GIScience & Remote Sensing, 54(5), 741–758. https://doi.org/10.1080/15481603.2017.1323377

[16]. Munir, K., Elahi, H., Ayub, A., Frezza, F., & Rizzi, A. (2019). Cancer diagnosis using deep learning: a bibliographic review. Cancers, 11(9), 1235.

[17]. Salem, M. A. M. (2018). Mammogram-Based cancer detection using deep convolutional neural networks. In 2018 13th International Conference on Computer Engineering and Systems (ICCES) (pp. 694-699). IEEE.

[18]. Dong, C., Loy, C. C., He, K., & Tang, X. (2015). Image super-resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelligence, 38(2), 295–307.

[19]. Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).

[20]. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

21. Li H, Meng X, Wang T, Tang Y, Yin Y. Breast masses in mammography classification with local contour features. Biomed Eng Online. 2017;16(1):44. doi:10.1186/s12938-017-0332-0

[22]. Xu M, Qi S, Yue Y, et al. Segmentation of lung parenchyma in CT images using CNN trained with the clustering algorithm generated dataset. Biomed Eng Online. 2019;18(1):2. doi:10.1186/s12938-018-0619-9

[23]. Robinson, M. D., Chiu, S. J., To th, C. A., Izatt, J. A., Lo, J. Y., & Farsiu, S. (2017). New applications of super-resolution in medical imaging. Book Chapet in Super-resolution imaging. Peyman Milanfar (Editor) (pp. 383-412). CRC Press.

[24]. Yang, W., Zhang, X., Tian, Y., Wang, W., Xue, J. H., & Liao, Q. (2019). Deep learning for single image super-resolution: A brief review. IEEE Transactions on Multimedia, 21(12), 3106-3121.

[25]. Wan, J., Yin, H., Chong, A. X., & Liu, Z. H. (2020). Progressive residual networks for image super-resolution. Applied Intelligence, 1-13.

[26]. Yang, C. Y., Ma, C., & Yang, M. H. (2014). Single-image super-resolution: A benchmark. In European Conference on Computer Vision (pp. 372-386). Springer, Cham.

[27]. Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on image data augmentation for deep learning. Journal of Big Data, 6(1), 1-48.

[28]. Motamed, S., Rogalla, P., & Khalvati, F. (2021). Data Augmentation Using Generative Adversarial Networks (GANs) For GAN-Based Detection Of Pneumonia And COVID-19 In Chest X-Ray Images.

[29]. Antoniou, A., Storkey, A., & Edwards, H. (2017). Data augmentation generative adversarial networks. arXiv preprint arXiv:1711.04340.

[30]. Tran, N. T., Tran, V. H., Nguyen, N. B., Nguyen, T. K., & Cheung, N. M. (2021). On data augmentation for GAN training. IEEE Transactions on Image Processing, 30, 1882-1897.

[31]. Yadav, S. S., & Jadhav, S. M. (2019). Deep convolutional neural network based medical image classification for disease diagnosis. Journal of Big Data, 6(1), 1-18

[32]. Gulati, T., Sengupta, S., & Lakshminarayanan, V. (2020). Application of an enhanced deep super-resolution network in retinal image analysis. In Ophthalmic Technologies XXX (Vol. 11218, p. 112181K). International Society for Optics and Photonics.

[33]. Freeman, W. T., Jones, T. R., & Pasztor, E. C. (2002). Example-based super-resolution. IEEE Computer graphics and Applications, 22(2), 56-65.

[34]. Chang, H., Yeung, D. Y., & Xiong, Y. (2004). Super-resolution through neighbor embedding. In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. (Vol. 1, pp. I-I). IEEE.

[35]. Gribbon, K. T., & Bailey, D. G. (2004). A novel approach to real-time bilinear interpolation. In Proceedings. DELTA 2004. Second IEEE International Workshop on Electronic Design, Test and Applications (pp. 126-131). IEEE.

[36]. Wang, Z., Chen, J., & Hoi, S. C. (2020). Deep learning for image super-resolution: A survey. IEEE transactions on pattern analysis and machine intelligence.4

[37]. Zhang, Yulun et al. (2018). "Residual Dense Network for Image Super-Resolution." : 0–10.Jiang, Y., Chen, L., Zhang, H., & Xiao, X. (2019). Breast cancer histopathological image classification using convolutional neural networks with small SE-ResNet module. PloS one, 14(3), e0214587.

[38]. Dong, C., Loy, C.C., He, K. and Tang, X. (2014). Learning a Deep Convolutional Network for Image Super-Resolution. Proceedings of European Conference on Computer Vision, Zurich, 6-12 September, 184-199.

[39]. Dong, C., Loy, C.C., He, K. and Tang, X. (2016). Image Super-Resolution using Deep Convolutional Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38, 295-307. https://doi.org/10.1109/TPAMI.2015.2439281

[40]. Umehara, K., Ota, J., & Ishida, T. (2017). Super-resolution imaging of mammograms based on the super-resolution convolutional neural network. Open Journal of Medical Imaging, 7(4), 180-195.

[41]. Jiang, Y., & Li, J. (2020). Generative Adversarial Network for Image Super-Resolution Combining Texture Loss. Applied Sciences, 10(5), 1729.

[42]. Tschuchnig, M. E., Oostingh, G. J., & Gadermayr, M. (2020). Generative Adversarial Networks in Digital Pathology: A Survey on Trends and Future Potential. arXiv preprint arXiv:2004.14936.

[43]. Ledig, C., Theis, L., Huszár, F., Caballero, J.,et al, (2017). Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4681-4690).

[44]. Sood, R., Topiwala, B., Choutagunta, K., Sood, R., & Rusu, M. (2018). An application of generative adversarial networks for super resolution medical imaging. In 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 326-331). IEEE.

[45]. Lim, B., Son, S., Kim, H., Nah, S., & Mu Lee, K. (2017). Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 136-144).

[46]. Abbass, M. Y. (2020). Residual Dense Convolutional Neural Network for Image Super-Resolution. Optik, 165341.

[47]. Wang, X., Yu, K., Wu, S., et al.(2018). Esrgan: Enhanced super-resolution generative adversarial networks. In Proceedings of the European Conference on Computer Vision (ECCV) (pp. 0-0).

[48]. Kim, J., Kwon Lee, J., & Mu Lee, K. (2016). Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1646-1654).

[49]. Shi, W., Caballero, J., Huszár, F., et al.(2016). Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1874-1883).

[50]. Wang, Z., Liu, D., Yang, J., Han, W., & Huang, T. (2015). Deep networks for image super-resolution with sparse prior. In Proceedings of the IEEE international conference on computer vision (pp. 370-378).

[51]. Kim, J., Kwon Lee, J., & Mu Lee, K. (2016). Deeply-recursive convolutional network for image super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1637-1645)

[52]. Tai, Y., Yang, J., & Liu, X. (2017). Image super-resolution via deep recursive residual network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3147-3155).

[53]. Al-Masni, M. A., Al-Antari, M. A., Park, J. M., Gi, G., et al (2018). Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system. Computer methods and programs in biomedicine, 157, 85-94.

[54]. Geras, K. J., Wolfson, S., Shen, Y., Wu, N., Kim, S., Kim, E. et al (2017). High-resolution breast cancer screening with multi-view deep convolutional neural networks. arXiv preprint arXiv:1703.07047.

[55]. Lee, J. G., Jun, S., Cho, Y. W., Lee, H., Kim, G. B., Seo, J. B., & Kim, N. (2017). Deep learning in medical imaging: general overview. Korean journal of radiology, 18(4), 570-584.

[56]. S. Sengupta, A. Singh, H. A. Leopold, T. Gulati, V. Lakshminarayanan (2020). Ophthalmic diagnosis using deep learning with fundus images – A critical review, Artificial Intelligence in Medicine, Volume 102,2020,101758,ISSN 0933-3657, https://doi.org/10.1016/j.artmed.2019.101758

[57]. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[58]. Singh, V. K., Rashwan, H. A., Romani, S., et al,. (2020). Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network. Expert Systems with Applications, 139, 112855.

[59]. Tong, T., Li, G., Liu, X., & Gao, Q. (2017). Image super-resolution using dense skip connections. In Proceedings of the IEEE International Conference on Computer Vision (pp. 4799-4807).

[60]. Yang, J., Wright, J., Huang, T. S., & Ma, Y. (2010). Image super-resolution via sparse representation. IEEE transactions on image processing, 19(11), 2861-2873.

[61]. Schulter, S., Leistner, C., & Bischof, H. (2015). Fast and accurate image upscaling with super-resolution forests. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3791-3799).

[62]. Kaji, S., & Kida, S. (2019). Overview of image-to-image translation by use of deep neural networks: denoising, super-resolution, modality conversion, and reconstruction in medical imaging. Radiological physics and technology, 12(3), 235-248.

[63]. Rajkumar, S., & Malathi, G. (2016). A comparative analysis on image quality assessment for real time satellite images. Indian J. Sci. Technol, 9(34).

[64]. Juhong, A., Li, B., Yao, C. Y., Yang, C. W., Agnew, D. W., Lei, Y. L., ... & Qiu, Z. (2023). Super-resolution and segmentation deep learning for breast cancer histopathology image analysis. Biomedical Optics Express, 14(1), 18-36.

[65].Sheba, K. U., & Raj, S. G. (2016). Objective quality assessment of image enhancement methods in digital mammography–a comparative study. Signal & Image processing: An International Journal, 7(4), 1-13.

[66]. Gao, X., Lu, W., Tao, D., & Li, X. (2010). Image quality assessment and human visual system. In Visual Communications and Image Processing 2010 (Vol. 7744, p. 77440Z). International Society for Optics and Photonics.

[67]. Abdelhafiz, D., Yang, C., Ammar, R., & Nabavi, S. (2019). Deep convolutional neural networks for mammography: advances, challenges and applications. BMC bioinformatics, 20(11), 281.

[68].Soulami, K. B., Kaabouch, N., Saidi, M. N., & Tamtaoui, A. (2021). Breast cancer: One-stage automated detection, segmentation, and classification of digital mammograms using UNet model based-semantic segmentation. Biomedical Signal Processing and Control, 66, 102481.

[69]. Valvano, G., Santini, G., Martini, N., Ripoli, A., Iacconi, C., Chiappino, D., & Della Latta, D. (2019). Convolutional neural networks for the segmentation of microcalcification in mammography imaging. Journal of Healthcare Engineering, 2019.

[70].Abdelhafiz, D., Bi, J., Ammar, R., Yang, C., & Nabavi, S. (2020). Convolutional neural network for automated mass segmentation in mammography. BMC bioinformatics, 21(1), 1-19.

[71].Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[72].He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[73].Almajalid, R., Shan, J., Du, Y., & Zhang, M. (2018). Development of a deep-learning-based method for breast ultrasound image segmentation. In 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 1103-1108). IEEE.

[74].Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

[75].Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence, 39(12), 2481-2495.

[76].Romera, E., Alvarez, J. M., Bergasa, L. M., & Arroyo, R. (2017). Erfnet: Efficient residual factorized convnet for real-time semantic segmentation. IEEE Transactions on Intelligent Transportation Systems, 19(1), 263-272.

[77]. Lazo, J. F., Moccia, S., Frontoni, E., & De Momi, E. (2020). Comparison of different CNNs for breast tumor classification from ultrasound images. arXiv preprint arXiv:2012.14517.

[78]. Alkhaleefah, M., Chittem, P. K., Achhannagari, V. P., Ma, S. C., & Chang, Y. L. (2020). The influence of image augmentation on breast lesion classification using transfer learning. In 2020 International Conference on Artificial Intelligence and Signal Processing (AISP) (pp. 1-5). IEEE.

[79]. Wu, N., Phang, J., Park, J., Shen, Y., Huang, Z., Zorin, M., ... & Geras, K. J. (2019). Deep neural networks improve radiologists' performance in breast cancer screening. IEEE transactions on medical imaging, 39(4), 1184-1194.

[80]. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708).

[81]. Ferreira, C. A., Melo, T., Sousa, P., Meyer, M. I., Shakibapour, E., Costa, P., & Campilho, A. (2018). Classification of breast cancer histology images through transfer learning using a pre-trained inception resnet v2. In International Conference Image Analysis and Recognition (pp. 763-770).

[82]. Lee, R., Gimenez, F., Hoogi, A. et al. A curated mammography data set for use in computer-aided detection and diagnosis research. Sci Data 4, 170177 (2017). https://doi.org/10.1038/sdata.2017.177

[83]. Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I., Boggis, C., Ricketts, I., et al. (2015). Mammographic Image Analysis Society (MIAS) database v1.21 [Dataset]. https://www.repository.cam.ac.uk/handle/1810/250394

[84].Li, S., Hatanaka, Y., Fujita, H., Hara, T., & Endo, T. (1999). Automated Detection of Mammographic Masses in MIAS Database. Medical Imaging Technology, 17(4), 427-428.

[85]. Moreira, I. C., Amaral, I., Domingues, I., Cardoso, A., Cardoso, M. J., & Cardoso, J. S. (2012). Inbreast: toward a full-field digital mammographic database. Academic radiology, 19(2), 236-248.

[86].Yu, T., & Zhu, H. (2020). Hyper-parameter optimization: A review of algorithms and applications. arXiv preprint arXiv:2003.05689.

[87]. Sara, U., Akter, M., & Uddin, M. S. (2019). Image quality assessment through FSIM, SSIM, MSE and PSNR—a comparative study. Journal of Computer and Communications, 7(3), 8-18.

[88] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2019). Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE transactions on medical imaging, 39(6), 1856-1867.

[89].Vianna, P., Farias, R., & de Albuquerque Pereira, W. C. (2021). U-Net and SegNet performances on lesion segmentation of breast ultrasonography images. Research on Biomedical Engineering, 37(2), 171-179.

[90].Jiménez Gaona Y., Rodriguez-Alvarez M.J., Espino-Morato H., Castillo Malla D., Lakshminarayanan V. (2021) DenseNet for Breast Tumor Classification in Mammographic Images. In: Rojas I., Castillo-Secilla D., Herrera L.J., Pomares H. (eds) Bioengineering and Biomedical Signal and Image Processing. BIOMESIP 2021. Lecture Notes in Computer Science, vol 12940. Springer, Cham. https://doi.org/10.1007/978-3-030-88163-4_16

[91]. Rybiałek, A., & Jeleń, Ł. (2020). Application of DenseNets for Classification of Breast Cancer Mammograms. In International Conference on Computer Information Systems and Industrial Management (pp. 266-277). Springer, Cham.

[92].Jahangeer, G. S. B., & Rajkumar, T. D. (2021). Early detection of breast cancer using hybrid of series network and VGG-16. Multimedia Tools and Applications, 80(5), 7853-7886.

[93].Khan, S., Islam, N., Jan, Z., Din, I. U., & Rodrigues, J. J. C. (2019). A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. Pattern Recognition Letters, 125, 1-6.

[94].Chen, Y., Zhang, Q., Wu, Y., Liu, B., Wang, M., & Lin, Y. (2018). Fine-tuning ResNet for breast cancer classification from mammography. In The International Conference on Healthcare Science and Engineering (pp. 83-96). Springer, Singapore.

[95].Lan, L., You, L., Zhang, Z., Fan, Z., Zhao, W., Zeng, N., ... & Zhou, X. (2020). Generative adversarial networks and its applications in biomedical informatics. Frontiers in Public Health, 8, 164.

[96]. Guan, S., & Loew, M. (2019). Breast cancer detection using synthetic mammograms from generative adversarial networks in convolutional neural networks. Journal of Medical Imaging, 6(3), 031411.

[96].Y. Jimenez-Gaona, "Breast Mass Regions Classification from Mammograms using Convolutional Neural Networks and Transfer Learning", figshare (2023),

https://opticapublishing.figshare.com/s/d3fb2f7112f41c58fa07

Chapter 4

ULTRASOUND BREAST IMAGES DENOISING USING GENERATIVE ADVERSARIAL NETWORKS (GANs)

Yuliana Jiménez Gaona, María José Rodríguez-Alvarez ,Líder Escudero, Carlos Sandoval, Vasudevan Lakshminarayanan

Intelligent Data Analysis, vol. Pre-press, no. Pre-press, pp. 1-18, 2024

Abstract

Introduction: Ultrasound in conjunction with mammography imaging, plays a vital role in the early detection and diagnosis of breast cancer. However, speckle noise affects medical ultrasound images and degrades visual radiological interpretation. Speckle carries information about the interactions of the ultrasound pulse with the tissue microstructure, which generally causes several difficulties in identifying malignant and benign regions. The application of deep learning in image denoising has gained more attention in recent years. **Objectives:** The main objective of this work is to reduce speckle noise while preserving features and details in breast ultrasound images using GAN models. **Methods:** We proposed two GANs models (Conditional GAN and Wasserstein GAN) for speckle-denoising public breast ultrasound databases: BUSI, DATASET A, AND UDIAT (DATASET B). The Conditional GAN model was trained using the Unet architecture, and the WGAN model was trained using the Resnet architecture. The image quality results in both algorithms were measured by Peak Signal to Noise Ratio (PSNR, 35-40 dB) and Structural Similarity Index (SSIM, 0.90-0.95) standard values. **Results:** The experimental analysis clearly shows that the Conditional GAN model achieves better breast ultrasound despeckling performance over the datasets in terms of PSNR=38.18 dB and SSIM=0.96 with respect to the WGAN model (PSNR=33.0068 dB and SSIM=0.91) on the small ultrasound training datasets. **Conclusions:** The observed performance differences between CGAN and WGAN will help to better implement new tasks in a computer-aided detection/diagnosis (CAD) system. In future work, these data can be used as CAD input training for image classification, reducing overfitting and improving the performance and accuracy of deep convolutional algorithms.

Keywords: Breast cancer; Ultrasound image denoising; Generative adversarial networt

1. Introduction

Medical image analysis plays an important role in breast cancer screening, feature extraction, segmentation, and classification breast lesions locally. There are several breast cancer detection methods, such as Positron Emission Tomography (PET) [1], Computer Tomography (CT) [2] and Magnetic Resonance Imaging (MRI) [3], which are usually used when women are at high risk of breast cancer. Other complementary techniques such as X-ray mammography [4] and ultrasound (US) [5] are more commonly used in screening programs, according to the American Cancer Society.

Among these modalities, US is used as a complementary imaging modality for further evaluation of lesions detected early by mammography due to its non-invasive nature, low cost, safety, portability, and low radiation dose. However, one of its main shortcomings is the poor quality of US image, which is corrupted by random noise added during its acquisition [6,7], i.e. low contrast and different brightness levels, resulting in increased noise and artifacts that can affect the radiologist's opinion and diagnosis. US images have a granular appearance called speckle noise, which degrades visual assessment [8], making it difficult for humans to distinguish normal from pathological tissue in diagnostic examinations.

Image denoising techniques, typically low-dose, address this problem [9]. The primary purpose of denoising is to restore the maximum detail of the image by removing excess noise [10], while preserving as much as possible the feature details to benefit the diagnosis and classification of benign, premalignant, and malignant abnormalities (microcalcifications, masses, nodules, tumors, cysts, fibroadenoma, adenosis, and lesions) that may be difficult to identify at first sight or early in the patient.

Thus, denoising medical images is essential before training a classifier based on deep-learning models. Recently, several US denoising techniques based on deep learning have been widely used, such as Convolutional Neural Networks (CNN) [11-14], Generative Adversarial Networks (GANs) [15-17], and Autoencoders (AEs) [18,19], which can recover the original dataset and make it noise-free with better robustness and precision [20]. Deep learning methods have obtained better results in medical imaging in comparison with previous methods such as Wavelet, Wiener, Gaussian [21], Multi-Layer perceptron [22], Dictionary Learning [23], Least Square, Bilateral Filter, Non-Local Mean [24]. Variational approaches [6,25], because these filters have presented some limitations such as smoothing problems, more computational cost, and inability to preserve information such as edges and textures of images as well as possible [25].

1.1 Related Work

Many traditional denoising filtering techniques have been proposed in the literature to reduce speckle noise [26-29], which can be categorized into three main types: 1) Spatial domain (Median filter, Mean filter, Adaptive Mean Filter, Frost, Total variation filter, Anisotropic Diffusion, Nonlocal means filter, Linear Minimum Mean Squared Error (LMMSE)). 2) Transform domain (Wiener filter, Low pass filter, Discrete wavelet transform), and 3) Deep learning-based techniques such as Convolutional Neural Networks (CNN), Generative Adversarial Networks (GAN), and Variational Autoencoders (VAEs).

The Spatial and Transform domain methods are computationally simple and fast but sometimes blur the image, and there can be a loss of resolution and low accuracy. Spatial domain filters also have size limitations and window shape problems [28].

However, Deep learning-based models can provide better results compared to these traditional methods, because deep models gives better visual quality by extracting various features of an image as example Li et al. proposed TP-Net [30] as 3D shape classification and segmentation tasks, on a wide range of common datasets, which main contribution is the design of dilated convolution strategy tailored for the irregular and non-uniform structure of 3D mesh data.

Several Generative models (GANs,VAEs) have been successfully used for medical image denoising and data augmentation to improve robustness and prevent overfitting in deep CNN image classification algorithms. Some relevant works are discussed in this section.

Wu et al. [31] implemented a perceptual metrics-guided GAN (PIGGAN) framework to intrinsically optimize generation processing, and experiments show that PIGGAN can produce photo-realistic results and quantitatively outperforms state-of-the-art (SOTA) methods. Pang et al. [32] implemented the TripleGAN model to augment the breast US images. These synthetic images were used to classify breast masses classification using the CNN model, achieving a classification accuracy of 90.41%, sensitivity of 87.94% and specificity of 85.86%. Al-Dhabyani et al. [33] first used breast US data augmentation with GAN and then two deep learning classification approaches: (i) CNN (AlexNet) and (ii) TL (VGG16, ResNet, Inception, and NASNet), achieving in the BUSI dataset an accuracy of 73%,84%,82%,89%,91% and in Dataset B (UDIAT) an accuracy of 75%, 80%, 77%, 86%,90% respectively.

Jain et al [34] found that CNN provided comparable and, in some cases, superior performance to Wavelet and Markov Random Field methods. Thus, the Resnet approach proposed by MRDGet al. [11] was used to improve mammography image quality with a peak signal-to-noise ratio (PSNR) of 36.18 and a similar structural index metrix (SSIM) of

0.841. Feng et al [13] implemented a hybrid neural network for US denoising based on the Gaussian noise distribution and VGGNet model to extract the structural boundary information, the results show a (PSNR=30.57, SSIM=0.90, Mean Square Error (MSE)=66.61) US denoising effectiveness.

Denoising autoencoders based on convolutional layers also perform well for their ability to extract spatial solid correlation [35]. Kaji et al. [9] present an overview describing encoder–decoder networks (pix-2-pix) and cycle GAN as image noise reduction.

Chen et al. [12] proposed the autoencoder and the residual encoder–decoder CNN for low-dose computer tomography (CT) imaging, achieving a good performance index (PSNR of 39.19/ SSIM of 0.93 and Root Mean Square Deviation (RMSD) of 0.0097), compared to with other methods in terms of noise suppression, structure preservation, and lesion detection.

However, the use of GANs is considered more stable than autoencoders. GANs are typically used when dealing with images or visual data and work better for signal image processing, such as anomaly detection; on the contrary, VAEs are used for predictive maintenance or security analysis applications [35]. For the previous reason, several GANs have recently been used for data augmentation [36-40], image super-resolution [21], image translation [9], and noise reduction in the medical field [41,42].

Zhou et al. [37] proposed a GAN + U-Net network (generator model) to achieve mapping between low-quality US images and corresponding high-quality images. In contrast to the traditional GAN method, U-Net is used to reconstruct the image's tissue structure, details, and speckles. The evaluation indices indicated that PSNR, SSIM, and MI (Mutual dependence index) values are increased by 48.3%, 205.0%, and 44.0% and that the proposed method can successfully reconstruct a high-quality image.

The most recent deep GAN models used for image denoising are Conditional GAN [43] and Wasserstein GAN [44], which have shown better performance than conventional denoising algorithms [45,46]. Kim et al. [43] implemented a CGAN network as a medical image denoising algorithm, where the SSIM metric was improved by 1.5 and 2.5 times over conventional methods (Nonlocal Means and Total Variation) respectively, demonstrating a superiority in quantitative evaluation. Vimala et al. [47] proposed an image noise removal in US breast images based on Hybrid Deep Learning Technique, where local speckle noise was destroyed, reaching a signal-to-noise ratios (SNRs) greater than 65 dB, PSNR ratios greater than 70 dB, edge preservation index values more significant than the experimental threshold of 0.48. Zou et al. [37] proposed a network model based on the Wasserstein GAN for image denoising, which improved the noise removal effect.

Based on the previous mentioned our propose integrates concepts from breast cancer research and ultrasound image denoising in a comparative study to evaluate the effect of image pre-processing in improving breast image quality. Improving image quality clarifies patterns, allowing the deep learning model to identify and classify features within the image more accurately. In this study, we explore a novel approach by combining fine-tuning techniques GANs + CNNs, providing new insights into breast cancer classification.

Denoising of medical images has been used to improve the performance of CNN segmentation and classification algorithms [48-50]. Ans several CNN methods for general image denoising have been studied ADNet, NERNet, SAnet, CDNet, DRCNN [51], but in this research, as a technical novelty, we combine Conditional GAN + Unet and WGAN + Resnet particularly focusing on the medical image quality improvement of breast ultrasound. The results will help to better implement new tasks in a computer-aided detection/diagnosis (CAD) system.

Consenquently, this study aims to: (i) to implement two types of GANs+CNNs architecture models as speckle denoising in ultrasound breast images, and (ii) to select the best architecture to generate new quality images based on the quantitative evaluation metrics (PSNR and SSIM).

2. Materials and Methods

2.1 Databases collection

Three publicly available breast US databases were used in this study: (i) The *Breast Ultrasound Images Dataset (BUSI,* https://scholar.cu.edu.eg/?q=afahmy/pages/dataset*)* [52]. This contains data from 600 female patients. The dataset consists of 780 images (133 normal, 437 benign and 210 malignant) with an average image size of 500x500 pixels. (ii) The *Dataset A* is obtained from Rodrigues et al. [53] (https://data.mendeley.com/datasets/wmy84gzngw/1) and contains 250 breast cancer images, 100 benign and 150 malignant. The *Dataset B* (Breast Ultrasound Lesions Dataset, http://www2.docm.mmu.ac.uk/STAFF/m.yap/dataset.php) collected in UDIAT-Centre Diagnóstic, Corporació Parc Taulí, Sabadell (Spain). The dataset consists of 163 images of different women with an average image size of 760 × 570 pixels, each of the images shows one or more lesions. Of the 163 images of lesions, 53 are images of cancerous masses and 110 with benign lesions [54].

A total of 1060 US images were used to train the GAN models; see Table 1.

Table 1. Breast ultrasound public databases

Dataset	Benign	Malignant	Total
BUSI	437	210	647
Dataset A	100	150	250
Dataset B	110	53	163
Total	647	413	**1060**

Figure 1 shows the workflow used in denoising breast ultrasound images, which is divided into the following steps: i) Acquisition of public ultrasound databases, ii) Dimensionality and cropping of regions of interest (RoIs), iii) Image denoising using two GANs + CNN models, and iv) Image quality evaluation.

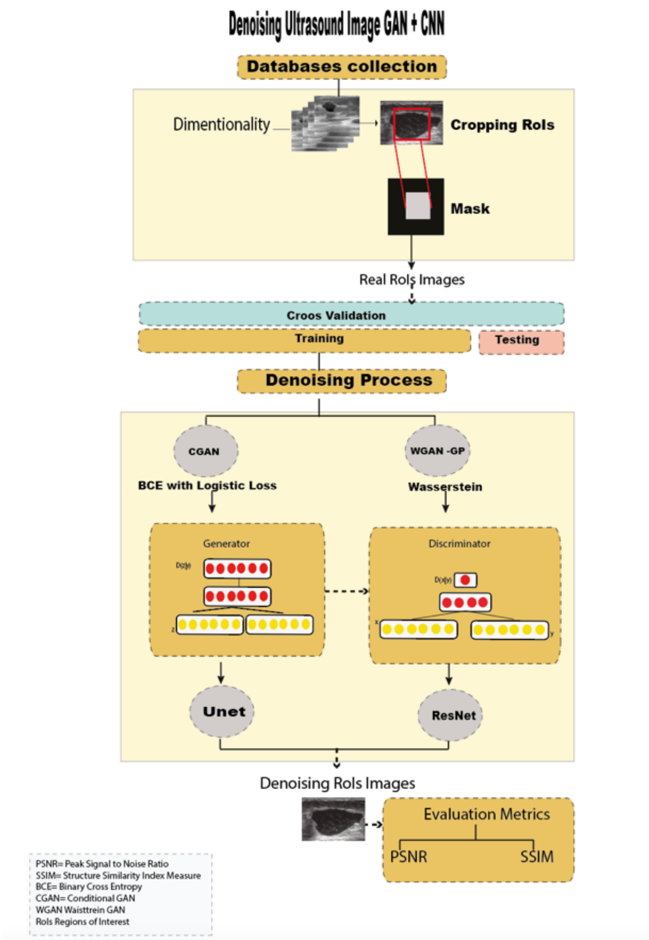


Figure 1. Workflow of GANs+CNN models implementation in breast ultrasound denoising.

2.2 Data Dimensionality and RoIs cropping

The torchvision (pytorch) library was used to perform transformations (preserving all features and structure of the images) and to standardize the images to a single dimension (256x256 pixels), which were acquired in different sizes (BUSI: 431x476, 765x590, 786x556; Dataset A: 153x87, 95x75, 93x57; Dataset B: 760x570).

According to Wu et al. [36], synthesizing a lesion into RoIs (regions of interest) gives advantages to the generative model, as it generates more realistic lesions, improving subsequent classification performance over traditional augmentation techniques. Thus, automatic RoI extraction was performed on all US images.

Then, using a cross-validation technique, the dataset was randomly divided (with the Sklearn library) into a training set (80%, 851 images) and a testing set (20%, 209 images) for training the GAN models (with the Tensorflow, Keras libraries).

2.3 Generative Adversarial Network

The GAN architecture is represented by a generative (G) network and a discriminator (D) network, which are trained simultaneously. While the G network is trained to produce realistic images G(z) from a random vector z, the D network is trained to discriminate between real and generated images [55]. In the original GAN the optimization function was formulated by the Eq. 1.

$$min_G \ max_D \ V\ (D, G) = \ E_{x \sim P_{r(x)}}[log \ log \ D(x)\ ] + E_{z \sim P_{z(z)}}[log \ log \ (1 - D\big(G(z)\big))\ ] \qquad (1)$$

Given random noise vector *z* and real image *x*, the generator attempts to minimize log (1 – D(G(z)) and the discriminator attempts to maximize $log \ log \ D(x)$. Whre, P_r and P_z are the real data distribution and the noise data distribution, x is the input variable, D(x) is the prediction label and D(z) is the generated sample.

In this work, we used two ultrasound denoising GANs; (i) conditional GAN and (ii) WGAN, both has been widely used in medical image reconstruction, denoising and data augmentation [56]. Especially CGAN model have been propose as new framework that can largely mitigate the biases and discriminations in machine learning systems while at the same time enhancing the prediction accuracy of these systems [57].

Conditional GAN (CGAN)

CGAN was introduced by Douzas et al. [58], as an extension of GAN with conditional information in D and G. GANs are generative models that learn a mapping from random noise vector z to output image y, (G: z → y) [59]. In contrast, conditional GANs learn a mapping from observed image x and random noise vector z to y, (G : {x, z} → y). The CGAN objective function is framed by Eq. 2, where G tries to minimize this objective function and D tries to maximize it.

$$L_{cGAN}(G, D) = E_{x,y}\,[log\ log\ D(x, y)\,] + E_{x,z}\,[D(x, G(x, z))\,] \qquad (2)$$

In this work, the generator and discriminator architectures were adapted from [60-61]. A manual exploration of different configurations in the general hyperparameters was performed to optimize the denoising of breast US images, before selecting and implementing our CGAN model. The selected hyperparameters are: Number of epochs=40, Buffer size=954, Batch size=80; Optimiser=Adam, Activation function=Binary Cross-Entropy Loss, Generator layers=48 and Discriminator layers=12.

The *denoiser generator* network is based on the U-Net [61] architecture, which consists of a contraction path and an expansion path. This is composed of 48 convolutional layers including the input layer, 8 contraction layers, 7 expansion layers, 6 concatenation layers spread over the expansion layers, and finally a transposed convolutional layer. Each encoder and decoder block is replaced by residual dense connectivity and batch normalization to remove speckle noise followed by the ReLU function (Figure 2, Appendix S.1 and S.2).

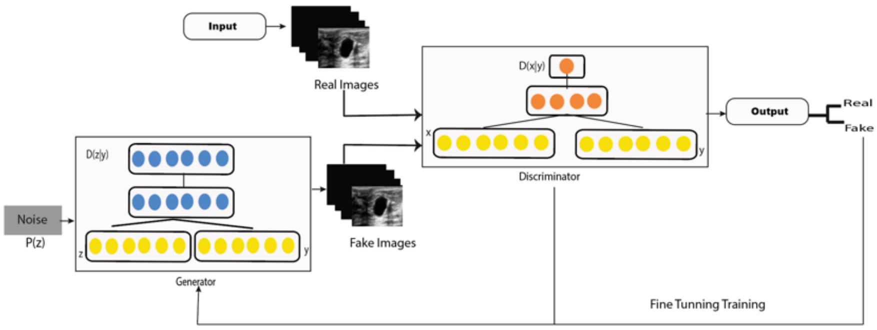


Figure 2. CGAN model.

The *denoiser discriminator* network is based on a Markovian random field (PatchGAN). This consists of an input convolutional layer and 24 convolutional layers followed by batch normalization and a ReLU function (Figure 2). The output consists of successive convolutional layers 256, 128, 64 and 1. This means that as the input image passes through each of the convolution blocks, the spatial dimension is reduced by a factor of two.

Wasserstein GAN (WGAN)

WGAN was introduced by Arjovsky et al [62], which uses a Wasserstein distance instead of a JS (Jensen-Shanon) or KL (Kullback-Leibler) divergence to evaluate the discrepancy between the distribution distance of noisy and denoised images. It provides a better approximation of the distribution of the observed data in the training data.

The Wassertein (W) model is defined as Eq. 3:

$$W\left(P_r, P_g\right) = inf_{\gamma} \sim \Pi \left(P_r, P_g\right) E(x, y) \sim \gamma[\|x - y\|] \qquad (3)$$

Where Π (P_r, P_g) denotes the set of all the joint distributions γ(x, y) based on the marginal values of P_r and P_g; γ(x,y) indicates how many "RoIs" must be transported from *x* to *y* in order to transform the distributions P_r into the distribution P_g; *x* and *y* denote the predicted and real actual values, respectively, and P denotes the probability distribution. The general hyperparameters implemented in this model are number of epochs= 130, buffer size = 954, batch size=60; optimizer=Adam, cctivation function=Wasserstein, generator layers=26 and discriminator layers=12.

The denoising generator, was trained by the Resnet model [63]. The generator contains 54 layers, including the input layer, 8 sequential layers of 3 layers each (convolutional layer, normalisation layer and LeakyReLU layer), 7 residual sequences of 4 layers each (transposed convolutional layer, normalisation layer, dropout layer and LeakyReLU layer) and finally a transposed convolutional layer (Figure 3, Appendix S.3 and S.4).

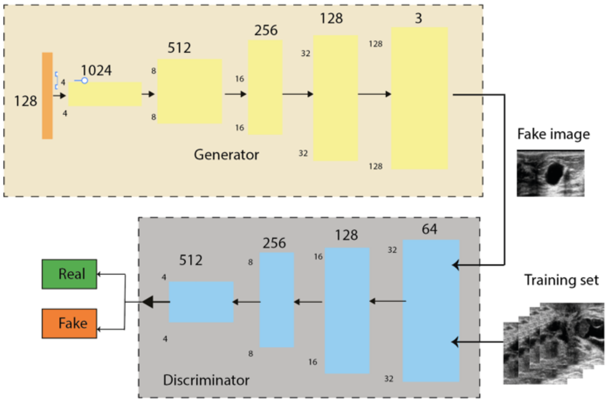


Figure 3. WGAN model. Adapted from Hao, Zhuangzhuang et al. (2022).

The *denoising discriminator* uses the PatchGAN model combined with the Res-Net architecture (convolutional layer, normalization layer and LeakyReLU layer), where the layers were connected directly in a single sequence instead of linking several sequences.

The training phase was carried out with the Google Colab GPU PRO environment, using the Tensorflow and Sklearn libraries for image pre-processing, and PyTorch (CUDA 10.2 graphics cores) to obtain more computational resources and minimise the algorithm execution time. The Tensorflow and Keras libraries were used to train the GAN models.

2.4 Evaluation metrics

In addition, most filter techniques use various evaluation metrics such as Mean Square Error (MSE), Root-Mean-Square Error (RMSE), Signal-to-Noise Ratio (SNR), Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) to assess image quality.

For quantitative comparison, the PSNR and SSIM [64,65] were introduced to measure image restoration quality, which is widely used in biomedical applications, especially in mammography and US diagnosis and cancer detection fields.

The PSNR is the metric used to measure the quality of the denoising image when it is corrupted due to noise and blur. A higher value of PSNR indicates a higher quality rate. The standard value of PSNR is 35 to 40 dB (Table 2). The PSNR is calculated by Eq. 4, where is the variance of noise evaluated over the RoI image and is the variance of the filtered image.

$PSNR = 10\ \log \log\ \left(\frac{\sigma_S^2}{\sigma_{\hat{S}}}\right)$ (4)

SSIM is a perception-based model that considers the image degradation as perceived change in contrast and structural information. Thus, we can apply this value to assess the quality of any images [66], which lies from 0 to 1 (table 2).

Table 2. PSNR and SSIM range values.

Quality	PSNR	SSIM
Low	< 30	< 0.90
Aceptable	35 - 40	0.90 - 0.95
High	40 - 50	0.95-1

SSIM index is computed using the correlation coefficient, see Eq. 5.

$SSIM\ (x, y) = \frac{(2\mu_{x+}\mu_y)(2\sigma_{xy})}{(\mu x^2 + \mu y^2)(\sigma x^2 + \sigma y^2)}$ (5)

Where,

$$u_x = \frac{1}{N}\sum_{i}^{N} = 1\ x_i$$

$$u_y = \frac{1}{N}\sum_{i}^{N} = 1\ y_i$$

$$\sigma_x = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N} (x_i - \mu_x)^2}$$

$$\sigma_y = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N} (y_i - \mu_y)^2}$$

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N} (x_i - \mu_x)(y_i - \mu_y)$$

N is the total number of pixels in the image. $x_{i,j}$ is the filtered image at *i* and *j* coordinates and $y_{i,j}$ is the noisy image at *i* and *j* coordinates. $\mu_x\ \mu_I$ is the mean of reference images, $\mu_y\ \mu_i$ is the mean of filtered images, $\sigma_x\sigma_I^2$ is the variance of reference images, $\sigma_y\sigma_i^2$ is the variance of filtered image, $cov_{Ii}\ \sigma_{xy}$ is the covariance of filtered image.

3. Results

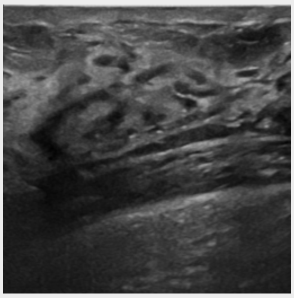
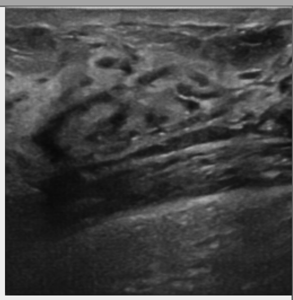
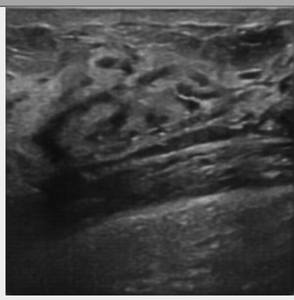
This section presents the most relevant numerical experiments obtained from speckle removal GAN algorithms. First, to improve the algorithm performance, the RoI images were used as GAN training models; in total, we denoising 1060 malignant and benign RoIs. The image quality of the generated data was evaluated with PSNR and SSIM metrics, which are expressed in terms of average value. The most relevant scores are displayed in Table 4; these indicate that the Conditional GAN model showed a significant improvement compared to the other model.

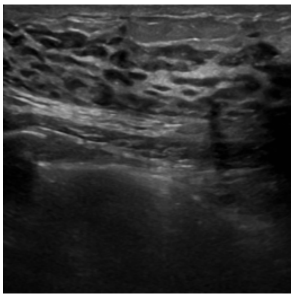
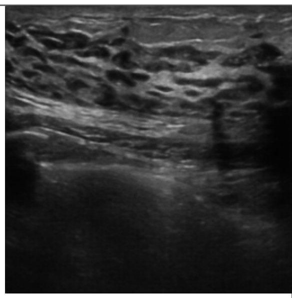
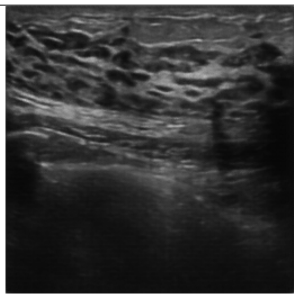
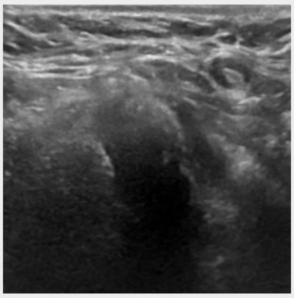
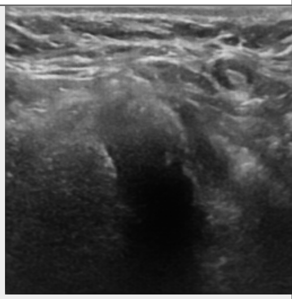
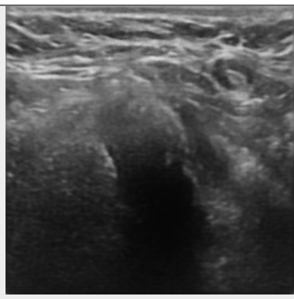
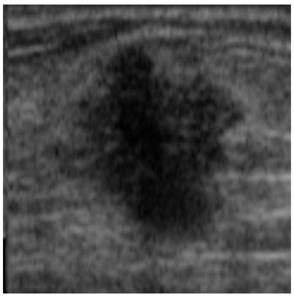
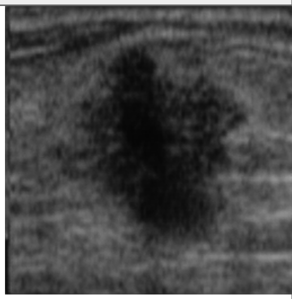
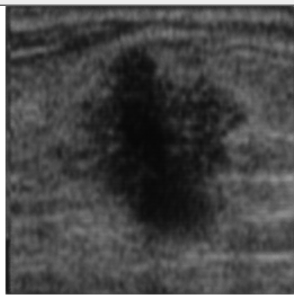
Table 3. Summary of the CGAN and WGAN average comparison results (PSNR and SSIM).

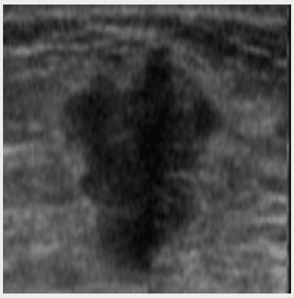
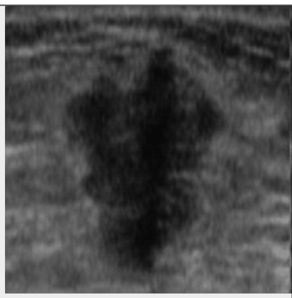
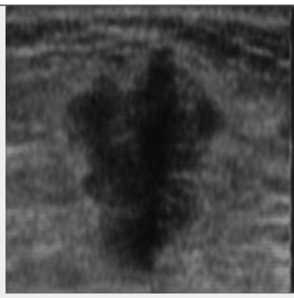
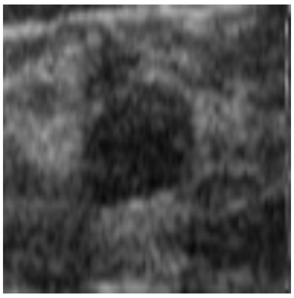
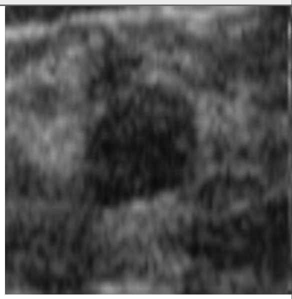
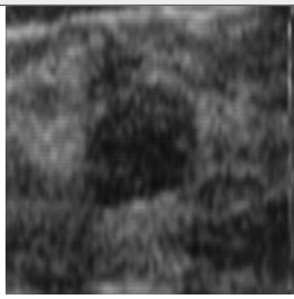
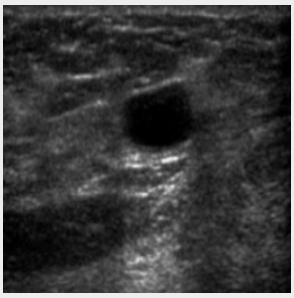
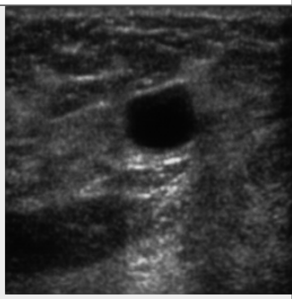
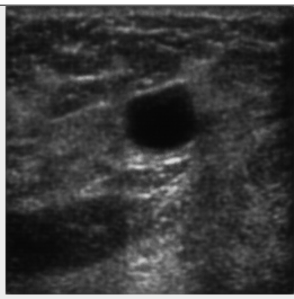
ID	CGAN		ID	WGAN	
	PSNR (dB)	SSIM		PSNR (dB)	SSIM
BUSI					
img_busi _7	39.8433	0.974624	**img_busi_7**	35.0476	0.930708
img_busi _56	39.8223	0.906241	**img_busi_56**	35.1609	0.818753
img_busi _58	39.8341	0.976325	**img_busi_58**	35.5627	0.952616
img_busi_60	40.1839	0.978979	**img_busi_60**	35.2361	0.931421
img_busi _70	39.7809	0.971730	**img_busi_70**	35.7736	0.943916
img_busi_175	39.4099	0.972768	**img_busi_175**	35.5431	0.942358
img_busi _199	39.7116	0.929269	**img_busi_199**	35.3159	0.939286
DATASET A					
img_datasetA_6	41.8245	0.977663	**img_datasetA_6**	38.2882	0.965505
img_datasetA_11	42.1565	0.977758	**img_datasetA_11**	37.7888	0.965114
img_datasetA_23	41.8171	0.978695	**img_datasetA_23**	38.2925	0.967823
img_datasetA_76	41.9047	0.977636	**img_datasetA_76**	38.4245	0.971207
img_datasetA_188	41.9888	0.977348	**img_datasetA_188**	37.2507	0.968667
img_datasetA_217	41.9424	0.978819	**img_datasetA_217**	37.7399	0.971379

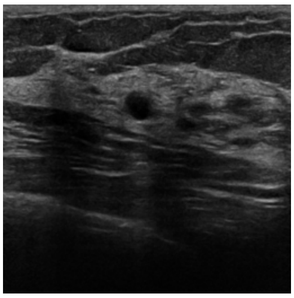
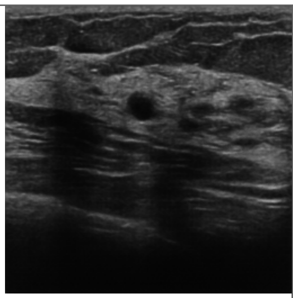
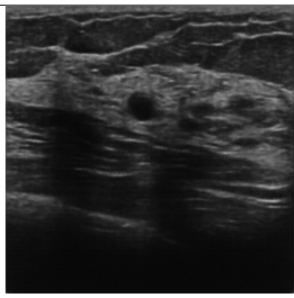
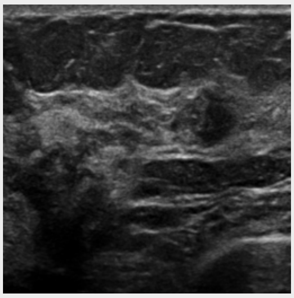
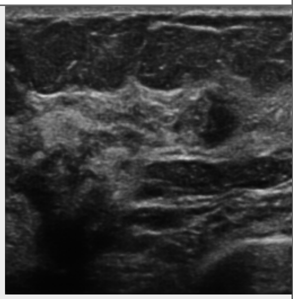
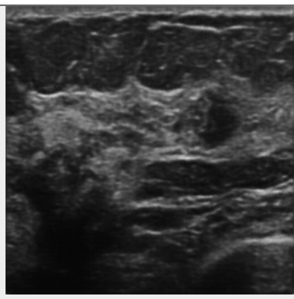
img_datasetA_222	42.6280	0.980217	**img_datasetA_222**	37.2250	0.967832
UDIAT					
img_udiat_55	38.0735	0.876853	**img_udiat_55**	34.1079	0.936932
img_udiat_77	40.4911	0.967255	**img_udiat_77**	36.4130	0.939990
img_udiat_102	36.9104	0.967851	**img_udiat_102**	34.5283	0.932152
img_udiat_114	36.8855	0.967821	**img_udiat_114**	34.1357	0.930100
img_udiat_135	36.9244	0.972911	**img_udiat_135**	33.3826	0.939381
img_udiat_165	38.8622	0.967638	**img_udiat_165**	34.3925	0.922628
img_udiat_200	37.9759	0.961544	**img_udiat_200**	33.7251	0.918583
Total average	38.1873	0.961547	**Total average**	33.0068	0.919955

Table 4. Visual comparison between original ultrasound RoI images and denoising images generated by Conditional GAN and WGAN.

ID	Original	CGAN PSNR/SSIM	WGAN PSNR/SSIM
img_busi_34		 40.18 dB / 0.9789	 34.35 dB / 0.9535

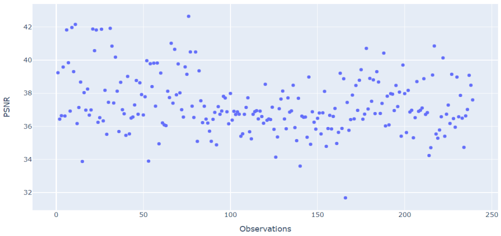
img_busi_70		 39.78 dB / 0.9717	 35.77 dB / 0.9439
img_busi_175		 39.40 dB / 0.9727	 35.54 dB / 0.9423
img_datasetA_6		 41.82 dB / 0.9776	 38.28 dB / 0.9655

img_datasetA_11		 42.15 dB / 0.9777	 38.29 dB / 0.9678
img_datasetA_76		 41.90 dB / 0.9776	 38.42 dB / 0.9712
img_udiat_77		 38.86 dB / 0.9676	 36.41 dB / 0.9399

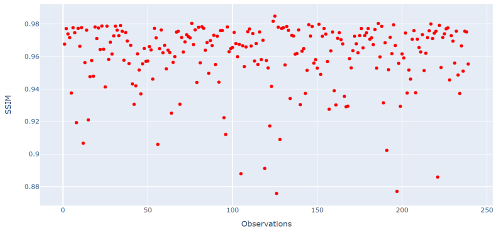
img_ udiat _165		 40.49 dB / 0.9672	 33.38 dB / 0.9393
img_ udiat _200		 37.97 dB / 0.9615	 33.72 dB / 0.9185

Although they are visually very similar according to Table 4, the quality values obtained define that the CGAN network achieves a higher mean value in PSNR=41.03 dB and SSIM=0.97 concerning the WGAN network values (PSNR=35.47 dB/ SSIM= 0.43).This indicates that the CGAN model is the network that best eliminates the speckle noise in ultrasound images while preserving the structural details and quality better than the WGAN model. Furthermore, we can see from Table 5 that the best visual results correspond mainly to dataset A, whose original images had the lowest resolution compared to the other datasets.

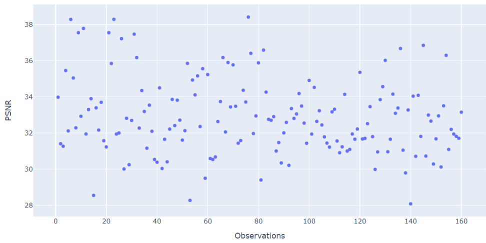
To confirm the previous information, the test dataset (239 US images) was used to evaluate the data dispersion of the CGAN and WGAN algorithms using the PSNR and SSIM metrics.



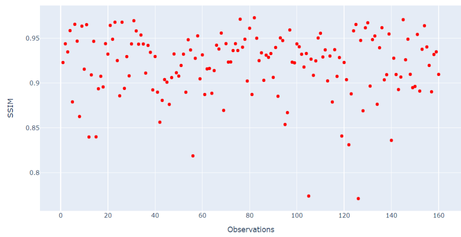
a.



b.



c.



d.

Fig. 4 Dispersion report for PSNR/SSIM metrics. a) CGAN network with PSNR metric. b). CGAN network with SSIM metric. c). WGAN network with PSNR metric. d. WGAN network with SSIM metric.

Figures 4a-4d show the statistical results obtained using R software, where a and b show the dispersion data obtained by CGAN. The blue points represent the PSNR metric, which ranges from 30 to 40 dB, and the red points represent the SSIM metric, which ranges from 0 to 1.

Figures 4a and 4b show more signal of better image quality using CGAN network, it means better luminance (PSNR 36-42dB/ SSIM 0.85 to 0.98), contrast and structural information in the restructured images by CGAN with respect to WGAN network (PSNR 36-48dB/ SSIM 0.85 to 0.95) Figures 4c and 4d.

4.Discussion

Ultrasound is a complementary technique to mammography and is used for breast cancer detection due to its sensitivity. However, the appearance of speckle noise in US is an interference mode that causes low contrast resolution [33], which makes it difficult to specialize in identifying abnormalities in the breast. In this paper, we trained a pair of GANs combined with CNN architectures as US image denoising, and then evaluated the quality of the denoised images using PSNR and SSIM metrics.

The quality of the denoising image in the Conditional GAN achieved a higher average PSNR (41.03 dB) and SSIM (0.97) in contrast to the average PSNR (35.47 dB) and SSIM (0.93) in the WGAN. Thus, according to the values given in Table 4, the CGAN is consistent with a higher quality image [63] and achieves success in ultrasound denoising images compared to the WGAN. This can be attributed to the fact that CGAN uses the Unet architecture as the generator model and Binary Cross Entropy (BCE) as the loss function (in addition to the L1 loss) [67,68] to generate real images and provide greater robustness to the model. The Unet has an encoder-decoder network to reconstruct the despeckled image by extracting features from the noisy image to effectively enhance the image features and suppress some speckle noise during the encoding phase [69].

In contrast, WGAN uses Wasserstein distance and Resnet architecture as the generator model with gradient clipping as the loss function to achieve a 1-Lipschitz function. Although this network sometimes avoids the mode collapse problem, resulting in more stable training and less sensitivity to hyperparameter settings (because it is trained based on image distribution loss, rather than image pixel loss) [69], in this work the results generated by WGAN are not statistically significantly better than those generated by CGAN. For the previous reason Gulrajani et al. [70] proposed a WGAN with gradient penalty (GP) to replace the gradient clipping and to enforce Lipschitz continuity, which performs better and more stable training than WGAN with almost no hyperparameter setting.

These performance differences in performance observed between the CGAN and the WGAN will also help to better implement new tasks in a computer system for detection/diagnosis of benign or malignant breast lesions. The pre-processing steps such as denoising, super resolution, or data augmentation based on deep learning algorithms help to improve the performance and accuracy in terms of clinical relevance in detection, diagnosis, segmentation, or image classification using CNN algorithms.

The main advantage of using GAN algorithms are the quality of the new images produced and the ability to generalize beyond the boundaries of the original dataset to produce new patterns.

Consequently, many researchers have been proposed a deep residual network structure based on GAN networks for image denoising.

Zhang et al. [71] used GANs Unet-based architecture as ultrasound image denoising, with residual dense connectivity and weighted joint loss (GAN-RW) to overcome the limitations of traditional denoising algorithms. The results demonstrated that the noise level (PSNR=3.08% and SSIM 1.84%) was effectively removed by the method, image detail was better preserved, and the subjective visual effect was improved. Lan et al. [69] implemented a mixed-attention mechanism (MARU) with UNet model for real-time ultrasound image despeckle, using an encoder-decoder network to reconstruct the despeckled image by extracting features from the noisy image. Visual comparison shows that the proposed method outperforms the compared despeckling methods (SBF, SRAD, NML) in terms of speckle noise reduction and detail preservation.

The GAN-based combination methods have been applied to different tasks, and have achieved better results. For example, [72] proposed a conditional GAN using a WGAN as an objective loss function in medical image denoising, the PSNR/SSIM values (29.4/0.88) demonstrated good results with respect to other state-of-the-art methods, perceiving the structure and details of the images.

Cantero J. [73] investigated two GANs (DCGAN and WGAN-GP) for the generation of synthetic PET (positron emission tomography) breast images. The visual results show that these two architectures can generate sinogram images that confound human evaluators. According to [74] the lower the amount of noise present in the real images the faster the DCGAN network learns to generate high fidelity images, but the results obtained here by WGAN-GP are not significantly better than those produced by DCGAN. In conclusion joint training of denoising and image classification significantly improves the performance of classification. A comparison of the accuracy of our work with more recent methods is shown in Table 6.

Table 5. Comparison of the accuracy of our denoising method with others GAN and CNN denoising methods.

Author	Method	Main Idea	PSNR /SNR(dB)	SSIM	Acc/Sen/ Spec (%)
Eckert et al. [11]	MRDGet	DL method based on CNNs for mammogram denoising to improve the image quality.	36.18	0.841	-
Feng et al. [13]	VGGNet	The network extracts the structure boundaries before and after US image de-speckling.	30.57	0.90	-
Pang et al. [32]	TripleGAN	Method to perform data augmentation in breast US images.	-	-	90.41/ 87.94/

		Then its images are used to classify breast masses using a CNN.			85.86
Al-Dhabyani et al. [33]	AlexNet+ GAN	US breast masses classification with data augmentation.			99/-/-
Vimala et al.[47]	Recurrent Neural Network	Hybrid deep learning technique to remove local speckle noise from breast US images.	70/ 65	-	-
Li et al. [72]	CGAN	WGAN loss are combined as the objective loss function to ensure the consistency of denoised image (lung and chest) and real image.	33.26	0.92	
Huang, et al. [76]	DUGAN +UNET	Deep learning-based model for Low-dose CT denoising	34.6	0.91	-
Elhoseny and Shankar [77]	CNN	Edge preservation and effective noise removal in MRI and CT images. Then, CNN classifier is used to classify the denoised image as normal or abnormal.	47.52	0.95	-
Ours	WGAN	Reduce speckle noise while preserving features and details in breast US images.	33.00	0.92	-
	CGAN		38.18	0.96	-

Finally, in this study, some limitations were presented, particularly in the availability of private data collection, because only public breast ultrasound databases were used. The implementation of hyperparameters in GAN training is very complex due to the sensitivity of their modification, generating some challenges (collapse mode, convergence, Nash equilibrium, and gradient), which are typical of generative networks. To minimize this problem during the training, it is essential to manually modify some hyperparameters (optimization functions, loss functions, number of epochs, layers, iterations), even to implement new alternatives based on deep convolutional networks to train the generator and the discriminator in a better way.

The research is reproducible, replicable and generalizable, and all code, data and materials have been deposited in the Mendeley repository [75], where the information can be accessed and used by others.

Conclusions

In conclusion, in this work CGAN proved to be a useful tool with a better-quality result for denoising breast ultrasound images than the WGAN model. This was obtained by comparing the mean statistical values (PSNR and SSIM) of the GAN models. The higher robustness demonstrated by CGAN is attributed to the fact that the generator uses U-Net encoder-decoder architecture with BCE loss function to remove the speckle noise in a better way than the Resnet architecture used in WGAN. The proposed CGAN technique is particularly useful for small data sets with low variance. These networks are widely used for image generation or data augmentation, but their application in US image denoising is still limited. In future work, other advanced deep learning methods for denoising such as convolutional neural networks and autoencoders will be used, and additional features will be considered in denoising breast images such as PET, thermal, CT, MRI to improve the performance of breast lesion classification algorithms.

Author Contributions: Conceptualization Y.J.-G. and V.L.; methodology Y.J.-G.; formal analysis, Y.J.-G., M.J.R.-Á., and V.L.; investigation Y.J.-G and O.V; resources, D.C, Y.S, L.E, A.S, C.S; writing original draft preparation Y.J.-G, , O.V; writing manuscript and editing, Y.J.-G., M.J.R.-Á., and V.L.; visualization, Y.J.-G.; supervision, M.J.R.-Á. and V.L.; project administration, M.J.R.-Á. and V.L.; funding acquisition, M.J. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data that support the findings of this study are openly available in the Mendeley repository (https://data.mendeley.com/drafts/g3cmj46xyx) [75]

Abbreviations:

BUSI	Breast Ultrasound Images Dataset
BCE	Binary cross entropy
CT	Computer Tomography
CGAN	Conditional GAN
CNN	Convolutional neural network
CNR	Contrast to-noise ratio
D	Discriminator
GAN	Generative adversarial network
G	Generator
JS	Jensen Shannon
KL	Kullback–Leibler
KID	Kernel inception distance
MRI	Magnetic Resonance Image
MSE	Mean Square Error
PET	Positron Emission Tomography

PSNR	Peak Signal-to-Noise Ratio
RMSE	Root-Mean-Square Error
SNR	Signal-to-Noise Ratio
SSIM	Structural Similarity Index
ReLu	Rectified Linear Unit
UDIAT	Diagnostic Centre of the Parc Tauli Corporation
US	Ultrasound
WGAN	Wasserstein GAN

3.1 References

1. Satoh, Y., Imokawa, T., Fujioka, T. et al. Deep learning for image classification in dedicated breast positron emission tomography (dbPET). Ann Nucl Med 36, 401–410 (2022). https://doi.org/10.1007/s12149-022-01719-7

2. Park, E. K., Lee, K. S., Seo, B. K., Cho, K. R., Woo, O. H., Son, G. S., ... & Chang, Y. W. (2019). Machine learning approaches to radiogenomics of breast cancer using low-dose perfusion computed tomography: predicting prognostic biomarkers and molecular subtypes. Scientific reports, 9(1), 17847.

3. Ji, Y., Li, H., Edwards, A. V., Papaioannou, J., Ma, W., Liu, P., & Giger, M. L. (2019). Independent validation of machine learning in diagnosing breast Cancer on magnetic resonance imaging within a single institution. Cancer Imaging, 19, 1-1

4. Salama, W. M., & Aly, M. H. (2021). Deep learning in mammography images segmentation and classification: Automated CNN approach. Alexandria Engineering Journal, 60(5), 4701-4709.

5. Xu, Y., Wang, Y., Yuan, J., Cheng, Q., Wang, X., & Carson, P. L. (2019). Medical breast ultrasound image segmentation by machine learning. Ultrasonics, 91, 1-9.

6. Szabo, T. L. (2004). Diagnostic ultrasound imaging:Inside Out. Elseivier Acdemin Press. ISBN: 0-12-680145-2

7. Tole, N. M. (2005). Basic physics of ultrasonographic imaging. Essential Health Technologies. Health Technology and Pharmaceuticals WORLD HEALTH ORGANIZATION.Geneva.ISBN 92 41592990

8. Wang S, Huang TZ, Zhao XL, Mei JJ, Huang J. Speckle noise removal in ultrasound images by first-and second-order total variation. Numerical Algorithms. 2018;78(2):513–33.

9. Kaji S, Kida S. Overview of image-to-image translation by use of deep neural networks: Denoising, super-resolution, modality conversion, and reconstruction in medical imaging. Igaku Butsuri [Internet]. 2020;40(4):139. Disponible en: http://dx.doi.org/10.11323/jjmp.40.4_139

10. Njeh I, Sassi OB, Chtourou K, Ben Ha Mida A. Speckle noise reduction in breast ultrasound images: SMU (SRAD median unsharp) approch. En: Eighth International Multi-Conference on Systems, Signals & Devices. IEEE; 2011.

11. Eckert D, Vesal S, Ritschl L, Kappler S, Maier A. Deep learning-based denoising of mammographic images using physics-driven data augmentation. En: Informatik aktuell. Wiesbaden: Springer Fachmedien Wiesbaden; 2020. p. 94–100.

12. Chen, H., Zhang, Y., Kalra, M. K., Lin, F., Chen, Y., Liao, P., ... & Wang, G. Low-dose CT with a residual encoder-decoder convolutional neural network. IEEE transactions on medical imaging. 2017;36(12), 2524-2535.

13. Feng, X., Huang, Q., & Li, X. Ultrasound image de-speckling by a hybrid deep network with transferred filtering and structural prior. Neurocomputing. 2020;414, 346-355.

14. Ilesanmi AE, Ilesanmi TO. Methods for image denoising using convolutional neural network: a review. Complex Intell Syst [Internet]. 2021;7(5):2179–98. Disponible en: http://dx.doi.org/10.1007/s40747-021-00428-4

15. Kang, E., Koo, H. J., Yang, D. H., Seo, J. B., & Ye, J. C. Cycle-consistent adversarial denoising network for multiphase coronary CT angiography. Medical physics. 2019;46(2), 550-562.

16. Li P, Li Z, Pang X, Wang H, Lin W, Wu W. Multi-scale residual denoising GAN model for producing super-resolution CTA images. J Ambient Intell Humaniz Comput [Internet]. 2022;13(3):1515–24. Disponible en: http://dx.doi.org/10.1007/s12652-021-03009-y

17. Yang, Q., Yan, P., Zhang, Y., Yu, H., Shi, Y., Mou, X., ... & Wang, G. Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. IEEE transactions on medical imaging, . 2018;37(6), 1348–57.

18. Ahmed AS, El-Behaidy WH, Youssif AA. Medical image denoising system based on stacked convolutional autoencoder for enhancing 2-dimensional gel electrophoresis noise reduction. Biomedical Signal Processing and Control. 2021;69.

19. Daoud MI, Saleh A, Hababeh I, Alazrai R. Content-based image retrieval for breast ultrasound images using convolutional autoencoders: A feasibility study. En: 2019 3rd International Conference on Bio-engineering for Smart Technologies (BioSMART). IEEE; 2019.

20. Ghosh, S. K., Biswas, B., & Ghosh, A. A novel stacked sparse denoising autoencoder for mammography restoration to visual interpretation of breast lesion. Evolutionary Intelligence. 2021;14(1), 133–49.

21. Jiménez YG, Rodríguez-Álvarez MJ, Freire J, Castillo D, Lakshminarayanan V. Preprocessing fast filters and mass segmentation for mammography images. International Society for Optics and Photonics. 2021;11842.

22. Lore, K. G., Akintayo, A., & Sarkar, S. LLNet: A deep autoencoder approach to natural low-light image enhancement. Pattern Recognition. 2017;61, 650-662.

23. Chen X, Shen Q. Medical image denoising based on dictionary learning. Biomedical Research. 2017;(20).

24. Huang J, Yang X. Fast reduction of speckle noise in real ultrasound images. Signal Processing [Internet]. 2013;93(4):684–94. Disponible en: http://dx.doi.org/10.1016/j.sigpro.2012.09.005

25. Kohan MN, Behnam H. Denoising Medical Images using Calculus of Variations. J Med Signals Sens [Internet]. 2011;1(3):184–90. Disponible en: http://dx.doi.org/10.4103/2228-7477.95413

26. Njeh, I., Sassi, O. B., Chtourou, K., & Mida, A. B. H. (2011). Speckle noise reduction in breast ultrasound images: SMU (SRAD median unsharp) approch. In Eighth International Multi-Conference on Systems, Signals & Devices (pp. 1-6). IEEE.

27. Dass, R. (2018). Speckle noise reduction of ultrasound images using BFO cascaded with wiener filter and discrete wavelet transform in homomorphic region. Procedia computer science, 132, 1543-1551.

28. Beevi, A. S., & Ratheesha, S. (2021, March). Speckle Noise Removal Using Spatial and Transform Domain Filters in Ultrasound Images. In 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS) (Vol. 1, pp. 291-297). IEEE.

29. Pradeep, S., & Nirmaladevi, P. (2021, February). A review on speckle noise reduction techniques in ultrasound medical images based on spatial domain, transform domain and CNN methods. In IOP Conference Series: Materials Science and Engineering (Vol. 1055, No. 1, p. 012116). IOP Publishing.

30. Li, P., He, F., Fan, B., & Song, Y. (2023). TPNet: A Novel Mesh Analysis Method via Topology Preservation and Perception Enhancement. Computer Aided Geometric Design, 102219.

31. Wu, H., He, F., Duan, Y., & Yan, X. (2022). Perceptual metric-guided human image generation. Integrated Computer-Aided Engineering, 29(2), 141-151.

32.Pang, T., Wong, J. H. D., Ng, W. L., & Chan, C. S. (2021). Semi-supervised GAN-based radiomics model for data augmentation in breast ultrasound mass classification. Computer Methods and Programs in Biomedicine, 203, 106018.

33. Al-Dhabyani, W., Gomaa, M., Khaled, H., & Aly, F. (2019). Deep learning approaches for data augmentation and classification of breast masses using ultrasound images. Int. J. Adv. Comput. Sci. Appl, 10(5), 1-11.

34. Jain V, Seung S. Natural image denoising with convolutional networks. En: Advances in neural information processing systems. 2008.

35. Wickramaratne SD, Mahmud MS. Conditional-GAN based data augmentation for deep learning task classifier improvement using fNIRS data. Front Big Data [Internet]. 2021;4:659146. Disponible en: http://dx.doi.org/10.3389/fdata.2021.659146

36. Wu E, Wu K, Cox D, Lotter W. Conditional infilling GANs for data augmentation in mammogram classification. En: Image Analysis for Moving Organ, Breast, and Thoracic Images. Cham: Springer International Publishing; 2018. p. 98–106.

37. Zhou Z, Wang Y, Guo Y, Qi Y, Yu J. Image quality improvement of hand-held ultrasound devices with a two-stage generative adversarial network. IEEE Transactions on Biomedical Engineering. 2019 Apr 24;67(1):298-311.

38. Bargsten L, Schlaefer A. SpeckleGAN: a generative adversarial network with an adaptive speckle layer to augment limited training data for ultrasound image processing. International journal of computer assisted radiology and surgery. 2020 Sep;15:1427-36.

39. Khor HG, Ning G, Zhang X, Liao H. Ultrasound speckle reduction using wavelet-based generative adversarial network. IEEE Journal of Biomedical and Health Informatics. 2022 Jan 25;26(7):3080-91.

40. Mishra D, Chaudhury S, Sarkar M, Soin AS. Ultrasound image enhancement using structure oriented adversarial network. IEEE Signal Processing Letters. 2018 Jul 23;25(9):1349-53.

41. Carrara F, Amato G, Brombin L, Falchi F, Gennaro C. Combining GANs and AutoEncoders for efficient anomaly detection. En: 2020 25th International Conference on Pattern Recognition (ICPR). IEEE; 2021.

42. Yao, Y., Wang, X., Ma, Y., Fang, H., Wei, J., Chen, L., ... & Braytee, A. (2022, October). Conditional Variational Autoencoder with Balanced Pre-training for Generative Adversarial Networks. In 2022 IEEE 9th International Conference on Data Science and Advanced Analytics (DSAA) (pp. 1-10). IEEE.

43. Kim, H. J., & Lee, D. (2020). Image denoising with conditional generative adversarial networks (CGAN) in low dose chest images. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 954, 161914.

44. Zou, X., Zhu, D., Huang, J., Lu, W., Yao, X., & Lian, Z. (2022). WGAN-Based Image Denoising Algorithm. Journal of Global Information Management (JGIM), 30(9), 1-20.

45.Singh, V. K., Romani, S., Rashwan, H. A., Akram, F., Pandey, N., Sarker, M. M. K., ... & Puig, D. (2018). Conditional generative adversarial and convolutional networks for X-ray breast mass segmentation and shape classification. In Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11 (pp. 833-840). Springer International Publishing.

46. Zhang, Y., Chen, H., & Ke, W. (2022). Medical image denoising. In Biomedical Image Synthesis and Simulation (pp. 255-278). Academic Press.

47.Vimala, B.B.; Srinivasan, S.; Mathivanan, S.K.; Muthukumaran, V.; Babu, J.C.; Herencsar, N.; Vilcekova, L. Image Noise Removal in Ultrasound Breast Images Based on Hybrid Deep Learning Technique. Sensors 2023, 23, 1167. https://doi.org/10.3390/s23031167

48.Khaledyan, D., Marini, T. J., M. Baran, T., O'Connell, A., & Parker, K. (2023). Enhancing breast ultrasound segmentation through fine-tuning and optimization techniques: Sharp attention UNet. Plos one, 18(12), e0289195.

49.Zama, S., Fujioka, T., Yamaga, E., Kubota, K., Mori, M., Katsuta, L., ... & Tateishi, U. (2024). Clinical Utility of Breast Ultrasound Images Synthesized by a Generative Adversarial Network. Medicina, 60(1), 14.

50. Li, M., Jiang, Y., Zhang, Y., & Zhu, H. (2023). Medical image analysis using deep learning algorithms. Frontiers in Public Health, 11, 1273253.

51.Ilesanmi, A. E., & Ilesanmi, T. O. (2021). Methods for image denoising using convolutional neural network: a review. Complex & Intelligent Systems, 7(5), 2179-2198.

52. Al-Dhabyani W, Gomaa M, Khaled H, Fahmy A. Dataset of breast ultrasound images. Data Brief [Internet]. 2020;28(104863):104863. Disponible en: http://dx.doi.org/10.1016/j.dib.2019.104863

53. Rodrigues PS. Breast Ultrasound Image [Internet]. Mendeley; 2017. Disponible en: http://dx.doi.org/10.17632/wmy84gzngw.1

54. Yap MH, Pons G, Marti J, Ganau S, Sentis M, Zwiggelaar R, et al. Automated breast ultrasound lesions detection using convolutional neural networks. IEEE J Biomed Health Inform [Internet]. 2018;22(4):1218–26. Disponible en: http://dx.doi.org/10.1109/jbhi.2017.2731873

55. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2020). Generative adversarial networks. Communications of the ACM, 63(11), 139-144.

56. Gong M, Chen S, Chen Q, Zeng Y, Zhang Y. Generative adversarial networks in medical image processing. Curr Pharm Des [Internet]. 2021;27(15):1856–68. Disponible en: http://dx.doi.org/10.2174/1381612826666201125110710

57. Abusitta, Adel & Aimeur, Esma & Wahab, Omar. (2020). Generative Adversarial Networks for Mitigating Biases in Machine Learning Systems.

58. Douzas, G., & Bacao, F. (2018). Effective data generation for imbalanced learning using conditional generative adversarial networks. Expert Systems with applications, 91, 464-471.

59. Yu, Y., Gong, Z., Zhong, P., & Shan, J. (2017). Unsupervised representation learning with deep convolutional neural network for remote sensing images. In Image and Graphics: 9th International Conference, ICIG 2017, Shanghai, China, September 13-15, 2017, Revised Selected Papers, Part II 9 (pp. 97-108). Springer International Publishing.

60. Isola P, Zhu J-Y, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. En: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; 2017.

61. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. En: Lecture Notes in Computer Science. Cham: Springer International Publishing; 2015. p. 234–41.

62. Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks. En: International conference on machine learning. PMLR; 2017. p. 214–23.

63. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. En: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; 2016.

64. Obukhov A, Krasnyanskiy M. Quality assessment method for GAN based on modified metrics inception score and Fréchet inception distance. En: Software Engineering Perspectives in Intelligent Systems. Cham: Springer International Publishing; 2020. p. 102–14.

65. A comparative analysis on image quality assessment for real time satellite images. Indian J Sci Technol, . 2016;9(34), 1–11.

66. Rajkumar, S., & Malathi, G. (2016). A comparative analysis on image quality assessment for real time satellite images. Indian J. Sci. Technol, 9(34).

67. Martinez MT, Heiner ON. Conditional Generative Adversarial Networks for Solving Heat Transfer Problems (No. SAND-2020-10569). En: Sandia National Lab(SNL-NM). Albuquerque, NM (United States; 2020.

68. Mohammadi N, Doyley MM, Cetin M. Regularization by adversarial learning for ultrasound elasticity imaging. En: 2021 29th European Signal Processing Conference (EUSIPCO). IEEE; 2021.

69. Lan, Y., & Zhang, X. (2020). Real-time ultrasound image despeckling using mixed-attention mechanism based residual UNet. IEEE Access, 8, 195327-195340.

70. Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville AC. Improved training of wasserstein gans. Advances in neural information processing systems. 2017.

71. Zhang L, Zhang J. Ultrasound image denoising using generative adversarial networks with residual dense connectivity and weighted joint loss. PeerJ Comput Sci [Internet]. 2022;8:e873. Disponible en: http://dx.doi.org/10.7717/peerj-cs.873

72. Li, Y., Zhang, K., Shi, W., Miao, Y., & Jiang, Z. (2021). A Novel Medical Image Denoising Method Based on Conditional Generative Adversarial Network. Computational and Mathematical Methods in Medicine, 2021, 1-11.

73. Cantero Lorenzo J. A GAN approach to synthetic PET imaging generation for breast cancer diagnosis. [España]: Universitat Oberta de Catalunya; 2021 [citado el 25 de enero de 2023]. Disponible en: https://github.com/jcantlord/PET-SinoGAN

74. Lei, Y., Zhang, J., & Shan, H. (2021). Strided self-supervised low-dose CT denoising for lung nodule classification. Phenomics, 1, 257-268.

75. Jimenez, Yuliana; Rodriguez-Alvarez, Maria Jose; Escudero, Lider; Sandoval, Carlos; Lakshminarayanan, Vasudevan (2023), "Ultrasound Breast images denoising using Generative Adversarial Networks (GANs)", Mendeley Data, V1, doi: 10.17632/g3cmj46xyx.1

76. Huang Z, Zhang J, Zhang Y, Shan H. DU-GAN: generative adversarial networks with dual-domain U-net-based discriminators for low-dose CT denoising. IEEE Trans Instrum Meas. (2021) 71:1–12. doi: 10.1109/TIM.2021.3128703

77. Elhoseny M, Shankar K (2019) Optimal bilateral flter and convolutional neural network based denoising method of medical image measurements. Measurement 143:125–135

Chapter 5

GAN-BASED DATA AUGMENTATION TO IMPROVE BREAST ULTRASOUND AND MAMMOGRAPHY MASS CLASSIFICATION

Yuliana Jiménez-Gaona, Diana Carrión-Figueroa, Vasudevan Lakshminarayanan María José Rodríguez-Álvarez

Biomedical Signal Processing and Control, Volume 94, 106255,

https://doi.org/10.1016/j.bspc.2024.106255

Abstract

Data imbalance is a common problem in breast cancer diagnosis, to address this challenge, the research explores the use of Generative Adversarial Networks (GANs) to generate synthetic medical data. Various GAN methods, including Wasserstein GAN with Gradient Penalty (WGAN-GP), Cycle GAN, Conditional GAN, and Spectral Normalization GAN (SNGAN), were tested for data augmentation in breast regions of interest (ROIs) using mammography and ultrasound databases. The study employed real, synthetic, and hybrid ROIs (128x128 pixels) to train a Resnet network for classifying as benign (B) or malignant (M) classes. The quality and diversity of the synthetic data were assessed using several metrics: Fréchet Inception Distance (FID), Kernel Inception Distance (KID), Structural Similarity Index (SSIM), Multi-Scale SSIM (MS-SSIM), Blind Reference Image Spatial Quality Evaluator (BRISQUE), Naturalness Image Quality Evaluator (NIQE), and Perception-based Image Quality Evaluator (PIQE).Results revealed that the SNGAN model (FID=52.89) was most effective for augmenting mammography data, while CGAN (FID=116.03) excelled with ultrasound data. Cycle GAN and WGAN-GP, though demonstrating lower KID values, did not perform better than SNGAN and CGAN. The lower average MS-SSIM values suggested that SNGAN and CGAN produced a high diversity of synthetic images. However, lower SSIM, BRISQUE, NIQE, and PIQE values indicated poor quality in both real and synthetic images. Classification results showed high accuracy without data augmentation in both US (93.1%B/94.9%M) and mammography (80.9%B/76.9%M). The research concludes that preprocessing and characterizing ROIs by abnormality type is crucial to generate diverse synthetic data and improve accuracy in the classification process using combined GANs and CNN models.

Key words: Breast cancer; Data augmentation; Deep learning algorithms; Generative Adversarial Networks (GAN); Mammography; Ultrasound.

1. Introduction

The timely and accurate diagnosis of breast cancer plays a central role in reducing mortality rates associated with this disease. Among the available diagnostic tools, digital mammography (DM) and ultrasound (US) imaging have emerged as crucial modalities for breast cancer screening and detection. These techniques offer several advantages over other imaging modalities such as MRI, PET, and CT, including non-invasiveness, lower radiation dose, lower cost, and greater accessibility, particularly in regions like Latin America, where economic and social factors can limit access to advanced imaging technologies [1].

While digital mammography and US are complementary imaging methods, US serves as a valuable supplementary modality, especially in cases where the presence of dense breast tissue may limit mammograms. However, digital mammography has inherent limitations when imaging dense breast tissue, characterized by a higher proportion of fibrous and glandular tissue and less fatty tissue. This composition results in a similar contrast between anomalies and surrounding tissue, potentially leading to difficulty in detecting tumors [2].

Consequently, analyzing breast medical images, particularly US images, presents clinical challenges. The manual assessment of morphological variations in anomalies poses difficulties in interpreting their malignant potential, contributing to increased false positives [3]. Furthermore, this manual process is time-consuming and labor-intensive.

To overcome these challenges, Computer-Aided Diagnosis (CAD) systems have emerged as valuable tools to assist radiologists in breast lesion segmentation and classification. However, the efficacy of these systems is hindered by limited access to comprehensive and well-balanced medical databases [4]. This limitation restricts the utilization of state-of-the-art Deep Learning (DL) algorithms, which often require large datasets to overcome the overfitting problem. Recent studies [3-5] have highlighted the potential of synthetic medical images generated through Convolutional Neural Networks (CNNs) to enhance image classification accuracy and early detection of breast lesions.

Various techniques have been proposed to mitigate overfitting in DL models, including dropout, normalization, transfer learning, and data augmentation. Data augmentation involves applying affine transformations, such as flipping, scaling, translating, rotating, cropping, blurring, and sharpening, to augment the size and diversity of the dataset [5]. However, these techniques are limited to generating synthetic images that mimic the original data distribution and do not effectively address the need for novel training data to enhance DL algorithm performance [6].

Researchers have focused on medical image data augmentation to address these limitations, particularly by leveraging the power of Generative Adversarial Networks (GANs) [2]. GANs have demonstrated promising results in generating realistic synthetic images and improving classification performance. Data augmentation using GANs holds promise as an effective oversampling solution and a valuable tool to alleviate overfitting during the training process, as evidenced by the accuracy improvements observed during each epoch of the validation phase [7,8].

However, training GANs for medical image data augmentation poses its own set of challenges. Convergence issues arise when the discriminator (D) outperforms the generator (G), making it difficult for D to distinguish between natural and synthetic samples, especially in high-dimensional images. The collapse mode may also occur, wherein the generator produces a limited variety of new data, hindering the learning process. Another challenge is gradient vanishing, where the generator becomes highly proficient, impeding the discriminator's ability to improve its discrimination between real and fake images. This phenomenon is characterized by a slight gradient decrease, which prevents effective weight adjustments, resulting in the neural network halting further training [9].

Alternative approaches can be explored to overcome these challenges and enhance the performance of GANs. These include (i) different hyperparameters during the training model: number of epochs and iterations, learning rate, optimization, and loss functions. (ii) alternative normalization techniques to that first used by Goodfellow et al., [10] such as Gradient penalty, Spectral Normalization, Batch normalization, and Least Squares, or other techniques [9].

1.1Related work

GANs, introduced by [10], consist of a generator (G) and a discriminator (D) that are trained adversarial using neural networks (Figure 1). The generator G generates synthetic images by taking random noise z as input, while the discriminator D classifies input binary samples x as real (1) or fake (0). The optimization of GANs involves minimizing the objective function V(G, D), which is a min-max equation. (Eq. 1).

$$min_G \, max_D \, V(D, G) = E_{x \sim p_{data\,(x)}}[log \, log \, D(x)\,] + \, E_{z \sim p_{z\,(z)}}[log \, log \, (1 - D(G(z)))\,]$$
(1)

$$min_G \, max_D \, V(D, G) = E_{x \sim p_{data\,(x)}}[log \, log \, D(x)\,] + \, E_{z \sim p_{z\,(z)}}[log \, log \, (1 - D(G(z)))\,]$$

The generator aims to minimize the objective function, whereas the discriminator aims to maximize it [9]. The goal is to make the distribution of the generated images (G) and the real image distribution *Pdata(x)* similar.

There are various GAN architectures that have been used in different applications such as object detection, super-resolution, image restoration, and medical data augmentation for solving the problem of the small dataset [11,12]. These architectures (figure 1) include Progressive GAN [13], Style GAN [14], MSGAN [15], Vanilla GAN[16], Deep Convolutional GAN (DCGAN) [17], Conditional GAN[18-19]; Wasserstein-WGAN[20], WGAN-GP[21], Alpha-GAN[22], Cycle GAN [23].

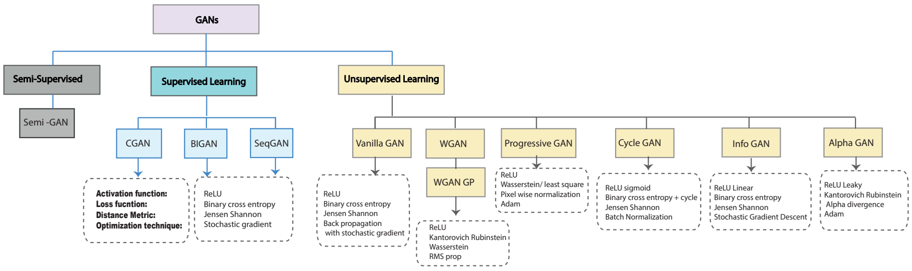


Fig 1. GAN architecture classification and its hyperparameters.

Baur et al. [24] investigated the differences between several GAN architectures and concluded that GANs are useful for mitigating imbalances dataset classes, warning of visually noticeable differences between generated and real images. However, Nielsen and Okoniewski [25] demonstrated that the benefit of using GANs to augment datasets is also for improving image classification tasks. The GAN model provides many solutions for general augmentation of medical images such as breast US [26,27] and breast mammography [28]. In breast mammography, Lin et al. [29] proposed the *blen-GAN* and Shen et al [30] the *Nc-GAN* as synthetic image generation models. Korkinof et al. [31] used *Progressive GAN* to synthesize mammograms (up to 1280x1024 pixels, the highest resolution achieved), enabling visualizations within standard mammographic protocols.

Escobar et al. [32] introduced a modified version of the GAN called the *UltraGAN*; the model was used to increase the number of US images using quality transfer while preserving structural information. Fujioka et al. [33] used a deep convolution generative adversarial network (*DCGAN*) in US generation to enhance the segmentation process and increase the number of US images. The *Speckle GAN* [34] was used as a novel architecture for augmenting synthetic ultra-sounds with speckle noise.

A modified version of GAN called stacked generative adversarial networks (*StackGAN*) [35] was also utilized to generated new samples of images from a smaller dataset. Despite its ability to generate high-resolution images, particularly US and brain images, it has several disadvantages, including the tendency for some images to be incoherent and the possibility of mode collapse. And *vanilla GANs* suffer from the training instability problem which is hard to make the model converge.

Figure 2 displays the general scheme of these architectures that are going to be used in this paper.

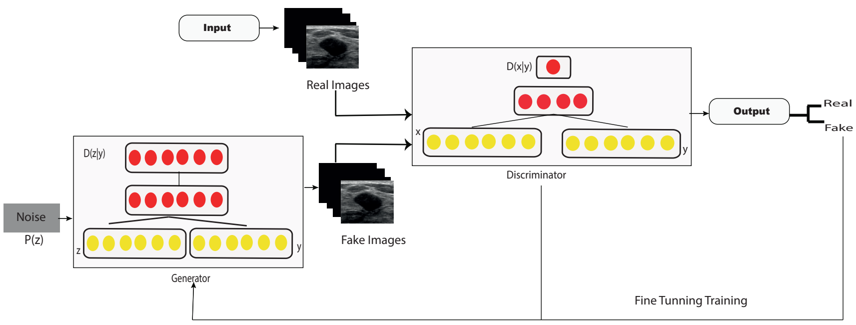


Fig 2. GAN architecture, where P(z)=random noise, x=input variable, y= real sample, G(z|y) = generated sample, D(x|y)= prediction label. The generator takes the noise vector 'P(z)' along with a condition variable 'y' and generates a sample 'G(z|y)' which is passed to a discriminator network, and it predicts whether the given sample is real or not.

During GAN training, the overall distance between the real and generated distributions plays a crucial role in evaluating the difference between the two distributions. Commonly used measures to calculate this distance include Jensen Shannon (JS) divergence, Kullback-Leibler (KL) divergence, Wasserstein distance (Earth Moving Distance), and Hellinger (f-divergence) [36,37]. Previous studies have shown that regular GANs using JS divergence based on KL divergence are not optimal [38].

To address this issue, WGAN was introduced by Arjovsky et al. [20] with a new objective function based on Wasserstein distance [21], which is derived from Kantorovich-Rubinstein duality. This approach offers better convergence performance compared to JS divergence. Even when two distributions do not overlap, Wasserstein distance can still reflect their distance. However, WGAN utilizes weight clipping as a technique to mitigate exploding gradients, but it leads to optimization difficulties. To enforce the Lipschitz constraint, its gradient clipping must be mathematically consistent (i.e., 1-Lipschitz continuous). The norm of the gradient must be less than or equal to 1 for all points in the function $f(x)$ (Eq. 2).

$$||\nabla f(x)||_2 \leq 1 \qquad ||\nabla \mathrm{f}(x)||_2 \leq 1 \tag{2}$$

To achieve this consistency can result in better samples or convergence failure. To overcome this problem, an alternative approach called WGAN-GP (Wasserstein GAN with gradient penalty) was proposed [21] to achieve Lipschitz continuity. This technique involves the inclusion of a gradient penalty in the discriminator's loss function, which improves network optimization, particularly in the distribution between real and fake samples. Other alternative loss functions for stabilizing GAN training include LS (Least Square) and SN (Spectral Normalization). Cycle GAN with spectral normalization and stacked GAN with Progressive GAN architecture are also used to generate high-quality images using multiple GANs consecutively [2,12].

Realistic data augmentation of public breast mammography and US datasets collection is essential to breast tumour segmentation and classification using DL algorithms. However, only some studies [39] focus on US and mammogram data augmentation.

Consequently, this research aims to (i) generate new synthetic breast ROIs (regions of interest) data using four modern GAN frameworks (WGAN-GP, Cycle GAN, Conditional GAN and SNGAN) with different normalization techniques. (ii) Evaluate the ROIs synthetic quality and diversity using different statistical metrics. (iii) Classify the synthetic, real, and hybrid data using the Resnet18 model, and (iv) evaluate the classification performance model using different statistical metrics.

2. Methods and experiments

The workflow of this paper is organized as follows (Figure 3).

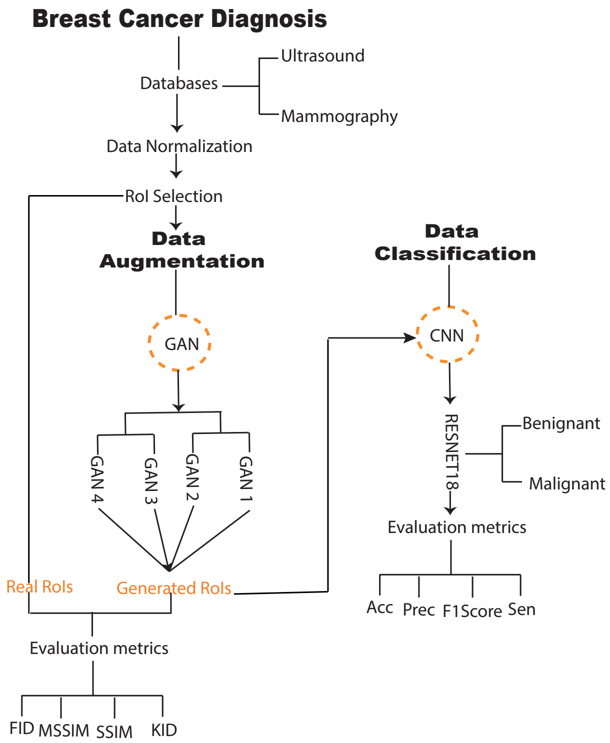


Fig 3. GAN-based workflow overview for breast ROIs data augmentation and classification: Step 1: Databases collection. Step 2: Data normalization and RoI mask extraction. Step 3: Data augmentation experiment settings using different GAN architectures. Step 3: RoIs classification using Resnet18 model and Step 4: Quantitative evaluation of the GAN and CNN models using several statistical metrics.

2.1 Databases and ROI Preprocessing

Six public breast databases (Table 1) were used to train all GAN models proposed in this research. The first three datasets contain US images, while the remaining datasets consist of mammography images. The details of the datasets are provided in Table 1.

(i) Breast Ultrasound Images Dataset (BUSI), collected by Al-Dhabyani et al. [3], contains 780 images (133 normal, 437 benign, and 210 malignant).

(ii) Dataset A, collected by Rodrigues et al. [40] (https://data.mendeley.com/datasets/wmy84gzngw/1), contains 250 breast cancer images, including 100 benign and 150 malignant.

(iii) Dataset B consists of 163 US images corresponding to 110 benign and 53 malignant breast masses. These data were acquired from the UDIAT Diagnostic Centre of the Parc Tauli Corporation, Sabadell, Spain [41].

(iv) CBIS-DDSM: Curated Breast Imaging Subset–Digital Database for Screening Mammography (https://n9.cl/qtl48) comprises 2620 cases, with 695 normal and 1925 abnormal mammograms [42].

(v) mini-MIAS:Mammographic Image Analysis Society (http://peipa.essex.ac.uk/info/mias.html), includes 322 Medio Lateral Oblique (MLO) mammograms from 161 patients, with 208 normal, 63 benign, and 51 malignant images [43,44].

(vi) Inbreast consists of a total of 115 cases. This dataset was not initially identified by classes, but an expert radiologist assisted in the respective identification of benign and malignant cases (https://biokeanos.com/source/INBreast) [45,46].

Table 1. US and mammography summary datasets.

Type	Database	Format	Resolution	Benignant	Malignant	Normal	Class	Total
Mammography	CBIS-DDSM	PNG	several	4683	5556	-	2	10239
	Mini-MIAS	PGM	1024x1024	63	52	207	3	322
	Inbreast	DCM	several	207	52	-	2	410
US	BUSI	PNG	500x 500	437	210	133	3	780
	Dataset A	BMP	several	100	150	109	3	250
	Dataset B	PNG	760x570	109	54	-	2	163

2.2 Data normalization

A total of 4,463 mammography and 1,060 benign and malignant US images were resized [42] to a common resolution (128px × 128px) prior to data enhancement. The torchvision (pytorch) library was used to perform transformations and standardise the images taken at different sizes to a single dimension (128x128x1 pixels). Finally, the cross-validation technique was used to randomly divide the data into 80% training set and 20% test set for training the GAN modes (using the Sklearn library).

Breast ROI annotation

Following the approach proposed by Wu et al. [18], lesion synthesis in ROIs and mask extraction was performed to enhance the generative model, eliminating the need to augment the entire image. ROI extraction was carried out automatically for the Mini-MIAS and Inbreast datasets, while the CBIS-DDSM dataset already contained annotations by the authors. RoIs were extracted from the BUSI dataset and Dataset B (UDIAT) for the US images, and the Dataset A already had annotated RoIs. Consequently, the GAN networks were trained using a total of 1,060 US RoIs and 4,463 mammography RoIs, as detailed in Table 2.

Table 2. US and Mammography ROIs.

Type	US/ DM	Benignant	Malignant
US	BUSI	437	210
	Dataset B	100	150
	BUS (UDIAT)	109	54
	Total 1060	646	414
DM	Mini-MIAS	118	91
	INbreast	2106	144
	CBIS-DDSM	1225	779
	Total 4463	3449	1014

2.3 Data augmentation and implementation details

In this experimental study, four GAN networks were trained to mitigate the limitations of GANs in generating synthetic breast ROIs. Different hyperparameters and loss functions for each architecture were chosen according to the references and works carried out in table 3. The specific equations (Eq. 3 to 6) and details can be found in Table 3.

Table 3.GAN loss variants.

Model	Loss Function	Equation
WGAN [16]	Wasserstein with gradient penalty	$min_G\ max_D\ V(D,G) = E_{x \sim p_{r\,(x)}}[D(x)] + E_{z \sim p_{z\,(z)}}[1 - D(G(z))]$ 3)

CGAN [28,29]	Binary Cross entropy with Logistic Loss	$min_G\ max_D\ V(D,G) = E_{x\sim p_{r\,(x)}}[log\ log\ D(x\|c)\ ] + E_{z\sim p_{g\,(z)}}[log\ log\ (1 - D(G(z\|c)))\ ]$	(4)
Cycle GAN [30]	Jensen Shannon divergence	$L(G,F,D_x,D_y) = L_{GAN}\left(G,D_y,X,Y\right) + L_{GAN}(F,D_x,X,Y) + \lambda L_{cyc}(G,F)$	(5)
SNGAN [31,32]	Hinge loss	$L_D = -E_{(x,y)\sim P_{data}}\left[\left(0,-1+D(x,y)\right)\right] - E_{z\sim P_{z,y\sim Pdata}}\left[\left(0,-1 - D(G(z),y)\right)\right] L_G = -E_{z\sim P_{z,y\sim Pdata}} D(G(z),y)$	6)

Conditional GAN (CGAN)

CGAN is an architecture that incorporates a specific loss function known as Binary Cross Entropy with Logistic Loss. During CGAN training, the conditioning information *y* is introduced as an additional input to both the generator (G) and the discriminator (D) [36].

The weights of the networks were initialized randomly from a normal distribution N (μ = 0, σ = 0.02). Detailed hyperparameters for both the US and digital mammography (DM) datasets can be found in Table 4.

Table 4. CGAN generator and discriminator hyperparameters in breast images

HIPERPARAMETER	US AND DM
OPTIMIZER	Adam
B1	0.3
B2	0.999 0.999
LATENT VECTOR	100

Learning rate	2e-4	
Batch size	64	
Optimization function (Discriminator)	LeakyReLU	

Cycle GAN. –

Cycle GAN is a model that operates on two image domains, denoted as x and y. It involves the use of two generators, G_{xy} and G_{yx}, as well as two discriminators, D_x and D_y. The purpose of G_{xy} is to perform image transfer from domain X to domain Y, while G_{yx} accomplishes the inverse transformation from Y to X. Each generator is associated with a corresponding discriminator, D_X and D_Y, which determine the domain of an image [23].

To assess the similarity between the probability distributions of the two domains, the Jensen-Shannon normalization method is employed. This method is based on the Kullback-Leibler (KL) divergence [33]. The configuration of hyperparameters used in the experiment is provided in detail in Table 5.

The generator model employed in this study was composed of three main sections. Firstly, an Encoder module was utilized, which consisted of three convolutional layers aimed at reducing the size of the input image at each layer. Secondly, a Transformer component was implemented, comprising nine residual blocks. The Transformer took the features generated by the Encoder layer and produced an output. Lastly, a Decoder module was employed, which consisted of several deconvolutional layers responsible for generating new images [47].

On the other hand, the discriminator model followed a different architecture. It comprised a 4x4 convolutional layer with the LeakyReLU activation function, which was subsequently followed by a sigmoid function. The purpose of the discriminator was to predict whether the input images were real or synthetic.

For further reference, Table 5 presents the specific hyperparameters employed for the Cycle GAN generator and discriminator in the context of breast images.

Table 5. Cycle GAN generator and discriminator hyperparameters in breast images

Hyperparameters	**Values**
Epochs number	100
Batch size	64

Learning rate	0.0002 (G) 0.0002 (D)
Image size	256 x256
Optimizator	Adam
Activation function	LeakyReLU
lambda_cycle	10

WGAN-GP.-

The architecture [20] WGAN-GP (Wasserstein GAN with Gradient Penalty) used a Wasserstein loss function plus a gradient penalty to achieve Lipschitz continuity. The architecture proposed in [21] employed a Wasserstein loss function along with a gradient penalty to ensure Lipschitz continuity.

Both the generator (G) and discriminator (D) networks utilized individual optimizers, as summarized in Table 6 and 7. The Adam optimizer with a momentum parameter (β1 = 0.5) was employed for both G and D. The learning rate values were set to 0.0002 for the Generator and 0.00005 for the Discriminator. The chosen loss function was Binary Cross-Entropy, and Batch Normalization was utilized. The activation function employed was LeakyReLU. The training process was conducted for 100 epochs, with a batch size of 64.

The *Generator's hyperparameters* included an input noise vector composed of 256 values within the range of 0 to 1. This noise vector passed through four fully connected layers, followed by additional 2D deep convolutional (Conv) layers and deconvolutional layers (Conv Up). The activation functions were applied to both the noise vector in the latent space and a coded vector used to determine the classes, as detailed in Table 6.

Table 6 summarizes the specific hyperparameters employed during the training of the WGAN-GP generator.

Table 6. The WGAN-GP generator training hyperparameters.

LAYER (TYPE)	**OUTPUT SHAPE**	**PARAM #**
input_1 (InputLayer)	(None, 256)	0

dense_1 (Dense)	(None, 4096)	104857 6
batch_normalization	(None, 4096)	16384
leaky_re_lu_7 (LeakyReLU)	(None, 4096)	0
reshape (Reshape)	(None, 4, 4, 256)	0
up_sampling2d (UpSampling2D)	(None, 8, 8, 256)	0
conv2d_7 (Conv2D)	(None, 8, 8, 128)	294912
batch_normalization_1	(None, 8, 8, 128)	512
leaky_re_lu_8 (LeakyReLU)	(None, 8, 8, 128)	0
dropout_8 (Dropout)	(None, 8, 8, 128)	0
up_sampling2d_1 (UpSampling 2D)	(None,16,16, 128)	0
conv2d_8 (Conv2D)	(None, 16, 16, 64)	73728
batch_normalization_2	(None, 16, 16, 64)	256
leaky_re_lu_9 (LeakyReLU)	(None, 16, 16, 64)	0
dropout_9 (Dropout)	(None, 16, 16, 64)	0
up_sampling2d_2 (UpSampling 2D)	(None, 32, 32, 64)	0
conv2d_9 (Conv2D)	(None, 32, 32, 32)	18432
batch_normalization_3	(None, 32, 32, 32)	128
leaky_re_lu_10 (LeakyReLU)	(None, 32, 32, 32)	0
dropout_10 (Dropout)	(None, 32, 32, 32)	0
up_sampling2d_3 (UpSampling 2D)	(None, 64, 64, 32)	0
conv2d_10 (Conv2D)	(None, 64, 64, 16)	4608
batch_normalization_4	(None, 64, 64, 16)	64
leaky_re_lu_11 (LeakyReLU)	(None, 64, 64, 16)	0
dropout_11 (Dropout)	(None, 64, 64, 16)	0
up_sampling2d_4 (UpSampling 2D)	(None,128,128,16)	0

conv2d_11 (Conv2D)	(None, 128, 128, 8)	1152
batch_normalization_5	(None, 128, 128, 8)	32
leaky_re_lu_12 (LeakyReLU)	(None, 128, 128, 8)	0
dropout_12 (Dropout)	(None, 128, 128, 8)	0
up_sampling2d_5 (UpSampling 2D)	(None, 256, 256, 8)	0
conv2d_12 (Conv2D)	(None, 256, 256, 1)	72
batch_normalization_6	(None, 256, 256, 1)	4
activation (Activation)	(None, 256, 256, 1)	0

*Discriminator hyperparameter*s, the input of the discriminator was the output of the generator, where the input was a patch of images, and the output was a prediction of real or fake images. The images were then passed through convolutional down sampling layers (there are no pooling layers in the model). The last layer was fully connected ones and were activated by ReLU function (table 7).

Table 7. The WGAN-GP discriminator training hyperparameters in breast images.

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	(None, 256, 256, 1)	0
conv2d (Conv2D)	(None, 128, 128, 8)	208
leaky_re_lu (LeakyReLU)	(None, 128, 128, 8)	0
dropout (Dropout)	(None, 128, 128, 8)	0
conv2d_1 (Conv2D)	(None, 64, 64, 16)	3216
leaky_re_lu_1 (LeakyReLU)	(None, 64, 64, 16)	0
dropout_1 (Dropout)	(None, 64, 64, 16))	0
conv2d_2 (Conv2D)	(None, 32, 32, 32)	12832
leaky_re_lu_2 (LeakyReLU)	(None, 32, 32, 32)	0
dropout_2 (Dropout)	(None, 32, 32, 32)	0

conv2d_3 (Conv2D)	(None,16,16, 64)	51264
leaky_re_lu_3 (LeakyReLU)	(None, 16, 16, 64)	0
dropout_3 (Dropout)	(None, 16, 16, 64)	0
conv2d_4 (Conv2D)	(None, 8, 8, 128)	204928
leaky_re_lu_4 (LeakyReLU)	(None, 8, 8, 128)	0
dropout_4 (Dropout)	(None, 8, 8, 128)	0
conv2d_5 (Conv2D)	(None, 4, 4, 256)	819456
leaky_re_lu_5 (LeakyReLU)	(None, 4, 4, 256)	0
dropout_5 (Dropout)	(None, 4, 4, 256)	0
conv2d_6 (Conv2D)	(None, 2, 2, 256)	1638656
leaky_re_lu_6 (LeakyReLU)	(None, 2, 2, 256)	0
dropout_6 (Dropout)	(None, 2, 2, 256)	0
flatten (Flatten)	(None, 1024)	0
dropout_7 (Dropout)	(None, 1024)	0
dense (Dense)	(None, 1)	1025

SNGAN.-

The proposed architecture introduced a novel weight normalization technique known as spectral normalization to enhance the training stability of the discriminator network, serving as the foundation for synthetic image generation. This technique involved normalizing the learned weights (as depicted in Figure 4) within each layer of the discriminator using the spectral norm of the weight matrix, denoted as σ, which corresponds to its largest singular value [48]. It ensured that each weight matrix satisfied the Lipschitz constraint, thereby promoting better convergence and optimization during training [49].

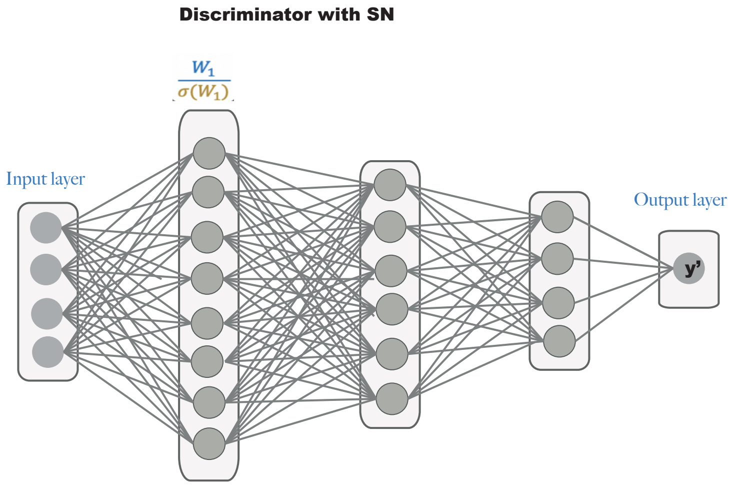


Fig 4. Diagram of SNGAN illustrating the application of spectral normalization, where the weights W_i are divided by their corresponding spectral norms $\sigma(W_i)$ (i.e., the largest singular value of W_i).

The training hyperparameters are: num_epochs = 100, batch_size = 16, image_size = 128, lr = 0.0002, beta1 = 0.1, beta2 = 0.9.

2.4 Diversity and Structural Quality of the synthetic data

The following quantitative techniques have been applied for evaluating the quality of real and synthesized ROIs generated by GANs models: (i) feature-based metrics: FID (Frechet inception distance) and KID (Kernel Inception Distance) to compute the distance between the vector representation of the synthesized and real images; (ii)nonreference-based metrics: BRISQUE (Blind referencess image spatial quality evaluator), NIQE (Naturalness Image Quality Evaluator), and PIQE (Perception-based Image Quality Evaluator) to evaluate the image quality. (iii) reference-based metrics: MS-SIM (Multi-scale Structural Similarity Index Measure) and SSIM (Structural Similarity Index Measure) to evaluate the diversity quality and structural information of a synthesized image against the real image [50].

Feature-based metrics

Fréchet Inception Distance (FID):

This metric is based on a pretrained Inception V3 model [51]. FID compares the distributions of the original and generated images to assess how well the generated images represent the training dataset. Lower FID scores indicate better quality images [41], and it is calculated as shown in Eq. (7):

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

$$FID = \left\|\mu_r - \mu_g\right\|^2 + Tr(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{1/2}) \quad (7)$$

Here, μ_r represents the mean of the feature vector calculated from the real images, μ_g is the mean of the feature vector calculated from the fake images, Σ_r is the covariance of the feature vector from the real images, and Σ_g is the covariance of the feature vector from the fake images.

While FID is more advanced than Inception Score (IS) as it can detect intraclass mode collapse, it has a drawback of high prediction error (bias) and requires a large sample size (above 50K) for accurate calculation. Smaller sample sizes can result in overestimation of the actual FID [52].

Kernel Inception Distance (KID):

This metric serves as an alternative to FID and is preferred because it is insensitive to the global structure of the data distribution and does not assume a parametric form for the distribution. KID employs the cubic kernel (Eq. 8) to compare the skewness, mean, and variance [53]. A lower KID value signifies a higher visual similarity between the actual and generated images. The cubic polynomial kernel is defined as shown in Eq. (8):

$$k(x, y) = \left(\frac{1}{d}\, x^T y + 1\right)^3$$ $$k(x, y) = \left(\frac{1}{d}\, x^T y + 1\right)^3 \quad (8)$$

where *d* represents the dimension of the feature space for vectors *x* and *y*.

Reference-based metrics

Multi-scale Structural Similarity Index Measure (MS-SSIM):

The diversity of synthetic images is assessed by this measure. It compares the score of the real and synthetic image datasets generated by GANs [54], using contrast, structure, and luminance features. A higher MS-SSIM score of the synthetic dataset indicates the occurrence of mode collapse in GANs. MS-SSIM can be computed between two image samples *a* and *b* as defined in Eq. 9.

$$MS - SSIM(a,b) = I_M(a,b)^{\propto M} \prod_{j=1}^{M} C_j(a,b)^{\beta j} S_j(a,b)^{\gamma j} \qquad (9)$$

Where, C =Contrast and S= structural features of images are computed at scale *j*. Luminance (I) is calculated at the coarsest scale (M). The α, β, and γ are the weight parameters as detailed in [52].

Structural Similarity Index Metric (SSIM):

SSIM is a perception-based model that considers the image degradation as perceived change in structural information where, structural information is the idea that the pixels have strong inter- dependencies especially when they are spatially close. The linear dependence factor is computed using the correlation coefficient in SSIM index [55].

The value ranges from 0 to 1, and higher the value of SSIM (0.95 to 1), indicates higher the quality rate. The standard value of SSIM (0.90 to 0.95) indicates acceptable quality and lower the value (< 0.90) indicates poor quality.

The SSIM index is defined as follows (Eq. 11):

$$SSIM\ (x,y) = \frac{(2u_x\mu_{y+c1})(2\sigma_{xy}+c2)}{(\mu x^2 + \mu y^2 + c1)(\sigma x^2 + \sigma y^2 + c2)} \qquad (11)$$

Where:

- u_x the pixel sample mean of *x*;
- u_y the pixel sample mean of *y*;
- σx^2 the variance of *x*;
- σy^2 the variance of *y*;
- σ_{xy} the covariance of *x* and *y*;
- $c_{1\ and\ c_2}$ two small variables to add numerical stability.

Nonreference-based metrics

The BRISQUE, metric uses a subjective quality score (high score has high quality) [56], and analyses images with similar distortions. The NIQE metric, computes the quality of images with arbitrary distortion (high score has low quality) [57]. Contrary to BRISQUE, NIQE does not use subjective quality scores, and hence the assessment of the comparison of BRISQUE and NIQE may not be readily obvious from a mere look. Whereas both BRISQUE and NIQE require a trained model for their computation, PIQE does not. PIQE computes the quality of a given image based on an arbitrary distortion in a blockwise approach (high score has low quality) [58].

ROIs classification using Resnet model.

The synthesized images were used as input for pretraining the ResNet-18 [59-60] classification model to discriminate between malignant and benign ROIs lesions, and to compare the effectiveness in to classify synthetic, real and hybrid data. The generator and discriminator models are fine-tuned using different hyperparameter details are outlined below:

Hyperparameter values to classify US RoIs: Optimizer:Adam, num_epochs = 100, batch_size = 16, image_size = 128, learning rate = 0.0002, beta1 = 0.1,beta2 = 0.9

Hyperparameter values to classify mammography RoIs: num_epochs = 100, batch_size = 16, image_size = 128, lr = 0.0002, beta1 = 0.1, beta2 = 0.9.

The training model was implemented by Python libraries such as the scikit-learn module, pytorch, keras, tensorflow and Google Colab PRO (GPU) cloud-based.

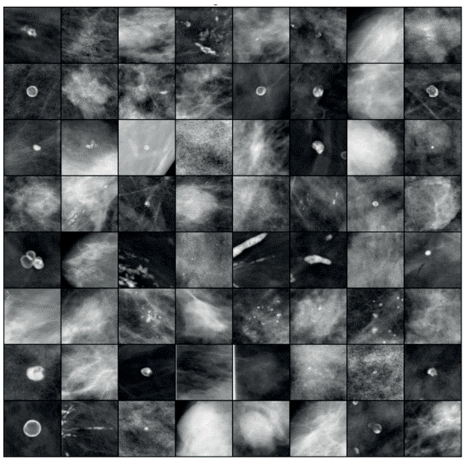
Classification performance model

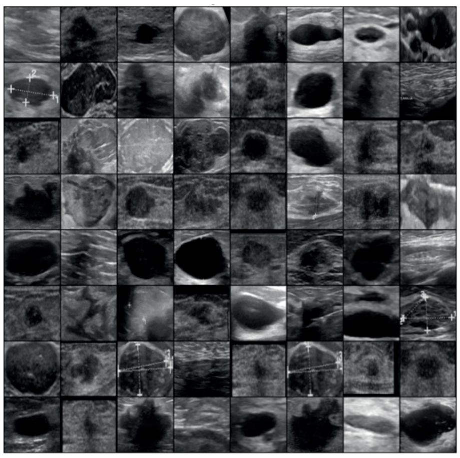
The statistical metrics used to evaluate the performance of our image classification model are accuracy (Acc), precision (Prec), F1-score and Recall [2].

3. Results

3.1 Breast ROIs data augmentation

Cycle GAN and WGAN-GP models generated a smaller number of synthetic RoIs in both mammography and US datasets, totaling 2,310 RoIs. In contrast, CGAN and SNGAN generated a significantly larger number of synthetic RoIs, with 10,000 synthetic mammography RoIs and 4,000 synthetic US RoIs. Figure 5a depicts the grid with real RoIs from the mammography dataset, while Figure 5b displays the grid with real RoIs from the US dataset.

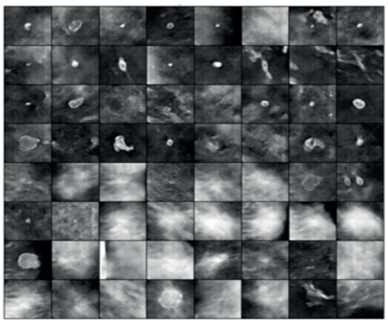


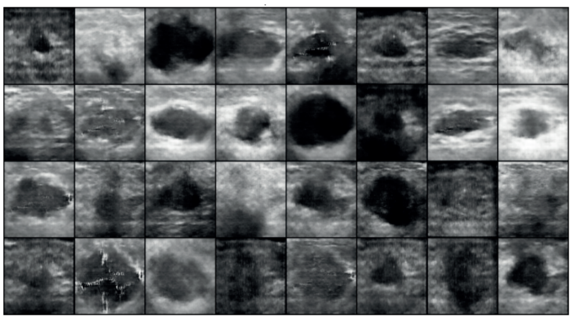


(a) (b)

Fig. 5. Sample of real ROIs a. mammography and b. ultrasound

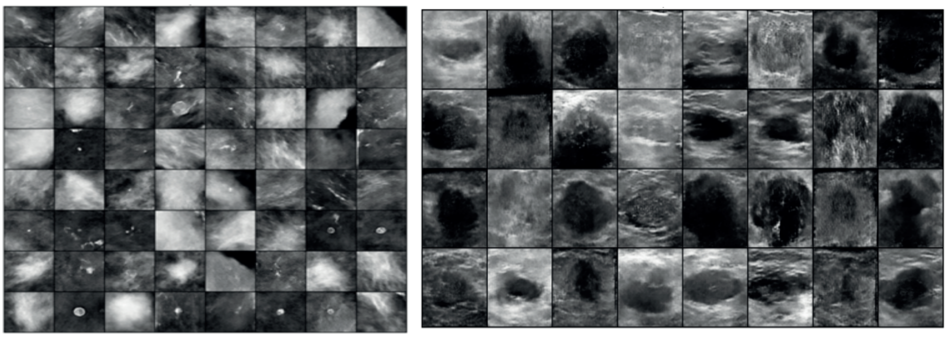
Figures 6 and 7 showcase the synthetic RoIs generated by CGAN and SNGAN, respectively. Each figure presents a grid of 64 example images, allowing for a detailed observation of the visual output produced by these models.





(a) (b)

Fig. 6. Synthetic RoIs samples obtained with CGAN a. mammography and b. ultrsound.

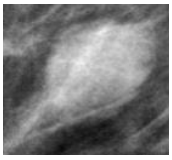
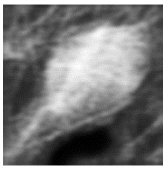
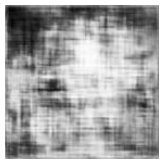


(a) *(b)*

Fig. 7. Synthetic RoI samples obtained with SNGAN a. mammography and b. ultrasound

Table 8 displays the synthetic RoIs obtained with Cycle GAN and WGAN-GP respectively.

Table 8. Cycle GAN and WGAN-GP synthetic RoI evolution with respect to mammography real image.

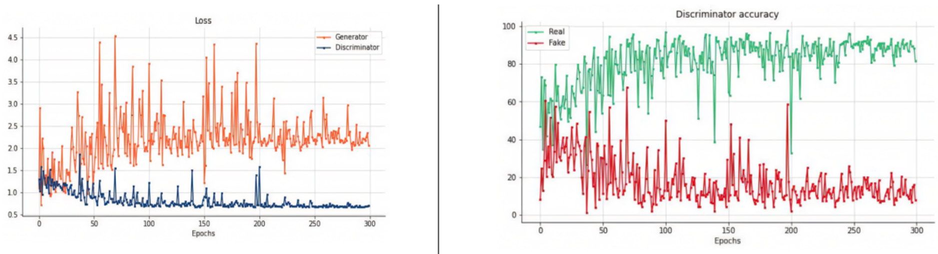
	Real Image	**CycleGAN**	**WGAN-GP**
Epoch 100			

3.1.1 Generator and discriminator loss and accuracy

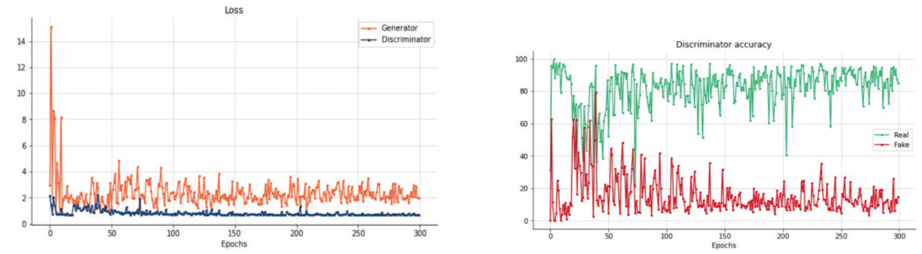
The loss function and accuracy of the generator and discriminator play a crucial role in assessing the training stability and performance of GANs. A stable GAN is characterized by a discriminator loss around 0.5 or higher than 0.7, while the generator loss typically ranges from 1.0 to higher values like 1.5, 2.0, or even more. The accuracy of the discriminator, both on real and synthetic images, is expected to hover around 70% to 80%.

Figures 8, 9, and 10 illustrate the loss and accuracy trends for CGAN, SNGAN, Cycle GAN, and WGAN-GP, respectively. The CGAN loss plot shows that both the discriminator and generator exhibit initial instability. As training progresses, the generator improves (reaching values above 0.5), while the discriminator performance deteriorates (around 0.5), eventually converging to stable points around epoch 75 and 220, respectively. These stable points indicate that either the generator or the discriminator has undergone sufficient training. Beyond these values, the network may fall into a collapse mode.

The accuracy plot for CGAN (Figures 8a-8d) and SNGAN (Figures 9a-9d) displays the discriminator's accuracy on real (green) and fake (red) images during training. Initially, the accuracy differs significantly between the two types of breast images but stabilizes between epochs 100 and 300 at approximately 80% to 85%. The time scales (number of iterations or epochs) vary depending on the image type and GAN models used.



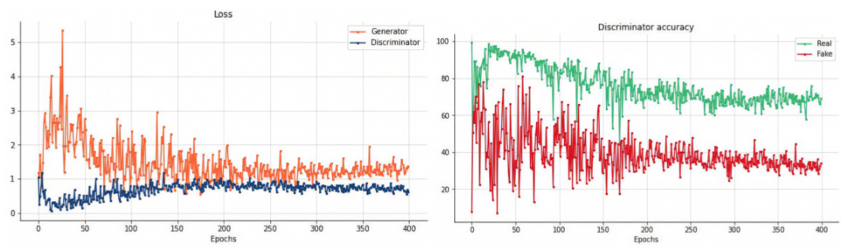
(a) (b)



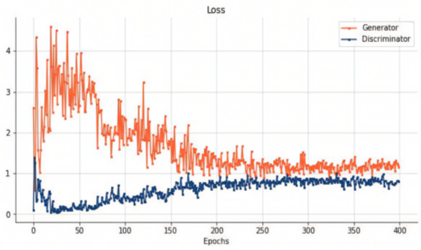
(c) (d)

Fig 8. a. CGAN loss in mammography images. b. CGAN accuracy in mammography images. c. CGAN loss in US images. d. CGAN accuracy in US images.

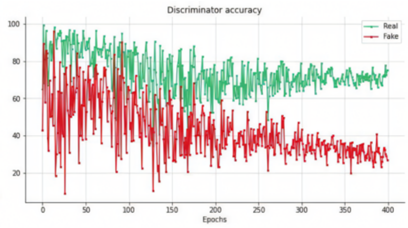
In the case of SNGAN, the discriminator loss increases until reaching a stable point, while the generator approaches a value close to 1. SNGAN encounters difficulty in discerning between real and fake images; however, it manages to avoid falling into a mode collapse.



(a) (b)



(c)



(d)

Fig 9. a. SNGAN loss function in mammography images. b. SNGAN accuracy in mammography images. c. SNGAN loss function in US images. d. SNGAN accuracy in US images.

Figure 10 demonstrates the loss values for the discriminators and generators in Cycle GAN, indicating a reduction in losses and high predictive ability from the early stages of training. The Cycle GAN model generates acceptable images between epochs 20 and 100. Conversely, the images generated by WGAN are of poor quality until epoch 100 (as indicated in Table 8).

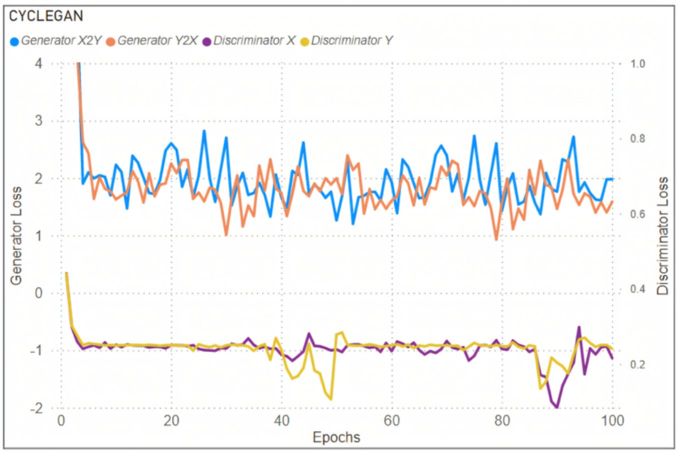


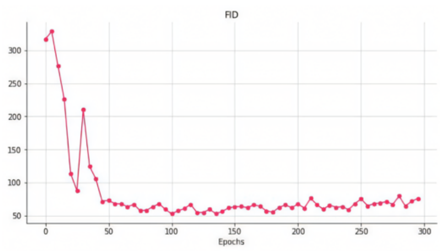
Fig 10. a. Cycle GAN loss function in mammography and US images.

3.1.2 Feature-based metrics

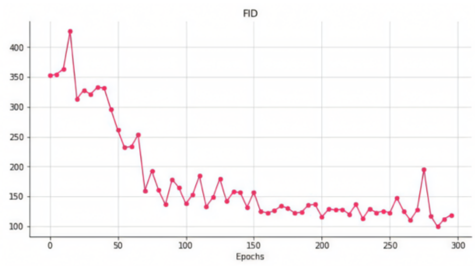
The performance of GANs in breast ROI generation was evaluated using the FID and KID metrics, which capture the distance between the feature distributions of real and generated images better than the inception score [59]. A lower FID and KID value indicates a higher visual similarity between the real and generated images, and higher FID score shows a larger distance between synthetic and real data distributions that indicates the occurrence of mode collapse.

First, the distances between CGAN and SNGAN were evaluated using FID. Figure 11a and 11b display the FID values obtained after training CGAN with mammography and US datasets, respectively. In the mammography dataset, the highest FID score occurs at epoch 5 (328.25), while the lowest score is observed at epoch 100 (52.89). For the US dataset, the FID values reach their highest point (352.25) at epoch 15 and the lowest point (133.03) at epoch 115. Figures 11c and 11d show the FID values in SNGAN. In the mammography dataset, the highest score is observed at epoch 10 (350), and the lowest score is reached at epoch 390 (58.80). In the US dataset, the network stabilizes around epoch 200, with the highest value (380.17) decreasing to its lowest point (116.85) at epoch 310.

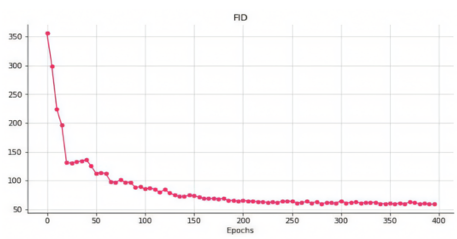
The average FID values for CGAN are 52.89 to mammography and 133.03 to US. In SNGAN, the FID values are 58.80 to mammography and 116.85 to US. The top-rated CGAN and SNGAN images can be observed in Figures 9a-c.

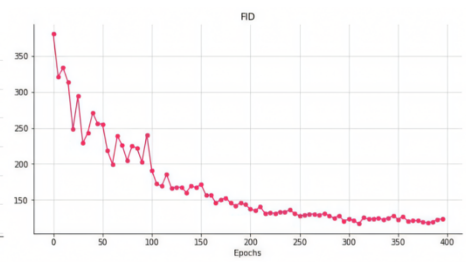


(a)



(b)

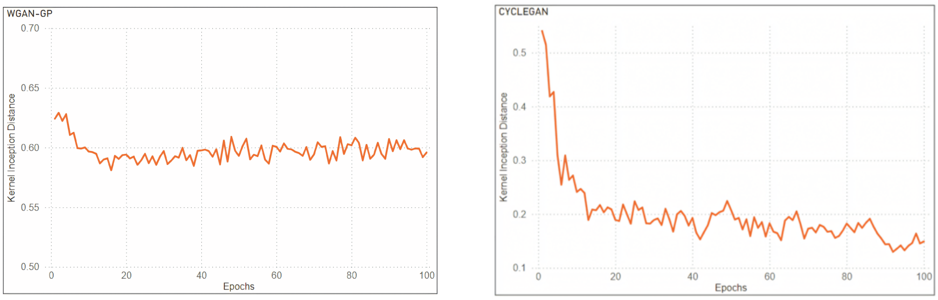




(c) (d)

Fig 11. a. FID values taken from mammography RoIs using CGAN. b. FID values taken from ultrasound RoIs using CGAN. c. FID values taken from mammography RoIs using SNGAN and d. FID values taken from US RoIs using CGAN.

Then, the KID metric was employed to compare real and synthetic images generated by Cycle GAN and WGAN-GP (Figure 12), as it is suitable for smaller datasets. It offers computational efficiency, numerical stability, and simplicity of implementation. From Figures 12a and 12b, it is evident that Cycle GAN produces high-quality and diverse synthetic images, achieving a lower KID value (0.20) compared to WGAN-GP with a KID value of 0.60, indicating a higher visual similarity between real and generated images.



(a) (b)

Figure 12. KID score evolution during Cycle GAN and WGAN-GP training.

Table 9 presents the average KID values for both GAN networks. The Cycle GAN network demonstrates smaller values, suggesting the generation of more diverse and higher-quality images. This observation aligns with the visual evaluation presented in Table 8.

Table 9. KID metric values for Cycle GAN and WGAN-GP networks.

Epoch	WGAN-GP	CYCLE GAN
1	0.6242	0.5405
5	0.6106	0.3093
10	0.5967	0.2413
15	0.591	0.2072
20	0.5941	0.1892

25	0.5948	0.2238
30	0.597	0.189
35	0.5998	0.1676
40	0.5976	0.1931
45	0.5859	0.202
50	0.5931	0.208
55	0.5928	0.1592
60	0.6006	0.1829
65	0.5965	0.1949
70	0.5942	0.1728
75	0.597	0.1673
80	0.6019	0.1825
85	0.5907	0.1839
90	0.6072	0.144
95	0.5992	0.133
100	0.5959	0.149

3.1.3 Reference and Nonreference-based metrics

Table 10 shows the average scores of MS-SSIM, SSIM, BRISQUE, PIQE, and NIQE metrics after evaluated real and synthetic image datasets (see Appendix A).

Table 10. Comparison of the average scores between the diversity and quality analysis of real and synthesized image pairs. The metrics range in reference-based, nonreference-based, and feature-based categories.

Dataset	*SSIM/MS-SSIM*	*SYNTHETIC IMAGES* ***NIQE/PIQE/BRISQUE***	*REAL IMAGES* **NIQE/PIQE/BRISQUE**
US	0.10/0.16	17.33/ 37.25 /40.75	17.34/24.79/32.89
DM	0.20/0.25	13.88/17.19/21.13	13.92/15.90/19.56

Classification of breast ROIs

Based on the SNGAN and CGAN previous results (which show the most efficient diversity and structural quality scores to produce more realistic synthetic breast mass), 10,0000 mammography and 4,000 US ROI samples were used as Resnet-18 training. The datasets were combined as follows: Data Augmentation (Real and Synthetic), without Data Augmentation (Real), and Hybrid Data (Real+Synthetic). Table 11 details the average scores of this comparative evaluation (see Appendix B).

Table 11. Resnet classification into B: Benning and M: Malignant RoIs.

	Dataset	RoI number Training (80%) Testing (20%)	Accuracy (%) B/M	Precision (%) B/M	Recall (%) B/M	F1-Score(%) B/M
Hybrid Data	US	4032/1009	43.9/55.4	49/50	44/55	46/52
Without Data augmentation		832/209	93.1/94.9	97/89	93/95	95/92
Data augmentation		3500/500	47.9/52.4	50/50	48/42	49/51
Hybrid Data	DM	3747/937	74.7/78.2	91/51	75/78	82/62
Without Data augmentation		3570/893	80.9/76.9	92/56	87/0.78	86/65
Data augmentation		8000/2000	62.4/33.8	48/48	62/34	54/40

Discussion

This synthetic data can be used as system input for various CNN-based segmentation [60] and classification [61] techniques. It is widely acknowledged that the quality and diversity of the training dataset greatly impact the training of deep learning models [36]. However, access to medical data is often restricted due to privacy policies, ethical considerations, and the high costs associated with data collection. This poses a challenge in computer vision, as imbalanced or limited diversity data can lead to overtraining on majority classes and poor generalization to testing samples. Conversely, a balanced training breast dataset enhances the performance and accuracy of CAD models.

Generative models, particularly GANs [11], have recently shown promising results in generating realistic images compared to traditional methods that only generate images with a similar distribution to the original dataset. Unlike generative models based on maximum likelihood estimation, GAN-based generative models have demonstrated success in generating synthetic medical images. By leveraging the latent space, these models can generate new realistic images based on the data distribution [22].

To the best of our knowledge, there are no comparative studies of several GAN models for generating breast synthetic data that combines US and mammogram RoIs. It is well-known that training GANs can be challenging due to training instability caused by the original objective function, the JS divergence. To address this, we implemented four modern GAN frameworks (WGAN-GP, Cycle GAN, CGAN, SNGAN) with different normalization techniques.

The metrics range in feature-based categories (FID and KID), reference-based (MS-SIM, SSIM) and nonreference-based (BRISQUE, NIQE and PIQE) were used to evaluate the diversity and the quality of the synthetic images generated by GAN models. A lower FID and KID value means a higher visual similarity between the real and generated images. The results indicate that the CNGAN model is more effective for mammography data augmentation (FID=52.89), while SNGAN is suitable for US data augmentation (FID=116.03).

The average KID value (0.20) demonstrates that Cycle GAN outperforms WGAN-GP (0.60) in generating high-quality and diverse synthetic images for both datasets. Cycle GAN exhibits a higher visual similarity between the real and generated images, whereas WGAN-GP produces visually unacceptable results (Table 9) with limited discriminator and generator training. These findings align with Cantero et al. [62], who explored two GANs (DCGAN and WGAN-GP) for generating synthetic PET breast imaging and found that the synthetic data generated by WGAN-GP did not significantly outperform those produced by DCGAN. Moreover, in [63-64], the authors concluded that Cycle GAN was the better technique for US data augmentation.

This disparity can be attributed to the fact that Cycle GAN utilizes a domain of real images as a reference for data augmentation, while WGAN-GP relies on theoretical image features for augmentation. Some researchers have replaced JS divergence in the original GAN with Wasserstein distance to achieve stability during training, Gulrajani et al. [21] proposed WGAN with gradient penalty (GP) to penalize the norm of the gradient, resulting in more stable training compared to WGAN with additional hyperparameter settings. However, even with these improvements, the network may still generate poor samples or fail to converge.

Guan et al. [17] cropped ROIs from DDSM mammography images using two data augmentation methods: (i) affine transform (AT) and (ii) Cycle GAN. Their results showed that synthetic ROIs produced by GANs more closely resembled real ROIs in terms of mean and entropy values.

Other quantitative measures were implemented to evaluate the structure, diversity (SSIM and MS-SSIM), and quality (BRISQUE, PIQUE and NIQUE) of the generated images. Table 10 shows the average scores of these metrics, after evaluated real and synthetic image datasets. The lower average MS-SSIM values shown in Table 10 indicate that the SNGAN and CGAN models generate a high diversity of synthetic images, and no mode collapse occurred during the model training. However, the lower SSIM values indicate that these synthetic images are poor quality. It can be attributed to the blurry and the image size (128 x 128 pixels), as indicated by Oyelade et al. [50], use ROIs digital mammography for characterizing abnormalities in breast cancer does not contain sufficient data with a fair distribution of all cases of abnormalities.

For instance, in super-resolution (SR) [52,55] algorithms, these values are usually higher (>0.40) because the distribution of the new images is compared over the same image. In contrast, in data augmentation these metrics are lower because the comparison is between the distribution differences of the real and generated images.

Table 10 also show the quality of the real and synthetic images by BRISQUE, PIQUE, and NIQUE. The lower values in NIQUE mean high generated image quality, but lower values in BRISQUE and PIQUE mean poor quality [50]. Thus, a similar distribution is demonstrated by the proposed GAN models, with BRISQUE and PIQUE lower outcomes (< 40%), which means poor image quality, resulting in architectural distortion taken from the real ROI images. As reflected these average values indicates that the distribution of real and synthetic images shows more outliers and broader distribution values.

GAN data augmentation has also been used as a training method in CNN-based classification approaches [3,17,39,65-66] to improve performance (accuracy of 79.8%) and prevent overfitting.

Wu et al. [18] trained a multiscale Conditional GAN to generate high-resolution synthetic DDSM mammograms and evaluated the ResNet50 classifier with this augmented dataset. The results demonstrated that GAN-based augmentation improved mammogram RoI-based classification with an AUC improvement of 0.014 over traditional GANs and 0.009 over traditional augmentation techniques. Wong et al. [36] proposed a Conditional GAN for BreaKHist histopathological image data augmentation.

The experiments showed that CGAN-based data augmentation achieved an average accuracy of 77.345%, outperforming standard augmentation by 0.678.

Additionally, the computational cost of using SN is relatively small compared to other normalization techniques [6].

Consistent with previous findings [2], SNGAN proves to be an effective technique for stabilizing the training of the GAN discriminator in mammography data augmentation. This technique can generate images of equal or better quality compared to previous training stabilization techniques such as weight normalization and gradient penalty.

However, in our research, the evaluation metrics (Table 11) based on Resnet-18 indicate that the values with high accuracy were shown by the dataset without data augmentation in US (93.1% B and 94.9 % M) and in mammography (80.9 % B and 76.9% M). According to Oyelade et al. [50] is better to focus the main attention in previously classifies and characterizing abnormalities into architectural distortion, asymmetry, mass, and microcalcification so that training distinctively learns the features of each abnormality and generates sufficient images for each category, before training a GAN model. However, in [50] no ROI classification process was performed, and no detail is given on how many synthetic images were assessed as real or synthetic.

In conclusion, GANs prove to be a useful tool for breast data augmentation. However, achieving these standards often involves meticulous data collection, preprocessing (like noise reduction and artifact removal), and expert domain knowledge to correctly classify and interpret anomalies. This process enhances the effectiveness and accuracy of anomaly detection systems. The availability of diverse and high-quality synthetic data addresses the challenges posed by limited access to medical datasets. Future research can further explore the application of GANs in different medical imaging domains and investigate additional normalization and regularization methods to improve GAN and CNN training stability and performance.

Conclusions

Due to the limited availability of mammography and US datasets, realistic medical data augmentation plays a crucial role in training deep learning systems to improve breast lesion segmentation and classification, as well as to prevent overfitting during the training process. In this study, we implemented four GAN models with different normalization techniques to generate synthetic images. The results demonstrated that SNGAN was effective for mammography data augmentation (FID=52.89), while CGAN performed well for US data augmentation (FID=116.03). Moreover, the Cycle GAN model proved to be more successful in generating high-quality and diverse synthetic images for both datasets (KID=0.20) compared to WGAN-GP (KID=0.60). Cycle GAN exhibited a higher visual similarity between the real and generated images, even in a reduced training environment. On the other hand, while WGAN-GP successfully addressed the mode collapse issue, it resulted in visually unacceptable outcomes and lower stability in discriminator and generator training for both datasets. Nevertheless, cycle GAN and WGAN do not demonstrate better results than SNGAN and CGAN. Resnet-18 demonstrated high accuracy values by the Benign and Malignant classes without data augmentation, in US (93.1%/ 94.9 %) and mammography (80.9 %/76.9%) with respect to the other two datasets with data augmentation and Hybrid data.

In conclusion, the selection of the most suitable GAN model for data augmentation depends on the specific problem, dataset characteristics, evaluation quality metrics and hyperparameter choices to enhance GAN training stability.

For future work, we aim to incorporate other types of breast images (PET, Thermal and MRI) that have been pre-processed with denoising, super-resolution, previously classify and characterize by abnormality type, to obtain a diverse range of synthetic data and further improve the classification process using convolutional algorithms.

Data Availability Statement: Data will be made available on request.

CRediT authorship contribution statement

Y.J: Investigation, Methodology, Software, Validation, Formal analysis, and Writing – original draft. M.J.A.R: Project administration, Resources, Writing –original draft. V.L: Conceptualization, Writing – review & editing original draft. D.F Data curation, Writing – original draft.

Abbreviations

BUSI:Breast Ultrasound Images Dataset

BCE:Binary cross entropy

CGAN:Conditional GAN

CAD:Computer aided system

CNN:Convolutional neural network

DDSM:Digital Database for Screening Mammography

DM:Digital mammography

DL:Deep learning

D:Discriminator

GAN:Generative adversarial network

FID:Fréchet Inception Distance

G:Generator
IS:Inception Score
JS:Jensen Shannon
KL:Kullback–Leibler
KID:Kernel inception distance
LS:Least Square
mini-MIAS:Mammographic Image Analysis Society
RoI:Region of interest
ReLu:Rectified Linear Unit
SNGAN:Spectral Normalization GAN
TL:Transfer Learning
UDIAT:Diagnostic Centre of the Parc Tauli Corporation
US:Ultrasound
WGAN-GP:Wasserstein GAN with gradient penalty

3.2 References

[1]. Shen Y, Shamout FE, Oliver JR, Witowski J, Kannan K, Park J, et al. Artificial intelligence system reduces false-positive findings in the interpretation of breast ultrasound exams. Nat Commun [Internet]. 2021;12(1):5645. [http://dx.doi.org/10.1038/s41467-021-26023-2]

[2]. Jiménez-Gaona, Y., Rodríguez-Álvarez, M. J., & Lakshminarayanan, V. Deep-learning-based computer-aided systems for breast cancer imaging: a critical review. Applied Sciences. 2020;10(22):8298.

[3]. Al-Dhabyani, W., Gomaa, M., Khaled, H., & Aly, F. Deep learning approaches for data augmentation and classification of breast masses using ultrasound images. Int J Adv Comput Sci Appl. 2019;10(5):1–11.

[4]. Saini M, Susan S. Deep transfer with minority data augmentation for imbalanced breast cancer dataset. Appl Soft Comput [Internet]. 2020;97(106759):106759. [http://dx.doi.org/10.1016/j.asoc.2020.106759]

[5]. Hussain Z, Gimenez F, Yi D, Rubin D. Differential data augmentation techniques for medical imaging classification tasks. En AMIA annual symposium proceedings. American Medical Informatics Association. 2017;

[6]. Jiménez Gaona, Y., Castillo Malla, D., Vega Crespo, B., Vicuña, M. J., Neira, V. A., Dávila, S., & Verhoeven, V. (2022). Radiomics diagnostic tool based on deep learning for colposcopy image classification. *Diagnostics*, *12*(7), 1694.

[7]. Chaudhari P, Agrawal H, Kotecha K. Data augmentation using MG-GAN for improved cancer classification on gene expression data. Soft Comput 2020;24(15):11381–91. [http://dx.doi.org/10.1007/s00500-019-04602-2].

[8]. Tran N-T, Tran V-H, Nguyen N-B, Nguyen T-K, Cheung N-M. On data Augmentation for GAN training. IEEE Trans Image Process. 2021;30:1882–97. [http://dx.doi.org/10.1109/TIP.2021.3049346]

[9]. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. (2016). Improved techniques for training gans. Advances in neural information processing systems, 29.

[10]. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial networks. Commun ACM. 2020;63(11):139–44. [http://dx.doi.org/10.1145/3422622]

[11]. Chen Y, Yang X-H, Wei Z, et al. Generative Adversarial Networks in Medical Image augmentation: A review. Comput Biol Med. 2022;144:105382. [doi:10.1016/j.compbiomed.2022.105382]

[12]. Goceri E. Medical image data augmentation: techniques, comparisons and interpretations. Artif Intell Rev. March 2023. [doi:10.1007/s10462-023-10453-z]

[13]. Kim, M., Kim, S., Kim, M., Bae, H. J., Park, J. W., & Kim, N. (2021). Realistic high-resolution lateral cephalometric radiography generated by progressive growing generative adversarial network and quality evaluations. *Scientific reports*, *11*(1), 12563. https://doi.org/10.1038/s41598-021-96793-8

[14]. Hong, S., et al.: 3D-StyleGAN: a style-based generative adversarial network for generative modeling of three-dimensional medical images. In: Engelhardt, S., et al. (eds.) DGM4MICCAI/DALI 2021. LNCS, vol. 13003, pp. 24–Springer, Cham (2021). https://doi.org/10.1007/978-3-030-88210-5 3

[15]. Molahasani Majdabadi, M., Choi, Y., Deivalakshmi, S., Ko, S.: Capsule GAN for prostate MRI super-resolution. Multimed. Tools Appl. 81(3), 4119–4141 (2022)

[16]. Gab Allah, A.M.; Sarhan, A.M.; Elshennawy, N.M. Classification of Brain MRI Tumor Images Based on Deep Learning PGGAN Augmentation. Diagnostics 2021, 11, 2343

[17]. Guan S. Breast cancer detection using synthetic mammograms from generative adversarial networks in convolutional neural networks. J Med Imaging. 2019;6(03):1. [doi:10.1117/1.JMI.6.3.031411]

[18]. Wu E, Wu K, Cox D, Lotter W. Conditional Infilling GANs for Data Augmentation in Mammogram Classification. In: ; 2018:98-106. [doi:10.1007/978-3-030-00946-5_11]

[19]. Li Y, Zhang K, Shi W, Miao Y, Jiang Z. A novel medical image denoising method based on conditional generative adversarial network. Comput Math Methods Med [Internet]. 2021;2021:9974017. [http://dx.doi.org/10.1155/2021/9974017]

[20]. Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks. In: International conference on machine learning. 2017. p. 214–23.

[21]. Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville A. Improved Training of Wasserstein GANs. March 2017. http://arxiv.org/abs/1704.00028.

[22]. Cai, L., Chen, Y., Cai, N., Cheng, W., & Wang, H. Utilizing amari-alpha divergence to stabilize the training of generative adversarial networks. Entropy, 2020, vol 22, no 4,. 22(4):410.
[23]. Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2223-2232).
[24]. C. Baur, S. Albarqouni and N. Navab, 'MelanoGANs: High resolution skin lesion synthesis with GANs', arXiv preprint arXiv:1804.04338,2018.
[25]. C. Nielsen and M. Okoniewski, 'GAN data augmentation through active learning inspired sample acquisition', in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops,2019, pp. 109–112
[26]. Alruily, M., Said, W., Mostafa, A. M., Ezz, M., & Elmezain, M. (2023). Breast Ultrasound Images Augmentation and Segmentation Using GAN with Identity Block and Modified U-Net 3+. Sensors, 23(20), 8599.
[27]. Goceri, E. (2023). Medical image data augmentation: techniques, comparisons and interpretations. Artificial Intelligence Review, 1-45.
[28]. Chen, Y., Yang, X. H., Wei, Z., Heidari, A. A., Zheng, N., Li, Z., ... & Guan, Q. (2022). Generative adversarial networks in medical image augmentation: A review. Computers in Biology and Medicine, 144, 105382.
[29]. C. Lin, et al. Breast mass detection in mammograms via blending adversarial learning International Workshop on Simulation and Synthesis in Medical Imaging, Springer (2019)
[30]. T. Shen, et al.Mass image synthesis in mammogram with contextual information based on GANs Comput. Methods Progr. Biomed., 202 (2021), p. 106019
[31]. Korkinof, D., Rijken, T., O'Neill, M., Yearsley, J., Harvey, H., & Glocker, B. High-resolution mammogram synthesis using progressive generative adversarial networks. arXiv preprint arXiv:180703401, 2018. 2018;
[32].Escobar, M.; Castillo, A.; Romero, A.; Arbeláez, P. UltraGAN: Ultrasound Enhancement Through Adversarial Generation. In Simulation and Synthesis in Medical Imaging; Springer International Publishing: Cham, Switzerland, 2020; pp. 120–130.
[33].Fujioka, T.; Mori, M.; Kubota, K.; Kikuchi, Y.; Katsuta, L.; Adachi, M.; Oda, G.; Nakagawa, T.; Kitazume, Y.; Tateishi, U. Breast Ultrasound Image Synthesis using Deep Convolutional Generative Adversarial Networks. Diagnostics 2019, 9, 176
[34]. Bargsten, L.; Schlaefer, A. SpeckleGAN: A generative adversarial network with an adaptive speckle layer to augment limited training data for ultrasound image processing. Int. J. Comput. Assist. Radiol. Surg. 2020, 15, 1427–1436
[35]. Zhang, H.; Xu, T.; Li, H.; Zhang, S.; Wang, X.; Huang, X.; Metaxas, D. StackGAN: Text to Photo-Realistic Image Synthesis with Stacked Generative Adversarial Networks. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 5908–5916.
[36]. Wong, W. S., Amer, M., Maul, T., Liao, I. Y., & Ahmed, A. (2020). Conditional generative adversarial networks for data augmentation in breast cancer classification. In Recent Advances on Soft Computing and Data Mining: Proceedings of the Fourth

International Conference on Soft Computing and Data Mining (SCDM 2020), Melaka, Malaysia, January 22– 23, 2020 (pp. 392-402). Springer International Publishing. [doi:10.1007/978-3-030-36056-6_37]
[37]. Chen H. Challenges and Corresponding Solutions of Generative Adversarial Networks (GANs): A Survey Study. J Phys Conf Ser. 2021;1827(1):012066. [doi:10.1088/1742-6596/1827/1/012066]
[38]. Mikolajczyk A, Grochowski M. Data augmentation for improving deep learning in image classification problem. In: 2018 International Interdisciplinary PhD Workshop (IIPhDW). IEEE; 2018:117-122. [doi:10.1109/IIPHDW.2018.8388338]
[39]. Pang T, Wong JHD, Ng WL, Chan CS. Semi-supervised GAN-based Radiomics Model for Data Augmentation in Breast Ultrasound Mass Classification. Comput Methods Programs Biomed. 2021;203:106018. [doi:10.1016/j.cmpb.2021.106018]
[40].Rodrigues, PS (2017), "Breast Ultrasound Image", Mendeley Data, V1, [doi: 10.17632/wmy84gzngw.1]
[41]. Yap MH, Pons G, Marti J, et al. Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks. IEEE J Biomed Heal Informatics. 2018;22(4):1218-1226. [doi:10.1109/JBHI.2017.2731873]
[42]. Heath M, Bowyer K, Kopans D, Kegelmeyer P, Moore R, Chang K, et al. Current status of the digital database for screening mammography. En Digital mammography. Dordrecht: Springer; 1998.
[43]. Suckling J, Parker J, Dance D, Astley S, Hutt I, Boggis C, et al. Mammographic Image Analysis Society (MIAS) database v1.21. 2015.
[44]. Li S, Hatanaka Y, Fujita H, Hara T, Endo T. Automated Detection of Mammographic Masses in MIAS Database. Medical imaging technology. 1999;17:427–8.
[45]. Moreira IC, Amaral I, Domingues I, Cardoso A, Cardoso MJ, Cardoso JS. INbreast. Acad Radiol. 2012;19(2):236-248. [doi:10.1016/j.acra.2011.09.014]
[46]. Huang M-L, Lin T-Y. Dataset of breast mammography images with masses. Data Br. 2020;31:105928. [doi:10.1016/j.dib.2020.105928]
[47].Ponti, M., Kittler, J., Riva, M., de Campos, T., & Zor, C. (2017). A decision cognizant Kullback–Leibler divergence. Pattern Recognition, 61, 470-478.
[48].Lin Z, Sekar V, Fanti G. Why Spectral Normalization Stabilizes GANs: Analysis and Improvements. September 2020. http://arxiv.org/abs/2009.02773.
[49]. Miyato T, Kataoka T, Koyama M, Yoshida Y. Spectral Normalization for Generative Adversarial Networks. February 2018. http://arxiv.org/abs/1802.05957.
[50]. Oyelade, O. N., Ezugwu, A. E., Almutairi, M. S., Saha, A. K., Abualigah, L., & Chiroma, H. (2022). A generative adversarial network for synthetization of regions of interest based on digital mammograms. Scientific Reports, 12(1), 6166.
[51]. Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the AAAI conference on artificial intelligence (Vol. 31, No. 1).
[52]. Borji A. Pros and cons of GAN evaluation measures. Comput Vis Image Underst. 2019;179:41-65. [doi:10.1016/j.cviu.2018.10.009]

[53]. Alqahtani, Hamed, et al. (2019). An analysis of evaluation metrics of GANs. In *International Conference on Information Technology and Applications (ICITA)* (Vol. 7).
[54]. Saad, M.M., Rehmani, M.H., O'Reilly, R. (2023). A Self-attention Guided Multi-scale Gradient GAN for Diversified X-ray Image Synthesis. In: Longo, L., O'Reilly, R. (eds) Artificial Intelligence and Cognitive Science. AICS 2022. Communications in Computer and Information Science, vol 1662. Springer, Cham. https://doi.org/10.1007/978-3-031-26438-2_2
[55]. Rajkumar, S., & Malathi, G. (2016). A comparative analysis on image quality assessment for real time satellite images. Indian J. Sci. Technol, 9(34), 1-11.
[56].Mittal, A., Moorthy, A. K., & Bovik, A. C. (2012). No-reference image quality assessment in the spatial domain. IEEE Transactions on image processing, 21(12), 4695-4708.
[57].Mittal, A., Soundararajan, R., & Bovik, A. C. (2012). Making a "completely blind" image quality analyzer. IEEE Signal processing letters, 20(3), 209-212.
[58].Venkatanath, N., Praneeth, D., Bh, M. C., Channappayya, S. S., & Medasani, S. S. (2015, February). Blind image quality evaluation using perception based features. In 2015 twenty first national conference on communications (NCC) (pp. 1-6). IEEE.
[59].Wu, N., Phang, J., Park, J., Shen, Y., Huang, Z., Zorin, M. et al. (2020). Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening. IEEE transactions on medical imaging, 39(4), 1184–1194. https://doi.org/10.1109/TMI.2019.2945514
[60]. Yi Z, Wang J, Li M. Deep image and feature prior algorithm based on U-ConformerNet structure. Phys Med. 2023;107(102535):102535. [http://dx.doi.org/10.1016/j.ejmp.2023.102535]
[61]. Ricciardi R, Mettivier G, Staffa M, Sarno A, Acampora G, Minelli S, et al. A deep learning classifier for digital breast tomosynthesis. Phys Med. 2021;83:184–93. [http://dx.doi.org/10.1016/j.ejmp.2021.03.021]
[62].Cantero Lorenzo J. (2021). A GAN approach to synthetic PET imaging generation for breast cancer diagnosis. Master's Degree in Data Science. Barcelona, Spain. Universitat Oberta de Catalunya.
[63]. Qi, C., Chen, J., Xu, G., Xu, Z., Lukasiewicz, T., & Liu, Y. (2020). SAG-GAN: Semi-supervised attention-guided GANs for data augmentation on medical images. *arXiv preprint arXiv:2011.07534*. [https://doi.org/10.48550/arXiv.2011.07534]
[64]. Fan Y, Jiao J. Mammography Data Augmentation Using ACGAN. En: Proceedings of the 2nd International Symposium on Artificial Intelligence for Medicine Sciences. New York, NY, USA: ACM; 2021.
[65] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, H. Greenspan, GANbased synthetic medical image augmentation for increased CNN performance in liver lesion classification, Neurocomputing 321 (2018) 321–331.
[66].Muramatsu, C., Nishio, M., Goto, T., Oiwa, M., Morita, T., Yakami, M., ... & Fujita, H. (2020). Improving breast mass classification by shared data with domain transformation using a generative adversarial network. Computers in biology and medicine, 119, 103698.

Chapter 6

BRANET: A MOBIL APPLICATION FOR BREAST IMAGE CLASSIFICATION BASED ON DEEP LEARNING ALGORITHMS

Yuliana Jiménez-Gaona, María José Rodríguez Álvarez, Darwin Castillo-Malla, Santiago García-Jaen, Diana Carrión-Figueroa, Patricio Corral-Domínguez, and Vasudevan Lakshminarayanan .

Medical & Biological Engineering & Computing, https://doi.org/10.1007/s11517-024-03084-1

Abstract

Background: Mobile health apps are widely used for breast cancer detection using artificial intelligence algorithms, providing radiologists with second opinions, and reducing false diagnoses. **Aim:** This study aims to develop an open-source mobile app named "BraNet" for 2D breast imaging segmentation and classification using deep learning algorithms. **Methods:** During the Phase off-line, an SNGAN model was previously trained for synthetic image generation, and subsequently, these images were used to pre-trained SAM segmentation and ResNet18 classification models. During Phase Online, the BraNet app was developed using the React Native framework, offering a modular deep-learning pipeline for mammography (DM) and ultrasound (US) breast imaging classification. This application operates on a client-server architecture and was implemented in Python for iOS and Android devices. Then, two diagnostic radiologists were given a reading test of 290 total original RoI images to assign the perceived breast tissue type. The reader's agreement was assessed using the kappa coefficient. **Results:** The BraNet App Mobil exhibited the highest accuracy in benign and malignant US images (94.7%/93.6%) classification compared to DM during Training I (80.9% /76.9%) and Training II (73.7 / 72.3 %). The information contrasts with radiological experts' accuracy, with DM classification being 29%, concerning US 70% for both readers, because they achieved a higher accuracy in US ROI classification than DM images. The kappa value indicates a fair agreement (0.3) for DM images and moderate agreement (0.4) for US images in both readers. **Conclusion:** It means that not only the amount of data is essential in training deep learning algorithms. Also, it is vital to consider the variety of abnormalities, especially in the mammography data, where several BI-RADS categories are present (microcalcifications, nodules, mass, asymmetry, and dense breasts) and can affect the API accuracy model.

Key Words: Breast cancer; Mobil app; deep learning; ultrasound; mammography

1.Introduction

Today, in the healthcare landscape, artificial intelligence tools hold great promise for clinicians by enhancing breast cancer diagnostics and tailoring treatment strategies to match the disease's characteristics [1-3]. However, in the same line, there are some alternatives, such as command-line tools with shell scripts [4] and manual, semi-automated, and fully automated methods for image processing [5]; these options are not user-friendly for specialists and researchers without a background in computer science. Furthermore, the available graphical interface tools are often task-specific [6-7], focusing on contour delineation, segmentation, or classification.

In this context, radiomics constitutes an emerging field in medical imaging and offers the potential to extract diagnostic and prognostic information from 2D grayscale images by analyzing lesion features [1-2]. Hence, graphical, and mobile tools are elevating the role of radiomics in biomedical research, potentially serving as a second opinion for radiologists in breast lesion detection. Specifically, Computer-Aided Diagnosis (CAD) systems based on deep/machine learning (DL/ML) play a crucial role in addressing various computer vision challenges, such as medical image pre-processing with super-resolution [8-11] and denoising, data augmentation [12-15], medical image segmentation [16-18] (e.g., NiftyNet [6], MIScnn [16] and NiftySeg [17]), image classification [19], computer-assisted interventions [5], image recognition [20], and annotation [5].

In the context of detecting cancer, there are several radiomic projects, CAD based on deep/machine learning (DL/ML) systems, and studies that propose different artificial intelligence techniques that help to provide decision support for many applications in the patient care processes, such as lesion detection, characterization, cancer staging, and treatment planning. The major challenge in this field of research is to build a fully automatic CAD system that can analyze large quantities of images to provide an accurate diagnosis and, at the same time, robust enough to handle the biological variations in humans [21].

The most successful DL algorithms used in the processing of medical images are Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), and Recurrent Neural Networks (RNNs), which play a crucial role in improving healthcare outcomes by providing accurate and efficient analysis in processing medical images, each offering unique capabilities in data augmentation, pattern recognition, and feature extraction. [22]. In the early detection of breast cancer, CAD systems have several stages: (i)image collection, (ii) annotation and detection of tumors based on the region of interest (ROI), (iii) segmentation, (iv) classification based on the ROI shape using deep learning models and (v) performance evaluation of the models [23-24].

Image collection and annotation are the main challenges in performing large-scale medical image analysis using DL algorithms. Some CNNs-based options to consider as mask segmentation for detected tumors in medical images are You Only Look Once (YOLO) [25], Region-Based Convolutional Neural Network (R-CNN) [26,27] and their variants (Mask R-CNN [26] and Faster R-CNN [27]), deep neural networks such as Natural Language Process (NLP), which can help us to automatically identify and extract relevant information from radiology clinical reports and images [28].

Although there is a variety of CAD systems developed concerning breast cancer, it is also important to mention that there are systems deployed in mobile applications for use in the smartphone, e.g., in [29] an automated breast cancer diagnosis system on mobile phones for taking photos of ultrasound reports was implemented. The authors include the automatic extraction of intricate image features by convolutional neural networks (CNNs) and the precise classification of breast masses. It eliminates the need for manual feature engineering and reduces human error. These applications streamline the diagnostic process, increase efficiency, and, most importantly, enhance patient outcomes by providing reliable, consistent, and accessible early breast cancer detection and treatment tools.

1.1Related Work

In this section, we will briefly introduce NLP and CNN modelling as more recent approaches using neural networks and discuss how several authors have used these models in radiomics and biomedical applications.

One of the main fundamentals of NLP is extracting image information using patterns such as the accession number, series number, and image number. Information about the imaging modality, magnetic resonance imaging (MRI), CT (computed tomography), positron emission tomography (PET), Ultrasound (US), and Mammography imaging can be relevant, too. It can be extracted from the accession number and image number, where the patient identification number (ID) can be appropriate if the patient's history is of interest.

Linna et al. [30] indicate that NLP tools in radiology and other medical settings have been used for information retrieval and classification. NLP-based algorithms have opened more possibilities for medical image processing, detecting findings, and giving possible diagnoses [31]. Wang et al. [32] suggested that cancers are the most common subject area in NLP-assisted medical research on diseases, with breast cancers (23.30%) and lung cancers (14.56%) with the highest proportion of studies. Also, Luo et al. [33] specified that NLP is useful for creating new automated tools that could improve clinical workflows and unlock unstructured textual information contained in radiology and clinical reports for the development of radiology and clinical artificial intelligence applications.

Otal et al. [34] proposed a machine learning system using WEKA algorithms to detect cancer staging classification. Buckley et al. [35] used NLP to extract clinical information from >76,000 breast pathology reports, the model of which demonstrated a sensitivity and specificity of 99.1% and 96.5% compared to expert humans. Chen et al. [36] proposed an NLP extraction pipeline system that accepts scanned images of operative and pathology reports. The system achieved 91.9% (operative) and 95.4% (pathology) accuracy. The pipeline accurately extracted outcomes data pertinent to breast cancer tumor characteristics, prognostic factors, and treatment-related variables. Liu et al. [37] implemented an NLP program to extract index lesions and their corresponding imaging features accurately from the text of breast MRI reports.

The NLP system demonstrated 91% recall and 99.6% precision in correctly identifying and extracting image features from the index lesions. The recall and precision for correctly identifying the BI-RADS categories were 96.6% and 94.8%, respectively. Kirillov et al. [38] created the NLP-based segment anything model (SAM) as a mask extraction and promptable segmentation task. Thus, it can transfer zero-shot [39] to new image distributions.

The NLP system demonstrated 91% recall and 99.6% precision in correctly identifying and extracting image features from the index lesions. The recall and precision for correctly identifying the BI-RADS categories were 96.6% and 94.8%, respectively. Kirillov et al. [38] created the NLP-based segment anything model (SAM) as a mask extraction and promptable segmentation task. Thus, it can transfer zero-shot [39] to new image distributions.

Likewise, in a CAD system, the classification task is an important step after the segmentation process. The most widely used deep learning-based algorithms for image classification are CNN models (ResNet [40], DenseNet [41], NasNet [42,43], VGG-16 [44], GoogLeNet [45], Inception-V3 [46]).

Several authors [47-51] have used models for benign and malignant breast mass classification. The CNN used for breast classification is divided into two main categories: (i) novo-trained model (e.g., Scratch) and (ii) Transfer Learning-based models that exploited previously trained networks (e.g., AlexNet, VGG-Net, GoogLeNet, and ResNet) [47]. In [48], the ResNet model was used as a classification training model using an original and synthetic mammography (DDSM) dataset, obtaining a performance of 67.6 and 72.5%, respectively.

In [49], several CNN models were proposed (GoogLeNet, Visual Geometry Group Network (VGGNet), and ResNet), to classify malignant and benign cells using average pooling classification. The results overcome all the other deep learning architectures in terms of accuracy (97.67%). However, the choice of architecture depends on the specific problem and involves commitments between factors such as model size, computational efficiency, and accuracy.

Despite the extensive availability of medical radiomic tool research and CAD-based deep learning systems [52-53], this technology has limited support within mobile app infrastructure for 2D breast medical image analysis. Consequently, the BraNet's workflow has two main phases Off-line and On-line, to achieve the following aims: (i) to develop a mobile app based on deep learning models for segmenting and classifying 2D breast images into benign and malignant lesions and (ii) to implement statistical metrics as a prediction performance evaluation tool.

2. Methods

2.1Data collection

We collected seven open-access breast image databases, including three datasets of breast ultrasound (US) images and four datasets of mammography images.

(i) Breast Ultrasound Images Dataset (BUSI): This dataset, gathered by [43], comprises 780 images (133 normal, 437 benign and 210 malignant).
(ii) Dataset A: collected by Rodrigues et al. [54] available at (https://data.mendeley.com/datasets/wmy84gzngw/1), Dataset A contains 250 breast US images (100 benign and 150 malignant).
(iii) Dataset B: Comprising 163 US images, these data were acquired from the UDIAT Diagnostic Centre of the Parc Tauli Corporation, Sabadell, Spain [55].
(iv) CBIS-DDSM: Curated Breast Imaging Subset–Digital Database for Screening Mammography, available at (https://n9.cl/qtl48), this database comprises 2620 cases [56].
(v) mini-MIAS (Mammographic Image Analysis Society): available at http://peipa.essex.ac.uk/info/mias.html), includes 322 (208 normal, 63 benign and 51 malignant images) Medio Lateral Oblique (MLO) mammograms from 161 patients [57-58].

(vi) Inbreast: This dataset comprises a total of 115 images and can be found at (https://biokeanos.com/source/INBreast) [59].

(vii) VinDr-Mammo: introduces a large-scale full-field digital mammography dataset of 5,000 four-view exams (https://physionet.org/content/vindr-mammo/1.0.0/) [60].

2.1 Pretraining models in Phase off-line

Data Normalization and Automatic ROI annotation

An ROI annotation is needed from a large dataset of US and mammography images from the above public database to improve the previously trained GAN and ResNet models and their computational performance. The breast images vary in size, see Table 1.

Thus, it is necessary to perform transformations and standardize the images taken at different sizes to a single dimension (128 x 128x 1 pixel). It was also necessary to transform it to a single channel (grayscale pixel) and normalize it in the range [-1,1] with a mean of 0.5 and a standard deviation of 0.5. The torch-vision (pytorch) library and Jupyter notebook algorithm (crop_vindr_images.ipynb) were used as the image annotation region processes to identify ROIs that may contain lesions.

Figure 1 details the overall process followed in this study.

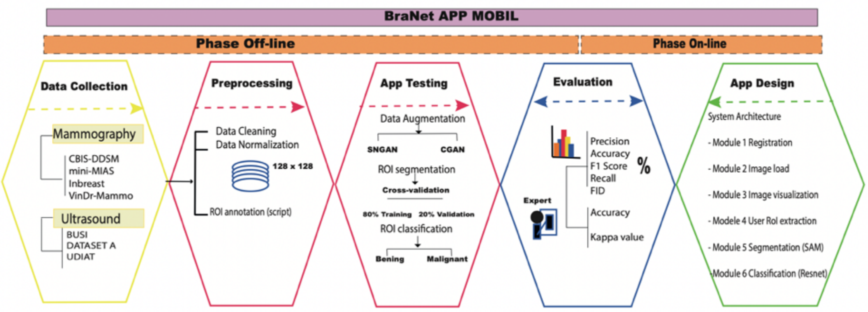


Fig. 1. The BraNet's workflow has two main phases Off-line: (i) Breast data collection from public databases. (ii) Data preparation and App testing and evaluation. On-line: (iii) System Architecture design. (iv) Statistical metrics comparison between the mobile application and human experts.

User ROI Extraction and Segmentation

As other studies have pointed out [61], to improve the detection accuracy, smaller patches (i.e., RoIs) where all breast masses and micros (e.g., cysts, calcification) are included inside this extracted area are generated from the original mammogram. In most mammogram images, 32% to 56% are background pixels, which do not contribute to breast cancer diagnosis.

In this research, the Segment Anything Model (SAM) [38] (an encoder-decoder architecture based on NLP prompt-based learning) [35] was trained as automatic ROI segmentation before being implemented in the Module 5 (BraNet application Phase On-line). SAM is an open-source software, and the quality of the segmentation masks was rigorously previously evaluated, with automatic masks deemed high quality and effective for training models, leading to the decision to include automatically generated masks.

NLP tasks include sentence boundary detection, tokenization, and problem-specific segmentations, and the *SamAutomaticMaskGenerato* function was used for automatic mask extraction [62]. SAM model is available under a permissive open license (Apache 2.0) at https://segment-anything.com.

The SAM predefined hyperparameters used are: points_per_side (32), points_per_batch (64), pred_iou_thresh (0.88), stability_score_thresh (0.95), stability_score_offset (1.0), box_nms_thresh (0.7), crop_n_layers (0), crop_nms_thresh (0.7), crop_overlap_ratio (512/1500), crop_n_points_downscale_factor (1), point_grids (Null), min_mask_region_area (0), output_mode ('binary_mask'). Also, SAM accuracy ROI segmentation is evaluated by the Intersection-Over-Union (Jaccard Index) metric [18] using *calculate_stability_score* function.

The model was trained on a large and diverse set of masks of Mammography and US images. These ROIs were previously extracted from the Mini-MIAS, Inbreast, and VinDr-Mammo databases (corresponding annotated bounding boxes are available in a .csv file). However, it is worth noting that its authors had already classified the ROIs from the CBIS-DDSM dataset; thus, no additional ROI extraction was required. Regarding US images, RoIs were extracted from the BUSI and Dataset B (UDIAT), excluding Dataset A, because it already contained RoIs.

A total of 6,592 breast ROI images were used for pre-trained SAM and GAN models, with 4,463 mammography images (benign and malignant) and 1,041 US images, as shown in Table 1.

Table 1. US and Mammography ROIs.

Type	Training	Database	Image size (pixels)	Benignant	Malignant
US	I	BUSI	500x500	427	201
		Dataset B	-	100	150
		BUS (UDIAT)	760 x570	109	54
		Total 1041		636	415
DM	I	Mini-MIAS	1024x1024	118	91
		INbreast	3328 x 4084 256 x 3328	2106	144
		CBIS-DDSM	3784 x 5912	1225	779
	II	VinDr-Mammo	3518 x 2800	893	1236
	Total 5892			4342	1550

Data Augmentation

This technique was previously employed in the Phase off-line to mitigate the risk of overfitting effectively. To generate new realistic images and improve BraNet's classification task performance, all ROIs were previously augmented by a GAN using the Spectral Normalization technique (SNGAN). SNGAN introduces a novel weight normalization technique known as spectral normalization to enhance the training stability of the discriminator network [63,64], serving as the foundation for synthetic image generation, which use Hinge Loss function (see Eq.1).

$$L_D = -E_{(x,y)\sim P_{data}}\left[\left(0,-1+D(x,y)\right)\right] - E_{z\sim P_{z,y\sim Pdata}}\left[\left(0,-1-D(G(z),y)\right)\right] L_G = -E_{z\sim P_{z,y\sim Pdata}} D(G(z),y) \quad \text{Eq. (1)}$$

Where P_{data} is the real data distribution, $P(z)$ is a prior distribution on noise vector z, $D(x)$denotes the probability that x comes from the real data rather than generated data, $E_{x \sim P_{data}}$ represents the expectation of *x* from real data distribution P_{data} and $E_{z \sim P_{(z)}}$ is the expectation of *z* sampled from noise.

The *clean-fid* library was used to obtain the FID value, using the Tensorflow and PyTorch libraries, some original implementations of the metric were taken from Parmar et al. [65]. The GAN model was trained using a cross-validation technique in Google Colab Pro 1 GPU model V100 with CUDA cores execution; with the hyperparameters detailed in the table 2.

Table 2. Hyperparameters tunning of deep learning models.

Hyperparameter	**SNGAN**	**ResNet**
	DM/US	**DM/US**
Batch Size	64/32	32/16
Image Size	128 x 128	128 x 128
Nro Epochs	100	100
Learning Rate	$2X10^{-4}/1.5X10^{-4}$	$2X10^{-5}$
Optimizer	Adam	Adam
Activation function	LeakyReLU	ReLu
β1	0.3	0.1
β2	0.999/ 0.75	0.9
Latent vector	100	-
Loss function	Hinge/ BCE	2.190
Optimization function (Discriminator)	LeakyReLU	-
Optimization function (Generator)	ReLU	-

Cross-validation analysis

The technique divides the dataset into multiple folds and trains (DM:Training I (4,463) and Training II(6,592) and US: Training I (1,041)) the model on different subsets while validating the remaining fold can provide a more robust estimate of the model's performance, effectively mitigating the risk of overfitting. It helps detect overfitting early and tune the model accordingly. A total of 6933 benign and malignant ROIs, were split into 80% training and 20% validation, using the Sklearn library from Pytorch. See Table 3.

Table 3. Training and Validation datasets of DM 80% (5273), 20% (1319) and US 80% (832), 20% (209) breast images.

Classes	**Training datasets**		**Validation dataset**	
Image type	**US**	**DM**	**US**	**DM**
Benning	505	3471	131	871
Malignant	327	1802	78	448
TOTAL	**832**	**5273**	**209**	**1319**

ROI classification process

Before implementing Module 6 (ROI Image Classification) in the BraNet mobil interface, the ResNet model was pre-trained on a large set of generated Mammography and US ROIs using also cross-validation technique.

ResNet18 training model

The ResNet18 CNN-deep learning-based classification model has been widely used in medical image classification, especially in breast lesion diagnosis and detection, and was chosen for its effectiveness in transfer learning, offering reduced training time and the automatic extraction of features [40]. This approach effectively mitigates the issues of vanishing or exploding gradients that can arise from increasing neural network depth, ultimately leading to improved accuracy [66-69].

Thus, to train the ResNet model and distinguish between malignant and benign breast lesions, the datasets were divided in two categories (i) DATASET A (Original + Synthetic ROIs) and (ii) DATASET B (Synthetic ROIs). The model consists of three convolutional layers and two fully connected layers. The kernel size for the first convolutional layer is 5 × 5, and, for the rest, 3 × 3. The size of the first and the second fully connected layers are 128 and 2 (the number of classes), with a dropout of 0.5. After the flattening and the first fully connected layers, the ReLU activation function for all layers except the output layer, where softmax was used. The model was pre-trained with the PyTorch library using a Google Colab Pro 1 GPU model V100 with CUDA cores execution; the training hyperparameters are outlined in Table 2.

2.2 System Architecture in Phase on-line

The system architecture consists of two primary components facilitating scalability and system maintenance: (i) The mobile application and (ii) The backend server, following a client-server architecture.

The mobile application is a client that communicates with the backend server to request services and image analysis. The backend server processes these requests and returns the results to the client for display (see Figure 2). The backend server was developed using React Native and was implemented in the Python programming language.

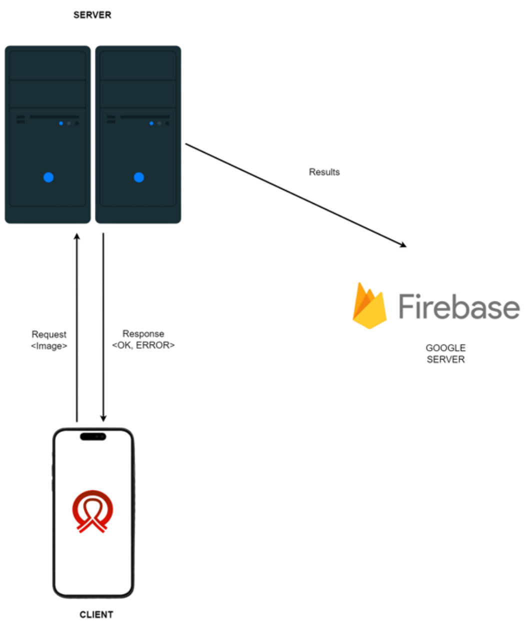


Fig. 2. Client-Server architecture to BraNet App

The mobile application comprises several interrelated components:

- Module 1: Registration, Login Synchronization, and User Profile Information. Data generated by the application, such as images and metadata, is stored in Firebase (a mobile and web application development platform). Firebase is also used for user authentication, mobile application registration, and log in.
- Module 2: Image Import allows users to upload breast ultrasound and mammography images in PNG, JPG, and JPEG formats, with a maximum file size of 10 MB.
- Module 3: Visualization Area (a history of image analysis results, image analysis capabilities, and user assistance).
- Module 4: User ROIs Extraction.
- Module 5: ROI Segmentation (See 2.2.2 section).
- Module 6: ROI Image Classification (See 2.2.4 section).

2.2.1 Evaluation metrics in Phase off-line and On-line

Quality of Synthetic Image

The FID and KID quantitative feature-based metrics have been applied to evaluate the quality of real and synthesized ROIs generated by GANs models and to compute the distance between the vector representation of the synthesized and authentic images.

Fréchet Inception Distance (FID): FID compares the distributions of the original and synthetic images to assess how well the generated images represent the training dataset. Lower FID scores indicate better quality images [70], and it is calculated as shown in Eq. (2):

$$FID = \|\mu_r - \mu_g\|^2 + Tr(\textstyle\sum_r + \sum_g - 2(\sum_r \sum_g)^{1/2}) \qquad \text{Eq (2)}$$

Here, μ_r represents the mean of the feature vector calculated from the real images, μ_g is the mean of the feature vector calculated from the fake images, Σ_r is the covariance of the feature vector from the real images, and Σ_g is the covariance of the feature vector from the fake images.

Kernel Inception Distance (KID): KID employs the cubic kernel to compare the skewness, mean, and variance [70]. A lower KID value signifies a higher visual similarity between the actual and generated images. The cubic polynomial kernel is defined as shown in Eq. 3):

$$k(x,y) = \left(\frac{1}{d}\ x^T y + 1\right)^3 k(x,y) = \left(\frac{1}{d}\ x^T y + 1\right)^3 \quad (3)$$

where *d* represents the dimension of the feature space for vectors *x* and *y*.

BraNet's classification performance evaluation

For assessing the *BraNet's* classifier's performance, we employed a confusion matrix, F1 score (Dice), accuracy (Acc), precision (Prec), sensitivity (Sen) o Recall, and specificity (Spec) [24] metrics (see Tables 4 and 5). The accuracy of the model was calculated using statistical score libraries such as the classification report and confusion matrix from the Python sci-kit-learn module.

Table 4. Confusion matrix to distinguish between two classes (benign, and malignant). TP: true positive; FN: false negative, FP: false positive, TN: true negative, PPV: positive predictive value, NPV: negative predictive value.

		Predicted Classes		
	Classes	**Positive**	**Negative**	**Measures**
		C_1 (Benign)	**C_2**(Malignant)	
Actual Classes	C_1 (Benign)	TP	FP	PPV
	C_2(Malignant)	FN	TN	NPV
	Measures	Sen	Spec	Acc

Table 5. Validation assessment metrics.

Model	**Equation**
Accuracy	$Acc = \left(\frac{TP+TN}{TP+TN+FP+FN}\right)$
Sensitivity	$Sen = \left(\frac{TP}{TP+FN}\right)$

Specificity	$Spec = \left(\frac{TN}{TN + FP}\right)$
Precision	$Prec = \left(\frac{TP}{TP + FP}\right)$
F1 score	$F1\ score = 2\text{ x}\left(\frac{Prec\ x\ recall}{prec+recall}\right)$

Human Expert *Evaluation*

Two senior radiologists were asked to assess, annotate, and classify images independently to ensure that BI-RADS categories are correctly assigned. Representative original ROI images for each breast type is available in https://drive.google.com/drive/folders/1HMeqPfI8qL58hAqwVpZupH6uq4W_kHrl.

A comparison between the images tested by human experts and those annotated in public databases was conducted using an independent test set of 212 mammography images (47 malignant and 165 non-malignant) and 78 US images (24 malignant and 54 non-malignant).

Two diagnostic radiologists (reader 1 with 20 years of experience and reader 2 with 13 years of experience) were given a reading test consisting of 290 total original RoI images to assign the perceived breast tissue type. The readers rated each image as (1) benign or (0) malignant.

Kappa coefficient *and overall accuracy*

Furthermore, the agreement between the two readers' answers (considering all elements of error matrix) was assessed by determining the Kappa coefficient (K), using the ranges between 0 (when there is no agreement) and 1 (when there is substantial agreement), and is calculated using the Eq. (4). The error matrix was calculated by comparing the two readers' choices from five possibilities and was interpreted as follows: <0.2 slight; 0.21–0.40 fair; 0.41–0.60 moderate; 0.61–0.80 high; 0.81–1.0 almost perfect [71].

$$k = \frac{P_0 - P_e}{1 - P_e} \qquad \text{Eq. (4)}$$

P_o is the correctly allocated samples (agreement cases), and P_e is the hypothetical random agreement.

The overall accuracy (Eq. 5) allows the description of model performance and is calculated by dividing the total number of correctly classified samples by the total number of samples.

$$Acc = \frac{C_S}{N_S} \quad \text{Eq. (5)}$$

C_s is the number of correct samples classified, and N_s is the total number of samples.

3. Results

The main results are categorized into two Phases, Off-line and On-line. First, we introduced the preprocessing and pre-training models section with data augmentation (GAN), segmentation (SAM) and classification (ResNet) algorithms. Then, we presented the On-line phase with a practical utility of BraNet's user interface, ant its modules.

3.1 Preprocessing and Pretrained models

3.1.1.1 Synthetic data to feed the classification network.

A significantly number of synthetic RoIs (10,000 (Training I) and 2,000 (Training II) mammography RoIs and 4,000 US RoIs (Training I)) were generated by SNGAN to feed the classifier. The loss function and accuracy of the generator and discriminator play a crucial role in assessing the training stability and performance of GANs. A stable GAN is characterized by a discriminator loss around 0.5 or higher than 0.7, while the generator loss typically ranges from 1.0 to higher values like 1.5, 2.0, or even more. The accuracy of the discriminator, both on real and synthetic images, is expected to hover around 70% to 80%. Appendix B presents the accuracy plot to SNGAN.

The average FID and KID values in SNGAN are 58.80/0.052 and 52.34/0.051 for mammography Training I and Training II respectively, and 116.85/0.06 for the Training I in US (see Appendix C). The lowest values indicate that SNGAN-generated synthetic images closely resemble to the original mammography and US images in clinical characteristics, suggesting their potential utility in clinical data augmentation and training, particularly for enhancing diagnostic skills in breast imaging.

3.1.1.2 Resnet training model

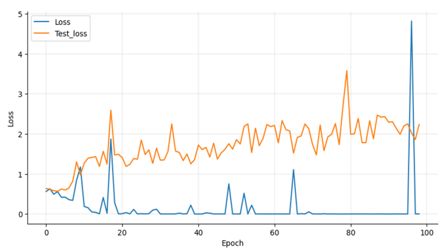
The model shows the highest accuracy in US image classification (see Table 6) concerning the mammography dataset. Although the network received more mammography images (6592) as input (Mini-MIAS, Inbreast, CBIS-DDSM, VinDr-Mammo) with respect to the small number (1041) of US data (BUSI, UDIAT, DATASET A).

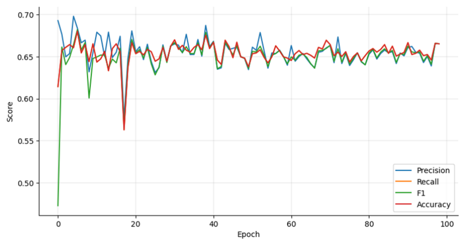
It means that not only the amount of the data is important to train deep learning algorithms. Also, it is important to considerer the variety of abnormalities especially in the mammography data, such as microcalcifications, nodules, mass, asymmetry, and dense breasts, because it can improve the accuracy of the ResNet training model.

Table 6. Resnet statistical performane evaluation in US and DM image classification.

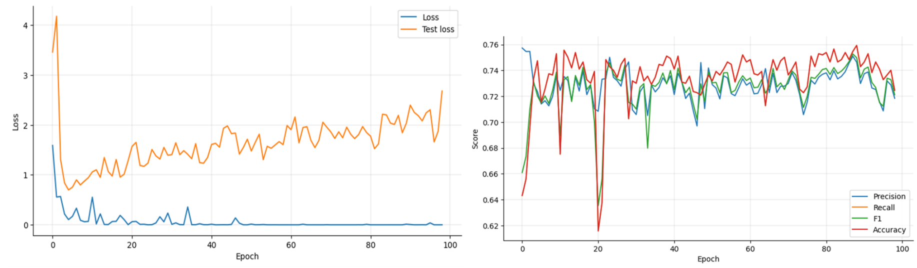
RESNET18	IMAGE MODALTY					
	TRAINING I (1041)		TRAINING I (4463)		TRAINING II (6592)	
IMAGE TYPE	US (%)		DM (%)		DM(%)	
CLASSES	Benignant	Malignant	Benignant	Malignant	Benignant	Malignant
ACCURACY	94.7	93.6	80.9	76.9	73.7	72.3
PRECISION	97	89	92	56	84	59
RECALL	93	95	81	77	74	72
F1-SCORE	95	92	86	65	78	65

Therefore, it is essential to monitor the evolution and performance of the models using training and validation datasets, see Figures 3 (a-d). Figure3aand 3c displays the loss and accuracy values concerning each epoch during the ResNet training and validation model using mammography and US images, while Figure 3b and 3d shows the accuracy, F1 score, recall, and precision by each epoch in both image types. Appendix D shows the details of the network training and validation dataset.





(a) (b)



(c) (d)

Figure 3. Training II and testing plots for Mammography images a. Loss vs Acc (Real ROI data). b. Loss vs Acc (Real + Data augmentation). c. Evaluation metrics (Real ROI data) and d. Evaluation metrics (Real + Data augmentation).

The BraNet App Mobil exhibited the highest accuracy in benign and malignant US images (94.7%/93.6%) classification compared to DM during Training I (80.9% /76.9%) and Training II (73.7 / 72.3 %). And the Resnet model does not improve the accuracy of benign and malignant ROI lesion classification during Training II compared to the previous Training I.

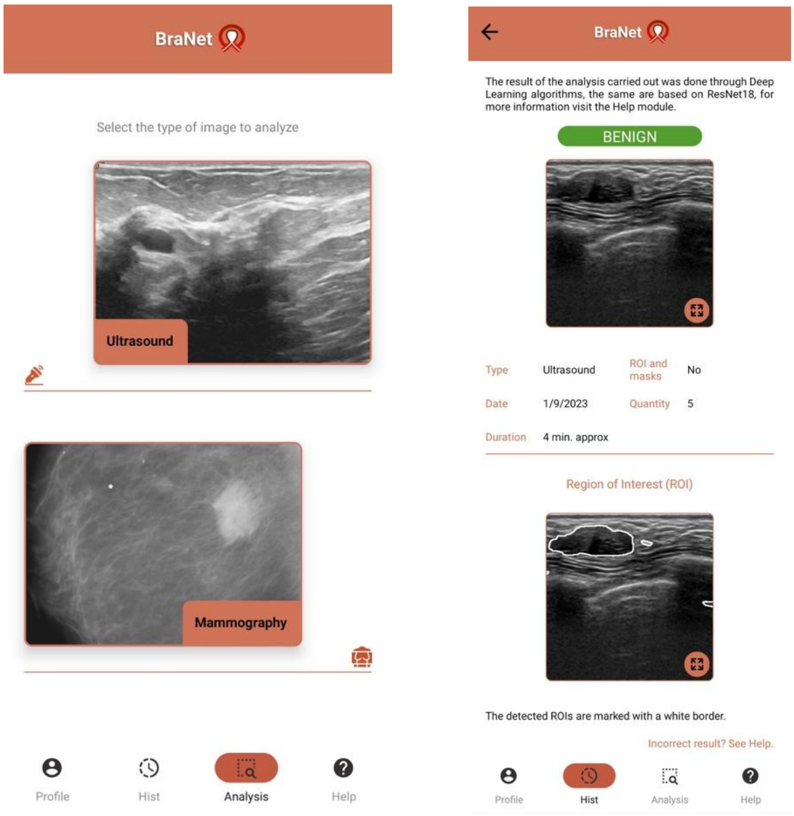
3.1.2 The BraNet and its Graphical User Interface (GUI)

The mobile app's user interface was developed using Python v3.11 and React Native as a JavaScript framework for creating native mobile applications compatible with iOS and Android platforms. The interface is composed of several modules, each serving distinct purposes:

Module 1: Registration, Login Synchronization, and User Profile Information: This module handles user registration and login functionalities, synchronizing user data and providing access to user profile information.

Module 2: Image Import: Users can import images in standard picture formats, such as JPG, JPEG, and PNG, with a maximum size limit of 10MB.

Module 3: Visualization Area: This area displays loaded images. The selected image is displayed in grayscale, preserving the original image's aspect ratio (see Figure 4a).



(a) (b)

Fig. 4. BraNet user interface within the toolbox. (a) Upload the breast image type. (b) US breast ROI selection and classification as benign class.

Module 4: Manual ROIs Extraction: This module allows users to manually or semi-automatically create masks and define ROIs within the selected image. Masks are represented as binary matrices with the same dimensions as the loaded mammography image, where true values indicate the ROIs. Users can customize the size and sampling method for RoIs.

Module 5: ROI Segmentation: Users can segment a subset or the entire set of features from the segmentation section (as shown in Figure 4b). Before performing calculations on the image, the user must add at least one ROI in the "Regions and Masks."

Module 6: ROI Image Classification: This module employs the ResNet18 model to classify ROI images into benign and malignant classes. Example output classes are provided in Figure 4b.

The BraNet's graphical user interface enhances the user experience by providing intuitive image analysis and classification tools, making it a valuable resource for medical professionals and researchers. The practical usage of the tool can be accessed via the following link: https://drive.google.com/file/d/1d1vnjQ6LqOd0fdz65eaVg791d7cFRPWO/view

3.1.2.1 Comparison of the BraNet with human experts' evaluation

The accuracy percentages of correct rates between benign and malign images classification for readers 1 and 2 were 29% and 42% respectively. The reader agreement was assessed using the kappa coefficient, which values are 70% and 71% in mammography and US classification, respectively. Table 7 indicates a fair agreement (0.3) for mammography images and moderate agreement (0.4) for ultrasound images in both readers, with a change in prevalence from the lowest in US images to the highest value in mammography images, resulting in a corresponding change in sensitivity (19.2/60) to specificity (51/84.4) percentage points. This effect was statistically significant ($p < 0.05$) for either sensitivity or specificity in both image types.

Table 7. Interrater Reliability, Cohen's Kappa, and statistical values for 2 Raters in both classes.

Method	***DM***	***US***
Subjects	212	78
Agreement %	70	71
Kappa	0.294	0.426
p-value	< .001	< .001
Sensitivity	19.2 %	60.0%
Specificity	51.6%	84.8%
Prevalence	69.5%	57.7%
Accuracy	29.0%	70.5%
PPV	47.5%	84.4%
NPV	21.9%	60.9%

Discussion

The pressing need to transition automatic medical image classification by CAD systems from research laboratories into practical clinical applications is evident. BraNet's aims to provide an API for setting up a breast image classification pipeline with ROI mask extraction and segmentation capabilities. The tool offers an open-source solution for processing US and mammography images, complete with statistical metrics for evaluating model performance.

There have been many published examples of AI algorithms that demonstrate excellent performance in cancer detection for screening mammography. These include several algorithms trained and evaluated on private and public data sets. Table 8 compares BraNet's performance against other state-of-the-art medical image classification applications.

Table 8. Comparison of the BraNet´s performance against other deep learning applications.

Author	Application name	Description	Acc/Sen/Spec/Prec/AUC (%)
Gibson et al. [6]	NiftyNet	A DL open-source platform used in three medical image analysis applications (MRI, CT and US). Including a Conditional GAN model as ultrasound image generation.	88/7.5/9.1/-/-
Pang et al. [72]	TripleGAN	Method to perform data augmentation in breast US images and feed a CNN mode to classify breast masses.	90.41/87.94/ 85.86/-
Al-Dhabyani et al. [43]	AlexNet+GAN (CNN) -BUSI -Dataset B -BUSI+ DatasetB	US breast classification with data augmentation. The model examines two different methods: a CNN approach and a Transfer Learning (TL). The results confirm an overall enhancement using augmentation methods with TL classification methods.	78/-/-/-/- 80/-/-/-/- 84/-/-/-/-
	TL -BUSI -Dataset B -BUSI+ DatasetB		94/-/-/-/- 92/-/-/-/- 99/-/-/-/-

Jiménez et al. [73]	Radiomic tool	Colposcopy image classification combining UNET+SVM as segmentation and classification cervix abnormalities.	80/70/48.8/-
Dihge et al. [74]	NILS	A web-based tool for noninvasive lymph node staging in breast cancer.	-/90/34/-/71
To T. et al. [75]	DUV-WSI	DUV-WSI Deep ultraviolet (DUV) fluorescence scanning microscopy provides rapid whole-surface imaging (WSI) of breast tissues. Images are split into small patches (512 x512), and features are extracted using a pre-trained Resnet 50 as patch classification.	81.7/91.7/66.7/-/-
Qi et al. [29]	**Deep-CAD system**	The breast cancer system is deployed on mobile phones, takes a photo of the US as input, and performs diagnosis on each image. Then the system to classify images into malignant and non- malignant using CNNs.	89.34/87.31/87.49
Ours	BraNet	A deep learning tool for breast regions classification using mammography and US images. DM (TRAINING I)	Acc 94.7/93.6 Prec 97/89 Recall 93/95 F1 score 95/92
		DM (TRAINING II)	Acc 93.7/ 72.3 Prec 84/ 59 Recall 74/72 F1 score 78/65
		US (TRAINING I)	Acc 80.9/76.9 Prec 92/56 Recall 81/77 F1 score 86/ 65

However, there is a significant gap in understanding how these AI applications will perform with multimodal images in the real world when radiologists use them in clinical practice [76-77].

The BraNet Mobile App is an open interface for classifying specific 2D breast image types using deep learning models. It is believed that this is the first system for breast cancer diagnosis deployed on mobile phones to both types of images. The API's development comprises two main phases: (i) Off-line to pre-train the deep algorithms and (ii) Online to release the app, which includes several modules, including Model Selection, Model Extraction (by a human expert), Segmentation (SAM model), Model Classification (ResNet18 model), and Model Evaluation.

During the Off-line phase the pre-trained GAN algorithm was implemented as synthetic image generation, and the image quality was evaluated by two feature-based metrics FID and KID. It is widely acknowledged that the preprocessing images, quality, and diversity of the training dataset greatly impact the training of GAN and CNN deep learning models [78-80]. The lower FID and KID values mean a higher visual similarity between the real and generated images. The results (Appendix C) indicate that the SNGAN model is suitable for mammography and US synthetic data generation with average values of FID =52.4/ KID =0.051 for mammography and FID =116.85/ KID =0.06 for US.

With these datasets DATASET A (Original + Synthetic ROIs) and DATASET B (Real ROIs) the classification model was trained. Table 6 and Figures 3 shows the accuracy results averaged in BraNet ROI classification are: (i) Training I in US (94.7 (B) /93.6 (M)) and DM (80.9 (B) /76.9 (M)) and (ii) Training II in DM (73.7 (B) /72.3 (M)).The result demonstrated that Resnet model during Training II with original +synthetic images (where the VirDrMammo database was added) did not improve the accuracy (73.7 / 72.3 %) concerning Training I (80.9/76.9). In comparison, with radiological experts, accuracy in DM was 29% concerning with 70% in DM for both readers. These results show that both API and Readers obtained a better percentage of accuracy in classifying the ROIs of mammography images than US images.

A final comparison between BraNet and radiological experts' evaluation demonstrates that for the all-breast image types, reader accuracy was higher with US images (75%) than with original ROI images from public databases. The reader agreement was 70% and 71% in mammography and US classification, respectively. The kappa value indicates a fair agreement (0.3) for mammography images and moderate agreement (0.4) for ultrasound images in both readers (Table 7). This can be contrasted with BraNet classification accuracy (Table 8), where the API shows the highest accuracy in US image classification (Table 6) concerning the mammography dataset. Although the network received more mammography images (5892) with respect to US (1041). It means that not only the amount of the data is important to train deep learning algorithms. Also, it is important to considerer the variety of abnormalities especially in the mammography data, where several BI-RADS categories are present (microcalcifications, nodules, mass, asymmetry, and dense breasts), and can be affect the accuracy in the ResNet training model.

According to the previous results, some limitations in implementing BraNet must be addressed in future work. One is the need to classify and characterize images based on different abnormalities, such as architectural distortion, asymmetry, mass, and microcalcification. BraNet no was trained using different breast tissue types and variations in mammography and US imaging techniques; the ROI classification process was performed only using two classes 1 (Benignant according to BI-RADS 1-3) and 0 (Malignant according to BI-RADS 4-6) categories. Oyelade et al. [81], indicates that is better to focus on previously classified and characterizing abnormalities into architectural distortion, asymmetry, mass, and microcalcification so that training distinctively learns the features of each abnormality. It generates sufficient images for each category before training a GAN model.

Thus, in future work, we plan to annotate the datasets with more fine-grained classes to get more targeted training in GAN and CNN models. Moving forward, we should consider pre-processing with denoising, super-resolution, improving the overall image quality and reducing blur and artifacts. Also, previous breast tissue types of classification are needed to obtain a diverse range of synthetic data, resulting in a more accurate image generation and classification process using GAN and convolutional algorithms. We must also compare our image classification with other TL models, such as Nasnet and Densenet, to ensure we use the most effective techniques.

An updated version of the BraNet application and prospectively explore the real AI/human interaction could be implemented, which can recognize full 2D images and not only resized images of 128x128 pixels. The App could be used for performance and load testing to assess how the application processes many images simultaneously. Simulate an increasing number of users or requests to see how the application performs under progressively higher loads.

Implement the App as a Web Server and realize scalability Testing. Incrementally increase resources (like CPU, GPU, memory) available to the application and measure performance improvement to determine how efficiently the application scales. Make full use of available CPU/GPU cores to process images in parallel, enhancing throughput. Utilize image compression techniques to reduce the size of high-resolution breast images without losing critical details necessary for analysis.

Finally, the use of IA in medical diagnosis brings about a range of ethical considerations that must be carefully navigated to ensure that the integration of these technologies benefits patients, healthcare providers, and the broader healthcare system responsibly and equitably. It is essential to highlight ethical considerations regarding using artificial intelligence in developing CAD systems in healthcare.

The patient's well-being is paramount, necessitating a comprehensive approach to protecting their data privacy and confidentiality [82-83]. This project ensures patient privacy through the anonymization and coding of training image databases during the application's first and second modules, which are also publicly available.

Another ethical consideration is the fairness of AI models [84], which requires providing equitable healthcare outcomes across various patient demographics. Thus, the developed application aims to contribute to medical service equity, particularly by facilitating pathology diagnosis in rural groups and sectors typically deprioritized in healthcare, especially in developing countries.

Finally, transparency regarding the capabilities and limitations of CAD systems is fundamental [85], ensuring that medical staff and patients know that decisions and outcomes adhere to ethical standards. In this context, the developed application is merely a test prototype that aspires to achieve the necessary maturity for use in a real healthcare setting, ensuring the requisite medical reliability.

Conclusions

In this paper, we have introduced BraNet, a Mobile App for Breast Image Classification based on Deep Learning algorithms. The API enables the rapid construction of breast image classification workflows, encompassing data input/output, ROI mask extraction, segmentation, and evaluation metrics. The Client-Server architecture, coupled with its open interface, empowers users to customize the pipeline and swiftly establish comprehensive medical image classification setups using Python libraries and the React Native framework for creating native mobile applications on iOS and Android. We have demonstrated the functionality of the BraNet app by conducting automatic cross-validation on data augmentation, ROI segmentation, and classification using public ultrasound and mammography datasets, resulting in a preclinical tool. After implementing some improvements and future updates, BraNet will facilitate the migration of medical image segmentation and classification from research laboratories to practical applications. Also, ensuring that the App complies with all regulations and standards governing data privacy and security in healthcare is essential. It is only a preclinical testing phase; thus, there is still work to be done in this area. BraNet currently offers a pipeline for breast image segmentation and classification, and it will continue to receive regular updates and extensions in the future. This data must be rigorously analyzed, reported, and often published in scientific journals to ensure its accuracy and reliability.

SUPPLEMENTARY MATERIALS

The following supporting information can be downloaded at: Video S1: https://drive.google.com/file/d/1d1vnjQ6LqOd0fdz65eaVg791d7cFRPWO/view The BraNet framework can be directly installed from: https://expo.dev/artifacts/eas/5BCL5XQNxZ1vGrRMaV4t1p.apk

DATA AVAILABILITY STATEMENT: All codes are available as Mendeley Data: https://data.mendeley.com/preview/jh9trvbjbv?a=57b040ca-ae6d-4ebb-bc04-ac8c27deae59 [86].

Author Contributions: Conceptualization, Y.J.G., M.J.R.-Á and V.L.; study conception and design: Y.J.G, M.J.R.-Á, D.C.M, VL; formal analysis, Y.J.G., M.J.R.-Á. and V.L.; investigation, Y.J.G.; Data Collection Y.J.G and S.G.J; analysis and interpretation of results: Y.J.G, S.G.J;D.C.F and P.C.D writing—original draft preparation, Y.J.G, M.J.R.A, D.C.M, VL writing—review and editing, Y.J.G, M.J.R.A, D.C.M, D.C.F, P.C.D and VL visualization, Y.J.G.; supervision, M.J.R.-Á. and V.L.; project administration, M.J.R.-Á.; funding acquisition M.J.R.-Á. All authors reviewed the results and approved the final version of the manuscript.

References

1. Lambin, P., Leijenaar, R. T., Deist, T. M., Peerlings, J., De Jong, E. E., Van Timmeren, J., ... & Walsh, S. (2017). Radiomics: the bridge between medical imaging and personalized medicine. Nature reviews Clinical oncology, 14(12), 749-762.

2. Papadimitroulas, P., Brocki, L., Chung, N. C., Marchadour, W., Vermet, F., Gaubert, L., ... & Hatt, M. (2021). Artificial intelligence: Deep learning in oncological radiomics and challenges of interpretability and data har-moni-zation. Physica Medica, 83, 108-121.

3. Castiglioni, I., Rundo, L., Codari, M., Di Leo, G., Salvatore, C., Interlenghi, M., ... & Sardanelli, F. (2021). AI applications to medical images: From machine learning to deep learning. Physica Medica, 83, 9-24.

4. Wollny, G., Kellman, P., Ledesma-Carbayo, M. J., Skinner, M. M., Hublin, J. J., & Hierl, T. (2013). MIA-A free and open source software for gray scale medical image analysis. Source Code for Biology and Medicine, 8(1), 1-20.

5. Philbrick, K. A., Weston, A. D., Akkus, Z., Kline, T. L., Korfiatis, P., Sakinis, T., ... & Erickson, B. J. (2019). RIL-contour: a medical imaging dataset annotation tool for and with deep learning. Journal of digital imaging, 32, 571-581.

6. Gibson, E., Li, W., Sudre, C., Fidon, L., Shakir, D. I., Wang, G., ... & Vercauteren, T. (2018). NiftyNet: a deep-learning platform for medical imaging. Computer methods and programs in biomedicine, 158, 113-122.

7. Papademetris, X., Jackowski, M. P., Rajeevan, N., DiStasio, M., Okuda, H., Constable, R. T., & Staib, L. H. (2006). BioImage Suite: An integrated medical image analysis suite: An update. The insight journal, 2006, 209.

8. Zhang L, Dai H, Sang Y. Med-SRNet: GAN-Based Medical Image Super-Resolution via High-Resolution Repre-sentation Learning. Comput Intell Neurosci. 2022 Jun 20;2022:1744969. doi: 10.1155/2022/1744969. PMID: 35747717; PMCID: PMC9210125.

9. Y. Zhu, Z. Zhou, G. Liao and K. Yuan, "Csrgan: Medical Image Super-Resolution Using A Generative Adversarial Network," 2020 IEEE 17th International Symposium on Biomedical Imaging Workshops (ISBI Workshops), Iowa City, IA, USA, 2020, pp. 1-4, doi: 10.1109/ISBIWorkshops50223.2020.9153436.

10. Zhang, K.; Hu, H.; Philbrick, K.; Conte, G.M.; Sobek, J.D.; Rouzrokh, P.; Erickson, B.J. SOUP-GAN: Su-per-Resolution MRI Using Generative Adversarial Networks. Tomography 2022, 8, 905-919. https://doi.org/10.3390/tomography8020073

11. Ahmad, W., Ali, H., Shah, Z. et al. A new generative adversarial network for medical images super resolution. Sci Rep 12, 9533 (2022). https://doi.org/10.1038/s41598-022-13658-4

12. Bargsten, L., Schlaefer, A. SpeckleGAN: a generative adversarial network with an adaptive speckle layer to augment limited training data for ultrasound image processing. Int J CARS 15, 1427–1436 (2020). https://doi.org/10.1007/s11548-020-02203-1

13. Haque, A. (2021, May). EC-GAN: Low-Sample Classification using Semi-Supervised Algorithms and GANs (Student Abstract). In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 35, No. 18, pp. 15797-15798).

14. Y. Sun, P. Yuan and Y. Sun, "MM-GAN: 3D MRI Data Augmentation for Medical Image Segmentation via Gen-erative Adversarial Networks," 2020 IEEE International Conference on Knowledge Graph (ICKG), Nanjing, China, 2020, pp. 227-234, doi: 10.1109/ICBK50248.2020.00041.

15. Bashar, M. A., & Nayak, R. (2020). TAnoGAN: Time series anomaly detection with generative adversarial net-works. In 2020 IEEE Symposium Series on Computational Intelligence (SSCI) (pp. 1778-1785). IEEE.

16. Müller, D., Kramer, F. MIScnn: a framework for medical image segmentation with convolutional neural networks and deep learning. BMC Med Imaging 21, 12 (2021). https://doi.org/10.1186/s12880-020-00543-7

17. Cardoso, M., Clarkson, M., Modat, M., & Ourselin, S. (2012). NiftySeg: open-source software for medical image segmentation, label fusion and cortical thickness estimation. In IEEE International Symposium on Biomedical Imaging, Barcelona, Spain.

18. Jiménez Gaona, Y.; Castillo Malla, D.; Vega Crespo, B.; Vicuña, M.J.; Neira, V.A.; Dávila, S.; Verhoeven, V. Ra-diomics Diagnostic Tool Based on Deep Learning for Col-poscopy Image Classification. Diagnostics 2022, 12, 1694. https://doi.org/10.3390/diagnostics12071694

19. Chen, Y., Zhang, Q., Wu, Y., Liu, B., Wang, M., & Lin, Y. (2018). Fine-tuning ResNet for breast cancer classification from mammography. In The International Conference on Healthcare Science and Engineering (pp. 83-96). Springer, Singapore.

20. C. M. J. M. Dourado, S. P. P. da Silva, R. V. M. da Nobrega, P. P. Rebouças Filho, K. Muhammad and V. H. C. de Albuquerque, "An Open IoHT-Based Deep Learning Framework for Online Medical Image Recognition," in IEEE Journal on Selected Areas in Communications, vol. 39, no. 2, pp. 541-548, Feb. 2021, doi: 10.1109/JSAC.2020.3020598.

21. Lee, H., & Chen, Y. P. P. (2015). Image based computer aided diagnosis system for cancer detection. Expert Systems with Applications, 42(12), 5356-5365.

22. Chowdhury, D., Das, A., Dey, A., Sarkar, S., Dwivedi, A. D., Rao Mukkamala, R., & Murmu, L. (2022). ABCanDroid: A cloud integrated android app for noninvasive early breast cancer detection using transfer learning. Sensors, 22(3), 832.

23. Jiménez-Gaona, Y., Rodríguez-Álvarez, M. J., & Lakshminarayanan, V. (2020). Deep-learning-based computer-aided systems for breast cancer imaging: a critical review. Applied Sciences, 10(22), 8298.

24. Guo, Z., Xie, J., Wan, Y., Zhang, M., Qiao, L., Yu, J., ... & Yao, Y. (2022). A review of the

current state of the computer-aided diagnosis (CAD) systems for breast cancer diagnosis. Open Life Sciences, 17(1), 1600-1611.

25. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.

26. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.

27. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In Proceedings of the Advances in Neural Information Processing Systems 28, Montreal, QC, Canada, 7–12 December 2015.

28. Shin, H. C., Lu, L., & Summers, R. M. (2017). Natural language processing for large-scale medical image analysis using deep learning. Deep learning for medical image analysis, 405-421.

29. Qi, X., Yi, F., Zhang, L., Chen, Y., Pi, Y., Chen, Y., ... & Yi, Z. (2022). Computer-aided diagnosis of breast cancer in ultrasonography images by deep learning. Neurocomputing, 472, 152-165.]

30. Linna, N., & Kahn Jr, C. E. (2022). Applications of natural language processing in radiology: A systematic review. International Journal of Medical Informatics, 163, 104779.

31. Yin C, Qian B, Wei J, et al. Automatic generation of medical imaging diagnostic report with hierarchical recurrent neural network. In: Wang J., Shim K., Wu X., eds. Proc. IEEE Int. Conf. Data Min. ICDM. Vol 2019-November. Institute of Electrical and Electronics Engineers Inc.; 2019:728-737. doi:10.1109/ICDM.2019.00083.

32. Wang, J., Deng, H., Liu, B., Hu, A., Liang, J., Fan et al. (2020). Systematic evaluation of research progress on natural language processing in medicine over the past 20 years: bibliometric study on PubMed. Journal of medical Internet research, 22(1), e16816.

33. Luo, J. W., & Chong, J. J. (2020). Review of natural language processing in radiology. Neuroimaging Clinics, 30(4), 447-458.

34. Otal, R. G., Guerra, J. L., Calderón, C. P., & Martínez, A. (2013). Application of Artificial Intelligence in Tumors Sizing Classification for Breast Cancer.

35. Buckley, J. M., Coopey, S. B., Sharko, J., Polubriaginof, F., Drohan et al. (2012). The

feasibility of using natural language processing to extract clinical information from breast pathology reports. Journal of pathology informatics, 3(1), 23.

36. Chen, Y., Hao, L., Zou, V. Z., Hollander, Z., Ng, R. T., & Isaac, K. V. (2022). Automated medical chart review for breast cancer outcomes research: a novel natural language processing extraction system. BMC medical research methodology, 22(1), 136.

37. Liu, Y., Liu, Q., Han, C., Zhang, X., & Wang, X. (2019). The implementation of natural language processing to extract index lesions from breast magnetic resonance imaging reports. BMC medical informatics and decision making, 19(1), 1-10.

38. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., ... & Girshick, R. (2023). Segment anything. arXiv preprint arXiv:2304.02643.

39. Keshari, R., Singh, R., & Vatsa, M. (2020). Generalized zero-shot learning via over-complete distribution. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 13300-13308).

40. Wu, N., Phang, J., Park, J., Shen, Y., Huang, Z., Zorin, M., ... & Geras, K. J. (2019). Deep neural networks improve radiologists' performance in breast cancer screening. IEEE transactions on medical imaging, 39(4), 1184-1194.

41. Huang, Y., Han, L., Dou, H., Luo, H., Yuan, Z., Liu, Q., ... & Yin, G. (2019). Two-stage CNNs for computerized BI-RADS categorization in breast ultrasound images. Biomedical engineering online, 18(1), 1-18.

42. Weng, Y., Zhou, T., Li, Y., & Qiu, X. (2019). Nas-unet: Neural architecture search for medical image seg-men-tation. IEEE Access, 7, 44247-44257.

43. Al-Dhabyani, W., Gomaa, M., Khaled, H., & Aly, F. (2019). Deep learning approaches for data augmentation and classification of breast masses using ultrasound images. Int. J. Adv. Comput. Sci. Appl, 10(5), 1-11.

44. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 3rd International Conference on Learning Representations, ICLR 2015–Conference Track Proceedings. 2015;1–14.

45. Rajakumari, R., & Kalaivani, L. (2022). Breast Cancer Detection and Classification Using Deep CNN Techniques. Intelligent Automation & Soft Computing, 32(2).

46. Szegedy C, Vanhoucke V, Iofe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. 2016 IEEE Conference on Computer Vision and Pattern.

47. Abhisheka, B., Biswas, S. K., & Purkayastha, B. (2023). A comprehensive review on breast cancer detection, classification and segmentation using deep learning. Archives of Computational Methods in Engineering, 30(8), 5023-5052.

48. Li, Y., Wu, J., & Wu, Q. (2019). Classification of breast cancer histology images using multi-size and discriminative patches based on deep learning. Ieee Access, 7, 21400-21408.

49. Murtaza, G., Shuib, L., Abdul Wahab, A. W., Mujtaba, G., Mujtaba, G., Nweke, H. F., ... & Azmi, N. A. (2020). Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges. Artificial Intelligence Review, 53, 1655-1720.

50. Muramatsu, C., Nishio, M., Goto, T., Oiwa, M., Morita, T., Yakami, M., ... & Fujita, H. (2020). Improving breast mass classification by shared data with domain transformation using a generative adversarial network. Computers in biology and medicine, 119, 103698

51. Khan, S., Islam, N., Jan, Z., Din, I. U., & Rodrigues, J. J. C. (2019). A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. Pattern Recognition Letters, 125, 1-6.

52. Jiménez-Gaona, Y., Rodríguez-Álvarez, M. J., Diana Carrión-Figueroa, Darwin Castillo-Malla & Lakshminarayanan, V. (2024). Breast Mass Regions Classification from Mammograms using Convolutional Neural Networks and Transfer Learning. Journal of Modern Optics TMOP. https://doi.org/10.1080/09500340.2024.2313724.

53. van Timmeren, J. E., Cester, D., Tanadini-Lang, S., Alkadhi, H., & Baessler, B. (2020). Radiomics in medical imaging "How-to" guide and critical reflection. Insights into Imaging, 11(1), 1-16.

54. Rodrigues, PS (2017), "Breast Ultrasound Image", Mendeley Data, V1, [doi: 10.17632/wmy84gzngw.1]

55. Yap MH, Pons G, Marti J, et al. Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks. IEEE J Biomed Heal Informatics. 2018;22(4):1218-1226. [doi:10.1109/JBHI.2017.2731873]

56. Heath M, Bowyer K, Kopans D, Kegelmeyer P, Moore R, Chang K, et al. Current status of the digital database for screening mammography. En Digital mammography. Dordrecht: Springer; 1998.

57. Suckling J, Parker J, Dance D, Astley S, Hutt I, Boggis C, et al. Mammographic

Image Analysis Society (MIAS) database v1.21. 2015.

58. Li S, Hatanaka Y, Fujita H, Hara T, Endo T. Automated Detection of Mammographic Masses in MIAS Database. Medical imaging technology. 1999;17:427–8.

59. Moreira IC, Amaral I, Domingues I, Cardoso A, Cardoso MJ, Cardoso JS. INbreast. Acad Radiol. 2012;19(2):236-248. [doi:10.1016/j.acra.2011.09.014]

60. Pham, H. H., Nguyen Trung, H., & Nguyen, H. Q. (2022). VinDr-Mammo: A large-scale benchmark dataset for computer-aided detection and diagnosis in full-field digital mammography (version 1.0.0). PhysioNet. https://doi.org/10.13026/br2v-7517.

61. Ibrokhimov, B., & Kang, J. Y. (2022). Two-stage deep learning method for breast cancer detection using high-resolution mammogram images. Applied Sciences, 12(9), 4616.

62. Shin, H. C., Lu, L., & Summers, R. M. (2017). Natural language processing for large-scale medical image analysis using deep learning. Deep learning for medical image analysis, 405-421.

63. Shao, S., Wang, P., & Yan, R. (2019). Generative adversarial networks for data augmentation in machine fault diagnosis. Computers in Industry, 106, 85-93.

64. Woldesellasse, H., & Tesfamariam, S. (2023). Data augmentation using conditional generative adversarial network (cGAN): Application for prediction of corrosion pit depth and testing using neural network. Journal of Pipeline Science and Engineering, 3(1), 100091.

65. Parmar, G., Zhang, R., & Zhu, J.-Y. (2022). On Aliased Resizing and Surprising Subtleties in GAN Evaluation. 11400–11410. https://doi.org/10.1109/cvpr52688.2022.01112

66. Gao, M., Song, P., Wang, F., Liu, J., Mandelis, A., & Qi, D. (2021). A novel deep convolutional neural network based on ResNet-18 and transfer learning for detection of wood knot defects. Journal of Sensors, 2021, 1-16.

67. Aljuaid, H., Alturki, N., Alsubaie, N., Cavallaro, L., & Liotta, A. (2022). Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning. Computer Methods and Programs in Biomedicine, 223, 106951.

68. Swarnambiga Ayyachamy, Varghese Alex, Mahendra Khened, and Ganapathy Krishnamurthi "Medical image retrieval using Resnet-18", Proc. SPIE 10954, Medical Imaging 2019: Imaging Informatics for Healthcare, Research, and Applications, 1095410

(15 March 2019); https://doi.org/10.1117/12.2515588

69. M. Guo and Y. Du, "Classification of Thyroid Ultrasound Standard Plane Images using ResNet-18 Networks," 2019 IEEE 13th International Conference on Anti-counterfeiting, Security, and Identification (ASID), Xiamen, China, 2019, pp. 324-328, doi: 10.1109/ICASID.2019.8925267.

70. Borji A. Pros and cons of GAN evaluation measures. Comput Vis Image Underst. 2019;179:41-65. [doi:10.1016/j.cviu.2018.10.009]

71. Zama, S., Fujioka, T., Yamaga, E., Kubota, K., Mori, M., Katsuta, L., ... & Tateishi, U. (2023). Clinical Utility of Breast Ultrasound Images Synthesized by a Generative Adversarial Network. Medicina, 60(1), 14.

72. Pang, T., Wong, J. H. D., Ng, W. L., & Chan, C. S. (2021). Semi-supervised GAN-based radiomics model for data augmentation in breast ultrasound mass classification. Computer Methods and Programs in Biomedicine, 203, 106018.

73. Jiménez Gaona, Y.; Castillo Malla, D.; Vega Crespo, B.; Vicuña, M.J.; Neira, V.A.; Dávila, S.; Verhoeven, V. Radiomics Diagnostic Tool Based on Deep Learning for Colposcopy Image Classification. Diagnostics 2022, 12, 1694. https://doi.org/10.3390/diagnostics12071694

74. Dihge, L., Bendahl, P. O., Skarping, I., Hjärtström, M., Ohlsson, M., & Rydén, L. (2023). The implementation of NILS: A web-based artificial neural network decision support tool for noninvasive lymph node staging in breast cancer. Frontiers in Oncology, 13, 1102254.

75. To, T., Lu, T., Jorns, J. M., Patton, M., Schmidt, T. G., Yen, T., ... & Ye, D. H. (2023). Deep learning classification of deep ultraviolet fluorescence images toward intra-operative margin assessment in breast cancer. Frontiers in Oncology, 13, 1179025.

76. Taylor, C. R., Monga, N., Johnson, C., Hawley, J. R., & Patel, M. (2023). Artificial Intelligence Applications in Breast Imaging: Current Status and Future Directions. Diagnostics, 13(12), 2041.

77. LE, E. P. V., WANG, Y., HUANG, Y., HICKMAN, S., & GILBERT, F. J. (2019). ARTIFICIAL INTELLIGENCE IN BREAST IMAGING. CLINICAL RADIOLOGY, 74(5), 357-366

78.Huynh, H. N., Tran, A. T., & Tran, T. N. (2023). Region-of-Interest Optimization for Deep-Learning-Based Breast Cancer Detection in Mammograms. Applied Sciences, 13(12), 6894.

79.Afrin, H., Larson, N. B., Fatemi, M., & Alizad, A. (2023). Deep Learning in Different Ultrasound Methods for Breast Cancer, from Diagnosis to Prognosis: Current Trends, Challenges, and an Analysis. Cancers, 15(12), 3139.

80.Prodan, M., Paraschiv, E., & Stanciu, A. (2023). Applying Deep Learning Methods for Mammography Analysis and Breast Cancer Detection. Applied Sciences, 13(7), 4272.

81. Oyelade, O. N., Ezugwu, A. E., Almutairi, M. S., Saha, A. K., Abualigah, L., & Chiroma, H. (2022). A generative adversarial network for synthetization of regions of interest based on digital mammograms. Scientific Reports, 12(1), 6166.

82.Herington, J., McCradden, M. D., Creel, K., Boellaard, R., Jones, E. C., Jha, A. K., ... & Saboury, B. (2023). Ethical considerations for artificial intelligence in medical imaging: data collection, development, and evaluation. Journal of Nuclear Medicine, 64(12), 1848-1854. , Herington, J., McCradden, M. D., Creel, K.

83.Boellaard, R., Jones, E. C., Jha, A. K., ... & Saboury, B. (2023). Ethical considerations for artificial intelligence in medical imaging: deployment and governance. Journal of Nuclear Medicine, 64(10), 1509-1515

84.Ueda, D., Kakinuma, T., Fujita, S., Kamagata, K., Fushimi, Y., Ito, R., ... & Naganawa, S. (2024). Fairness of artificial intelligence in healthcare: review and recommendations. Japanese Journal of Radiology, 42(1), 3-15.

85.Drabiak, K., Kyzer, S., Nemov, V., & El Naqa, I. (2023). AI and machine learning ethics, law, diversity, and global impact. The British Journal of Radiology, 96, 20220934.

86.Jimenez, Yuliana; Rodriguez-Alvarez, Maria Jose; Castillo-Malla, Darwin; Garcia, Santiago; Carrión-Figueroa, Diana; Lakshminarayanan, Vasudevan (2024), "BraNet: A mobil Application for Breast image classification based on Deep Learning algorithms.", Mendeley Data, V1, doi: 10.17632/jh9trvbjbv

Chapter 7
General discussion

1.Introduction

The discussion is divided into five points, each corresponding to each published article.

- The first point, corresponding to a critical review discussion about Deep Learning-Based Computer-Aided Systems for Breast Cancer Imaging (Chapter 2), here several aspects related to the state of the art are addressed: (i) Systematic compilation of literature related to breast cancer definitions and available mammography and ultrasound public databases, (ii) Implementation of different network architectures deep learning-based for breast images pre-processing and post-processing, and (ii) Evaluation of the accuracy and performance of these models by means of the most common statistical methods.

- The second point, devoted to Breast Mass Regions Classification from Mammograms using Convolutional Neural Networks is presented; its research focuses on improving the super-resolution of mammography images, specifically from public databases such as mini-MIAS, Inbreast and CBIS-DDSM. It is structured into three sub-points (i) delineate the regions of interest (RoIs) to improve image quality using EDSR and SR-RDN super-resolution algorithms. Then (ii) RoIs segmentation and classification using convolutional neural network-based algorithms such as Unet, SegNet and Resnet. (iii) Finally, to evaluate the image quality, accuracy, precision, and performance of these models using metrics such as PSNR, SSIM, IoU, Dice, accuracy, specificity, and sensitivity.

- The third point concerns implementing GAN models to reduce speckle noise in breast ultrasound images while preserving features and details. Two GANs models (Conditional GAN and Wasserstein GAN) were tested as speckle-denoising reduction using several breast public ultrasound databases: BUSI, DATASET A, UDIAT (DATASET B). The image quality results in both algorithms were measured by Peak Signal Noise Ratio (PSNR) and Structural Similarity Index (SSIM) values.

- In a fourth point, the problem of data augmentation to improve breast cancer classification and validation processes using real and synthetic data is presented. The focus is on the implementation of algorithms based on Generative Adversarial Networks (GANs) to perform data augmentation on public mammography and ultrasound images (CBIS-DDSM, Mini-MIAS, Inbreast, BUSI, UDIAT). This research presents three different approaches: (i) A database collection, (ii) GANs models implementation (CGAN, WGAN, SNGAN, Cycle GAN) for data augmentation, (iii) Data classification using ResNet model and (iii) Models performance evaluation using feature-based, nonreference-based and reference-based metrics and several statistical metrics as classification evaluation performance.

- Finally, using deep learning algorithms, a five-point, mobile graphical interface, “BraNet,” was developed as US and mammography breast imaging segmentation and classification. Two phases were implemented during APP development: (i) Phase off-line and (ii) Phase on-line. During the off-line section, SAM NLP-based, GAN, and CNN models were previously trained for synthetic image generation, segmentation, and classification. Then, during the On-line section, the BraNet app was developed using the React Native framework, offering a modular deep-learning pipeline on a client-server architecture implemented in Python for iOS and Android devices. The information was contrasted with two radiological experts using the kappa value as an agreement between them.

2. Deep Learning-Based Computer-Aided Systems for Breast Cancer Imaging (Chapter 2).

In the first article (chapter 2), we consider two sets of breast tumors screening images: 1.Digital mammography [66] and 2. US [67]. US is used as a complementary imaging modality to mammography in the assessment of breast health. For example, US imaging is one of the most effective tools in breast cancer detection, as it has been shown to achieve high accuracy in mass classification [68] and in distinguishing abnormalities between fluid-filled cysts and solid masses in dense breasts.

The effectiveness of breast cancer detection ultimately depends on a combination of imaging modalities and clinical assessment, and decisions about which imaging tests to use are often made on a case-by-case basis. Medical professionals consider a patient's individual risk factors, breast density, and specific clinical findings when determining the most appropriate approach to breast cancer detection and diagnosis.

For the abovementioned reasons, we have addressed the first article review using both kinds of breast images, focusing on different Machine Learning and Deep Learning architectures applied in breast tumor processing, and offering a general overview of the most commonly public breast databases used for CNNs training, including their relation and efficacy in performing segmentation, feature extraction, selection, and classification tasks [69].

Thus, according to the research the most utilized public databases for mammography images are mini-MIAS and CBIS-DDSM, and for US image classification are BUSI, DDBUI, UDIAT and OASBUD. The mammography images contributed to 110 and 168 published conference papers for the DDSM and MIAS databases, respectively [70]. However, the databases report some limitations and advantages; for example, the mini-MIAS database contains a limited number of images, strong noise, and low-resolution images. In contrast, the CBIS-DDSM contains a big dataset. Likewise, Inbreast contains high-resolution images but has a small data size. BCDR, in comparison with DDSM, has been used in a few studies. Some details about the other strengths and limitations of these databases are discussed in Abdelhafiz et al. [71].

Based on the most popular datasets, CNN seems to perform rather well [72], as demonstrated by Chiao et al., Samala et al. and Yap et al. [73-75]. Furthermore, [76-77] used several preprocessing and postprocessing techniques for high-resolution [62-63] data augmentation, segmentation, and classification. The most commonly CNNs used are AlexNet, VGG, ResNet, DenseNet, Inception (GoogleNet), LeNet, and UNet, which employ recent Python libraries for implementing CNNs, such as Tensorflow, Caffe, and Keras, with different hyper-parameters to training the network [9].

Most of these deep learning architectures use a large data set; thus, it is required to apply an augmentation technique to avoid overfitting and to have better performance during classification. Furthermore, various studies [78-80] prove that those CNN methods that compare images from CC and MLO views improve the accuracy of detection and reduce the false positive rate.

Indeed, different evaluation metrics are used to measure the performance and accuracy of Convolutional Neural Networks (CNNs) in medical imaging, including techniques like mammography and US imaging. These metrics help assess the quality and effectiveness of the CNN models in various aspects. Some common evaluation metrics used in medical image analysis are Precision, Accuracy, Sensitivity, Specificity, Dice Coefficient (F1-Score), Intersection over Union (IoU), Frechet Inception Distance (FID) and Kernel Inception Distance (KID), Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), Receiver Operating Characteristic (ROC) Curve, Area Under the ROC Curve (AUC-ROC).

The choice of evaluation metrics depends on the specific task and goals of the CNN model in medical imaging. For instance, in disease detection, sensitivity and specificity may be more critical, while segmentation tasks may prioritize Dice coefficient or IoU. In image quality PSNR and SSIM, in synthetic data evaluation KID and FID are often used to provide a comprehensive assessment of model performance [81-82]. These metrics demonstrate that in most cases, the deep learning architectures outperformed traditional methodologies.

To continue this work, we have chosen all the public ultrasound and mammography databases suggested from this literature review results. Also, the most significant convolutional neural networks in performance and accuracy were selected for segmentation and classification training, using relevant statistical metrics to evaluated the performance of the algorithms taken from this review.

3. Breast Mass Regions Classification from Mammograms using Convolutional Neural Networks and transfer learning (Chapter 3).

The second article (chapter 3) provides a comprehensive study on the utilization of convolutional neural networks (CNNs) and transfer learning techniques to enhance the accuracy of breast cancer detection through mammography (DM). DM images are often acquired at lower resolutions to minimize radiation exposure while maintaining adequate diagnostic quality. However, low-resolution images can compromise the ability to detect subtle features or abnormalities, such as microcalcifications in breast cancer lesions. That CNN training for high-resolution image enhancement and segmentation can indeed be valuable and significantly improve the visibility of fine details, making it easier for radiologists to identify and classify breast cancer lesions, especially in low-resource settings where DM images may have poor resolution.

The study leverages three well-known public mammography databases: CBIS-DDSM, Mini-MIAS, and Inbreast, highlighting the significance of data quality and the impact of pre-processing in clinical decision-making. The authors introduce novel CNN-based algorithms for super-resolution, notably the Enhanced Deep Residual Network (EDSR) and the Residual Dense Network (RDN), which demonstrated superior performance in improving image quality measured by metrics such as PSNR and SSIM.

ROI segmentation was then required using pre-trained segmentation models (Unet and SegNet). Finally, ROI image classification process was required with ResNet50 model as clinical decision aimed at improving breast diagnosis. Where, the EDSR successfully reconstructed the detailed textures and edges in the RoIs and exhibited better quality output where PSNR and SSIM index (39.05 dB and 0.90) over (32.68 dB and 0.82) to RDN model. UNet demonstrated superior performance over SegNet, boasting an average Intersection over Union (IoU) of 0.862, an average Dice coefficient of 0.991, and an accuracy rate of 0.947 in RoI segmentation results. By analyzing the segmented regions, the ResNet50 provides good performance (Accuracy of 75%) in breast feature extraction and classification using data augmentation by affine transformation compared to a training model without data augmentation (Accuracy of 68%).

This accuracy could be because generated images by traditional augmentation methods, which share a similar distribution with the original ones, are not suitable for medical images processing (Lan et al. [83]). Guan et al. [84] demonstrated that RoIs generated by GAN models are more like real RoIs than affine transformed RoIs in terms of mean, standard deviation, skewness, and entropy.

Finally, we have concluded that:

(i) Breast tumor classification using mammography images has some consequences and limitations with respect to using traditional data augmentation because the model did not improve the accuracy in comparison with other state-of-the-art deep CNN-based classifiers [85-88].

(ii) Breast tumor classification can be improved using GAN models as generation of synthetic data.

Additionally, the research delves into the challenges associated with deep learning models, particularly the need for extensive datasets for training, which is often limited due to privacy and data protection concerns.

This study is significant for its comprehensive approach to addressing the challenges in mammography image analysis and its contribution to developing more accurate and reliable computer-aided diagnostic systems for breast cancer.

In summary, employing CNNs for super-resolution image enhancement, segmentation, and breast cancer lesion classification can help improve diagnostic capabilities in breast lesion detection, especially in settings where resource constraints might limit traditional diagnostic approaches.

4. Ultrasound Breast images denoising using Generative Adversarial Networks (GANs) (Chapter 4).

Chapter 4 presents a solution to the speckle noise problem in breast US images, caused by interference patterns in the reflected US waves. Understanding and mitigating this noise is essential for improving the accuracy and reliability of breast imaging for diagnostic and therapeutic purposes. US imaging utilizes high-frequency sound waves to create images of internal body structures. Speckle noise arises from the interference patterns produced by the reflected ultrasound waves. These interference patterns occur due to the constructive and destructive interference of the waves as they bounce off different tissue structures within the breast, and speckle noise appears as a grainy or granular pattern superimposed on the US images.

It can obscure fine details, distorts edges, and reduces the contrast between structures within the breast. This noise presents challenges for radiologists and clinicians who interpret breast US images. It can make it difficult to distinguish between normal and abnormal tissue, identify subtle features such as microcalcifications or small masses, and accurately assess the extent of lesions, affecting the accuracy of breast cancer detection, characterization, and treatment planning. It may lead to false-positive or false-negative findings, potentially impacting patient care and outcomes.

Various image processing techniques can reduce speckle noise in breast ultrasound images, including filtering algorithms, wavelet transforms, and machine learning approaches such as deep learning and GANs. These techniques aim to preserve crucial diagnostic information while suppressing noise artifacts. The research utilizes two public breast ultrasound databases, BUSI DATASET A and UDIAT (DATASET B), to train the GAN models. Here, two GANs (Conditional GAN and Wasserstein GAN) were proposed for image noise reduction in breast US images. The GANs are trained using RoIs of US breast images. The CGAN model uses the Unet architecture, while the WGAN model employs the Resnet architecture. The quality of the denoised images is measured using standard values of Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM).

The generator learns to generate images that resemble the real US images, while the discriminator learns to differentiate between real and generated images. The objective of the GAN is to minimize the difference between the generated and real images. In the case of speckle noise reduction, the generator learns to remove speckle noise while preserving important features of the breast tissue. Once trained, the generator can be used to denoise new US images of the breast in real-time or in batch processing. The higher robustness demonstrated by CGAN is attributed to the generator using U-Net encoder-decoder architecture with BCE loss function to remove the speckle noise better than the Resnet architecture used in WGAN.

The proposed CGAN technique is beneficial for small data sets with low variance. These networks are widely used for image generation or data augmentation, but their application in US image denoising still needs to be investigated. Overall, using GANs for speckle noise reduction in US breast images can improve the quality of the images, aiding in more accurate diagnosis and treatment planning in medical applications.

In summary, the article underscores the importance of image denoising in medical imaging, representing a significant advancement in radiomics, offering a promising approach to enhancing the quality of ultrasound images and potentially improving the accuracy of breast cancer detection.

5. GAN-based data augmentation to improve breast Ultrasound and Mammography Mass Classification (Chapter 5).

Chapter 5 addresses the challenge of imbalanced data in medical datasets by using various GAN models to augment data in breast Ultrasound and Mammography Regions of Interest (ROIs) Classification. The paper highlights the challenge of limited and imbalanced datasets applying deep learning algorithms for breast cancer diagnosis. GANs were used to generate synthetic medical images to augment the available datasets, thus enhancing the performance of these algorithms. The research evaluated four GAN models WGAN-GP, Cycle GAN, Conditional GAN, and Spectral Normalization GAN (SNGAN) for their effectiveness in augmenting breast imaging data. Here, the quality and diversity of the synthetic data were assessed using various metrics such as FID, KID, SSIM, MS-SSIM, BRISQUE, NIQE, and PIQE.

The study found that SNGAN was most effective for mammography data augmentation, while CGAN was best suited for ultrasound data. Both models produced high-quality synthetic images, improving classification performance when training a Resnet network. The study also emphasized the importance of preprocessing and characterizing ROIs by abnormality type to generate diverse and compelling synthetic data. The paper concludes that GAN-based data augmentation holds significant potential for improving the accuracy of breast cancer diagnosis by generating diverse synthetic data. It suggests that further research should explore the integration of other breast imaging modalities and investigate additional normalization and regularization methods to enhance GAN training stability and performance.

One of the paper's strengths is its detailed analysis and comparison of different GAN models, providing valuable insights into their relative effectiveness for specific types of data (mammography vs. ultrasound). It can guide future research and application of GANs in medical imaging. However, the paper also highlights the challenges associated with training GANs, such as convergence issues and the risk of mode collapse. These challenges underline the need for ongoing research to refine GAN models and training techniques, ensuring the generation of high-quality, diverse synthetic images that can truly enhance the training of deep learning algorithms for medical diagnosis. Overall, the paper significantly contributes to the medical imaging field and artificial intelligence, offering a promising avenue for addressing the challenge of data limitations in developing AI-driven diagnostic tools.

6. BraNet: A mobil Application for Breast image classification based on Deep Learning algorithms (Chapter 6).

In the last article (chapter 6), a mobile application for Breast Images named "BraNet" is presented as an innovative open-source mobile application aimed at enhancing breast cancer detection through advanced image classification techniques. The application uses deep learning algorithms, specifically SNGAN for synthetic image generation, Segment Anything, and ResNet18 for segmentation and classification, respectively, to aid in accurately and efficiently analyzing 2D breast imaging, encompassing mammography and US images.

The study highlights the potential of mobile health apps in providing valuable second opinions to radiologists, thereby reducing false diagnoses and advancing breast cancer detection methodologies. The application, developed using the React Native framework, features a client-server architecture and is compatible with iOS and Android platforms. The research emphasizes the importance of the quantity and diversity of data in training deep learning algorithms, particularly underlining the challenges posed by various abnormalities in mammography images.

The results demonstrate the application's high accuracy in classifying benign and malignant lesions in US images, outperforming traditional mammography classification by radiological experts. It indicates a significant advancement in leveraging deep learning for medical image analysis, potentially enhancing diagnostic precision and efficiency.

The research also discusses the ethical considerations and challenges associated with deploying artificial intelligence (AI) in medical diagnosis, emphasizing the need for transparency, data privacy, and the equitable use of AI technologies in healthcare. Future directions include improving the application by addressing limitations, such as classifying different abnormalities and enhancing image quality through advanced pre-processing techniques.

In conclusion, "BraNet" represents a significant stride towards integrating deep learning in clinical practices for breast cancer detection. It offers a preclinical user-friendly tool that enhances diagnostic accuracy and supports radiologists' decision-making processes.

Chapter 8
Conclusions

This thesis provides five main conclusions according to the experimental studies (Chapters 2 to 6), and are summarized here:

~~1.~~ Deep learning algorithms could be a promising new technique for obtaining key features for automatic breast tumor classification, even in dense breasts. DL provides a mechanism to automatically extract features through a self-learning network, thereby increasing classification accuracy.

2. The second experimental study proposed a deep learning-based CAD system framework for breast mammography data augmentation, super-resolution, segmentation, and classification using the transfer learning concept. Enhanced Super-Resolution (EDSR) demonstrated superior performance in image quality enhancement for digital mammography images over Super-Resolution Residual Dense Network (SR-RDN), as evidenced by higher Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Metric (SSIM) values. This suggests EDSR's better capability in reconstructing detailed textures and edges in Regions of Interest (RoIs), contributing to more accurate breast lesion segmentation and classification. UNet outperformed SegNet in RoIs segmentation, achieving higher Intersection over Union (IoU), Dice similarity coefficient, and accuracy metrics. This indicates that UNet is more effective in delineating potential breast cancer lesions within mammograms, which is crucial for subsequent classification tasks. The ResNet-50 model improved classification accuracy to 75% when using enhanced RoI data augmentation through affine transformation, surpassing traditional machine learning algorithms. However, it did not achieve the same accuracy as other state-of-the-art deep Convolutional Neural Network (CNN)-based classifiers, which could be due to the limitations of traditional data augmentation techniques in accurately simulating the real distribution of medical images in comparison with the generative adversarial models.

3. In this study, we have explored the application of advanced image processing techniques, with a particular focus on the cutting-edge computer vision algorithms like Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs). These methodologies have

significantly advanced the field of breast imaging by offering superior solutions for image segmentation, super-resolution, enhancement, and denoising. This progress is critical in elevating the precision of breast anomalies detection, enabling more accurate tumor characterization, and facilitating more effective treatment planning strategies. Specifically, the application of GANs has shown remarkable potential in mitigating speckle noise prevalent in ultrasound breast images. This noise reduction capability not only enhances the overall image quality but also plays a pivotal role in preserving vital diagnostic information. The resultant clarity and detail in the imagery substantially contribute to more accurate diagnoses and informed treatment planning in medical practice. Looking forward, the integration of these sophisticated image processing techniques opens new horizons for research and development in medical imaging.

4. Generative Adversarial Network (GAN) models with different normalization techniques to generate synthetic images have been implemented: Among these models, SNGAN was found to be effective for mammography data augmentation, while CGAN performed well for ultrasound US data augmentation. Additionally, the Cycle GAN model was highlighted for its success in generating high-quality and diverse synthetic images for both datasets, showing higher visual similarity between real and generated images even in a reduced training environment. However, WGAN-GP, despite successfully addressing the mode collapse issue, resulted in visually unacceptable outcomes and lower stability in discriminator and generator training for both datasets. The selection of the most suitable GAN model for data augmentation is contingent upon the specific problem, dataset characteristics, evaluation quality metrics, and hyperparameter choices to enhance GAN training stability.

5. The experimental studies showed that deep learning techniques can be successfully applied to breast image classification and assessment of malignant and benign breast lesions. The main contribution was the ability to augment, segment, and classify regions of interest using generative adversarial networks and convolutional neural networks. This finding opens a line of research aiming to accurately classify breast imaging using other types of images such as thermography, MRI, and PET and develop a CAD radiomic tool that will have potential benefits if successfully validated and applied in clinical routine.

6. Thus, BraNet's API facilitates the construction of breast image classification workflows, encompassing data input/output, ROI mask extraction, segmentation, and evaluation metrics. BraNet is currently in a preclinical testing phase, indicating that more work is required. The API will continue

receiving regular updates and extensions, emphasizing the need for rigorous analysis to ensure its accuracy and reliability.

7. These results could serve as a starting point for the development of other breast screening applications, including the thermal patch as a breast scanner linking with the BraNet API mobile, that would have a significant medical impact because it is non-invasive screening and widely accessible (even in regions with limited access to advanced medical facilities). It also has the potential to bring significant advances in breast health monitoring and early detection of breast cancer.

Limitations and Future Work

The main limitation of this work is that some databases are not available in the open literature due to patient privacy and proprietary intellectual issues. Here we envisage using synthetic breast data or data augmentation for training deep learning techniques based on GANs (Generative Adversarial Network) models as input data to train alternative breast mass classifiers based on convolutional networks (Densenet, Nasnet, VGGnet), aiming at improving the accuracy of breast lesion classification and reducing the overfitting percentage.

We also found that super-resolution studies based on CNN models using 2D breast images, such as mammography and US, are limited because most of the literature uses urban and natural images. Here, we developed a deep CNN approach for mammography super-resolution, segmentation, and classification of RoIs, resulting in good indices and quality scores.

There is still a need for better architectures, more extensive datasets that overcome class imbalance problems, and better optimization hypermeters and methods, to further enhance the performance of DL in breast tumor classification. A significant limitation identified in the study is the unavailability of comparison of our results (same datasets) with other algorithms and results in the open literature due to proprietary intellectual property issues.

We also develop a preclinical mobile application to implement different modules with the deep learning algorithms previously trained during the on-line phase for the early mammography and US breast lesions detection. In future research, we pretend extending these algorithms to other types of breast imaging, such as thermography and PET.

Finally, we will develop a thermal breast patch (BraPatch) to link it with the BraNet application as complementary preclinical tool for early cancer detection and prevention, which involves a complex and multidisciplinary approach, including medical expertise, smart biotextiles, bioengineering (sensors), and computer vision based on deep learning models.

Bibliography

1. Alejandro Di Sibio, Graciela Abriata, David Forman, Mónica S. Sierra, Female breast cancer in Central and South America, Cancer Epidemiology,Volume 44, Supplement 1, 2016,Pages S110-S120,ISSN 1877-7821.
2. Lim CN, Suliong C, Rao CV, Aung T, Sieman J, Tin W, et al. Recent Advances in Breast Cancer Diagnosis Entering an Era of Precision Medicine. 2019.
3. DeSantis CE, Bray F, Ferlay J, Lortet-Tieulent J, Anderson BO, Jemal A. International Variation in Female Breast Cancer Incidence and Mortality Rates. Cancer Epidemiol Biomarkers Prev. 2015; 24(10): 1495-506. http://www.ncbi.nlm.nih.gov/pubmed/26359465
4. Ginsburg O, Yip CH, Brooks A, Cabanes A, Caleffi M, Dunstan Yataco JA, et al. Breast cancer early detection: A phased approach to implementation. Cancer. 2020; 126 Suppl 10: 2379-93. http://www.ncbi.nlm.nih.gov/pubmed/32348566
5. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. CA: a cancer journal for clinicians. 2017; 67(1):7–30.\
6. Zaidi, Z., & Dib, H. A. (2019). The worldwide female breast cancer incidence and survival, 2018.
7. Justo, N., Wilking, N., Jönsson, B., Luciani, S., & Cazap, E. (2013). A review of breast cancer care and outcomes in Latin America. The oncologist, 18(3), 248–256. https://doi.org/10.1634/theoncologist.2012-0373
8. Ragab DA, Sharkas M, Marshall S, Ren J. 2019. Breast cancer detection using deep convolutional neural networks and support vector machines. PeerJ 7:e6201 http://doi.org/10.7717/peerj.6201
9. Jiménez-Gaona, Yuliana, María J. Rodríguez-Álvarez, and Vasudevan Lakshminarayanan. 2020. "Deep-Learning-Based Computer-Aided Systems for Breast Cancer Imaging: A Critical Review" Applied Sciences 10, no. 22: 8298. https://doi.org/10.3390/app10228298
10. J. E. Skandalakis, Embryology and Anatomy of the Breast, Springer BerlinHeidelberg, Berlin,Heidelberg,Germany, 2009.
11. Abdullah-Al Nahid and Yinan Kong, "Involvement of Machine Learning for Breast Cancer Image Classification: A Survey," Computational and Mathematical Methods in Medicine, vol. 2017, Article ID 3781951, 29 pages, 2017. https://doi.org/10.1155/2017/3781951.

12. Łuczyńska, E., Anioł, J., Dyczek, S., Mituś, J., & Stelmach, A. (2006). Ductal carcinoma in situ (DCIS) and invasive ductal carcinoma (IDC) in patients with breast lesion marked by the localized needle. Polish Journal of Radiology, 71(4).

13. Vuong, D., Simpson, P. T., Green, B., Cummings, M. C., & Lakhani, S. R. (2014). Molecular classification of breast cancer. Virchows Archiv, 465, 1-14.

14. Lim CN, Suliong C, Rao CV, Aung T, Sieman J, Tin W, et al. Recent Advances in Breast Cancer Diagnosis Entering an Era of Precision Medicine. 2019.

15. Jalalian, A.; Mashohor, S.B.; Mahmud, H.R.; Saripan, M.I.B.; Ramli, A.R.B.; Karasfi, B. Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: A review. Clin. Imaging 2013, 37, 420–426.

16. Flanagan, F. L., Dehdashti, F., & Siegel, B. A. (1998). PET in breast cancer. In Seminars in nuclear medicine , 28(4), 290-302.

17. Ha, R., Mutasa, S., Karcich, J., Gupta, N., et al. (2019). Predicting Breast Cancer Molecular Subtype with MRI Dataset Utilizing Convolutional Neural Network Algorithm. Journal of digital imaging, 32(2), 276–282

18. Michell, M. J., & Batohi, B. (2018). Role of tomosynthesis in breast imaging going forward. Clinical Radiology, 73(4), 358–371. https://doi.org/10.1016/j.crad.2018.01.00

19. Singh, D., & Singh, A. K. (2020). Role of image thermography in early breast cancer detection-Past, present and future. Computer methods and programs in biomedicine, 183, 105074.

20. Araújo, T., Aresta, G., Castro, E., et al. (2017). Classification of breast cancer histology images using Convolutional Neural Networks. PLoS One, 12(6), e0177544. https://doi.org/10.1371/journal.pone.0177544

21. Das, K., Conjeti, S., Roy, A. G., Chatterjee, J., & Sheet, D. (2018). Multiple instance learning of deep convolutional neural networks for breast histopathology whole slide classification. In 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018) (pp. 578-581). IEEE. https://doi.org/10.1109/ISBI.2018.8363642

22. Iranmakani, S., Mortezazadeh, T., Sajadian, F. et al. A review of various modalities in breast imaging: technical aspects and clinical outcomes. Egypt J Radiol Nucl Med 51, 57 (2020). https://doi.org/10.1186/s43055-020-00175-5

23. Ekici, S., & Jawzal, H. (2020). Breast cancer diagnosis using thermography and convolutional neural networks. Medical hypotheses, 137, 109542.

24. Al Husaini, M. A. S., Habaebi, M. H., Hameed, S. A., Islam, M. R., & Gunawan, T. S. (2020). A systematic review of breast cancer detection using thermography and neural networks. IEEE Access, 8, 208922-208937.

25. Timmers, J. M. H., Van Doorne-Nagtegaal, H. J., Zonderland, H. M., Van Tinteren, H., Visser, O., Verbeek, A. L. M., Den Heeten, G. J., & Broeders, M. J. M. (2012). The Breast Imaging Reporting and Data System (BI-RADS) in the Dutch breast cancer screening programme: its role as an assessment and stratification tool. European Radiology, 22(8), 1717–1723. https://doi.org/10.1007/S00330-012-2409-2

26. Han Z, Wei B, Zheng Y, Yin Y, Li K, Li S. Breast cancer multi-classification from histopathological images with structured deep learning model. Sci Rep. 2017;7(1):4172.

27. Zhu, C., Song, F., Wang, Y., Dong, H., Guo, Y., & Liu, J. (2019). Breast cancer histopathology image classification through assembling multiple compact CNNs. BMC medical informatics and decision making, 19(1), 198. https://doi.org/10.1186/s12911-019-0913-x

28. Shengfeng Liu, Yi Wang, Xin Yang, Baiying Lei, Li Liu, Shawn Xiang Li, Dong Ni, Tianfu Wang (2019). Deep Learning in Medical Ultrasound Analysis: A Review, Engineering, Volume 5, Issue 2, 2019, Pages 261-275, ISSN 2095-8099, https://doi.org/10.1016/j.eng.2018.11.020.

29. Zhang, Q.; Xiao, Y.; Dai, W.; Suo, J.; Wang, C.; Shi, J.; Zheng, H. Deep learning based classification of breast tumors with shear-wave elastography. Ultrasonics 2016, 72, 150–157.

30. Cheng, J.-Z.; Ni, D.; Chou, Y.-H.; Qin, J.; Tiu, C.-M.; Chang, Y.-C.; Huang, C.-S.; Shen, D.; Chen, C.-M.Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans. Sci. Rep. 2016, 6, 24454.

31. Shin, S.Y.; Lee, S.; Yun, I.D.; Kim, S.M.; Lee, K.M. Joint Weakly and Semi-Supervised Deep Learning for Localization and Classification of Masses in Breast Ultrasound Images. IEEE Trans. Med. Imaging 2018, 38,762–774.

32. Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2018). Generative adversarial networks: An overview. IEEE Signal Processing Magazine, 35(1), 53–65. https://doi.org/10.1109/MSP.2017.2765202

33. Muramatsu, C., Nishio, M., Goto, T., Oiwa, M., Morita, T., Yakami, M., Kubo, T., Togashi, K., & Fujita, H. (2020). Improving breast mass classification by shared data with domain transformation using a generative adversarial network. Computers in Biology and Medicine, 119, 103698. https://doi.org/10.1016/j.compbiomed.2020.103698

34. Tschuchnig, M. E., Oostingh, G. J., & Gadermayr, M. (2020). Generative adversarial networks in digital pathology: a survey on trends and future potential. Patterns, 1(6), 100089.

35. Elbattah, M., Loughnane, C., Guérin, J. L., Carette, R., Cilia, F., & Dequen, G. (2021). Variational Autoencoder for Image-Based Augmentation of Eye-Tracking Data. Journal of Imaging, 7(5), 83.

36. Gondara, L. (2016). Medical image denoising using convolutional denoising autoencoders. In 2016 IEEE 16th international conference on data mining workshops (ICDMW) (pp. 241-246). IEEE.

37. Yao, H., Zhang, X., Zhou, X., & Liu, S. (2019). Parallel structure deep neural network using CNN and RNN with an attention mechanism for breast cancer histology image classification. Cancers, 11(12), 1901. https://doi.org/10.3390/cancers11121901

38. Geras, K. J., Wolfson, S., Shen, Y., et al. (2017). High-resolution breast cancer screening with multi-view deep convolutional neural networks. arXiv preprint arXiv:1703.07047.

39. Duggento, A., Aiello, M., Cavaliere, C., et al. (2019). An Ad Hoc Random Initialization Deep Neural Network Architecture for Discriminating Malignant Breast Cancer Lesions in Mammographic Images. Contrast Media & Molecular Imaging, 2019, 5982834. https://doi.org/10.1155/2019/5982834

40. Samala, R. K., Chan, H.-P., Hadjiiski, L., Helvie, M. A., Wei, J., & Cha, K. (2016). Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography. Medical Physics, 43, 6654–6666. https://doi.org/10.1118/1.4967345

41. Dhungel, N., Carneiro, G., & Bradley, A. P. (2015). Automated mass detection in mammograms using cascaded deep learning and random forests. In 2015 international conference on digital image computing: techniques and applications (DICTA) (pp. 1-8). IEEE. doi: 10.1109/DICTA.2015.7371234.

42. Chougrad, H., Zouaki, H., & Alheyane, O. (2018). Deep Convolutional Neural Networks for breast cancer screening. Computer Methods and Programs in Biomedicine, 157, 19–30. https://doi.org/10.1016/j.cmpb.2018.01.011

43. Moon, W. K., Lee, Y.-W., Ke, H.-H., Lee, S. H., Huang, C.-S., & Chang, R.-F. (2020). Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks. Computer Methods and Programs in Biomedicine,190, 105361. https://doi.org/10.1016/j.cmpb.2020.1053

44. Zhang, J., Saha, A., Zhu, Z., & Mazurowski, M. A. (2019). Hierarchical convolutional neural networks for segmentation of breast tumors in MRI with application to radiogenomics. IEEE Transactions on Medical Imaging, 38(2), 435–447. https://doi.org/10.1109/TMI.2018.2865671

45. Byra, M., Jarosik, P., Szubert, A., Galperin, M., Ojeda-Fournier, H., Olson, L., O'Boyle, M., Comstock, C., & Andre, M. (2020). Breast mass segmentation in ultrasound with selective kernel U-Net convolutional neural network. Biomedical Signal Processing and Control, 61, 102027. https://doi.org/10.1016/j.bspc.2020.102027

46. Khoshdel, V., Ashraf, A., & LoVetri, J. (2019). Enhancement of multimodal microwave-ultrasound breast imaging using a deep-learning technique. Sensors, 19(18), 4050. https://doi.org/10.3390/s19184050

47. Dhungel, N., Carneiro, G., & Bradley, A. P. (2015). Automated mass detection in mammograms using cascaded deep learning and random forests. In 2015 international conference on digital image computing: techniques and applications (DICTA) (pp. 1-8). IEEE. doi: 10.1109/DICTA.2015.7371234.

48. Al-Masni, M. A., Al-Antari, M. A., Park, J.-M., et al. (2018). Simultaneous detection and classification of breast masses in digital mammograms via a deep learning YOLO-based CAD system. Computer Methods and Programs in Biomedicine, 157, 85–94. https://doi.org/10.1016/j.cmpb.2018.01.017

49. Solanki, A., Nayyar, A., & Naved, M. (Eds.). (2021). Generative Adversarial Networks for Image-to-Image Translation. Academic Press.

50. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y.

51. Guan, S., & Loew, M. (2019). Breast cancer detection using synthetic mammograms from generative adversarial networks in convolutional neural networks. Journal of Medical Imaging, 6(3), 031411. Generative Adversarial Networks in Digital Pathology: A Survey on Trends and Future Potential Maximilian E.

52. Tschuchnig, M. E., Oostingh, G. J., & Gadermayr, M. (2020). Generative adversarial networks in digital pathology: a survey on trends and future potential. Patterns, 1(6), 100089.

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

54. Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 4401-4410).

55. Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.

56. Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2223-2232).

57. Mikołajczyk, A., & Grochowski, M. (2018). Data augmentation for improving deep learning in image classification problem. In 2018 international interdisciplinary PhD workshop (IIPhDW) (pp. 117-122). IEEE.

58. Wang, Z., Wang, L., Duan, S., & Li, Y. (2020). An Image Denoising Method Based on Deep Residual GAN. In Journal of Physics: Conference Series (Vol. 1550, No. 3, p. 032127). IOP Publishing.

59. Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein generative adversarial networks. In International conference on machine learning (pp. 214-223). PMLR

60. Cai, L., Chen, Y., Cai, N., Cheng, W., & Wang, H. (2020). Utilizing amari-alpha divergence to stabilize the training of generative adversarial networks. Entropy, 22(4), 410.

61. Kang, E., Koo, H. J., Yang, D. H., Seo, J. B., & Ye, J. C. (2019). Cycle-consistent adversarial denoising network for multiphase coronary CT angiography. Medical physics, 46(2), 550-562.

62. Li, P., Li, Z., Pang, X., Wang, H., Lin, W., & Wu, W. (2021). Multi-scale residual denoising GAN model for producing super-resolution CTA images. Journal of Ambient Intelligence and Humanized Computing, 1-10.

63. Yang, Q., Yan, P., Zhang, Y., Yu, H., Shi, Y., Mou, X., ... & Wang, G. (2018). Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. IEEE transactions on medical imaging, 37(6), 1348-1357.

64. Ahmed, A. S., El-Behaidy, W. H., & Youssif, A. A. (2021). Medical image denoising system based on stacked convolutional autoencoder for enhancing 2-dimensional gel electrophoresis noise reduction. Biomedical Signal Processing and Control, 69, 102842.

65. Daoud, M. I., Saleh, A., Hababeh, I., & Alazrai, R. (2019). Content-based image retrieval for breast ultrasound images using convolutional autoencoders: A Feasibility study. In 2019 3rd International Conference on Bio-engineering for Smart Technologies (BioSMART) (pp. 1-4). IEEE

66. Miglioretti, D. L., Lange, J., Van Den Broek, J. J., Lee, C. I., Van Ravesteyn, N. T., Ritley, D., ... & Hubbard, R. A. (2016). Radiation-induced breast cancer incidence and mortality from digital mammography screening: a modeling study. Annals of internal medicine, 164(4), 205-214.

67. Rodríguez-Ruiz, A., Krupinski, E., Mordang, J. J., Schilling, K., Heywang-Köbrunner, S. H., Sechopoulos, I., & Mann, R. M. (2019). Detection of breast cancer with mammography: effect of an artificial intelligence support system. Radiology, 290(2), 305-314.

68. Azzam, H., Kamal, R. M., Hanafy, M. M., Youssef, A., & Hashem, L. M. B. (2020). Comparative study between contrast-enhanced mammography, tomosynthesis, and breast ultrasound as complementary techniques to mammography in dense breast parenchyma. Egyptian Journal of Radiology and Nuclear Medicine, 51(1), 1-9.

69. Harada-Shoji, N., Suzuki, A., Ishida, T., Zheng, Y. F., Narikawa-Shiono, Y., Sato-Tadano, A., ... & Ohuchi, N. (2021). Evaluation of adjunctive ultrasonography for breast cancer detection among women aged 40-49 years with varying breast density undergoing screening mammography: a secondary analysis of a randomized clinical trial. JAMA network open, 4(8), e2121505-e2121505.

70. Zou, L.; Yu, S.; Meng, T.; Zhang, Z.; Liang, X.; Xie, Y. A Technical Review of Convolutional Neural Network-Based Mammographic Breast Cancer Diagnosis. Comput. Math. Methods Med. 2019, 2019, 6509357

71. Abdelhafiz, D., Yang, C., Ammar, R., & Nabavi, S. (2019). Deep convolutional neural networks for mammography: advances, challenges and applications. BMC bioinformatics, 20(11), 281.

72. Jalalian, A., Mashohor, S. B., Mahmud, H. R., et al. (2013). Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review. Clinical imaging, 37(3), 420-426.

73. Chiao, J. Y., Chen, K. Y., Liao, K. Y. K., Hsieh, P. H., Zhang, G., & Huang, T. C. (2019). Detection and classification the breast tumors using mask R-CNN on sonograms. Medicine, 98(19).

74. Samala, R. K., Chan, H.-P., Hadjiiski, L., Helvie, M. A., Wei, J., & Cha, K. (2016). Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography. Medical Physics, 43, 6654–6666. https://doi.org/10.1118/1.4967345

75. Yap, M. H., Pons, G., Marti, J., et al. (2018). Automated Breast Ultrasound Lesions Detection Using Convolutional Neural Networks. IEEE Journal of Biomedical and Health Informatics,22, 1218–1226. https://doi.org/10.1109/JBHI.2017.2731873

76. Salama, W. M., & Aly, M. H. (2021). Deep learning in mammography images segmentation and classification: Automated CNN approach. Alexandria Engineering Journal, 60(5), 4701-4709.

77. Salvi, M., Acharya, U. R., Molinari, F., & Meiburger, K. M. (2021). The impact of pre-and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis. Computers in Biology and Medicine, 128, 104129.

78. Mayo, R. C., Kent, D., Sen, L. C., Kapoor, M., Leung, J. W., & Watanabe, A. T. (2019). Reduction of false-positive markings on mammograms: a retrospective comparison study using an artificial intelligence-based CAD. Journal of digital imaging, 32, 618-624.

79. Yan, Y., Conze, P. H., Lamard, M., Quellec, G., Cochener, B., & Coatrieux, G. (2021). Towards improved breast mass detection using dual-view mammogram matching. Medical image analysis, 71, 102083.

80. Jafari, Z., & Karami, E. (2023). Breast Cancer Detection in Mammography Images: A CNN-Based Approach with Feature Selection.

81. Sheba, K. U., & Raj, S. G. (2016). Objective quality assessment of image enhancement methods in digital mammography–a comparative study. Signal & Image processing: An International Journal, 7(4), 1-13.

82. Gao, X., Lu, W., Tao, D., & Li, X. (2010). Image quality assessment and human visual system. In Visual Communications and Image Processing 2010 (Vol. 7744, p. 77440Z). International Society for Optics and Photonics.

83. Lan, L., You, L., Zhang, Z., Fan, Z., Zhao, W., Zeng, N., ... & Zhou, X. (2020). Generative adversarial networks and its applications in biomedical informatics. Frontiers in Public Health, 8, 164.

84. Guan, S., & Loew, M. (2019). Breast cancer detection using synthetic mammograms from generative adversarial networks in convolutional neural networks. Journal of Medical Imaging, 6(3), 031411.

85. Rybiałek, A., & Jeleń, Ł. (2020). Application of DenseNets for Classification of Breast Cancer Mammograms. In International Conference on Computer Information Systems and Industrial Management (pp. 266-277). Springer, Cham.

86. Jahangeer, G. S. B., & Rajkumar, T. D. (2021). Early detection of breast cancer using hybrid of series network and VGG-16. Multimedia Tools and Applications, 80(5), 7853-7886.

87. Khan, S., Islam, N., Jan, Z., Din, I. U., & Rodrigues, J. J. C. (2019). A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. Pattern Recognition Letters, 125, 1-6.

88. Chen, Y., Zhang, Q., Wu, Y., Liu, B., Wang, M., & Lin, Y. (2018). Fine-tuning ResNet for breast cancer classification from mammography. In The International Conference on Healthcare Science and Engineering (pp. 83-96). Springer, Singapore.

Appendix I. Documentation related to the articles of this thesis

- Solicitud para la presentación de la tesis por compendio de artículos

- Documentos de aceptación de coautoría