Deep Learning Algorithms Design for Medical Imaging in Computer-Aided Diagnosis Tasks

Karen Yaneth Sánchez Quiroga

M.S. in Electronic Engineering

Universidad Industrial de Santander

Facultad De Ingenierías Fisicomecánicas

Escuela De Ingenierías Eléctrica, Electrónica y de Telecomunicaciones

Bucaramanga

2023

Deep Learning Algorithms Design for Medical Imaging in Computer-Aided Diagnosis Tasks

Karen Yaneth Sánchez Quiroga

M.S. in Electronic Engineering

Doctoral thesis to qualify for the title of Doctor of Philosophy in Engineering

Advisor:

Ph.D. Henry Arguello

Professor at HDSP Research Group, Universidad Industrial de Santander, Colombia.

Co-Advisor:

Ph.D. Adrian Basarab

Professor at CREATIS Research Group, University Claude Bernard Lyon 1, France.

Universidad Industrial de Santander

Facultad De Ingenierías Fisicomecánicas

Escuela De Ingenierías Eléctrica, Electrónica y de Telecomunicaciones

Bucaramanga

2023

DEDICATION

Dedico mi tesis doctoral:

A mis padres, Yaneth y Carlos Enrique, por su gran amor y educación. Y sobre todo, por el regalo más grande que siempre me han hecho: confiar en mí. Su confianza ha hecho que todos los sueños parezcan posibles.

A mis hermanos Karol y Carlos Enrique Jr., que con su cariño me han acompañado en este camino y con sus habilidades me han animado a ser mi mejor versión.

A Carlos Alberto, el coautor del libro de mi vida, por su gran apoyo y cariño.

A una parte de mi ascendencia: los pacientes de lepra de Contratación, Santander. Estoy orgullosa de ser parte de esta historia.

A todas las niñas y jóvenes que creen en sus capacidades y en el poder de sus sueños.

ACKNOWLEDGEMENTS

I express my gratitude:

To the Universidad Industrial de Santander and the High Dimensional Signal Processing (HDSP) group for their valuable academic support over these years.

To the IRIT and CREATIS labs from the University of Toulouse and the University of Lyon, respectively, for their technological support and training during my research internship in France.

To my co-advisor, Professor Adrian Basarab, for his teachings, example, and dedication. Thank you for helping me become a better professional.

To my HDSP colleagues and co-authors for their friendship and company on this path. Especially to Andres Jerez.

To the examination committee of this Ph.D. thesis, for the dedicated time, effort, and feedback.

Content

List of Figures

List of Tables

page

RESUMEN

TÍTULO: DISEÑO DE ALGORITMOS DE APRENDIZAJE PROFUNDO PARA IMÁGENES MÉDICAS EN TAREAS DE DIÁGNOSTICO ASISTIDO POR COMPUTADOR [*]

AUTOR: KAREN YANETH SÁNCHEZ QUIROGA [**]

PALABRAS CLAVE: APRENDIZAJE PROFUNDO, IMÁGENES MÉDICAS, ADAPTACIÓN DE DOMINIO, & DIAGNÓSTICO ASISTIDO POR COMPUTADOR.

DESCRIPCIÓN:

Esta tesis doctoral examina la aplicación de la inteligencia artificial, específicamente modelos de aprendizaje profundo, en tareas de diagnóstico asistido por computadora (CAD) dentro de imágenes médicas. Si bien los modelos de aprendizaje profundo han revolucionado el campo médico, siguen dependiendo de grandes volúmenes de datos etiquetados, a menudo escasos y privados, y que varían entre los centros médicos. Esta tesis explora los conceptos de "adaptación de dominio" y "aumento de datos generativos" para abordar el problema de sobreajuste que surge de la falta de datos disponibles y afecta la precisión y generalización de los modelos. El primero aprovecha el conocimiento de un dominio de origen etiquetado para mejorar el rendimiento del modelo en un dominio de destino con datos limitados o sin datos etiquetados. El último se centra en la creación de datos sintéticos para aumentar el conjunto de entrenamiento, mejorando la generalización y precisión del modelo. En una contribución doble, esta tesis presenta primero un método para la selección inteligente, transformación e incorporación de radiografías de tórax de un conjunto de datos públicos en una red neuronal para mejorar la precisión de la clasificación de la neumonía. Este método mitiga los desafíos de trabajar con conjuntos de datos

---

[*] Tesis de doctorado

[**] Facultad de Ingenierías Fisicomecánicas. Escuela de Ingenierías Eléctrica, Electrónica y telecomunicaciones. Director: Ph.D. Henry Arguello Fuentes. Co-Director: Ph.D Adrian Basarab.

pequeños y variables entre diferentes hospitales. En segundo lugar, esta tesis presenta un método novedoso para el aumento de datos generativos para mejorar la precisión de la segmentación de tumores hepáticos en imágenes de resonancia magnética de múltiples contrastes. Al crear datos sintéticos para aumentar el conjunto de entrenamiento, este método busca mejorar la precisión y confiabilidad de la segmentación de tumores, una tarea vital para un diagnóstico preciso y una planificación del tratamiento. La investigación, desarrollada en colaboración con múltiples instituciones académicas y de investigación, tiene como objetivo en última instancia superar los desafíos en el análisis de imágenes médicas presentados por la escasez de datos etiquetados y mejorar las tareas CAD en diversas aplicaciones médicas.

ABSTRACT

TITLE: DEEP LEARNING ALGORITHMS DESIGN FOR MEDICAL IMAGING IN COMPUTER-AIDED DIAGNOSIS TASKS [*]

AUTHOR: KAREN YANETH SÁNCHEZ QUIROGA [**]

KEYWORDS: DEEP LEARNING; MEDICAL IMAGING, DOMAIN ADAPTATION, & COMPUTER-AIDED DIAGNOSIS.

DESCRIPTION:

This Ph.D. thesis examines the application of artificial intelligence, specifically deep learning models, in computer-aided diagnostic (CAD) tasks within medical imaging. While deep learning models have revolutionized the medical field, they remain dependent on large volumes of labeled data, often scarce and private, and that vary among medical centers. This thesis explores the concepts of "domain adaptation" and "generative data augmentation" to tackle the overfitting problem that arises from the lack of available data and affects the accuracy and generalization of models. The former leverages knowledge from a labeled source domain to improve the model's performance in a target domain with limited or no labeled data. The latter focuses on creating synthetic data to augment the training set, enhancing the model's generalizability and accuracy. In a two-pronged contribution, this thesis first presents a method for smart selection, transformation, and incorporation of chest X-rays from a public dataset into a neural network to improve the accuracy of pneumonia classification. This method mitigates the challenges of working with small, varying data sets across different hospitals. Secondly, this thesis introduces a novel method for generative data augmentation to enhance the

---

[*] Doctoral thesis

[**] Facultad de Ingenierías Fisicomecánicas. Escuela de Ingenierías Eléctrica, Electrónica y telecomunicaciones. Advisor: Ph.D. Henry Arguello Fuentes. Co-Advisor: Ph.D. Adrian Basarab.

accuracy of liver tumor segmentation in multi-contrast magnetic resonance images. By creating synthetic data to augment the training set, this method seeks to improve the precision and reliability of tumor segmentation, a vital task for accurate diagnosis and treatment planning. The research, developed in collaboration with multiple academic and research institutions, ultimately aims to overcome the challenges in medical imaging analysis presented by the scarcity of labeled data and enhance CAD tasks in various medical applications.

1. Introduction

Medical imaging plays a crucial role in modern healthcare, providing valuable insights into the human body and diagnosing multiple diseases. The rapid advancements in technology have led to the emergence of artificial intelligence (AI) as a powerful tool in medical imaging, particularly in computer-aided diagnosis (CAD) tasks. AI techniques, such as deep learning, have shown promising results in analyzing medical images and assisting healthcare professionals in the prediction, diagnosis, or treatment of diseases (Doi, 2007; Cheng et al., 2016).

Medical imaging analysis strategies include Gaussian mixture models, conditional random fields, statistical atlases, logistic regression, nearest-neighbour methods, support vector machines, random forests, recurrent neural networks, convolutional neural networks (CNNs), auto-encoders, deep reinforcement learning, among others (Riaz et al., 2020; Erickson et al., 2017; El-Naqa et al., 2002; Htay and Maung, 2018). The development of deep learning models has revolutionized the medical imaging computing field, offering the capability to process large amounts of data with high speed and extract complex characteristics not visible to the human eye. This progress has established deep learning as the most effective tool for analyzing medical images in recent years (Ravì et al., 2016). In a wide array of problem domains, ranging from image reconstruction to segmentation and registration, the current state-of-the-art solutions are based on deep neural networks (Abdeltawab et al., 2020; Esteva et al., 2017; Raman et al., 2019; Rahman et al., 2020; Hashmi et al., 2020; Toğaçar et al., 2020).

However, healthcare-oriented deep learning models face several challenges (Shen et al.,

2020; Lee et al., 2017). One significant obstacle is the accuracy of deep learning models depends heavily on large amounts of labeled data. Medical data, however, is often scarce, isolated, and private, and varies in appearance across different medical centers (Frid-Adar et al., 2018). This challenge limits the generalization of models, causing overfitting problems to a particular data domain source in medical imaging. For instance, a deep learning model trained on a large dataset from one medical center may not maintain its accuracy when tested on a dataset from another center, because of image distribution discrepancies (Choudhary et al., 2020).

In response to these challenges, this Ph.D. thesis proposes novel algorithms that leverage domain adaptation (Sanchez et al., 2022, 2021a) and generative data augmentation techniques (Sanchez et al., 2023a) to enhance the accuracy of CAD tasks in medical imaging. The first contribution addresses the limitation of over-fitting to a particular data domain. A novel method for smart selection, transformation, and incorporation of chest X-rays from a public dataset into a neural network system has been proposed (Sanchez et al., 2022). This method aims to improve the accuracy of classifying pneumonia and normal X-rays from a small and private dataset acquired at a clinical center located in a different continent.

The second contribution tackles the challenge of scarcity of labeled medical data. A novel method for generative data augmentation has been proposed, specifically, to enhance the accuracy of liver tumor segmentation in multi-contrast magnetic resonance images (Sanchez et al., 2023a). By generating synthetic data to augment the training set, the proposed method aims to improve the precision and reliability of tumor segmentation, which is a critical task for accurate diagnosis and treatment planning. The proposed method in this contribution was among the 2021 Artificial

Intelligence Data Challenge winners of the Radiology French Society SFR.

Furthermore, this thesis addresses the broader challenge of increasing the accuracy of computer vision tasks in medical applications. It covers the segmentation and follow-up of chronic ulcers in RGB photography (Sanchez et al., 2023b), the classification of skin lesions in RGB images (Calderón et al., 2021), and the classification of gastrointestinal anomalies in endoscopy photograms (Escobar et al., 2021). By exploring innovative approaches and algorithms, this research aims to contribute to the advancement of AI in medical imaging, ultimately improving patient care and outcomes.

In summary, this Ph.D. thesis focuses on leveraging domain adaptation and generative data augmentation techniques to overcome the challenges posed by the lack of labeled medical images and the heterogeneous nature of clinical data processing. It explores the impact of datasets size and multi-contrast images on the classification problem in CAD tasks, including pneumonia classification in chest X-rays and liver tumor segmentation in multi-contrast magnetic resonance images, while also addressing various computer vision tasks in medical applications. The overarching aim is to contribute to the refinement and enhancement of AI applications in healthcare.

1.1. Background

For a comprehensive and detailed understanding of this research thesis, it is necessary to address various integral subjects. These involve exploring medical imaging modalities, and techniques in computer-aided diagnosis, becoming familiar with domain adaptation principles, and data augmentation approaches in medical imaging. In the following sections, these aforementioned subjects, along with additional theoretical background, will be described. This will establish a conceptual framework to guide the subsequent discourse, evaluation, and analyses within this thesis.

1.1.1. Medical Imaging Modalities. Medical imaging is the set of techniques and processes used to create images of the human body, or parts of it, for clinical purposes. The healthcare area widely uses these techniques to diagnose and predict diseases and treatments. Usually, a radiologist reviews the patient's acquired medical images and writes a report with the findings. Then, the receiving physician defines a diagnosis and treatment plan based on the images and the radiologist's report. Besides, these types of images are essential for surgical planning and real-time monitoring of surgeries.

As shown in Fig. 1, according to the range of the electromagnetic spectrum used to illuminate, the type of sensor, and the technology of the acquisition equipment, medical images can be in various modalities. Each type of image is remarkable for specific applications; the modalities used in this thesis (X-ray and Magnetic Resonance) are detailed below:

X-ray/Radiography: The earliest form of medical imaging was the radiograph or X-ray.

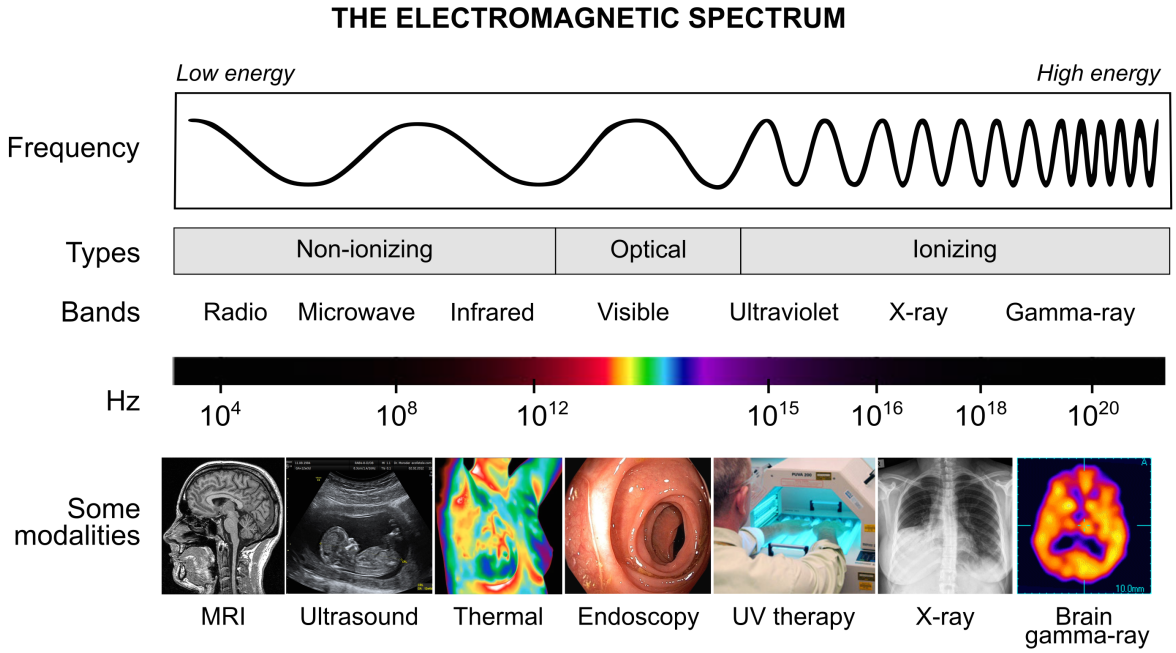


Figure 1. Medical imaging modalities according to the electromagnetic spectrum.

Even with all the new complex imaging techniques available, radiography is still an invaluable tool, particularly for imaging the skeleton. In radiography, the production of an image starts with a high-voltage electric current which creates a stream of electrons that are fired at a metal plate. The resulting interaction is the creation of X-rays which are collimated into a beam. This source produces X-rays which are directed toward the desired object to be imaged, in this case, the patient. The X-ray could pass through the patient, be absorbed by the patient, and/or be attenuated according to the density of each tissue and its respective atomic weights. The remaining X-rays are acquired on a film or digital detector (DXR), as shown in Fig. 2 (Ahmad et al., 2014). Despite its limitation as a 2D image with only a spectrum of black to white, X-ray remains one of the most helpful imaging techniques in clinical practice with the major advantages, disadvantages,

and applications listed in Table 1.

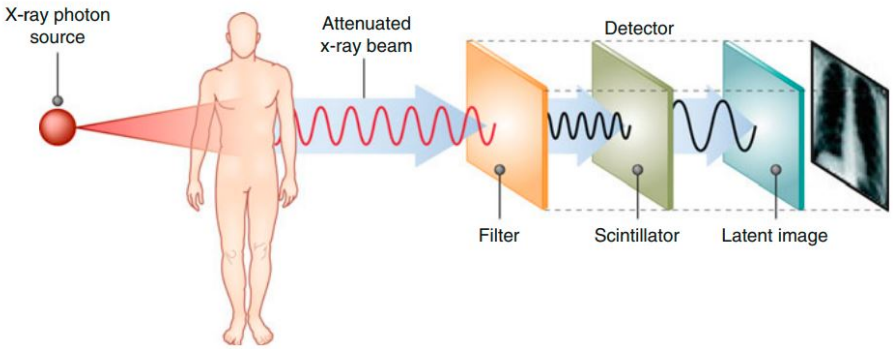


Figure 2. X-ray from source to image. Source: (Ahmad et al., 2014).

Table 1
Radiography: applications, advantages, and disadvantages

Applications	Fractures; bone diseases; pneumonia; pulmonary edema; intestinal obstructions; renal or gallblader stones.
Advantages	Low cost; widely available; portable; bedside.
Disadvantages	Radiation; limited color spectrum; 2D information.

Some applications of X-ray images, such as diagnosis of fractures, bone diseases, pneumonia, pulmonary edema, intestinal obstruction, or Hirschsprung's disease, are shown in Fig. 3.

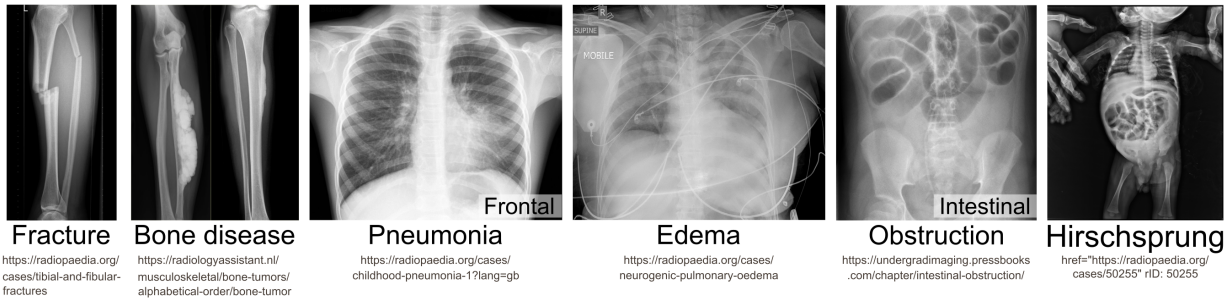


Figure 3. Some applications of X-ray images. Source: Adapted from (Dai et al., 2023).

Magnetic Resonance Imaging (MRI): is a powerful and non-invasive imaging technique

widely used in medical diagnosis and research. It provides detailed anatomical and functional information about the human body, enabling clinicians and researchers to visualize internal structures with exceptional clarity and precision.

At its core, MRI relies on the principle of nuclear magnetic resonance (NMR), a phenomenon observed in the 1940s. NMR occurs when atomic nuclei with an odd number of protons or neutrons are exposed to a strong magnetic field and radiofrequency pulses. In response to these stimuli, the nuclei align with the magnetic field and subsequently emit radiofrequency signals as they return to their original state.

A computer processes the detected signals to construct highly detailed images of the body's internal structures. MRI can distinguish between different types of tissues based on variations in the behavior of hydrogen atoms. For instance, hydrogen atoms in water molecules produce strong signals, allowing the differentiation of various soft tissues. By manipulating parameters such as the timing of radiofrequency pulses and the direction of magnetic gradients, MRI can generate different types of images, such as T1-weighted, T2-weighted, and proton density images, providing specific information about the tissue characteristics.

One of the significant advantages of MRI is its ability to visualize soft tissues, such as the brain, spinal cord, muscles, and internal organs, without the use of ionizing radiation. This makes MRI particularly valuable in diagnosing and monitoring a wide range of medical conditions, including neurological disorders, musculoskeletal injuries, cardiovascular diseases, and cancer.

MRI has revolutionized medical imaging and has become an indispensable tool in clinical practice and research. Its non-invasive nature, exceptional image quality, and versatility make it

invaluable for understanding and diagnosing human health and disease. Other imaging modalities such as computed tomography (CT), ultrasound, thermal imaging, visible spectrum imaging (endoscopic or dermatological), and gamma-ray imaging can be found in (Suetens, 2017).

1.1.2. Computer-aided Diagnosis: Traditional and Data-driven Techniques. The automated interpretation of medical images, facilitated by algorithms that strive to deduce mathematical or physical correlations between input and output parameters (commonly known as *computer-aided diagnosis*), has been a focal point of investigation for several decades. Some **traditional model-based approaches** that have been applied to this end are:

1. Generic models.
    - Mathematical models.
    - Biological or physics-based models.
2. Probabilistic models.
    - Gaussian mixture models.
    - Graphical models, including Markov (MRFs) and conditional random fields.
3. Population-based models.
    - Single-subject atlases.
    - Probabilistic atlases.
    - Statistical atlases (shape and appearance models).

The introduction of deep learning models has revolutionized the field of medical image computing. The following learning or **data-driven approaches** are some of the newer techniques

for computer-aided diagnosis:

1. Shallow learning models.
    - Regression.
    - Nearest-neighbor methods.
    - Support vector machines (SVMs).
    - Random forests
2. Deep learning (DL) models.
    - Recurrent neural networks.
    - Convolutional neural networks (CNNs).
    - Autoencoders.
    - Deep reinforcement learning.

In almost all medical imaging domains' tasks, including image reconstruction, organ segmentation, image registration, and interpretation, the current state-of-the-art, with better performance, is based on deep neural networks. Applications of deep learning to medical image analysis began to appear first in workshops and conferences and later in journals and have grown rapidly since 2015. Figure 4 shows a review of the number of articles published per year, the type of neural network, the task undertaken, imaging modality, and application area. Deep learning is now a dominant topic in medical conferences and in leading medical journals. In (Litjens et al., 2017), the authors published a comprehensive and dedicated review on deep learning in medical imaging. Although the report in (Litjens et al., 2017) shows the literature review until 2017, the same trend is

evidenced nowadays. Table 2 lists some recent medical imaging applications where deep learning has achieved state-of-the-art results.

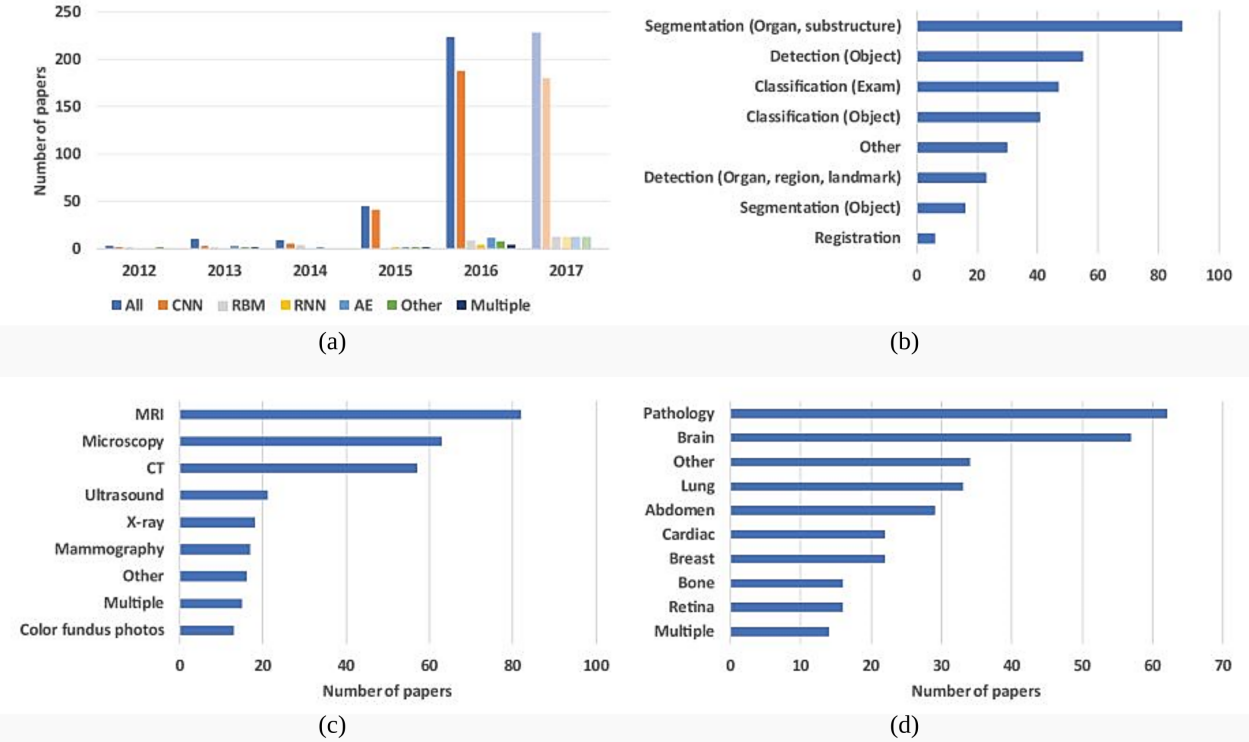


Figure 4. Breakdown of deep learning in medical image analysis. In quadrant (a) at the top-left, the chart displays the number of papers published by year and technique. Blue bars represent the total papers for each year, while orange, gray, yellow, and sky blue bars indicate papers using specific techniques like CNN (Convolutional Neural Networks), RBM (Restricted Boltzmann Machine), RNN (Recurrent Neural Networks), and AE (Autoencoders), respectively. Green bars denote other papers, and dark blue bars represent those utilizing multiple techniques. Quadrant (b), at the top right, shows the number of papers per task addressed. Quadrant (c) at the bottom left discriminates papers by medical imaging modality. Finally, quadrant (d) at the bottom right exhibits the number of papers per medical application area.

Source: (Litjens et al., 2017).

1.2. Related Works - Computational Techniques

1.2.1. Domain Adaptation in Medical Imaging.

In recent years, medical imaging computing has made great strides due to the rapid development of deep learning techniques.

Table 2
Some relevant applications in the health area based on deep learning methods.

Ref	Title	Year	Journal	Application
(Kooi et al., 2017)	Large scale deep learning for computer aided detection on mammographic lessions	2017	Medical Image Analysis	Mammography
(Raman et al., 2019)	Fundus photograph-based deep learning algorithms in detecting diabetic retinopathy	2019	The Eye	Optometry
(Esteva et al., 2017)	Dermatologist-level classification of skin cancer with deep neural networks	2017	Nature	Dermatology
(Chouhan et al., 2020)	A Novel Transfer Learning Based Approach for Pneumonia Detection in Chest X-ray Images	2020	Applied Sciences	Chest X-ray
(Yu et al., 2017)	Volumetric ConvNets with mixed residual connections for automated prostate segmentation from 3D MR images	2017	AAAI Conference	Prostate segmentation

However, access to large annotated medical image datasets is limited due to the tedious labeling process and privacy concerns. While multi-center datasets can increase the amount of annotated data, these sets vary due to different hospital technology configurations and diverse patient populations. Medical image databases with expert annotations are still at least one order of magnitude smaller than comparable databases in computer vision, e.g., ImageNet or MS-COCO. A disadvantage of deep learning approaches is that they often do not generalize well beyond data that are very similar to the training data. Additionally, the changes in distribution between the available training dataset and the dataset found in clinical practice lead to that the model previously trained by one dataset may fail when evaluating another dataset. In particular, the generalization ability of deep learning approaches is difficult to predict, often leading to failures during clinical deployment. Domain Adaptation (DA) is a transductive transfer learning approach that aims to transfer knowledge between domains by learning domain invariant transformations, which align the domain distributions (Figure 5). DA assumes that the source data is labeled, while the target domain can consist

of a fully-labeled dataset (supervised learning), a small set of labeled data (semi-supervised), or a completely unlabeled dataset (unsupervised).

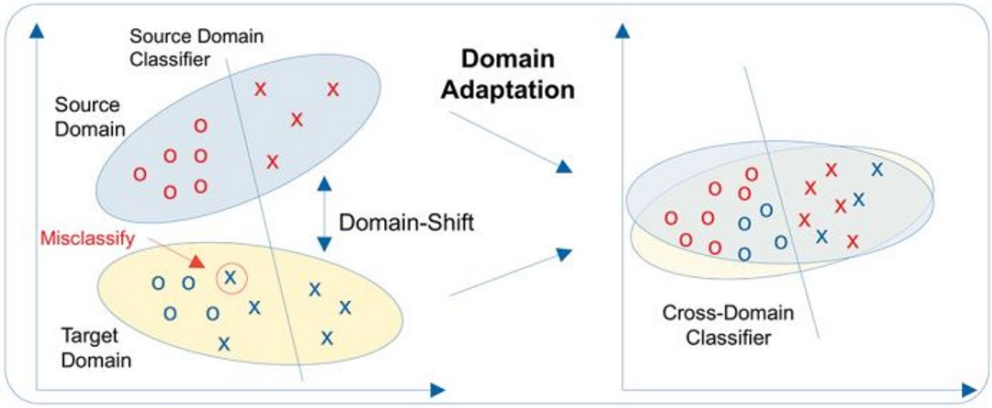


Figure 5. An overview of domain adaptation methods. Source: (Choudhary et al., 2020)

Figure 6 shows a categorization of medical imaging DA publications as per a) imaging modality, b) anatomy, and c) learning scenarios. Until today, MRI is the most used modality, followed by CT and X-rays. Anatomically, the brain, lungs, and heart are the most studied organs in this type of work. Regarding labeling approaches, a large percentage of the works correspond to applications on unlabeled target databases.

Some of the most outstanding domain adaptation works for classification in the health area are presented in Table 3.

This thesis addresses the problem of chest X-ray image classification on a small dataset and proposes a solution based on domain adaptation (Chapter 2).

1.2.2. Data Augmentation in Medical Imaging. Synthetic image generation methods have been developed in the literature to improve segmentation task results in medical imaging

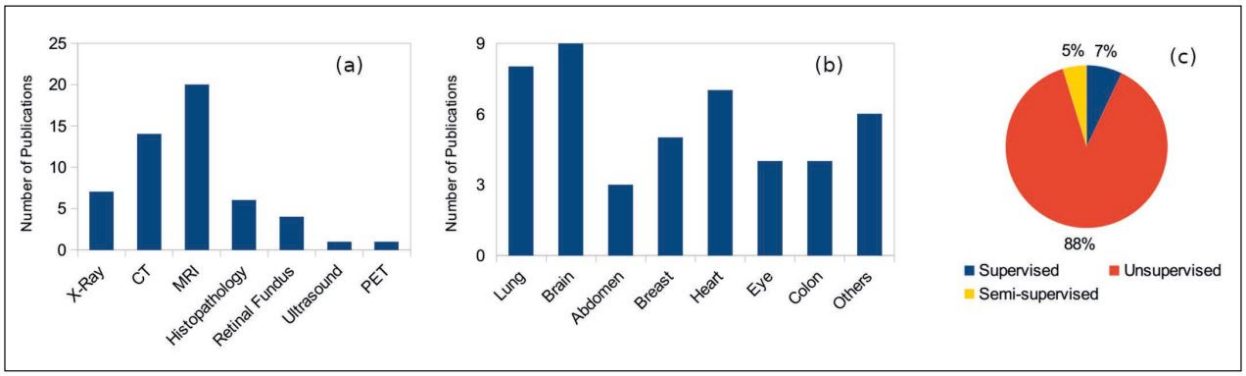


Figure 6. Categorization of medical imaging Domain Adaptation publications as per a) imaging modality; b) anatomy; c) learning scenarios. Source: (Choudhary et al., 2020)

Table 3
Summary of recent domain adaptation studies in medical imaging.

Ref	Short Title	Year	Journal	Application
(Tang et al., 2019)	TUNA-Net	2019	MICCAI Conference	Lung X-ray (different demographics)
(Pan et al., 2018)	MRI to PET	2018	MICCAI Conference	Brain supervised.
(Lafarge et al., 2017)	Variability of Histopathology	2017	DLMIA Workshop	Variability of Histopathology Images for Breast supervised
(Zhang et al., 2020b)	DA for Medical Diagnosis	2020	IEEE Trans. on Image Processing	Colon unsupervised from microscopy images
(Zhang et al., 2020a)	COVID-DA	2020	Arxiv (pending)	Chest X-ray (different diseases)

applications (Kebaili et al., 2023). State-of-the-art generative adversarial network (GAN) based architectures for medical image augmentation in segmentation tasks have been developed for different types of medical imaging modalities, including Fundus photography (Platscher et al., 2020), X-ray (Neff et al., 2017; Shen et al., 2023), computed tomography (CT) (Sandfort et al., 2019; Jiang et al., 2020; Shi et al., 2020), and mostly MRI. For instance, (Mok and Chung, 2019) proposes a coarse-to-fine generator architecture to capture the diversity of training sets and generate

augmented data, which led to a 3.5% improvement in Dice's coefficient on the Multimodal Brain Tumor Image Segmentation Benchmark *BRATS2015* MRI dataset (Bakas et al., 2017) compared to traditional data augmentation methods. Another technique for producing artificial MRI images with brain tumors is suggested in (Shin et al., 2018), where a GAN is trained using two public sets of brain MRI, the BRATS2015 dataset and the Alzheimer's Disease Neuroimaging Initiative *ADNI* dataset (Jack Jr et al., 2008). In (Jiang et al., 2019), the authors propose a cross-modality model that encodes the transformation of CT to pseudo-MRIs, achieving a segmentation DSC of 0.75 on a private dataset of 81 MRI scans of 28 patients with non-small cell lung cancers. Moreover, (Qasim et al., 2020) presents a DA protocol based on GANs that conditions networks at pixel-level and global-level information and injects synthetic images into the training set, significantly improving segmentation accuracy. This approach was validated on two medical datasets: BRATS2015 and the International Skin Imaging Collaboration *ISIC* (Codella et al., 2019). Lastly, (Platscher et al., 2022) combines image-to-image translation models (Pix2Pix, SPADE) and a CycleGAN to create a large database of synthetic stroke images using a small private dataset. This method outperforms the model trained on clinical images alone and yields significant improvements even with a small clinical dataset.

As was mentioned before, other methods for medical image augmentation in segmentation tasks are based on variational autoencoders (VAE). Some methods combine GAN and VAE for segmentation tasks. For instance, (Liang and Chen, 2021) generates realistic images of thyroid ultrasound and helps train a U-Net model to get better segmentation results. In (Gan and Wang, 2022), they propose an adversarial learned variational autoencoder method composed of only one

encoder and one generator. The discriminate objective is added to the encoder so that the method requires no extra discriminators. Results are evaluated on esophageal optical coherence tomography segmentation. The authors of (Huo et al., 2022) propose a brain lesion synthesis framework to expand both the quantity and diversity of the training dataset in 47 MRI scans of brain tumors, achieving 74.18 DICE in segmentation.

Finally, hybrid methods combine three deep learning strategies: GAN, VAE, and diffusion models. In MRI applications, some examples include Denoising Diffusion Probabilistic Models (DDPM) (Khader et al., 2022), which utilizes a two-stage process to encode the images into a low-dimensional latent space before training a diffusion probabilistic model on the latent representation of the data. Another method, brainSPADE (Fernandez et al., 2022), combines a synthetic diffusion-based label generator with a semantic image generator to produce fully synthetic brain labels on-demand for use in segmentation models, with or without specific pathologies of interest. Lastly, IITM-Diffusion (Wolleb et al., 2022) proposes a method for image-to-image translation using denoising diffusion implicit models, with both a regression and segmentation problem incorporated to guide image generation to the desired output. That approach was evaluated on both facial photos and MRI scans of brain tumors.

1.3. Related Works - Clinical Context

1.3.1. Chronic wounds medical assessment.

Chronic wounds treatment involves periodic visual inspection by medical staff for infection control, and moisture balance, where edge and size analysis are typically used to determine wound evolution (Bowers and Franco, 2020). However, incorrect wound management is common (Bowers and Franco, 2020), and can lead to

limb amputations, infection or even mortality (Yazdanpanah et al., 2015). For instance, (Gupta et al., 2021) shows that the incidence of wound complications increases in developing countries, as access to medical centers in rural areas is troublesome due to insufficient medical infrastructures and poor rural transportation, which leads to infrequent patient visits, inadequate treatment, and interrupted wounds tracking.

Computational methods (CM) have emerged as an alternative to support medical staff in the prognosis, diagnosis, and treatment of diseases through automatic medical image (Calderón et al., 2021), (Escobar et al., 2021). In the particular case of chronic wound analysis, image processing algorithms have been successfully employed to detect the region of interest containing the wound, or find segmentation maps that highlight the image pixels where the wound is present (Hsu et al., 2019; Mukherjee et al., 2017; Chairat et al., 2021). For instance, in (Song and Sacan, 2012) wound images are segmented using different algorithms like K-means clustering, region growing, edge detection, and thresholding. The outputs of these algorithms are then used as input to a multi-layer perception (MLP) and a radial basis function, which find the best segmentation map. Further, the work in (Fauzi et al., 2015) generates a Red-Yellow-Black-White (RYKW) probability map, that provides a segmentation according to granulation, slough, and eschar tissues. These probability maps work as a support to two segmentation algorithms, optimal thresholding and region growing.

More recently, fully convolutional networks with a transfer learning scheme have been used to address the segmentation task (Goyal et al., 2017; Long et al., 2015). They initialize its weights from generic image datasets like Imagenet or Pascal VOC segmentation dataset, and a refinement training is later performed with a dataset containing wound images (Goyal et al., 2017). Other

approaches adapt state-of-the-art networks like the VGG net or GoogleNet to perform the segmentation task (Long et al., 2015), build a two-step process to locate and segment the wound (Liu et al., 2017) or use specifically developed architectures like the U-net (Ronneberger et al., 2015). It is worth noting that most of these works focus on just obtaining a wound segmentation from the full image, i.e., without any pre-processing. In consequence, background details may negatively affect the network's performance. For this reason, the most recent studies on chronic wound analysis use a wound detection stage before segmentation to improve the results (Scebba et al., 2022), (Anisuzzaman et al., 2020), reducing errors induced by the background.

Even though the reported works provide accurate segmentation maps, the wound analysis is limited to individual images, which prevents medical staff from tracking the evolution/involution of the wound. Further, a general drawback of the deep learning segmentation methods lies in the fact that they are designed to work with a particular dataset (Goyal et al., 2017), which introduces a typical limitation for medical imaging applications, known as overfitting to a particular data domain (Sanchez et al., 2022). This problem implies that methods trained on a particular dataset exhibit lower performance when tested on other datasets acquired for different clinical centers or populations, due to substantial differences related to wound features, etiology of the wound, and image acquisition protocols. Chapter 4 presents the contribution of this thesis to this problem.

1.3.2. Skin lesions classification. Over the last decades, the research on computer-aided diagnostic (CAD) techniques has intensified to verify, support, or provide a second opinion to physicians' decisions (Haggenmüller et al., 2021; Adegun and Viriri, 2021; Göçeri, 2020b, 2021; Monroy et al., 2021).

Several works have used convolutional neural networks (CNN) to extract features and identify anomaly types on the HAM10000 dataset (Tschandl et al., 2018), which contains seven skin lesions labels tagged by experts. Miglani and Bathia (Miglani and Bhatia, 2020) used a pre-trained EfficientNet-B0 CNN architecture to train and classify via fine-tuning seven skin lesion types with 0.89 precision and 0.97 AUC. Mohapatra *et al.* (Mohapatra et al., 2020) used the lightweight MobileNetV1 model to classify skin lesions of the same dataset with 0.86 precision and incorporated it into an online platform. Chaturvedi *et al.* (Chaturvedi et al., 2020) also used a pre-trained MobileNet model to classify the HAM10000 dataset; they achieved 0.89 of precision. Emara *et al.* (Emara et al., 2019) proposed a modified version of the Inception-V4 architecture, adding a residual connection to fuse low-level to high-level features and achieving an accuracy of up to 0.8617 and 0.88 AUC. Chopade *et al.* (Chopade, 2020) proposed a lightweight CNN built from scratch with which they reach 0.89 of precision. Finally, Rishu Garg *et al.* (Garg et al., 2019) used image processing to remove noise, add resolution, make previous segmentation, including data augmentation and transfer learning on a pre-trained ResNet model achieving 0.88 of precision and 0.905 of accuracy. For comparison purposes, the accuracy classification results of the HAM10000 dataset with non-deep learning-based methods are 0.659, 0.6515, and 0.6586 for the random forest, XGBoost, and support vector classifiers, respectively. In Chapter 5, we present the state-of-the-art method for skin lesion classification on the HAM10000 dataset as part of the contribution of this thesis.

1.3.3. Gastrointestinal diseases classification. The automatic detection of diseases and gastrointestinal tract anomalies is challenging for medical experts, affecting patient

treatment decisions. Therefore, it is essential to implement CAD systems (Pang et al., 2021) that support the detection of anomalies and diseases in endoscopic images. Therefore, in (Pogorelov et al., 2017), the authors propose the classification of the Kvasir V1 dataset using three approaches, using global features, deep learning in CNNs, and transfer learning in deep learning. The best result was obtained by training from scratch a three-layer CNN, with an accuracy of 95.9%. In (Cogan et al., 2019), they performed a pre-processing of edge removal, contrast enhancement, filtering, color mapping, and scaling to each image in the Kvasir-V2 dataset and used the data augmentation technique. These images were used to train and test three CNNs: Inception-v4, Inception-ResNet-v2, and NASNet, obtaining the best result with the CNN of Inception-ResNet-v2 with an accuracy of 98.48% accuracy. On the other hand, taking advantage of the transfer learning technique, in (KahsayGebreslassie et al., 2019), the authors implemented a transfer learning technique with fine-tuning on two deep CNNs: ResNet50 and DenseNet121, pre-trained with the ImageNet dataset. Then, they classified the Kvasir V1 dataset, resulting in an accuracy of 87.8% in the residual network and 86.9% in the dense model. An approach of combining features extracted by several CNNs was addressed in (Gamage et al., 2019), they proposed to classify the Kvasir-V2 dataset using a set of six CNNs: DenseNet-201, ResNet-18, VGG-16, InceptionV3, Xception, InceptionResnetV2, with a global average pooling layer to obtain feature vectors. Then, they obtained the final feature vector for the classification task by adding the vectors generated in each CNN. Finally, they feed a single layer of decision that allows them to obtain an accuracy of 97.38%. Chapter 6 presents an efficient and highly accurate method that uses only one-fifth of the trainable parameters compared to the state-of-the-art methods.

1.4. Evaluation Metrics

The developed algorithms and methods in this dissertation for classification, segmentation, and image synthesis tasks are evaluated using the following metrics. In these equations, *TP* corresponds to true positives, *FP* to false positives, *TN* to true negatives, and *FN* to false negatives. The datasets used in all studies throughout this thesis were divided into train, validation, and test sets, with specific percentages detailed in each chapter. Consistently, for all chapters, we employed a k-fold cross-validation approach during the ablation studies and hyper-parameter selection, conducting this on both the train and validation sets. Whereas, the final reported results are based on a single run performed on the test sets. In chapters 1 to 5, we used a 10-fold strategy for k-fold cross-validation, while in the 6th chapter, we employed a 5-fold strategy.

Accuracy measures the ratio of correct predictions over the total number of samples evaluated. The Accuracy is calculated as

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}. \tag{1}$$

Precision measures samples correctly identified as positive between the total identified positive samples. High precision relates to the low false positive rate. The Precision is calculated as follows

$$Precision = \frac{TP}{TP + FP}. \tag{2}$$

Recall is the ratio of samples correctly identified as positive among all existing positive

samples. The Recall is calculated as

$$Recall = \frac{TP}{TP+FN}. \tag{3}$$

F1 score is defined as the harmonic mean between precision and recall. Therefore, this score takes both false positives and false negatives into account. The F1 score is calculated as

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall}. \tag{4}$$

Specificity measures the proportion of actual negatives that are correctly identified as such. The Specificity is calculated as follows

$$\frac{TN}{TN+FP} \tag{5}$$

AUC. The Area under the ROC Curve assesses the ability of the model to distinguish between classes. The Receiver operating characteristic (ROC) curve represents the True positive rate (TPR) versus False Positive Rate (FPR) parameters. TPR is the ratio between true positive and all positive data points, calculated as

$$TPR = \frac{TP}{FN+TP}. \tag{6}$$

FPR is the ratio between the negative data points mistakenly considered as positive and all negative data points, calculated as

$$FPR = \frac{FP}{FP + TN}. \tag{7}$$

Matthews correlation coefficient (MCC) is a measure for the quality of binary classifications, known as a correlation coefficient between all predicted and true values. The MCC is calculated as follows

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}. \tag{8}$$

Furthermore, the quality of the generated synthetic images in this thesis is evaluated via the ***Frechet Inception Distance (FID) score*** (Heusel et al., 2017). FID quantifies the distance between original and synthetic images created by a model generator. Mathematically, the quality assessment of the generated images can be expressed as follows

$$\ell_{FID} = \|\mu_R - \mu_F\|^2 - Tr\left(\Sigma_R + \Sigma_F - 2\Sigma_R\Sigma_F\right), \tag{9}$$

where $Tr(\cdot)$ is the trace of the argument matrix, the pair (R,F) denotes the distributions of the real and generated data, (μ_R, μ_F) represents the mean of each distribution, and (Σ_R, Σ_F) indicates their covariance matrices, respectively. A lower FID score indicates better quality images, while a high FID stands for lower quality synthetic images in a nearly linear relationship with those in the training dataset.

1.5. Research Overview

General objective

To develop computational algorithms through deep learning approaches to address the challenge raised by the lack of labeled medical images, the heterogeneous clinical data processing, and the classification problem in computer-aided diagnostic tasks.

Specific objectives

1. To process a small chest X-ray database acquired in the Toulouse Hospital and to select a freely available chest X-ray database to be used in a proposed classification system.

2. To design computational algorithms for the selection, synthetic generation, and classification of chest X-ray images with the statistical properties of a small database and a freely available database.

3. To develop a methodology to arrange and process heterogeneous clinical data that include medical images and feed a deep learning algorithm in a computer-aided diagnosis task.

4. To evaluate the algorithms designed in objectives 2 and 3 through simulations and compare the obtained classification results against other state-of-the-art approaches.

5. To verify the algorithms developed in objectives 2 and 3 in a real scenario with a medical specialist in the area.

1.6. Summary and Thesis Overview

This research thesis, consisting of six chapters, embodies two primary novel contributions in chapters 2 and 3. It also includes a research project with a societal perspective carried out in a Colombian hospital in chapter 4. As well as the application of deep learning in two practical medical imaging scenarios explored in chapters 5 and 6.

Specifically, this research thesis addresses the challenges presented in Chapter 1 by incorporating a carefully designed domain adaptation technique in Chapter 2. This technique will enable the selection, transformation, and integration of a limited number of X-ray images from a public dataset, thereby enhancing the classification task results of a smaller target dataset. Additionally, an investigation into a mask-guided data augmentation technique will be undertaken in 3 to improve the segmentation of liver tumors in multi-contrast magnetic resonance scans (MRI). This will involve integrating *realistic-looking* synthetic images generated using the proposed method. Moreover, this thesis will present three data-driven approaches to applying deep learning in the field of medicine. The extensions detailed in Chapters 4, 5, and 6 enable the measurement of chronic wounds resulting from leprosy, the classification of skin lesions, and the detection of gastrointestinal anomalies, respectively. The evaluation of these methods will primarily utilize RGB images of ulcers (chronic wounds), skin lesions, and endoscopic RGB images captured from within the human body.

1.6.1. Contributions and Outline. In summary, the main contributions of this thesis can be summarized as follows:

- **Chapter 2**: This chapter proposes the CX-DaGAN architecture, a three-step domain adaptation technique to overcome the overfitting problem of scarce medical datasets availability. This method uses a small chest X-ray dataset acquired during this thesis in the Hospital of Toulouse in France as the target domain and a public-large labelled dataset from China as the source domain. It is important to remark that an expert clinician in radiology validated the collected dataset, method, and results. Content of Chapter 2 includes the development of the Specific Objectives 1 and 2 of this PhD thesis and the partial development of Specific Objectives 4 and 5.

- **Chapter 3**: This chapter introduces the LT-SyGAN architecture, an innovative mask-guided synthetic image generation technique designed to tackle the challenge of scarcity of labeled data in medical imaging segmentation tasks. The method demonstrates enhanced accuracy in the segmentation of tumors in multiparametric MRIs and was recognized as one of the winners in the 2021 Data Augmentation Challenge hosted by the French Society of Radiology in Paris. Chapter 3 presents the development of Specific Objective 3 of this PhD thesis and the partial development of Specific Objectives 4 and 5.

- **Chapter 4**: This chapter presents a new algorithm for chronic wound tracking based on deep learning, which works on RGB images captured with smartphones, avoiding bulky and complicated acquisition setups. The framework integrates mainstream algorithms for medical image processing, including wound detection, segmentation, and quantitative analysis of area and perimeter. In addition, a new data set of chronic wounds of leprosy patients

residing in Contratación, Santander, Colombia, is provided to the scientific community. Currently, this algorithm is used by the medical staff of the Sanatorio de Contratación to support the treatment of their patients. The extension work presented in Chapter 4 contributes significantly to the development of Specific Objective 5 and broadens the scope of the general objective of this thesis.

- **Chapter 5**: This chapter presents a bilinear CNN approach capable of classifying seven skin lesion classes with the highest state-of-the-art accuracy and low computational cost. The framework proposed includes a data augmentation step to correct the data imbalance problem, transfer learning and fine-tuning to improve the classification performance while reducing the computational cost. Several simulations were executed over the HAM10000 dataset. The extension work presented in this Chapter 5 is related to Specific Objective 4 and addresses the classification problem raised in the general objective of this thesis.

- **Chapter 6**: This chapter presents a methodology to classify diseases and anomalies of the gastrointestinal tract using image processing and a transfer learning strategy. The proposed method is tested on the Kvasir-V2 dataset, containing 8000 endoscopic images divided into eight classes. The experiments show that the proposed approach achieves more than 98% accuracy during testing by only using one-fifth of the trainable parameters compared to the state-of-the-art methods. The extension work presented in this Chapter 6 is related to Specific Objective 4 and addresses the lack of labeled medical images and the classification problem raised in the general objective of this thesis.

All the algorithms and methods developed in this research were published in international conferences and journals. In addition to the principal contributions listed above, the development of novel methods to address other different research problems that are out of this thesis' objectives were published by the author of this dissertation in (Sanchez et al., 2021b), (Sanchez et al., 2021a), (Hinojosa et al., 2022), (Marquez et al., 2022), (Monsalve et al., 2022). The following section lists all the publications and author contributions during her doctoral studies.

1.7. List of Publications

International Journals

1. **K. Sanchez**, C. Hinojosa, H. Arguello, D. Kouamé, O. Meyrignac, and A. Basarab. (2022). "CX-DaGAN: Domain Adaptation for Pneumonia Diagnosis on a Small Chest X-ray Dataset". IEEE Transactions on Medical Imaging, vol. 41, pp. 3278 - 3288. `https://doi.org/10.1109/TMI.2022.3182168` *(Published)*

2. B. Monroy, **K. Sanchez**, P. Arguello, J. Estupiñan, J. Bacca, C. V. Correa, L. Valencia, J. C. Castillo, O. Mieles, H. Arguello, S. Castillo, and F. Rojas-Morales. (2023). "Automated Chronic Wounds Medical Assessment and Tracking Framework Based on Deep Learning". Computers in Biology and Medicine. `https://doi.org/10.1016/j.compbiomed.2023.107335` *(Published)*

3. **K. Sanchez**, C. Calderón, and H. Arguello. (2021). "BILSK: A Bilinear Convolutional Neural Network Approach for Skin Lesion Classification", in Computer Methods and Programs

in Biomedicine Update, vol. 1, pp. 100036. https://doi.org/10.1016/j.cmpbup.2021.100036 *(Published)*

4. C. Hinojosa, **K. Sanchez**, H. Garcia, and H. Arguello. (2022). "C-3SPCD: Coded aperture similarity constrained design for Spatio-spectral Classification of Single-Pixel Measurements", in Applied Optics, vol. 61, pp. E21-E32. https://doi.org/10.1364/AO.445326 *(Published)*

5. **K. Sanchez**, J. Bacca, L. Arévalo-Sánchez, H. Arguello, and S. Castillo. (2021). "Classification of cocoa beans based on their level of fermentation using spectral information", in Tecno-Lógica, vol. 24, no. 50, pp. 172-188. https://doi.org/10.22430/22565337.1654 *(Published)*

International Conferences with Proceedings

1. **K. Sanchez**, C. Hinojosa, K. Arias, H. Arguello, D. Kouamé, O. Meyrignac, and A. Basarab. “Mask-guided Data Augmentation for Multiparametric MRI Generation with a Rare Hepatocellular Carcinoma", in Proceedings of the IEEE 20th International Symposium on Biomedical Imaging ISBI, 2023. https://doi.org/10.1109/ISBI53787.2023.10230758

2. L. Valencia, O. Mieles, J.C. Castillo, B. Monroy, **K. Sanchez**, P. Arguello, J. Estupiñán, J. Bacca, C. Correa-Pugliese, H. Arguello, S. Castillo, and F. Rojas-Morales. “Plataforma de evaluación remota, objetiva y automatizada de la evolución de lesiones crónicas en extremidades por Enfermedad de Hansen", 48 Congreso Semana Quirúrgica Nacional, 2022.

3. **K. Sanchez**, C. Hinojosa, H. Arguello, S. Freiss, N. Sans, D. Kouamé, O. Meyrignac, and A. Basarab. “Subspace-based Domain Adaptation Using Similarity Constraints for Pneumonia Diagnosis within a Small Chest X-ray Image Dataset", in Proceedings of the IEEE 18th International Symposium on Biomedical Imaging ISBI, pp. 1232-1235, 2021. https://doi.org/10.1109/ISBI48211.2021.9434173

4. B. Monroy, J. Bacca, **K. Sanchez**, H. Arguello, and S. Castillo. “Two-step Deep Learning Framework for Chronic Wounds Detection and Segmentation: A Case Study in Colombia", in XXIII Symposium on Image, Signal Processing and Artificial Vision (STSIVA), 2021. https://doi.org/10.1109/STSIVA53688.2021.9592008

5. J. Escobar, **K. Sanchez**, C. Hinojosa, and H. Arguello. “Accurate Deep Learning based Gastrointestinal Disease Classification via Transfer Learning Strategy", in XXIII Symposium on Image, Signal Processing and Artificial Vision (STSIVA), 2021. https://doi.org/10.1109/STSIVA53688.2021.9591995

6. J. Escobar, N. Gomez, **K. Sanchez**, H. Arguello. “Transfer Learning with Convolutional Neural Network for Gastrointestinal Diseases Detection using Endoscopic Images", in IEEE Colombian Conference on Applications of Computational Intelligence (ColCACI), 2020. https://doi.org/10.1109/ColCACI50549.2020.9247847

7. Monsalve, J., Marquez, M., **K. Sanchez**, Hinojosa, C., Esnaola, I. and Arguello, H., 2022, September. “Cocosvi: Single Snapshot Compressive Spectral Video Via Covariance Matrix Estimation", in 2022 12th Workshop on Hyperspectral Imaging and Signal Processing: Evo-

lution in Remote Sensing (WHISPERS) (pp. 1-6). IEEE.

https://doi.org/10.1109/WHISPERS56178.2022.9955136

8. Marquez, M., Monsalve, J., Arias, K., **K. Sanchez** , Hinojosa, C. and Arguello, H., 2022, September. “Hierarchical Compressed Subspace Clustering Of Infrared Single-Pixel Measurements", in 2022 12th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS) (pp. 1-6). IEEE.

   https://doi.org/10.1109/WHISPERS56178.2022.9955096

9. **K. Sanchez**, C. Hinojosa, H. Garcia, and H. Arguello. “Compressed-domain Classification Algorithm for Spectral Imaging Based on Designed Single-Pixel Camera Codification", in 2021 OSA Imaging and Applied Optics Congress.

   https://doi.org/10.1364/COSI.2021.CTu2F.5

1.8. Undergraduate Advising

Thesis title:	Detection of pneumonia in X-ray images from image processing and neural networks
Advised student:	Natalia Andrea Gómez Albiadez
Advisor:	MSc. Karen Sanchez
Co-advisor:	Ph.D. Henry Arguello
Status:	Approved
Program:	Systems Engineering
University:	Universidad Industrial de Santander
Year:	2021

Thesis title:	Computer tool for assistance to the treatment of chronic wounds in patients from Hansen in the department of Santander
Advised student:	Juan Sebastian Estupiñan Cobos
Advisor:	MSc. Karen Sanchez
Co-advisor:	Ph.D. Henry Arguello
Status:	Approved
Award:	Outstanding undergraduate thesis
Program:	Systems Engineering
University:	Universidad Industrial de Santander
Year:	2023

1.9. List of Abbreviations

Abbreviations and notations in Chapter 2:

S Source dataset (large)

T Target dataset (small)

S_P, S_N Pneumonia and normal classes images, respectively, in S

T_P, T_N Pneumonia and normal classes images, respectively, in T

D Number of pixels in one image

$\mathbf{y}$ An image sample in $D \times 1$ vector form

d Number of eigenvalues from T

U_P, U_N Subspaces of T by classes

I_d Identity matrix

E_P, E_N Similarity scores between an image and the class subspace

n_1, n_2 Number of images from S_P, S_N

Abbreviations and notations in Chapter 3:

MRI Magnetic Resonance Imaging

CNN Convolutional Neural Network

mpMRI Multiparametric Magnetic Resonance Imaging

GAN Generative Adversarial Network

DA Data Augmentation

VAE Variational autoencoders

MMHCC Macrotrabecular-Massive Hepatocellular Carcinoma

Abbreviations and notations in Chapter 4:

TL Transfer learning

DA Domain Adaptation

R, *A* Radius and Area

$\mathscr{X} \in \mathscr{R}^{N \times M \times 3}$ input RGB image

c_w, c_p Bounding boxes for the wound region and the calibration pattern

D_θ Detection network

$\mathscr{X}_w, \mathscr{X}_p$ Crop wound and calibration pattern region

S_ϕ Segmentation network

$\hat{Y}$ Estimated segmentation map

A_w, P_w Calculated wound area and perimeter

Abbreviations and notations in Chapter 5:

$\mathscr{A}$ and $\mathscr{B}$ The output matrix from the first and the second model

lr Learning rate

$\mathscr{X} \in \mathbb{R}^{b \times L_1 \times L_2}$ The output computed from the two model outputs

m, n Spatial dimensions

Abbreviations and notations in Chapter 6:

CAD computer-aided diagnosis

2. Domain Adaptation Method for Pneumonia Diagnosis on a Small Chest X-Ray Dataset

This chapter presents CX-DaGAN, a new three-step domain adaptation technique to overcome the scarcity of labeled data in classification tasks. This method uses as the target domain a small dataset of 573 chest X-ray images acquired and labeled during this thesis in the Hospital of Toulouse in France. In contrast, a public-large labeled dataset of 5,849 chest X-ray images from pediatric patients aged one to five years old at the Guangzhou Women and Children's Medical Center in China is used as the source domain. It is important to remark that three expert clinicians in radiology validated the collected dataset, method, and results. Notice that the differences between these domains primarily pertain to the patient populations they represent and the geographic origins of the data. An in-depth analysis of factors such as image acquisition systems, image quality, and acquisition protocols was beyond the scope of this study but could be explored in future research.

Part of this section has been adapted from the journal paper (Sanchez et al., 2022) and the conference paper (Sanchez et al., 2021a). Section 1.9 summarises the notation used in this chapter.

Chapter contribution. Recent advances in deep learning led to several algorithms for the accurate diagnosis of pneumonia from chest X-rays. However, these models require large training medical datasets, which are sparse, isolated, and generally private. Furthermore, these models in medical imaging are known to over-fit to a particular data domain source, i.e., these algorithms do not conserve the same accuracy when tested on a dataset from another medical center, because of image distribution discrepancies. In this chapter, a domain adaptation and classification technique

is proposed to overcome the over-fit challenges on a small dataset.

This method proposes a new data augmentation technique to classify a small chest X-ray dataset taking advantage of a large public dataset acquired in different clinical center. The resulting algorithm is called CX-DaGAN, meaning Chest X-rays Domain adaptation with Generative Adversarial Network. The proposed technique is composed of three stages. First, a data selection from a source dataset through similarity constraints with the target dataset is performed. Then, a translation of the selected source images to the target domain with a generative adversarial network approach is processed. Finally, training of a CNN using both sets, target and translated sets, is performed in order to classify the target set. For testing, a sub-set from the target set is used. As we will detail later, the mean and standard deviation of the reported results are calculated using a 10-fold cross-validation strategy. The proposed approach achieved a notable increase in the target dataset F1-score, reaching up to 96.91% compared to 90.03% by standard transfer learning. In terms of Area under the curve (AUC), the proposed method reaches up to 0.96 against 0.87 obtained by standard transfer learning.

2.1. Proposed Method: CX-DaGAN

Overall, the proposed approach consists of three stages: first, it selects from the source dataset the images which are the most similar to the ones of the target domain; these images are chosen based on a similarity function that measures the subspace-projection error obtained by projecting the source data onto the target subspaces of each class: pathological and normal images. Second, it uses the selected source images as input for a Cycle-GAN to generate images in the

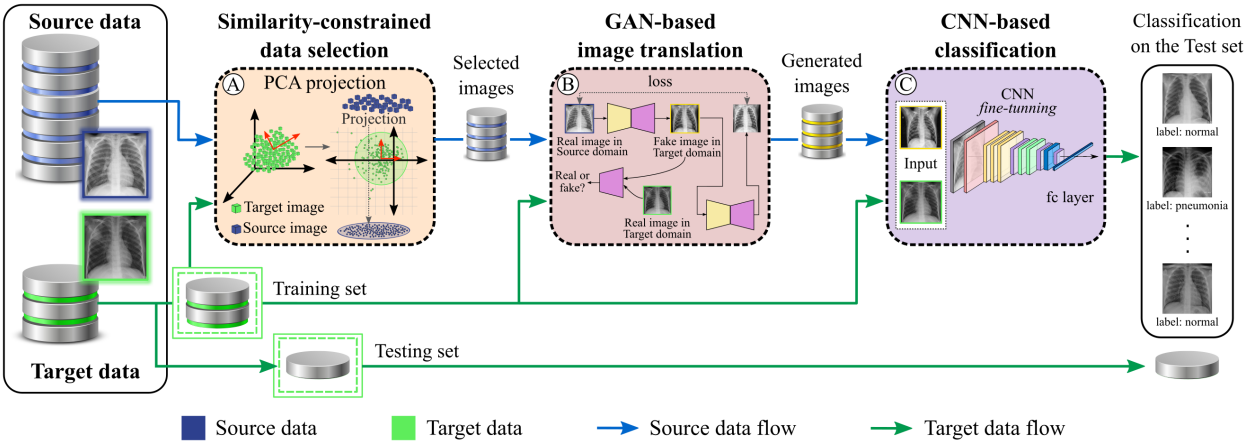


Figure 7. Proposed domain adaptation and classification framework. First, the chest X-ray images from the source and training target domains are fed into step (A) for the similarity-constrained data selection process. In step (A), images from the source domain are selected using a similarity function that measures the subspace-projection error obtained by projecting the source data onto the training target data domain. Then, the selected source images and the train target set are used as input to the proposed GAN-based image-to-image translation (Step B). The output of step (B) consists of synthetic images generated from the GAN that follows the target image distribution. Finally, we fine-tune a pre-trained CNN-based classification network for pneumonia diagnosis using the generated images in step (B) and the training target set as input. The performance of the proposed workflow is evaluated on the testing target set.

target domain. This second step uses images from the training target set to discriminate between the real and images generated by the GAN. Finally, in the third stage, the translated images and the small target train set are used to feed a CNN pneumonia/normal classification network and test it on the test set from the target dataset. The network parameters are further reduced by proposing a fine-tuning strategy. The proposed method is depicted in Fig. 7. The following sections provide more details of each stage (A, B, and C). Note that stages A and B are preprocessing steps before training a CNN network for pneumonia/normal classification of the X-ray images in step C.

2.1.1. Step A: Similarity-constrained Data Selection. This section introduces step A of the proposed method shown in Fig. 7 and detailed in Fig. 8. Let us denote by S the set

of medium to large labeled dataset (source), and by T the small labeled dataset (target). Throughout this chapter, we refer to "small" a dataset with less than 1,000 images, "medium" a dataset containing between 1,000 and 5,000 images, and "large" a dataset with more than 5,000 images. This choice is based on the number of images needed to train an X-ray-based pneumonia classifier from scratch with excellent (98%), good (95%), and insufficient (90%) accuracy (Barbedo, 2018). Furthermore, let us denote by $S_P \subseteq S$, and $S_N \subseteq S$ the subsets of images labeled as pneumonia and normal, respectively, in the source domain. Similarly, $T_P \subseteq T$ and $T_N \subseteq T$ denote the subsets of images labeled as pneumonia and normal in the target domain, respectively. Also, $\mathbf{y}_{S_P} \in \mathbb{R}^{D\times 1}$ represents an image sample from the S_P subset after reshaping it in a D-dimensional vector form, i.e., D corresponds to the total amount of pixels in the image. Similarly, $\mathbf{y}_{S_N} \in S_N, \mathbf{y}_{T_P} \in T_P$, and $\mathbf{y}_{T_N} \in T_N$ denote vectorized images from the corresponding subsets.

First, every source and target data from T_P and T_N is normalized to have zero mean and unit standard deviation. Then, principal component analysis (PCA) is applied to select, for each domain, d eigenvectors corresponding to the d largest eigenvalues. These eigenvectors are used as bases of the subspace for each subset. Specifically, the matrices $U_P \in \mathbb{R}^{D\times d}$ and $U_N \in \mathbb{R}^{D\times d}$ are obtained, used as the subspaces. Note that U_P and U_N are semi-orthonormal, thus $U_P' U_P = I_d$ and $U_N' U_N = I_d$, where I_d is the identity matrix of size d^2 and $'$ denote the transpose of the matrix. Furthermore, two types of projections are performed: (1) project every image from each source class onto the target subspace of the same class, i.e., $\mathbf{y}_{S_P}$ is projected onto U_P and $\mathbf{y}_{S_N}$ onto U_N; (2) project every image from each source class onto the target subspace of the opposite class, i.e., $\mathbf{y}_{S_P}$ is projected onto U_N and $\mathbf{y}_{S_N}$ onto U_P. Based on these projections, the following similarity functions considering the

projection errors are defined

$$E_P(\mathbf{y}) = \|U_P U_P' \mathbf{y} - \mathbf{y}\|_2, \tag{10}$$

$$E_N(\mathbf{y}) = \|U_N U_N' \mathbf{y} - \mathbf{y}\|_2, \tag{11}$$

where $||*||_2$ stands for the ℓ_2-norm. (10) and (11) are used to project all images in S_P and S_N and build four error vectors: $\mathbf{q}_1 \in \mathbb{R}^{n_1}, \mathbf{q}_2 \in \mathbb{R}^{n_2}, \mathbf{q}_3 \in \mathbb{R}^{n_1}$ and $\mathbf{q}_4 \in \mathbb{R}^{n_2}$. Specifically, the vector $\mathbf{q}_1$ is built as $\mathbf{q}_1 = \left\{E_P(\mathbf{y}_{S_P}^1), \cdots, E_P(\mathbf{y}_{S_P}^{n_1})\right\}$ using all images (n_1) from S_P. Similarly, $\mathbf{q}_2 = \left\{E_N(\mathbf{y}_{S_N}^1), \cdots, E_N(\mathbf{y}_{S_N}^{n_2})\right\}$ is formed using all images (n_2) from S_N; $\mathbf{q}_3 = \left\{E_N(\mathbf{y}_{S_P}^1), \cdots, E_N(\mathbf{y}_{S_P}^{n_1})\right\}$; and $\mathbf{q}_4 = \left\{E_P(\mathbf{y}_{S_N}^1), \cdots, E_P(\mathbf{y}_{S_N}^{n_2})\right\}$.

The vectors $\mathbf{q}_1$ and $\mathbf{q}_2$ are sorted in ascending order, and vectors $\mathbf{q}_3$ and $\mathbf{q}_4$ in descending order. Finally, considering the first k values from each error vector, the corresponding k images from the source domain (S_N and S_P) are selected and used as input for the proposed cycle-GAN-based network shown in step (B) of Fig. 7. Note that in this document, "similarity" is named for the minimum mathematical difference between pixel values of two spatial sets. A human inspection of the images was not considered to establish the similarity. This way of selecting images from the source set is guided by the idea of choosing the images that are the most similar intra-class to those from the target set, and the most different inter-class between the two domains.

2.1.2. Step B: GAN-based Image-to-Image Translation. After selecting the most similar images from the source dataset (with respect to the target domain images) using the proposed subspace-based approach, a multi-domain and unpaired image-to-image (I2I) translation

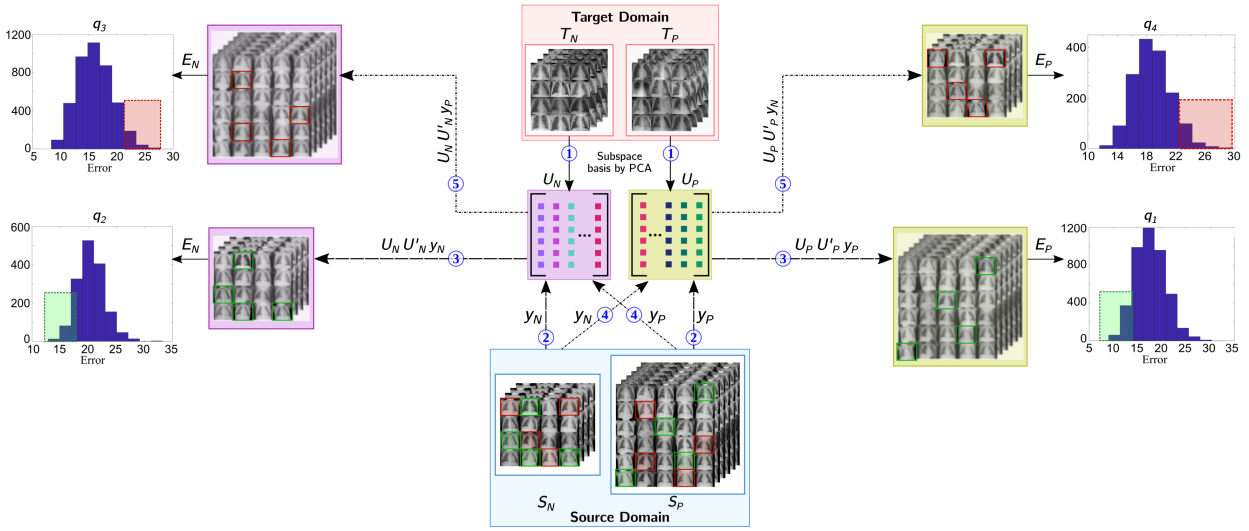


Figure 8. Step A of Fig 7. Similarity-constrained data selection via subspace projection error. The circled blue numbers represent the order of the stages in the figure. First, we calculate a subspace basis for each target class using principal component analysis (PCA). Second, the images from each class in the source domain are projected onto the subspaces of **T** within their corresponding classes. Third, the results of these projections are obtained. Fourth, each source image is projected onto the opposite class subspace basis, and the results are obtained in the fifth line. Sixth, for each case, we calculate the projection error between the projected and original corresponding image through the similarity functions shown in Eq. (1)-(2). Finally, we select the images with the lowest projection error when projecting the images onto subspaces of the same classes and the largest error when projecting the images onto the subspaces of the opposite classes.

network is used to generate images following the target domain distribution. Specifically, Step B generates the same number of images that were selected from the source dataset by Step A. The proposed network is depicted in Fig. 7 (step B) and detailed in Fig. 9. Specifically, the I2I translation strategy (Zhu et al., 2017) is adopted to map images from the two domains corresponding to the same class ($\mathbf{y}_{S_P} \rightleftarrows \mathbf{y}_{T_P}$, $\mathbf{y}_{S_N} \rightleftarrows \mathbf{y}_{T_N}$).

In this work, the Cycle-GAN (Zhu et al., 2017) is adopted to learn two mappings: $S \rightarrow T$, and $T \rightarrow S$, with generators $G_{S\rightarrow T}(\mathbf{y}_S)$ and $G_{T\rightarrow S}(\mathbf{y}_T)$, so that discriminators D_T and D_S cannot distinguish between real and synthetic images generated by the generators. In a Cycle-GAN network, $G_{S\rightarrow T}$ and its discriminator D_T are used to define the adversarial learning objective loss as

$$\begin{aligned}\mathscr{L}_{\text{adv}}\left(G_{S\rightarrow T}, \mathscr{D}_T\right) = \mathbb{E}_{y_s\sim\mathbf{y}_S}\left[\log(1-D_T(G_{S\rightarrow T}(y_s))\right] \\ +\mathbb{E}_{y_t\sim\mathbf{y}_T}\left[\log D_T(y_t)\right],\end{aligned} \tag{12}$$

where $\mathbb{E}[\cdot]$ denotes the expected value over the data instances specified in the subindex.

A similar adversarial loss can be designed for mapping $G_{T\rightarrow S}$ and its discriminator D_S as well, i.e., $\min_{G_{T\rightarrow S}}\max_{D_S}\mathscr{L}_{adv}(G_{T\rightarrow S}, D_S)$. To preserve sufficient low-level content information, we use the *cycle-consistency* loss (Zhu et al., 2017) to force the reconstructed synthetic images y'_s

and y'_p to resemble their inputs y_s and y_t

$$\mathscr{L}_{\text{cyc}}(G_{S\to T}, G_{T\to S}) = \mathbb{E}_{y_s\sim \mathbf{y}_S}\left[\|y'_s - y_s\|_1\right] + \mathbb{E}_{y_t\sim \mathbf{y}_T}\left[\|y'_t - y_t\|_1\right], \tag{13}$$

where $y'_s = G_{T\to S}(G_{S\to T}(y_s))$, $y'_t = G_{S\to T}(G_{T\to S}(y_t))$, and $\|\cdot\|_1$ is the ℓ_1-norm. The generative adversarial training with cycle consistency enables synthesizing realistic-looking radiographs across domains. However, there is no guarantee that high-level semantics would be preserved during translation, thus decreasing the classification accuracy.

To improve the classification accuracy on the generated synthetic target images, a classification model $\mathscr{F}$ was included in the GAN-based network to guide the training by considering the classification loss. Specifically, the classification model $\mathscr{F}$ is learned on the synthetic target data $\bar{T} = \{G_{S\to T}(\mathbf{y}_S), \bar{L}_S\}$, where $\bar{L}_S$ represent the corresponding labels (Normal or Pneumonia) of the synthetic $\bar{T}$ data. The binary cross-entropy loss was used to classify the two categories:

$$\mathscr{L}_{\text{cls}}(\mathscr{F}, \bar{T}) = -\mathbb{E}_{\bar{t}\sim\bar{T}} \sum_{c=1}^{C} \mathbb{1}_c \log\left(\sigma(\mathscr{F}^{(c)}(y_{\bar{t}})\right), \tag{14}$$

where σ is the softmax function, $\mathbb{1}_c = 1$ if an input image $y_{\bar{t}}$ belongs to class $c \in C = \{\text{Normal}, \text{Pneumonia}\}$, otherwise $\mathbb{1}_c = 0$. The final objective of our proposed GAN-based network for synthetic target images generation is the sum of adversarial learning losses, cycle consistency loss, and classification

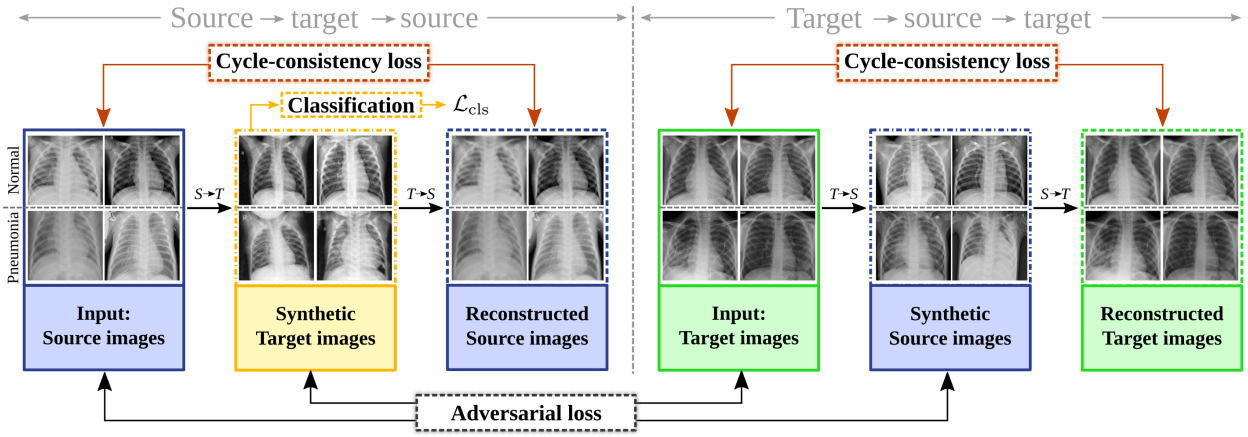


Figure 9. Step B of Fig 7. Proposed GAN-based I2I architecture used to translate chest X-ray images from the source domain (Normal or Pneumonia) to its corresponding class in the target domain. The network also translates back the generated images to the source domain to maintain cycle consistency. To ensure the generated synthetic images maintain the high-level semantics after the transformation and improve classification accuracy, a classification model was incorporated to guide the training by considering the classification loss. The training set of the target domain was used to measure the adversarial loss during training.

loss:

$$\mathscr{L} = \mathscr{L}_{\text{adv}}\left(G_{S\to T}, \mathscr{D}_T\right) + \mathscr{L}_{\text{adv}}\left(G_{T\to S}, \mathscr{D}_S\right) \tag{15}$$

$$+\lambda \mathscr{L}_{\text{cyc}}(G_{S\to T}, G_{T\to S}) + \mathscr{L}_{\text{cls}}(\mathscr{F}, \bar{T}).$$

It is worth mentioning that, for ease of notation, the above equations were developed without distinguishing between the two classes. However, during implementation, four generators described the mappings from source/target images with pneumonia/normal to target/source images, respectively. Similarly, four discriminators were associated with each generator output. Also, the same classification network ($\mathscr{F}$) was used in the last step of the proposed framework shown in Fig. 7, which is described in the following section.

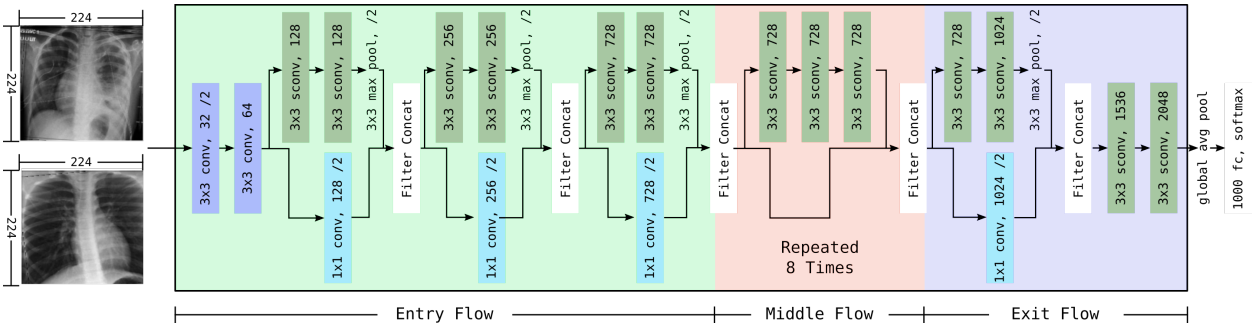


Figure 10. Step C of Fig 7. It starts by using the pre-trained weights of the Xception architecture on ImageNet and investigating different fine-tuning settings to achieve the highest accuracy while training fewer parameters. It uses the generated images obtained by following steps (A) and (B) of the proposed workflow as input to this architecture.

2.1.3. Step C: CNN-based Classification. The augmented training dataset obtained following steps A and B detailed in the previous sections is used to feed a convolutional neural network (CNN) trained to perform the final classification. In this work, the Xception CNN was adopted as the backbone (see ablation study in Table 7) to extract features and used a fully connected layer at the end of the network to perform the classification. The Xception (Chollet, 2017a) is an extension of the Inception architecture which replaces the standard Inception modules with depthwise separable convolutions. Instead of partitioning input data into several compressed chunks, it maps the spatial correlations for each output channel separately and then performs a 1×1 depthwise convolution to capture cross-channel correlation. This is essentially equivalent to an existing operation known as a "depthwise separable convolution", which consists of a depthwise convolution (a spatial convolution performed independently for each channel) followed by a pointwise convolution (a 1×1 convolution across channels). The Xception architecture is shown in Fig. 10. In general, the network can be divided into three sections: the entry, middle, and exit

flow, where the middle flow is repeated eight times. Given the limited size of the medical training dataset, pre-trained weights from the large ImageNet visual dataset were used to initialize the network, and the layers of the Xception network were tuned to adapt to the specific pneumonia detection task, avoiding training from scratch. Section 2.2.3 investigates different fine-tuning settings to achieve high accuracy while training fewer parameters.

2.2. Simulations and Results

This section illustrates the efficiency of the proposed CX-DaGAN classification algorithm for normal and pneumonia images on a small chest X-rays dataset. All simulations were implemented in Python with Tensorflow 2.3 and ran on an Nvidia Quadro RTX 6000 GPU with 24 GB of memory.

2.2.1. Datasets. The proposed CX-DaGAN algorithm was tested using two datasets for domain adaptation: a large source dataset (S) from which we extracted and transformed a selected number of images; second, a small target dataset (T) from which we performed the classification. This work used a private dataset as T and a publicly available dataset as S.

Specifically, the "Chest X-ray Images (Pneumonia) dataset"[1] was used as S which consists of $5,849$ labeled images acquired in the Guangzhou Women and Children's Medical Center in China (Kermany et al., 2018). The 8-bit X-ray grayscale images are separated into $4,266$ pneumonia (S_P) and $1,583$ normal (S_N).

On the other hand, 573 chest X-ray images acquired at the Toulouse University Hospital

[1] The dataset is available for free download at https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia

in France were used as the target dataset T. Two expert radiologists labeled each image of T as pneumonia or normal class. The dataset T is divided into 275 normal and 298 pneumonia images. In the following experiments, we split T in 400 images for training and the remaining 173 for testing, corresponding to $69,8\%$ and $30,2\%$ of the data, respectively. To train and evaluate the proposed CX-DaGAN method, images with a fixed size of 224×224 pixels were considered.

Metrics. To quantitatively evaluate the performance of the proposed method, three metrics (Hossin and Sulaiman, 2015) were computed: Accuracy (ACC), F1 score (F1), and Area under the ROC curve (AUC).

2.2.2. Quantitative Classification Results. For the testing chest X-ray images, the CX-DaGAN algorithm was used to predict the probability of pneumonia. By comparing with the binary ground-truth labels, the overall accuracy of the proposed method was calculated in extensive simulations, as shown in Table 5. In this table, each value corresponds to the average and standard deviation of a 10-fold cross-validation strategy of the proposed method, evaluated on 173 images from the target dataset. In this Table 5, the number of images used for training throughout the entire framework was varied. Specifically, between columns and rows, the number of images from the target and source, respectively, changes. Since the small size of the target dataset, the maximum number of target images used for training was 400 (see columns in Table 5). On the other hand, the number of images from the source domain chosen by the similarity-constrained data selection step was simultaneously varied. It is worth noting that this method was designed for selecting the same amount of images from each class in S, i.e., after the similarity phase (A), the proposed method ensures 50% of normal (disease-free) and 50% of pneumonia images to feed the step B.

Thus, considering the composition of the unbalanced public dataset used in these experiments, the maximum number of X-rays from *S* was 2400, corresponding to 1200 images of each class (see rows in Table 5).

In deep learning, especially for CNNs, it is well known that a greater number of labeled training samples leads to better classification (Diker et al., 2019), (Barbedo, 2018). In general terms, Table 5 shows that the classification accuracy is lower using fewer target images for training the CX-DaGAN algorithm. Conversely, the best results are concentrated in the right area of Table 5, where we used more images from the target dataset. However, it can also be observed vertically that using more images from the source domain does not necessarily imply better accuracy. Instead, there is a central area with combinations of data that are particularly interesting.

In the highest case of average precision, the data showed that the pneumonia prediction accuracy obtained by the CX-DaGAN proposed algorithm is higher than 97% when training using 250 images from the target and 400 images from the source set. Similarly, high accuracy results can be achieved by training the algorithm with 400 images from the target and 200 images from the source. For a deeper analysis of results in Table 5, Fig. 11 presents the 20 highest average classification results, without considering their standard deviation (STD), organized in descending order and graphed with their respective STD. Note that the best classification average value is 97,78. However, its corresponding STD is 0,7. On the other hand, the third-highest average rating in this plot is 96,75. In this case, the STD is lower (0,4), which can be interpreted as a statistically more stable and reliable result. Therefore, selecting only one combination of target/source data as the "best" is impossible since results can vary between the ranges defined by the STD. Note that

this table aims to provide results with different source/target data combinations to allow the user to select the best combination according to the data availability.

Figure 12 shows a histogram with all the error values of Table 5 to visualize their behavior. Note that the mean of the error values is 0,732, and the standard deviation of the error values is 0,231.

The ten highest average accuracy results from Table 5, their STD, and the corresponding target/source data combination for training are shown in Fig. 13. Note that the "x" and "y" axis correspond to the number of "target" and "source" images used for training, respectively. The size of each dot is associated with the accuracy value: larger dots indicate better precision, as conventions dictate; the color of the dot indicates the value of the standard deviation, where red corresponds to a higher STD and green to a lower one. The aim of this plot is to expose the most accurate options achieved by the proposed method, according to the required number of images for training and their STD.

The results shown in Tables 5, 6, and 7 were obtained using $d = 200$ in step A. Conversely, OA results of the CX-DaGAN method using other values of $d = 60, 100, 140, 180$ for step A are presented in Table 4. Note that subspaces U_P and U_N have dimensions $\mathbb{R}^{D\times d}$ and consequently the upper bound of d in this example is 298 for U_P and 275 for U_N, corresponding to T_P and T_N, respectively. One may remark from the OA values in Table 4 that the choice of d has limited influence on the accuracy of the CX-DaGAN method. This observation is sustained by the fact that 88% of the images selected from S within step A were the same for all d values evaluated. Accordingly, $d = 200$ eigenvectors ($\approx 70\%$) were chosen to run the following simulations, given that this value leads to the best balance between classification accuracy and algorithm performance.

Table 4
Classification results in terms of Overall Accuracy (OA) of the CX-DaGAN algorithm for different values of d in Step A (average over 10 realizations).

d	**60**	**100**	**140**	**180**	**200**
	(20,9%)	(34,9%)	(48,8%)	(62,8%)	(69,8%)
OA	95,0 ± 0,3	96,2 ± 0,3	95,5 ± 0,4	96,8 ± 0,3	97,6 ± 0,4

Table 5
Quantitative classification results for different data combinations (target/source) in the training (average over 10 realizations).

		(↑) **Overall accuracy** (%)								
Images	**Target →**	**0**	**50**	**100**	**150**	**200**	**250**	**300**	**350**	**400**
↓ Source	**%**	0%	12,5%	25%	37,5%	50%	62,5%	75%	87,5%	100%
0	0%	51,33 ± 1,2	74,92 ± 1,1	84,51 ± 1,2	89,75 ± 1,0	90,56 ± 0,8	92,04 ± 0,9	92,81 ± 0,8	94,69 ± 0,8	91,09 ± 0,9
50	0,9%	52,33 ± 1,0	80,39 ± 1,1	81,32 ± 0,8	88,48 ± 0,9	90,92 ± 0,9	91,08 ± 0,6	94,09 ± 0,6	94,75 ± 0,9	95,75 ± 0,7
100	1,7%	52,23 ± 1,1	78,85 ± 0,8	90,15 ± 1,2	92,81 ± 1,0	90,81 ± 0,8	95,21 ± 1,1	94,12 ± 0,4	94,90 ± 0,8	94,90 ± 0,6
150	2,6%	58,64 ± 0,9	83,15 ± 1,0	84,84 ± 0,8	86,71 ± 0,8	89,90 ± 1,0	93,89 ± 0,8	93,81 ± 0,9	95,58 ± 0,7	95,56 ± 0,9
200	3,4%	59,34 ± 1,1	80,42 ± 1,0	89,27 ± 0,9	91,01 ± 0,9	89,38 ± 0,8	94,68 ± 0,9	94,85 ± 0,7	95,62 ± 0,6	97,02 ± 0,8
250	4,3%	59,77 ± 0,9	86,23 ± 0,7	83,07 ± 0,8	84,34 ± 1,2	89,38 ± 1,0	91,82 ± 1,2	94,95 ± 0,6	95,75 ± 0,6	95,66 ± 0,7
300	5,1%	62,44 ± 1,0	81,19 ± 0,8	84,84 ± 0,7	87,38 ± 0,6	90,06 ± 0,7	93,41 ± 0,4	94,35 ± 0,6	94,89 ± 0,5	95,46 ± 0,5
350	6,0%	61,88 ± 1,0	80,61 ± 0,9	84,84 ± 0,6	89,37 ± 0,7	91,04 ± 0,6	91,06 ± 0,7	94,51 ± 0,4	95,01 ± 0,5	96,70 ± 0,6
400	6,8%	63,11 ± 1,2	80,51 ± 0,9	83,96 ± 1,0	86,91 ± 0,8	92,61 ± 0,5	97,78 ± 0,7	95,02 ± 0,4	96,75 ± 0,4	95,85 ± 0,3
600	10,3%	66,37 ± 0,9	76,11 ± 1,0	83,19 ± 0,6	85,84 ± 0,7	87,61 ± 0,6	84,07 ± 0,6	87,61 ± 0,5	90,27 ± 0,4	92,04 ± 0,4
1000	17,1%	67,26 ± 0,9	79,65 ± 0,6	87,61 ± 0,8	85,84 ± 0,5	87,61 ± 0,4	88,5 ± 0,7	90,27 ± 0,5	86,73 ± 0,5	93,81 ± 0,5
1600	27,4%	68,14 ± 0,8	75,22 ± 0,6	76,99 ± 0,7	80,53 ± 0,5	87,61 ± 0,4	84,96 ± 0,4	86.73 ± 0,4	92,04 ± 0,4	90,27 ± 0,4
2400	41,0%	70,80 ± 0.7	82,3 ± 0,6	83,19 ± 0,5	80,53 ± 0,6	80,53 ± 0,7	89,38 ± 0,5	86,73 ± 0,5	88,5 ± 0,4	91,15 ± 0,3

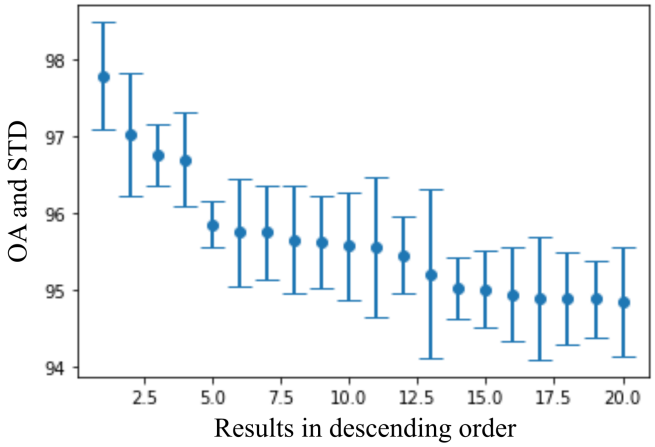


Figure 11. Twenty highest overall accuracy results from Table 5 in descending order and their STD.

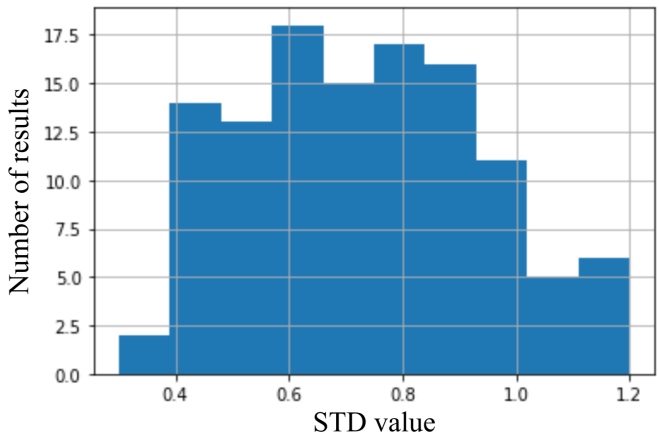


Figure 12. Histogram of the error values from the results presented in Table 5.

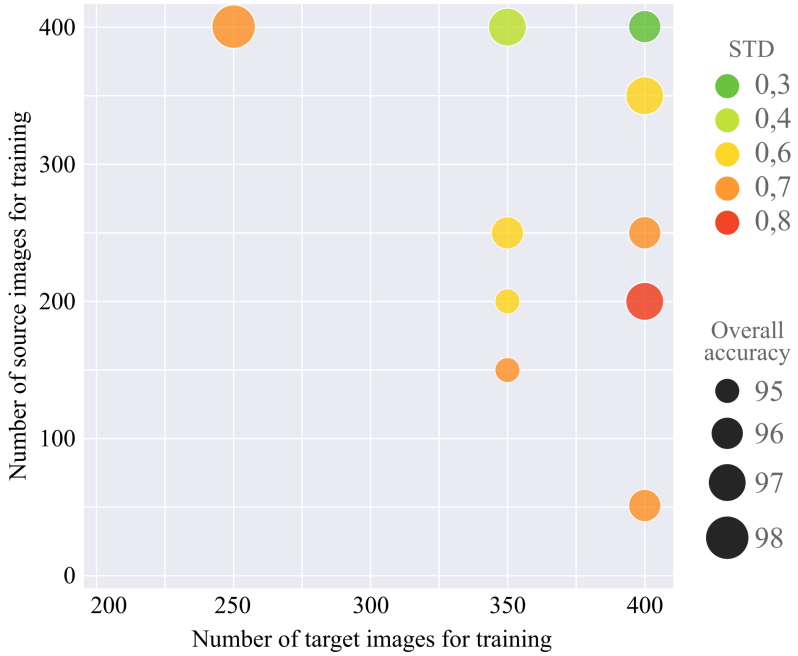


Figure 13. Ten highest overall accuracy average results achieved with the proposed method (dot size), STD (dot color), and their corresponding data combinations of target/source in the training, presented in the "x" and "y" axis, respectively.

2.2.3. Ablation Studies. Two ablation experiments were conducted to evaluate the configuration of the proposed training pipeline. The first ablation study (Ablation study 1) evalu-

ates the influence of fine-tuning training in step C of the proposed approach. The second ablation study (Ablation study 2) validates the importance of each step of the CX-DaGAN algorithm in the training procedure, evaluating the separate use of one or two of the three steps. Furthermore, in the Ablation study 2, the number of training images of the two sets, target and source, are updated simultaneously.

Ablation Study 1: In this experiment, the fine-tuning training on the Xception architecture shown in Section 2.1.3 was performed. The pre-trained weights from the ImageNet dataset were first loaded and then each block (entry, middle, and exit flow) was fine-tuned while freezing the other layers. The network was trained using Adam optimizer with a learning rate of 0.0001, a batch size of 16, a dropout of 0.2 before the decision layer, and 100 epochs. It is important to note that we employed a training duration of 100 epochs to ensure thorough model convergence and robustness, ultimately selecting the weights from the epoch with the highest performance for the final model. The mean F1-score results when fine-tuning different blocks of the Xception are shown in Fig. 14. Notice that, fine-tuning the middle flow block of the Xception architecture leads to a very similar performance compared to fine-tuning the network; hence, it is unnecessary to retrain all the Xception network. Consequently, in the following experiments, only the middle flow was fine-tuned as it provides the best results.

Ablation Study 2: For this experiment, combinations of some number of images from target/source data were selected from Table 5 to perform an ablation study presented in Table 4. Specifically, the accuracy results of the proposed method were calculated for seven target/source data combinations by eliminating one or two of the three main steps (A, B, and C) of our proposed

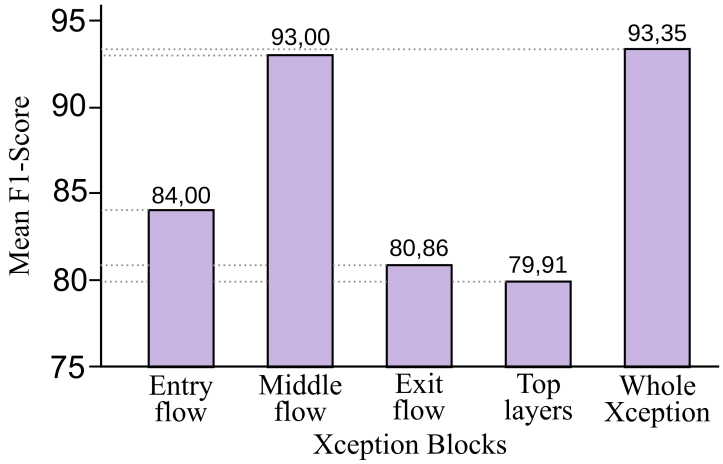


Figure 14. Mean results when fine-tuning different blocks of Xception. Fine-tuning the middle flow block of the Xception architecture leads to a very similar performance compared to fine-tuning the network.

CX-DaGAN algorithm.

The column *Step C*, which gives the worst results in Table 6, presents the classification of the test images of the small/target dataset through the Xception neural network. For this, the CNN was fine-tuned with the number of source and target images indicated in each row. In this case, source images used to complement the training dataset are randomly selected. Note that the most accurate result is obtained in this column with the largest number of images from each dataset. It should be noted that although 400 images from the target dataset are used in two cases, the highest precision is obtained when more images (350) are incorporated from the source dataset. Thus, combining a total of 750 tagged images for retraining, the accuracy was of 88.78%.

On the other hand, the results are better when we use two steps of our method. For instance, in *"Steps A + C"*, the similarity-constrained stage selects the source images to train the CNN. In this way, an increase of up to 7.91% in the average accuracy was achieved.

Then, the "*Steps B + C*" column shows the results of randomly selecting images from target and source, generating synthetic source images based on Cycle-GAN, and training the Xception. In this case, an increase of up to 4.73% compared to using only stage C was observed. Finally, the last row in Table 6 shows the result using the three steps (A + B + C). It is evident that the simultaneous combination of all the steps allows a better performance of our proposed method, compared to using a part of it, thus proving the importance of each of these steps. It is worth mentioning that the results reported in the Tables 5, 6, and 7 include the cross-validation technique, which is used to evaluate the results of statistical analysis and ensure that they are independent of the partition between training and test data. Note that, as indicated in the table titles, each result in Tables 4 and 5 included 10 realizations, and each result in Tables 6 and 7 included 30 realizations.

Table 6
Quantitative classification results of the Ablation study 1. Importance of each step in the performance of the CX-DaGAN algorithm (average over 30 realizations).

Images for Train		(↑) **Overall accuracy** (%)			
target	**source**	**Step C**	**Steps A + C**	**Steps B + C**	**Proposed method: A + B + C**
100	100	84.72±2.5	88.93±1.6	87.78±1.7	**90.15 ± 1.2**
250	100	83.78±3.6	90.86±0.4	89.09±1.1	**95,21 ± 1,1**
250	400	85.55±1.1	89.50±0.1	88.79±2.3	**97,78 ± 1,5**
350	250	82.60±1.7	94.40±1.7	89.09±1.1	**95,75 ± 0,6**
350	400	87.32±0.4	93.22±1.5	90.56±1.7	**96,75 ± 1,0**
400	150	87.91±0.4	94.69±0.7	90.27±0.7	**97,02 ± 1,8**
400	350	88.78±4.6	96.69±0.7	93.51±1.1	**96,70 ± 1,4**

2.2.4. Comparison Results. Other classification methods. In order to compare the performance of our proposed CX-DaGAN algorithm with other state-of-the-art methods, the

above reference approaches were implemented and tested:

- **TL:** Transfer learning with two chest X-rays datasets. This approach consists in considering a CNN previously trained with the ImageNet dataset and retraining it with all available source images (5216 samples) + target images (400 samples). The resulting training network was used to classify the test target set.
- **NO-S:** No source images. In this experiment, a CNN pre-trained with the ImageNet dataset is re-trained with target images (400 samples), assuming no access to a second (source) X-ray dataset.
- **RAND-S:** Random selection of source images. In this case, the CNN is retrained on a training dataset consisting in target images and randomly selected source images. This experiment aims at the contrast of increasing the training dataset randomly compared to our source image selection method.
- **SDASC:** Subspace-based Domain Adaptation using Similarity Constraints (Sanchez et al., 2021a), a recent method of augmenting a target dataset with source images to improve classification results.

In all cases, the average of over 30 realizations and the classification of 173 samples from the target dataset are reported. To ensure a fair comparison, all the methods used the same network backbone but with different optimization procedures. However, to broaden the comparison and to evaluate the consistency of our method, a discussion with other backbones is included using VGG-16, ResNet-50, and Xception networks. These results are reported in Table 7.

Table 7
Quantitative classification results of five data-based methods (including our proposed method) for three different CNNs. All methods include 400 images from the target domain for training (average over 30 realizations).

Metric	**S**	**VGG-16**			**ResNet-50**			**Xception**		
Method	**Images**	**ACC**	**F1**	**AUC**	**ACC**	**F1**	**AUC**	**ACC**	**F1**	**AUC**
TL	5216	75.05	74.15	0.74	63.72	62.48	0.62	88.36	88.47	0.87
NO-S	0	85.84	85.83	0.85	90.27	90.26	0.90	90.03	90.03	0.88
RAND-S	100	85.84	85.84	0.85	87.61	87.61	0.88	89.52	88.19	0.90
	200	81.42	81.41	0.82	90.27	90.27	0.90	89.98	89.68	0.89
SDASC	100	84.96	84.96	0.84	91.15	91.15	0.91	93.25	92.54	0.92
	200	88.50	88.50	0.88	93.69	93.48	0.93	96.18	95.96	0.95
CX-DaGAN	100	85.25	86.22	0.86	92.04	92.03	0.92	94.90	92.03	0.92
	200	**90.02**	**90.12**	**0.90**	**94.12**	**93.97**	**0.94**	**97.02**	**96.91**	**0.96**

In general, all the methods shown in Table 7 perform better using the Xception network for the classification task, except for the NO-S method. NO-S method provides the best result by using only the target dataset. Overall, the results shown in Table 7 suggest that the classification accuracy is improved by adding data from another (source) dataset. However, the samples used to increase the size of the training database should be adequately selected. In particular, the RAND-S method, consisting in selecting randomly images from the source dataset, is shown to degrade the classifier's accuracy. On the other hand, the methods SDASC and CX-DaGAN achieve the best results due to careful data selection. Indeed, the proposed CX-DaGAN method presents a significant advantage in classification accuracy over the other algorithms in terms of ACC, F1, and AUC results. It is worth highlighting that SDASC and the proposed CX-DaGAN method were designed for binary classification. Therefore, a disadvantage of the proposed method, specifically for step A, which is based on data selection through equal and crossed classes projection, is that it

can not be applied directly to a multi-class problem.

2.2.5. Visual Results. This section presents the visual results of each stage of the proposed method.

Step A: First, images from the source dataset are selected based on error metrics that account for the similarity between these images and the target domain when projected on their subspaces. To get a deeper understanding of our CX-DaGAN, we visualized some of the images selected by Step A. In Fig.15 the best and worst projected images are presented according to (1) and (2). Therefore, the figure is divided into four parts, one for each error metric. The two top rows depict the projection of each class into the same category of the target PCA subspaces. The first row presents original images from the source domain, while the second row shows the result of the projections. Specifically, we show the images with the best (a and c) and the worst (b and d) projection error for each class, normal (left) and pneumonia (right). Note that the projected image is visually more distorted when the error is larger than when the projection error is smaller. On the other hand, the two bottom rows correspond to the projection of the source images on the target subspaces with opposite classes. In this case, the error numbers are similar to those in the upper part due to the remarkable similarity between all chest radiographs, regardless of their pathology. However, our proposed method for cross-class projection involves this time, selecting the source images with the highest error projection (b and d) as shown in the pink boxes.

Step B: The source samples selected in step A were used together with the training target images (400 images) to train the GAN-based image-to-image translation proposed in Section 2.1.2. New synthetic images are generated within the target domain from the transformation of

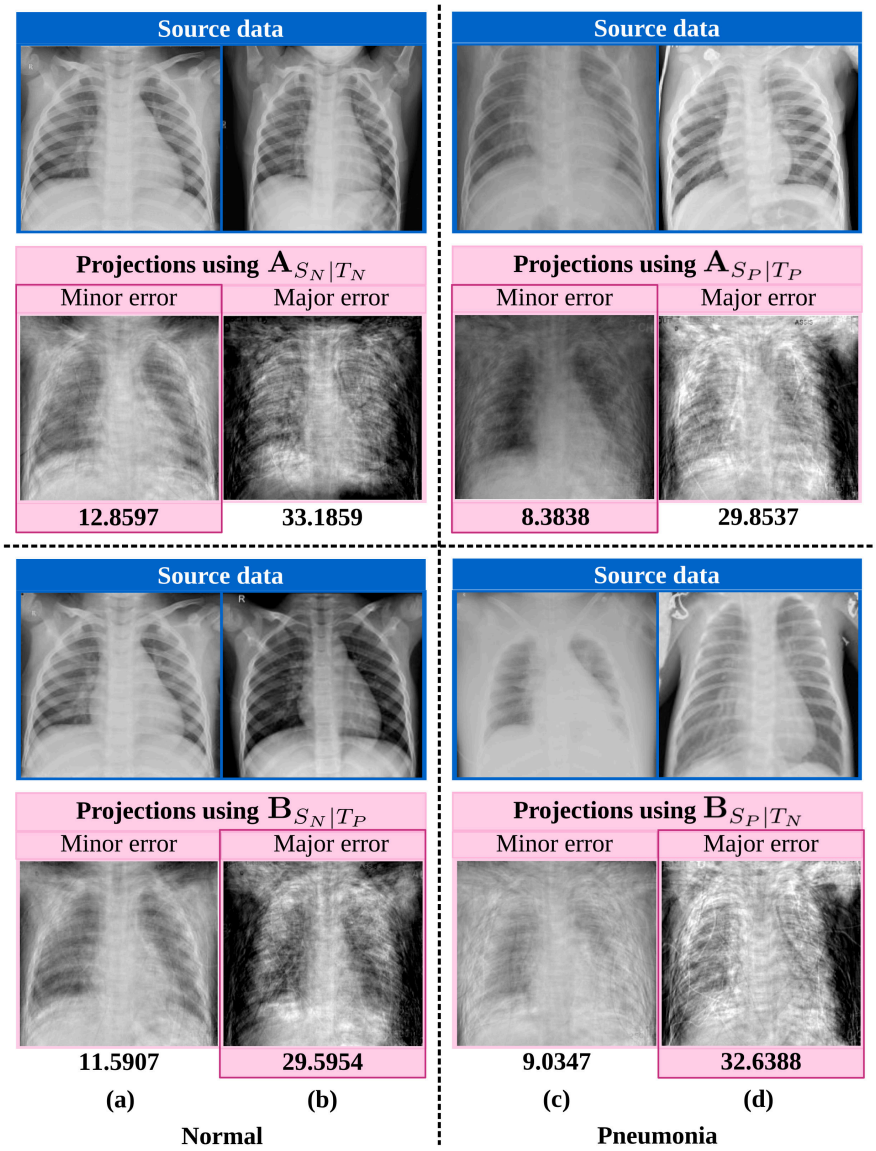


Figure 15. Selected X-ray images from the source domain considering the projection error when projected onto target subspaces obtained by PCA. The selected images are highlighted in each case with pink color. The first two rows depict the projection of each class into the same category of the target PCA subspaces. The last two rows depict the projection of the source images on the target subspaces with opposite classes.

the previously chosen source images. In such a way, the number of source images that feed the GAN network is equal to the number of output synthetic images. Fig. 16 shows four random

source images selected by step A that entered the GAN, and in the row below, their respective transformations to the target domain.

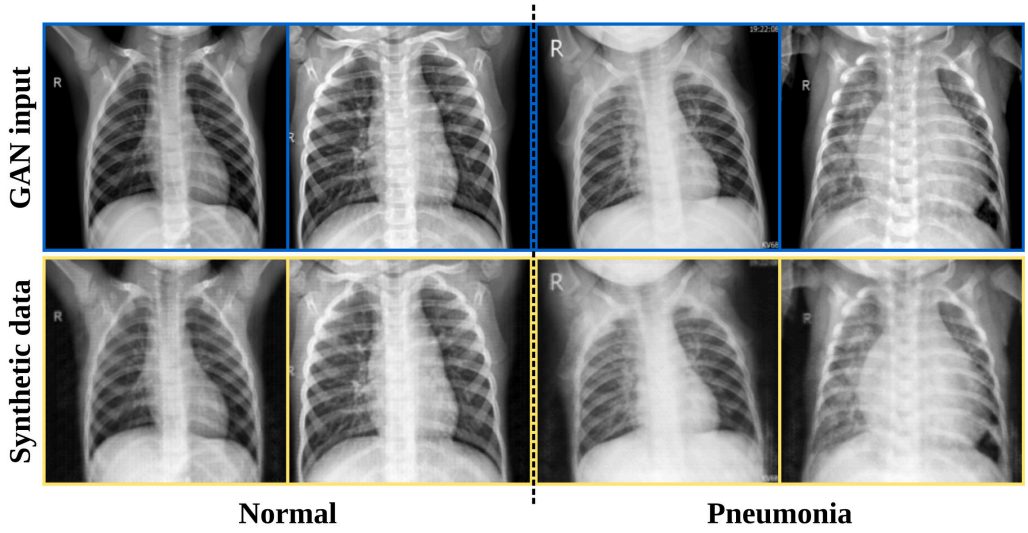


Figure 16. Input (source domain) and output (synthetic data) examples of our GAN network.

Step C: Synthetic images and the target samples were used to train the Xception network with a particular proposed fine-tuning strategy. Fig. 17 depicts some examples of the classification results obtained when testing our method with step C on 173 images of the target domain. The correct predictions are presented with black labels, and erroneous predictions are shown with a red label.

2.3. Conclusions

The main contribution of the proposed method is to take advantage of information from an extensive labeled public dataset to improve the classification accuracy of a small X-ray dataset acquired in a different hospital. Specifically, the main goal of the proposed approach is to select from a large dataset the images that best fit the small target dataset in the sense of their intra-

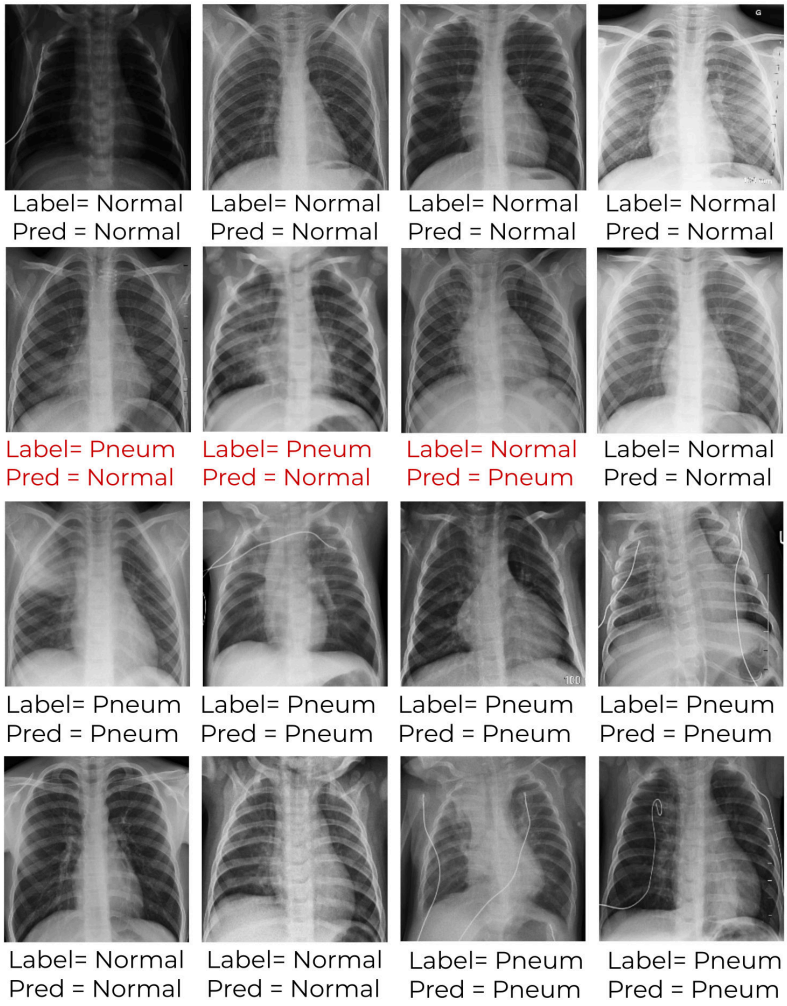


Figure 17. Classification results of some chest X-ray images from the test subset. Correctly classified samples are shown with black labels and incorrect ones with red.

class similarity and inter-class dissimilarity. In addition, a classification improvement is achieved by generating new images through a GAN network that follows the target data distribution. This chapter introduced the CX-DaGAN algorithm, an original method to address the problem of chest X-ray pneumonia diagnosis on a small target dataset. To achieve this purpose, we propose to use information extracted from a larger and publicly available chest X-ray source dataset. Specifically, our proposed algorithm is a complete domain adaptation workflow that consists of three stages.

First, we proposed a subspace-based domain adaptation method to select images from the large dataset (source domain). We then used the selected images and the train set of the small dataset (target domain) to train our proposed GAN-based image-to-image translation network. We finally used the synthetic images generated from GAN, which follow the target domain distribution, and the training set of the target dataset to fine-tune a pre-trained CNN classification network to achieve the final classification accuracy. During the experiments, we observed that training on target data without performing our proposed domain adaptation workflow led to an overall accuracy of 88.36%. However, when we used our proposed workflow to augment the training set of targets and carefully fine-tune the Xception network, we achieved an overall accuracy of up to 97.78%. Future studies will consist of evaluating the performance of this new domain adaptation method for the classification of small datasets in other related medical tasks and involving other medical imaging modalities. Furthermore, it would be interesting to address the fundamental ideas behind the CX-DaGAN algorithm to extend its scopes to multi-class classification tasks.

3. Synthetic Imaging Approach To Improve Segmentation of Liver Tumors in Multiparametric MRI

This chapter introduces a new data augmentation technique through synthetic images to address the challenge of scarcity of labeled data in medical imaging segmentation tasks. The proposed approach was recognized as one of the winners in the 2021 Data Augmentation Challenge, an event hosted by the French Society of Radiology in Paris. Some of the material featured in this section has been previously published in the international conference paper (Sanchez et al., 2023a). Section 1.9 summarises the notation used in this chapter.

Chapter contribution. Automated segmentation of liver tumors in Magnetic Resonance Images (MRI) is essential in medical image processing, supporting the detection, diagnosis, and treatment planning for liver cancer patients. However, ensuring accuracy and reliability in tumor segmentation remains a challenge due to the heterogeneity of tumor appearances in MRI and the scarcity of labeled data for training robust machine learning models in the medical imaging field. In this chapter, we present LT-SyGAN, a novel three-step methodology designed to enhance the precision of liver and tumor segmentation in Multiparametric (T1 arterial, T1 portal, and T2) Magnetic Resonance Imaging (mpMRI) through the generation of realistic-looking synthetic images. In particular, the LT-SyGAN framework produces synthetic mpMRI images with Massive Macrotrabecular Subtype Hepatocellular Carcinoma (MMHCC), along with their corresponding tumor masks, via a novel mask-guided procedure.

The proposed method first generates abdominal edges and liver tumor masks derived from

geometric transformations of a source dataset. Then, three Generative Adversarial Networks (GANs) – each assigned to a specific MRI contrast – utilize the edge masks to create synthetic data. This procedure results in the generation of the desired number of MRI triplets within liver tumors. The final stage leverages the synthetic data to train a U-net segmentation network, enabling the prediction of pixel-wise tumor and liver labels with heightened accuracy.

Extensive simulations were conducted on a private dataset of 89 real patients, curated by the French Society of Radiology for the 2021 challenge on this topic. Using the proposed method, $5,000$ synthetic MMHCC mpMRI cases, along with their segmentation masks, were generated. Additionally, our method was utilized for the segmentation of the liver and MMHCC tumor in the test set of the dataset. The diversity and fidelity of the synthetic images were assessed both qualitatively and quantitatively, utilizing the Frechet Inception Distance metric. These results were compared with outcomes from other data generation strategies and with different segmentation networks used as the backbone. The contributions of the proposed strategy are as follows:

1. This method generates diverse and realistic multiparametric (T1 arterial, T1 portal, and T2) MRIs. Essentially, it preserves the overall structure, appearance, and interrelationship between tissues within the data distribution, enhancing the authenticity of the synthetic images.

2. It also yields the corresponding tumor segmentation mask for each synthetically generated patient case. This not only aids in training machine learning models but also provides a reference point for evaluating segmentation accuracy.

3. Experimental results underline the efficacy of the generated synthetic images. They con-

tribute to increasing the accuracy of liver tumor segmentation in real MRIs across three types of contrast (T1 arterial, T1 portal, and T2), demonstrating the practical implications and utility of our proposed strategy.

3.1. Proposed Method: LT-SyGAN

This section describes LT-SyGAN, the proposed strategy for synthetic multiparametric magnetic resonance imaging of liver tumors. This method consists of three stages: (1) first, using a source dataset, it generates new tumor and abdominal edge masks; (2) in the second stage, the proposed approach leverages the use of GANs to produce synthetic triplets (T1 arterial, T1 portal, and T2) of MMHCC MRI from the input masks generated at the stage (1); (3) finally, the synthetic images are used to train a segmentation network that validates the fidelity and usefulness of the generated mpMRIs. The overall method is depicted in Fig. 18.

3.1.1. Step A: Mask Generation. This section introduces step A of the proposed method shown in Fig. 18 as "Mask Generation". The aim of this step A is to create valid anatomical borders and tumor locations.

(**i**) **Geometric transformations of tumor masks**. Let us denote by $X1_m, X2_m, X3_m$ the training data available, *i.e.*, the three MRIs available for each patient affected by MMHCC. Furthermore, let us denote the tumor segmentation masks of the training dataset as X_m, for $m = 1...M$, where M is the total number of patients in the training database. Let $s = 1...S$ be the subscript of synthetic cases generated from each patient, where S is their total number. First, each X_m is modified with multiple-random geometric operations, such as zoom, rotation, flip, and translation to generate new tumor segmentation masks denoted as $\bar{X}_{ms}$ and named "Output tumor mask" in Fig 18. In

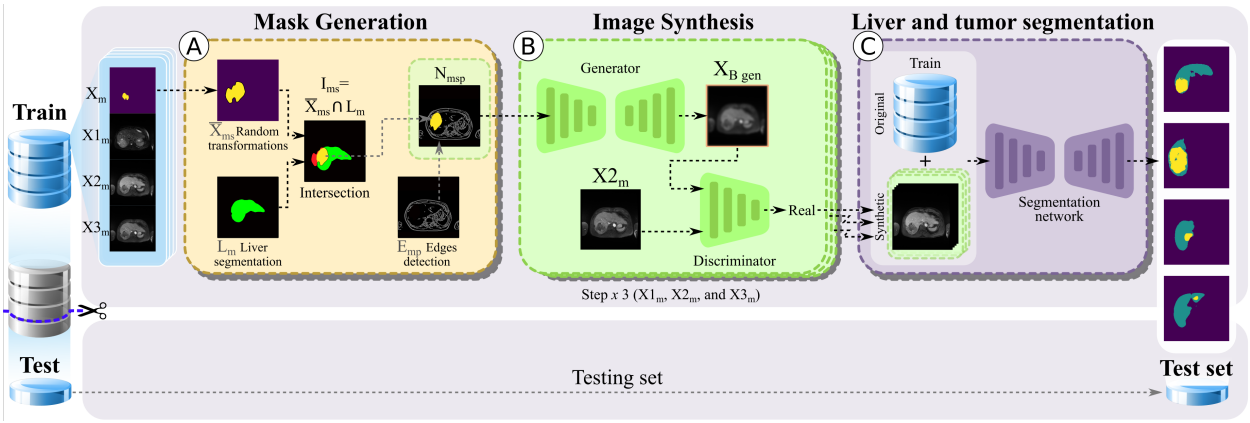


Figure 18. Proposed data augmentation and segmentation framework. First, the training set is fed into step (A) for the novel tumor mask creation. In step (A), tumor masks from the training set are geometrically transformed, intersected with liver masks, and superimposed on the computed abdominal edges of the training MRIs. Then, the new masks are used as input to the GAN-based image synthesis stage (Step B) to define the spatial distribution and liver tumor location of the triplet-generated MRIs. Finally, a pre-trained U-net segmentation network is fine-tuned for pixel-level liver and tumor classification using both the synthetic and original training images (Step C). The performance of the proposed workflow for tumor and liver segmentation is evaluated on the testing set.

this step, the *zoom* transformation for creating new masks is set within the range of minimum and maximum tumor sizes of the original dataset. The *translation* operation was spatially limited by the maximum area occupied by the centroids of the tumors in the original data. This ensured that the centroid of the synthetic tumor is inscribed in the region that can anatomically correspond to the liver of an average person, avoiding situs inversus. Note that this step provides the first condition for generating our synthetic images since the mask created at this stage determines the position of the tumor.

(**ii**) **Manual liver segmentation**. An experienced radiologist manually segmented the liver on all $X1_m, X2_m, X3_m$ MRIs using a freely available computer vision annotation tool (CVAT)(CVAT.ai Corporation, 2022). Output liver segmentation images are denoted as L_m.

(**iii**) **Algorithm for edge detection.** The proposed method uses the traditional Canny edge detector algorithm (Rong et al., 2014) to estimate all the contours inside $X1_m, X2_m$, and $X3_m$ images. In this step, the method normalizes the input image pixels between 0 and 255. Then, the Canny operator uses a multi-stage algorithm to detect a wide range of edges in the images with a randomly set of P thresholds between 30 and 120, a heuristically defined range. Each contour image, denoted by E_{mp}, will be used to delimit the distribution of organs and tissues in the synthetic images. Since the threshold to calculate the borders is variable, this method can generate multiple contour images from the same MRI, which enhances the variability of the generated synthetic data.

(**iv**) **Tumor and liver intersection mask.** This fourth step aims to ensure that the synthetic MRIs generated have a tumor located strictly inside the liver. For this purpose, the proposed method calculates an intersection image, denoted as I_s, between the tumor $\bar{X}_{ms}$ and the liver masks L_m, i.e., $I_{ms} = \bar{X}_{ms} \cap L_m$.

(**v**) **New tumor and edges mask.** The last step of the first stage overlays the tumor masks I_{ms} from the previous step on the edge detection images E_{mp}. Let us denote these new images as N_{msp}, which will be used as input for the adversarial generative network of the second stage.

3.1.2. Step B: Image Synthesis. As illustrated in the second step of Fig. 18, Step B covers synthetic image generation through an adversarial network. The data augmentation strategy proposed in this method uses the Pix2Pix (Isola et al., 2017) architecture. In this network, the output image generation is conditional on an input one, *i.e.*, image-to-image translation. Specifically, three Pix2Pix networks are configured in parallel, the three generators are trained to receive as input the same image N_{msp} resulting from the first stage, and each network produces a different

type of MRI contrast (T1-FS arterial, T1-FS portal, and T2) as output. On the other hand, the discriminator part of each Pix2Pix network is fed with both the output images of the generator (target) and the real training set (source) to determine through the loss function if the target is a plausible transformation of the source image. Therefore, a different set of training data ($X1_m$, $X2_m$, or $X3_m$) is used as a source for each Pix2Pix. The GAN that receives the $X1_m$ images can generate the synthetic $Y1_{msp}$. Similarly, networks trained with the images $X2_m$ and $X3_m$ will produce $Y2_{msp}$ and $Y3_{msp}$, respectively. Note that the N_{msp} images were designed to delimit the distribution of organs, tissues, and the location of MMHCC tumors on the synthetic images. This guarantees an adequate anatomical composition of the new MRI, which is important for our final task, and traditional GANs do not set this condition by default. The quality of the generated synthetic images was evaluated through the Frechet Inception Distance (FID) score (see Section 1.4).

3.1.3. Step C: Liver and tumor segmentation. The synthetic training dataset, generated following the proposed approach in Steps A and B, is now used to train the final computer-aided diagnosis task of this study: liver and tumor segmentation. This work examines the segmentation accuracy through a comparative analysis of various state-of-the-art networks integrated into Step C of the LT-SyGAN framework. More precisely, in the following section, we compare the results obtained from six segmentation networks: PSPNet (Zhao et al., 2017), DeepLabV3 (Chen et al., 2018), FPN (Lin et al., 2017), Linknet (Chaurasia and Culurciello, 2017), U-net++ (Zhou et al., 2018), and U-net (Ronneberger et al., 2015).

3.2. Experiments and Results

This section illustrates the efficiency of the proposed LT-SyGAN algorithm. All simulations were implemented using Python and PyTorch, and executed on an Nvidia TITAN RTX GPU with 24 GB of memory.

3.2.1. Dataset. The proposed approach was evaluated through a private dataset of 267 multiparametric MRIs corresponding to 89 patients with MMHCC, made available by the French Society of Radiology within the "GAN-based data augmentation of rare liver cancers: The SFR 2021 Artificial Intelligence Data Challenge" (Mulé et al., 2022). The objective of this challenge was to design deep learning-based methods capable of generating any number of synthetic MRI triplets ($1,000$ was the initial target of the challenge) from this available small dataset of MR images. Specifically, a tumor segmentation mask and three acquisitions (MRI contrast images) were available for each image: T1-FS arterial, T1-FS portal, and T2-weighted.

Implementation details. The Pix2Pix networks for the experiments reported in the next section were trained using the Adam solver with 200 epochs, a learning rate of 0.0001, a batch size of 24, and momentum parameters $\beta_1 = 0.5, \beta_2 = 0.9$. We utilized 200 epochs to ensure model convergence, validated through visual inspection, and indeed, for our final model, we employed the weights from the epoch with the highest performance, while the hyperparameters were heuristically defined. Before the training, all MRIs were resized to 256×256 pixels.

3.2.2. Quantitative Results. ***Assessment of generated synthetic images.*** In this subsection, the effectiveness of the proposed data augmentation framework to create realistic-

looking MRIs is evaluated quantitatively and compared to state-of-the-art methods. All the methods were trained on the dataset described previously and used to generate $1,000$ synthetic multiparametric cases of MMHCC MRIs. Table 8 presents quantitative results with four methods: a Pix2Pix network (Isola et al., 2017), a CycleGAN (Zhu et al., 2017), the proposed method, using a CycleGAN in the second stage (LT-SyGAN-CycleGAN), and the proposed method using Pix2Pix (LT-SyGAN-Pix2Pix) as described previously, in this chapter. Table 8 showcases the FID scores, a crucial metric for measuring the quality of generated MRIs, achieved by four GAN-based methods. This table comprises results for each MRI contrast, as well as an aggregate average across all generated MRIs. The FID is calculated from the distance between the training and the generated data. Therefore, 89 real and $1,000$ synthetic MRIs of each MRI type are used to calculate the results for the first three columns. The score in the last column results from the 267 real images and $3,000$ synthetic ones. The best result is shown in bold, and the second best is underlined. It is evident from the results that the LT-SyGAN proposed method outperforms the others, achieving the lowest FID scores, thus indicating superior quality synthetic MRIs. The same framework with CycleGAN instead of Pix2Pix provides the second-best results. Conversely, well-known GANs such as Pix2Pix and CycleGAN provide higher FID, thus less realistic synthetic MRIs, which effectively highlights the superiority of our proposed method in generating high-quality synthetic MRIs. Finally, another major drawback of traditional GANs compared to our LT-SyGAN framework is that they do not provide the tumor segmentation mask corresponding to the synthetic images.

Assessment of liver and tumor segmentation. To evaluate the usability of the synthetic images created with the proposed method, different segmentation networks were used. In particular,

Table 8
Quantitative evaluation of generation of 3,000 synthetic MRI, measured by FID scores, comparing two state-of-the-art methods with the proposed architecture in this study, employing CycleGAN and Pix2PixGAN.

(↓) Frechet Inception Distance score (FID)				
MRI type → **↓ Method**	**T1 Arterial**	**T1 Portal**	**T2**	**All**
Pix2Pix (Isola et al., 2017)	274.87	284.95	267.77	234.39
CycleGAN (Zhu et al., 2017)	251.12	241.27	258.33	210.32
(ours) LT-SyGAN-CycleGAN	226.76	211.32	229.22	193.12
(ours) LT-SyGAN-Pix2Pix	**115.13**	**102.47**	**123.37**	**86.55**

Table 9 presents an ablation study of the proposed LT-SyGAN method, detailing the performance of six segmentation networks as step C of the proposed framework, including PSPNet (Zhao et al., 2017), DeepLabV3 (Chen et al., 2018), U-net (Ronneberger et al., 2015), Linknet (Chaurasia and Culurciello, 2017), U-net++ (Zhou et al., 2018), and FPN (Lin et al., 2017). Each row in Table 9 corresponds to a different segmentation network, with its respective performance on both training and testing sets. Performance is measured in terms of F1-score and Intersection over Union (IOU). Furthermore, for comparison, we present the segmentation results, training the networks with two numbers of synthetic images: 15,000 and 9,000, which correspond to 5,000 and 3,000 triplets, respectively. Based on the results showcased in Table 9, Linknet emerged as the most effective segmentation network within the LT-SyGAN framework, delivering the highest F1-score and IOU results in the testing phase. Notably, these results are close to the performance of the U-net network. It is worth noting from Table 9 that superior results are achieved when a greater number of synthetic images are used. This underscores the significant role that synthetic image augmentation plays in enhancing the performance of the segmentation task.

Table 9
Quantitative tumor segmentation results with different CNNs in Step C of the LT-SyGAN proposed method.

Synthetic images	**Backbone segmentation**	**Training**		**Testing**	
		F1-Score	**IOU**	**F1-Score**	**IOU**
15,000	PSPNet (Zhao et al., 2017)	89.29	85.06	79.19	71.94
	DeepLabV3 (Chen et al., 2018)	92.83	89.23	81.53	74.47
	U-net (Ronneberger et al., 2015)	**94.93**	**92.10**	83.26	76.19
	Linknet (Chaurasia and Culurciello, 2017)	94.19	91.19	**83.38**	**76.49**
9,000	U-net++ (Zhou et al., 2018)	88.76	84.56	79.37	71.98
	DeepLabV3 (Chen et al., 2018)	88.02	83.77	80.21	73.73
	FPN (Lin et al., 2017)	90.04	87.06	81.96	**75.69**
	U-net (Ronneberger et al., 2015)	89.30	85.32	81.32	75.07
	Linknet (Chaurasia and Culurciello, 2017)	**92.28**	**88.72**	**82.89**	74.42

Table 10
Quantitative tumor segmentation results using a traditional data augmentation (DA) approach and the LT-SyGAN proposed method with two numbers of synthetic images.

Results → **↓ Method**	**Number of synthetic images**	**Accuracy segmentation** **F1 Score**	**IOU**
Traditional DA	15000	77.90	71.03
	9000	76.40	69.21
LT-SyGAN	15000	83.38	76.49
	9000	82.89	74.42

Table 10 highlights the best segmentation outcomes on the test set achieved by our proposed method when generating and employing $15,000$ and $9,000$ synthetic images, respectively. These results were derived from Table 9, specifically from the rows associated with the Linknet network. To facilitate result repeatability, we offer a thorough description of the implementation parameters applied in Step C for achieving the results presented in Table 10.

The setup of the segmentation network used Python version 3.10.9, with a batch size of 33

and a test size of 0.2. Depending on the number of synthetic images utilized, certain parameters varied: - For **15,000** synthetic images: learning rate of 0.00024425, 125 epochs, and a training duration of 1 day, 3 hours, 35 minutes, and 47 seconds. - For **9,000** synthetic images: learning rate of 0.0008752, 220 epochs, and a training duration of 19 hours, 17 minutes, and 8 seconds. Figures 19 and 20 depict the accuracy and loss curves during the training of the Linknet architecture. Specifically, Figure 19 relates to a 125-epoch training with 15,000 images, while Figure 20 corresponds to a 220-epoch training with 9,000 images (top and bottom row of Table 10, respectively).

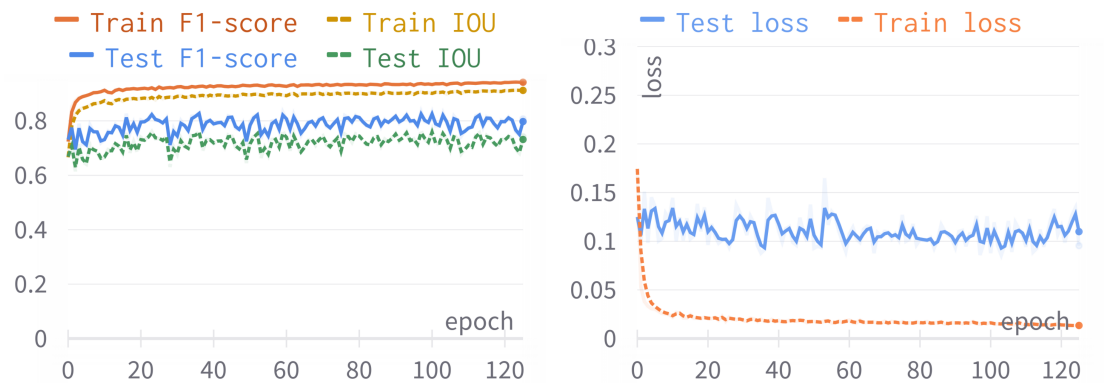


Figure 19. (Left) Accuracy curves of Train and Test, and (Right) Loss curves of Train and Test. In this experiment, the Linknet architecture was trained using 15,000 synthetic images.

Moreover, Figures 21 and 22 display the accuracy and loss curves for all architectures listed in Table 9. Specifically, Figure 21 showcases the results of architectures trained with 15,000 synthetic images, while Figure 22 demonstrates the performance of networks trained with 9,000 synthetic images. It is important to underscore that, during the training of various neural networks explored for stage C (segmentation) of the method presented in this chapter, the early stopping condition was employed, as is demonstrated in Figures 19 to 22.

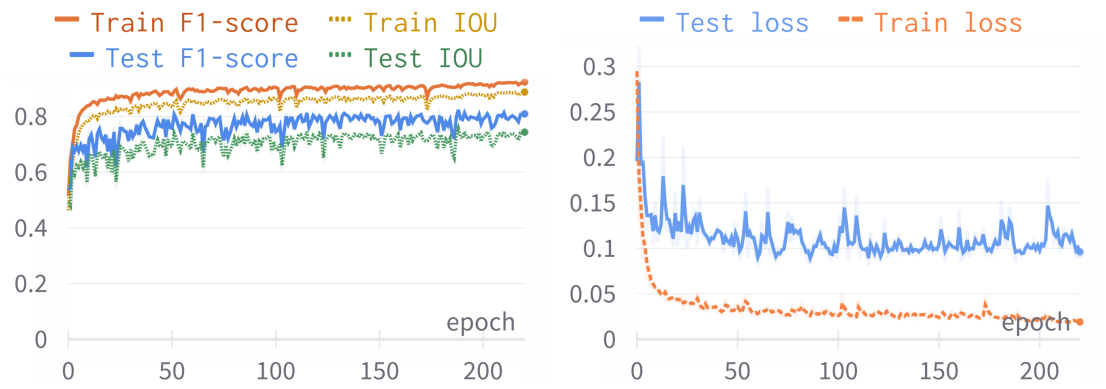


Figure 20. (Left) Accuracy curves in Train and Test, and (Right) Loss curves in Train and Test. In this experiment, the Linknet architecture was trained using 9,000 synthetic images.

3.2.3. Visual Results. This section presents the visual results of each stage of the proposed method.

Step A: First, we detail one visual result of the first step of the proposed method, which corresponds to the generation of masks through the overlapping of tumor-transformed shapes and anatomical edges, as illustrated in Figure 23. From left to right, the Figure contains a real tumor mask X_m of the original dataset, its transformation, and intersection with the liver mask I_{ms}, and the overlay of the tumor and the contour image N_{msp}. Notice that the center image in this Figure shows the result of an intersection mask I_{ms}, where the liver manual segmentation L_m is in green, the section of the tumor located inside the liver in yellow, and the tumor outside the liver in red. In the right-hand side of Fig. 23, the intersecting tumor region I_{ms} is overlaid with an edge image E_{mp} to generate a N_{msp} image.

Step B: Figure 24 illustrates four random cases of synthetic patients, one per row, generated by the LT-SyGAN proposed method. The first column refers to the tumor segmentation mask

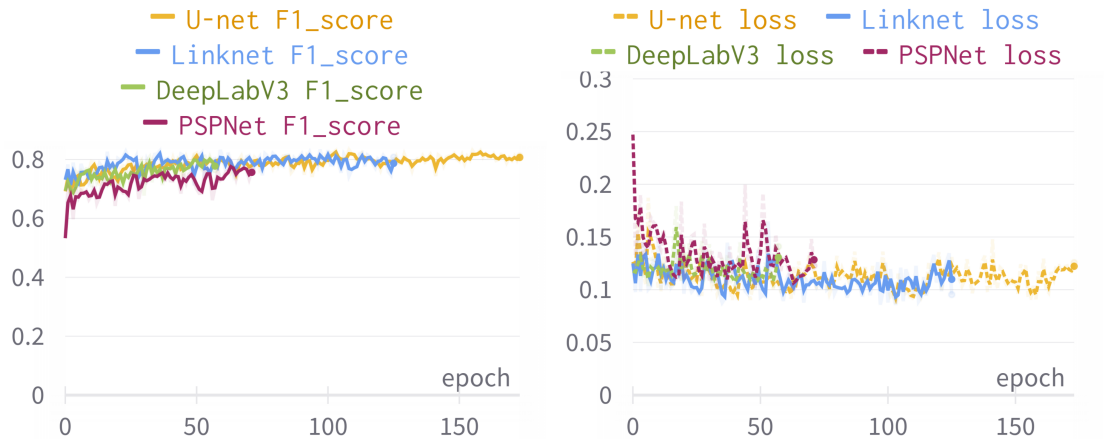


Figure 21. (Left) Accuracy curves in Test, and (Right) Loss curves using 15,000 synthetic images to train the Linknet, U-net, DeepLabV3, and PSPNet networks.

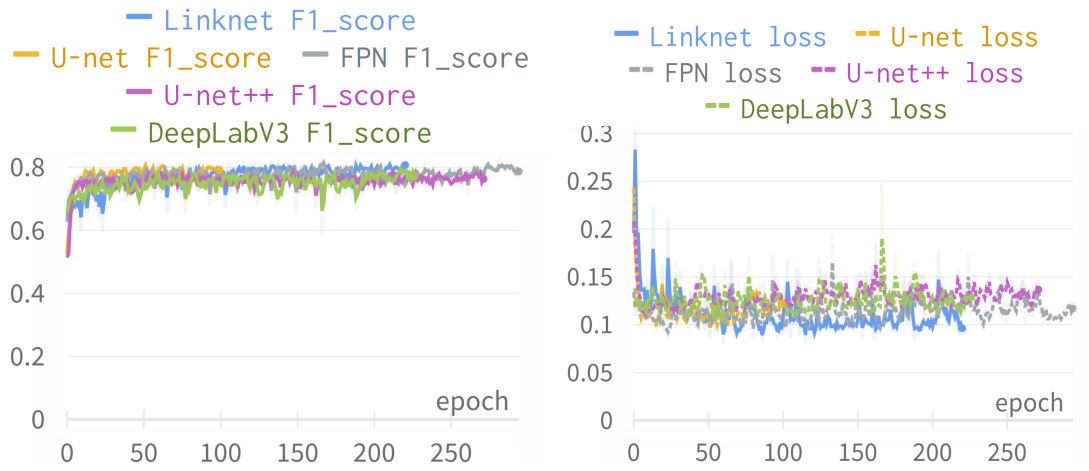


Figure 22. (Left) Accuracy curves in Test, and (Right) Loss curves using 9,000 synthetic images to train the Linknet, U-net, FPN, U-net++, and DeepLabV3 networks.

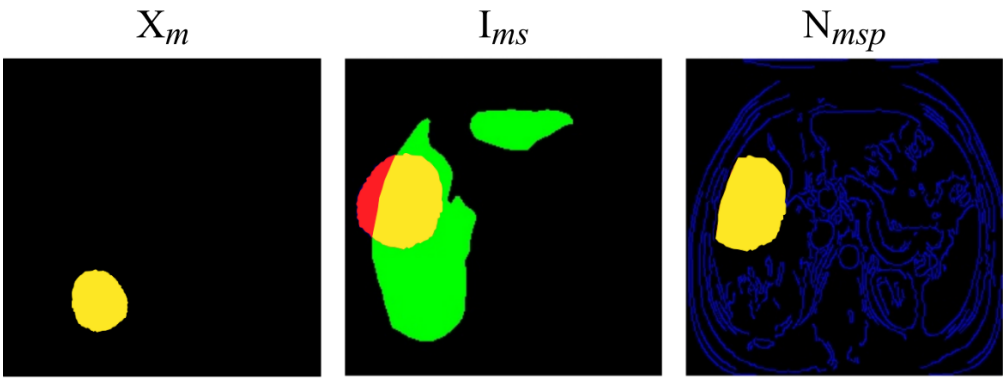


Figure 23. Visual results of Step A with one sample. The sequence, from left to right, showcases the original tumor mask from the training set, the intersection between the modified tumor and the liver mask, and the anatomical liver contour with the tumor mask.

created by our method (I_{ms}), the second column to the synthetic T1-FS arterial $Y1_{msp}$, the third corresponds to the synthetic T1-FS portal $Y2_{msp}$, and the last column to the synthetic T2-weighted $Y3_{msp}$. The three synthetic images by row are associated with the tumor mask shown in the first column. $1,000$ results of triplets of image synthetic with their respective tumor masks are available on the project website, showing the diversity and fidelity of the generated MRI. To assess the efficacy of this method, we generated up to $5,000$ triplets, as presented in Table 10. It is important to highlight that this method can produce an indefinite number of unique triplets, thanks to the diverse masks created in step A through geometric transformations of the tumor mask and adjustable edge thresholding.

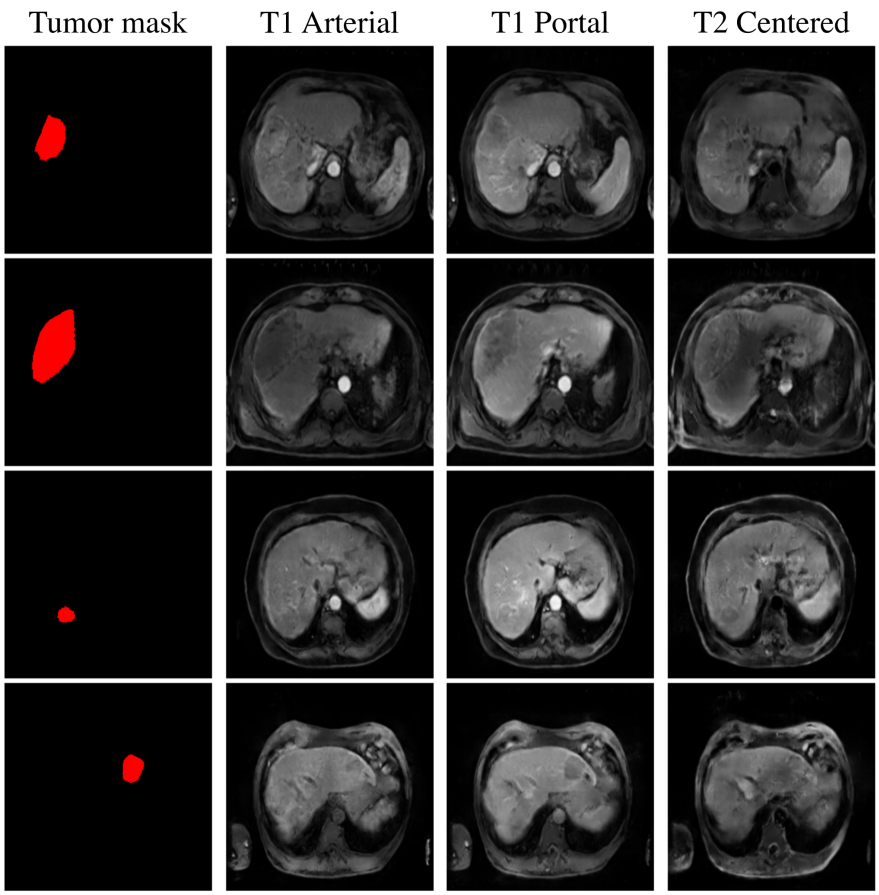


Figure 24. Visual results of Step B of the LT-SyGAN proposed method. In the figure, four triplets of synthetic images, one per row. Each case (row) corresponds to three different contrast-enhanced MRIs (arterial T1, portal T1, and T2-weighted) and their corresponding segmentation tumor mask represented in the first column.

Step C: Both synthetic images and source training samples were utilized to train the Linknet network. Figure 25 illustrates six examples of the liver (highlighted in green) and MMHCC tumor (indicated in yellow) segmentation results obtained through testing our LT-SyGAN method on a subset of 54 images from the source dataset—equivalent to 20% of the actual data. The top row presents the test images, the middle row serves as a reference by displaying the ground truth, and the bottom row shows the segmentation results achieved using LT-SyGAN.

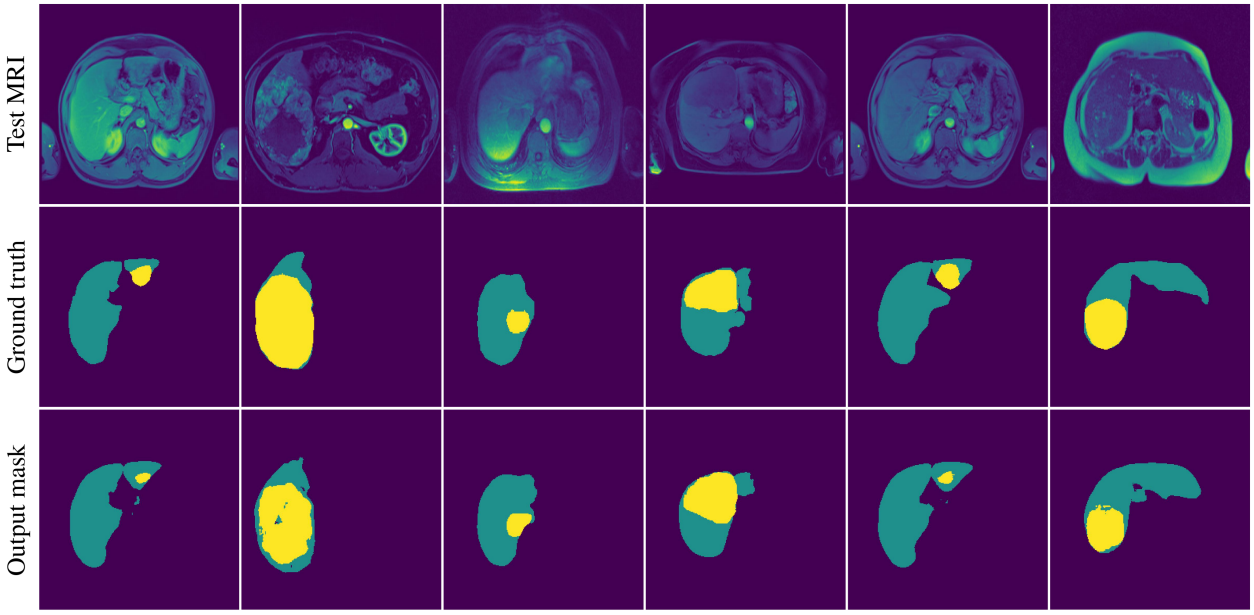


Figure 25. Visual liver and tumor segmentation results of some MRI images from the test subset using the LT-SyGAN proposed method. (Top) original MRI. (Middle) ground truth. (Bottom) segmentation result.

3.3. Conclusions

This chapter introduced an innovative framework for the generation of synthetic multiparametric MRIs, encapsulating two different sequence types (T1-FS and T2) and three different acquisitions (arterial, portal, and delayed), specific to patients with a rare liver tumor. A crucial element of our approach involved generating novel tumor masks via the intersection of liver segmentation and augmented tumor masks, which was essential in ensuring anatomically accurate tumor placements within the synthetically generated images.

From the use of new tumor masks, superimposed on real abdominal MRI edges, we initialized three Pix2Pix networks trained to create three interrelated multiparametric MRIs, analogous to a synthetic patient. Our method offers the capacity to generate an infinite number of cases, given

its ability to produce limitless segmentation masks by incorporating various thresholds for Canny-based edge detection and multiple geometric transformations of the tumor masks. The simulation results demonstrated satisfactory FID scores and visually compelling outcomes.

Crucially, the synthetic images produced by our approach were employed to train a segmentation network. It was demonstrated that these synthetic images significantly enhanced the accuracy of the segmentation process, compared to the performance when training data was not expanded. This establishes the synthetic images generated by our method as an effective tool for augmenting training data and improving performance in image segmentation tasks.

In the future, we plan to evaluate the impact of replacing the manual liver segmentation step of this chapter by using an additional MRI dataset from liver masks, which would improve the automation of the method and increase the dependency on additional data.

4. Automated Chronic Wounds Medical Assessment and Tracking Framework Based on Deep Learning

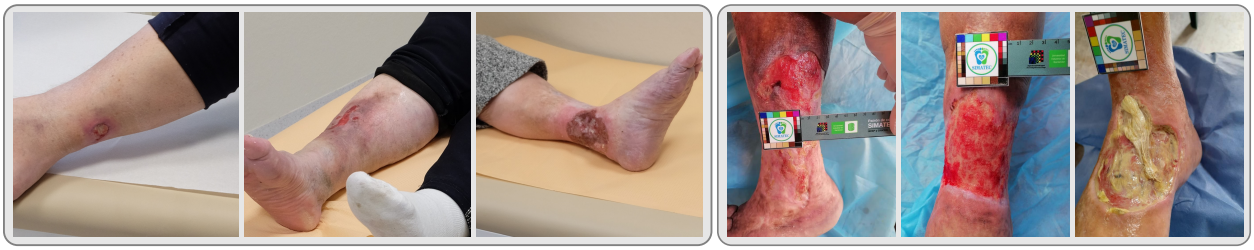
This chapter presents the first extension of this doctoral thesis, a framework to effectively acquire, detect, segment, and measure chronic wounds. Furthermore, it presents a new dataset comprising 164 RGB images from 69 Colombian leprosy patients. Each image is accompanied by its corresponding boundary coordinates and segmentation maps. The methodology proposed in this study has been integrated into a web-based platform and implemented at the *Sanatorio de Contratación ESE*, a Colombian hospital recognized for its specialization in treating Hansen's disease (leprosy) patients. Currently, the algorithm has become an integral component of the toolkit employed by healthcare professionals at that hospital in their daily management of chronic wounds.

Part of this chapter has been adapted from the conference paper (Monroy et al., 2021) and the accepted journal paper (Sanchez et al., 2023b). Section 1.9 summarises the notation used in this chapter.

Chapter contribution. Chronic wounds are a latent health problem worldwide, due to the high incidence of diseases such as diabetes and Hansen. Typically, wound evolution is tracked by medical staff through visual inspection, which becomes problematic for patients in rural areas with poor transportation and medical infrastructure. Alternatively, the design of software platforms for medical imaging applications has been increasingly prioritized.

However, since the predominant cause of chronic wounds worldwide is type 2 diabetes, there are a few public diabetes datasets, mainly from European and/or North American patients.

This means that any neural model trained on public databases cannot be directly evaluated on patients from different locations, including developing countries, due to differences in skin features such as color or wound severity. This performance is even worse for different aetiologies like leprosy, for which no public image datasets are available. Figure 26 illustrates a comparison between images from the publicly available "Chronic wounds database (CW-DB)" (Krecichwost et al., 2021), acquired in Poland from diabetes type-2 patients, and images captured in this study from leprosy patients in the Contratación town, Colombia. Differences in wound characteristics and skin type of the patients can be easily noted.



(a) (b)

Figure 26. Comparison of chronic wound images from (a) European diabetes patients (Chronic wounds public dataset (Krecichwost et al., 2021)), and (b) Colombian patients with wounds caused by leprosy.

Although several algorithms have tackled chronic wound analysis, to the best of the authors' knowledge, a computational framework for automatic wound tracking, including leprosy ulcers, has not been to date developed. Therefore, this chapter tackles this issue, employing deep learning methods to sequentially detect, segment, and provide quantitative measures of the wound caused by leprosy, allowing evolution tracking to support decision-making related to the effectiveness of the medical treatment. The proposed "CO2Dnet" framework works on RGB images captured with mobile devices like a smartphone, following a proposed protocol, which avoids the requirement

of specialized and bulky acquisition setups, facilitating access to medical monitoring of patients in rural communities from developing countries. To calculate quantitative measures of the wound, i.e., area and perimeter, a calibration pattern was specifically designed to work as a reference so that image pixels can be associated with metric units. Comprehensive tests in the collected data show that the proposed framework overcomes the results of state-of-the-art methods on the collected dataset by up to 16% in the F1-score metric. The developed framework was deployed to an online platform so that medical staff from a local hospital and their patients could benefit from it. The online-available framework enables a temporal evolution analysis of each specific wound, allowing a personalized follow-up of patients' conditions. The full implementation is available at `https://github.com/simatec-uis/CO2Dnet`. To summarize, the specific contributions of this work are as follows:

- **The CO2Dnet framework for automated chronic wound tracking.** A supervised deep learning-based framework to detect, segment, and measure the area and perimeter of chronic skin ulcers in metric units of the international system, and monitor the condition of a single wound over time on leprosy patients, using RGB images.

- **A general data acquisition protocol** using smartphone cameras, which includes a designed calibration pattern required to calculate quantitative wound measurements directly on the RGB image.

- **A chronic wounds dataset "CO2Wounds",** which contains 164 RGB images of chronic wounds acquired from 69 leprosy patients, including the calibration pattern designed by the

authors, the detection boxes of wounds and calibration patterns, and the segmentation masks of the wounds.

- **Qualitative and Quantitative Study.** Extensive cross-validation experiments were conducted to evaluate and validate each step of the framework and the global proposed model performance.

4.1. Proposed Method: CO2Dnet

The proposed CO2Dnet framework receives RGB images of chronic wounds as input, acquired with commercial smartphones, and it mainly consists of four steps: A) data acquisition, B) wound and calibration pattern detection, C) wound segmentation, and D) area and perimeter calculation.

These steps are conducted so that the region of interest that contains the wound is cropped and resized to obtain a standardized centered wound image, that is then segmented to extract the wound. The calibration pattern is fully characterized in advance so that it is possible to calculate the area and perimeter of the wound. Moreover, the estimated metrics from multiple captures at different dates enable a temporal analysis of the evolution of the chronic wound and, allow the medical staff to monitor the patient's condition as well as the efficiency of the prescribed treatment. Figure 27 depicts a general overview of the CO2Dnet proposed framework, and the key aspects are summarized in Algorithm 1. The following subsections describe the procedure conducted at each step of the framework.

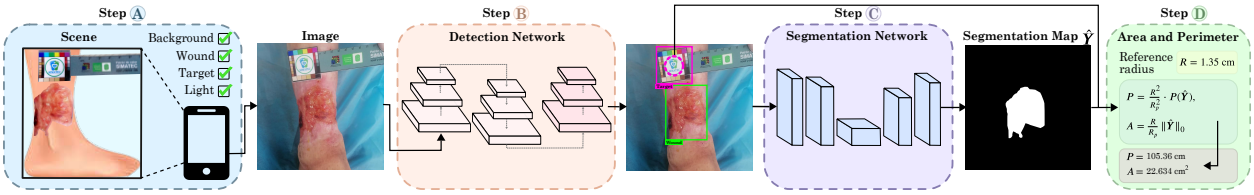


Figure 27. General scheme of the proposed CO2Dnet deep learning-based framework for automatic segmentation and measurement of chronic wounds in skin ulcers RGB images acquired with traditional smartphones.

4.1.1. Step A: Data Acquisition Protocol. The first step of the framework involves the image acquisition process, in order to build a chronic wound dataset from Colombian leprosy patients. A customized calibration pattern was designed, as illustrated in Fig. 28, which works as a calibration tool to calculate the area and perimeter of ulcers in the International System of Units. More specifically, the inner circle was designed to have radius $R = 1.35$cm, and the square side is $A = 3.5$cm. These figures are used in the detection and metrics calculation processes, as will be detailed in Sections 4.1.2 and 4.1.4 (Steps B and D, respectively). Additional characteristics of the calibration pattern, i.e., colored squares were carefully selected, such that the detection network can easily detect the pattern, due to color contrast. Further, colored squares are also intended for color calibration purposes, that account for different smartphone camera responses. This process, however, is relegated to future work. The top row includes the RGB and CMYK colors. The bottom row contains six colors corresponding to the Fitzpatrick Skin Color Scale (Sachdeva et al., 2009). This scale categorizes a person's skin according to their complexion, hair color, propensity to tan, and tolerance to sunlight, so-called skin phototypes. The squares of the two left columns and the inner right column correspond to 22 colors specially defined and selected to match ulcer tissue

representation as in (Wannous et al., 2012). Finally, the outer right column contains a grayscale, traditionally used in pattern systems to calibrate the white and black sensor responses. It is worth recalling that the acquisition of this dataset does not require a specialized setup, as this study is intended to provide wound monitoring to patients in communities with insufficient medical assistance. Therefore, images are acquired with smartphones, and they should include the calibration pattern, located next to the wound. In addition, the following instructions were provided to health personnel to complement the proposed acquisition protocol, so as to obtain good quality images: i) the wound area to be photographed should be horizontally arranged, preferably on a background of homogeneous color; ii) illumination should face the wound, and natural light is preferred; iii) the calibration pattern should be located close to the wound, without covering it; iv) position the smartphone camera parallel to (in front of) the wound; v) the complete wound and the calibration pattern must be visible on the smartphone screen, without shadows over the wound or objects in the background. Once all these conditions are satisfied, the image can be captured.

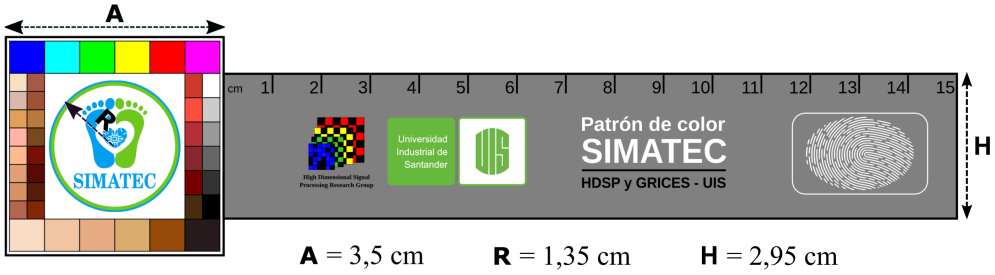


Figure 28. Calibration pattern to be used in Step A during the image acquisition and then in the calculation of measurements in Stage D.

4.1.2. Step B: Wound and Calibration Pattern Detection. Captured wound images usually contain unnecessary information around the region of interest, i.e., the actual wound.

This implies a class imbalance scenario, where a high percentage of image pixels correspond to the background, while a small percentage actually belongs to the wound (Johnson and Khoshgoftaar, 2019). Previous work has shown that working only with the region of interest improves the performance of segmentation models (Monroy et al., 2021). Therefore, this work employs a detection model YoloV4 (Bochkovskiy et al., 2020) to locate and crop the wound region from the images, aiming to mitigate the impact of unnecessary information and improve segmentation results. Specifically, the YoloV4 network is trained as a detection model, with the purpose of detecting not only the chronic wound but also the calibration pattern. It is worth noting that the pattern information is critical in the proposed framework because it works as a size reference object to estimate the area and perimeter of the wounds.

The YoloV4 detection network receives the RGB images acquired in Step A as input, and provides the bounding boxes of the detected regions of interest, i.e., the wound and the calibration pattern, as illustrated by the green and magenta squares in Fig. 27. The detection YoloV4 architecture is composed of three sub-networks: a CSPDarknet53 (Wang et al., 2020b) as backbone, which is pre-trained on the ImageNet dataset; an SPP module (He et al., 2015) and a PANet path aggregation (Liu et al., 2018) are used to collect feature maps from different hidden layers of the backbone network; and a YoloV3 (Redmon and Farhadi, 2018) as head, which is used to predict classes and bounding boxes of objects.

The process can be mathematically modeled as follows. Let $\mathscr{X} \in \mathscr{R}^{N\times M\times 3}$ be the input RGB image with $N \times M$ pixels, and $\mathbf{c} = [i_1, j_1, i_2, j_2]$ is the boundary box provided by the YoloV4 network, surrounding the detected object, where (i_1, j_1) and (i_2, j_2) are the top-left and bottom-

right coordinates of the box, respectively. Since, in this work, the detection network $\boldsymbol{D}_{\theta}(\cdot)$ with parameters θ are adjusted to detect two objects: the wound and the calibration pattern, the current output of the YoloV4 network is

$$\mathbf{c}_w, \mathbf{c}_p = \mathscr{D}_{\theta}(\mathscr{X}), \tag{16}$$

where $\mathbf{c}_w$ and $\mathbf{c}_p$ are the coordinates of the boundary boxes for the wound region and calibration pattern, respectively. Then, the boundary boxes are used to crop the image $\mathscr{X}$, so as to obtain two independent sub-images: the wound $\mathscr{X}_w \in \mathbb{R}^{H\times W\times 3}$ and the calibration pattern $\mathscr{X}_p \in \mathbb{R}^{H\times W\times 3}$ as

$$\mathscr{X}_w = \text{CropResize}(\mathscr{X}, \mathbf{c}_w), \tag{17a}$$

$$\mathscr{X}_p = \text{Crop}(\mathscr{X}, \mathbf{c}_p), \tag{17b}$$

where CropResize$(\cdot)$ represents the operator that crops an image according to the bounding box coordinates, and provides a resized output of $H \times W$ pixels, with $H = W = 320$ as input parameters. Further, considering that task-relevant information varies for different image resolutions (Van Noord and Postma, 2017), the size parameters $H \times W$ for which the wound image is resized must match the size of the images used in the segmentation model training.

4.1.3. Step C: Wound Segmentation. The segmentation step can be seen as a classification task at the pixel level of the wound image $\mathscr{X}_w \in \mathbb{R}^{H\times W\times 3}$, where each pixel in the image is associated with a class (wound/background), yielding a binary map of the wound region. Given a segmentation network $\boldsymbol{S}_{\phi}(\cdot)$ and its networks parameters ϕ, the estimated segmentation

map can be mathematically modeled as

$$\hat{\boldsymbol{Y}} = \boldsymbol{S}_{\phi}(\mathscr{X}_w), \tag{18}$$

where $\hat{\boldsymbol{Y}} \in \{0,1\}^{H \times W}$ is the estimated segmentation map, with one value indicating that a pixel corresponds to the wound region, and zero otherwise. Based on our previous analysis (Monroy et al., 2021), the segmentation model used in this work follows a U-net architecture. The U-net is divided into three parts: 1) a max-pooling path composed of down-scaling blocks, 2) a bottleneck composed of convolutional layers, and 3) an up-sampling path composed of up-scaling blocks. As illustrated in Fig. 29, a U-net is composed of L levels of pairs of down-scaling and up-scaling blocks, with each pair connected by a skip connection. For the network configuration, we set the number of levels to $L = 5$ with 24 features for the first level, and a multiplier factor of 2 in each level, i.e., 48 features for the second level, 96 features for the third level, etc. Finally, a convolution layer with a Sigmoid activation function is used to obtain the binary segmentation map.

4.1.4. Step D: Calculation of Wound Area and Perimeter. At the request of the medical staff involved in this project, it was decided to quantify two key variables for wound assessment: area and perimeter. These measurements are routinely used to document the current condition of wounds, enabling the monitoring of their progression or regression over time, in accordance with established medical procedures. For this aim, the estimated binary segmentation map $\hat{\boldsymbol{Y}}$ of the wound region and the calibration pattern region $\mathscr{X}_p$ are used. This process requires the conversion factor from image pixels to centimeters, so as to estimate the wound area (A_w)

and perimeter (P_w). Since the dimensions of the calibration pattern are known a priori (See Fig. 28), as well as the input image size, the relation between camera pixels and centimeters can be estimated via cross-multiplication. Specifically, we first employ a circle detection algorithm to measure (in pixels) the radius R_p of the circle in the detected calibration pattern $\mathscr{X}_p$, as $R_p = \text{calculateRadius}(\mathscr{X}_p)$. Then, since the actual radius of the circle is known, i.e., $R = 1.35$ cm, the conversion factor can be calculated as $C_f = \frac{R}{R_p}$. Thus, the conversion factor is used to calculate the area A_w and perimeter P_w of the wounds from the segmented images $\hat{\mathbf{Y}}$ as

$$A_w = C_f \|\hat{\boldsymbol{Y}}\|_0, \quad P_w = C_f^2 \cdot P(\hat{\boldsymbol{Y}}), \tag{19}$$

where, $\|\cdot\|_0$ is the ℓ_0 norm, namely, the amount of non-zero values, and $P(\cdot)$ denotes the contour perimeter of the binary segmentation map. To compute the contour perimeter of a closed shape, we employ the arcLength$(\cdot)$ function included in the OpenCV library. The input parameters

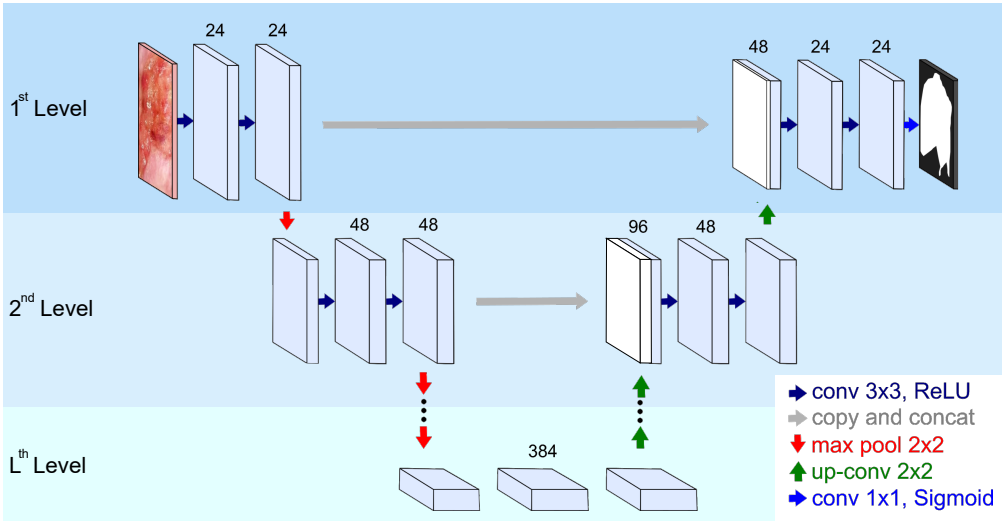


Figure 29. Illustration of the U-net architecture used for the wound segmentation task in Step C.

of arcLength$(\cdot)$ function are an input vector of 2D points that describes the shape to analyze and a flag indicating whether the shape is closed or not.

Algorithm 1 Chronic Wounds Analysis Algorithm

Require: $\mathscr{X}$ ▷ RGB input image
Ensure: $\hat{\boldsymbol{Y}}$

1: $\mathbf{c}_w, \mathbf{c}_p = \boldsymbol{D}_\theta(\mathscr{X})$ ▷ Wound and calibration pattern detection with YoloV4
2: $\mathscr{X}_w = \text{CropResize}(\mathscr{X}, \mathbf{c}_w)$ ▷ Crop wound region
3: $\mathscr{X}_p = \text{Crop}(\mathscr{X}, \mathbf{c}_p)$ ▷ Crop calibration pattern region
4: $\hat{\boldsymbol{Y}} = \boldsymbol{S}_\phi(\mathscr{X}_w)$ ▷ Wound segmentation
5: R_p = calculateRadius$(\mathscr{X}_p)$ ▷ Measure circle radius in pixels
6: $C_f = \frac{R}{R_p}$ ▷ Compute pixels to cm conversion ratio
7: $A_w = C_f \|\hat{\boldsymbol{Y}}\|_0, \quad P_w = C_f^2 \cdot P(\hat{\boldsymbol{Y}})$ ▷ Calculate wound area and perimeter

Training Procedure and Data Augmentation. Network training involves the adjustment of parameters so that the error between network predictions and original labels is reduced. Thus, a successful implementation is obtained for a particular task. During the training procedure of a specific network $\mathscr{M}_\alpha(\cdot)$, the error between the model predictions for input data x, $\hat{y} = \mathscr{M}_\alpha(x)$ and the ground-truth labels y is measured given an objective loss function $\mathscr{L}_{task}(\cdot)$, accordingly selected to the desired task. Thus, network parameters are iteratively updated during training so that the loss function is minimized. The training process is conducted until optimal performance is achieved or a stopping criterion is reached. In general, the training procedure consists of solving the following optimization problem

$$\alpha^* \in \arg\min_{\alpha} \quad \mathbb{E}[\mathscr{L}_{task}(\mathscr{M}_\alpha(x^{(i)}), y^{(i)})], \tag{20}$$

where the expected value $\mathbb{E}[\cdot]$ is calculated for the training data set that contains labeled

image pairs, i.e., $\{x^{(i)}, y^{(i)}\}_{i=1}^{N}$. Specifically, for the training procedure of step B (Section 4.1.2), corresponding to the YOLO network, the parameters are initialized from pre-trained values using the ImageNet database, and the parameters configuration suggested by the authors in (Bochkovskiy et al., 2020). This strategy is known as transfer learning (TL). Mathematically, the training of the YOLO detection model $\mathscr{D}_\theta$ is described as follows

$$\theta^* = \arg\min_{\theta} \quad \mathbb{E}[\mathscr{L}_{yolo}(\mathscr{D}_\theta(\mathscr{X}), \mathbf{c}_w, \mathbf{c}_p)], \tag{21}$$

where $\mathscr{L}_{yolo}$ is the detection loss function (YOLO MSE-based loss) proposed by authors in (Redmon and Farhadi, 2018). Analogously, the training of the U-net segmentation model $\mathscr{S}_\phi$ in step C, with binary cross-entropy loss is described as follows

$$\phi^* = \arg\min_{\phi} \quad \mathbb{E}[\mathscr{L}_{BC}(\mathscr{S}_\phi(\mathscr{X}_w), \boldsymbol{Y})]. \tag{22}$$

A data augmentation (DA) scheme was implemented for both the detection and segmentation models to add robustness to the networks against image rotations and close-ups, caused by different camera positions across acquisitions. For the detection model, random values for saturation and exposure were used following the default configuration provided by the Darknet framework (Bochkovskiy et al., 2020; Wang et al., 2021). Conversely, for the segmentation model, the data augmentation scheme accounts for varying distances between the wound and the camera. To this end, the original data set was randomly cropped following the detection box labels, simulating zoom alterations in the images, comparable to different acquisition distances.

4.2. Experiments and Results

This section illustrates the performance of the CO2Dnet proposed framework for segmentation and tracking of chronic wound images on an RGB data set acquired from leprosy patients. All simulations were implemented in Python with Tensorflow 2.3 and ran on an Nvidia Quadro Tesla T4 GPU with 16 GB of memory. Image data sets and evaluation metrics employed in these experiments are described below.

4.2.1. Datasets. The performance of the proposed framework was evaluated using two chronic wounds datasets: first, the new "CO2Wounds" dataset constructed, manually labeled, and introduced in this chapter; second, the public database "Chronic wounds" (CW-DB) (Krecichwost et al., 2021). Specifically, the CO2Wounds, whose images contain the designed calibration pattern, were used to evaluate all steps of the proposed framework, i.e., detection, segmentation, and area and perimeter estimation. While the CW-DB was used to evaluate the segmentation step. In the following, a detailed description of both datasets is provided.

The ***new CO2Wounds dataset*** contains 164 chronic wound RGB images from leprosy patients, compiled in this study, following the acquisition process described in Step A of the proposed framework (Section 4.1.1). The images were acquired by the Leprosy program control from the Sanatorio de Contratación in Colombia. Images from 69 consenting patients were collected by medical staff for 7 months (November 2021 to June 2022), using different smartphone references and the provided calibration pattern from Fig. 28. The dataset includes 164 RGB images with

their corresponding binary segmentation maps. The public dataset is available[2] and it is being constantly updated.

The ***Chronic wounds database*** (CW-DB) contains images of diabetes patients from Poland, including four different modalities: color (RGB) photography, thermovision, stereovision, and depth perception, recorded along the same axis. The binary segmentation maps of the RGB ulcer images are also provided by the authors. Moreover, the binary ulcer segmentation maps are delineated by an experienced surgeon providing a ground truth for image segmentation algorithms. In total, 737 images of 47 patients are available. However, this work only employed the 188 RGB images and their corresponding segmentation maps.

4.2.2. Quantitative Segmentation Results. A cross-dataset validation experiment was carried out to verify the overfitting domain assumption in the segmentation model from Step C (Section 4.1.3). Considering that the two datasets (CW-DB and CO2Wounds) are available to test the proposed segmentation model, this experiment evaluates its accuracy for the four possible training and validation data combinations as indicated in Table 11. Here, a 10-fold configuration was employed, where the original data set was randomly partitioned into ten equal-sized sub-datasets so that a single sub-dataset is used for testing and the others for training. The framework performance was evaluated using four metrics: precision, recall, Intersection over union (IoU), and F1-score, calculated given the ground truth and the predicted segmentation masks $\boldsymbol{Y}$ and $\hat{\boldsymbol{Y}}$, respectively.

---

[2] CO2Wounds database: `https://doi.org/10.17632/nkw5gx57hw.1`

Thus, a 10-fold configuration was employed in the cross-validation process, where the original data set was randomly partitioned into ten equal-sized sub-datasets so that a single sub-dataset is used for testing and the others for training. The model was trained from scratch with the corresponding dataset at each run. The results in Table 11 show that, as expected, the proposed framework obtains a suitable performance when training and validation data belong to the same data set, reaching an F1-score value of 87.2% when CW-DB was used, and 64.3% in the case for CO2Wounds. Otherwise, poor metrics were obtained as a consequence of domain overfitting.

Further, it should be noted that the highest metrics were obtained when CW-DB was used for training and validation, because of the homogeneity of the images in this data set, as it was acquired under controlled illumination conditions and a fixed camera setup. In contrast, CO2Wounds is a more heterogeneous data set, containing images acquired with several smartphone cameras under different illumination conditions in uncontrolled environments. Nonetheless, these characteristics should not be considered shortcomings since they are inherent to the nature of the proposed framework, intended to be accessible for hard-to-reach communities with limited medical infrastructure. Moreover, the experiments in the subsequent sections will show that fully employing the proposed framework improves the results of just using the segmentation model.

To illustrate the cross-dataset results, Fig. 30 presents a qualitative and quantitative comparison of the segmentation maps of four different images from the CO2Wounds data set. The first column contains the original input image of the wound; the second column shows the ground truth of the binary segmentation map; the third column illustrates the results obtained from training with CW-DB and evaluation with CO2wounds; and the fourth column depicts the results of training and

evaluation with the CO2Wounds data set. Wounds at each result image are mapped to 4 colors: green representing the true positives, blue the false positives, red the false negatives, and black the true negatives. The F1 Score, precision, and recall are indicated in the black, dark gray, and light gray rectangles. Note that the fourth column exhibits more true positive (green) and true negative (black) pixels, with fewer false negative (red) and false positive (blue) pixels, in comparison to the results in the third column. Also, the CO2W/CO2W column results are closer to the ground truth images, with better F1-score, precision, and recall metrics. Conversely, the results of evaluating the CO2W data set on a model trained on CW-DB show demonstrate the impact of domain overfitting.

Table 11
Cross-dataset validation results. Metrics are the result of 10-fold cross-validation.

Dataset		(↑) **Metrics** (mean±std)		
Training	**Validation**	**F1-Score**	**Precision**	**Recall**
CW-DB	CW-DB	87.2±1.43	89.3±1.42	88.6±1.33
CW-DB	CO2Wounds	43.6±4.34	79.8±8.75	34.0±5.60
CO2wounds	CW-DB	21.7±5.44	93.2±4.07	12.7±3.33
CO2wounds	CO2wounds	64.3±5.92	82.7±7.02	66.7±8.45

Estimation Error for Area and Perimeter. This section provides a general error estimation for calculating the wound area and perimeter in units of the international metric system using the Step D of the CO2Dnet framework. It should be pointed out that real wound area and perimeter measurements are not available, because manually measuring chronic wounds typically yields erratic values. These errors are related to the difficulties of appropriately locating a metric instrument close to the wound. Therefore, the error evaluation performed in this work consists of digitally crafted binary masks of a known object, i.e., rectangles in the calibration pattern, and uses

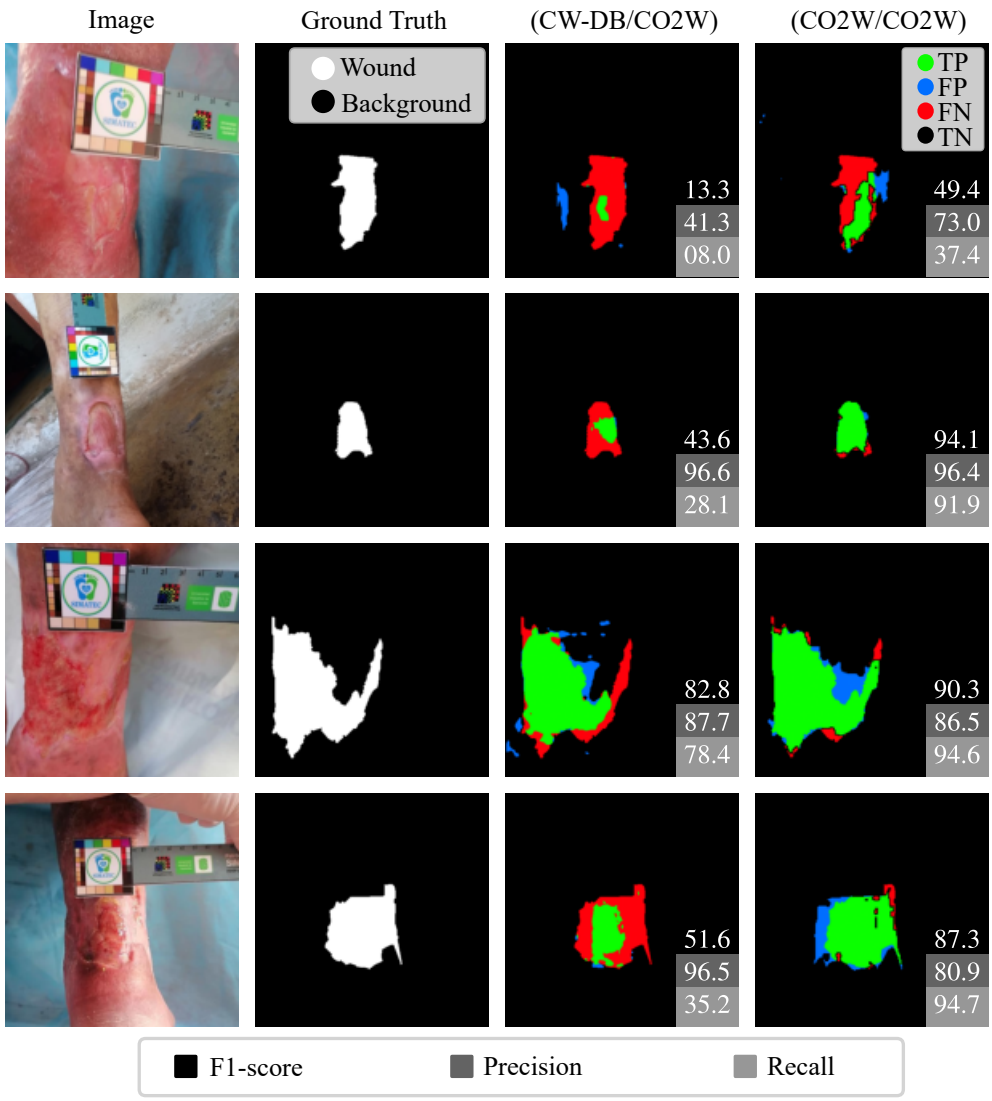


Figure 30. Segmentation results of example images from the test subset in the CO2Wounds (CO2W) database. (First column) Input image; (Second column) Segmentation ground-truth maps; (Third-Fourth columns) Segmentations from the (train/test) dataset evaluation case. Color conventions in the top-right indicate True positives (TP), False positives (FP), False negatives (FN), and True negatives (TN) pixels.

the resulting binary masks and photos as inputs for Step D . In this evaluation, we are working with regular figures, allowing us to calculate the error by comparing the area and perimeter values obtained through Step D with the actual known measurements of this object, described in Fig. 28.

Specifically, five images of the calibration pattern were acquired with five different smartphone cameras. Then, the CO2Dnet framework was used to estimate the area and perimeter of fifteen elements within the pattern. Specifically, the ruler with the metric units included in the calibration pattern was used to define fifteen rectangular pieces of different sizes on each image, as follows. Each of the n rectangles has a width of $1+n$ cm, with $n = \{0, \cdots, 14\}$, along the grey ruler of the pattern, and a fixed height 2.95 cm. Figure 31(a-e) shows the five images of the calibration pattern used in this procedure, and Figure 31(1-15) illustrates the rectangular elements of Fig. 31(b) used for error calculation.

For each image, the area and perimeter of the 15 rectangles were calculated using Equation 19, where C_f was automatically obtained by CO2Dnet using the area and perimeter in pixels of the inner circle from the calibration pattern detected at each image (Fig. 31(a)). The mean squared error (MSE) between the real and estimated area, and perimeter of the fifteen pieces was calculated for each of the five images. The average error results along with the standard deviation are shown in Fig. 32.

Note that these numerical results show that the error in the area estimation is inversely proportional to the element size, i.e., smaller pieces are more prone to larger errors in the area estimation. Numerically, area relative errors vary between 7% and 11%, reaching a maximum of 15.6%. On the other hand, the error in the estimation of the perimeter is smaller, more stable, and slightly independent of the object size, as observed in Fig. 32 (Right). Specifically, this error varies between 4.6% and 6.5%, reaching a maximum of 8%. Consequently, these results show that area and perimeter calculations are useful for quantitative wound tracking, providing more confident

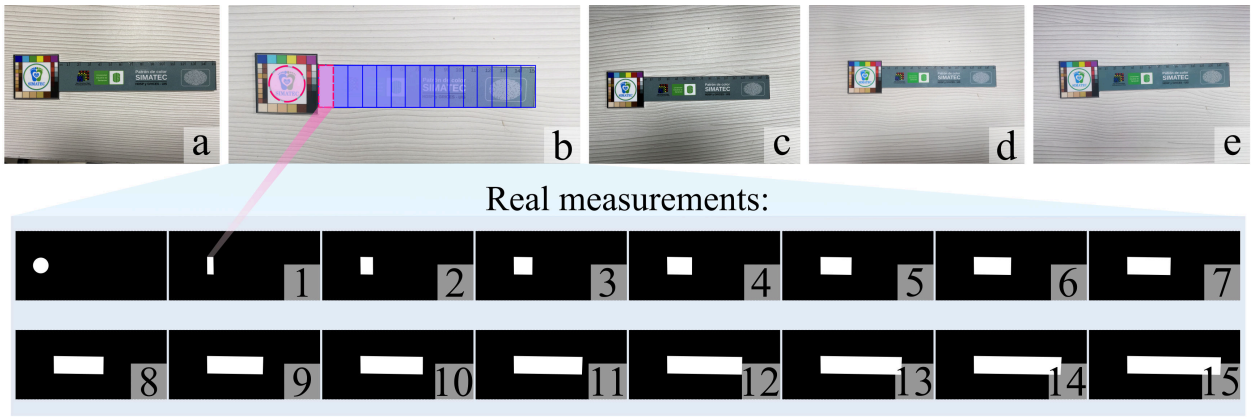


Figure 31. (Top) Images of the calibration pattern acquired with five different smartphones. (b) Rectangular pieces of different known sizes are highlighted in blue. (Bottom) Ground truth of the circle in the calibration pattern, and (1-15) Fifteen rectangles of known size along the ruler, used to determine the estimation error of area and perimeter calculation on each image. Rectangles are 2.95 cm tall and $n+1$ cm wide, with $n = \{0, \cdots, 14\}$.

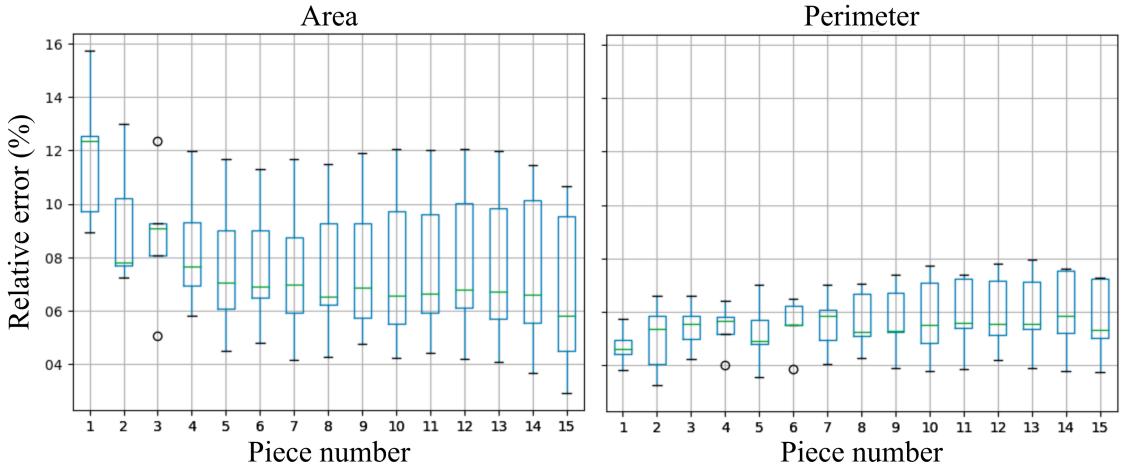


Figure 32. Average MSE along with standard deviation (Left) area and (Right) perimeter calculation using Step D (Section 4.1.4) of CO2Dnet framework. The method was evaluated in fifteen pieces of known size in the calibration pattern through images acquired with five different smartphone cameras.

results for larger wounds that, in turn, represent medically critical scenarios.

4.2.3. Ablation Study. In this section, an ablation study was conducted to assess the contribution of the detection step (Step B in Section 4.1.2) to the performance of the overall segmentation process. In addition, the impact of typical deep learning strategies on the segmentation neural network was also evaluated. Specifically, we consider the contribution of transfer learning (TL) and data augmentation (DA) strategies to the overall pipeline, as explained below:

i) Transfer learning (TL): This strategy, commonly used to improve the performance of neural networks, avoids training a model from scratch by initializing its weights with those from a pre-trained network using a different data set. In this work, the segmentation network was initially trained with the CW-DB data set, so this pre-trained model was afterward retrained with the CO2Wounds images.

ii) Detection (D) step (Sec. 4.1.2): enables image cropping and calibration pattern detection for subsequent use in the area and perimeter calculation step. Here, the contribution of this step is evaluated by comparing the obtained results when detection is employed with respect to those from the segmentation without detection, i.e., the full image.

iii) Data augmentation (DA) scheme: This strategy was included to consider image rotations and close-ups, caused by image cropping performed by the YOLOV4 detection model. Essentially, DA aims at preserving performance and reducing over-fitting.

This ablation study consists in evaluating the segmentation performance using different combinations of the evaluated processes in the network training, i.e., TL, D, and DA, as indicated in Table 12. It is worth noting that these processes should follow a particular sequence since, for

instance, the data-augmentation stage requires the results of the detection stage (first detecting the calibration pattern). For this reason, the data-augmentation without detection scenario was not considered. The results are summarized in Table 12, in terms of F1-score, Precision, and Recall, for which higher values indicate better results. A cross-validation strategy was used to statistically evaluate the results and ensure they were independent of the data partition for training and testing. Thus, the dataset is split into 90% data for training and 10% for testing. Further, each result is the average of ten different model runs. The results in bold correspond to the best configuration, and the underlined ones are the second best. It can be noted that combining the segmentation with transfer learning, detection, and data augmentation provides the best results, while the second best results were obtained for different configurations depending on the evaluation metric. Figure 33 presents visual results of the segmentation step C in Section 4.1.3, for 3 testing images of the CO2Wound data set. From left to right, each column contains the RGB images, the wound segmentation ground truth from a specialist, and the segmentation results obtained with transfer learning (TL), transfer learning with the detection step (TL+D), detection step (D), detection and data augmentation (D+DA), and all processes together (TL+D+DA). Each image includes the three metrics from top to bottom, i.e., F1-score, Precision, and Recall. Further, true positive (TP), false positive (FP), false negative (FN), and true negative (TN) values are represented by green, blue, red, and black, respectively. In particular, green represents the properly segmented wound tissue (TP); blue corresponds to tissue segmented as wound in healthy (no-wound) areas (FP); red indicates wound tissue that was segmented as no-wound (FN); and black represents correctly segmented regions where in fact no wound was present (TN). In summary, better-segmented images should

contain a few blue and red pixels and more green and black pixels. Thus, it can be seen that this behavior occurs for the case TL+D+DA in the last column.

Table 12
Quantitative segmentation results of the Ablation study 1. Importance of each step in the performance of the proposed framework (10-fold cross-validation).

TL=Transfer-Learning, D=Detection, DA=Data-Augmentation.

Ablation Settings			(↑) **Metrics** (mean±std)		
TL	**D**	**DA**	**F1-score**	**Precision**	**Recall**
✓	-	-	71.2±6.36	79.8±11.3	66.6±8.69
✓	✓	-	79.2±3.96	77.7±5.79	81.0±4.68
-	✓	-	77.4±5.59	75.3±6.58	80.4±4.24
-	✓	✓	76.7±4.44	80.7±6.55	79.7±5.39
✓	✓	✓	**83.2±3.81**	**84.5±4.51**	**81.9±4.71**

4.2.4. Visual Results. To provide a visual representation of the results obtained by each step of the proposed framework, Fig. 34 illustrates the images resulting from Steps A, B, C, and D for six different wound images in the CO2Wounds data set. Specifically, the first column presents the wound images acquired following the data protocol proposed in Step A (Section 4.1.1), which are the input to the proposed framework. The second column illustrates the results of Step B (Section 4.1.2), where the wound and calibration pattern is detected; specifically, the calibration pattern is highlighted by the magenta box, while the wound is enclosed by the lime green box. The third column depicts the resulting binary wound segmentation maps from Step C (Section 4.1.3). Finally, the fourth column presents the numerical results for wound area and perimeter estimation from Step D (Section 4.1.4). These results show that the proposed framework is able to work on images of different wound severity and locations within the lower limbs. Also, in all cases, the wound and the calibration pattern were appropriately detected. Further, when more

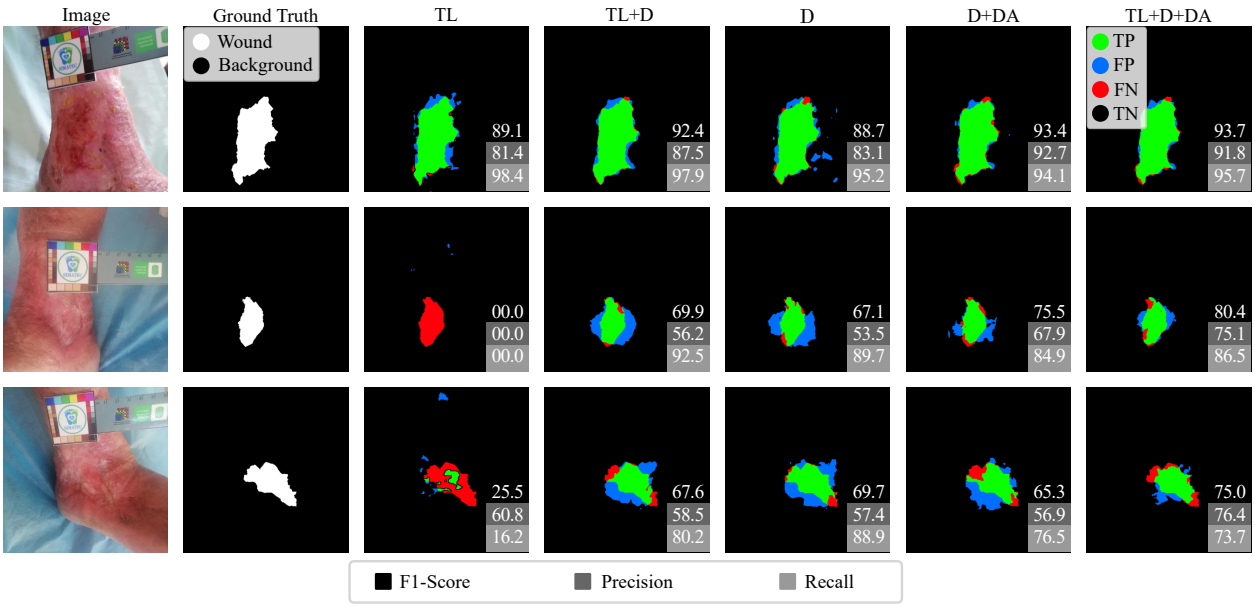


Figure 33. Visual results of step C -wound segmentation. (First column) RGB testing images from the CO2Wound database. (Second column) Expected ground truth wound segmentation; (Remaining columns) Segmentation results from the proposed framework with different combinations of transfer learning (TL), detection (D) and data augmentation (DA) schemes.

than one wound is present in the image, as in the last row of Fig. 34, the proposed framework is able to detect and segment both of them.

In addition, the performance of the proposed framework was compared with respect to other state-of-the-art segmentation methods for the analysis of chronic wounds. Specifically, the following segmentation methods were considered: VGG16 (Goyal et al., 2017), SegNet (Wang et al., 2015), MobileNetV2 (Wang et al., 2020a), and Unet (Ronneberger et al., 2015). All methods were implemented and tested using the CO2Wounds data set acquired in this work, following the network configurations suggested by the authors. Hyperparameters were tuned to achieve the best performance for the data set. Average results for 10 runs of each model are reported in Table 13, where it can be seen that the proposed framework outperforms its counterparts by at least 16%,

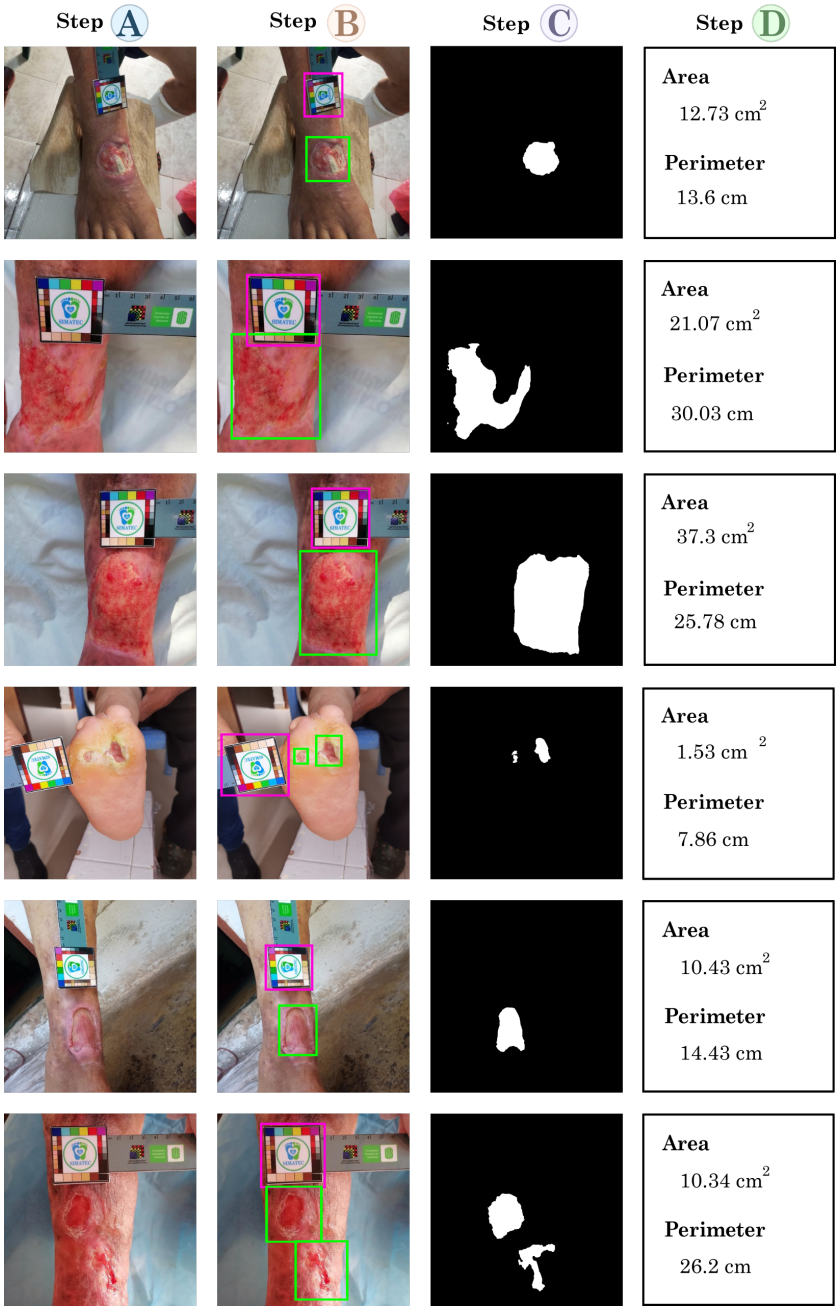


Figure 34. Step-by-step visual results of the proposed framework for six images from the CO2Wounds data set (rows). Each column corresponds to one step of the framework.

1.8%, and 11.2% in IoU, precision, and recall, respectively. It is worth highlighting the robustness of the proposed framework, as indicated by the standard deviation values, which are the lowest

overall. These results validate the positive influence of the detection, transfer learning, and data augmentation tasks in the proposed framework.

Table 13
Segmentation comparison with state-of-the-art and deep-learning-based segmentation methods on the CO2Wounds data set.

Methods	(↑) **Metrics** (mean±std)			
	IoU	**F1-Score**	**Precision**	**Recall**
VGG16 (Goyal et al., 2017)	53.6±7.25	60.6±7.33	68.7±12.7	53.9±8.71
SegNet (Wang et al., 2015)	48.4±5.45	58.80±5.23	68.4±11.4	51.5±7.17
MobileNetV2 (Wang et al., 2020a)	60.1±7.72	68.2±8.12	65.9±12.5	70.7±12.6
Unet (Ronneberger et al., 2015)	64.3±5.92	73.9±7.19	82.7±7.02	66.7±8.45
CO2Dnet framework (Ours)	**80.3±2.83**	**83.2±3.81**	**84.5±4.51**	**81.9±4.71**

4.2.5. Clinical Tracking Validation. Considering that, the proposed framework was designed as an alternative for chronic wound assessment in rural areas where medical infrastructure is insufficient, or poor transportation prevents appropriate treatment and tracking, we deploy it on an online platform so that medical staff can benefit from its functionality. For continuous integration, scalability, and integrity, the proposed framework for the segmentation and measurement of chronic wounds was deployed in a cloud computing service and integrated with the web platform via an API gateway connection. To date, nurse practitioners from the Leprosy Control program of *Sanatorio de Contratación*, in the Colombian town of Contratación, Santander, are the main users of this platform. Each user can manage their patients, as well as upload wound images that are automatically processed by the framework, which provides a visual comparison of the original and the segmented wound with the corresponding measures of area and perimeter, calculated as in Section 4.1.4. Thus, images of the same wound uploaded at different dates are

grouped in a temporal set, so that a temporal analysis can be performed. In essence, the temporal analysis provides insights into the evolution of the wound, with visual and numerical results that determine whether the wound is improving or progressing with respect to the oldest image. In this way, medical staff can make informed decisions related to the treatment.

Figure 35 illustrates examples of the wound tacking results with the proposed framework deployed online. In particular, the evolution of four different ulcers is presented in terms of the area and perimeter, in cm^2 and cm, respectively. Also, the segmented wound image is included at the top for each date, corresponding to the activation maps of each chronic wound image in the last convolutional layer of the segmentation neural network (step C in Section 4.1.3). Since image acquisition and uploading dates are independent of the algorithm, the top two cases have five records each, while the cases at the bottom have four records each.

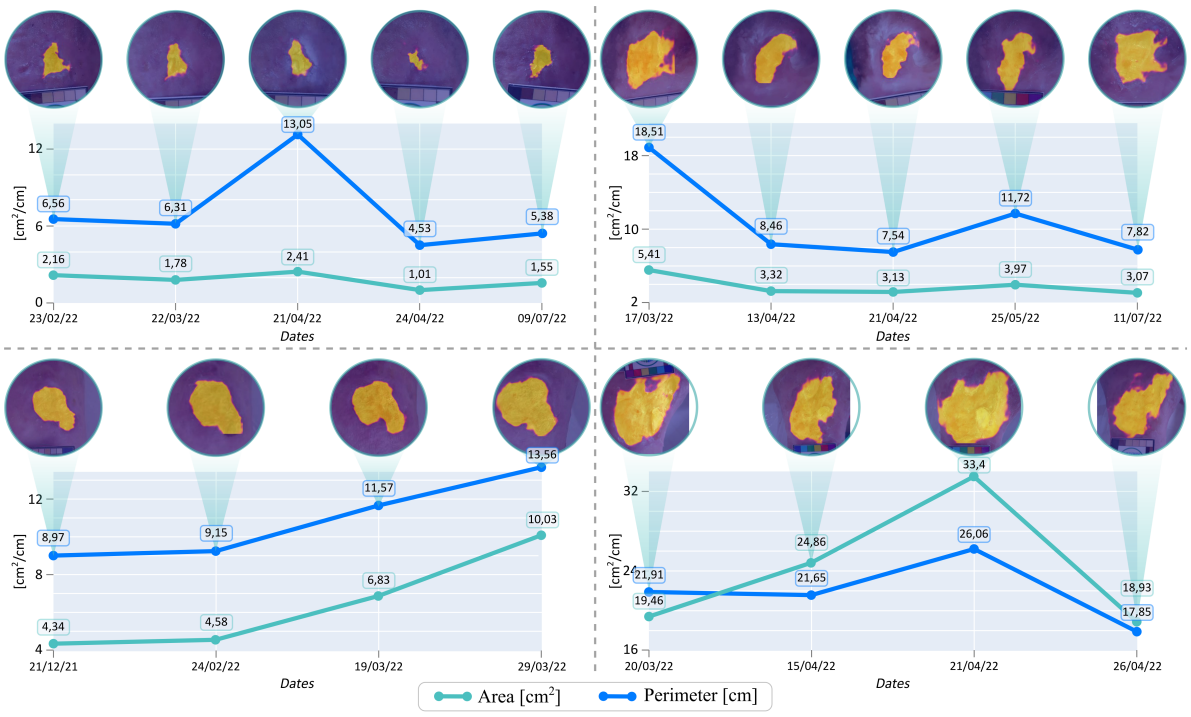


Figure 35. Tracking results in four leprosy chronic wounds using the proposed framework.

4.3. Conclusions

The CO2Dnet, a deep learning-based framework for chronic wounds assessment and tracking has been presented. The proposed framework includes four steps that include: image acquisition, wound and calibration pattern detection, wound segmentation, and wound area and perimeter estimation, where detection and segmentation employ deep neural networks with transfer learning and data augmentation schemes to avoid overfitting and improve performance. In addition, a new leprosy chronic wound data set (CO2 Wounds) from a developing country has been built. This new data set is intended to provide more variability in chronic wound data sets so that to enable future works in this context. Experimental results validate the proposed framework, with 80.3% IoU, 84.5% precision, and 81.9% recall over the CO2Wounds dataset. These results overcome

those from state-of-the-art segmentation methods (VGG16, SegNet, MobileNetV2, and U-net). Moreover, temporal analysis of the wounds is allowed by the proposed calibration pattern, which enables wound area and perimeter estimation. In this way, wounds can be monitored over time. Future works can take advantage of the calibration pattern for color calibration purposes, aiming at considering skin and ulcer color variations. Also, a deep-skin approach could be considered, where analysis of the wounds contemplates underlying tissues and below-skin information that can support wound healing.

5. A bilinear convolutional neural network approach for skin lesion classification

This chapter presents the second extension of this doctoral thesis, an accurate method based on deep learning for the classification of skin lesions. Specifically, this section introduces a framework to classify seven classes of skin lesions on RGB photographs through a bilinear convolutional neural architecture, achieving the highest state-of-the-art accuracy -as far as the authors know- over the well-known HAM10000 dataset.

Part of this chapter has been adapted from the journal paper (Calderón et al., 2021). Section 1.9 summarises the notation used in this chapter.

Chapter contribution. Skin lesions are dermis areas with abnormal growth or appearance compared to the surrounding area. They can be harmless such as a small scrape or severe skin cancer. These dermis abnormalities increase in size over time and cause morbidity problems. Correctly diagnosing these skin lesions is crucial for successful treatment, which is generally too expensive. Convolutional neural networks (CNN) have been investigated for classifying skin lesions with different training methods and techniques. However, the best results obtained by previous works show that there is a wide range to be achieved regarding detection, precision, and computational costs. The HAM10000 dataset is one of the best-known public data sets of skin lesion images (Tschandl et al., 2018), containing seven experts tagged classes. Some works that have addressed the classification of this data set are described below. In (Miglani and Bhatia, 2020), the authors fine-tuned the pre-trained EfficientNet-B0 architecture allowing 0.89 of precision and 0.97 of area under the curve (AUC). In (Mohapatra et al., 2020), they used the lightweight Mo-

bileNetV1 model to classify skin lesions of the same dataset with 0.86 of precision. In (Chaturvedi et al., 2020), they used a pre-trained MobileNet, achieving a similar 0.89 precision. Authors in (Emara et al., 2019) proposed a modified version of the Inception-V4 network, adding a residual connection to fuse low-level to high-level features and achieving 0.8617 of accuracy and 0.88 of AUC. (Chopade, 2020) proposed a lightweight CNN built from scratch with which they reach 0.89 of precision. Finally, in (Garg et al., 2019), they proposed a framework with a pre-processing step, which includes removing noise, adding resolution, pre-segmentation, data augmentation and transfer learning over a ResNet architecture, achieving 0.88 of precision and 0.905 of accuracy. For comparison purposes, the accuracy classification results of the HAM10000 dataset with non-deep learning-based methods are 0.659, 0.6515, and 0.6586 for the random forest, XGBoost, and support vector classifiers, respectively.

This chapter presents a bilinear CNN approach capable of classifying the seven skin lesion classes of the HAM10000 dataset (Tschandl et al., 2018), achieving the highest state-of-the-art accuracy and low computational cost. Specifically, this work proposes a framework of three steps: (i) image pre-processing (preparation, data augmentation, resize), (ii) transfer learning via fine-tuning on a bilinear CNN model composed by the pre-trained ResNet50 + VGG16 architectures working in parallel, and (iii) features extraction for classification. Several simulations were executed over the HAM10000 dataset. The results show that a bilinear approach composed of the ResNet50 and the VGG16 architectures increases the accuracy by up to 2.7% compared to the state-of-the-art. Specifically, the proposed approach achieves 0.9321 in accuracy, 0.9292 in precision, 0.9300 in recall, 0.9321 in F1 score, and requires 238.6 minutes for training. This performance increase

can help support the clinicians' diagnosis to provide a second opinion, reducing morbidity and treatment costs.

5.1. Proposed Method: BILSK

This method focuses on classifying skin lesion images through a deep-learning approach with a nontraditional bilinear CNN architecture. The model is trained in a transfer learning and fine-tuning way.

5.1.1. Step A: Image preprocessing. The image preprocessing step for our experiments includes preparation, data augmentation, and resizing.

First, we randomly divided the dataset into three subsets: 80%, 15%, and 5% for training, testing, and validation, respectively. Second, in the training of CNNs, it is essential to have a large amount of data available to obtain a good classification performance and avoid over-fitting (Sanchez et al., 2021a; Göçeri, 2020c). Therefore, in this work, data augmentation is applied to increase the size of the training subset through six transformations: flip horizontal, vertical, displacement width, height, rotation, and zoom. Third, we resize each training image to 224×224 pixels.

The original number of images by each class and images by set (Train/ Validation/ Test) in the HAM10000 dataset is shown in Table 14.

5.1.2. Step B: Bilinear CNN Architecture. A bilinear CNN architecture consists of the disposition in parallel of two equal or different CNN models connected at the top and end. Figure 36 shows a bilinear architecture of two CNN models, A and B. An input feed both networks and a training step calculates the weights for each neuron; the two model tensor outputs are for the

Table 14
Images by each class and set (Train/Validation/Test) in the HAM10000 dataset.

Class	Train (80%)	Validation (5%)	Test (15%)	Total per Class	% per Class
Akiec	272	14	41	327	3.26%
Bcc	450	16	48	514	5.13%
Bkl	939	40	120	1099	10.97%
Df	101	3	11	115	1.14%
Mel	1030	21	62	1113	11.11%
Nv	5101	401	1203	6705	66.94%
Vasc	119	6	17	142	1.41%
Total	**8012**	**501**	**1502**	**10015**	**100%**

features at each image location and pooling to obtain an image descriptor, as denoted in 23. The outer product and sum pooling for two model tensor outputs are denoted as

$$\mathscr{X}_{b,L_1,L_2} = \sum_{m,n} \mathscr{A}_{b,m,n,L_1} \mathscr{B}_{b,m,n,L_2}, \tag{23}$$

where $\mathscr{A} \in \mathbb{R}^{b\times m\times n\times L_1}$ and $\mathscr{B} \in \mathbb{R}^{b\times m\times n\times L_2}$ are the output matrix from the first and the second model. Further, b denotes the batch size, m and n the spatial dimensions, and L_1 and L_2 the features. The resulting matrix $\mathscr{X} \in \mathbb{R}^{b\times L_1\times L_2}$ captures pairwise correlations between the feature channels and can provide richer representations than linear models (Ustinova et al., 2017). The resulting matrix $\mathbf{X}$ can be reshaped to $b \times L_1 L_2$ and then is normalized by signed square root normalization given by

$$\mathbf{Y} = sign(\mathbf{X})\sqrt{|\mathbf{X}|}, \tag{24}$$

which is also normalized by ℓ_2 norm

$$\mathbf{Z} = \frac{\mathbf{Y}}{\|\mathbf{Y}\|_2}. \tag{25}$$

Finally, the output is passed to a fully connected layer with a softmax activation to obtain predictions.

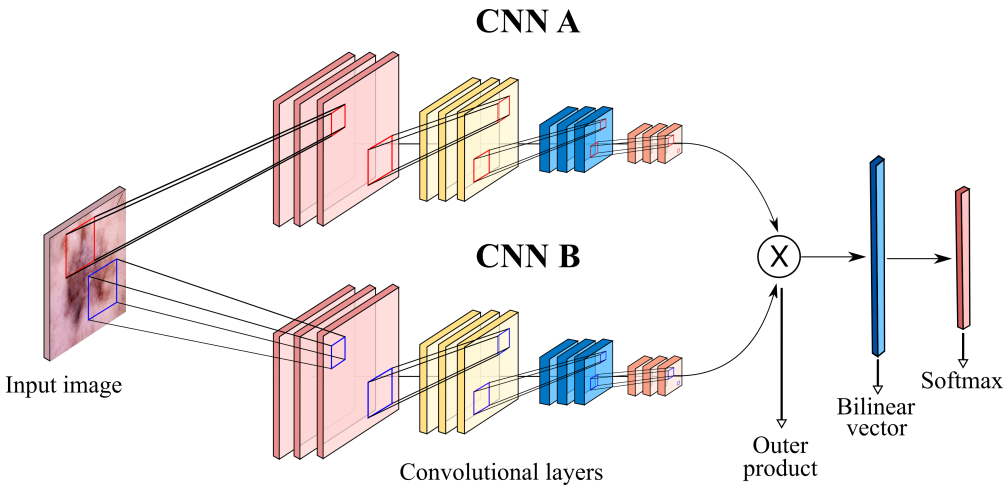


Figure 36. Bilinear CNN for skin lesion classification.

Three CNN architectures were used in this approach to compose and compare two different bilinear architectures. The selected CNNs have demonstrated strong performance in image classification tasks. Each model is described in the following sub-sections.

This approach uses two bilinear CNN models; the first combines the ResNet50 and the VGG16 architectures, and the second combines the ResNet50 and the Xception architectures. The fully connected layers were removed from each model before combining them into the bilinear CNN, and after outer product operation, a single fully connected layer was adjusted to the classification task, i.e., seven neurons were added instead of 1000 neurons.

ResNet-50 with VGG16 The CNNs used in the first bilinear model addressed in this work

are described below. **ResNet-50** (He et al., 2016) is a deep CNN with high performance in computer vision tasks, winning the ImageNet classification challenge in 2015. ResNet learns the residual representation functions instead of directly learning the signal representation. This network uses the skip connection, which helps avoid the vanishing gradient problem and increases the model convergence, forwarding the output of a layer to the input in subsequent next layers. In particular, we use for this work the ResNet-50 model since it provides better performance for this purpose than other ResNet architectures. This CNN performs the initial convolution with a kernel size of 7×7 followed by a Max-pooling with 3×3 kernel size. Then, the model consists of four stages with 3, 4, 6, and 3 residual blocks, respectively. Each residual block has three convolutional layers of kernel sizes 1×1, 3×3, and 1×1. The 1×1 convolution layer is used for dimension reducing and restoring, and the 3×3 layer is used as a bottleneck with smaller input/output dimensions. Finally, the network has an average pooling layer followed by a fully connected layer of 1000 neurons. In total, the ResNet-50 has over 23 million trainable parameters. On the other hand, **VGG16** was proposed in (Simonyan and Zisserman, 2014). The model obtained 92.7% top-5 test accuracy in an ImageNet Classification Challenge. This CNN has 16 layers using 13 convolutional layers and three fully connected layers, reaching 138 million trainable parameters. The input layer size is 224x224, followed by five convolutional blocks. The first and second blocks have two layers of 3×3 kernel size, with a max-pooling with size 2×2. The third, fourth, and fifth blocks have three layers with 3×3 kernel size and a max-pooling of 2×2. Finally, the model has three fully connected layers of 4096, 4096, and 1000 neurons. Activation functions should be chosen carefully in deep networks with residual blocks (Göçeri, 2019). Although various activation functions

have been applied in recent works (Yu et al., 2020; Göçeri, 2021; Tanaka, 2020; Goceri, 2021), ReLU has been used in the proposed architecture due to its efficiency.

ResNet-50 with Xception The second bilinear CNN approach addressed in this work is composed of a ResNet-50 in one arm and an Xception network in the other one. ResNet-50 was detailed in 2.3.1. and Xception is described as: **Xception** (Chollet, 2017b) is a deep CNN with 36 convolutional layers, which outperformed the ResNet and the VGG-16 in the ImageNet classification challenge. The Xception model has 14 modules divided into three stages, the entry flow, the middle flow, and the exit flow, reaching up to 22 million parameters. The entry flow has four modules; the first one has an input layer of size 299×299, followed by two convolutional layers with a 3×3 kernel size. The second to the fourth modules have two separable convolutional layers with a 3×3 kernel size followed by a 3×3 max pooling and a linear residual connection around each module. In the middle flow, a three separable convolutional layer module is repeated eight times, each layer with a 3×3 kernel size, and a residual connection around the module. The exit flow starts with two separable convolutional layer modules with a 3×3 kernel size, a 3×3 max pooling, and a residual connection. Finally, two separable convolutional layers are followed by a global average pooling.

5.1.3. Step C: Transfer Learning and Fine-Tuning. Training a deep CNN from scratch is challenging, especially in medical image classification due to a lack of large labeled datasets. Therefore, the transfer learning technique emerges as an alternative to reuse a pre-trained model as a starting point to generate a model for a new task of interest. This technique is useful as an optimization that allows rapid progress or improved performance when modeling the second

task.

For the purposes of this work, we use the ResNet50, VGG16, and Xception convolutional neural networks previously trained with the ImageNet dataset, which are available in the Keras library. ImageNet (see subsection 4.1.2) is a large visual database designed for visual object recognition software research (Deng et al., 2009). Then, the fully connected layer with 1000 neurons is replaced by a seven neurons layer; and the average pooling layer is removed to retrain the weights to the target task, without the initialization being random.

For improving performance, we use the fine-tuning technique of freezing 70% of the layers for the ResNet50 model, 75% of the layers for the VGG16 model, 70% of the layers for the Xception model and adjust the weights of the remaining ones, this can potentially achieve meaningful improvements, by incrementally adapting the pre trained features to the new data (Tajbakhsh et al., 2016).

Hyperparameters. Table 15 describes the hyperparameters used for training the models. Images were resized to 224x224 in order to have a balance in a good quality resolution and model efficiency and the batch size was set in 32 for both models. Also, a decay function was used to monitor the validation loss and decrease the learning rate (lr) when this metric stopped improving after a certain number of epochs determined by the patience parameter, see Eq. 26.

$$Decreased_{lr} = lr \cdot factor, \tag{26}$$

Adam optimizer was used in both models to improve the learning process and enhance the

performance, and the epochs number was set at 10. Although several optimizers have been implemented in deep networks to obtain results with high performance, such as Lagrangian optimizer (Kervadec et al., 2019) or Sobolev gradient-based optimizer (Göçeri, 2020a), they generally cause high computational costs. Therefore, we applied the Adam optimizer in this work.

Table 15
Hyperparameters for each proposed bilinear CNN.

Parameter	**ResNet50 & VGG16**	**ResNet50 & Xception**
Image Size	224x224	224x224
Batch Size	32	32
Learning Rate	0.0001	0.001
Decay Factor	0.3	0.5
Patience	2	2
Optimizer	Adam	Adam
Epochs	10	10

5.2. Simulations and Results

This section evaluates the performance of the BILSK skin lesion classification proposed method. We classify the lesions from the HAM10000 dataset described below, perform the data augmentation technique, and quantitatively evaluate performance using six standard metrics in medical image classification. This proposed approach gets the highest accuracy against current state-of-the-art techniques in the classification of this skin lesions dataset. We compare the results obtained against other methods of state-of-the-art. To evaluate the classification results, we use the metrics: accuracy, precision, F1 score, recall, the area under the ROC curve (AUC), and the Matthews correlation coefficient (MCC).

5.2.1. Dataset. The HAM10000 dataset is an extensive collection of 10,015 dermoscopic images of skin-pigmented lesions, annotated into seven classes (Tschandl et al., 2018). This dataset became available in 2018 through the ISIC Challenge by Canfield Scientific, a benchmarking initiative that gives tasks of diagnosis, detection, and segmentation to the research community. The classes included in the dataset are actinic keratoses and intraepithelial carcinoma (Akiec), basal cell carcinoma (Bcc), benign keratosis (Bkl), dermatofibroma (Df), melanoma (Mel), melanocytic nevi (Nv), and vascular lesion (Vasc). Figure 37 shows a random sample of each skin lesion type from the dataset, with abbreviations at the top representing the name of each class wound. The images for HAM10000 were taken on men and women ages 5 to 80 at two locations, the Department of Dermatology at the Medical University of Vienna, Austria, and the Cliff's Skin Cancer Practice Rosendahl from Queensland, Australia. More than 50 % of the lesions labeled in the dataset have been confirmed by pathology. Meanwhile, the ground truth of the rest of the cases was determined by follow-up, expert consensus, or confirmation by in vivo confocal microscopy.

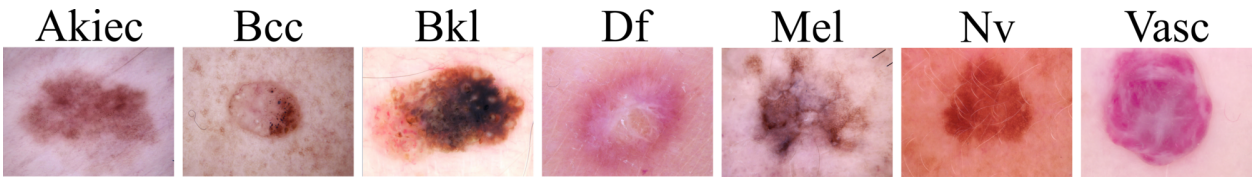


Figure 37. Seven skin lesions classes in the HAM10000 dataset.

Regarding each skin lesion in the dataset, the *Akiec* is the most common precancer in skin damaged by chronic exposure to ultraviolet rays (Goldenberg and Perl, 2014). *Bcc* is a slow-growing, locally invasive malignant epidermal skin tumor (Rubin et al., 2005). *Bkl* is one of the

most common non-cancerous skin neoplasms in older adults (Roh et al., 2016). *Df* is a common benign fibrous nodule that most often arises on the skin of the lower legs (Cohen et al., 2019). *Mel* is the most serious type of skin cancer, it develops in the melanocyte cells that produce melanin (Miller and Mihm Jr, 2006). *Nv* are benign neoplasms or hamartomas composed of melanocytes, the pigment producing cells that constitutively colonize the epidermis (Damsky and Bosenberg, 2017). Lastly, *Vasc* are relatively common abnormalities of the skin and underlying tissues, more commonly known as birthmarks (Syed et al., 2016).

Table 14 shows that the distribution of the training image set is unbalanced; that is, there are more images of some classes than others. This can cause the classifier to be biased and assign the majority classes' labels (e.g., melanocytic nevi or melanoma). For this reason, the data augmentation technique was applied. To achieve this task, several transformations such as rotation, zoom, width shift, height shift, horizontal flip, and vertical flip were applied to the original training images in order to generate new training samples to the training set.Random rotations up to 180 degrees were applied to the images. Besides, random zoom was applied in a factor of 0.1; width and height shift in a factor of 0.1; random horizontal and vertical flip were also performed. Furthermore, input images were normalized into a standard range from [0,1]. The results of the data augmentation for a random sample of the data set are shown in Fig. 38.

After applying the data augmentation technique, the training data is more balanced. The new amount of data per class is shown in Table 16.

5.2.2. Quantitative Classification Results. The ROC curves of all seven classes of the HAM10000 dataset, classified with the bilinear approach of CNN ResNet50 and VGG16,

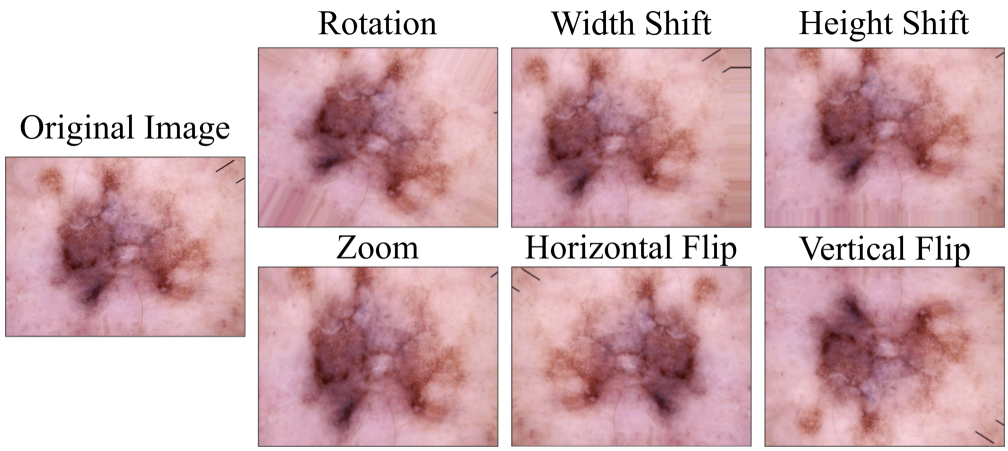


Figure 38. Data augmentation results with six transformations for a random sample of the HAM10000 dataset.

Table 16
Distribution of the HAM10000 training dataset by classes after data augmentation.

Class	Original Train Images	Augmented Images
Akiec	272	5696
Bcc	450	5656
Bkl	939	5890
Df	101	4747
Mel	1030	5886
Nv	5101	5997
Vasc	119	5570
Total	**8012**	**39442**

are shown in Fig. 39. Note that the corresponding AUC values for each curve are recorded in the labels of the conventions. Also, in these labels are reported the micro and macro-averages. Micro-average corresponds to the contributions of all classes to compute the average metric, whereas macro-average computes the metric independently for each class and then calculates the average.

In the same sense, the results with the ResNet50 and Xception bilinear approach and their respective AUC values are shown in Fig. 40. Note that in both bilinear approaches, the extreme

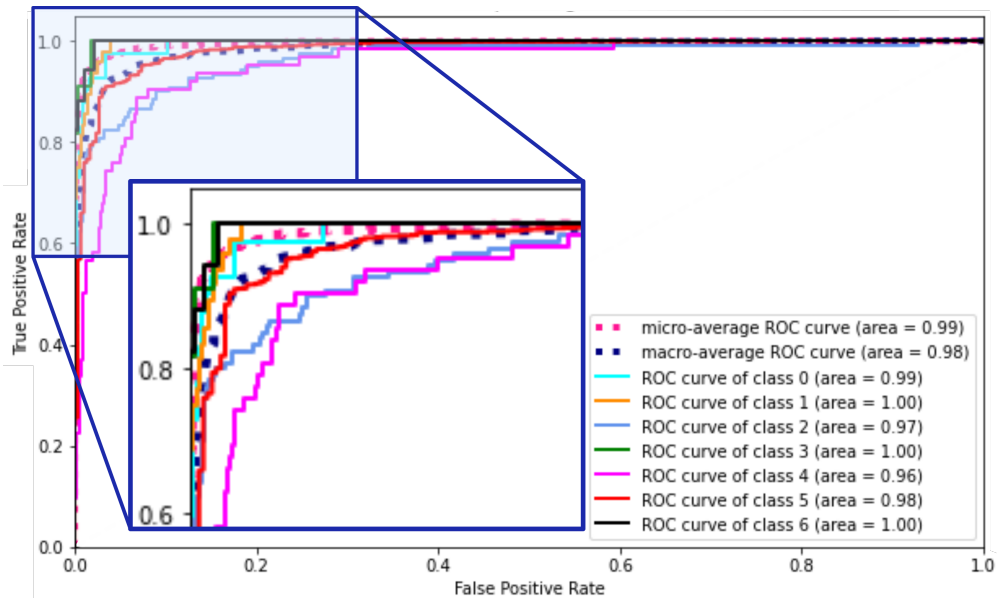


Figure 39. ROC curves for each class of the HAM10000 dataset using the ResNet50 and VGG16 bilinear approach.

point of curvature of the graphs is very close to the upper left corner. If the point of curvature were to form a right angle, the AUC would be 1. Therefore, the curves obtained through this work, with vertices close to forming angles of 90 degrees, demonstrate a good separation capacity of the evaluated model.

The quantitative results of the proposed approaches are presented in Tables 4 and 5. In particular, Table 4 shows the numerical results for each of the seven skin lesion classes in HAM10000 using the ResNet50 and VGG16 bilinear approach. Meanwhile, Table 5 exposes the ResNet50 and Xception bilinear structure results. These tables show the main classification results for each class, the macro, and the weighted average obtained by the model.

Note that the results are very similar in both cases based on the metrics evaluated in the tables. However, the precision, recall, and F1-Score are slightly higher in the VGG16 approach.

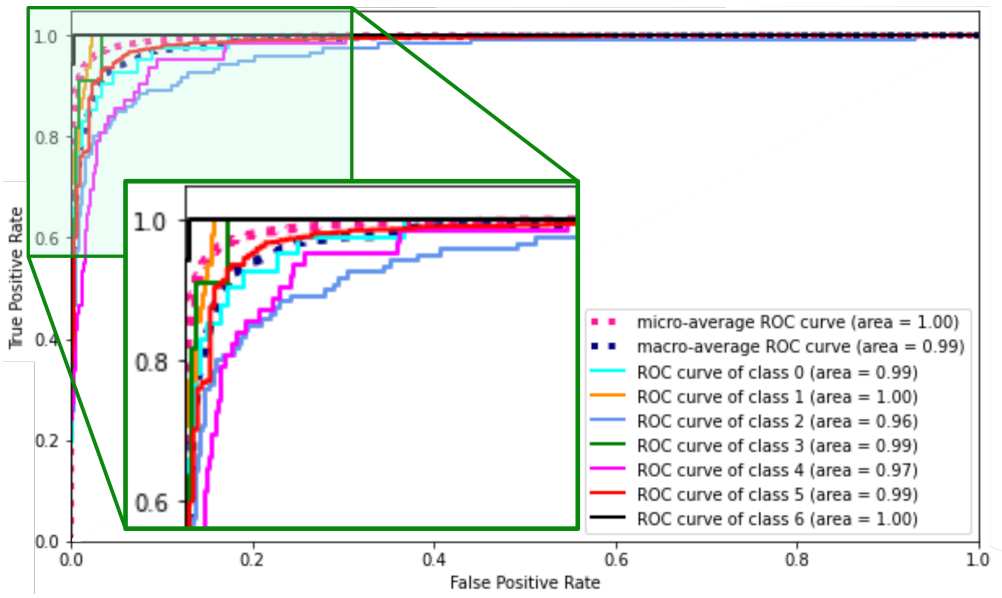


Figure 40. ROC curves for each class of the HAM10000 dataset using the ResNet50 and Xception bilinear approach.

The confusion matrix is a quantitative-graphical way of analyzing the classification results. In Figs. 41 and 42, the confusion matrices of each of the two bilinear approaches are shown when evaluating the images in the Test column of Table 2. In these matrices, the vertical axis corresponds to the true labels and the horizontal axis to the predicted labels.

Identity matrices are expected as a result of these graphs, which would indicate perfect classification of all images, while values outside the main diagonal are the errors of the classification. In Fig. 6 a value of 0.55 was obtained for the mel class (melanoma), due to several images corresponding to this class were predicted by the model as nv (melanocytic nevi) class images. In Fig. 7 the same confusion was also presented between the mel and nv classes. Also, with this model there were some errors in the labeling of df (dermatofibroma) images, since they were predicted as akiec (intraepithelial carcinoma), bkl (benign keratosis), or nv. However, in general it can be

Table 17
ResNet50 and VGG16 Classification Report

Class	Precision	Recall	F1-Score
Akiec	0.83	0.73	0.78
Bcc	0.78	0.81	0.80
Bkl	0.83	0.78	0.80
Df	1.00	0.73	0.84
Mel	0.65	0.55	0.60
Nv	0.96	0.98	0.97
Vasc	0.93	0.82	0.87
Macro Avg	**0.86**	**0.77**	**0.81**
Weighted Avg	**0.9292**	**0.9321**	**0.9300**

Table 18
ResNet50 and Xception Classification Report

Class	Precision	Recall	F1-Score
Akiec	0.65	0.78	0.71
Bcc	0.86	0.77	0.81
Bkl	0.82	0.70	0.75
Df	0.67	0.55	0.60
Mel	0.62	0.55	0.58
Nv	0.97	0.99	0.98
Vasc	0.94	0.88	0.91
Macro Avg	**0.79**	**0.74**	**0.76**
Weighted Avg	**0.9207**	**0.9261**	**0.9221**

deduced from both confusion matrices that the results are highly accurate. The proposed models were trained using Google Colab free GPU. Table 19 shows the computational costs in terms of RAM and GPU used in Gigabytes, time per epoch, and total training time in seconds. The Table shows that the ResNet50 and VGG16 models are less computationally expensive than the other approach. This was expected since the combination of the ResNet50 and Xception architectures of the second model has a significantly higher number of layers and trainable parameters than the first approach.

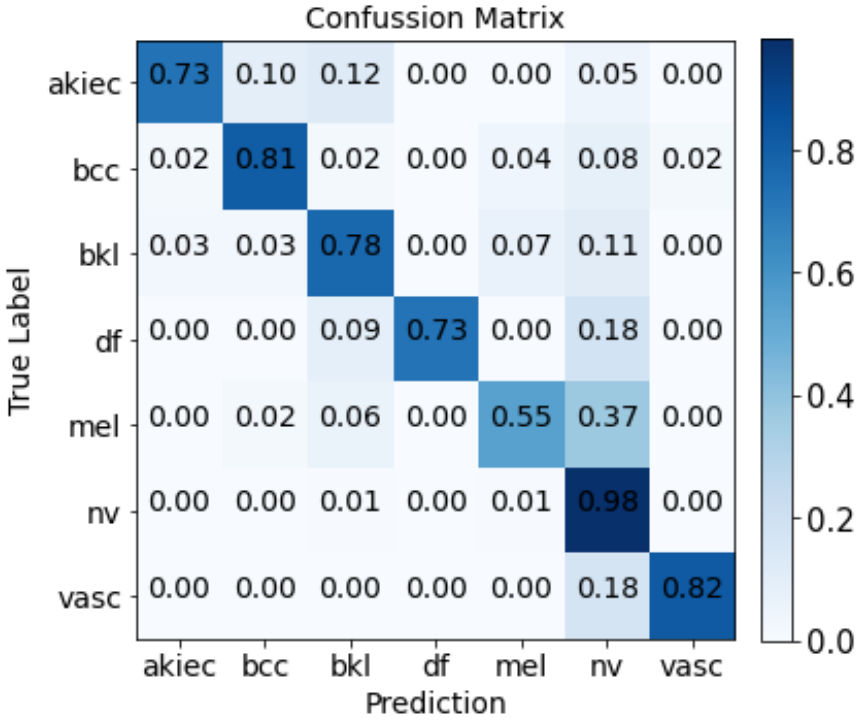


Figure 41. ResNet50 and VGG16 Confusion Matrix.

Table 19
Computational Costs

Model	**ResNet50 and VGG16**	**ResNet50 and Xception**
RAM (GB)	2.24	3.07
GPU (GB)	9.60	14.24
Time per epoch (s)	1432	1956
Training time (s)	14320	19560
Trainable Parameters (Million)	22.37	23.33

For comparison purposes, the most outstanding state of the art works in the classification of the data set HAM10000 have been compiled in Table 7. These methods have been described in the Introduction Section (Chopade, 2020; Mohapatra et al., 2020; Chaturvedi et al., 2020; Miglani and Bhatia, 2020; Garg et al., 2019; Emara et al., 2019). Table 7 includes the results of the six state-of-

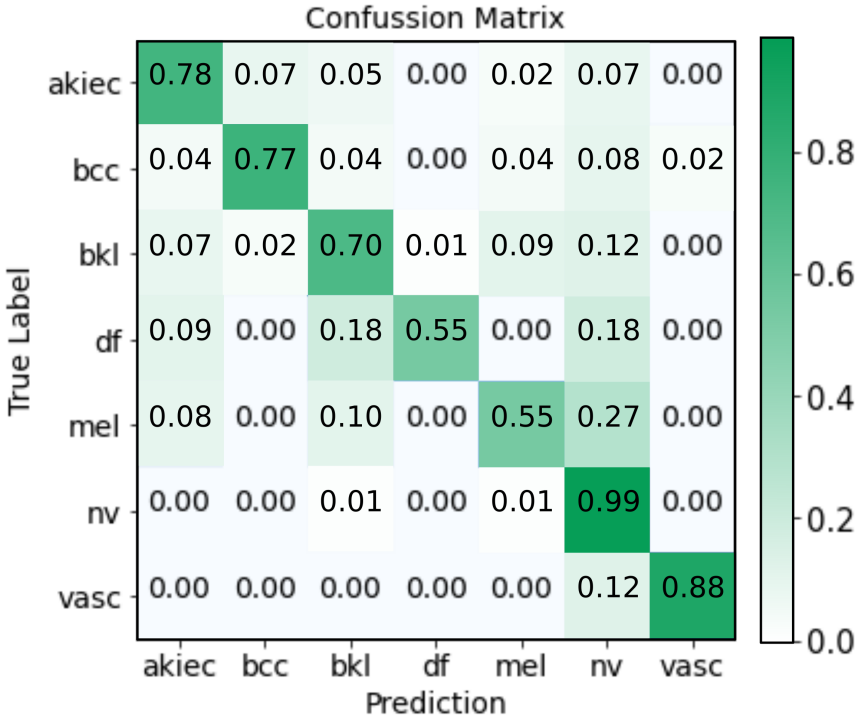


Figure 42. ResNet50 and Xception Confusion Matrix.

the-art works, from lowest to highest accuracy, and of the two bilinear approaches addressed in this chapter. The metrics that were not reported in the original documents of each work are indicated with a dash (-) in the table and the best result of each column is highlighted in bold. Note that, in general, all the best results were obtained with one of the proposed approaches. Specifically, in the first four metrics Accuracy, Precision, Recall and F1-Score, the best results were achieved with the CNN ResNet50 and VGG16, which is also the lightest model in computation terms. Finally, the CNN ResNet50 and VGG16 achieved a higher MCC compared to the other approach.

5.3. Conclusions

Developing CAD systems for medical image classification tasks using bilinear CNN models represents a significant performance improvement over single CNN, furthermore, balancing the

Table 20
Quantitative results of different classification methods for the HAM10000 dataset.

Method	Accuracy	Precision	Recall	F1-Score	MCC
EfficientNets (Miglani and Bhatia, 2020)	-	0.89	0.89	-	-
S-CNN (Mohapatra et al., 2020)	0.80	0.86	0.82	0.85	-
MobileNet (Chaturvedi et al., 2020)	0.8315	0.89	0.83	0.83	-
Modified Inception-v4 (Emara et al., 2019)	0.8617	-	-	0.717	-
Incremental CNN (Chopade, 2020)	0.9026	0.89	0.89	0.88	-
Decision Support System (Garg et al., 2019)	0.905	0.88	0.77	0.74	-
ResNet50 and Xception	0.9261	0.9207	0.9261	0.9221	0.8210
BILSK: ResNet50 and VGG16	**0.9321**	**0.9292**	**0.9321**	**0.9300**	**0.8330**

data, using pre-trained models, and adapting them to new data through transfer learning and fine-tuning improve model generalization. In general, the work carried out in this chapter overcomes the state-of-the-art methods in terms of Accuracy in the classification of the public dataset HAM10000. Fig. 42 shows that the ResNet50 and Xception improved the vasc and akieck class precision in comparison with the results presented in Fig 41. However, the model is less accurate in other classes, such as bcc, bkl, and df. It is observed that the use of bilinear approaches is a strategy that can provide better results in the analysis or classification of medical images. Regarding the two combinations addressed in this chapter, although there is no critical difference between the two models' precision, a distinction was evidenced in terms of precision by class. Therefore, if a single specific lesion type is looked at, the most suitable model could be selected for the required class.

6. Accurate Deep Learning-Based Gastrointestinal Diseases Classification via Transfer Learning Strategy

This chapter presents the latest extension of this doctoral thesis, a lightweight, efficient, and high-precision neural method for classifying diseases and abnormalities of the gastrointestinal tract. The novelty and usefulness of the proposed method lie in the fact that it uses only one-fifth of the trainable parameters compared to other more complex methods that achieve a similar precision.

Part of this chapter has been adapted from the conference paper (Escobar et al., 2021).

Chapter contribution. The automatic detection of diseases and gastrointestinal tract anomalies is challenging for medical experts, affecting patient treatment decisions. Therefore, it is essential to implement deep learning-based systems (Pang et al., 2021) that support the detection of anomalies and diseases in endoscopic images.

For instance, in (Pogorelov et al., 2017), they propose the classification of the Kvasir-V1 dataset (endoscopic images) using global features and transfer learning, obtaining an accuracy of 95.9%. In (Cogan et al., 2019), they pre-processed the Kvasir-V2 dataset with edge removal, contrast enhancement, filtering, color mapping, resizing, and data augmentation. These images were used to train and test three CNNs: Inception-v4, Inception-ResNet-v2, and NASNet, obtaining the best result, 98.48% accuracy, with the Inception-ResNet-v2. In (KahsayGebreslassie et al., 2019), they classified the Kvasir-V1 dataset using transfer learning and fine-tuning on the ResNet50 and DenseNet121 models, achieving an accuracy of 87.8% and 86.9%, respectively. An approach of combining features extracted by several CNNs was addressed in (Gamage et al., 2019). They

classified the Kvasir-V2 with DenseNet-201, ResNet-18, VGG-16, InceptionV3, Xception, and Inception-Resnet-V2, connected by a global average pooling layer to obtain feature vectors. This method allows them an accuracy of 97.38%.

Notice that state-of-the-art methods for gastrointestinal anomaly classification have complex architectures, which require multiple parameters to be trained. Then, there is a scope for developing a light and easily replicable deep-learning-based method that maintains the high precision of more complex models in gastrointestinal anomaly classification.

This chapter addresses the Kvasir-V2 dataset classification problem using pre-processing, transfer learning, and hyperparameter fitting. Due to the lack of labeled medical images, we first use data augmentation with geometric transformations to generate new images. Then, we use the VGG-16 network with previously trained weights for the Imagenet dataset to extract features. Instead of retraining all the architecture from scratch, we explore different fine-tuning settings that lead us to achieve the best accuracy with significantly fewer parameters. We found that performing a fine-tuning of the weights from the convolutional layer 3 in block 4 of the VGG-16 architecture to the fully connected layer led to the best results in terms of accuracy. Our proposed framework provides high classification precision from light and a computationally inexpensive neural model. In this way, we can show that it is unnecessary to readjust all the pre-trained weights of a neural model for a long retraining period to achieve high performance in classifying endoscopic images in eight classes of diseases or anomalies of the gastrointestinal tract. Our proposed approach achieves 98.20% accuracy during testing by only using the fifth part of trainable parameters compared to the state-of-the-art methods.

6.1. Proposed Method: GT-Net

In this section, we described our proposed approach for diseases and anomalies classification of the GI tract. We separate the steps and explain them in the following subsections.

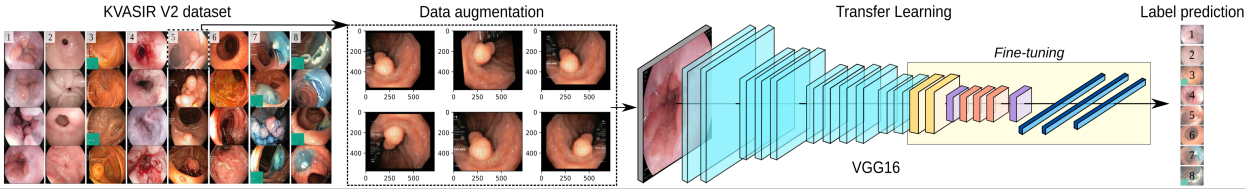


Figure 43. The proposed approach's workflow to detect gastrointestinal anomalies and diseases from endoscopic images using CNN and transfer learning.

6.1.1. Step A: Image Preprocessing. In the training of CNNs, it is essential to have a large amount of data available to obtain a good classification performance and avoid overfitting. Therefore, in this work, data augmentation is applied to increase the size of the training set by performing different geometric transformations. In our experiments, we randomly divided the dataset into three subsets: 80 %, 15 %, and 5 % for training, testing, and validation, respectively. Second, we perform data augmentation to the training subset through six transformations: flip horizontal, vertical, displacement width, height, rotation, and zoom. Finally, we resize each training image to 400×400 pixels before feeding the images to the CNN.

6.1.2. Step B: Transfer Learning Strategy. One of the main challenges of employing deep learning models in the medical area is the lack of training data (Alzubaidi et al., 2020). Furthermore, computationally expensive models are more challenging to scale. For these reasons, transfer learning is advantageous in the medical field. From a CNN model previously trained with natural image data sets, the classification task can be efficiently transferred to a do-

main with medical images avoiding the expensive training from the scratch process.

Consequently, we propose to use the transfer learning technique by comparing five different CNN models to choose the best-performing network. Also, we use fine-tuning to re-train only some layers of the neural networks using the preprocessed training data in II.A. For comparison purposes, we retrain the layers of the last convolutional block for each of the five models and keep all the other weights from their ImageNet pretraining unaltered. We replace the previous layer of the five networks with a new layer of eight units representing the number of classes in the Kvasir dataset. Figure 43 shows the pipeline of the proposed approach to detecting gastrointestinal anomalies and diseases from endoscopic images using CNNs via a transfer learning strategy.

6.2. Simulations and Results

This section presents the numerical results of the proposed framework for gastrointestinal disease classification through endoscopic images using the Kvasir-V2 dataset. We trained the proposed model using the *Adam* optimizer with a batch size of 32 and an initial learning rate of 1×10^{-4}. We applied an inverse time-decay learning rate schedule with a decay factor of $1 \times 10^{-5}/\text{epochs}$ that was triggered every epoch. We trained the network for 15 epochs, which took about 45 minutes on an Nvidia Tesla T4 GPU.

6.2.1. Dataset. The Kvasir-V2 dataset (Pogorelov et al., 2017) is composed of endoscopic images from inside the gastrointestinal tract. This version of the dataset consists of 8,000 images grouped into 8 different classes (i.e., 1000 per class), which have been annotated and verified by experienced endoscopists. The classes are based on three anatomical landmarks (z-line, pylorus, cecum), three pathological findings (esophagitis, polyps, ulcerative colitis), and two other

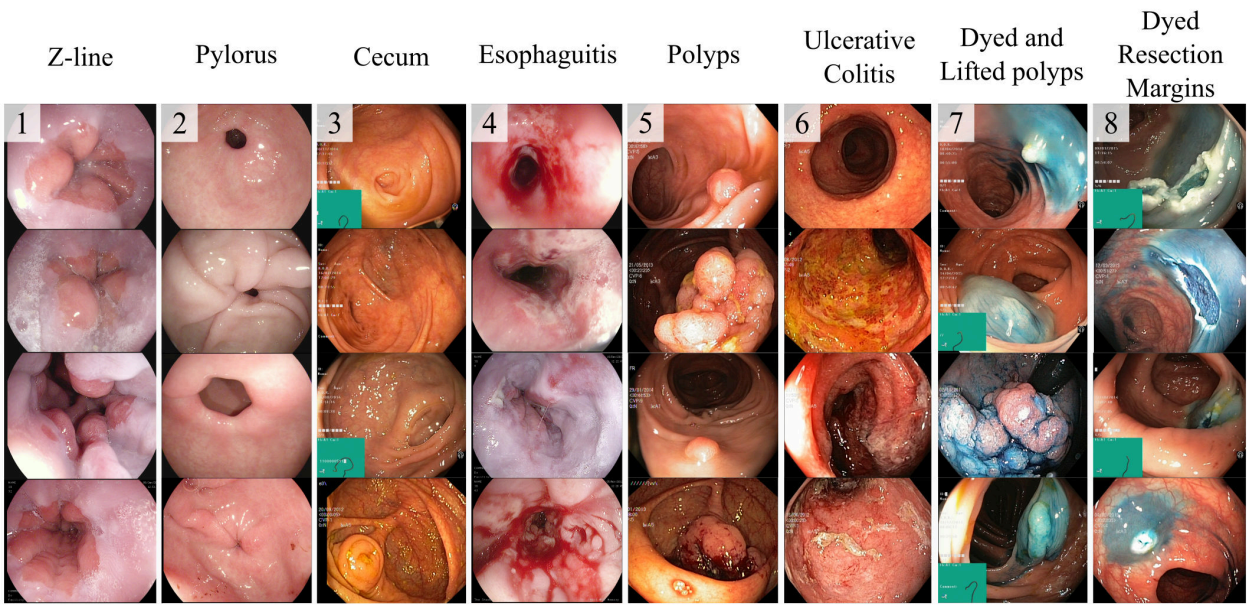


Figure 44. Four random samples from each class in the Kvasir-V2 dataset.

classes related to the removal of polyps (dyed and lifted polyps, dyed resection margins) as shown in Fig. 44. The dataset splits utilized in this study were recommended by the dataset owners. For the ablation studies, we conducted a 5-fold cross-validation on the train and validation sets, while the final reported results are based on a single run performed on the test set.

Data Augmentation. We perform data augmentation on the Kvasir-V2 dataset using the geometrical transformations exposed in Section II.A. (see Fig. 45). Then, the training images were doubled with the data augmentation, from 800 to 1600 training images for each class, obtaining a total of 12800 training images, considering that there are 8 classes and 1600 training images in each class, as seen in Table 21. Note that the number of images for testing and validation was not changed.

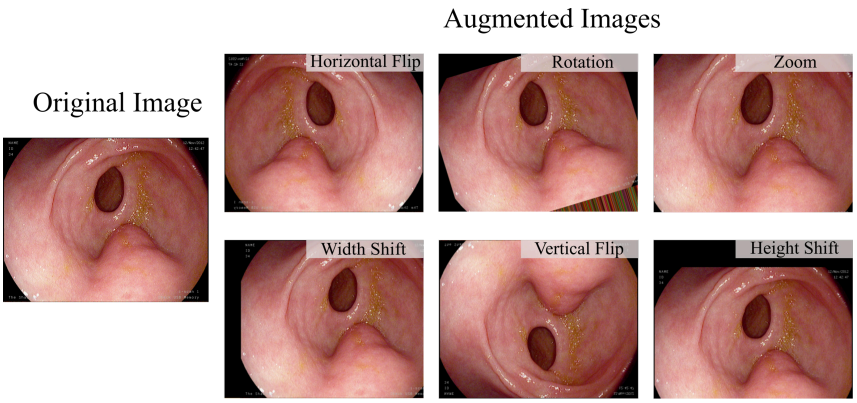


Figure 45. Geometric transformations in data augmentation.

Table 21
Images distribution before and after augmentations.

	Before		**After**	
	Images per class	Total images	Images per class	Total images
Train	800	6400	1600	12800
Test	150	1200	150	1200
Validation	50	400	50	400
Total	1000	**8000**	1800	**14400**

6.2.2. Quantitative Results. The results obtained for our proposed method are presented in the last row of Table 25, in which we calculated the metrics: accuracy (ACC), precision (PREC), recall (REC), specificity (SPEC), f1-score (F1), and Matthews's correlation coefficient (MCC), from the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) cases, which are reported in the confusion matrix of Fig. 46. The AUC for each class, macro, and micro-average of the proposed framework are reported in the ROC curves conventions in Fig. 47. Macro-average focuses on aggregations and totals, representing the arithmetic mean or average of precision values across all classes. In contrast, micro-average delves into individual class assess-

ment. It calculates the sum of all true positives divided by the sum of both true positives and false positives, effectively measuring the ratio of correctly identified predictions to the total number of predictions. Otherwise, Fig. 48 shows the accuracy and loss curves of our model along the training epochs. These curves show a training model converging smoothly and steadily towards weights that allow high precision and low error in training and validation subsets. Besides, in Fig. 47, the AUC of this model is 0.99, which shows a high capacity to distinguish among the eight classes.

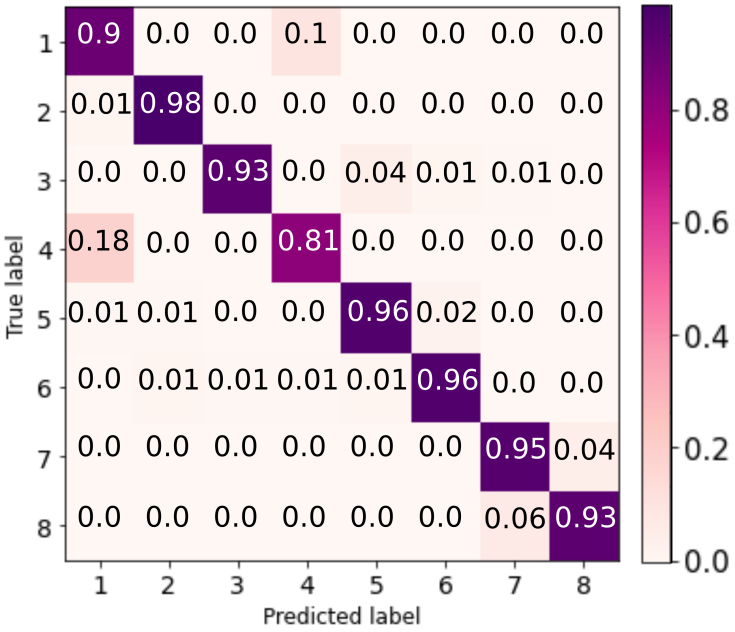


Figure 46. Confusion matrix of the proposed classification approach for the Kvasir-V2 dataset.

6.2.3. Ablation Studies. We evaluate the performance of the proposed pipeline by comparing our method with the baselines based on the same experimental setting. First, we tested five CNNs pre-trained with the ImageNet dataset: DenseNet201, ResNet50, Xception, VGG19, and VGG16. The results obtained are shown in Table 7. We reported the *accuracy* score for each experiment and found the best result using the VGG16 network. Therefore, we selected this

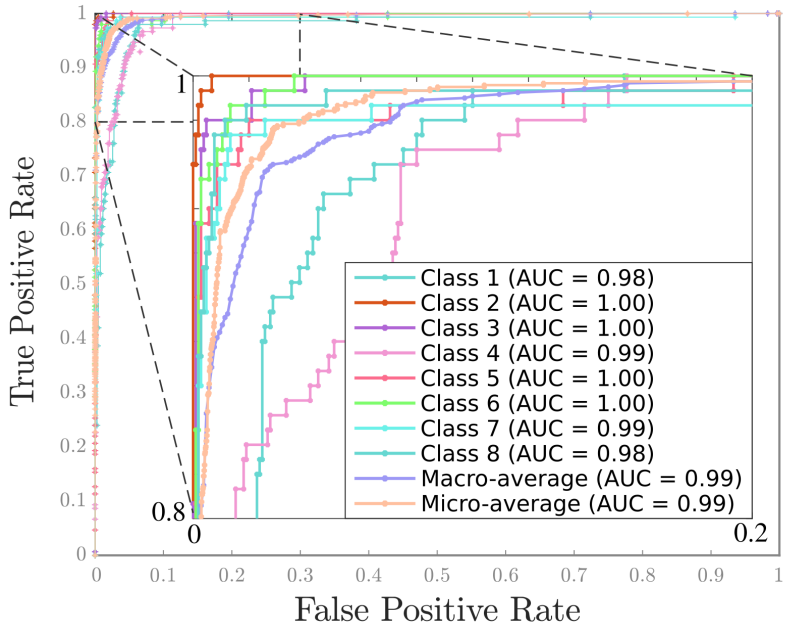


Figure 47. AUC values and ROC curves for each class in the Kvasir-V2 dataset using the proposed classification framework.

Table 22
Accuracy for different tested CNN architectures.

Pretrained CNNs	**Accuracy**
DenseNet201	78.55
ResNet50	90.42
Xception	78.26
VGG19	97.64
VGG16	**98.20**

network to continue with our study.

Once the network was defined, we tested the fine-tuning in the different convolutional blocks of the VGG16 network. As shown in Table 23, the best performance was obtained by training from "block4_conv1". Next, we performed the tests described in Table 24, where we applied the fine-tuning from each of the three internal layers of block 4. We observed that the best

Table 23
Accuracy for Different VGG16 blocks.

Block with Fine-tuning	Trainable Params	Non Trainable Params	Computacional Time [s]	Accuracy
block1_conv1	14,714,688	0	5849.254	0.9764
block2_conv1	14,675,968	38,720	4395.321	0.9781
block3_conv1	14,454,528	260,160	2253.218	0.9792
block4_conv1	12,979,200	1,735,488	2877.035	**0.9808**
block5_conv1	7,079,424	7,635,264	2363.919	0.9803

Table 24
Accuracy for the different layers of block 4.

Layer with Fine-tuning	Trainable Params	Non Trainable Params	Computacional Time [s]	Accuracy
block4_conv1	12,979,200	1,735,488	2877.035	0.9808
block4_conv2	11,799,040	2,915,648	2607.204	**0.9820**
block4_conv3	9,439,232	5,275,456	2462.092	0.9791

accuracy was obtained by training from "block4_conv2". Therefore, in our proposal, we use the VGG16 architecture previously trained with the ImageNet dataset, and the fine-tuning technique is applied by training from "block 4_conv2".

6.2.4. Comparison Results. We compare the performance of our proposed method with existing related methods, which have been reported for endoscopic image classification and described in the first part of this chapter. As shown in Table 25, our method outperforms three of the baseline methods in terms of *accuracy*. The best results are shown in bold font, and the second-best is underlined. Although our method did not outperform the work in (Cogan et al., 2019) by a difference of 0.28%, our method is significantly faster and less computationally complex.

Specifically, the authors in (Cogan et al., 2019) proposed a bilinear architecture that fuses extracted features with both Inception and ResNet networks. This method needs training over 55.8 million parameters to achieve the reported results. On the other hand, our method requires only 11.7 million trainable parameters to achieve very similar accuracy results.

Table 25
Applications of machine learning in GI tract analysis research using the Kvasir-V2 dataset.

Year	Method	Dataset Distribution			ACC	PREC	REC	SPEC	F1	MCC
		Train	Test	Validation						
2017	3 Layer CNN (Pogorelov et al., 2017)	50%	50%	-	0.959	0.589	0.408	0.890	0.453	0.430
2019	Inception-ResNet-v2 (Cogan et al., 2019)	85%	15%	-	**0.9848**	0.940	0.939	**0.991**	0.939	**0.930**
2019	ResNet50 CNN with Transfer Learning (KahsayGebreslassie et al., 2019)	60%	30%	10%	0.878	-	-	-	-	-
2019	ANN with pre-trained CNN feature extractors (Gamage et al., 2019)	80%	20%	-	0.9738	**0.9715**	**0.9727**	-	**0.9721**	-
2021	Proposed Method	80%	15%	5%	0.9820	0.9286	0.9275	0.99	0.9276	0.9173

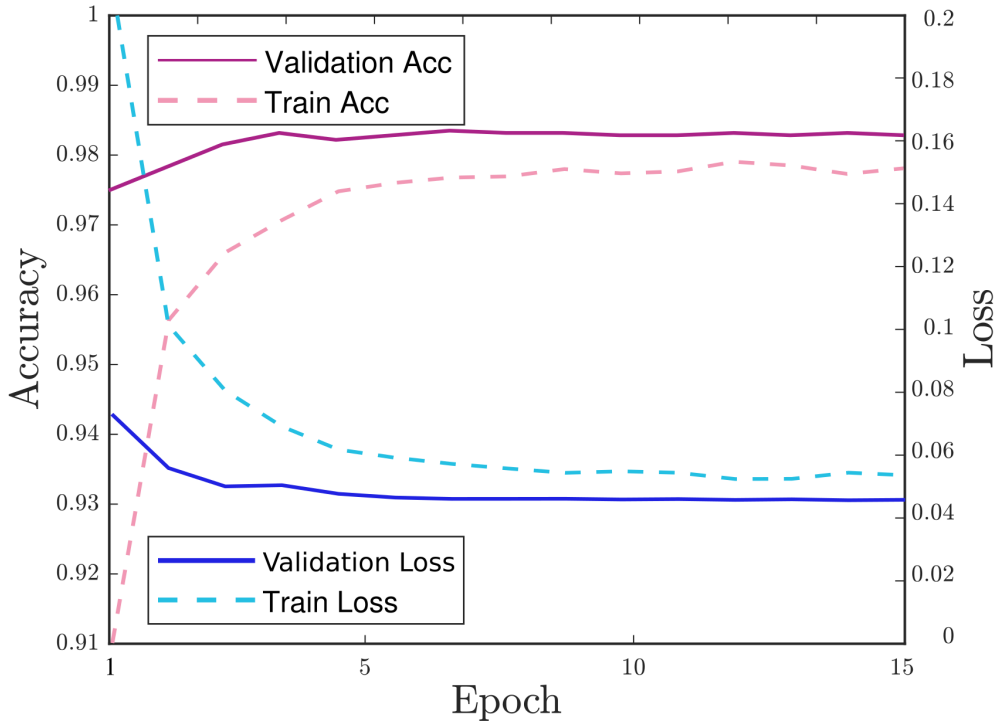


Figure 48. Epochs vs accuracy and loss classification curves

6.3. Conclusions

We studied convolutional neural networks and transferred learning to classify gastrointestinal endoscopic images. We showed that a method based on pre-processing and fine-tuning transfer learning on a light convolutional neural network, such as the VGG-16, could classify eight categories of diseases and abnormalities in the gastrointestinal tract with high precision. Experiments on the Kvasir-V2 dataset show that our proposed framework for endoscopic image processing and classification outperforms other state-of-the-art works. Furthermore, our method achieves an accuracy very similar to the most accurate work of the state-of-the-art, using only a fifth of the number of trainable parameters.

Conclusions

Throughout this PhD research, we have extensively explored the challenges posed by the lack of labeled medical images, the heterogeneous nature of clinical data processing, and the classification problem in computer-aided diagnostic (CAD) tasks. The conclusions and the explorations undertaken are:

Scarcity of Labeled Medical Images: Through developing computational algorithms that utilized deep learning, we have successfully addressed one of the most pressing challenges in the field – the scarcity of labeled medical data. Our novel algorithms for domain adaptation and generative data augmentation have shown promise in overcoming the limitations of dataset sizes, providing a feasible solution to this widespread problem.

Heterogeneous Clinical Data Processing: The methodology developed to arrange and process multiparametric MRIs, was used effectively to feed this data to a deep learning algorithm. We have taken significant steps towards making a CAD system for liver tumor segmentation more accurate through the synthetic image generation of heterogeneous contrast.

Overcoming Overfitting in CAD: The smart selection, transformation, and incorporation of chest X-rays from a public dataset have addressed the challenge of models overfitting to a specific data domain. Our approach enhanced the accuracy of classifying pneumonia in chest X-rays and showcased the methodology's broader applicability for other medical imaging challenges.

Addressing Various Medical Imaging Challenges: Beyond the primary objectives, this research explored the realms of chronic ulcer segmentation in RGB photography, skin lesion clas-

sification, and gastrointestinal anomaly detection in endoscopy photograms. These additional studies underscore the breadth of the challenges in medical imaging and emphasize the importance of diverse approaches to address them.

Comparative Overcome: The evaluation of our algorithms against state-of-the-art approaches showcased the advancements made in the field during this thesis. Through rigorous simulations and real-world verifications with medical specialists, the proposed solutions exhibited superior or competitive performance.

Potential contributions to Healthcare: This thesis has introduced algorithms with the potential for future integration into clinical scenarios for Computer-Aided Diagnostic (CAD) evaluation, although such integration is beyond the scope of this work. The AI applications developed herein offer promising avenues for advancing AI in medical imaging and, consequently, the broader healthcare sector.

Impact of Exogenous Variables: The algorithms presented in this thesis have been evaluated with specific data sets mentioned in each chapter. It is essential to recognize that the performance of these models can be negatively affected when applied to images acquired under different technological, professional, and demographic conditions. The numerical results shown here serve as valuable benchmarks, and practitioners should exercise caution and consider contextual factors when applying these algorithms to diverse data sets. The analysis of the influence of each exogenous variable on the impact of the models could be a topic of future research and exploration.

In retrospect, this PhD thesis has achieved the planned objectives and provided a foundation for potential further studies in the field of AI in medical imaging. The methodologies and algo-

rithms developed here offer a stepping stone for additional applications, contributing to ongoing efforts to enhance healthcare through artificial intelligence.

Compliance with Ethical Standards

Chapter 2: This research was retrospectively conducted by using the human subject data made openly accessible by Kermany et al. (2018). The attached license CC BY 4.0 confirms that ethical approval was not required. The Toulouse data acquisition of the small dataset and its use adhere to the Declaration of Helsinki principles. The Hospital of Toulouse in France granted the necessary ethical approval.

Chapter 3: The MRI dataset for this chapter was provided privately by the French Society of Radiology, specifically for participants of the 2021 Data Augmentation Challenge. As this is a numerical simulation-based study, ethical approval was not required.

Chapter 4: This research employed the open-access human subject data from the Chronic wounds database as provided by (Krecichwost et al., 2021). The attached license CC BY 4.0 verifies that ethical approval was not needed. The CO2Wounds database acquisition and its use are in alignment with the Declaration of Helsinki principles. Ethical approval was obtained both from the Ethics Committee of the *Sanatorio de Contratación ESE* hospital (Act 05-21) and the Scientific Research Ethics Committee of the Universidad Industrial de Santander (Act No. 19-13/11/20).

Chapter 5: This numerical simulation study used the open-access HAM10000 data collection, which received approval from the ethics review committee at both the Medical University of Vienna and the University of Queensland, as declared by (Tschandl et al., 2018).

Chapter 6: This study, based on numerical simulations, utilized the Kvasir dataset (Pogorelov et al., 2017). It was exempt from ethical approval as the dataset is sanctioned for research and educational use.

Credit author statement

This section presents the Contributor Roles Taxonomy (CRediT) of each author involved in the journal articles and conferences that are products related to this doctoral thesis. Authors are listed in the same order as they appear in academic publications.

Chapter 2 and 3, identically: **Karen Sanchez:** *PhD student.* Conceptualization, Methodology, Visualization, Data Curation, Software, Writing - Original Draft. **Carlos Hinojosa:** *PhD student.* Software, Validation. **Henry Arguello:** *UIS Professor*, Supervision, Project administration. **Denis Kouamé:** *U Toulouse Professor*, Supervision, Writing - Review & Editing. **Olivier Meyrignac:** *MD-PhD Radiology in Paris*, Resources and Validation. **Adrian Basarab:** *U de Lyon Professor*. Supervision, Writing - Review & Editing, Project administration.

Chapter 4: **Karen Sanchez:** *PhD student*, Project administration, Conceptualization, Methodology, Validation, Data Curation, Visualization, Formal analysis, Resources, Writing - Original Draft, Writing - Review & Editing. **Brayan Monroy:** *Master student*, Software, Visualization, Writing - Original Draft. **Paula Arguello:** *Bachelor student*, Software, Visualization, Writing - Original Draft. **Juan Estupiñán:** *Bachelor student*, Validation, Data Curation, Writing - Original Draft. **Jorge Bacca:** *PhD UIS*, Supervision, Writing - Original Draft. **Claudia V Correa:** *PhD UIS*, Methodology, Conceptualization, Formal analysis, Writing - Original Draft, Writing - Review & Editing. **Laura Valencia:** *MD. UIS*, Conceptualization.

Juan C Castillo: *MD. UIS*, Conceptualization. **Olinto Mieles:** *MD. Contratación*, Conceptualization, Resources, Data Curation. **Henry Arguello:** *UIS Professor*, Supervision. **Sergio Castillo:** *UIS Professor*, Supervision, Funding acquisition. **Fernando Rojas-Morales:** *UIS Professor*, Supervision.

Chapter 5: **Camilo Calderón:** *Bachelor student*, Software, Writing - Original Draft. **Karen Sanchez:** *PhD student*, Conceptualization, Methodology, Validation, Software, Writing - Review & Editing. **Sergio Castillo:** *UIS Professor*, Supervision. **Henry Arguello:** *UIS Professor*, Supervision.

Chapter 6: **Jessica Escobar:** *Bachelor student*, Software, Writing - Original Draft. **Karen Sanchez:** *PhD student*, Conceptualization, Methodology, Software, Writing - Review & Editing. **Carlos Hinojosa:** *PhD student*, Conceptualization, Validation, Formal analysis, Software. **Henry Arguello:** *UIS Professor*, Supervision. **Sergio Castillo:** *UIS Professor*, Supervision.

Bibliography

Abdeltawab, H., Khalifa, F., Taher, F., Alghamdi, N. S., Ghazal, M., Beache, G., Mohamed, T., Keynton, R., and El-Baz, A. (2020). A deep learning-based approach for automatic segmentation and quantification of the left ventricle from cardiac cine mr images. In *Computerized Medical Imaging and Graphics*, volume 81, page 101717. Elsevier.

Adegun, A. and Viriri, S. (2021). Deep learning techniques for skin lesion analysis and melanoma cancer detection: a survey of state-of-the-art. In *Artificial Intelligence Review*, volume 54, pages 811–841. Springer.

Ahmad, H. A., Yu, H. J., and Miller, C. G. (2014). Medical imaging modalities. In *Medical Imaging in Clinical Trials*, pages 3–26. Springer.

Alzubaidi, L., Fadhel, M. A., Al-Shamma, O., Zhang, J., Santamaría, J., Duan, Y., and Oleiwi, S. R. (2020). Towards a better understanding of transfer learning for medical imaging: A case study. *Applied Sciences*, 10(13):4523.

Anisuzzaman, D. M., Patel, Y., Niezgoda, J., Gopalakrishnan, S., and Yu, Z. (2020). A mobile app for wound localization using deep learning. *CoRR*, abs/2009.07133.

Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J. S., Freymann, J. B., Farahani, K., and Davatzikos, C. (2017). Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data*, 4(1):1–13.

Barbedo, J. G. A. (2018). Impact of dataset size and variety on the effectiveness of deep learning and transfer learning for plant disease classification. 153:46–53.

Bochkovskiy, A., Wang, C.-Y., and Liao, H.-Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*.

Bowers, S. and Franco, E. (2020). Chronic wounds: evaluation and management. *American family physician*, 101(3):159–166.

Calderón, C., Sanchez, K., Castillo, S., and Arguello, H. (2021). Bilsk: A bilinear convolutional neural network approach for skin lesion classification. In *Computer Methods and Programs in Biomedicine Update*, volume 1, page 100036. Elsevier.

Chairat, S., Dissaneewate, T., Wangkulangkul, P., Kongpanichakul, L., and Chaichulee, S. (2021). Non-contact chronic wound analysis using deep learning. In *2021 13th Biomedical Engineering International Conference (BMEiCON)*, pages 1–5. IEEE.

Chaturvedi, S. S., Gupta, K., and Prasad, P. S. (2020). Skin lesion analyser: An efficient seven-way multi-class skin cancer classification using mobilenet. In *International Conference on Advanced Machine Learning Technologies and Applications*, pages 165–176. Springer.

Chaurasia, A. and Culurciello, E. (2017). Linknet: Exploiting encoder representations for efficient semantic segmentation. In *2017 IEEE visual communications and image processing (VCIP)*, pages 1–4. IEEE.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., and Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818.

Cheng, J.-Z., Ni, D., Chou, Y.-H., Qin, J., Tiu, C.-M., Chang, Y.-C., Huang, C.-S., Shen, D., and Chen, C.-M. (2016). Computer-aided diagnosis with deep learning architecture: applications to breast lesions in us images and pulmonary nodules in ct scans. In *Scientific reports*, volume 6, pages 1–13. Nature Publishing Group.

Chollet, F. (2017a). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258.

Chollet, F. (2017b). Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258.

Chopade, A. (2020). Seven class classification of skin lesions by using incremental convolutional neural network in python. In *Available at SSRN 3574612*.

Choudhary, A., Tong, L., Zhu, Y., and Wang, M. D. (2020). Advancing medical imaging informatics by deep learning-based domain adaptation. In *Yearbook of medical informatics*, volume 29, page 129. Thieme Medical Publishers.

Chouhan, V., Singh, S. K., Khamparia, A., Gupta, D., Tiwari, P., Moreira, C., Damaševičius, R., and De Albuquerque, V. H. C. (2020). A novel transfer learning based approach for pneumonia

detection in chest x-ray images. In *Applied Sciences*, volume 10, page 559. Multidisciplinary Digital Publishing Institute.

Codella, N., Rotemberg, V., Tschandl, P., Celebi, M. E., Dusza, S., Gutman, D., Helba, B., Kalloo, A., Liopyris, K., Marchetti, M., et al. (2019). Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). *arXiv preprint arXiv:1902.03368*.

Cogan, T., Cogan, M., and Tamil, L. (2019). Mapgi: Accurate identification of anatomical landmarks and diseased tissue in gastrointestinal tract using deep learning. In *Computers in biology and medicine*, volume 111, page 103351. Elsevier.

Cohen, P. R., Erickson, C. P., and Calame, A. (2019). Atrophic dermatofibroma: A comprehensive literature review. In *Dermatology and therapy*, pages 1–20. Springer.

CVAT.ai Corporation (2022). Computer Vision Annotation Tool (CVAT).

Dai, J., Lian, X., Mo, J., Li, X., Mo, W., Wang, H., and Jiang, J. (2023). Case report: A clinical case study of six patients with chlamydia psittaci pneumonia. *Frontiers in Cellular and Infection Microbiology*, 13:1084882.

Damsky, W. and Bosenberg, M. (2017). Melanocytic nevi and melanoma: Unraveling a complex relationship. In *Oncogene*, volume 36, pages 5771–5792. Nature Publishing Group.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-

scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee.

Diker, A., Cömert, Z., Avcı, E., Toğaçar, M., and Ergen, B. (2019). A novel application based on spectrogram and convolutional neural network for ecg classification. In *2019 1st International Informatics and Software Engineering Conference (UBMYK)*, pages 1–6. IEEE.

Doi, K. (2007). Computer-aided diagnosis in medical imaging: historical review, current status and future potential. In *Computerized medical imaging and graphics*, volume 31, pages 198–211. Elsevier.

El-Naqa, I., Yang, Y., Wernick, M. N., Galatsanos, N. P., and Nishikawa, R. M. (2002). A support vector machine approach for detection of microcalcifications. In *IEEE transactions on medical imaging*, volume 21, pages 1552–1563. IEEE.

Emara, T., Afify, H. M., Ismail, F. H., and Hassanien, A. E. (2019). A modified inception-v4 for imbalanced skin cancer classification dataset. In *2019 14th International Conference on Computer Engineering and Systems (ICCES)*, pages 28–33. IEEE.

Erickson, B. J., Korfiatis, P., Akkus, Z., and Kline, T. L. (2017). Machine learning for medical imaging. In *Radiographics*, volume 37, pages 505–515. Radiological Society of North America.

Escobar, J., Sanchez, K., Hinojosa, C., Arguello, H., and Castillo, S. (2021). Accurate deep learning-based gastrointestinal disease classification via transfer learning strategy. In *2021 XXIII Symposium on Image, Signal Processing and Artificial Vision (STSIVA)*, pages 1–5.

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., and Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. In *nature*, volume 542, pages 115–118. Nature Publishing Group.

Fauzi, M. F. A., Khansa, I., Catignani, K., Gordillo, G., Sen, C. K., and Gurcan, M. N. (2015). Computerized segmentation and measurement of chronic wound images. *Computers in biology and medicine*, 60:74–85.

Fernandez, V., Pinaya, W. H. L., Borges, P., Tudosiu, P.-D., Graham, M. S., Vercauteren, T., and Cardoso, M. J. (2022). Can segmentation models be trained with fully synthetically generated data? In *Simulation and Synthesis in Medical Imaging: 7th International Workshop, SASHIMI 2022, Held in Conjunction with MICCAI 2022, Singapore, September 18, 2022, Proceedings*, pages 79–90. Springer.

Frid-Adar, M., Klang, E., Amitai, M., Goldberger, J., and Greenspan, H. (2018). Synthetic data augmentation using gan for improved liver lesion classification. In *2018 IEEE 15th international symposium on biomedical imaging ISBI*, pages 289–293. IEEE.

Gamage, C., Wijesinghe, I., Chitraranjan, C., and Perera, I. (2019). Gi-net: Anomalies classification in gastrointestinal tract through endoscopic imagery with deep learning. In *2019 Moratuwa Engineering Research Conference (MERCon)*, pages 66–71. IEEE.

Gan, M. and Wang, C. (2022). Esophageal optical coherence tomography image synthesis using an adversarially learned variational autoencoder. *Biomedical Optics Express*, 13(3):1188–1201.

Garg, R., Maheshwari, S., and Shukla, A. (2019). Decision support system for detection and classification of skin cancer using cnn. In *arXiv preprint arXiv:1912.03798*.

Göçeri, E. (2019). Analysis of deep networks with residual blocks and different activation functions: classification of skin diseases. In *2019 Ninth international conference on image processing theory, tools and applications (IPTA)*, pages 1–6. IEEE.

Göçeri, E. (2020a). Capsnet topology to classify tumours from brain images and comparative evaluation. In *IET Image Processing*, volume 14, pages 882–889. The Institution of Engineering and Technology.

Göçeri, E. (2020b). Convolutional neural network based desktop applications to classify dermatological diseases. In *2020 IEEE 4th International Conference on Image Processing, Applications and Systems (IPAS)*, pages 138–143. IEEE.

Göçeri, E. (2020c). Image augmentation for deep learning based lesion classification from skin images. In *2020 IEEE 4th International Conference on Image Processing, Applications and Systems (IPAS)*, pages 144–148. IEEE.

Göçeri, E. (2021). Automated skin cancer detection: Where we are and the way to the future. In *2021 44th International Conference on Telecommunications and Signal Processing (TSP)*, pages 48–51. IEEE.

Goceri, E. (2021). Deep learning based classification of facial dermatological disorders. In *Computers in Biology and Medicine*, volume 128, page 104118. Elsevier.

Göçeri, E. (2021). Diagnosis of skin diseases in the era of deep learning and mobile technology. In *Computers in Biology and Medicine*, volume 134, page 104458. Elsevier.

Goldenberg, G. and Perl, M. (2014). Actinic keratosis: Update on field therapy. In *The Journal of clinical and aesthetic dermatology*, volume 7, page 28. Matrix Medical Communications.

Goyal, M., Yap, M. H., Reeves, N. D., Rajbhandari, S., and Spragg, J. (2017). Fully convolutional networks for diabetic foot ulcer segmentation. In *2017 IEEE international conference on systems, man, and cybernetics (SMC)*, pages 618–623. IEEE.

Gupta, S., Sagar, S., Maheshwari, G., Kisaka, T., and Tripathi, S. (2021). Chronic wounds: Magnitude, socioeconomic burden and consequences. *Wounds Asia*, 4:8–14.

Haggenmüller, S., Maron, R. C., Hekler, A., Utikal, J. S., Barata, C., Barnhill, R. L., Beltraminelli, H., Berking, C., Betz-Stablein, B., Blum, A., et al. (2021). Skin cancer classification via convolutional neural networks: systematic review of studies involving human experts. In *European Journal of Cancer*, volume 156, pages 202–216. Elsevier.

Hashmi, M. F., Katiyar, S., Keskar, A. G., Bokde, N. D., and Geem, Z. W. (2020). Efficient pneumonia detection in chest xray images using deep transfer learning. In *Diagnostics*, volume 10, page 417. Multidisciplinary Digital Publishing Institute.

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 37(9):1904–1916.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30.

Hinojosa, C., Sanchez, K., Garcia, H., and Arguello, H. (2022). C-3spcd: coded aperture similarity constrained design for spatio-spectral classification of single-pixel measurements. *Applied Optics*, 61(8):E21–E32.

Hossin, M. and Sulaiman, M. (2015). A review on evaluation metrics for data classification evaluations. In *International Journal of Data Mining & Knowledge Management Process*, volume 5, page 1. Academy & Industry Research Collaboration Center (AIRCC).

Hsu, J.-T., Chen, Y.-W., Ho, T.-W., Tai, H.-C., Wu, J.-M., Sun, H.-Y., Hung, C.-S., Zeng, Y.-C., Kuo, S.-Y., and Lai, F. (2019). Chronic wound assessment and infection detection method. *BMC medical informatics and decision making*, 19(1):1–20.

Htay, T. T. and Maung, S. S. (2018). Early stage breast cancer detection system using glcm feature extraction and k-nearest neighbor (k-nn) on mammography image. In *2018 18th International Symposium on Communications and Information Technologies (ISCIT)*, pages 171–175. IEEE.

Huo, J., Vakharia, V., Wu, C., Sharan, A., Ko, A., Ourselin, S., and Sparks, R. (2022). Brain

lesion synthesis via progressive adversarial variational auto-encoder. In *Simulation and Synthesis in Medical Imaging: 7th International Workshop, SASHIMI 2022, Held in Conjunction with MICCAI 2022, Singapore, September 18, 2022, Proceedings*, pages 101–111. Springer.

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134.

Jack Jr, C. R., Bernstein, M. A., Fox, N. C., Thompson, P., Alexander, G., Harvey, D., Borowski, B., Britson, P. J., L. Whitwell, J., Ward, C., et al. (2008). The alzheimer's disease neuroimaging initiative (adni): Mri methods. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 27(4):685–691.

Jiang, J., Hu, Y.-C., Tyagi, N., Zhang, P., Rimner, A., Deasy, J. O., and Veeraraghavan, H. (2019). Cross-modality (ct-mri) prior augmented deep learning for robust lung tumor segmentation from small mr datasets. *Medical physics*, 46(10):4392–4404.

Jiang, Y., Chen, H., Loew, M., and Ko, H. (2020). Covid-19 ct image synthesis with a conditional generative adversarial network. *IEEE Journal of Biomedical and Health Informatics*, 25(2):441–452.

Johnson, J. M. and Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, 6(1):1–54.

KahsayGebreslassie, A., Hagos, M. T., et al. (2019). Automated gastrointestinal disease recogni-

tion for endoscopic images. In *2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, pages 312–316. IEEE.

Kebaili, A., Lapuyade-Lahorgue, J., and Ruan, S. (2023). Deep learning approaches for data augmentation in medical imaging: A review. *Journal of Imaging*, 9(4):81.

Kermany, D., Zhang, K., and Goldbaum, M. (2018). Labeled optical coherence tomography (oct) and chest x-ray images for classification. In *Mendeley data*.

Kervadec, H., Dolz, J., Yuan, J., Desrosiers, C., Granger, E., and Ayed, I. B. (2019). Constrained deep networks: Lagrangian optimization via log-barrier extensions. In *arXiv preprint arXiv:1904.04205*.

Khader, F., Mueller-Franzes, G., Arasteh, S. T., Han, T., Haarburger, C., Schulze-Hagen, M., Schad, P., Engelhardt, S., Baessler, B., Foersch, S., et al. (2022). Medical diffusion–denoising diffusion probabilistic models for 3d medical image generation. *arXiv preprint arXiv:2211.03364*.

Kooi, T., Litjens, G., Van Ginneken, B., Gubern-Mérida, A., Sánchez, C. I., Mann, R., den Heeten, A., and Karssemeijer, N. (2017). Large scale deep learning for computer aided detection of mammographic lesions. In *Medical image analysis*, volume 35, pages 303–312. Elsevier.

Krecichwost, M., Czajkowska, J., Wijata, A., Juszczyk, J., Pyciński, B., Biesok, M., Rudzki, M., Majewski, J., Kostecki, J., and Pietka, E. (2021). Chronic wounds multimodal image database. *Computerized Medical Imaging and Graphics*, 88:101844.

Lafarge, M. W., Pluim, J. P., Eppenhof, K. A., Moeskops, P., and Veta, M. (2017). Domain-adversarial neural networks to address the appearance variability of histopathology images. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3*, pages 83–91. Springer.

Lee, J.-G., Jun, S., Cho, Y.-W., Lee, H., Kim, G. B., Seo, J. B., and Kim, N. (2017). Deep learning in medical imaging: general overview. In *Korean journal of radiology*, volume 18, page 570. Korean Society of Radiology.

Liang, J. and Chen, J. (2021). Data augmentation of thyroid ultrasound images using generative adversarial network. In *2021 IEEE International Ultrasonics Symposium (IUS)*, pages 1–4. IEEE.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B., and Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. In *Medical image analysis*, volume 42, pages 60–88. Elsevier.

Liu, S., Qi, L., Qin, H., Shi, J., and Jia, J. (2018). Path aggregation network for instance segmenta-

tion. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8759–8768.

Liu, X., Wang, C., Li, F., Zhao, X., Zhu, E., and Peng, Y. (2017). A framework of wound segmentation based on deep convolutional networks. In *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pages 1–7. IEEE.

Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440.

Marquez, M., Monsalve, J., Arias, K., Sanchez, K., Hinojosa, C., and Arguello, H. (2022). Hierarchical compressed subspace clustering of infrared single-pixel measurements. In *2022 12th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pages 1–6. IEEE.

Miglani, V. and Bhatia, M. (2020). Skin lesion classification: A transfer learning approach using efficientnets. In *International Conference on Advanced Machine Learning Technologies and Applications*, pages 315–324. Springer.

Miller, A. J. and Mihm Jr, M. C. (2006). Melanoma. In *New England Journal of Medicine*, volume 355, pages 51–65. Mass Medical Soc.

Mohapatra, S., Abhishek, N., Bardhan, D., Ghosh, A. A., and Mohanty, S. (2020). Skin cancer

classification using convolution neural networks. In *Advances in Distributed Computing and Machine Learning*, pages 433–442. Springer.

Mok, T. C. and Chung, A. C. (2019). Learning data augmentation for brain tumor segmentation with coarse-to-fine generative adversarial networks. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part I 4*, pages 70–80. Springer.

Monroy, B., Bacca, J., Sanchez, K., Arguello, H., and Castillo, S. (2021). Two-step deep learning framework for chronic wounds detection and segmentation: A case study in colombia. In *2021 XXIII Symposium on Image, Signal Processing and Artificial Vision (STSIVA)*, pages 1–6. IEEE.

Monsalve, J., Marquez, M., Sanchez, K., Hinojosa, C., Esnaola, I., and Arguello, H. (2022). Cocosvi: Single snapshot compressive spectral video via covariance matrix estimation. In *2022 12th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pages 1–6. IEEE.

Mukherjee, R., Tewary, S., and Routray, A. (2017). Diagnostic and prognostic utility of non-invasive multimodal imaging in chronic wound monitoring: a systematic review. *Journal of medical systems*, 41(3):1–17.

Mulé, S., Lawrance, L., Belkouchi, Y., Vilgrain, V., Lewin, M., Trillaud, H., Hoeffel, C., Laurent, V., Ammari, S., Morand, E., et al. (2022). Generative adversarial networks (gan)-based data

augmentation of rare liver cancers: The sfr 2021 artificial intelligence data challenge. *Diagnostic and Interventional Imaging*.

Neff, T., Payer, C., Stern, D., and Urschler, M. (2017). Generative adversarial network based synthesis for supervised medical image segmentation. In *Proc. OAGM and ARW joint Workshop*, volume 3, page 4.

Pan, Y., Liu, M., Lian, C., Zhou, T., Xia, Y., and Shen, D. (2018). Synthesizing missing pet from mri with cycle-consistent generative adversarial networks for alzheimer's disease diagnosis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 455–463. Springer.

Pang, X., Zhao, Z., and Weng, Y. (2021). The role and impact of deep learning methods in computer-aided diagnosis using gastrointestinal endoscopy. *Diagnostics*, 11(4):694.

Platscher, M., Zopes, J., and Federau, C. (2020). Image translation for medical image generation–ischemic stroke lesions. *arXiv preprint arXiv:2010.02745*.

Platscher, M., Zopes, J., and Federau, C. (2022). Image translation for medical image generation: Ischemic stroke lesion segmentation. *Biomedical Signal Processing and Control*, 72:103283.

Pogorelov, K., Randel, K. R., Griwodz, C., Eskeland, S. L., de Lange, T., Johansen, D., Spampinato, C., Dang-Nguyen, D.-T., Lux, M., Schmidt, P. T., et al. (2017). Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection. In *Proceedings of the 8th ACM on Multimedia Systems Conference*, pages 164–169.

Qasim, A. B., Ezhov, I., Shit, S., Schoppe, O., Paetzold, J. C., Sekuboyina, A., Kofler, F., Lipkova, J., Li, H., and Menze, B. (2020). Red-gan: Attacking class imbalance via conditioned generation. yet another medical imaging perspective. In *Medical Imaging with Deep Learning*, pages 655–668. PMLR.

Rahman, T., Chowdhury, M. E., Khandakar, A., Islam, K. R., Islam, K. F., Mahbub, Z. B., Kadir, M. A., and Kashem, S. (2020). Transfer learning with deep convolutional neural network (cnn) for pneumonia detection using chest x-ray. In *Applied Sciences*, volume 10, page 3233. Multidisciplinary Digital Publishing Institute.

Raman, R., Srinivasan, S., Virmani, S., Sivaprasad, S., Rao, C., and Rajalakshmi, R. (2019). Fundus photograph-based deep learning algorithms in detecting diabetic retinopathy. In *Eye*, volume 33, pages 97–109. Nature Publishing Group.

Ravì, D., Wong, C., Deligianni, F., Berthelot, M., Andreu-Perez, J., Lo, B., and Yang, G.-Z. (2016). Deep learning for health informatics. In *IEEE journal of biomedical and health informatics*, volume 21, pages 4–21. IEEE.

Redmon, J. and Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.

Riaz, F., Rehman, S., Ajmal, M., Hafiz, R., Hassan, A., Aljohani, N. R., Nawaz, R., Young, R., and Coimbra, M. (2020). Gaussian mixture model based probabilistic modeling of images for medical image segmentation. In *IEEE Access*, volume 8, pages 16846–16856. IEEE.

Roh, N. K., Hahn, H. J., Lee, Y. W., Choe, Y. B., and Ahn, K. J. (2016). Clinical and histopathological investigation of seborrheic keratosis. In *Annals of dermatology*, volume 28, pages 152–158.

Rong, W., Li, Z., Zhang, W., and Sun, L. (2014). An improved canny edge detection algorithm. In *2014 IEEE international conference on mechatronics and automation*, pages 577–582. IEEE.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.

Rubin, A. I., Chen, E. H., and Ratner, D. (2005). Basal-cell carcinoma. In *New England Journal of Medicine*, volume 353, pages 2262–2269. Mass Medical Soc.

Sachdeva, S. et al. (2009). Fitzpatrick skin typing: Applications in dermatology. *Indian journal of dermatology, venereology and leprology*, 75(1):93.

Sanchez, K., Hinojosa, C., Arguello, H., Freiss, S., Sans, N., Kouamé, D., Meyrignac, O., and Basarab, A. (2021a). Subspace-based domain adaptation using similarity constraints for pneumonia diagnosis within a small chest x-ray image dataset. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 1232–1235. IEEE.

Sanchez, K., Hinojosa, C., Arguello, H., Kouamé, D., Meyrignac, O., and Basarab, A. (2022). Cx-dagan: Domain adaptation for pneumonia diagnosis on a small chest x-ray dataset. In *IEEE Transactions on Medical Imaging*. IEEE.

Sanchez, K., Hinojosa, C., Arias, K., Arguello, H., Kouamé, D., Meyrignac, O., and Basarab, A. (2023a). Mask-guided data augmentation for multiparametric mri generation with a rare hepatocellular carcinoma. In *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, pages 1–5. IEEE.

Sanchez, K., Hinojosa, C., Garcia, H., Arguello, H., and Castillo, S. (2021b). Compressed-domain classification algorithm for spectral imaging based on designed single-pixel camera codification. In *OSA Imaging and Applied Optics Congress 2021 (3D, COSI, DH, ISA, pcAOP)*, page CTu2F.5. Optical Society of America.

Sanchez, K., Monroy, B., Arguello, P., Estupiñán, J., Bacca, J., Correa, C. V., Valencia, L., Castillo, J. C., Mieles, O., Arguello, H., Castillo, S., and Rojas-Morales, F. (2023b). Automated chronic wounds medical assessment and tracking framework based on deep learning. In *Computers in Biology and Medicine*. Elsevier.

Sandfort, V., Yan, K., Pickhardt, P. J., and Summers, R. M. (2019). Data augmentation using generative adversarial networks (cyclegan) to improve generalizability in ct segmentation tasks. *Scientific reports*, 9(1):1–9.

Scebba, G., Zhang, J., Catanzaro, S., Mihai, C., Distler, O., Berli, M., and Karlen, W. (2022). Detect-and-segment: a deep learning approach to automate wound image segmentation. *Informatics in Medicine Unlocked*, 29:100884.

Shen, C., Nguyen, D., Zhou, Z., Jiang, S. B., Dong, B., and Jia, X. (2020). An introduction to deep

learning in medical physics: advantages, potential, and challenges. volume 65, page 05TR01. IOP Publishing.

Shen, Z., Ouyang, X., Xiao, B., Cheng, J.-Z., Shen, D., and Wang, Q. (2023). Image synthesis with disentangled attributes for chest x-ray nodule augmentation and detection. *Medical Image Analysis*, 84:102708.

Shi, H., Lu, J., and Zhou, Q. (2020). A novel data augmentation method using style-based gan for robust pulmonary nodule segmentation. In *2020 Chinese Control and Decision Conference (CCDC)*, pages 2486–2491. IEEE.

Shin, H.-C., Tenenholtz, N. A., Rogers, J. K., Schwarz, C. G., Senjem, M. L., Gunter, J. L., Andriole, K. P., and Michalski, M. (2018). Medical image synthesis for data augmentation and anonymization using generative adversarial networks. In *Simulation and Synthesis in Medical Imaging: Third International Workshop, SASHIMI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 3*, pages 1–11. Springer.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. In *arXiv preprint arXiv:1409.1556*.

Song, B. and Sacan, A. (2012). Automated wound identification system based on image segmentation and artificial neural networks. In *2012 IEEE International Conference on Bioinformatics and Biomedicine*, pages 1–4. IEEE.

Suetens, P. (2017). *Fundamentals of medical imaging*. Cambridge university press.

Syed, N. M. et al. (2016). Vascular lesions of head and neck: A literature review. In *Indian Journal of dental sciences*, volume 8, page 176. Medknow Publications.

Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., and Liang, J. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning? In *IEEE transactions on medical imaging*, volume 35, pages 1299–1312. IEEE.

Tanaka, M. (2020). Weighted sigmoid gate unit for an activation function of deep neural network. In *Pattern Recognition Letters*, volume 135, pages 354–359. Elsevier.

Tang, Y., Tang, Y., Sandfort, V., Xiao, J., and Summers, R. M. (2019). Tuna-net: Task-oriented unsupervised adversarial network for disease recognition in cross-domain chest x-rays. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 431–440. Springer.

Toğaçar, M., Ergen, B., Cömert, Z., and Özyurt, F. (2020). A deep feature learning model for pneumonia detection applying a combination of mrmr feature selection and machine learning models. In *IRBM*, volume 41, pages 212–222. Elsevier.

Tschandl, P., Rosendahl, C., and Kittler, H. (2018). The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. In *Scientific data*, volume 5, page 180161. Nature Publishing Group.

Ustinova, E., Ganin, Y., and Lempitsky, V. (2017). Multi-region bilinear convolutional neural

networks for person re-identification. In *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–6. IEEE.

Van Noord, N. and Postma, E. (2017). Learning scale-variant and scale-invariant features for deep image classification. *Pattern Recognition*, 61:583–592.

Wang, C., Anisuzzaman, D., Williamson, V., Dhar, M. K., Rostami, B., Niezgoda, J., Gopalakrishnan, S., and Yu, Z. (2020a). Fully automatic wound segmentation with deep convolutional neural networks. *Scientific Reports*, 10(1):1–9.

Wang, C., Yan, X., Smith, M., Kochhar, K., Rubin, M., Warren, S. M., Wrobel, J., and Lee, H. (2015). A unified framework for automatic wound segmentation and analysis with deep convolutional neural networks. In *2015 37th annual international conference of the ieee engineering in medicine and biology society (EMBC)*, pages 2415–2418. IEEE.

Wang, C.-Y., Bochkovskiy, A., and Liao, H.-Y. M. (2021). Scaled-YOLOv4: Scaling cross stage partial network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13029–13038.

Wang, C.-Y., Liao, H.-Y. M., Wu, Y.-H., Chen, P.-Y., Hsieh, J.-W., and Yeh, I.-H. (2020b). Cspnet: A new backbone that can enhance learning capability of cnn. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 390–391.

Wannous, H., Lucas, Y., Treuillet, S., Mansouri, A., and Voisin, Y. (2012). Improving color

correction across camera and illumination changes by contextual sample selection. *Journal of Electronic Imaging*, 21(2):023015.

Wolleb, J., Sandkühler, R., Bieder, F., and Cattin, P. C. (2022). The swiss army knife for image-to-image translation: Multi-task diffusion models. *arXiv preprint arXiv:2204.02641*.

Yazdanpanah, L., Nasiri, M., and Adarvishi, S. (2015). Literature review on the management of diabetic foot ulcer. *World journal of diabetes*, 6(1):37.

Yu, L., Yang, X., Chen, H., Qin, J., and Heng, P. A. (2017). Volumetric convnets with mixed residual connections for automated prostate segmentation from 3d mr images. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31.

Yu, Y., Adu, K., Tashi, N., Anokye, P., Wang, X., and Ayidzoe, M. A. (2020). Rmaf: Relu-memristor-like activation function for deep learning. In *IEEE Access*, volume 8, pages 72727–72741. IEEE.

Zhang, Y., Niu, S., Qiu, Z., Wei, Y., Zhao, P., Yao, J., Huang, J., Wu, Q., and Tan, M. (2020a). Covid-da: Deep domain adaptation from typical pneumonia to covid-19. In *arXiv preprint arXiv:2005.01577*.

Zhang, Y., Wei, Y., Wu, Q., Zhao, P., Niu, S., Huang, J., and Tan, M. (2020b). Collaborative unsupervised domain adaptation for medical image diagnosis. *IEEE Transactions on Image Processing*, 29:7834–7844.

Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J. (2017). Pyramid scene parsing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2881–2890.

Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., and Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*, pages 3–11. Springer.

Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232.