Universidad de Guanajuato
Campus León

DIVISIÓN DE CIENCIAS E INGENIERÍAS

AI-Driven Deep Learning Models for Enhanced Medical Image Diagnosis

A

T H E S I S

SUBMITTED FOR THE DEGREE OF:

DOCTOR EN FÍSICA

BY:

ARON HERNANDEZ TRINIDAD

SUPERVISED BY:

DR. TEODORO CORDOVA FRAGA
DR. RAFAEL GUZMAN CABRERA

Leon, GTO, 2024

Dedicated to my parents and grandmother. To those who are and are no longer.

Acknowledgments

I would like to express my sincere gratitude to Dr. Teodoro Córdova Fraga and Dr. Rafael Guzmán Cabrera for their expert guidance and invaluable support throughout the research process. I also appreciate the administrative staff of the Department of Physics for their assistance and collaboration. This work would not have been possible without the generous funding provided by CONAHCyT scholarship CVU 893699. I also thank my research colleagues for their stimulating ideas and discussions. Last but not least, I want to express my profound gratitude to my family and friends for their love, patience, and constant encouragement throughout this journey.

Abstract

Computer-aided diagnosis (CAD) using Artificial Intelligence (AI) has emerged as a powerful tool to improve accuracy and efficiency in medical image interpretation. This thesis investigates AI techniques for diagnosing various pathologies through medical imaging, covering a broad spectrum of diseases, from tuberculosis to post-COVID conditions.

The primary objective of this study was to develop and evaluate AI algorithms for the automatic detection and classification of diseases using medical images. MRI datasets, chest X-rays, and other imaging modalities were employed to train and test the proposed AI models. Various deep learning techniques, such as convolutional neural networks, were implemented to perform classification and detection tasks on medical images.

The results demonstrate a high degree of accuracy and sensitivity in disease detection, often surpassing the performance of traditional diagnostic methods. A significant improvement in early disease detection was observed, which could lead to more timely treatment and better patient outcomes.

This research has significant clinical implications and highlights AI's potential to contribute to medical diagnosis. Current limitations are discussed, and future research directions are proposed in this exciting and promising field of study.

Contents

List of Figures

Preface

Medicine has witnessed significant advances throughout history, transforming how we diagnose and treat diseases. In the current era, the convergence of artificial intelligence and medical imaging has opened new possibilities for improving accuracy and efficiency in medical diagnosis. This foreword serves as an introduction to a research endeavor that explores the potential of artificial intelligence to revolutionize the field of computer-aided medical diagnosis.

In this thesis, we delve into the world of artificial intelligence applied to medicine, with a particular focus on disease diagnosis through medical imaging. In the following chapters, we will explore the challenges and opportunities that arise when combining the power of artificial intelligence with the wealth of information provided by medical images.

This work addresses a wide range of topics, from tuberculosis to post-COVID conditions, from the design of deep learning algorithms to the experimental evaluation of their performance on clinical datasets. The clinical implications of our findings are discussed, and future research directions in this field of study are proposed.

Our work not only serves as a significant contribution to the field of computer-aided medical diagnosis but also as a source of inspiration for future researchers interested in exploring the transformative potential of artificial intelligence in medicine.

Chapter 1

Introduction

1.1 Background

1.1.1 Advancements in medical imaging and artificial intelligence

Medical imaging has undergone significant advancements over the past few decades, transforming how diseases are diagnosed and treated. Techniques such as X-rays, magnetic resonance imaging (MRI), computed tomography (CT), and ultrasound have become indispensable tools in modern medicine. These modalities provide non-invasive means to visualize internal structures, detect abnormalities, and guide therapeutic interventions. Despite their immense utility, interpreting medical images remains a complex task requiring extensive expertise and experience. Human error, image quality variability, and certain pathological signs' subtlety can all contribute to diagnostic inaccuracies [1].

In parallel with the evolution of medical imaging, the field of artificial intelligence (AI) has made remarkable strides. AI encompasses a broad range of computational techniques designed to mimic human intelligence, including machine learning, deep learning, and neural networks. These technologies have demonstrated exceptional performance in tasks such as image recognition, natural language processing, and predictive analytics [2].

1.1.2 The emergence of deep learning in medical diagnostics

Deep learning, a subset of machine learning, has garnered significant attention for its ability to learn hierarchical representations from data automatically. Convolutional neural networks (CNNs), a type of deep learning architecture, have proven particularly effective for image-related tasks. CNNs consist of multiple layers that can learn to detect increasingly complex features, from simple edges to intricate patterns, directly from raw pixel data [3].

The application of deep learning in medical diagnostics has opened new horizons for enhancing diagnostic accuracy and efficiency. CNNs have been successfully employed to analyze various types of medical images, including radiographs, MRIs, and CT scans. These models can assist in identifying diseases such as cancer, pneumonia, tuberculosis, and neurological disorders with high precision. The ability of deep learning models to process vast amounts of data and recognize patterns beyond human perception makes them valuable tools for augmenting clinical decision-making [4].

1.1.3 Challenges in medical image diagnosis

Despite AI's promise in healthcare, several challenges remain in deploying deep learning models for medical image diagnosis [5]. Variability in image quality, presence of noise, and differences in imaging protocols across institutions can affect model performance. Additionally, the complexity of medical images, which often contain subtle and heterogeneous features, poses significant challenges for accurate disease detection [6].

Moreover, integrating AI into clinical workflows requires addressing model interpretability, trustworthiness, and regulatory compliance issues. Ensuring that AI models provide transparent and explainable results is crucial for gaining the trust of healthcare professionals and patients. Addressing these challenges necessitates a comprehensive approach that includes robust preprocessing techniques, advanced feature extraction methods, and reliable classification algorithms [7].

1.1.4 The role of preprocessing, feature extraction, and classification

In this research, we propose a three-stage model for computer-aided medical diagnosis using deep learning, encompassing preprocessing [8], feature extraction [9], and classification [10].

1. **Preprocessing**: The preprocessing stage focuses on enhancing image quality by reducing noise and optimizing images for analysis. This step is crucial for ensuring that the subsequent stages operate on the best possible data.

2. **Feature extraction**: Feature extraction stage is fine-tuned. The feature extraction stage utilizes pre-trained CNNs such as ResNet50 and VGG16 through transfer learning to identify the most relevant features from the medical images. These pre-trained networks are fine-tuned to specific medical datasets and enable efficient and accurate extraction of intricate image features.

3. **Classification**: The classification stage involves applying advanced machine learning classifiers to the extracted features to diagnose diseases accurately. This stage also includes evaluating model performance using various metrics to ensure reliability and effectiveness.

The modular design of this model allows each stage to be independently improved without affecting the other components. This flexibility facilitates the incorporation of advancements in preprocessing techniques, feature extraction methods, and classification algorithms, ensuring that the model remains at the forefront of technological developments in medical diagnostics [11].

1.2 Problem statement

Medical diagnosis is a critical and complex task that significantly influences patient outcomes [12]. The advent of advanced imaging techniques such as X-rays, MRIs, CT

scans, and ultrasounds have revolutionized the diagnostic process by providing detailed visualizations of internal body structures. However, interpreting these medical images remains a formidable challenge due to several factors.

1.2.1 Variability in image quality

Medical images often suffer from variability in quality due to differences in imaging equipment, protocols, and patient conditions. This variability can introduce noise and artifacts, making it difficult for radiologists to interpret the images accurately [13], as inconsistent image quality can lead to diagnostic errors, potentially resulting in misdiagnosis or delayed treatment.

1.2.2 Subtle and heterogeneous pathological features

Many diseases manifest with subtle and heterogeneous features that are difficult to detect even for experienced radiologists. For example, early-stage cancers or microbleeds in brain imaging may present with minute changes that can be easily overlooked [14]. The complexity of visual patterns in medical images requires advanced techniques capable of discerning these subtle features to ensure accurate diagnosis.

1.2.3 High cognitive load on radiologists

The increasing volume of medical imaging studies presents a significant cognitive load for radiologists, who must analyze numerous images with high precision under time constraints. This workload can lead to fatigue and reduced diagnostic accuracy [15]. Automated diagnostic tools that assist in image analysis can alleviate this burden and enhance the overall efficiency and accuracy of the diagnostic process.

1.2.4 Need for standardization in image analysis

The lack of standardized protocols for image analysis contributes to variability in diagnostic outcomes, as differences in interpretation methods among radiologists can

lead to inconsistent diagnoses. This highlights the need for standardized, automated tools that ensure consistent and reliable analyses across medical institutions [16].

1.2.5 Integrating AI into clinical workflows

While AI, particularly deep learning, has shown great promise in medical diagnostics, integrating these technologies into clinical workflows presents several challenges. These include ensuring the interpretability and trustworthiness of AI models, complying with regulatory standards, and gaining acceptance from healthcare professionals [17]. Models must be transparent, providing explainable results that clinicians can trust and act upon.

1.2.6 Research gap in modular AI models

Current AI models for medical diagnosis often lack modularity, making it challenging to incorporate improvements in specific stages of the diagnostic process without affecting the entire system. A modular approach that allows independent enhancements in preprocessing, feature extraction, and classification stages can provide a more flexible and adaptable solution [18].

1.2.7 Addressing the problem

To address these challenges, this research focuses on developing a robust, modular deep-learning model for computer-aided medical diagnosis. The proposed model is designed to improve diagnostic accuracy and efficiency by integrating three key components: preprocessing, feature extraction, and classification. Each component can be independently optimized to incorporate the latest advancements, ensuring the model remains at the cutting edge of medical diagnostics [19].

- ***Preprocessing:*** Enhances image quality by reducing noise and optimizing images, providing a solid foundation for subsequent analysis.

- ***Feature extraction:*** Utilizes pre-trained convolutional neural networks (CNNs) through transfer learning to extract relevant features, leveraging state-of-the-art deep learning techniques.

- ***Classification:*** Applies advanced machine learning classifiers to the extracted features, delivering accurate and reliable disease diagnoses.

1.3 Objectives of the study

This research aims to improve medical image diagnosis accuracy, consistency, and efficiency by addressing the aforementioned issues through a comprehensive and adaptable approach. The ultimate goal is to develop a tool that supports healthcare professionals in delivering better patient care through precise and timely diagnoses [20]. Specifically, the aims are as follows:

1. **Develop a robust deep learning model for automated medical diagnosis**: To create a comprehensive deep learning model for automatically diagnosing various pathologies using medical imaging. This model will integrate three main components: preprocessing, feature extraction, and classification.

2. **Enhance image quality through preprocessing techniques**: To improve the quality of medical images by reducing noise and optimizing them for subsequent analysis. This stage will ensure that the images used for feature extraction and classification are of high quality.

3. **Utilize pre-trained convolutional neural networks for feature extraction**: Use pre-trained convolutional neural networks (CNNs) such as ResNet50, VGG16, and others to extract the most relevant features from medical images. Transfer learning will facilitate extracting high-quality features, enhancing the model's diagnostic capabilities.

4. **Implement advanced classification algorithms**: This stage will apply advanced machine learning classifiers to the features extracted by the CNNs, en-

suring accurate and reliable disease classification. It will also involve evaluating the model using various performance metrics.

5. **Ensure modular design for independent component improvement**: To design the model so that each of its three stages (preprocessing, feature extraction, and classification) can be independently improved without affecting the other stages. This modularity will allow for the seamless integration of better techniques as they become available.

1.4 Research hypotheses

1. **Enhanced image quality improves diagnostic accuracy**: Preprocessing techniques that reduce noise and optimize image quality will improve feature extraction and diagnostic accuracy.

2. **Transfer learning enhances feature extraction**: Utilizing pre-trained CNNs through transfer learning will extract more relevant features from medical images than traditional methods, leading to better classification performance.

3. **Modular design facilitates continuous improvement**: The deep learning model's modular design will allow for independent improvements in preprocessing, feature extraction, and classification stages, leading to a more adaptable and robust diagnostic tool over time.

4. **Advanced classifiers improve disease detection**: Implementing advanced machine learning classifiers on the features extracted by CNNs will result in high accuracy and reliability in the automatic diagnosis of various diseases.

1.5 Justification of the study

Integrating artificial intelligence, particularly deep learning, into medical diagnostics represents a significant advancement in the healthcare field [21]. The ability to accurately and efficiently diagnose diseases using medical images can have profound

implications for patient outcomes. The three-stage model proposed in this research comprising preprocessing, feature extraction, and classification addresses several critical aspects of automated medical diagnosis:

1. **Preprocessing** [22]: Improving the quality of medical images is essential for accurate diagnosis. By reducing noise and optimizing images, preprocessing ensures that the subsequent stages of the model work with the best possible data, ultimately leading to more reliable diagnostic results.

2. **Feature extraction** [23]: Using pre-trained CNNs for feature extraction leverages the power of transfer learning, which has been proven highly effective in various image analysis tasks. These pre-trained models can extract intricate and relevant features from medical images, which are crucial for accurate disease classification.

3. **Classification** [24]: Advanced machine learning classifiers are vital for interpreting the extracted features and making precise diagnostic decisions. Using robust classifiers ensures that the model can handle the complexity and variability of medical images, providing accurate and reliable diagnoses.

4. **Modular design**: A modular design allows for continuous model improvement. As new and better techniques are developed in image preprocessing, feature extraction, and classification, they can be integrated into the existing model without disrupting its overall functionality. This adaptability is crucial for keeping the model up-to-date with the latest advancements in AI and medical imaging.

In summary, this research aims to develop a state-of-the-art deep-learning model for computer-aided medical diagnosis that is accurate, reliable, and adaptable [25]. This model can significantly enhance medical practitioners' diagnostic capabilities and improve patient outcomes by addressing key aspects of image quality, feature extraction, and classification.

1.6 Thesis structure

This thesis is organized into six chapters, each addressing a key aspect of the research conducted. Below is a detailed overview of each chapter:

- **Chapter 2**: Literature review – A comprehensive review of existing research on AI in medical diagnostics, including deep learning techniques and their applications in medical imaging.
- **Chapter 3**: Methodology – Detailed explanation of the methods used in developing and evaluating the deep learning model, including preprocessing techniques, feature extraction, and classification algorithms.
- **Chapter 4**: Results – Presentation and analysis of the results obtained from the experimental evaluation of the model.
- **Chapter 5**: Discussion – Interpretation of the results, discussion of their implications, and comparison with existing studies.
- **Chapter 6**: Conclusions and future work – Summary of the main findings, conclusions drawn from the research, and suggestions for future research directly.

1.7 Summary

In this introductory chapter, we have laid the foundation for understanding the context and significance of our research on computer-aided medical diagnosis using artificial intelligence [26]. We began by highlighting the remarkable advancements in medical imaging techniques and the transformative potential of artificial intelligence, particularly deep learning, in enhancing diagnostic accuracy and efficiency. We then identified the key challenges that persist in medical image diagnosis, including variability in image quality, the complexity of pathological features, and the high cognitive load on radiologists [27].

We presented our proposed solution, a modular deep learning model comprising three stages: preprocessing, feature extraction, and classification. This model

addresses the identified challenges by enhancing image quality, extracting relevant features using pre-trained convolutional neural networks, and accurately classifying medical conditions using advanced machine learning classifiers. The flexibility of this modular approach allows for independent improvements in each stage, ensuring the model remains adaptable to future advancements.

This research's objectives, hypotheses, and justification were clearly defined, outlining our goals to develop a robust diagnostic tool that improves clinical decision-making. The significance of this research lies in its potential to augment the capabilities of healthcare professionals, reduce diagnostic errors, and ultimately enhance patient outcomes.

With a comprehensive understanding of the problem statement and the proposed approach, we are now prepared to delve deeper into the existing body of research that informs and supports our work. In the next chapter, we will conduct a thorough literature review, examining the current state of AI in medical diagnostics, exploring various deep learning techniques, and understanding their applications in medical imaging. This review will provide the necessary background and context, setting the stage for the detailed methodology and experimental results that follow.

As we transition into Chapter 2, the focus will shift to reviewing and synthesizing relevant literature, identifying gaps in existing research, and positioning our work within the broader landscape of AI-driven medical diagnostics. This will highlight the novelty and contributions of our research and ensure that our approach is grounded in a solid understanding of current technologies and practices.

Chapter 2

Literature review

2.1 Introduction

Artificial Intelligence (AI) applications in medical diagnostics have gained significant attention in recent years due to their potential to revolutionize how medical images are analyzed and interpreted. Traditional diagnostic methods, while effective, often require substantial human expertise and can be prone to variability depending on the individual interpreting the data. Integrating AI, particularly deep learning techniques, offers the ability to automate the diagnostic process, improve accuracy, and reduce the workload on medical professionals [28].

This chapter provides a comprehensive review of the current state of AI in medical diagnostics, with a particular emphasis on the deep learning models developed for medical image analysis. AI-driven approaches have been applied across various imaging modalities, such as X-rays, computed tomography, and magnetic resonance imaging, to aid in the detection and classification of various pathologies, including lung diseases, brain tumors, and cardiovascular conditions [29].

As the field evolves, the role of deep learning, particularly convolutional neural networks, has become increasingly important due to their success in image classification, segmentation, and feature extraction. This chapter explores the existing literature on these techniques and highlights their applications, challenges, and limitations within the context of medical imaging. Additionally, the review seeks to identify the gaps in

the current research and position the contributions of this thesis within the broader landscape of AI-driven diagnostics [30].

2.2 Brief overview of AI in medical diagnostics

AI has emerged as a transformative technology in various sectors, and its applications in healthcare, particularly in medical diagnostics, have shown immense potential. AI in medical diagnostics refers to using algorithms, primarily machine learning and deep learning, to assist in identifying and diagnosing diseases from medical data [31]. These technologies have advanced the capabilities of medical professionals by providing faster, more accurate, and reproducible results, especially in medical imaging. Figure 2.1 shows in a simple and general way the auxiliary process for the detection, classification, and segmentation of medical images using Deep Learning (DL) in various medical pathologies, intending to optimize and generate early, timely, and precise treatment [32].

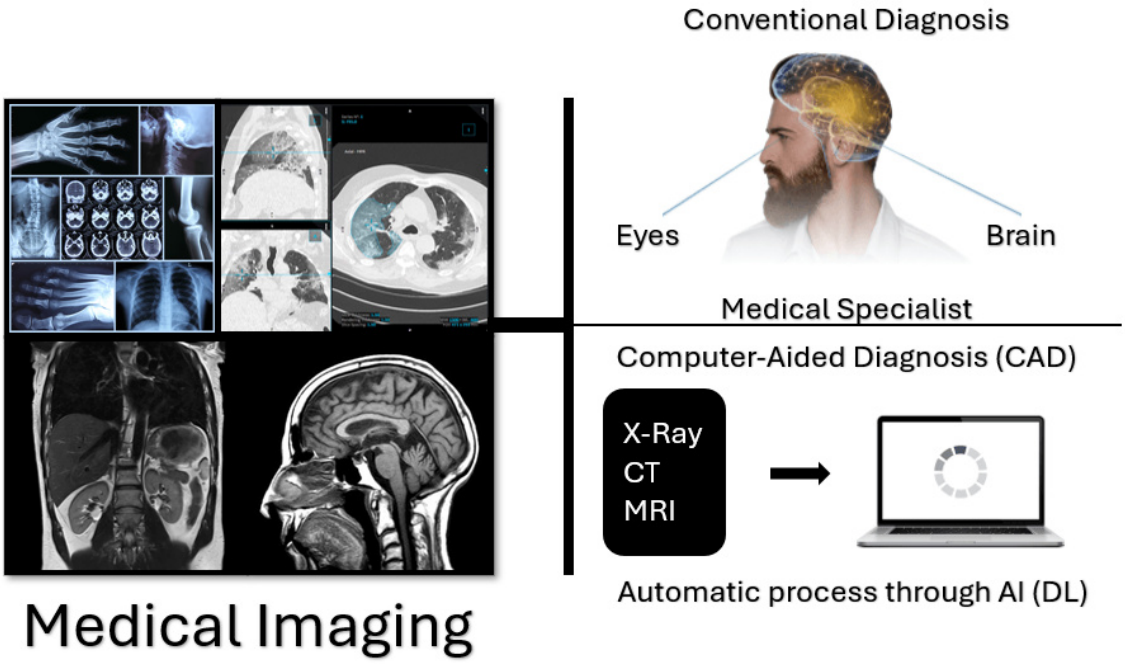


Figure 2.1: Automatic classification model as an auxiliary tool for conventional diagnosis by the specialist physician.

The integration of AI into medical diagnostics primarily focuses on processing

large datasets generated by imaging technologies like X-ray, CT, MRI, and ultrasound. These datasets are often complex and difficult to interpret manually, especially when subtle patterns or abnormalities are present. AI algorithms, particularly deep learning-based ones, can automatically analyze these images, detect patterns, and provide diagnostic outputs that can support clinical decision-making [33].

2.2.1 Evolution of AI in diagnostics

The journey of AI in diagnostics can be traced back to the early development of machine learning techniques, where statistical methods were applied to clinical data to predict patient outcomes. However, it was not until the advent of deep learning in the early 2010s that AI made significant strides in medical imaging [34]. CNNs, a type of deep learning architecture specifically designed for image analysis, have proven particularly effective for tasks such as image classification, segmentation, and detection [35], as shown in Figure 2.2.

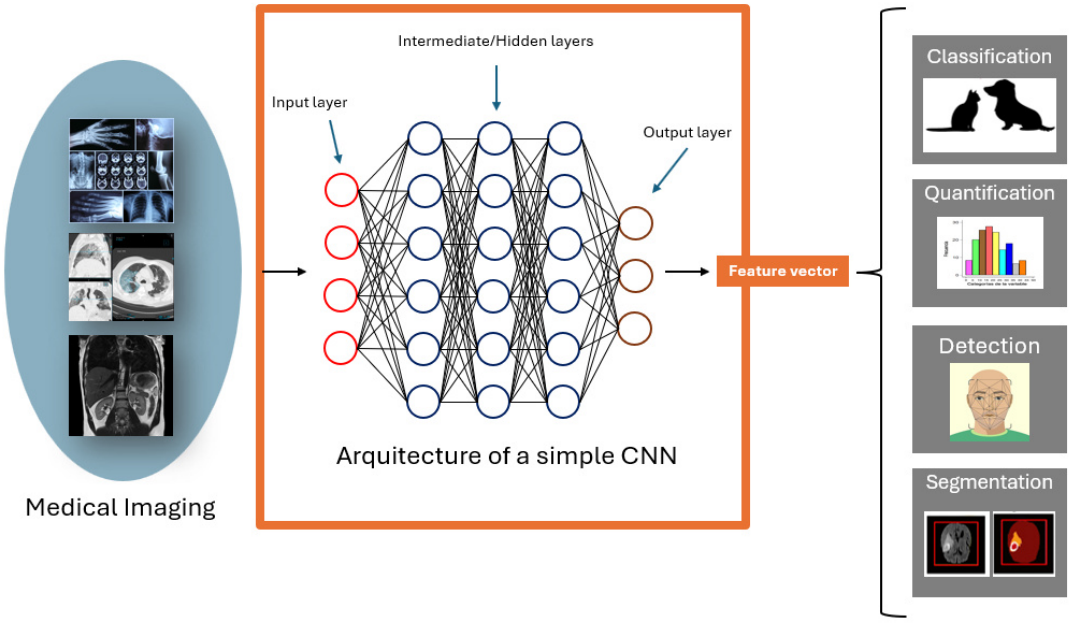


Figure 2.2: The simple process of a medical image passing through a simple three-hidden layer CNN, whose output is a feature vector that can be used for different systems.

The use of CNNs in medical diagnostics allows for the automatic extraction of features from raw images, bypassing the need for manual feature engineering, a limitation of earlier machine-learning techniques. This has been especially useful in diagnostic tasks involving radiographic images, where detecting minute anomalies can be critical. As a result, deep learning has been successfully applied to detect a range of pathologies, including pneumonia, tuberculosis, lung cancer, and brain tumors, among others [36].

2.2.2 AI techniques in diagnostics

The two primary categories of AI techniques in medical diagnostics are traditional machine learning and deep learning.

- **Traditional machine learning** [37]: These algorithms require manual feature extraction and selection, followed by classification or regression techniques to diagnose diseases. Support Vector Machines (SVM), Decision Trees, and Random Forests have been commonly used in this regard. While effective in structured data, these methods often struggle with unstructured data, such as images, where feature extraction is not straightforward.

- **Deep learning** [38]: In contrast, deep learning methods, especially CNNs, have revolutionized medical diagnostics by automating feature extraction from raw medical images. CNNs consist of layers of filters that learn to identify various features of an image, such as edges, textures, and patterns, which are crucial for medical diagnoses. Deep learning has enabled the development of end-to-end models that take raw medical images as input and output diagnostic decisions without needing handcrafted features.

Among the most widely used architectures are:

1. **Convolutional neural networks** [39]: Ideal for image analysis, CNNs are commonly used in tasks such as detecting tumors in MRI scans or abnormalities in X-rays.

2. **Recurrent neural networks (RNNs)** [40]: Often applied to time-series medical data such as ECG or EEG signals.

3. **Generative adversarial networks (GANs)** [41]: Used in image generation and enhancement, GANs have been applied to improve image resolution and reduce noise in diagnostic images.

2.3 Applications of AI in medical imaging diagnosis

AI has been integrated into diagnostic workflows in several imaging modalities:

a) **X-ray** [42]: AI has been used extensively to analyze chest X-rays to detect pneumonia, tuberculosis, and even COVID-19. By learning from vast datasets, CNN-based models can accurately distinguish between healthy and abnormal chest X-rays, often identifying conditions that human radiologists might miss.

b) **Computed tomography (CT)** [43]: CT scans generate detailed cross-sectional images of the body, which are used in diagnosing a wide array of conditions, including cancers and vascular diseases. AI models have demonstrated high accuracy in identifying lung nodules, and hemorrhages, and even predicting patient outcomes based on CT data.

c) **Magnetic resonance imaging (MRI)** [44]: MRI is frequently used to detect brain tumors, spinal injuries, and cardiac conditions. Deep learning models have been developed to segment tumors, classify brain lesions, and even predict disease progression from MRI scans.

AI models in these applications are often designed to work in tandem with radiologists, providing a second opinion or identifying suspicious regions for further investigation. Studies have shown that AI systems can reduce diagnostic errors and improve patient outcomes when combined with human expertise.

2.3.1 Ethical consideration and limitations

While AI offers numerous advantages, its integration into medical diagnostics is not without challenges. One of the primary concerns is the black-box nature of deep learning models, which can make it difficult to understand how decisions are made. In critical applications like healthcare, transparency, and interpretability are essential for gaining the trust of medical professionals and patients [45].

Another limitation is the availability of high-quality, annotated data for training these models. Deep learning requires large amounts of labeled data, and obtaining such datasets in the medical field can be challenging due to privacy concerns and the need for expert annotations [46].

Lastly, the risk of bias in AI models is a growing concern. AI models trained on data from specific populations may not generalize well to other populations, leading to disparities in healthcare outcomes. Therefore, ensuring that AI models are trained on diverse datasets representing various demographic and clinical factors is crucial [47].

2.3.2 Future directions

The future of AI in medical diagnostics is promising, with ongoing research focused on improving the robustness, transparency, and generalizability of AI models. Techniques such as explainable AI (XAI) are being developed to make AI models more interpretable, provide insights into how decisions are made, and ensure that they align with medical knowledge [48].

Moreover, advancements in transfer learning, where pre-trained models are fine-tuned on smaller medical datasets, have enabled leveraging the power of deep learning even in cases with limited data. The integration of AI into clinical practice will continue to grow, with AI-driven diagnostics playing an increasingly central role in improving patient outcomes and reducing the burden on healthcare systems [49].

2.4 Techniques for medical imaging

2.4.1 X-ray imaging

X-ray imaging is one of the most fundamental and widely used techniques in medical diagnostics [50]. At its core, X-ray imaging leverages the physics of electromagnetic radiation. X-rays are a form of high-energy electromagnetic radiation, with wavelengths in the range of 0.01 to 10 nanometers, corresponding to frequencies between 30 petahertz (PHz) and 30 exahertz (EHz) [51]. This high energy enables X-rays to penetrate various human body tissues, making them invaluable for creating inner images.

2.4.2 Physical principles of X-rays

X-rays are a form of high-energy electromagnetic radiation. In many languages, it is referred to as Röntgen radiation, after the German scientist Wilhelm Conrad Röntgen, who discovered it in 1895 and named it X-radiation to signify an unknown type of radiation [52]. Figure 2.3 shows the print of Wilhelm Rönteng's first medical X-ray of his wife's hand, taken on 22 December 1895 and presented to Ludwig Zehnder of the Physik Institut, University of Freiburg, on 1 January 1896 [53].

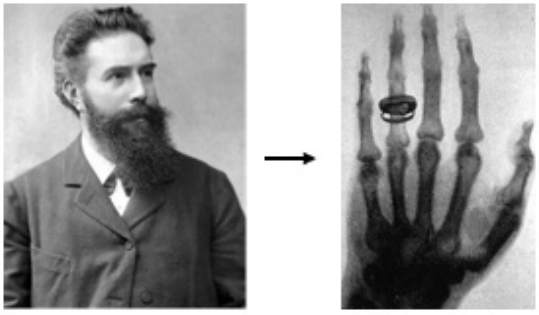


Figure 2.3: First medical image taken from X-rays.

X-rays with high photon energies above 5-10 keV (below 0.2–0.1 nm wavelength) are called hard X-rays, while those with lower energy (and longer wavelength) are called soft X-rays. The intermediate range with photon energies of several keV is often

referred to as tender X-rays. Due to their penetrating ability, hard X-rays are widely used to image the inside of objects (e.g. in medical radiography and airport security). The term X-ray is metonymically used to refer to a radiographic image produced using this method, in addition to the method itself. Since the wavelengths of hard X-rays are similar to the size of atoms, they are also useful for determining crystal structures by X-ray crystallography. By contrast, soft X-rays are easily absorbed in the air; with an attenuation length of 600 eV (2 nm) X-rays in water are less than 1-micrometer [54] (see Figure 2.4).

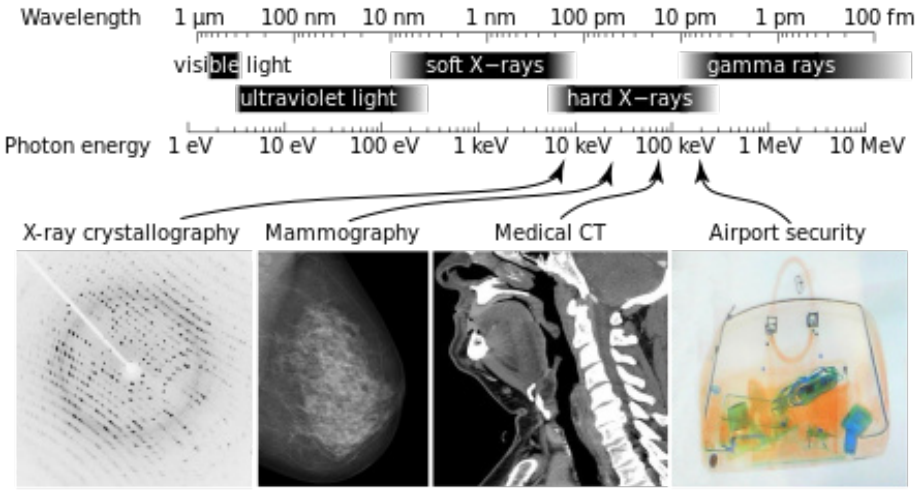


Figure 2.4: X-rays are part of the electromagnetic spectrum, with wavelengths shorter than UV light. Different applications use different parts of the X-ray spectrum.

X-rays are generated when high-energy electrons collide with a metal target, inside an X-ray tube, typically tungsten. This process is known as "bremsstrahlung," or braking radiation. As electrons decelerate upon striking the metal target, their kinetic energy is converted into electromagnetic radiation, which manifests as X-rays. The energy of the emitted X-rays is proportional to the deceleration and the atomic number of the target material. The resulting X-rays consist of a spectrum of energies, with higher-energy X-rays capable of penetrating denser tissues, such as bones, and lower-energy X-rays being absorbed by softer tissues. Based on tissue density, the varying degrees of absorption form the contrast in the resulting image. Radiation can be classified by wavelength λ_p, the length of one wave period. The wavelength can

also be represented by frequency f_p and the wave propagation speed, i.e., the speed of light c [55].

$$\lambda_p = \frac{c}{f_p}\,. \tag{2.1}$$

In Eq. (2.2) the energy of photons is given, where h denotes Planck's constant ($\approx 6.626069 \times 10^{-34} J \cdot s$) and c is the speed of light ($\approx 2.99792 \times 10^5 m/s$). The energy is directly related to the photon's wavelength λ_p or its frequency f_p and is given by the unit electron volt [eV]. We can easily obtain that the photon energy is proportional to its frequency and inversely proportional to its wavelength, which means the higher its frequency, the higher its energy [56].

$$E_p = \frac{hc}{\lambda_p} = f_p h\,. \tag{2.2}$$

When a monochromatic X-ray beam traverses a homogeneous object with absorption coefficient μ, according to Lambert-Beer's law, the observed intensity I is related to the intersection length of the object x and the ray:

$$I = I_0 \cdot e^{-\mu \cdot x}, \tag{2.3}$$

here, I_0 is the X-ray intensity at the X-ray source. As mentioned, it is a digital radiological diagnostic technique. The image is obtained by exposing the radiographic image receptor to a high-energy radiation source, commonly X-rays or gamma radiation (from radioactive isotopes, e.g., iridium-192, cobalt-60, cesium-137); this can be seen in Figure 2.5. The interaction of X-rays with biological tissues follows several key mechanisms (see Figure 2.6):

- **Photoelectric effect** [57]: An atom absorbs X-rays, its energy can eject an electron from one of the inner shells of the atom. This effect is more significant in denser tissues (like bone), which results in greater X-ray absorption and a whiter appearance on the image. The photoelectric effect is directly proportional to the tissue's atomic number (Z), making it the primary mechanism responsible

for bone visibility in X-rays.

- **Compton scattering** [58]: X-rays can scatter off electrons in a tissue, changing direction while losing some energy. This scattering can lead to image noise, but it also helps differentiate soft tissues. Compton scattering dominates in lower-density tissues, such as muscle or fat.

- **Rayleigh scattering** [59]: This involves the elastic scattering of X-rays by atoms without energy loss. Though less common in medical X-rays, it plays a small role in image formation.

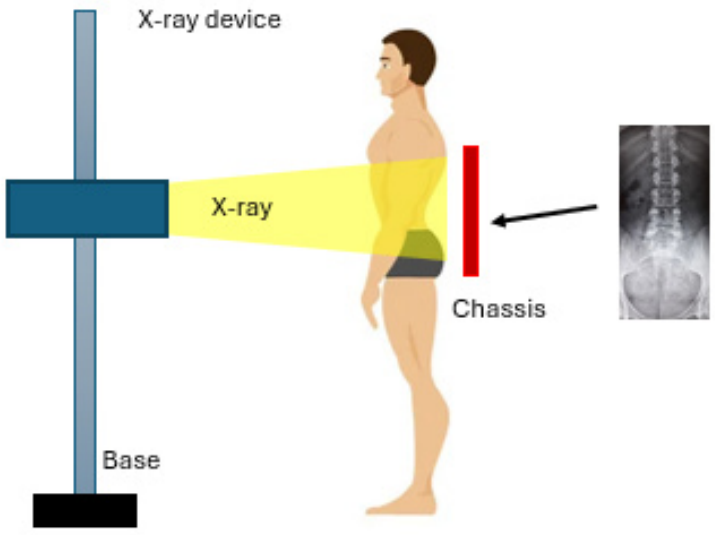


Figure 2.5: Simple diagram of obtaining digital X-ray images.

The differential absorption of X-rays by various tissues allows for the detailed visualization of internal structures. Dense structures like bones absorb a significant portion of the X-rays, leading to fewer X-rays reaching the detector, resulting in a bright image appearance. In contrast, soft tissues absorb fewer X-rays, allowing more to pass through and reach the detector, creating darker areas on the image. Modern X-ray systems have advanced with digital detectors, replacing traditional film. These detectors convert X-rays into electrical signals, which are then processed

to generate digital images. This process reduces radiation doses and improves image quality by enabling post-processing adjustments, such as contrast enhancement or noise reduction.

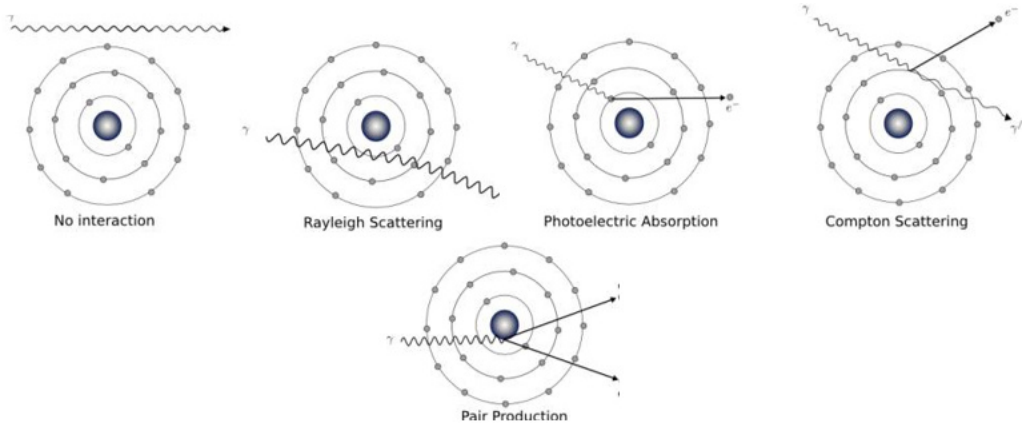


Figure 2.6: Principles of photon-matter interactions.

X-ray imaging is particularly useful in diagnosing fractures, infections, and tumors. Beyond basic radiography, specialized forms of X-ray imaging, such as fluoroscopy and computed tomography (CT), offer dynamic and cross-sectional body views, providing more detailed insights into internal pathologies [60]. X-ray imaging remains a cornerstone in medical diagnostics due to its simplicity, speed, and effectiveness in providing clear images of bony structures and some soft tissues. However, its limitations, including exposure to ionizing radiation and reduced contrast for soft tissues, make it necessary to complement it with other imaging techniques such as CT or MRI, for more comprehensive diagnoses.

2.4.3 Computed tomography

Computed Tomography (CT), also known as CAT (Computerized Axial Tomography) scanning, is an advanced imaging technique that builds upon the basic principles of X-ray imaging to produce detailed three-dimensional images of the body [61]. CT combines the physics of X-rays with computational algorithms to provide cross-sectional images, offering a much more comprehensive view of internal structures than traditional X-ray imaging.

2.4.4 Physics behind CT imaging

CT scanners take multiple X-ray images from different angles around the patient and use computer algorithms to reconstruct these into a detailed 3D representation of the body's internal structures. Since its development in the 1970s, CT scanning has proven to be a versatile imaging technique [62]. While CT is most prominently used in medical diagnosis, it can also form images of non-living objects. The 1979 Nobel Prize in Physiology or Medicine was awarded jointly to South African-American physicist Allan MacLeod Cormack and a British electrical engineer Godfrey Hounsfield "for the development of computer-assisted tomography" [63] (see Figure 2.7).

Figure 2.7: 1979 Nobel Prize winners for inventing X-ray CT.

A CT scanner consists of an X-ray tube that rotates around the patient and a set of detectors positioned on the opposite side. As the X-ray tube rotates, it emits a narrow beam of X-rays that passes through the body. The detectors measure the intensity of X-rays after they have passed through the body, which varies depending on the tissue density. Dense tissues, like bone, absorb more X-rays, while softer tissues, like organs, absorb fewer [64]. Based on image acquisition and procedures, various types of scanners are available in the market: sequential, spiral, electron beam, dual-energy, perfusion imaging, and positron emission. In Figure 2.8, you can see a simple diagram of how this device captures images.

A CT scanner consists of an X-ray tube that rotates around the patient and a set of detectors positioned on the opposite side. As the X-ray tube spins, it emits a narrow beam of X-rays that passes through the body. The detectors measure the

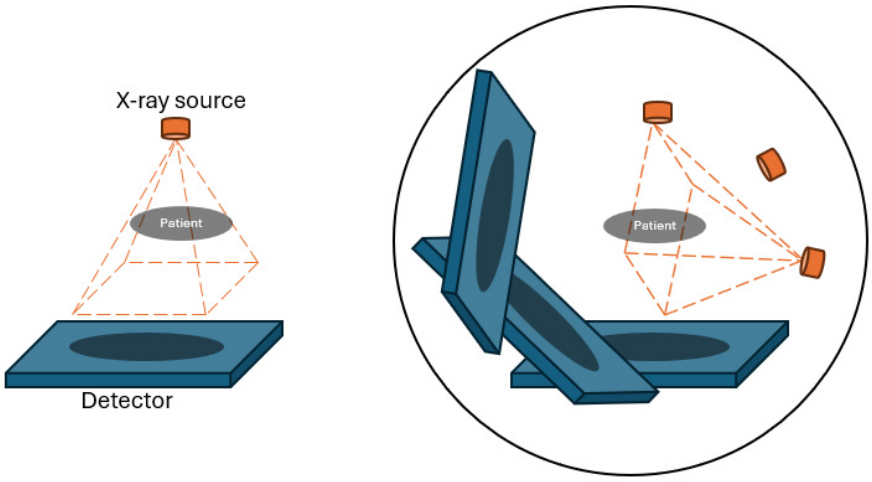


Figure 2.8: Outline of the basis of obtaining tomography images.

intensity of X-rays after they have passed through the body, which varies depending on the tissue density. Dense tissues, like bone, absorb more X-rays, while softer tissues, like organs, absorb fewer. The X-ray source takes images around the patient at multiple angles (typically 360 degrees). For each angle, the detectors record an X-ray attenuation profile, producing what is known as a "projection". These projections contain information about the body's internal structure from that particular angle. The key to CT imaging is mathematical reconstruction algorithms, which are used to transform the numerous 2D projections into cross-sectional "slices" of the body. The Filtered Back Projection method is one of the most commonly used algorithms [65]. This technique mathematically processes the projections to reconstruct the internal structure at each slice, combining all the data points into a coherent 2D image of the body's cross-section.

More advanced methods, like Iterative Reconstruction, are now being employed to reduce noise and improve image quality, which is especially useful in lowering the radiation dose during CT scanning. Image reconstruction in CT is a mathematical process that calculates the 2D cross-sectional attenuation distribution function $f(x, y)$ from a series of one-dimensional (1D) X-ray projections $P(r, \theta)$ acquired as the line integrals at many different angles around the three-dimensional (3D) object [66], as shown in Figure 2.9.

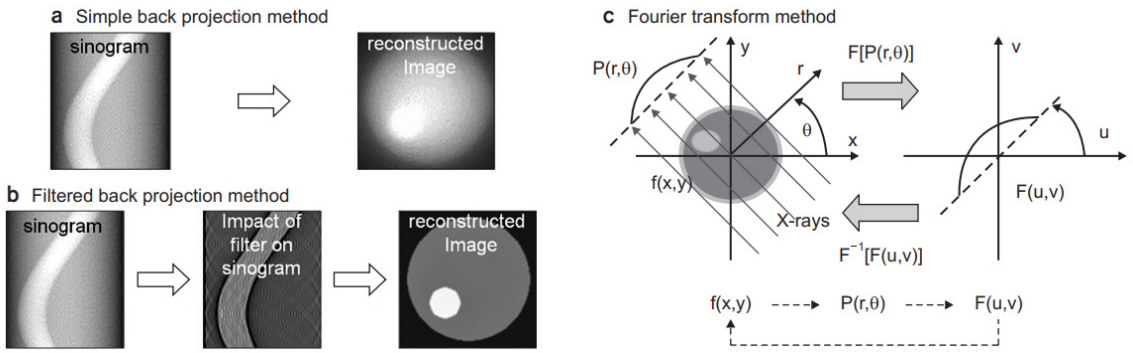


Figure 2.9: Computed tomography image reconstruction methods. (a) Simple back projection algorithm method, (b) filtered back projection algorithm method, and (c) Fourier transform algorithm method.

A projection, $P(r, \theta)$ is formed by the attenuated set of parallel X-rays (or isotropic X-rays emitted from a point source) through a 2D object of interest. A collection of projections at several angles during a single rotation of the X-ray source-detection system is called a sinogram. Effectively, this is a linear transform of the cross-sectional image of the object. Sinograms simply display all of the different projections for any slices stacked together. In reality, we can know the intensity I_0 of the incident X-ray beam and measure the intensity I of the detected beam. As a result, we aim to understand the distribution of individual X-ray attenuation coefficients of the object (see Figure 2.9). Image reconstruction in CT involves the calculation of these individual coefficients in the internal structure of the object [67]. Where $F\{P(r, \theta)\}$ is:

$$F\{P(r,\theta)\} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) e^{-ixw(u,v)} dxdy \,. \tag{2.4}$$

The typical CT image is constructed on a square matrix of 512×512 $(262,144$ pixels) and has an image depth of 12 bits $(4,096$ gray levels) but has increased to matrices composed of $1,024 \times 1,024$ or $2,046 \times 2,048$ (ultrahigh-resolution CT) pixels in recent years. CT numbers (Hounsfield units [HU]) are defined as the attenuation values of the imaged tissues normalized to that of water [68]:

$$CTnumber(HU) = 1,000 \times \frac{\mu_{pixel} - \mu_{water}}{\mu_{water}}, \tag{2.5}$$

where μ_{water} is water's linear attenuation coefficient, μ_{pixel} is a given pixel's linear attenuation coefficient. According to the definition of CT number (HU), the CT numbers of air and water range from $-1,000$ and 0, respectively. The CT numbers for various organs in humans are shown in Figure 2.10:

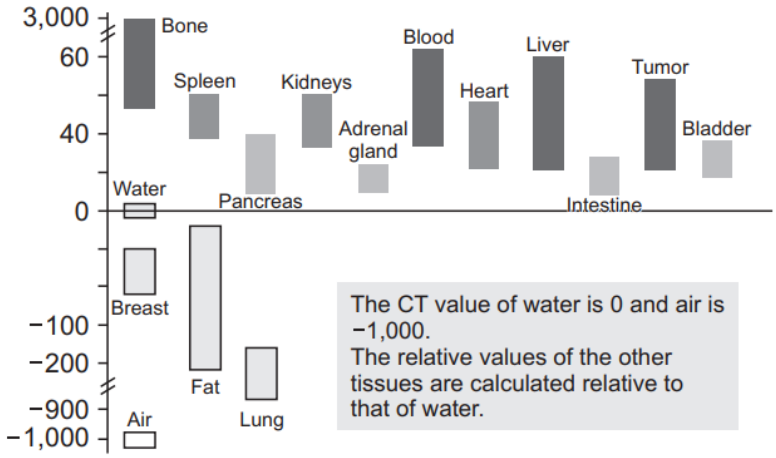


Figure 2.10: Hounsfield scale of computed tomography (CT) numbers for various tissues.

The most common application of X-ray CT refers to cross-sectional (axial) images used for diagnostic and therapeutic purposes in various medical fields [69]. Modern CT workstations are specifically designed to navigate and visualize various reformatted images, such as multiplanar reconstruction (MPR) techniques. The MPR technique involves the image processing tasks needed to convert thin-slice data from volumetric CT in the axial plane from another plane (such as coronal, sagittal, or oblique) to control the window level (WW) and center the image to view whichever structures needed with the aid of various software and a workstation, as shown in Figure 2.11. From the 3D reconstruction of a sequence of tomographic images, the MPR display can generate interactive slices in Cartesian planes (axial, sagittal, and coronal), or even in oblique planes [70]. This means the physicians can obtain views of the patient's internal structures that provide greater clarification than the images

from the original sequence.

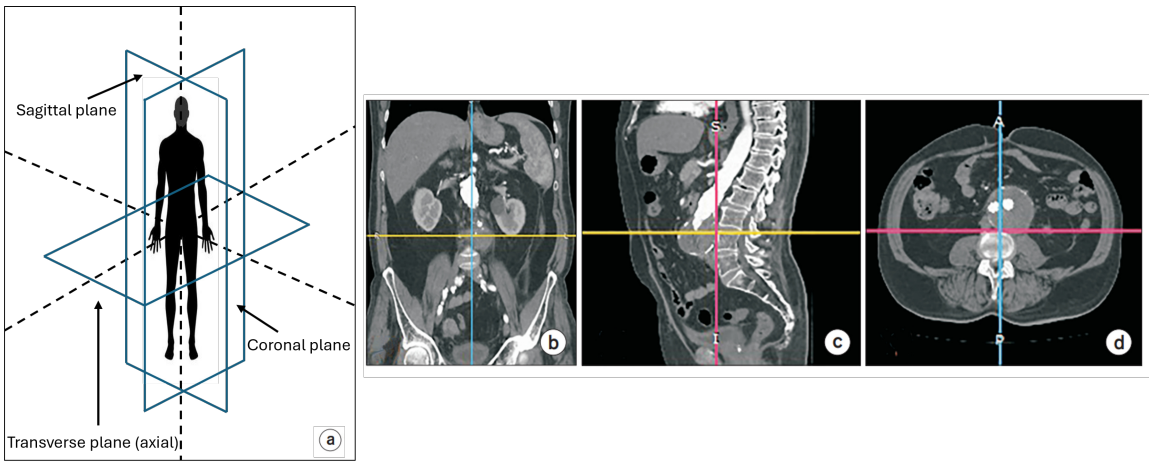


Figure 2.11: (a) Anatomical planes in human, multiplanar reconstruction computed tomography images, such as (b) coronal and (c) sagittal converted from (d) axial.

CT scans, while powerful, face several challenges, especially concerning image quality, radiation exposure, and the complexity of image interpretation. These challenges have opened avenues for deep learning techniques, significantly enhancing segmentation, classification, and diagnostic accuracy in CT imaging. CT uses ionizing radiation, which poses a health risk, particularly with repeated exposures. Reducing radiation dose while maintaining image quality is a significant challenge in the medical use of CT. Deep learning techniques, such as convolutional neural networks (CNNs), have helped address this by enhancing low-dose CT images [71]. Noise reduction algorithms, powered by AI, allow for clearer images at reduced doses, thus minimizing patient exposure to harmful radiation. The sheer amount of data generated by CT scans makes manual interpretation by radiologists time-consuming and prone to error, particularly in identifying subtle anomalies. Deep learning models, specifically CNNs, have revolutionized segmentation in CT images, allowing for automated and highly accurate delineation of structures like organs, tumors, and blood vessels. These models can learn to identify patterns and distinguish between different tissue types, which is especially useful in oncology for delineating tumor boundaries.

In classification, deep learning algorithms have been employed to identify specific disease patterns in CT scans, such as lung nodules, liver tumors, or coronary artery

calcifications. The ability of CNNs to process large volumes of data and extract complex features from CT images has improved diagnostic accuracy and consistency. Deep learning techniques have significantly contributed to overcoming many of the challenges associated with CT imaging. By automating the segmentation and classification processes, AI reduces the workload of radiologists and increases diagnostic accuracy. Moreover, noise-reduction algorithms and artifact removal techniques ensure clearer images, even at lower radiation doses. The future of CT imaging is closely intertwined with AI, with ongoing research aimed at further enhancing image quality, reducing radiation exposure, and improving diagnostic precision [72].

2.4.5 Magnetic resonance imaging (MRI)

Magnetic Resonance Imaging (MRI) is one of the most advanced medical imaging techniques available, offering exceptional detail in soft tissue imaging. Unlike CT, which relies on ionizing radiation, MRI uses strong magnetic fields and radiofrequency (RF) waves to produce images, making it a safer option for repeated imaging. MRI is widely used in diagnosing various conditions, particularly those involving soft tissue, such as brain injuries, cancers, and neurological diseases [73].

2.4.6 MRI physics

MRI operates on the principle of nuclear magnetic resonance (NMR), where atomic nuclei with an odd number of protons or neutrons (such as hydrogen) align in a magnetic field, and respond to radiofrequency waves. Hydrogen atoms, abundant in water molecules within the body, are most commonly targeted because of their high sensitivity to the MRI process. When the radiofrequency field is turned off, the MRI sensors can detect the energy released as the protons realign with the magnetic field. The time it takes for the protons to realign with the magnetic field, and the amount of energy released, changes depending on the environment and the chemical nature of the molecules [74]. Physicians can tell the difference between various types of tissues based on these magnetic properties. To obtain an MRI image, a patient is placed

inside a large magnet and must remain still during the imaging process so as not to blur the image. Contrast agents (often containing the element Gadolinium) may be given to a patient intravenously before or during the MRI to increase the speed at which protons realign with the magnetic field. The faster the protons realign, the brighter the image. Figure 2.12 shows the axial section of how these devices are constructed [75].

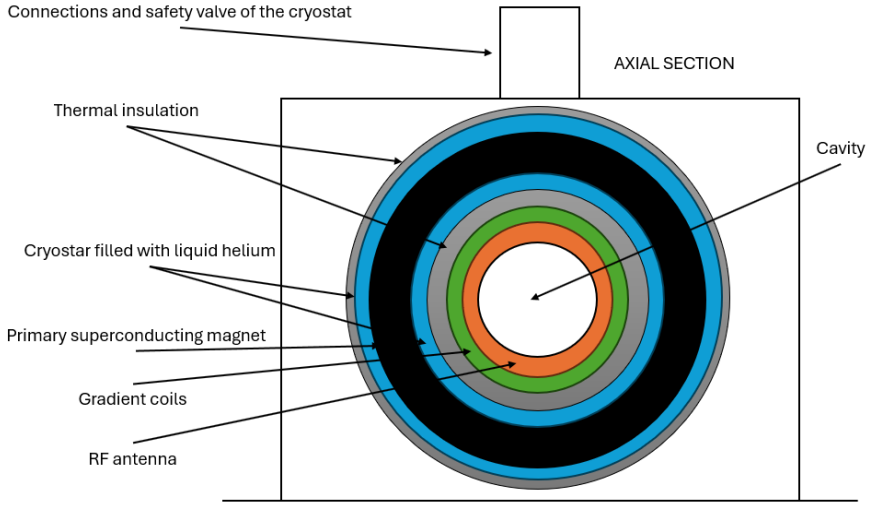


Figure 2.12: Axial section, Clinical MRI Scanner.

Figure 2.13 shows the longitudinal section.

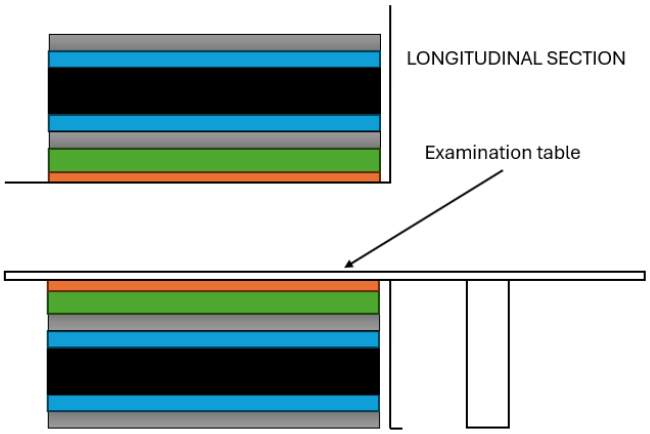


Figure 2.13: Longitudinal section, Clinical MRI Scanner.

Nuclear magnetic resonance (NMR) was first described and measured in molecular

beams by Isidor Rabi in 1938, by extending the Stern–Gerlach experiment, and in 1944, Rabi was awarded the Nobel Prize in Physics for this work [76]. In 1946, Felix Bloch and Edward Mills Purcell expanded the technique for use on liquids and solids, for which they shared the Nobel Prize in Physics in 1952. Vladislav Ivanov proposed the idea of using the newly discovered phenomenon for imaging. However, it wasn't until 1977 that Raymond Andrew developed the MRI Scanner based on this research [77]. The full-body scanner was created in 1978.

Subatomic particles have the quantum mechanical property of spin. Certain nuclei such as $^{1,2}H$ (protons), ^{3}He, ^{23}Na, or ^{31}P have a non-zero spin and, therefore, a magnetic moment. In the case of the so-called spin $-\frac{1}{2}$ nuclei, in fact, ^{1}H there are two spin states, sometimes referred to as up and down [78]. When these spins are placed in a strong external magnetic field, they precess around an axis along the direction of the field. Protrons align in two energy eigenstates (the Zeeman effect), low-energy and high-energy, which are separated by very small splitting energy [79].

A non-zero spin $\vec{S}$ is associated with a non-zero magnetic dipole moment, $\vec{\mu}$, via the relation:

$$\vec{\mu} = \gamma \vec{S}, \tag{2.6}$$

where γ is the gyromagnetic ratio, classically, this corresponds to the proportionality between the angular momentum and the magnetic dipole moment of a spinning charged sphere, both are vectors parallel to the rotation axis whose length increases proportional to the spinning frequency. The magnetic moment and its interaction with magnetic fields allow the observation of NMR signal associated with transitions between nuclear spin levels during resonant RF irradiation or caused by Larmor precession of the average magnetic moment after resonant irradiation [80]. Electron spin resonance (ESR) is a related technique in which transitions between electronic rather than nuclear spin levels are detected. The basic principles are similar, but the instrumentation, data analysis, and detailed theory differ significantly. Nuclear spin is an intrinsic angular momentum that is quantized. This indicates that the magnitude of

angular momentum is quantized (i.e., the spin S can only assume a limited range of values). Additionally, the angular momentum's x, y, and z components are quantized, constrained to integer or half-integer multiples of $\hbar$, the reduced Planck constant. The integer or half-integer quantum number associated with the spin component along the z-axis or the applied magnetic field is known as the magnetic quantum number, m, and can take values from $+S$ to $-S$, in integer steps. Hence, there are a total of angular momentum states for any given nucleus $2S + 1$. Therefore, the z-component of the angular momentum vector ($\vec{S}$) is $S_z = m\hbar$. The z-component of the magnetic moment is simply [81]:

$$\mu_z = \gamma S_z = \gamma m \hbar \,. \tag{2.7}$$

Consider a nucleus with a spin of one-half, like ^1H; its nucleus has two linearly independent spin states, with $m = \frac{1}{2}$ or $m = -\frac{1}{2}$ (also referred to as spin-up and spin-down, or sometimes α and β spin states, respectively) for the z-component of spin. In the absence of a magnetic field, these states are degenerate; that is, they have the same energy. Hence the number of nuclei in these two states will be equal at thermal equilibrium.

Suppose a nucleus with spin is placed in a magnetic field. However, the two states no longer have the same energy due to the interaction between the nuclear magnetic dipole moment and the external magnetic field. The energy of a magnetic dipole moment $m\vec{u}$ in a magnetic field $\vec{B}_0$ is given by [82]:

$$E = -\vec{\mu} \cdot \vec{B}_0 = -\mu_x B_{0x} - \mu_y B_{0y} - \mu_z B_{0z} \,. \tag{2.8}$$

Usually the z-axis is chosen to be along $\vec{B}_0$, and the above expression reduces to:

$$E = -\mu_z B_0, \tag{2.9}$$

or alternatively:

$$E = -\gamma m \hbar B_0 \,. \tag{2.10}$$

As a result, the different nuclear spin states have different energies in a non-zero magnetic field. In less formal language, we can talk about the two spin states of a spin $\frac{1}{2}$ as being aligned either with or against the magnetic field. If γ is positive (true for most isotopes used in NMR) then $m = -\frac{1}{2}$ (spin-up) is the lower energy state. The energy difference between the two states is:

$$\Delta E = \gamma \hbar B_0. \tag{2.11}$$

This results in a small population bias favoring the lower energy state in thermal equilibrium. With more spins pointing up than down, a net spin magnetization along the magnetic field $\vec{B}_0$ results. A central concept in NMR is the precession of the spin magnetization around the magnetic field at the nucleus, with the angular frequency:

$$\omega = -\gamma B, \tag{2.12}$$

where $\omega = 2\pi\nu$ relates to the oscillation frequency ν and B is the magnitude of the field. This means that the spin magnetization, which is proportional to the sum of the spin vectors of nuclei in magnetically equivalent sites (the expectation value of the spin vector in quantum mechanics), moves in a conical path around the B field. This is analogous to the precessional motion of the axis of a tilted spinning top around the gravitational field. In quantum mechanics, ω is the Bohr frequency $\Delta E/\hbar$ of the S_x and S_y expectation values [83]. Precession of non-equilibrium magnetization in the applied magnetic field $\vec{B}_0$ occurs with the Larmor frequency:

$$\omega_L = 2\pi\nu_L = -\gamma B_0\,. \tag{2.13}$$

Without change in the populations of the energy levels because energy is constant (time-independent Hamiltonian).

The MRI machine generates a very strong static magnetic field, typically in the range of 1.5 to 3 Tesla (T), although some machines can go up to 7T for research purposes. When a patient is placed inside this magnetic field, the hydrogen protons in their body align either with or against the direction of the magnetic field. Since

there are more protons aligned with the field, this creates a net magnetization. Next, the MRI scanner emits a pulse of RF energy at a specific frequency, temporarily disrupting the hydrogen protons' alignment. This pulse causes the protons to absorb energy and flip their alignment from a low-energy state to a high-energy one. When the RF pulse is turned off, the protons return to their original alignment, releasing the absorbed energy as a detectable signal. Two critical properties define how the hydrogen protons release this absorbed energy, allowing the MRI scanner to create images [84]:

- T1 Relaxation (longitudinal relaxation): This is the time it takes for protons to realign with the magnetic field after turning off the RF pulse. T1 relaxation times vary between different tissue types and provide contrast between them.
- T2 Relaxation (transverse relaxation): This is the time it takes for protons to lose coherence with each other, dephasing after the RF pulse is turned off. T2 relaxation times also vary between tissues and contribute to image contrast, especially in tissues with high water content.

As the protons return to their low-energy state, they emit RF signals. These signals are detected by a coil in the MRI machine and processed by a computer to form images. The differences in relaxation times between tissues (e.g., fat, muscle, water, etc.) generate contrast in the images. Sequences such as T1 and T2-weighted imaging highlight different tissue types and abnormalities. Magnetic field gradients vary the magnetic field across the patient's body. These gradients enable spatial encoding, allowing the MRI scanner to localize the signal from different body areas. By adjusting the gradients, images can be acquired in various planes (axial, sagittal, coronal) and at different resolutions.

In 1983, Ljunggren and Twieg introduced the k-space formalism, which proved invaluable in unifying different MR imaging techniques [85]. They showed that the demodulated MR signal $S(t)$ generated by the interaction between an ensemble of freely precessing nuclear spins in the presence of a linear magnetic field gradient G

and a receiver-coil equals the Fourier transform of the effective spin density, $\rho(\vec{x})$. Fundamentally, the signal is derived from Faraday's law of induction [86]:

$$S(t) = \widetilde{\rho}_{eff}\left(\vec{k}(t)\right) = \int_{-\infty}^{\infty} d\vec{x}\rho(\vec{x}) \cdot e^{2\pi i\vec{k}(t)\cdot\vec{x}}, \tag{2.14}$$

where:

$$\vec{k}(t) = \int_0^t \vec{G}(\tau)d\tau\,. \tag{2.15}$$

In other words, as time progresses the signal traces out a trajectory in k-space with the velocity vector of the trajectory proportional to the vector of the applied magnetic field gradient. By the term effective spin density we mean the true spin density $\rho(\vec{x})$ corrected for the effects of T_1 preparation, T_2 decay, dephasing due to field inhomogeneity, flow, diffusion, etc., and any other phenomena that affect that amount of transverse magnetization available to induce a signal in the RF probe or its phase concerning the receiving coil's electromagnetic field. From the basic k-space formula, it follows immediately that we reconstruct an image $I(\vec{x})$ by taking the inverse Fourier transform of the sampled data [87]:

$$I(\vec{x}) = \int_{-\infty}^{\infty} d\vec{k}S\left(\vec{k}(t)\right) \cdot e^{-2\pi\vec{k}(t)\cdot\vec{x}}\,. \tag{2.16}$$

Using the k-space formalism, several seemingly complex ideas became simple. For example, it becomes very easy (for physicists, in particular) to understand the role of phase encoding (the so-called spin-warp method). In a standard spin echo or gradient echo scan, where the readout (or view) gradient is constant (e.g., G), a single k-space line is scanned per RF excitation. When the phase encoding gradient is zero, the line scanned is the k_x axis. When a non-zero phase-encoding pulse is added in between the RF excitation and the commencement of the readout gradient, this line moves up or down in k-space, i.e., indicating that we are scanning the line $k_y = constant$ [88].

The importance of the center of k-space in determining image contrast can be exploited in more advanced imaging techniques. One such technique is spiral acquisition, where a rotating magnetic field gradient is applied, causing the trajectory in

k-space to spiral out from the center to the edge. Due to T_2 and T_2^* decay, the signal is greatest at the start of the acquisition; hence, acquiring the center of k-space first improves contrast to noise ratio (CNR) compared to conventional zig-zag acquisitions, especially in rapid movement. Since $\vec{x}$ and $\vec{k}$ are conjugate variables (concerning the Fourier transform) we can use the Nyquist theorem to show that a step in k-space determines the field of view of the image (the maximum frequency that is correctly sampled) and the maximum value of k sampled determines the resolution; i.e. [89]:

$$FOV(Field\,of\,View) \propto \frac{1}{\Delta k}\ ,\ Resolution \propto |k_{max}|\,. \tag{2.17}$$

While MRI is an invaluable tool in modern diagnostics, it presents several challenges, particularly in terms of image interpretation and variability in tissue quality. Deep learning techniques have become increasingly important in addressing these issues, improving image quality, and automating image analysis. MRI images are rich in detail, but their interpretation can be difficult due to the variability in tissue quality and subtle differences between healthy and diseased tissue. This complexity can lead to differences in diagnosis between radiologists, particularly in difficult cases such as small tumors or early-stage disease. Deep learning models, especially CNNs, have addressed this variability by providing automated and consistent segmentation and classification of MRI images [90]. These models can learn from large datasets of labeled images to identify patterns that might not be obvious to the human eye, improving diagnostic accuracy. Deep learning has revolutionized MRI analysis by automating the interpretation of complex images, improving segmentation accuracy, and enhancing image quality. Noise reduction, super-resolution imaging, and artifact removal techniques driven by AI have all contributed to making MRI a more efficient and accurate diagnostic tool. As the technology continues to evolve, the role of AI in MRI is expected to grow, further improving diagnostic outcomes and patient care [91].

2.5 Deep learning (DL) techniques

Over the past decade, DL has transformed the field of medical imaging, offering unparalleled accuracy in diagnosing diseases by automatically analyzing complex medical images. With its ability to handle large amounts of data and extract intricate features, deep learning has made significant strides in addressing some of the long-standing challenges in medical diagnostics, such as segmentation, classification, and anomaly detection [92].

2.5.1 Objective of DL in medical imaging

The primary objective of deep learning in medical imaging is to enhance the diagnostic process by automating the analysis of medical images. Traditionally, medical imaging interpretation has relied on the expertise of radiologists, who must manually examine and interpret the images. This process is time-consuming, subject to inter-observer variability, and prone to error, especially when dealing with subtle abnormalities [93].

Deep learning, particularly convolutional neural networks, has revolutionized this area by automating tasks such as identifying tumors, segmenting organs, and classifying diseases. These models can learn from vast datasets of labeled medical images, enabling them to generalize and perform tasks with high accuracy across various medical conditions.

2.5.2 Applications of DL in medical imaging

Deep learning has been successfully applied across multiple imaging modalities, including X-ray, CT, MRI, ultrasound, and histopathology images. Below are some key applications where CNNs and other deep learning architectures have been effectively used in medical diagnostics:

- **Segmentation** [94]: Image segmentation is a critical task in medical imaging where the goal is to delineate anatomical structures, lesions, or tumors. Accurate segmentation is essential for diagnosis, treatment planning, and monitoring

of disease progression. Deep learning techniques, particularly the U-Net architecture, have been highly successful in medical image segmentation. The U-Net, with its encoder-decoder structure, is designed to capture high-level semantic information and fine-grained details of the input image. This makes it ideal for tasks like segmenting organs, tumors, and blood vessels from CT and MRI scans. In one of our studies, we applied the U-Net architecture to segment kidneys from CT scans of Mexican patients. The model's performance was evaluated based on the Hounsfield Unit similarity coefficient and organs intersection over union (IoU) from CT scans, achieving high accuracy in delineating the kidneys' boundaries. This segmentation task was crucial for the subsequent classification and analysis of kidney pathologies.

- **Classification** [95]: Medical image classification is another significant application of deep learning. CNNs, with their hierarchical structure, can automatically learn features from raw image data, making them well-suited for classification. These networks can differentiate between various classes of diseases, such as distinguishing between benign and malignant tumors or identifying different stages of disease progression. In this research, a classification model was developed and used a pre-trained CNNs such as ResNet50 and VGG16 as feature extractors. These networks were fine-tuned using transfer learning, allowing them to leverage knowledge from large image datasets while adapting to the specific features of the medical images in my dataset. The classification system was applied to detect kidney abnormalities in CT scans, achieving promising accuracy, sensitivity, and specificity results. By using CNN-based feature extraction followed by advanced machine learning classifiers such as support vector machines (SVM), the model demonstrated robustness in identifying the presence of abnormalities.

- **Anomaly detection** [96]: Anomaly detection involves identifying rare and unusual patterns in medical images, which may indicate disease. In many medical datasets, abnormal cases (e.g., cancerous lesions) are relatively rare compared to normal cases, making this a challenging problem. Autoencoders and gen-

erative adversarial networks (GANs) are often used for anomaly detection, as they can learn to represent the normal data distribution and flag deviations as anomalies. A deep learning model was developed and utilized a variational autoencoder (VAE) for detecting kidney abnormalities from CT images. The model was trained on a large dataset of normal kidney images and evaluated on images containing abnormalities. The model successfully identified anomalous regions by comparing the original image with its reconstructed version, highlighting areas that deviated from the learned normal patterns.

2.5.3 Key DL techniques and approaches

Several deep learning techniques and methodologies have been instrumental in improving medical image analysis. These techniques focus on image preprocessing, feature extraction, and classification, forming a complete pipeline for disease diagnosis.

a) **Preprocessing:** Medical images often contain noise, artifacts, and variability in contrast that can hinder deep learning models' performance. Effective preprocessing techniques are essential to enhance image quality and ensure that relevant features are highlighted. In our research, we applied several preprocessing techniques to CT and MRI images, including *noise reduction:* Gaussian filtering and median filtering were used to remove noise and improve the clarity of the images. *Contrast enhancement:* Techniques such as histogram equalization and adaptive contrast enhancement were applied to improve image contrast, making abnormalities more visible. *Normalization:* Image pixel values were normalized to ensure consistent input data for the deep learning model, reducing variability due to lighting or acquisition conditions. Preprocessing was a crucial step in my kidney segmentation project, where removing noise and enhancing contrast led to improved model performance in detecting kidney structures.

b) **Feature extraction:** Feature extraction is one of the core strengths of deep learning models. CNNs, in particular, are adept at automatically learning hierarchical feature representations from raw image data. In this research, sev-

eral pre-trained CNNs were used such as ResNet50, VGG16, MobileNet, InceptionV3, and others as feature extractors for medical images. These networks, trained on large datasets like ImageNet, were fine-tuned using transfer learning to adapt to the specific characteristics of medical images. The convolutional layers of these networks act as filters, detecting edges, textures, and more complex structures as the network goes deeper. For example, in kidney CT scans, the CNNs automatically learn to detect the kidney's shape, size, and texture, which are crucial for identifying abnormalities. The extracted features were then passed to a classification system for the final diagnostic decision.

c) **Classification algorithms:** After feature extraction, the next step is classification, where the goal is to assign the image to a specific category (e.g., normal or abnormal). In our research, we experimented with various classifiers, including *Support vector machines (SVM):* This algorithm is highly effective for binary classification tasks and was used to classify CT images of kidneys into normal or abnormal categories. *Random forests:* This ensemble learning method was applied in some of my studies to classify images based on the features extracted from CNNs. Random forests are robust enough to overfit and work well with high-dimensional data. *Fully connected layers:* In some cases, we used fully connected layers directly on top of the CNN feature extractor for end-to-end classification. These layers output a probability distribution over possible classes, allowing multi-class classification tasks.

2.5.4 Challenges in DL for medical imaging

Despite the significant advancements made possible by deep learning, several challenges remain. Medical image datasets are often limited in size due to privacy concerns and the high cost of acquiring labeled data. Data augmentation and synthetic data generation are often used to mitigate this issue, but the lack of diverse data can still impact the generalizability of deep learning models. Furthermore, the interpretability of deep learning models is another critical challenge. CNNs, while highly effective,

often function as "black boxes," making it difficult to understand how they arrive at specific conclusions. AS such, developing explainable AI (XAI) methods remains a priority in medical AI research [97].

2.5.5 Transfer learning in medical imaging

Transfer learning has become a powerful technique in the field of deep learning, particularly in medical imaging, where labeled datasets are often limited, and computational resources are constrained [98]. By leveraging pre-trained models, transfer learning enables adapting deep learning models trained on large, general datasets to specific medical imaging tasks. This approach significantly enhances segmentation, classification, and anomaly detection performance. In traditional deep learning models, training a network from scratch requires large amounts of labeled data to learn the optimal feature representations. However, medical imaging datasets are often small and specialized, making it difficult to train deep-learning models without overfitting. Transfer learning mitigates this problem by using models pre-trained on large, general datasets (e.g., ImageNet) and adapting them to the medical imaging domain [99]. This approach allows deep learning models to "transfer" the knowledge acquired from solving a general task to a more specific one, significantly improving performance and reducing training time (see Figure 2.14).

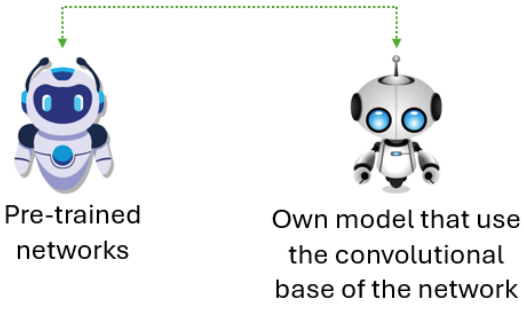


Figure 2.14: Several models use the convolutional layers of pre-trained networks to obtain a characteristic vector of the images in DL projects.

Pre-trained networks such as ResNet, VGG, Inception, and DenseNet have been

widely used in medical imaging tasks [100]. These networks are initially trained on massive datasets containing millions of labeled images across thousands of categories. The initial layers of these models learn low-level features, such as edges, textures, and gradients, while the deeper layers capture more complex and abstract representations. These pre-trained networks can be fine-tuned for specific tasks, making them highly effective for medical image analysis, where feature extraction is critical.

2.5.6 Benefits of transfer learning in medical imaging

Transfer learning provides several key advantages in the field of medical imaging [101]:

1. *Improved feature extraction:* The pre-trained convolutional layers of deep networks, such as those found in ResNet and VGG, have already learned to detect general visual features. By fine-tuning these layers for specific medical imaging tasks, the model can extract relevant features from medical images without extensive retraining. This is particularly useful when working with small medical datasets, where the model's ability to learn features from scratch is limited.

2. *Faster training and convergence:* Training deep learning models from scratch can be computationally expensive and time-consuming, especially when dealing with high-resolution medical images. Transfer learning reduces training time by allowing the network to start with pre-trained weights rather than random initialization. This means the model can converge faster and perform better with fewer training iterations.

3. *Overcoming limited data:* Medical imaging datasets are often limited due to privacy concerns and the difficulty of labeling images. Transfer learning addresses this issue by enabling models to learn from a small dataset while retaining the generalization ability. Pre-trained networks already possess a wealth of learned knowledge from large datasets like ImageNet, which helps them perform well even on small medical datasets.

While transfer learning has proven highly effective in medical imaging, challenges

still need to be addressed. One major issue is the domain shift between natural images (such as those in ImageNet [102]) and medical images. Medical images are often grayscale, have different spatial structures, and may contain more subtle features than natural images. This domain shift can limit the effectiveness of transfer learning, as the pre-trained model's features may not fully capture the nuances of medical images [103].

Moreover, while transfer learning reduces the need for large amounts of labeled data, fine-tuning still requires careful selection of which layers to freeze or adjust. Overfitting can occur if the pre-trained model is too rigid or fine-tuned too aggressively on a small dataset. Careful experimentation is needed to strike the right balance between transfer learning and overfitting.

2.6 Supervised learning

Supervised learning is a type of machine learning where a model is trained using labeled data, meaning each input is paired with the correct output. The goal is for the model to learn the mapping from inputs to outputs to predict the output for new, unseen data accurately. Key components are: labeled data, training process and prediction [104]. In supervised learning, the dataset consists of input-output pairs. The input (features) is the data we want to analyze, and the output (label) is the correct answer or category that corresponds to the input. The model learns by analyzing the input and the labeled output relationships. During training, the model makes predictions and compares them to the true labels. Based on the difference (error) between the prediction and the actual label, the model adjusts its parameters to improve accuracy. Once trained, the model can make predictions on new, unseen data. Since the training data includes the correct answers, the model generalizes from the examples to make informed predictions on similar data. There are types of supervised learning such as classification and regression. In medical diagnostics, supervised learning can classify images of organs or tissues as healthy or diseased based on labeled examples provided by radiologists. For example, a model trained on

CT scans of lungs labeled as cancerous or non-cancerous could learn to detect lung cancer in new patients' scans [105].

Suppose we have a Biomedical dataset, we can learn to predict the pathologies using this one for making predictions. To establish notation for future use, we will use $x^{(i)}$ to denote the input variables (pathologies in this example), also called input features, and $y^{(i)}$ to denote the output or target variable that we are trying to predict. A pair $\left(x^{(i)}, y^{(i)}\right)$ is called a training example, and the dataset that we will use to learn, a list of n training examples, $\left\{(x^{(i)}, y^{(i)}); i = 1, ..., n\right\}$ is called a training set. Note that the superscript (i) in the notation is simply an index into the training set, and has nothing to do with exponentiation. We will also use χ to denote the space of input values and Y to the space of output values. In this example, $\chi = Y = \Re$. To describe the supervised learning problem slightly more formally, our goal is, given a training set, to learn a function $h : \chi \rightarrow Y$ so that $h(x)$ is a good predictor for the corresponding value of y. For historical reasons, this function h is called a hypothesis. The process is, therefore, seen pictorially like this (Figure 2.15). When the target variable that we are trying to predict is continuous, such as in our biomedical dataset example, we call the learning problem a regression problem. When y can take on only a small number of discrete values, we term it a classification problem.

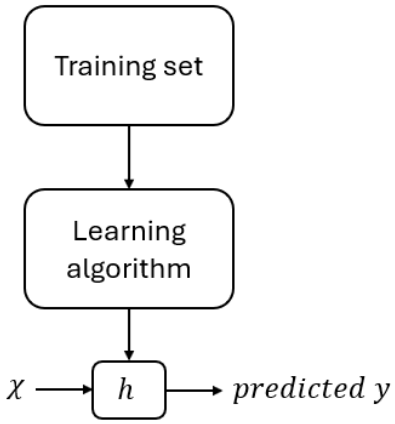


Figure 2.15: Process to obtain the prediction of the dataset.

2.6.1 Linear regression

To make our Biomedical dataset example more interesting, let us consider a slightly richer dataset in which we also know some special features in the pathologies. Hence, the x's are two-dimensional vectors in $\Re^2$. For instance, $x_1^{(i)}$ is the pathology of the i-th pathology in the training set, and $x_2^{(i)}$ is its special feature. In general, when designing a learning problem, it will be up to the user to decide what features to choose. To perform supervised learning, we must decide how to represent functions/hypotheses h in a computer. As an initial choice, let us say we decide to approximate y as a linear function of x [106]:

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2, \tag{2.18}$$

here, the θ_i's are the parameters (also called weights) that parameterize the space of linear functions mapping from χ to Y. When there is no risk of confusion, we will drop the θ subscript in $h_\theta(x)$, and write it more simply as $h(x)$. To simplify our notation, we also introduce the convention of letting $x_0 = 1$ (this is the intercept term), so that:

$$h(x) = \sum_{i=0}^{d} \theta_i x_i = \theta^T x, \tag{2.19}$$

where on the right-hand side above we view θ and x both as vectors, and here, d is the number of input variables (not counting x_0). Given a training set, how do we pick or learn the parameters θ? One reasonable method seems to make $h(x)$ close to y, at least for our training examples. To formalize this, we will define a function that measures, for each value of the θ's how close the $h\left(x^{(i)}\right)$'s are to the corresponding $y^{(i)}$'s. We define the cost function [107]:

$$J(\theta) = \frac{1}{2} \sum_{i=1}^{n} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2 . \tag{2.20}$$

This is familiar; the least-squares cost function seems to give rise to the ordinary least-squares regression model.

2.6.2 Least-mean squares (LMS) algorithm

We want to choose θ to minimize $J(\theta)$. To do so, let us use a search algorithm that starts with some initial guess for θ, and that repeatedly changes θ to make $J(\theta)$ smaller until, hopefully, we converge to a value of θ that minimizes $J(\theta)$. Specifically, let us consider the gradient descent algorithm, which starts with some initial θ, and repeatedly performs the update [108]:

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta), \tag{2.21}$$

here, α is called the learning rate. This is a very natural algorithm that repeatedly takes a step in the direction of the steepest decrease of J. To implement this algorithm, we must determine the partial derivative term on the right-hand side. Let uss first work it out for the case of if we have only one training example (x, y) so that we can neglect the sum in the definition of J. We have:

$$\begin{aligned}
\frac{\partial}{\partial \theta_j} J(\theta) &= \frac{\partial}{\partial \theta_j} \frac{1}{2} \left(h_\theta(x) - y\right)^2, \\
&= 2 \cdot \frac{1}{2} \left(h_\theta(x) - y\right) \cdot \frac{\partial}{\partial \theta_j} \left(h_\theta(x) - y\right), \\
&= \left(h_\theta(x) - y\right) \cdot \frac{\partial}{\partial \theta_j} \left(\sum_{i=0}^{d} \theta_i x_i - y \right), \\
&= \left(h_\theta(x) - y\right) x_j .
\end{aligned} \tag{2.22}$$

This gives the update rule for a single training example: We use the notation $a := b$ to denote an operation (in a computer program) in which we set the value of a variable a to be equal to the value of b. In other words, this operation overwrites a with the value of b. In contrast, we will write "$a = b$" when we are asserting a statement of fact, that the value of a is equal to the value of b.

$$\theta_j := \theta_j + \alpha \left(y^{(i)} - h_\theta(x^{(i)})\right) x_j^{(i)} . \tag{2.23}$$

The rule is called the LMS update rule and is also known as the Widrow-Hoff

learning rule [109]. This rule has several properties that seem natural and intuitive. For instance, the magnitude of the update is proportional to the error term $(y^{(i)} - h_\theta(x^{(i)}))$; thus, for instance, if we are encountering a training example on which our prediction nearly matches the actual value of $y^{(i)}$, then we find that there is little need to change the parameters; in contrast, a larger change to the parameters will be made if our prediction $h_\theta(x^{(i)})$ has a large error (i.e. if it is very far from $y^{(i)}$). We would have derived the LMS rule for when there was only a single training example. There are two ways to modify this method for a training set of more than one example. The first is to replace it with the following algorithm:

Repeat until convergence {

$$\theta_j := \theta_j + \alpha \sum_{i=1}^{n} \left(y^{(i)} - h_\theta(x^{(i)})\right) x_j^{(i)} \ , \ (for\ every\ j), \tag{2.24}$$

}

By grouping the updates of the coordinates into an update of the vector θ, we can rewrite the update Equation (2.24) in a slightly shorter way:

$$\theta := \theta + \alpha \sum_{i=1}^{n} \left(y^{(i)} - h_\theta(x^{(i)})\right) x^{(i)} . \tag{2.25}$$

The quantity in the summation in Equation (2.25) is just the Equation (2.22) (for the original definition of J). So, this is simply a gradient descent on the original cost function J. This method looks at every example in the entire training set on every step and is called batch gradient descent. Note that while gradient descent can be susceptible to local minima in general, the optimization problem we have posed here for linear regression has only one global, and no other local optima. Thus, gradient descent always converges (assuming the learning rate α is not too large) to the global minimum. Indeed, J is a convex quadratic function. Figure 2.16 shows an example of gradient descent as it is run to minimize a quadratic function [110].

The ellipses shown above are the contours of a quadratic function. Also shown is the trajectory taken by gradient descent, which was initialized at $(48, 30)$. The x's in the figure (joined by straight lines) mark the successive values of θ that the

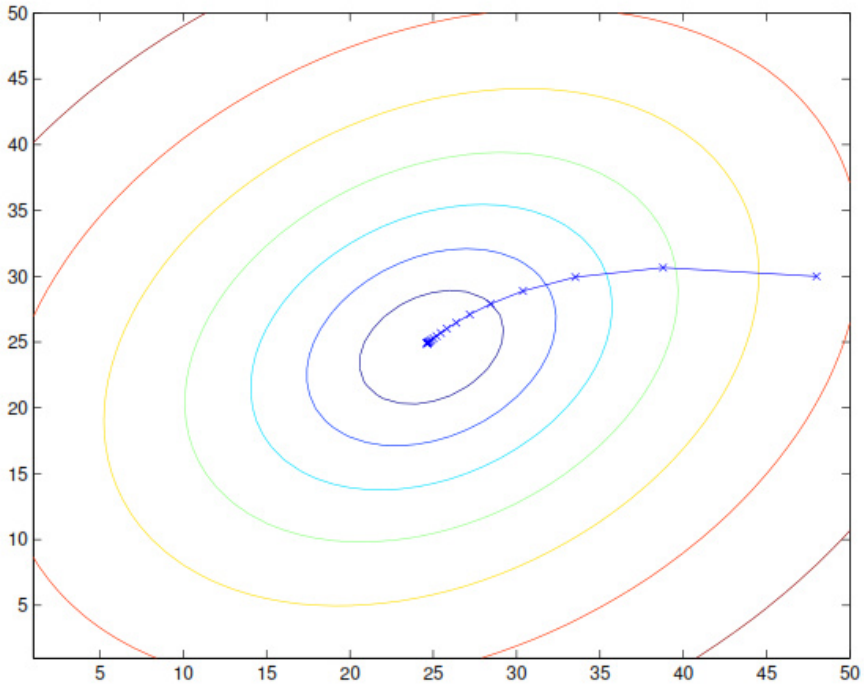


Figure 2.16: Gradient descent.

gradient descent went through. Gradient descent gives one way of minimizing J. Let us discuss a second way. This time the minimization is performed explicitly without an iterative algorithm. In this method, we will minimize J by explicitly taking its derivatives for the θ_j's and setting them to zero. To enable us to do this without having to write reams of algebra and pages full of matrices of derivatives, let us introduce some notation for doing calculus with matrices.

2.6.3 Matrix derivatives

For a function $f : \Re^{n \times d} \to \Re$ mapping from n by d matrices to the real numbers, we define the derivative of f concerning A to be [111]:

$$\nabla_A f(A) = \begin{bmatrix} \frac{\partial f}{\partial A_{11}} & \cdots & \frac{\partial f}{\partial A_{1d}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f}{\partial A_{n1}} & \cdots & \frac{\partial f}{\partial A_{1n}} \end{bmatrix}, \tag{2.26}$$

Thus, the gradient $\nabla_A f(A)$ is an n by d matrix, whose (i, j)-element is $\partial f / \partial A_{ij}$.

For example, suppose $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$, is a 2 by 2 matrix, and the function $f : \Re^{2\times 2} \to \Re$ is given by:

$$f(A) = \frac{3}{2}A_{11} + 5A_{12}^2 + A_{21}A_{22}, \tag{2.27}$$

here, A_{ij} denotes the (i, j) entry of the matrix A. We then have:

$$\nabla_A f(A) = \begin{bmatrix} \frac{3}{2} & 10A_{12} \\ A_{22} & A_{21} \end{bmatrix}. \tag{2.28}$$

Armed with the tools of matrix derivatives, let us now proceed to find in closed form value of θ that minimizes $J(\theta)$. We begin by re-writing J in matrix-vectorial notation. Given a training set, define the design matrix X to be the n by d matrix (actually n-by-d +1, if we include the intercept term) that contains the training examples' input values in its rows [112]:

$$X = \begin{bmatrix} -(x^{(1)})^T- \\ -(x^{(2)})^T- \\ \vdots \\ -(x^{(n)})^T- \end{bmatrix}. \tag{2.29}$$

Also, let $\vec{y}$ be the n-dimensional vector containing all the target values from the training set:

$$\vec{y} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(n)} \end{bmatrix}, \tag{2.30}$$

now, since $h_\theta(x^{(i)}) = (x^{(i)})^T\theta$, we can easily verify that:

$$
\begin{aligned}
X\theta - \vec{y} &= \begin{bmatrix} (x^{(1)})^T\theta \\ \vdots \\ (x^{(n)})^T\theta \end{bmatrix} - \begin{bmatrix} y^{(1)} \\ \vdots \\ y^{(n)} \end{bmatrix}, \\
&= \begin{bmatrix} h_\theta(x^{(1)}) - y^{(1)} \\ \vdots \\ h_\theta(x^{(n)}) - y^{(n)} \end{bmatrix}.
\end{aligned} \tag{2.31}
$$

Thus, using the fact that for a vector z, we have that $z^T z = \sum_i z_i^2$:

$$
\begin{aligned}
\frac{1}{2}(X\theta - \vec{y})^T(X\theta - \vec{y}) &= \frac{1}{2}\sum_{i=1}^{n}(h_\theta(x^{(i)}) - y^{(i)})^2, \\
&= J(\theta)\,.
\end{aligned} \tag{2.32}
$$

Finally, to minimize J, let us find its derivatives concerning θ. Hence:

$$
\begin{aligned}
\nabla_\theta J(\theta) &= \nabla_\theta \frac{1}{2}(X\theta - \vec{y})^T(X\theta - \vec{y}), \\
&= \frac{1}{2}\nabla_\theta \left((X\theta)^T X\theta - (X\theta)^T\vec{y} - \vec{y}^T(X\theta) + \vec{y}^T\vec{y}\right), \\
&= \frac{1}{2}\nabla_\theta \left(\theta^T(X^TX)\theta - \vec{y}^T(X\theta) - \vec{y}^T(X\theta)\right), \\
&= \frac{1}{2}\nabla_\theta(\theta^T(X^TX)\theta - 2(X^T\vec{y})\theta), \\
&= \frac{1}{2}(2X^TX\theta - 2X^T\vec{y}), \\
&= X^TX\theta - X^T\vec{y}\,.
\end{aligned} \tag{2.33}
$$

In the third step, we used the fact that $a^Tb = b^Ta$, and in the fifth step we used the facts $\nabla_x b^T x = b$ and $\nabla_x x^T A x = 2Ax$ for symmetric matrix A. To minimize J, we set its derivatives to zero and obtain the normal equations:

$$
X^TX\theta = X^T\vec{y}. \tag{2.34}
$$

Thus, the value of θ that minimizes $J(\theta)$ is given in closed form by the equation:

$$\theta = (X^T X)^{-1} X^T \vec{y}. \tag{2.35}$$

2.6.4 Probabilistic interpretation

When faced with a regression problem, why might linear regression, and specifically why might the least-squares cost function J, be a reasonable choice? In this section, we will give a set of probabilistic assumptions, under which least-squares regression is derived as a very natural algorithm. Let us assume that the target variables and the inputs are related via the equation [113]:

$$y^{(i)} = \theta^T x^{(i)} + \epsilon^{(i)}, \tag{2.36}$$

where $\epsilon^{(i)}$ is an error term that captures either unmodeled effects (such as if some features are pertinent to predict, but that we had left out of the regression) or random noise. Let us further assume that the $\epsilon^{(i)}$ are distributed IID (independently and identically distributed) according to a Gaussian distribution (also called a Normal distribution) with mean zero and some variance σ^2. We can write this assumption as $\epsilon^{(i)} \sim \aleph(0, \sigma^2)$. In other words, the density is given by:

$$p\left(\epsilon^{(i)}\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(e^{(i)})^2}{2\sigma^2}\right), \tag{2.37}$$

this implies that [114]:

$$p(y^{(i)}|x^{(i)};\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^2}\right). \tag{2.38}$$

The notation $p(y^{(i)}|x^{(i)};\theta)$ indicates that this is the distribution of $y^{(i)}$ given $x^{(i)}$ and parameterized by θ. Note that we should not condition on $p(y^{(i)}|x^{(i)},\theta)$, since θ is not a random variable. We can also write the distribution of $y^{(i)}$ as $y^{(i)}|x^{(i)};\theta \sim \aleph(\theta^T x^{(i)}, \sigma^2)$. Given X (the design matrix, which contains all the $x^{(i)}$'s) and θ, what is the distribution of the $y^{(i)}$'s? The probability of the data is given by $p(\vec{y}|X;\theta)$.

This quantity is typically viewed as a function of $\vec{y}$ (and perhaps X) for a fixed value of θ when we wish to view this as a function of θ explicitly, we will instead call it the likelihood function:

$$L(\theta) = L(\theta; X; \vec{y}) = p(\vec{y}|X; \theta)\,. \tag{2.39}$$

Note that by the independence assumption on the $\epsilon^{(i)}$'s (and hence also the $y^{(i)}$'s given the $x^{(i)}$'s), this can also be written:

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} p(y^{(i)}|x^{(i)}; \theta), \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^2}\right)\,. \end{aligned} \tag{2.40}$$

Now, given this probabilistic model relating the $y^{(i)}$'s and the $x^{(i)}$'s, what is a reasonable way of choosing our best guess of the parameters θ? maximum likelihood principle says that we should choose θ to make the data as high probability as possible, i.e., we should choose θ to maximize $L(\theta)$.

Instead of maximizing $L(\theta)$, we can also maximize any strictly increasing function of it. In particular, the derivations will be a bit simpler if we instead maximize the log-likelihood $l(\theta)$ [115]:

$$\begin{aligned} l(\theta) &= \log L(\theta), \\ &= \log \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^2}\right), \\ &= \sum_{i=1}^{n} \log \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y^{(i)} - \theta^T x^{(i)})^2}{2\sigma^2}\right), \\ &= n \log \frac{1}{\sqrt{2\pi}\sigma} - \frac{1}{\sigma^2} \cdot \frac{1}{2} \sum_{i=1}^{n} (y^{(i)} - \theta^T x^{(i)})^2, \end{aligned} \tag{2.41}$$

hence, maximizing $l(\theta)$ gives the same answer as minimizing:

$$\frac{1}{2}\sum_{i=1}^{n}(y^{(i)} - \theta^T x^{(i)})^2 \,. \tag{2.42}$$

Which we recognize to be $J(\theta)$, our original least-squares cost function. Under the previous probabilistic assumptions on the data, least-squares regression corresponds to finding the maximum likelihood estimate of θ. Thus, this is one set of assumptions under which least-squares regression can be justified as a natural method that only estimates maximum likelihood. Note, however, that the probabilistic assumptions are by no means necessary for least-squares to be a perfectly good and rational procedure, and there may be, and indeed there are other natural assumptions that can also be used to justify it.

2.7 Classification and logistic regression

Let us now talk about the classification problem. This is just like the regression problem, except that the values y we now want to predict take on only a small number of discrete values. For now, we will focus on the binary classification problem in which y can take on only two values, 0 and 1. For instance, if we are trying to build a spam classifier for email, then $x^{(i)}$ may be some features of a piece of email, and y may be 1 if it is a piece of spam mail, and 0 otherwise. 0 is also called the negative class, and 1 the positive class, and they are sometimes also denoted by the symbols "$-$" and "$+$". Given $x^{(i)}$, the corresponding $y^{(i)}$ is also called the label for the training example.

2.7.1 Logistic regression

We could approach the classification problem by ignoring that y is discrete-valued, and using our old linear regression algorithm to predict y given x. However, it is easy to construct examples where this method performs poorly. Intuitively, it also does not make sense for $h_\theta(x)$ to take values larger than 1 or smaller than 0 when we know that $y \in \{0, 1\}$. To fix this, let us change the form for our hypotheses $h_\theta(x)$. We will

choose [116]:

$$h_\theta(x) = g\left(\theta^T x\right) = \frac{1}{1 + e^{-\theta^T x}}, \tag{2.43}$$

where:

$$g(z) = \frac{1}{1 + e^{-z}}\,. \tag{2.44}$$

Equation 2.44 is c called the logistic function or the sigmoid function [117]. Here is a plot showing $g(z)$ in Figure 2.17. Notice that $g(z)$ tends towards 1 as $z \to \infty$, and $g(z)$, and hence also $h(x)$ is always bounded between 0 and 1. As before, we are keeping the convention of letting $x_0 = 1$, so that $\theta^T x = \theta_0 + \sum_{j=1}^{d} \theta_j x_j$.

For now, let us take the choice of g as given. Other functions that smoothly increase from 0 to 1 can also be used, but the choice of the logistic function is fairly natural. Before moving on, here is a useful property of the derivative of the sigmoid function, which we write as g':

$$\begin{aligned} g'(z) &= \frac{d}{dz}\frac{1}{1 + e^{-z}}, \\ &= \frac{1}{(1 + e^{-z})^2}\left(e^{-z}\right), \\ &= \frac{1}{1 + e^{-z}} \cdot \left(1 - \frac{1}{1 + e^{-z}}\right), \\ &= g(z)(1 - g(z))\,. \end{aligned} \tag{2.45}$$

So, given the logistic regression model, how do we fit θ for it? Following how we saw least squares regression could be derived as the maximum likelihood estimator under a set of assumptions, let us endow our classification model with a set of probabilistic assumptions, and then fit the parameters via maximum likelihood.

Let us assume that:

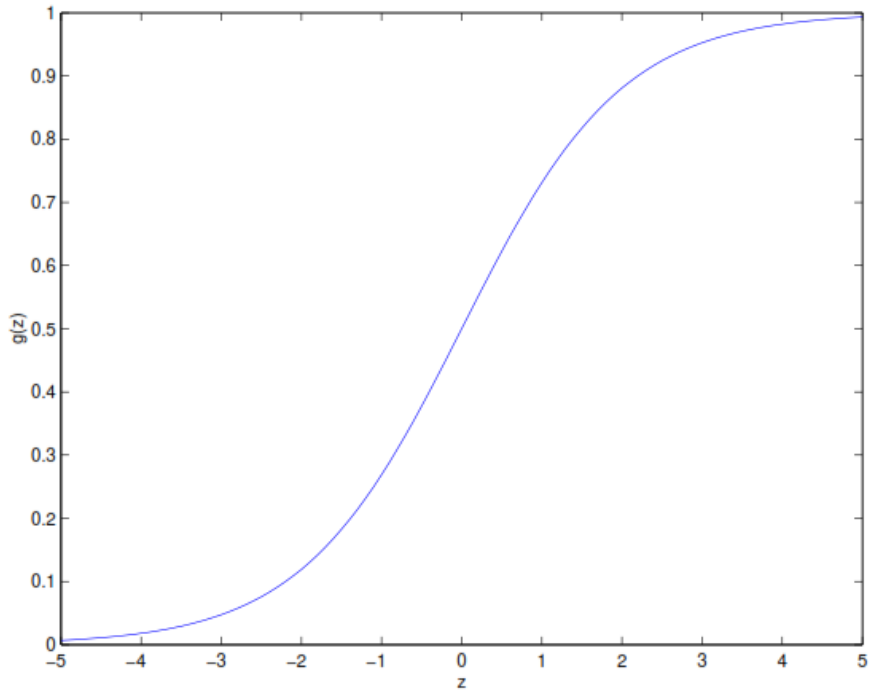


Figure 2.17: Graph of $g(z)$: sigmoid function.

$$\begin{aligned} P(y = 1|x; \theta) &= h_\theta(x), \\ P(y = 0|x; \theta) &= 1 - h_\theta(x), \end{aligned} \tag{2.46}$$

note that this can be written more compactly as:

$$p(y|x; \theta) = (h_\theta(x)^y (1 - h_\theta(x))^{1-y}\,. \tag{2.47}$$

Assuming that the n training examples were generated independently, we can then write down the likelihood of the parameters as:

$$\begin{aligned} L(\theta) &= p(\vec{y}|X; \theta), \\ &= \prod_{i=1}^{n} p(y^{(i)}|x^{(i)}; \theta), \\ &= \prod_{i=1}^{n} \left(h_\theta(x^{(i)})\right)^{y^{(i)}} \left(1 - h_\theta(x^{(i)})\right)^{1-y^{(i)}}, \end{aligned} \tag{2.48}$$

as before, it will be easier to maximize the log-likelihood:

$$l(\theta) = \log L(\theta) = \sum_{i=1}^{n} y^{(i)} \log h(x^{(i)}) + (1 - y^{(i)}) \log(1 - h(x^{(i)})) \,. \tag{2.49}$$

How do we maximize the likelihood? Similar to our derivation in the case of linear regression, we can use gradient ascent. In vectorial notation, our updates will be given by $\theta := \theta + \alpha \nabla_\theta l(\theta)$. Note the positive rather than negative sign in the update formula, since we are now maximizing, rather than minimizing a function. Let us start by working with just one training example (x, y) and take derivatives to derive the stochastic gradient ascent rule:

$$\begin{aligned}
\frac{\partial}{\partial \theta_j} l(\theta) &= \left(y \frac{1}{g(\theta^T x)} - (1 - y) \frac{1}{1 - g(\theta^T x)} \right) \frac{\partial}{\partial \theta_j} g(\theta^T x), \\
&= \left(y \frac{1}{g(\theta^T x)} - (1 - y) \frac{1}{1 - g(\theta^T x)} \right) g(\theta^T x)(1 - g(\theta^T x)) \frac{\partial}{\partial \theta_j} \theta^T x, \\
&= (y(1 - g(\theta^T x)) - (1 - y) g(\theta^T x)) x_j, \\
&= (y - h_\theta(x)) x_j \,.
\end{aligned} \tag{2.50}$$

Above, we used the fact of the Equation 2.45. This therefore gives us the stochastic gradient ascent rule:

$$\theta_j := \theta_j + \alpha (y^{(i)} - h_\theta(x^{(i)})) x_j^{(i)} \,. \tag{2.51}$$

If we compare this to the LMS update rule, we see that it looks identical; but this is not the same algorithm, because $h_\theta(x^{(i)})$ is now defined as a non-linear function of $\theta^T x^{(i)}$. Nonetheless, it is surprising that we end up with the same update rule for a rather different algorithm and learning problem.

2.7.2 The perceptron learning algorithm

We now digress to talk briefly about an algorithm of some historical interest, which we will also return to later when we talk about learning theory. Consider modifying the logistic regression method to "force" it to output values that are either 0 or 1. To do so, it seems natural to change the definition of g to be the threshold function [118]:

$$g(z) = \begin{cases} 1, & \text{if } z \geq 0 \\ 0, & \text{if } z < 0 \end{cases}, \tag{2.52}$$

if we then let $h_\theta = g(\theta^T x)$ as before but use this modified definition of g, and if we use the update rule:

$$\theta_j := \theta_j + \alpha(y^{(i)} - h_\theta(x^{(i)}))x_j^{(i)}\,. \tag{2.53}$$

Then we have the perceptron learning algorithm. In the 1960s, this "perceptron" was argued to be a rough model for how individual neurons in the brain work. Note however that even though the perceptron may be cosmetically similar to the other algorithms we talked about, it is a very different type of algorithm than logistic regression and least squares linear regression; in particular, it is difficult to endow the perceptron's predictions with meaningful probabilistic interpretations or derive the perceptron as a maximum likelihood estimation algorithm.

2.8 Neural networks

Neural networks refer to a broad type of non-linear models/parametrizations $\bar{h}_\theta(x)$ that involve combinations of matrix multiplications and other entrywise non-linear operations. To have a unified treatment for regression and classification problems, here we consider $\bar{h}_\theta(x)$ as the neural network output [119]. For the regression problem, the final prediction $h_\theta(x) = \bar{h}_\theta(x)$, and for the classification problem, $\bar{h}_\theta(x)$ is the logits and the predicted probability will be $h_\theta(x) = 1/(1 + \exp(-\bar{h}_\theta(x)))$ for binary classification or $h_\theta(x) = softmax(\bar{h}_\theta(x))$ for multi-class classification. We will start

small and slowly build up a neural network, step by step.

2.8.1 A neural networks with a single neuron

Let us propose a housing price prediction problem: given the house's size, we want to predict the price. We will use it as a running example. Figure 2.18 shows a graph of size versus housing price. If we fit a straight line to the graph we can observe the trend. Now, instead of fitting a straight line, we wish to prevent negative housing prices by setting the absolute minimum price as zero. This produces a "kink" in the graph. How do we represent such a function with a single kink as $\bar{h}_\theta(x)$ with an unknown parameter? We define a parameterized function $h_\theta(x)$ with input x, parameterized by θ, which outputs house y's price. Formally, $\bar{h}_\theta : x \rightarrow y$.

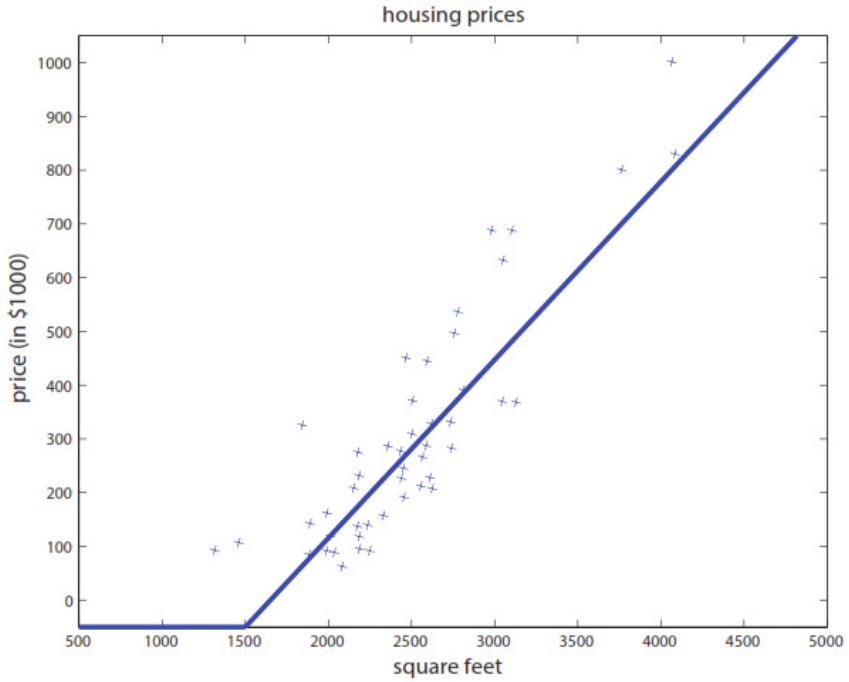


Figure 2.18: Housing prices with a "kink" in the graph.

Perhaps one of the simplest parametrizations would be [120]:

$$\bar{h}_\theta(x) = \max(\omega x + b, 0), \quad \text{where } \theta = (\omega, b) \in \Re^2, \tag{2.54}$$

here, $\bar{h}_\theta(x)$ returns a single value: $(\omega x + b)$ or zero, whichever is greater. In the

context of neural networks, the function $\max\{t, 0\}$ is called a ReLU (pronounced "ray-lu"), or rectified linear unit, and often denoted by $ReLU(t) = \max\{t, 0\}$. Generally, a one-dimensional non-linear function that maps $\Re$ to $\Re$, such as ReLU, is often referred to as an activation function. The model $\bar{h}_\theta(x)$ is said to have a single neuron partly because it has a single non-linear activation function. When the input $x \in \Re^d$ has multiple dimensions, a neural network with a single neuron can be written as:

$$\bar{h}_\theta(x) = ReLU(\omega^\top x + b) \text{ where } \omega \in \Re^d, b \in \Re, \text{ and } \theta = (\omega, b)\,. \tag{2.55}$$

The term b is often referred to as the bias, and the vector ω is referred to as the weight vector. Such a neural network has a 1 layer. A more complex neural network may take the single neuron described above and "stack" them together such that one neuron passes its output as input into the next neuron, resulting in a more complex function. Let us now deepen the housing prediction example. In addition to the size of the house, suppose that you know the number of bedrooms, the zip code, and the wealth of the neighborhood. Building neural networks is analogous to Lego bricks: you take individual bricks together to build complex structures. The same applies to neural networks: we stack individual neurons together to create complex neural networks. Given these features (size, number of bedrooms, zip code, and wealth), we might then decide that the price of the house depends on the maximum family size it can accommodate. Suppose the family size is a function of the size of the house and the number of bedrooms (see Figure 2.19).

The zip code may provide additional information such as how walkable the neighborhood is (e.g., can you walk to the grocery store or do you need to drive everywhere?). Combining the zip code with the wealth of the neighborhood may predict the quality of the local elementary school. Given these three derived features (family size, walkability, and school quality), we may conclude that the price of the home ultimately depends on these three features.

Formally, the input to a neural network is a set of input features x_1, x_2, x_3, x_4. We

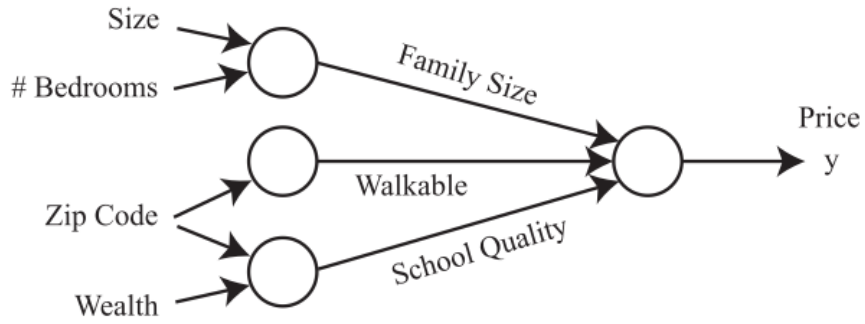


Figure 2.19: Diagram of a small neural network for predicting housing prices.

denote the intermediate variables for "family size", "walkable", and "school quality" by a_1, a_2, a_3 (these a_i's are often referred to as "hidden units" or "hidden neurons"). We represent each of the a_i's as a neural network with a single neuron with a subset of $x_1, ..., x_4$ as inputs. Then, as in Figure 2.18, we will have the parameterization:

$$a_1 = ReLU(\theta_1 x_1 + \theta_2 x_2 + \theta_3), \tag{2.56}$$

$$a_2 = ReLU(\theta_4 x_3 + \theta_5), \tag{2.57}$$

$$a_3 = ReLu(\theta_6 x_3 + \theta_7 x_4 + \theta_8), \tag{2.58}$$

where $(\theta_1, ..., \theta_8)$ are parameters. Now we represent the final output $h_\theta(x)$ as another linear function with a_1, a_2, a_3 as inputs, and we get (Typically, for multi-layer neural network, at the end, near the output, we do not apply ReLU, especially when the output is not necessarily a positive number):

$$\bar{h}_\theta(x) = \theta_9 a_1 + \theta_{10} a_2 + \theta_{11} a_3 + \theta_{12}, \tag{2.59}$$

where θ contains all the parameters $(\theta_1, ..., \theta_{12})$. Now we represent the output as a quite complex function of x with parameters θ.

2.8.2 Inspiration from biological neural networks

As the name suggests, biological neural networks inspired artificial. The hidden units $a_1, ..., a_m$ correspond to the neurons in a biological neural network, and the parameters θ_i's correspond to the synapses. However, it is unclear how similar the modern deep artificial neural networks are to the biological ones. For example, perhaps not many neuroscientists think biological neural networks could have 1000 layers, while some modern artificial neural networks do. Moreover, it is an open question whether human brains update their neural networks in a way similar to how computer scientists learn artificial neural networks [121]. In Figure 2.20, we can observe: (a) It consists of a cell body (soma), a branching tree or dendrites, and an axon. The input information channel is the dendrites, the processor is the body or soma and the axon represents the output channel. The connection between them is given by the synapse electrically and chemically. (b) Where x_n represents input signals, ω_{in} synaptic weights, $\sum$ is the summing junction (bias) followed by the activation function, and y_i is the output.

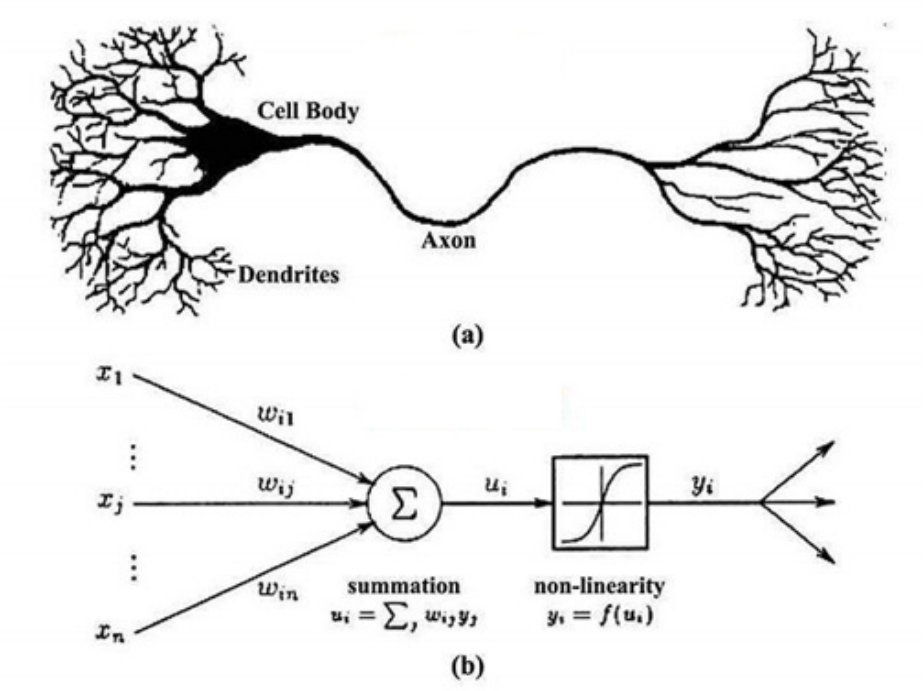


Figure 2.20: Comparison between a biological neuron (a) and an artificial neuron (b).

2.8.3 Two-layer fully-connected neural networks

We have constructed the neural network in Equation 2.59 using a significant amount of prior knowledge/belief about how the "family size", "walkable", and "school quality" are determined by the inputs. We implicitly assumed that we know the family size is an important quantity to look at and it can be determined by only the "size" and "# bedrooms". Such prior knowledge might not be available for other applications. It would be more flexible and general to have a generic parameterization. A simple way would be to write the intermediate variable a_1 as a function of all $x_1, ..., x_4$:

$$a_1 = ReLU(\omega_1^\top x + b_1), \text{ where } \omega_1 \in \Re^4 \text{ and } b_1 \in \Re, \tag{2.60}$$

$$a_2 = ReLU(\omega_2^\top x + b_2), \text{ where } \omega_2 \in \Re^4 \text{ and } b_2 \in \Re, \tag{2.61}$$

$$a_3 = ReLu(\omega_3^\top x + b_3), \text{ where } \omega_3 \in \Re^4 \text{ and } b_3 \in \Re\,. \tag{2.62}$$

We still define $\bar{h}_\theta(x)$ using Equation 2.59 with a_1, a_2, a_3 defined as above. Thus we have a so-called fully connected neural network because all the intermediate variables a_i's depend on all the inputs x_i's. For full generality, a two-layer fully-connected neural network with m hidden units and d dimensional input $x \in \Re^d$ is defined as [122]:

$$\forall j \in [1, ..., m], \quad z_j = {\omega_j^{[1]}}^\top x + b_j^{[1]} \text{ where } \omega_j^{[1]} \in \Re^d,\, b_j^{[1]} \in \Re, \tag{2.63}$$

$$a_j = ReLU(z_j), \tag{2.64}$$

$$a = [a_1, ..., a_m]^\top \in \Re^m, \tag{2.65}$$

$$\bar{h}_\theta(x) = {\omega^{[2]}}^\top a + b^{[2]} \text{ where } \omega^{[2]} \in \Re^m,\, b^{[2]} \in \Re, \tag{2.66}$$

note that by default the vectors in $\Re^d$ are viewed as column vectors, and in particular a is a column vector with components $a_1, a_2, ..., a_m$. The indices $^{[1]}$ and $^{[2]}$ are used to distinguish two sets of parameters: the $w_j^{[1]}$'s (each of which is a vector

in $\Re^d$) and $\omega^{[2]}$ (which is a vector in $\Re^m$). We will have more of these later.

Before introducing neural networks with more layers and complex structures, we will simplify the expressions for neural networks with more matrix and vector notations. Another important motivation for vectorization is the speed perspective in the implementation. To implement a neural network efficiently, one must be careful when using for loops. The most natural way to implement Equation 2.66 in code is perhaps to use a for loop. In practice, the dimensionalities of the inputs and hidden units are high. As a result, the code will run very slowly if you use it for loops. Leveraging parallelism in GPUs is crucial for the progress of deep learning. This gives rise to vectorization. Instead of using for loops, vectorization uses matrix algebra and highly optimized numerical linear algebra packages (e.g., BLAS) to make neural network computations run quickly. Before the deep learning era, a for loop may have been sufficient on smaller datasets, but modern deep networks and state-of-the-art datasets are infeasible to run with for loops. We vectorize the two-layer fully connected neural network as below. We define a weight matrix $W^{[1]}$ in $\Re^{m\times d}$ as the concatenation of all the vectors $\omega_j^{[1]}$'s in the following way [123]:

$$W^{[1]} = \begin{bmatrix} -\omega_1^{[1]^T}- \\ -\omega_2^{[1]^T}- \\ \vdots \\ -\omega_m^{[1]^T}- \end{bmatrix} \in \Re^{m\times d}, \tag{2.67}$$

now by the definition of matrix-vector multiplication, we can write $z = [z_1, ..., z_m]^\top \in \Re^m$ as:

$$\underbrace{\begin{bmatrix} z_1 \\ \vdots \\ \vdots \\ z_m \end{bmatrix}}_{z\in\Re^{m\times 1}} = \underbrace{\begin{bmatrix} -\omega_1^{[1]^T}- \\ -\omega_2^{[1]^T}- \\ \vdots \\ -\omega_m^{[1]^T}- \end{bmatrix}}_{W^{[1]}\in\Re^{m\times d}} \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}}_{x\in\Re^{d\times 1}} + \underbrace{\begin{bmatrix} b_1^{[1]} \\ b_2^{[1]} \\ \vdots \\ b_m^{[1]} \end{bmatrix}}_{b^{[1]}\in\Re^{m\times 1}}, \tag{2.68}$$

in other words:

$$z = W^{[1]}x + b^{[1]}\,. \tag{2.69}$$

We remark again that a vector in $\Re^d$ in this notes, following the conventions previously established, is automatically viewed as a column vector, and can also be viewed as a $d \times 1$ dimensional matrix. Note that this differs from Numpy where a vector is viewed as a row vector in broadcasting. Computing the activations $a \in \Re^m$ from $z \in \Re^m$ involves an elementwise non-linear application of the ReLU function, which can be computed in parallel efficiently. Overloading ReLU for element-wise application of ReLU (meaning, for a vector $t \in \Re^d$, ReLU(t) is a vector such that ReLU$(t)_i$ =ReLU(t)), we have [124]:

$$a = ReLU(z)\,. \tag{2.70}$$

Define $W^{[2]} = [\omega^{[2]^\top}] \in \Re^{1\times m}$ similarly. Then, the model in Equation 2.66 can be summarized as follows:

$$a = ReLU(W^{[1]}x + b^{[1]}), \tag{2.71}$$

$$\bar{h}_\theta(x) = W^{[2]}a + b^{[2]}, \tag{2.72}$$

here θ consists of $W^{[1]}$, $W^{[2]}$ (often referred to as the weight matrices) and $b^{[1]}$, $b^{[2]}$ (referred to as the biases). The collection of $W^{[1]}$, $b^{[1]}$ is referred to as the first layer, and $W^{[2]}$, $b^{[2]}$ the second layer. The activation a is referred to as the hidden layer. A two-layer neural network is also called a one-hidden-layer neural network.

2.8.4 Multi-layer fully-connected neural networks

With these succinct notations, we can stack more layers to get a deeper fully connected neural network. Let r be the number of layers (weight matrices). Let $W^{[1]}, ..., W^{[r]}, b^{[1]}, ..., b^{[r]}$ be the weight matrices and biases of all the layers. Then a multi-layer neural network can be written as [125]:

$$a^{[1]} = ReLU(W^{[1]}x + b^{[1]}), \tag{2.73}$$

$$a^{[2]} = ReLU(W^{[2]}a^{[1]} + b^{[2]}), \tag{2.74}$$

$$..., \tag{2.75}$$

$$a^{[r-1]} = ReLU(W^{[r-1]}a^{[r-2]} + b^{[r-1]}), \tag{2.76}$$

$$\bar{h}_\theta(x) = W^{[r]}a^{[r-1]} + b^{[r]}\,. \tag{2.77}$$

We note that the weight matrices and biases need compatible dimensions for the equations above to make sense. If $a^{[k]}$ has dimension m_k, then the weight matrix $W^{[k]}$ should be of dimension $m_k \times m_{k-1}$, and the bias $b^{[k]} \in \Re^{m_k}$. Moreover, $W^{[1]} \in \Re^{m_1 \times d}$ and $W^{[r]} \in \Re^{1 \times m_{r-1}}$. The total number of neurons in the network is $m_1 + \cdots + m_r$, and the total number of parameters in this network is $(d+1)m_1 + (m_1+1)m_2 + \cdots + (m_{r-1}+1)m_r$. Sometimes for notational consistency, we also write $a^{[0]} = x$, and $a^{[r]} = h_\theta(x)$. Then we have a simple recursion that:

$$a^{[k]} = ReLu(W^{[k]}a^{[k-1]} + b^{[k]}), \forall k = 1, ..., r-1\,. \tag{2.78}$$

Note that this would have been true for $k = r$ if there were an additional ReLU in Equation 2.77, but often people like to make the last layer linear (aka without a ReLU) so that negative outputs are possible, and it is easier to interpret the last layer as a linear model. The activation function ReLU can be replaced by many other non-linear functions $\sigma(\cdot)$ that maps $\Re$ to $\Re$ such as [126]:

$$\sigma(z) = \frac{1}{1+e^{-z}}, \quad \text{(sigmoid)} \tag{2.79}$$

$$\sigma(z) = \frac{e^{z}-e^{-z}}{e^{z}+e^{-z}}, \quad \text{(tanh)} \tag{2.80}$$

$$\sigma(z) = \max\{z, \gamma z\}, \gamma \in (0,1), \quad \text{(leaky ReLU)} \tag{2.81}$$

$$\sigma(z) = \frac{z}{2}\left[1 + erf\left(\frac{z}{\sqrt{2}}\right)\right], \quad \text{(GELU)} \tag{2.82}$$

$$\sigma(z) = \frac{1}{\beta}\log(1+\exp(\beta z)), \beta > 0, \quad \text{(Softplus)} \tag{2.83}$$

The activation functions are plotted in Figure 2.21. Sigmoid and tanh are less and less used these days partly because they are bounded from both sides, and the gradient vanishes as z goes to both positive and negative infinity, whereas all the other activation functions still have gradients as the input goes to positive infinity. Softplus is not used very often either in practice and can be viewed as a smoothing of the ReLU so that it has a proper second-order derivative. GELU and leaky ReLU are both variants of ReLU, but they have some non-zero gradient even when the input is negative.

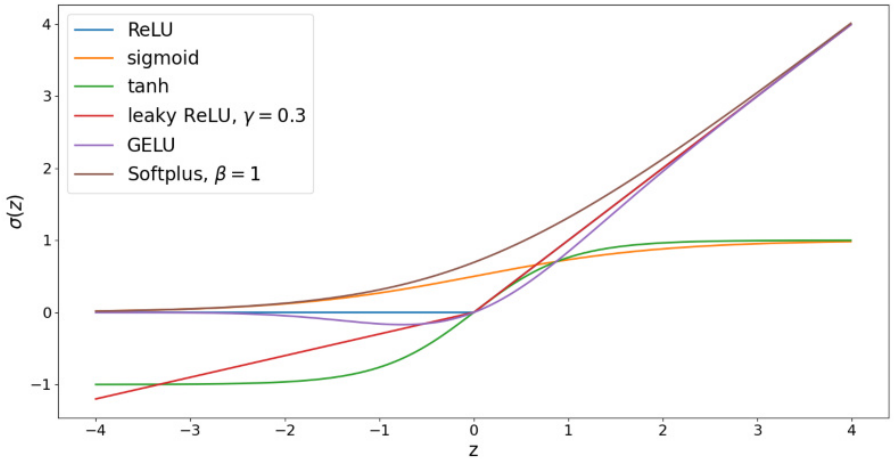


Figure 2.21: Activation functions in deep learning.

In summary, if we graph an artificial neural network (ANN) and stack the neurons, we obtain neural structures that become more complex as the neural layers increase

(stack of neurons), reaching the construction of neural network architectures with millions of neurons. In Figure 2.22, where the network formations are observed, it is worth mentioning that a deep network is considered to have more than three intermediate/hidden layers.

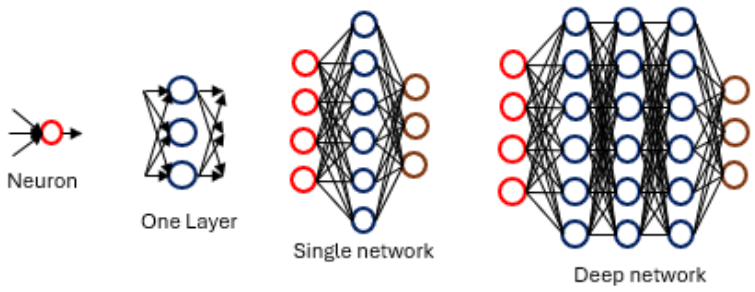


Figure 2.22: Circles represent artificial neurons. Stacking neurons makes up a layer. Red circles are part of the input layer, blue circles are part of the hidden layers, and brown circles are part of the output layer, for simple and deep neural networks.

2.8.5 Convolutional layers

CNN are neural networks that consist of convolution layers (and many other modules) and are particularly useful for computer vision applications [127]. For the simplicity of exposition, we focus on 1-D convolution in this text and only briefly mention 2-D convolution informally at the end of this subsection. (2-D convolution is more suitable for images that have two dimensions. 1-D convolution is also used in natural language processing). We start by introducing a simplified version of the 1-D convolution layer, denoted by Conv1D-S$(\cdot)$ which is a type of matrix multiplication layer with a special structure. The parameters of Conv1D-S are a filter vector $\omega \in \Re^k$ where k is called the filter size (oftentimes $k \ll m$), and a bias scalar b. Oftentimes, the filter is also called a kernel, but it does not have much to do with the kernel in the kernel method. For simplicity, we assume $k = 2l + 1$ is an odd number. We first pad zeros to the input vector z in the sense that we let $z_{i-l} = z_{i-l+1} = \cdots = z_0 = 0$ and $z_{m+1} = z_{m+2} = \cdots = z_{m+l} = 0$ and treat z as an $(m + 2l)$-dimension vector.

Con1D-S outputs a vector of dimension $\Re^m$ where each output dimension is a linear combination of subsets of z_j's with coefficients from ω:

$$\text{Conv1D-S}(z)_i = \omega_1 z_{i-l} + \omega_2 z_{i-l+1} + \cdots + \omega_{2l+1} z_{i+l} = \sum_{j=1}^{2l+1} \omega_j z_{i-l+(j-1)} \,. \tag{2.84}$$

Therefore, one can view Conv1D-S as a matrix multiplication with shared parameters: Conv1D-S$(z) = Qz$, where:

$$Q = \begin{bmatrix} \omega_{l+1} & \cdots & \omega_{2l+1} & 0 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ \omega_l & \cdots & \omega_{2l} & \omega_{2l+1} & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \omega_1 & \cdots & \omega_{l+1} & \cdots & \cdots & \cdots & \omega_{2l+1} & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \omega_1 & \cdots & \cdots & \cdots & \cdots & \omega_{2l} & \omega_{2l+1} & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & \omega_1 & \cdots & \cdots & \cdots & \omega_{2l+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & \omega_1 & \cdots & \omega_{l+1} \end{bmatrix}, \tag{2.85}$$

Note that $Q_{ij} = Q_{i-1,j-1}$ for all $i, j \in \{2, ..., m\}$; thus, convolution is a matrix multiplication with parameter sharing. We also note that computing the convolution only takes $O(km)$ times, but computing a generic matrix multiplication takes $O(m^2)$ time. Convolution has k parameters but generic matrix multiplication has m^2 parameters. Thus, convolution is supposed to be much more efficient than a generic matrix multiplication, as long as the additional structure imposed does not hurt the model's flexibility to fit the data. We also note that in practice there are many variants of the convolutional layers that we define here, e.g., there are other ways to pad zeros, or sometimes the dimension of the output of the convolutional layers could be different from the input. We omit some of these subtleties here for simplicity. The convo-

lutional layers used in practice also have many channels, and the simplified version above corresponds to the 1-channel version. Formally, Conv1D takes in C vectors $z_1, ..., z_C \in \Re^m$ as inputs, where C is referred to as the number of channels. In other words, the more general version, denoted by Conv1D, takes in a matrix as input, which is the concatenation of $z_1, ..., z_C$ and has dimension $m \times C$. It can output C' vectors of dimension m, denoted by Conv1D$(z)_1$, ..., Conv1D$(z)_{C'}$, where C' is referred to as the output channel, or equivalently a matrix of dimension $m \times C'$. Each output is a sum of the simplified convolutions applied on various channels [128].

$$\forall i \in [C'], \text{Conv1D}(z)_i = \sum_{j=1}^{C} \text{Conv1D-S}_{i,j}(z_j)\,. \tag{2.86}$$

Note that each Conv1D-S$_{i,j}$ is a module with different parameters, and thus the total number of parameters is k (the number of parameters in a Conv1D-S) $\times CC'$ (the number of Conv1D-S$_{i,j}$'s) $= kCC'$. In contrast, a generic linear mapping from $\Re^{m \times C}$ and $\Re^{m \times C'}$ has m^2CC' parameters. The parameters can also be represented as a three-dimensional tensor of dimension $k \times C \times C'$.

A 2-D convolution with one channel, denoted by Conv2D-S, is analogous to the Conv1D-S, but takes a 2-dimensional input $z \in \Re^{m \times m}$ and applies a filter of size $k \times k$, and outputs Conv2D-S$(z) \in \Re^{m \times m}$. The full 2-D convolutional layer, denoted by Conv2D, takes in a sequence of matrices $z_1, ..., z_C \in \Re^{m \times m}$ or equivalently a 3-D tensor $z = (z_1, ..., z_C) \in \Re^{m \times m \times C}$ and outputs a sequence of matrices, Conv2D$(z)_1$, ..., Conv2D$(z)_{C'} \in \Re^{m \times m}$, which can also be viewed as a 3D tensor in $\Re^{m \times m \times C}$. Each output channel is the sum of the outcomes of applying Conv2D-S layers on all the input channels.

$$\forall i \in [C'], \text{Conv2D}(z)_i = \sum_{j=1}^{C} \text{Conv2D-S}_{i,j}(z_j)\,. \tag{2.87}$$

Because there is a CC' number of Conv2D-S modules and each of the Conv2D-S modules has k^2 parameters, the total number of parameters is $CC'k^2$. The parameters can also be viewed as a 4D tensor of dimension $C \times C' \times k \times k'$.

So, as we mentioned, a CNN is a type of ANN used primarily for image processing

and is capable of recognizing complex patterns, making it useful in image classification tasks, object detection, and facial recognition, among others. Let us consider an example, according to Figure 2.23. The layers of convolutional blocks are described below:

- *Convolutional layers:* These layers are responsible for extracting important features from the image. They use special filters called "kernels" that slide across the image and perform mathematical operations to highlight patterns such as edges, textures, and shapes.
- *Activation layers:* After convolution, an activation function is applied to introduce nonlinearity into the network. The activation function determines whether a neuron is activated, or not, which helps capture complex relationships in the data.
- *Pooling layers:* These layers reduce the dimensionality of the features obtained by the convolutional layers. The pooling process combines the information into small regions to summarize it and retain only the most relevant information.

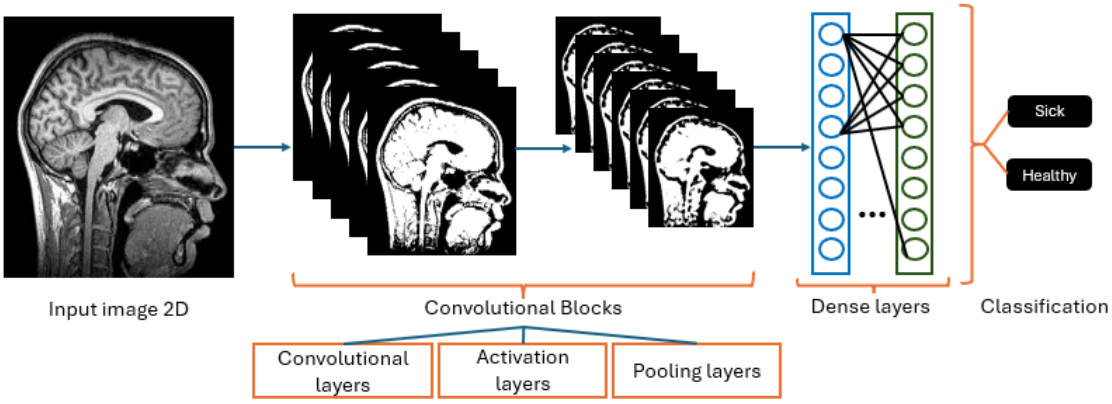


Figure 2.23: Structure of a convolutional neural network.

Chapter 3

Methodology

3.1 Introduction

The methodology employed in this research is centered around the development and evaluation of AI-driven deep-learning models designed to enhance medical image diagnosis. The primary focus is on the integration of convolutional neural networks (CNNs) and transfer learning techniques to classify, segment, and analyze medical images, including magnetic resonance imaging (MRI), computed tomography (CT), and X-rays.

This chapter details the step-by-step process used to construct and validate the models, beginning with image preprocessing to improve image quality, followed by feature extraction using pre-trained networks such as ResNet50 and VGG16. Finally, the classification and segmentation tasks were performed using advanced machine learning algorithms. The experiment, the datasets used, and the evaluation metrics for model performance are outlined here.

The methodology described in this chapter builds on techniques discussed in previously published research articles and presents a cohesive framework for automated medical diagnosis using artificial intelligence.

3.2 Modular model

In this section, we present the modular approach that forms the foundation of our deep learning model, designed for medical image analysis. This modular structure comprises three key components: preprocessing, feature extraction using pre-trained convolutional networks, and a robust classification system. By modularizing the model, each section can be optimized independently, allowing flexibility and adaptability to new datasets and techniques while maintaining competitive performance with the state of the art. Figure 3.1 shows these three sections, where the feature extractor, is the convolutional basis of the pre-trained networks.

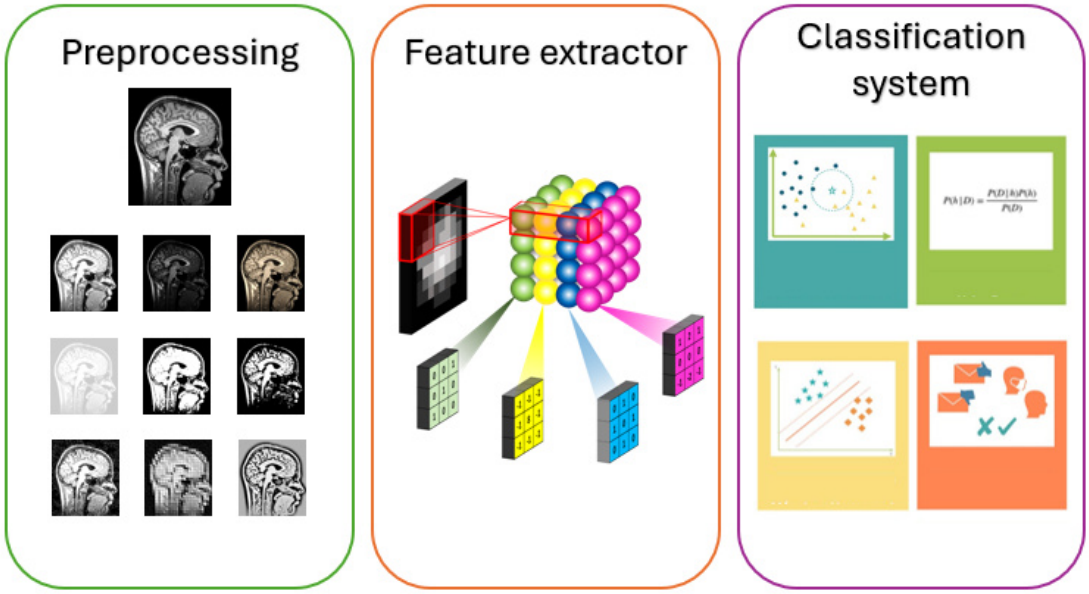


Figure 3.1: Proposed model for segmentation, classification, and detection of Biomedical images.

3.2.1 Preprocessing

Preprocessing plays a vital role in preparing medical images for analysis. Medical images, such as MRI and CT scans, often contain noise and artifacts that can impair the performance of deep learning models. The preprocessing stage is designed to enhance the image quality and remove irrelevant features. Common techniques include:

1. **Normalization:** Rescaling pixel intensity values to a standard range.

2. **Denoising:** Using filters such as Gaussian or median filters to reduce image noise.

3. **Data Augmentation:** Applying transformations like rotation, flipping, and zooming increases the dataset's variability and prevents overfitting.

4. **Resizing:** Ensuring images are resized to a uniform dimension suitable for input to the convolutional network

We have worked with various sets of images, that is, datasets. Each is described in Figure 3.2, with images of each set attached. It is worth mentioning that we have worked with more datasets, and they are mainly divided into two groups: scientific repositories such as Kaggle or our [129–132].

Database	Labelled	Repository
Parkinson's Drawings	Healthy: Spiral 51 images, Wave: 51 images Parkinson: Spiral 51 images, Wave 51 images	Kaggle (Zham, Kumar, Dabnichki, Poosapadi Arjunan, & Raghav, 2017)
Brain MRI Images	Tumor: 155 images No tumor: 98 images	Kaggle (Chakrabarty, 2019)
Chest X-ray: Montgomery County	Tuberculosis: 58 images Normal: 80 images	Kaggle (Jaeger et al., 2013)
Chest X-Rays in Patients with Pneumonia	Pneumonia: 1,190 images Normal: 1,034 images	Kaggle (Kermany et al., 2018)
Long COVID-19, Mexican Institute of Social Security T1 (IMSS T1, for its acronym in Spanish)	Post-COVID conditions: 57 CT images No Post-COVID conditions: 57 CT images	IMSS T1 (Provided by the Institute)

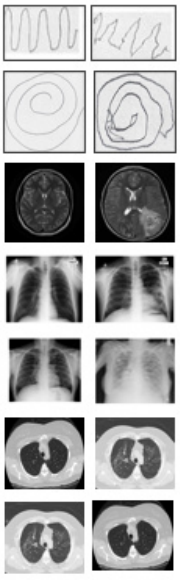


Figure 3.2: Description of the databases of medical pathologies.

Image filters. An image filter is a procedure applied to an image to highlight or improve some of its features. To create a filter, we need a kernel convolution matrix, which will be square and whose sizes are normally: 3X3, 5X5, and 7X7.

There are various kernels (filters) in the literature, including sharpening, blurring, edge enhancement, Gaussian, Sobel, Sharpen, and Canny. These matrices are also called masks (see Figure 3.3). Given a matrix $A_{m\times n}$ and a matrix $C_{(2N+1)\times(2N+1)}$ with $2N+1 < m, n$ the convolution of matrices A and C is defined as a new matrix $D = A * C$ defined from the expression [133]:

$$D[m,n] = \sum_i \sum_j A[m,n] * C[i-n, j-n]\,. \tag{3.1}$$

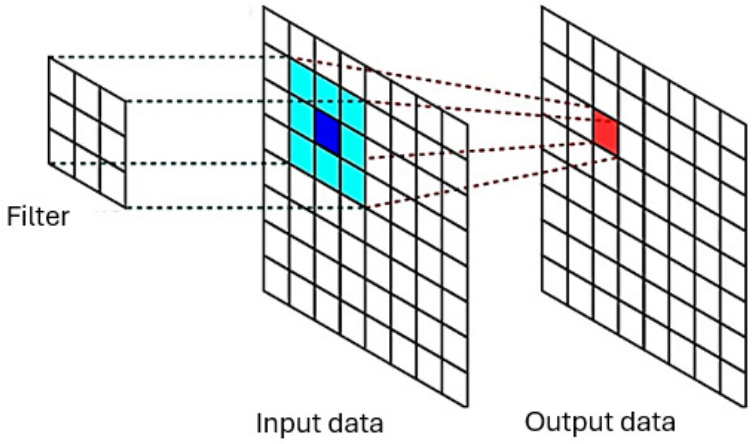


Figure 3.3: The multiplication between a filter and an image is known as convolution, and the output image is known as the filtered image.

Various preprocessing techniques have been performed in each of the data sets in Figure 3.2, this depends on the type of pathology to be studied. However, we can group the preprocessing into three main stages. The first stage is to resize, standardize, and normalize. The second is the set of images is grouped by labels and classifications previously diagnosed by a medical expert, such as a radiologist, which is why our model is based on supervised learning. The last and third stage, highlighting the main characteristics of the image with filters, this will depend greatly on the pathology to be diagnosed. For example, in Figure 3.4, the three stages are shown in block form on chest X-rays (pneumonia and tuberculosis) and computed axial tomography (pulmonary fibrosis derived from SARS-CoV-2). As we can see, preprocessing in medical imaging is crucial for enhancing image quality, reducing noise, and improv-

ing the overall performance of feature extraction and classification tasks. We have described the image preprocessing steps used, such as resizing, normalization, contrast enhancement, and noise reduction (depending on the type of image used—CT, MRI, or X-ray).

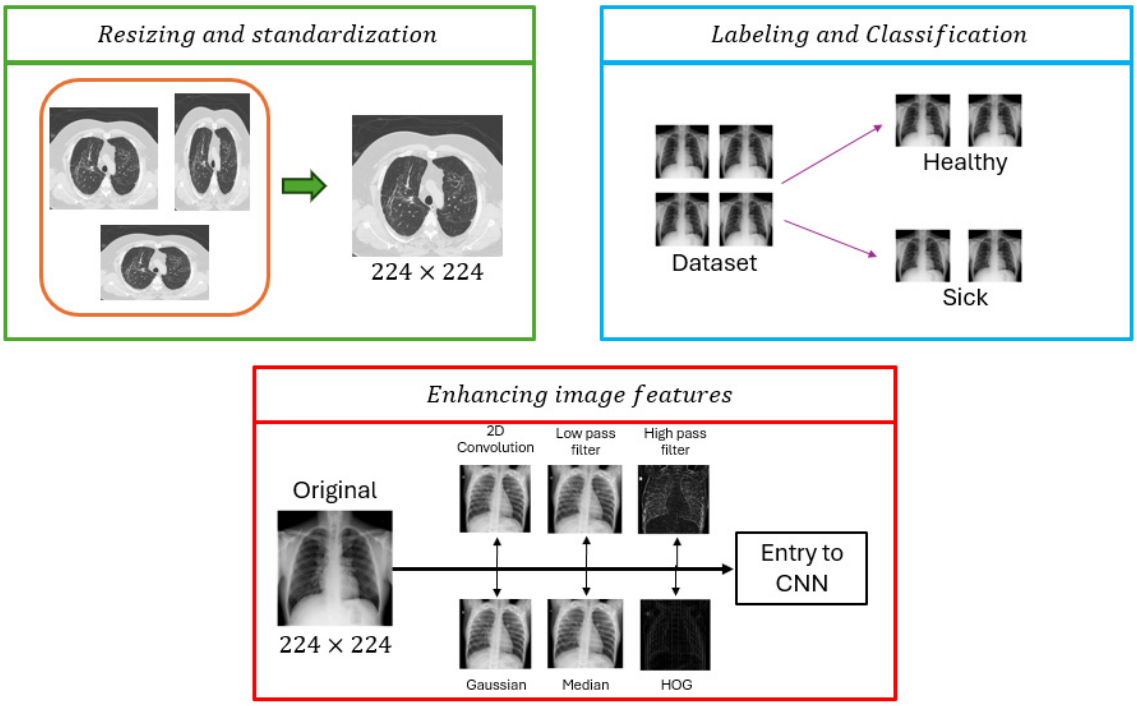


Figure 3.4: Processes performed on the set of images of different pathologies to classify, detect, or segment a medical condition to obtain an optimal model for medical diagnosis.

In summary, the preprocessing stage is essential in preparing medical images for analysis by deep learning models. These preprocessing methods improve the model's overall performance by reducing noise and irrelevant variations and contribute to a more robust and generalized system. By addressing challenges specific to medical images, such as image acquisition and noise variability, preprocessing helps ensure that subsequent stages in the modular model, particularly feature extraction and classification, operate on optimal input data, ultimately leading to more accurate and reliable diagnostic outcomes. The next step of the model is to extract the features from the set of already processed images. For this we move on to the feature extraction block.

3.2.2 Feature extraction

Feature extraction is a crucial stage in the modular deep learning pipeline, bridging raw medical images and their corresponding diagnostic predictions. In this step, we identify and extract the most significant and relevant patterns from the medical images contributing to accurate classification or segmentation. Feature extraction ensures that the model focuses on the vital information necessary for diagnosis while discarding redundant or irrelevant data.

3.2.3 Pre-trained CNN versus creating new one

Using the convolutional block of a pre-trained network as a feature extractor offers significant advantages, particularly when technological resources are limited. Pre-trained networks, such as ResNet, VGG, and others, have been trained on massive datasets like ImageNet, learning to recognize a wide variety of features from a vast number of images. These networks capture complex patterns and textures, making them ideal for feature extraction in medical imaging, where subtle variations in textures and shapes are crucial for accurate diagnosis.

Creating a neural network from scratch requires substantial computational power, large datasets, and extensive training time. Without the necessary technological tools, this process becomes unfeasible. Pre-trained models avoid this limitation by providing a robust and well-validated architecture. They significantly reduce the amount of training data required. Since the early layers of these models capture universal features such as edges and textures, they can easily be adapted to specific tasks in medical imaging.

Thus, using the convolutional block of a pre-trained network as a feature extractor allows for high-quality results while optimizing computational resources, making it a more practical and efficient solution for researchers with limited infrastructure.

3.2.4 Convolutional neural networks and pre-trained models

CNNs are the most commonly used architectures for medical imaging due to their proven ability to capture spatial hierarchies and complex features in image data. CNNs apply convolutional filters that learn to detect essential features such as edges, textures, and shapes through multiple layers of abstraction. However, training CNNs from scratch can be computationally expensive and time-consuming, especially with limited labeled data, as is often the case in medical imaging.

To overcome these challenges, transfer learning has become an essential technique. Pre-trained CNN models like VGG16 (see Figure 3.5), ResNet50 (see Figure 3.6), InceptionV3, Mobile-Net, and others have been widely adopted for feature extraction in medical image analysis [134]. These networks, pre-trained on large datasets such as ImageNet, capture generalized patterns in visual data. By leveraging transfer learning, we repurpose these pre-trained models to extract meaningful features from medical images, with minimal adjustment to the specific task. This not only reduces the training time but also improves the model's generalization, as the pre-trained layers already encapsulate valuable, low-level feature representations.

In the modular model, the feature extraction process occurs in the convolutional block, where the pre-trained network's layers are used to extract features from the input medical images. This is particularly advantageous when dealing with complex medical data like MRI, CT, or X-ray images, where patterns may be subtle or require a high degree of expertise to discern.

The advantages of transfer learning for feature extraction are described below:

1. **Improved accuracy:** Using pre-trained models makes the feature extraction process highly accurate, as the layers have already been optimized to detect intricate patterns.

2. **Reduced computational costs:** Transfer learning significantly reduces training time and computational resources.

3. **Handling small datasets:** Medical datasets are often smaller than those in

other fields, and transfer learning allows the model to achieve strong performance even with limited training data.

4. **Customization:** While the pre-trained networks serve as a foundation, the final layers can be fine-tuned or re-trained to tailor feature extraction to the specific medical imaging task, whether detecting tumors, classifying lung diseases, or segmenting organs.

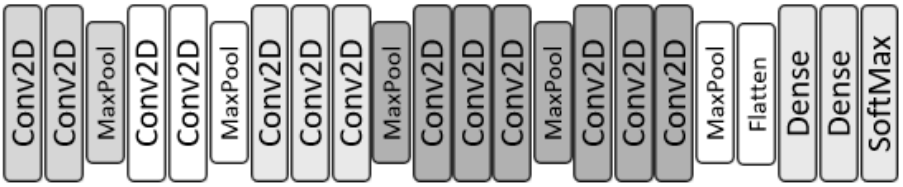


Figure 3.5: The architecture of the VGG16 network is composed of 16 layers, 13 convolutional (Conv2D) and three classification layers. The network's convolutional block is called the feature extractor and is used in our modular model. In the case of this network, the base starts from the first layer (Conv2D) and ends after the last convolutional layer (Conv2D), being more precise in flatten. The input image size of the neural network is 224 by 224

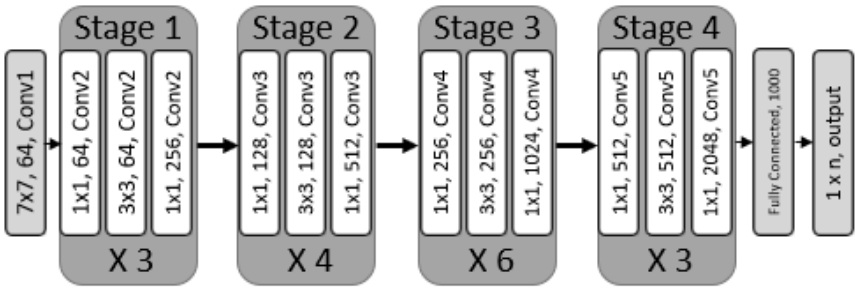


Figure 3.6: ResNet-50 is a convolutional neural network used for image classification. It is a network with 50 layers, deeper and more complex. It is a ResNet (Residual Networks) family model, designed to solve the problems that arise when training deep neural networks. The input image size of the neural network is 224 by 224.

In practical terms, the extracted features often include textures, shapes, edges, and specific anatomical details unique to the medical condition being analyzed. For instance, in pulmonary fibrosis classification, the feature extraction stage might identify patterns associated with fibrosis, such as specific textures in the lung tissues.

For brain tumor detection, the model may focus on identifying abnormal growths or deviations from typical brain structures.

In summary, feature extraction is pivotal in transforming raw medical images into meaningful representations that can be effectively used for diagnosis. We can extract highly relevant features in a computationally efficient by utilizing pre-trained CNN models and applying transfer learning. These features then serve as the input for the classification system, forming the basis for accurate predictions in medical image diagnostics. The feature extraction process's flexibility and modularity allow for performance improvements while maintaining compatibility with the rest of the model, making it a key component of modern deep-learning pipelines for medical imaging.

3.2.5 Classification system

The classification phase is crucial in evaluating how well the model can generalize to unseen data. We have implemented two main classification strategies to ensure robust evaluation: training and testing set, and cross-validation.

3.2.6 Training and testing set

For the first system, the dataset is divided into two parts: a training set and a testing set. The training set typically consists of 80% of the data, while the remaining 20% is used for testing (sometimes 70% and 30%, respectively, according to the size of the dataset) [135]. The division ensures that the model is trained on a majority of the data while being tested on an independent set to evaluate its performance. This system effectively measures the model's generalization capability to unseen data. Once the model is trained enough with the relevant training data, it is tested with the test data. We can understand the whole process of training and testing in three steps, which are as follows:

a) **Feed:** Firstly, we need to train the model by feeding it with training input data.

b) **Define:** Now, training data is tagged with the corresponding outputs (in Supervised Learning), and the model transforms the training data into text vectors or a number of data features.

c) **Test:** In the last step, we test the model by feeding it with the test data/unseen dataset. This step ensures that the model is trained efficiently and can be generalized well.

The above process is explained using a flowchart given below in Figure 3.7:

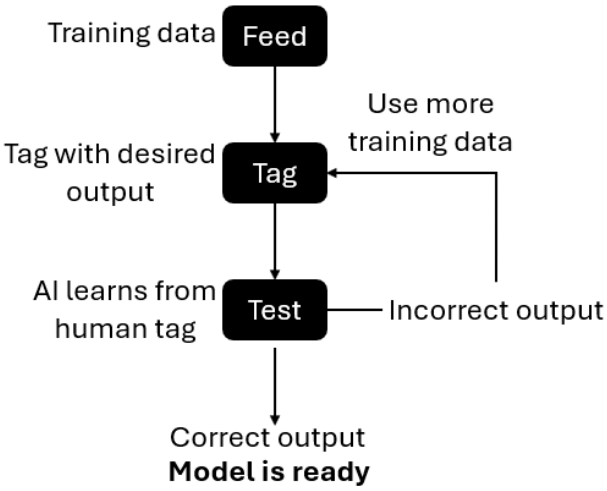


Figure 3.7: Training and Testing Classification System Process.

3.2.7 Cross-Validation

In addition to the training and testing split, we utilized cross-validation, which subdivides the dataset into smaller partitions, or folds. The most commonly used is 10-fold cross-validation, where the data is split into 10 equal parts [136]. The model is trained on nine folds and tested on the remaining one, rotating through all folds. In some cases, we experimented with 20 or 30 splits to refine the evaluation further (see Figure 3.8). Cross-validation provides a more thorough assessment of the model's robustness by mitigating the risk of overfitting or relying too much on specific data subsets.

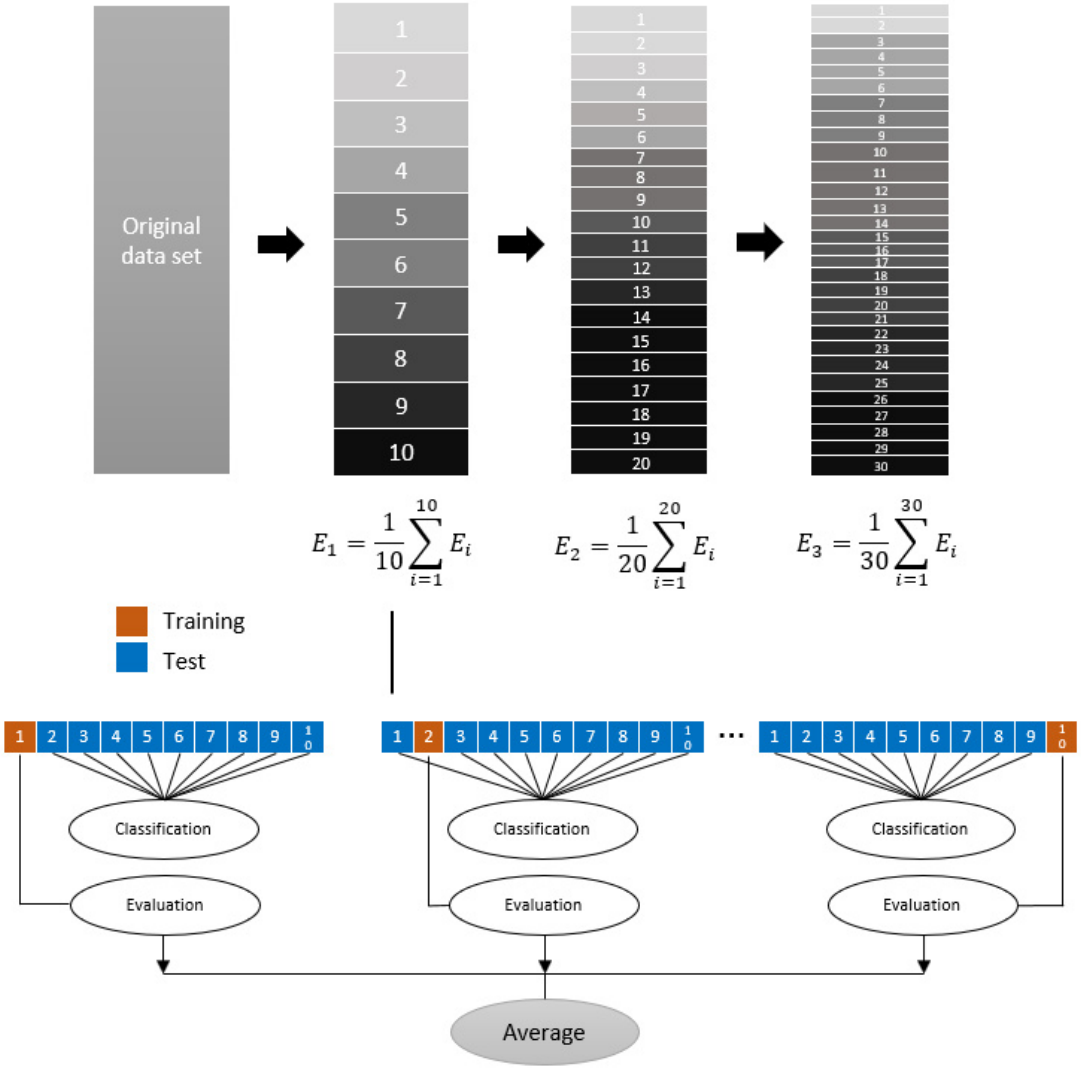


Figure 3.8: Process of the cross-validation classification scenario for the case.

Where the performance E of the model is evaluated as [136]:

$$E = \frac{1}{K}\sum_{i=1}^{K} E_i, \tag{3.2}$$

hence, K is the number of splits. We have used K-folds with $K = 10, 20$, and 30.

3.2.8 Classifier algorithms

After preparing the dataset, we employed the following advanced machine-learning algorithms to classify the medical images:

- *Support vector machine (SVM)* [137]: SVM is used for its efficiency in handling high-dimensional spaces and performing well in both linear and non-linear classification tasks. A data point is viewed as a p-dimensional vector, and if we want to know whether we can separate such points with a $(p-1)$-dimensional hyperplane. We are given a training dataset of n points of the form: $(x_1, y_1), ..., (x_n, y_n)$. Where the y_i are either 1 or -1, each indicating the class to which the point x_i belongs. Each x_i is a p-dimensional real vector. We want to find the "maximum-margin hyperplane" that divides the group of points x_i for which $y_i = 1$ from the group of points for which $y_i = -1$, which is defined so that the distance between the hyperplane and the nearest point x_i from either group is maximized (see Figure 3.9). Any hyperplane can be written as the set of points x satisfying $w^\top x - b = 0$, where w is the normal vector to the hyperplane. This is much like Hesse's normal form, except that w is not necessarily a unit vector. The parameter $b/\|w\|$ determines the offset of the hyperplane from the origin along the normal vector w.

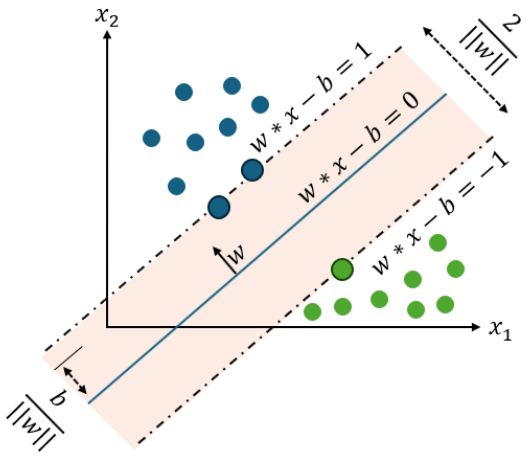


Figure 3.9: Maximum-margin hyperplane and margins for an SVM trained with samples from two classes. Samples on the margin are called the support vectors.

- *Logistic regression (LR)* [138]: This model estimates the probability that a given instance belongs to a particular class, making it ideal for binary and multi-class

classification problems. The logistic function is:

$$p(x) = \frac{1}{1 + e^{-(x-\mu)/s}}, \tag{3.3}$$

where μ is a location parameter (the midpoint of the curve, where $p(\mu) = 1/2$) and s is a scale parameter. This expression may be rewritten as:

$$p(x) = \frac{1}{1 + e^{-(\beta_0+\beta_1 x)}}, \tag{3.4}$$

where $\beta_0 = -\mu/s$ and is known as the intercept (it is the vertical intercept or y-intercept of the line $y = \beta_0 + \beta_1 x$), and $\beta_1 = 1/s$ (inverse scale parameter or rate parameter): these are the y-intercept and slope of the log-odds as a function of x. Conversely, $\mu = -\beta_0/\beta_1$ and $s = 1/\beta_1$.

- *K-nearest neighbors (KNN)* [139]: KNN is a simple, yet effective, algorithm based on instance-based learning. It assigns a class based on the majority vote of its nearest neighbors. Suppose we have pairs $(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n)$ taking values in $\Re^d \times \{1, 2\}$, where Y is the class label of X, so that $X|Y = r \sim P_r$ for $r = 1, 2$ (and probability distributions P_r). Given some norm $\|\cdot\|$ on $\Re^d$ and a point $x \in \Re^d$, let $(X_{(1)}, Y_{(1)}), ..., (X_{(n)}, Y_{(n)})$ be a reordering of the training data such that $\|X_{(1)} - x\| \leq \cdots \leq \|X_{(n)} - x\|$. The training examples are vectors in a multidimensional feature space, each with a class label. The algorithm's training phase consists only of storing the training samples' feature vectors and class labels. In the classification phase, k is a user-defined constant, and an unlabeled vector (a query or test point) is classified by assigning the most frequent label among the k training samples nearest to that query point. A commonly used distance metric for continuous variables is Euclidean distance. For discrete variables, such as text classification, another metric, such as the overlap metric (or Hamming distance), can be used. For example, k-NN has been employed with correlation coefficients, such as Pearson and Spearman, as a metric in the context of gene expression microarray data. In Figure 3.10 an

example is presented.

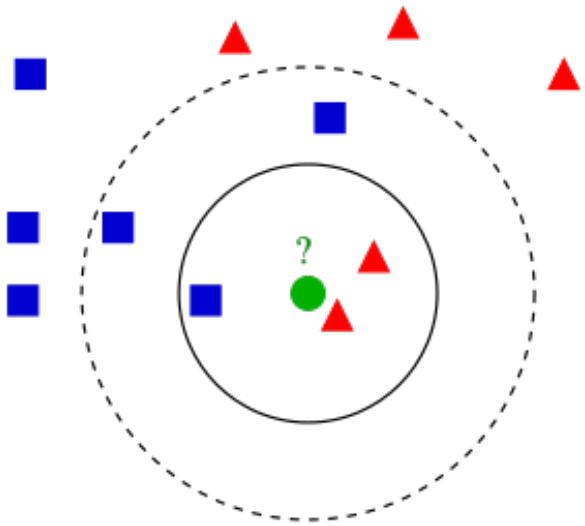


Figure 3.10: Example of k-NN classification. The test sample (green dot) should be classified as blue squares or red triangles. If $k = 3$ (solid line circle) it is assigned to the red triangles because there are two triangles and only 1 square inside the inner circle. If $k = 5$ (dashed line circle) it is assigned to the blue squares (three squares vs. two triangles inside the outer circle).

- *Naïve-Bayes probability (NB)* [140]: This algorithm is chosen for its simplicity and speed, particularly when dealing with large datasets. Despite its simplicity, it often yields surprisingly good performance in classification tasks. Abstractly, NB is a conditional probability model: it assigns probabilities $p(C_k|x_1, ..., x_n)$ for each of the K possible outcomes or classes C_k given a problem instance to be classified, represented by a vector $\vec{x} = (x_1, .., x_n)$ encoding some n features (independent variables). The problem with the above formulation is that if the number of features n is large or if a feature can take on a large number of values, then basing, such a model on probability tables is infeasible. The model must, therefore, be reformulated to make it more tractable. Using Bayes' theorem, the conditional probability can be decomposed as:

$$p(C_k|\vec{x}) = \frac{p(C_k)p(\vec{x}|C_k)}{p(\vec{x})}, \tag{3.5}$$

in plain English, using Bayesian probability terminology, the above equation can be written as:

$$posterior = \frac{prior \times likelihood}{evidence}\,. \tag{3.6}$$

In practice, there is interest only in the numerator of that fraction, because the denominator does not depend on C, and the values of the features x_i are given so that the denominator is effectively constant. The numerator is equivalent to the joint probability model: $p(C_k, x_1, ..., x_n)$ which can be rewritten as follows, using the chain rule for repeated applications of the definition of conditional probability:

$$p(C_k, x_1, ..., x_n) = p(x_1, ..., x_n, C_k), \tag{3.7}$$

$$p(C_k, x_1, ..., x_n) = p(x_1|x_2, ..., x_n, C_k)p(x_2, ..., x_n, C_k), \tag{3.8}$$

$$p(C_k, x_1, ..., x_n) = p(x_1|x_2, ..., x_n, C_k)p(x_2|x_3, ..., x_n, C_k)p(x_3, ..., x_n, C_k), \tag{3.9}$$

$$p(C_k, x_1, ..., x_n) = p(x_1|x_2, ..., x_n, C_k)p(x_2|x_3, ..., x_n, C_k)\cdots p(x_{n-1}|x_n, C_k)p(x_n|C_k)p(C_k), \tag{3.10}$$

now the Naïve conditional independence assumptions come into play: assume that all features in $\vec{x}$ are mutually independent, conditional on the category C_k. Under this assumption: $p(x_i|x_{i+1}, ..., x_n, C_k) = p(x_i|C_k)$. Thus, the joint model can be expressed as:

$$p(C_k|x_1, ..., x_n) \propto p(C_k, x_1, ..., x_n), \tag{3.11}$$

$$p(C_k|x_1, ..., x_n) = p(C_k)p(x_1|C_k)p(x_2|C_k)p(x_3|C_k)\cdots\,, \tag{3.12}$$

$$p(C_k|x_1, ..., x_n) = p(C_k)\prod_{i=1}^{n} p(x_i|C_k), \tag{3.13}$$

where $\propto$ denotes proportionality since the denominator $p(\vec{x})$ is omitted. This means that under the above independence assumptions, the conditional distribution over the class variable C is:

$$p(C_k|x_1, ..., x_n) = \frac{1}{Z}p(C_k)\prod_{i=1}^{n} p(x_i|C_k), \tag{3.14}$$

where the evidence $Z = p(\vec{x}) = \sum_k p(C_k)p(\vec{x}|C_k)$, is a scaling factor dependent only on $x_1, ..., x_n$, that is, a constant if the values of the feature variables are known.

3.2.9 Evaluation metrics

We used evaluation metrics derived from the confusion matrix to quantify the model's performance. A confusion matrix is a crucial tool used in classification problems to visualize the performance of a machine-learning model [141]. It provides a comprehensive breakdown of how the model's predictions match the actual labels in the dataset, offering insight into the correct classifications and the errors made by the model. The matrix is typically structured as a square table where each row corresponds to the actual class, and each column corresponds to the predicted class. For a binary classification problem, the confusion matrix has four key components:

Table 3.1: Confusion Matrix.

	Predicted Positive	**Predicted Negative**
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

i) **True Positive (TP):** These are instances where the model correctly predicted the positive class.

ii **True Negative (TN):** These are instances where the model correctly predicted the negative class.

iii **False Positive (FP):** These are instances where the model incorrectly predicted the positive class when it should have predicted the negative. This is often referred to as a "Type I error."

iv) **False Negative (FN):** These are instances where the model incorrectly predicted the negative class when it should have predicted positive. This is often referred to as a "Type II error.

The confusion matrix serves as the foundation for calculating several evaluation metrics, which are essential for understanding the strengths and weaknesses of the model. The confusion matrix allows us to compute several key evaluation metrics, each providing a different perspective on model performance. The most commonly used metrics are [142]:

1. *Accuracy:* The proportion of correctly classified instances (positives and negatives) out of the total number of instances. It gives a general sense of how often the model is right, but it can be misleading in cases where the classes are imbalanced. Accuracy gives a quick overview but should be interpreted cautiously, especially for imbalanced datasets.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{3.15}$$

   while accuracy is a widely reported metric, it may not be the best measure for imbalanced datasets, where one class significantly outnumbers the other. In such cases, accuracy might hide the minority class's poor performance.

2. *Precision:* Also termed Positive Predictive Value; it focuses on the quality of positive predictions. It is the ratio of true positive predictions to the total number of instances predicted as positive, helping evaluate how many predicted positives were correct. Precision is critical when the cost of false positives is high, such as in medical diagnostics where unnecessary treatment should be avoided.

$$\text{Precision} = \frac{TP}{TP + FP}, \tag{3.16}$$

   high precision means the model is usually correct when predicting a positive outcome. This metric is particularly useful in scenarios where the cost of false positives is high (e.g., diagnosing a disease when it is not present).

3. *Recall:* Also known as Sensitivity or True Positive Rate. It measures the model's ability to identify all the actual positive cases correctly. It is the ratio of true positives to the total number of actual positive instances. Recall is important

when it is crucial to identify all positive cases, as in early disease detection.

$$\text{Recall} = \frac{TP}{TP + FN}, \tag{3.17}$$

high recall indicates that the model is good at finding all the relevant positive cases. This is especially important in applications like medical diagnosis, where missing a positive case (i.e., a false negative) could have serious consequences.

4. *F1-score:* The F1-Score is the harmonic mean of precision and recall, offering a single metric that balances the two. This is particularly useful when there is an imbalance between precision and recall, as it provides a more comprehensive measure of a model's performance. F1-Score balances precision and recall, which is key when both types of errors (false positives and false negatives) are costly.

$$\text{F1-Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall}, \tag{3.18}$$

F1-Score is a good metric to use when both false positives and false negatives must be minimized. It helps evaluate models that perform well across precision and recall rather than excelling at one at the expense of the other.

5. *Specificity:* Also known as True Negative Rate. It measures the model's ability to identify negative instances correctly. It is the ratio of true negatives to the total number of actual negatives. Specificity complements recall by showing how well the model handles the negative class, which is useful in cases where avoiding false positives is important.

$$\text{Specificity} = \frac{TN}{TN + FP}\,. \tag{3.19}$$

Specificity is essential in cases where the negative class is important, such as ensuring healthy individuals are correctly identified as not having a disease.

These metrics ensure that the model is evaluated from various perspectives, allowing for a thorough understanding of its effectiveness in real-world applications.

By balancing precision, recall, and other metrics, the model's predictions can be fine-tuned to meet the specific needs of a given application, particularly in sensitive fields like medical image analysis. Through these methods, we have achieved classification, detection, and segmentation models with accuracies of 97%, 98%, and 99%, which are highly competitive with the state of the art. These results have been documented and published in both national and international journals with recognized impact factors, DOI, and ISSN.

3.2.10 IMSS T1 ethics committee

The Local Health Research Committee 1001, SPECIALTY HOSPITAL NUM.1, Bajío, León, GTO., approved the data set of CT scans of Mexican patients, with the research protocol number: Institutional Registration Number R-2023-1001-022, with the work of post-covid conditions

Chapter 4

Results

4.1 Introduction

In this chapter, the results of the developed modular model for medical image classification, detection, and segmentation are presented. The models were trained and evaluated using multiple biomedical datasets, including those related to lung disease, brain tumors, and post-COVID conditions. The results focus on key performance metrics such as accuracy, precision, recall, F1-score, and segmentation metrics like the dice coefficient, providing a detailed comparison between various classifiers, including Support Vector Machines (SVM), logistic regression (LR), k-nearest neighbors (KNN), and Naïve-Bayes (NB).

The results are divided into sections based on the different tasks the model was designed to handle, such as classification, detection, and segmentation. For each task, tables and figures are presented to illustrate the model's performance and how it compares to existing state-of-the-art techniques. In addition to numerical results, confusion matrices, and segmented image outputs are displayed to represent the model's effectiveness visually.

The datasets used for evaluation are diverse, spanning different medical conditions and imaging modalities (CT, MRI, and X-ray), to ensure the model's generalizability and robustness. The results not only demonstrate the model's high accuracy but also highlight its strengths and limitations across different medical image applications.

The results will be presented in greater detail in the following sections, starting with classification performance, followed by detection and segmentation outcomes, and concluding with a comparative analysis against current methods in the literature.

4.2 Results by imaging modality and pathology

4.2.1 X-ray imaging results

4.2.2 Pneumonia

Pneumonia is an inflammatory lung condition primarily affecting the alveoli (small air sacs). Physical examination with imaging is vital for the detection of pneumonia. Chest X-rays are frequently used in validation procedures and represent a rapid alternative to determine the characteristics and extent of lung inflammation. Areas of opacity on the X-ray image are commonly correlated with regions affected by pneumonia (see Figure 4.1).

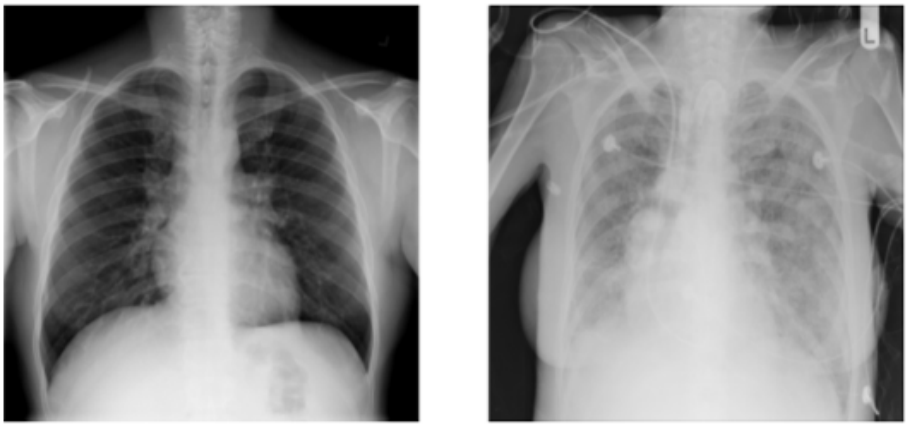


Figure 4.1: Normal (left) and abnormal (right) chest radiographs. Increased opacity suggests pneumonia.

The dataset chosen for this machine learning exercise is a modified version of Paul Mooney's Kaggle submission, "Chest X-ray Images (Pneumonia). According to the author, the images are manually labeled by experts in the field, and the dataset states, "...Chest X-ray images (anterior-posterior) were selected from retrospective

cohorts of pediatric patients aged one to five years at Women and Children's Medical Center... All chest x-ray images performed were part of the patient's routine clinical care...". The original dataset contains 5,863 observations, split into training (1,341 normal cases, 3,875 pneumonia cases), validation (8 normal cases, 8 pneumonia cases), and testing (234 normal cases, 390 pneumonia cases) folders. The modified dataset contains 2,224 observations, where a more balanced split between training and test images is proposed: Training observations: 1,600 (800 normal cases, 800 lung opacity cases). Test observations: 624 (234 normal cases, 390 lung opacity cases).

The methodologies discussed in Chapter 3 were used. The most relevant results are presented below in graphs showing the evaluation metrics and classification algorithm.

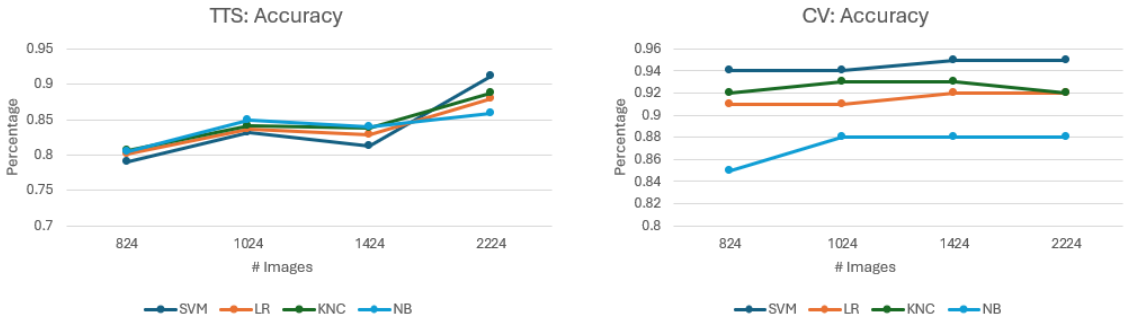


Figure 4.2: Values of the accuracy metric evaluation are TTS, which is the training and testing set, and CV, which is cross-validation. Likewise, KNC is KNN. The range of the ordinates is in decimals, representing the percentage scale; 0.5 represents 50%.

The results show that the cross-validation model has a high-reliability index when classifying chest X-rays to predict pneumonia. However, as the number of instances in the training set increases, the training and test set acquires competitive values with those reported in the literature, higher than 90% (specifically 95%). The CNN ResNet50 network has been used as a feature extractor for this dataset. This modular model is presented as an alternative to classifying X-rays of this type, unlike traditional CNN approaches where networks are created and trained that report values in the literature. However, our optimal values compete with the state of the art, reaching 95% accuracy in our model, so our model where a CNN is used as a feature extractor, represents an opportunity to experiment with different pathologies in X-rays. The

extraction of the features and the processing of images, without any refinement to the instances, are different from any other previously trained model; thus, this reported work represents an alternative to classify chest X-ray images, previously manually labeled.

As can be seen, increasing the number of images and iterations for the classification system results in better performance, as reported in the results. Now, applying filters (Figure 4.3 and 4.4) to the data set that highlights the characteristics of pneumonia in the X-ray can also result in better results.

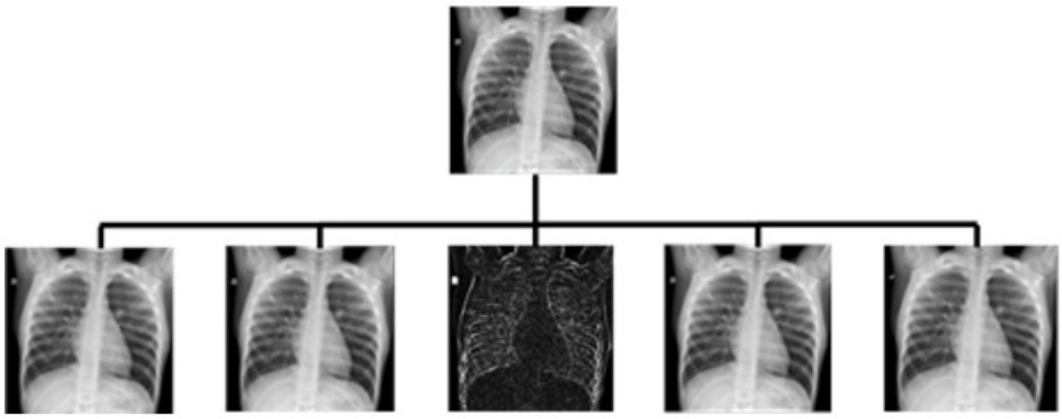


Figure 4.3: Results of filtering the original image. (From left to right) 2D convolution, low-pass filter, high-pass filter, Gaussian and median.

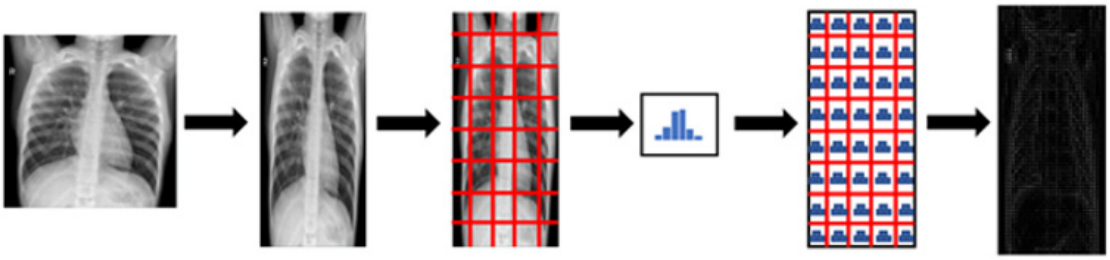


Figure 4.4: Oriented gradient histogram process: taking the original image resized to 128x64 pixels to perform the corresponding steps as input and the oriented gradients as output.

By applying a filter to the original images, we obtain better results in the evaluation metrics. In fact, as the number of images increases, the best values are found

with the low-pass filter with 96% accuracy in the cross-validation model in the full experiment. What improves the model is this preprocessing block. Figure 4.5 shows the complete model with the three blocks of the modular model for detecting pneumonia in chest X-rays.

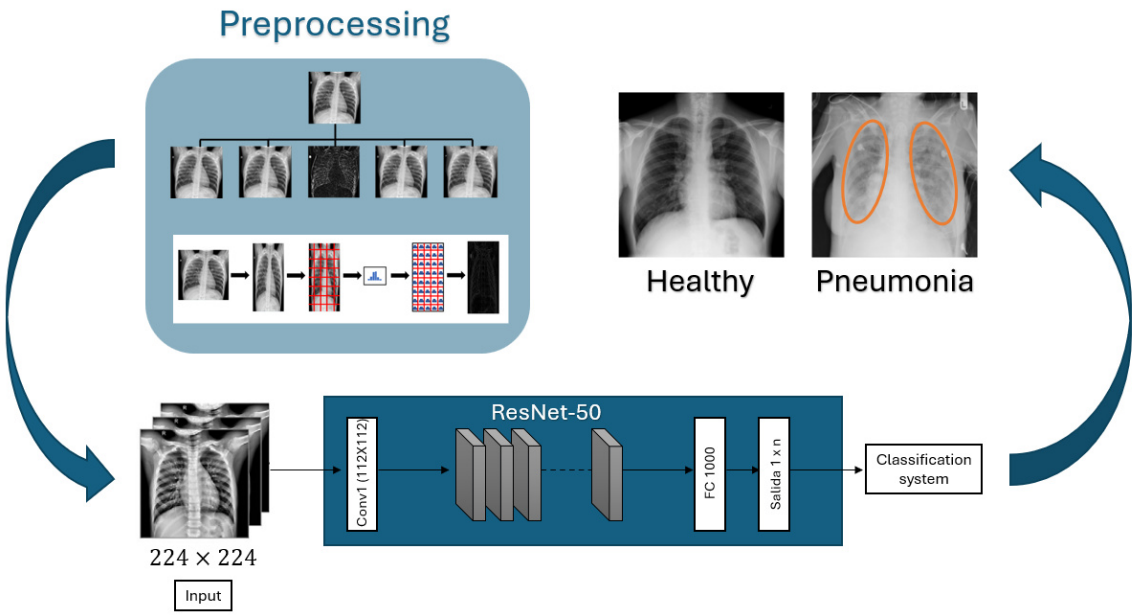


Figure 4.5: Proposed model to detect pneumonia using the convolutional base of the ResNet50 network.

The results show a high-reliability index in the cross-validation model when classifying chest X-rays to predict pneumonia. However, as the instances in the training set increase, the training and test set acquire competitive values, similar to those reported in the literature. Feature extraction and image processing, without any refinement to the instances, are different from any other model, so this work has been reported as an alternative to classifying chest X-ray images under supervised learning. We can also conclude that the SVM classifier has reported the best value in the model, considering the same for future projects with X-rays.

4.2.3 Tuberculosis

According to the World Health Organization (WHO), tuberculosis (TB) is one of the ten leading causes of death in the world. This disease is caused by Mycobacterium

tuberculosis, which usually affects the lungs, but can also affect other body parts. Without treatment, TB disease has a high mortality rate (about 50%). According to the WHO, Mexico in 2019 had between 23,000 and 37,000 new cases of TB, with a rate of 23 cases per 100,000 people, mostly young adults. On the other hand, official data from the Tuberculosis National Program (PNT) of the Mexico Secretary of Health registered 22,285 new cases, of which 80% were from Pulmonary Tuberculosis (PTB). Chest X-ray is useful to detect early lesions in PTB with the main disadvantage of its low specificity; therefore, it is used to correlate chest radiograph findings with bacteriological confirmation.

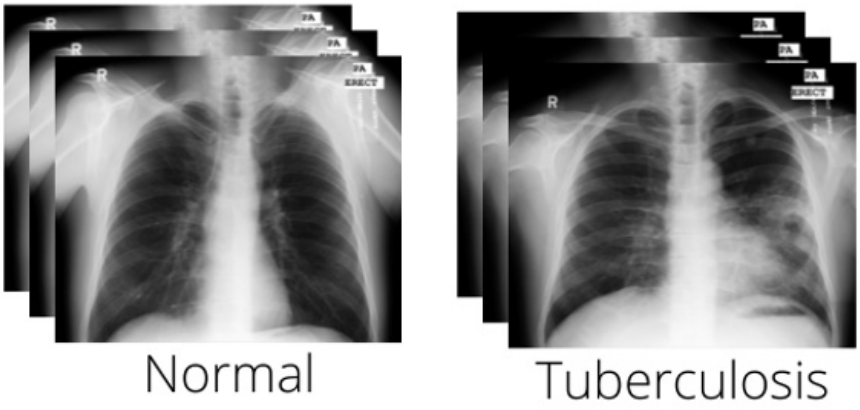


Figure 4.6: Montgomery County's Tuberculosis screening program in the Kaggle repository.

Recently, the use of deep learning as an aid in medical diagnosis has increased with the implementation of convolutional neural networks; biological neurons inspire these neural networks and consist of multiple layers that extract and learn relevant features without human supervision. For this project, a small dataset from Montgomery County's Tuberculosis screening program was used (see Figure 4.6). This set contained 138 frontal chest X-rays that were manually labeled by experts in the field, of which 80 were normal cases, and 58 were cases with manifestations of TB. Figure 4.7 shows the graphical diagram of the proposed methodology. Subsequently, the convolutional base of architecture VGG16 was applied as the feature extractor to obtain a vector of characteristics. Then, two methods of classification were imple-

mented: Cross-Validation (CV) using 10 and 20 folds and Train/Test with a training set of 80%, and 20% for testing.

In the Train/Test method, the best results were obtained using the SVM classifier, with 97% precision. The results are shown in Table 4.1. The columns show the classifiers with the respective evaluation metrics. Table 4.2 shows the results for the CV method, comparing the increase of folds, with better results in the SVM classifier using 20 folds, with 89% precision.

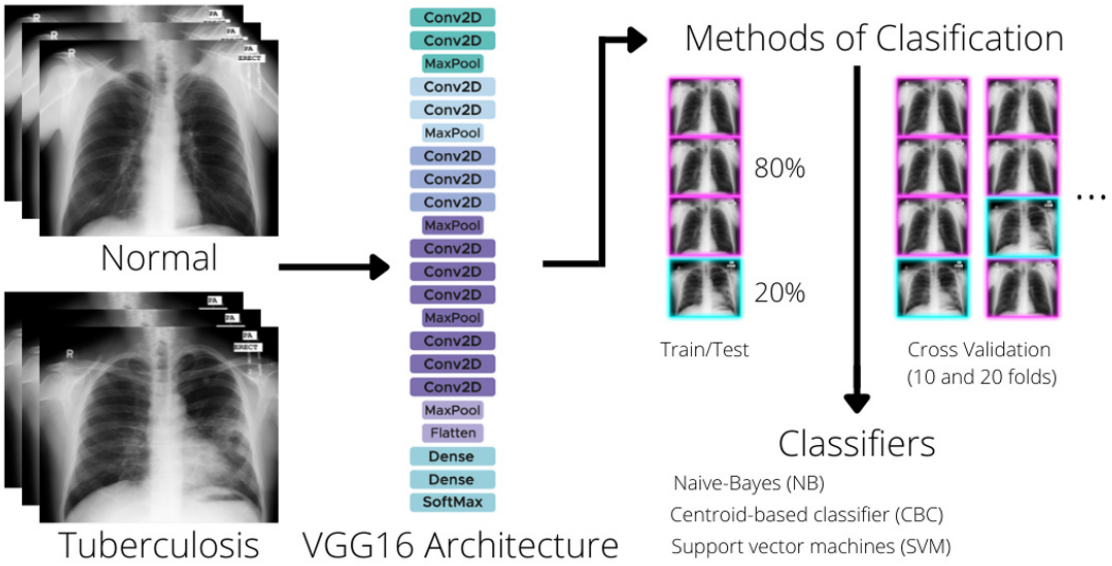


Figure 4.7: Proposed model for automatic classification of chest radiographs to detect tuberculosis.

Both methods used three classifiers: Support Vector Machine (SVM), Naïve Bayes (NB), and Centroid-Based Classifier (CBC), each widely used in state-of-the-art applications. To assess the performance of the proposed methodology, evaluation metrics such as accuracy and precision were used, based on the confusion matrix. These metrics enabled the identification of true and false predictions.

The proposed model for automatic classification of chest X-rays to detect tuberculosis obtained a 97% accuracy. This value is due to the improvement shown in Table 4.1 and Table 4.2. The reported model is competitive with the state-of-the-art model and is a support tool for diagnosing tuberculosis.

Table 4.1: Results with training and testing set scenario.

Evaluation metrics	SVM	NB	CBC
Accuracy	0.96	0.86	0.86
Precision	**0.97**	0.87	0.87
Recall	0.96	0.84	0.84
F1	0.96	0.85	0.85

Table 4.2: Results with cross-validation scenario.

Folds	Evaluation metrics	SVM	NB	CBC
20	**Accuracy**	0.86	0.76	0.74
	Precision	**0.89**	0.77	0.76
	Recall	0.85	0.75	0.73
	F1	0.85	0.73	0.72

4.2.4 CT imaging results

4.2.5 Post-COVID conditions

Coronavirus disease 2019, better known as COVID-19, is an infectious disease caused by the SARS-CoV-2 virus. This produces severe clinical pictures and serious cases such as pneumonia and acute respiratory distress syndrome. Infected patients have had post-COVID symptoms that worsen their physical and mental state. Using diagnostic imaging (radiography, tomography, and ultrasound) findings such as bronchial dilatations, pulmonary fibrosis, and opacities have been found. These conditions, considered sequelae of infection, are usually diagnosed three months after suffering from the disease and cause general symptoms, mainly affecting the brain, lungs, heart, and kidneys.

The present research proposes a supervised learning classification methodology through image processing using a deep neural network (50 intermediate layers, ResNet50) to identify post-COVID conditions, as shown in Figure 4.8. The dataset is classified into two categories: post-COVID findings and non-post-COVID findings. Due to its applications, image processing is a technique used in various research articles. It is used to extract regions of interest with properties potentially related to computer-aided medical diagnosis.

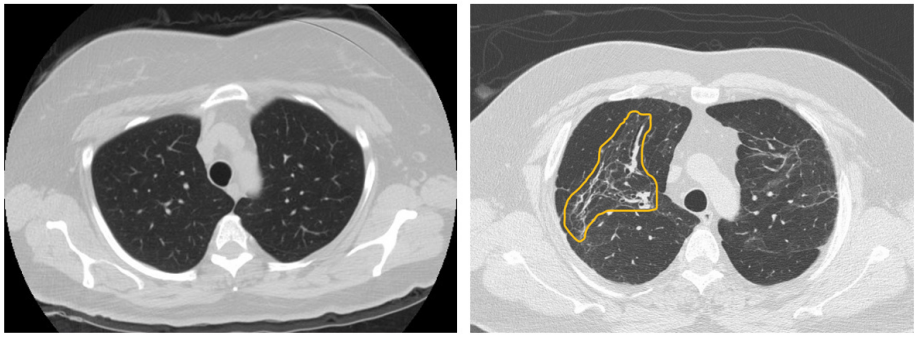


Figure 4.8: Simple chest tomography with the pulmonary window in the axial section, where a) no finding is observed and b) a typical pattern of pulmonary fibrosis (distortion of the pulmonary architecture) derived from COVID-19 (yellow outline).

The dataset used in the present investigation was provided by the Unidad Médica de Alta Especialidad No. 1 (High Specialty Medical Unit) of the Instituto Mexicano del Seguro Social (Mexican Institute of Social Security), T1-IMSS of León, Guanajuato, Mexico, from Mexican patients infected with COVID-19. The post-COVID diagnosis was made to 16 patients: approximately one year after their infections. The imaging study obtained the medical diagnosis through computed tomography (CT) of the chest, which is more sensitive than conventional radiography. For this reason, it was the main diagnostic method used in the pandemic. In the tomography, two types of cases were observed: normal and abnormal. 57 CT images with post-COVID findings and 57 images without findings were obtained (symmetrical set). The most common findings were ground glass image, which can be translated into an acute inflammatory process and possibly represents immature fibrosis that can resolve or progress over time, and pulmonary fibrosis; mainly made up of findings such as architectural distortion, traction bronchiectasis, and cobblestone pattern (see Figure 4.9).

The proposed model was based on the cross-validation classification scenario. The number of images with and without findings post-COVID was homogeneous and symmetric, with 57-57 tomographic. This type of classification is appropriate when there are not enough images in the dataset. As a result, we can see that the KNN classifier obtains the highest values in the precision metric, which increases as the

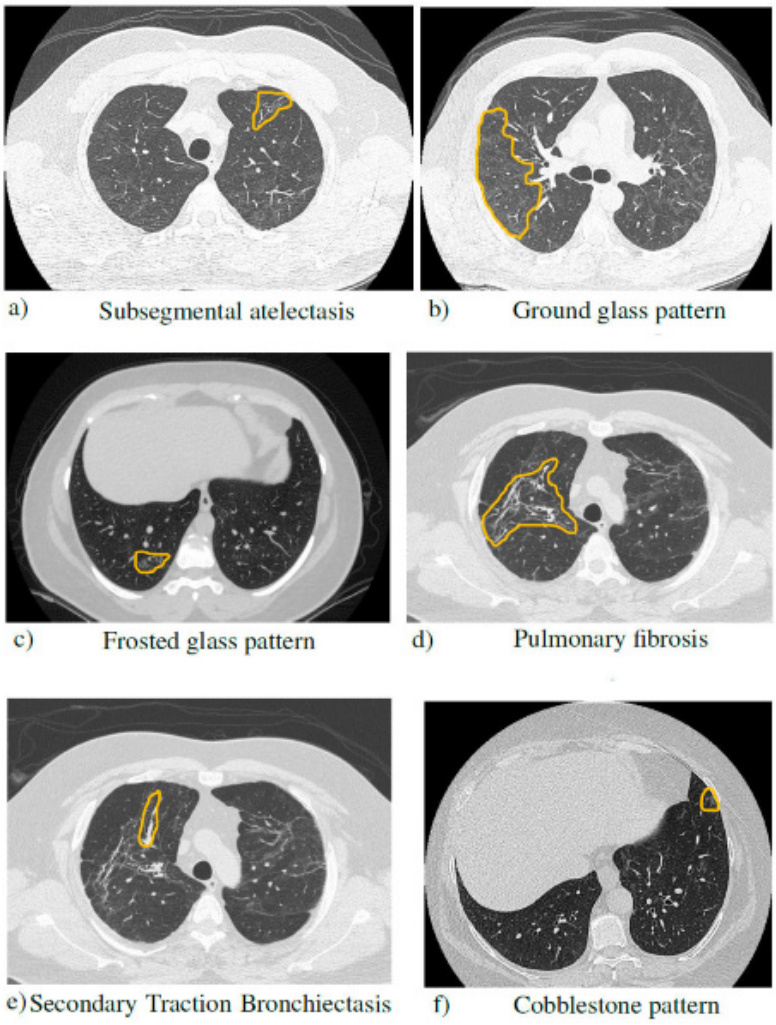


Figure 4.9: Simple chest tomography with lung window in axial section, where the findings of the pathologies are observed.

fold increases. Figure 4.10 presents the results. This indicates that the classification percentage in 10 divisions is 95%, in 20 divisions, it is 96%, and finally, in 30 divisions, it is 97%. The precision gives us the quality of the classification scenario and tells us the percentage of the positive class predicted by the model and the actual positive class of the dataset. The classifier with the best values tends to work optimally in small datasets, as with the present set, with 114 tomography images with and without post-COVID findings.

The ResNet50 convolutional neural network is an essential part of the methodology, since, from it, the characteristics are extracted through deep learning of the

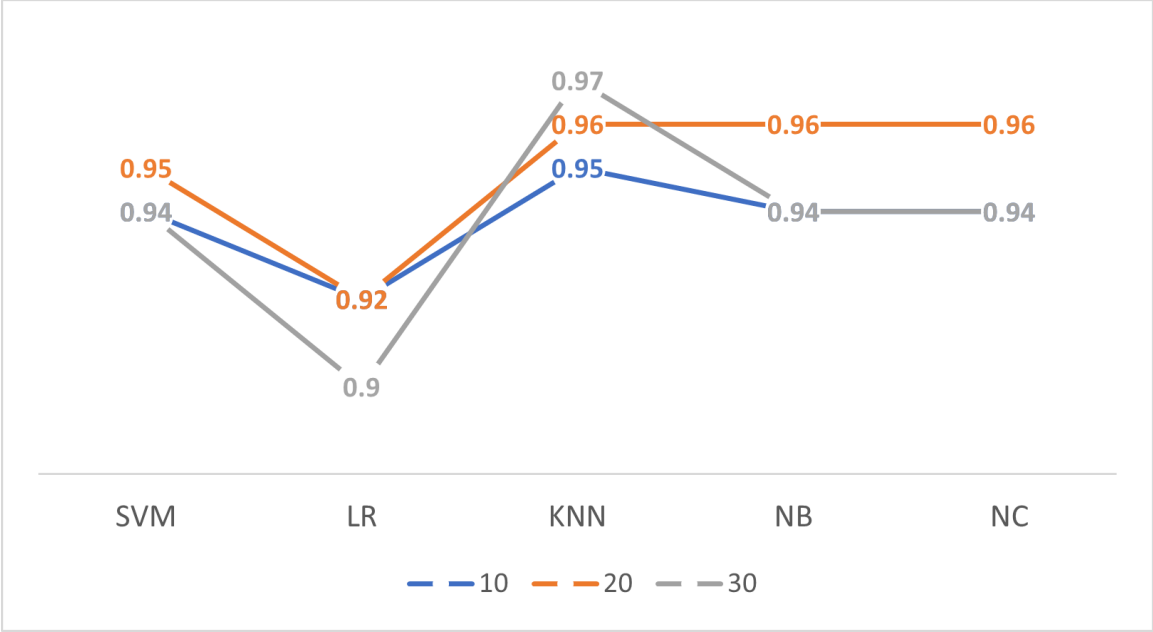


Figure 4.10: Learning methods in the three division experiments in the precision evaluation metric.

post-COVID findings in medical images and thus can be compared with those that do not have lesion findings through classifiers that are evaluated using metrics that provide us with the performance of the model. The efficacy of the proposed method, based on a dataset of Mexican patients, demonstrates that using the nearest neighbor classifier achieves a classification precision of 97% for binary labels indicating the presence or absence of post-COVID lung lesions.

4.2.6 Osteoporosis

Osteoporosis is a disease of adults that can affect both genders. The gold standard technique is bone densitometry, although it is inferred that it can be diagnosed through digital image processing. This work implemented and manipulated image classification methods using artificial intelligence algorithms. From the public repository Kaggle, with X-ray images of healthy patients and patients with osteoporosis, digital image processing techniques such as filters, neural networks for feature extraction, and machine learning techniques were developed. An X-ray image classifier obtained an Accuracy of 77%. It is concluded that it is a result with potential since

osteoporosis has been diagnosed in a human knee. In the first instance, a search was made for a database that would be useful for this research work. Several database pages were reviewed, including Kaggle, where the Osteoporosis Knee X-ray Dataset by Steve Python was found.

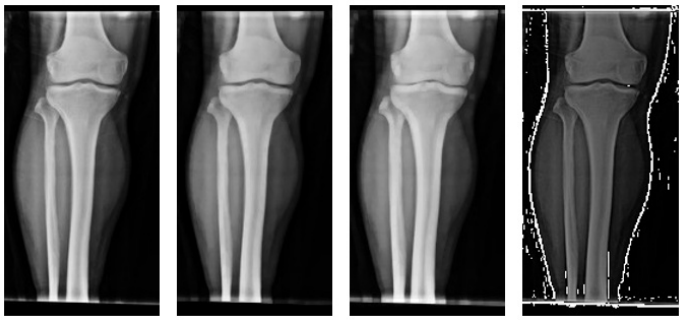


Figure 4.11: X-ray images of a person without osteoporosis, from left to right: unfiltered, with aperture filter, with dilatation filter, and with a high-pass filter.

All images in the dataset (185 with osteoporosis and 185 healthy subjects) were subjected to preprocessing, where three types of filters were applied to highlight characteristics of X-ray images such as lines and edges. The three filters applied were two morphological filters: aperture and dilatation filters, and a high-pass filter. The dilatation filter has the effect of increasing the size of objects or causing small holes to close, while the aperture filter eliminates elements that protrude from a given object, eliminates fine connections, and softens contours. The high-pass filter detects edges or sharp edges in an image. These filters resulted in images with marked visual characteristics that serve the neural network as a classification aid (see Figures 4.11 and 4.12).

After preprocessing by filtering, the images were subjected to feature extraction using different neural networks, the first of which was VGG16, followed by ResNet50 and InceptionResNetV2. With each of these networks, vectors were generated from the 185 images of healthy patients and the 185 images of patients with osteoporosis, these vectors could be visualized as two text files for each network in .txt format. Individual .txt files were also generated, in which all the vectors of the images were placed for later use, placing the vectors of characteristics of normal patients first, followed by the vectors of patients with osteoporosis.

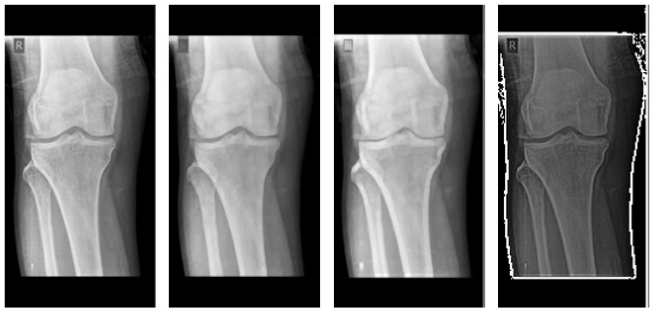


Figure 4.12: X-ray images of a person with osteoporosis, from left to right: unfiltered, with aperture filter, with dilatation filter, and with a high-pass filter.

The results were satisfactory, considering the quantity and quality of the images we worked with. The preprocessing of the images showed an evident influence on the classification. In the same way, the SVM classification method yielded favorable results that show the preference for its use in the works of other researchers. Table 4.3 shows the results. Using neural networks and classification methods in the medical field is very useful in enhancing technological advances around health, reducing human error, or supporting non-specialized doctors in cases of osteoporosis, such as general practitioners.

Table 4.3: Classification results.

Filter	**CNN**	**Classification system**	**Classifier**	**Accuracy**	**Precision**
Dilatation	ResNetV2	TTS	KNN	**0.77**	**0.77**
Aperture	ResNet50	CV	SVM	0.75	0.76

The osteoporosis images used laid the groundwork for the work done, while preprocessing helped to strengthen the model so that the accuracy and precision of this classification could be increased. The application of morphological aperture and dilatation filters proved to be an enhancer for recognizing lines and edges in healthy and osteoporosis images, helping the networks to extract features that correctly identified images of their respective class. Likewise, a satisfactory result was observed with the K Neighbors Classifier and Support Vector Machine classifier, highlighting the three best results of all the work executed. These results, which were almost 80%

demonstrated that a good starting point was achieved in this classification exercise, and an opportunity for improvement in its subsequent application to real-life clinical cases in health systems (see Figure 4.13).

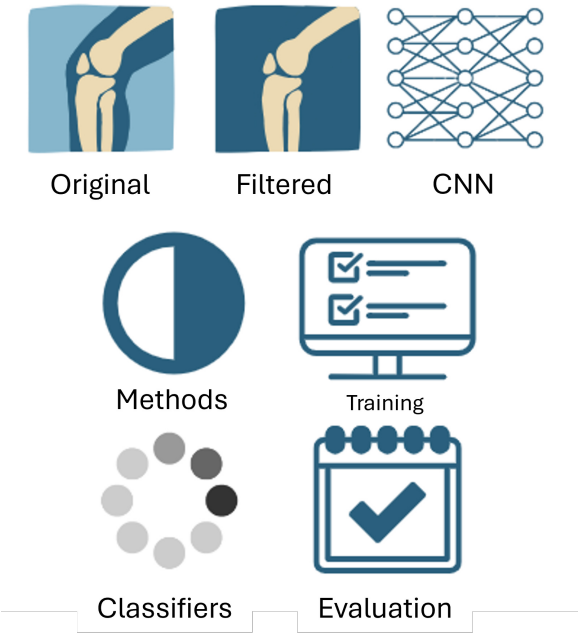


Figure 4.13: Model proposed.

4.2.7 Osteoarthritis

Osteoarthritis (OA) is a degenerative joint condition characterized by the breakdown of cartilage, leading to pain, stiffness, and swelling. It is a common condition, particularly in older adults, and can significantly impact quality of life. Imaging techniques play a crucial role in diagnosing OA. Ultrasound is suitable for assessing superficial structures but has limitations in evaluating deeper ones. MRI is the gold standard for visualizing bone and soft tissues and accurately detecting early-stage cartilage degeneration. CT can quantify joint space width and detect early structural changes, although its use is generally reserved for specific cases. Emerging options like PET-MRI offer a comprehensive assessment of joint alterations but are limited by high cost and time-intensive nature. While radiographs are valuable for visualizing bony changes, they have limitations in detecting early-stage OA and cannot directly visu-

alize cartilage.

It is worth pointing out that radiographs are of great value for visualizing bone changes; however, this technique has limitations regarding detecting early OA and cannot visualize cartilage directly. In developing countries, where MRI and CT techniques are unavailable because of their high cost and infrastructural demand, radiography remains one of the accessible and inexpensive methods. The Kellgren-Lawrence (KL) classification system, is widely used to assess knee OA. KL defines OA in five grades based on key changes in the knee joint. These grades are (see Figure 4.14):

- Grade 0: Normal, no radiographic signs of OA.
- Grade 1: Doubtful narrowing of joint space and possible osteophytic lipping.
- Grade 2: Definite osteophytes and possible narrowing of joint space.
- Grade 3: Moderate multiple osteophytes, definite narrowing of joint space, some sclerosis, and possible deformity of bone ends.
- Grade 4: Large osteophytes, marked narrowing of joint space, severe sclerosis, and definite deformity of bone ends.

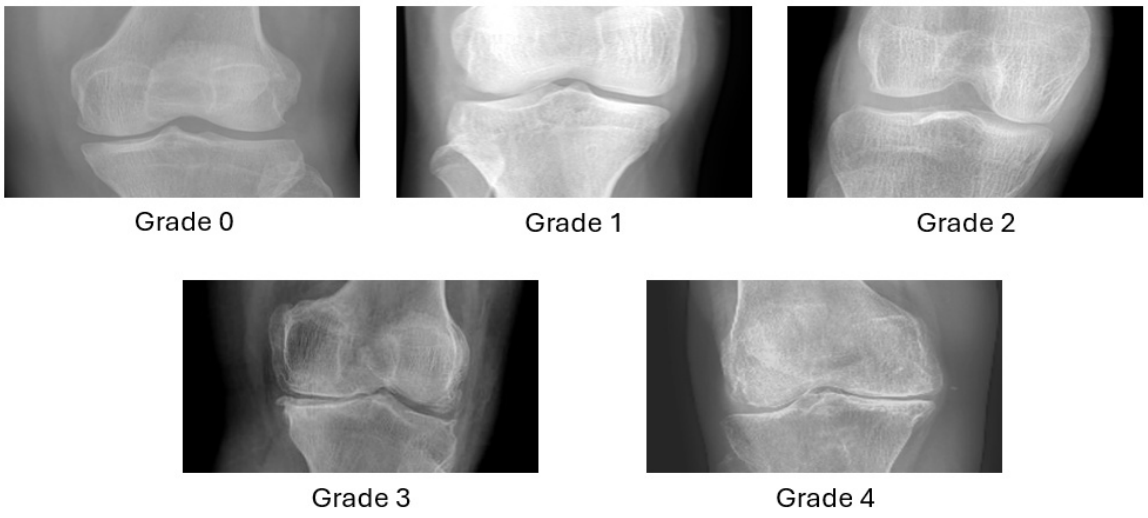


Figure 4.14: Levels of osteoarthritis disease according to KL criteria.

The dataset used for OA classification is located in the Kaggle repository, under the name Annotated Dataset for Knee Arthritis Detection. It has 1650 images of clinical examinations divided, as shown in Figure 4.14. The VGG16 neural network was used as a feature extractor, and different experiments were performed, which gave rise to the results presented in Table 4.4.

Table 4.4: Results of paired experiments in OA disease.

Classification system	**Classifier**	**Experiment**	**Accuracy**
TTS	SVM	Grade 0 vs Grade 1	**0.71**
CV			0.70
TTS		Grade 0 vs Grade 2	**0.84**
CV			**0.84**
TTS		Grade 0 vs Grade 3	0.65
CV			**0.89**
TTS		Grade 0 vs Grade 4	**0.98**
CV			**0.98**

As we can see, the VGG16 convolutional base provides great robustness to the experiment as it manages to obtain the most relevant features of each experiment. As the grade of OA disease progresses, the model acquires better results, reaching a classification of 98%. This fact is important since the model can improve with the preprocessing block, which indicates greater relevance and optimal values in a robust model for classifying OA in the knee according to its grade of advancement.

4.2.8 Pulmonary fibrosis

In this study, we focus on addressing the prolonged effects of COVID-19, specifically pulmonary fibrosis, through automated classification of computed tomography (CT) images from patients in León, GTO, Mexico. We employed a convolutional neural network (CNN) VGG16 and image enhancement filters like Meijering and Roberts, to optimize image quality. Prolonged COVID has been linked to complications, including pulmonary fibrosis, underscoring the need for early detection. Our model, supported by a comprehensive dataset, achieved classification accuracy exceeding 97%, successfully distinguishing between patients with and without pulmonary fi-

brosis. The combination of enhancement filters and CNN VGG16 proved crucial in this success, highlighting the potential of our model for early detection and effective management of pulmonary fibrosis in long-term COVID patients. This promising approach may significantly impact clinical practice, enhancing outcomes and enabling timely interventions. In summary, we present an effective method contributing to the timely diagnosis and treatment of pulmonary fibrosis in long-term COVID patients.

In Figure 4.15, we can see that pulmonary fibrosis was the most frequent of all the findings of the most important sequelae in the dataset obtained. For this reason, experiments were conducted to optimize the model to classify fibrosis from the rest of the dataset presented in the post-COVID conditions section. Pulmonary fibrosis occurs when lung tissue is damaged and scarred. This thickened and stiffened tissue makes breathing more difficult and prevents the blood from receiving enough oxygen. Below is the development carried out to detect pulmonary fibrosis in a data set of Mexican patients.

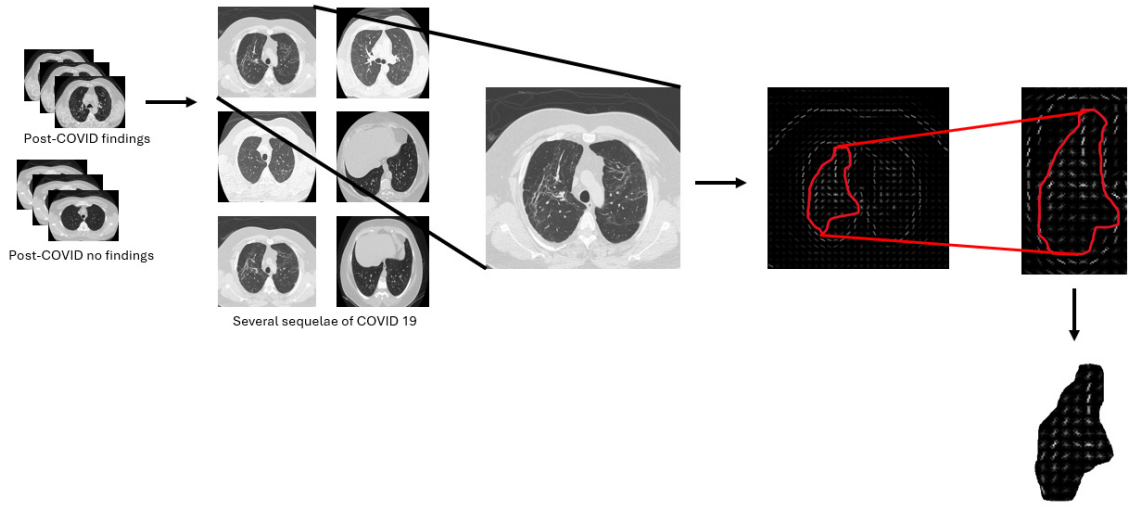


Figure 4.15: Region of interest extraction process from the image set with 96 CT scans, labeled in two classes: (left) 57 healthy axial CT scans and (right) 39 axial CT scans with pulmonary fibrosis findings.

Two experiments were performed, one without masks and one with masks. However, in both experiments, three filters were used to highlight the scars of the alveoli. Figure 4.16 shows the results of filtering the original dataset.

As mentioned, the second experiment consisted of circular masks that eliminated

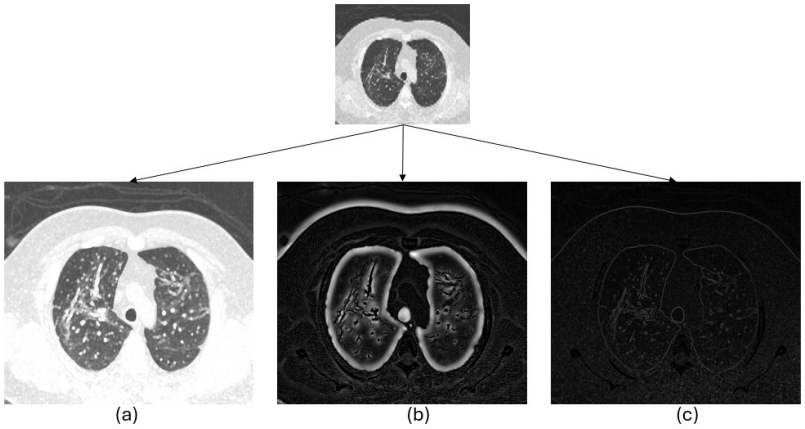


Figure 4.16: Preprocessing of the images from the dataset to highlight the features and quality of the classes. (a) Dilatational morphological filter, (b) Meijering filter, and (c) Roberts filter.

the body tract seen in the tomography. It is worth noting that in experiment one, better results were obtained from one filter than from the other filters (Meijering). Figure 4.17 shows the mask used for tomography to eliminate the tract and mark the region of interest for the alveoli, to obtain better results in the model.

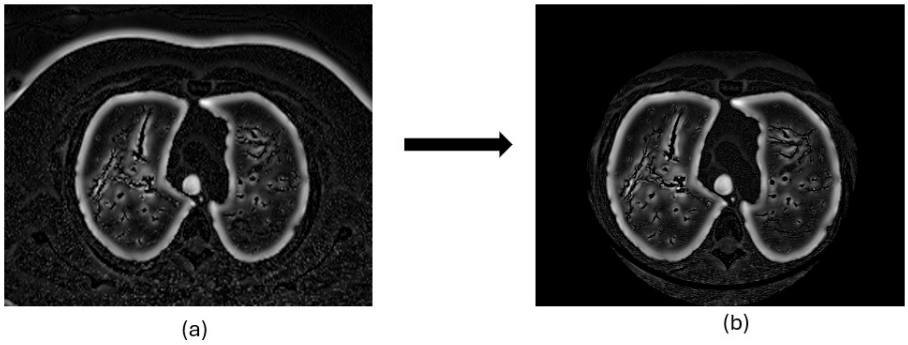


Figure 4.17: Application of masks to highlight the region of interest. (a) Experiment 1, with the original images and the filters; (b) Experiment 2, with the results of the best filter and circular/ellipse mask.

The cross-validation classification scenario is a technique that allows for evaluating the statistical analysis results and ensures that they are independent of the partition between training and test data. The results are shown in Tables 4.5 and 4.6, with two evaluation metrics: accuracy and precision. The first indicates how often, in general, the classifier is correct, and the second how often it predicts fibrosis. The classifier

algorithm used was k-NN, which obtained the optimal values of the classification model (97%). Therefore, we have generated a model that classifies pulmonary fibrosis with a 3% error.

Table 4.5: Result of the first experiment.

Image	Accuracy	Precision
Original	0.87(+/- 0.14)	0.87 (+/- 0.16)
Roberts	0.81(+/- 0.12)	0.87 (+/- 0.12)
Meijering	**0.92 (+/- 0.12)**	**0.92 (+/- 0.12)**
Dilatation	0.87 (+/- 0.14)	0.87 (+/- 0.14)

Table 4.6: Result of the second experiment.

Meijering filter	Accuracy	Precision
No mask	0.92(+/- 0.12)	0.92 (+/- 0.12)
Mask	**0.95(+/- 0.09)**	**0.97 (+/- 0.05)**

4.2.9 Automated detection of free fluid in the abdominal cavity

An algorithm for the automated, non-invasive detection of free fluid in the abdominal cavity using Computed Tomography (CT) images was developed, aiming to significantly optimize detection procedures and provide an efficient solution for medical diagnosis. The algorithm identifies the presence of fluid in CT images and generates a detailed three-dimensional reconstruction of the abdominal cavity and the fluid, facilitating the medical diagnostic process. Using Hounsfield Units (HU), the program accurately segments areas corresponding to free fluid in the abdominal cavity by recognizing pixel intensities in the images and distinguishing between anatomical structures. An essential achievement of this approach is its ability to identify free fluid presence using only an abdominal CT scan, providing detailed information about its specific location and distribution, and enabling the visualization of the severity of the condition. This approach represents a significant advancement in automated free fluid detection in the abdominal cavity, offering healthcare professionals a valuable

tool for precise and rapid condition identification. The detailed three-dimensional representation enhances the medical diagnostic process in clinical settings.

In Chapter 1, Hounsfield units were described. CT imaging provides cross-sectional views of the abdominal cavity, with each scan offering data on density variations among tissues. The organs within the abdominal cavity; liver, pancreas, spleen, and kidneys, display distinct Hounsfield unit values, enabling precise anatomical mapping. These values help differentiate free fluid from surrounding tissues, which is crucial for identifying acute pathologies. The free fluid typically registers within the 0 to 30 HU range depending on its composition, while denser fluids, such as blood, present higher values. The 90-110 HU range mentioned in this study corresponds to a subset of free fluid types but may vary based on clinical context (see Figure 4.18). Recent studies corroborate this variability, reinforcing the need for flexible threshold adjustments during segmentation. A detailed segmentation process involves edge detection and thresholding to capture relevant HU values for fluid identification.

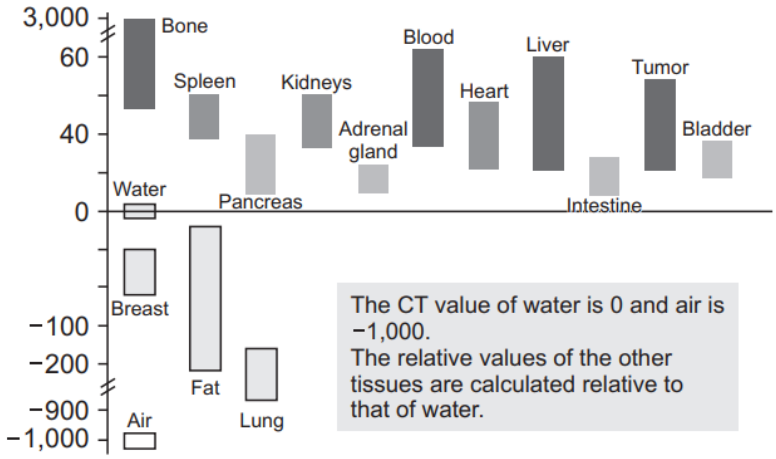


Figure 4.18: Hounsfield scale of computed tomography (CT) numbers for various tissues.

CT segmentation techniques for free fluid detection employ advanced algorithms based on density thresholds, active contour methods, or convolutional neural networks. These methods accurately identify and delineate areas with characteristic free fluid densities, even in the presence of adjacent presence structures with similar densities. Specific HU ranges are applied to distinguish free fluid from other abdominal

tissues or fluids. Adapting these ranges according to pathology or anatomical location allows for more precise and reliable free fluid identification, mitigating false positives or negatives. The segmentation techniques utilized in this study include thresholding based on HU values and edge detection algorithms. The first phase identifies potential fluid regions by isolating areas with HU values indicative of free fluid. The second phase involves active contour methods to delineate fluid boundaries. Post-processing filters, such as morphological operations, are applied to refine the segmentation and eliminate noise.

Although the dataset includes 23 studies, additional validation is planned with larger datasets to account for diverse clinical presentations (see Figure 4.19). This study utilized a dataset of abdominal computed tomography images obtained through collaboration with the IMSS UMAE T-1 Hospital in León, Guanajuato.

Table 4.7: Workflow of the free fluid detection algorithm using CT imaging.

Step	**Description**
1. Image acquisition	CT images are acquired in Hospital IMSS T1.
2. Preprocessing	Images are converted from DICOM format to a standard gray-scale image.
3. Thresholding	Pixel values are thresholded based on HU to isolate fluid densities (0 to 30 HU).
4. Morphological operations	Erosion and dilatation are applied to remove noise and refine the segmented regions.
5. Segmentation	The final segmentation identifies areas of potential free fluid within the abdominal cavity.
6. Comparison with Radiologist Annotations	The segmented regions are compared with manual annotations by radiologist for validation.

The dataset consisted of 23 CT studies from various adult patients of both sexes. The characteristics of each study varied, including different numbers of slices and varying slice thicknesses. Each case exhibited the accumulation of free fluid in different quantities and locations within the abdominal cavity. In addition to the image datasets, the studies were accompanied by annotations made by trained professionals in free fluid detection, intended for comparison purposes. In Table 4.7, the workflow

of the model is observed. Results indicate that regions identified as free fluid displayed homogeneous pixel intensity distributions, corresponding to Hounsfield unit values ranging from 0 to 30 HU. In contrast, regions corresponding to water transit displayed more heterogeneous patterns, reflecting the dynamic nature of water accumulation and movement within the abdominal cavity (see Figure 4.20).

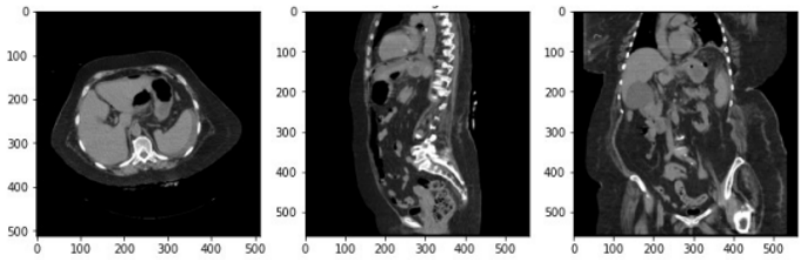


Figure 4.19: Example of dataset images. An axial section on the left, a sagittal section in the center, and a coronal section on the right.

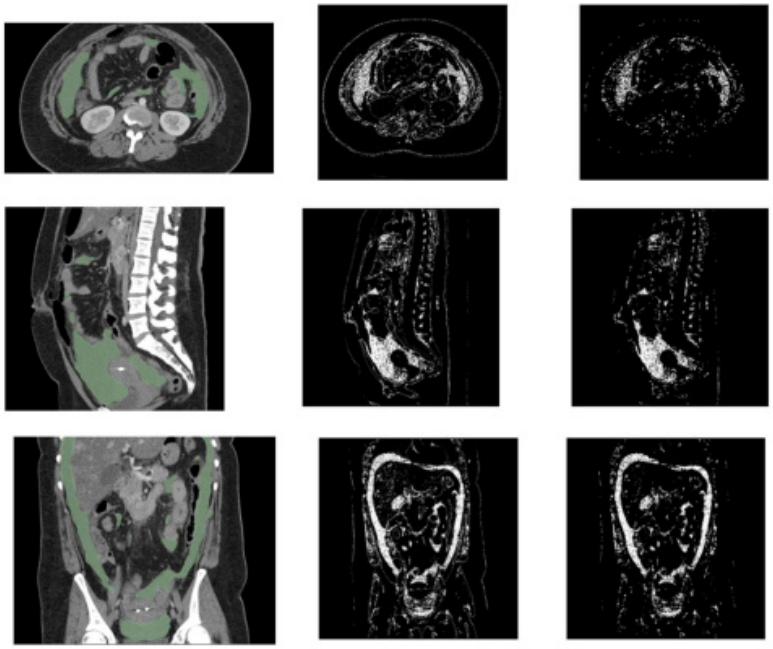


Figure 4.20: In the first column, there are the reference images of the free fluid in the abdominal cavity previously identified by the medical expert. In the second column, the free fluid is segmented based on image processing techniques in conjunction with the HU units. Lastly, a morphological aperture filter is applied in the third column to reduce noise and improve liquid detection in the three sections.

4.2.10 Renal lithiasis

Renal lithiasis is characterized by the presence of kidney stones, which are chronic and morbid. This study used samples of CT scans on renal lithiasis from patients with and without renal lithiasis, provided by the Instituto Mexicano del Seguro Social (IMSS), León, GTO. The studies are classified by stone size: less than 10 mm, ranging from 10 mm to 19 mm, and larger than 20 mm. The studies were contrasted with the control sample of patients without renal lithiasis. Two methodologies were employed for the classification model using CNN: samples with augmentation techniques using Convex Hull contours and without Convex Hull to homogenize sample sizes. The CNNs used in this study are VGG16, ResNet 50, and Inception ResNet V2. The results show better accuracy in the models that employ a Convex Hull contour, reaching an accuracy of 94% and precision of 93% for the Naïve Bayes Multinomial statistical test, the highest among all classifiers.

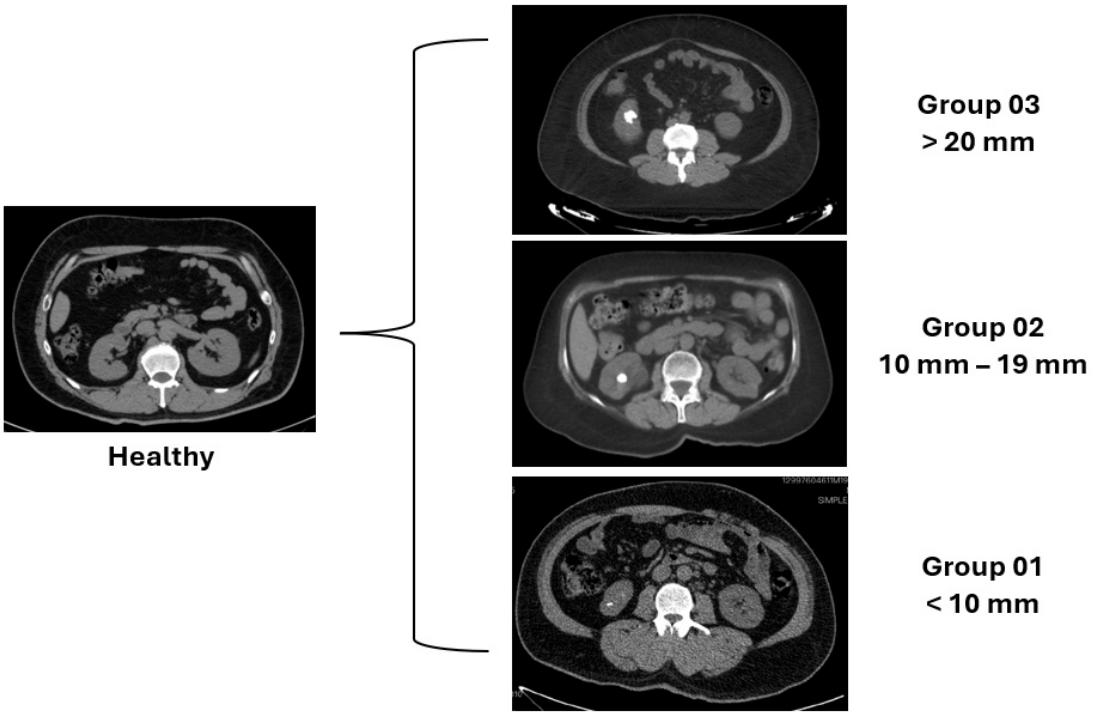


Figure 4.21: Dataset of renal lithiasis.

The studies included patients in three classifications: Patients with stones smaller than 10 mm, those between 10 and 20 mm, and those with stones larger than 20 mm.

Figure 4.21 shows the CT scans provided by the IMSS. The sample sizes provided are small and non-homogeneous. Data augmentation techniques were employed to homogenize the sample to a size of 56, (rotation 1°, zoom, and cut).

We implemented the Grad-CAM technique to produce visual explanations of the convolutional neural networks used and verify whether they effectively identified the areas where the models are concentrated, the area of interest, and kidney stones. Figure 4.22 shows the results.

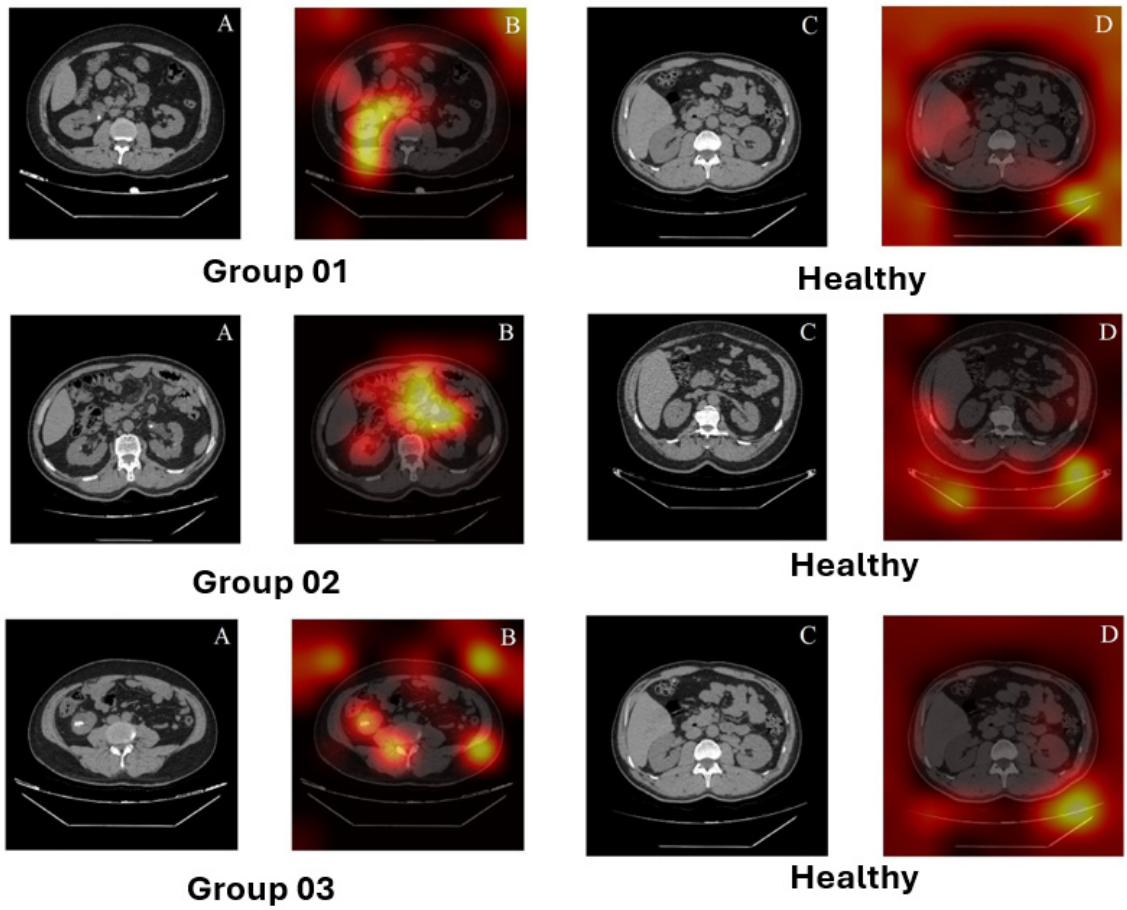


Figure 4.22: The Grad-CAM technique shows that for each group, the model detects kidney stones with a larger diameter better than healthy patients, where the model does not detect any region of interest.

4.2.11 MRI imaging results

4.2.12 Brain MRI

A brain tumor is a mass or lump of abnormal cells found in the brain. There are two main types of tumors: cancerous (malignant) and benign. Cancerous tumors can be divided into primary tumors, which start within the brain, and secondary tumors, which have spread from elsewhere, known as brain metastasis tumors. Today, deep learning, a subfield of AI, can classify, diagnose, segment, and provide a medical diagnosis; in other words, it is a computer-aided diagnosis.

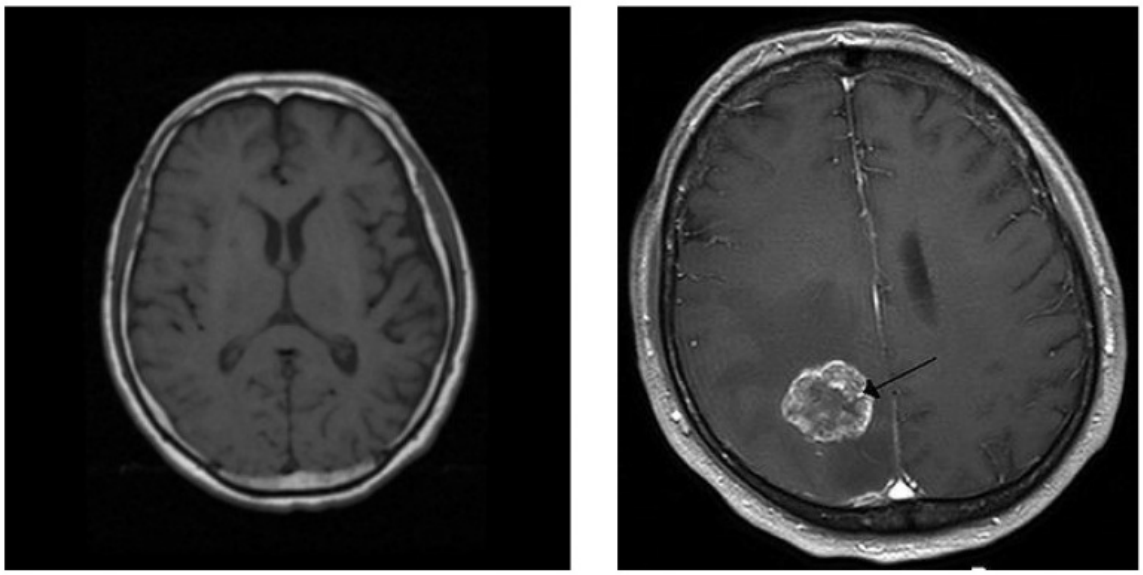


Figure 4.23: Set of images divided into benign (left) and malignant (right) tumors.

The proposed model consists of three parts: Data set acquisition and preprocessing, feature vector and classification system, and evaluation and diagnostic metrics. The images are collected from Kaggle datasets of Brain MRI Images. The dataset is labeled into two classes of YES and NO based on the presence of tumors, as shown in Figure 4.23. There are 155 images with brain tumors; the remaining 98 images are of normal brains. The dataset is in the Kaggle repository; Brain MRI Images for Brain Tumor Detection. It consists of MRI scans of two classes: no tumor, benign, and tumor, malignant. Transfer learning uses the convolutional base of a pre-trained

network called VGG16 as a feature extractor to obtain the feature vector that enters the classification system. The model is illustrated in Figure 4.24.

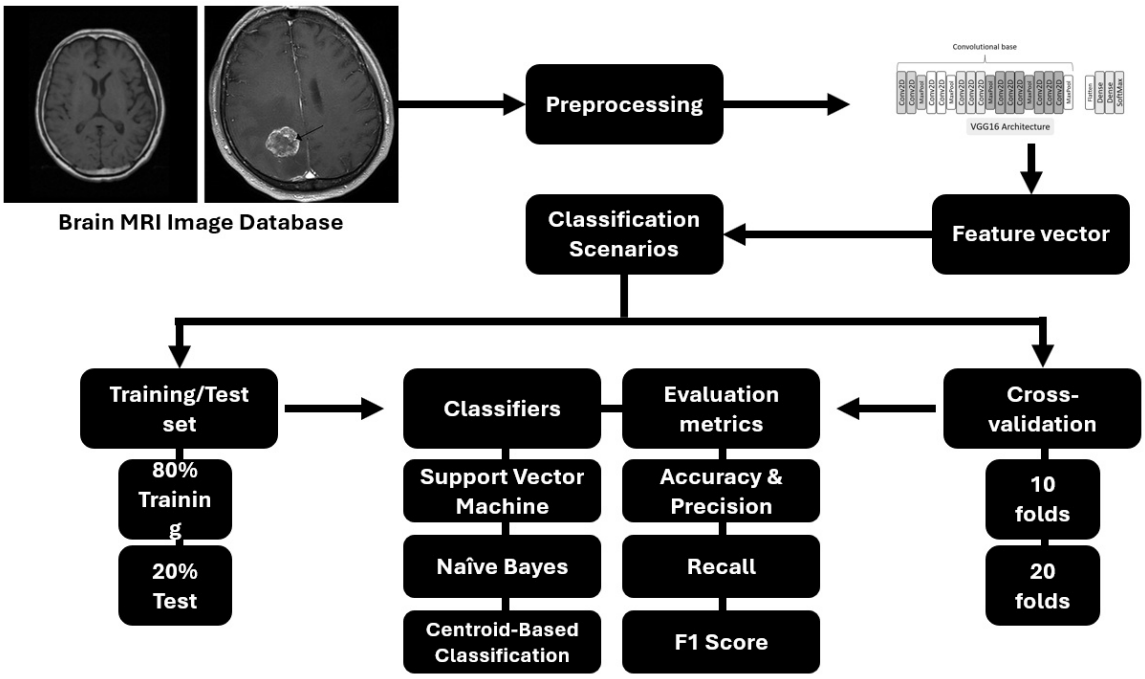


Figure 4.24: Proposed model to detect brain tumors.

The results of the model are presented in Table 4.8:

Table 4.8: Most relevant results of the proposed model.

Classification system	Classifier	Accuracy	Precision
TTS	SVM	0.89	0.90
CV	SVM	0.90	0.91

Regarding the results, it is important to highlight that the Support Vector Machine (SVM) classifier obtained the best values in the training and test set technique, and in particular, a high level of precision of 90% and 91% with CV. These results suggest that our model is effective in this task.

4.2.13 Identification of microbleeds

The detection of microbleeds in MRI studies using the susceptibility-weighted image (SWI) technique is presented. The SWI technique has been shown to play a relevant

role in the identification of microbleeds, unlike conventional MRI techniques, the sensitivity is higher for detecting microbleeds and iron deposits. This work presents a technique for the quantitative detection of microbleeds through the implementation of Shannon entropy. A statistical analysis of the results complements it and establishes a specific range that allows for the early detection of these structures in future research. Preliminary results suggest that this method represents a significant advancement in the accurate and timely detection of cerebral microbleeds, offering an alternative to traditional Magnetic Resonance approaches, suggesting that artificial intelligence is a promising path for deeper investigations in this field. Mathematically, Shannon entropy, denoted as $H(x)$ for a discrete variable X with n possible outcomes, is calculated using the formula:

$$H(x) = -\sum_{i=1}^{n} P(x_i) \cdot \log_2(P(x_i)), \tag{4.1}$$

where $P(x_i)$ represents the probability of outcome x_i occurring, Shannon entropy can also be interpreted as the minimum average amount of bits required to represent the information contained in the random variable X. The proposed method consists of the following general steps:

1. Define a function to calculate the entropy of the 50 images from healthy patients.
2. Divide each image into sections and calculate the entropy for each section.
3. Calculate the means and standard deviations per section.
4. Load a new image for analysis where the patient has microbleeds.
5. Divide the new image into sections and calculate the entropy for each section, repeating steps 2 and 3 but for the new image.
6. Compare the entropies of the new image with the means and standard deviations.

Shannon entropy was implemented to assess the complexity of information in each image segment. This analysis was applied to 50 images, obtaining entropies for each

section in each image. Below are some results obtained according to the selected images, along with their respective entropy values and Z-scores. CSIRO provided the database and the Biomedical Informatics Group ResDevCons under the supervision of Jurgen Mejan-Fripp by CSIRO (Australia).

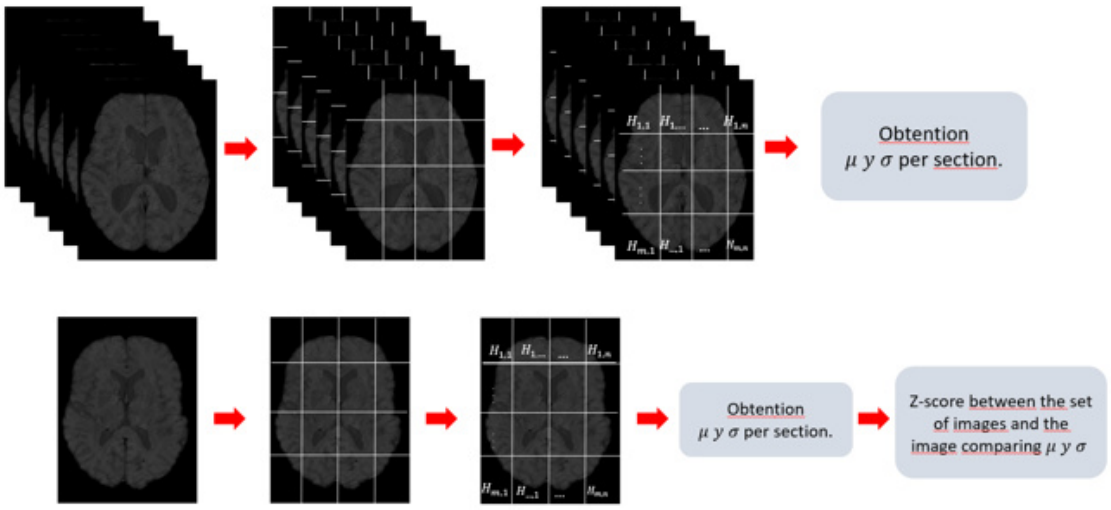


Figure 4.25: Diagram outlining Phase I and Phase II.

Phase I: Each image was divided into 16 equal sections. Phase II: Each image was divided into 32 equal sections. The Z-score was applied to quantify the standard deviations of entropy concerning the mean of images from healthy patients, allowing for an accurate assessment of variations in image complexity.

A Pearson correlation analysis was conducted between entropy and Z-score values, providing a deeper understanding of the relationship between these parameters. The similarity in entropy values confirmed the robustness of the results obtained. Two scatterplots with regression lines were analyzed, representing entropy data and their respective Z-score values, revealing correlations of 0.9418 and 0.8295, respectively (see Figure 4.26). Furthermore, according to the obtained data, an interval allows us to detect a possible cerebral microbleed through its entropy. This interval is described using a standard deviation of the data. This is considered an important step towards the statistical quantification of a microbleed. Considering that this could be a significant step towards early detection of cerebral microbleeds for neurodegenerative

diseases is important.

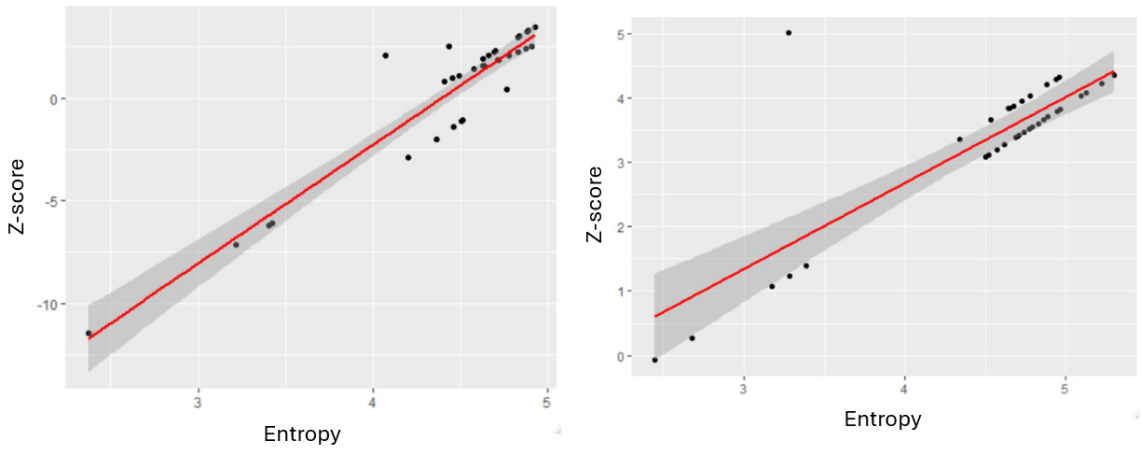


Figure 4.26: Statistical analysis of Shannon entropy versus Z-score.

This method achieves the quantitative detection of cerebral microbleeds through the statistical analysis of entropy, obtaining a range for potential microbleed detection. The values obtained provide valuable information that can significantly contribute to the early detection of these anomalies in the human brain. Despite being a preliminary work, its relevance is highlighted as it represents an important step towards accurately identifying microbleeds.

4.2.14 Applications of AI in the classification of MRI

Medical imaging is pivotal in diagnosing and treating diseases, offering intricate visual insights into the human body. Among the array of available imaging techniques, MRI has witnessed substantial growth in adoption due to its capacity for capturing high-resolution images that exhibit exceptional contrast between soft tissues. The accessibility of magnetic resonance imaging has surged thanks to technological advancements and heightened recognition of its clinical value.

These images (see Figure 4.28), obtained from various anatomical regions and under diverse protocols, furnish indispensable information about anatomical structures, functions, and potential abnormalities. Nevertheless, the interpretation of these images presents formidable challenges. Manual analysis by radiologists can

be labor-intensive, reliant on expertise, and vulnerable to inter-observer variations. Furthermore, each patient's burgeoning volume of images underscores the imperative for precise and efficient analysis to bolster clinical decision-making. In this context, the application of AI in classifying magnetic resonance images has emerged as a promising solution.

AI holds the potential to process large volumes of images swiftly and accurately, thereby bolstering clinicians in the early detection, characterization, and ongoing monitoring of diseases. Leveraging machine learning techniques and convolutional neural networks, the development of automatic classification models for medical images has demonstrated their competitiveness compared to traditional methods.

These models excel in discerning subtle patterns and features within MRI, thus facilitating precise diagnoses and prognoses for conditions. Figure 4.27 illustrates the applications of AI in the classification of MRI.

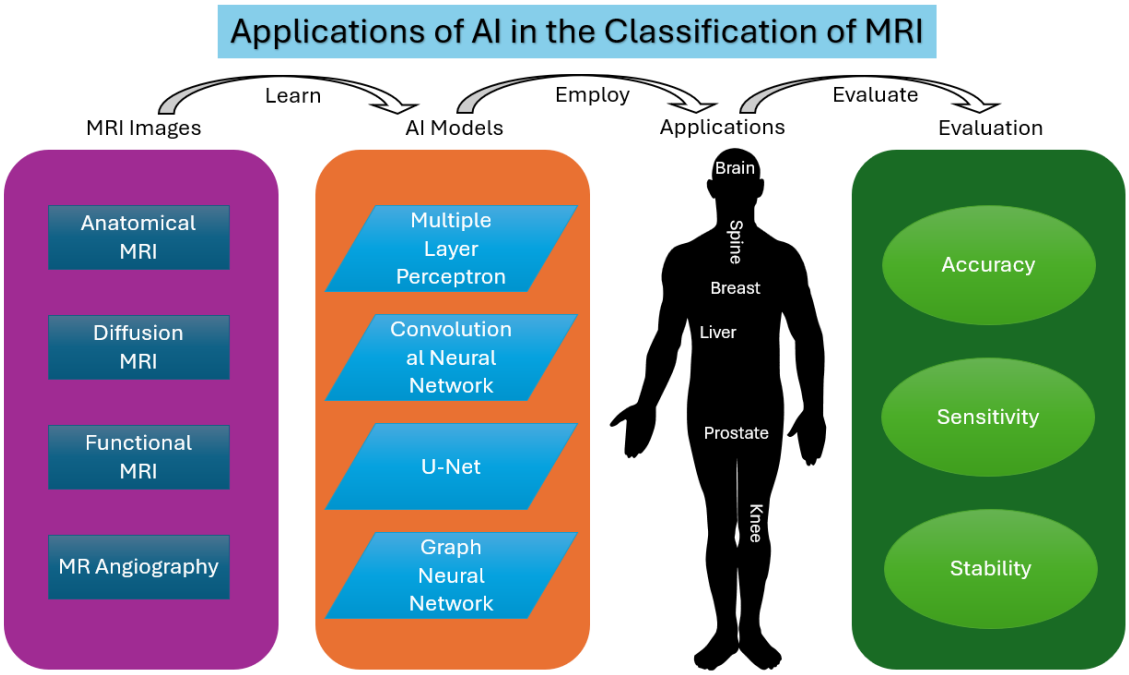


Figure 4.27: First review MRI images and select the previously labeled dataset by an expert. Next, select AI models applied to learn those MRI images. Then, applied those to MRI applications that employ AI models. Finally, discuss the evaluation metrics proposed to evaluate these AI models' performance.

In summation, given the current landscape of medical imaging with the expanding

availability of magnetic resonance images and the compelling need for precise and efficient analysis to underpin clinical decisions, the application of artificial intelligence in image classification is a field of research and development of profound significance.

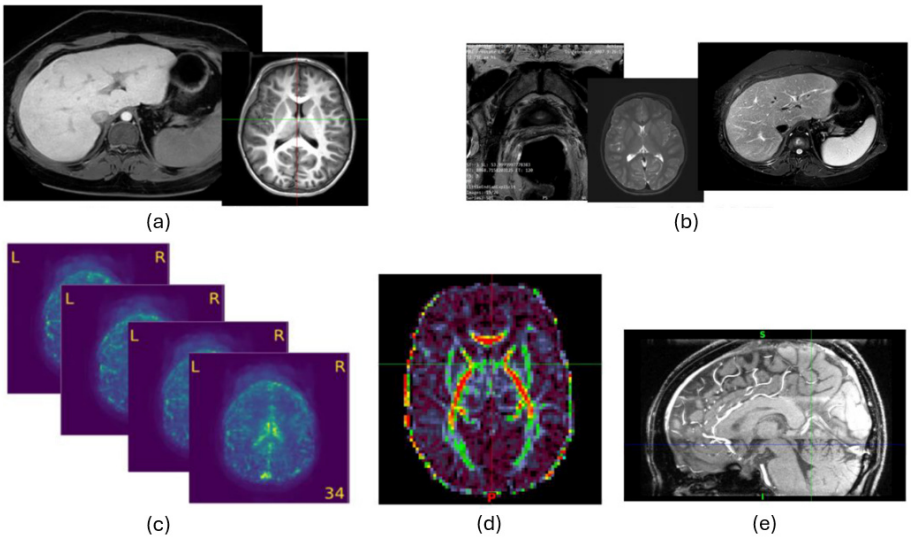


Figure 4.28: Illustration of common MRI images. (a) T1-weighted MRI; left: Liver; right: Brain [neonate], (b) T2-weighted MRI; left: Prostate; middle: Brain [neonate]; right: Liver, (c) functional MRI, (d) diffusion tensor imaging, and (e) MR angiography.

Chapter 5

Discussion

The application of artificial intelligence in the realm of biomedical imaging has shown tremendous promise in enhancing diagnostic accuracy, especially for a wide range of pathologies, including pneumonia, tuberculosis, post-COVID conditions, osteoporosis, osteoarthritis, and pulmonary fibrosis. The modular model proposed in this study serves as a robust framework for addressing the complexities and variations inherent in medical imaging data, allowing for effective classification, segmentation, and detection of these conditions across various imaging modalities. The results of our study on automated free fluid detection in abdominal CT imaging, renal lithiasis classification, brain MRI tumor detection, and the identification of microbleeds highlight the significant advancements made in the field of medical imaging through the application of AI and machine learning techniques.

5.1 X-ray imaging

Pneumonia and tuberculosis are significant public health concerns characterized by inflammatory processes affecting the lungs. The modular model's ability to leverage deep learning techniques, particularly convolutional neural networks (CNNs), demonstrates notable efficacy in distinguishing between various forms of pneumonia and identifying early signs of tuberculosis. The integration of Hounsfield unit (HU) thresholds for computed tomography (CT) imaging aids in differentiating between

these conditions are based on density variations. Given the potential overlap in radiographic findings, this capability is crucial where pneumonia may mimic the lesions seen in tuberculosis.

5.2 CT imaging

The post-COVID landscape presents unique challenges, including accurately assessing lung damage and associated sequelae. The proposed modular model has been adapted to identify pulmonary changes characteristic of post-COVID conditions, utilizing advanced image segmentation techniques to highlight areas of fibrosis, ground-glass opacities, and other alterations in lung architecture. This facilitates better clinical decision-making and provides a quantifiable method for monitoring disease progression and therapeutic outcomes.

In musculoskeletal imaging, osteoporosis and osteoarthritis are prevalent conditions requiring precise assessment for effective management. The modular model's application to identify bone density changes associated with osteoporosis, alongside cartilage degradation indicative of osteoarthritis, underscores its versatility. The segmentation of anatomical structures, supported by the morphological operations detailed in the results chapter, enables clinicians to visualize and quantify the extent of these conditions, fostering improved patient outcomes.

Accurate detection and characterization of pulmonary fibrosis are critical for timely intervention. The modular model's capability to integrate various imaging modalities allows for a comprehensive assessment of the lung, identifying patterns associated with fibrosis. The findings from the dataset reveal a significant correlation between the model's output and expert annotations, highlighting its reliability as a clinical tool. Furthermore, the statistical analyses conducted on entropy values provide a new dimension for evaluating the complexity of pulmonary images, aiding in the early identification of fibrotic changes.

The high degree of agreement observed between expert annotations and algorithm-generated outcomes underscores the reliability and accuracy of the developed algo-

rithm for free fluid detection. Using density thresholds, active contour methods, and post-processing filters enabled effective identification and delineation of areas containing free fluid in CT images. This demonstrates the algorithm's potential as a valuable adjunct to clinical diagnosis and emphasizes the importance of adaptive thresholds considering pathology and anatomical location. The need for further validation with larger datasets remains crucial to ensure the algorithm's robustness across diverse clinical presentations.

In the classification of renal lithiasis, the study demonstrated that CNNs effectively distinguished between varying stone sizes, with an accuracy of 94% achieved through the Convex Hull augmentation technique. This highlights the efficacy of data augmentation in addressing sample size limitations and improving model performance. The superior results obtained using the Naive Bayes Multinomial statistical test point towards the effectiveness of probabilistic classifiers in this domain. However, the small and non-homogeneous sample sizes suggest that future studies should aim to incorporate more extensive and diverse datasets to enhance the generalizability of the findings.

5.3 MRI imaging

The application of deep learning techniques in classifying brain MRI images revealed promising results, particularly with the Support Vector Machine (SVM) classifier achieving high accuracy and precision. The use of transfer learning with a pre-trained model, such as VGG16, proved effective in leveraging existing knowledge to enhance the classification capabilities of the model. The distribution of tumor types and the imbalance in sample sizes between benign and malignant cases warrant further investigation, as these factors can influence model performance and bias. Future research should focus on increasing dataset size and diversity to strengthen the model's applicability across different clinical settings.

The innovative use of Shannon entropy for quantitatively detecting cerebral microbleeds represents a significant advancement in neuroimaging. The ability to es-

tablish a specific entropy range for early microbleed detection could aid in the timely diagnosis of neurodegenerative diseases. The strong correlation observed between entropy and Z-score values reinforces the method's reliability. However, further validation and refinement of this approach are necessary to address potential false positives and improve overall detection accuracy. Additionally, integrating machine learning algorithms with entropy analysis could enhance the detection capabilities and allow for more sophisticated assessments of brain abnormalities.

5.4 Modular model

The strength of the modular model lies in its structured approach to medical imaging, facilitating the integration of multiple imaging modalities, including CT and MRI. By employing a pre-trained network as a feature extractor, the model efficiently utilizes transfer learning to adapt to the unique characteristics of each imaging dataset. This modularity enhances the model's generalizability across various pathologies and allows for scalability in future studies involving larger datasets.

While the current study demonstrated promising results across various conditions, further validation with larger, more diverse datasets is essential. The incorporation of additional machine learning techniques, such as ensemble methods and unsupervised learning approaches, may enhance the model's performance, especially in challenging cases with complex presentations. Future research should also focus on refining the modular architecture to accommodate real-time processing capabilities, enabling its implementation in clinical settings.

The integration of AI in medical imaging not only improves diagnostic accuracy but also streamlines the workflow for radiologists, potentially reducing the time required for image interpretation. The methodologies developed in this study offer valuable tools for enhancing clinical decision-making and patient management. The automation of complex analysis tasks allows healthcare professionals to focus on higher-level clinical reasoning and patient care, ultimately leading to improved patient outcomes.

5.5 Published works

This section presents a compilation of our published articles that contribute to the body of knowledge surrounding computer-aided medical diagnosis and medical imaging. Each publication reflects a unique aspect of our research journey, highlighting the methodologies, results, and implications of our work in the context of medical image analysis. These studies support the findings presented in this thesis and offer insights into the challenges and advancements in the field. By sharing these works, we aim to demonstrate scientific research's collaborative and iterative nature and the practical applications of artificial intelligence and deep learning techniques in improving diagnostic accuracy and patient outcomes.

Table 5.1: Articles published from the thesis project in national and international journals with impact factor, indexed, DOI, and ISSN.

Biomedical Datasets	**Accuracy**	**Status**
Pneumonia	95%, 96%	Published: [143], [144], [145]
Tuberculosis	97%	Published: [146]
Post-COVID	97%	Published: [147]
Osteoporosis	77%	Accepted (in press)
Osteoarthritis	98%	In process
Pulmonary fibrosis	97%	Published: [148]
Free Fluid	85%	Accepted (in process)
Renal Lithiasis	94%	Accepted (in press)
Brain MRI	90%	Published: [149]
Microbleeds	94%	Accepted (in press)
AI in MRI	98%	Published: [150]

Chapter 6

Conclusions

This thesis has explored the application of deep learning techniques for computer-aided medical image diagnosis, presenting a modular approach for classification, detection, and segmentation tasks. The research demonstrates how convolutional blocks of pre-trained networks, coupled with various classifiers, can effectively identify and assess pathologies using biomedical images from multiple modalities, including X-ray, CT, and MRI.

One of the key achievements of this work has been the successful development and deployment of AI models that reach competitive levels of accuracy, precision, recall, and F1-score, as evidenced in the classification of diseases such as pneumonia, tuberculosis, osteoporosis, osteoarthritis, post-COVID conditions, and pulmonary fibrosis. Across these applications, the models provided robust and consistent performance, achieving accuracy rates up to 99%, which is crucial for enhancing diagnostic accuracy in clinical settings.

The use of modular models allowed for flexibility and adaptability in applying different classifiers and optimization techniques to medical imaging tasks. Notably, the convolutional neural networks (CNNs) integrated with classifiers like SVM and Naïve Bayes yielded superior results, particularly in CT imaging of renal lithiasis and brain MRI tumor detection. The Grad-CAM technique further enabled explainability in model decision-making, enhancing the interpretability of kidney stone detection models and other medical imaging tasks.

In addition to the primary focus of the thesis, the exploration of entropy-based methods for the early detection of cerebral microbleeds using Shannon entropy represents an innovative contribution to the field of neuroimaging. This method showcased how statistical and AI-driven approaches can complement traditional radiological practices.

Overall, the proposed modular model approach has proven to be a valuable tool for tackling challenges in medical image analysis, and it holds significant potential for improving diagnostic workflows, reducing interobserver variability, and increasing diagnostic precision.

In summary, the proposed modular model represents a significant advancement in Biomedical imaging, offering a comprehensive framework for detecting and analyzing diverse pathologies. Its ability to adapt to different imaging modalities and its reliance on robust statistical methodologies underscore its potential as a valuable tool for clinicians, ultimately leading to improved diagnostic accuracy and patient care. The findings from this study lay the groundwork for future explorations in artificial intelligence applications within medical imaging, contributing to the ongoing evolution of diagnostic methodologies in healthcare.

Despite the promising results achieved throughout this research, there are several areas for future investigation and improvement: expansion to new pathologies and imaging modalities, incorporation of more advanced techniques for model explainability, integration with clinical decision support systems, continued research on neuroimaging and microbleed detection, and extending the Modular Model for Multimodal Imaging.

This research contributes significantly to the ongoing efforts to advance AI-driven medical image analysis methods. By demonstrating the effectiveness of modular deep learning approaches, this thesis opens avenues for further innovation in the intersection of AI and healthcare. With continued exploration and refinement, AI has the potential to transform diagnostic processes, offering faster, more accurate, and more reliable assessments that can ultimately improve patient outcomes.

Despite the promising results achieved, this research is not without limitations.

The primary constraint lies in the generalizability of the proposed modular model, as it was evaluated using specific datasets from particular imaging modalities (X-ray, CT, MRI) and pathologies (e.g., pneumonia, tuberculosis, osteoarthritis). This may limit its applicability to other diseases or imaging techniques without further validation. Additionally, the reliance on pre-trained convolutional neural networks, while beneficial for efficiency, may introduce biases inherent to the original training datasets. Future studies should explore the application of the model to more diverse datasets, incorporate other imaging modalities such as ultrasound or PET scans, and investigate potential biases to enhance robustness and generalizability across broader clinical scenarios.

Appendix A

Supplementary research contributions

A.1 Diagnosing Parkinson's disease through drawing patterns

According to the World Health Organization (WHO), Parkinson's disease (PD) is a degenerative brain condition associated with motor symptoms as well as non-motor complications. In Mexico, PD usually begins between the ages of 50 and 65; the rate is estimated at 50 cases per 100,000 inhabitants. Artificial intelligence (AI), drawing and writing patterns characteristic of motor symptoms, can be a noninvasive method for diagnosing this condition. An automatic classification processing model is proposed for diagnosing the disease based on spiral drawings, whose results obtained 95% accuracy, while for wave drawings 89% accuracy was obtained. The reported model is competitive with the state-of-the-art and is a support tool for diagnosing PD.

A dataset provided by collaborators of the NIATS (Centre for Innovation and Technology Assessment in Health) of the Federal University of Uberlândia was used to carry out this study. This dataset was collected specifically to detect Parkinson's disease from images of spiral and wave drawings. The dataset is composed of two subsets: one with 102 spiral images and another with 102 wave images.

Image processing was performed using the Visual Geometry Group (VGG16) ar-

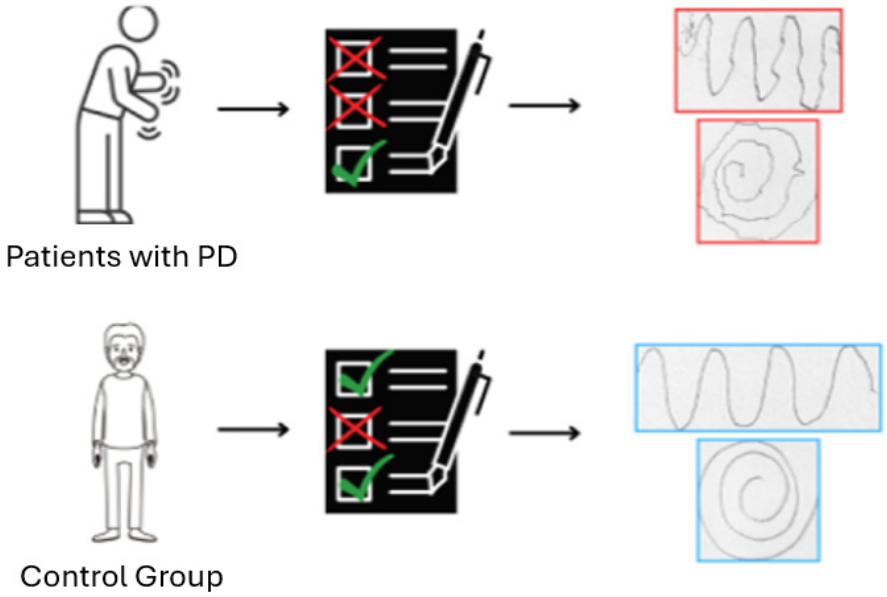


Figure A.1: Image set provided by NIATS.

chitecture, a well-established convolutional neural network widely used in computer vision tasks. VGG16 was used as a feature extractor, allowing us to obtain meaningful representations of the drawings for further classification. In Table A.1, we can see the results.

Table A.1: Values obtained from the proposed model.

Drawing	**Classification system**	**Classifier**	**Accuracy**
Spiral	TTS	SVM	0.95
Wave	CV	SVM	0.91

For spiral drawings, the model achieved 95% accuracy, demonstrating its ability to distinguish between healthy individuals and patients with Parkinson's disease based on characteristic drawing patterns. For wave drawings, a class sensitivity of 91% was obtained.

A.2 Ancient Purépecha petroglyphs

This research introduces an innovative approach to analyzing Purépecha petroglyphs through interdisciplinary collaboration between computer vision and archaeology. The research addresses the practical needs of archaeologists and scholars who require critical support for searching and recovering collections of ancient Purépecha images. The paper begins by providing an overview of the interdisciplinary approach to enhance the documentation, analysis, and preservation of Purépecha pictographic data. Subsequently, it objectively evaluates the contextual descriptor's performance in image retrieval tasks, utilizing a set of 14 images featuring Purépecha petroglyphs. In summary, the proposed approach shows promise by enhancing performance in retrieval tasks and can be validated from an epigraphic perspective, offering the potential for novel insights into archaeology and practical solutions for academia. The current study comprises 14 images of Purépecha petroglyphs captured in Tzintzuntzan, Michoacán (see Figure A.2).

Figure A.2: Purépecha Petroglyphs. A significant testament to the culturally rich heritage of this pre-Columbian civilization.

The accurate delineation of image contours is crucial, yet certain engravings show signs of erosion over time, while others exhibit less distinct shapes. Therefore, an image processing phase has been implemented to enhance these characteristics. The goal is to enable the Contextual Shape Descriptor (CSD) to identify edges and contours swiftly, precisely, and definitively. This procedure entails the application of two stages of digital filtering, aimed at enhancing the image's characteristics and detecting its contours. The initial stage involves the application of edge detection filters to amplify contours and enhance overall image quality. In the initial stage of our exper-

iments, an analysis was conducted to evaluate how well the outlines of the Purépecha petroglyphs are represented from a Mayan glyph in different samplings.

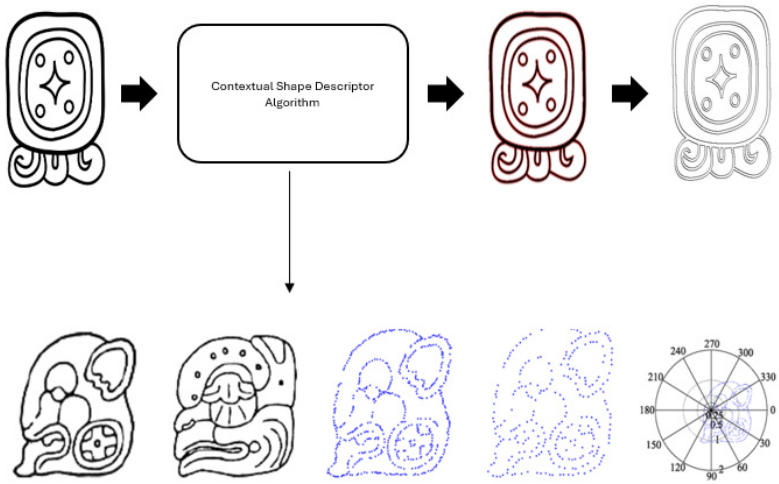


Figure A.3: Proposed Model showcasing the outcomes of the curve recovery process applied to the detected contours of Mayan glyphs. (From left to right) The original image undergoes the CSD algorithm, identifying the image contours; the resulting output image displays the contours previously detected by the algorithm, facilitating the recovery of the original image.

The model (Figure A.3) was employed on a series of images featuring Purépecha petroglyphs. By applying morphological filters to the original images, results reveal effects that are not discernible to the naked eye. In general, with all filters, the following results were obtained (Figure A.4):

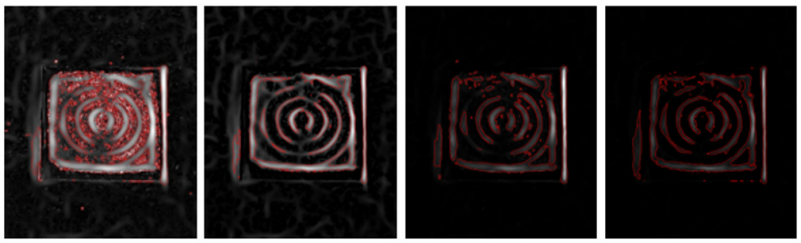


Figure A.4: Implementation of the model with DFC for contour detection in images of Purépecha petroglyphs. (First two) Contours were detected using the Meijering filter and the morphological dilation and erosion filters, respectively. (Last two) Contours were detected using the Frangi filter and the morphological dilation and erosion filters, respectively. The ability to detect contours in petroglyphs is a significant achievement, capturing details that cannot be seen with the naked eye.

A.3 Plasma for medical applications

Plasma Desorption Ionization (PDI) stands as a robust analytical technique facilitating identifying and quantifying a diverse array of compounds, including drugs, metabolites, and biomolecules within biological specimens. This study delves into exploring efficient and cost-effective ionization methodologies geared toward medical applications and introducing an innovative avenue known as Americium-241 Plasma Desorption Ionization (AmDI), leveraging the radioactive element Americium-241 found in smoke detectors, boasting a lengthy half-life of 432.2 years. Desorption ionization experiments were executed employing an ion chamber extracted from a smoke detector, accompanied by sample tablets (ibuprofen and paracetamol) positioned near a Tesla coil, serving as the plasma ionization source. The findings affirm the viability and efficacy of the AmDI approach in characterizing this specific category of samples. This development represents a significant stride in medical diagnostics, offering a pragmatic and budget-friendly resolution for comprehensive analysis. A disassembled ionization chamber from a smoke detector (Figure A.5), widely available online at low cost, was used as the basis for the Americium-241 Plasma Desorption Ionization (Source AmDI).

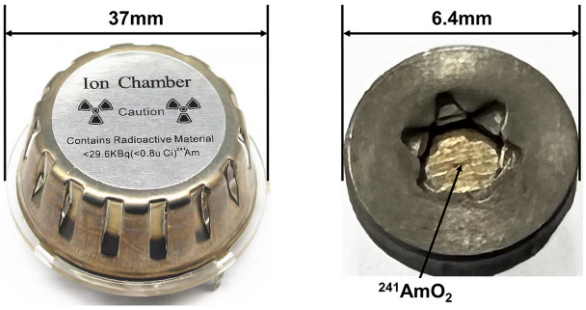


Figure A.5: Smoke detector covered in a gold sheet.

This research undertakes a practical and preliminary exploration, where desorption ionization assessments are conducted utilizing a smoke detector-extracted ion chamber and sample tablets (ibuprofen and paracetamol), strategically positioned near a Tesla coil functioning as the plasma ionization source. The results unearth the viability and efficacy of the AmDI technique, showcasing its prowess in charac-

terizing these specific types of samples. As a pioneering stride, this advancement contributes significantly to medical diagnostics, furnishing an accessible cost-effective solution for comprehensive analytical pursuits. As the demand for effective and economical ionization methods for medical applications escalates, this study delves into innovation, proposing a novel approach named Americium-241 Plasma Desorption Ionization (AmDI).

This study implemented a flyback as the main power source to achieve plasma generation. The americium desorption ionization (AmDI) process was carried out by the interaction of americium with the plasma arc generated by the flyback. By positioning the americium within the arc, ionization of the surrounding air molecules occurs, which causes the desorption of the ions from the surface and their subsequent ionization in the generated plasma. In Figure A.6 the assembly diagram used is shown:

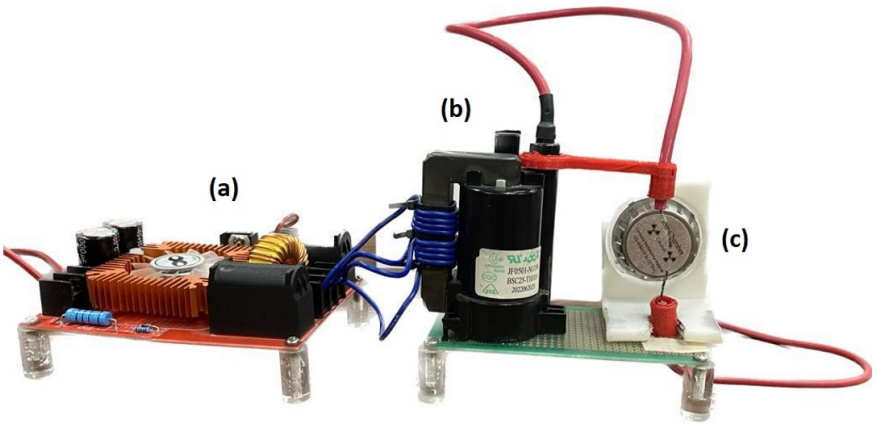


Figure A.6: Assembly diagram of the AmDI system. (a) Drive plate, (b) flyback, and (c) americium-241.

The results obtained after exposing ibuprofen and paracetamol to the plasma source with Americium-241 were analyzed using the Fourier Transform Infrared Spectroscopy (FTIR) technique (Figure A.7). This technique made it possible to compare the infrared spectra of the samples before and after contact with the plasma, providing valuable information on the composition and structure of the drugs. These results indicate that paracetamol did not undergo major structural changes due to exposure

to plasma generated by americium-241. These findings are relevant, as they support the safety of the AmDI method and suggest that this ionization technique does not affect the molecular structure of paracetamol during analysis. These observations are consistent with the known molecular structure of ibuprofen and allow confirmation of its identification using FTIR spectroscopy. Each of these absorption bands is characteristic and specific to the molecular vibrations of ibuprofen, providing additional evidence for the effectiveness of the AmDI method in accurately analyzing this drug. The results obtained after FTIR analysis of plasma-irradiated ibuprofen indicated no significant alteration in its molecular structure due to exposure to plasma generated by Americium-241. This information is valuable in supporting the reliability of the AmDI method in drug characterization and its potential application in various fields of research and clinical analysis.

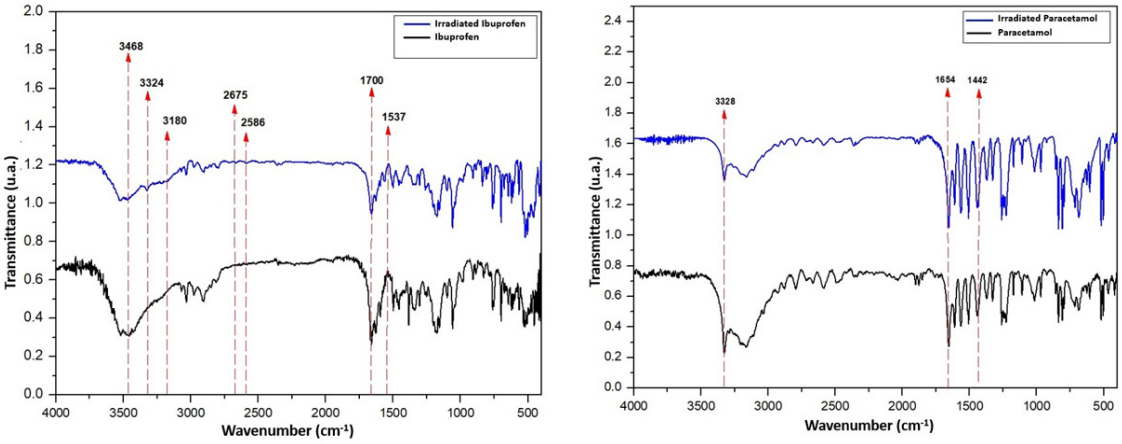


Figure A.7: FTIR of irradiated and normal ibuprofen and paracetamol, respectively.

In summary, the AmDI approach offers a promising outlook for future research and applications in the field of mass spectrometry and ambient ionization. The combination of various analytical techniques will allow further progress in the understanding and application of this method, fostering significant advances in medical diagnostics, materials studies, and other relevant scientific areas.

A.4 Electrical insulators in power transmission

This study presents a simple and efficient methodology for implementing a fault detection process in electrical insulators in transmission lines (Figure A.8). The methodology consists of creating a data set with faultless insulators, insulators with broken disks, and insulators with damage from discharges due to contamination, thus separating the insulators in good condition and the damaged insulators; as a second step, a preprocessing was carried out that allows the standardization of the size of the images.

Figure A.8: Electrical insulator in good condition (top), (bottom left) broken electrical insulator, and (bottom right) insulator damaged by electrical discharge defect due to contamination.

In this research, the VGG16 architecture was adopted to carry out the feature extraction. Two classification systems were used, where the first consists of creating a training and test set (70/30), and the second in cross-validation, both with the same evaluations and metrics. Two experiments were carried out. The first determined

which insulators are in good condition vs. damaged insulators, the second experiment dealt with the specific cases of insulators in good condition vs. insulators with broken disks, and finally, the case of insulators in good condition vs. insulators damaged by electrical discharges due to contamination. In terms of precision and accuracy, experiment 2 with the SVM classifier seems more successful and robust in classifying images of faulty electrical insulators.

The dataset used in this study focuses on detecting defects in electrical insulators in transmission lines. It consists of 1600 original images collected from various sources including Baidu, Google, and public datasets, and was obtained from the Ninja Dataset repository. To ensure a homogeneous and balanced set, 100 images were randomly selected from the three classes of insulators: normal, broken, and pollution-flashover.

The methodology consisted of resizing the images to a uniform resolution of 224x224 pixels since a VGG16-CNN was used as an extractor of image features. The prediction and evaluation of results are performed, using the trained classifier to predict the labels of the test set and calculate various evaluation metrics such as precision, accuracy, and confusion matrix.

Table A.2: Best results from combining experiments between electrical insulators.

Insulator Experiments	**Classification System**	**Classifier**	**Accuracy**
Good vs Damaged	TTS	SVM	0.93
	CV		0.97
Good vs Broken	TTS		0.98
	CV		0.99
Good vs Pollution-flashover	TTS		0.93
	CV		0.99

We have obtained up to 99% classification among the three groups of electrical insulators. We can thus conclude that the methodology we used in the project was successful in detecting faults in electrical insulators in power transmission lines. In fact, filters can be applied to obtain segmentations in the image and obtain a region of interest.

A.5 Published works

From the complementary research carried out during the PhD, the projects were published in national and international journals with impact factor, indexed, DOI, and ISSN.

Table A.3: Articles published, supplementary research.

Datasets	**Accuracy**	**Status**
Parkinson	95%	Published: [151]
Petroglyphs	85%	Published: [152], [153]
Plasma	safety	Accepted (in press)
Electrical Insulators	99%	Accepted (in press)

Bibliography

[1] Pranav Rajpurkar and Matthew P Lungren. The current and future state of ai interpretation of medical images. *New England Journal of Medicine*, 388(21):1981–1990, 2023.

[2] Isabella Castiglioni, Leonardo Rundo, Marina Codari, Giovanni Di Leo, Christian Salvatore, Matteo Interlenghi, Francesca Gallivanone, Andrea Cozzi, Natascha Claudia D'Amico, and Francesco Sardanelli. Ai applications to medical images: From machine learning to deep learning. *Physica medica*, 83:9–24, 2021.

[3] Daniel SW Ting, Yong Liu, Philippe Burlina, Xinxing Xu, Neil M Bressler, and Tien Y Wong. Ai for medical imaging goes deep. *Nature medicine*, 24(5):539–540, 2018.

[4] John C Gore. Artificial intelligence in medical imaging, 2020.

[5] Andreas S Panayides, Amir Amini, Nenad D Filipovic, Ashish Sharma, Sotirios A Tsaftaris, Alistair Young, David Foran, Nhan Do, Spyretta Golemati, Tahsin Kurc, et al. Ai in medical imaging informatics: current challenges and future directions. *IEEE journal of biomedical and health informatics*, 24(7):1837–1857, 2020.

[6] Shier Nee Saw and Kwan Hoong Ng. Current challenges of implementing artificial intelligence in medical imaging. *Physica Medica*, 100:12–17, 2022.

[7] Luciano M Prevedello, Safwan S Halabi, George Shih, Carol C Wu, Marc D Kohli, Falgun H Chokshi, Bradley J Erickson, Jayashree Kalpathy-Cramer, Katherine P Andriole, and Adam E Flanders. Challenges related to artificial intelligence research in medical imaging and the importance of image analysis competitions. *Radiology: Artificial Intelligence*, 1(1):e180031, 2019.

[8] Yasser Alginahi et al. Preprocessing techniques in character recognition. *Character recognition*, 1:1–19, 2010.

[9] UK Lopes and J. Francisco Valiati. Pre-trained convolutional neural networks as feature extractors for tuberculosis detection. *Computers in biology and medicine*, 89:135–143, 2017.

[10] Sotiris B Kotsiantis, Ioannis D Zaharakis, and Panayiotis E Pintelas. Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review*, 26:159–190, 2006.

[11] Badih Ghazi, Rina Panigrahy, and Joshua Wang. Recursive sketches for modular deep learning. In *International Conference on Machine Learning*, pages 2211–2220. PMLR, 2019.

[12] Igor Kononenko. Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in medicine*, 23(1):89–109, 2001.

[13] Harrison H Barrett. Objective assessment of image quality: effects of quantum noise and object variability. *JOSA A*, 7(7):1266–1278, 1990.

[14] Andreas Heindl, Sidra Nawaz, and Yinyin Yuan. Mapping spatial heterogeneity in the tumor microenvironment: a new era for digital pathology. *Laboratory investigation*, 95(4):377–384, 2015.

[15] Cindy S Lee, Paul G Nagy, Sallie J Weaver, and David E Newman-Toker. Cognitive and system factors contributing to diagnostic errors in radiology. *American Journal of Roentgenology*, 201(3):611–617, 2013.

[16] Wendy A Wells, Robert O Rainer, and Vincent A Memoli. Equipment, standardization, and applications of image processing. *American journal of clinical pathology*, 99(1):48–56, 1993.

[17] Krishna Juluru, Hao-Hsin Shih, Krishna Nand Keshava Murthy, Pierre Elnajjar, Amin El-Rowmeim, Christopher Roth, Brad Genereaux, Josef Fox, Eliot Siegel, and Daniel L Rubin. Integrating al algorithms into the clinical workflow. *Radiology: Artificial Intelligence*, 3(6):e210013, 2021.

[18] Michael van Bekkum, Maaike de Boer, Frank van Harmelen, André Meyer-Vitali, and Annette ten Teije. Modular design patterns for hybrid learning and reasoning systems: a taxonomy, patterns and use cases. *Applied Intelligence*, 51(9):6528–6546, 2021.

[19] David M Bradley. *Learning in modular systems*. Carnegie Mellon University, 2010.

[20] Kenji Suzuki. Overview of deep learning in medical imaging. *Radiological physics and technology*, 10(3):257–273, 2017.

[21] June-Goo Lee, Sanghoon Jun, Young-Won Cho, Hyunna Lee, Guk Bae Kim, Joon Beom Seo, and Namkug Kim. Deep learning in medical imaging: general overview. *Korean journal of radiology*, 18(4):570–584, 2017.

[22] Diah Anggraeni Pitaloka, Ajeng Wulandari, Tjan Basaruddin, and Dewi Yanti Liliana. Enhancing cnn with preprocessing stage in automatic emotion recognition. *Procedia computer science*, 116:523–529, 2017.

[23] Manjunath Jogin, MS Madhulika, GD Divya, RK Meghana, S Apoorva, et al. Feature extraction using convolution neural networks (cnn) and deep learning. In *2018 3rd IEEE international conference on recent trends in electronics, information & communication technology (RTEICT)*, pages 2319–2323. IEEE, 2018.

[24] Sotiris B Kotsiantis, Ioannis Zaharakis, P Pintelas, et al. Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, 160(1):3–24, 2007.

[25] Kunio Doi. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized medical imaging and graphics*, 31(4-5):198–211, 2007.

[26] Kunio Doi. Current status and future potential of computer-aided diagnosis in medical imaging. *The British journal of radiology*, 78(suppl_1):s3–s19, 2005.

[27] Shouvik Chakraborty and Kalyani Mali. An overview of biomedical image analysis from the deep learning perspective. *Applications of advanced machine intelligence in computer vision and object recognition: emerging research and opportunities*, pages 197–218, 2020.

[28] James S Duncan, Michael F Insana, and Nicholas Ayache. Biomedical imaging and analysis in the age of big data and deep learning [scanning the issue]. *Proceedings of the IEEE*, 108(1):3–10, 2019.

[29] Peter Savadjiev, Jaron Chong, Anthony Dohan, Maria Vakalopoulou, Caroline Reinhold, Nikos Paragios, and Benoit Gallix. Demystification of ai-driven medical image interpretation: past, present and future. *European radiology*, 29:1616–1624, 2019.

[30] Suresh Dodda, Sathishkumar Chintala, Satyanarayan Kanungo, Tolu Adedoja, Sourabh Sharma, et al. Exploring ai-driven innovations in image communication systems for enhanced medical imaging applications. *Journal of Electrical Systems*, 20(3s):949–959, 2024.

[31] Peter Szolovits, Ramesh S Patil, and William B Schwartz. Artificial intelligence in medical diagnosis. *Annals of internal medicine*, 108(1):80–87, 1988.

[32] Mugahed A Al-Antari. Artificial intelligence for medical diagnostics—existing and future ai technology!, 2023.

[33] Pavel Hamet and Johanne Tremblay. Artificial intelligence in medicine. *metabolism*, 69:S36–S40, 2017.

[34] Emefa Surprize Deborah Buaka and Md Zubab Ibne Moid. Ai and medical imaging technology: evolution, impacts, and economic insights. *The Journal of Technology Transfer*, pages 1–13, 2024.

[35] Manisha Singh, Gurubasavaraj Veeranna Pujar, Sethu Arun Kumar, Meduri Bhagyalalitha, Handattu Shankaranarayana Akshatha, Belal Abuhaija, Anas Ratib Alsoud, Laith Abualigah, Narasimha M Beeraka, and Amir H Gandomi. Evolution of machine learning in tuberculosis diagnosis: a review of deep learning-based medical applications. *Electronics*, 11(17):2634, 2022.

[36] Yunqiang Zhu Jia Song, Shaohua Gao and Chenyan Ma. A survey of remote sensing image classification based on cnns. *Big Earth Data*, 3(3):232–254, 2019.

[37] Pin Wang, En Fan, and Peng Wang. Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. *Pattern recognition letters*, 141:61–67, 2021.

[38] Aykut Çayir, Işil Yenidoğan, and Hasan Dağ. Feature extraction based on deep learning for some traditional machine learning methods. In *2018 3rd International conference on computer science and engineering (UBMK)*, pages 494–497. IEEE, 2018.

[39] Zewen Li, Fan Liu, Wenjie Yang, Shouheng Peng, and Jun Zhou. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*, 33(12):6999–7019, 2021.

[40] Stephen Grossberg. Recurrent neural networks. *Scholarpedia*, 8(2):1888, 2013.

[41] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

[42] Junko Yano and Vittal K Yachandra. X-ray absorption spectroscopy. *Photosynthesis research*, 102:241–254, 2009.

[43] Philip J Withers, Charles Bouman, Simone Carmignato, Veerle Cnudde, David Grimaldi, Charlotte K Hagen, Eric Maire, Marena Manley, Anton Du Plessis, and Stuart R Stock. X-ray computed tomography. *Nature Reviews Methods Primers*, 1(1):18, 2021.

[44] Girish Katti, Syeda Arshiya Ara, and Ayesha Shireen. Magnetic resonance imaging (mri)–a review. *International journal of dental clinics*, 3(1):65–70, 2011.

[45] Bifa Fan. Limitations and ethical implications of artificial intelligence. In *Artificial Intelligence in Anesthesiology*, pages 109–113. Springer, 2023.

[46] Vijaya Lakshmi Pavani Molli. Ethical considerations in ai-assisted diagnosis: Balancing privacy, accuracy, and patient autonomy. *International Journal of Machine Learning and Artificial Intelligence*, 2(2):1–10, 2021.

[47] Jonathan Herington, Melissa D McCradden, Kathleen Creel, Ronald Boellaard, Elizabeth C Jones, Abhinav K Jha, Arman Rahmim, Peter JH Scott, John J Sunderland, Richard L Wahl, et al. Ethical considerations for artificial intelligence in medical imaging: deployment and governance. *Journal of Nuclear Medicine*, 64(10):1509–1515, 2023.

[48] Plamen P Angelov, Eduardo A Soares, Richard Jiang, Nicholas I Arnold, and Peter M Atkinson. Explainable artificial intelligence: an analytical review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 11(5):e1424, 2021.

[49] Xu Han, Zhengyan Zhang, Ning Ding, Yuxian Gu, Xiao Liu, Yuqi Huo, Jiezhong Qiu, Yuan Yao, Ao Zhang, Liang Zhang, et al. Pre-trained models: Past, present and future. *AI Open*, 2:225–250, 2021.

[50] Anne Sakdinawat and David Attwood. Nanoscale x-ray imaging. *Nature photonics*, 4(12):840–848, 2010.

[51] Jie Wang, Tuo Xu, Li Zhang, Tianying Chang, Jin Zhang, Shihan Yan, and Hong-Liang Cui. Nondestructive damage evaluation of composites based on terahertz and x-ray image fusion. *Ndt & E International*, 127:102616, 2022.

[52] RF Mould. Röntgen and the discovery of x-rays. *The British journal of radiology*, 68(815):1145–1176, 1995.

[53] Arati S Panchbhai. Wilhelm conrad röntgen and the discovery of x-rays: Revisited after centennial. *Journal of Indian academy of oral medicine and radiology*, 27(1):90–95, 2015.

[54] Gottfried Landwehr, Axel Haase, and Eberhard Umbach. *Rontgen Centennial-X-rays Today In Natural And Life Sicences*. World Scientific, 1997.

[55] Gerald J Hine and Gordon L Brownell. *Radiation dosimetry*. Elsevier, 2013.

[56] Ervin B Podgoršak et al. *Radiation physics for medical physicists*, volume 1. Springer, 2006.

[57] Stephen Klassen. The photoelectric effect: Reconstructing the story for the physics classroom. *Science & Education*, 20:719–731, 2011.

[58] Malcolm Cooper, Peter Mijnarends, Nobuhiro Shiotani, Nobuhiko Sakai, and Arun Bansil. *X-ray Compton scattering*, volume 5. OUP Oxford, 2004.

[59] Andrew T Young. Rayleigh scattering. *Applied optics*, 20(4):533–535, 1981.

[60] Henry N Chapman. X-ray imaging beyond the limits. *Nature materials*, 8(4):299–301, 2009.

[61] Thorsten M Buzug. Computed tomography. In *Springer handbook of medical technology*, pages 311–342. Springer, 2011.

[62] Euclid Seeram. Computed tomography: physical principles and recent technical advances. *Journal of Medical Imaging and Radiation Sciences*, 41(2):87–109, 2010.

[63] Allan M Cormack and Godfrey N Hounsfield. The nobel prize in physiology or medicine 1979. *Nobelprize. org*, 9, 2011.

[64] Adam M Alessio, Paul E Kinahan, Phillip M Cheng, Hubert Vesselle, and Joel S Karp. Pet/ct scanner instrumentation, challenges, and solutions. *Radiologic Clinics*, 42(6):1017–1032, 2004.

[65] David W Townsend and Thomas Beyer. A combined pet/ct scanner: the path to true image fusion. *The British journal of radiology*, 75(suppl_9):S24–S30, 2002.

[66] Vibhu Kapoor, Barry M McCook, and Frank S Torok. An introduction to pet-ct imaging. *Radiographics*, 24(2):523–543, 2004.

[67] Martin J Willemink and Peter B Noël. The evolution of image reconstruction for ct—from filtered back projection to artificial intelligence. *European radiology*, 29:2185–2195, 2019.

[68] Uwe Schneider, Eros Pedroni, and Antony Lomax. The calibration of ct hounsfield units for radiotherapy treatment planning. *Physics in Medicine & Biology*, 41(1):111, 1996.

[69] Michel M Ter-Pogossian. Basic principles of computed axial tomography. In *Seminars in nuclear medicine*, pages 109–127. Elsevier, 1977.

[70] Poonam Batra, Kathleen Brown, Richard J Steckel, James D Collins, Carl O Ovenfors, and Denise Aberle. Mr imaging of the thorax: a comparison of axial, coronal, and sagittal imaging planes. *Journal of computer assisted tomography*, 12(1):75–81, 1988.

[71] Inês Domingues, Gisèle Pereira, Pedro Martins, Hugo Duarte, João Santos, and Pedro Henriques Abreu. Using deep learning techniques in medical imaging: a systematic review of applications on ct and pet. *Artificial Intelligence Review*, 53:4093–4160, 2020.

[72] CM McLeavy, MH Chunara, RJ Gravell, A Rauf, A Cushnie, C Staley Talbot, and RM Hawkins. The future of ct: deep learning reconstruction. *Clinical radiology*, 76(6):407–415, 2021.

[73] Girish Katti, Syeda Arshiya Ara, and Ayesha Shireen. Magnetic resonance imaging (mri)–a review. *International journal of dental clinics*, 3(1):65–70, 2011.

[74] Donald B Plewes and Walter Kucharczyk. Physics of mri: a primer. *Journal of magnetic resonance imaging*, 35(5):1038–1054, 2012.

[75] Terry O Woods. Standards for medical devices in mri: present and future. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 26(5):1186–1189, 2007.

[76] LD Hall. Nuclear magnetic resonance. *Advances in carbohydrate chemistry*, 19:51–93, 1964.

[77] Chris Boesch. Nobel prizes for nuclear magnetic resonance: 2003 and historical perspectives. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 20(2):177–179, 2004.

[78] Peter J Hore. *Nuclear magnetic resonance*. Oxford University Press, USA, 2015.

[79] Anne J Kox. The discovery of the electron: Ii. the zeeman effect. *European journal of physics*, 18(3):139, 1997.

[80] Vladimír Mlynárik. Introduction to nuclear magnetic resonance. *Analytical Biochemistry*, 529:4–9, 2017.

[81] Ray Freeman. *Handbook of nuclear magnetic resonance*. John Wiley and Sons, New York, NY, 1987.

[82] Robin Kingsley Harris. *Nuclear magnetic resonance spectroscopy*. John Wiley and Sons Inc., New York, NY, 1986.

[83] Valerii E Ogluzdin. The role of bohr frequencies in the scattering, luminescence, and generation of radiation in different media. *Physics-Uspekhi*, 49(4):401, 2006.

[84] Marcos Wolf, Anneloes de Boer, Kanishka Sharma, Peter Boor, Tim Leiner, Gere Sunder-Plassmann, Ewald Moser, Anna Caroli, and Neil Peter Jerome. Magnetic resonance imaging t1-and t2-mapping to assess renal structure and function: a systematic review and statement paper. *Nephrology Dialysis Transplantation*, 33(suppl_2):ii41–ii50, 2018.

[85] JS Petersson, J-O Christoffersson, and K Golman. Mri simulation using the k-space formalism. *Magnetic Resonance Imaging*, 11(4):557–568, 1993.

[86] Paul Peter Urone and Roger Hinrichs. 23.2 faraday's law of induction: Lenz's law. *College Physics*, 2016.

[87] Michael Loecher and Oliver Wieben. k-space. *Basic Principles of Cardiovascular MRI: Physics and Imaging Technique*, pages 13–23, 2015.

[88] JS Petersson and J-O Christoffersson. Multidimensional k-space model for analysis of flow-related phenomena in mr imaging. *International journal of imaging systems and technology*, 10(2):115–127, 1999.

[89] Peter Börnert and Bernd Aldefeld. On spatially selective rf excitation and its analogy with spiral mr image acquisition. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 7:166–178, 1998.

[90] Jin Liu, Yi Pan, Min Li, Ziyue Chen, Lu Tang, Chengqian Lu, and Jianxin Wang. Applications of deep learning to mri images: A survey. *Big Data Mining and Analytics*, 1(1):1–18, 2018.

[91] Alexander Selvikvåg Lundervold and Arvid Lundervold. An overview of deep learning in medical imaging focusing on mri. *Zeitschrift für Medizinische Physik*, 29(2):102–127, 2019.

[92] Amitha Mathew, P Amudha, and S Sivakumari. Deep learning techniques: an overview. *Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2020*, pages 599–608, 2021.

[93] Sushama Tanwar and Jeevanandam Jotheeswaran. Survey on deep learning for medical imaging. *Journal of Applied Science and Computations*, 5:1608–1620, 2018.

[94] Shervin Minaee, Yuri Boykov, Fatih Porikli, Antonio Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. Image segmentation using deep learning: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(7):3523–3542, 2021.

[95] Hee E Kim, Alejandro Cosa-Linan, Nandhini Santhanam, Mahboubeh Jannesari, Mate E Maros, and Thomas Ganslandt. Transfer learning for medical image classification: a literature review. *BMC medical imaging*, 22(1):69, 2022.

[96] Federico Di Mattia, Paolo Galeone, Michele De Simoni, and Emanuele Ghelfi. A survey on gans for anomaly detection. *arXiv preprint arXiv:1906.11632*, 2019.

[97] Roberto Confalonieri, Ludovik Coba, Benedikt Wagner, and Tarek R Besold. A historical perspective of explainable artificial intelligence. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 11(1):e1391, 2021.

[98] Maithra Raghu, Chiyuan Zhang, Jon Kleinberg, and Samy Bengio. Transfusion: Understanding transfer learning for medical imaging. *Advances in neural information processing systems*, 32, 2019.

[99] Padmavathi Kora, Chui Ping Ooi, Oliver Faust, U Raghavendra, Anjan Gudigar, Wai Yee Chan, K Meenakshi, K Swaraja, Pawel Plawiak, and U Rajendra

Acharya. Transfer learning techniques for medical image analysis: A review. *Biocybernetics and Biomedical Engineering*, 42(1):79–107, 2022.

[100] Pedro Marcelino. Transfer learning from pre-trained models. *Towards data science*, 10(330):23, 2018.

[101] Souleyman Chaib, Hongxun Yao, Yanfeng Gu, and Moussa Amrani. Deep feature extraction and combination for remote sensing image classification based on pre-trained cnn models. In *Ninth international conference on digital image processing (ICDIP 2017)*, volume 10420, pages 712–716. SPIE, 2017.

[102] Benjamin Recht, Rebecca Roelofs, Ludwig Schmidt, and Vaishaal Shankar. Do imagenet classifiers generalize to imagenet? In *International conference on machine learning*, pages 5389–5400. PMLR, 2019.

[103] Ross W Filice, Anouk Stein, Carol C Wu, Veronica A Arteaga, Stephen Borstelmann, Ramya Gaddikeri, Maya Galperin-Aizenberg, Ritu R Gill, Myrna C Godoy, Stephen B Hobbs, et al. Crowdsourcing pneumothorax annotations using machine learning annotations on the nih chest x-ray dataset. *Journal of digital imaging*, 33:490–496, 2020.

[104] Pádraig Cunningham, Matthieu Cord, and Sarah Jane Delany. Supervised learning. In *Machine learning techniques for multimedia: case studies on organization and retrieval*, pages 21–49. Springer, 2008.

[105] Trevor Hastie, Robert Tibshirani, Jerome Friedman, Trevor Hastie, Robert Tibshirani, and Jerome Friedman. Overview of supervised learning. *The elements of statistical learning: Data mining, inference, and prediction*, pages 9–41, 2009.

[106] Xiaogang Su, Xin Yan, and Chih-Ling Tsai. Linear regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(3):275–294, 2012.

[107] Douglas C Montgomery, Elizabeth A Peck, and G Geoffrey Vining. *Introduction to linear regression analysis*. John Wiley & Sons, 2021.

[108] Weifeng Liu, Puskal P Pokharel, and Jose C Principe. The kernel least-mean-square algorithm. *IEEE Transactions on signal processing*, 56(2):543–554, 2008.

[109] Hervé Abdi, Dominique Valentin, Betty Edelman, and Alice J O'Toole. A widrow–hoff learning rule for a generalization of the linear auto-associator. *Journal of Mathematical Psychology*, 40(2):175–182, 1996.

[110] Shun-ichi Amari. Backpropagation and stochastic gradient descent method. *Neurocomputing*, 5(4-5):185–196, 1993.

[111] Paul S Dwyer and MS Macphail. Symbolic matrix derivatives. *The annals of mathematical statistics*, pages 517–534, 1948.

[112] Paul S Dwyer. Some applications of matrix derivatives in multivariate analysis. *Journal of the American Statistical Association*, 62(318):607–625, 1967.

[113] Guthrie Miller. *Probabilistic interpretation of data*. Lulu. com, 2013.

[114] Richard S Zemel, Peter Dayan, and Alexandre Pouget. Probabilistic interpretation of population codes. *Neural computation*, 10(2):403–430, 1998.

[115] John S Bridle. Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition. In *Neurocomputing: Algorithms, architectures and applications*, pages 227–236. Springer, 1990.

[116] Michael P LaValley. Logistic regression. *Circulation*, 117(18):2395–2399, 2008.

[117] Xinyou Yin, JAN Goudriaan, Egbert A Lantinga, JAN Vos, and Huub J Spiertz. A flexible sigmoid function of determinate growth. *Annals of botany*, 91(3):361–371, 2003.

[118] Stephen I Gallant et al. Perceptron-based learning algorithms. *IEEE Transactions on neural networks*, 1(2):179–191, 1990.

[119] Yu-chen Wu and Jun-wen Feng. Development and application of artificial neural network. *Wireless Personal Communications*, 102:1645–1656, 2018.

[120] Thomas M McKenna, Joel L Davis, and Steven F Zornetzer. *Single neuron computation*. Academic Press, 2014.

[121] Andrea Soltoggio, Kenneth O Stanley, and Sebastian Risi. Born to learn: the inspiration, progress, and future of evolved plastic artificial neural networks. *Neural Networks*, 108:48–67, 2018.

[122] SH Shabbeer Basha, Shiv Ram Dubey, Viswanath Pulabaigari, and Snehasis Mukherjee. Impact of fully connected layers on performance of convolutional neural networks for image classification. *Neurocomputing*, 378:112–119, 2020.

[123] Yuchen Zhang, Jason Lee, Martin Wainwright, and Michael I Jordan. On the learnability of fully-connected neural networks. In *Artificial Intelligence and Statistics*, pages 83–91. PMLR, 2017.

[124] Yuhan Bai. Relu-function and derived function review. In *SHS Web of Conferences*, volume 144, page 02006. EDP Sciences, 2022.

[125] Marius-Constantin Popescu, Valentina E Balas, Liliana Perescu-Popescu, and Nikos Mastorakis. Multilayer perceptron and neural networks. *WSEAS Transactions on Circuits and Systems*, 8(7):579–588, 2009.

[126] Serhat Kiliçarslan and Mete Celik. Rsigelu: A nonlinear activation function for deep neural networks. *Expert Systems with Applications*, 174:114805, 2021.

[127] K O'Shea. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015.

[128] Rikiya Yamashita, Mizuho Nishio, Richard Kinh Gian Do, and Kaori Togashi. Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, 9:611–629, 2018.

[129] Poonam Zham, Dinesh K Kumar, Peter Dabnichki, Sridhar Poosapadi Arjunan, and Sanjay Raghav. Distinguishing different stages of parkinson's disease using

composite index of speed and pen-pressure of sketching a spiral. *Frontiers in neurology*, 8:435, 2017.

[130] Satrajit Chakrabarty, Aristeidis Sotiras, Mikhail Milchenko, Pamela LaMontagne, Michael Hileman, and Daniel Marcus. Mri-based identification and classification of major intracranial tumor types by using a 3d convolutional neural network: a retrospective multi-institutional analysis. *Radiology: Artificial Intelligence*, 3(5):e200301, 2021.

[131] Stefan Jaeger, Alexandros Karargyris, Sema Candemir, Jenifer Siegelman, Les Folio, Sameer Antani, and George Thoma. Automatic screening for tuberculosis in chest radiographs: a survey. *Quantitative imaging in medicine and surgery*, 3(2):89, 2013.

[132] Daniel S Kermany, Michael Goldbaum, Wenjia Cai, Carolina CS Valentim, Huiying Liang, Sally L Baxter, Alex McKeown, Ge Yang, Xiaokang Wu, Fangbing Yan, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *cell*, 172(5):1122–1131, 2018.

[133] Mohamed Najim. *Digital filters design for signal and image processing*. John Wiley & Sons, 2013.

[134] Sheldon Mascarenhas and Mukul Agarwal. A comparison between vgg16, vgg19 and resnet50 architecture frameworks for image classification. In *2021 International conference on disruptive technologies for multi-disciplinary research and applications (CENTCON)*, volume 1, pages 96–99. IEEE, 2021.

[135] Afshin Gholamy, Vladik Kreinovich, and Olga Kosheleva. Why 70/30 or 80/20 relation between training and testing sets: A pedagogical explanation. *Int. J. Intell. Technol. Appl. Stat*, 11(2):105–111, 2018.

[136] Mervyn Stone. Cross-validation: A review. *Statistics: A Journal of Theoretical and Applied Statistics*, 9(1):127–139, 1978.

[137] William S Noble. What is a support vector machine? *Nature biotechnology*, 24(12):1565–1567, 2006.

[138] David W Hosmer Jr, Stanley Lemeshow, and Rodney X Sturdivant. *Applied logistic regression*. John Wiley & Sons, 2013.

[139] Kashvi Taunk, Sanjukta De, Srishti Verma, and Aleena Swetapadma. A brief review of nearest neighbor algorithm for learning and classification. In *2019 international conference on intelligent computing and control systems (ICCS)*, pages 1255–1260. IEEE, 2019.

[140] Eguturi Manjith Kumar Reddy, Akash Gurrala, Vasireddy Bindu Hasitha, and Korupalli V Rajesh Kumar. Introduction to naive bayes and a review on its subtypes with applications. *Bayesian reasoning and gaussian processes for machine learning applications*, pages 1–14, 2022.

[141] Mohammadreza Heydarian, Thomas E Doyle, and Reza Samavi. Mlcm: Multi-label confusion matrix. *IEEE Access*, 10:19083–19095, 2022.

[142] Margherita Grandini, Enrico Bagli, and Giorgio Visani. Metrics for multi-class classification: an overview. *arXiv preprint arXiv:2008.05756*, 2020.

[143] R. Guzman-Cabrera Aron Hernandez-Trinidad, T. Cordova Fraga. Clasificacion automatica de radiografias de torax usando inteligencia artificial. *Identidad Energetica*, 5:27–33, 2022.

[144] Aron Hernandez-Trinidad, Eduardo Perez Careta, Angelica Hernandez Rayas, Teodoro Cordova-Fraga, and Rafael Guzman-Cabrera. Improved medical diagnosis of chest x-rays using deep learning with incremental iterations. *DYNA*, 97(5):522–527, 2022.

[145] Aron Hernandez-Trinidad, Huetzin A Perez-Olivas, Blanca O Murillo-Ortiz, Rafael Guzman-Cabrera, and Teodoro Cordova-Fraga. Deep learning-based approach to classify pneumonia on chest radiographs. *Preprint: Research Square*, 2023.

[146] Omar Villalpando-Vargas, Aron Hernández-Trinidad, Blanca Olivia Murillo-Ortiz, Rafael Guzmán-Cabrera, Luis Carlos Padierna-Garcıa, and Teodoro Córdova-Fraga. Automatic processing of chest x-rays to identify tuberculosis. *Research in Computing Science*, page 13, 2024.

[147] Arón Hernández Trinidad, Teodoro Córdova Fraga, Luis Carlos Padierna García, José Luis López Hernández, Blanca Olivia Murillo Ortiz, and Rafael Guzman-Cabrera. Automatic image processing to identify post-covid conditions by using deep learning. *Revista Mexicana de Física*, 69(6 Nov-Dec):061101–1, 2023.

[148] Aron Hernandez-Trinidad, Rafael Guzman-Cabrera, José Ruiz Pinales, and Teodoro Cordova-Fraga. Classification of long covid pulmonary fibrosis based on computed tomography in mexican patients. *International Journal of Combinatorial Optimization Problems & Informatics*, 15(2), 2024.

[149] Teodoro Cordova-Fraga Omar Villalpando-Vargas, Aron Hernandez-Trinidad and Rafael Guzman-Cabrera. Brain mri images for brain tumor detection using deep learning. *International Journal of Applied Mathematics and Machine Learning*, 17(2):83–97, 2023.

[150] Aron Hernandez-Trinidad, Blanca Olivia Murillo-Ortiz, Rafael Guzman-Cabrera, and Teodoro Cordova-Fraga. Applications of artificial intelligence in the classification of magnetic resonance images: Advances and perspectives. *New Advances in Magnetic Resonance Imaging*, 2023.

[151] Teodoro Cordova-Fraga Omar Villalpando-Vargas, Aron Hernandez-Trinidad and Rafael Guzman-Cabrera. Uso de la inteligencia artificial para el diagnostico de la enfermedad de parkinson a traves de patrones de dibujo. *Identidad Energetica*, 6:103–107, 2022.

[152] Aron Hernandez-Trinidad, Armando Perez-Crespo, Ricardo Espinosa-Thirion, Rafael Guzman-Cabrera, and Teodoro Cordova-Fraga. Análisis de colec-

ciones de petroglifos purépechas antiguos con descriptor de forma contextual. *JÓVENES EN LA CIENCIA*, 25:1–8, 2024.

[153] Aron Hernandez-Trinidad, Armando Perez-Crespo, Rafael Guzman-Cabrera, and Teodoro Cordova-Fraga. Ancient purépecha petroglyphs analysis using contextual shape descriptor. *Journal of Scientific and Technical Applications*, 9(24):7–12, 2023.

UNIVERSIDAD DE GUANAJUATO

Carta de Aceptación de TESIS
León, GTO, 28 de Octubre de 2024

Dr. Modesto Antonio Sosa Aquino
Director de la División de Ciencias e Ingenierías
Universidad de Guanajuato campus León

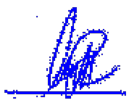
Estimado Dr. Sosa:

Por este medio le informo que el trabajo de tesis de doctorado con título: ***AI-Driven Deep Learning Models for Enhanced Medical Image Diagnosis***, me fue compartida por el MCA ***Arón Hernández Trinidad***

En mi opinión, el trabajo está completo, con el nivel de una tesis de doctorado en Física, además he de enfatizar que el MCA Arón Hernández Trinidad, de quien participe además en su comité de seguimiento académico, siempre mostró un gran dominio del tema y una amplia disponibilidad para atender, en tiempo y forma, las observaciones y sugerencias realizadas, por lo cual sugiero se le autorice su defensa a finde poder obtener el grado de Doctor en Física

Sin más por el momento, me despido enviando un cordial y afectuoso saludo.

ATENTAMENTE
"LA VERDAD OS HARÁ LIBRES"



Dra. Angélica Hernández Rayas
Técnico en Investigación D
SNI nivel I

Carta de Aceptación de TESIS
León, GTO, 31 de octubre de 2024

Dr. Modesto Antonio Sosa Aquino
Director de la División de Ciencias e Ingenierías
Universidad de Guanajuato campus León

Estimado Dr. Sosa:

Por este medio le informo que el trabajo de tesis de doctorado con título: "*AI-Driven Deep Learning Models for Enhanced Medical Image Diagnosis*" me fue entregado por el Fís. *Arón Hernández Trinidad*.

En mi opinión, el trabajo está completo, con el nivel de una tesis de Doctorado en Física, además he de enfatizar que el Fís. Hernández Trinidad, de quien participe además en su Comité de Seguimiento Académico, siempre mostró un gran dominio del tema y una amplia disponibilidad para atender, en tiempo y forma, las observaciones y sugerencias realizadas, asimismo hemos trabajado en proyectos de investigaciones que se han publicado en diferentes revistas JCR, por lo cual sugiero se le autorice su defensa a fin de poder obtener el grado de Doctor en Física

Sin más por el momento, me despido enviando un cordial y afectuoso saludo.

ATENTAMENTE
"LA VERDAD OS HARÁ LIBRES"

Dra. Blanca Olivia Murillo Ortiz
Doctora en Ciencias Médicas
Unidad de Alta Especialidad UMAE No.1

UNIVERSIDAD DE GUANAJUATO

Carta de Aceptación de TESIS
León, GTO, 30 de octubre de 2024

Dr. Modesto Antonio Sosa Aquino
Director de la División de Ciencias e Ingenierías
Universidad de Guanajuato campus León

Estimado Dr. Sosa:

Por este medio le informo que el trabajo de tesis de doctorado con título: **AI-Driven Deep Learning Models for Enhanced Medical Image Diagnosis** me fue entregado por el **MCA. Arón Hernández Trinidad.**

En mi opinión, el trabajo está completo, con el nivel de una tesis de doctorado en Física, además he de enfatizar que he trabajado con el MCA Hernández Trinidad, en proyectos de investigación sobre Petroglifos Purépechas, y hemos publicado un par de artículos sobre procesamiento digital de estos. Siempre mostró un gran dominio del tema y una amplia disponibilidad para atender, en tiempo y forma, las observaciones y sugerencias realizadas, por lo cual sugiero se le autorice su defensa a finde poder obtener el grado de Doctor en Física

Sin más por el momento, me despido enviando un cordial y afectuoso saludo.

ATENTAMENTE
"LA VERDAD OS HARÁ LIBRES"

Dr. José Armando Pérez Crespo
Profesor Titular
armando.perez@ugto.mx

THESIS Acceptance Letter
León, GTO, October 30th, 2024

Dr. Modesto Antonio Sosa Aquino
Director of the Division of Sciences and Engineering
University of Guanajuato, Leon Campus

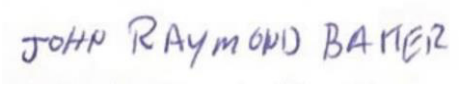
Dear, Dr. Sosa:

I hereby inform you that the doctoral thesis entitled: **AI-Driven Deep Learning models for Enhanced Medical Image Diagnosis**, was shared with me by **MCA Aron Hernandez Trinidad**.

In my opinion, the work is complete, at the level of a PhD thesis in Physics, for which reason I suggest that he be authorized to defend it to obtain the degree of Doctor in Physics.

Without further ado, I say goodbye with my warmest and most affectionate greetings.

SINCERELY
"LA VERDAD OS HARÁ LIBRE"



John R. Baker, PhD
Shinawatra University
Bangkok, Thailand
drjohnrbaker@yahoo.com

UNIVERSIDAD DE
GUANAJUATO

Carta de Aceptación de TESIS
León, GTO, 01 de noviembre de 2024

Dr. Carlos Herman Wiechers Medina
Coordinador del Posgrado en Física
Universidad de Guanajuato campus León

Estimado Dr. Wiechers.

Por este medio le informo que el trabajo de tesis de doctorado con título: "**AI-Driven Deep Learning Models for Enhanced Medical Image Diagnosis**", me fue compartido por el **MCA. Arón Hernández Trinidad**.

En mi opinión, el trabajo está completo, con el nivel de una tesis de doctorado en Física, además he de enfatizar que el MCA Hernández Trinidad, de quien participe además en su Comité de Seguimiento Académico, siempre mostró un gran dominio del tema y una amplia disponibilidad para atender, en tiempo y forma, las observaciones y sugerencias realizadas, por lo cual sugiero se le autorice su defensa a fin de poder obtener el grado de Doctor en Física.

Sin más por el momento, me despido enviando un cordial y afectuoso saludo.

ATENTAMENTE
"LA VERDAD OS HARÁ LIBRE"

Dr. Modesto Antonio Sosa Aquino
Profesor Titular C
Profesor con Perfil Deseable
SNI Nivel III

División de Ciencias e Ingenierías, campus León
Loma del Bosque 103, Lomas del Campestre, 37150 León, GTO, MX.
Tel. 477 788-5100 ext. 8415 modesto@fisica.ugto.mx, http://www.ifug.ugto.mx/

Asunto: Carta de Aceptación Tesis

León, Gto. 04 de Noviembre del 2024

Dr. Modesto Antonio Sosa Aquino
Director de la División de Ciencias e Ingenierías
Universidad de Guanajuato Campus León

Estimado Dr. Sosa:

Por este medio de la presente me permito enviarle un cordial saludo. Sirva además la presente para hacer de su conocimiento que he revisado cuidadosamente el trabajo de tesis titulado: **"AI- Driven deep learning models for enhanced medical image diagnosis"** desarrollado por el **MCA Aron Hernández Trinidad** para obtener el grado de **Doctor en Física**.

Le comunico que el estudiante atendió de forma completa mis comentarios y correcciones sobre el documento. Considero que tanto el contenido como los resultados presentados en la tesis son muy completos y que Aron tiene los conocimientos y las aptitudes para defender su trabajo de tesis sin ningún inconveniente.

Por lo tanto, confirmo que el MCA Hernández puede realizar los trámites para presentar su trabajo de tesis en las fechas que se establezcan.

Sin otro particular, me despido quedando a sus órdenes ante cualquier información adicional.

Dra. Marysol García Pérez
Investigadora Tiempo Completo
Facultad Ingenierías y Tecnologías / Dirección de Investigación y Doctorado
La Salle Bajío, Campus Campestre

UNIVERSIDAD DE GUANAJUATO

Oficio: JMBO/2024-041

Asunto: Revisión de Tesis de Doctorado

León, Gto., 25 de octubre de 2024

DR. MODESTO SOSA AQUINO
DIRECTOR DE LA DIVISIÓN DE CIENCIAS E INGENIERÍAS
UNIVERSIDAD DE GUANAJUATO – CAMPUS LEÓN
PRESENTE

Por este medio HAGO CONSTAR el trabajo de investigación del estudiante Aron Hernández Trinidad Intitulado: "**AI-Driven Deep Learning Models for Enhanced Medical Image Diagnosis**", en donde no solicité cambios y que sustentará para obtener el grado de Doctor en Física. Por tal motivo, sugiero que le permita realizar la defensa de su trabajo de investigación

Sin más por el momento, agradezco de antemano su amable atención.

ATENTAMENTE
"LA VERDAD OS HARÁ LIBRES"

Dr. José Marco Balleza Ordaz
Responsable del CA de Ingeniería Física de Procesos Biológicos y Clínico-Médicos
Departamento de Ingeniería Física
División de Ciencias e Ingenierías
Universidad de Guanajuato – Campus León