

Bachelor Thesis

"System for identifying and interpreting pulmonary pathologies from X-ray images using Deep Learning."

Coordinator

Fernando Pozo Ocampo

Academic tutor

Manuel Vázquez López

Leganés, 2022



This work is licensed under the Creative Commons **Attribution - Non-Commercial - No Derivative Works** license.

ABSTRACT

Chest X-rays are often the first opportunity to diagnose the most common pulmonary pathologies. Here we show a system for automatic processing, identification, and interpretation of clinical pulmonary manifestations from chest X-rays using various artificial intelligence techniques.

The study will analyze the different methodologies implemented for the diagnosis of pneumonia, the detection of possible pulmonary pathologies (such as cardiomegaly or pleural effusion), and the automatic generation of clinical reports on these pathologies, demonstrating with quantitative results the current capacity of deep learning to help diagnose pulmonary pathological manifestations from chest X-rays.

The state of the art of the proposed intelligent systems has been evaluated, both in the detection of pneumonia and in the multiple classifications of pulmonary pathologies employing computer vision and transfer learning techniques, as well as their interpretation through the latest techniques based on saliency maps, and the automatic generation of clinical reports from natural language processing techniques. In addition, a web application has been developed, translated into over one hundred languages, and ready to be used by clinicians to explore, classify and interpret possible pulmonary pathologies, proposing a diagnostic methodology based on the interaction between an artificial intelligence system and the health professional himself.

Methodologies such as the one presented below will speed up and improve the entire diagnostic process, leaving the clinician more time for other tasks such as personalized patient care, interpretation of medical tests, or manual curation of available databases, whose number of images has increased exponentially since the COVID19 pandemic began just two years ago.

Keywords: Chest X-ray image, Deep learning, computer vision, natural language processing.

RESUMEN

Las radiografías de tórax son, en diversas ocasiones la primera oportunidad para diagnosticar las patologías pulmonares más comunes. Aquí se muestra un sistema de procesamiento, identificación e interpretación automática de manifestaciones clínicas pulmonares a partir de radiografías de tórax mediante diversas técnicas de inteligencia artificial.

Dentro del estudio se analizarán las diferentes metodologías implementadas para el diagnóstico de neumonía, la detección de posibles patologías pulmonares (como pueden ser la cardiomegalia o la efusión pleural) y la generación automática de informes clínicos sobre dichas patologías, demostrando con resultados cuantitativos la capacidad actual que tiene el aprendizaje profundo para servir de ayuda en diagnosticar manifestaciones patológicas pulmonares a partir de radiografías de tórax.

Se ha evaluado el estado del arte de los sistemas inteligentes propuestos, tanto en la detección de neumonía como en la clasificación múltiple de patologías pulmonares mediante técnicas de visión por ordenador y aprendizaje por transferencia, así como también su interpretación mediante las técnicas más novedosas basadas en mapas de saliencia, y la generación automática de informes clínicos a partir de técnicas de procesamiento del lenguaje natural. Además, se ha desarrollado una aplicación web traducida en más de cien lenguajes y preparada para ser utilizada por los clínicos para explorar, clasificar e interpretar posibles patologías pulmonares, proponiendo una metodología de diagnóstico basada en la interacción entre un sistema de inteligencia artificial y el propio profesional sanitario.

Metodologías como la presentada a continuación, acelerarán y mejorarán el proceso completo de diagnóstico, dejando más tiempo al clínico para otras tareas como la atención personalizada del paciente, la interpretación de las pruebas médicas o la curación manual de bases de datos disponibles, cuyo número de imágenes ha aumentado exponencialmente desde la pandemia de COVID19 que comenzó hace escasos dos años.

Palabras clave: Imagen de radiografía de tórax, Aprendizaje profundo, visión por ordenador, procesamiento del lenguaje natural.

ACKNOWLEDGEMENTS

First of all, I want to thank my family for their constant support during all the days of my life; they have been the fundamental support of help, love, and thanks forever.

I would also like to thank Fernando Pozo for collaborating on this project with me and supporting me in the most challenging moments of my study. I am also grateful for all that I have been able to learn from him on a technical, professional, and personal level; he is a great tutor and person.

Finally, I would like to express my gratitude to all my classmates for the opportunity to share all kinds of educational and personal experiences and thank them for the company, cooperation, and support I have received from them.

TABLE OF CONTENTS

1.INTRODUCTION	1
1.1 Motivation	1
1.2 Objectives	2
2.METHODS AND APPLICATIONS	3
2.1 Current situation	3
2.2 Definitions and basic concepts	4
2.3 Neural networks	6
2.3.1 Simple Perceptron	6
2.3.2 Multilayer perceptron	7
2.3.3 Backward Propagation, explanation, and conceptual example	8
2.3.4 Convolutional Networks	8
2.3.4.1 Image properties	8
2.3.4.2 Structure of a supervised convolutional network	9
2.3.4.3 Densenet121	11
2.4 Saliency Maps	12
2.5 Transfer learning	14
2.6 Torchxrayvision	15
2.7 Evaluation metrics for rating models	16
2.7.1. Main evaluation definitions	16
2.7.2 Definition of evaluation metrics	16
2.8 Natural language processing and state of the art of the field	17
2.9 Chest radiography	18
3.PROJECT STRUCTURE	19
3.1 Introduction and structure	19
3.2 Database search	20
3.3 Image preprocessing	21
3.4. Training of classification models	22
3.4.1 Detection of pneumonia	22
3.4.2 Identification of pulmonary pathologies	22
3.5 Lung pathology interpretation systems	24
3.6 Automatic diagnostic report generation	25
3.7 Web production	26
3.8 Code repository	27
4.RESULTS AND DISCUSSION OF THE STUDY	28
4.1 Database exploration and analysis	28
4.1.1 Analysis of databases with different pathologies	28
4.1.2 Definition of a database for the identification of pneumonia	31

4.2 Results of the training models	31
4.2.1 Evaluation of Pneumonia Screening Models	31
4.2.2 Evaluation of the pathology detection model	33
4.3 Evaluation of the intelligent models and interpretation of the saliency maps with the support of a radiologist	35
4.3.1 Pneumonia screening	35
4.3.1.1 Patient 1	36
4.3.1.2 Patient 2	37
4.3.1.3 Patient 3	38
4.3.1.4 Patient 4	39
4.3.1.5 Patient 5	40
4.3.1.6 Patient 6	41
4.3.2 Detection of pathologies	42
4.3.2.1 Diseases	42
4.3.2.1.1 Patient 1 (Pneumothorax)	42
4.3.2.2.1.2 Patient 2 (Fracture)	43
4.3.2.2 Pulmonary Lesions	44
4.3.2.2.1 Patient 3 (Consolidation)	44
4.3.2.2.2 Patient 4 (Atelectasis)	45
4.3.2.3 Radiological Findings	46
4.3.2.3.1 Patient 5 (Cardiomegaly)	46
4.3.3 Automatic report generation	47
4.3.3.1 Report 1	48
4.3.3.2 Report 2	49
4.3.3.3 Report 3	50
4.3.3.4 Report 4	51
4.3.3.5 Report 5	52
4.3.3.6 Report 6	53
5. STUDY MANAGEMENT	54
5.1 Socioeconomic impact	54
5.2 Budget	55
5.2.1 Labor costs	55
5.2.2 Software costs	55
5.2.3 Hardware costs	56
5.2.4 Material cost	57
5.2.5 Total costs	57
5.3 Regulatory framework	58
6. CONCLUSIONS	59
6.1 Main conclusions	59
6.2 Future Work	60
7. REFERENCES	62

INDEX OF FIGURES

Fig. 1 Diagram of artificial intelligence [11].....	5
Fig. 2 Diagram of a simple perceptron [12].	6
Fig. 3 Diagram of a multilayer perceptron [13].....	7
Fig. 4 Grayscale color depth [14].	9
Fig. 5 Image with different resolutions [15].....	9
Fig. 6 Convolution of a matrix with a 2x2 filter [17].....	10
Fig. 7 <i>Max pooling</i> operation [18].....	10
Fig. 8 Convolutional network for dermatological diagnosis [19].	11
Fig. 9 Dense block with five convolutional layers [20].	12
Fig. 10 Diagram of a <i>DenseNet</i> network [20].	12
Fig. 11 Identification of birds with saliency maps [21].....	13
Fig. 12 GradCAM for the identification of pneumonia [22].	14
Fig. 13 Transfer learning strategies [25].	15
Fig. 14 Confusion matrix. Visualization of different metrics for the evaluation of classification models [27].....	17
Fig. 15 Image preprocessing diagram.....	22
Fig. 16 Diagram of the different MIMIC-CXR-JPG lung pathologies [43].....	23
Fig. 17 Distribution of pathologies in the training set.....	24
Fig. 18 Architecture of the diagnostic report generation model [9]	25
Fig. 19 Diagram of the application functionality.	27
Fig. 20 Bar chart of pathologies defined in <i>Chexpert</i> , <i>CC-CXRI</i> , <i>NIH Chest X-Ray</i> , and <i>MIMIC-CXR-JPG</i>	29
Fig. 21 Percentage of uncertainty in the different studies associated with <i>Chexpert</i> and <i>MIMIC-CXR-JPG</i>	30
Fig. 22 Pearson correlation of labels defined in <i>Chexpert</i> and <i>CC-CXRI-P</i>	30
Fig. 23 Bar chart with images of pneumonia and normal controls in different databases.	31
Fig. 24 Area under the curve and confusion matrix for the <i>CC-CXRI-P</i> test set. Results from the <i>All</i> base model.....	32
Fig. 25 Balance data partitioning in training and testing for the classification of multiple pathologies.....	33
Fig. 26 Average <i>AUC</i> of the models previously trained in <i>Torchxrayvision</i> [46].....	34
Fig. 27 Pneumonia Diagnosis (patient 1) radiologist and heat map evaluation.	36
Fig. 28 Pneumonia Diagnosis (patient 2) radiologist and heat map evaluation.	37
Fig. 29 Normal diagnosis (patient 3) radiologist and heat map evaluation.	38
Fig. 30 Normal diagnosis (patient 4) physician and heat map evaluation.....	39
Fig. 31 Pneumonia diagnosis (patient 5) evaluation of the model and heat map.....	40
Fig. 32 Pneumonia diagnosis (patient 6) physician and heat map assessment.....	41

TABLE INDEX

Table 1. EVALUATION OF THE TRAINING MODEL WITH PREVIOUSLY TRAINED MODELS FROM DIFFERENT DATABASES.	32
Table 2 EVALUATION OF THE ADJUSTED MULTICLASSIFICATION LUNG DISEASES TRAINING MODEL	33
Table 3 EVALUATION OF PATIENT 1 WITH PNEUMOTORAX LABELED. EVALUATION OF THE RADIOLOGIST AND THE MODEL.	42
Table 4 EVALUATION OF PATIENT 2 WITH FRACTURE LABELED. RADIOLOGIST AND MODEL EVALUATION.	43
Table 5 EVALUATION OF PATIENT 3 WITH CONSOLIDATION LABELED. RADIOLOGIST AND MODEL EVALUATION.	44
Table 6 EVALUATION OF PATIENT 4 WITH ATELECTASIS LABELED. EVALUATION OF THE RADIOLOGIST AND THE MODEL.	45
Table 7 EVALUATION OF PATIENT 5 WITH CARDIOMEGALY LABELED. EVALUATION OF THE RADIOLOGIST AND THE MODEL.	46
Table 8 EVALUATION OF THE REPORT 1. PERFORMANCE OF THE RADIOLOGIST AND THE CLINICAL REPORT MODEL.	48
Table 9 EVALUATION OF THE REPORT 2. PERFORMANCE OF THE RADIOLOGIST AND THE CLINICAL REPORT MODEL.	49
Table 10 EVALUATION OF REPORT 3. PERFORMANCE OF THE RADIOLOGIST AND THE CLINICAL REPORT MODEL.	50
Table 11 EVALUATION OF THE REPORT 4. PERFORMANCE OF THE RADIOLOGIST AND THE CLINICAL REPORT MODEL.	51
Table 12 EVALUATION OF THE REPORT 5. PERFORMANCE OF THE RADIOLOGIST AND THE CLINICAL REPORT MODEL.	52
Table 13 EVALUATION OF REPORT 6. PERFORMANCE OF THE RADIOLOGIST AND THE CLINICAL REPORT TEMPLATE.	53
Table 14 LABOR COSTS.	55
Table 15 SOFTWARE COSTS.	56
Table 16 HARDWARE COSTS.	56
Table 17 MATERIAL COSTS.	57
Table 18 TOTAL COSTS.	57

MATHEMATICAL EXPRESSIONS

Equation 1. Parameters for <i>GradCAM</i> calculation.....	13
Equation 2. Calculation of <i>GradCAM</i> concerning the class to be attended.	14

1. INTRODUCTION

The introduction explains the main reason for proposing a work-oriented application of intelligent systems in personalized medicine sciences. In the same way, the subject of the study is presented, and the objectives to be achieved are established, listing the tasks performed and mentioning the main methods used in the project's development.

1.1 Motivation

The use of software tools for medicine based on artificial intelligence techniques is already demonstrating a direct impact in different areas of clinical and health application, among others, concerning the diagnosis of various existing diseases and pathologies [1].

Thus, the revolution of personalized medicine requires the emerging analysis of collecting a lot of data in the most accurate way and in the shortest possible time to be used for its application. One of the most common measures to solve this problem is employing cutting-edge technologies to perform computational data calculations at the quantitative and qualitative health science information levels [1]. On the one hand, it is worth highlighting the term *Big Data*, a set of methodologies that allow the analysis of samples on an extensive experimental scale, offering techniques to compute massive storage volume data. At the same time, numerical optimization systems stand out, providing a rational sense to the calculations made, demonstrating observational intelligence with making a decision or capturing relationships of specific data with others, and providing solutions to some real-life issues. Currently, medicine is opting to use these technologies to contribute to better quality health. Thanks to these technologies, patient data can be managed in the most efficient way possible in different hospitals around the world, diagnostic reports can be generated with voice recordings, patient medical history information can be accessed standardized, etc.

In the future, it is intended to know the limit to which artificial intelligence can act in the field of health and if, at some point, it will be able to surpass human intelligence (see section 5.1 for the socio-economical impact of the study). At the moment, the investment in artificial intelligence and its evolution are growing exponentially until today [2], offering the opportunity to generate new ideas that not only help people's social, mental, and physical well-being but also in other fields such as the economy or industry. At the same time, this development offers to obtain more complex health solutions related to a deep understanding of the functioning of the human body.

1.2 Objectives

The work's main objective consists of creating an artificial intelligence web ecosystem of free access and uncomplicated use to assist clinicians in diagnosing pulmonary pathologies. In the same way that the ecosystem helps physicians, the second objective proposed is to achieve reciprocal help from expert clinicians, recognizing misdiagnoses made by the artificial intelligence system and being able to label or modify the diagnosis based on health knowledge.

The assistance that the intelligent system provides to the clinician consists of three functionalities that perform different lung diagnostic tasks:

1. Pneumonia detection is based on detecting whether the lung X-ray contains pneumonia or falls within the usual standards, locating the lung areas affected by pneumonia, and recognizing the regular regions of the image.
2. Detection of pulmonary pathologies: the second objective attempts to detect whether the chest X-ray contains different pathologies, including possible diseases (pneumonia, pneumothorax), lung lesions (atelectasis, consolidation), or radiological findings proposing to locate the areas affected by each of the pathologies spatially.
3. Automatic diagnostic report: the objective is to generate a radiological diagnostic report from a lung X-ray image.

These functionalities are the main reason for being part of a web application and how the clinician assists. The application aims to be straightforward to use. It also includes a translator to complement the application, making it an accessible and universal tool for clinicians worldwide.

For the creation of an intelligent ecosystem, databases of pulmonary radiographs are compiled, refined, and analyzed in compliance with the legal regulations of the established frameworks (see section 5.3) as well as the financial resources necessary for the development of the study (see section 5.2).

The main methods to implement the intelligent systems that execute the different functionalities are mainly based on various fields of artificial intelligence and image processing, including the most innovative designs of state of the art to meet the established objectives.

The results of the implemented intelligent systems are evaluated in quantitative terms by using metrics that report the performance of the systems in statistical terms, making a comparison with state of the art. The evaluation of these systems by an expert radiologist is also proposed, obtaining a more critical and technical opinion.

2. METHODS AND APPLICATIONS

This section describes the current advances in applying artificial intelligence systems to the health field and the techniques used to develop the design to understand the project better.

2.1 Current situation

Artificial intelligence has advanced with systems, tools, and research applied to health sciences in recent years. One of the most innovative developments in the field of health sciences is the tool *AlphaFold 2* [3], a technology whose means is deep learning to predict the three-dimensional structure of a protein from its amino acid sequence. For decades this problem has been approached from the biological point of view without achieving a transcendent solution. However, *AlphaFold 2* [3] has managed to propose the best solution to the problem of protein structure prediction, according to the *Critical Assessment Techniques for Protein Structure (CASP)* [4].

Also important is the use of technologies for computerized medical records management, such as *MedKnowts* [5], which allows the display of intuitive visualizations of patients' medical data, using a simple, fast tool and attempting to ensure interoperability of the different types of clinical data.

On the other hand, artificial intelligence takes advantage of web applications to ensure an intelligent system whose use is functional for the physician or the consumer of the application. *SEMIC RF Electronic GmbH* has implemented a web application on multiplatform mobile systems that can identify through an artificial intelligence system if a person has COVID from a photo of the eye [6]. This application manages with a specificity of 97 percent to predict that the patient is healthy, achieving a faster and cheaper method to diagnose the virus (see section 2.7 for the definition of specificity).

Concerning the reproduction of the study, different research projects that apply artificial intelligence to analyze the X-ray image to diagnose certain lung diseases have been taken into account. In 2020, a research project developed a deep learning system to detect COVID and normal controls from an image, achieving remarkable numerical results [7]. There has also been a desire to go further, trying to localize different lung diseases and pathologies to extend the results of diagnostic radiographic imaging and assess lung health depending on its condition [8].

Finally, the possibility of generating an automatic report from an X-ray image [9] should be highlighted as the latest novelty. This last work is very premature, and there is hardly as much research done as in the previous two. Therefore, it is a novel project to continue developing and providing new ideas for the future.

All the state-of-the-art collected aims to link artificial intelligence with professional clinicians and develop an accessible, free, and easy-to-use web system [10].

2.2 Definitions and basic concepts

In order to understand the process and the methodologies implemented, prior knowledge of theoretical concepts related to artificial intelligence is required. Some of the most important ideas associated with the development of the study are listed below:

- **Observation:** it is the set of several data of the different semantic types that belong to or are identified with a subject. For example, in a data table where the weight and height of the family members are represented, the weight and size of a family member are considered one observation. Another example would be to have a list of images of animals. In this case, each observation would correspond to the idea of an animal.
- **Variable:** it is a characteristic of the set of observations in a data table. Regarding the two previous examples, two different variables or factors are considered: each family member's weight and height. In the second example, the variables would be the set of pixels that characterize the image of the animal.
- **Training:** this is the state by which the artificial intelligence algorithm learns to make the decisions or approaches proposed. A considerable number of observations are required to train an intelligent system. The greater the number of observations (considering that the observations are not duplicated), the more experience and learning it will get when performing that task. It does not mean that the results will be better in terms of accuracy, but that you will know in a more realistic way how accurately the algorithm will be able to perform.
- **Artificial Intelligence:** it is the set of techniques, tools, methods, and processes that allow computers to have an autonomous behavior to approximate that of a human being. Its structure comprises different fields; not all are mentioned, only those related to the project. Within artificial intelligence, there are separate fields of study for the development of intelligent tasks:
 - **Big Data:** it is a field of artificial intelligence responsible for calculating and efficiently managing large volumes of data.
 - **Machine Learning:** it is a branch of artificial intelligence in charge of making a series of decisions or objectives established from observing data information [11]. They are usually algorithms applied to data tables, and the procedures to reach the solution are based on mathematical optimization problems. Within machine learning, we distinguish two types of methods, depending on whether there is a variable to predict or not (supervised and unsupervised learning).
 - **Supervised learning** requires a variable as a target to make an intelligent decision. For example, if we have weight and height information, we could consider height as a variable to be guessed from weight information. Depending on the type of variable to predict, we believe in two different types of algorithms [11].
 - **Regression:** the objective to be predicted is of a numerical or continuous nature. An example would be to predict the height of a person.

- Classification: the target to predict consists of discrete or dichotomous data. An example would be to decide whether there is a fish in an image or not.
- Unsupervised learning captures groups or patterns containing the different observations of a data table employing the other numerical relationships that may exist between them, without the need to establish the objective of taking a variable to predict. Therefore, these methods do not guarantee to learn to make a decision sought a priori [11].
- Deep learning: Deep learning is a field of machine learning also aimed at making intelligent decisions [11]. The main difference in the application of both is in the dimensionality of the data. A clear example of a data type on which deep learning is applied could be images, where each image represents an observation in the database, defined by different pixels that constitute the characteristics or variables of the image and determine a high dimensionality or range of variables in the observation. When the observation is represented in high dimensions, deep learning is usually applied to have algorithms with a more complex structure that allows capturing all the information of this data.

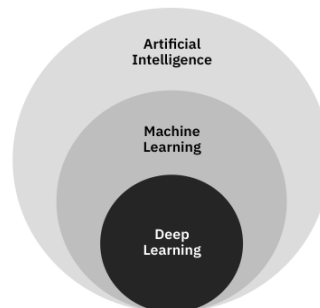


Fig. 1 Diagram of artificial intelligence [11].

2.3 Neural networks

Neural networks are a subset of the field of machine learning whose architecture is based on connections between units known as *neurons*, which capture patterns or characteristics of the experience of observing data. Within this field, different network structures vary depending on the problem. The section will discuss different types of neural networks to understand the project's design.

2.3.1 Simple Perceptron

The simple perceptron is a neural network structure to predict or monitor a task represented by a value. The perceptron structure is constituted by several neurons defined and adjusted to the number of variables (input information to be considered for predicting the output value). Each neuron is usually many numerical, dichotomous, or discrete characters. The input information suffers a transformation through a mathematical function resulting in a scalar number. This first step makes it possible to project a data set to a numerical value that will later represent a decision or an approximation to the target variable. Finally, a non-linear function, known as the *activation function*, is responsible for making a decision or approximation.

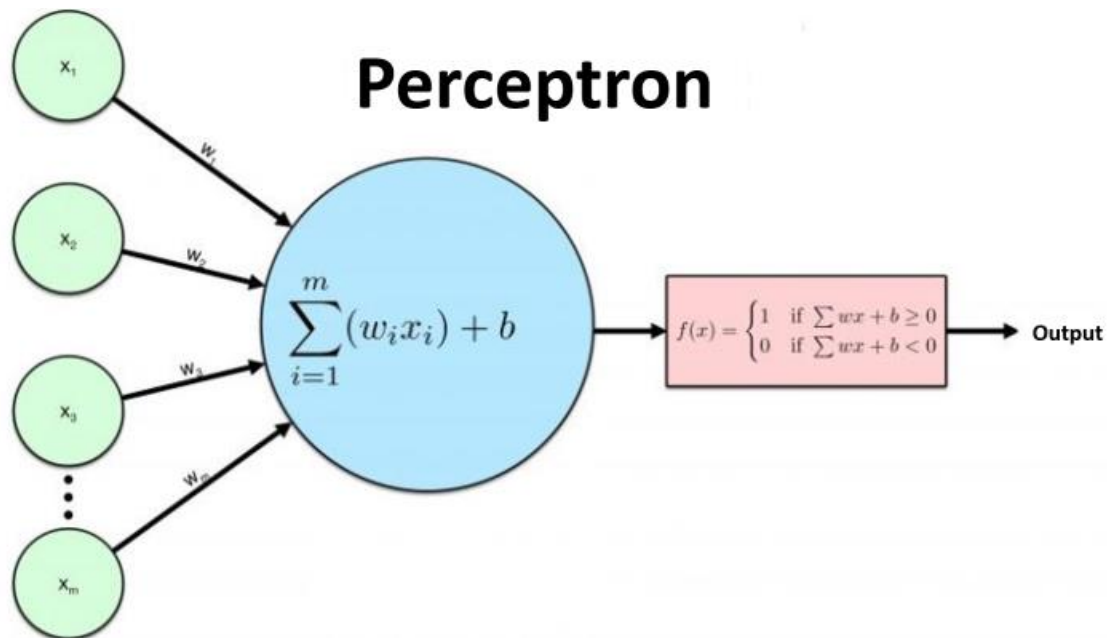


Fig. 2 Diagram of a simple perceptron [12].

The connections represented by the letter w are the parameters that allow the system to learn to make decisions regarding the target to be predicted and compose the machine intelligence. In the learning moment known as *training*, parameters will adjust their value to perform the task as well as possible, given the architecture established in the structure of the perceptron.

Most supervised machine learning algorithms are simple perceptrons where the input data is a table of variables used to predict a numerical (regression) or discrete (classification) target.

2.3.2 Multilayer perceptron

The multilayer perceptron is a more complex network structure than the previous one in which its composition consists of layers of neurons. These layers capture the information from the input data through a directional connection between them. Once the data traverses all layers, knowledge of the last layer is projected to a number or a vector of numbers. Finally, an *activation function* is applied as output to make a decision or approximate a value [13]. In the transition between layers, activation functions can also be given in each neuron, capturing nonlinear relationships of the input data [13].

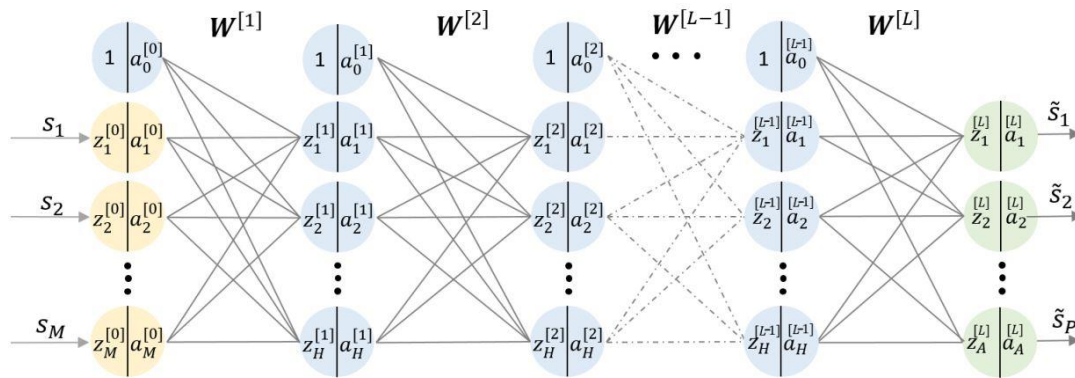


Fig. 3 Diagram of a multilayer perceptron [13].

In Figure 3, the connections of neurons from one layer to the next are denoted by the letter W . The number of links of the multilayer perceptron depends on the number of neurons in each layer and the layer's depth constituting the network.

Usually, this algorithmic structure is applied when a large number of variables are found in the structural information of the data to capture in the design of the network patterns of information.

2.3.3 Backward Propagation, explanation, and conceptual example

In the previous chapters, we discussed the transformations that drive a data input to a value representing a task or a conceived approximation. The question in this section is how we indicate to the network the job it has to identify or the value it has to approximate from the data input.

Back propagation is the computational method applied to train neural networks. The optimization of the computation is performed through partial derivatives from the output neuron (which is the task to learn) to the neurons of the network structure, setting an objective. The objective could be to minimize the error between the actual price of a supermarket product (for example, a bag of potato chips) with the price proposed by the output neuron, based on the information of specific characteristics of the product (for example, the ingredients required to create the bag of potato chips), which in its case would be the input information to the neural network.

The objective is optimized by a mathematical function known as a *loss function* and varies depending on the task to teach to the artificially intelligent system. In the example above, the *loss function* could minimize the average error between the product's actual price and the prediction made about the product price. The partial derivatives search for the parameters of the neurons of the network that help the value predicted by the neural network to be as close as possible to the actual value. These parameters are updated according to the input information, capturing patterns and characteristics that help minimize the price prediction error, trying to learn by experience a system for predicting the price of supermarket products.

2.3.4 Convolutional Networks

Convolutional networks are a subfield of neural networks based on systems specialized in supervised and unsupervised image training. In our case, the operation of supervised convolutional networks is explained. First, a brief description of the properties of the image, which is the input of the information of the convolutional neural network, is given. Subsequently, the structure of the convolutional network is explained.

2.3.4.1 Image properties

In technical terms, the image (information input from the convolutional network) is a matrix or array of matrices whose numbers are known as *pixels* and represent the details of the picture. Each pixel contains a finite range of values where each pixel defines a distinct color. Commonly, 8 bits of depth are used, corresponding to 255 different shades of color, and characterize the range of numerical values of the pixel.

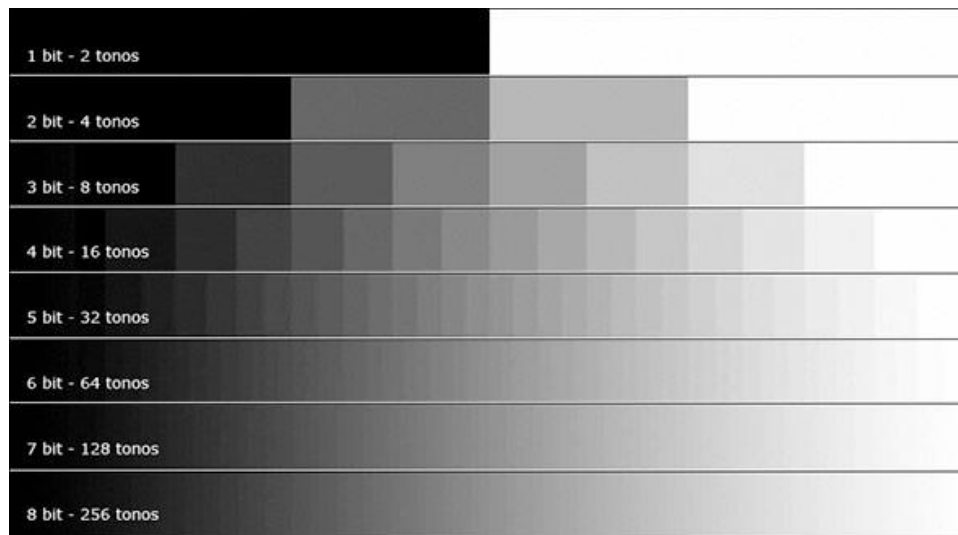


Fig. 4 Grayscale color depth [14].

On the other hand, the number of rows and columns in the matrix that define the *image's resolution* is essential. The *resolution* determines the quality of the image in spatial terms.

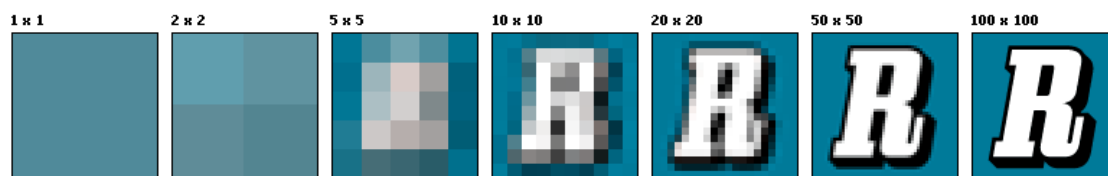


Fig. 5 Image with different resolutions [15].

Both image characteristics determine the amount of information (pixel sample and image color depth) that the convolutional network can learn, and both the depth of the network and the number of parameters to be taught are adjusted according to these two characteristics.

2.3.4.2 Structure of a supervised convolutional network

The structure of the convolutional network is composed of an input layer based on an image, layers' links that identify the patterns of the image, and an output layer represented in a number or a vector of numbers which is the decision or approximation to take into account.

The input matrix undergoes a dimensional reduction as the information goes through the structure of each layer that makes up the network. This linear reduction is achieved through a linear transformation known as convolutions, a mathematical operation between the input matrix and another matrix called *filters* which are responsible for capturing the spatial information of the image pixels [16]. The *filters* perform a linear

function concerning each of the pixels of the input image, achieving an information output with fewer rows and columns than the input matrix [16]. The convolutional operations and their respective transformations compose a layer structure. It can be seen in figure 6 how the filter goes through the input matrix, completing the convolution process and obtaining an output of three rows and three columns.

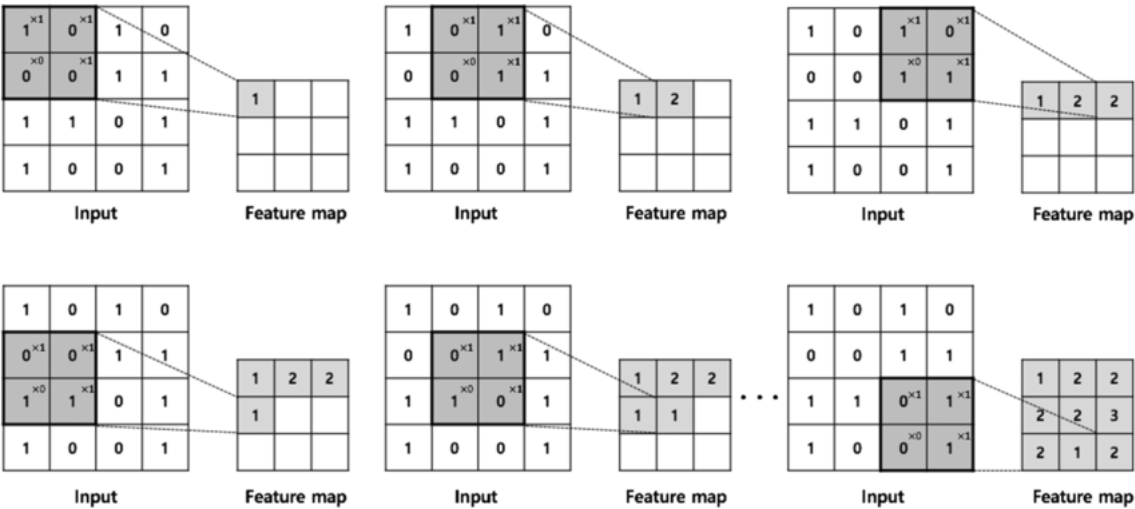


Fig. 6 Convolution of a matrix with a 2x2 filter [17].

After convolution, other dimensional reduction techniques are often applied, such as *max-pooling*, which interpolates the maximum value of an attached local region of the image [16]. Next to the output, operations such as an activation function or a linear function can be applied.

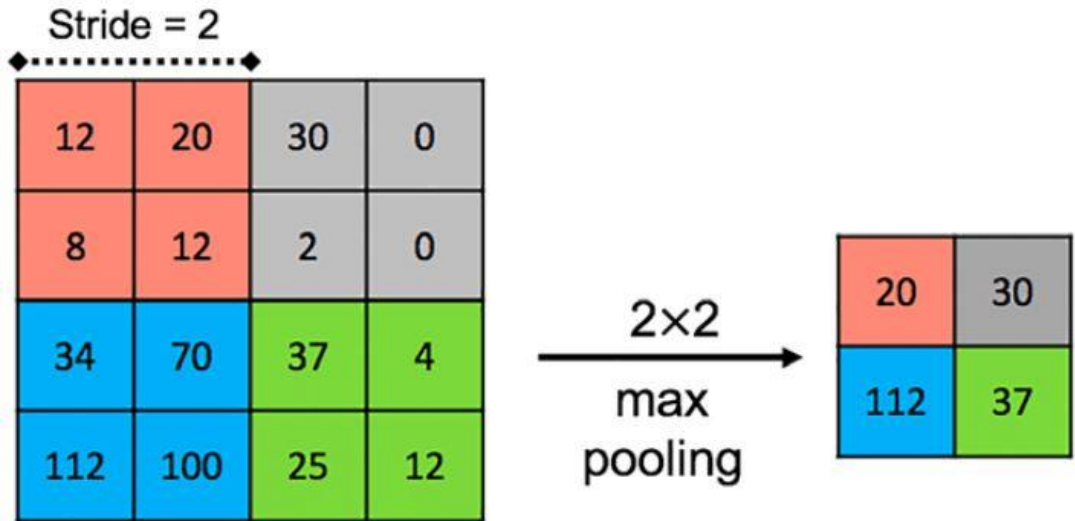


Fig. 7 *Max pooling* operation [18].

Once the image is dimensionally reduced, a transformation of the matrix of N rows and M columns to a vector with the information defined in the matrix of size $N \times M$ is performed. This vector is passed through a multilayer perceptron to obtain a vector or a scalar for decision making.

It should be noted that the first layers of the convolutional network usually capture the most general information of the images. As the data moves through the network layers, the structure computes patterns more related to the decision.

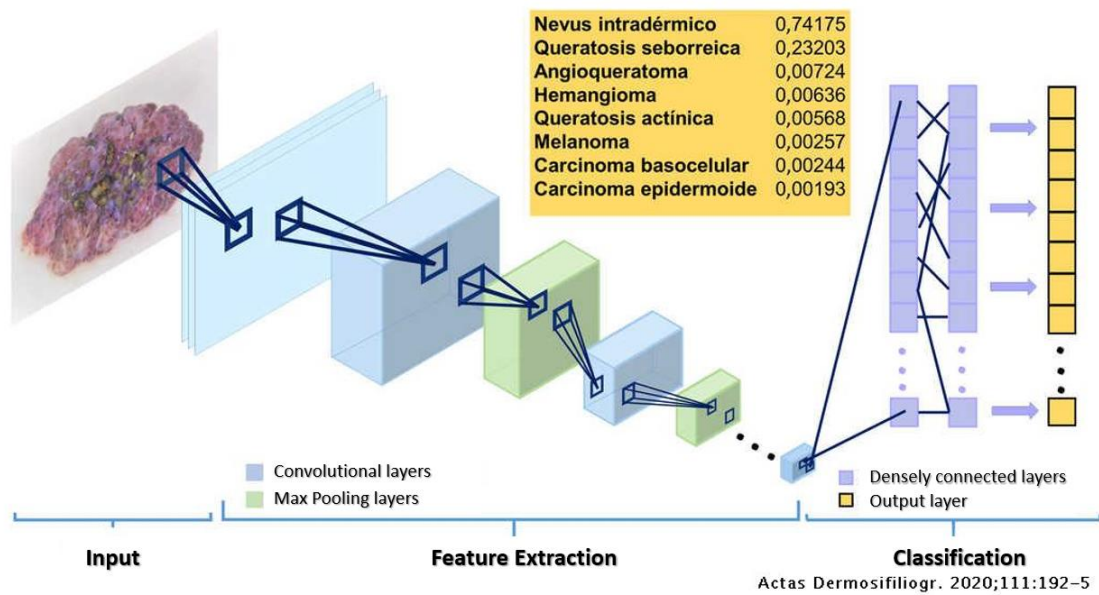


Fig. 8 Convolutional network for dermatological diagnosis [19].

2.3.4.3 Densenet121

Densenet121 [20] is a variation of the traditional convolutional network architecture. While the layer connections of traditional convolutional networks are sequential, this architecture proposes a relationship between the subsequent layer and its consecutive layers. The network whose layers are connected between the next layer and its successive layers is known as a *dense block* and constitutes a subset of the *Densenet121* network architecture [20]. Figure 9 shows the spatial structure of a *dense block* formed by five convolutional layers.

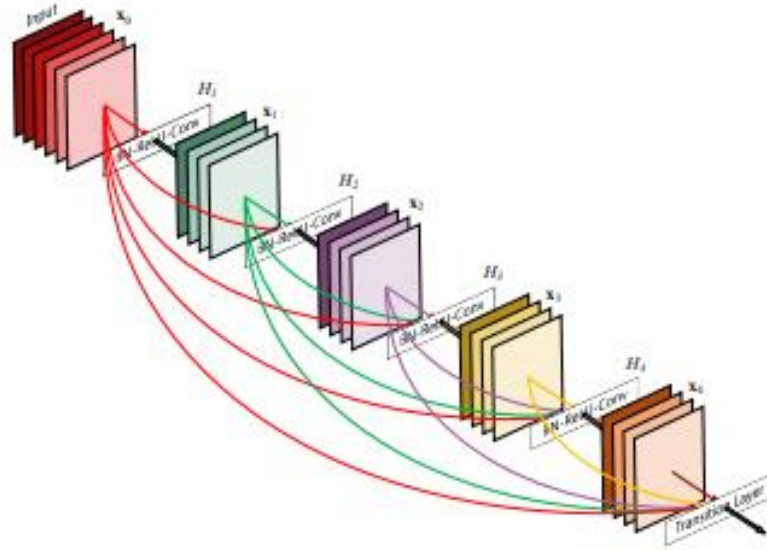


Fig. 9 Dense block with five convolutional layers [20].

The *Densenet121* network structure consists of four dense blocks with a convolution at the transition from the input image connected to the first dense block, convolution, and a *max-pooling* as a connection transition between dense blocks in blocks 1 and 2, and 3. After block 4, a *global average pooling* (a variation of the *max pooling* operation) is applied, transforming the spatial information from a matrix array to a numeric vector. The network connects to a linear layer which can be a vector of numbers or a numeric value representing the task to be decided [20]. An intuitive example of the connections of a *DenseNet* type of network is shown in figure 10.

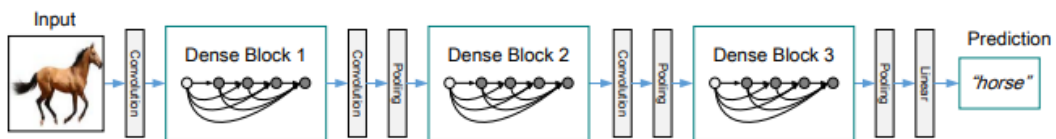


Fig. 10 Diagram of a *DenseNet* network [20].

2.4 Saliency Maps

Saliency maps are novel algorithms in the state of the art of computer vision and serve to know which areas of the image the convolutional network is attending to decide on a proposed task. These types of techniques add value to the trained models in an intuitive way. A commonly used conceptual example is neural networks that identify various animals from an input image. Saliency maps capture information from network layers that attend to patterns in the image that they consider relevant to locating the animal. This attention is expressed utilizing *heat maps*, color maps that identify which areas of the image the network has attended for decision-making. In figure 11, the reddest areas are the bird's wings, a characteristic that the network considers vital to identify the bird.

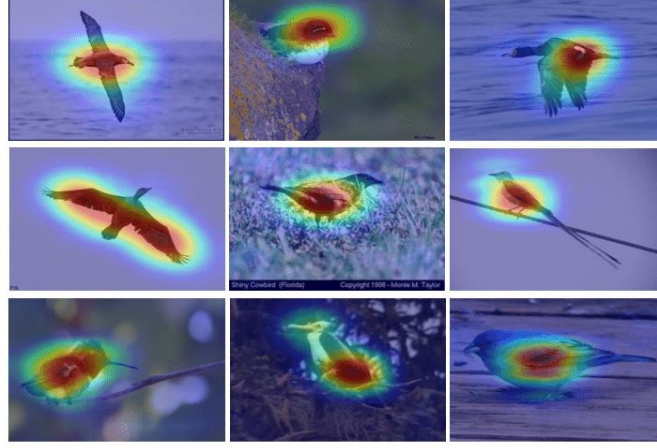


Fig. 11 Identification of birds with saliency maps [21].

In technical terms, the saliency map is the calculation of partial derivatives of the output neuron (which identifies the animal) concerning the last convolutional layer of the network structure. Three different variations of saliency maps are proposed for the development of the project:

- *GradCAM*: the algorithm calculates the partial derivative of the class of interest concerning the pixels of the last convolutional layer. The final convolutional layer has three dimensions (N rows, M columns, and K matrices). For each of the matrices of size $N \times M$, the partial derivative of the class of interest with respect to each of the pixels of the matrix is calculated by averaging the partial results obtained.

Equation 1. Parameters for *GradCAM* calculation.

$$w_k^c = \frac{1}{NM} \sum_i \sum_j \frac{\delta y^c}{\delta A_{ij}} \quad (1)$$

The result of the average obtained in each of the matrices of the last convolutional layer is denoted by w_k^c . This parameter determines the importance matrix k has to attend class c spatially. This calculation is applied to all the matrices of the last convolutional layer; therefore, the result will be a vector of parameters of length K .

Once this operation is performed, the saliency map is calculated by summing the linear multiplication of each of the parameters by its corresponding matrix, obtaining a matrix of $N \times M$ resolutions. Subsequently, the matrix goes through an activation function known as *Relu*, which transforms the negative values of the matrix by zeros, obtaining the pixels that constitute the *GradCAM* attention map.

Equation 2. Calculation of *GradCAM* concerning the class to be attended.

$$GradCAM^c = Relu\left(\sum_k w_k^c A^k\right) \quad (2)$$

Once the saliency map is obtained, the resolution of the resulting matrix is lower than the input image. To project the saliency map to the resolution of the input image, the resolution of the saliency map needs to be equal to the original image. Therefore interpolation techniques are used to achieve the exact resolution as the input image, giving an intuitive idea of how relevant each pixel is in predicting class c .

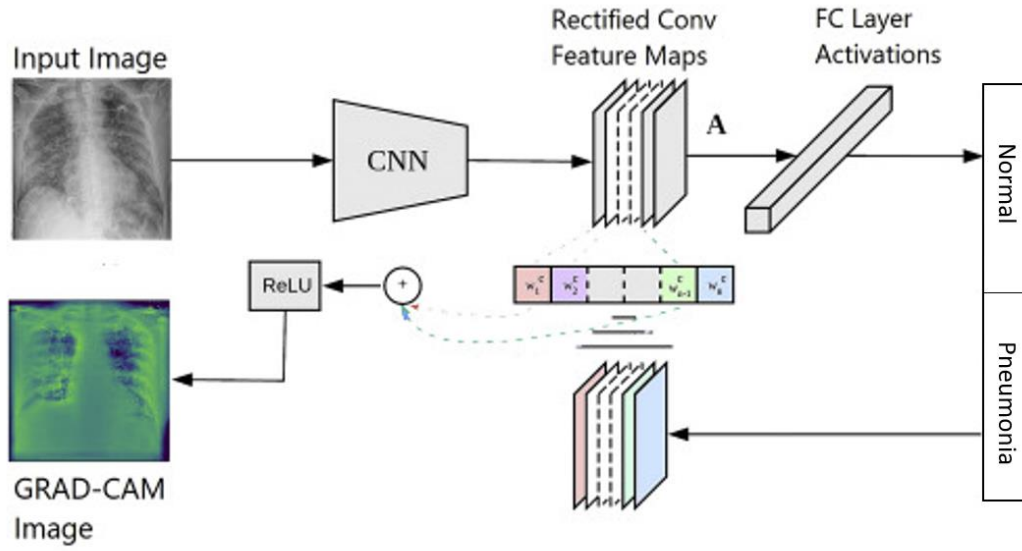


Fig. 12 GradCAM for the identification of pneumonia [22].

It should also be noted that *GradCAM++* and *XGradCAM*, two variations of the *GradCAM* saliency map, have been used to develop the study.

2.5 Transfer learning

Transfer learning is a technique whose objective is storing the learning of a model that performs a task and applying it to another model whose aim is to solve a different problem [23]. These models that transfer the learned parameters are trained from a massive volume of image data. A typical example of a base model on which learning is transferred is the *ImageNet* network [24], a convolutional network trained with 14,197,122 images to perform object classification tasks.

For transfer learning, it is possible to modify which neural network layers will be retrained and which will remain fixed. Depending on the transfer setting applied, the parameters learned by these networks will be used to a greater or lesser extent. By *freezing the layers*,

we mean that the layers of the transfer learning model remain fixed; therefore, at the time of transfer learning, the values of the frozen parameters are never modified. Depending on the layers to be frozen, more or less new parameters are learned. The more layers are frozen in the transfer learning network's structure, the fewer parameters are modified again; therefore, the calculation process is faster. Conversely, the fewer layers are frozen, the more parameters are changed in the new training, and therefore, the computation is more expensive in computational terms.

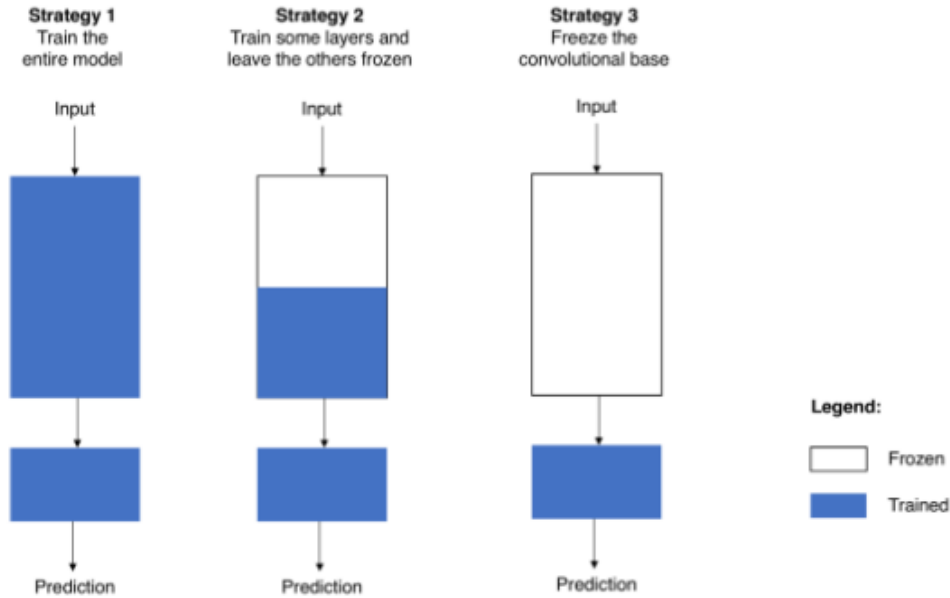


Fig. 13 Transfer learning strategies [25].

2.6 Torchxrayvision

Torchxrayvision is an open-source repository designed in *Python* for X-ray image analysis, modeling, and interpretation [26]. This framework stores models trained on lung radiographs for detecting different lung pathologies. Its use provides excellent value to take advantage of the learning of these models and transfer its experience to the creation of other new models proposed in the development of the project aimed at diagnosing pulmonary pathologies.

2.7 Evaluation metrics for rating models

Once a classification model has been trained, a set of observations that the artificial intelligence model has not seen is proposed, known as the *test set*, to quantitatively show how well the model can classify the proposed targets. The actual values are compared with the model's predicted values, and the prediction's success and failure concerning the actual values are quantified.

2.7.1. Main evaluation definitions

In this section, the explanation of each evaluation metric presented in the design of the work is proposed. We assume that from a radiograph, it is identified that there is pneumonia or that the image is typical. Therefore, in the test set, we have the actual values of which images contain pneumonia and which are normal. At the same time, you have the information on which images the model has diagnosed with pneumonia and which images are typical. The true positives (VP) set is the number of test observations with pneumonia, and the model predicts pneumonia. Similarly, the true negatives (TN) are the set of normal specimens that the model predicts as normal. The model error would consist of the collection of samples with pneumonia that are predicted as normal radiographs (FN) and the typical images that are predicted as radiographs with pneumonia (FP), known as false negatives and false positives, respectively.

2.7.2 Definition of evaluation metrics

Once the main terms related to the evaluation are known, the results are expressed through a real value to quantify the model's performance from different perspectives. In the following, the various common perspectives for evaluating the performance of a classification model are proposed:

- **Accuracy:** *Accuracy* computes the sum of true positives and negatives in proportion to the total number of observations in the sample. Regarding the example offered in section 2.7.1, the accuracy would be the total of pneumonia and normal controls correctly classified by the model in proportion to the total observations tested.
- **Sensitivity:** *Sensitivity* evaluates the proportion of positive labels that are well classified or detected as positive. In the example in section 2.7.1, the completeness consists of the ratio of observations the model sees with pneumonia out of the total number of sample observations with pneumonia.
- **Specificity:** analogous to sensitivity, specificity consists of the proportion of observations classified as usual by the model concerning the total number of observations with normal radiographs.
- **Precision:** the metric evaluates the number of positive values in proportion to the values positively predicted by the model.
- **F1-Score:** this metric consists of a harmonized average of the model's sensitivity and precision to know how well the model generally performs in classifying different possibilities.

- The area under the curve: the area under the curve is a method for quantitatively knowing the volume of false positives and false negatives in the evaluation. An area under the curve of one would mean that there are no false positives or false negatives; therefore, it would be a perfect model with no errors concerning the set of test observations evaluated.

		Predicted Class		
		Positive	Negative	
Actual Class	Positive	True Positive (TP)	False Negative (FN) Type II Error	Sensitivity $\frac{TP}{(TP + FN)}$
	Negative	False Positive (FP) Type I Error	True Negative (TN)	Specificity $\frac{TN}{(TN + FP)}$
		Precision $\frac{TP}{(TP + FP)}$	Negative Predictive Value $\frac{TN}{(TN + FN)}$	Accuracy $\frac{TP + TN}{(TP + TN + FP + FN)}$

Fig. 14 Confusion matrix. Visualization of different metrics for the evaluation of classification models [27].

2.8 Natural language processing and state of the art of the field

Natural language processing is a branch of artificial intelligence dedicated to dealing with text, intending to develop systems that try to interpret, understand and translate the human language. Natural language processing takes ideas and knowledge from other artificial intelligence sectors, such as *Big Data*, to process massive natural language data with deep learning networks to develop intelligent text processing tasks.

One of the most novel natural language processing intelligence systems in state of the art is *GPT-3* [28], a neural network model trained with the purpose of simulating the textual behavior of a human being. This technology has had its evolution with different versions released from 2018 to the present day (*GPT-1*, *GPT-2*, AND *GPT-3*). These models are also helpful in performing transfer learning to other tasks, such as the one reproduced in the project design (see section 3.6), in which *GPT-2* [28] is used as a training model to design a correct syntactic structure of sentences that constitute the generation of the clinical report.

GPT-2 [28] is a deep learning model trained from a database known as *WebText* [28] that stores text from millions of web pages. Its mechanism and structure have impacted the world of artificial intelligence since the model is not explicitly supervised and manages to generate coherent paragraphs of text. This model has presented a new version released in 2020, *GPT-3* [28], whose functionalities have a more robust performance than *GPT-2*.

"GPT-3 achieves high performance on many NLP datasets, such as translation tasks, question answering, and gap-filling tasks, as well as several tasks that require on-the-fly reasoning or domain adaptation, such as decoding words, using a new word in a sentence, or performing 3-digit arithmetic operations" [28]. Another very novel technology that competes with *GPT-3* [28] is the *LaMDA* technology [29], an artificial intelligence system developed by *Facebook* designed to simulate human responses according to a proposed topic, as well as to serve as a translator, calculator, or other specific tasks related to obtaining an information response system from the text.

Other tasks developed in natural language processing that have not been mentioned are summarizing texts, analyzing the feelings expressed in a text, translating the text into other languages, or classifying texts by different disciplinary fields.

2.9 Chest radiography

Chest X-rays are images created from ionizing radiation, distinguishing the organs and capturing the inside of the chest. They are used to evaluate the health of the lungs, heart, and chest wall, helping to diagnose a variety of possible pathologies such as pneumonia, cancer, or a heart problem [30].

From a technical approach to image creation, X-rays are a form of radiation that passes through the patient's thoracic area and visualizes a photographic image applied for its generation [30]. Generally, for the diagnosis of chest radiography, two images are used, one from a lateral perspective of the body and another from the patient's back [30]. Many of the advantages of X-ray imaging are that its economic cost is cheaper than other methods used for diagnosing chest pathologies, such as computed tomography, which is more expensive in terms of time and financial expense because its technology is more sophisticated [31].

Unlike the procedure applied by clinicians to report chest X-rays, one of the objectives proposed in the study design is to use only one X-ray image in training artificial intelligence systems to facilitate the diagnosis of pulmonary pathologies (see sections 3.4.1 and 3.4.2). Consequently, economic costs are reduced, avoiding the use of more expensive techniques and easing the workload of clinicians.

3. PROJECT STRUCTURE

The chronological design of the project is presented below, defining each of the tasks proposed for X-ray diagnostic imaging assistance. The databases used, the image preprocessing, the creation of the three proposed models, and the web ecosystem are explained.

3.1 Introduction and structure

This section explains how the study has been developed and what direct application methodologies have been implemented to identify and diagnose pulmonary pathologies. The structure is presented in temporal order distinguished in different sections:

- Database search: the need to diagnose different lung pathologies from lung radiographs required a database search with lung X-ray images and labeling of various lung pathologies.
- Image preprocessing: with previously selected databases, the treatment of X-ray images ensures a minimum of quality and curation of information for our training models.
- Training and adjustment of the image models: subsequently, and once a curated data set has been ensured, the different tasks proposed for diagnosing pulmonary pathology have been trained. The two proposed classification models and the objective to be solved for each of them have been discussed.
- Interpretation of the models: once the training was achieved, the need was sought to explain a visual result of the diagnosis gained through different saliency maps (see section 2.4).
- Natural language processing for the generation of clinical reports: another of the tasks proposed for X-ray imaging has been based on the reproduction of research work to automatically generate a clinical description of the chest X-ray.
- Web production of the project: finally, it has been decided to put all the project's development into a web application, translated into different languages, to create an open-source tool for radiologists worldwide. The production of the web page has been uploaded to the Internet in an open-source repository through the *Github* tool [32] so that clinicians can access it, and the application can aid chest radiodiagnosis.

3.2 Database search

The search for databases has been essential to train the image models, evaluate the model results and compare them with state of the art. The different databases that have been collected about X-ray images have been the following:

- *Chexpert*: a database that stores 226316 lung radiographs with the labeling of different lung pathologies, managed by a group from Stanford University whose objective is to diagnose various lung diseases employing deep learning [33]. Each study is stored in a data table with the possible pathologies detected. Each pathology contains three possible values, each with a corresponding interpretation.
 - o The pathology in the study is "1" if it is present in the patient's imaging study [34].
 - o The pathology in the study is "0" if it is not present in the patient's imaging study [34].
 - o The pathology in the study is "-1" if it is presented indecisively in the study of the patient's image. Images with any "doubtful" label in their pathologies are not considered for training the various artificial intelligence models developed and have therefore been discarded from the process [34].

For the use of *Chexpert*, access permissions have been required to comply with the legislative aspects of the study development (see section 5.3).

- *China National Center for Bioinformation (CC-CXRI)*: stores an open-source database with chest X-ray images labeled with different lung pathologies. Its creation aimed to build artificial intelligence systems for training fourteen pulmonary pathologies. It was also designed to diagnose viral, bacterial pneumonia, and lung diseases [35]. Within the set of images in *CC-CXRI*, a database known as *CC-CXRI-P* has been chosen, which stores information regarding patients with pneumonia and normal controls, and has been used for the training of the pneumonia diagnostic model (see section 3.4.1).
- *MIMIC-CXR-JPG*: the *Medical Information Mart for Intensive Care* provides a database of chest radiographs labeled with fourteen different pulmonary pathologies. The creation of the database stems from another database known as *MIMIC-CXR*, which stores images in *DICOM* format, a prevalent format for storing X-ray images. *MIMIC-CXR-JPG* [36] holds its pictures in *jpg* format, which occupies less storage space than the *DICOM* format, which is why *MIMIC-CXR-JPG* was chosen for the development of the study. Each of the pathologies of the studies in the database contains four possible values, each with a corresponding interpretation:
 - o The pathology in the study is labeled with a "1" if it is present in the patient's imaging study [37].
 - o The pathology in the study is labeled with a "-1" if it is not present in the patient's imaging study [37].
 - o The pathology in the study is labeled with a "0" if it is undecided in the survey [37]. These images are not considered for training the models proposed in the development of the study since we do not know if the pathology is present.

- The pathology in the study is labeled as "*missing*" if it is not marked in the database. Image studies with "missing" labels have also been discarded for training the proposed models [37].

For the use of *MIMIC-CXR-JPG*, database access permissions have been required in compliance with the regulatory framework of the research (see section 5.3).

- *Kaggle*: *Kaggle* is a company of data scientists sharing machine learning models and databases through competitions [38]. Thanks to the platform, three other open-source and free-to-use databases have been collected:
 - *Covid19 PA Dataset*: this database stores 6939 images labeled with typical and pneumonia radiographs [39].
 - *NIH Chest X-Ray*: the database originates from another database known as the *National Institutes of Health (NIH)*, which has 112000 chest X-ray images of 30,000 different patients with different labeled pathologies [40].
 - *Chest-X-Ray (Pneumonia)*: the database contains lung X-ray images with labels of images with viral pneumonia, bacterial pneumonia, and normal controls [41].

3.3 Image preprocessing

The input information requires a series of modifications to improve image quality, eliminating redundant information for model diagnosis (noise) and standardizing image resolution. The different preprocessing techniques that have been applied are defined in temporal order:

1. *Otsu's threshold*: in many cases, chest X-rays are taken to minimize the possible radiation exposure to which the patient may be subjected. The systems that make this minimization possible are called *collimators* and reduce the amount of radiation that reaches patients by reducing the size of the image taken, causing the rest of the non-irradiated areas to have black borders, which are irrelevant information for the diagnostic model. An *otsu thresholding* technique is used to eliminate black edges, which is helpful for their erasure.
2. *Center Cropping*: consists of cropping the image taking the center of the image as a reference point.
3. *Image scaling* is a technique that standardizes the original image's resolution to the proposed one. The images have been scaled to 224 by 224 pixels by adjusting the input resolution to the input layer of the neural network architectures used for training the proposed models.

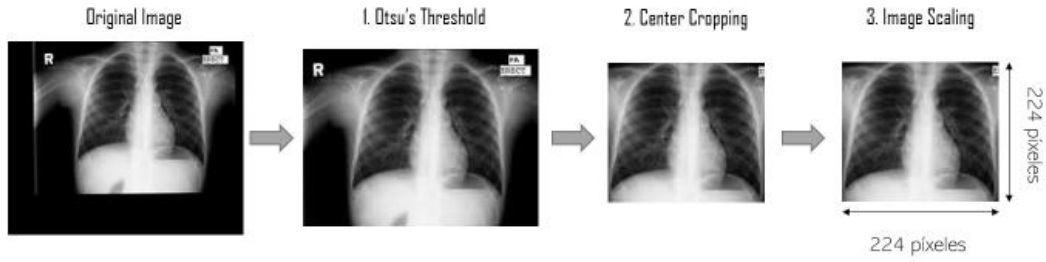


Fig. 15 Image preprocessing diagram.

3.4. Training of classification models

In this section, two main models have been proposed for identifying pulmonary pathologies. The first one is based on binary image classification to identify pneumonia in the image or detect it as usual. The second model proposes the detection of different lung pathologies. For this purpose, the *Torchxrayvision* models [26] have played a key role in training both models. These models are built on a tool specialized in designing deep learning systems known as *Pytorch* [42].

3.4.1 Detection of pneumonia

The procedure aims to use X-ray images as input to decide whether the patient has pneumonia or not. The image input comes from the China National Center for Bioinformation (*CC-CXRI-P*) database, which explicitly contains patients labeled with pneumonia and patients with normal controls. For training, 3004 pneumonia images and 2540 normal radiographs have been taken into account.

The convolutional neuronal network architecture proposed is a *DenseNet121* network architecture [20] previously trained on different X-ray imaging databases (*NIH*, *ChexPert*, *MIMIC IV*, *RSNA*, *PadChest*), as well as a model previously taught on the mixture of all the mentioned databases (*All*).

The architecture has been modified by adding an output neuron passed through an activation function known as *Sigmoid* that constrains the output value of the network to a number between zero and one. This output number represents the probability that the radiograph has pneumonia or not.

3.4.2 Identification of pulmonary pathologies

Identifying other pulmonary pathologies allows for a more detailed description of the state of the lung, providing radiologists with complex diagnostic assistance. Therefore, it is proposed to detect different lung pathologies from a lung X-ray.

The input information comes from the *MIMIC-CXR-JPG* database, which contains fourteen labels of possible pathologies to be identified in the image. The labels are divided into different types according to the severity of each label:

- Direct diseases: these are pathologies that directly express a pulmonary disease. This group includes pneumothorax, pleural effusion, other effusions, fracture, and pneumonia.
- Pulmonary lesions: are pathologies that can indirectly produce some disease based on lesions of the lung wall. These include consolidation, atelectasis, pulmonary edema, or a general lesion.
- Radiological findings: these are pathologies that may provoke direct lesions or disease. Among this group, cardiomegaly and elongated cardio mediastinum stand out.
- Finally, identifying a group of unrelated labels to any disease is worth noting. For example, normal radiographs or radiographs with electronic devices belong to this group.

A brief diagram of the different database tags can be seen in figure 16.

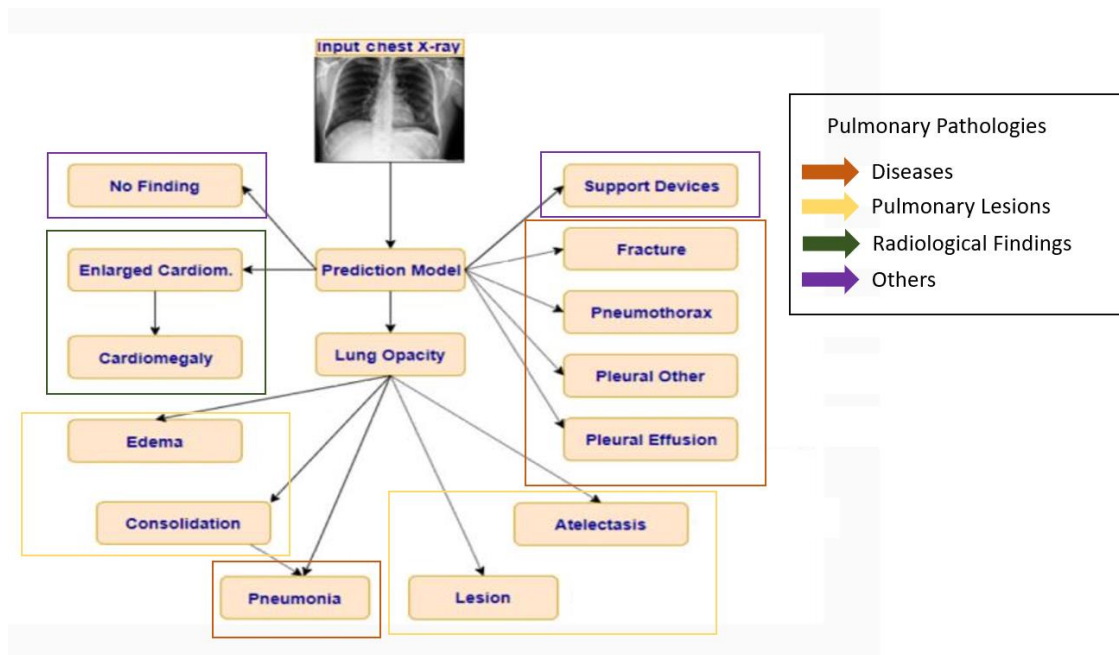


Fig. 16 Diagram of the different MIMIC-CXR-JPG lung pathologies [43].

Images from the MIMIC-CXR-JPG database have been used for the multiclassification training model. Within the set of images to train, we have defined the condition that each radiograph has at most one pathology to be identified. This condition is not proposed to predict future observations (which needs to identify one or more pathologies from an image in case there are any) but to force the model to distinguish the different pathologies and not to cause confusion between pathologies that are similar in the same study. Figure 17 shows the distribution of the number of observations per pathology used in the training set.

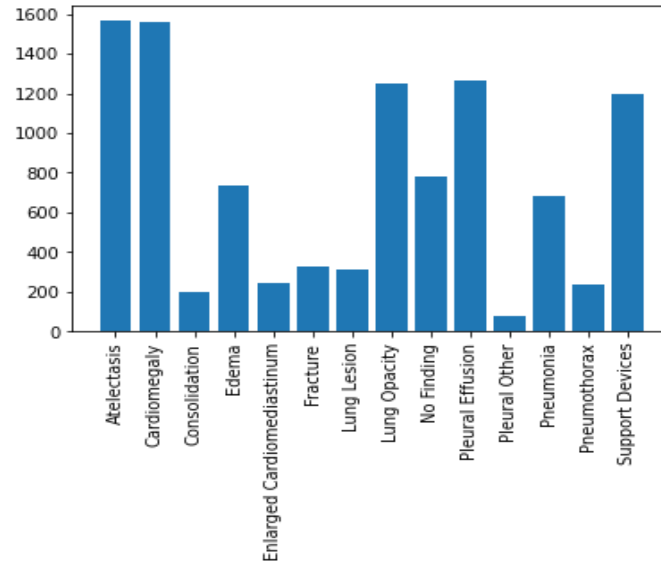


Fig. 17 Distribution of pathologies in the training set.

For training, a total of 10432 imaging radiographs with different lung pathologies with an unbalanced ratio of images per pathology are proposed.

The proposed convolutional network architecture is a *Densenet121* [20] with the parameters previously trained on the *NIH database*, applying transfer learning and adding 14 neurons to the output layer. Finally, a *Softmax* function [44] is used in the architecture to bound each output value to a number between zero and one. Each number represents the probability that the input image has one of the pathologies mentioned above.

3.5 Lung pathology interpretation systems

For the spatial interpretation of pulmonary pathologies in X-ray images, heat maps have been used to capture the areas that the artificial intelligence model considers relevant information to predict the pathology.

For the interpretation of the pneumonia classification model and the pulmonary pathology identification model, the different variations of the saliency map presented in section 2.4 have been proposed.

The attention offered by the saliency map depends on the diagnosis or prediction made by the model. For example, suppose the model detects pneumonia in the patient's image. In that case, a saliency map is proposed that activates the spatial zones of the image essential for diagnosing pneumonia and the same for diagnosing typical images. Each color of the saliency map determines the importance of each area of the radiograph for detecting the pathology. It is worth mentioning that different color palettes have been developed to generate a saliency map according to the visual taste of the clinician.

For a better understanding of the saliency maps performance, go to document sections 4.3.1.2 and 4.3.1.3, where the performance of the saliency map to detect pneumonia and standard control are shown, respectively, with the representation of the meaning of each color range.

3.6 Automatic diagnostic report generation

This section proposes reproducing a deep learning system that generates a diagnostic report of possible pulmonary pathologies from an X-ray image. It should be noted that no learning or training has been performed on the defined databases because it has been possible to get a neural network architecture with parameters already trained for the automatic generation of clinical reports. These architectures have been found in an open-source repository through *Github* [9].

The architecture of the model put into reproduction is made up of three distinct networks, each with a different task to generate a clinical report from a lung radiograph:

1. *Visual features*: the system first proposes a model previously trained on *Chexnet*, a *Densenet121* architecture network trained on X-ray images. The architecture is modified, changing the last linear layer of 14 neurons with a linear layer of 105 neurons. Each output neuron has a numerical value between zero and one [9].
2. *Semantic features*: on the other hand, the model labels are trained from a set of biomedical texts stored in the *MEDLINE/Pubmed database* to transform the biomedical words into matrices [9]. The label vector of length 105 is multiplied by the words from the biomedical texts and transformed into a vector matrix, obtaining an information matrix that caters to the essential semantic features of the image [9].
3. *Decoder*: finally, the matrix and partial visual features information are passed through a model previously trained on a reduced version of *GPT-2* known as *distilGPT2* with a network structure composed of attention mechanisms (a type of neural network) to generate the clinician report[9].

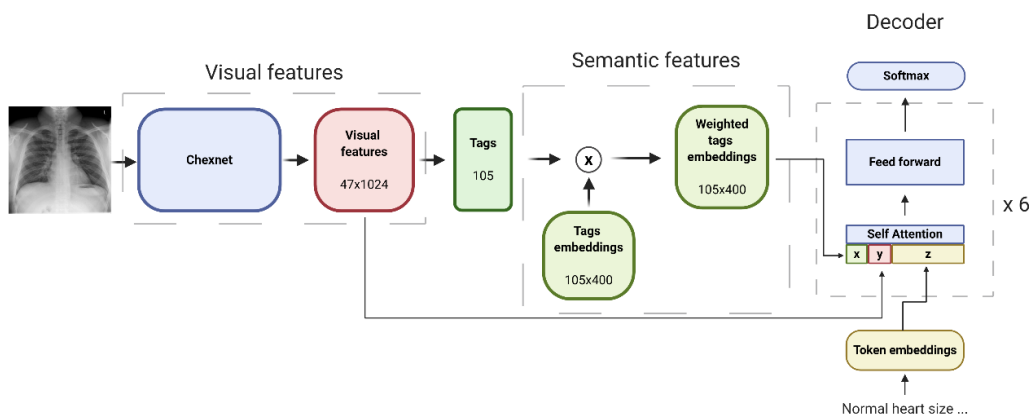


Fig. 18 Architecture of the diagnostic report generation model [9]

3.7 Web production

Once the three pathology diagnostic models were created, the development of the process was put on the *Streamlit* web platform [45] (see section 3.8), proposing an open-use page for physicians with a fast and functional aid for patient radiodiagnosis, translated into all the languages enabled by the Google Translator API. The platform's operating mechanism can enter a lung X-ray and display model predictions. The website has been sectioned into four main menus, each with different functionality.

1. Pneumonia identification: the functionality consists of predicting whether the patient has pneumonia or not. To do so, first, the likelihood of the diagnosis is displayed. Depending on the prediction made, the application shows an attention map for the predicted label (if the model predicts that the image is typical, an attention map displays the areas that the model considers relevant to predict the X-ray as usual and respectively with the pneumonia label). The page provides different parameters to be customized so that the physician can evaluate the model according to their taste. One can choose various trained models, apply several saliency maps, choose the color intensity of the saliency map, and restrict the saliency map information that the model considers most important or be more flexible with the different types of information it can provide. Finally, it is possible to choose and decide the color scale the user wants to visualize the saliency map to let the physician determine the color palette they prefer to visualize the color maps.
2. Identification of pulmonary pathologies: the functionality consists of predicting one or several possible pathologies to alarm the physician of possible diseases or thoracic lesions. For this purpose, threshold probability limits are established for each pathology defined. If the model detects that the probability of the pathology exceeds its respective threshold, the likelihood that the pathology exists is displayed, as well as the saliency map that identifies the areas that are relevant for detecting that possible pathology. The page incorporates several parameters for customization, like choosing the range of colors proposed in the care model, increasing or decreasing the intensity of the saliency map colors, and restricting the areas the model considers most important for predicting the pathology.
3. Automatic report generation: the purpose is to get a report of the X-ray image. For this purpose, the automated medical report template is reproduced. Subsequently, this report has an automatically created process to eliminate wrong symbols and words, repeated sentences, etc. Finally, the information is displayed in sentences delimited by dots for better understanding.
4. Diagnostic system for the clinician: the fourth menu's functionality encompasses all the diagnostic artificial intelligent models created to display the complete diagnosis in just one page.

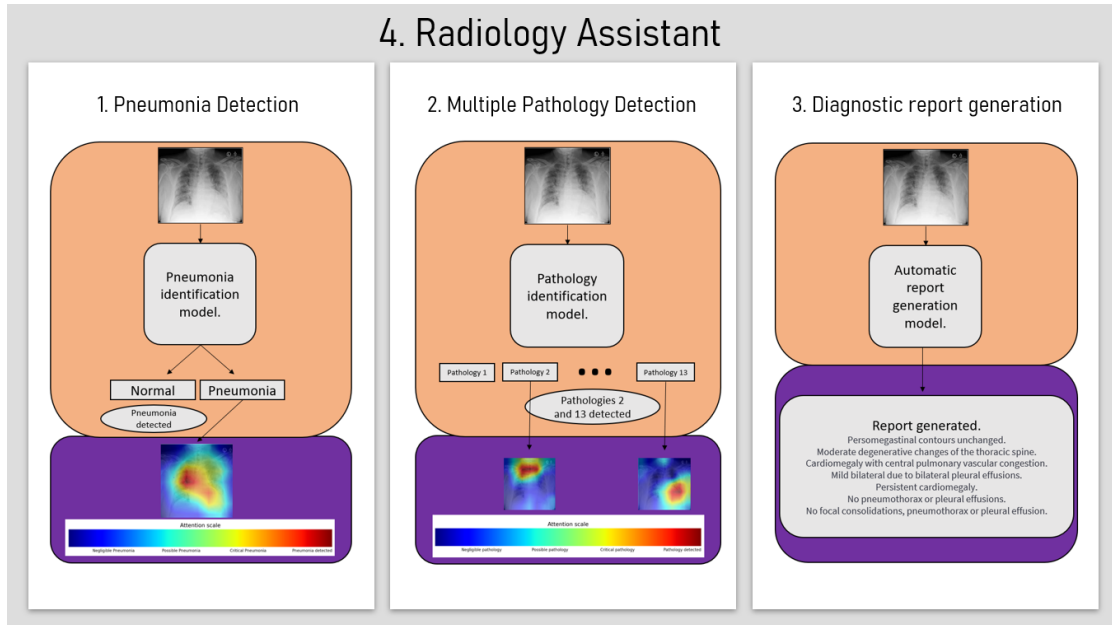


Fig. 19 Diagram of the application functionality.

3.8 Code repository

The organization of the content worked on during the research development has been structured in folders. The meaning and scope of each of the folders are mentioned below:

- *Data*: the different databases defined in section 3.2 have been stored in this folder.
- *Models*: the "models" folder has stored the parameters of the models trained for pneumonia detection, the multiple pathology classification models, and the models reproduced for the generation of the clinical report (see sections 3.4.1, 3.4.2, and 3.6, respectively).
- *Notebooks*: in this folder, it has been created the code for training the different models, the treatment of the databases, and the code that generates the saliency maps, mainly.
- *Src*: the folder stores the script's menus and model architectures of the web application.

Finally, in the root of the folder directory, there is a file called "app.py" which is responsible for initializing the contents of the application.

The entire folder structure of the web program has been uploaded on the Internet using the *Github* tool [32]. In the root of the directory, a file in *Markdown* format has been added, indicating the installation instructions for the web application. For more details of the repository, access the following link: <https://github.com/Rules99/Chest-X-Ray-with-Radiologist-AI>.

The *Github* repository created has been connected to the *Streamlit* virtual machine, putting the web platform on the Internet. To visit the web page and interact with the application, access the following link: <https://share.streamlit.io/rules99/chest-x-ray-with-radiologist-ai/main/app.py>.

4. RESULTS AND DISCUSSION OF THE STUDY

This section evaluates the project design results, analyzing the objectives achieved. An analysis of databases established can be appreciated. The performance of models for X-ray image diagnosis is discussed, as well as the performance of the proposed saliency maps comparing them with the diagnosis of a clinician specialist in chest radiography.

4.1 Database exploration and analysis

The search for possible databases has led us to manage the different databases according to the various existing labels, conducting an exploratory analysis to filter the essential and reliable pathologies. Some critical requirements have been proposed when using the convenient observations for the training of our models later on. The first requirement was based on the assumption that each radiograph had only one pathology to be diagnosed, so the model would train the classification of only one possible pathology. Secondly, in some databases such as *MIMIC-CXR-JPG* and *Chexpert*, there were labels associated with studies in which the pathology detection was unclear. These uncertain radiographs have been discarded, structuring a database with correctly labeled radiographs. Finally, the database with the different labeled pathologies has been analyzed through visualizations, defining a database with many observations to perform larger training models for the next steps. Therefore, it should be clarified that not all the database observations described for either of the two models have been trained (see sections 3.4.1 and 3.4.2).

4.1.1 Analysis of databases with different pathologies

A bar chart of the various radiographs obtained by each database with their respective labeled pathologies has been visualized to analyze these databases. The *CC-CXRI-P* and *NIH Chest X-Ray* contain positively or negatively defined pathologies. However, *MIMIC-CXR-JPG* and *Chexpert* store studies in which the pathologies have uncertain labels, giving rise to doubts about a positive or negative pathology identification. It can be seen in figure 20 that the most significant number of positively mentioned labels is the identification of electronic devices in *Chexpert*. A large number of normal controls are also found in *CC-CXRI-P*, *NIH Chest X-Ray*, and *MIMIC-CXR*, having an unbalanced database in the proportion of each pathology against normal radiographs.

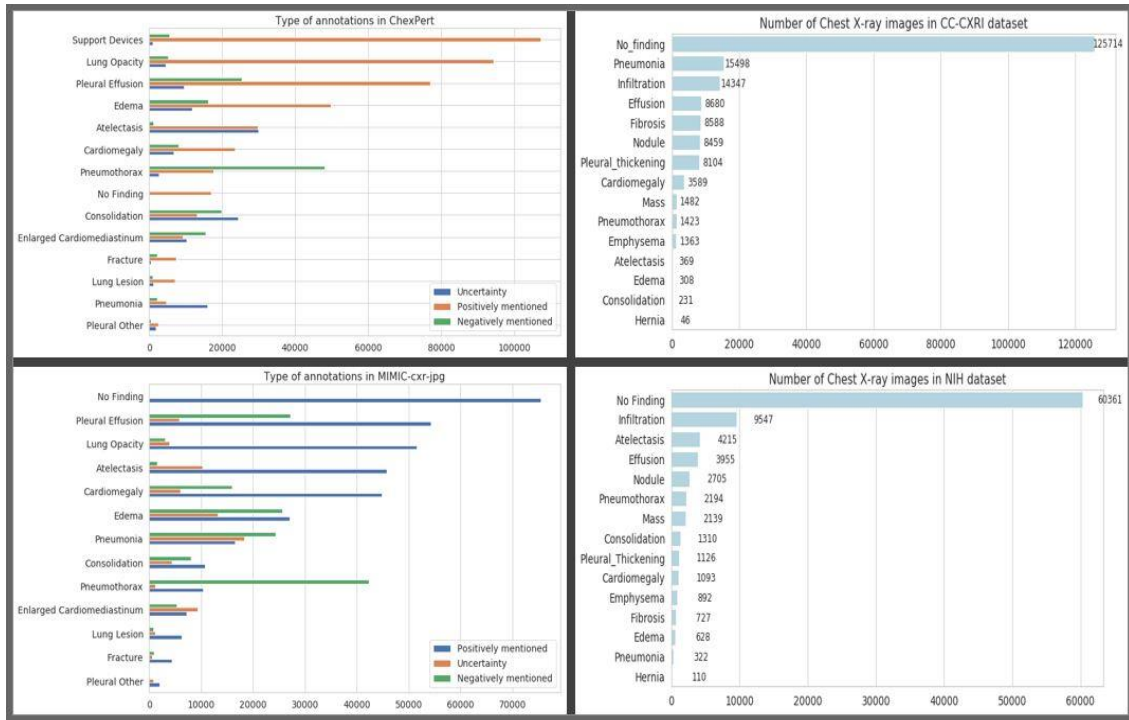


Fig. 20 Bar chart of pathologies defined in *Chexpert*, *CC-CXRI*, *NIH Chest X-Ray*, and *MIMIC-CXR-JPG*.

It is also worth noting in figure 21 how electronic devices and normal controls are easier to detect by radiologists because the uncertainty of identification is very low. Pneumonia is the pathology that has caused more doubts in diagnosis. Chexpert has more than 65 percent of images whose diagnoses are doubtful concerning the total number of images labeled as positive, negative, or uncertain. Respectively, *MIMIC-CXR-JPG* reaches more than 30 percent uncertainty, the second pathology with more uncertainty. The uncertainty of pneumonia diagnosis has led to training a model for diagnosing pneumonia.

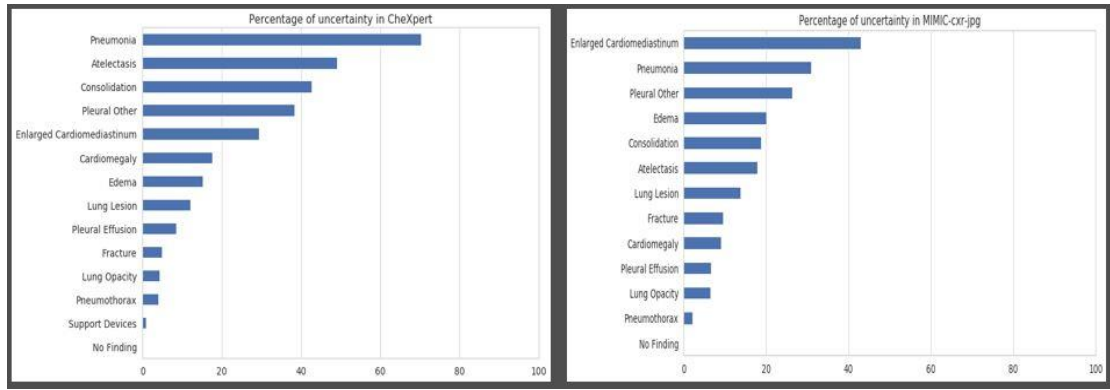


Fig. 21 Percentage of uncertainty in the different studies associated with *CheXpert* and *MIMIC-CXR-JPG*.

On the other hand, a Pearson correlation matrix of the different labels defined in the *CC-CXRI-P* and *CheXpert* databases has been visualized. The images without any pathology negatively correlate with tags that are diseases, radiological findings, or pulmonary diseases, meaning that typical images (*No Finding*) usually do not present signs of existing pathologies. The consolidation is not strongly correlated with pneumonia (8 percent in *CheXpert* and 7 percent in *CC-CXRI-P*). Although a consolidation may induce pneumonia (see figure 22), looking at the sample of both databases, it is not very common for this to happen. It should also be noted that lung lesions are not usually correlated with each other (*Atelectasis*, *Edema*, *Consolidation*), and taking into account the sample of the databases, they are generally independent of each other. A high positive correlation is obtained between pulmonary infiltration and pneumonia (*Infiltration*, *Pneumonia*), so it can mean statistically and only taking into account the observation of our sample, a possible direct relationship between the two.

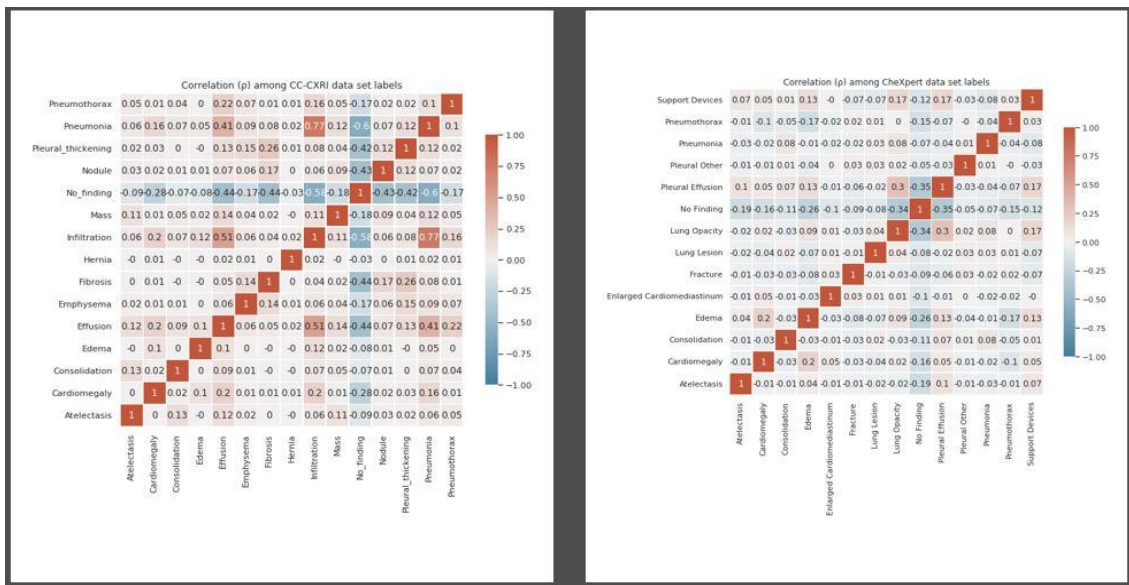


Fig. 22 Pearson correlation of labels defined in *CheXpert* and *CC-CXRI-P*.

4.1.2 Definition of a database for the identification of pneumonia

This section proposes a collection of images with pneumonia and normal controls with the different databases mentioned. This data collection has been defined to implement a pneumonia identification model that generalizes to other data distributions. The database has already been described. However, due to lack of time, the model trained on different databases will be implemented in future work.

Figure 23 shows the number of images obtained by each of the databases. Each database has been balanced to have approximately 50 percent normal and 50 percent of pneumonia images, getting a total of 69184 radiographs.

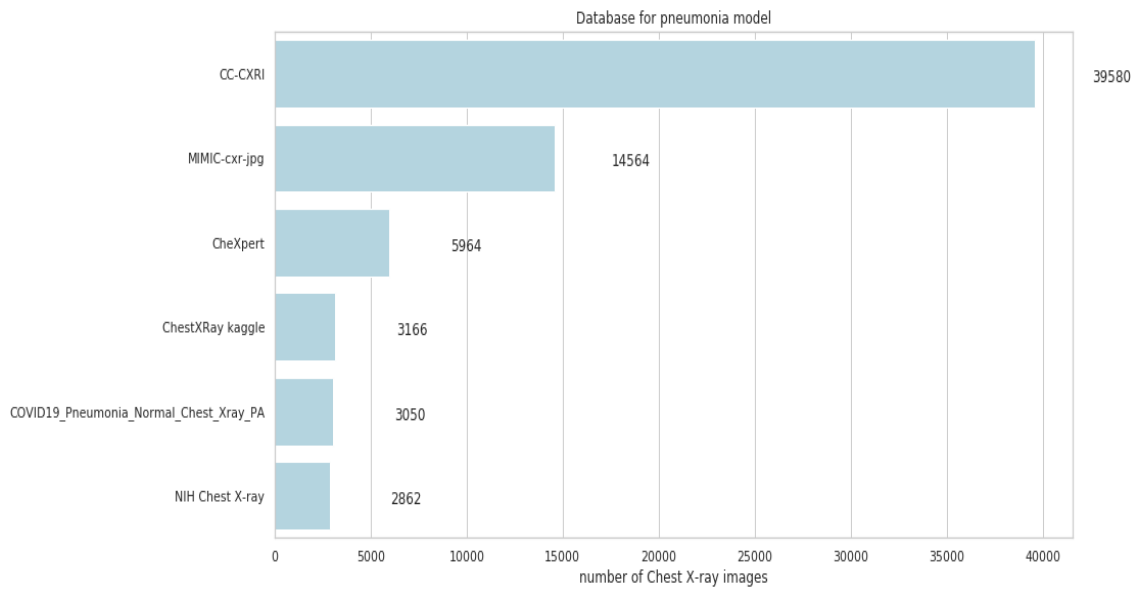


Fig. 23 Bar chart with images of pneumonia and normal controls in different databases.

4.2 Results of the training models

Within the results of the training model, it is essential to note that the results are not extrapolated to other databases, performing the evaluation with a partition of test data not seen by the models but belonging to the same database with which the models have been trained.

4.2.1 Evaluation of Pneumonia Screening Models

For the evaluation of the pneumonia detection models, 387 pneumonia images and 327 typical images were proposed as an unobserved test set from a database from the *China National Center for Bioinformation (CC-CXRI-P)*. Different evaluation metrics have been applied to observe the prediction results at a quantitative level for pneumonia and normal controls (*AUC, precision, recall, specificity, VPN, F1-score, accuracy*). The

decision probability threshold is set at 0.5 (if the likelihood is greater than the threshold, the image is identified as pneumonia; otherwise, it is detected as a normal radiograph).

Table 1. EVALUATION OF THE TRAINING MODEL WITH PREVIOUSLY TRAINED MODELS FROM DIFFERENT DATABASES.

Base model	AUC	accuracy	precision	recall	VPN	specificity	f1_score
All	0,9869	0,9482	0,9581	0,9457	0,9367	0,9511	0,9519
RSNA	0,9859	0,9412	0,9504	0,9406	0,9305	0,9419	0,9455
NIH	0,984	0,937	0,9407	0,9432	0,9325	0,9297	0,9419
PadChest	0,9853	0,9398	0,9433	0,9457	0,9356	0,9327	0,9445
Chexpert	0,9889	0,9524	0,938	0,9767	0,9711	0,9235	0,957
MIMIC IV	0,9771	0,9328	0,927	0,9509	0,9401	0,9113	0,9388

In Table 1, it is appreciated that the *Torchxrayvision* model previously trained on all databases (*All*) achieves the best result, obtaining an almost 96 percent chance of detecting the radiographs with pneumonia and a harmonized average precision and sensitiveness of 95 percent (*F1-Score*). As can be seen in figure 24, there is a deficient number of false positives and negatives, which leads to the conclusion of an impressive performance in the pneumonia identification model:

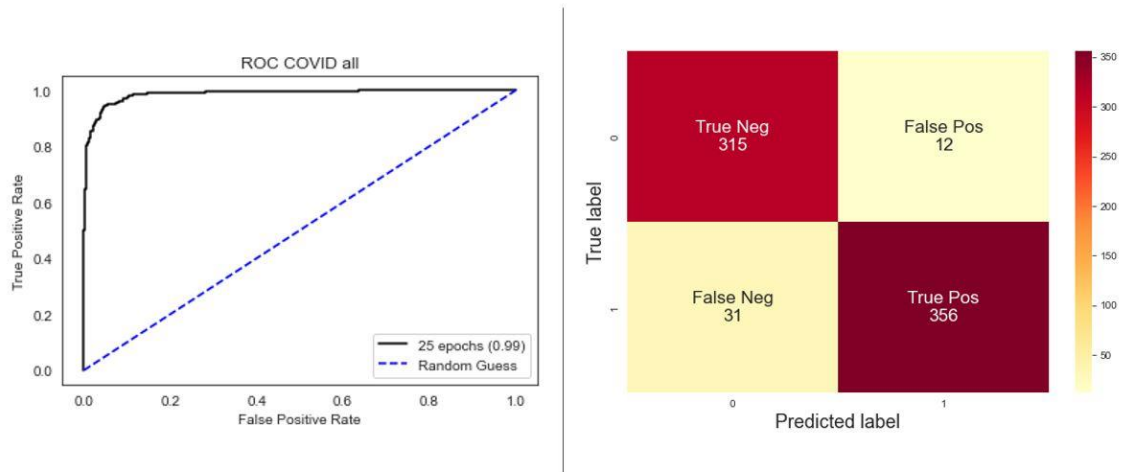


Fig. 24 Area under the curve and confusion matrix for the *CC-CXRI-P* test set. Results from the *All* base model.

4.2.2 Evaluation of the pathology detection model

A sample of test X-ray images from the *MIMIC-CXR-JPG* database was used to evaluate the pathology model. The number of tests per pathology was different due to the imbalance in the data distribution. However, an approximate balance between the percentage of training samples per label and test samples, respectively, has been applied (Fig. 25).

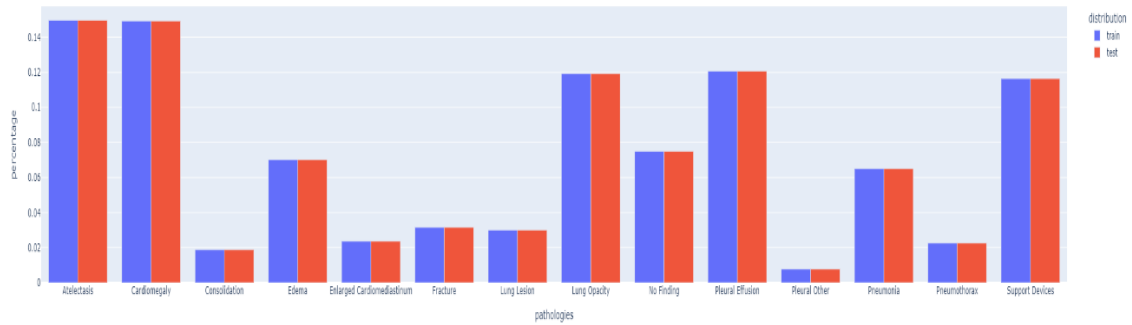


Fig. 25 Balance data partitioning in training and testing for the classification of multiple pathologies.

The same evaluation metrics of the pneumonia detection model have been used to check the model's performance.

Table 2 EVALUATION OF THE ADJUSTED MULTICLASSIFICATION LUNG DISEASES TRAINING MODEL

Pathology	AUC	accuracy	precision	recall	VPN	specificity	f1_score
Atelectasis	0,75	0,65	0,44	0,72	0,85	0,62	0,54
Cardiomegaly	0,77	0,65	0,44	0,82	0,89	0,57	0,57
Consolidation	0,69	0,82	0,10	0,5	0,98	0,83	0,17
Edema	0,87	0,87	0,53	0,51	0,93	0,93	0,52
Enlarged Cardiomediastinum	0,54	0,26	0,05	0,80	0,96	0,23	0,09
Fracture	0,65	0,78	0,14	0,52	0,97	0,80	0,23
Lung Injury	0,68	0,83	0,13	0,36	0,96	0,86	0,20
Pulmonary Opacity	0,65	0,47	0,27	0,82	0,87	0,37	0,42
Normal	0,83	0,82	0,42	0,69	0,94	0,84	0,52
Pleural effusion	0,86	0,80	0,55	0,75	0,91	0,81	0,63
Other Spills	0,73	0,80	0,02	0,17	0,99	0,80	0,02
Pneumonia	0,65	0,65	0,20	0,60	0,92	0,66	0,30
Supporting devices	0,76	0,63	0,36	0,71	0,87	0,60	0,48

It can be seen that the quantitative results are not striking, with the mean performance balance achieving a maximum harmonized mean of 0.57 for detecting cardiomegaly in the image. In addition, the results of the area under the curve of the pathologies are worse compared to the mean area under the curve of the *Torchxrayvision* models themselves (Fig. 26).

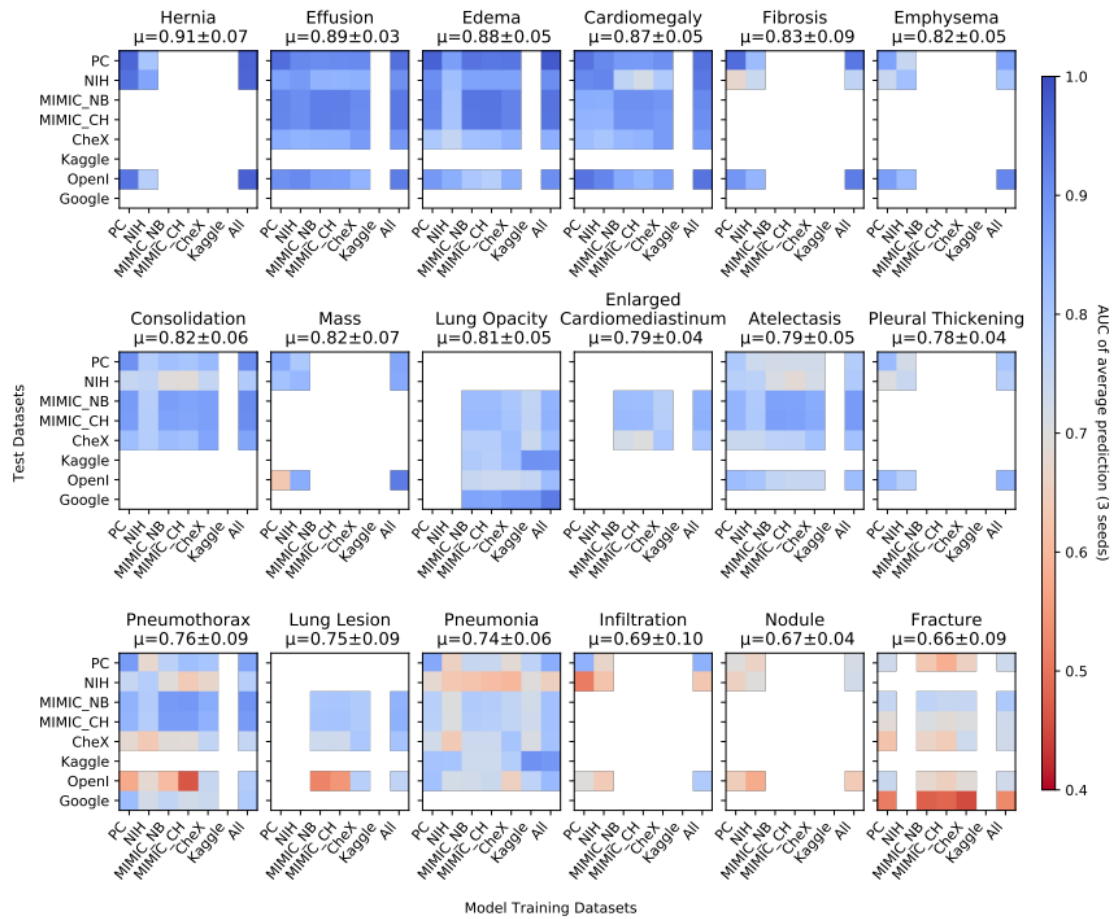


Fig. 26 Average AUC of the models previously trained in *Torchxrayvision* [46].

4.3 Evaluation of the intelligent models and interpretation of the saliency maps with the support of a radiologist

This section evaluates the reliability of heat maps for pneumonia detection and the pathologies' prediction models. For this purpose, Socorro Martín Barón [47], dedicated to thoracic radiology, with 18 years of experience in the sector, has helped us to contrast the model's results with her proposed diagnosis. A sample of tests not seen by the model is used to determine whether the model generates the heat maps correctly.

4.3.1 Pneumonia screening

For the evaluation of pneumonia, six images from the test sample of the *CC-CXRI-P database* with two normal radiographs and six pneumonia radiographs have been used, contrasting the radiologist's opinion with the performance of the model and the saliency maps. For heat map generation, the *jet* color scale is applied where red is the most relevant area for the model to interpret a specific pneumonia diagnosis.

It should be noted in the following figures (figure 27, figure 28, figure 29, figure 30, figure 31, figure 32) that the left image is the original image, and the spatial site of the image where the physician locates the pneumonia is shown with a red circle. The physician interprets the typical areas with a green circle if the X-ray is normal. The heat map generated by the model is shown on the right side. Below the two images is shown the *jet* color scale intuitively indicating the meaning of each of the color ranges representing the attention paid by the model to pneumonia or usual areas, depending on the diagnosis made by the model. It is also worth mentioning that both the model and the physician are unaware of the value of the labeled radiograph. Still, it is explicitly mentioned in each report to compare the actual against the model and the radiologist's diagnosis.

4.3.1.1 Patient 1

Patient one contains a diagnosis of pneumonia, so the model performance is expected to make a prediction identifying the radiograph as pneumonia. According to the radiologist, the pneumonia is located in the patient's left base. The heat map points to the same area the physician indicates, in addition to detecting that the plaque has pneumonia.

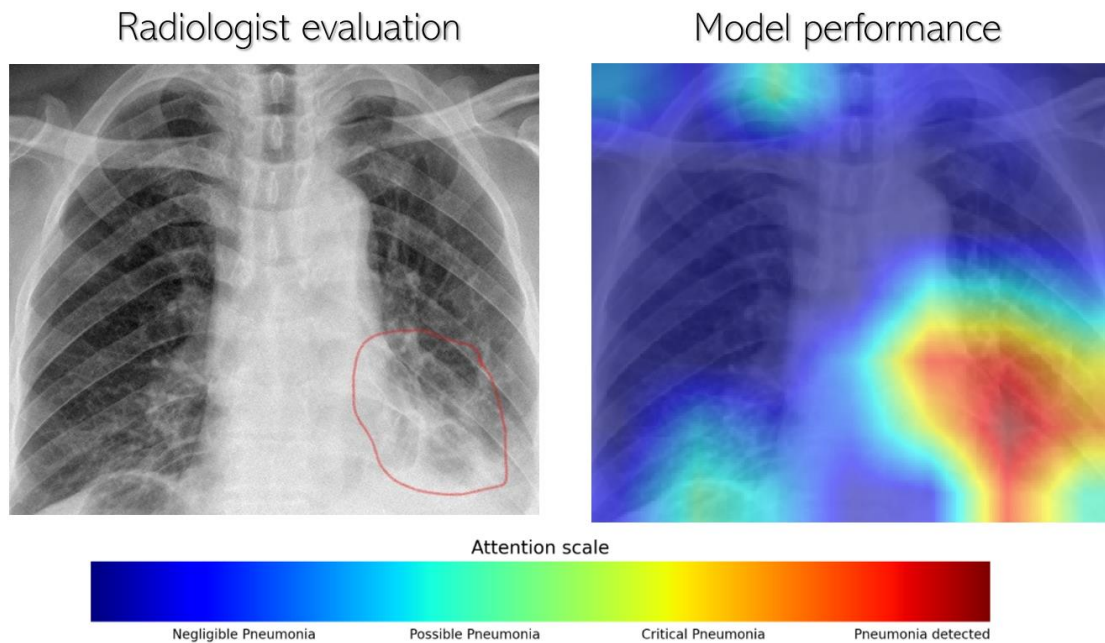


Fig. 27 Pneumonia Diagnosis (patient 1) radiologist and heat map evaluation.

4.3.1.2 Patient 2

Patient two has a pneumonia diagnosis, so the model is expected to detect pneumonia on the radiograph. According to the radiologist, there is pneumonia on the patient's right side, which the heat map detects. However, no pneumonia is seen in the left central part of the patient, which the model points to as a possible affected area. The model's diagnosis predicts that the radiograph has pneumonia.

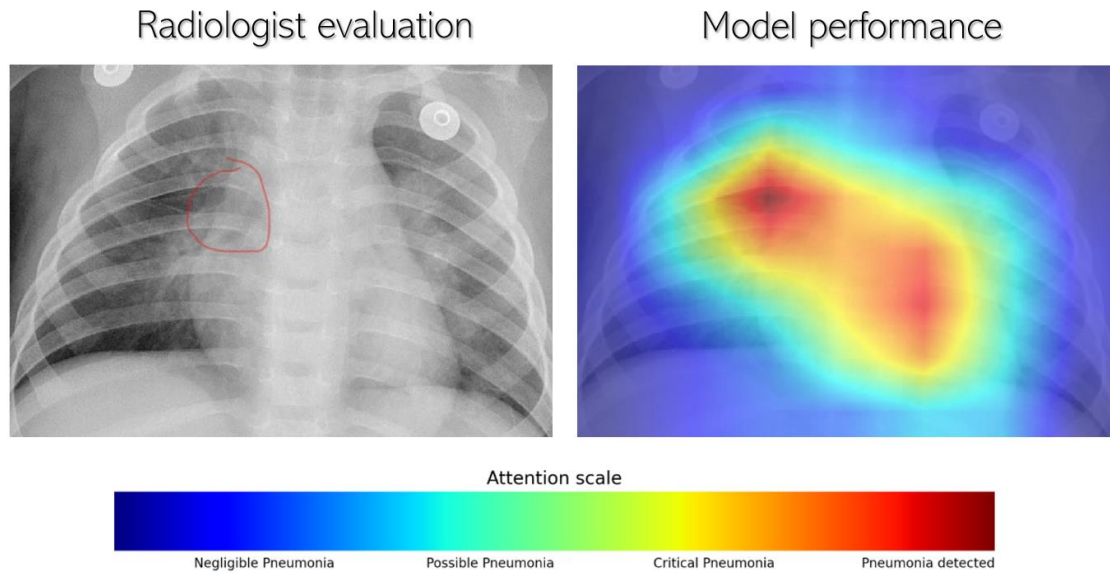


Fig. 28 Pneumonia Diagnosis (patient 2) radiologist and heat map evaluation.

4.3.1.3 Patient 3

Patient three has a normal control X-ray, so the expected prediction is that the diagnosis will be normal. The exciting thing about the diagnosis is that the model is expected to look at relevant information to diagnose the radiograph as normal. According to the radiologist, the X-ray is free of disease, and the heat map captures the cardiac silhouette within normal limits.

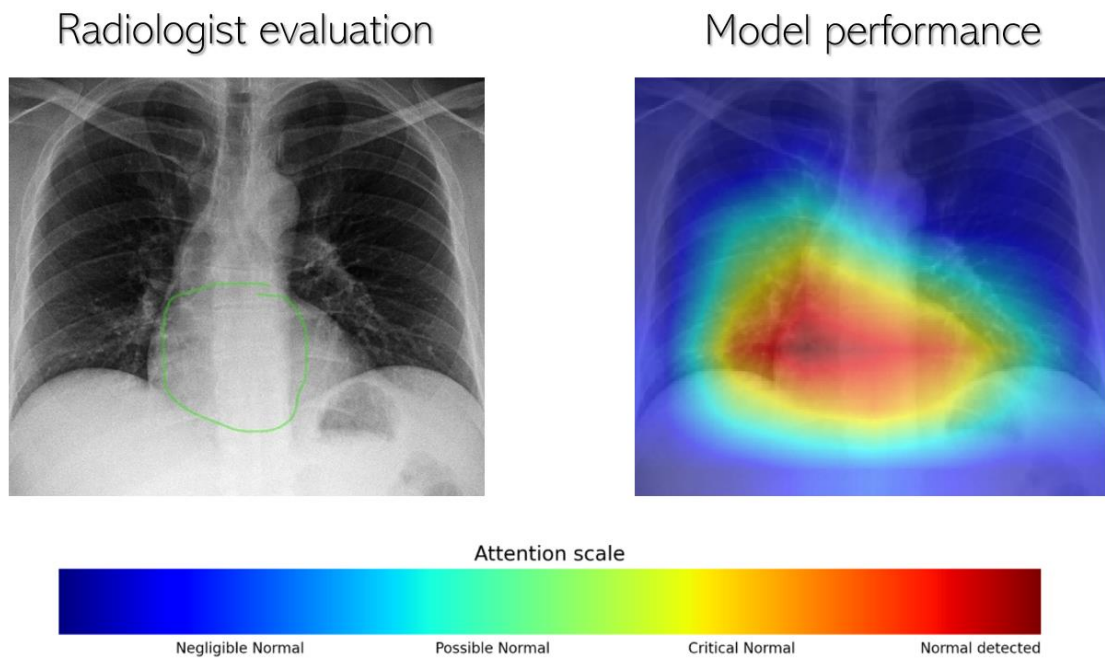


Fig. 29 Normal diagnosis (patient 3) radiologist and heat map evaluation.

4.3.1.4 Patient 4

Patient four's image is labeled as a normal control on his radiograph, so the expected prediction is that the diagnosis is normal. It can be seen again that the model points to the cardiac silhouette to predict that the radiograph is normal. According to the radiologist, the image is within normal limits, and she again mentions that the model points to the heart silhouette.

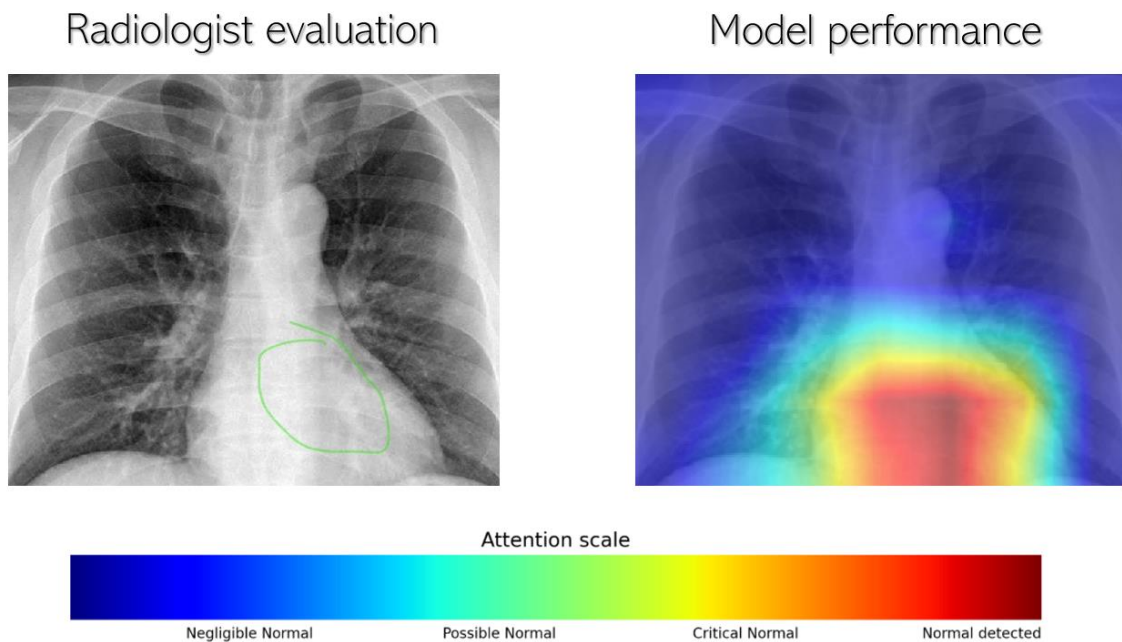


Fig. 30 Normal diagnosis (patient 4) physician and heat map evaluation.

4.3.1.5 Patient 5

The image of patient five is labeled with pneumonia on his radiograph. According to the radiologist, pneumonia is identified in the patient's right upper lobe. The heat map results approximate and matches the location given by the radiologist, determining pneumonia on the image.

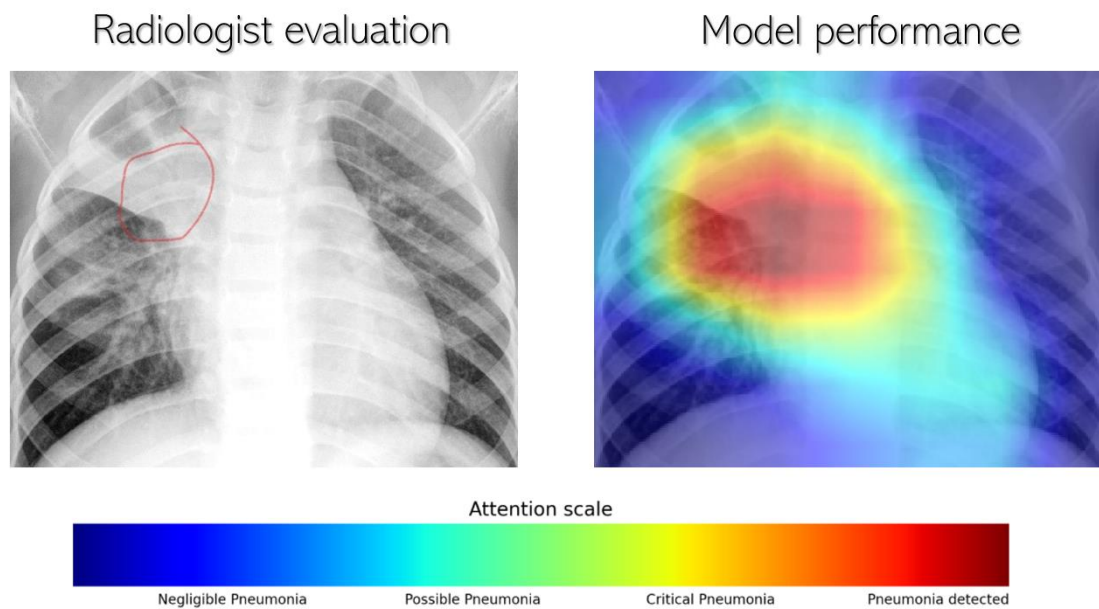


Fig. 31 Pneumonia diagnosis (patient 5) evaluation of the model and heat map.

4.3.1.6 Patient 6

Patient six's image is labeled with pneumonia on his x-ray. According to the radiologist, the image is difficult to diagnose as the radiograph is not seen well. Despite the difficulty of diagnosis, he does not rule out pneumonia on the patient's right side due to the increased image density of the right lung compared to the patient's left lung. The heat map attends to the right lobe of the lung primarily. Although the opinions of the model and the physician are close, there is no clear evidence of pneumonia considering the visibility of the radiograph.

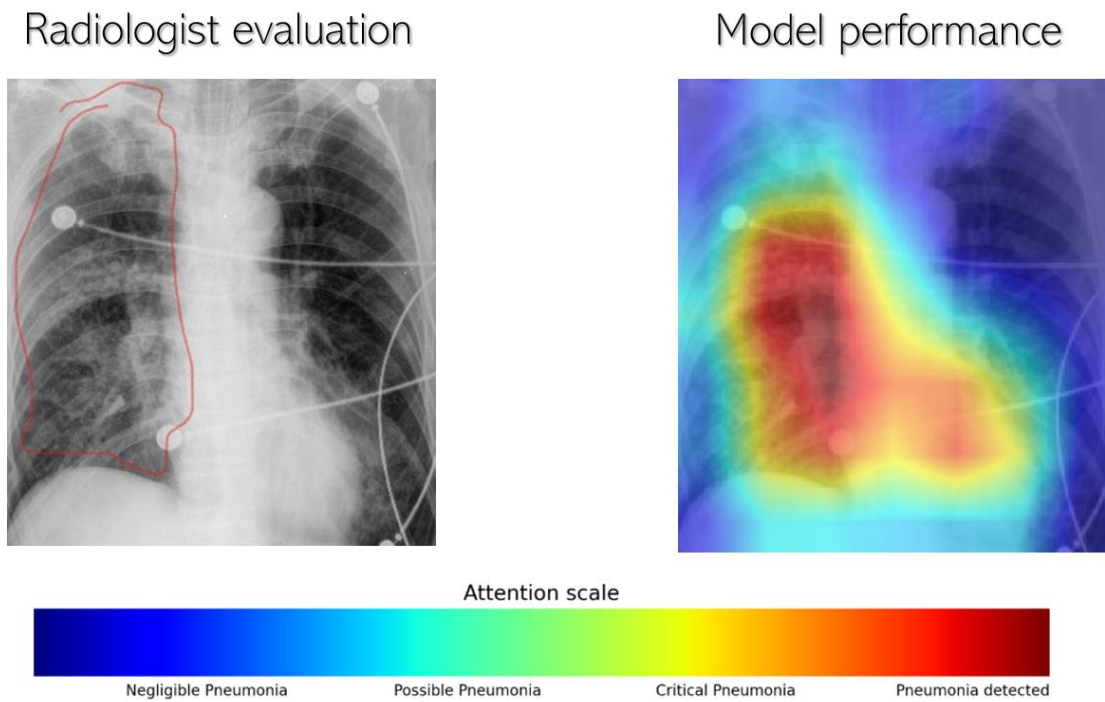


Fig. 32 Pneumonia diagnosis (patient 6) physician and heat map assessment.

The diagnostic model was agreed in the six experiments with the clinician in diagnosing the radiograph and the localization of the pneumonic areas of the lung. However, the results with only six experiments are not extrapolated to indicate that the model generalizes well with images from other databases or distributions and even the rest of the test images from the *CC-CXRI-P* database.

4.3.2 Detection of pathologies

For the evaluation of pathology detection, five *MIMIC-CXR-JPG* images have been considered, of which two are labeled with direct diseases, two are labeled with lung lesions, and one is a radiological finding. Initially, each of these samples has only one labeled disease. However, the expert radiologist has located more than one pathology in most of the evaluated samples, so it can be assumed that the images are not correctly labeled.

The evaluation considers the aforementioned label, the detection of possible pathologies by the radiologist, and the detection of the model. Finally, whether the saliency map locates the pathologies where the radiologist detects them on the radiograph or cannot see the pathology is considered.

4.3.2.1 Diseases

Among the diseases mentioned in figure 16 of the document, a pneumothorax and a fracture are considered.

4.3.2.1.1 Patient 1 (Pneumothorax)

The radiologist and the model locate the pneumothorax, labeled in the database. Besides, the clinician finds atelectasis on the patient's left side and multiple pathology model alarms for atelectasis without correctly locating the disease on the saliency map. The model detects cardiomegaly that the physician does not discover, nor is it labeled in the study.

Table 3 EVALUATION OF PATIENT 1 WITH PNEUMOTORAX LABELED. EVALUATION OF THE RADIOLOGIST AND THE MODEL.

Diagnosis of the model	Pathology statement	Radiologist evaluation	Pathology localization
Pneumothorax	Yes	Small right upper right pneumothorax	Yes
Cardiomegaly	No	No	No
Atelectasis	No	A little bit of atelectasis	No

4.3.2.2.1.2 Patient 2 (Fracture)

The radiologist and the model locate the labeled fracture in the upper left of the observer on the saliency map. However, the model detects an enlarged cardiac mediastinum that is not marked and that the physician fails to locate, as well as alarms that the radiograph is within normal limits when it contains a fracture.

Table 4 EVALUATION OF PATIENT 2 WITH FRACTURE LABELED. RADIOLOGIST AND MODEL EVALUATION.

Diagnosis of the model	Pathology statement	Radiologist evaluation	Pathology localization
Enlarged Cardiac Mediastinum	No	No enlarged cardiac mediastinum	No
Fracture	Yes	Fracture of the upper left side of the observer	Yes
Normal	No	No	No

4.3.2.2 Pulmonary Lesions

The proposed lung lesions are two samples, one with consolidation and the other with atelectasis.

4.3.2.2.1 Patient 3 (Consolidation)

The model and the radiologist diagnose the labeled consolidation, however, the pathology is not located on the saliency map. On the other hand, the image does not have a labeled atelectasis. Still, the radiologist finds atelectasis in both parts of the lung, as well as the model correctly detects atelectasis in the saliency map. The model also correctly localizes pneumonia coinciding with the radiologist's assessment without being labeled pneumonia in the database. Other pathologies not marked, but the model and the expert radiologist detect are pulmonary opacity, pleural effusion, and support devices without locating the pathologies in the saliency map. Finally, the model alarms for a possible enlarged cardiac mediastinum that is not labeled and does not coincide with the radiologist's opinion.

Table 5 EVALUATION OF PATIENT 3 WITH CONSOLIDATION LABELED. RADIOLOGIST AND MODEL EVALUATION.

Diagnosis of the model	Pathology statement	Radiologist evaluation	Pathology localization
Consolidation	Yes	Consolidation detected	No
Atelectasis	No	Possible atelectasis on the left and right sides.	Yes
Enlarged Cardiac Mediastinum	No	No enlarged cardiomeastinum	No
Pulmonary Opacity	No	Pulmonary opacity was detected in the areas of consolidation.	No
Pleural effusion	No	Pleural effusion detected	No
Pneumonia	No	Pneumonia detected	Yes
Support Devices	No	Support devices detected	No

4.3.2.2.2 Patient 4 (Atelectasis)

The model diagnosed and detected the atelectasis labeled in the database, coinciding with the radiologist's opinion. Other pathologies not labeled and yet interpreted by the model and radiologist are pulmonary opacity and support devices without correctly localized by the saliency map. The model diagnoses a possible enlarged cardiac mediastinum and a pleural effusion that are non-existent in the radiologist's opinion.

Table 6 EVALUATION OF PATIENT 4 WITH ATELECTASIS LABELED. EVALUATION OF THE RADIOLOGIST AND THE MODEL.

Diagnosis of the model	Pathology statement	Radiologist evaluation	Pathology localization
Atelectasis	Yes	Captured atelectasis	Yes
Cardiomegaly	No	No cardiomegaly	No
Enlarged Cardiac Mediastinum	No	No	No
Pulmonary Opacity	No	Pulmonary opacity was detected in the area of atelectasis, the right side of the observer.	No
Pleural effusion	No	No-spill	No
Support devices	No	Yes	No

4.3.2.3 Radiological Findings

A cardiomegaly sample was taken into account in the radiological findings section.

4.3.2.3.1 Patient 5 (Cardiomegaly)

Patient five's labeled cardiomegaly is diagnosed by the radiologist and the model, failing to localize the pathology on the saliency map.

Table 7 EVALUATION OF PATIENT 5 WITH CARDIOMEGALY LABELED. EVALUATION OF THE RADIOLOGIST AND THE MODEL.

Diagnosis of the model	Pathology statement	Radiologist evaluation	Pathology localization
Cardiomegaly	Yes	Cardiomegaly detected	No

It can be observed in general terms that of the five images evaluated, many pathologies are not correctly labeled, the model being able to alarm unlabeled pathologies and coinciding with the radiologist's opinion. However, a hasty diagnosis by the model can be appreciated in detecting an enlarged cardiac mediastinum, which the expert radiologist does not observe in any of the experiments in which this pathology is diagnosed. The performance of the saliency map in proportion to the labeled pathologies is engaging, locating four of five pathologies in the image and diagnosing the five labeled pathologies.

4.3.3 Automatic report generation

For the evaluation of the automatic report, we have taken the same patient test from section 4.1, where pneumonia detection is evaluated. It should not be forgotten that the report generation model has not had transfer learning applied to it and is a reproduction based on open source code from *Github*.

To know the system's performance, the radiologist briefly described the model and compared the coherence of the sentences with the possible affected areas of the image and the paradoxes reported by the model. In each of the following sections, the physician's textual description is provided with the generated report and a table with the total number of correctly and incorrectly reported sentences. Any sentence that includes assistance to the patient is scored with a one in case that helps the clinician and a minus one in case it could be a result that drastically casts doubt on the physician or is an absolute contradiction concerning the diagnosis that would hypothetically be expected. The inconsequential sentences do not provide us with information; therefore, their score is zero. Insignificant sentences will be discarded, so correct and incorrectly reported sentences will be contributed. Thus the final score of each report will obtain a result between -1 and 1, indicating that all sentences are incorrectly or correctly generated, respectively.

It should be mentioned that the spatial description of the model is made concerning the clinician's position and not the patient's position. For example, if the model reports a "lung lesion on the left side", it means that it finds the left side a lung lesion taking into account the clinician's observation; therefore, the lesion would be located in the patient's right lung.

4.3.3.1 Report 1

- Radiologist's description: "opacity in the right lung base with minimal atelectasis, compatible with pneumonia. "

Table 8 EVALUATION OF THE REPORT 1. PERFORMANCE OF THE RADIOLOGIST AND THE CLINICAL REPORT MODEL.

MODEL	RADIOLOGIST	EVALUATION
The lungs are clear.	The patient's left lung has an opacity.	-1
There is a bottleneck of the right costophrenic xxxx.	The report is not understood.	0
The heart is of normal size.	Yes.	1
No pneumothorax.	Yes.	1
No focalconsolidation is seen.	Focal consolidation is seen in the left base of the patient.	-1
The mediastinum is not noticeable.	Yes.	1
Lungs are hypoinflated with renewed costophrenic xxxx.	It is not understood.	0
There is flaccidity on the right.	It is not understood.	0
Bibasilar atelectasis.	Yes	1
TOTAL		2/6

An apparent contradiction can be observed in the first sentence of the model, as the lung contains an opacity when the model generates a sentence describing that the lungs are clear, assuming that the radiograph is within normal limits for lung disease. The same occurs in the fifth sentence mentioned above, in which the model does not appreciate focal consolidation when the physician detects consolidation. The rest of the sentences are redundant or describe typical lung characteristics.

4.3.3.2 Report 2

- Radiologist's description: "ill-defined right retrocardiac opacity compatible with pneumonia. "

Table 9 EVALUATION OF THE REPORT 2. PERFORMANCE OF THE RADIOLOGIST AND THE CLINICAL REPORT MODEL.

MODEL	RADIOLOGIST	EVALUATION
Mild elevation of the left hemidiaphragm.	There is an elevation of the left diaphragm from the observer.	1
No acute radiographic cardiopulmonary process.	No. There is an acute cardiopulmonary process due to pneumonia.	-1
No free subdiaphragmatic air.	Yes.	1
Not free.	It is not understood.	0
The cardiac silhouette is within normal limits for size and contour.	Yes.	1
The lungs are normally inflated without evidence of focal airspace disease, pleural effusion, or pneumothorax.	Partially: the lungs are inflated, however, there is evidence of airspace disease. Indeed, there is no effusion or pneumothorax.	0
No free xxx.	It is not understood.	0
TOTAL		2/4

The model detects a mild elevation in the left hemidiaphragm coinciding with the radiologist's description and the radiologist's third and fifth sentence of the evaluation table. However, an apparent contradiction of the model's description with the radiologist can be observed in sentence two, implying that the radiograph does not have pneumonia when there is pneumonia. In sentence six, a hit and a miss is followed in the same sentence describing the radiologist scoring neutrally.

4.3.3.3 Report 3

- Radiologist's description: "There are no opacities in the pulmonary parenchyma. There are no findings of pleural effusion. Radiologically typical cardiac silhouette. "

Table 10 EVALUATION OF REPORT 3. PERFORMANCE OF THE RADIOLOGIST AND THE CLINICAL REPORT MODEL.

MODEL	RADIOLOGIST	EVALUATION
There are no acute bone findings.	Yes.	1
No pneumothorax.	Yes.	1
There is a mild elevation of the left hemidiaphragm. There are no acute surgical clips.	No. The diaphragms are at the same level.	-1
The lungs are clear of focal airspace disease, pneumothorax, or pleural effusion.	Not well understood, but interpreted to be true. Translation problem.	1
The cardiac silhouette and pulmonary vasculature are within normal size limits.	Not well understood, but interpreted to be true. Translation problem.	1
There is the mild elevation of the left hemidiaphragm.	No. Reiteration of the previous sentence.	-1
No acute cardiopulmonary findings.	Yes.	1
There are.	It is not understood.	0
TOTAL		3/7

Five hits and two contradictions are observed compared to the radiologist's description. The hits are mainly based on the normal radiographic description. However, there is no elevation of the left hemidiaphragm that the model describes in his report, reiterating the same sentence twice.

4.3.3.4 Report 4

- Radiologist's description: "There are no opacities in the pulmonary parenchyma. There are no findings of pleural effusion. Radiologically typical cardiac silhouette. "

Table 11 EVALUATION OF THE REPORT 4. PERFORMANCE OF THE RADIOLOGIST AND THE CLINICAL REPORT MODEL.

MODEL	RADIOLOGIST	EVALUATION
No acute radiographic cardiopulmonary process.	Yes.	1
No acute bone abnormality.	Yes.	1
Mild degenerative changes of the thoracic spine.	No degenerative changes are seen.	-1
Without intrathoracic trachea.	Serious contradiction; the trachea is intrathoracic.	-1
No intrathoracic abnormality. No intrathoracic abnormalities.	Yes	1
The cardiac silhouette is within normal limits for size and contour.	Yes	1
The lungs are generally inflated without evidence of focal airspace disease, pleural effusion, or pneumothorax.	Yes	1
No.	It is not understood	0
Bone structures are within normal limits for the patient's age.	Yes	1
TOTAL		4/8

The model generates a serious literal contradiction in sentence four, causing a drastic confusion in the diagnosis and warning of non-existent degenerative changes. The rest of the sentences describe a pathology within normal limits, coinciding with the physician's opinion.

4.3.3.5 Report 5

- Radiologist's description: "Pulmonary opacity in the right upper lobe, with signs of consolidation and mild volume loss. Compatible with pneumonia. "

Table 12 EVALUATION OF THE REPORT 5. PERFORMANCE OF THE RADIOLOGIST AND THE CLINICAL REPORT MODEL.

MODEL	RADIOLOGIST	EVALUATION
Stemotomy xxxx sternotomy xxxx	It is not understood.	0
The lungs are free of focal airspace disease.	Pulmonary opacity is seen, the lungs are not free of disease.	-1
Stemotomy sutures and shunt markers are.	It is not understood.	0
Heart size and pulmonary vascularization appear within normal limits.	Vascularization is normal, but the cardiac silhouette is slightly enlarged.	0
No pleural effusion or pneumothorax is identified.	Yes.	1
The xxxx xxxx sternotomy sutures and the bypass graft are identified.	No sutures or grafts are identified.	-1
Stemotomy sutures and shunt graft markers are identified.	No sutures or grafts are identified.	-1
Interval development of left lower lobe airspace disease.	It is not understood	0
Sternotomy sutures xxxx and shunt graft markers are identified.	No	-1
TOTAL		-3/5

A contradictory result can be observed in general terms, resulting in evaluating a report with a poor diagnosis. The contradictions are profound because the model reports a description of a normal thorax when there is a pulmonary opacity described by the physician (sentence 2). Conversely, the model repeatedly says non-existent lung problems (sentences 6, 7, and 9). The model does not identify any pleural effusion coinciding with the physician's opinion.

4.3.3.6 Report 6

- Description: "asymmetry in the density of both hemithorax with significant attenuation of the patient's right hemithorax. It may correspond to pneumonia or, if the patient was lying down, to a pleural effusion. "

Table 13 EVALUATION OF REPORT 6. PERFORMANCE OF THE RADIOLOGIST AND THE CLINICAL REPORT TEMPLATE.

MODEL	RADIOLOGIST	EVALUATION
Xxxx left effusion is present.	Pleural effusion is possible. Language translation problem.	0
The lungs are free of focal airspace disease.	Probably not.	0
Heart size and pulmonary vascularization appear within normal limits.	Yes.	1
Fracture deformity fracture of the proximal right humerus.	The humerus is not visible.	-1
The left effusion is evident. Xxxx left effusion.	Pleural effusion is possible. Language translation problem.	0
No pneumothorax is identified.	Yes	1
Effusion is observed on the left.	It is possible.	0
No pneumothorax or pleural effusion was observed.	No pneumothorax is seen, but there is likely an effusion.	0
TOTAL		1/3

The model correctly describes the reports that diagnose the thorax within normal limits. However, it can be seen in sentence four that it reports the visualization of the humerus when it is impossible to locate since it does not appear within the limits of the image.

In conclusion to the six reports, it can be considered that the model contains profound contradictions, which may cause doubts for non-expert clinicians in radiology. However, the model could help generate judgments concerning the description of chest radiographs that a priori are known to be normal, according to the samples seen in reports 3 and 4, in which the physician's opinion coincides quite well with that of the model. It should be noted that only six images have been evaluated and the discussion is described according to the statement of the samples only seen.

5. STUDY MANAGEMENT

This chapter discusses all the operational resources that have been applied for the development of the study. The socioeconomic impact that the platform in production may have on the field of health sciences is discussed, and the different legislative regulations that have been applied to the regulatory framework of the project.

5.1 Socioeconomic impact

As mentioned, one of the study's main objectives was to establish a web-based ecosystem in production to serve as an accessible tool for diagnostic chest radiography imaging. Currently, the application is in a state of research and development, so its use for automatic diagnosis of pulmonary manifestations without clinician review or supervision is not recommended.

However, the development may cause technological, social, and economic impacts on the health sciences applied to technology, specifically medicine. The potential benefits the full effect of the study may cause are mentioned.

- Assistance and support for pathology detection and reporting, ensuring a universal and open-source application for hospitals around the world, providing a tool for those countries where socio-economic conditions are worse in terms of resources, an insufficient number of doctors, etc.
- Acceleration of chest X rays diagnostic processes, proposing an intelligent system that aids in detecting diseases and normal controls.
- Creation of a scientific community to improve possible bugs and errors of the application, propose the code in an open-source so that improvements are possible, and establish interactive workflows between the organizers (us) and the collaborators of the work repository.
- Computed tomography (scanner) generates a more costly image in economic terms. The time to obtain the image takes longer than in an X-ray, leading to lower salary costs for hospitals and simultaneously speeding up the radio diagnostic process.
- Radiodiagnosis executed by two bits of intelligence, the artificial intelligence machine and the human one, allows contrasting opinions and enables radiology experts to label predictions poorly made by the models put into production, building a database marked by physicians with the assistance of artificial intelligence.

5.2 Budget

The budget of the work has been based mainly on the operational environment used, the hardware and software tools applied for the project development, and the cost of the work performed.

5.2.1 Labor costs

We have implemented the work of a Data Engineer for the definition, treatment, and preprocessing of data, a Data Scientist Engineer for the training and evaluation of the results of the models, and finally, a Web Developer to put the work into production. According to the average established in Spain per year for each job, the process's labor costs are hypothetically defined.

Table 14 LABOR COSTS.

Employee	Cost/hour	Hour/week	Weeks	Hours	Total
Data Engineer	13 €	10	2	20	260 €
Data Scientist	14 €	7.5	6	45	630 €
Web Developer	11 €	10	4	40	440 €
TOTAL					1330 €

5.2.2 Software costs

There is no application in the cost of the software with the tools used because all the tools are open source. However, a two-month payment was made in *Github Enterprise*, which allows us to transfer heavy storage files to the platform with greater flexibility in the number of transactions to uploading and downloading files to the repository. Among the platforms and tools that are worth mentioning are mainly the following:

- *Python*: is a programming language oriented to data processing, analysis, and training, providing a set of open-source libraries. Within the *Python* environment, it is worth highlighting the main libraries used for code development.
 - o *Torchxrayvision*: free-to-use library with convolutional network architectures trained on different X-ray image databases for transfer learning.
 - o *Pytorch*: free use library for the training and designing of deep learning models used to execute transfer learning.
 - o *Tensorflow*: open-source library used to put the automatic report generation model into production.
 - o *Streamlit*: free use library to create the web application.

- *Cv2, Skimage*: free use libraries dedicated to image preprocessing.
- *Pandas*: library for managing the databases used in the project.
- *Visual Studio Code*: free code editor to facilitate the programmer's language syntax, code debugging, and code version control tasks.
- *Github*: is a platform mainly based on implementing collaborative code to have an open-source information source code. The platform is free, however, it has an integrated section for uploading large files to the platform (*Git LFS*) that does not allow unlimited upload and download transactions of large files to the repositories. For more flexibility in uploading large storage models, we have subscribed to a 2-month paid plan, which has made it easier to control the repository version with large files.

Table 15 SOFTWARE COSTS.

Software Tools	Total
Python 3.8 + Dependencies	0 €
Visual Studio Code	0 €
Github + Git LFS with 2 month subscription	10 €
TOTAL	10€

5.2.3 Hardware costs

Hardware tools have been based mainly on the physical components of the computer to process and store data.

- *NVIDIA GPU 3050*: this graphics card has allowed us to accelerate the model training processes and advance the work's development more efficiently.
- *AMD Ryzen 5600H*: processor used for information filtering and database preprocessing.
- *16GB RAM*: this has allowed us to store data in cache memory so that the reading and writing of information have fast execution.
- *HP Victus 16-e0067ns*: computer model on which the study was conducted.
- *Samsung T7 SSD external hard disk, 500 GB*: external hard disk used to store databases.

Table 16 HARDWARE COSTS.

Hardware components	Total
HP Victus 16-e0066ns AMD Ryzen 5 5600H/16GB/RTX 3500/	780 €
Samsung T7/SSD/500GB	96 €
TOTAL	876 €

5.2.4 Material cost

The cost of supplementary material has been allocated concerning electricity and monthly fiber used for the computer power supply and access to web information.

Table 17 MATERIAL COSTS.

Material	Cost	Weather	Total
Electricity	0.3 €/ hour	105 hours	31,5 €
Fiber Optics	30,95 €/month	6 months	185,7 €
TOTAL			217,2 €

5.2.5 Total costs

Total costs include labor, hardware, and material costs used in the development of the process.

Table 18 TOTAL COSTS.

Type of Cost	Total
Software costs	10 €
Hardware costs	876 €
Labor costs	1330 €
Material costs	217,2 €
TOTAL	2.433,2 €

5.3 Regulatory framework

Within the scope of current legislation, the adequacy of the development and implementation of the methods applied to the General Data Protection Regulation (EU) 2016/679 has been considered. The databases come anonymized by default, complying with the definitions of the RGPD according to Article 4.5 (data anonymization) and also Article 4.1 (protection of personal data), as well as also the *Organic Law 3/2018, of December 5, on the Protection of Personal Data and Guarantee of Digital Rights (LOPDGDD)*, which protects individuals concerning the processing of images. It also establishes the purposes of treatment with the information and access to personal data complying with Article 13 of the RGPD.

The treatment, definition, and analysis of the data, have had the authorization conceived of each one of the databases according to the *Law 5/1998* with the protection of the databases. As for the requirements of each of the databases, the necessary compliances for the use of these are defined:

- *Chexpert*: to access the *Chexpert* database, the data use agreements are complied with, and registration for the use of the database has been made [34].
- *China National Center for Bioinformation (CC-CXRI)*: the database is freely accessible to support physicians [35].
- *MIMIC-CXR-JPG*: access has been based on the authorization of user credentials imposed by *Physionet*, complying with the credential agreements for its use [36].
- *Kaggle databases*: all *Kaggle* databases are freely accessible, making it possible to download anonymized images. The only requirement for downloading is to subscribe to the web platform.

On the other hand, the *European Commission* regulation 21/4 complies with the principle of artificial intelligence systems to not create intelligent systems that cause an adverse risk of violating people's safety.

The reproduction of the work based on ideas, code, and documentation has been cited in the IEEE reference style to protect the copyright and intellectual property rights according to the *Intellectual Property Law* established by the *Royal Legislative Decree 1/1996* of April 12 and modified by *Law 5/1998*, as well as in compliance with the *Royal Decree-Law 24/2021* that changes it to adapt the copyright to digital content.

Regarding technical standards, no conventional procedure has been followed with code structure and format. However, the code is documented so that contributors can continue to improve the system.

6. CONCLUSIONS

This chapter discusses the results obtained concerning the proposed objectives of section 1.2, contrasting the purpose to be achieved with the work developed. On the other hand, new possible ideas for the future to improve the medical care system are put forward.

6.1 Main conclusions

The main objective of the work is to present a possible mechanism of radiological assistance to detect and localize pulmonary pathologies by employing an artificial intelligence system that allows for automation of the reporting and diagnostic tasks. Within the range of proposed tasks, most of the objectives have been fulfilled, discussing possible improvements regarding the obtained results:

- We have defined a set of databases of X-ray images with their possible pathologies, making an exploratory analysis and a set of visualizations on these, as well as a set of stored data with a large volume to be able to use these in future work.
- The pneumonia detection model achieves quantitatively interesting results predicting pneumonia and typical images. The saliency map of the pneumonia detection model succeeds in pinpointing areas with pneumonia in the four experiments performed with test specimens and identifying standard regions in the radiograph image in the other two experiments (section 4.3.1).
- The pathology detection model and the state-of-the-art models obtain exciting results, being quantitatively improvable. However, in the performed test experiments, the pathology detection model (section 4.3.2) successfully detects each labeled pathologies. Furthermore, each of the test samples is labeled with only one pathology, and yet the expert radiologist manages to see more than one pathology in the example, as well as in many cases, her conclusion coincides with that of the trained model, demonstrating that the *MIMIC-CXR-JPG* database shows mislabeled image symbols, being a poorly curated database to perform multiple pathology training. It is possibly one of the reasons why state-of-the-art and the model do not obtain remarkably accurate results relative to the models trained on the *MIMIC-CXR-JPG* database.
- We have managed to reproduce the automatic report through an open-source code, adapting it to our training dataset and working to have an automated code flow to introduce an image and generate a clinical report.
- A web ecosystem has been created in production through *Streamlit*, a platform that allows you to create your application using *Python* code. Within the application, it has been possible to integrate a menu to diagnose pneumonia, another to classify multiple pathologies, another to generate a diagnostic report, and, finally, one that performs the diagnosis of the three previous menus mentioned above. The application web also presents adjustable parameters, intending to facilitate the tasks of assistance and radio diagnostic support to the clinician. The web page also incorporates a system that allows the visualization of care maps, locating possible areas of pathological infection, and

exposing them to the physician. Finally, a language translator has been incorporated, enabling the application to be read in more than one hundred languages.

- We put the web page on the Internet to reach all the public health administrations in the world, as well as the entire community of data scientists, with the possibility of proposing possible improvements to our work for the future. The code to collaborate on the project is available on *Github*, and there is also Internet access for the first release of the application.
- It has not been possible to provide the possibility on the web page itself for the physician to modify the diagnosis proposed by the models. This objective was established exclusively for expert radiologists to relabel ill-defined images from the stored databases to have a web intermediary to improve the quality of the databases.

6.2 Future Work

Considering the different objectives achieved and not achieved in the development of the project, it is worth highlighting possible innovative ideas for future work to improve the medical support system:

- The described complexity of the analyzed databases opens the door for a more comprehensive manual curation, in which clinicians and data scientists collaborate to improve the study results. In addition, although the project has attempted to cover the most important databases of chest radiographs, other projects could also be added.
- Regarding the diagnosis of pneumonia detection, it is proposed to create a model with the distribution of data from *CC-CXRI-P* and the different databases defined. For this purpose, a database has already been described with images from *MIMIC-CXR-JPG*, *Chexpert*, *Kaggle*, and *CC-CXRI-P* (see section 4.1.2) to create a pneumonia model that generalizes in different distributions.
- Concerning the multiclass classification, it is proposed to change the training strategy of the model, making three different detection models for other groups of pathologies (see section 3.4.2). These models would serve as diagnostics for detecting diseases, radiological findings, or lung lesions. The classification of each model would be binary, and the objective would be to identify whether or not the patient has the disease, whether or not they have radiological findings, and whether or not they have lung lesions from a chest image. Once these models could solve these tasks, they would play the role of alarming physicians regarding whether they have a disease, lung lesions, or radiological findings.
- Regarding the website, it should be noted that the translator's performance is understandable but poor in grammatical terms. An innovative idea would be to incorporate better translation services to make the documentation more accurate.
- Finally, another proposal would be to incorporate a menu dedicated to expert radiologists so that they could modify radiographs from mislabeled databases, to have

a web interface to update those changes and report possible errors in the models put into production.

7. REFERENCES

- [1] A. Esteva *et al.*, “A guide to deep learning in healthcare”, *Nat. Med.*, vol. 25, núm. 1, pp. 24–29, 2019, doi: 10.1038/s41591-018-0316-z.
- [2] Europa Press, “El mercado de la Inteligencia Artificial alcanzará los 1.400 millones de euros invertidos en 2025”, 2022. Available: <https://www.europapress.es/portaltic/empresas/noticia-mercado-inteligencia-artificial-alcanzara-1400-millones-euros-invertidos-2025-20220329180652.html>
- [3] J. Jumper *et al.*, “Highly accurate protein structure prediction with AlphaFold”, *Nature*, vol. 596, núm. 7873, pp. 583–589, 2021, doi: 10.1038/s41586-021-03819-2.
- [4] A. Kryshchak, T. Schwede, M. Topf, K. Fidelis, y J. Moult, “Critical assessment of methods of protein structure prediction (CASP)-Round XIV”, *Proteins*, vol. 89, núm. 12, pp. 1607–1617, 2021, doi: 10.1002/prot.26237.
- [5] L. Murray, D. Gopinath, M. Agrawal, S. Horng, D. Sontag, y D. R. Karger, “MedKnowts: Unified documentation and information retrieval for electronic health records”, 2021. doi: 10.1145/3472749.3474814.
- [6] “Startseite - SEMIC RF Electronic GmbH - HF & Mikrowelle - Künstliche Intelligenz - KI - SaaS”, *Semic.de*. Available: <https://www.semic.de/de/startseite>.
- [7] L. Sarker, M. M. Islam, T. Hannan, y Z. Ahmed, “COVID-DenseNet: A deep learning architecture to detect COVID-19 from chest radiology images”, *Preprints*, 2020. doi: 10.20944/preprints202005.0151.v1.
- [8] J. P. Cohen *et al.*, “Predicting COVID-19 pneumonia severity on chest X-ray with deep learning”, 2020, doi: 10.48550/ARXIV.2005.11856.
- [9] O. Alfarghaly, R. Khaled, A. Elkorany, M. Helal, y A. Fahmy, “Automated radiology report generation using conditioned transformers”, *Inform. Med. Unlocked*, vol. 24, núm. 100557, p. 100557, 2021, doi: 10.1016/j.imu.2021.100557.
- [10] J. P. Cohen, P. Bertin, y V. Frappier, “Chester: A web delivered locally computed chest X-ray disease prediction system”, *arXiv [cs.CV]*, 2019, doi: 10.48550/ARXIV.1901.11210.
- [11] IBM, “Machine learning”, *IBM Design for AI*. Available: <https://www.ibm.com/design/ai/basics/ml/>.
- [12] J. M. Alvarez, “El perceptrón como neurona artificial”, *Blog de Jose Mariano Alvarez*, el 11 de junio de 2018. Available: <http://blog.josemarianoalvarez.com/2018/06/10/el-perceptron-como-neurona-artificial>.

- [13] M. Schaedler, C. Bluemm, M. Kuschnerov, F. Pittalà, S. Calabrò, y S. Pachnicke, “Deep neural network equalization for optical short reach communication”, *Appl. Sci. (Basel)*, vol. 9, núm. 21, p. 4675, 2019, doi: 10.3390/app9214675.
- [14] “CURSO BÁSICO DE EDICIÓN DE IMÁGENES”, *Upv.es*. Available: https://yosedo.webs.upv.es/curso_edicion_de_imagenes_nivel_basico/1_7.html.
- [15] Wikipedia contributors, “Image resolution”, *Wikipedia, The Free Encyclopedia*, el 30 de mayo de 2022. Available: https://en.wikipedia.org/w/index.php?title=Image_resolution&oldid=1090639796.
- [16] I. Goodfellow, B. Yoshua, y C. Aaron, *Deep Learning*. Cambridge, Mass., Estados Unidos de América: MIT Press, 2016. Available: <http://www.deeplearningbook.org>.
- [17] H. Chung y K.-S. Shin, “Genetic algorithm-optimized multi-channel convolutional neural network for stock market prediction”, *Neural Comput. Appl.*, vol. 32, núm. 12, pp. 7897–7914, 2020, doi: 10.1007/s00521-019-04236-3.
- [18] A. Gupta *et al.*, “Deep learning in image cytometry: A review: Deep learning in image cytometry”, *Cytometry A*, vol. 95, núm. 4, pp. 366–380, 2019, doi: 10.1002/cyto.a.23701.
- [19] Á. Iglesias-Puzas y P. Boixeda, “Deep learning y DerMATología”, *Actas Dermosifiliogr.*, vol. 111, núm. 3, pp. 192–195, 2020, doi: 10.1016/j.ad.2019.01.014.
- [20] G. Huang, Z. Liu, L. van der Maaten, y K. Q. Weinberger, “Densely connected convolutional networks”, *arXiv [cs.CV]*, 2016, doi: 10.48550/ARXIV.1608.06993.
- [21] X. He, Y. Peng, y J. Zhao, “Fine-grained discriminative localization via saliency-guided faster R-CNN”, 2017. doi: 10.1145/3123266.3123319.
- [22] C. V. Aravinda, M. Lin, K. R. Udaya Kumar Reddy, y G. Amar Prabhu, “A demystifying convolutional neural networks using Grad-CAM for prediction of coronavirus disease (COVID-19) on X-ray images”, pp. 429–450, 2021. doi: 10.1016/b978-0-12-824536-1.00037-x.
- [23] S. Thrun y L. Pratt, “Learning to learn: Introduction and overview”, en *Learning to Learn*, Boston, MA: Springer US, 1998, pp. 3–17. doi: 10.1007/978-1-4615-5529-2_1.
- [24] O. Russakovsky *et al.*, “ImageNet large scale visual recognition challenge”, *Int. J. Comput. Vis.*, vol. 115, núm. 3, pp. 211–252, 2015, doi: 10.1007/s11263-015-0816-y.

- [25] P. Marcelino, “Transfer learning from pre-trained models”, *Towards Data Science*, el 23 de octubre de 2018. Available: <https://towardsdatascience.com/transfer-learning-from-pre-trained-models-f2393f124751>.
- [26] J. P. Cohen *et al.*, “TorchXRyVision: A library of chest X-ray datasets and models”, *arXiv [eess.IV]*, 2021. Available: <http://arxiv.org/abs/2111.00595>
- [27] D. Chakravorty, “Confusion matrix”, *Debadityachakravorty.com*. Available: <https://www.debadityachakravorty.com/ai-ml/cmatrix/>.
- [28] T. B. Brown *et al.*, “Language Models are Few-Shot Learners”, 2020, doi: 10.48550/ARXIV.2005.14165.
- [29] R. Thoppilan *et al.*, “LaMDA: Language models for Dialog Applications”, 2022, doi: 10.48550/ARXIV.2201.08239.
- [30] Radiological Society of North America (RSNA) y American College of Radiology (ACR), “Rayos X del tórax”, *Radiologyinfo.org*. Available: <https://www.radiologyinfo.org/es/info/chestrad>.
- [31] “Radiografías, tomografías computarizadas (TC) e imágenes por resonancia magnética (IRM) (X-rays, CT Scans, and MRIs) - OrthoInfo - AAOS”, *Aaos.org*. Available: <https://orthoinfo.aaos.org/es/treatment/radiografias-tomografias-computarizadas-tc-e-imagenes-por-resonancia-magnetica-irm-x-rays-ct-scans-and-mris/>.
- [32] T. Preston-Werner, S. Chacon, C. Wanstrath, y P. J. Hyett, “Github”, *Github*. Available: <https://github.com/>.
- [33] J. Irvin *et al.*, “CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison”, 2019, doi: 10.48550/ARXIV.1901.07031.
- [34] “CheXpert: A large dataset of chest X-rays and competition for automated chest X-ray interpretation”, *Github.io*. Available: <https://stanfordmlgroup.github.io/competitions/chexpert/>.
- [35] CNCB-NGDC Members and Partners, “Database resources of the National Genomics Data Center, China National Center for Bioinformation in 2021”, *Nucleic Acids Res.*, vol. 49, núm. D1, pp. D18–D28, 2021, doi: 10.1093/nar/gkaa1022.
- [36] A. E. W. Johnson *et al.*, “MIMIC-CXR-JPG, a large publicly available database of labeled chest radiographs”, 2019, doi: 10.48550/ARXIV.1901.07042.

- [37] A. Johnson *et al.*, “MIMIC-CXR-JPG - chest radiographs with structured labels”. el 14 de noviembre de 2019. doi: 10.13026/8360-t248.
- [38] A. Goldbloom, “Kaggle: Your machine learning and data science community”, *Kaggle.com*. Available: <https://www.kaggle.com/>.
- [39] A. Asraf, “COVID19_Pneumonia_Normal_Chest_Xray_PA_Dataset”. el 13 de julio de 2020. Consultado: el 20 de junio de 2022. Available: <https://www.kaggle.com/datasets/amanullahasraf/covid19-pneumonia-normal-chest-xray-pa-dataset>
- [40] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, y R. M. Summers, “ChestX-Ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases”, 2017. doi: 10.1109/cvpr.2017.369.
- [41] P. Mooney, “Chest X-ray images (pneumonia)”. el 24 de marzo de 2018. Consultado: el 20 de junio de 2022. Available: <https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia>
- [42] A. Paszke *et al.*, “PyTorch: An imperative style, high-performance deep learning library”, 2019, doi: 10.48550/ARXIV.1912.01703.
- [43] H. H. Pham, T. T. Le, D. Q. Tran, D. T. Ngo, y H. Q. Nguyen, “Interpreting chest X-rays via CNNs that exploit hierarchical disease dependencies and uncertainty labels”, 2019, doi: 10.48550/ARXIV.1911.06475.
- [44] M. Wang, S. Lu, D. Zhu, J. Lin, y Z. Wang, “A high-speed and low-complexity architecture for softmax function in deep learning”, 2018. doi: 10.1109/apccas.2018.8605654.
- [45] “Streamlit • The fastest way to build and share data apps”, *Streamlit*. Available: <https://streamlit.io/>.
- [46] J. P. Cohen, M. Hashir, R. Brooks, y H. Bertrand, “On the limits of cross-domain generalization in automated X-ray prediction”, 2020, doi: 10.48550/ARXIV.2002.02497.
- [47] “SEICAT: Sociedad Española de Imagen Cardiorádica : Congreso 2012”, *Seicat.org*. Available: <http://seicat.org/index.php?seccion=congreso2012>.

