Universidad Autónoma de Madrid
Escuela politécnica superior

Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación

Trabajo Fin de Grado

EARLY DETECTION OF LUNG CANCER THROUGH NODULE CHARACTERIZATION BY DEEP LEARNING

Autor: Andrea González Rodríguez
Tutor: Marcos Escudero Viñolo
Ponente: Jesús Bescós Cano

Junio 2019

EARLY DETECTION OF LUNG CANCER THROUGH NODULE CHARACTERIZATION BY DEEP LEARNING

Autor: Andrea González Rodríguez
Tutor: Marcos Escudero Viñolo
Ponente: Jesús Bescós Cano

Video Processing and Understanding Lab
Departamento de Tecnología Electrónica y de las Comunicaciones
Escuela Politécnica Superior
Universidad Autónoma de Madrid
Junio 2019

Abstract

Resumen

El cáncer de pulmón es uno de los cánceres más comunes del mundo con 1.8 millones de casos nuevos reportados en el 2012, suponiendo el 12.9% de todos los cánceres en el mundo, representando 1.4 millones de muertes hasta 2008.

La importancia de la detección temprana y de la clasificación de nódulos malignos y benignos usando tomografías computarizadas (TC) puede facilitar a los radiólogos las tareas de evaluación de la estadificación de nódulos y la planificación terapéutica individual. Asimismo, si se detectan nódulos malignos potenciales en las TC, los tratamientos podrían ser menos agresivos, incluso no llegando a necesitar quimioterapia o radioterapia tras la cirugía.

Este Proyecto se centra en la exploración de los métodos y bases de datos existentes para la clasificación automática de nódulos pulmonares basados en imágenes TC. Para ello, comenzamos por reunir, estudiar y analizar algunos estudios de última generación en detección, caracterización y clasificación de nódulos pulmonares. Además, reportamos y contextualizamos arquitecturas de aprendizaje profundo de última generación adecuadas para la clasificación de nódulos pulmonares. De los conjuntos de datos estudiados, seleccionamos un conjunto de datos, grande y ampliamente usado, en imagenes TC de nódulos pulmonares, y lo usamos para ajustar una red neural convolucional de última generación. Comparamos esta estrategia con la construcción -desde cero- de una nueva red neuronal menos profunda.

La evaluación inicial sugiere que: (1) La transferencia de aprendizaje no es capaz de funcionar correctamente debido a su incapacidad para adaptarse a los dominios entre las imágenes naturales y los escáneres TC. (2) El aprendizaje desde cero no puede entrenarse con un número pequeño de muestras. Sin embargo, esta primera evaluación facilita el camino hacia la creación de mejores métodos de clasificación alimentados por bases de datos públicas mejor anotadas.

En general, este Proyecto es una primera etapa obligatoria sobre un tema de investigación candente.

Palabras Clave

Nódulos pulmonares, base de datos LIDC-IDRI, trasnferencia de aprendizaje, aprendizaje desde cero, AlexNet, TC.

Abstract

Lung cancer is one of the most frequent cancers in the world with 1.8 million new cases reported in 2012, representing 12.9% of all new cancers worldwide, accounting 1.4 million deaths up to 2008.

The importance of early detection and classification of malignant and benign nodules using computed tomography (CT) scans, may facilitate radiologists the tasks of nodule staging assessment and individual therapeutic planning. However, if potential malignant nodules are detected on CT scans, treatments may be less aggressive, not even requiring chemotherapy or radiation therapy after surgery.

This Bachelor Thesis focus on the exploration of existing methods and data sets for the automatic classification of lung nodules based on CT images. To this aim, we start by assembling, studying and analyzing some state-of-the-art studies in lung nodule detection, characterization and classification. Furthermore, we report and contextualize state-of-the-art deep learning architectures suited for lung nodule classification. From the public datasets researched, we select a widely used and large data set of lung nodules CT scans, and use it to fine-tune a state-of-the-art convolutional neural network. We compare this strategy with training-from-scratch a new shallower neuronal network.

Initial evaluation suggests that: (1) Transfer learning is unable to perform correctly due to its inability to adapt between natural images and CT scans domains. (2) Learning from scratch is unable to learn from a small number of samples. However, this first evaluation paves the road towards the design of better classification methods fed by better annotated public-available data sets.

In overall, this Project is a mandatory first stage on a hot research topic.

Key words

Lung nodule, LIDC-IDRI dataset, transfer learning, learning from scratch, AlexNet, CT.

Agradecimientos

En primer lugar quiero dar las gracias a mi tutor Marcos por su compañía, su energía y su apoyo durante el desarrollo de este proyecto. Ante todas las dificultades que se han presentado, siempre ha estado dispuesto a dedicarme su tiempo y su conocimiento. Se lo agradezco de corazón. He aprendido mucho gracias a él.

Tambien quiero agradecer a mi madre haberme aportado las competencias médicas necesarias para este trabajo, además de su apoyo absoluto durante estos cuatro años de carrera. De igual manera a mi padre, que siempre nos saca una sonrisa incluso en las situaciones más estresantes. Gracias a mis padres y a mis dos hermanos por creer en mí y por todos los consejos y buenos momentos juntos. Han hecho mucho más amenos estos años.

A mis amigos y compañeros de viaje durante la carrera: Marta Fernández, Berta Fernández y Manuel Moyano, por ser mi pilar fundamental desde que les conozco. No podría haberlo conseguido sin ellos, es una suerte haberlos conocido.

Gracias a mis amigos: por las risas, los buenos momentos, el apoyo, la paciencia y el amor incondicional que siempre me dan. Son mi familia también. Tan importantes como un vaso de agua en el desierto.

Mi más sincero agradecimiento a Nicolás García, mi apoyo fundamental: gracias por tu fortaleza, confianza, ayuda, paciencia y cariño.

Y por último, a todos aquellos profesores de la Escuela Politécnica Superior de la UAM, que me han ayudado durante el camino y siguen haciendolo. Les tengo un gran aprecio y son una parte importante también de este trabajo.

Contents

List of figures **ix**

List of tables **xi**

1 Introduction **1**

1.1 Motivation 1

1.2 Objectives 5

1.3 Document Organization 5

2 Related Work **7**

2.1 Automatic Learning Techniques 7

2.1.1 The neuron 7

2.1.2 Artificial Neural Networks 8

2.1.3 Optimization methods for deep learning 15

2.1.4 Learning strategies 16

2.1.5 Augmented Learning 18

2.2 Detection and Classification of lung nodules 19

2.2.1 Detection of lung nodules 19

2.2.2 Classification of lung nodules 21

3 Material and Methods **23**

3.1 Dataset: LIDC-IDRI 23

3.1.1 Specification of the dataset 23

3.1.2 Definition of the annotation structure 25

3.1.3 Annotated characteristics 26

3.1.4 Alternative datasets 27

3.2 Explored CNN architecture 27

3.2.1 Defining a CNN architecture 28

4 Experiments **31**

4.1 Experimental context: dataset 31

4.2 Experimental setup 33
4.2.1 Experiments description 33
4.2.2 Transfer learning setup 34
4.2.3 Learning from scratch setup 36
4.2.4 Contextualization of experimental results and general discussion 38

5 Conclusions and future work **39**
5.1 Conclusions 39
5.2 Future work 40

List of figures

1.1 Mortality standardized by age for lung cancer in Spain, men (left) and women (right) (period 1951-2011). Extracted from [1]. 3

1.2 Estimation of the number of new cancer cases in Spain by 2019 (excluding non-melanoma skin tumours). Extracted from [1]. 3

1.3 "Collage of some medical imaging applications in which deep learning has achieved state-of-the-art results. From top-left to bottom-right: mammographic mass classification, extracted from [2], segmentation of lesions in the brain (top ranking in BRATS, ISLES and MRBrains challenges, extracted from [3] , leak detection in airway tree segmentation, extracted from [4]), diabetic retinopathy classification, extracted from [5], prostate segmentation (top rank in PROMISE12 challenge), nodule classification (top ranking in LUNA16 challenge), breast cancer metastases detection in lymph nodes (top ranking and human expert performance in CAMELYON16), human expert performance in skin lesion classification, extracted from [6], and state-of-the-art bone suppression in x-rays, extracted from [7]." Figure and caption extracted from [8]. 4

2.1 Neuron scheme. 8

2.2 CNN scheme. Extracted from [9]. 9

2.3 Example of convolution operation. Extracted from [10]. 11

2.4 Classification results through time. Extracted from [11]. 13

2.5 Transfer Learning scheme. Extracted from [12]. 17

2.6 Example of dropout. Extracted from [13]. 18

2.7 Computer-Aided Detection (CAD) system process. 20

2.8 "(a) Node graphs of 1D representation of architectures commonly used in medical imaging. (b) Auto-encoder, (c) Restricted Boltzmann machine, (d) Recurrent Neural Network (RNN), (e) Multi-stream convolutional neural network, (f) U-net (with a single down-sampling stage)." Figure and caption extracted from [8]. . . . 21

3.1 "Examples of lesions (nodules and non-nodules) in axial, sagittal, and coronal view. Lesions are in the center of the box (50 50 mm). The left set of images are nodules with a wide range of morphological characteristic: a) solid nodule, b) pleural nodule, c)–d) large nodules with irregular shape, e) –f) sub-solid nodules. The right set of images are irregular lesions that are not related with nodules or cancers. These examples illustrate that designing features for accurate detection and classification of nodules may not be trivial." Figure and caption extracted from [14]. 24

3.2 Examples of the six types of lung tissue in the LIDC-IDRI dataset [15]. "Disease tissue types are mapped with orange arrows. a): healthy; b): emphysema; c): ground glass; d): fibrosis; e): micronodules; f): consolidation." Figure and caption extracted from [16]. 25

3.3 Degree of malignancy of three lung nodules: a) Malignancy: 5 b) Malignancy: 3, c) Malignancy: 1. 26

3.4 AlexNet architecture. Adapted from [17]. 27

3.5 Comparative of different convolutional neural networks on ImageNet. Extracted from [18]. 28

4.1 Dataset node statistics. 32

4.2 Dataset roi statistics. 32

4.3 Malignancy statistics in Training and Validation dataset. 33

4.4 An example of validation loss function result (Optimizer: Adam, Initial Learning Rate: 0.0005, MaxEpochs: 40). 34

4.5 t-SNE 2D Train and Validation of transfer learning. 35

4.6 t-SNE 2D Train and Validation of learning from scratch. 36

4.7 Network from scratch architecture. 37

5.1 Gantt's diagram for the 2019 Bachelor Thesis. 40

List of tables

3.1 AlexNet architecture parameters. Adapted from [17]. 30

4.1 Transfer Learning results. 34

4.2 Learning from scratch results. 36

1 Introduction

In this work we start by introducing the rationale of the study and define the partial objectives that may lead to the accomplishment of our main goal: advance on the acquaintance of techniques and datasets of interest that may be useful for early detection of lung cancer. The chapter concludes by presenting the structure of this document.

1.1 Motivation

Lung cancer is one of the most frequent cancers in the world with 1.8 million new cases reported in 2012, representing 12.9% of all new cancers worldwide, [19]. In terms of mortality, lung cancer accounted 1.4 million deaths in 2008 worldwide, [20]. Being the principal cause of cancer death in the United States (US) and in the European Union (EU). A recently study reported in US showed an estimated 234,030 new cases of lung cancer assessed by end of 2018, and 154,050 deaths, [21]. In the EU, lung cancer is the fourth most frequent cancer; almost 313,000 new cases were assessed and 268,000 deaths in 2012, [19]. In Japan, lung cancer is the third most frequent cancer; approximately 113,000 new cases were diagnosed in 2012 and approximately 71,500 deaths in that year (Center for Cancer Control and Information Services, National Cancer Center, 2017). Five-year survival rates remain low: 17.7% and 13% in the US and Europe, respectively, [22], [23]. On another note, the number of tumors diagnosed in Spain in 2019 will reach 277,234, according to REDECAN (Spanish Network of Cancer Registries) estimates, compared to 247,771 cases diagnosed in 2015: 148,827 in men and 98,944 in women. The most frequent cancers diagnosed in Spain in 2019 will be those of the colon and rectum (44,937 new cases), prostate (34,394 new cases), breast (32,536 new cases), lung (29,503 new cases) and urinary bladder (23,819 new cases). The absolute number of cancers diagnosed in Spain has continued to rise for decades in relation to the probable increase in population (until 2012), the aging of the population (age is a fundamental risk factor in the evolution of cancer) and the exposure to other risk factors, [1].

Historically, lung cancers have been classified on the strength of histologic criteria, i.e., microscopic examination of tissue biopsy, and are divided into two categories: small-cell lung cancer (SCLC) and non–small-cell lung cancer (NSCLC). This distinction is essential for staging, treatment, and diagnosis because it have different therapeutic approaches. NSCLC is the most prevalent histologic class, accounting for nearly 85% of all lung cancers, [24], [25], and enclose

different subtypes as well as adenocarcinoma, squamous-cell carcinoma, large-cell carcinoma and bronchioloalveolar carcinoma, [26].

The principal risk factor for the development of lung cancer is tobacco use, accountable for approximately 90% of all lung cancers, [27]. As it can be observe in the graphs of the Figure 1.1, the incorporation of women into the habit of smoking and the latency time necessary to develop lung cancer in the smoking population means that we are now seeing an increment in the incidence of lung cancer in women that started to smoked at 60s. Therefore, avoidance of smoking and smoking abeyance provides the most important way to diminishing the morbidity and mortality identified with this disease. Because of this, we are now seeing a reduction in the incidence of men who began smoking in the 40s and 50s of the twentieth century. In this way, this parallelism can be observed in the data on both incidence and mortality in lung cancer in men and women in the different age groups. Additionally, as we can see in the Figure 1.2, the data indicates that age is a risk factor for all types of cancer in general, i.e., ageing processes increase the incidence of cancer, specially in the population exposed to a carcinogen like tobacco. Other risk factors are: previous radiation therapy, exposure to environmental toxins: exposure to smoke (passive smokers), asbestos, radon, metals (arsenic, chromium, and nickel), ionizing radiation, polycyclic aromatic hydrocarbon), HIV infection and pulmonary fibrosis.

Clinical manifestations Most lung cancer cases are diagnosed at advanced stages and about 78% of newly diagnosed patients have regional/distant disease, which means that cancer has disseminated to nearby lymph nodes or distant organs, [19], [23]. Due to the aggressive biology of the disease, no symptoms are observed until locally advanced or metastatic disease is present. Symptoms may be a consequence from the local effects of the tumor, from regional or distant spread, or from the distant effects unrelated to metastases (paraneoplastic syndromes). Almost 25% of non-screened patients have one or more symptoms at the time of diagnosis. The most frequent symptoms at at the time of submission were cough (55%), dyspnea (45%), pain (38%), and weight loss (36%), as discussed in a study of 2293 patients with NSCLC which the mean age was 64 years, [28].

The standard treatment algorithm for advance NSCLC patients has historically involved frontline treatment with chemotherapy. In recent years, attention has turned to stratifying lung cancer patients on the basis of molecular alterations since it has become clear that histologically identical tumors are driven by different oncogenes and are therefore likely to respond differently to therapeutic intervention, such as targeted therapies against mutated epidermal growth factor receptor (EGFR), anaplastic lymphoma kinase-rearranged (ALK+), c-ros oncogene 1 (ROS1) rearranged, or BRAF V600E mutant NSCLC and others.

If potential malignant nodules are detected on computed tomography (CT) scans, treatments may be less aggressive, not even requiring chemotherapy or radiation therapy after surgery.

The evaluation of lung cancer is mainly based on the assessment of patients with symptoms. Traditional chest radiography and sputum cytology were not shown to be effective in reducing mortality, therefore screening for lung cancer was not commonly used. However, the implantation of other radiological approaches such as CT, in screening trials, has changed the scenario. The National Lung Cancer Screening Program contrast the use of CT images with chest x-ray screening, yielding results that demonstrate a 20% reduction in lung cancer mortality in regular smokers who were screened per year in the course of three years with CT, [29], [30]. These previous results are due to experimental studies have shown that, through CT screening, it is possible to detect lung nodules in a early stage, which have a encouraged prognosis.

Image diagnosis is the gold-standard for establishing the diagnostic of suspicion of a neoplasm, as well as for establishing the extent of the disease. These images can be obtained by various techniques based on X-ray. The greatest sensitivity and specificity of the image is required in order to avoid false positives and negatives and also to avoid invasive tests. The final

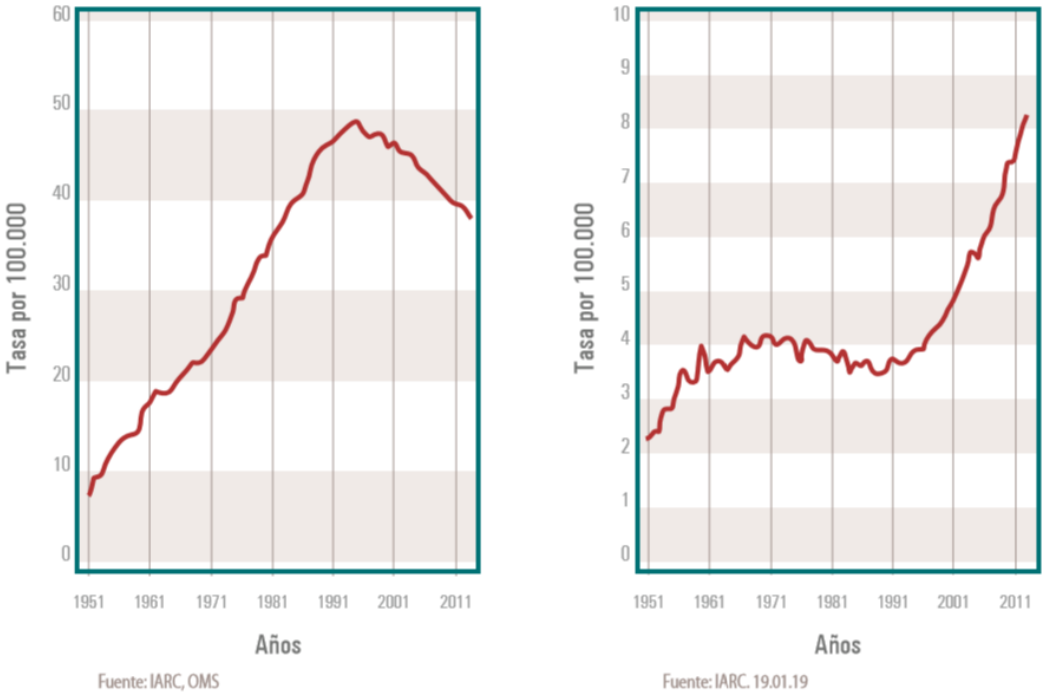

Figure 1.1: Mortality standardized by age for lung cancer in Spain, men (left) and women (right) (period 1951-2011). Extracted from [1].

	Hombres	Mujeres	Ambos sexos
< 65 años	54.148	54.223	108.371
> = 65 años	106.916	61.947	168.863
Todas las edades	161.064	116.170	277.234

Fuente: Red Española de Registros de Cáncer (REDECAN).

Figure 1.2: Estimation of the number of new cancer cases in Spain by 2019 (excluding non-melanoma skin tumours). Extracted from [1].

diagnostic of neoplasm is make by histopathological examination, [31].

In Figure 1.3, it can be observed examples of previous successful uses of deep-learning grounded techniques for the detection of several types of cancers. It is relevant to say that image resolution is a key factor for the performance of these techniques.

The design of automatic methods for the assessment of nodule status (malignancy, calcification, specularity, etc.) is a complex task, as it is hard to find well-suited generic and discriminative cues. Moreover, the accuracy of existing datasets is unreliable, as the definition of nodule morphology is subjected to radiologist subjectivity (e.g., boundary annotation), [32].

The importance of early detection and classification of malignant and benign nodules using CT images may simplify radiologists the responsibility of nodule staging evaluation and individual therapeutic strategy.

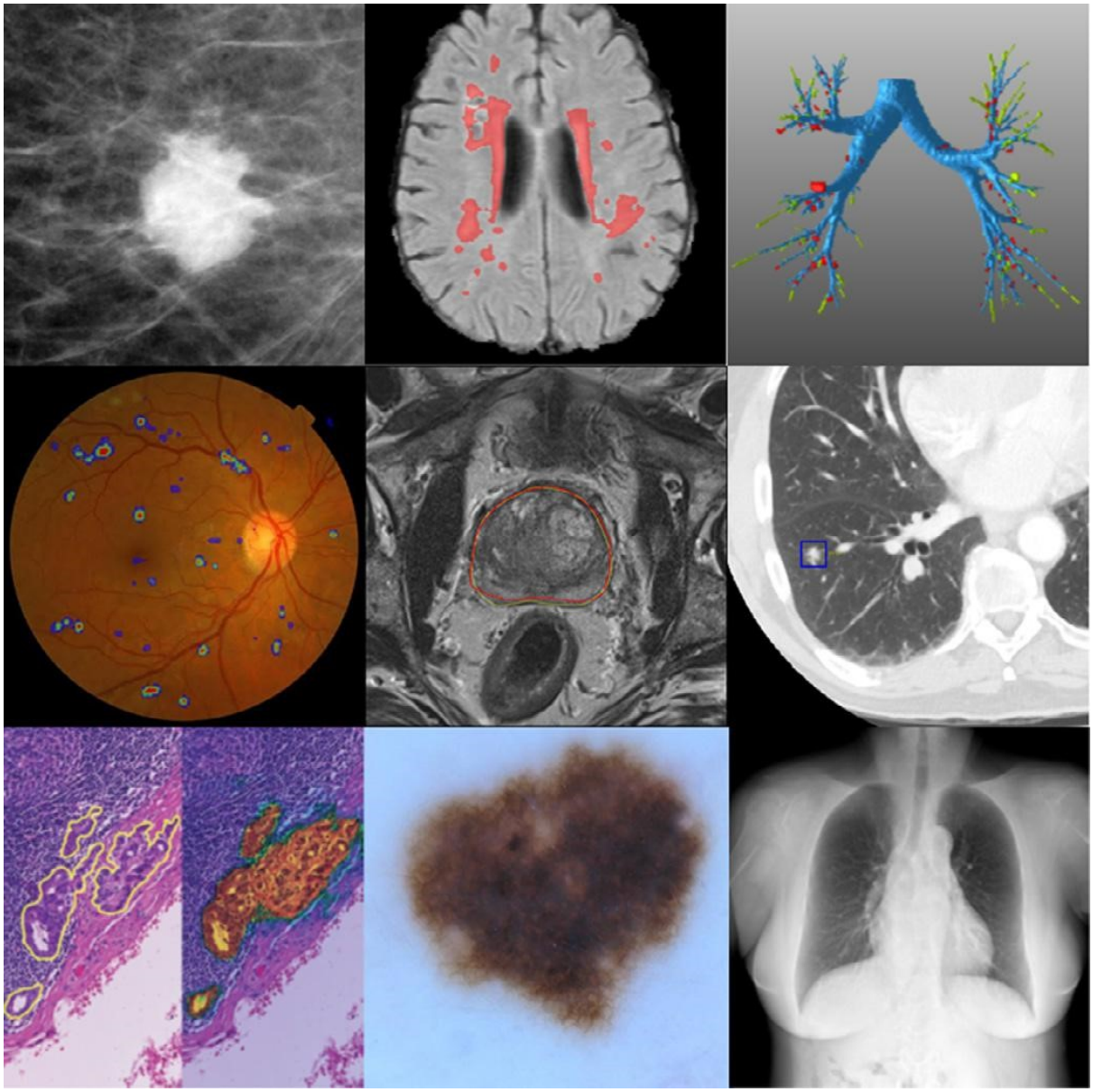

Figure 1.3: "Collage of some medical imaging applications in which deep learning has achieved state-of-the-art results. From top-left to bottom-right: mammographic mass classification, extracted from [2], segmentation of lesions in the brain (top ranking in BRATS, ISLES and MR-Brains challenges, extracted from [3] , leak detection in airway tree segmentation, extracted from [4]), diabetic retinopathy classification, extracted from [5], prostate segmentation (top rank in PROMISE12 challenge), nodule classification (top ranking in LUNA16 challenge), breast cancer metastases detection in lymph nodes (top ranking and human expert performance in CAMELYON16), human expert performance in skin lesion classification, extracted from [6], and state-of-the-art bone suppression in x-rays, extracted from [7]." Figure and caption extracted from [8].

1.2 Objectives

This Project focus on the exploration of existing methods and datasets for the automatic classification of lung nodules based on CT images. To complete this main objective, the following sub-objectives are proposed:

1. Assembling, study, organization and analysis of state-of-the-art studies in lung nodule detection, characterization and classification as well as of deep learning methods suited for these tasks.
2. Assembling, organization and assessment of existing indexed datasets of lung nodule CT images.
3. Initial evaluation of deep learning solutions trained to assess the malignancy of lung nodules.

1.3 Document Organization

The rest of this document is organized as follows: the second chapter provides a brief overview of automatic learning techniques and reviews existing literature on detection and classification of lung nodules. The material and methods used during the project are described in the third chapter, followed by the experiments description and results in chapter four. Our conclusions and future work are drawn in the fifth chapter.

2

Related Work

2.1 Automatic Learning Techniques

The rise of machine learning techniques, and specially of neural network models, has led to a paradigm shift in many areas of medical-oriented computer vision, such as: image, segmentation, nodule detection and classification and assessment of lesion's evolution. One of the key differences with respect to previous approaches is learning from data: a classification for a new unobserved nodule is learned given a set of data and, in the case of supervised learning, its corresponding labels (the information of interest). Learning is driven by a cost function and obtaining the best classification model becomes an optimization problem on this function.

2.1.1 The neuron

Summarizing, a neuron is a structure composed of a series of inputs, a ensemble of weights and an activation function [33]. The role of the neuron is to transform these series of inputs into a single output value, which can be taken as input from other neurons in posterior layers.

As can be observed in Figure 2.1, the primary elements of the neuron are:

- Inputs: They can be the initial numerical data or the outputs (also numerical) of other neurons.
- Weights: The weighting factor by which the corresponding input will be multiplied. These weights are specific and learned for each neuron.
- Output: The numerical value that the neuron produces after performing different calculations.
- Weighted sum: Each entry is multiplied by its respective weight, and the sum of all the above-mentioned multiplications is performed. The resulting value is the input to the activation function. Sometimes the bias is added to the result of the summation. Which is, essentially, a new entry with value 1 and weight *w0*.

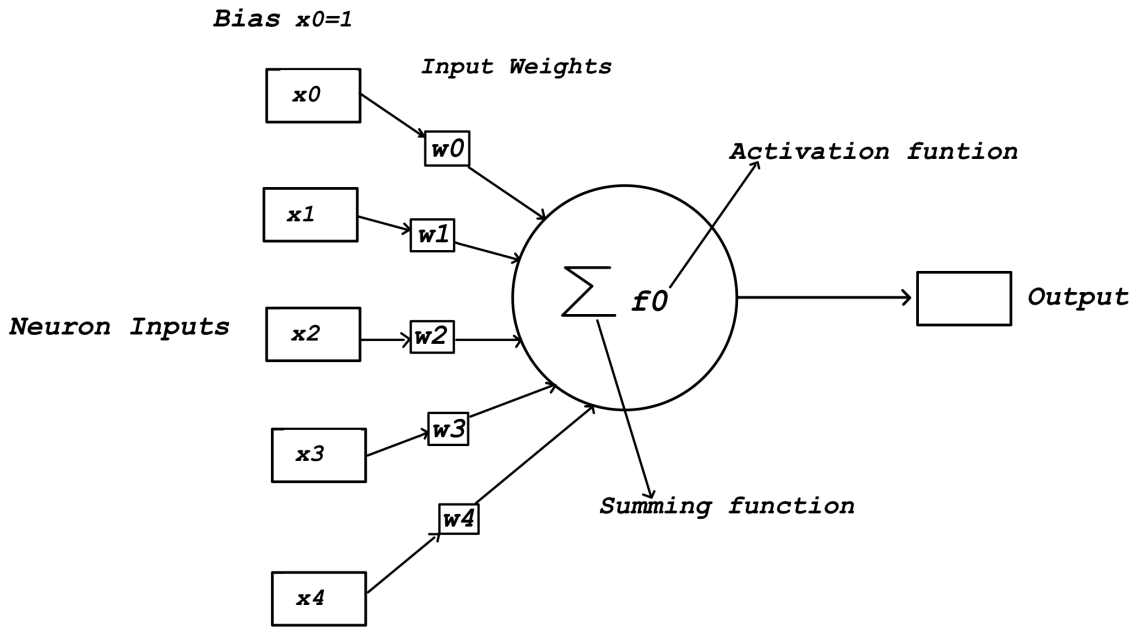

Figure 2.1: Neuron scheme.

- Activation function: The leading assignment of activation function is to maintain the values produced by the neuron within a reasonable range (Many times 0 and 1). There are two main types of activation functions: linear and non-linear. The linear functions give rise to simple models, which most of the times are unable to complex models (non-linear functions, highlighting among them, the Gaussian and the sigmoidal).

2.1.2 Artificial Neural Networks

Artificial neural networks (ANN) may be understood as computational models which is trying to shape biological performance model, [33]. It is constituted of a of a ensemble of units, called artificial neurons, linked together to transmit signals to each other. The input information goes through the neural network (where it is submitted to different operations) yielding some output values. The ANN is going to learn hierarchical knowledge, i.e., the neurons of the first layer are going to learn different characteristics and this acquaintance is going to be transmitted and combined in the following neurons, to be able to learn and solve the overall assignment.

Generic architecture: Multilayer perceptron

Multilayer perceptron is an ANN composed of multiple layers, which objective is to solve problems that are not linearly separable, which is the principal perceptron restriction (also called simple perceptron). The multilayer perceptron can be totally or locally connected.

The basic unit of an ANN is the neuron, represented in Figure 2.1. A network is composed of several neurons, which are arranged in layers as follows:

- Input layer: These neurons are the ones that receive the numerical data that the network processes.
- Output layer: Layer in charge of containing the final result of the mathematical operations that have taken place inside the network.

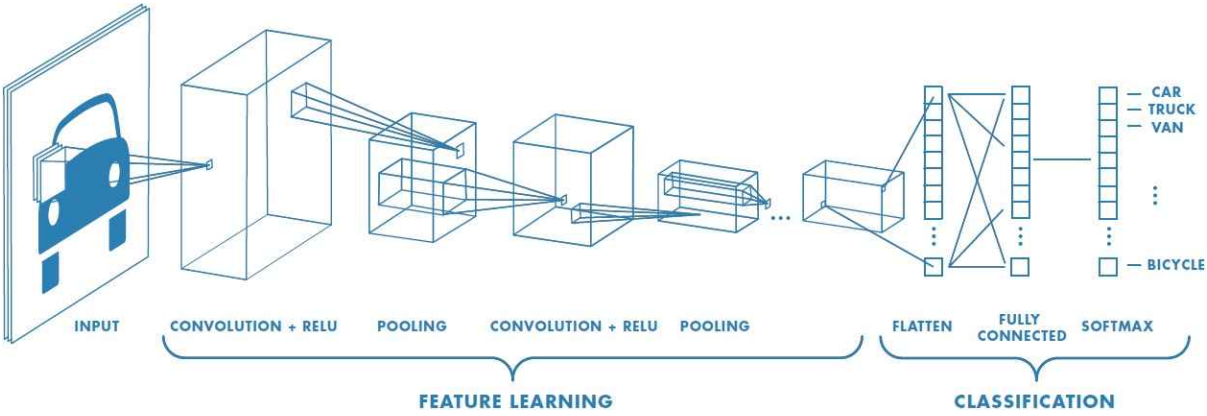

Figure 2.2: CNN scheme. Extracted from [9].

- Hidden layers: Its neurons contain the calculations and values of intermediate weights of the network.

The collection of hidden layers, located between the input and output layers, the number of neurons in each layer (the layer's capacity) and the nature of their activation functions are hyperparameters of the model design, that, together with the different types of architectures (fully-connected, convolutional, LSTM, among others), the learning rate and the use of regularization techniques, involves a multitude of design decisions that complicate the search for the best configuration. Usually, it depends on the concrete architecture, e.g., in fully-connected model, all neurons in one layer are associated to all the neurons in the subsequent layer, where weights are used in the main calculation of neural networks: multiply the output value of a neuron by the weight of a connection, the result of this operation being an input for the next neuron, [33]

Architecture of Convolutional Neural Networks

Convolutional neuronal networks (CNNs) [33] are especially used in artificial vision's field, where the neurons they use, belong to receptive domains in a similar way to neurons in the visual cortex of the brain. This inspiration is due to the work carried out by Hubel and Wiesel in 1959, [34]. Thanks to this work, the performance of the visual cortex was largely understood, especially cell's work responsible for orientation selectivity and edge detection in visual stimuli. Hubel and Wiesel discovered two types of neurons: simple and complex. Although, they are different, they share the characteristic of being excited if the visual stimulus they receive is aligned with the patterns that these cells have. This will be a feature that we will also find in the neurons of this type of architecture.

It is noteworthy that this architecture of neural network is a modification, by pruning connections, of the multilayer perceptron. However, thanks to its two-dimensional application, these networks are especially effective in classification and segmentation of images, and of any type of data that is spatially distributed continuously throughout the entry.

CNNs consists of different layers with convolutional filters which have one or more dimensions. Subsequently convolutional layer, it is common to add a function to include non-linear causal mapping.

As shown in Figure 2.2, CNNs can be interpreted to be divided into two stages: feature extraction and classification. In the first stage, the input data characteristics are extracted (feature maps). Feature maps have two layers: the convolution layer and the pooling layer.

Once the feature maps have been extracted, a fully connected network acts as a sorter. This layer is responsible for separating and classifying each feature. This stage is defined by one or several fully connected layers and, usually, a Softmax layer, which configures an activation function.

Types of layers and neurons in convolutional networks

Convolutional Layers A convolutional layer is a group of neurons. Neurons connect little areas of the input or of the previous layer. This areas can be understood as filters (kernels). These filters, of pre-defined dimension, have weights which value is learned in the training phase by (automatically) adjusting them to solve the target problem. Each filter is applied throughout the input vertically and horizontally, repeating the scalar product of the weights and the input. The amount of weights of a filter is defined by $h \times w \times c$, being h the height of the filter, being w the width of the filter and being c the number of input channels. The filter moves along the input, shaping the feature map.

In a layer of K squared filters of dimension *(2d + 1)* $\times$ *(2d + 1)*, being d the size of the neighborhood on which the formula will be applied, being X the input signal and being H the filter, the convolution operation is expressed by:

$$Y_k[i,j] = \sum_{n=i-d}^{i+d} \sum_{m=j-d}^{j+d} X[n,m]H_k[n-i,m-j], k = 1,...,K \tag{2.1}$$

Thus, a convolutional layer provides an output matrix for each filter, Y. Convolution operation is depicted in Figure 2.3.

Usually, CNNs are based on the consecutive application of several convolutional layers, which, are usually interspersed with pooling layers (see Section 2.1.2), which purpose is to sub-sample the data using a certain criteria.

- **Convolutional Layer's Setup**

  Filters and Stride A convolution layer is a neuron cluster connected to input images' sub-regions or outputs's sub-regions of the previous layer. The size of the step at which filters move through the image is called stride, [35].

  Feature Map The filter moves along the input, shaping the feature map. Furthermore, it uses the same collection of weights and biases for the convolution, [35].

  Zero Padding Technique based on adding zeros to the edges of the input image, in order to adapt the output size, [35].

  Output Size The output size of a convolution layer is determined by the image input size, the filter size and the stride. In certain software, if the output size does not completely cover the image, the software omits the remaining part of the image throughout the right and bottom convolution edge, [35].

  Number of neurons It is calculated as the product of the height and width of the output. The number of neurons in a convolution layer is defined by the multiplication of the map size and the number of filters, [35].

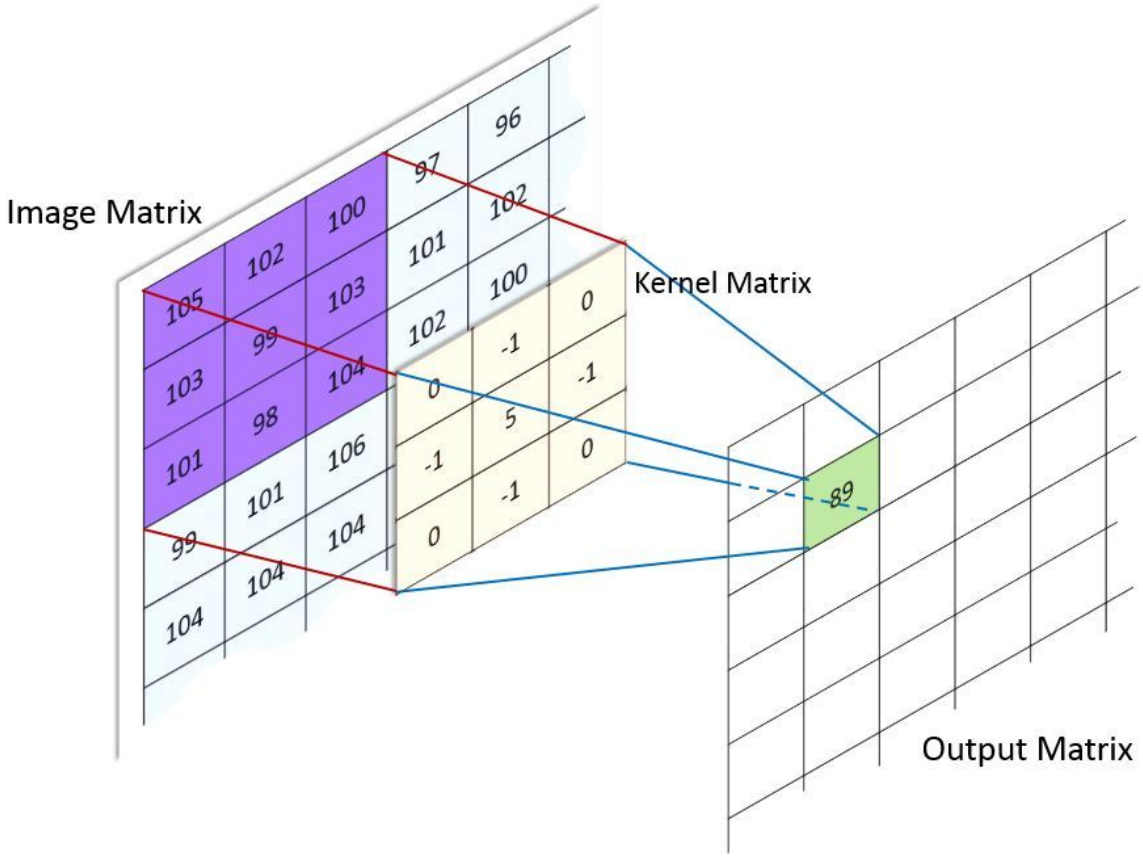

Figure 2.3: Example of convolution operation. Extracted from [10].

Including non-linearity dependence The activation function of a layer defines the range of possible outputs of its neurons, and its first derivative plays a key role in optimizing the cost function, [36]. The most frequent activation functions in neural networks are:

The **sigmoid function**, also called logistic function, is frequently used in other deep learning models such as logistic regression. Its output is bounded between 0 and 1, it is monotonically increasing and its derivative is larger for arguments closer to 0.

The **hyperbolic tangent** of a real number x is designated by $\tanh(x)$ and it is characterized as the quotient between the hyperbolic sine and the hyperbolic cosine of the real number x. With properties similar to the sigmoid function, it has an output range is (-1,1).

The **ReLu (Rectified Linear Unit)** function executes a threshold operation for each element, where every input value less than zero is reset to zero and positive values remain unchanged.

The **SoftMax function** is the activation function of the output unit after the fully-connected layer for classification problems. In the output classification layer, the SoftMax function associates each input to one of the different classes, generally employing the cross-entropy function. Also it is known as normalized exponential and it is a generalization of the sigmoid function to multidimensional vectors. The layer's output using SoftMax function is bounded between 0 and 1, and all the outputs of the layer add up to 1. Because of this, it is an adequate function to represent probability density functions.

Pooling Layer It is usually located after the convolutional layer. Its main utility is the reduction of the spatial dimensions of the input volume data for the next layer, a reduction which also allows to explore different scales of the input. In the resolution reduction process, the number of parameters is reduced by keeping the most common characteristics and it have not effect over depth dimension of the data, [35].

Fully-connected Layers Every neuron in a fully connected layer is associated to whole neurons in the previous layer. This layer merges all the characteristics learned by the previous layer, [35].

Evolution of Artificial Neural Networks

ANNs

The history of ANNs [37] starts in 1943, the year in which Warren Sturgis McCulloch and Walter Harry Pitts shaped the first neural network model by using mathematics and algorithms called threshold logic [38]. Shortly after his advances, Donald Hebb explained neuronal learning in broad strokes, knowing his law the *Hebb's rule* [39], which is the precursor to the neural networks' learning techniques use nowadays. Subsequently, in 1956, a conference held in Dartmouth was focused on the ability of computing to simulate the learning of biological brains. Shortly after 1960, the first application of ANN came to real problems: The elimination of echoes in telephone lines through the Adaline (Adaptative Linear Elements) and Madaline (Multiple Adaptative Linear Elements).

In 1969, Marvin Minsky and Seymour Papert wrote and published the book *Perceptrons: An introduction to Computational Geometry* [40], nonetheless the development in neural networks decreased suddenly since the book showed important deficiencies in ANN models that were developed up to that moment, especially of the perceptron. However, this fact did not hinder research in this topic, highlighting in 1977 the work done by James Anderson denominated *Brain-State-in-a-Box* [41], which allowed to model more complex functions. The resurgence of neural networks occurred in 1985 with the publication by John Hopfield of the book *Neural computation of decisions in optimization problems* [42]. A year later, in 1986, David Rumelhart introduced the Backpropagation algorithm (see Section 2.1.2) in its current version, which caused a new and much more encouraging panorama in the research and development of neural networks. Since then, we have observed numerous efforts in the development of neural networks, leading to substantial advance in diverse and varied assignments.

CNNs

Up to our knowledge, the first reported CNN is the *Neocognitron* [43], acquainted by Kunihiko Fukushima in 1980. This model was sharpened by Yann Lecun in 1998, by including learning based on Backpropagation algorithm, [44] (see Section 2.1.2). In 2012, these types of networks were refined by Dan Ciresan and they were implemented on a Graphics Processing Unit (GPU), achieving a better computational performance than those obtained until then [45]. In addition, the quality of the results obtained by this CNN (AlexNet) was higher than previous techniques in image classification, as Support Vector Machines (SVMs) [46] or Haar wavelet [47]. Since that year, more deeper models were research due to they have a memory footprint during interference, making it possible to use them in digital computing devices such as smartphones.

In the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) of 2012, AlexNet managed to improve its achievement over other non-deep learning methods. From that model,

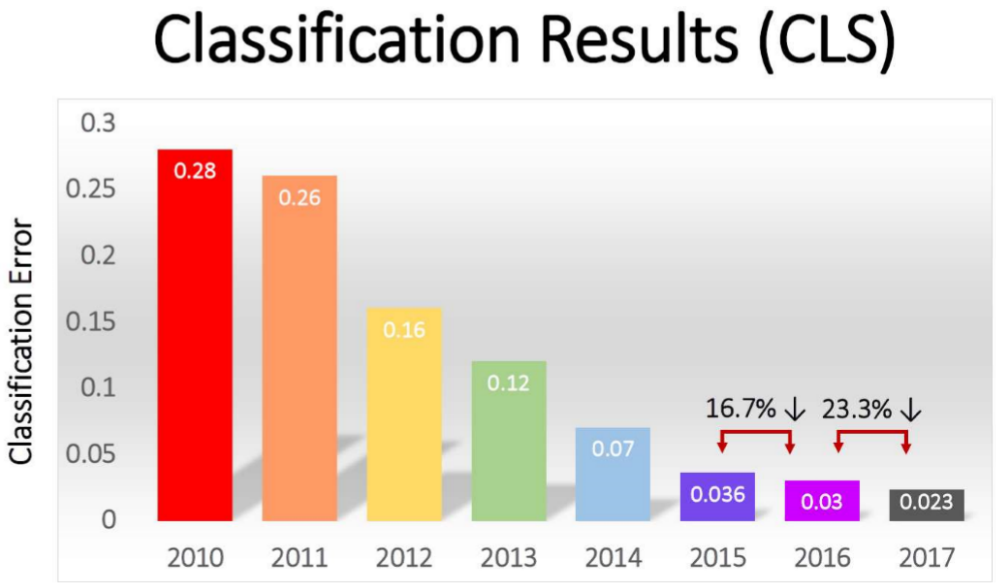

Figure 2.4: Classification results through time. Extracted from [11].

which achieved 57% accuracy at that time, progress has been made to a scenario in which the main models move around 75% accuracy in the ImageNet dataset (Top-1).

It was not until 2014, thanks to the work published by Simonyan and Zisserman, that the deepest networks began to be understood. They were the first to use fixed size filters in each layer. The ILSVRC competition of this year, VGG19 network or also called OxfordNet was presented and won the challenge.

The creation of deeper and deeper networks arose from the necessity to classify images with small and dark objects. Another deep networks improvement was that they introduced more complex blocks to increase the achievement of the training process and decrease the number of parameters to train.

In the same year (2014), Szegedy and coworkers [48] introduced GoogleNet, also called Inception, which is a 22-layer network, using blocks with a collection of convolutional filters with various sizes.

The ILSVRC challenge of 2015 was won by ResNet, introduced by Microsoft. The idea, very summarized, is based on increasing the number of layers by introducing a residual connection (with an identity layer). The identity layer passes directly to the next one, improving the learning process.

For medical applications, models such as AlexNet or VGG19 are commonly used. However, recent reference studies use a GoogleNet version called Inception v3 [8]. It is difficult to asses if the use of GoogleNet is due to a superior architecture or because the model is default option in the most successful software packages.

CNNs are the most generally used neural networks in the research of medical images although recurrent neural networks (RNNs) are beginning to gain popularity in this field. This kind of networks have been used for medical image processing since 1995.

At first, between the 1970s and 1990s, medical image evaluation was performed with the consecutive application of low-level pixel processing, i.e., they were used as edge and line filter detectors and to detect the growing region of images. Mathematical modeling was also used.

Symbolic artificial intelligence is the assembled name for all artificial intelligence research methods based on high-level "symbolic" representations of problems and mathematical logic. John Haugeland in 1985 coined the term GOFAI (good old-fashioned artificial intelligence) to describe this type of artificial intelligence. But this method remained imprecise with respect to image analysis.

Supervised techniques gained popularity in the late 1990s for developing medical image evaluation systems. A few examples involve active shape modes (for segmentation) and used of feature extraction and statistical classifiers (for Computer-Aided Detection (see Section 2.2.1) and examination). There was also a significant change in using computers to extract feature vectors from training data. This type of feature extraction algorithms is very important in order to obtain data that is sufficiently discriminative for the images. Nowadays, this kind of methods are still being developed to replace manual labelling with human beings.

To refer to these data, the Figure 2.4 represents the data in reference to the classification results over time of ImageNet challenge story using supervised learning (ImageNet have 1000 categories where there are 1000 images of training samples for each category and 100,000 images of testing samples). Observing how the classification error has decreased so far.

For all these networks, the training stage can be accelerated through the use of GPUs. Solutions based on deep learning techniques are generally dependent on NVIDIA's GPU-based computing. GPUs work very well processing in parallel and accelerate the execution of a CNN between 10 and 20 times, [35].

Training of Artificial Neural Networks

One of the most important ANNs and CNNs characteristics, is their ability to learn autonomously via a learning or training stage in which the network learns the most appropriate sets of values (weights and biases) to produce an output in agreement with a given assignment. The learning process is an iterative procedure where the quoted parameters are gradually refined —usually starting from random values—in order to reach a sufficiently good level of performance.

Training can be defined as the process that seeks to minimize the loss function. In many cases, the algorithm responsible for helping us look for these parameters is the Backpropagation algorithm [49] (see Section 2.1.2). The idea is to calculate, through it, the partial derivatives of the cost function for each of the parameters. Then, in the following iteration, the parameters are updated to progressively reduce the cost function.

Being W^t the weights' vector or model parameters at iteration t, and being $L(W)$ a cost function, the weights at each iteration are updated as:

$$W^{t+1} = W^t - \rho \nabla L(W) \tag{2.2}$$

, where the parameter ρ is the learning rate, and it regulates the update's rate of the weights based on the evolution of the cost function.

Cross-entropy loss In classification scenarios, training is usually guided by a cross-entropy loss function (see equations 2.3 and 2.4) that measures the performance of the network based on the predictions ($\hat{y}$) with respect to the assemblage of ground-truth annotations or labels (y). The value taken by the cost function will be lower the better the model's predictions are [36].

$$H(y, \hat{y}) = \sum^{i} y_i \log \frac{1}{\hat{y}_i} \tag{2.3}$$

$$H(y, \hat{y}) = \sum_{i}^{} H(y, \hat{y}) \tag{2.4}$$

A lower cross-entropy loss does not have a higher hit rate for a dataset but it gives more approximate probability distributions to the tags. In order to carry out Backpropagation algorithm (see Section 2.1.2), the cost function must be derivable concerning to the parameters of the model (weights).

Backpropagation Algorithm

The Backpropagation algorithm is used to calculate the output gradient (given by the cost function) with a set of weights, that can be updated in each iteration through optimization algorithms as gradient descent algorithm [33]. It is the core of supervised learning methods in which we can distinguish two phases:

- Propagation phase: In this phase, a certain input is given to the network, in such manner, it propagates from the first layer to the last of them, producing a certain output. Then, this output is compared with the expected output through the loss function—also called cost or error—.
- Adaptation phase: With the error calculated in the previous phase, an inverse propagation begins, from the last to the first layer. The propagation ends up reaching all the neurons that contribute directly to the output. However, it should be noted that each neuron receives only a portion of the error corresponding to its relative contribution to the original output. Knowing the contribution of each neuron to the error, in order to reduce the error in future iterations, weights can be adjusted.Then, The algorithm is effective in training neural networks, as it corrects the weights for the desired outputs are produced. During training stage, a training set is propagated through the network from the deeper to the shallower layers. A complete iteration of the training assemblage to the neural network is denominated an epoch. In this way, the learning process is repeated along several epochs until the parameters stabilize or a maximum number of epochs is reached.

The algorithm optimization based on gradient descent has received several modifications and revisions.

2.1.3 Optimization methods for deep learning

SGD Algorithm

The stochastic gradient descent (SGD) optimizer [50], also known as incremental gradient descent, is an iterative procedure for optimizing a differential objective function, a stochastic approximation of gradient descent. Stochastic algorithm does not need to remember which examples were viewed during the previous iterations, therefore it can process examples meanwhile the training stage.

SGDM Algorithm

The stochastic gradient descent with momentum (SGDM) optimizer [51] is a variation of the stochastic gradient descent used for faster convergence of the loss function. It uses a unique learning rate for entire parameters. This provides the gradient descent with a *momentum* to avoid being trapped at local minimum. However, the use of (too) large *momentum* together with an also large learning rate may produce a skip of a suitable solution (local minimum) due to a large updating step.

RMSProp Algorithm

The root mean square propagation (RMSProp) optimizer [9] use different learning rates for each parameter. Learning rates can automatically adapt to the loss function being optimized with the aim of improving network training. It manages a moving average of the element-wise squares of the settings gradients.

Adam's Algorithm

The derived from adaptive moment estimation (Adam) optimizer [52] insert *momentum* term and it uses a parameter update similar to RMSProp. The method estimates individual adaptive learning rates for diverse parameters from computes of first and second moments of the gradients. Some of Adam's benefits are that the amounts of parameter updates are invariant to re-scaling of the gradient, its steps sizes are almost bounded by the stride size hyperparameter.

2.1.4 Learning strategies

The accuracy of a CNN is widely conditioned by the amount of data used for training. That is, for a more accurate model, several training samples (i.e., thousands or millions) are needed, resulting in a training time that, depending on the hardware used and the complexity of the problem, may last for hours, days or even weeks.

Learning from Scratch

Learning from scratch stands for creating something starting with basic ingredients (the *Tabula rasa* concept). Whole the parameters of a CNN model are initialized with random Gaussian distributions, when learning from scratch.

Transfer Learning

Transfer learning [53] is a method where a pre-trained neural network is fine-tuned to solve a specific problem (target assignment) differently from the one used to train it (source assignment). This method is a practical way to apply deep learning without the need for a large dataset and a lot of training time, [8]. It can be observed the scheme of how the transfer learning works in Figure 2.5, using the Backpropagation algorithm (see Section 2.1.2). Sometimes, transfer learning is arranged in two different strategies:

- Using pre-trained network such as feature extractor.
- Fine-Tuning a pre-trained network on dataset.

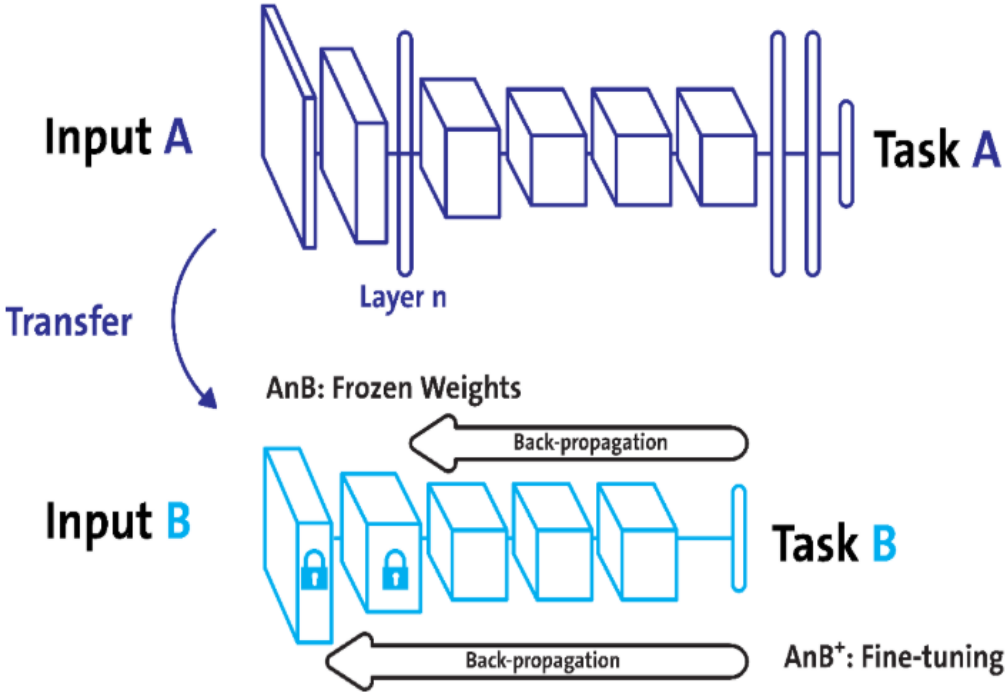

Figure 2.5: Transfer Learning scheme. Extracted from [12].

The benefits of these strategies depends on the source and target tasks. On one hand, in multi-class grade assessment of knee osteoarthritis, fine-tuning has achived 57,6% accuracy in contrast to feature extraction with 53,4%, [54]. On the other hand, in cytopathology image classification, feature extractor has obtained 70,5% accuracy meanwhile fine-tuning has obtained a 69,1%, [55]. However, sometimes is preferable training a CNN from scratch instead of transfer learning. For instance, for a small dataset of approximately a 1000 images of skin lesions training from scratch performs better than fine-tuning, [56].

Alternative Techniques

- **Domain Adaptation** It is used to adapt a distribution of target data (e.g., natural images) in the domain of distribution source data (e.g., synthetic images), and thus prepare networks to be efficient with images that have not been trained with. There are three main approaches for domain adaptation, that differ in the information examined for the target assignment: unsupervised, semi-supervised and supervised. In unsupervised domain adaptation, we try to create a network that extracts features which are the same for all domains. Therefore, the network tries to make the characteristics of the images sufficiently discriminatory to be representative of the origin domain. Differently, in semi-supervised domain adaptation, it is also considered a "small" collection of labeled target examples, whereas in supervised domain adaptation, all the images contemplated are supposed to be labeled, [53].

- **Self-Supervised** Self-supervised learning is an autonomous supervised learning. It is a depiction learning approach that purpose to eliminate the pre-requisite of involving humans to label data. It is a supervised learning because its objective is to learn a function from couple of inputs and labeled outputs. However, explicit use of labeled input-outputs combination in self-supervised learning is not needed, [57]. Self-supervised learning sys-

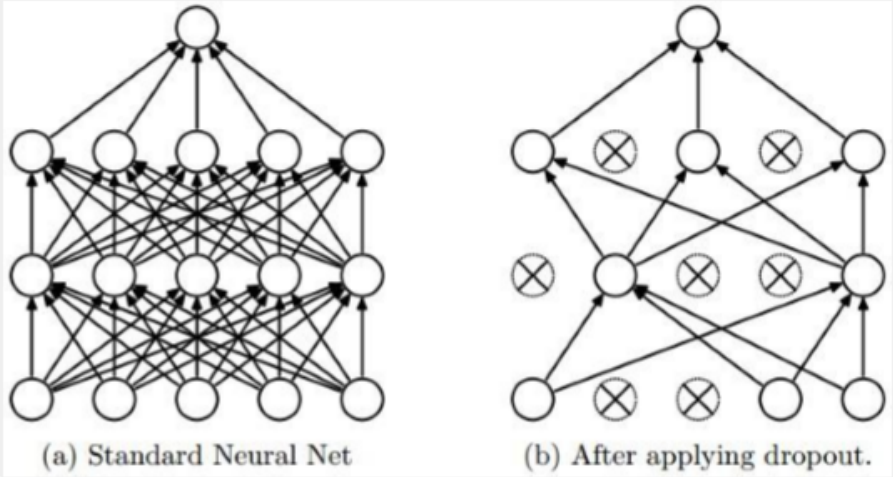

Figure 2.6: Example of dropout. Extracted from [13].

tems extract and use the naturally available relevant context and embedded meta-data as supervisory signals shaping, [58].

2.1.5 Augmented Learning

Overfitting is the process that takes place when the learning model is (too) adjusted to the training data, hindering its ability to generalize to new unobserved data. To find out if there is overfitting meanwhile training stage, the training dataset is usually divided into a train and a validation set. A efficient indicator of overfitting can be derived from the study of the loss function: the loss function increases for the validation data, but still decreases for the training data.

One of the most effective methods to avoid overfitting is to increase the amount of training data. To this aim, a common strategy is to apply minor alterations to the data (data augmentation). These minor changes include flips, translations and rotations, [59]. Through these process, new images, which inherit the labels from the originals, are created. The alterations needs to be varied and diverse, as otherwise, the neural network may learn the transformations instead of the goal task.

Thanks to data augmentation we can reduce overfitting in model by increasing the volume of training data.

Dropout

A dropout strategy randomly places the input elements of a given layer at zero with a previously established probability. In this way, for each element of the input, a subset of neurons is randomly selected, implicitly creating different connection patterns in the architecture, as it can be seen in Figure 2.6. Dropout helps preventing overfitting, because learning process does not lean on neurons and specific connections, [35].

2.2 Detection and Classification of lung nodules

Research into the development of a system capable of detecting, segmenting, and diagnosing isolated lung nodules from CT began in 1980. Nodule morphology (malignancy, size, internal structure, etc), location, and type of cancer are significant in the classification. Thus, the detection and classification of nodules has become a challenge, often with different methods, each handling an aspect of the problem.

2.2.1 Detection of lung nodules

A nodule is described as a mass of cells measuring between 3 mm and 30 mm in diameter. The concept *micro-nodule* is used for tumours of less than 3 mm in diameter and the term *mass* is used for tumours with a diameter larger than 30 mm. The nodule diameter is usually related to malignancy, i.e., malignant nodules are related to a large size. For instance, according to Igor Rafael S.Valente et al. [14] "In ELCAP dataset, the percentage of malignancy associated with nodules is as follows: 1% for nodules with a diameter smaller than 5 mm, 24% for nodules with a diameter between 6 and 10 mm, 33% with a diameter between 11 and 20 mm and 80% for nodules with a diameter greater than 20 mm".

Physicians measure the diameter of the nodules in order to make a prognosis of the cancer. Therefore, it is very common to make mistakes when calculating the diameter of asymmetric or non-spherical nodules. In addition, if the nodule is very small, the image must be maximized to be able to measure it accurately, [60].

Despite initiatives to encourage the use of early diagnosis systems in hospitals, radiologist can make mistakes when using them. Factors like fatigue, distraction, inexperience, or limitations of the human visual system can affect physicians, making an inaccurate analysis. For the above reasons, the development of automatic nodule detection methods is necessary.

The objective of screening is to adapt the images, highlighting the existing anomalies to help radiologists, aiming to aid cancer detection in an earlier stage when therapeutic treatment offers better perspectives. The employment of screening would imply a symbolic decrease of reading effort beneficial to radiologists, [14].

The number of network parameters and the memory required for the computation are fundamental requirements for increased accuracy in nodule detection.

In this context, the development of deep architectures where the input are images with a low-scale representation and a high resolution is emerging. To improve image generation, CNNs are used to transform an input image into another, [8].

Computer-Aided Detection (CAD)

To assist radiologists in the reading process of CT scans, the Computer-Aided Detection (CAD) systems [60] have been developed.

The CAD system commonly involve two stages (as it can be seen in Figure 2.7): (1) nodule candidates' detection called detection system (CADe), (2) nodule candidates' classification called classification system (CADx).

The CADe system is used to identify abnormalities in CT images. In this way, radiologists could evaluate these regions of interest, [16].

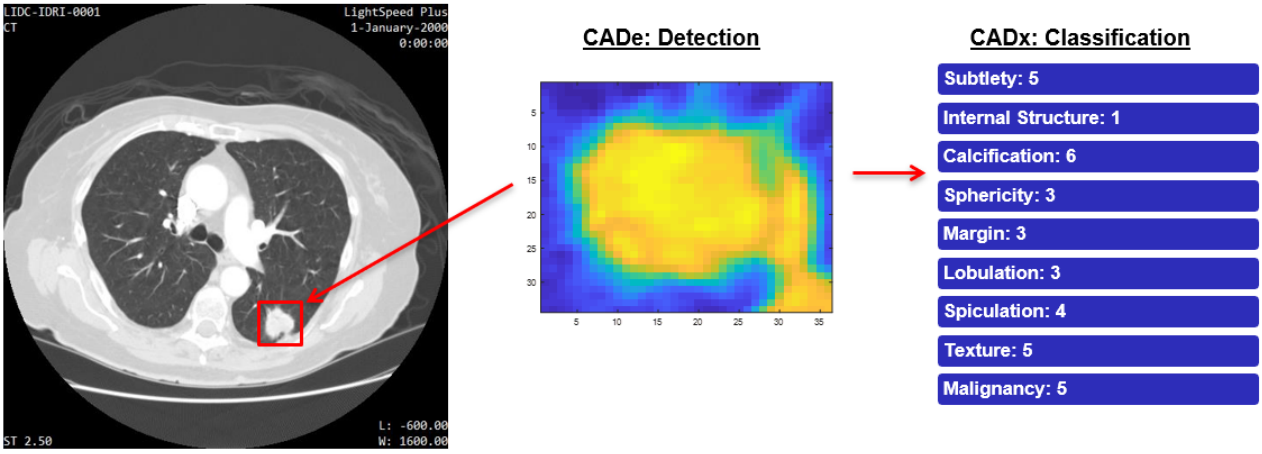

Figure 2.7: Computer-Aided Detection (CAD) system process.

The objectives of CADe can be summarized as follows:

- Enhance evaluation efficiency and assist in obtaining high accuracy.
- Decrease the number of false positives (FP) as result of errors in CT scan analysis. Consequently, represents a reduction in human examination time by radiologists.
- Provide fast processing.
- Provide an automated process where manual procedures are not necessary, e.g., the system must be able to automatically take the Digital Imaging and Communications in Medicine (DICOM) exam files, and develop and save the final results in standard files.
- Submit low implementation costs, training, backing and supply.
- Identify distinct types, sizes and shapes of nodules, included isolated nodules, micronodules, sub-solid nodules or nodules associated to lung edges.
- Increase system security to prevent data loss, inaccurate results, unavailability or incorrect use of data.

A CADx system is designed to help radiologists in the identification of the disease as well as the type, severity, stage, progression or regression of lung cancer. A CADx system can either use only image describing or it can straddle other relevant information to the diagnosis, [14].

Despite improving the reading efficiency of radiologists, CAD systems do not detect a considerable number of nodules, making it impossible to use them in daily medical practice.

It should be noted that there are other types of lesions that are visible on CT scan and they are not nodules. Due to the absence of symptoms until the disease is in an advanced stage, datasets often have nodules with a high malignancy. In this way, the different categories of nodules are not balanced.

In summary, extracting sufficiently discriminatory features for the detection algorithms is very difficult. It is necessary to develop techniques that allow the detection of any type of lesion and thus improve the performance of CAD systems.

In supervised learning surveys, CNNs are trained to learn highly discriminatory features, eliminating the requirement to label lung nodules in CT image manually. It has been shown to reduce the number of false positives in CAD systems, [60].

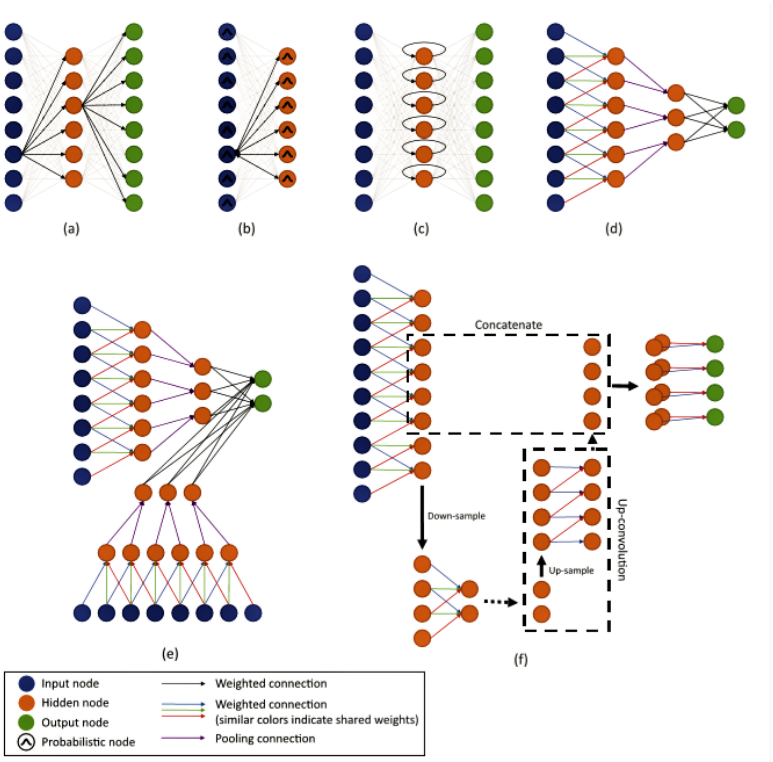

Figure 2.8: "(a) Node graphs of 1D representation of architectures commonly used in medical imaging. (b) Auto-encoder, (c) Restricted Boltzmann machine, (d) Recurrent Neural Network (RNN), (e) Multi-stream convolutional neural network, (f) U-net (with a single down-sampling stage)." Figure and caption extracted from [8].

2.2.2 Classification of lung nodules

The main contribution of deep learning in the medical image analysis field was the classification of images. The most common approach is to introduce multiple labeled images into the network and as an output have a single diagnostic variable (e.g., existence of the disease or not). The medical dataset is usually small compared to the datasets used in computer vision (e.g., hundreds/thousands vs. millions of samples). To solve this problem, it is used transfer learning (see Section 2.1.4), [61].

The evolution of computer image analysis has been parallel to deep networks progression. The first study employing deep neural networks for image classification arose in 2013 and was applied on neuroimaging, [62].

In 2015, 2016 and 2017, 47 were published studies related to the classification of medical images. Among these studies, 36 used CNNs, 5 used AEs (Autoencoder) and 6 used RBMs (Restricted Boltzmann Machine). These mentioned methods were used to classify nodules in brain MRI (Magnetic resonance imaging), in retinal imaging and in CT scans (lung cancer), [8].

Some of the methods used to classify nodules using the LIDC-IDRI dataset (Section 3.1) are the following, in chronological order:

In the study [63] raised in 2013, the authors use the Haralick method on 2D nodule images to calculate the texture characteristics of the nodules. The algorithm was proposed in 1973 by Robert M. Haralick and it is based on the calculation of a set of 14 texture measurements based on the spatial dependence of gray tones. The extraction of the spatial dependence matrices of gray tones of the CT image is performed through the angular relationships and distances between pixel pairs of the image (or between cell pairs) on each slice of the CT scan. As mentioned in

this section, the morphology of nodules is fundamental in distinguishing between benign and malignant nodules (e.g., malignant nodules have a large size, spiculated edges and are very contrasted with the environment). In order to obtain more robust nodule characteristics, it is proposed to combine boundary annotations of the same nodule of at least three radiologists to create a volumetric nodule. Therefore, they overlap the areas of the tagged pixels inside the region of interest defined by each radiologist for the same 2D nodule, creating the 3D nodule.

The proposed research with RNN was in 2015 and its objective was to assemble new images related to the samples of the training assemblage. In the natural images *pixelRNNs* are used as auto-regressive models. This type of network has been used in image segmentation problems in the area of medical analysis, [8].

The use of a novel 3D-CNN for the classification of 3D volumetric data (i.e., 3D nodes) is proposed in 2015. It was used to categorize candidates on lung nodules candidates on circulating tumor cell (CTC) [64]. In the studies [65] [66] [67], it is proposed that CNNs use supervised feature learning to classify medical images with labels.

In the research [65], authors describe a Multi-scale Convolutional Neural Networks (MCNN) architecture in 2015. MCNNs are CNNs with the difference that pooling layers adopt different methods with different areas throughout the network. It is proposed to capture the diversity of nodules by extracting discriminatory characteristics from the stacked layers of the net. For this purpose, multi-scale nodule patches are used to learn a specific set of characteristics simultaneously by chaining response neuron activations obtained in the last layer from each different scale of the input image.

In the research [68], authors introduced a deep learning architecture based on a convolutional auto-encoder neural network (CANN) for the classification of lung nodules. According to these method, subsections of the input image (also called patches images) are extracted. Each patch achieves a classification result after extracting the characteristic of a particular area through the network. This approach achieves faster tagging of medical data by subsections and it is also very effective in extracting features from the image. Auto-encoder with unsupervised learning is used to determine network parameters and a small amount of tagged data is used to fine-tuning the settings of CANN and training the classifier.

In the study [69], the authors describe an method based on the structural co-occurrence matrix (SCM) to classify lung nodules according to their malignancy. The images on which SCM is applied can be in grayscale (common CT scan) or scanners can be filtered by one of the following four types of filters: mean, Laplace, Gaussian, and Sobel.The classifiers used have been: multilayer perceptron, support vector machine, and k-nearest neighbors algorithm. The objectives of this study were: to classify nodules according to their malignancy (i.e., malignant or benign) and also according to their stage of malignancy (from 1 to 5).

In the research [70], authors used transfer learning over the AlexNet network. They modified network'structure by eliminating the last fully connected layers and modifying the classification layer to catalogue lung nodules into three classes. The authors used the confusion matrix on a different dataset to assess network performance.

Some of the methods explained in this Section are represented in Figure 2.8.

3
Material and Methods

In this chapter we describe the dataset and the convolutional neural network used in this project.

3.1 Dataset: LIDC-IDRI

3.1.1 Specification of the dataset

Lung Image Dataset Consortium image collection (LIDC-IDRI) is a public dataset of lung cancer screening thoracic computed tomography (CT) scans with annotated lesions. LIDC-IDRI contains a set of 1,018 patients. This dataset has been created thanks to the collaboration of public and private medical institutions with the aim of improving medical research in the area of lung cancer. In 2004, Image dataset Resource Initiative (IDRI) was created by the National Cancer Institute (NCI) and the Foundation for the National Institutes of Health (FNIH). Five institutions, two academic centers and eight medical imaging companies joined the initiative, supplying the resources to increase the number of cases in the dataset, finally creating the LIDC-IDRI dataset. According to Igor Rafael S.Valente et al. [14] "The dataset consists of 7371 nodules and 2669 nodules $\geq$ 3 mm marked by at least one radiologist of which 928 (34.7%) received the same ratings from all four radiologists."

The scan depth of CT images ranges from 0.6 mm to 5.0 mm with a median of 2.0 mm. Every patient folder includes a related XML file. Each XML file have the conclusions of a two-phase image annotation process evaluated by four experienced thoracic radiologists. The first phase consists in blinded-read of the nodules of each CT. In other words, each radiologist independently examined the CT scans and classified marked lesions in one of three different types ("nodule $\geq$ 3 mm", "nodule $<$ 3 mm", and "non-nodule $\geq$ 3 mm"). In Figure 3.1 we can recognize the different variety of nodules and not nodules. After that, on the second phase (unblinded-read phase), each radiologist individually reviewed their own annotations and the anonymous marks of the three other radiologists. Finally, each radiologist decided to either accept or reject the annotations.

Nodules may be primary lung cancers, metastatic disease, or non-cancerous nodules; since a decision on whether a lesion is a "nodule" should not be made based on presumed histology, the presence of calcification should not be a factor in determining the "nodule" status of a lesion.

For the following features: malignancy, internal structure, lobulation, margin, sphericity,

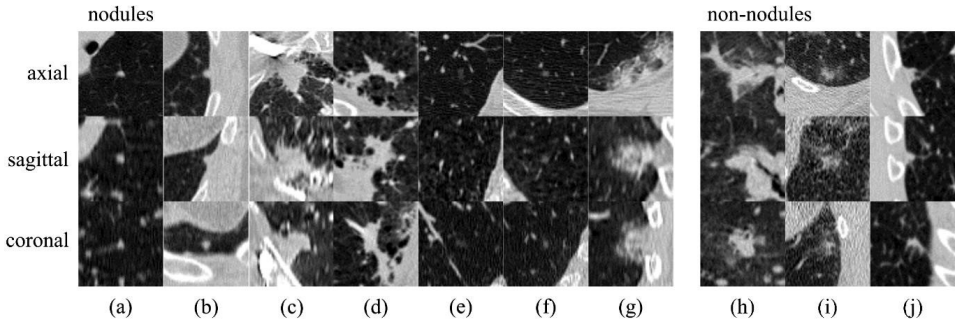

Figure 3.1: "Examples of lesions (nodules and non-nodules) in axial, sagittal, and coronal view. Lesions are in the center of the box (50 50 mm). The left set of images are nodules with a wide range of morphological characteristic: a) solid nodule, b) pleural nodule, c)–d) large nodules with irregular shape, e) –f) sub-solid nodules. The right set of images are irregular lesions that are not related with nodules or cancers. These examples illustrate that designing features for accurate detection and classification of nodules may not be trivial." Figure and caption extracted from [14].

spiculation, subtlety and texture, lung nodules are classified with levels from 1 to 5 (1 denotes the smallest degree and 5 denotes highest degree), but for the calcification characteristic the classification range is from 1 to 6, [15].

In addition to nodules and non-nodules, six lung tissue types can be found in the dataset. They are represented in Figure 3.2.

The lesions are classified in four different categories [15]:

Nodules which diameter is greater than 3 mm, but smaller than 30 mm. For these lesions, each radiologist drew a complete region of interest (roi) enclosing, but generally not touching, the nodule. Annotations of the nine characteristics are provided for each nodule and each radiologist. Since clinical information is not provided, radiologist define potential malignancy using a 60-year-old male smoker as reference.

Nodules which diameter is smaller than 3 mm. For these lesions, if the opacity is clearly benign (in radiologists' opinion), none information is provided. Otherwise, only an approximated three-dimensional center-of-mass is provided. No subjective assessment of characteristics is provided for both cases.

Non-nodules which diameter is greater or equal to 3 mm. Again, for these lesions, only an approximated three-dimensional center-of-mass is provided and no subjective assessment of characteristics is given. To distinguish these nodules from the previous ones, a special tag is provided. A mass that exceeds 30 mm should be marked as one of this type.

Non-Nodules which diameter is smaller than 3 mm. These lesions were not considered.

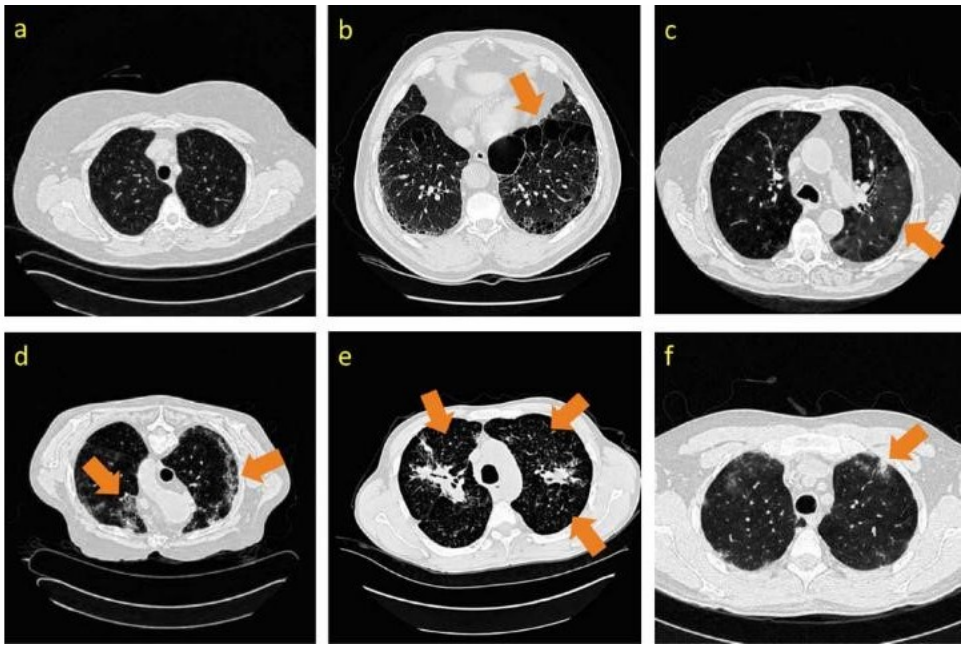

Figure 3.2: Examples of the six types of lung tissue in the LIDC-IDRI dataset [15]. "Disease tissue types are mapped with orange arrows. a): healthy; b): emphysema; c): ground glass; d): fibrosis; e): micronodules; f): consolidation." Figure and caption extracted from [16].

3.1.2 Definition of the annotation structure

As mentioned in the previous section, four radiologists did two-phase image annotation and results were saved in XML files using the following format [15]:

- <readingSession> indicates the index of the reading session. There will be up to four, one for each radiologist of each CT series.
- <Version> indicates version number and <StudyInstanceUID> indicates reader id. All reader ids are set to anonymous, but there are only four.
- For each nodule $\geq$ 3 mm:
  - The field <unblindedReadNodule> indicates the beginning of the section.
  - The field <noduleID> is a unique id for the nodule marked by each reader.
  - In the field <characteristics> a radiologist assessed characteristics of the nodule.
  - <3droi> indicates the complete description of the three dimensional contour of the nodule. This information was not available at the beginning of the project.
  - <imageZposition> is used to report the list of z positions (longitudinal direction) on which the nodule can be visualized.
  - A nodule's view (one per slice at each reported z position) is described by:
    a <roi> contains the roi enclosing the two-dimensional nodule's view on the slice.
    b <ImageZposition> reports the current slice.
    c <imageSOP_UID> specifies the name of the corresponding folder where image is contained.
    d <inclusion> This tag indicates if the nodule' view has a clear and well-contrasted contour ("inclusion = false") or a poorly contrasted blurry contour ("inclusion = true").

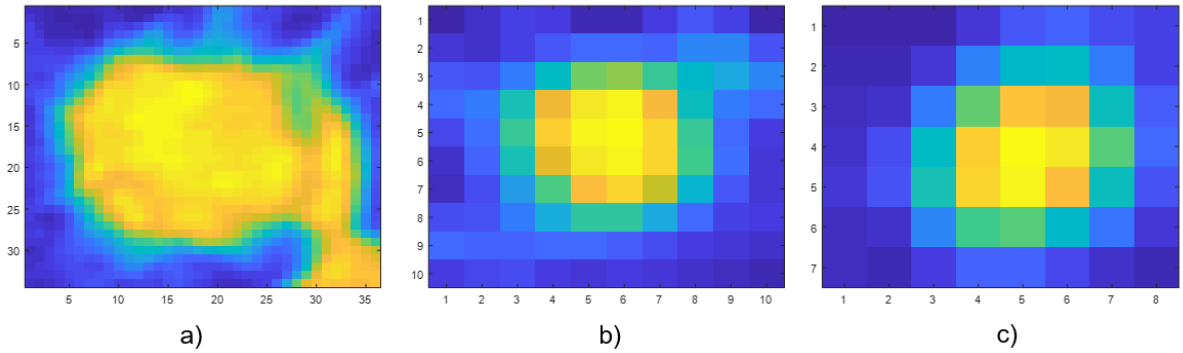

Figure 3.3: Degree of malignancy of three lung nodules: a) Malignancy: 5 b) Malignancy: 3, c) Malignancy: 1.

e <EdgeMap> includes the list of (x,y) pairs that describes the connected set of points shaping the nodule contour.

- For each nodule which diameter is smaller than 3 mm, only a single point is reported (in the roi label) and there are not radiologist-assessed characteristics. Special fields of these nodules include:

  a <roi> contains a single z position and a single spatial point representing the approximate centroid of that nodule.

  b <EdgeMap> contains the same information than roi.

- For each non-nodule, also only a single point is described and there are not radiologist-assessed characteristics. The data structure recorded for this type of lesions are similar to the information described for the previous nodules but have a unique identification: <nonNodule> tag.

3.1.3 Annotated characteristics

The characteristics [71] analyzed and assessed by the radiologists in the dataset are the following:

- Calcification appearance of the nodule: Examples of benign patterns of calcification are diffuse, central, laminated or popcorn. The remaining patterns can be considered as a sign of malignancy. The exception to the rule is when it is known to be a primary tumor.

- Internal structure of the nodule, describing if it is solid, semisolid, cystio or cavitated.

- Lobulation or lobular shape indicates, for instance, if the nodule has in-petals-like contour shape. Not-lobulated contours are an indication of benignity.

- Malignancy or likelihood of malignancy is associated with huge nodule size. The nodule is more likely to be benign if it have small size. The Figure 3.3 shows different degrees of malignancy. We can observe that some characteristics (nodule size, texture and spiculated or lobed margins) clearly indicate the degree of malignancy of a lung nodule.

- Margin describes the contour contrast. Examples of margin descriptors are: Corona radiata sign —close associated with malignancy—, scalloped margins —intermediate probability of malignancy —and smooth margins that are more likely associated with benign nodules, unless the patient suffers metastatic in origin.

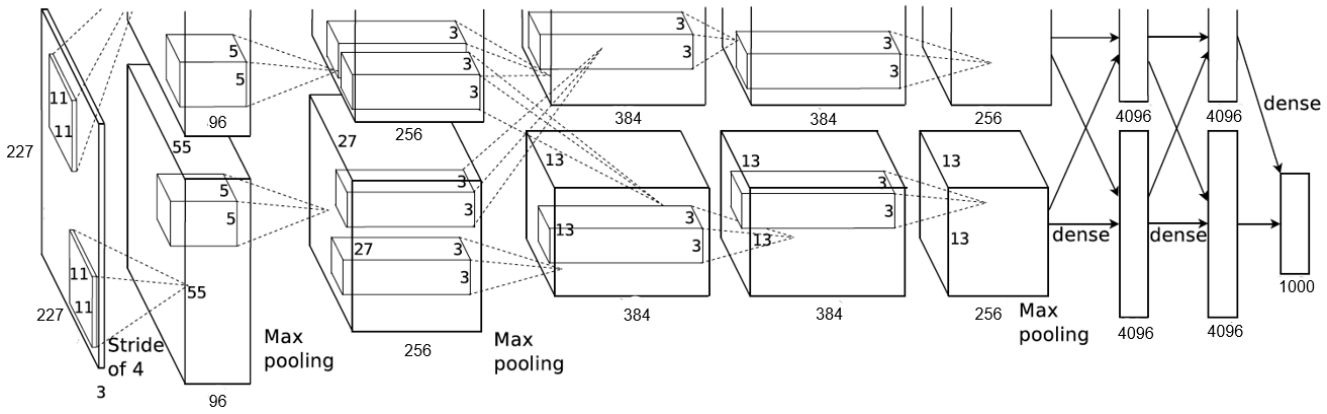

Figure 3.4: AlexNet architecture. Adapted from [17].

- Sphericity defines how spherical is the nodule shape.
- Spiculation indicates the degree of spike-like structures. For instance a border—spiculated margin is an signal of malignancy.
- Subtlety or distinction in detection based on the contrast between the lung and its surroundings.
- Texture defines the internal density of the nodule-texture. It is an important factor when attempting to classify a nodule. Accurate delineation of nodule boundary may be complicated if the nodule have part-solid and non-solid texture.

3.1.4 Alternative datasets

During the work we found other public datasets, interesting to process in the future. The alternative more complete public datasets available include: the Early Lung Cancer Action Program (ELCAP) [72], Nederlands Leuvens Longkanker Screeningsonderzoek (NELSON) [73] and the Automatic Nodule Detection 2009 (ANODE09), [74], [75].

3.2 Explored CNN architecture

AlexNet [61] network was created in 2012 quite similar to the one developed by Yann LeCun et al. called LeNet [76]. Both networks were relatively shallow. The main achievements of AlexNet neural network are: the incorporation of ReLUs activation functions instead of hyperbolic tangents for the incorporation of non-linearity to data, dropout as a regularization method to avoid overfitting and overlapping of pooling layers to reduce network size.

AlexNet has a total of twenty-five layers, of which five are convolutional layers and three are fully-connected layers. ReLU activation function is applied after all convolutional layers and dropout regularization is applied just before the first and second fully-connected layer.

To design and develop CNN, AlexNet was originally written in CUDA (Compute Unified Device Architecture) to be executed on GPUs. AlexNet CNN input is designed to receive 227x227 pixel images and three RGB channels. Therefore, there will be combinations of convolution layers (conv), ReLU (relu), Croos Channel Normalization (norm) and Max Pooling (pool) to carry out the learning of the characteristics. The first two convolution layers (conv1, conv2) are

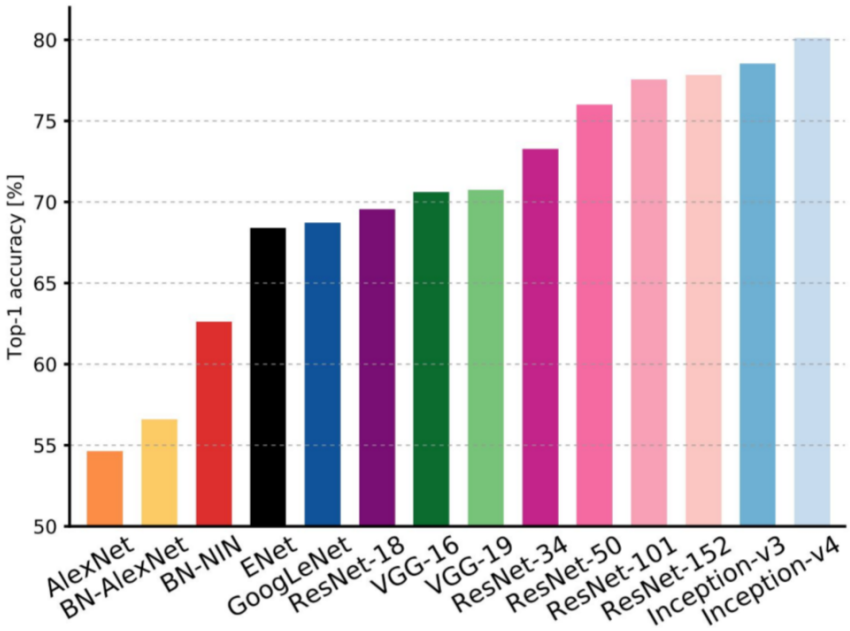

Figure 3.5: Comparative of different convolutional neural networks on ImageNet. Extracted from [18].

followed by ReLU layer (relu1, relu2) and a max pooling layer (pool1, pool2) and the following three convolutional layers (conv3, conv4 and conv5) are directly connected with three ReLU layers after each convolutional layer(relu3, relu4 and relu5). The last convolution layer (conv5) is followed by a single ReLU layer (relu5) single max pooling layer (pool5). Then a series of three fully-connected layers (fc6, fc7 and fc8) with output to a Softmax classifier (prob) with 1000 class labels.

Table 3.1 agglutinates the parameters of each AlexNet neural network layer. The architecture described above can be seen in Figure 3.4.

AlexNet was trained with the ImageNet dataset, specially designed for use in object recognition systems and containing more than one million images and 20,000 classes.

The Figure 3.5 shows a comparative of several CNNs on the ImageNet image classification task. The one selected, AlexNet, is the least complex and also the one with the worst performance [17]. However, AlexNet has a low capacity and can be considered a shallow CNN, which turns it a valuable and easy-to-test architecture to carry the initial experiments out.

3.2.1 Defining a CNN architecture

To create a CNN architecture [77] one should be aware of the parameters that define and relate each of the networks' layer. Assuming a spatial square input to a convolutional layer, the side length of its output, O, is given by:

$$O = \frac{I - K + 2P}{S} + 1$$

, where:

- I is the side length of the input.
- K is the side length of the square-kernels used in the convolutional layer.
- N is the number of kernels.
- S is the stride of the convolution operation.
- P is the (optional) Padding used on the input.

The number of channels in the output is equal to the number of kernels N. For instance, for AlexNet (Section 3.2), the input image size is 227x227x3. The first convolutional layer has 96 kernels of size 11x11x3. The stride is 4 and padding is 0. Therefore the output image size just after the first set of convolutional layers is 55x55x96 (one channel for each kernel).

The Size of the Output Tensor of a pooling layer is given by

$$O = \frac{I - P_s}{S} + 1$$

, where:

- I is the side length of the input.
- S is the stride of the convolution operation.
- P $_s$ is the pooling factor.

The AlexNet pooling layer after the set of convolution filters, has an input size of 55x55x96, pool size of 3 and stride of 2. The output after the pooling layer is of size 27x27x96.

Finally, a fully-connected layer achievement a vector of length equal to the number of neurons in this layer.

Size / Operation	Filter	Depth	Stride	Padding	Number of Parameters	Forward Computation
3* 227 * 227						
Conv1 + Relu	11 * 11	96	4		(11*11*3 + 1) * 96=34944	(11*11*3 + 1) * 96 * 55 * 55=105705600
96 * 55 * 55						
Max Pooling	3 * 3		2			
96 * 27 * 27						
Norm						
Conv2 + Relu	5 * 5	256	1	2	(5 * 5 * 96 + 1) * 256=614656	(5 * 5 * 96 + 1) * 256 * 27 * 27=448084224
256 * 27 * 27						
Max Pooling	3 * 3		2			
256 * 13 * 13						
Norm						
Conv3 + Relu	3 * 3	384	1	1	(3 * 3 * 256 + 1) * 384=885120	(3 * 3 * 256 + 1) * 384 * 13 * 13=149585280
384 * 13 * 13						
Conv4 + Relu	3 * 3	384	1	1	(3 * 3 * 384 + 1) * 384=1327488	(3 * 3 * 384 + 1) * 384 * 13 * 13=224345472
384 * 13 * 13						
Conv5 + Relu	3 * 3	256	1	1	(3 * 3 * 384 + 1) * 256=884992	(3 * 3 * 384 + 1) * 256 * 13 * 13=149563648
256 * 13 * 13						
Max Pooling	3 * 3		2			
256 * 6 * 6						
Dropout (rate 0.5)						
FC6 + Relu					256 * 6 * 6 * 4096=37748736	256 * 6 * 6 * 4096=37748736
4096						
Dropout (rate 0.5)						
FC7 + Relu					4096 * 4096=16777216	4096 * 4096=16777216
4096						
FC8 + Relu					4096 * 1000=4096000	4096 * 1000=4096000
1000 classes						
Overall					62369152=62.3 million	1135906176=1.1 billion
Conv VS FC					Conv:3.7million (6%) , FC: 58.6 million (94%)	Conv: 1.08 billion (95%) , FC: 58.6 million (5%)

Table 3.1: AlexNet architecture parameters. Adapted from [17].

4 Experiments

4.1 Experimental context: dataset

The experiments described in this section have been performed on the isolated nodules in the LIDC-IDRI dataset described in Section 3.1. However, the dataset has been adapted for this purpose due to several reasons:

- Mislabeled patients: some patients are missing patient labels.
- Missing annotations: some XML files from some patients are missing.
- Removal of small nodules (less than 3 mm of diameter): these nodules are not labeled to any of the dataset characteristics (e.g., no malignancy information is provided for these nodules).

After the adaption, the dataset is composed of a total of 589 patients (out of 1018 registered cases). Patient's nodules are assigned a different identification and labels by each of the four radiologist (see Section 3.1.2), therefore, we consider each annotation as an independent nodule, ending up with a total of 6225 *nodules*. As each nodule is visible in several slices these *nodules* produce a total of 26982 *annotated nodule images* (e.i. roi annotations).

Labels or characteristics of the nodules and the nodule images are arranged as depicted in the Figures 4.1 and 4.2 respectively. Both figures seem very similar but are in fact different as the number of slices per nodule changes with the nodule.

We randomly divide these samples into two subsets, training (70% of the samples) and validation set (30% of the samples). In the experiments we focus on the malignancy characteristic, as it is the one better balanced (see Figure 4.3).

Discussion on dataset Due to the privacy and security statutes, it is a huge effort to develop an appropriate labeled screening dataset for CT analysis assistance. Also mention that the preprocessing of the dataset was a difficult task because we have to understand how to read a CT and how the machine works in order to takes the images (it takes images of transverse cuts of the internal organs from different angles).

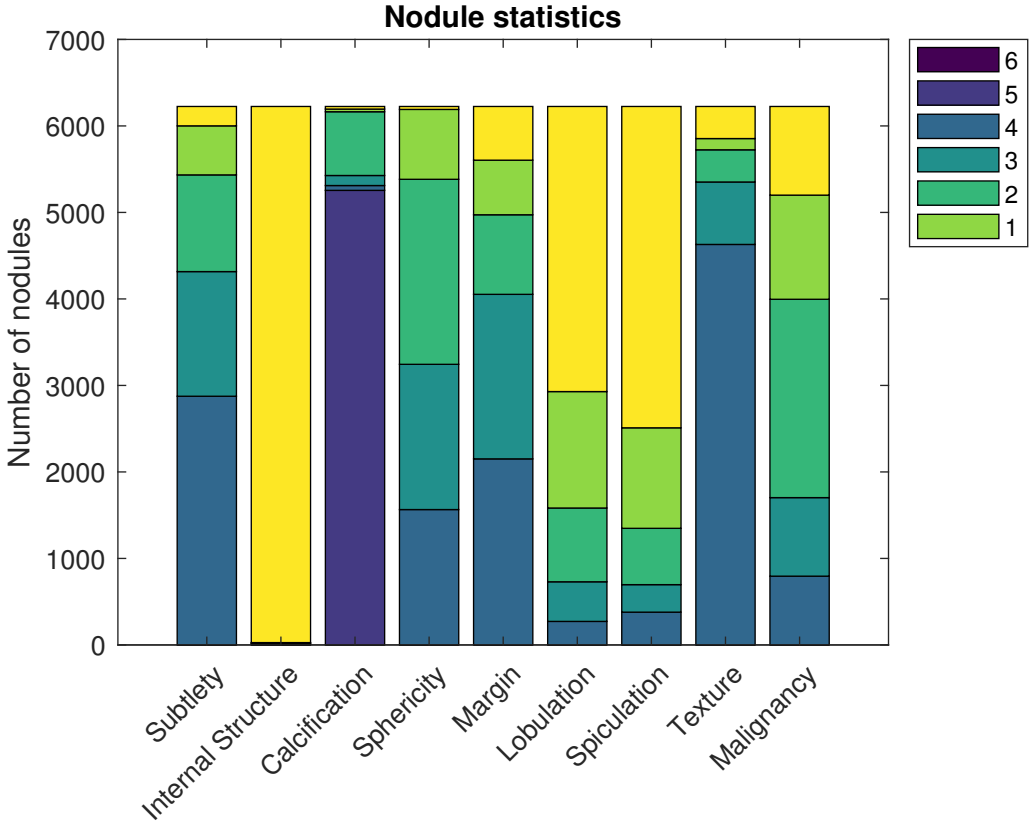

Figure 4.1: Dataset node statistics.

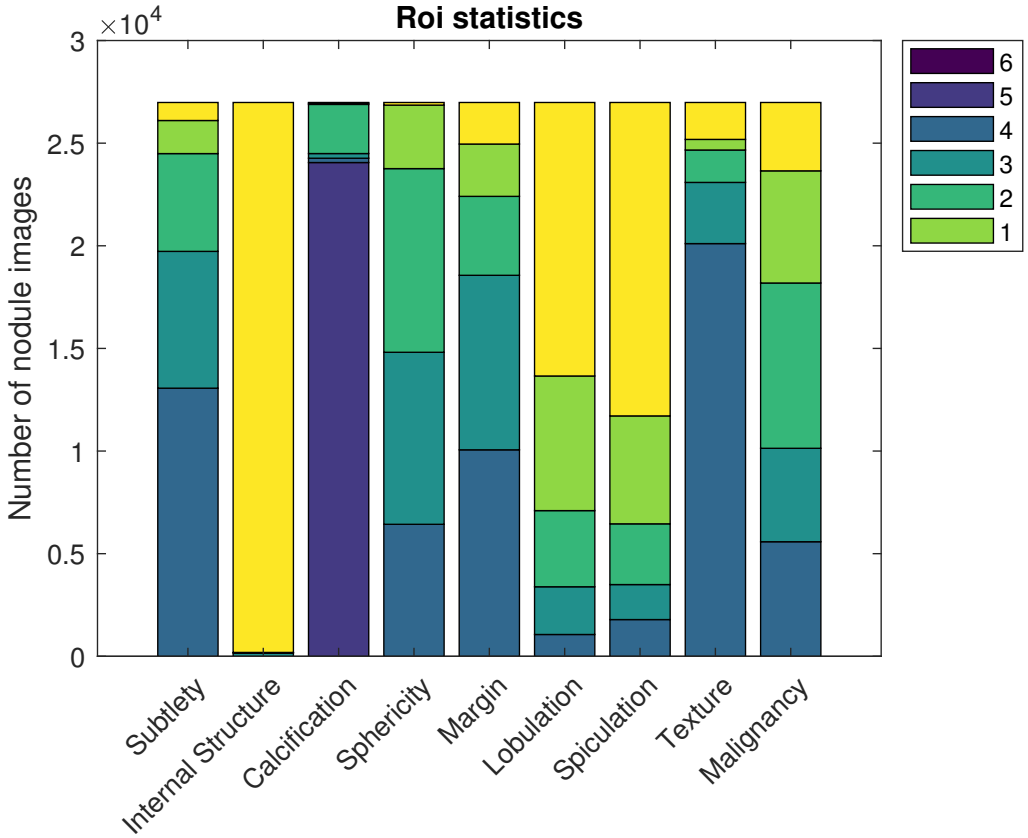

Figure 4.2: Dataset roi statistics.

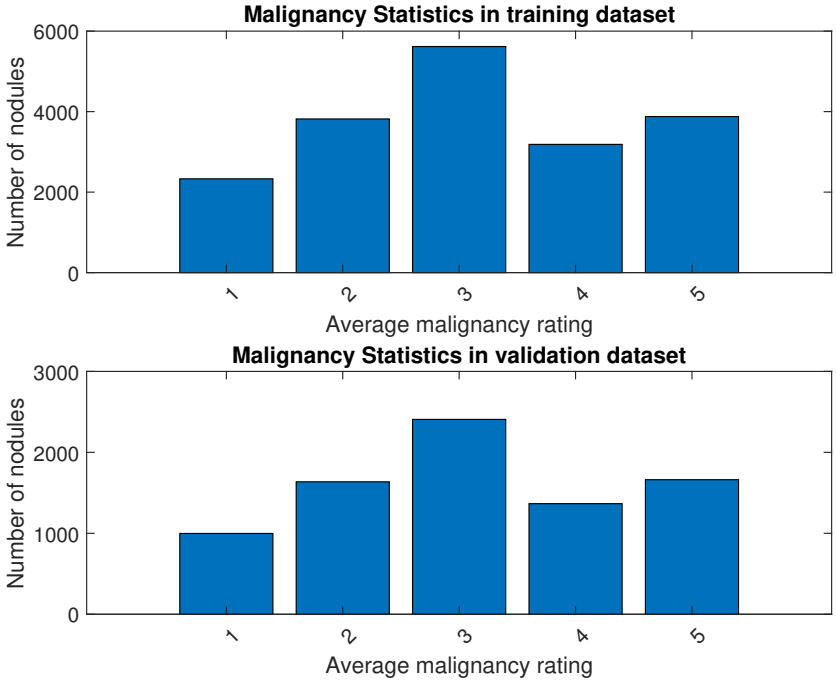

Figure 4.3: Malignancy statistics in Training and Validation dataset.

4.2 Experimental setup

4.2.1 Experiments description

The purpose of this study is to initially evaluate the performance of simple deep learning methods in predicting nodules' malignancy by analyzing CT images. Prior to classification, we assume that nodules have been previously segregated (segmented). To bypass this process, we make use of roi annotations, which are provided in the shape of enclosing bounding-boxes (see chapter 3.1).

We propose to compare two approaches: transfer learning and learning from scratch. The former evaluates the feasibility of adapting an AlexNet architecture trained on ImageNet to the classification of lung nodules, whereas the latter focuses on training a CNN initialized with random weights.

For each of these two approaches we also evaluate some of the effects of changing their configuration. Moreover, for each approach we perform a different pre-processing of images.

Achievement of each network is evaluated through the classification accuracy of the validation data at the last epoch (the result does not have to be the highest acurracy of the entire training process). Validation loss at last epoch is also included to assess the quality of the learning process.

To evaluate the neural network used in each case, we use t-SNE (t-Distributed Stochastic Neighbor Embedding). The t-sne creates a set of low-dimensionality points from high-dimension data in order to visualize the data in clusters. Try to map high-dimensionality objects into low-dimensionality objects, preserving the kinship structure of the neighborhood. The main concept is that nearby points (in multidimensional space) are attracted and distant points are repelled. The points are spread along the two dimensions, being able to observe how the data are divided between the five classes in which we classify the malignancy. Ideally, five clusters should be

SETUP					RESULTS	
Initial Learning Rate	MiniBatchSize	MaxEpochs	Data Augmentation	Optimizer	Validation Accuracy	Validation Loss
0.05	20	40	NO	ADAM	36.27	1.4765
0.005	20	40	NO	ADAM	33.36	1.4738
0.0005	20	40	NO	ADAM	37.12	7.7912
0.0005	20	80	NO	ADAM	34.09	5.9964
0.05	20	40	NO	SGDM	35.52	1.7175
0.005	20	40	NO	SGDM	36.34	1.644
0.0005	20	40	NO	SGDM	33.17	5.4059
0.0005	20	80	NO	SGDM	36.57	5.3719
0.05	20	40	NO	RMSPROP	13.86	1.8354
0.005	20	40	NO	RMSPROP	15.05	1.8525
0.0005	20	40	NO	RMSPROP	37.06	8.6773
0.0005	20	80	NO	RMSPROP	34.64	6.8032

Table 4.1: Transfer Learning results.

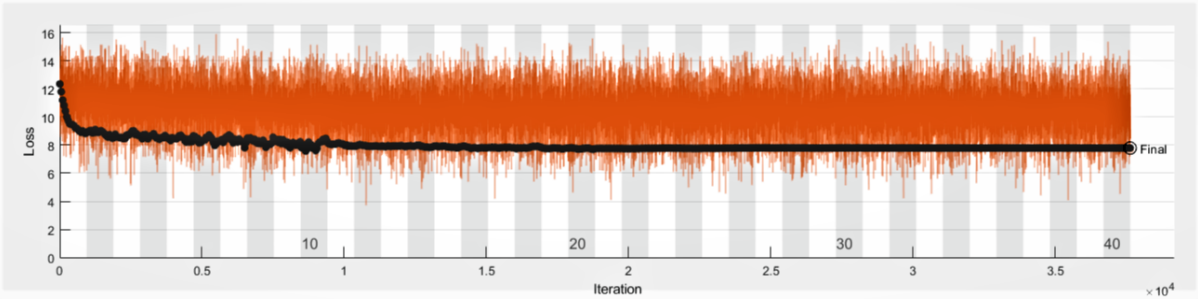

Figure 4.4: An example of validation loss function result (Optimizer: Adam, Initial Learning Rate: 0.0005, MaxEpochs: 40).

created.

To carry out the experiments we use Matlab2018b and the Deep Learning Toolbox Model for AlexNet Network and the Deep Learning Toolbox.

4.2.2 Transfer learning setup

The experiment is inspired by the one performed in [16].

Pre-processing Pre-processing operations are performed on the lung CT slices (images) in order to adapt their resolution to the one expected by AlexNet network. On the images of original nodules trimmed, we have made a padding of zeros to the size of 227 × 227 that is the input size in AlexNet network. In addition, we resize the images and thus they have three dimensions as requested by AlexNet input. This choice was preferred over re-sizing to avoid image artifacts and aliasing and better conserve the shape and texture of nodules. Note that the lungs also contain several structures that can be confused with nodules.

Finally, nodule images are normalized to zero-mean and variance one in the context of the dataset (i.e., instead of using default ImageNet normalization).

Network configuration The last fully-connected layer (fc8) and the Softmax layer (prob) of AlexNet network are adapted to the target dataset. The other twenty-two layers of the network are frozen, i.e., during training, the parameters of the frozen layers are not updated. If the target dataset is small, as is our case, then freezing earlier network layers can also prevent those

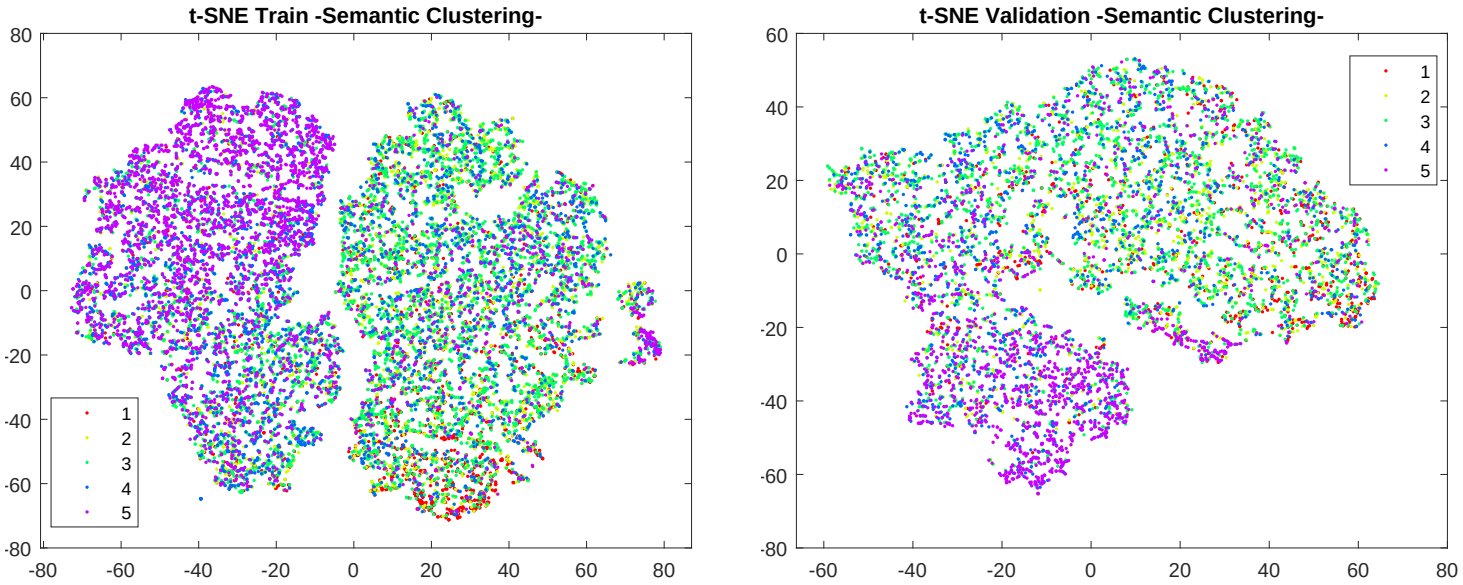

Figure 4.5: t-SNE 2D Train and Validation of transfer learning.

layers from overfitting to the new dataset. Besides, freezing the weights of these initial layers can significantly speed up network training.

Parameter setup We swap on the optimizer (SGDM, RMSProp and ADAM), the maximum number of epochs to use for training (40 and 80) and the learning rate $(0'05, 0'005, 0'0005)$. We fix the mini-batch size to 20 for all the experiments. In addition, we use an option for dropping the learning rate during training, in particular the learning rate is updated by a 0.1 factor every 10 epochs.

Results of this experiment are included in Table 4.1 and Figure 4.5.

Discussion on transfer learning experiment Based on the results, the Adam optimizer is the one that provides the top accuracy. Lower learning rates seem to help learning up to a limit. Differently, the number of epochs and the batch size do not seem to have a clear impact on performance.

We hypothesize that the moderate performance of all the models trained is due to the inability of transfer knowledge from a network trained on a source task driven by natural images (ImageNet) to a target task where images are representations on a different domain (CT scans of lung nodules). Moreover, the amount of data used for training is clearly not enough to unfreeze additional layers. The collection of a large number of medical images and their subsequent annotation remains a difficult task. AlexNet contain tens of millions of free parameters to train, and thus require a sufficiently large dataset. For instance, ImageNet has more than 1.2 million 256×256 images categorized under 1000 categories where there are 1000 images of training samples for each category and 100,000 images of testing samples. Above all, the fundamental problem we encountered during training was the size of the nodules. The nodules are small with respect to the input size of AlexNet; hence, most of the nodule information may be lost after the pooling layers.

We have not used Data Augmentation because looking at the training figures, we did not consider there was overfitting (see an example in Figure 4.4).

SETUP					RESULTS	
Initial Learning Rate	MiniBatchSize	MaxEpochs	Data Augmentation	Optimizer	Validation Accuracy	Validation Loss
0.05	20	40	NO	ADAM	29.83	1.569
0.005	20	40	NO	ADAM	36.11	1.6743
0.005	20	40	SI	ADAM	40.92	1.4092
0.0005	20	40	NO	ADAM	35.15	1.8469
0.0005	20	40	SI	ADAM	40.23	1.4121

Table 4.2: Learning from scratch results.

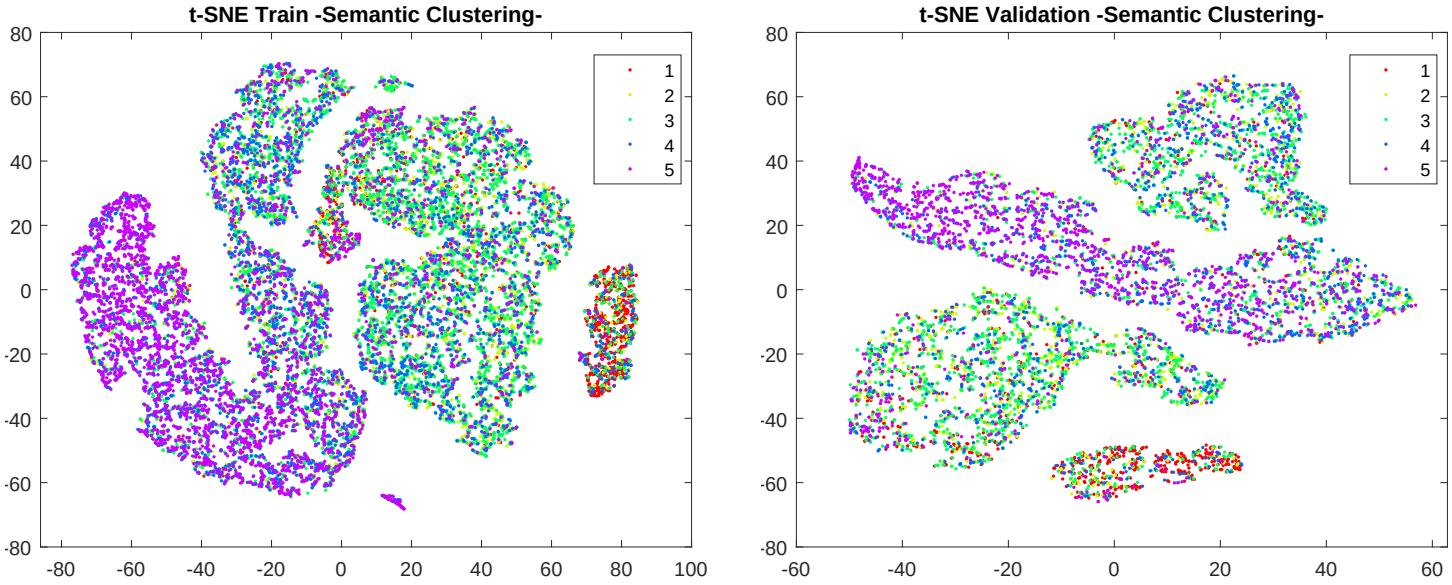

Figure 4.6: t-SNE 2D Train and Validation of learning from scratch.

In Figure 4.5, we can observe how two clusters are formed in the t-sne train, being the classes 3 and 5 (that more cases there are) the ones that are better classified. The other classes find more or less concentrated in a certain point of space.

4.2.3 Learning from scratch setup

Pre-processing On the images of original nodules trimmed, we have made a padding of zeros to the size of 78×78 that is the size of the largest node of the dataset. This choice was preferred over re-sizing to avoid image artifacts and aliasing and better conserve the shape and texture of nodules. Nodule images are normalized to zero-mean and variance one in the context of the dataset (i.e., instead of using default ImageNet normalization). In addition, images have a single dimension and not three as requested by AlexNet at the input.

Network configuration Our network from scratch has two convolutional layers. The first one arranged 16-dimensional filters of size 11×11. The second one poses 32 dimensional filters of 3×3. We relied on AlexNet's filter sizes to define those in this network. The network have one Max-pooling layer, one fully-connected layer, two Relu layers after the Convolutional layer, one Softmax layer and finally a Classification layer. The architecture of the network can be observed in Figure 4.7. Overall, the network architecture comprises eleven layers up to a total of 2224 neurons.

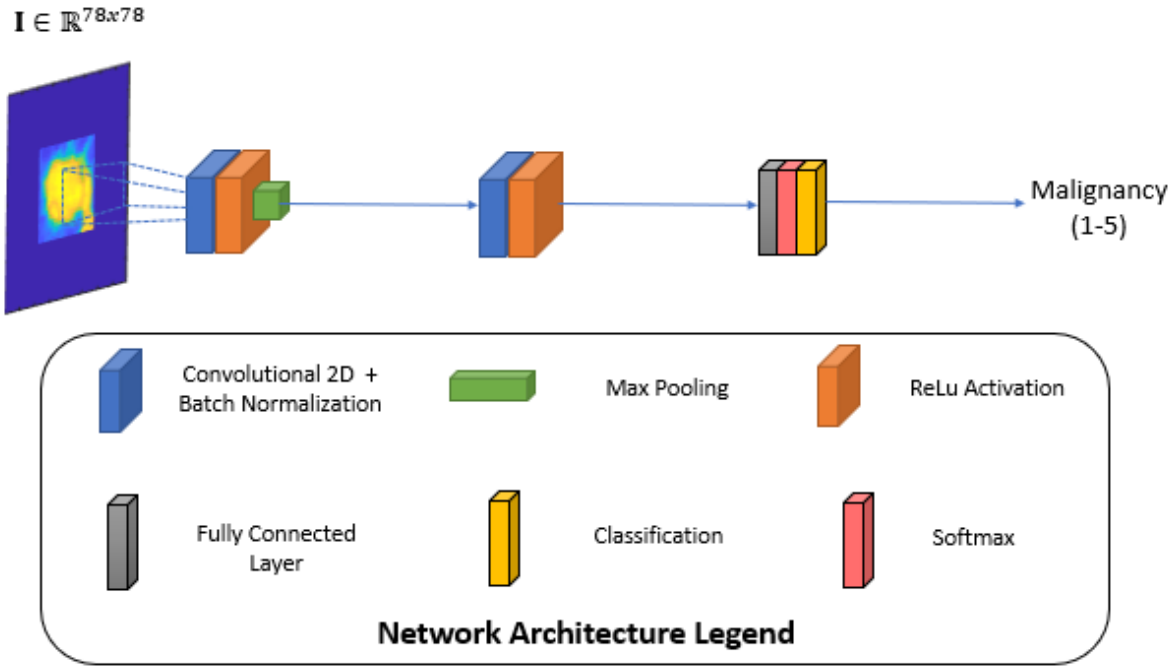

Figure 4.7: Network from scratch architecture.

Parameter setup We swap on the maximum number of epochs to use for training (40 and 80) and the learning rate $(0'05, 0'005, 0'0005)$. We fix the mini-batch size to 20 and use the Adam optimizer for all the experiments. In addition, we use an option for dropping the learning rate during training, in particular the learning rate is updated by a 0.1 factor every 10 epochs. We also evaluate the benefits of using Data Augmentation, to cope with overfitting produced during the training.

Results for this experiment are included in Table 4.2 and Figure 4.6.

Discussion on learning from scratch experiment Results suggest that this architecture gives us better accuracy than transfer learning from AlexNet architecture. We decide to use Adam optimizer because it was the one that gave us top accuracy in the previous experiment.

Comparing the results of this experiment with those of the previous one, the use of a small number of pooling layers may help to better conserve information of small nodules as suggested by the higher validation accuracy obtained for this experiment. It can also be observed that data augmentation give us a clear improvement of validation accuracy.

In Figure 4.6 we can observe four clusters are formed in t-sne validation but there are points of level three that are distributed in two of them, giving errors in the classification. As we have mentioned, this method gives us better results than transfer learning, where we can observe only two clusters. Although the malignancy characteristic is the most balanced compared to the other eight characteristics, we still observe that level 3 and 5 have many more samples than level 1, 2 and 4.

In our opinion, the advantages of implementing algorithms from scratch are also formative:

- Comprehension: We achieve a deep acknowledge for the mathematical description and how the algorithm works. Also we learn how all of the parameters are used and their effects, e.g., over and under training.

- Starting Point: Instead of using an opaque off-the-shelf library (AlexNet) where we do not have knowledge about the complexity of the code inside, learning from scratch give us a absolute comprehension of the network architecture.

The problem of *learning from scratch* is that it is usually time consuming (involves in-depth research into how neural networks and how are their different architectures), [78].

4.2.4 Contextualization of experimental results and general discussion

In Section 2.2.2, we report different classification methods posed by state-of-the-art papers which yield a (substantially) higher performance than the methods we propose.

To sum up, [68] obtained an accuracy of 75.01% with a sensitivity of 83.35%. 3D-CNNs achieved an an average sensitivity of 96.85%, [65] [66] [67]. In paper [65] authors propose a Multi-scale Convolutional Neural Networks (MCNN) method that achieved 86.84% for nodule classification.

In our opinion, one of the main reasons for our low performance is the small number of images used for training. The reported classification methods generally combine information from several public datasets for training. Furthermore, their annotations are better than those used in this paper, as they generally combine information from different radiologist by computing common intersection areas. We did not manage to relate nodules between radiologist, so we were not able to obtain these enhanced annotations. Differently to previous approaches, our performance is for nodule images, not for nodules themselves. Moreover, previous approaches proposed more complex architectures than those posed in this work.

On the correctness of annotations and use of automatic systems for diagnosis

Heisenberg's Uncertainty Principle [79] states that it is impossible to measure simultaneously, and with absolute precision, the value of the position and the amount of motion of a particle. Of course, the principle applies only to physics but its philosophical implications go further. Medicine is a science of probabilities where medical interventions are based on experience, in this way, part of the measurement process is subjective. For this reason, clinical data have a part of uncertainty. An incomplete medical record, misunderstandings or imprecise information provided by the patient in combination with the non-deterministic perception of doctors to each patient individually may provide data that do not allow us to create automatic processing techniques based on deep learning with sufficiently generalist.

On the relevance of the medical record

To suspect a neoplastic disease is determinant the information about the subject: the personal history of tobacco intake, exposure to other carcinogens, other previous illness and the family history of cancer. It is important to achieve a complete medical record that include the symptoms and the physical examination of the patient in order to identify signs of the local or distant extension of the disease. With this information, the physicians start the diagnostic process and the complementary test needed to complete the study. Finally, the histopathology of the tumour establish the confirmatory diagnostic of a neoplasm and, in this way, create a more reliable and comprehensive automatic classification system.

5

Conclusions and future work

In this chapter we provide our conclusions on this work and we discuss possible future work.

5.1 Conclusions

The conclusion we get from the comparison of transfer learning and learning from scratch techniques is the following: the transfer learning problem is its inability to adapt between natural images and CT scans domains by contrast learning from scratch is unable to learn from a small number of samples. Despite this, leaning from scratch gives us better results (higher validation accuracy) than transfer learning from AlexNet architecture. These results are due to the nodules are so small with respect to the input size of AlexNet, most of the nodule information may be also lost after the first pooling layers, and the learning from scratch architecture includes a smaller number of pooling layers, better preservation of lung nodule information.

Many of the studies analyzed during this work have shown the potential of automatic learning methods for the classification of lung nodules from CT images. Because of this, to get more data and help to promote the development of detection and classification of medical images, it is necessary to strengthen the relationships between researchers, physicians, governments and engineers. We believe that the efforts of all involved will encourage the use of these methods in daily medical practice in hospitals.

In addition, it is necessary to mention that the deficit of labeled public datasets, supposes an obstacle at the time of creating neural networks that allow the classification of the nodules. The lack of information on the patient's medical record means that the neuronal network for classification does not generalize adequately for use in medical practice, for which reason we must insist on the creation of public datasets.

As a conclusion of the work dedicated to this Bachelor Thesis, I also attach a Gantt chart that illustrates my project schedule.

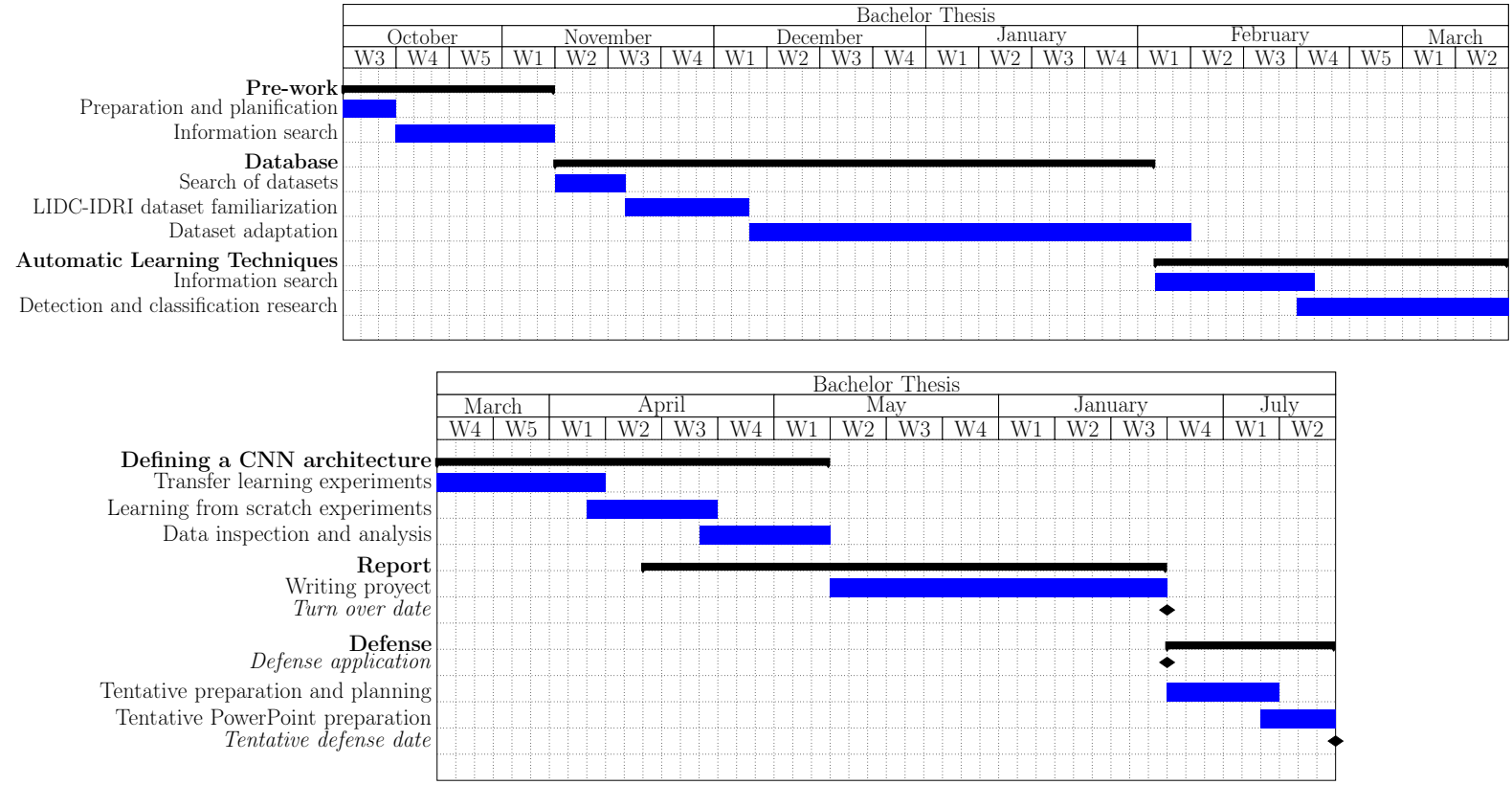

Figure 5.1: Gantt's diagram for the 2019 Bachelor Thesis.

5.2 Future work

In our opinion these results represent the beginning on early detection and classification of lung cancer. Particularly, it would be interesting to analyze the performance of methods exploiting alternative pre-treatment of CT images and thus achieve better network performance.

Results show that learning from scratch is a better method against to transfer learning due to the size of the nodes is so small and they are not robust enough to pass through the filters of the AlexNet network which a mandatory input much larger than the size of these lung nodules. One possible improvement would be obtaining more data for the training process and,in this way, the learning from scratch network can achieve a greater sensibility.

These conclusions may help resolve the limitations of existing methods for classifying lung nodules regardless of size. We have learned the advantages and disadvantages of both methods employed and the benefits of the database we have worked with.

This Bachelor Thesis is therefore the starting point for an upcoming end-of-master Project.

Bibliography

[1] SEOM Sociedad Española de Oncología Médica. Las cifras del cancer en españa 2019. 2019.

[2] Thijs Kooi, Albert Gubern-Merida, Jan-Jurre Mordang, Ritse Mann, Ruud Pijnappel, Klaas Schuur, Ard den Heeten, and Nico Karssemeijer. A comparison between a deep convolutional neural network and radiologists for classifying regions of interest in mammography. pages 51–56, 2016.

[3] Snehashis Roy, John A Butman, Daniel S Reich, Peter A Calabresi, and Dzung L Pham. Multiple sclerosis lesion segmentation from brain mri via fully convolutional neural networks. *arXiv preprint arXiv:1803.09172*, 2018.

[4] Jean-Paul Charbonnier, Eva M Van Rikxoort, Arnaud AA Setio, Cornelia M Schaefer-Prokop, Bram van Ginneken, and Francesco Ciompi. Improving airway segmentation in computed tomography using leak detection with convolutional networks. *Medical image analysis*, 36:52–60, 2017.

[5] Gwenolé Quellec, Katia Charrière, Yassine Boudi, Béatrice Cochener, and Mathieu Lamard. Deep image mining for diabetic retinopathy screening. *Medical image analysis*, 39:178–193, 2017.

[6] Andre Esteva, Brett Kuprel, Roberto A Novoa, Justin Ko, Susan M Swetter, Helen M Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115, 2017.

[7] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*, 2017.

[8] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.

[9] MathWorks. Redes neuronales convolucionales. `https://es.mathworks.com/solutions/deep-learning/convolutional-neural-network.html`, 2017.

[10] Jyo Pari Nameer Hirschkind and Jimin Khim contributed. Convolutional neural network. `https://brilliant.org/wiki/convolutional-neural-network/`, 2017.

[11] Diego Ortego. Self-supervised learning. *Deep learning workshop*, 2019.

[12] Md Rezaul Karim. *Java Deep Learning Projects: Implement 10 real-world deep learning applications using Deeplearning4j and open source APIs*. Packt Publishing Ltd, 2018.

[13] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929–1958, 2014.

[14] Igor Rafael S. Valente, Paulo César Cortez, Edson Cavalcanti Neto, José Marques Soares, Victor Hugo C de Albuquerque, and João Manuel RS Tavares. Automatic 3d pulmonary nodule detection in ct images: a survey. *Computer methods and programs in biomedicine*, 124:91–107, 2016.

[15] McLennan Geoffrey Bidaut Luc McNitt-Gray Michael F. Meyer Charles R. Reeves Anthony P. Clarke Laurence P. Armato III, Samuel G. et al. The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans. *Medical Physics*, 38:915–931, 2015.

[16] Hoo-Chang Shin, Holger R Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Nogues, Jianhua Yao, Daniel Mollura, and Ronald M Summers. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging*, 35(5):1285–1298, 2016.

[17] Hao Gao. A walk-through of alexnet. `https://medium.com/@smallfishbigsea/a-walk-through-of-alexnet-6cbd137a5637`, 2017.

[18] Alfredo Canziani, Adam Paszke, and Eugenio Culurciello. An analysis of deep neural network models for practical applications. *arXiv preprint arXiv:1605.07678*, 2016.

[19] Jaques Ferlay, Eva Steliarova-Foucher, Joannie Lortet-Tieulent, Sonia Rosso, Jan-Willem W Coebergh, Harry Comber, David Forman, and Freddie Bray. Cancer incidence and mortality patterns in europe: estimates for 40 countries in 2012. *European journal of cancer*, 49(6):1374–1403, 2013.

[20] Ahmedin Jemal, Freddie Bray, Melissa M Center, Jacques Ferlay, Elizabeth Ward, and David Forman. Global cancer statistics. *A cancer journal for clinicians*, 61(2):69–90, 2011.

[21] V Cokkinides, J Albano, A Samuels, ME Ward, and JM Thum. American cancer society: Cancer facts and figures. *Atlanta: American Cancer Society*, 2005.

[22] Silvia Francisci, Pamela Minicozzi, Daniela Pierannunzio, Eva Ardanaz, Andrea Eberle, Tom K Grimsrud, Arnold Knijn, Ugo Pastorino, Diego Salmerón, Annalisa Trama, et al. Survival patterns in lung and pleural cancer in europe 1999–2007: results from the eurocare-5 study. *European Journal of Cancer*, 51(15):2242–2253, 2015.

[23] N Howlader, AM Noone, M Krapcho, D Miller, K Bishop, SF Altekruse, CL Kosary, M Yu, J Ruhl, Z Tatalovich, et al. Seer cancer statistics review, 1975-2013, national cancer institute. bethesda, md, 2016.

[24] National Comprehensive Cancer Network et al. Nccn clinical practice guidelines in oncology: Small cell lung cancer (version 2.2015). *Fort Washington, PA, National Comprehensive Cancer Network*, 2015.

[25] M Reck, S Popat, N Reinmuth, Dirk De Ruysscher, KM Kerr, and S Peters. Metastatic non-small-cell lung cancer (nsclc): Esmo clinical practice guidelines for diagnosis, treatment and follow-up. *Annals of oncology*, 25(3):27–39, 2014.

[26] Julian R Molina, Ping Yang, Stephen D Cassivi, Steven E Schild, and Alex A Adjei. Non-small cell lung cancer: epidemiology, risk factors, treatment, and survivorship. *Mayo Clinic Proceedings*, 83(5):584–594, 2008.

[27] Anthony J Alberg and Jonathan M Samet. Epidemiology of lung cancer. *Chest*, 123(1):21S–49S, 2003.

[28] Florian Kocher, Wolfgang Hilbe, Andreas Seeber, Andreas Pircher, Thomas Schmid, Richard Greil, Jutta Auberger, Meinhard Nevinny-Stickel, William Sterlacci, Alexandar Tzankov, et al. Longitudinal analysis of 2293 nsclc patients: a comprehensive study from the tyrol registry. *Lung Cancer*, 87(2):193–200, 2015.

[29] National Lung Screening Trial Research Team. Reduced lung-cancer mortality with low-dose computed tomographic screening. *New England Journal of Medicine*, 365(5):395–409, 2011.

[30] Linda L Humphrey, Mark Deffebach, Miranda Pappas, Christina Baumann, Kathryn Artis, Jennifer Priest Mitchell, Bernadette Zakher, Rongwei Fu, and Christopher G Slatore. Screening for lung cancer with low-dose computed tomography: a systematic review to update the us preventive services task force recommendation. *Annals of internal medicine*, 159(6):411–420, 2013.

[31] Dear Ms Syrek Jensen. National coverage analysis for lung cancer screening with low dose computed tomography. 2014.

[32] Wei Shen, Mu Zhou, Feng Yang, Caiyun Yang, and Jie Tian. Multi-scale convolutional neural networks for lung nodule classification. *International Conference on Information Processing in Medical Imaging*, pages 588–599, 2015.

[33] Pablo Pastor Martín. *Usando Redes Neuronales Convolucionales Para Convertir Características Visuales en Estímulos Sonoros.* San Cristóbal de La Laguna, Tenerife, 2018.

[34] David H Hubel and Torsten N Wiesel. Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology*, 148(3):574–591, 1959.

[35] José Luis Galárraga. *Clasificador de hojas mediante Deep Learning.* Madrid, España, 2017.

[36] Diego de Benito Gorrón. *Detección de voz y música en un corpus a gran escala de eventos de audio.* Universidad Autónoma de Madrid, España, 2018.

[37] Damián Jorge Matich. Redes neuronales: Conceptos básicos y aplicaciones. *Universidad Tecnológica Nacional, México*, 2001.

[38] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133, 1943.

[39] Donald O Hebb. The organization of behavior; a neuropsycholocigal theory. *A Wiley Book in Clinical Psychology.*, pages 62–78, 1949.

[40] Marvin Minsky and Seymour Papert. An introduction to computational geometry. *Cambridge tiass., HIT*, 1969.

[41] James A Anderson, Jack W Silverstein, Stephen A Ritz, and Randall S Jones. Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological review*, 84(5):413, 1977.

[42] John J Hopfield and David W Tank. "neural" computation of decisions in optimization problems. *Biological cybernetics*, 52(3):141–152, 1985.

[43] Kunihiko Fukushima, Sei Miyake, and Takayuki Ito. Neocognitron: A neural network model for a mechanism of visual pattern recognition. *IEEE transactions on systems, man, and cybernetics*, (5):826–834, 1983.

[44] Yoshua Bengio et al. Learning deep architectures for ai. *Foundations and trends® in Machine Learning*, 2(1):1–127, 2009.

[45] Dan CireşAn, Ueli Meier, Jonathan Masci, and Jürgen Schmidhuber. Multi-column deep neural network for traffic sign classification. *Neural networks*, 32:333–338, 2012.

[46] Johan AK Suykens and Joos Vandewalle. Least squares support vector machine classifiers. *Neural processing letters*, 9(3):293–300, 1999.

[47] Wikipedia contributors. Haar wavelet. `https://en.wikipedia.org/w/index.php?title=Haar_wavelet&oldid=877612973`, 2019.

[48] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[49] David E Rumelhart, Geoffrey E Hinton, Ronald J Williams, et al. Learning representations by back-propagating errors. *Cognitive modeling*, 5(3):1, 1988.

[50] Léon Bottou. Large-scale machine learning with stochastic gradient descent. *Proceedings of COMPSTAT*, pages 177–186, 2010.

[51] Sebastian Ruder. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*, 2016.

[52] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[53] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.

[54] Joseph Antony, Kevin McGuinness, Noel E O'Connor, and Kieran Moran. Quantifying radiographic knee osteoarthritis severity using deep convolutional neural networks. *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 1195–1200, 2016.

[55] Edward Kim, Miguel Corte-Real, and Zubair Baloch. A deep semantic mobile application for thyroid cytopathology. *Medical Imaging 2016: PACS and Imaging Informatics: Next Generation and Innovations*, 9789:97890A, 2016.

[56] Afonso Menegola, Michel Fornaciali, Ramon Pires, Sandra Avila, and Eduardo Valle. Towards automated melanoma screening: Exploring transfer learning schemes. *arXiv preprint arXiv:1609.01228*, 2016.

[57] Dickey Singh. Self-supervised learning gets us closer to autonomous learning, 2018.

[58] Olivia Wiles, A Koepke, and Andrew Zisserman. Self-supervised learning of a facial attribute embedding from video. *arXiv preprint arXiv:1808.06882*, 2018.

[59] Luis Perez and Jason Wang. The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*, 2017.

[60] Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Geert Litjens, Paul Gerke, Colin Jacobs, Sarah J Van Riel, Mathilde Marie Winkler Wille, Matiullah Naqibullah, Clara I Sánchez, and Bram van Ginneken. Pulmonary nodule detection in ct images: false positive reduction using multi-view convolutional networks. *IEEE transactions on medical imaging*, 35(5):1160–1169, 2016.

[61] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, pages 1097–1105, 2012.

[62] Tom Brosch, Roger Tam, Alzheimer's Disease Neuroimaging Initiative, et al. Manifold learning of brain mris by deep learning. pages 633–640, 2013.

[63] Fangfang Han, Guopeng Zhang, Huafeng Wang, Bowen Song, Hongbing Lu, Dazhe Zhao, Hong Zhao, and Zhengrong Liang. *A texture feature analysis for diagnosis of pulmonary nodules using LIDC-IDRI database*. IEEE, 2013.

[64] H. Kim R. Tachibana T. Hironaka J. Nappi T. Uemura, H. Lu and H. Yoshida. Classification of polyp candidates on ctc based on 3d-cnn. *International Forum on Medical Imaging in Asia*, pages 103–105, 2017.

[65] F. Yang C. Yang W. Shen, M. Zhou and J. Tian. Multi-scale convolutional neural networks for lung nodule classification. *International Conference on Information Processing in Medical Imaging*, pages 588–599, 2015.

[66] F. Yang D. Yu D. Dong C. Yang Y. Zang W. Shen, M. Zhou and J. Tian. Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification. *Pattern Recognition*, 61:663–673, 2016.

[67] W. Shen X. Li X. Li, Y. Kao and G. Xie. Lung nodule malignancy prediction using multi-task convolutional neural network. *SPIE Medical Imaging. International Society for Optics and Photonics*, 10134:241–247, 2017.

[68] Min Chen, Xiaobo Shi, Yin Zhang, Di Wu, and Mohsen Guizani. Deep features learning for medical image analysis with convolutional autoencoder neural network. *IEEE Transactions on Big Data*, 2017.

[69] Murillo B Rodrigues, Raul Victor M Da Nóbrega, Shara Shami A Alves, Pedro Pedrosa Rebouças Filho, Joao Batista F Duarte, Arun K Sangaiah, and Victor Hugo C De Albuquerque. Health of things algorithms for malignancy level classification of lung nodules. *IEEE Access*, 6:18592–18601, 2018.

[70] Haritha Sathyan and J Vinitha Panicker. *Lung Nodule Classification Using Deep ConvNets on CT Images*. IEEE, 2018.

[71] Robin Smithuis and Otto van Delden. The radiology assistant: Solitary pulmonary nodule: benign versus malignant. `http://www.radiologyassistant.nl/en/p460f9fcd50637/solitary-pulmonary-nodule-benign-versus-malignant.html`, 2019.

[72] CI Henschke, DI McCauley, DF Yankelevitz, DP Naidich, G McGuinness, and DM Libby. The early lung cancer action program (elcap): Baseline results using low-dose screening ct for lung cancer. 209:222–222, 1998.

[73] Ying Ru Zhao, Xueqian Xie, Harry J de Koning, Willem P Mali, Rozemarijn Vliegenthart, and Matthijs Oudkerk. Nelson lung cancer screening study. *Cancer Imaging*, 11(1A):79, 2011.

[74] Consortium for Open Medical Image Computing. Automatic nodule detection 2009. `https://anode09.grand-challenge.org/`, 2009.

[75] Bram Van Ginneken, Samuel G Armato III, Bartjan de Hoop, Saskia van Amelsvoort-van de Vorst, Thomas Duindam, Meindert Niemeijer, Keelin Murphy, Arnold Schilham,

Alessandra Retico, Maria Evelina Fantacci, et al. Comparing and combining algorithms for computer-aided detection of pulmonary nodules in computed tomography scans: the anode09 study. *Medical image analysis*, 14(6):707–722, 2010.

[76] Yann LeCun, Léon Bottou, Yoshua Bengio, Patrick Haffner, et al. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[77] Sunita Nayak. Number of parameters and tensor sizes in a convolutional neural network (CNN). `https://www.learnopencv.com/`, 2018.

[78] Jason Brownlee. Benefits of implementing machine learning algorithms from scratch. `https://machinelearningmastery.com`, 2014.

[79] Klaus-Peter Adlassnig. Fuzzy set theory in medical diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics*, 16(2):260–265, 1986.