UNIVERSITÀ POLITECNICA DELLE MARCHE
FACULTY OF ENGINEERING

MASTER'S DEGREE IN BIOMEDICAL ENGINEERING

A New Deep-Learning Method for 3D Lung Cancer Delineation

Advisor:
Dr. Agnese Sbrollini

Candidate:
Mattia Carletti

Coadvisor:
Prof.ssa Laura Burattini

Accademic Year 2023/2024

To my beloved grandmother,

whose love continues to live within me.

Acknowledgements

This thesis represents years of study, research, and support that so many people gave to me throughout my academic journey. Without their guidance, patience, and encouragement, this work would never have seen light.

First of all, I would like to thank my advisor, Dr. Agnese Sbrollini, for her mentorship and guidance throughout the process. In fact, her great expertise and commitment have driven me to shape and achieve the approach and results of the research.

I am also deeply thankful to PHD Giulia Bruschi, the doctoral student who guided me throughout my work. Her advice and attention to detail have been highly relevant for the completion of this project.

I would also like to extend my heartfelt gratitude to my coadvisor, Prof. Laura Burattini, as well as the other members of the Cardiovascular Bioengineering Lab, who made me feel at home during my internship, with a special thanks to MHD Jafar Mortada for the help extended valuably in the Python coding. The patience and technical support meant a lot.

I would like to thank my fellow students and friends, who made all the lessons, exams, and projects at the university so enjoyable and memorable. To be sure, the level of support and teamwork extended in working on this project made it a worthwhile experience.

Lastly, I would like to thank my family members for the endless love, support, and encouragement provided to me since always. It was because of their belief in me that I got motivated to stick to my passion and have reached this day.

Thanks to one and all who helped in making this possible.

Mattia Carletti

Abstract

Lung cancer is one of the most common malignancies worldwide, responsible for a great number of cancer deaths due to late-stage diagnosis, which is typical of the asymptomatic nature of the disease in its early course. It is diagnosed in conjunction with Computed Tomography (CT) to generate high-resolution images that are able to outline tumors. Manual segmentation and analysis of the lung nodules in CT images, on the other hand, is a very time-consuming process and susceptible to subjective error. Recently, deep learning methods, mostly convolutional neural networks (CNN), have had much success in biomedical image segmentation with architectures such as U-Net and 3D U-Net. This thesis will mainly focus on a new approach using a 2D U-Net architecture for automatic segmenting of lung nodules within CT images. This is followed by post-processing continuity analysis, which constructs 3D models of the cancerous regions in a way that clinically relevant tumor delineation can be obtained without more complex and resource-consuming models that work with heavier 3D data and features. The model is trained and tested on the NSCLC-radiomics reference dataset of lung cancer segmentation, which contains CT images and their corresponding manual cancer delineation. The performance of the model evaluated using standard metrics to make sure it is clinically viable on the test set shows promising results. In fact, it provided a very good performance on the test set with a high segmentation accuracy (0.9608) of the lung nodules and a Dice Coefficient (DC) of 0.9315. On the entire dataset, for clinical evaluation, the model provides DCs of 0.7997, 0.9123, and 0.9590 for small, medium, and large tumors, respectively. Lastly, DC in 3D reconstruction tests increased from 0.8448 to 0.8850 after post-processing application, hence appropriate volume delineation and cleaning. To our knowledge, this thesis presents a pioneering approach for lung cancer delineation, which for the first time uses the faster and less consuming 2D network to reconstruct data on the 3D level, which provide more clinically relevant information than single-slice imaging. Therefore, also according to the promising results, this work could be considered a forerunner in this line of research, marking a significant advancement in automated medical image analysis.

Contents

List of Figures

List of Tables

Introduction

Lung cancer is one of the most prevalent malignancies worldwide and among the leading causes of cancer-related deaths, accounting for approximately 1.8 million deaths annually. As with all the other cancers, it is lumps of tissue created by abnormal or damaged cells that have grown and multiplicated uncontrollably. In particular, lung cancers arise through a multistep process involving the development of multiple genetic and epigenetic alterations, the particular activation of growth promoting proteins, and the inhibition of tumor suppressor genes. The disease is usually classified into two major subtypes, which differ in terms of treatment and prognosis: non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC); among them, NSCLC represents the most common type.

Signs and symptoms vary depending on tumor type and presence of metastases; however, the disease is detectable, and survival rates are high when treated early. The main problem is that over half of lung cancer cases are diagnosed at stage IV, due to the non-specificity of most of the symptoms. There exist different diagnostic techniques for lung cancer, conventional and emerging, but, traditionally, lung cancer diagnosis is done through Computed Tomography (CT). CT generates high-resolution images of the lungs that unrestrictedly capture minute changes in the lung tissue and are therefore able to detect potential tumors. However, manual interpretation of such scans is usually subjective and prone to error, so automated systems are needed for enhancing accuracy and objectivity in tumor detection and delineation.

Recent deep learning (DL) achievements, mainly in the scope of convolutional neural networks (CNN), have altered medical image analysis by powerful methods of feature extraction and pattern recognition, outperforming the state-of-the-art in many visual recognition tasks. Among the many DL architectures, Fully Convolutional Network (FCN), U-Net, and 3D U-Net have shown great success in medical imaging segmentation tasks. These models learn intricate patterns in images through large and annotated datasets (in supervised learning settings) that enable them to differentiate between

cancerous and non-cancerous tissues. These recent improvements don't exclude challenges linked to the necessity of high-quality training data, overfitting, and generalization of the models across populations and modalities. In the last years, a lot of studies tried to improve lung cancer segmentation by adding to the state-of-the-art network new algorithm, obtaining good performances and promising results.

This thesis proposes a deep learning model based on 2D U-Net for the segmentation of lung nodules in CT images, with a subsequent postprocessing technique for the 3D cancer volume reconstruction, introducing a new approach to obtain clinically relevant results even with a less complex and resource consuming model.

Chapter 1

Clinical Background

1.1 Anatomy and Phisiology of Respiratory System

The respiratory system provides us with the fundamental ability to breathe: inhale and exhale air from our lungs. The respiratory system consists of two divisions: upper airways and lower airways, shown in Figure 1.1.

The upper airway system comprises the nose and the sinuses, the pharynx, and the larynx. The lower airway system consists of the trachea, the bronchi-stem, and all the airways that branch extensively within the lungs [1].

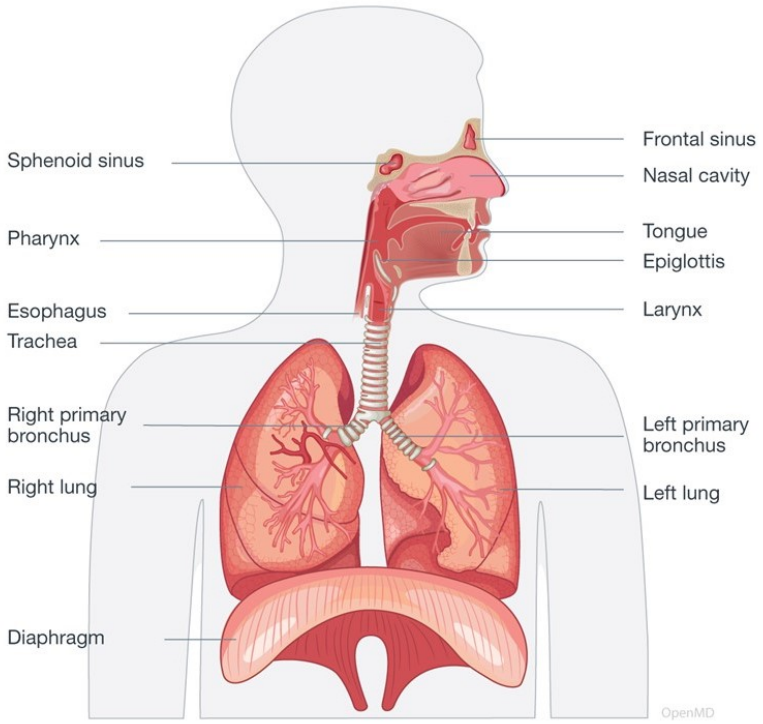

FIGURE 1.1: The Respiratory System Anatomy.

The nose, Figure 1.2a, is the primary entrance to the respiratory tract and consists of an external and internal part. The internal portion has nasal cavities separated by a septum and lined with structures called nasal conchae that increase the efficiency of airflow for cleaning, warming, and humidifying the air [1, 2].

The pharynx, Figure 1.2b, is the muscular tube following the nose subdivided into three areas: the nasopharynx, a passageway that connects to the nasal cavity involved, together with tonsils, in immunity; the oropharynx, where both air and food travel; and the laryngopharynx, which forms the lower part through which food goes after separating from airways [1, 2].

The larynx, Figure 1.2c, is a tube conducting the airways and a sound-producing organ below the pharynx, consisting of various cartilages, of which the most prominent is the thyroid cartilage with the attached epiglottis, which guards against food entering the airways during swallowing [1, 2].

The next one, trachea, Figure 1.2d, extends from larynx to lungs, supported by the ring of the C-shaped cartilage reinforced along its length, and lined with ciliated epithelium for trapping debris [1, 2]. The trachea divides, at the end, into the right and left main bronchi, further dividing into smaller bronchi and eventually into bronchioles, facilitating air distribution to the lungs and leading to alveoli where gas exchange occurs [3].

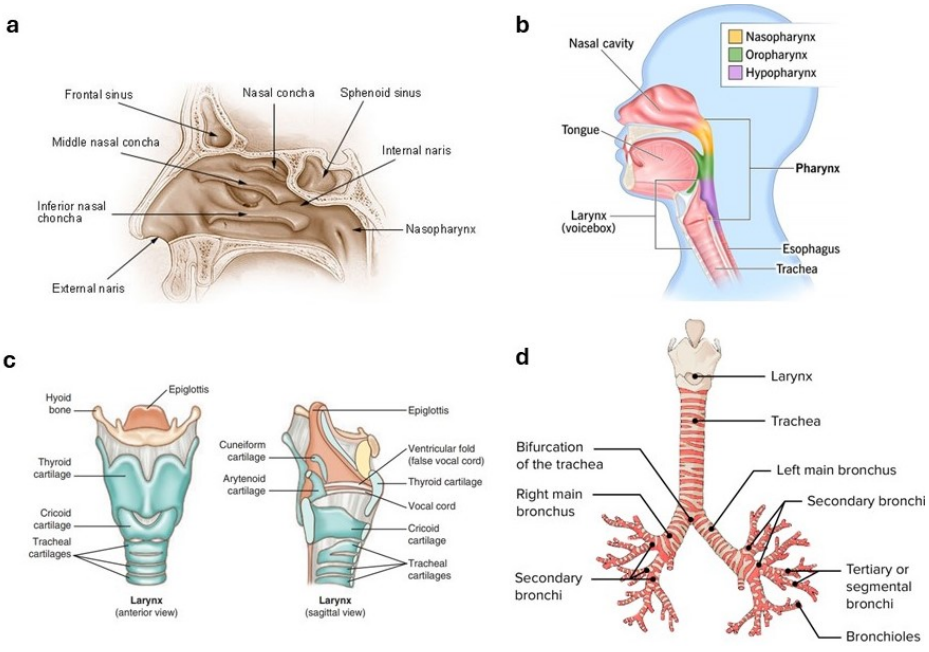

FIGURE 1.2: The Airway of the Respiratory System Anatomy.
a. Nose; **b**. Pharynx; **c**. Larynx; **d**. Trachea and Bronchi

1.1.1 Lungs Structural Organization

The lungs are simple organs within the human respiratory system, which contain air and are principally accountable for exchanging gases, adding oxygen, and removing carbon dioxide from the blood.

Included in the human thoracic cavity, the right and left lungs are enclosed by a slimy membrane known as the pleura [1, 2]. There are two layers within the pleura: the visceral pleura, which covers the surface of the lung directly and extends into the fissures between the lobes, and the parietal pleura, which lines the thoracic wall and mediastinum. This structure allows for smooth movement during breathing, whereas the pleural cavity, which is between the two pleural layers, houses pleural fluid that lubricates the membrane and also maintains a negative pressure that keeps the lungs expanded against the thoracic wall [1, 3].

The right lung is larger and wider anatomically compared to the left lung. This is fully attributed to the presence of the heart that presses the airfield of the left lung; otherwise, this lung is dented by a concavity referred to as the cardiac notch [2, 3]. The lobes of the right lung are three: superior, middle, and inferior. In between, there are two fissures: the horizontal and oblique fissures. The left lung is constructed of two lobes separated by the oblique fissure [1, 3]. Each further divides into bronchopulmonary segments, with the right lung having ten compared to the left, which has between eight and ten, depending on classification. Each of these segments is supplied by an individual bronchial and arterial system, which in turn permits localized surgical interventions during the management of disease [1, 2, 4]. They are shown in Figure 1.3.

The connective tissue structure of the lungs supports their elasticity and, therefore, their integrative function in the respiratory process, which occurs about 16 times per minute. The lungs are mechanically coupled to the thoracic cavity; at the base lies the diaphragm, which is a dome-shaped muscle that mechanically couples with it and has a quite important role in ventilation [3, 4]. When the diaphragm contracts, it enlarges the thoracic cavity, thus creating a negative pressure that pulls air into the lungs, which fill not only their lobes but also, most importantly, the pleural recesses, spaces that enable the lungs to expand during deep breaths [1, 3].

The lungs are also dual in terms of blood circulation: the low-pressure pulmonary circulation and the bronchial circulation for the nutritional needs of lung tissues with oxygenated blood [1, 2]. The pulmonary arteries transport deoxygenated blood from the right ventricle to the lungs, where they

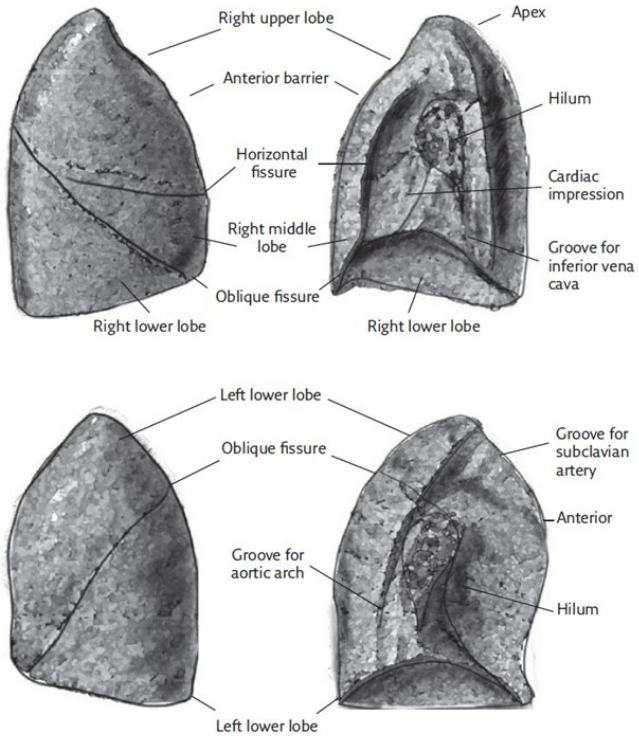

FIGURE 1.3: Lungs Anatomy.

branch extensively to form the pulmonary capillary network around the alveoli, enabling gaseous exchange [2, 4], while, through small pulmonary veins that drain into the left atrium of the heart, oxygenated blood is collected [1, 3].

Lymphatic drainage from the lungs is equally important in fluid balance and in immune function. This includes superficial and deep lymphatic vessels that converge about the bronchial and vascular structures before draining into hilar and mediastinal lymph nodes [1, 2]. Fluid from mesothelial cells is known as pleural fluid, providing lubrication of the pleural surfaces and thereby reducing friction during respiratory movements [2, 3].

Neural control of the lungs is shared by both divisions of the autonomic nervous system, including sympathetic and parasympathetic. Parasympathetic input from the vagus nerve leads to broncoconstriction, whereas sympathetic input favors broncodilation, allowing airflow to be adjusted according to changing physiological requirements. This complex neural input controlling respiratory reflexes, such as coughing, comes from the pulmonary plexus, being built around the stem bronchi [1, 2].

Such information will enable the identification of possible pathological

changes very quickly, as will often require surgical interventions that specifically affect individual lung segments without causing damage to the neighboring tissues. Anatomy is also very important to understand how structural changes or anomalies may result in functional impairment, pointing out the dynamic nature of the lungs in the respiratory system [1, 2, 4].

1.1.2 Respiratory System Physiology

The physiology of the respiratory system incorporates neural control, pressure differentials, and muscular mechanics, interplaying to ensure efficient gas exchange for the sustenance of cellular metabolism.

The neural control centers in breathing are essentially found in the pons and the medulla oblongata of the brainstem. This coordination results in an automated process of breathing, allowing for inspiration and expiration by means of controlled pressure fluctuations within the thoracic cavity [1].

In inspiration, the diaphragm contracts and descends downward toward the abdominal cavity. At the same time, the external intercostal muscles contract to pull the ribs upward and outward. As the volume of the thoracic cavity increases through this combined action, the pressure inside the alveoli falls below the atmospheric pressure. This pressure gradient is instantaneously equalized by air flowing into the lungs. The physics behind it, however, adheres to Boyle's Law, which states that pressure in a gas is inversely related to volume. Thus, with the increase in lung volume during inspiration, there is a transient decrease in pressure within the alveoli such that atmospheric air rushes in. In contrast to inspiration, expiration is normally a passive process because of the elastic recoil of the lung tissue when the diaphragm and intercostal muscles relax. The recoil causes the lung volume to decrease and the intra-alveolar pressure to increase, becoming larger than the atmospheric pressure, which allows the air to be expelled out of the lungs [2]. However, during forced expiration, such as in vigorous exercise or blowing hard, other accessory muscles, such as the abdominal muscles and internal intercostals, actively contract to further compress the thoracic cavity, thereby further reinforcing the process of exhalation. The inspiration and expiration are shown in Figure 1.4.

Ventilation also must increase its activity continuously in response to changes in the metabolic demands. This is achieved by means of highly developed feedback mechanisms involving various types of sensory receptors that are located throughout the body. Chemoreceptors monitor blood concentrations

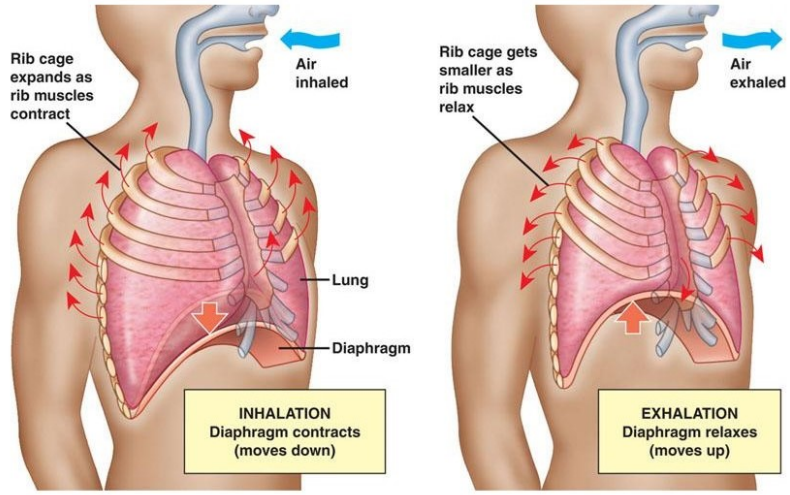

FIGURE 1.4: The Respiratory System Physiology: Pulmonary Ventilation.

of CO_2 and O_2, while mechanoreceptors provide information about lung expansion and the resistance encountered in the airways to breathing. The respiratory centers receive this information and allow the body to make changes to the ventilation pattern that are appropriate for exercising or resting. For instance, when the oxygen content of the blood is low (hypoxia), the respiratory system increases the rate and depth of respiration to enhance the intake of oxygen while removing more carbon dioxide [3].

Besides the basic changes in pressure, the mechanics of ventilation include a huge number of muscle groups working to augment the action of the diaphragm and intercostals. For example, during times of intense exercise, accessory muscles such as the scalene and sternocleidomastoid can also make quite a contribution, giving increased lung capacity for intake of oxygen. Beyond these, vocal and pharyngeal muscles change resistance in the upper respiratory tract during breathing, an example of how structures involved in respiration often tend to serve multiple purposes [1].

With such mechanisms at work, the respiratory system is then very flexible and resistant. Hence, it maintains all the necessary gas exchanges in the face of changing physical demands and environmental changes. This ensures sufficient oxygen for metabolism and, in turn, the efficient removal of carbon dioxide formed during the metabolic process as a waste product[3].

1.1.3 Gas Exchange

It is the basic activity of the respiratory system, taking place primarily at the level of the alveoli in the lungs. For oxygen-rich air to be inhaled and the carbon dioxide produced from cellular metabolism to be removed, this vital exchange has to take place for the proper supply of tissues in the body.

This exchange occurs in the respiratory membrane, where the walls of the alveoli come into contact with the capillary walls of the pulmonary capillaries. Here, the movement of the gases takes place through simple diffusion [1, 2]. Thus, the structure of the lungs is suited to ensure maximum efficiency in exchanging gases; this is made possible by a huge surface area of about 160 square meters in adults. The respiratory membrane is very thin, about 0.5 micrometers, which increases the rate of gas diffusion. Thus, due to the large surface area and high membrane permeability, gas diffusion happens rapidly [1]. As the blood flows through the capillaries, oxygen from the alveoli diffuses into the blood, and carbon dioxide from the blood diffuses into the alveoli, out of the body during exhalation [2].

The partial pressure principle, however, is what drives the movement of gases; independent of other gases' concentrations, a gas will diffuse from an area of higher partial pressure to that with lower partial pressure [1]. Actually, this means that, during external respiration, there is a far greater partial pressure in the oxygen within the alveoli, about 104 mmHg, compared to the 40 mmHg of oxygen present in capillary blood. This large gradient gives a rapid rate of diffusion of the oxygen into the blood. On the other hand, the partial pressure of carbon dioxide is higher in blood, approximately 45 mmHg, than in the alveoli, about 40mmHg. Therefore, carbon dioxide will diffuse down this gradient into the alveolar space from the blood. However, this does not suggest that the partial pressure gradient is the only determinant of gas exchange efficiency; also, the different solubilities in blood of carbon dioxide and oxygen play an important role. Although the solubility of oxygen in plasma is poor, essentially, oxygen complexes with hemoglobin in red blood cells to increase by several times the volume of oxygen to be borne. This hemoglobin-oxygen binding is a critical adaptation that allows blood to carry enough oxygen to meet metabolic demands. In contrast, carbon dioxide is more readily soluble in blood and can be transported in several forms: dissolved in plasma, bound to hemoglobin, or changed into bicarbonate ions in the blood plasma.

Moreover, gas exchange has to take place at two sites: within the lungs, which is the external respiration, and at the tissues, which is the internal

respiration (Figure 1.5). While external respiration happens at the alveolar level with the intake of oxygen and the release of carbon dioxide, internal respiration happens at the cellular level in tissues of the body. In other words, the oxygen is released from the hemoglobin at the tissue level and diffuses into cells whose partial pressure of oxygen is typically low, about 40 mmHg, enough for cellular respiration. This is simultaneously diffused out of the cells into the blood with carbon dioxide, which is produced as a byproduct of metabolic processes, where its partial pressure is lower than that in the tissues, hence closing the cycle of gas exchange [2].

Additionally, pulmonary ventilation has an important role in providing air to the alveoli to maintain appropriate pressure gradients for both oxygen and carbon dioxide exchange. While ventilation takes place in a cyclical manner, blood flow through the pulmonary circulation is continuous and therefore allows nearly all the blood passing through the lungs to participate in efficient gas exchange. In healthy lungs, the relationship between ventilation and perfusion is regulated; proper distribution leads to optimum exchange of gases between them [1].

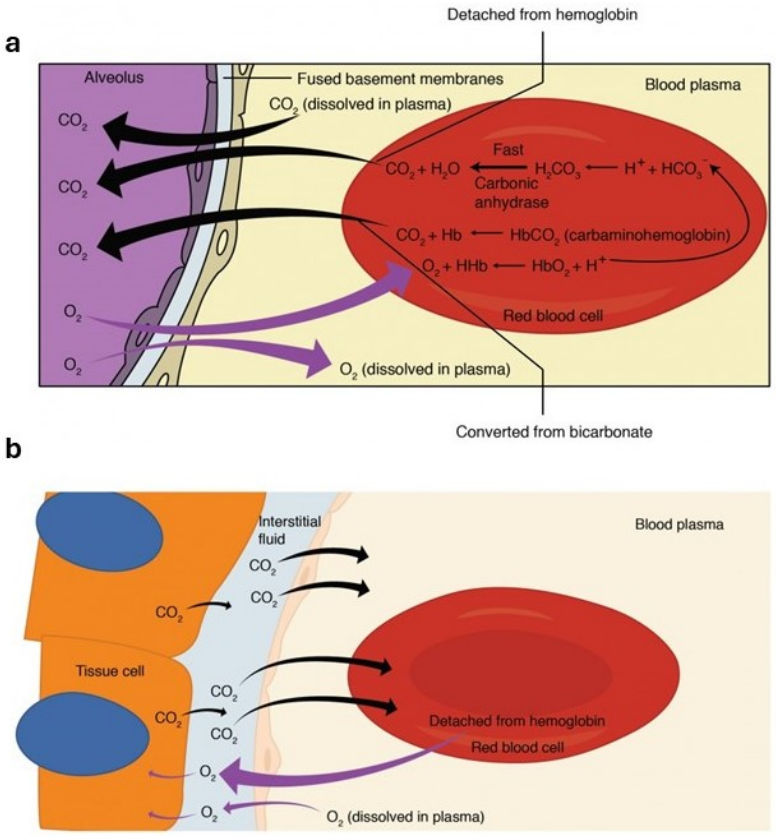

FIGURE 1.5: **a.** External respiration; **b.** Internal Ventilation.

Chapter 2

Lung Cancer

2.1 Epidemiology and Pathogenesy

The human body is made up of trillions of cells. Normally, they grow and multiply to form new cells to take place of those cells that grow old or become damaged. Sometimes, this process breaks down, and abnormal or damaged cells grow and multiply, creating tumors or lumps of tissue. If these tumors can spread or invade nearby tissues and can travel to distant places in the body to form new tumors, they are called cancerous tumors. Thus, cancer is a disease in which some of the body's cells grow uncontrollably and spread to other parts of the body. It can start almost anywhere in the human body [5]. In particular, lung cancers arise through a multistep process involving the development of multiple genetic and epigenetic alterations, particularly activation of growth-promoting proteins and inhibition of tumor suppressor genes. There is great genetic diversity in lung cancer due to the fact that lung cancers have a highly complex genome. Genomic studies have confirmed previously well known alterations in lung cancer (KRAS, EGFR, and BRAF) and have also identified low-frequency but recurrent mutations that are novel in lung cancer, including potentially targetable alterations in JAK2, ERBB4, and RET [6]. Nowadays, lung cancer is the leading cause of global cancer-related mortality [7].

2.1.1 Lung Cancer Statistic

According to the latest GLOBOCAN estimates, Figure 2.1, 2,480,675 new cases of lung cancer were diagnosed globally in 2022, making the lung cancer the most frequently diagnosed cancer in 2022, responsible for one in eight cancers worldwide (12.4% of all cancers globally). Specifically, in men, lung cancer remains the most frequent one with 1,572,045 cases (15.2% of all cancers in males), followed by prostate cancer, while, in women, it is in second

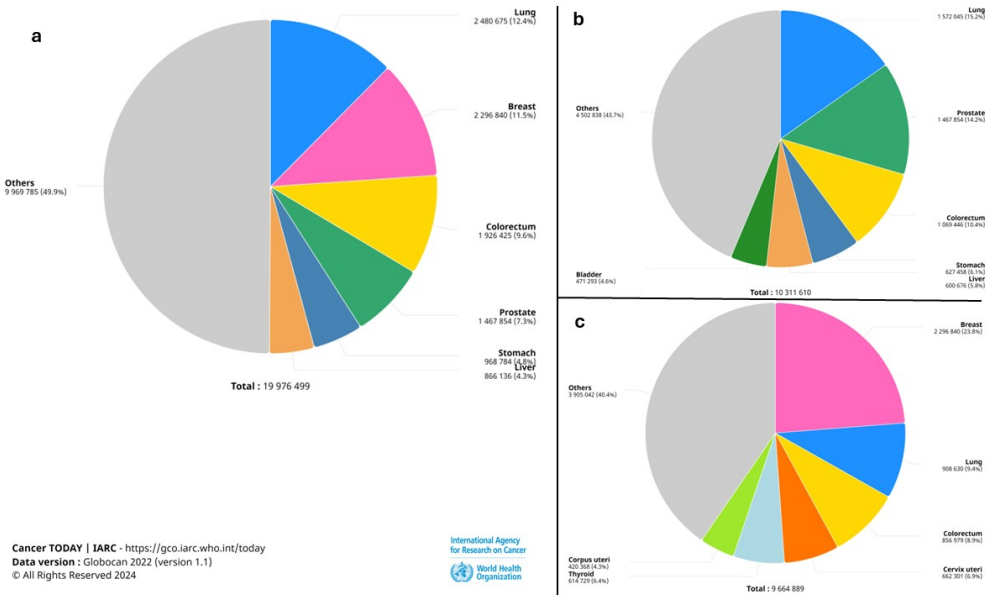

FIGURE 2.1: **a.** New cases in global population; **b.** New cases in male population; **c.** New cases in female population; by GLOBOCAN 2022.

place with 908,630 cases (9.4% of all cancers in females), following breast cancer. [7].

Lung cancer is also the leading cause of cancer mortality worldwide, among both men and women separately, as shown in Figure 2.2. Globally, lung cancer is responsible for 1,817,172 deaths (18.7% of the total cancer deaths), in men of 1,233,241 deaths (22.7% of the total cancer deaths in males), being at the top, while in women of 584,228 deaths (13.5% of the total cancer deaths in females), second only to breast cancer [7].

In 2022, five-year survival from lung cancer tends to be below 20% in most countries, with little differences according to human development, with factors like treatment, health care systems, and the extent of comorbidity playing important roles in survival rate. Such a low survival rate is linked to the fact that most lung cancers are diagnosed at a later stage, when curative treatment is not anymore possible [7].

2.1.2 Risk Factors

The factors that increase the risk of developing lung cancers could be divided into two macro-categories: non-modifiable risk factors and modifiable risk factors. Among the first group we can find:

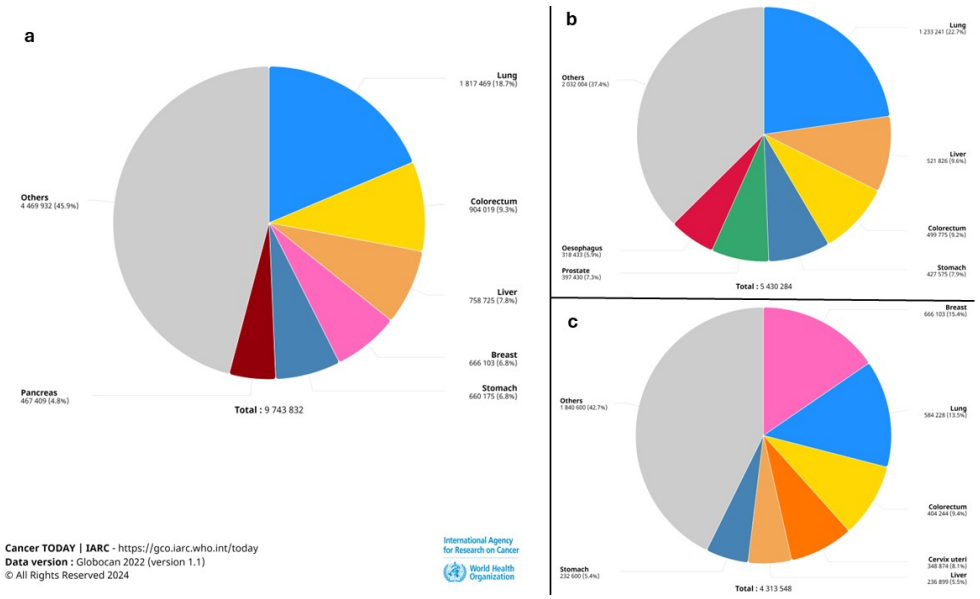

FIGURE 2.2: **a.** Deaths in global population; **b.** Deaths in male population; **c.** Deaths in female population; by GLOBOCAN 2022.

- *Age*. With biological aging, the cells lose their ability to withstand and repair DNA damage as well as surveil for aberrant cells. Young patients are more likely to receive aggressive treatment and report better survival too [8].

- *Gender*. As we can see from the data above, men are over double more than twice as likely to be diagnosed with, and die of, lung cancer. This disparity is mainly due to the higher tendency of men to smoke tobacco, thus is not completely related to gender. But, in non-smoking people, there is a higher rate of lung cancer in women due to hormonal influence [8].

- *Race/ethnicity*. There are some variations of race and ethnicity in the association between the main three susceptibility loci (chromosomes involved in the development of lung cancer) and lung cancer risk [9].

- *Family history*. A positive family history increases the risk of lung cancer by 1.7 times, according to meta-analyses from Central and Eastern Europe. If the history is among first-degree relatives, the risk is increased to 2-4 times, even after careful adjustment for smoking [8].

Among the modifiable risk factors, there are:

- *Tobacco (and Cannabis) smoking*. Tobacco smoking is the major cause of all major histological types of lung cancer. A carcinogenic effect of tobacco smoke on the lung was demonstrated in epidemiological studies conducted since the early 1950s [9]. More than 80% of lung cancer cases in the Western world are attributable to cigarette smoking. The combustion of tobacco produces over 60 known carcinogens. Second-hand smoke exposure likewise has shown a dose-dependent relationship with lung cancer risk. Certain carcinogens in second-hand smoke are inhaled in higher concentrations than by the smoker due to the filters on the user end of cigarettes. For what concern cannabis smoking, the combustion of marijuana is known to produce carcinogenic substances, with levels of some higher than those in tobacco [8].

- *Diet*. There is evidence from case–control studies that a diet rich in vegetables and fruits, especially cruciferous vegetables, may exert some protective effect against lung cancer. High intake of fried or well-done red meat may increase the risk of lung cancer, and this may be related to the formation of nitrosamines during cooking. There is evidence from observational studies that low levels of vitamin D are associated with lung cancer risk [9].

- *Chronic inflammations and infections*. COPD is the most common independent factor, other than smoking, that increases the risk of lung cancer [8]. Patients with pulmonary tuberculosis have been found to be at increased risk of lung cancer [9]. HIV also increases the risk of lung cancer by up to 2.5 times, considering that lung cancer has also become the leading cause of death among HIV patients [8].

- *Ionizing radiation*. Exposure to ionizing radiation increases the risk of lung cancer, reported in atomic bomb survivors, in patients treated with radiotherapy, and in underground miners exposed to radioactive radon and its decay products [9]. Residential radon exposure (basements in geographic regions with high Uranium concentrations) is the second greatest risk factor for lung cancer in the Western world, accounting for an estimated 10% of cases [8].

- *Occupational exposures*. The risk of lung cancer is increased among workers employed in industries and occupations. The most important occupational lung carcinogens are reported to be asbestos (a naturally occurring mineral used in construction [8]), silica, heavy metals (like arsenic,

chromium, high nickel, and cadmium), and polycyclic aromatic hydrocarbons (chemicals formed during combustion of organic material) [9].

- *Air pollution*. There are risk factors for lung cancer related to air quality: carcinogens from the combustion of fossil fuels and particular matter suspended in the air. Indoor air pollution from combustion products, wood and charcoal, commonly used for cooking and heating. Studies have found that ventilation of such cooking areas can reduce lung cancer risk by up to 50% [8].

2.2 Histopathology

The histopathological classification is helpful in defining the prognosis, facilitating the treatment, and predicting of successful results. The main categories are two: small cell lung carninoma (SCLC) and non-small cell carcinoma (NSCC) [10].

The NSCC tumors account for 80% of the cases, and they are managed by a combination of surgery and adjuvant therapy. A subclassification of the NSCC tumors defines three main types: Adenocarcinoma, Squamos cell carcinoma, and Large cell carcinoma.

The SCLC tumors account for the remaining 20% of the cases and are treated non-surgically in most cases [11].

2.2.1 Adenocarcinoma

Adenocarcinoma is the most common type of lung cancer, accounting for more than 40% of lung cancers and 60% of the NSCC [11]. It is considered the commonest subtype in young women and nonsmokers [10].

It is defined as a malignant epithelial tumor with glandular differentiation or mucin production, showing acinar, papillary, bronchioloalveolar, or solid mucin growth patterns or a mixture of these patterns. Adenocarcinomas are most frequently peripheral nodules under 4.0 cm in size, with the most frequent pleura and chest involvement. The recognized patterns of adenocarcinoma are solid nodules (solid density), ground glass opacities (non-solid, air-containing), and mixed solid/ground glass (part solid, subsolid) opacities. The borders of cell clusters are typically sharply delineated, and the cytoplasm is usually abundant, varying in volume. It is typically cyanophilic

and translucent, and it may be single or arranged in three-dimensional morulae, acini, pseudopapillae, and true papillae with fibrovascular cores and/or sheets of cells [12].

This tumor frequently causes metastasis to the liver, bone, and nervous system within the same lung or the contralateral lung and adrenal glands. The prognosis of this type of tumor is significantly better than other lung cancers [10].

The major individual histologic patterns/subtypes are lepidic acinar, papillary, micropapillary, and solid adenocarcinoma, shown in Figure 2.3. The lepidic growth pattern denotes tumor cells spreading along preexisting alveolar structures. Acinar adenocarcinoma is a common type of adenocarcinoma with tumor cells arranged in a classic glandular structure on a fibroelastic stroma. The papillary pattern is formed by tumor cells lining the surface of branching fibrovascular cores; the presence of fibrovascular cores separates this tumor type from micropapillary adenocarcinoma. Solid adenocarcinomas form any other recognizable patterns with a poorly carcinoma expression [11].

2.2.2 Squamos Cell Carcinoma

Squamous cell carcinoma represents 30% of all lung cancers. It arises from altered bronchial epithelium and growths in situ [10]. Over 90% of squamous

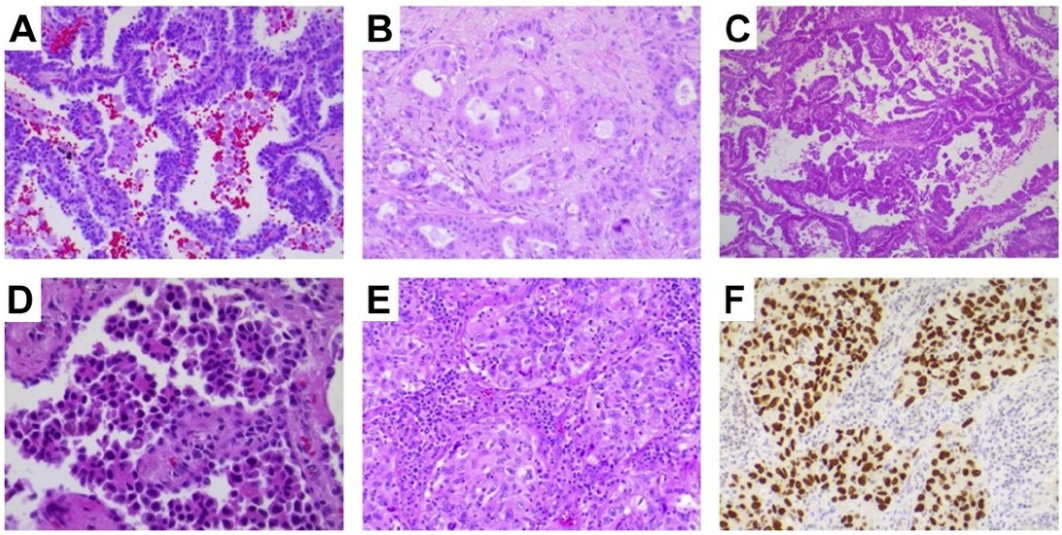

FIGURE 2.3: **a.** Lepidic Adenocarcinama; **b.** Acinar Adenocarcinoma; **c.** Papillary Adenocarcinoma; **d.** Micropapillary Adenocarcinoma; **e.** and **f.** Solid Adenocarcinoma. [11]

cell lung carcinomas occur in cigarette smokers.

It is defined as a malignant epithelial tumor showing keratinization and/or intercellular bridges that arises from the bronchial epithelium. The tumors are usually white or gray and may grow to a large size and may cavitate. Central, segmental, or subsegmental tumors can extend into regional lymph nodes and appear as hilar, perihilar, or mediastinal masses with or without lobar collapse. Peripheral tumors present as solitary pulmonary nodules, smaller than 3 cm, or masses, bigger than 3 cm [12]. The tumor cells usually have hyperchromatic nuclei, visible to inconspicuous nucleoli, and moderate to abundant cytoplasm with delineated intercellular bridges [11]. They frequently cause segmental or lobar lung collapse due to their central location [10]. Figure 2.4 shows a Squamos cell carcinoma with and without keratinizazion.

2.2.3 Large Cell Carcinoma

Large cell carcinoma is strongly associated with smoking [10] and represents a minority of NSCC cases (approximately 9%) [11].

It is an undifferentiated non-small cell carcinoma that lacks the cytologic and architectural features of small cell carcinoma and glandular or squamous differentiation. Large cell carcinomas typically present as large, peripheral masses and often invade the visceral pleura, chest wall, or adjacent structures. It is frequently formed by cellular aggregation in prominent nuclei with round or extremely irregular shapes [12]. It is usually peripherally located, bulky, and necrotic in appearance [11]. An example is shown in Figure 2.5.

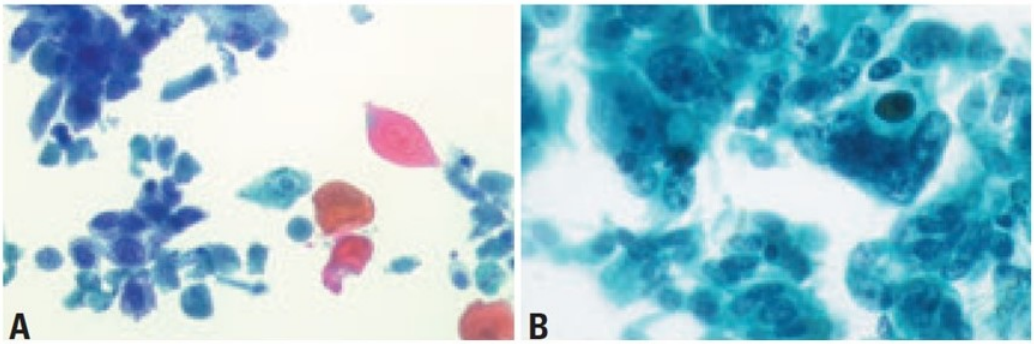

FIGURE 2.4: **a.** Squamos Cell Carcinoma; **b.** Squamos Cell Carcinoma with keratinization. [12]

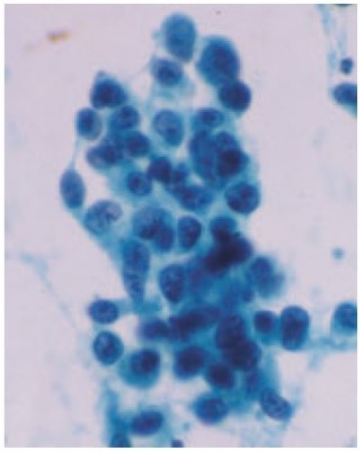

FIGURE 2.5: Large Cell Carcinoma. [12]

2.2.4 Small Cell Lung Carcinoma

SCLCs comprise slightly more than 10% of all lung cancers [11]. It is among the most aggressive primary lung cancers. More than 70% of patients show evidence of extrathoracic metastatic disease at the time of diagnosis [10].

It is a malignant epithelial tumor consisting of small cells with scant cytoplasm, not defined cell borders, finely granular nuclear chromatin, and absent or inconspicuous nucleoli. They are typically white-tan, soft, friable perihilar masses that show extensive necrosis and frequent nodal involvement. Tumor cells are usually less than the size of three small resting lymphocytes and have round, ovoid, or spindled nuclei and scant cytoplasm. The SCLC is combined with an additional NSLC component, usually adenocarcinoma, squamous cell carcinoma, or large cell carcinoma. Both tumors are shown in Figure 2.6.

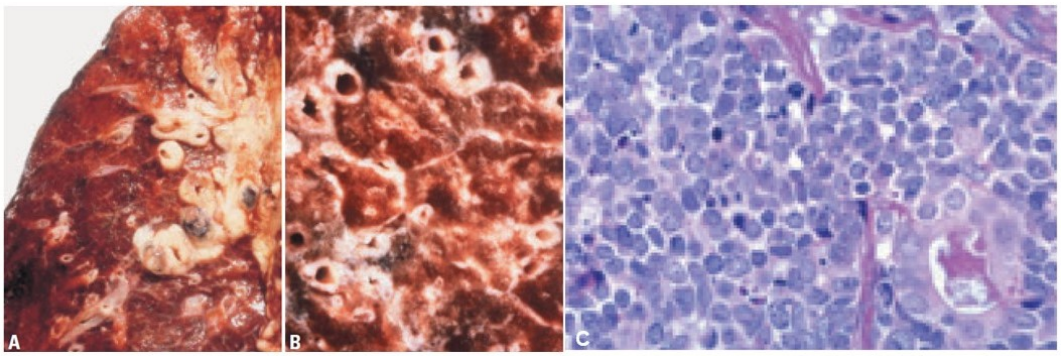

FIGURE 2.6: **a.** Central SCLC; **b.** Peripheral SCLC; **c.** Combined SCLC and Adenocarcinoma. [12]

2.3 Prognosis and Staging

The prognosis of lung cancer is critically dependent on appropriate staging [13]. The fundamental purpose of stage classification is to provide a nomenclature about the anatomic extent of disease that is used consistently around the world to enable reliable communication about a particular patient.

Staging of NSCLC utilizes three anatomical extents of tumors: T for extent of the primary tumor, N for involvement of lymph nodes, and M for distant metastases, such as brain, bones, adrenal glands, liver, the pleural fluid, or the other lung. Each category is then divided into subgroups by other characteristics, named descriptors [14]. Specific combinations of T, N, and M categories are grouped together into stage groups, summarized in Figure 2.7 and in Table 2.1 [15]. In Figure 2.8 is shown a visual explanation of the different stages. SCLC is biologically distinct from NSCLC; thus, its staging schema is simpler, divided into limited and extensive stages. Limited stage is defined as disease confined to one hemithorax, including ipsilateral

T (Primary Tumor)		Label
T0	No primary tumor	
Tis	Carcinoma in situ (Squamous or Adenocarcinoma)	Tis
T1	Tumor ≤3 cm,	
T1a(mi)	Minimally Invasive Adenocarcinoma	T1a*(mi)*
T1a	Superficial spreading tumor in central airways[a]	T1a *SS*
T1a	Tumor ≤1 cm	T1a *≤1*
T1b	Tumor >1 but ≤2 cm	T1b *>1-2*
T1c	Tumor >2 but ≤3 cm	T1c *>2-3*
T2	Tumor >3 but ≤5 cm or tumor involving:	
	visceral pleura[b],	T2 *Visc Pl*
	main bronchus (not carina), atelectasis to hilum[b]	T2 *Centr*
T2a	Tumor >3 but ≤4 cm	T2a *>3-4*
T2b	Tumor >4 but ≤5 cm	T2b *>4-5*
T3	Tumor >5 but ≤7 cm	T3 *>5--7*
	or invading chest wall, pericardium, phrenic nerve	T3 *Inv*
	or separate tumor nodule(s) in the same lobe	T3 *Satell*
T4	Tumor >7 cm	T4 *>7*
	or tumor invading: mediastinum, diaphragm, heart, great vessels, recurrent laryngeal nerve, carina, trachea, esophagus, spine;	T4 *Inv*
	or tumor nodule(s) in a different ipsilateral lobe	T4 *Ipsi Nod*
N (Regional Lymph Nodes)		
N0	No regional node metastasis	
N1	Metastasis in ipsilateral pulmonary or hilar nodes	
N2	Metastasis in ipsilateral mediastinal/subcarinal nodes	
N3	Metastasis in contralateral mediastinal/hilar, or supraclavicular nodes	
M (Distant Metastasis)		
M0	No distant metastasis	
M1a	Malignant pleural/pericardial effusion[c] or pleural /pericardial nodules	M1a *Pl Dissem*
	or separate tumor nodule(s) in a contralateral lobe;	M1a *Contr Nod*
M1b	Single extrathoracic metastasis	M1b *Single*
M1c	Multiple extrathoracic metastases (1 or >1 organ)	M1c *Multi*

FIGURE 2.7: TNM Tumor Classification. [14, 15]

TABLE 2.1: Correlation between TNM Classification and Staging.

Stage	**TNM Grouping**		
	T	**N**	**M**
Occult (hidden) cancer	TX	N0	M0
0	T0, Tis	N0	M0
IA1	T1a(mi)	N0	M0
	T1a	N0	M0
IA2	T1b	N0	M0
IA3	T1c	N0	M0
IB	T2a	N0	M0
IIA	T2B	N0	M0
IIB	T1a, T1b, T1c	N1	M0
	T2a, T2b	N1	M0
	T3	N0	M0
IIIA	T1a, T1b, T1c	N2	M0
	T2a, T2b	N2	M0
	T3	N1	M0
	T4	N0, N1	M0
IIIB	T1a, T1b, T1c	N3	M0
	T2a, T2b	N3	M0
	T3	N2	M0
	T4	N2	M0
IIIC	T3	N3	M0
	T4	N3	M0
IVA	AnyT	AnyN	M1a
	AnyT	AnyN	M1b
IVB	AnyT	AnyN	M1c

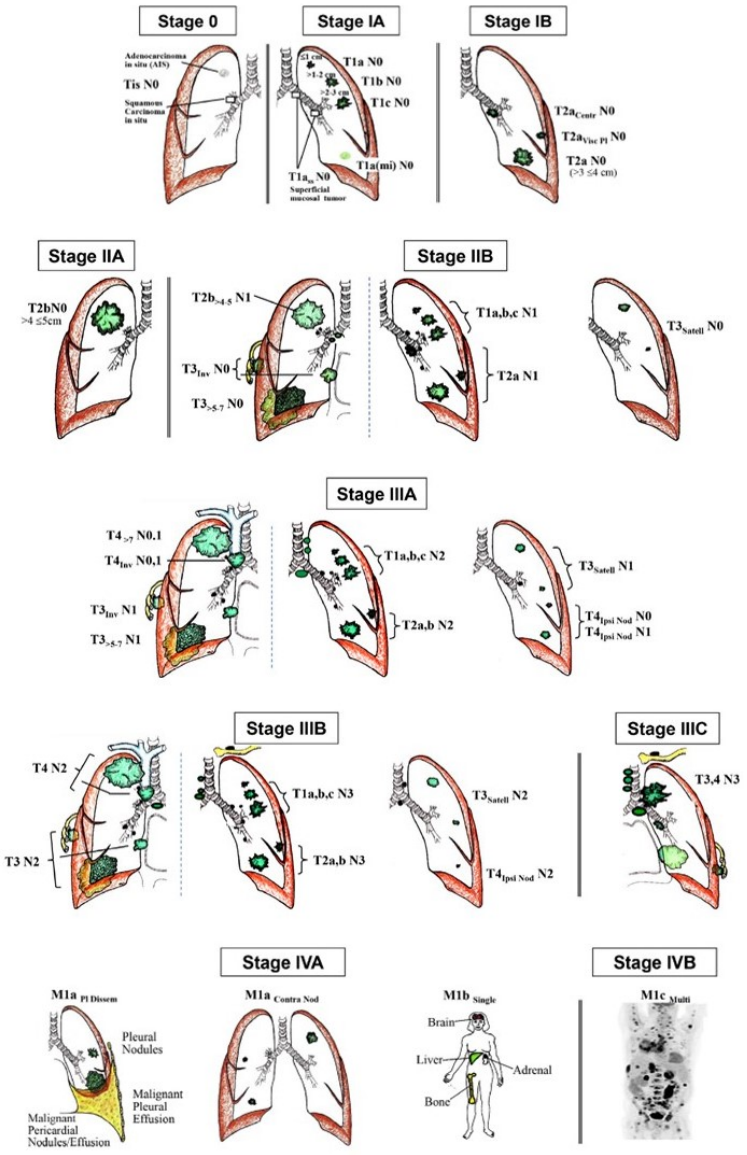

FIGURE 2.8: Stages of Lung Tumors.

mediastinal and/or supraclavicular disease, excluding malignant pleural effusion. All other tumors are characterized as extensive [13].

2.4 Diagnosis

Over half of lung cancer cases are diagnosed in stage IV. This high percentage of late diagnoses is largely due to the non-specificity of most of the symptoms (no clearly predominant lung cancer symptom), to the fact that they are not as frequent as would be expected, or to patients's delay in consulting their physician [16]. Patients may present with the nonspecific systemic

symptoms of fatigue, anorexia, and weight loss, or with direct signs and symptoms caused by the primary tumor (chest discomfort, cough, dyspnea, and hemoptysis), intrathoracic spread (laryngeal nerve paralysis, Pancoast's tumor, pleura effusions and pain, and superior vena cava obstruction), or extrathoracic spread (metastatic sites include bones, liver, adrenal glands, lymph nodes, brain, and spinal cord) [17].

According to a new nationwide study [16], the most frequent symptom in NSCLC in the early stages is cough, followed by chest pain, while in stage IV they are reversed. Generally, the presence of symptoms increases across stages; however, some patients could have only one or two, or even none, symptoms in stage IV too. There is no difference between smokers and never smokers in the presence of symptoms. This study highlighted that approximately 30% of all patients diagnosed in stages III and IV had no lung cancer symptoms at diagnosis.

For what concern SCLC, the patients with extended SCLC showed cough, dyspnea and pain as the most occuring symptoms, while patients with limited SCLC are mostly asymptomatic.

Thus, the clinicians should not rule out the presence of lung cancer in cases where no symptoms are present, due to the lack of specificity of lung cancer symptoms. There exist different diagnostic techniques for lung cancer, both conventional and emerging:

- *Biopsy*. It is a highly invasive technique based on the removal of a part of a tissue or a whole nodule or lump, as a sample, with the use of a needle and a small incision. The most common one in lung cancer detection is the needle biopsy, used if the nodule is bigger than 2 cm. It is used to determine the tumor present in the lung pleura, mediastinum, or in the lung parenchyma [18].

- *Bronchoscopy*. A flexible tube, with a camera to the end, let the visualization of the air passages. The tumors are usually less refractive hence are black in color and shown as a disruption in the white part of the tissue (higher refractive) [18].

- *Sputum Citology*. Sputum cytology involves the examination of sputum, mucus coughed up from the lungs, which is induced by artificial techniques or naturally produced. It is a non-invasive technique used for early detection of lung cancer [18].

- *Biomarkers and Biosensors*. Biomarkers are naturally occurring molecules, genes, or other biological entities that undergo certain changes during

a disease condition. The most important biomarker for lung cancer detection is neuron-specific enolase (NSE), considering that the levels of NSE increase in patients with lung cancer. A biosensor is an analytical tool that converts chemical signals into electronic response and is used to detect and quantify the biomarkers [18].

- *Medical Imaging*. It is a technique within which it is possible to obtain a 2D, or even 3D, representation of an internal anatomical structure or its functional processes. There exist a lot of imaging techniques with different quality and outcomes in detecting and characterizing lung cancer. These techniques are chest radiography (CXR), computer tomography (CT), magnetic resonance (MR), position emission tomography (PET), and endobronchial ultrasound (EBUS). They are better discussed in the following subsections.

  CXR represents the first line of investigation in cases suspected to have cancer, considering that it is a simple, relatively inexpensive technique and very accessible [19, 10]. It is excellent for the delineation of lung lesions into either benign or malignant; it furnishes staging descriptive features, such as the size of the lesion and resultant complications, but it cannot adequately detect chest wall or mediastinal invasion and nodal involvement [20]. Recent innovations, like digital radiography, have enhanced image quality and the accuracy of nodule detection while allowing dose reduction and cost-benefit efficiency [19]. An example of CXR used to detect lung cancer is shown in Figure 2.9a.

  CT will be explored in the Section 2.4.1, as it is the imaging technique on which this thesis is focused.

  MRI is uniquely useful for diseases of the central nervous and musculoskeletal systems but faces serious difficulties in the lungs because of motion artifacts (due to breathing), low proton density in lung parenchyma, and attenuation by air-soft tissue interfaces [19, 21]. It does complement CT in determining tumor invasion of the chest wall, evaluating diaphragmatic abnormalities, and in mediastinal lymphoma; however, practical use awaits further acquisition speed and spatial resolution improvement [19]. An example of MRI used to detect lung cancer is shown in Figure 2.9b.

  PET represents a very sensitive and specific modality for in vivo imaging of metabolic pathways. The main tracer used is FDG, which underlines the enhanced glucose metabolism of the neoplastic cells [19,

20]. When this technique is combined with CT, diagnostic accuracy significantly improves, determining important data about TNM staging, mediastinal infiltration, and nodal metastasis. It also allows for the detection of small neoplastic lymph nodes, ensuring an accurate delineation of the tumor from surrounding structures [22]. An example of PET used to detect lung cancer is shown in Figure 2.9c.

Finally, EBUS revolutionized bronchoscopy by facilitating the assessment of airway walls, enabling guided biopsies of lymph nodes and tumors, and allowing for real-time transbronchial needle aspiration [23]. The technique has proven superior compared to conventional methods in lung cancer staging with improved cost savings and enhanced patient outcomes [24]. An example of EBUS used to detect lung cancer is shown in Figure 2.9d.

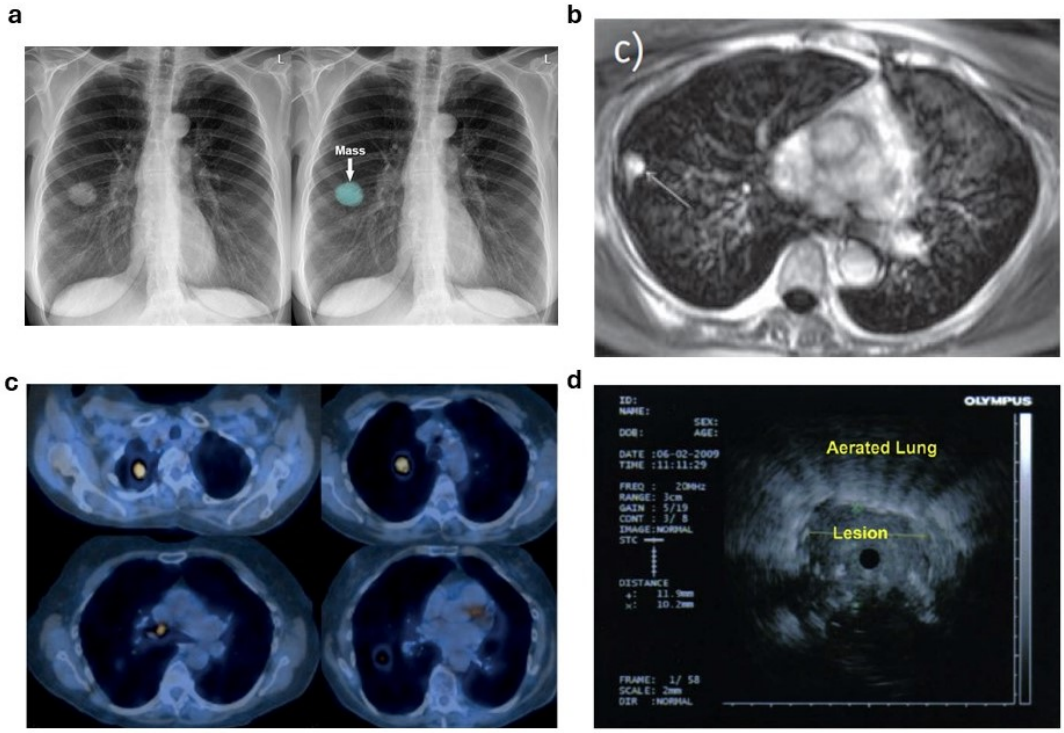

FIGURE 2.9: Lung Cancer in Imaging techniques. **a**. CXR; **b**. MRI [19]; **c**. PET [22]; **d**. EBUS [23]

2.4.1 Computed Tomography (CT)

CT is the imaging technique with the highest sensitivity for the detection of pulmonary nodules [19]. When the CXR raises the suspicion for malignancy, CT with contrast should be performed for complete staging. CT assists in finding abnormalities, highlights signs of disease, monitors the response to treatment, and supports the planning therapy [10]. Increased use of CT has led to improved identification of small peripheral nodules, many of which prove to be adenocarcinomas; adenocarcinoma is often distinct from the other histologic subtypes of lung cancer [12].

A fan-shaped X-ray beam irradiates a narrow section of the body, and the transmitted radiation is giving the X-ray attenuation in one dimension along a projection through the body. Many sets of attenuation data are acquired at different angles by rotating the X-ray tube, and a two-dimensional image of a slice through the body can then be reconstructed [20]. This is the conventional method, while recently new CT methods were introduced. With helical CT, a single volumetric dataset while the patient is moved through the CT gantry is produced, increasing drastically the thoracic image quality. More recently multi-detector CT (MDCT) uses multiple rows of detectors, increasing temporal resolution, fastening the scanning, and increasing spatial resolution. The main reconstruction methods used in MDCT are multiplanar reconstruction (MPR), which improved the image quality and provided a supplementary staging tool; maximum intensity projection (MIP), which reduced the number of overlooked small cancers and increased the sensitivity; and 3-dimensional reconstruction (3DR), which provided more accurate information on the shape, length, and severity. MPR converts the thin axial plan slices acquired in non-axial plane images, such as oblique, sagittal, and coronal; MIP projects the pixels with the highest attenuation values in a 2D format; and 3DR is like CT bronchography with external rendering and bronchoscopy with internal rendering. Finally, MDCT may help reduce the radiation dose: low-dose CT (LDCT) can identify small lung cancers in an at-risk population [19].

Generally, CT scan is the most used imaging modality for T staging, characterized by sensitivity between 38% and 87%, and a specificity between 40% and 90% for chest wall invasion and accuracy of 56-89% for predicting mediastinal invasion. For N staging, which consists of the evaluation of mediastinal disease, CT is not considered the optimal modality; however, it has a sensitivity of 60-83%, specificity of 77-82%, and accuracy of 75-80%. In the

end, in M staging, CT can detect intrathoracic metastases, even with a degree of uncertainty [25]. An example of CT showing lung cancer is shown in Figure 2.10.

2.5 Treatments

Treatment differs according to the histologic type of cancer, the stage at presentation, and the patient's functional evaluation, as shown in Table 2.2.

For NSCC, surgery or resection is the treatment for patients with stage I until resectable stage IIIA, with adjuvant chemotherapy for those undergoing complete resection [17]. Some patients presenting with stage III disease, particularly those with stage IIIA, may be offered preoperative chemotherapy or chemoradiotherapy to downstage the disease to render the disease surgically resectable. In addition, comorbid conditions, especially those caused by tobacco smoking, namely coronary artery disease and COPD, may make surgical resection technically difficult or impossible [13]. For unresectable and higher NSCC may involve radiotherapy and chemotherapy [17]. In particular, radiation therapy is an excellent palliative option for control of pain, hemoptysis, and bronchial obstruction with post-obstructive pneumonia [13]. Chemotherapy (combined with radiotherapy in limited stage disease) is the main treatment for SCLC [17].

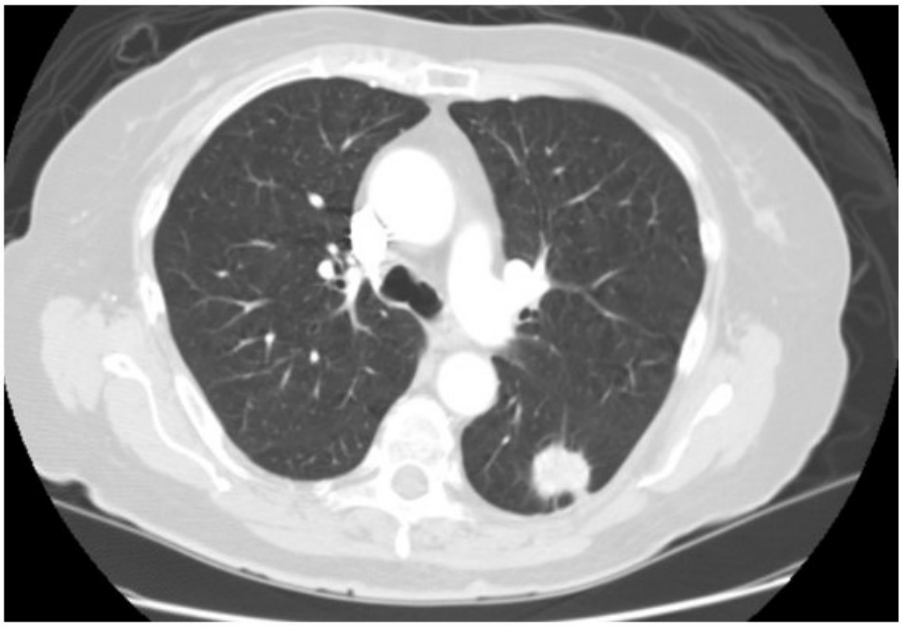

FIGURE 2.10: Lung Cancer in CT.

TABLE 2.2: Treatment associated to the type and stage of lung cancer. [17]

Type	**Stage**	**Primary Treatment**	**Adjuvant Treatment**	**5-Year Survival**
NSCC	I	Resection	Chemotherapy	60 to 70
	II	Resection	Chemotherapy with or without radiootherapy	40 to 50
	IIIA (resectable)	Resection with or without preoperative chemotherapy	Chemotherapy with or without radiotherapy	15 to 30
	IIIA (unresectable) or IIIB (involvement of controlateral or supraclavicular lymph nodes	Chemotherapy with concurrent or subsequent radiotherapy	None	10 to 20
	IIIB (pleural effusion) or IV	Chemotherapy or resection of primary brain metastasis and primary T1 tumor	None	10 to 15 (2-Year)
SCLC	Limited Disease	Chemotherapy with concurrent radiotherapy	None	15 to 25
	Extensive Disease	Chemotherapy	None	<5

Chapter 3

Deep Learning

3.1 Introduction to Machine Learning

Artificial intelligence (AI) refers to the capability of machines to simulate cognitive functions typically associated with human minds, such as learning and problem-solving. The AI research includes reasoning, knowledge, planning, learning, communication, perception, and object manipulation [26]. Today, AI is integrated into our daily lives in many forms, starting to be also incorporated into medicine, improving patient care via earlier detection and diagnosis, improved workflow, reducing medical errors, costs, morbidity, and mortality. It's not meant to replace human physicians but rather assists or augments the medical care [27].

Machine learning (ML) is a subfield of AI based on the ability of a computer system to adapt its processing based on newly acquired information learned from experience [27]. This technology powers many aspects of modern society: they are used to identify objects in images, transcribe speech in text, match new items, and even to select results of searches. Anyway, conventional ML techniques were limited in their ability to process natural data in their raw form; thus, these applications started and increased to use a class of techniques called deep learning (DL), a human-brain-like processing [27, 28].

DL are representation-learning methods with multiple levels of representation, composed of simple but non-linear modules. Each one of these modules transforms the representation at one level, starting with the raw input, into a representation at a higher, slightly more abstract level. If enough transformations are considered, even complex functions can be learned. It has turned out to be very good at discovering intricate structure in high-dimensional data and to be applicable to various science domains [28]. DL algorithms can be based on "supervised", "unsupervised" and "reinforcement" learning, better discussed in Section 3.2.2. The DL architectures have

been applied to fields like computer vision, automatic speech recognition, natural language processing, audio recognition, and bioinformatics [26]. Relevant clinical-ready successes obtained in health care as well, such as cancer classification in biomedical imaging [29]. It is demonstrated that DL algorithms have an ability to outperform other ML algorithms [26].

3.2 Artificial Neural Network

Artificial neural networks (ANN) are inspired by the 1959 biological model proposed by Nobel laureates David H. Hubel & Torsten Wiesel, who found two types of cells in the primary visual cortex: simple cells and complex cells. Many artificial neural networks can be viewed as cascading models of cell types inspired by these biological observations [26]. Thus, we can say that a neural network is a type of artificial intelligence that attempts to imitate the way a human brain works.

A biological brain is a huge collection of neurons, which take electrical and chemical signals as input, elaborate on them, and transmit the output through connections, called synapsis. ANN works analogously: a collection of connected units, called artificial neurons. Each connection carries a real number value, which determines the weight/strength of the signal [30]. Each neuron computes a weighted sum, called pre-activation, whose strength reflects the overall strength of the input and the match between the input pattern and the weight pattern. The pre-activation forms the input to the unit's activation function, discussed in Section 3.2.1 [31]. The weight can be positive (excitation) and negative (inhibition), and the higher it is, the higher will be the influence of one unit on another [30]. Each neuron adds to the pre-activation a bias, enabling the unit to shift its activation function horizontally [31]. A representation of an artificial neuron is shown in Figure 3.1.

Units can be assembled into networks in many different configurations [31], arriving even to thousands or millions of artificial neurons arranged in a series of layers, each of which connects to the layers on either side. Some units, called input units, are specialized to receive different forms of information from the outside, while others, called output units, are designed to respond to the information it's learned. In the middle, one or more layers of hidden units are present, which just process the input received. Most neural networks are fully connected, which means each hidden unit and each output unit is connected to every unit in the layers on either side [30]. This type of architecture is known as a feed-forward network, as shown in Figure 3.2,

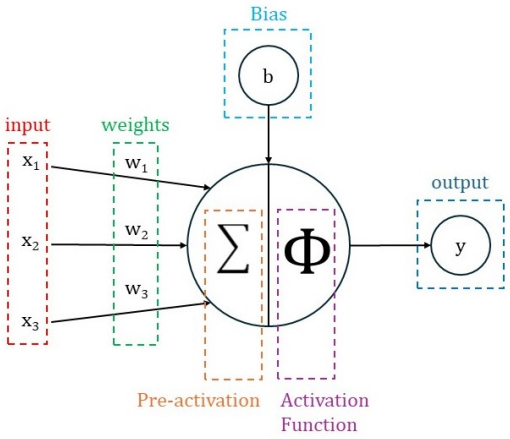

FIGURE 3.1: Representation of an Artificial Neuron.

for the reason that successive layers feed into one another in the forward direction from input and output.

Thus, an ANN is typically defined by three types of parameters: the interconnection design, the learning method for updating the weights, and the activation function [32].

3.2.1 Activation Function

In an ANN, activation functions are very important as they help in learning and making sense of non-linear and complicated mappings between the inputs and corresponding outputs. Without an activation function, the ANN

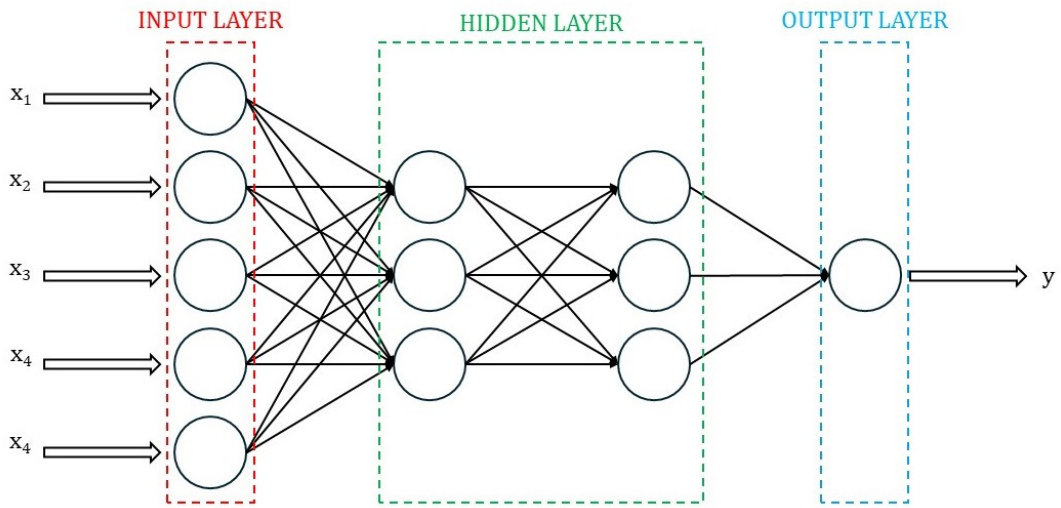

FIGURE 3.2: Architecture of a Feed-Forward network.

acts as a linear regression model, where the output would simply be a linear function, which is just a polynomial of degree one. Then, even if a linear equation is easy to solve, its complexity is limited, and it does not have the ability to learn and recognize complex mappings from data. Thus, we need to apply an activation function to make the network dynamic, add the ability to it to extract complex and complicated information from data, and represent non-linear convoluted random functional mappings between input and output [33].

There exist a lot of types of activation functions in literature, going from the simplest ones, such as the binary step function and the linear one, to the most complex ones with learnable parameters, used in deep learning. The main advantage of these more complex activation functions is that they can better learn the abstract features through nonlinear transformations, but, as disadvantages, they need large datasets for training [34].

Next, the commonly used activation functions in deep learning will be analyzed.

- *Sigmoid function*. It is one of the most common forms of activation function, and it transforms the values in the range 0 to 1 [33]. It is defined by Eq. 3.1.

$$g(x) = \frac{1}{1 + e^{-x}} \tag{3.1}$$

  The sigmoid is a continuous function, which means that it is differentiable everywhere. The derivative is defined by Eq. 3.2.

$$g(x) = \frac{e^{-x}}{(1 + e^{-x})^2} \tag{3.2}$$

  It is commonly used in shallow neural networks, mainly employed on the output level since its soft saturation. This soft saturation results in the difficulties of training a deep neural network, generating the vanishing gradient [35]. Also, the sigmoid function is not symmetric about zero, which means that the signs of all output values of neurons will be the same. This issue can be improved by scaling the function [33]. The Figure 3.3 shows the function plot.

- *Hyperbolic Tangent*. Hyperbolic tangent is similar to a sigmoid function, and it can be easily defined as the ratio between the sine and the cosine

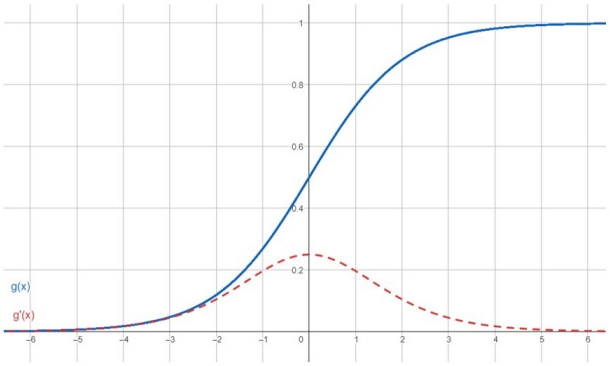

FIGURE 3.3: Sigmoid Function (blue) and its derivative (red).

functions, as shown by Eq. 3.3.

$$tanh(x) = 2sigmoid(2x) - 1 = \frac{e^x - e^{-x}}{e^x + e^{-x}} = \frac{sinh(x)}{cosh(x)} \tag{3.3}$$

The hyperbolic tangent is symmetric about the origin; thus, the outputs could have different signs with values ranging between -1 and 1. Its derivative is defined by Eq. 3.4.

$$tanh'(x) = 2sigmoid'(2x) - 1 = \frac{4e^{-2x}}{(1 + e^{-2x})^2} \tag{3.4}$$

It has the same soft saturation as the sigmoid function, which also has the vanishing gradient problem [35]. The Figure 3.4 shows the function plot.

- *Rectified Linear Unit (ReLU)*. In sigmoid and hyperbolic tangent functions, almost one half of the neuron units are activated at the same time,

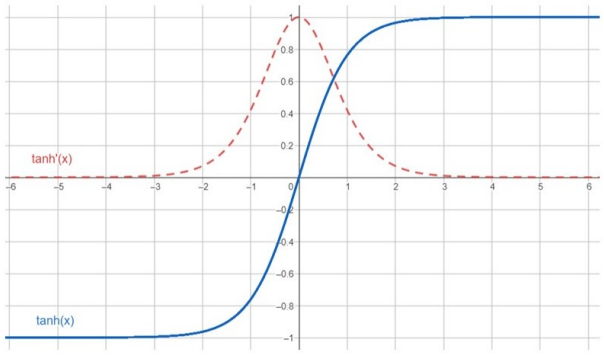

FIGURE 3.4: Hyperbolic Tangent Function (blue) and its Derivative (red).

which is inconsistent with neuroscience research that indicates only 1% to 4% of neurons in the brain can be activated simultaneously. ReLU helps the hidden layer obtain the sparse output matrix, improving the efficiency. It is defined by Eq. 3.5, while its derivative is by Eq. 3.6.

$$g(x) = max(0, x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{3.5}$$

$$g'(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{3.6}$$

The main advantages of ReLU can be summarized as:

- Cheaper computation because there is no need for computing the exponential functions;
- The ANN converges faster, being a non-saturated function;
- ANN obtains easy sparse representation;
- The constant value of the derivative avoids resolves the vanishing gradient effect [35];

The Figure 3.5 shows the function plot.

- *ReLU Variants*. The compulsive operation of letting $g(x) = 0$ when $x < 0$ with ReLu uses the death of some neuron units. Thus, to alleviate this problem, it was introduced the Leaky Rectified Linear Units (LReLU) function, which allows for a small, non-zero gradient when the unit is saturated and not active. The function and its derivative are defined by

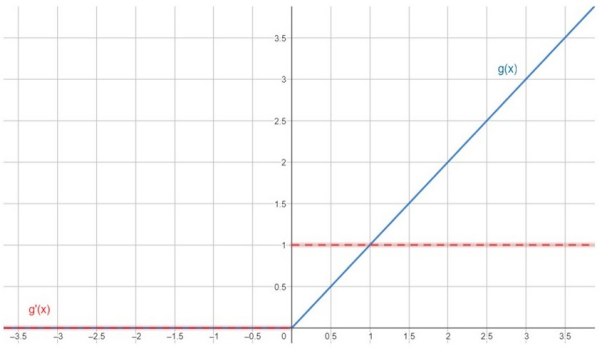

FIGURE 3.5: Rectified Linear Unit Function (blue) and its Derivative (red).

Eq. 3.7 and Eq. 3.8.

$$g(x) = max(0, x) = \begin{cases} x, & x \geq 0 \\ 0.01x, & x < 0 \end{cases} \tag{3.7}$$

$$g'(x) = \begin{cases} 1, & x \geq 0 \\ 0.01, & x < 0 \end{cases} \tag{3.8}$$

The main disadvantage of the LReLU is the loss of the sparse.

Another variant of the ReLu is the Parametric Rectified Linear Unit (PReLU), which is obtained by replacing the constant 0.01 with a learnable parameter a. This means that PReLU can learn the parameter from the data, resulting in faster convergence and lower train error. It seems also that the use of the PReLU solves the overfitting problem. The PReLU is defined by Eq. 3.9.

$$g(x) = \begin{cases} x, & x \geq 0 \\ ax, & x < 0 \end{cases} \tag{3.9}$$

Another improvement of ReLU is the Randomized Rectified Linear Unit (RReLU), where the slopes are randomized in a given range of training, sampled from a uniform distribution $U(A, B)$, and then fixed in the testing as the average of all parameters taken during the training: $(A + B)/2$. The RReLU is defined as the PReLU by Eq. 3.9, considering a random number and not a learnable parameter [35]. The Figure 3.6 shows these functions plot.

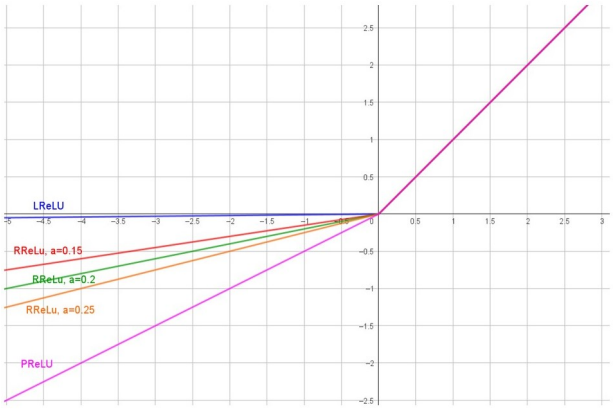

FIGURE 3.6: Rectified Linear Unit Variants.

- *Exponential Linear Unit (ELU).* ELU is also a variant of ReLU, which pushes the activation means closer to zero to decrease the bias shift effect of ReLU. It is defined by Eq. 3.10, while its derivative is Eq. 3.11. Both have $\alpha > 0$.

$$g(x) = \begin{cases} x, & x > 0 \\ \alpha(e^x - 1), & x \leq 0 \end{cases} \tag{3.10}$$

$$g'(x) = \begin{cases} 1, & x > 0 \\ \alpha e^x, & x \leq 0 \end{cases} \tag{3.11}$$

  The parameter α manages the value to which an ELU saturates for negative network inputs. The layers of the deep neural network with ELU can enable faster learning and better generalization performance than ReLU and LReLU [35]. The Figure 3.7 shows this function plot.

- *Swish function.* The Swish function is a relatively new activation function that is not monotonic. This means that the value of a function may decrease even though the values of inputs are increasing. It can outperform even the ReLU [33]. It is defined by Eq. 3.12.

$$g(x) = \frac{x}{1 + e^{-x}} \tag{3.12}$$

  The Figure 3.8 shows its plot.

- *SoftMax function.* It is a combination of multiple sigmoid functions; thus, it can be used for multiclass classification problems, while the sigmoid function is only for binary classification. Softmax returns the probability for every data point of all the individual classes [33]. It is

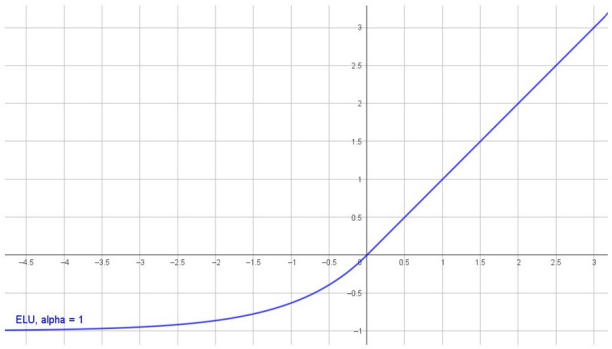

FIGURE 3.7: Exponential Linear Unit Function.

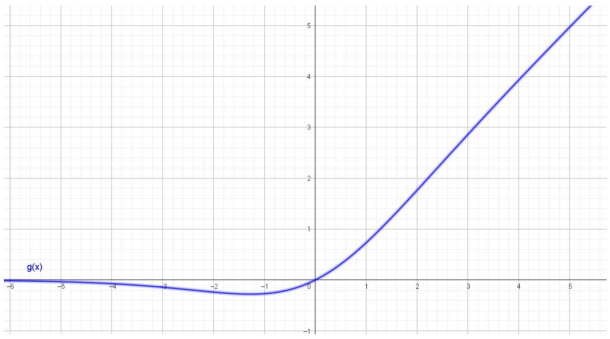

FIGURE 3.8: Swish Function.

defined by Eq. 3.13.

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \tag{3.13}$$

The Figure 3.9 shows its plot.

3.2.2 Learning and Training

Learning is a comprehensive term: the system changes itself to adapt. A neural network can learn by developing or deleting connections, changing connections weight and neurons' threshold values, varying the activation function, and developing or deleting new neurons [36].

There exist three main learning methods:

- *Unsupervised*. The training set only consists of unlabeled data; the network tries by itself to detect similarities and to generate pattern classes [36]. Data has no labels, and the "right answer" is not known [26].

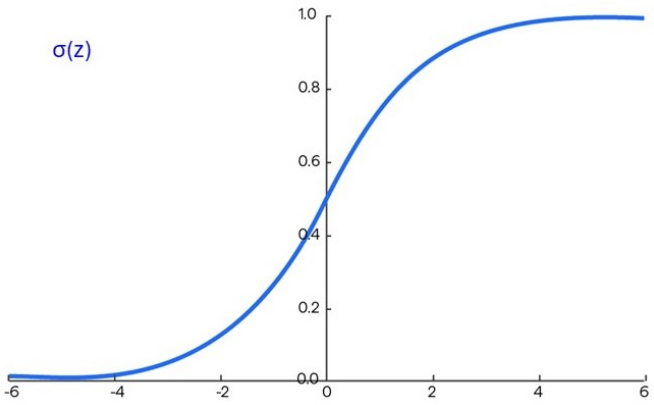

FIGURE 3.9: SoftMax Function .

- *Reinforcement*. The training set consists of input data, and, after completion of a sequence, a logical or real value is returned to the network indicating whether the result was right or wrong and, possibly, how right or wrong it was [36].

- *Supervised*. The training set consists of input labeled data so that the network can directly compare its own output with the correct solution. The network weights can be changed according to their differences. It is not always biologically plausible, as the unsupervised one, but it is extremely effective. The procedure on which the supervised learning is based is: entering the input data, forward propagation of the input and generation of the output, comparing it with the desired one, computing the error vector, and applying corrections based on this vector [36].

Supervised learning is the most used learning method for DL. It is useful to divide the set of training samples into a training set, used to train the model, and a testing set, with a proportion of 70/30% randomly chosen.

The difference between the predicted and the true value is computed and measured by the loss function or objective function. A lot of loss functions can be found in literature, and their choice is critical in defining the outputs in a way that is sensitive to the application at hand [37].

The learning curve indicates the progress of the error, thus, whether the network is progressing or not. A perfect learning curve looks like a negative exponential function (Figure 3.10) [36]. This is because the training procedure is just an iterative computation that must minimize the loss function.

The training of a neural network specifically contains two phases: forward and back propagation, as shown in Figure 3.11. The forward propagation phase was already partially described with the supervised learning. It

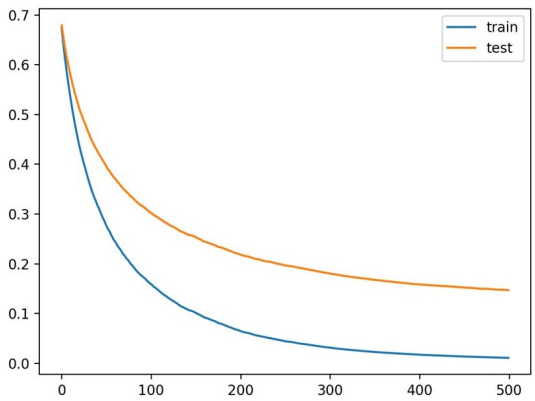

FIGURE 3.10: Learning Curves.

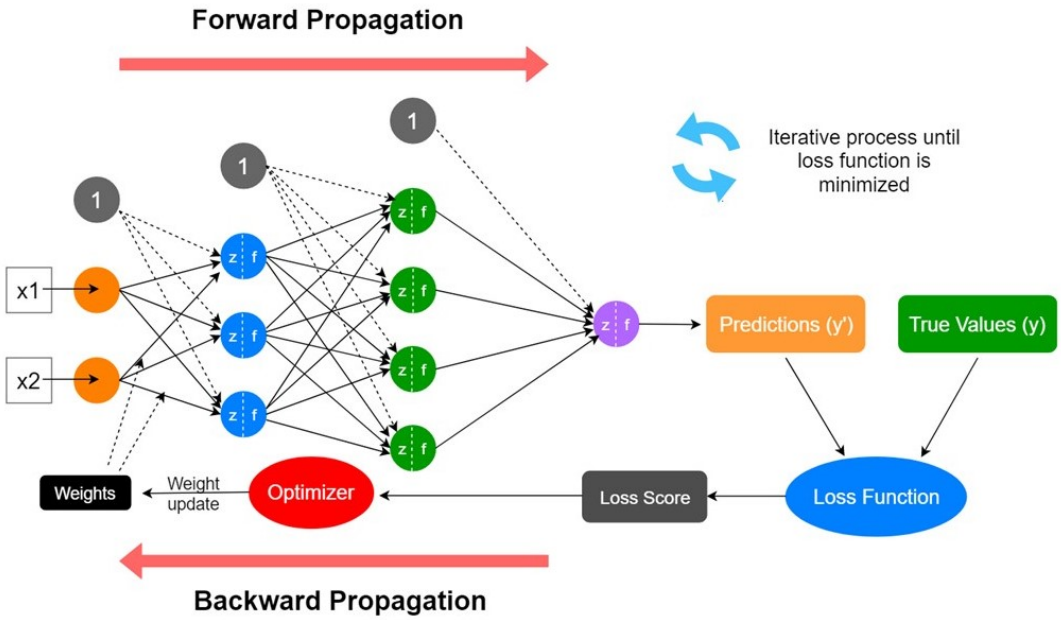

FIGURE 3.11: Forward and Back Propagation Network Architecture.

consists in feeding the inputs for the training into the network, resulting in a forward cascade of computations across the layers using the current set of weights. The loss function, its derivative with respect to the output, and its derivative with respect to the weights on each layer are computed.

The back propagation phase has the main goal of learning the gradient of the loss function with respect to the different weights by using the chain rule of differential calculus. These gradients are then used to update the weights and biases [37]. The back propagation, also known as the back propagation of errors, was introduced in 1986 by Rumelhart et al. (1986). In the DL context, it is a gradient descent optimization algorithm [38]. Gradient descent means to go downhill in small steps from any starting point of our function towards the gradient g, with the size (learning rate) of the steps being proportional to $|g|$. There are three main problems sources of error on gradient descent procedure (Figure 3.12):

- Get stuck within a local minima;
- Descent stop when passing a flat plateau;
- Miss a good minima due to too large step size;
- Oscillations due to a strong positive and negative gradient alternation [36].

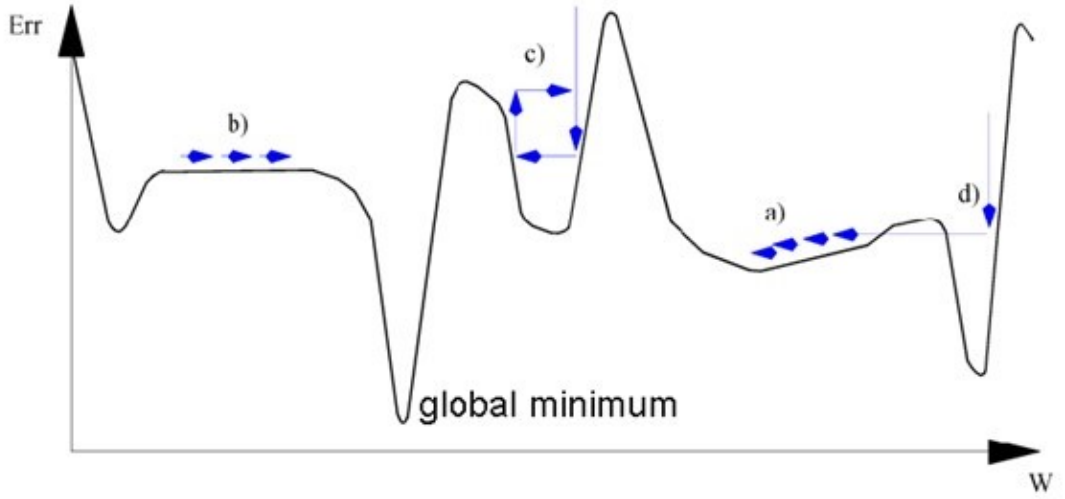

FIGURE 3.12: **a.** Detecting bad minima;**b.** Quasi-standstill with small gradient; **c.** Osscillation in canyons; **d.** Leaving the good minima. [36]

3.2.3 Generalization and Optimization

Training neural networks is a complex procedure; thus, several issues could occur. The most important ones are the overfitting and underfitting, related to the concept of generalization.

Generalization is the ability to perform well on previously unobserved inputs. Typically, during the training, as described before, it is computed the training error, and it is applied the optimization problem to minimize it. The distinction between ML and simple optimization lies in the focus on minimizing generalization error or test error. It is defined as the expected value of the error on an unseen input, the test set. The factors determining how well a model will perform are its ability to make the training error small and make the gap between training and test error small. They are connected to the two problems mentioned at the beginning.

Underfitting occurs when the model is not able to obtain a sufficiently low error value on the training set. Underfit networks are not suitable models with poor performance. This issue could be solved by redesigning the network.

Overfitting occurs when the gap between the training error and test error is too large [38]. This means that the model trained on a particular dataset does not guarantee that it will provide good performance on unseen test data, even if the model predicted the target in the training set perfectly [37].

Typically, training error decreases until it asymptotes to the minimum

possible error value as model capacity (ability to fit a wide variety of functions) increases, while, typically, generalization error has a U-shaped curve as a function of model capacity, as shown in Figure 3.13. The objective is to find the case of optimal capacity. The strategies used to reduce the test error, possibly at the expense of increased training error, are known as regularization [38]. The widely used ones are:

- *Dataset Augmentation*. Create fake data and add it to the training set to augment the limited amount of starting data. It is very effective for the object recognition classification problem because the images contain a lot of factors of variation that are easy to simulate [38].

- *Validation*. The starting dataset is not anymore divided into training and testing sets, but in three different sets. Two of them remain the training (build the model) and the test (test the accuracy of the tuned model), while the third one is a subset of the training set, called the validation set. It is used for model selection and parameter tuning. The validation data can be viewed as a kind of test data set to tune the parameters or to select the best design choice [37].

- *Early Stopping*. It uses the concept of validation. It evaluates at each epoch the training and the validation errors. Even if the training error is still decreasing, when the validation error starts to increase, it means that further training will cause overfitting. This point is called "early stopping". Before it, the training error is too high, thus the network is underfitted, while after it it's overfitted. The early point is the right point to stop training [37, 38]. Figure 3.13 shows the early stopping principle, coinciding with the point of optimal capacity.

- *Ensemble methods*. Bagging or Bootstrap aggregation is based on sampling the training dataset with replacement. The idea is to train several models separately, then apply to each model the test set, obtaining as many predictions as the number of models. Finally, the predictions are then averaged to yield a single prediction [37, 38].

  Subsampling is based on sampling the training dataset without replacement. Then act exactly like in the bootstrap. The only constraint is to select the number of samples in each model that is mandatory lower than the dimension of the starting dataset; otherwise, all the training sets of the models will be the same [37].

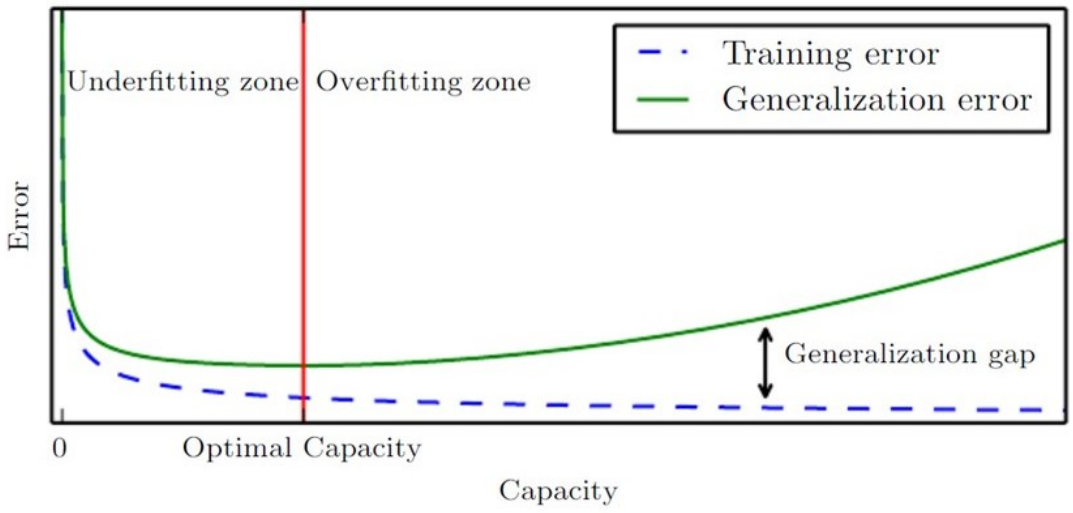

FIGURE 3.13: Relationship between Error and Capacity. [38]

Another issue in training a neural network is the vanishing and exploding gradient. They occur in networks with many layers, affecting their stability, and it is caused by the chain-rule product computation. In particular, the updates in earlier layers can either be really small (vanishing gradient), preventing the updating of the weights, or they can be really large (exploding gradient), causing a big update of the weights.

A method to address these two problems is batch normalization (BN). It causes activation gradients in successive layers to either increase or reduce in magnitude. Basically, it reparametrizes the model to make some units always be standardized by definition. It could be used just before or after an activation function, and it is applied as a layer in the architecture so that the input of this layer is normalized with a fixed mean and a fixed variance.

Another issue is the convergence difficulties: the loss function does not converge to the best result due to the "resistance" to the training process of the deep networks.

The last issue discussed is the local optima. It occurs when the optimization converges not in the global optima but in a local one. Lots of local optima are present due to the non-linearity of the neural network. A possible solution is the pretraining, so improving the initialization of the model [37].

3.3 Convolutional Neural Network

The CNN functions much like a feed-forward neural network, except that the operations in its layers are spatially organized with sparse connections

between layers [37].

The training of a CNN is based on finding the kernel size and the fully connected layer weights, discussed below [39]. The overall network (Figure 3.14) is built stacking in repetition of three types of layers:

- *Convolution Layer*. The convolution layer is fundamental. It performs feature extraction, which typically consists of a combination of linear (convolution) and non-linear (activation function) operations [39]. The convolution is an operation on two functions of a real-valued argument [38]. A small array of numbers, called a kernel, is applied across the array of numbers in input, called a tensor, to compute an element-wise product at each location of the tensor, then summed to obtain the output value, called a feature map. The usual step movement of the kernel, called a stride, is 1 (Figure 3.15). This procedure can be repeated with different kernels, obtaining different feature maps that represent different characteristics of the tensor (Figure 3.16). Fundamental is to define the size (3x3, 5x5, 7x7) and the number of kernels to determine the depth of output feature maps. A zero padding on each side of the tensor could be applied to allow the center of the kernel to overlap the outermost element of the tensor (Figure 3.17). The linear output of the convolution passed through a non-linear activation function, typically ReLU, discussed previously [39]. The role of the convolutional layer is to detect local conjunctions of features from the previous layer [28].

- *Pooling Layer*. A pooling layer performs a downsampling operation that reduces the in-plane dimensionality of the feature maps to introduce a

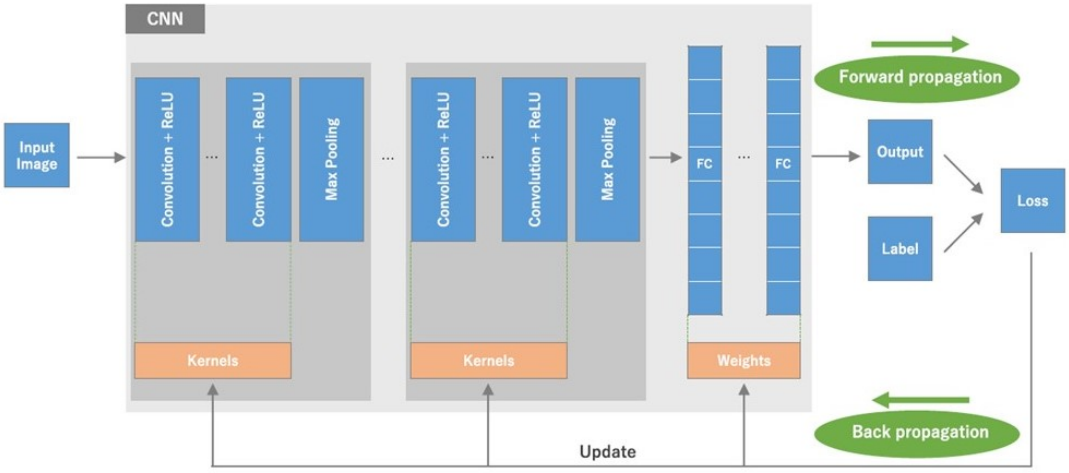

FIGURE 3.14: CNN Architecture. [39]

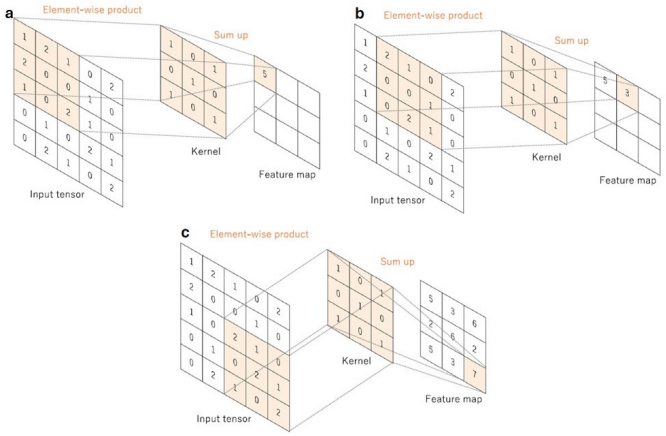

FIGURE 3.15: **a-c.** An example of convolution operation with a kernel size of 3 × 3, no padding, and a stride of 1. [39]

translation invariance to small shifts and distortions and decreases the number of learnable parameters. There exist two types of pooling. The Max Pooling is the most popular one. It subdivides the feature maps into patches, and from each patch select the maximum value and discard the others. It reduces the height and width of the images but not the depth dimension (Figure 3.18). The Global Average Pooling is an extreme operation, applied only once before the fully connected layer. It downsamples a feature map into a 1x1 array, taking the average of the elements (the depth remains the same as before). It reduces the number

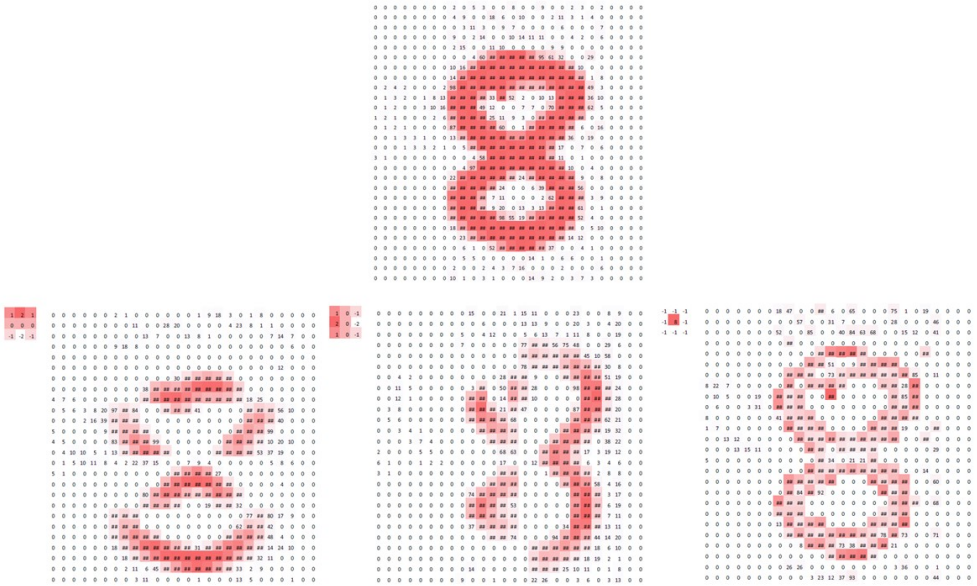

FIGURE 3.16: A convolution operation with zero padding. [39]

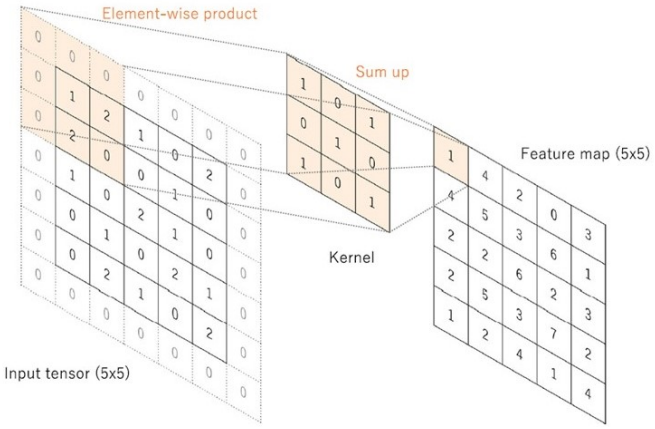

FIGURE 3.17: A convolution operation with zero padding. [39]

of learnable parameters and enables the CNN to accept variable size inputs [39]. The role of the pooling layer is to merge semantically similar

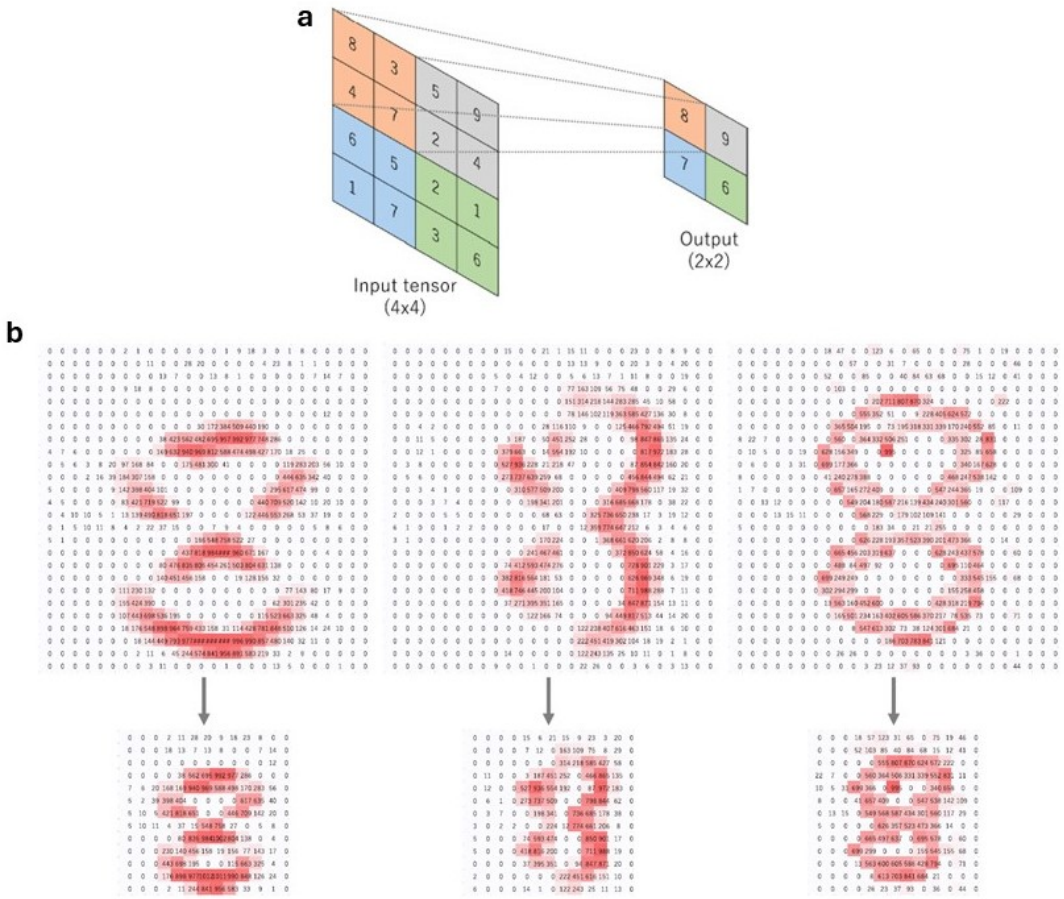

FIGURE 3.18: **a.** An example of max pooling and (**b.**) an example of max pooling applied on the same images of Figure 3.16. [39]

features into one [28].

- *Fully Connected Layer*. The output of the previous layer is flattened, reduced to a 1D vector, and connected to one or more fully connected layers, which work exactly as a feed-forward network. They map the features extracted before in the final outputs, such as the probabilities for each class in classification tasks. Each layer is followed by a ReLU, while the last one has an appropriate activation function for the task. In multiclass classification is often used a softmax [39].

The discussion and the images above refer to a CNN with a 2D tensor, but it could be extended simply also to a CNN with volumetric 3D inputs.

The CNN can be applied in healthcare to facilitate the classification, segmentation, or detection of anatomical structures in biomedical images. In particular, following the aim of this study, the segmentation field will be discussed in more detail.

3.4 Biomedical Image Segmentation

Segmentation of organs or anatomical structures is a fundamental image processing technique for medical image analysis. It is the process of partitioning the digital images into multiple segments, used to identify the object of interest or the related information. It could be performed manually by clinicians, but it could be used as a less-consuming process thanks to the CNN, which will reduce the time and the personnel needed.

To train a segmentation network, the training data consists of the medical images of the organ of interest and the segmentation masks. The process is based on predicting the probability of each pixel or voxel belonging to a specific anatomical structure using a CNN. Firstly, the probability map was constructed, and then the global context of the images was used to refine the output [39]. An example of the segmentation is shown in Figure 3.19. In the following subsection, some state-of-teh-art CNNs for segmentation are discussed.

3.4.1 Fully Convolutional Network

The Fully Convolutional Network (FCN) was proposed by Long et al. in 2015 [40], who wanted to build a network that takes input of arbitrary size and produces correspondingly sized output with efficient inference and learning.

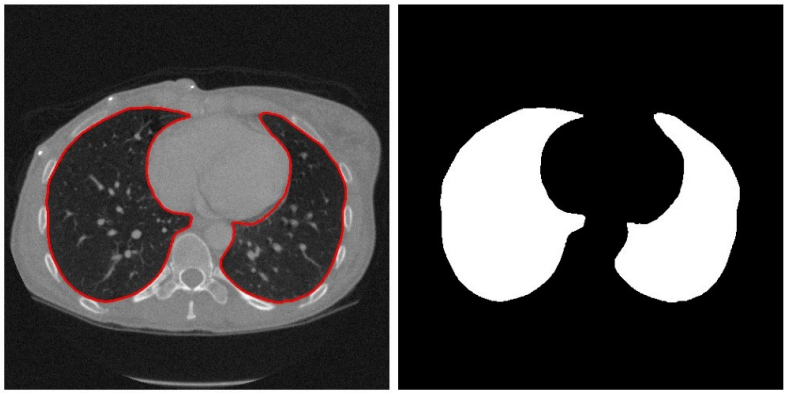

FIGURE 3.19: Biomedical Image Segmentation.

It was trained end-to-end, pixels-to-pixels on semantic segmentation, and it exceeded the at-time state-of-the-art, thus the CNN. This model transfers recent success in classification to dense prediction by reinterpreting classification nets as fully convolutional and fine-tuning from their learned representations.

The fully connected layers of typical recognition nets were "transformed" into convolution layers, seeing them as convolutions with kernels that cover their entire input regions, to preserve spatial information. Then the introduction of upsampling layers in substitution of some pooling operators increases the image resolution of the output. Finally, the output will be a dense prediction map, with the same size of the input but with the desired number of output channels, thus the classes in the segmentation problem. The FCN is shown in Figure 3.20

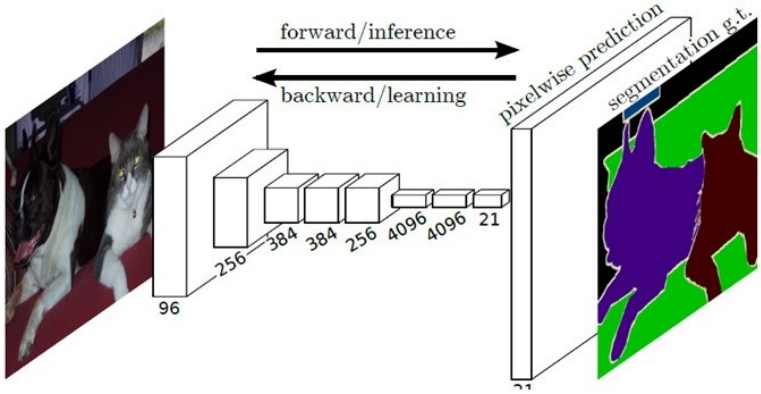

FIGURE 3.20: Fully Convolutional Network. [40]

3.4.2 U-Net

The U-Net (Figure 3.21) was proposed by Ronneberg et al. in 2015 [41], introducing one of the most widely used network architectures for biomedical image segmentation.

This architecture consists of a contracting path to capture context information and a symmetric expanding path that enables precise localization and can be trained end-to-end from very few images. It was an extension of the FCN [40] such that it works with very few training images and yields more precise segmentation; the main modification was the adding of feature channels in the up-sampling part, which allow the network to propagate context information to higher resolution layers.

The input images and their corresponding segmentation maps are used to train the network with the stochastic gradient descent implementation. To reach the training with few images, it is necessary the data augmentation to teach the network the desired invariance and robustness. Especially, random elastic deformations of the training samples are the key concept to training a segmentation network with very few annotated images.

The architecture is composed of a contracting path, which follows the traditional CNN architecture, and an expansive path. The contracting path is

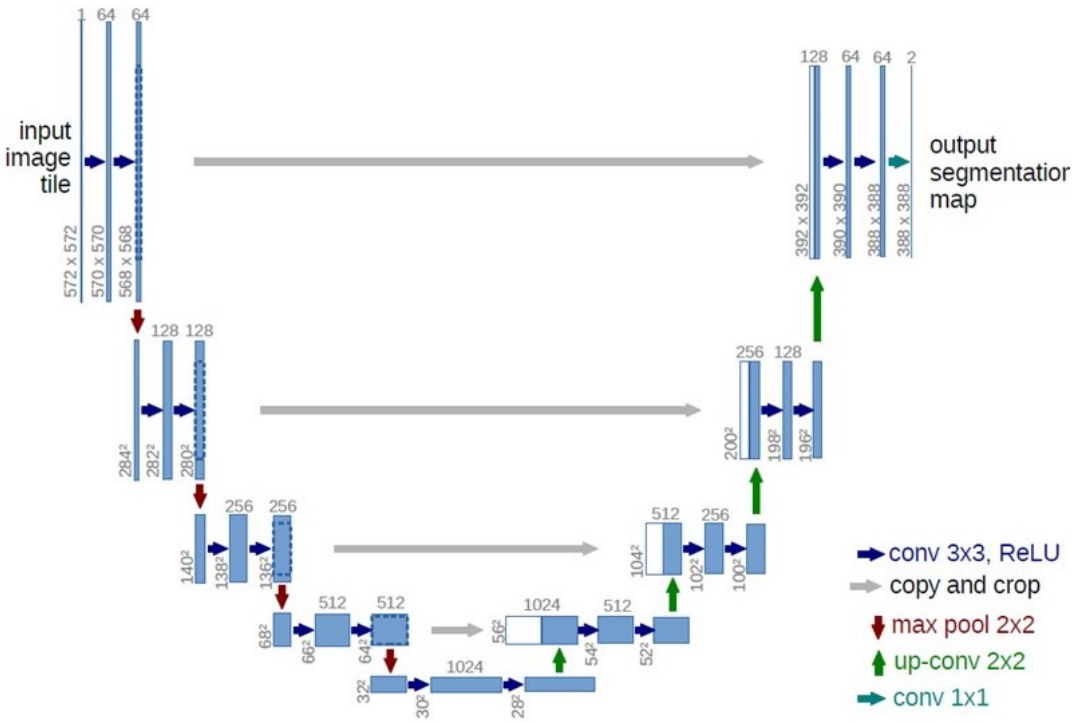

FIGURE 3.21: U-Net Architecture. [41]

the repetition of two 3x3 convolutions, each one followed by a ReLU, and a 2x2 max pooling as down sampling. At each down sampling, the feature channels are doubled. The expansive path consists of an upsampling followed by a 2x2 convolution, where the number of feature channels is halved, a concatenation with the correspondingly cropped feature map from the contracting path, necessary for the border pixels loss in every convolution, and two 3x3 convolutions, each one followed by a ReLU. A final 1x1 convolution maps each 64-component feature vector to the desired number of classes.

3.4.3 3D U-Net

The 3D U-Net was proposed by Cicek et al. in 2016 [42] and introduced an extension of the U-Net by replacing all 2D operations with their 3D counterparts (3D convolutions, 3D max pooling, 3D up-convolution layers), and adding a BN before each ReLU.

This network learns to generate dense volumetric segmentations only requiring some annotated 2D slices for training; it aims on densification of a sparsely annotated dataset, or it learns from multiple sparsely annotated datasets to generalize to new data. The two different ways the network can be used are shown in Figure 3.22.

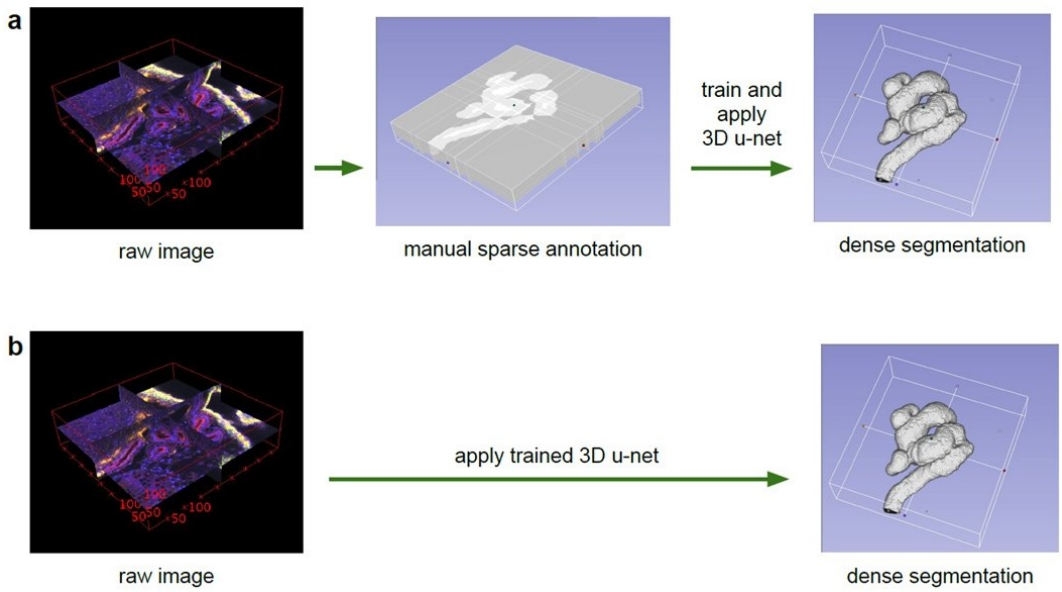

FIGURE 3.22: Volumetric segmentation with the 3D U-Net. **a.** Semi-automated segmentation; **b.** Fully-automated segmentation. [42]

Chapter 4

Literature Review

4.1 Introduction

In the last years, DL techniques have changed the approach to medical imaging analysis. A lot of studies focused their work on creating a DL model for lung nodule segmentation in CT images.

Most recent works started to apply CNN to that task, introducing the numerous advantages of these types of networks. However, this field is not without challenges, due mainly to the heterogeneity of the CT images and acquisition protocol and to the high dimension of data for the training.

Since this study aims to create a deep learning model for lung nodule segmentation in CT images, this literature review will provide a review focused on the most recent works in this field.

4.2 Methods

The literature review was conducted using a query-searching strategy on three electronic bibliographic databases, namely, PubMed, Scopus, and Web of Science.

The root 'lung cancer' was used to search for studies related to that specific tumor; the root 'deep learning' was used to search for studies using machine learning techniques; the roots 'ct' and 'computed tomography' were used to focus the search on studies with computed tomography image datasets; and the root 'segmentation' was used to search for studies concerning the segmentation of lung nodules. All these roots were limited in the 'Title and Abstract' fields of search; terms within each concept were combined with the boolean operator 'OR' and then combined with the boolean operator 'AND'. Moreover, two roots ('classification' and 'extraction') were used to exclude

the studies that do not focus only on segmentation techniques. Free access and English papers were the only ones considered.

The papers were collected and managed in the Zotero reference management system; the duplicates were removed, and title and abstract analyses were performed to select only the interesting documents.

4.3 Results

Overall, 89 articles were identified in the bibliographic databases. 40 were duplicated, 21 were excluded after the title analyses, and 13 after the abstract one. At the end, 15 studies were selected. The study inclusion procedure is shown in Figure 4.1

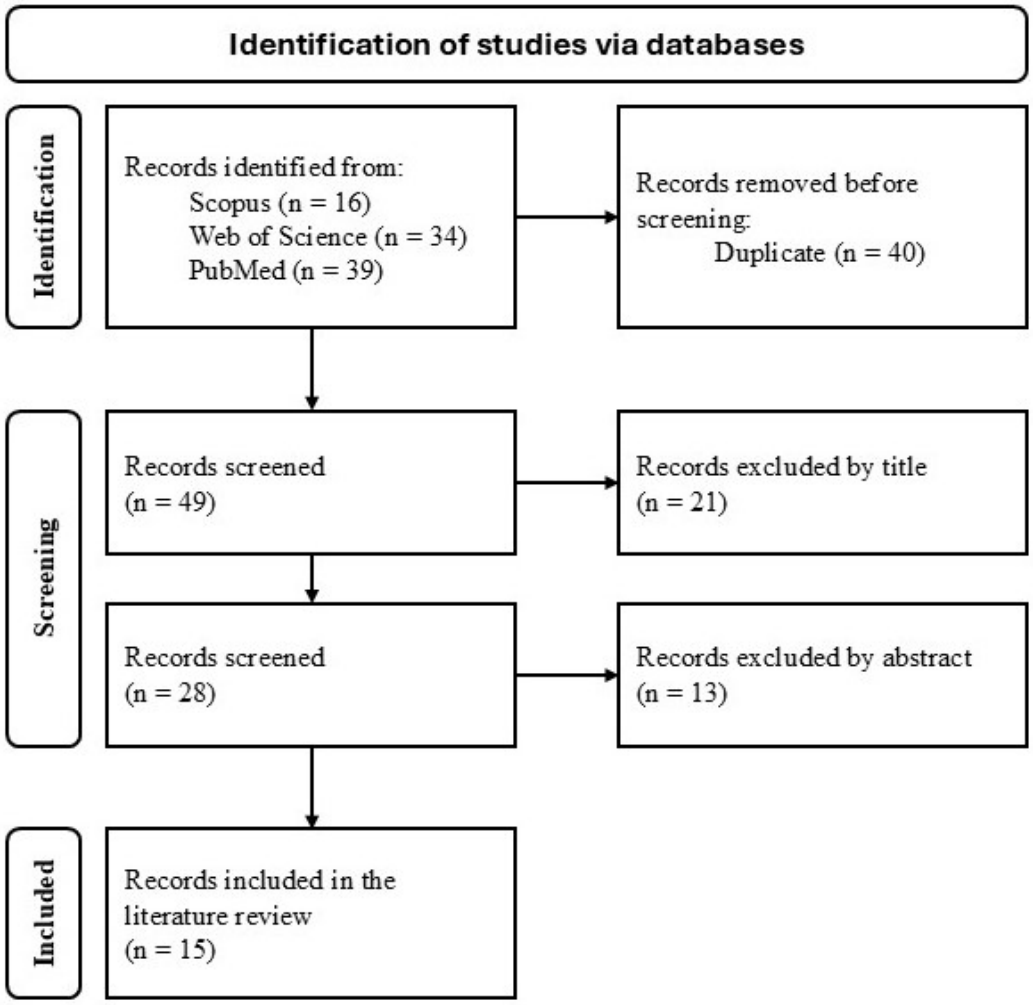

FIGURE 4.1: Performed literature search and study selection

4.3.1 Shuo Wang et al. (2017)

This study [43] proposed a data-driven model, named Central Focused Convolutional Neural Network (CF-CNN), to perform lung nodule segmentation on heterogeneous CT images, capturing a set of nodule-sensitive features from 3D and 2D CT images simultaneously.

The architecture, shown in Figure 4.2a, includes two identical deep branches, composed of six convolutional layers (C1 to C6), subdivided in three blocks, with each block including two 3x3 convolutional layers and aPReLU after each convolutional layer. Between two blocks is present a central pooling layer, while after the last a fully connected layer (F7) is applied and combined with the fully connected layer of the other branch through another fully connected layer (F8). The two parallel branches are relative to 3D and 2D CNN, respectively. The central pooling layer is proposed by the study and shown in Figure 4.2c. The main characteristics are that the kernel size varies according to the pooling position and is non-uniformly distributed on the input image, while in traditional max pooling the kernels are of the same size and uniformly distributed, as in Figure 4.2b.

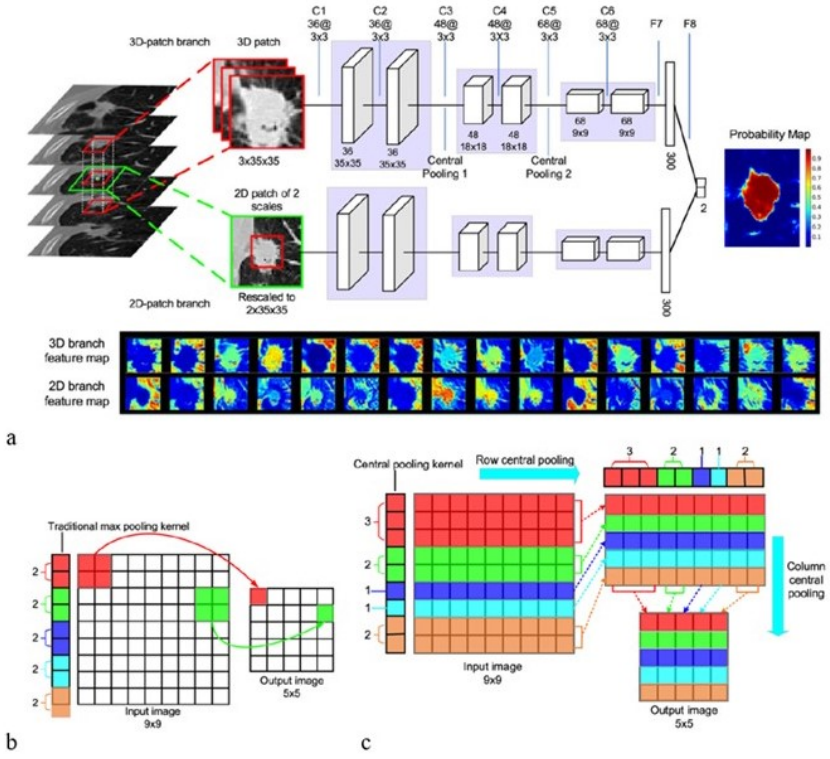

FIGURE 4.2: The network architecture proposed by Shuo Wang et al. (2017); **a.** Architecture; **b.** Tradtional pooling process; **c.** Central pooling process. [43]

The network was tested with the public dataset LIDC, which includes 893 nodules, and an independent dataset from Guangdong General Hospital, which includes 74 nodules. The first dataset was divided into 350 nodules for the training set, 50 for the validation set, and 493 for the testing set, while the second dataset was completely used as the testing set.

The results suggest that the CF-CNN achieved superior segmentation performance in the two datasets compared to several widely used lung nodule segmentation methods. It has high-performance segmenting nodules attached to pleura.

4.3.2 Zisha Zhong et al. (2018)

In this paper [44], it was proposed a novel approach for the segmentation of lung tumors. It combines the 3D U-Net, applied separately to PET and CT images to extract high-level discriminative features and generate tumor masks and probability maps, and the graph-cut-based co-segmentation model to obtain the final segmentation result. This combination gives the advantage of automatized localization of tumors.

The architecture is composed of two 3D U-Nets, which take as inputs pre-processed images. The encoder contains four 3x3x3 convolutional and 2x2x2 max pooling layers with 32, 64, 128, and 256 feature maps, while the decoder contains four deconvolutional, with factor 2, and convolutional layers with 256, 128, 64, and 32 feature maps. After each deconvolutional layer, the map is concatenated with the corresponding features in the encoder. At the end, the probability maps are obtained with a Softmax classifier, and a co-segmentation model generates the output as described in [45]. This architecture is shown in Figure 4.3.

The network was tested with a private dataset composed of 32 co-registered PET-CT scans, split into 20 for the training set and 12 for the testing set.

The quantitative results showed a better performance than the previous semi-automatic approach, using or not the co-segmentation model, demonstrating the learning ability of the 3D U-Net. The segmentation performance is however better in CT images than PET images. When combined with the co-segmentation model, the performance improves.

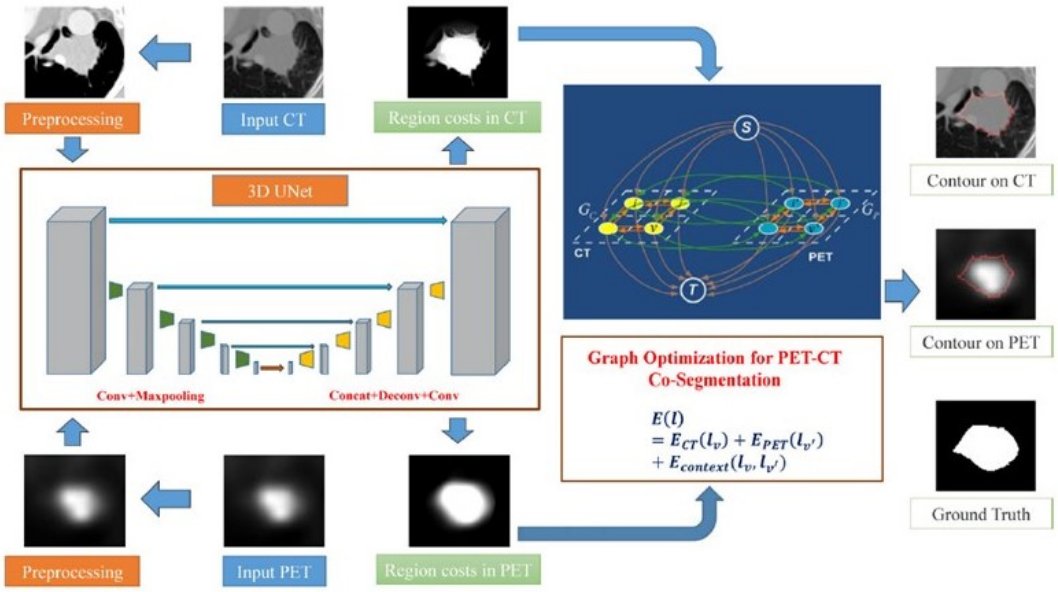

FIGURE 4.3: The network architecture proposed by Zisha Zhong et al. (2018) [44]

4.3.3 Liang Zhao (2020)

This paper [46] proposed a 3-dimensional densely connected convolution neural network for the segmentation of lung parenchyma in CT images, considering this type of deep learning approach better in accuracy results than those based on 2-dimensional segmentation. This is due to the CT images, which are a kind of 3D image with a lot of 3D information, lost and not used by the 2D segmentation network.

The architecture proposed, shown in Figure 4.4a, is a 3D FCN with three densely connected blocks. The first one, DenseBlock1, is composed of 4 densely connected layers, where, in each layer, there are a BN, a ReLU, and a 3x3x3 convolutional layer, with a growth rate of 16. DenseBlock2 and DenseBlock3 are similar to DenseBlock1, but with 8 and 16 dense layers, respectively. Their dense connectivity is represented by Figure 4.4b, showing that each layer has as inputs all the features extracted from all the previous layers. The input of the first dense block is a 64 convolutional filter 3x3x3 and a Max Pooling, while between two consecutive dense layers there are a BN, a ReLU, a convolutional layer 1x1x1 and a 2x2x2 max pooling layer. The downsampling path is formed by a BN, a ReLU, 2 convolutional layer 1x1x1 and 3 deconvolutional layer 3x3x3, to have the output images sized as the input ones.

The network was tested with the public dataset LIDC-IDRI, composed of

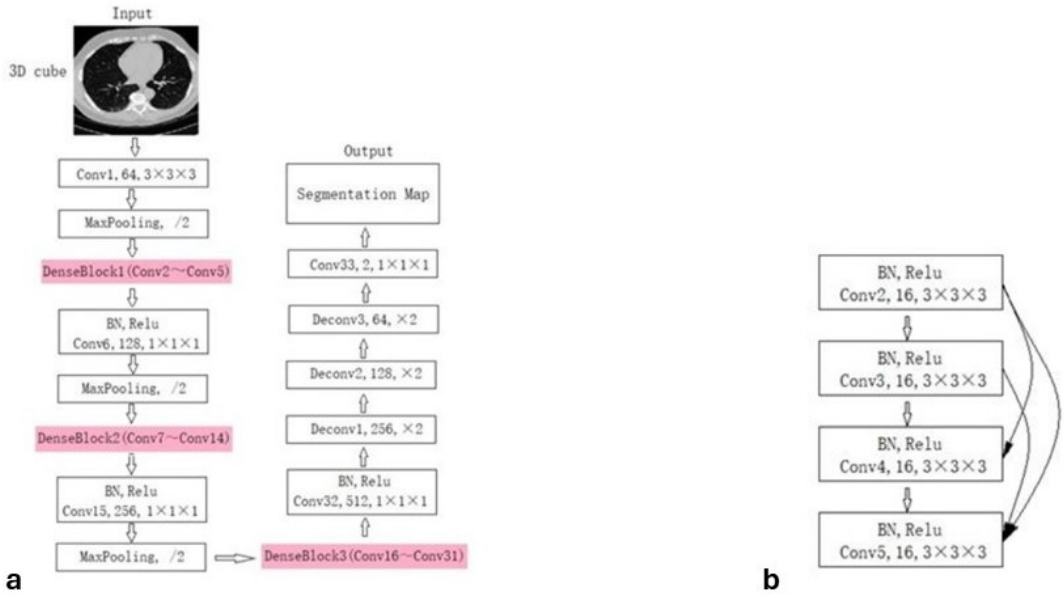

FIGURE 4.4: The network architecture proposed by Liang Zhao (2020); **a.** Architecture; **b.** DenseBlock structure. [46]

888 samples, divided into 708 for the training set and 90 for the validation and testing sets.

The results obtained were compared to those obtained with a 3D U-Net on the test set, showing that the network proposed can generally achieve better performance, despite the fact that the number of parameters is half.

4.3.4 Boris Shirokikh et al. (2021)

This paper [47] proposed a new accelerated segmentation method with a human-like technique to segment a 3D study. It is based on a rough analysis of the whole images to identify firstly the area of interest, lung nodules, and then locally segment each small part independently.

The architecture, LowRes, shown in Figure 4.5, is a 3D implementation of U-Net, with a residual block (ResBlocks), a BN, and a ReLu after each convolution, except the output one. It includes a low-resolution segmentation, a CNN segmentation model to predict the 8x8x8 times down sampled probability map, and a detailed segmentation, which iteratively and locally aggregates features from the first stage and predicts, in the original solution, the segmentation map.

The model was tested with the public dataset LIDC-IDRI, including 888 3D chest scans, divided into 534 for the training set, 178 for the validation set and 174 for the testing set.

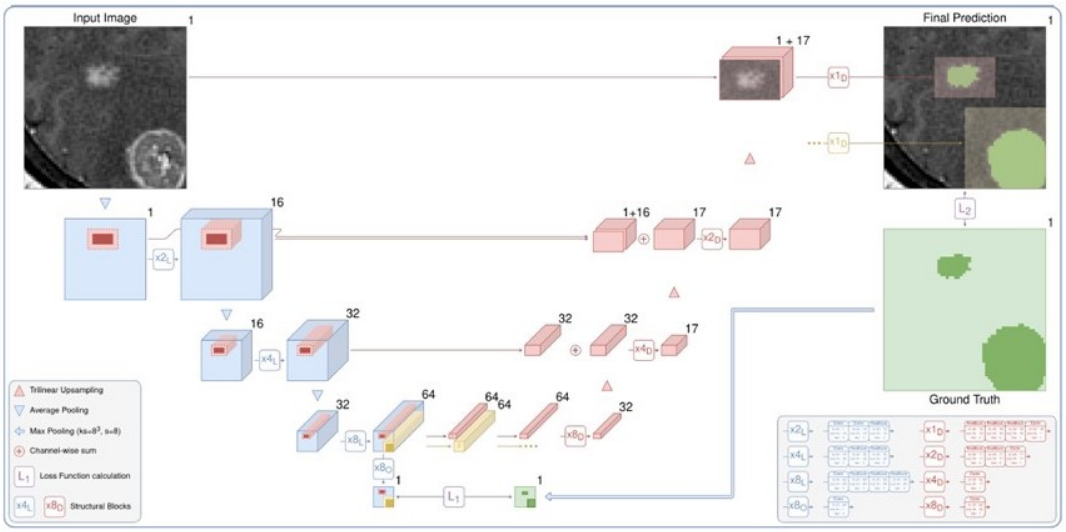

FIGURE 4.5: The network architecture proposed by Boris Shirokikh et al. (2021) [47]

The results indicate that this architecture achieves an inference speed close to that of mobile networks and simultaneously preserves or even increases the performance of the state-of-the-art segmentation network.

4.3.5 Wei Chen et al. (2021)

This paper [48] presented a novel deep learning-based approach for lung cancer segmentation from CT images, called Multiple Attention 3D U-Net (MAU-Net).

The architecture is based on a base U-Net, shown in Figure 4.6a, which has four levels in both encoder and decoder. In each level, a stride convolution reduces the feature's map resolution starting from 160x256x40, while the number of channels increases at a ratio of 2 starting from 32. At the end of the encoder is placed the Dual Attention Module (DAM). Firstly, a 3x3x3 convolution layer with 128 channels is applied, generating two compressed feature maps, which are fed one into a spatial attention block (SAB) and one into a channel attention block (CAB). The first block applies three 1x1x1 convolutions in parallel and reshapes them. The first, after their multiplication, are fed into a Softmax layer to generate the attention map, then multiplied with the third reshape, and finally multiplied with the starting feature map. The CAB models the relationship between different channels, performing the same model as SAB without the staring convolution. The output of the

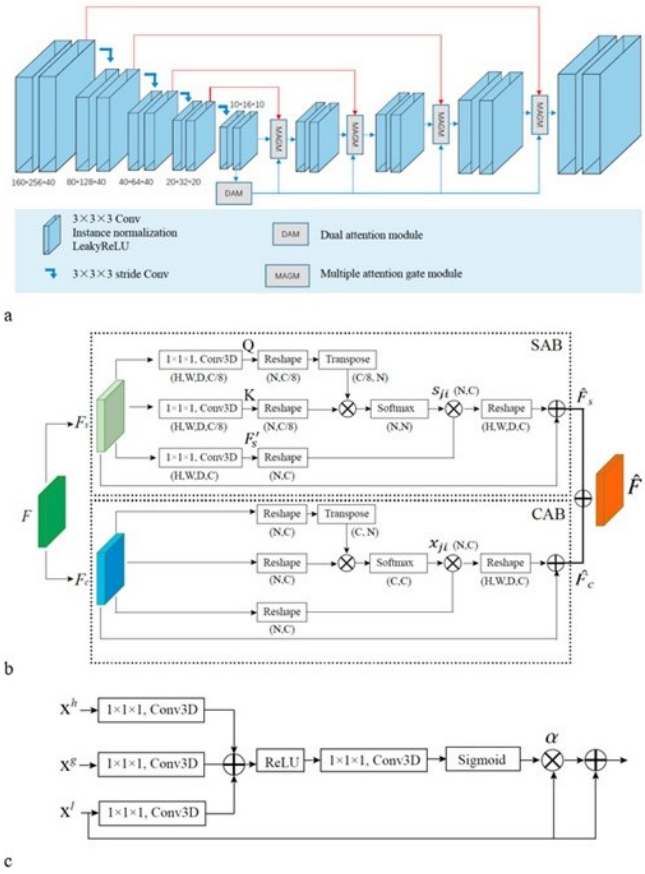

FIGURE 4.6: The network architecture proposed by Wei Chen et al. (2021); **a.** Architecture **b.** DAM **c.** MAGM. [48]

two blocks is combined for the global output. The DAM is added only between the encoder and decoder for computational reason; thus, to alleviate the noise, a Multiple Attention Gated Module (MAGM) is introduced between each layer of the decoder. It combines, after a 1x1x1 convolution, three feature information: the ones corresponding to the decoder layer, the ones corresponding to the encoder layer, and the feature of dual attention. After a ReLU activation function and another 1x1x1 convolution, the output is fed into a Sigmoid function to obtain the final feature maps. The architecture of the DAM and MAGM is shown in Figure 4.6b and Figure 4.6c.

The network was tested with a private dataset composed of 322 patients' images collection with tumor contouring performed by radiologists with software. The dataset was divided with a ratio of 7:1:2 in training, validation, and testing sets.

The results showed an improvement in segmentation accuracy than the base 3D U-Net.

4.3.6 Syeda Furruka Banu et al. (2021)

This study [49] proposed an end-to-end encoder-decoder deep learning approach to accurately segment lung nodules in CT images, called WEU-Net.

The overall architecture is shown in Figure 4.7. It is an encoder-decoder model, where the encoder part is composed by a collection of layers. Each layer includes a weight excitation-based CNN (WE-CNN) [50] and ReLU activation to capture the contextual features from the input image, followed by a Max Pooling layer. The size of the feature map gradually decreases while going deeper. The decoder is a collection of transposed WE-CNN with an increasing size of feature map with the decreasing of the depth. At the final layer, a 1x1 WE-CNN is used to map the final feature vector to the segmentation classes.

This network was tested with the public dataset LIDC-IDRI, composed of CT scans of 888 patients with a total of 1166 CT images with corresponding ground truth masks, split in 922 for the training set and 244 for the testing set.

The results showed that the weight excitation-based CNN significantly improved the performance of U-Net for lung nodule segmentation. Then, the proposed model can also segment the tiny region precisely when the other networks fail.

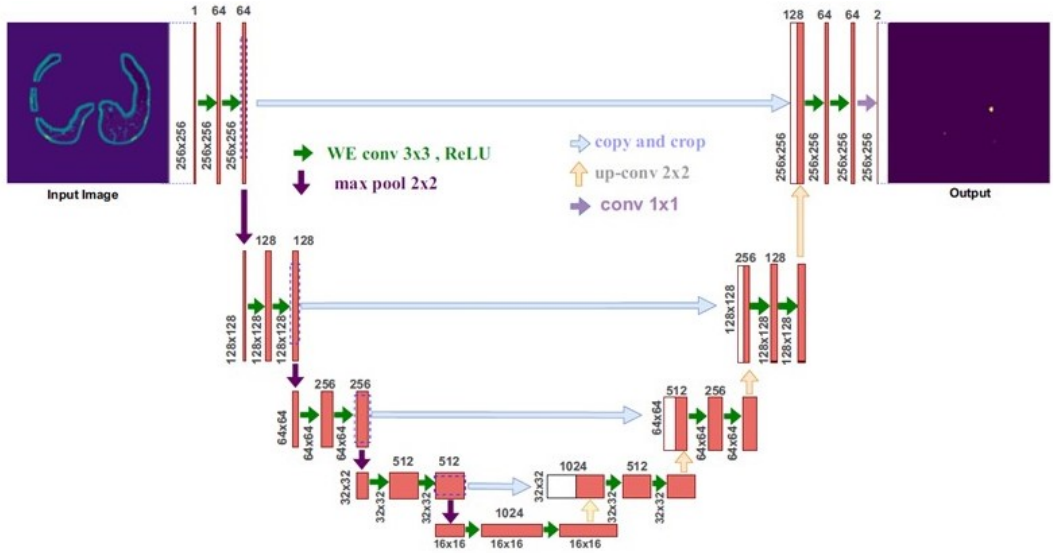

FIGURE 4.7: The network architecture proposed by Syeda Furruka Banu et al. (2021) [49]

4.3.7 Shoji Kido et al. (2022)

In this paper [51], it was proposed a method for robust and accurate three-dimensional segmentation of lung nodule regions using deep learning.

The architecture is composed of a single encoder network that extracts the image features, and it is shown in Figure 4.8a. The outputs include the deepest encoder network (e_1 to e_5), the deepest decoder network (d_5-1 to d_5-4) and the region map created by o_4. The loss function used all the created region maps (o_1 to o_4) to calculate the loss value. The different structures of the layer are shown in the 4.8a legend. The encoder and the decoder are connected by concatenation. The residual unit, Figure 4.8b, has a skip connection where the input does not pass through the convolutional layer.

The network was tested with a private database from Saiseikai Yamaguchi General Hospital, composed of 330 lung nodules. It was split in five parts: four of them, augmented (96 lung nodule images were generated from one lung nodule image) were used as the training set, and the remaining one as the testing set. This process was performed five times for a 5-fold cross-validation.

The results showed that the proposed method is significantly superior to well-known deep learning models for the lung nodule segmentation and significantly superior to conventional image processing methods for the lung nodule detection. This means that it may be useful for accurate and robust segmentation of lung nodules to assist radiologists in the diagnosis.

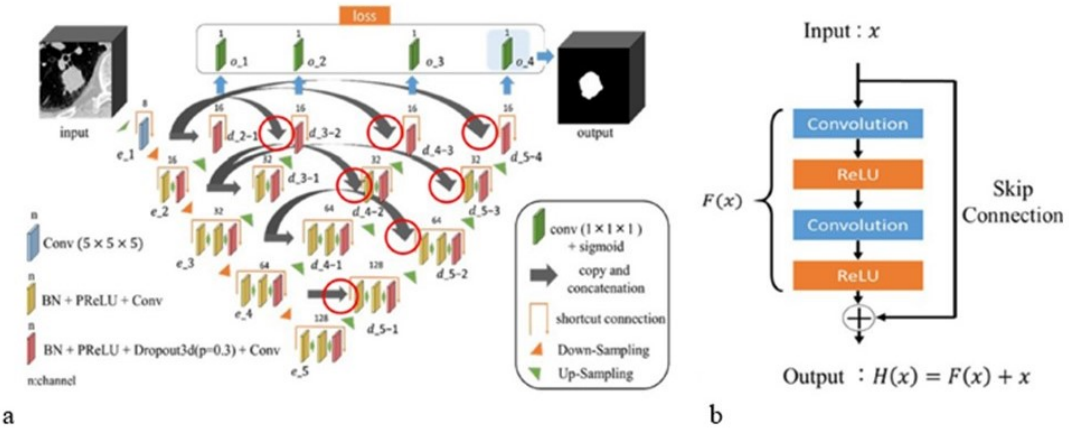

FIGURE 4.8: The network architecture proposed by Shoji Kido et al. (2022); **a.** Architecture; **b.** Residual Unit. [51]

4.3.8 Sundaresan A. Agnes et al. (2022)

In this paper [52], a multiscale fully convolutional three-dimensional U-Net (MF-3D U-Net) was proposed for automatic segmentation of lung nodules in CT images. It fuses a multiscale feature, a Maxout aggregation, and trainable down sampling. To retain the most important features and suppress the low-contribution ones.

The architecture uses four customized encoder blocks along the down sampling path to increase the efficiency of the traditional U-Net. The first convolutional block starts with 16 filters, while the subsequent blocks will double the number of these filters. The overall network is shown in Figure 4.9a, while the inside of each convolutional layer is shown in Figure 4.9b. Each convolutional layer uses a multiscale convolution of kernels, applying two different filter sizes (3x3x3 and 5x5x5), and is followed by BN and a ReLU non-linear activation operation. Then, a Maxout approach is applied to aggregate multiscale features; it joins the obtained features from both 3x3x3 and 5x5x5 convolutions, preserving the highly competitive feature maps. The aggregate maps are then subsampled to half-resolution with a learnable down sampling process performed by a large stride convolution operation. The expansion comprises a stack of deconvolution layers, using the feature

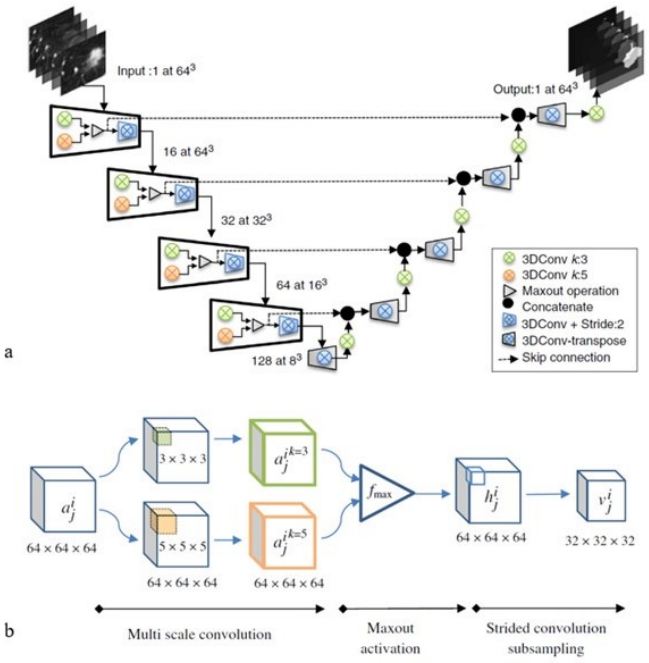

FIGURE 4.9: The network architecture proposed by Sundaresan A. Agnes et al. (2022); **a.** Architecture; **b.** Encoder. [52]

maps produced by the Maxout block through a skip connection. At the end, the Softmax layer computes the classification over the final expanded feature map.

The network was tested with the public dataset LIDC-IDRI, which contains 926 CT scans, added to the 84 CT scans from the previous LIDC dataset. For the training and the testing dataset, 300 CT scans were randomly selected.

The results were divided into a quantitative analysis that showed an accurate segmentation of different types of nodules, including solitary pulmonary and non-solid nodules, and a comparative analysis, which showed good reliability in terms of segmentation with respect to other, widely used, methods.

4.3.9 Yifan Wang et al. (2022)

In this paper [53] it was proposed a hybrid deep learning model (H-DL) for the segmentation of lung nodules with a wide variety of sizes, shapes, margins, and opacities.

The network is based on the redesign of the architectures of the encoders and decoders of a deep convolutional neural network (DCNN) in two separately trained U-shaped networks, combined then into an H-DL model to improve the learning capabilities for segmentation. Each of the two U-DL models consisted of six levels of nine convolution layers in both the contracting and expanding paths. In one it is used a shallow DCNN structure with 16 convolutional layers, while in the other a deep DCNN structure with 200 layers organized in five dense blocks as in a DenseNet. Then, the decoder was reduced to only one convolution layer at each level to reduce computational and memory costs. Both the encoding and the decoding paths used a 3x3 convolution, followed by a ReLU and a 2x2 max pooling with stride 2. It was also added a series of nests and dense skip structures to provide alternative pathways. At the end, the probabilities predicted by the two U-DL networks were combined into the H-DL model with an ensemble layer followed by a 3x3 convolutional layer with the sigmoid activation function. The overall architecture is shown in Figure 4.10.

The network was tested with a public dataset, LIDC, which contains 847 cases, randomly split into 683 for the training and validation sets, and 164 for the testing set.

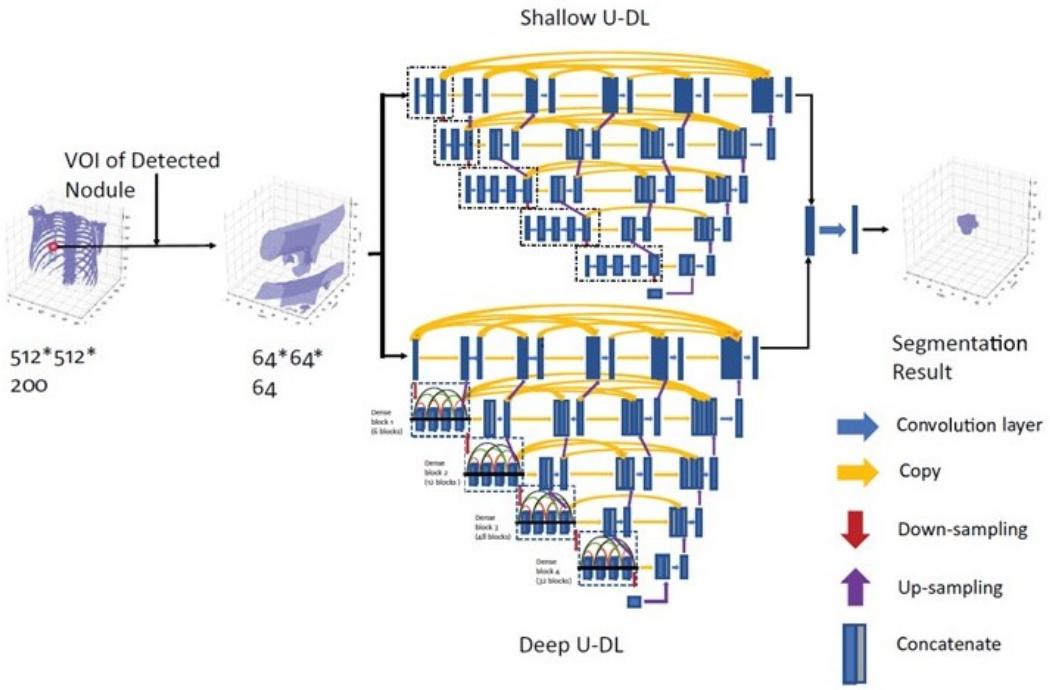

FIGURE 4.10: The network architecture proposed by Yifan Wang et al. (2022). [53]

The results showed that the proposed model outperformed the individual shallow or deep U-DL models and achieved a segmentation accuracy comparable to radiologists' segmentation for nodules.

4.3.10 Dechuan Lu et al. (2022)

In this study [54], a novel network for pulmonary nodule segmentation from CT images based on U-Net was proposed. It uses a dense connection that enhances the transmission and the utilization of the features and mitigates the class imbalance problem due to the small size of pulmonary nodules, and a new loss function, which is tolerance on the pixels near the borders of the nodule.

The architecture is a Dense U-Net, a new network that uses dense connections to combine the features of the current layer with those of all previous layers and transmit them to all the subsequent ones. Each dense connection has two 3x3 convolution layers, two BN, two ReLU, and two feature fusion operations, which fuse the generated feature graph with the original one, becoming the input of the next dense connection. This is shown in Figure 4.11a, while in Figure 4.11b is the whole architecture, composed of an encoder, a decoder, a classifier, and a skip connection. The encoder contains 4 dense

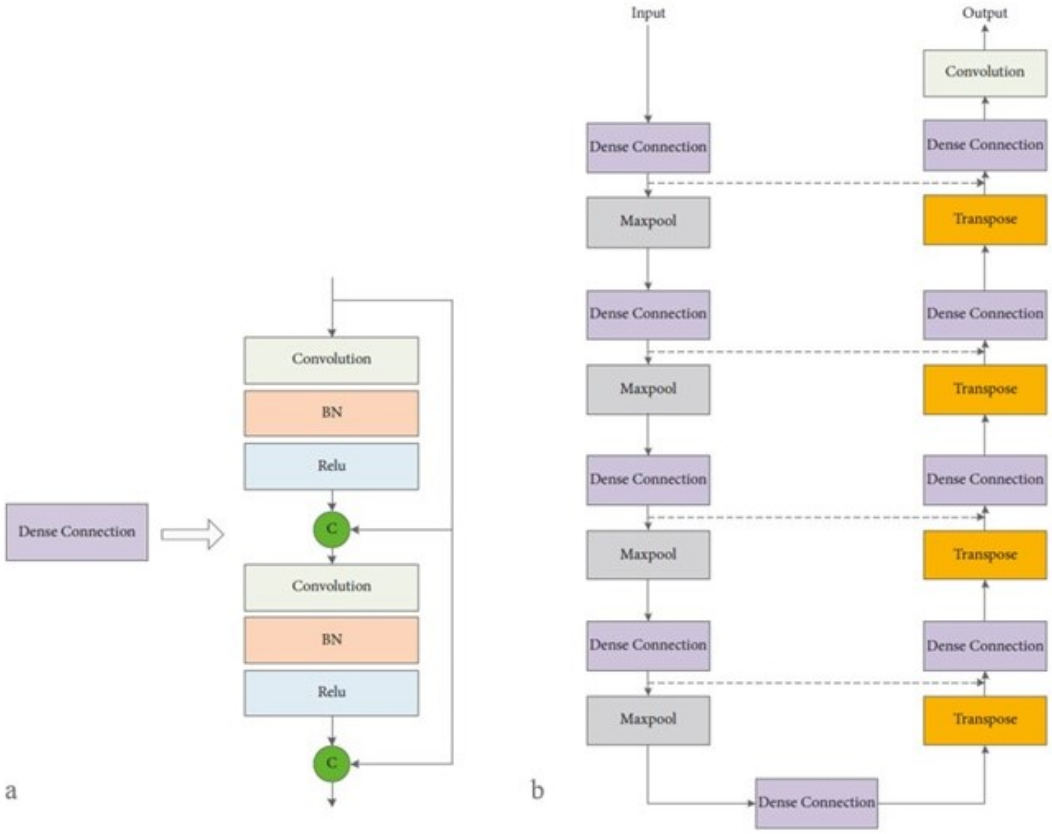

FIGURE 4.11: The network architecture proposed by Dechuan Lu et al. (2022); **a.** Dense Connectio structure **b.** Architecture. [54]

connection, and a maximum pooling layer, generating a 4x4 feature map. The decoder is composed of a dense connection and a deconvolution layer, which generates the final feature map with the same size as the input image. The encoder and the decoder are linked by a dense connection. The classifier is a 1x1 convolutional layer and a sigmoid activation layer. The loss function introduced is tolerant on the pixels near the borders of the nodule, measuring the loss between one pixel and all the pixels around the corresponding one in the ground truth and selecting the minimum value.

The network was tested with the public dataset LIDC-IDRI. 4000 images were used as the training set, 500 as the validation set, and 200 as the testing set. The test set was taken from another dataset from Jiangdu People's Hospital.

The results say that the Dense U-Net has a stronger learning ability for the small or fuzzy boundary pulmonary nodules, can alleviate the gradient disappearance problems, and obtains more accurate segmentation results than the other networks.

4.3.11 Chandra Sekhara Rao Annavarapu et al. (2023)

This study [55] proposed an end-to-end deep learning approach for lung nodule segmentation, which incorporates a bidirectional feature network (Bi-FPN) between an encoder and a decoder architecture. It introduced a Mish activation function and class weights of masks to enhance the efficiency of the segmentation.

The architecture is an encoder-decoder U-Net backbone combined with a Bi-FPN; thus, it is composed of three sections: contraction, Bi-FPN, and expansion. The first section applies two 3x3 convolutions followed by a non-linear Mish activation function, proposed by [56], to perform a strong regularization and a 2x2 max pooling. At the end, a dropout layer performs the regularization and feeds the second section, the Bi-FPN. This one is based on conventional top-down Feature Pyramid Networks (FPN) [57], which fuse features at different resolutions to obtain an efficient feature extraction, thanks to bidirectional cross-scale connections and weighted feature fusion. For improving efficiency, separable convolution, followed by a BN and a ReLU, was implemented. The output of the Bi-FPN is fed into the expansion section, where each step consists of a 2x2 up-convolution followed by two 3x3 convolutions and a Mish activation function. At the end, a 1x1 convolution block and a Sigmoid activation function were applied. The architecture is shown in Figure 4.12.

The network was tested on the public dataset LUNA-16, derived from the LIDC-IDRI dataset. After a pre-processing phase, the dataset was composed of 1166 CT images, divided into 922 for the training set and 244 for the testing set, applying a K-fold cross-validation of 4-folds.

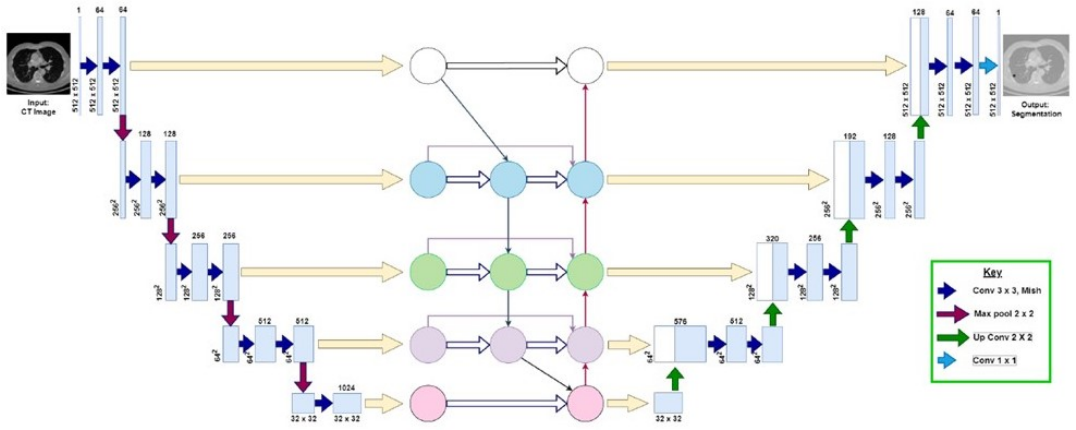

FIGURE 4.12: The network architecture proposed by Chandra Sekhara Rao Annavarapu et al. (2023). [55]

The results showed that the proposed architecture outperformed existing deep learning models as U-Net, demonstrating precision in segmentation of lung nodules and also achieving efficient performance of segmentation in daunting cases such as cavitary nodules and small nodules of less than 6 mm.

4.3.12 Junyoung Park et al. (2023)

This paper [58] proposed a two stage U-Net architecture to enhance the performance of lung cancer segmentation using PET/CT images.

The model architecture is composed of two stages:

- Stage 1, shown in Figure 4.13a, is a global 3D U-Net that receives the images as inputs and extracts the preliminary tumor area, generating a 3D binary volume as output. Each convolution block has a 3x3x3 convolution layer, a BN, and a leaky ReLU with a negative slop of 0.2 as the activation function, with a 2x2x2 max pooling after.
- Stage 2, shown in Figure 4.13b, is a regional U-Net based on DenseNET [59]. It receives as input eight consecutive slices centered on the one predicted to have lung cancer in stage 1 and generates a 2D binary image as output. The max pooling is a two-dimensional 2x2.

The network was tested with a private dataset composed of 887 sample images, divided into 730 for the training set, 81 for the validation set, and 76 for the testing set. The ground-truth volume of interests was drawn semi-automatically.

The results showed that this network outperformed the one-stage 3D U-Net in segmentation of primary lung cancer and predicted correctly the detailed margin of the tumors. The quantitative analysis confirmed the advantages of using a two-stage U-Net: the proposed method reduces the time and the effort required for lung cancer segmentation in PET/CT images.

4.3.13 T. Weikert et al. (2023)

In this paper, [60] proposed a Retina U-Net algorithm for the detection of primary lung tumors and associated metastasis stages on PET/CT images.

The architecture is based on a Retina U-Net, a state-of-the art approach in medical detection [61], characterized by additional branches in the lower decoder levels for end-to-end object classification (CL) and bounding box

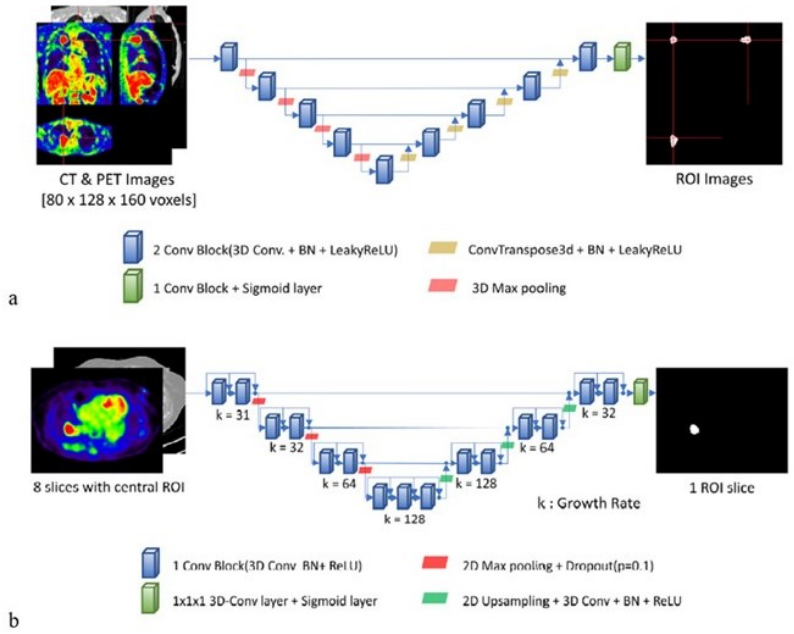

FIGURE 4.13: The network architecture proposed by Junyoung Park et al. (2023); **a.** Global Architecture; **b.** Regional Architecture. [58]

regression (BB). The network is an encoder-decoder structure that resembles a U-Net, complemented by CL and BB at the lower levels of the architecture to exploit object level features. It is shown in Figure 4.14.

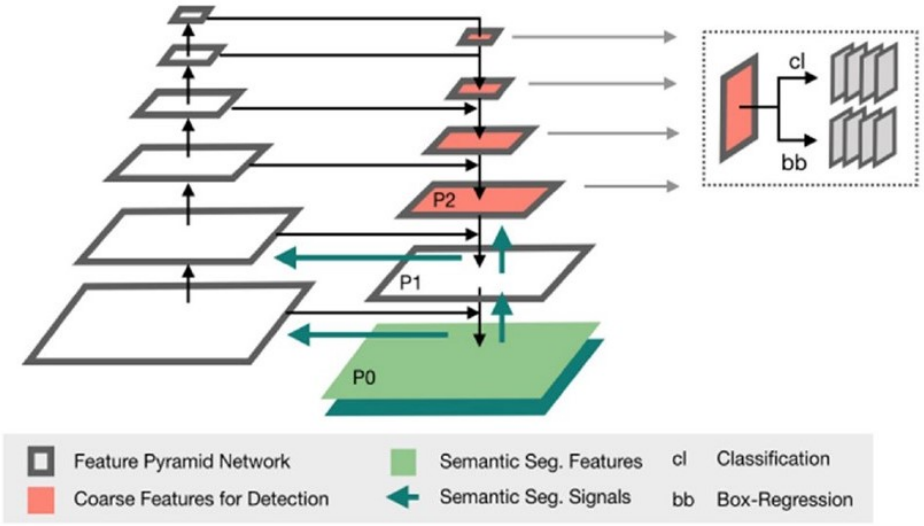

FIGURE 4.14: The network architecture proposed by T. Weikert et al. (2023). [60]

The network was tested with two private datasets with the images acquired at the authors institution. The first is called the internal dataset, composed of 364 samples and divided into the testing set (216), the validation set (74) and the testing set (74). The second is the external dataset composed of 20 samples as an external testing set.

The results revealed a good detection rate of lesions, with higher performance on larger tumors with respect to smaller ones, due to the fact that the small are affected by the partial volume effect.

4.3.14 Tenzin Kunkyab et al. (2024)

In this paper [62], it was proposed a novel deep learning architecture based on CNN, residual blocks, and transformers (Co-ReTR) to auto-segment the gross tumor volume (GTV) in CT images. This combination was introduced to overcome the CNN limitations in learning long-range spatial dependencies due to the locality of the convolutional layer.

The architecture, shown in Figure 4.15a, has three key components:

- *Encoder*. It is composed of two CNNs, a deep 3D network and a shallow 3D network, and its main function is extracting features. The low-resolution images are processed by the deep CNN, while the high-resolution images are processed by the shallow one to reduce the computational complexity. The shallow CNN has five 3D convolutional layers (Conv), interleaved with instance normalization (IN) and ReLU, complemented by six stages of 3D residual blocks, which incorporate Conv-IN-ReLU units. The output of the first onvolutional layer undergoes two convolution filters, 1x1x1 and 3x3x3. The deep CNN is similar, but with nine residual blocks. This structure is shown in Figure 4.15b.

- *Transformer*. It is composed by an inputs-to-sequence layer and a series of stacked deformable transformer layers (DeTrans), and it must capture and model the long-range contextual information. The positional encoding reintroduces the spatial information into the flattened sequence, while the DeTrans enhances the representation learning process by incorporating a combination of deformable self-attention (scan the feature maps to find all the possible locations around a reference one), a feed-forward network (introduce non-linearities and let the model catch more complex data relationships), and a layer normalization (standardize the activations within layers). Each DeTrans has a skip connection, facilitating the challenges with gradient vanishing.

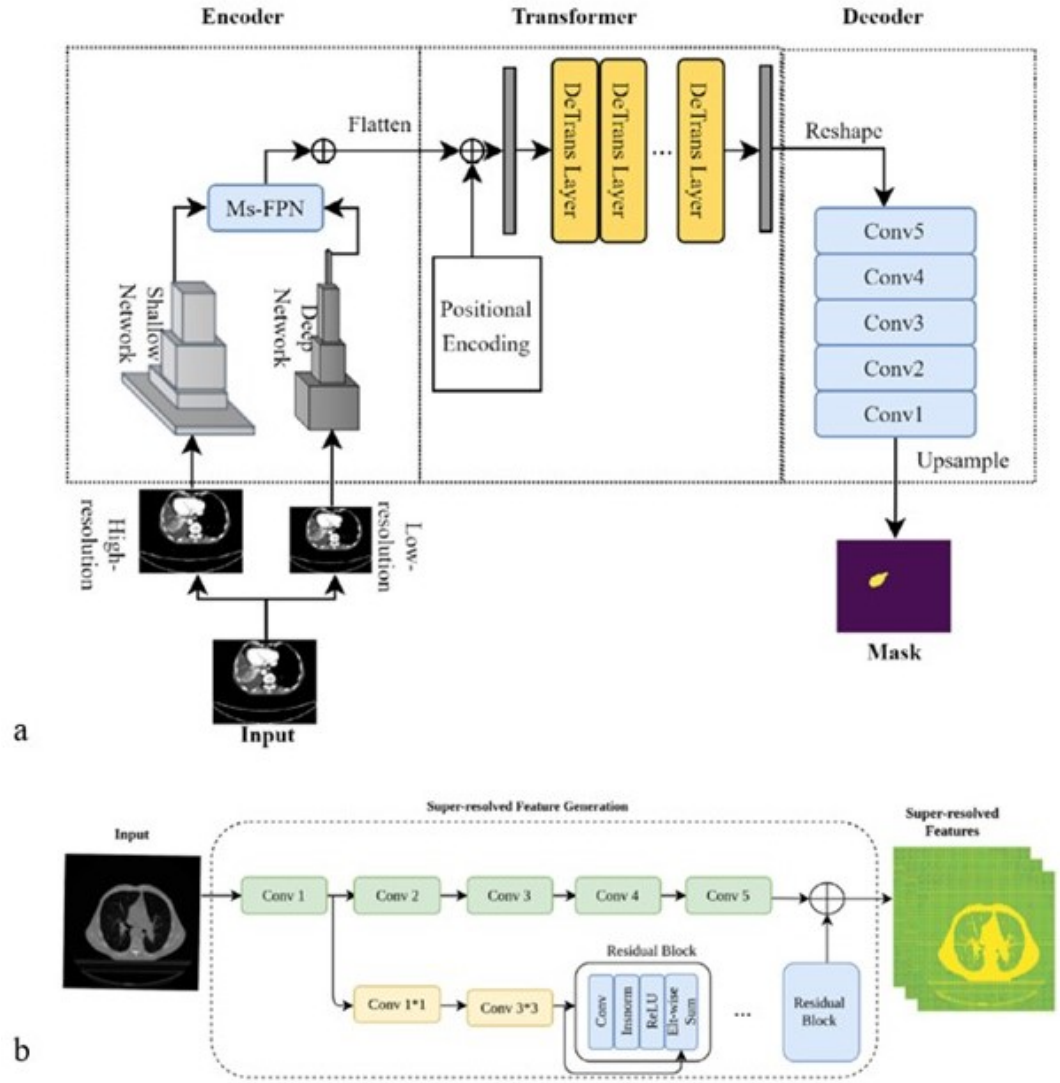

FIGURE 4.15: The network architecture proposed by Tenzin Kunkyab et al. (2024); **a.** Architecture; **b.** Shallow and deep CNN. [62]

- *Decoder*. It is a CNN block to upsample the feature maps back to the original image resolution. Through deconvolution. Then it uses residual blocks to capture details and enhance the segmentation quality output.

The architecture was tested with 676 CT images, taken from three public datasets: NSCLC radiomics, NSCLC radio genomics, and the paper authors clinical database. The first two, composed of 563 samples, were used as training and validation sets; instead, the third one, 113 samples, was used as the test set.

The results were compared against five other auto-segmentation techniques: U-Net, Att-U-Net, ResNet-U-Net, CoTr, and U-NetR, showing a higher segmentation accuracy by capturing local and global information. Other results were conducted to evaluate the ability of the proposed network to contour both single and multiple tumors, still resulting in a higher accuracy than the other five networks.

4.3.15 Fuli Zhang et al. (2024)

This study [63] proposed a unique two-stage deep learning method for lung cancer segmentation. The first stage is a coarse network of segmentation to detect the harsh region of lesions, while the second stage involves two distinct segmentation networks, trained separately, for two categories of images containing large-sized and small-sized tumors.

The architecture, as already said, is divided into two steps:

- *Coarse segmentation network*. It uses a U-Net with four convolution and down sampling operations and four up sampling and convolution operations to obtain the segmentation outcomes. Each convolution layer exhibits the same number of convolutions with an equally sized kernel. To well connect the output obtained to the second step, a processing phase is necessary, and it is shown in Figure 4.16a.

- *Fine segmentation network*, Figure 4.16b. It is used a TransU-Net, which combines transformers and U-Net. The transformer encodes tokenized image patches for capturing global context, while, simultaneously, the decoder carries the up sampling of the encoded features, then fused with the high-resolution CNN feature maps. To enhance the feature extraction capability, residual convolution blocks were used. In particular, a great number of these blocks and convolution channels were employed in the initial layer, with a decreased number for further layers.

The network was tested with the public dataset NSCLC, considering 200 cases for the training set, 40 cases for the validation set, and 60 cases for the testing set.

The quantitative results showed that the proposed method has a higher accuracy and efficiency in tumor segmentation than the other methods, which have under and over segmentation issues. Then, the model demonstrated

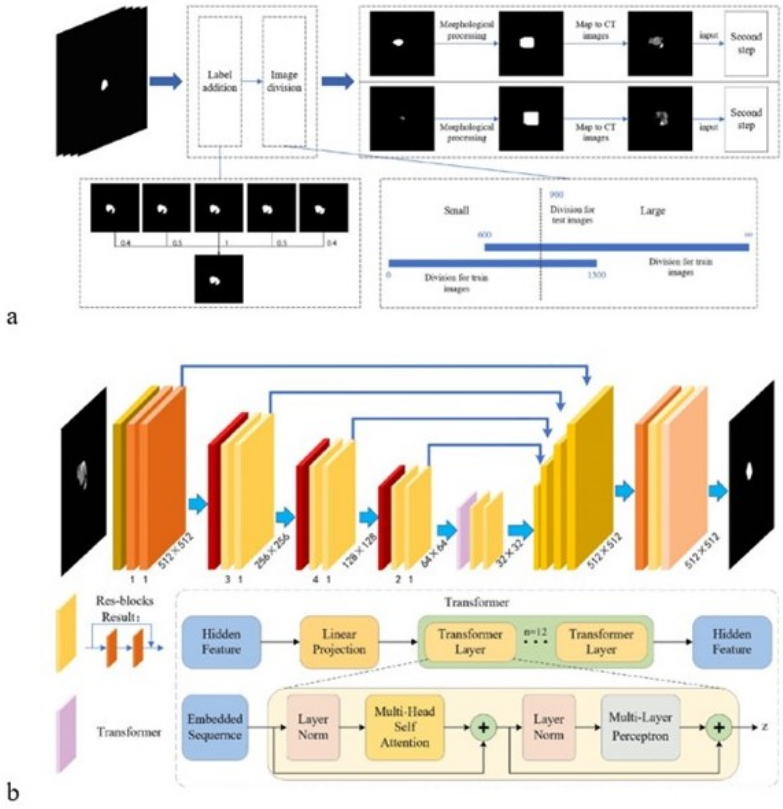

FIGURE 4.16: The network architecture proposed by Fuli Zhang et al. (2024); **a.** Processing of the first step's results.; **b.** Fine segmentation network. [63]

also good performance in CT images with small GTVs, which have poor performance in other methods, but generally, thanks to the training in two different groups of images, the network has exhibit significance enhancement in segmentation on both types of CT images.

4.3.16 Comparison Tables

Table 4.1, Table 4.2, Table 4.3, Table 4.4 summarize both the general and the specific information about each article resulted from this literature review.

It is possible to notice that most of the studies used the same databases but different preprocessing techniques, even if some of them are the same. The most used ones are: rescaling, resizing, and cropping the images and centering them to the nodule position; and the intensity range resizing according to the Hounsfield scale (HU), followed by a normalization.

As expected, each paper added different characteristics and features to the base network models to study the possible upgrade and improvements for the state-of-the-art model already discussed (FCN, U-Net, and 3D U-Net). Some studies started even from the simple CNN.

As it is possible to notice, each paper bases the evaluation of the performances of the proposed networks on the use of different metrics, but all of them except [60] used the Dice Similarity Coefficient (DSC). All the other evaluation metrics are not used by more than three studies. Table 4.5 shows the comparison of the evaluation metrics results obtained from the proposed articles. For proper knowledge, all the metrics are reported, even if the DSC is the best and the only one useful for the comparison.
List of the Evaluation Matrix:

- Dice Similarity Coefficient (DSC);
- Symmetric Average Surface Distance (ASD);
- Sensitivity (SEN);
- Positive Predictive Value (PPV);
- Inference Time (IT);
- Hausdorff Distance (HD);
- Relative Absolute Volume Difference (RAVD);
- Accuracy (ACC);
- Intersection over Union (IoU);
- Precision (PRE);
- False Positive Rate (FPR);
- True Positive Rate (TPR);

TABLE 4.1: Comparison table among the literature studies. [1/4]

Ref	Title	Author	Year	Dataset	Preprocessing	Base model	Advanced feature	Evaluation metrics
[43]	Central focused convolutional neural networks: Developing a datadriven model for lung nodule segmentation	Shuo Wang, Mu Zhou, Zaiyi Liu, Zhenyu Liu, Dongsheng Gu, Yali Zang, Di Dong, Olivier Gevaert, Jie Tian	2017	LIDC, Guangdong General Hospital dataset	Rescaling	CNN	Two CNN branches. Kernel size varying	DSC, ASD, SEN, PPV
[44]	3D Fully Convolutional Networks for Co-Segmentation Of Tumors On Pet-Ct Images	Zisha Zhong, Yusung im, Laixin Zhou, Kristin Plichta, Bryan ALLen, John Buatti, Xiaodong Wu	2018	private dataset	Resampling, Cropping centered on the tumors, Intensity Thresholding, Data augmentation	3D U-Net	Two 3D U-Net for PET/CT co-segmentation	DSC
[46]	3D Densely Connected Convolution Neural Networks for Pulmonary Parenchyma Segmentation from CT Images	Liang Zhao	2020	LIDC-IDRI	Image resizing, Gray value standardization	FCN	3 densely connected block	DSC
[47]	Accelerating 3D Medical Image Segmentation by Adaptive Small-Scale Target Localization	Boris Shirokikh, Alexey Shevtsov, Alezandra Dalechina, Egor Krivov, Valery Kostjuchenko,Andrey Golanov, Victor Gombolevskiy, Sergey Morozov, Mikail Belyaev	2021	LIDC-ICRI	Intesity ranging (-1000 and 300 Hounsfield units (HU)), Averaging , Normalization	3D U-Net	Low-resolution segmentation followed by a detailed segmentation	IT, REC, DSC

TABLE 4.2: Comparison table among the literature studies. [2/4]

Ref	Title	Author	Year	Dataset	Preprocessing	Base model	Advanced feature	Evaluation metrics
[48]	MAU-Net: Multiple Attention 3D U-Net for Lung Cancer Segmentation on CT Images	Wei Chen, Fengchang Yang, Xianru Zhang, Xin Xu, Xu Qiao	2021	Private dataset	Not reported	U-Net	DAM and MAGM	HD, RAVD
[49]	WEU-Net: A Weight Excitation U-Net for Lung Nodule Segmentation	Syeda Furruka Banu, Md. Mostafa Kamal Sarker, Mohamed Abdel-Nasser, Hatem A. Rashwan, Domenec Puig	2021	LIDC-ICRI	Cleaning (not specified) and Data augmentation	U-Net	Weight excitation based CNN	ACC, IoU, PRE, REC, DSC
[51]	Segmentation of Lung Nodules on CT Images Using a Nested Three-Dimensional Fully Connected Convolutional Network	Shoji Kido, Shunske Kidera, Yasushi Hirano, Shingo Mabu, Tohru Kamiya, Nobuyuki Tanaka, Yuki Suzuki, Masahiro Yanagawa, Noriyuki Tomiyama	2022	Private database from Saiseikai Yamaguchi General Hospital	Cropping	FCN	Encoder and decoder connected by concatenation	DSC, IoU
[52]	Efficient multiscale fully convolutional U-Net model for segmentation of 3D lung nodule from CT image	Sundaresan A. Agnes, Jeevanayagam Anitha	2022	LIDC-ICRI	Cropping, Intesity ranging (-1000 and 1000 HU) followed by a normalization	3D U-Net	Customized encoder, Multiscale convolution of kernels, Maxout aggregate layer, Subsampling with strided convolution	IoU, DSC

TABLE 4.3: Comparison table among the literature studies. [3/4]

Ref	Title	Author	Year	Dataset	Preprocessing	Base model	Advanced feature	Evaluation metrics
[53]	Hybrid U-Net-based deep learning model for volume segmentation of lung nodules in CT images	Yifan Wang, Chung Zhou, Heang-Ping Chan, Lubomir M. Hadjiiski, Aamer Chughtai, Ella A. Kazerooni	2022	LIDC-ICRI	HU scaling, resampling, Cropping centered to the tumors	CNN	Encoder and decoder two separate U-Net then combined	DSC, IoU
[54]	A Novel Deep Learning Network and Its Application for Pulmonary Nodule Segmentation	Dechuan Lu, Junfeng Chu, Rongrong Zhao, Yuanpeng Zhang, Guangyu Tian	2022	LIDC-ICRI	Clipping the image size	U-Net	Dense connection to transfer and utilize features, New loss function tolerant to the border pixels	PRE, REC, DSC
[55]	A Bi-FPN-Based Encoder–Decoder Model for Lung Nodule Image Segmentation	Chandra Sekhara Rao Annavarapu, Samson Anosh Babu Parisapogu, Nikhil Varma Keetha, Praveen Kumur Donta, Gurindapalli Rajita	2023	LUNA-16 (LIDC-ICRI)	Cropping and Data augmentation	U-Net	Bidirectional feature network, Mish activation function	DSC, SEN, PPV
[58]	Automatic Lung Cancer Segmentation in [18F]FDG PET/CT Using a Two-Stage Deep Learning Approach	Junyoung Park, Seung Kwan Kang, Donghwi Hwang, Hongyoon Choi, Seunggyun Ha, Jong Mo Seo, Jae Seon Eo, Jae Sung Lee	2023	Private dataset	Image resizing	U-Net	Two stage U-Net: Global and Regional	DSC

TABLE 4.4: Comparison table among the literature studies. [4/4]

Ref	Title	Author	Year	Dataset	Preprocessing	Base model	Advanced feature	Evaluation metrics
[60]	Automated lung cancer assessment on 18F-PET/CT using Retina U-Net and anatomical region segmentation	T. Weikert, P. F. Jaeger, S. Yang, M. Baumgartner, H. C. Breit, D. J. Winkel, G. Sommer, B. Stieltjes, W. Thaiss, J. Bremerich, K. H. Maier-Hein, A. W. Sauter	2023	Private dataset	Cropping, Intesity ranging (-1000 and 1000 HU) followed by a normalization between 0 and 1	U-Net	Retina U-Net	SEN
[62]	A deep learning-based framework (Co-ReTr) for auto-segmentation of non-small cell-lung cancer in computed tomography images	Tenzin Kunkyab, Zhila Bahrami, Heqing Zhang, Zheng Liu, Derek Hyde	2024	NSCLC radiomics, NSCLC radio genomics and the paper authors clinical database	Intesity ranging (-1024 and 3068 HU) followed by a min-max normalization	CNN	CNN combined with residul Blocks and Transformes (CO-ReTR)	HD, IT, DSC
[63]	Enhancing non-small cell lung cancer tumor segmentation with a novel two-step deep learning approach	Fuli Zhang, Qiusheng Wang, Enyu Fan, Na Lu, Diandian Chen, Huayong Jiang, Yanjun Yu	2024	NSCLC	Not reported	CNN	Two steps: coarse network segmentation, and fine network segmentation	DSC, HD, FPR, TPR

TABLE 4.5: Comparison of the evaluation metrics (mean) among the literature studies.

Ref	**DSC**	**ASD**(mm)	**SEN**(%)	**PPV**(%)	**IT**(s)	**HD (mm)**	**RAVD**	**ACC**(%)	**IoU**	**PRE**(%)	**FPR**	**TPR**
[43]	0.8215	0.1700	92.7500	75.8400								
[44]	0.8690											
[46]	0.9720											
[47]	0.7500		80.0000		13.0000							
[48]	0.8667					13.0036	0.1552					
[49]	0.8283		86.5300					99.9900	0.7055	81.6600		
[51]	0.8450								0.7380			
[52]	0.8300								0.7200			
[53]	0.7500								0.6170			
[54]	0.7442		72.5400							75.5100		
[55]	0.8282		92.2400	78.9200								
[58]	0.7800											
[60]			88.8000									
[62]	0.9200				12.0300	1.3330						
[63]	0.8000					9.4300					0.0005	0.8000

Chapter 5

Deep learning lung cancer detection

5.1 Introduction

Early diagnosis of the lung cancer is important in making positive changes to the patient's condition and in deciding on an appropriate treatment strategy. The current standards of oncology ensure that an accurate diagnosis largely depends upon the availability of good-quality medical images such as CT scans, which give clear visual details of abnormalities in lung tissues. Clinicians must deal with the reading and interpretation of complicated imaging data, where factors like anatomical structures overlaying the region of interest and poor image quality may often obscure the identification of malignant lesions. Various traditional methods of image processing have been considered for the detection of lung cancer; however, these usually lack sufficient sensitivity and specificity that would enable the reliable diagnosis of lung cancer. The recently considered deep learning algorithms, especially the ones using CNNs in the analysis of medical images, have shown high performance in the segmentation and classification of lung cancer-related lesions. This chapter will present a proposed deep learning model, based on a 2D U-Net architecture, for the segmentation of lung cancer in CT images, with a subsequent 3D reconstruction of the cancer volume.

5.2 Materials and methods

The model and its evaluation has been implemented in Python Code on the Google Colab environment, using the L4 GPU runtime with high RAM.

5.2.1 Dataset

The data used for this project comes from the NSCLC-Radiomics dataset available online on the "The Cancer Imaging Archive" database. The dataset contains images from 422 NSCLC patients. For each patient, pretreatment CT scans and manual delineation by a radiation oncologist of the 3D volume of the gross tumor volume and of the lungs are available. Also, clinical outcome data and other organ manual delineation (heart, exophages, and spinal cord) are present but not considered for the purpose of this study. In Figure 5.1, an example of the data present in the chosen dataset is shown.

5.2.2 Preprocessing

The above dataset was preprocessed to prepare it for the training and evaluation of the model. The entire preprocessing pipeline and an example of the resulting preprocessed images are shown in Figure 5.2. The dataset consists of 422 subject CT DICOM files along with their related segmentation DICOM file, except for one subject. In the remaining 421 segmentation files, 11 lacked lung segmentation masks. First, those CT files from DICOM format were converted to NIfTI format to make the imaging data easy to manipulate and analyze. 19 failed due to lack of missing slices. The subsequent steps were

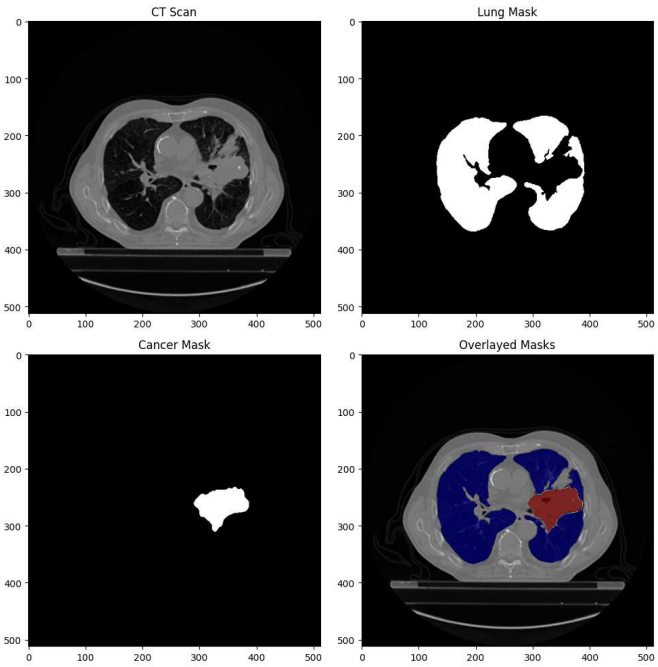

FIGURE 5.1: Example of data presents into the NSCLC-Radiomics dataset

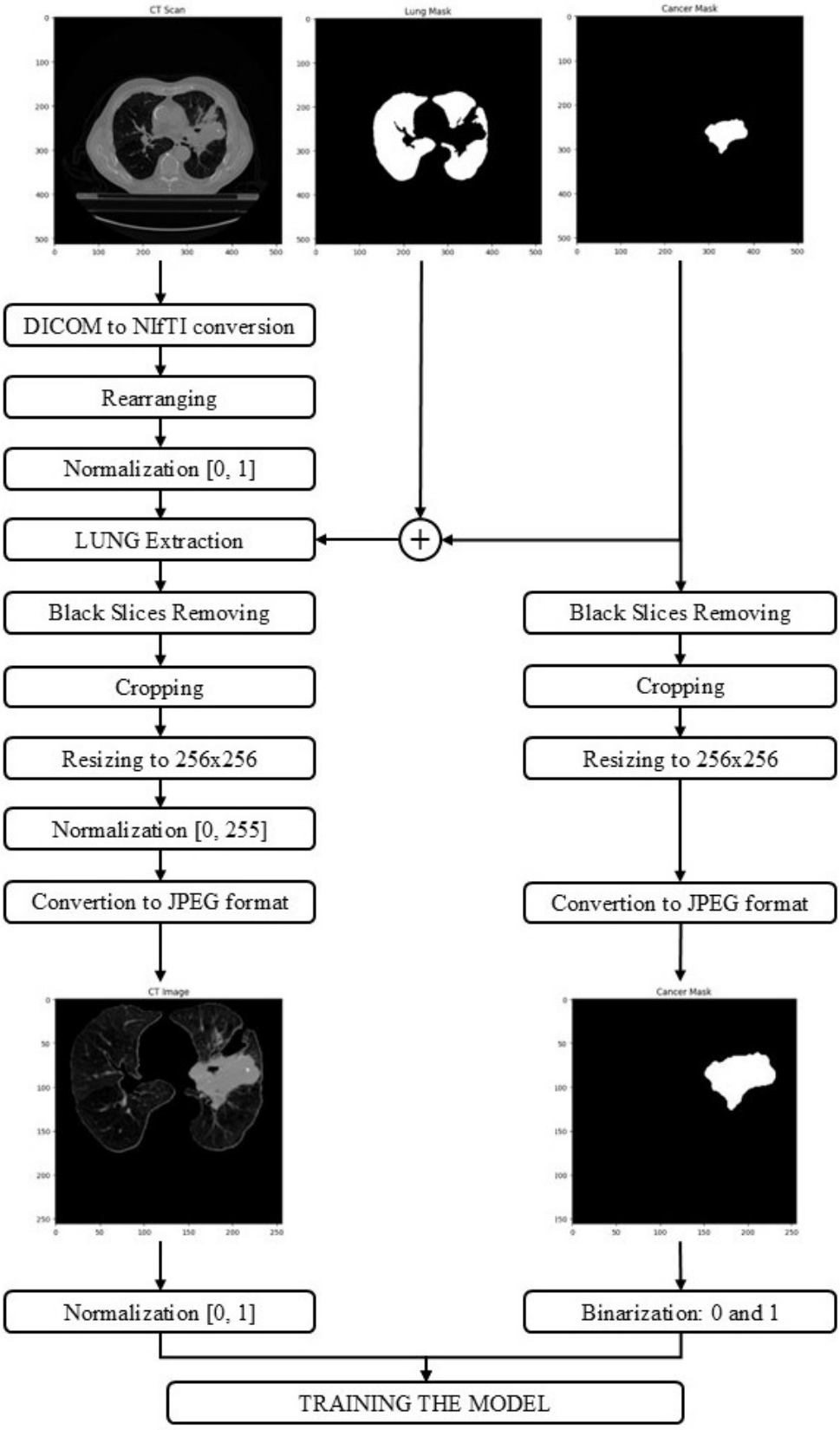

FIGURE 5.2: Preprocessing pipeline

the re-arrangement, so that their dimensions were consistent in the depth, width, and height orientations of all data entries, and the grayscale pixel intensity values windowing of the CT images in the range of -1000 and +1000 Hounsfield Unit (HU), which is good to intensify the voxels within the lung region [52]. Further, the data was normalized into the range of 0 to 1, enhancing suitability for further analysis. Then, the segmentation files were analyzed and only those segments that corresponded to lung and cancerous regions were extracted. These segmentation files were used to extract the lung from the CT slices with a targeted analysis that removed all the pixels of the slices that did not contain lung or cancer related information. Then, completely black slices without informative data were removed to further refine the data. Slices from each subject were cropped to create an optimal square window, which minimized the number of non-informative black pixels in the final images. To standardize the input dimensions, all slices were resized to 256x256 pixels, through an interpolation algorythm. After this, however, a renormalization between 0 and 255 was necessary in order not to introduce possible changes due to the resizing algorithm, important for the final conversion to JPEG format. The very last steps were the final normalization between 0 and 1 of the CT images and the binarization of the ground truth cancer masks to be used in the training and testing processes. After these preprocessing steps, the dataset was finally reduced to 391 scans.

5.2.3 Model Architecture

The proposed U-Net, Figure 5.3, consists of an encoder-decoder architecture that is able to capture the contextual information while maintaining spatial accuracy for the execution of precise segmentation tasks.

The encoder path is made up of five blocks. Each block comprises two convolutional layers, with LReLU as the activation function, followed by BN. Remarkably, LReLU allows a small gradient when the unit is assumed to be inactive; this aids in training the model by preventing the vanishing gradient problem that generally arises in deeper networks. All convolutional blocks employ a kernel size of (3,3) with "same" padding to preserve the spatial dimensions of the feature maps. The BN of the second convolution of each block is followed by a spatial dropout (SD), set experimentally at 0.2 after trials, through which a fraction of neurons is dropped out temporarily while training to help reduce overfitting and improve generalization. The

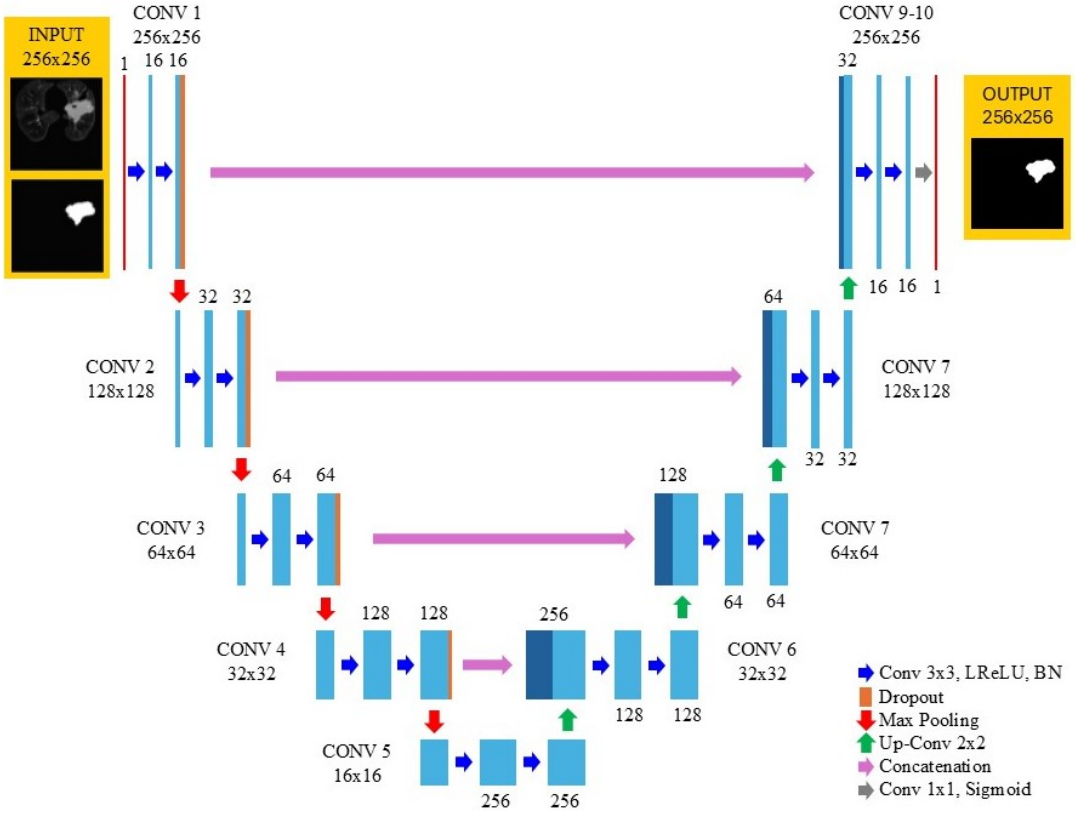

FIGURE 5.3: Architecture of the proposed U-Net

MaxPooling layers are added to the end of the convolutional blocks to further downsample the feature maps, reducing the spatial dimensions by half, from 256x256 to 16x16. This contraction helps to model higher states of abstraction since the network will be made deeper. This forms the bottleneck layer, where the size of the feature maps is the smallest, and from which the Grad-CAM will be computed. The bottleneck uses 2D convolution layers with a greater filter size of 256, thus increasing the model's capacity to learn more complex patterns associated with lung cancer. Next, the transposed convolutions mark the beginning of the expansive path to upsample the feature maps, concatenating them with the corresponding encoder outputs using the skip connections. In this manner, fine-grained spatial information that may be lost through downsampling is preserved, which is pretty important for high-resolution segmentation. The final output layer is a 2D convolution with kernel size (1, 1) followed by a sigmoid activation function. Those multi-channel feature maps are converted to a single-channel output map representing the presence of the tumor as a binary mask.

5.2.4 Training

After preprocessing, the dataset was divided into three sets: training, validation, and testing datasets. Firstly, the dataset was split into the training/validation set, which contains 80% of the subjects (312) that were used to train the implemented model, and the testing set, which contains 20% of the subjects (79), used to evaluate it. Lastly, the initial training/validation set was split again into a training set, which contains 80% of those subjects (249), and a validation set, which contains 20% of those subjects (63), used to determine when to stop the training process to avoid overfitting and guaranteeing generalization. In conclusion, the training set includes 20961 2D images, the validation 5271, and the testing 6679. The data splitting described above is shown in Figure 5.4.

A data generator was implemented with a batch size of 128 and a random shuffle of the training and validation sets. The described choice of the batch size allows finding the optimal balance between computational efficiency and stability of convergence.

The optimizer used in this work is Adam with a set learning rate of 0.001, with a learning rate scheduler that will reduce the learning rate with a 0.5 factor in case of stalls in progress. All choices were made experimentally. Besides that, the early stopping concerning the best validation loss and the restoration of the best weights are implemented to further enhance the training. All that let the model not overfit and to learn as much as possible from the given data.

Dice Coefficient loss was chosen as loss function. It is based on the computation of the Dice Coefficient (DC), used to evaluate the similarity of two

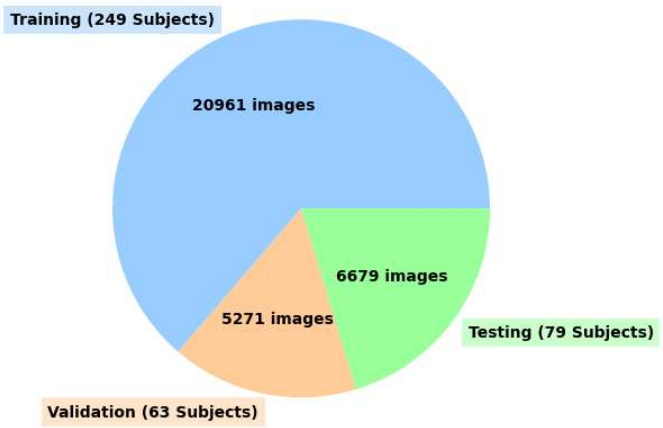

FIGURE 5.4: Data splitting

sets. It is mathematically defined by Eq. 5.1:

$$DC = \frac{2|A \cap B|}{|A| + |B|} \tag{5.1}$$

where A is the set of pixels in the predicted segmentation mask, B the set of pixels in the ground-truth mask. Here, $|A \cap B|$ is the number of pixels correctly predicted as positive, i.e., tumor pixels; $|A|$ and $|B|$ are the total number of pixels in the predicted and ground truth masks, respectively. For implementation of this in a loss function format suitable for training, we take the complement of the Dice Coefficient as represented by Eq. 5.2:

$$LossDC = 1 - DC \tag{5.2}$$

This provides a measure of similarity between two objects based on determining a ratio between the overlap versus the sum of all pixels in two masks. A Dice Coefficient of 1 reflects perfect agreement between the predicted and true segmentation, while a score of 0 indicates no overlap. By using the Dice Coefficient as a loss function, we concentrate the training process on maximizing the overlap of the predicted segmentation and actual tumor pixels.

The Dice Coefficient loss is good segmentation tasks, especially within medical imaging, for its sensitivity for small and irregular structures, its robustness to class imbalance, where background pixels outnumber the tumor pixels by a large margin, and its gradient behavior, smoother and more informative compared to the standard losses, which helps in smooth training and better convergence in hard segmentation tasks.

5.2.5 Testing and Evaluation Procedures

After the preprocessing and the following data splitting, the U-Net was trained and then tested. The procedure of the testing followed two routes. The first one is the simplest, consisting in the simple testing and the subsequent evaluation of the results in two different approaches: 2D analysis and 2.5 analysis for the final 3D evaluation and volume reconstruction. Here, the 2.5D analysis involves evaluating slices of data in a way that takes into account contextual information from adjacent slices. This approach bridges pure 2D and 3D analyses. The second route was based on including the gradient-weighted class activation mapping (Grad-CAM), where the most important input areas for the classification are detected [64]. The Grad-CAMS were obtained

from the bottle neck layer of the starting U-Net, Figure 5.5, and then concatenated to their correspondent CT image to create a new 2-channel input. The same U-Net architecture was then retrained on the new datasets and retested with the same analysis of the first route. Finally, the two procedures are then re-evaluated on a 3D level after a postprocessing step, consisting of a continuity analysis of the mask predictions. This analysis removed all those slices that were isolated in one of the three dimensions: if a mask prediction does not have any precedent and successive slice, that slice was removed; and if a mask prediction has successive or precedent slice with a different located in space mask prediction, the slice not connected to the continuity structure, or the smaller one, was removed. The testing and evaluation procedures are summarized in Figure 5.6.

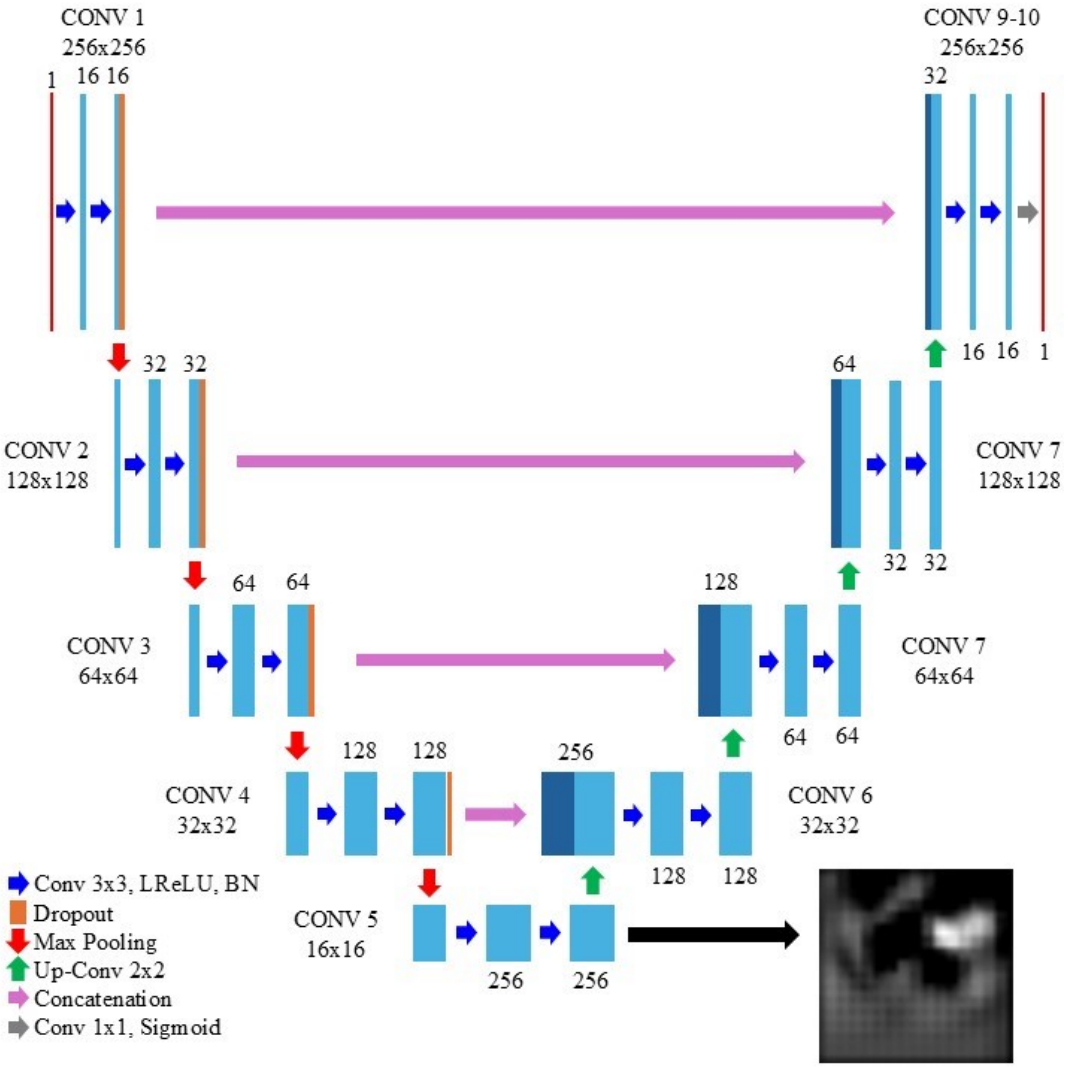

FIGURE 5.5: Grad-CAM computing process

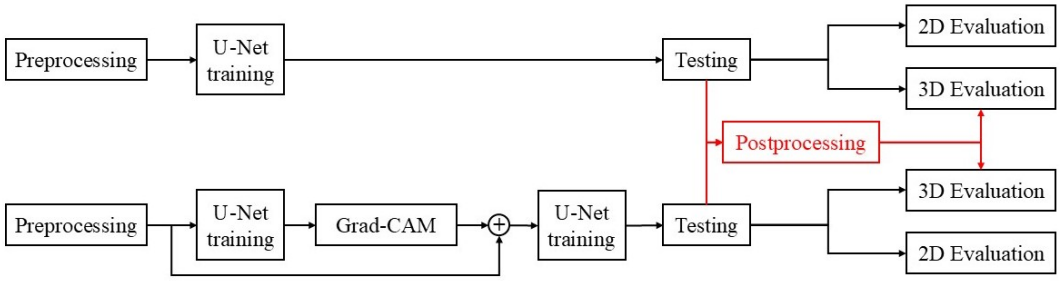

FIGURE 5.6: Pipeline for the testing and the evaluation. The postprocess (red) is applied on a second evaluation and only on the 3D prediction

5.2.6 Evaluation metrics and methods

The 2D U-Net has been evaluated on two different levels: the 2D level, evaluating directly the performances on single images, and the 3D level, evaluating the performances on the volumes reconstructed after the single slices' predictions.

For the 2D evaluation, the metrics used were the *DC* and the *Intersection over Union (IoU)*. The DC is the metric used also for building the loss function of the model, and it is represented by Eq. 5.1. IoU is one of the commonly used metrics for the performance evaluation of segmentation models because it essentially tells how well the predicted segmentation overlaps with the ground truth segmentation, similarly to the DC. IoU takes both false positives and false negatives into account; therefore, it gives a better insight into model performance in the case of imbalanced classes. IoU is more informative in tasks requiring object localization, as it puts great stress on the correct pixel prediction. Mathematically, IoU is defined by Eq. 5.3.

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{5.3}$$

where A is the set of predicted pixels in a class, B is the set of true pixels in that class, $|A \cap B|$ is the number of pixels in the intersection correctly predicted as positive, and $|A \cup B|$ is the number of pixels in the union of all predicted and true positive pixels. IoU values range from 0 to 1. When it is 1, it gives the perfect overlap, which is a very important metric for segmentation results in terms of exact and fine quality. Together with these two metrics, for the 2D evaluation, other performance indexes were used. *Sensitivity (SEN) or Recall (REC)* is the ratio of the true positive (TP) cases that

have been correctly caught by the model. This is defined by Eq. 5.4.

$$SEN = \frac{TP}{TP + FalseNegative} \tag{5.4}$$

This metric is essential in applications that require the identification of positive instances where a failure to do might be extremely costly, such as in medical diagnosis. *Specificity (SPE) or True Negative Rate*, defined by Eq. 5.5, is the measure of the proportion of true negatives (TN) that are correctly identified.

$$SPE = \frac{TN}{TN + FalsePositive} \tag{5.5}$$

Accuracy (ACC) is the overall ratio of the sum of correctly predicted positive and negative cases to the total cases, and it is defined Eq. 5.6.

$$ACC = \frac{TP + TN}{TotalCases} \tag{5.6}$$

ACC is a general performance measure; thus, it can be highly misleading in imbalanced datasets since it doesn't discriminate among the types of error. Therefore, SEN, SPE and ACC together are most likely to convey a detailed level picture of model performance. Finally, the 2D evaluation was performed from a statistical point of view, through the computing of the *True Area* of the cancer mask and the *Predicted Area*, and the performing of the Wilcoxon test and Bland-Altman (difference between ground truth and prediction distribution) on general data, on each independent set, and on each pixel surface set. This last analysis focuses on dividing the whole dataset into three sets based on the pixel dimension of the ground truth cancer mask for obtaining a sort of clinical analysis evaluating the ability of the model to predict small, medium, and large cancer areas. The sets were divided into smaller than 200 pixels, between 200 and 1000 pixels, and larger than 1000 pixels, without a clinical reference considering that the cancer size classification takes into account the volumes of the tumors and not the single surfaces of each slice.

For the 3D evaluation, the same two metrics were used, sensitivity as the only performance index, and, instead of the statistical analysis, a volume 3D reconstruction was performed. The 3D evaluation was then re-performed after the implementation of the postprocessing step.

5.3 Results

Table 5.1 and Table 5.2 report the 2D evaluation metrics and performance indexes of the proposed method, related to general segmentation performance and clinical utility performance, respectively. Figure 5.7, Figure 5.8, Figure 5.9, Figure 5.10, Figure 5.11 and Figure 5.12 show some examples of 2D lung cancer segmentation prediction performed by the proposed method. In each one the CT image and the ground truth used for testing, and the model prediction are reported. Figure 5.13, Figure 5.14, Figure 5.15 and Figure 5.16 show the Bland Altman plots obtained from the general segmentation performance and the clinical utility performance of the implemented model.

Table 5.3 reports the 3D evaluation metrics of the proposed method, computed before and after the implementation of the postprocess. Figure 5.17 and Figure 5.18 show some examples of the 3D volume reconstruction of the lung nodules from the 2D prediction output of the model, before and after the application of the postprocess technique. Table 5.4 report the comparison of the DC obtained as results from the proposed method and the others present in literature

TABLE 5.1: 2D performance according dataset split

Model	Split	DC	IoU	P-Value	SEN	SPE	ACC
U-Net	TRAIN	0.9306	0.8883	$6.77e^{-163}$	0.8628	0.9901	0.9648
	VAL	0.9193	0.8742	$6.28e^{-85}$	0.8711	0.9700	0.9491
	TEST	0.9315	0.8902	$1.23e^{-57}$	0.8619	0.9835	0.9573
	OVERALL	0.9289	0.8864	$2.88e^{-240}$	0.8640	0.9856	0.9608
Grad-CAM	TRAIN	0.9220	0.8730	$1.41e^{-152}$	0.8389	0.9892	0.9593
	VAL	0.9217	0.8714	$3.75e^{-32}$	0.8365	0.9736	0.9446
	TEST	0.9127	0.8626	$1.76e^{-72}$	0.8414	0.9828	0.9522
	OVERALL	0.9203	0.8703	$7.66e^{-252}$	0.8391	0.9855	0.9555

TABLE 5.2: 2D performance according to tumor surface

Model	Tumor Surface (pixels)	DC	IoU	P-Value
U-Net	SMALL (<200)	0.7997	0.7303	$3.32e^{-112}$
	MEDIUM (200><1000)	0.9123	0.8597	$1.90e^{-75}$
	LARGE (>1000)	0.9590	0.9283	$1.30e^{-70}$
Grad-CAM	SMALL (<200)	0.7735	0.6931	$2.19e^{-115}$
	MEDIUM (200><1000)	0.9021	0.8412	$3.34e^{-56}$
	LARGE (>1000)	0.9489	0.9109	$5.52e^{-101}$

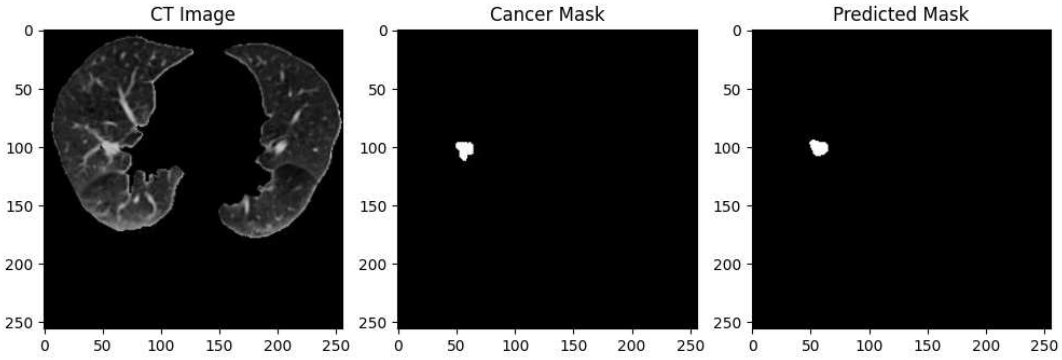

FIGURE 5.7: Example of the prediction of a small surface lung nodule by the proposed U-Net

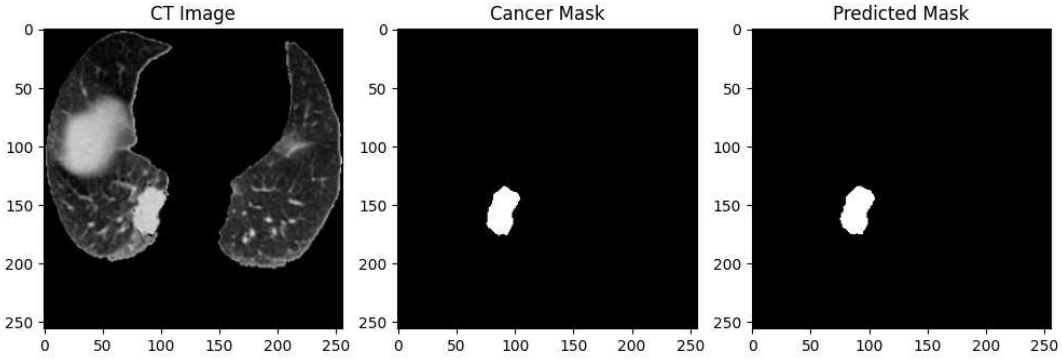

FIGURE 5.8: Example of the prediction of a medium surface lung nodule by the proposed U-Net

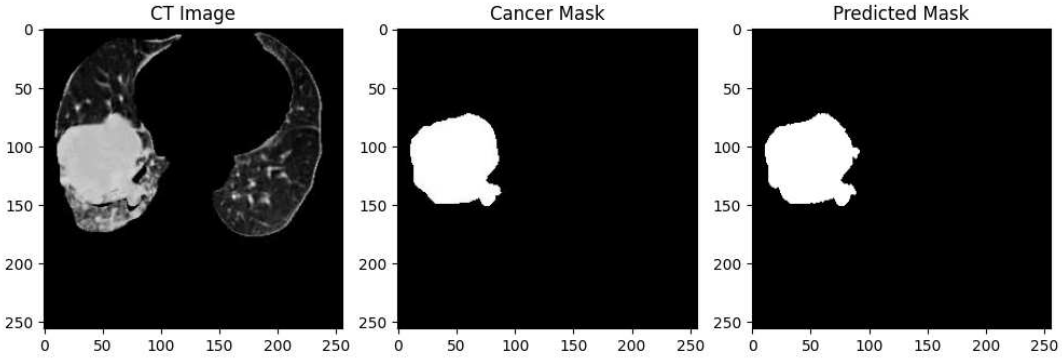

FIGURE 5.9: Example of the prediction of a large surface lung nodule by the proposed U-Net

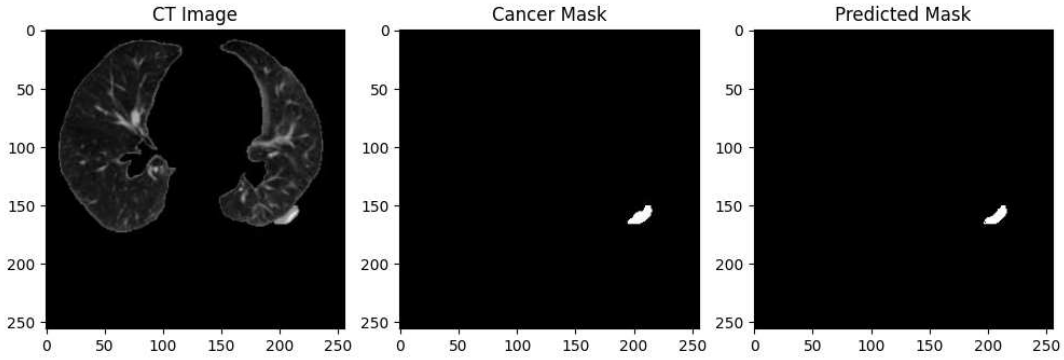

FIGURE 5.10: Example of the prediction of a small surface lung nodule by the proposed Grad-CAM U-Net

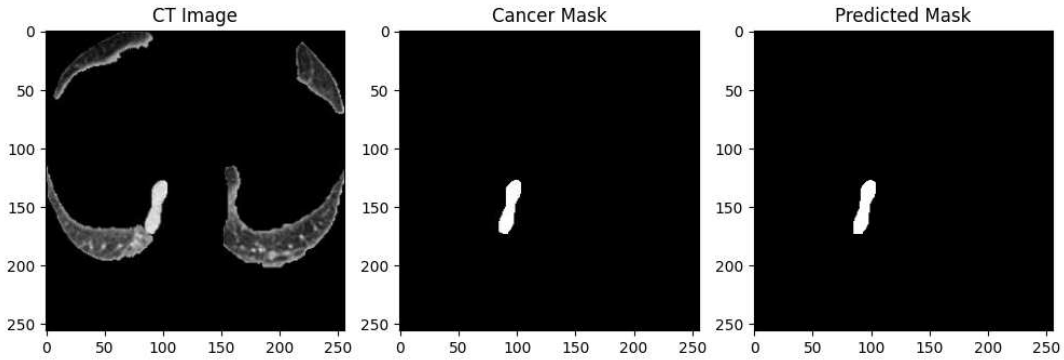

FIGURE 5.11: Example of the prediction of a medium surface lung nodule by the proposed Grad-CAM U-Net

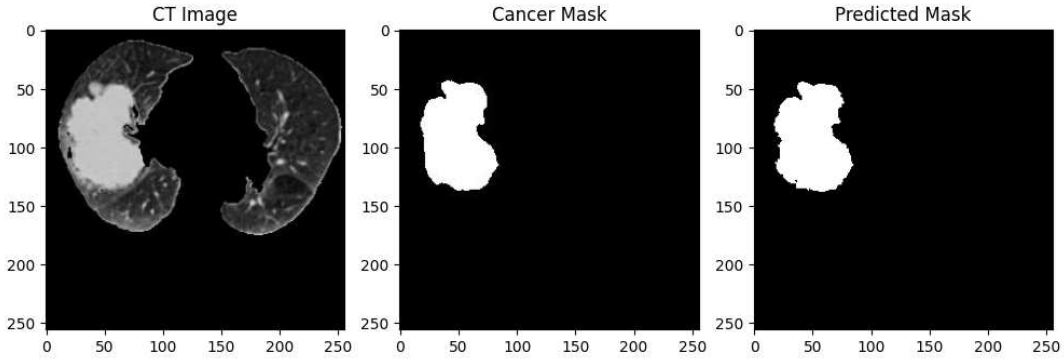

FIGURE 5.12: Example of the prediction of a large surface lung nodule by the proposed Grad-CAM U-Net

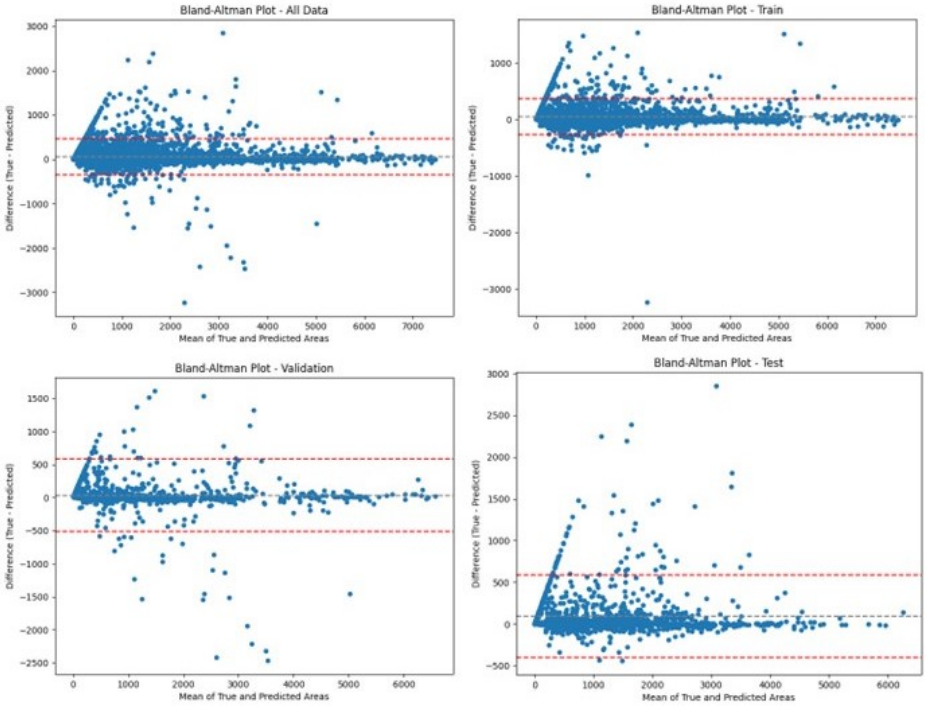

FIGURE 5.13: Bland Altman plot by dataset type for the proposed U-Net

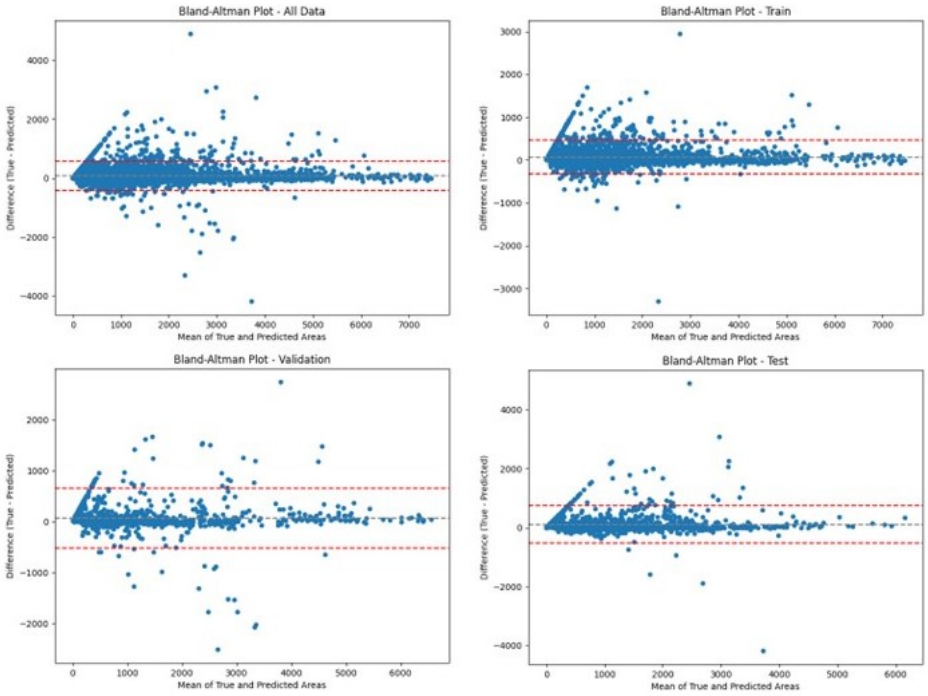

FIGURE 5.14: Bland Altman plot by dataset type for the proposed Grad-CAM U-Net

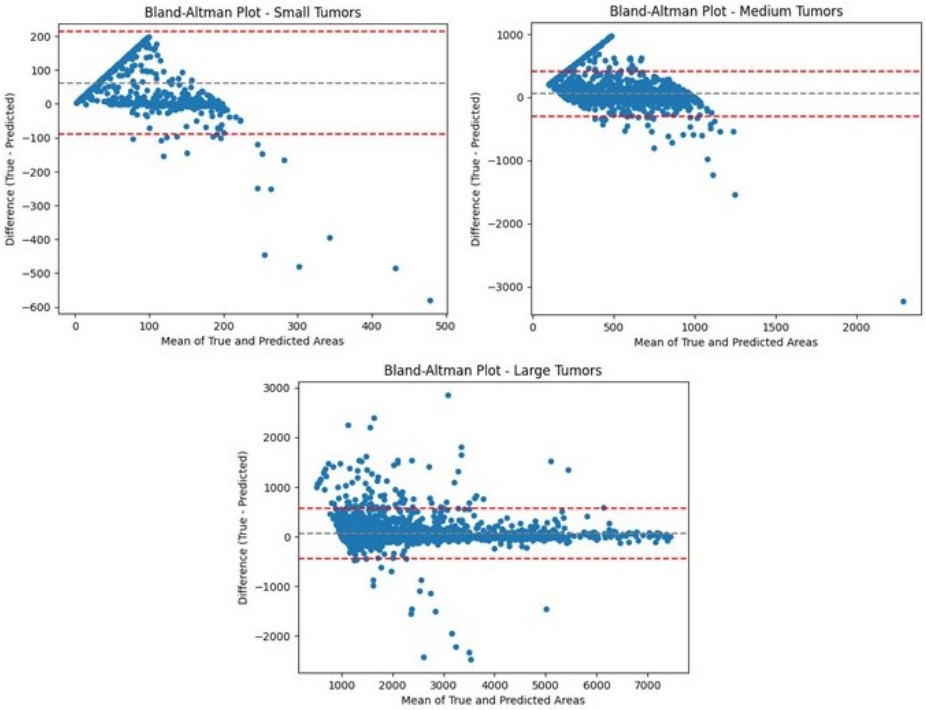

FIGURE 5.15: Bland Altman plot by nodule surface dimension for the proposed U-Net

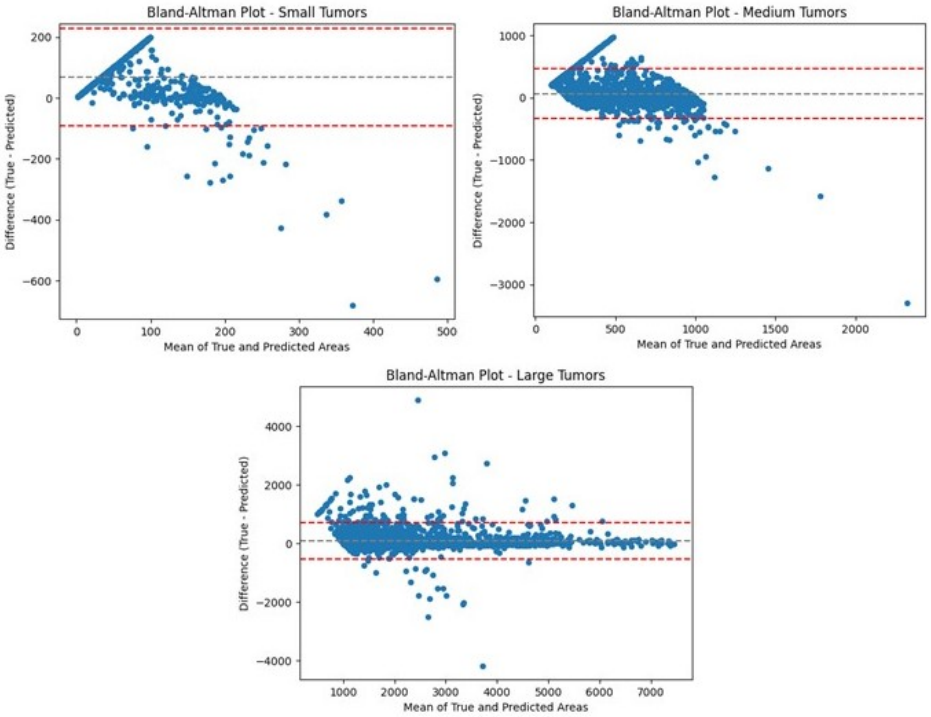

FIGURE 5.16: Bland Altman plot by nodule surface dimension for the proposed Grad-CAM U-Net

TABLE 5.3: 3D Performance Analysis with percentage increment after postprocessing

		Standard			**Postprocessed**		
Model	Split	DC	IoU	SEN	DC	IoU	SEN
U-Net	TRAIN	0.8504	0.7750	0.8229	0.8745 (+2.83%)	0.8005 (+3.30%)	0.8368 (+1.69%)
	VAL	0.8083	0.7203	0.8038	0.8262 (+2.21%)	0.7420 (+3.01%)	0.8123 (+1.06%)
	TEST	0.8448	0.7702	0.8278	0.8850 (+4.76%)	0.8123 (+5.47%)	0.8548 (+3.26%)
	OVERALL	0.8423	0.7653	0.8207	0.8685 (+3.11%)	0.7930 (+3.62%)	0.8362 (+1.89%)
Grad-CAM	TRAIN	0.8224	0.7357	0.8012	0.8567 (+4.17%)	0.7740 (+5.20%)	0.8183 (+2.14%)
	VAL	0.7998	0.7089	0.8047	0.8470 (+5.89%)	0.7591 (+7.10%)	0.8313 (+3.31%)
	TEST	0.8377	0.7530	0.8173	0.8645 (+3.20%)	0.7832 (+4.01%)	0.8346 (+2.12%)
	OVERALL	0.8218	0.7349	0.8050	0.8567 (+4.25%)	0.7735 (+5.26%)	0.8236 (+2.31%)

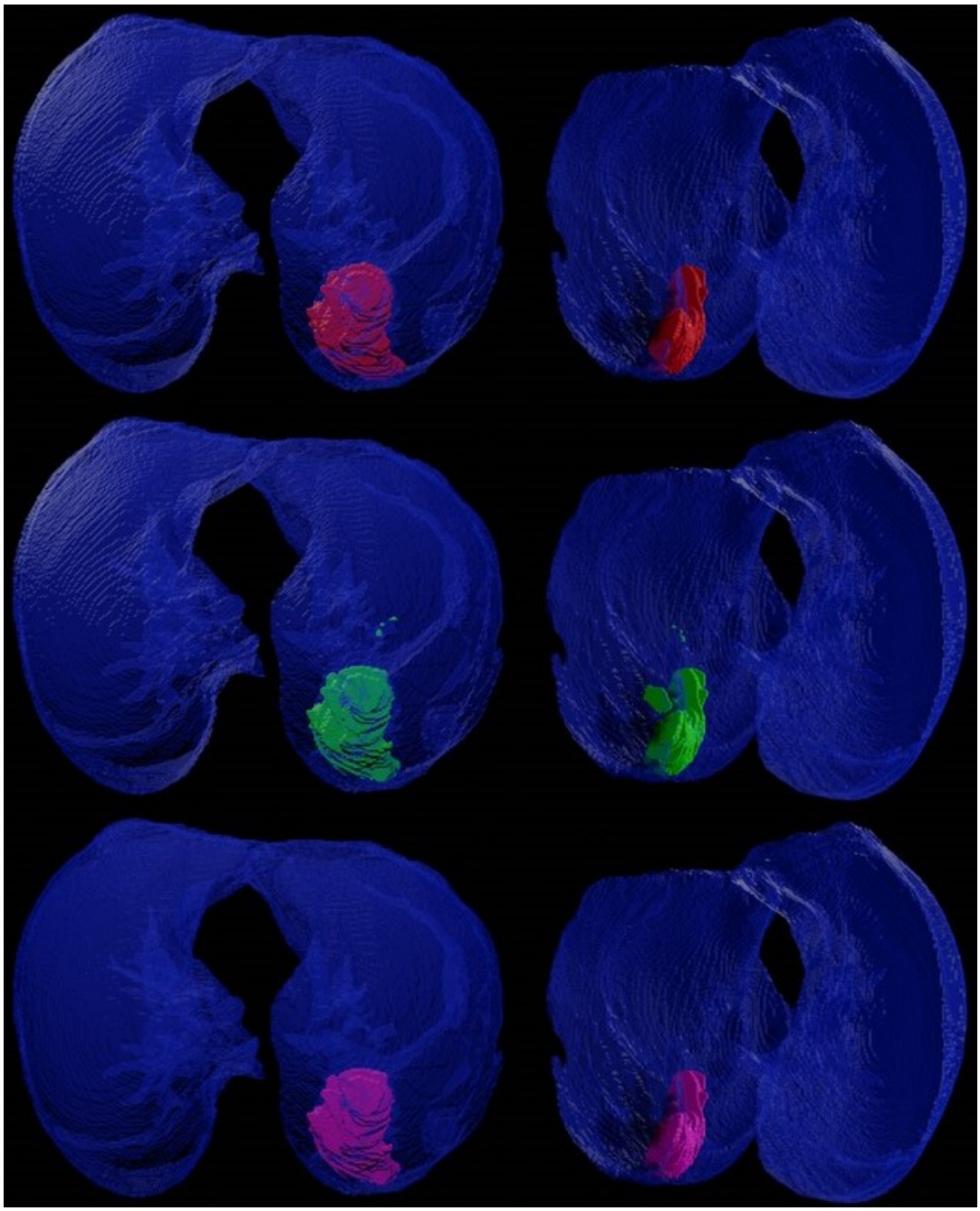

FIGURE 5.17: Example of the 3D volume reconstruction with the proposed U-Net prediction and the subsequent postprocess application; **(red)** True 3D volume, **(green)** Predicted 3D volume, **(fucsia)** Postprocessed 3D volume

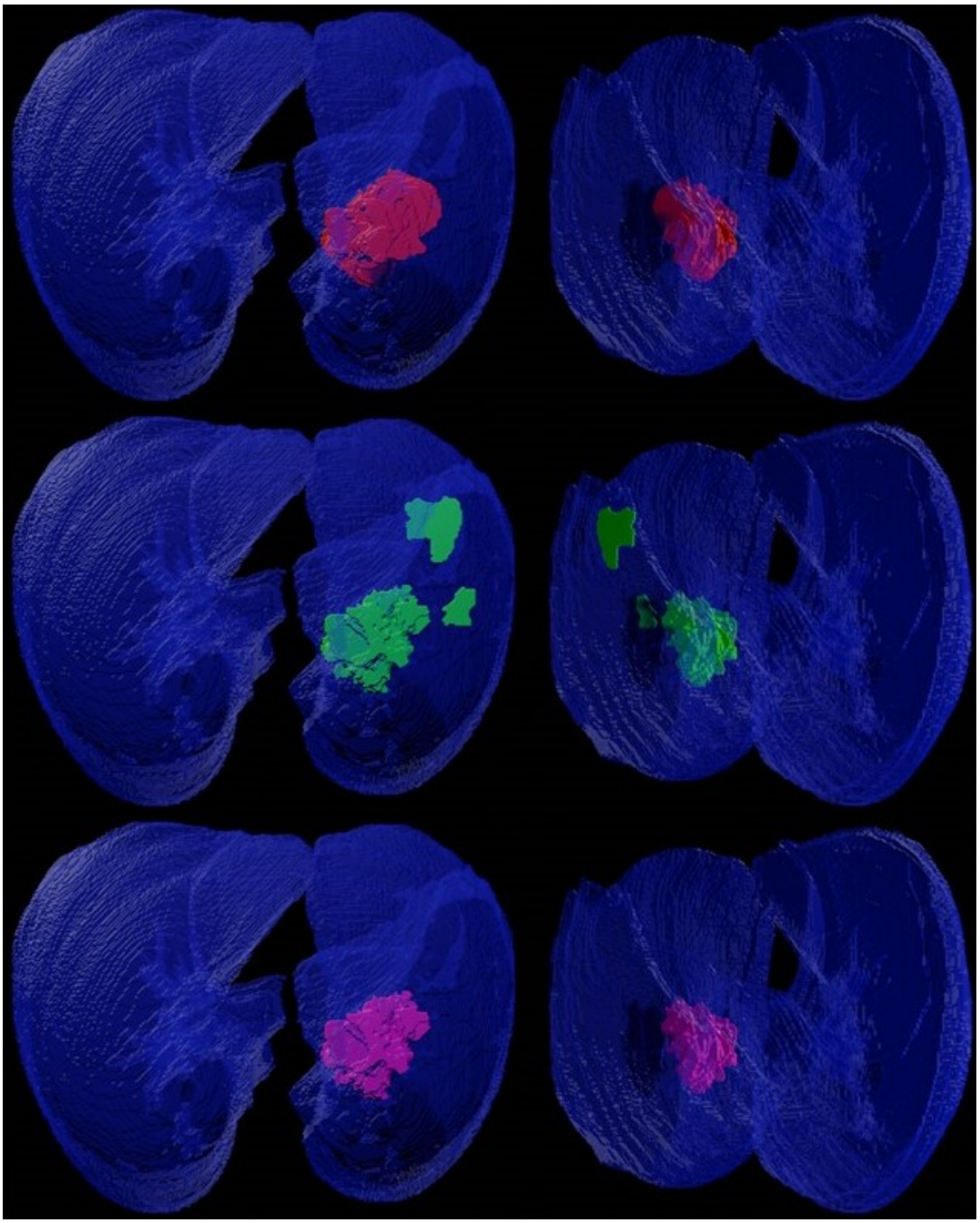

FIGURE 5.18: Example of the 3D volume reconstruction with the proposed Grad-CAM U-Net prediction and the subsequent postprocess application; **(red)** True 3D volume, **(green)** Predicted 3D volume, **(fucsia)** Postprocessed 3D volume

TABLE 5.4: Comparison with the other literature methods

Type	Method	DC
2D	[49]	0.8283
	[54]	0.7442
	[55]	0.8282
	[62]	0.9200
	[63]	0.8000
2D and 3D	[43]	0.8215
	[53]	0.7500
3D	[44]	0.8690
	[46]	0.9720
	[47]	0.7500
	[48]	0.8667
	[51]	0.8450
	[52]	0.8300
	[58]	0.7800
	[60]	NaN
Proposed Method	U-Net	0.9315
	3D U-Net Reconstruction	0.8448
	3D U-Net Reconstruction + Postprocess	0.8850
	2D Grad-CAM	0.9127
	3D Grad-CAM Reconstruction	0.8377
	3D Grad-CAM Reconstruction + Postprocess	0.8645

5.4 Discussion

In this project, a 2D U-Net model is proposed for the segmentation of the lung cancer in CT images. The architecture is based on a five-layer encoder-decoder structure that guarantees the effective extraction of contextual information and the spatial accuracy of precise segmentation. The idea of including the Grad-CAM block into the model architecture comes from trying to help the model focus on the right spot in the image to detect and then segment the cancer. Even though, as it has been seen, the results from the single U-Net model were good as well. From the comparison of all the evaluation methods included in the work, it can be noticed that the introduction of Grad-CAM did not improve the performance of the U-Net in absolute terms. However, the Grad-CAM model reduced the difference between the performances on the training and validation datasets, indicating that the model had less overfitting than the simple U-Net. The lack of improvements due to the introduction of Grad-CAM could be related to the fact that the activation maps were computed from the bottleneck layer of the same U-Net model on which they were used as a second channel for its training. Thus, their use did not provide any additional information. Focusing on the first level of evaluation, which is 2D, it is possible to see that the performances on all the dataset types and the entire dataset are similar, suggesting that the model was well-trained to avoid the overfitting. One important aspect to highlight is the differences between sensitivity and specificity. Sensitivity is much lower than specificity in all datasets, indicating that the model finds more false negatives than false positives. This means that the cancer not detected by the model has a higher percentage than the cancer detected where there is none. It's as if, when the model is in doubt, it doesn't take risk. This consideration is further supported by the clinical results. In both cases, the model's performance is much lower, but still acceptable, when it must find small surface cancers compared to medium and large ones. This means that the model is strong enough to handle various cancer dimensions but has more difficulty with those that have small surfaces. However, in all three cases, the correct cancer detection shows some distortions at the segmentation borders, while the shape and the location are quite good. This could be attributed to the preprocessing pipeline. The images were originally 512x512, and after cropping, they were resized with an interpolation algorithm to 256x256. This resizing could have distorted the segment borders of the CT images, forcing the model to follow this distortion. The statistical analysis, using the Wilcoxon,

shows that the difference between true and predicted tumor areas is statistically significant within each dataset and each size category. This consistency across datasets is a good sign, suggesting that the goodness of the model's performance is not limited to a specific subset of the data. The significant p-values across all tumor sizes reinforce the previous findings that there is a general difference between true and predicted tumor areas. The Bland Altman analysis shows that there is a slight systematic bias between the true and the predicted area, but most samples remain within the limits of agreement. By combining the statistical information within the metrics, it can be said that the model is capable of detecting the shape and the location of the cancer to segment but is not perfect in matching its actual dimensions and borders. Regarding the 3D evaluation, the metrics and indices used report good but lower results with respect to the 2D analysis. This is understandable given that the 3D evaluation was derived from the reconstruction of the 3D volumes based on slice-by-slice predictions. In particular, the comparison of metrics before and after the continuity analysis post-process indicates an improvement in the overlap of the 3D cancer prediction with the ground truth. This improvement is due to the removal of isolated false positive slices that were not connected to the volume. This analysis improved the segmentation of cancers that were quite well predicted, while volumes that were already poorly predicted, such as some small cancer volumes, did not provide valid improvements. However, considering the significant improvements that the post-process provides for a clearer 3D cancer volume reconstruction, it could be clinically valuable, especially with enhancements to the U-Net model regarding its hyperparameter tuning, loss functions, or architecture, to achieve better delineated 2D segmentation and a higher detection rate for small surface 2D cancers, which will help with 3D volume reconstruction. Compared to the methods in literature, both the single 2D U-Net and the Grad-CAM achieved similar results, even higher in some cases. Most important is the performance on 3D reconstruction; it seems to be on par with the other 3D methods presented in literature. Thanks to the results obtained, it is possible to reconstruct the 3D volume of the lung cancer from the CT images effectively, even after utilizing a less resource consuming and computational 2D deep learning model instead of performing a more complex 3D analysis.

Conclusion

In this thesis, a 2D U-Net model for the segmentation of lung cancer in CT images was implemented and evaluated. More importantly, it has sought to establish an effective method for 3D reconstruction. The proposed model architecture, based on a five-layer encoder-decoder structure, demonstrated effective performance and computationally efficient in segmenting lung cancer by capturing spatial details and contextual features. Despite the attempt to incorporate Grad-CAM into the model in order to improve it by focusing attention on parts of the image relevant for segmentation, the model's performance did not improve at all. The incorporation of Grad-CAM helped, however, reduce overfitting by bridging performance between the training and validation sets. This finding underlines the necessity of carefully incorporating additional modules, especially those providing no new information to the network. The results showed that the model was returning very strong 2D segmentation, maintaining consistent performance across the training and validation datasets into the test dataset with very minimal overfitting. However, sensitivity always remained lower than specificity because the model seems to adopt a conservative attitude in making segmentation decisions, leading to small areas that might remain undetected. This bias has shown a very important direction for enhancement, especially regarding how to improve the model's ability to detect smaller nodules, which in most cases prove crucial in the diagnosis of early-stage cancers. Also, the model is capable of underlining the general shape and the location of the cancer, but with some issues on the perfect overlapping and precision of the borders. The 3D evaluation, obtained by reconstructing the volumes of cancer from 2D slice predictions, had promising results, though a bit limited compared to the 2D analysis. Post-processing further improved the accuracy of volumes reconstructed from cancers by eliminating isolated false positives and highlighting the importance of continuity analysis for the derivation of a more realistic 3D representation. However, this post-processing step provided limited improvements for poorly segmented cases-mostly those with small cancer surface areas-while it was refining well-predicted volumes. In

conclusion, this work has proved that effective 3D reconstruction of lung cancer volumes can be obtained from a 2D model, thus offering a computationally efficient alternative to full 3D models. This technique is also clinically useful, especially with further improvements in the detection of small lesions and border precision. These results suggest that targeted improvements in hyperparameters, loss functions, and model architecture can make the 2D U-Net an effective tool for segmenting lung cancers and thereby enable 3D volume analysis with reduced computational resource use.

Bibliography

[1] Kara Rogers. *The Respiratory System (The Human Body)*. Britannica Educational Pub, Aug. 2010.

[2] J. Gordon Betts; Kelly A. Young; James A. Wise; Eddie Johnson; Brandon Poe; Dean H. Kruse; Oksana Korol; Jody E. Johnson; Mark Womble; Peter DeSaix. *Anatomy and Physiology*. Openstax, Apr. 2022, pp. 1033–1084.

[3] Paolo Palange; Gernot Rhode. *Handbook of Respiratory Medicine*. European Respiratory Society (ERS); third edition, Sept. 2019.

[4] Joseph F. Tomashefski Jr. MD and Carol F. Farver MD. "Anatomy and Histology of the Lung". In: *Dail and Hammar's Pulmonary Pathology* (2009), pp. 20–48. URL: https://doi.org/10.1007/978-0-387-68792-6_2.

[5] "National Center Cancer Institute (NCI)". In: *NIH* (). URL: https://www.cancer.gov/about-cancer/understanding/what-is-cancer.

[6] Wendy A. Cooper; David C. L. Lam; Sandra A. O'Toole; John D. Minna. "Molecular biology of lung cancer". In: *Journal Of Thoracic Disease* 5 (Supplement 5) (Oct. 2013). URL: https://jtd.amegroups.org/article/view/1597.

[7] Freddie Bray BSc; Mathieu Laversanne; Hyuna Sung; Jacques Ferlay; Rebecca L. Siegel; Isabelle Soerjomataram; Ahmedin Jemal. "Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries". In: *A Cancer Journal for Clinicians* 74 (Issue 3) (Apr. 2024), pp. 229–263. URL: https://doi.org/10.3322/caac.21834.

[8] Krishna Chaitanya Thandra; Adam Barsouk; Kalyan Saginala; John Sukumar Aluru; Alexander Barsouk. "Epidemiology of lung cancer". In: *Contemporary Oncology* 25 (Feb. 2021), pp. 45–52. URL: https://doi.org/10.5114/wo.2021.103829.

[9] Jyoti Malhotra; Matteo Malvezzi; Eva Negri; Carlo La Vecchia; Paolo Boffetta. "Risk factors for lung cancer worldwide". In: *European Respiratory Journal* 48 (May 2016), pp. 889–902. URL: https://doi.org/10.1183/13993003.00359-2016.

[10] Annalori Panunzio; Paolo Sartori. "Lung Cancer and Radiological Imaging". In: *Curr Radiopharm* 13 (2020), pp. 238–242. URL: https://doi.org/10.2174/1874471013666200523161849.

[11] Min Zheng. "Classification and Pathology of Lung Cancer". In: *Surgical Oncology Clinics* 25 (July 2016), pp. 447–468. URL: https://doi.org/10.1016/j.soc.2016.02.003.

[12] The International Agency for Research on Cancer. *Pathology and Genetics of Tumours of the Lung, Pleura, Thymus and Heart (IARC WHO Classification of Tumours)*. World Health Organization; 1st Edition, Sept. 2004.

[13] Irfan Maghfoor; Michael C Perry. "Lung cancer". In: *Annals of Saudi medicine* 25 (Feb. 2005), pp. 1–12. URL: https://doi.org/10.5144/0256-4947.2005.1.

[14] Frank C Detterbeck; Daniel J Boffa; Anthony W Kim; Lynn T Tanoue. "The Eighth Edition Lung Cancer Stage Classification". In: *Chest* 151 (Jan. 2017), pp. 193–203. URL: https://doi.org/10.1016/j.chest.2016.10.010.

[15] "American Cancer Society". In: *ACS* (). URL: https://www.cancer.org/cancer/types/lung-cancer.html.

[16] Alberto Ruano-Raviña; Mariano Provencio; Virginia Calvo de Juan; Enric Carcereny; Teresa Moran; Delvys Rodriguez-Abreu; Rafael López-Castro; Eugenio Cuadrado Albite; María Guirado; Lucía Gómez González; Bartomeu Massutí; Ana Laura Ortega Granados; Ana Blasco; Manuel Cobo; Rosario Garcia-Campelo; Joaquim Bosch; José Trigo; Óscar Juan; Carlos Aguado de la Rosa; Manuel Dómine; María Sala; Juana Oramas; Joaquín Casal-Rubio; Sara Cerezo. "Lung cancer symptoms at diagnosis: results of a nationwide registry study". In: *National Library of Medicine* 5 (2020), pp. 1–21. URL: https://doi.org/10.1136%2Fesmoopen-2020-001021.

[17] Lauren G Collins; Christopher Haines; Robert Perkel; Robert E Enck. "Lung cancer: diagnosis and management". In: *American Family Physician* 75 (Jan. 2007), pp. 56–63.

[18] Bala Prabhakar; Pravin Shende; Steffi Augustine. "Current trends and emerging diagnostic techniques for lung cancer". In: *Biomedicine and Pharmacotherapy* 106 (Oct. 2018), pp. 1586–1599. URL: https://doi.org/10.1016/j.biopha.2018.07.145.

[19] W De Wever; Johan Coolen; Johny Verschakelen. "Imaging techniques in lung cancer". In: *ERS Journal* 7 (June 2011), pp. 338–346. URL: http://dx.doi.org/10.1183/20734735.022110.

[20] M. A. Hayat. *Cancer Imaging: Lung and Breast Carcinomas)*. Elsevier Science Publishing Co Inc, 2007.

[21] Yi-Xiang J. Wang; Gladys G. Lo; Jing Yuan; Peder E. Z. Larson; Xiaoliang Zhang. "Magnetic resonance imaging for lung cancer screen". In: *Journal of Thoracic Disease* 6 (Sept. 2014), 1340–1348. URL: https://doi.org/10.3978%2Fj.issn.2072-1439.2014.08.43.

[22] Bruno Hochhegger; Giordano Rafael Tronco Alves; Klaus Loureiro Irion; Carlos Cezar Fritscher; Leandro Genehr Fritscher; Natália Henz Concatto; and Edson Marchiori. "PET/CT imaging in lung cancer: indications and findings". In: *Journal Brasileiro di Pneumologia* 41 (June 2015), 264–274. URL: https://doi.org/10.1590%2FS1806-37132015000004479.

[23] Armin Ernst Devanand Anantham; Mariko Siyue Koh. "Endobronchial ultrasound". In: *Respiratory Madicine* 103 (Oct. 2009), pp. 1406–1414. URL: https://doi.org/10.1016/j.rmed.2009.04.010.

[24] Fotios Sampsonas; Loukas Kakoullis; Dimosthenis Lykouras; Kiriakos Karkoulias; Konstantinos Spiropoulos. "EBUS: Faster, cheaper and most effective in lung cancer staging". In: *International Journal of Clinical Practice* 72 (Feb. 2018). URL: https://doi.org/10.1111/ijcp.13053.

[25] Nilendu C Purandare; Venkatesh Rangarajan. "Imaging of lung cancer: Implications on staging and management". In: *Indian J Radiol Imaging* 25 (June 2015), pp. 109–120. URL: https://doi.org/10.4103%2F0971-3026.155831.

[26] Pariwat Ongsulee. "Artificial intelligence, machine learning and deep learning". In: *International Conference on ICT and Knowledge Engineering (ICTandKE)* (2017), pp. 1–6. URL: https://doi.org/10.1109/ICTKE.2017.8259629.

[27] Yoav Mintz; Ronit Brodie. "Introduction to artificial intelligence in medicine". In: *Minimally Invasive Therapy and Allied Technologies* 28 (Feb. 2019), pp. 73–81. URL: https://doi.org/10.1080/13645706.2019.1575882.

[28] Yann LeCun; Yoshua Bengio; Geoffrey Hinton. "Deep learning". In: *Nature* 521 (May 2015), pp. 436–444. URL: http://dx.doi.org/10.1038/nature14539.

[29] Riccardo Miotto; Fei Wang; Shuang Wang; Xiaoqian Jiang; Joel T Dudley. "Deep learning for healthcare: review, opportunities and challenges". In: *Briefings in Bioinformatics* 19 (Nov. 2018), pp. 1236–1246. URL: http://dx.doi.org/10.1093/bib/bbx044.

[30] Mohaiminul Islam; Guorong Chen; Shangzhu JinShangzhu Jin. "An Overview of Neural Network". In: *American Journal of Neural Networks and Applications* 5 (May 2019). URL: http://dx.doi.org/10.11648/j.ajnna.20190501.12.

[31] Nikolaus Kriegeskorte; Tal Golan. "Neural network models and deep learning". In: *Current Biology* 29 (Apr. 2019), pp. 231–236. URL: https://doi.org/10.1016/j.cub.2019.02.034.

[32] Bulus Bali; Nachandiya Nathan; David Tufe Nzadon. "A study on medical applications of artificial neural networks". In: *International Journal of Research and Analytical Reviews* 7 (Sept. 2020). URL: http://dx.doi.org/10.13140/RG.2.2.23777.02404.

[33] Siddharth Sharma; Simone Sharma; Anidhya Athaiya. "Activation functions in neural networks". In: *International Journal of Engineering Applied Sciences and Technology* 4 (May 2020), pp. 310–316. URL: http://dx.doi.org/10.33564/IJEAST.2020.v04i12.054.

[34] Loris Nanni; Sheryl Brahnam; Michelangelo Paci; Stefano Ghidoni. "Comparison of Different Convolutional Neural Network Activation Functions and Methods for Building Ensembles for Small to Midsize Medical Data Sets". In: *MDPI Sensors* 22 (Aug. 2022). URL: https://doi.org/10.3390/s22166129.

[35] Bin Ding; Huimin Qian; Jun Zhou. "Activation functions and their characteristics in deep neural networks". In: *Chinese Control And Decision Conference (CCDC)* (June 2018), pp. 836–1841. URL: https://doi.org/10.1109/CCDC.2018.8407425.

[36] David Kriese. *A Brief Introduction to Neural Networks*. Kriesel2007NeuralNetworks, 2007. URL: http://www.dkriesel.com/.

[37] Charu C. Aggarwal. *Neural Networks and Deep Learning*. Springer, 2023.

[38] Ian Goodfellow; Yoshua Bengio; Aaron Courville. *Deep Learning*. The MIT Press, 2016.

[39] Rikiya Yamashita; Mizuho Nishio; Richard Kinh Gian Do; Kaori Togashi. "Convolutional neural networks: an overview and application in radiology". In: *Insights into Imaging* 9 (June 2018), pp. 611–629. URL: https://doi.org/10.1007/s13244-018-0639-9.

[40] Jonathan Long; Evan Shelhamer; Trevor Darrell. "Fully convolutional networks for semantic segmentation". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2015), pp. 3431–3440. URL: https://doi.org/10.1109/CVPR.2015.7298965.

[41] Olaf Ronneberger; Philipp Fischer; Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* 9351 (Nov. 2015), 234–241. URL: https://doi.org/10.1007/978-3-319-24574-4_28.

[42] Özgün Çiçek; Ahmed Abdulkadir; Soeren S. Lienkamp; Thomas Brox; Olaf Ronneberger. "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation". In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016* 9901 (Oct. 2016), 424–432. URL: https://doi.org/10.1007/978-3-319-46723-8_49.

[43] Shuo Wang; Mu Zhou; Zaiyi Liu; Zhenyu Liu; Dongsheng Gu; Yali Zang; Di Dong; Olivier Gevaert; Jie Tian. "Central Focused Convolutional Neural Networks: Developing a Data-driven Model for Lung Nodule Segmentation". In: *Medical Image Analysis* 40 (Aug. 2017), pp. 172–183. URL: https://doi.org/10.1016/j.media.2017.06.014.

[44] Zisha Zhong; Yusung Kim; Leixin Zhou; Kristin Plichta; Bryan Allen; John Buatti; and Xiaodong Wu. "3d fully convolutional networks for co-segmentation of tumors on pet-ct images". In: *IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (Apr. 2018), pp. 228–231. URL: http://dx.doi.org/10.1109/ISBI.2018.8363561.

[45] Qi Song; Junjie Bai; Dongfeng Han; Sudershan Bhatia; Wenqing Sun; William Rockey; John E Bayouth; John M Buatti; Xiaodong Wu. "Optimal Co-Segmentation of Tumor in PET-CT Images With Context Information". In: *IEEE Transactions on Medical Imaging* 32 (Sept. 2013), pp. 1685–1697. URL: http://dx.doi.org/10.1109/TMI.2013.2263388.

[46] Liang Zhao. "3D Densely Connected Convolution Neural Networks for Pulmonary Parenchyma Segmentation from CT Images". In: *Journal of Physics: Conference Series* 1631 (Sept. 2020), pp. 12–49. URL: http://dx.doi.org/10.1088/1742-6596/1631/1/012049.

[47] Boris Shirokikh; Alexey Shevtsov; Alexandra Dalechina; Egor Krivov; Valery Kostjuchenko; Andrey Golanov; Victor Gombolevskiy; Sergey Morozov; Mikhail Belyaev. "Accelerating 3D medical image segmentation by adaptive small-scale target localization". In: *Journal of imaging* 7 (Feb. 2021). URL: http://dx.doi.org/10.1088/1742-6596/1631/1/012049.

[48] Wei Chen; Fengchang Yang; Xianru Zhang; Xin Xu; Xu Qiao. "MAU-Net: Multiple Attention 3D U-Net for Lung Cancer Segmentation on CT Images". In: *Procedia Computer Science* 192 (2021), pp. 543–552. URL: https://doi.org/10.1016/j.procs.2021.08.056.

[49] Syeda Furruka Banu; Md Mostafa Kamal Sarker; Mohamed Abdel nasser; Hatem Rashwan; Domenec Puig. "WEU-Net: A Weight Excitation U-Net for Lung Nodule Segmentation". In: *Applied Sciences* 101 (Oct. 2021), pp. 349–356. URL: http://dx.doi.org/10.3233/FAIA210154.

[50] Niamul Quader; Md Mafijul Islam Bhuiyan; Juwei Lu; Peng Dai; Wei Li. "Weight Excitation: Built-in Attention Mechanisms in Convolutional Neural Networks". In: *European Conference on Computer Vision; Springer* 12375 (Sept. 2020), 87–103. URL: http://dx.doi.org/10.1007/978-3-030-58577-8_6.

[51] Shoji Kido; Shunske Kidera; Yasushi Hirano; Shingo Mabu; Tohru Kamiya; Nobuyuki Tanaka; Yuki Suzuki; Masahiro Yanagawa; Noriyuki Tomiyama. "Segmentation of lung nodules on CT images using a nested three-dimensional fully connected convolutional network". In: *Frontiers in Artificial Intelligence* 5 (Feb. 2022). URL: https://doi.org/10.3389/frai.2022.782225.

[52] Sundaresan A Agnes; Jeevanayagam Anitha. "Efficient multiscale fully convolutional UNet model for segmentation of 3D lung nodule from CT image". In: *Journal of Medical Imaging* 9 (Sept. 2022). URL: https://doi.org/10.1117/1.JMI.9.5.052402.

[53] Yifan Wang; Chuan Zhou; Heang-Ping Chan; Lubomir M Hadjiiski; Aamer Chughtai; Ella A Kazerooni. "Hybrid U-Net-based deep learning model for volume segmentation of lung nodules in CT images". In: *Medical physics* 49 (Sept. 2022), pp. 7287–7302. URL: https://doi.org/10.1002/mp.15810.

[54] Dechuan Lu; Junfeng Chu; Rongrong Zhao; Yuanpeng Zhang; Guangyu Tian. "A novel deep learning network and its application for pulmonary nodule segmentation". In: *Computational intelligence and neuroscience* (May 2022). URL: https://doi.org/10.1155/2022/7124902.

[55] Chandra Sekhara Rao Annavarapu; Samson Anosh Babu Parisapogu; Nikhil Varma Keetha; Praveen Kumar Donta; Gurindapalli Rajita. "A Bi-FPN-based encoder-decoder model for lung nodule image segmentation". In: *Diagnostics (Basel, Switzerland)* 13 (Apr. 2023). URL: https://doi.org/10.3390/diagnostics13081406.

[56] Diganta Misra. "Mish: A Self Regularized Non-Monotonic Neural Activation Function". In: *arXiv* (Aug. 2019). URL: https://doi.org/10.48550/arXiv.1908.08681.

[57] Mingxing Tan; Ruoming Pang; Quoc V. Le. "Efficientdet: Scalable and efficient object detection". In: *arXiv* (Nov. 2019). URL: https://doi.org/10.48550/arXiv.1911.09070.

[58] Junyoung Park; Seung Kwan Kang; Donghwi Hwang; Hongyoon Choi; Seunggyun Ha; Jong Mo Seo; Jae Seon Eo; Jae Sung Lee. "Automatic lung cancer segmentation in [18F]FDG PET/CT using a two-stage deep learning approach". In: *Nuclear medicine and molecular imaging* 57 (Apr. 2023), pp. 86–93. URL: https://doi.org/10.1007/s13139-022-00745-7.

[59] Chang Wang; Zongya Zhao; Qiongqiong Ren; Yongtao Xu; Yi Yu. "Dense U-net based on patchbased learning for retinal vessel segmentation". In: *Entropy 2019* 21 (Feb. 2019). URL: https://doi.org/10.3390/e21020168.

[60] T Weikert; P F Jaeger; S Yang; M Baumgartner; H C Breit; D J Winkel; G Sommer; B Stieltjes; W Thaiss; J Bremerich; K H Maier-Hein; A W Sauter. "Automated lung cancer assessment on 18F-PET/CT using Retina U-Net and anatomical region segmentation". In: *European radiology* 33 (June 2023), pp. 4270–4279. URL: https://doi.org/10.1007/s00330-022-09332-y.

[61] Paul F. Jaeger; Simon A. A. Kohl; Sebastian Bickelhaupt; Fabian Isensee; Tristan Anselm Kuder; Heinz-Peter Schlemmer; Klaus H. Maier-Hein. "Retina U-Net: Embarrassingly Simple Exploitation of Segmentation Supervision for Medical Object Detection". In: *Medicine, Computer Science* 116 (Nov. 20184), p. 161.183. URL: https://doi.org/10.48550/arXiv.1811.08661.

[62] Tenzin Kunkyab; Zhila Bahrami; Heqing Zhang; Zheng Liu; Derek Hyde. "A deep learning-based framework (Co-ReTr) for auto-segmentation of non-small cell-lung cancer in computed tomography images". In: *Journal of applied clinical medical physics* 25 (Mar. 2024). URL: https://doi.org/10.48550/arXiv.1811.08661.

[63] Fuli Zhang; Qiusheng Wang; Enyu Fan; Na Lu; Diandian Chen; Huayong Jiang; Yanjun Yu. "Enhancing non-small cell lung cancer tumor segmentation with a novel two-step deep learning approach". In: *Journal of Radiation Research and Applied Sciences* 17 (Mar. 2024). URL: https://doi.org/10.1016/j.jrras.2023.100775.

[64] Robert Kozma; Cesare Alippi; Yoonsuck Choe; Francesco Carlo Morabito. *Artificial Intelligence in the Age of Neural Networks and Brain Computing*. Academic Press; 2° edizione, Oct. 2023.