UNIVERSITÀ POLITECNICA DELLE MARCHE

FACULTY OF ENGINEERING

Master's Degree in Biomedical Engineering

Cloud-YLung Classification of Pulmonary Cancer Histotypes from PET Scans

Supervisor:
***Prof.* Laura Burattini**

Co-supervisors:
***Dr.* Agnese Sbrollini**
***Dr.* Selene Tomassini**

Candidate:
Di Marco Erica

Academic Year: 2021-2022

Abstract

Nowadays, lung cancer is one of the riskiest tumors in the world and its incidence is rapidly growing due to increased smoking and air pollution. Since only 20% of all cases are diagnosed early, lung cancer appears to be one of the cancers with the highest mortality rate in both men and women.

Clinically, many radiological imaging methods such as computed tomography and positron emission tomography are available to diagnose lung cancer. However, to verify the actual malignancy of the tumor mass and to find out which histological type it belongs to, a biopsy is required.

Deep learning algorithms for lung cancer diagnosis applied to both planar images and volumetric scans are growing in popularity, however most of them are focused on classifying the image or scan based on whether or not the cancer is present. Only few studies, instead, are focused on the very demanding task of identifying the histological type of lung cancer directly from the radiological datum. One of them is that of Tomassini et al. in which a neural network called Cloud-YLung has been trained with computed tomography whole-lung scans to recognize two histological types of lung cancer.

The aim of this thesis is to train Cloud-YLung with positron emission tomography scans to evaluate which kind of radiological data between positron emission tomography and computed tomography scans is more suitable in non-invasive lung cancer histological type classification.

According to the results obtained in this study and those obtained by Tomassini et al., it was possible to conclude that computed tomography data seem to be more informative for this task, achieving a higher accuracy in distinguishing the histological types. Nevertheless, Cloud-YLung was suitably built to be trained with computed tomography data, so tweaking it to better fit positron emission tomography data would likely increase the performances.

In conclusion, the task addressed in this thesis is highly challenging and the results obtained are however promising for future studies focused on deep learning algorithms for the automatic detection of the histological type of lung cancer, in order to avoid the need for biopsy, thus reducing the cost and time for lung cancer diagnosis.

Index

Introduction

Lung cancer is one of the most threatening forms of tumor existing nowadays, being the one with the highest mortality rate in man and the second highest mortality rate in women, ranking lower only in relation to breast cancer. It is so risky mainly because it is usually diagnosed in advanced stage, since this disease doesn't cause any signs or symptoms in the early stage, and this precludes many curative treatments. Thus, an early diagnosis is crucial to reduce the mortality rate of this disease.

The main causes that may lead to the onset of lung cancer are smoking, air pollution or genetic causes and the possible treatments for this disease are various, from surgical removement of the tumor mass to less invasive treatments, such as chemotherapy or radiotherapy, according to the severity of the condition.

Lung cancer constitutes a group of heterogenous tumors, that include a variety of different histological types. Generally, lung cancer can be divided into non-small cell carcinoma (NSCC) and small cell lung carcinoma (SCLC). NSCCs are more common and can be further divided into adenocarcinoma (ADC), squamous cell carcinoma (SCC) and large cell carcinoma (LCC).

There are several ways to diagnose the presence of lung cancer, such as through a chest radiography, a computed tomography (CT) or a positron emission tomography (PET). However, to check the malignancy or benignity of the tumor mass and possibly to find out which histological group the tumor belongs to, a biopsy is required.

Based on this, it would be very useful to find a way to identify the histological type of lung cancer directly from CT or PET scans, avoiding the need for biopsy, thus reducing the costs and time for the diagnosis.

Nowadays different studies are focused on the development of neural networks able to automatically identify a lung tumor from a CT or from a PET scan, discriminating between an image in which the tumoral mass is present and an image in which it is absent. However, the field of recognizing the histological type of lung cancer directly from the scan in an automatic way through a neural network is still not so developed.

According to that, the purpose of this thesis is to test the abilities of a neural network in distinguishing two histological types of lung cancer, namely ADC and SCC, from properly preprocessed PET scans, understanding if this kind of approach may be accurate enough to avoid the need for biopsy, or if this field should be further investigated.

In the following, a brief description of all the chapters of this study is reported, so that the reader can easily follow the thesis.

The first chapters introduce the main notions necessary to understand the research topics. Particularly, Chapter 1 illustrates the anatomy and physiology of the respiratory system, focusing on the anatomical parts that make up the respiratory system and on some physical aspects underlying the mechanics of breathing. In Chapter 2 lung cancer is described, highlighting the main causes, signs, symptoms, and histological types that compose this heterogeneous group, together with the description of how to diagnose this disease and possible therapies or preventive treatments. Chapter 3 clears up the most used diagnostic imaging modalities for lung cancer, centering on CT, PET, and the hybrid modality PET/CT. Chapter 4 reports a theoretical overview of neural networks, focusing on the convolutional neural networks for planar and volumetric data classification, whereas Chapter 5 illustrates a literature review on the deep learning techniques applied to the field of automatic recognition of lung cancer, centered on the methods for planar and volumetric data classification for both CT and PET/CT acquisition techniques.

The last chapters, instead, are focused on illustrating the work done in this thesis. Particularly, Chapter 6 describes the methods used in this study to reach the mentioned purpose, focusing on the dataset, the methodology and the statistic used. Chapter 7, instead, illustrates the results obtained which are then discussed in Chapter 8. Finally, in Chapter 9 the reader can appreciate the conclusions drawn in this thesis work.

1. Anatomy and physiology of the respiratory system

The respiratory system is a set of specialized organs involved in supplying oxygen (O_2) to the blood and removing carbon dioxide (CO_2) from the blood and it has more different functions than is commonly assumed, such as:

- *Gas exchange*: it provides for the exchange of O_2 and CO_2 between the blood and the air.
- *Oxygen supplier*: the respiratory system keeps the body constantly supplied with oxygen.
- *Elimination*: it eliminates carbon dioxide from the circulation.
- *Humidifier*: the respiratory system purifies, humidifies, and warms incoming air.
- *Communication*: it allows people to talk, laugh, cry.
- *Smell*: it provides for the sense of smell.
- *Acid-base balance*: by eliminating CO_2, it helps to control the pH of the blood. Excess CO_2 reacts with water and releases hydrogen ions (H^+), so, if the respiratory system fails to eliminate all the CO_2 produced, the H^+ ions accumulate in the blood, causing a lowering of pH (acidosis).
- *Regulation of blood pressure*: the lungs perform a leading role in the synthesis of angiotensin II, an important vasoconstrictor.
- *Blood and lymphatic flow*: the act of breathing creates a gradient between the rib cage and the abdomen which favors the lymphatic flow and venous return.

1.1 Path of the air within the respiratory system

The main organs of the respiratory system include nose, pharynx, larynx, trachea, bronchi and their smaller branches, and the lungs.

The *conducting portion* of the respiratory system consists of those ducts that are only used for the passage of the air flow from the nostrils to the bronchioles. In these passages there are no gas exchanges with blood because the walls of those ducts are too thick to allow gas diffusion quickly enough. The *respiratory portion* is made up of the alveoli and other regions placed distally, where gas exchanges occur.

The airways that go from the nose to the larynx are often referred to as the *upper respiratory tract*, whereas the portion that goes from the trachea to the lungs constitutes the *lower respiratory tract* (the respiratory organs located in the chest). In *Figure 1* the overall respiratory system is briefly illustrated.

1.1.1 Upper airways: Nose

The nose, see *Figure 2*, is the only externally visible part of the respiratory system, and it has several functions:

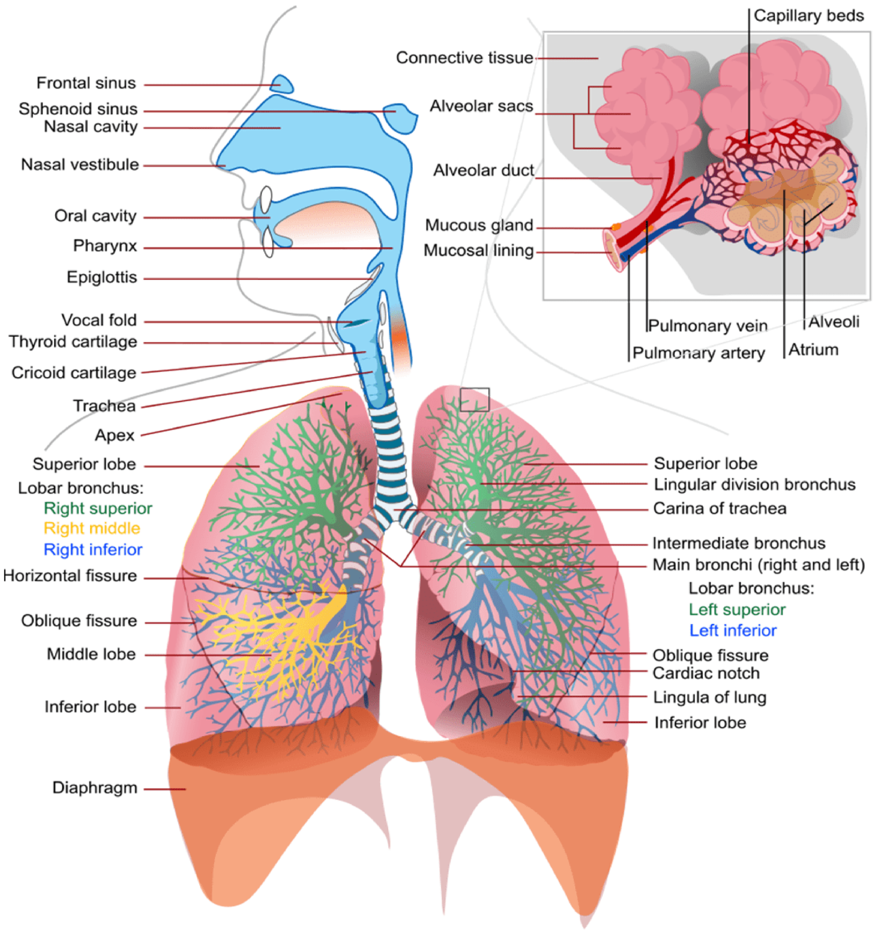

Figure 1: Respiratory system

- It heats, purifies, and humidifies the inhaled air.
- It detects odors in the air.
- It serves as a room of resonance that amplifies the voice.

The nose has four openings, two anterior called nostrils, or nares, and two posteriors called choanae. During breathing, air enters the nose by passing through the nostrils and, thanks to the presence of protective hairs or “vibrissae”, the entry of corpuscle material or insects into the respiratory tract is blocked.

The lateral walls of the nasal cavity are uneven due to three mucosa-covered projections, or lobes called conchae, which greatly increase the surface area of the mucosa exposed to the air and increase the air turbulence in the nasal cavity. The narrowness of the conchae and the turbulence caused by them causes the air along its course to encounter the nasal mucosa, which humidifies, cleanses, and heats the air. The nasal mucosa also contains mucous glands, localized in the lamina propria, which produce more mucus. Inhaled dust, pollen, bacteria, and other foreign substances are trapped by the mucus and swallowed; they are then digested, or they pass through the digestive tract rather than contaminating the lungs.

1.1.2 Upper airways: Pharynx

The pharynx is a funnel-shaped muscular organ that extends approximately 13 cm from the choanae to the larynx and serves as a common passageway for food and air. It is divided into three regions: nasopharynx, oropharynx, laryngopharynx. Air enters the superior portion, the nasopharynx, from the nasal cavity and then descends through the oropharynx and laryngopharynx to enter the larynx below. The peculiarity of the nasopharynx is that the inhaled air deviates 90° downward as it passes into it, such that relatively large particles generally collide with the posterior wall of the nasopharynx and attach to the mucosa near the tonsil, which respond to air-inhaled pathogens.

In the nasopharynx passes only air, while in the oropharynx and laryngopharynx it passes air, food, and drink. The muscles of the pharynx play the necessary roles in swallowing and speaking.

1.1.3 Upper airways: Larynx

The larynx, illustrated in *Figure 2*, is a cartilage cavity about 4 cm long. Its primary function is to prevent food and liquids from entering the airways, but another important function it has is the ability to produce sounds (phonation). The upper opening of the larynx is characterized by the presence of a particular leaf-like structure called the epiglottis. At rest, the epiglottis is positioned almost vertically. During swallowing, however, the extrinsic muscles of the larynx pull the larynx upwards, towards the epiglottis, while the tongue pushes the epiglottis downwards, so that the epiglottis closes the airways and directs food and liquids into the esophagus which is located behind the airways.

1.1.4 Lower airways: Trachea

The trachea, represented in *Figure 2*, is a rigid tube approximately 12 cm long with a diameter of 2.5 cm, anterior to the esophagus and it is supported by 16-20 C-shaped rings of hyaline cartilage. The open part of the C-shaped rings faces posteriorly and leaves space for the esophagus to dilate when the swallowed food passes.

The inner lining of the trachea presents muco-secerning goblet cells. The mucus traps inhaled particles, and the movement of upward cilia pushes debris and mucus itself towards the pharynx, where they are ingested through a mechanism called “mucociliary escalator”.

At the sternal angle, the trachea forks to form the left and right main bronchus. The lower tracheal cartilage has an internal median ridge called the carina, which directs the air flow to the right and left.

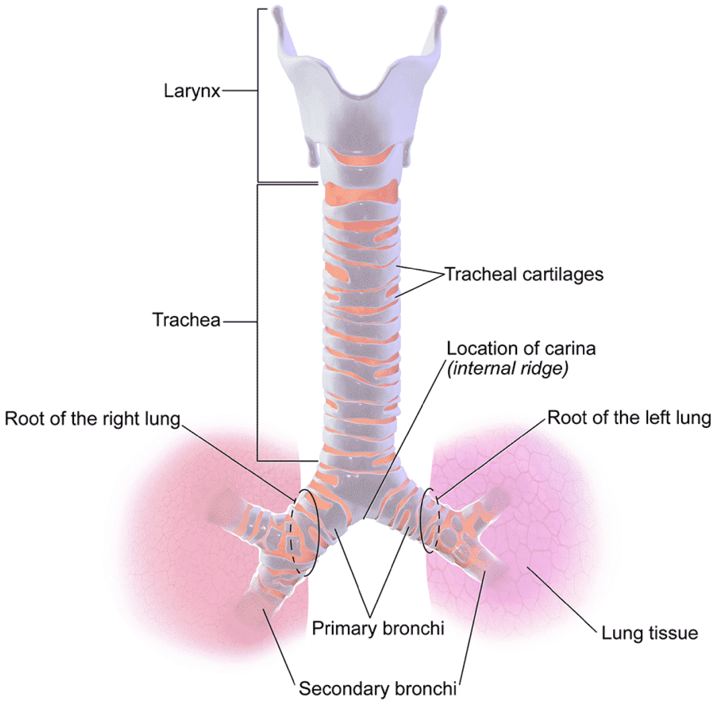

Figure 2: Anatomy of the trachea

1.1.5 Lower Airways: Bronchial Tree

Each lung has a branched air conduction system called the bronchial tree, which extends from the main bronchi, left and right, to the terminal bronchioles. In both lungs, each lobar bronchus divides into segmental bronchi, each of which ventilates a functionally independent unit of lung tissue. The main bronchi are supported, like the trachea, by C-shaped rings of hyaline cartilage, while the lobar and segmental bronchi are supported by cartilaginous plates arranged in a crescent shape. The bronchioles are the continuation of the airways that have lost the supporting cartilage. Each bronchiole divides into 50-80 terminal bronchioles, the final branches of the division into ducts. From each terminal bronchiole, two or more smaller respiratory bronchioles are formed, which have alveoli developing from their walls.

1.1.5 Lower airways: Lungs

The lung, shown in *Figure 3*, is a conical organ with a broad concave base resting on the diaphragm and a blunt vertex called apex that protrudes slightly above the collarbone.
The lungs are surrounded by the adjacent viscera, they do not occupy the entire rib cage and they are not symmetrical. The right lung is shorter than the left, due to the presence of the liver on the right. The left lung, although taller, is narrower than the right one, because the heart tilts to the left and takes up more space on this side of the mediastinum.

The right lung has three lobes: upper, middle, and lower. The left lung, instead, has only one upper and one lower lobe.

Each lung is surrounded by a double layer of a serous membrane, called the pleura. The layer on the surface of the lung is called the visceral pleura and extends into the fissures between the lobes of the lung. At the hilum, the visceral pleura folds in on itself and forms the parietal pleura, which adheres to the mediastinum, the inner surface of the rib cage and the upper face of the diaphragm. The space between the parietal pleura and the visceral pleura is called the pleural cavity and it is filled with a thin layer of slimy pleural fluid. Under pathological conditions this space can fill with air or liquid, separating the membranes and compressing the lungs. The pleura and pleural fluid have three functions:

1. *Reduction of friction*. The pleural fluid acts as a lubricant that allows the lungs to expand and contract with minimal friction.
2. *Creating a pressure gradient*. During inhalation, the rib cage expands and pulls the parietal pleura outwards. The visceral pleura adheres tightly to the parietal pleura and its outward movement expands the lung. The air pressure inside the lung falls under the atmospheric pressure outside the body and, due to this pressure gradient, the outside air enters the lung.

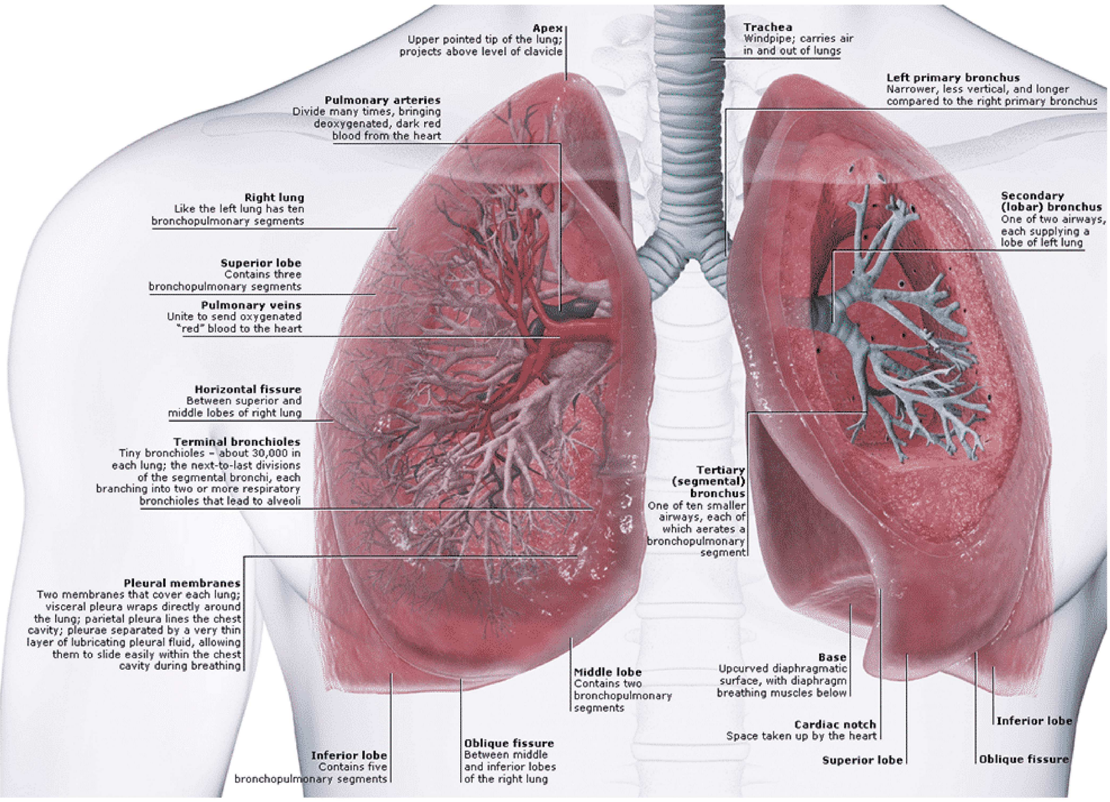

Figure 3: Anatomy of the lungs

3. *Compartmentalization*. The pleura, mediastinum, and pericardium divide the thoracic organs into compartments and prevent the injection from spreading from one organ to neighboring ones.

1.1.6 Lower airways: Alveoli

An alveolus, illustrated in *Figure 4*, is a pouch with a diameter of approximately 0.2-0.5 mm. Approximately 95% of the alveolar surface is covered by pneumocytes (type I), which, being very thin, allow the rapid diffusion of gases between the alveoli and the blood. The remaining 5% of the alveolar surface is covered by round or cuboidal pneumocytes (type II), that have two functions: (1) they repair the alveolar epithelium when the squamous alveolar cells are damaged, and (2) they secrete pulmonary surfactant, a mixture that coats the smaller alveoli and bronchioles and prevents them collapsing when you exhale.

Each alveolus is surrounded by many blood capillaries supplied by the pulmonary artery. The barrier between the air contained in the lungs and the blood is called the respiratory membrane.

1.2 Exchange of gas at the level of alveoli

As described before, the function of the respiratory system is to move two gases: oxygen and carbon dioxide. Gas exchange takes place in the millions of alveoli placed in the lungs together with the capillaries that envelop them, see *Figure 4*. Inhaled oxygen moves from the alveoli to the blood in the capillaries, and carbon dioxide moves from the blood in the capillaries to the air in the alveoli.

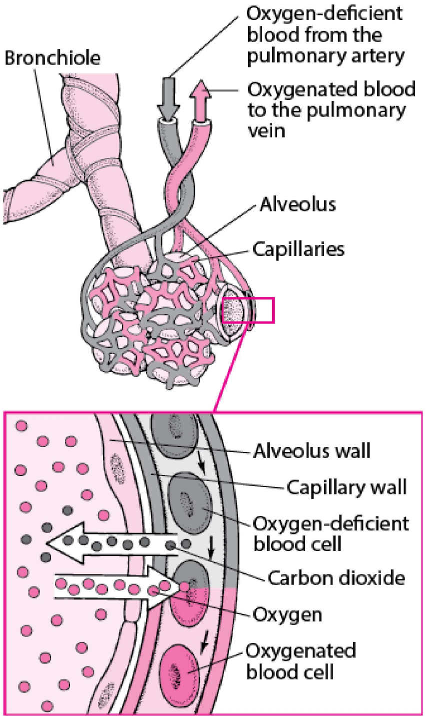

Figure 4: Exchange of gas between alveoli and capillaries

The layers of cells lining the alveoli and the surrounding capillaries are each only one cell thick and are in very close contact with each other. Thus, oxygen passes quickly through this air-blood barrier into the blood in the capillaries. Similarly, carbon dioxide passes from the blood into the alveoli and is then exhaled. Thus, oxygen is constantly being supplied through breathing, but it is also being taken away from the site at which the gas exchange occurs by the capillaries, because of the blood flow. This means that we are constantly maintaining a concentration gradient, which is what allows the gas exchange to continue.

1.2 Mechanics of breathing

Breathing is an essential physiological function which occurs when the contraction or relaxation of muscles around the lungs changes the total volume of air within the air passages inside the lungs. The processes of *inspiration* (breathing in) and *expiration* (breathing out) are vital for providing O_2 to tissues and removing CO_2 from the body. Inspiration occurs via active contraction of muscles, whereas expiration tends to be passive, unless it is forced.

1.2.1 Boyle's law

The mechanics of breathing can be explained in terms of gas laws. Boyle's Law describes the relationship between the pressure (p) and the volume (V) of a gas. This law states that the pressure and the volume of a gas in a closed container are inversely proportional: as the volume of the container decreases, the pressure of the contained gas increases; as the volume increases, the pressure decreases, see *Figure 5*. Mathematically, Boyle's law can be stated this way:

$$pV = k$$

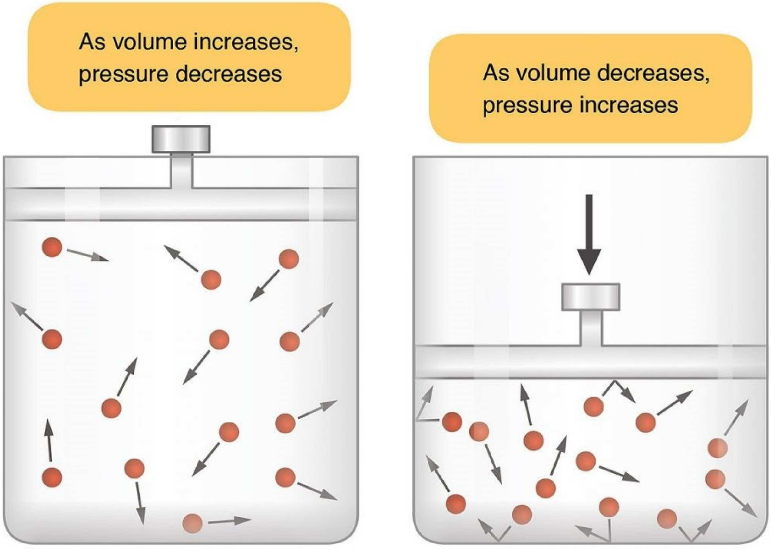

Figure 5: Representation of the Boyle's law

where p is the pressure, V is the volume and k is a constant. Thus, this law states that if the volume in a closed container is changed, the pressure changes together with the volume. The product of the two remains the same before and after the change:

$$p_1V_1 = p_2V_2$$

where p_1 is the initial pressure, V_1 is the initial volume, p_2 is the final pressure and V_2 is the final volume of the closed container.

1.2.2 Inhalation and exhalation

As explained in *Figure 6*, during inhalation, air is drawn in from the surroundings by a change in the volume of the lungs. This change in lung volume is due to two inspiratory muscles: the diaphragm and the external intercostal muscles.

The diaphragm separates the thoracic cavity (also known as the chest cavity) from the abdominal and as it contracts and is pulled down, the ribs expand, increasing the volume of the cavity.

The external intercostal muscles, instead, are placed between the ribs and as they contract, they elevate the ribs and the sternum, thus increasing the chest cavity. As the lungs are held against the inner thoracic wall by the pleural seal, they also undergo an increase in volume.

As explained by Boyle's law, an increase in lung volume results in a decrease in the pressure within the lungs. The pressure of the environment external to the lungs is now greater than that of the environment within the lungs, meaning that air moves into the lungs due to the pressure gradient. Once pressure equilibrates, inhalation ends.

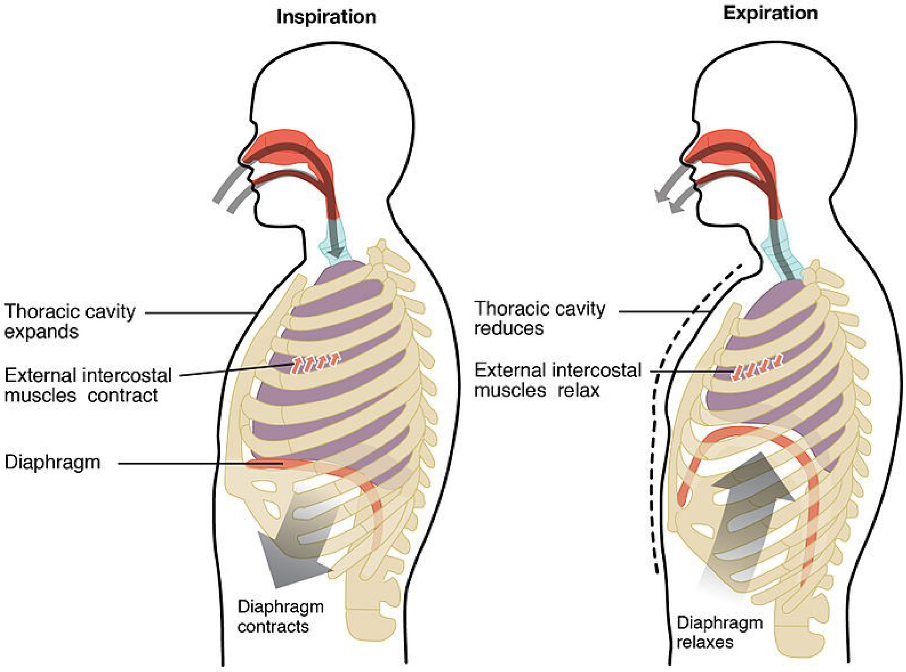

Figure 6: Mechanics of inspiration and expiration

Expiration occurs when the diaphragm and the external intercostal muscles relax, and the volume of the thoracic cavity and of the lungs decreases. Another pressure gradient forms, this time with higher pressure inside of the body (rather than outside of it). Air moves from the lungs out of the mouth or nose and into the surroundings during exhalation. Breathing depends on this back-and-forth pressure gradient caused by volume changes due to contraction and relaxation of the diaphragm.

1.2.3 Respiratory volumes and capacities

As explained before, in the respiratory system changes in the volume of the thoracic cavity during ventilation generate pressure gradients that determine the air flow. Ventilation is the air exchange by mass flow between the atmosphere and the alveoli. The amount of air moved during the ventilation can be divided into four pulmonary volumes, see *Figure 7*:

1) *Tidal volume (TV)*: it is the air volume that moves during a normal breathing act at rest, and it corresponds to about 500 ml.
2) *Inspiratory reserve volume (IRV):* it is the maximum amount of air forcibly inhaled after a normal inspiration.
3) *Expiratory reserve volume (ERV):* it is the maximum amount of air forcibly expired after a normal expiration.
4) *Residual volume (RV):* it is the air that remains into the lungs after a forced expiration.

The respiratory capacities, instead, are obtained summing the respiratory volumes and are four, see *Figure 7*:

1) *Vital capacity (VC):* it is the sum of the tidal volume, the inspiratory reserve volume, and the expiratory reserve volume and represents the maximum air volume that can be voluntarily moved in and out of the respiratory system during a ventilatory act.

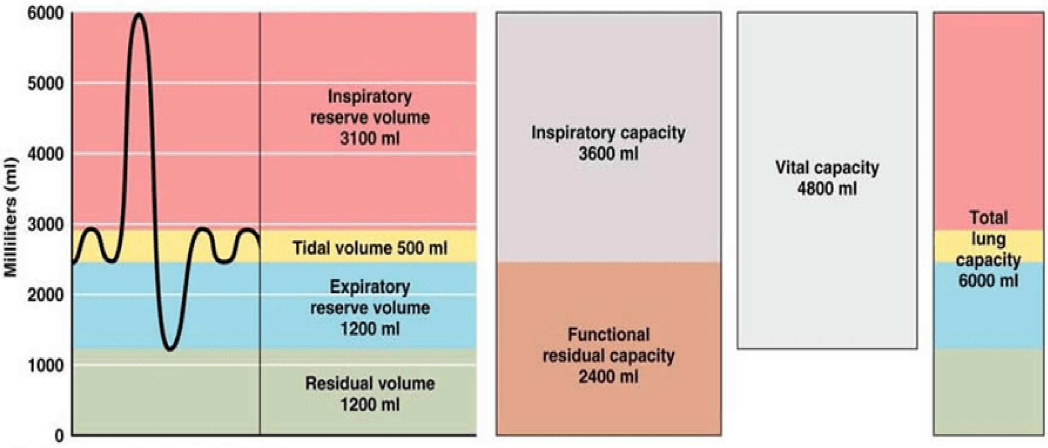

Figure 7: Respiratory volumes and capacities

2) *Total lung capacity (TLC):* it is the sum of the vital capacity and the residual volume.
3) *Inspiratory capacity (IC):* it is the sum of the tidal volume and the inspiratory reserve volume.
4) *Functional residual capacity (FRC):* it is the sum of the expiratory reserve and the residual volume.

2. Lung cancer

Lung cancer (*Figure 8*), also known as lung carcinoma, is a malignant lung tumor derived from transformed, malignant cells that originate as epithelial cells, or from tissues composed of epithelial cells that uncontrollably grow in the tissue of the lung. In time, this uncontrolled growth can spread beyond the lung, either by direct extension, by entering the lymphatic circulation, or via the hematogenous, bloodborne spread, the process called metastasis, into other tissues of the body.

Worldwide, lung cancer is currently the malignant tumor with the highest mortality rate in men and the second highest mortality rate in women, ranking lower only in relation to breast cancer [1]. Despite being more common in male individuals, also its incidence in women has drastically increased in recent decades, probably due to the smoking habit, which has become increasingly common among women [1]. Over the past 20 years, the incidence of lung cancer in women has increased by 134%, making it the second leading cause of death from cancer among females, whereas it has increased by only 57% in men.

Only 20% of the cases are diagnosed in the early stages, in fact, the diagnosis is usually delayed and is made only when the disease is already in the advanced stages, which precludes curative treatment.

2.1 Main causes

Some of the most known causes of lung cancer are:

- ***Smoking***

  Tobacco smoking is by far the main contributor to lung cancer, since cigarette smoke contains a variety of carcinogens. Passive smoking, i.e., the inhalation of smoke from another's

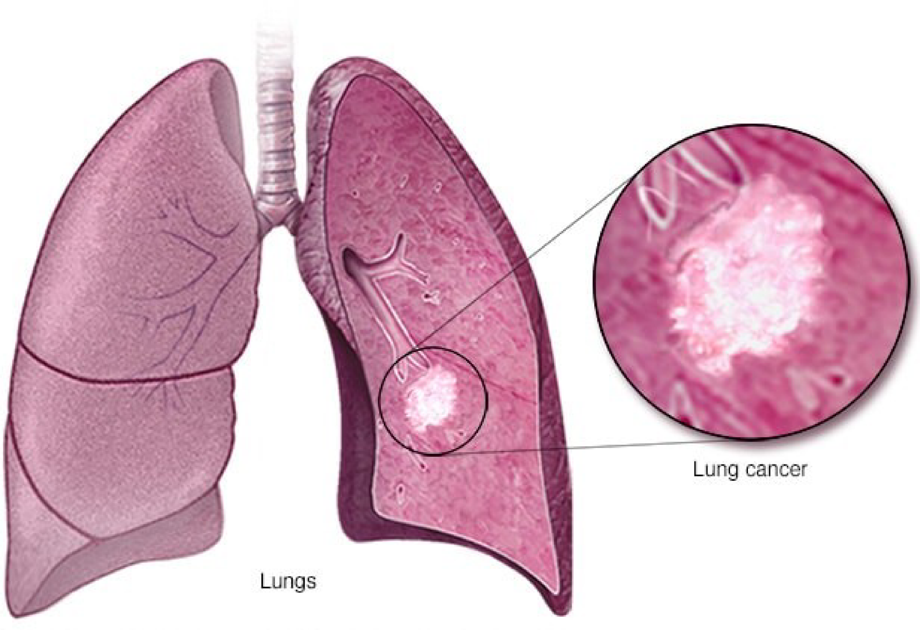

Figure 8: Lung cancer

smoking, is also a cause of lung cancer in non-smokers. However, lung cancer also occurs in people who never smoked and in those who never had prolonged exposure to secondhand smoke.

- ***Radon gas***
  Radon is a colorless and odorless gas generated by the breakdown of radioactive radium, which in turn is the decay product of uranium, found in the Earth's crust. The radiation decay products ionize genetic material, causing mutations that sometimes become cancerous.

- ***Asbestos***
  Asbestos is a naturally occurring fibrous silicate mineral, that has been used for construction and fireproofing, before becoming illegal in many countries. However, many older buildings still contain asbestos, and the inhalation of asbestos fibers can cause a variety of lung diseases such as lung cancer.

- ***Air pollution***
  Outdoor air pollutants, especially chemicals released from the burning of fossil fuels, increase the risk of lung cancer.

- ***Genetics***
  A small amount of lung cancer cases is caused by inherited (genetic) factors. Nevertheless, in relatives of people who are diagnosed with lung cancer, the risk is higher, likely due to a combination of genes.

- ***Other causes***
  Numerous other substances, occupations, and environmental exposures have been linked to lung cancer, such as some metals (aluminum production, cadmium and cadmium compounds, chromium compounds), some products of combustion, ionizing radiations (X-ray and gamma radiations) and some toxic gases.

2.2 Signs and symptoms

Lung cancer typically doesn't cause signs and symptoms in its earliest stages, but they typically occur when the disease is advanced. Signs and symptoms that may suggest lung cancer include:

- *Respiratory symptoms:* coughing, coughing up blood, wheezing, or shortness of breath.

- *Systemic symptoms:* weight loss, weakness, fever, or clubbing of the fingernails.
- *Symptoms due to the cancer mass pressing on adjacent structures:* chest pain, bone pain, superior vena cava obstruction, or difficulty swallowing. If the cancer grows in the airways, it may obstruct airflow causing breathing difficulties.

In many people, the cancer has already spread beyond the original site by the time they have symptoms and seek medical attention.

2.3 Lung cancer histological types

Lung cancer constitutes a group of heterogeneous tumors, with several differentiation types, recognized by the WHO classification of lung tumors. Thus, lung cancers are mainly classified according to their histological type, i.e., the size and appearance of the malignant cells seen by a histopathologist under a microscope.

Lung cancers are traditionally divided into non–small cell carcinoma (NSCC) and small cell carcinoma (small cell lung carcinoma, SCLC), with the former accounting for 80% of the cases and the latter accounting for the remaining 20% [2]. SCLCs behave aggressively and are treated non-surgically in most cases, whereas NSCCs are managed by a combination of surgery and adjuvant therapy. NSCCs are subclassified into adenocarcinoma, squamous cell carcinoma (SSC), and large cell carcinoma (LCC).

2.3.1 Small cell lung cancer

Small cell carcinomas or SCLCs comprise slightly less than 20% of all lung cancers and is a highly aggressive malignancy. Smoking history is present in virtually all cases of SCLC. Patients usually have metastatic disease at the time of presentation. Moreover, most patients relapse within the first 2 years after treatment and the 2-year survival rate is less than 10% in metastatic patients [2]. SCLC is commonly centrally located in the major airway and the tumor cells are small in size compared with other types of lung cancers, hence the name “small” cell lung cancer. WHO classification divides SCLC into 2 subtypes: pure SCLC and combined SCLC containing a component of NSCC.

2.3.2 Non-small cell lung cancer

Non-small cell lung cancers account for the majority of the cases and can be subdivided into adenocarcinoma, squamous cell carcinoma and large cell carcinoma.

2.3.2.1 Adenocarcinoma

Adenocarcinoma is the most common type of lung cancer, accounting for more than 40% of lung cancers, 60% of the NSCC, and more than 70% of surgically resected cases [2], being also the most common form of lung cancer among non-smoker people and ex-smokers.

Lung adenocarcinoma commonly forms a peripherally located mass with central fibrosis and pleural puckering. It can also have a variety of other gross appearances, including centrally located mass, diffuse lobar consolidation, bilateral multi-nodular distribution, and pleural thickening.

Multiple gene alterations can occur in adenocarcinomas, leading to a subclassification into minimally invasive adenocarcinoma (MIA), invasive non-mucinous adenocarcinoma, invasive mucinous adenocarcinoma (IMA), colloid adenocarcinoma, fetal adenocarcinoma, and enteric-type adenocarcinoma [3]. Invasive non-mucinous adenocarcinomas are the most common subtype of lung cancer, whereas colloid, fetal, and enteric-type variants are rare subtypes.

2.3.2.2 Squamous cell carcinoma

Squamous cell carcinomas (SCCs) make up about 20% of lung cancers and usually occurs in the central portion of the lung, along major airways, forming cavities when it achieves a large size. Thus, Survival rate for SCC is significantly better than that of adenocarcinoma [2]. SCCs are further divided into keratinizing, non-keratinizing, and basaloid subtypes.

2.3.2.3 Large cell carcinoma

Large cell carcinomas (LCCs) represent a minority of NSCCs, accounting for less that 3% of the lung cancers. LCC is usually peripherally located, bulky, and necrotic in appearance [2].

2.3.3 Neuroendocrine tumors

Neuroendocrine tumors (NETs) are relatively common lung tumors, accounting for about 20% to 25% of lung cancers. Their common morphologic, immunohistochemical, and ultra-structural features set them apart from other lung tumors.

The 2021 WHO classification categorizes NETs of the lung as a single group of tumors, which includes low- and intermediate-grade typical carcinoid and atypical carcinoid, respectively, and the high-grade neuroendocrine carcinomas (NECs), including large cell neuroendocrine carcinoma (LCNEC) and small cell carcinoma (SCLC). Nevertheless, it is recognized that typical and atypical carcinoids are clinically, epidemiologically, histologically, immunohistochemically, and genetically very different from LCNEC and SCLC [3].

- ***Typical Carcinoid***

  Carcinoid tumors are rare, accounting for 1% to 2% of all lung tumors, even if they are more common in the pediatric population. Indeed, typical carcinoids are different from other types of lung cancers in their presentation at a relatively younger age and more frequent presentation at an earlier stage, as well as good prognosis (more than 90% 5-year survival rate) [2].

- ***Atypical Carcinoid***

  Like typical carcinoids, atypical carcinoids are relatively common in the younger age group compared with other types of lung cancers and are frequently presented as early staged disease [2]. The prognosis of atypical carcinoid is significantly lower than typical carcinoid, with 5-year overall survival rate less than 80%.

- ***SCLC***

  SCLCs comprise slightly more than 10% of all lung cancers and is a highly aggressive malignancy [2], in fact the survival rate is less than 10% in metastatic patients.

- ***LCNEC***

  LCNEC, like SCLC, is associated with heavy smoking history and it is usually peripherally located in the lung. LCNEC is a highly aggressive neuroendocrine carcinoma, and 5-year survival rate is reported close to 30%, significantly worse than other types of NSCC [2].

2.4 Lung cancer diagnosis

Lung cancer may be seen on chest radiographs and computed tomography (CT) scans. The diagnosis is confirmed by biopsy, which is usually performed by bronchoscopy or CT-guidance.

Performing a chest radiograph (x-ray) is one of the first investigative steps if a person reports symptoms suggesting the presence of lung cancer. Computed tomography (CT) imaging of the chest is often used for diagnosis and may reveal a spiculated mass which is highly suggestive of lung cancer. CT imaging is also used to provide more information about the type and extent of disease.

The definitive diagnosis of lung cancer is based on the histological examination of the suspicious tissue in the context of the clinical and radiological features, called immunohistochemistry (IHC). The significance of applying IHC includes aiding effective and accurate classification of tumors, minimizing potential diagnostic errors, improving delineation of tumor types suitable for molecular testing, and utilizing lineage-specific markers for the exclusion of metastatic origin of the tumors.

2.5 Therapies

Treatment for lung cancer depends on the cancer's specific cell type, how far it has spread, and the person's performance status. Common treatments include:

- ***Surgery***

  When the tumor is still localized, the treatment for lung cancers is surgery and, in such cases, chances for cure are high. In most cases of early-stage NSCLC, removal of a lobe of lung (lobectomy) is the surgical treatment of choice. Rarely, removal of a whole lung (pneumonectomy) is performed. Surgery might improve outcomes when added to chemotherapy and radiation therapies in early-stage SCLC.

- ***Radiotherapy***

  Radiotherapy is often given together with chemotherapy and may be used with curative intent in people with NSCLC that are not eligible for surgery. For potentially curable SCLC cases treated with surgery, post-operative chest radiotherapy is recommended.

  Recent improvements in targeting and imaging have led to the development of "stereotactic radiation" in the treatment of early-stage lung cancer. In this form of radiotherapy, high doses are delivered over several sessions using stereotactic targeting techniques. For both NSCLC and SCLC patients, smaller doses of radiation to the chest may be used for symptom control (palliative radiotherapy).

- ***Chemotherapy***

  The chemotherapy regimen depends on the tumor type. SCLC is usually treated primarily with chemotherapy and radiation. In advanced NSCLC, chemotherapy improves survival and is used as first-line treatment. Adjuvant chemotherapy refers to the use of chemotherapy after apparently curative surgery to improve the outcome. Moreover, chemotherapy before surgery in NSCLC that can be removed surgically may improve outcomes.

- ***Targeted and immunotherapy***

  Several drugs that target molecular pathways in lung cancer are available, especially for the treatment of advanced disease and may be used for both SCLC and NSCLC.

- ***Bronchoscopy***

  Several treatments can be provided via bronchoscopy for the management of airway obstruction or bleeding, i.e., if an airway becomes obstructed by cancer growth.

- ***Palliative care***

  Palliative care, when added to usual cancer care, benefits people even when they are still receiving chemotherapy.

- ***Noninvasive interventions***

  The most effective intervention for avoiding death from lung cancer is to stop smoking. Some weak evidence suggests that certain supportive care interventions (non-invasive) that focus on well-being for people with lung cancer may improve quality of life [4]. Exercise training may benefit people with NSCLC who are recovering from lung surgery [5]. In addition, exercise training may benefit people with NSCLC who have received radiotherapy, chemotherapy, chemoradiotherapy, or palliative care [6].

2.6 Prevention

There's no sure way to prevent lung cancer, but there are some prevention factors that may reduce the risk:

1. ***Tobacco control***

   Eliminating tobacco smoking is a primary goal in the prevention of lung cancer, and smoking cessation is an important preventive tool in this process.

2. ***Screening***

   For individuals with high risk of developing lung cancer, computed tomography (CT) screening can detect cancer and give a person some options to respond to it in a way that prolongs life, this the chance of death from lung cancer.

3. ***Other prevention strategies***

   Vitamin C supplementation might reduce the risk of lung cancer. However, several rigorous studies have not demonstrated a clear association between diet and lung cancer risk, although meta-analysis that accounts for smoking status may show benefit from a healthy diet.

3. Diagnostic imaging modalities for lung cancer

3.1 Lung cancer screening

Despite advances in clinical care and diagnostic imaging, most lung cancer patients present with advanced-stage disease, for which a cure remains elusive, and the prognosis is therefore generally poor. However, low-dose computed tomography (LDCT) has recently been shown to result in the detection of earlier-stage lung cancers [7]. LDCT produces images of sufficient quality to detect many abnormalities, while involving 10–30% lower radiation doses that does the standard CT examination [8].

The effectiveness of various lung cancer screening programs in high-risk patients has been assessed in multiple studies in the last decade. Some of these programs were based on chest radiography (CXR), while others on low dose computed tomography (LDCT) [8]. In 2011, a high-quality trial, the National Lung Screening Trial (NLST), compared LDCT to CXR in a large sample of high-risk adults and showed a 20% relative reduction in lung cancer mortality for LDCT [9]. In September 2018, new data from the largest European trial (NELSON) showed an even bigger reduction in deaths from lung cancer than was seen in NLST [9].

Lung cancer screening has many benefits:

- Because CT scans can detect even very small nodules in the lungs, LDCT of the chest is especially effective for diagnosing lung cancer at its earliest, most treatable stage.
- CT is painless, noninvasive, and fast, which is important for patients who have trouble holding their breath.
- No radiation remains in a patient's body after a CT exam.
- X-rays used in LDCT of the chest have no immediate side effects and do not affect any metal parts in your body, such as pacemakers or artificial joints.
- LDCT scans of the chest produce images of high enough quality to detect many abnormalities while using up to 90% less ionizing radiation than a conventional chest CT scan.
- Studies prove that lung cancer screening with LDCT reduces the number of deaths from lung cancer in patients at high risk.

On the other hand, however, lung cancer screening has also some risks:

- False positive results may occur when a test appears to be abnormal, but no lung cancer is found. Those abnormal findings may require additional testing to determine whether cancer is present, thus involving risks and possibly causing patient anxiety.
- Test results that appear to be normal even when lung cancer is present are called false-negative results. A person who receives a false-negative test result may delay seeking medical care.

- Not all cancers detected by LDCT will be found in the early stage of the disease.
- Overdiagnosis occurs when screening identifies histologically confirmed lung cancer that would not have resulted in a patient's death if left untreated. Potential harmful effects of overdiagnosis include the psychological stress that accompanies a diagnosis of cancer as well as the morbidity and mortality that may accompany unnecessary medical procedures.
- Despite the use of a low-dose screening technique, the risk of radiation-induced malignancy is still a consideration. Furthermore, the risk of developing radiation-induced lung cancer may be increased in patients who smoke [7].

3.2 Screening techniques

The most common screening techniques for lung cancer detection are computed tomography (CT) and positron emission tomography/computed tomography (PET/CT).

3.2.1 Computed tomography

Computed Tomography was formally introduced in 1972 by a British engineer, Sir Godfrey Hounsfield, and it was a revolutionary invention in terms of medical imaging, because, for the first time, physicians were able to obtain high-quality tomographic (cross-sectional) images of internal structures of the body. With the first CT scanner, the projection data were acquired in approximately 5 minutes and the tomographic image was reconstructed in approximately 20 min. Since then, CT technology has developed dramatically and nowadays projection data are typically acquired in approximately 1 second, or even less, while the image is reconstructed in 3 to 5 sec [10].

The fundamental task of CT systems is to make an extremely large number of highly accurate measurements of x-ray transmission through the patient in a precisely controlled geometry. The final image of a CT, thus, is made by viewing the patient via X-ray imaging from numerous angles, by mathematically reconstructing the detailed structures and displaying the reconstructed image on a video monitor [11]. The X-ray source and detectors are mounted opposite to each other in a rigid gantry with the patient lying in between on a motorized couch, that is moved into the aperture of the gantry, see *Figure 9*.

Nowadays, there are mainly six generations of CT scanners which differ on the motion of the gantry during the acquisition and on the type of X-ray source used. These generations are:

1. ***First Generation: Parallel-Beam Geometry***

   First-generation CT systems are characterized by a single X-ray source, called pencil beam, directing across the object and a single detector. Both the source and the detector translate

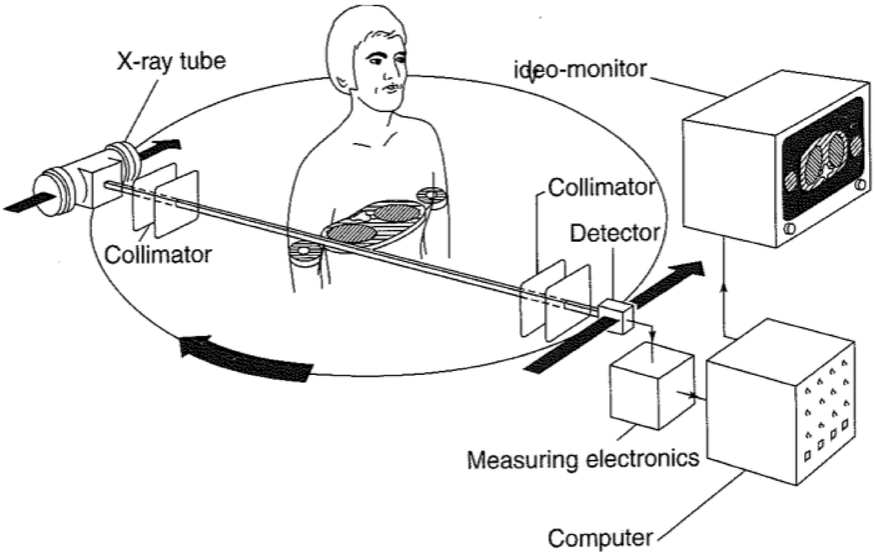

Figure 9: CT system

simultaneously in a scan plane for a given number of angular rotations [12]. The main disadvantage of this generation are long times for acquisition.

2. ***Second Generation: Fan Beam, Multiple Detectors***

   Second-generation CT systems use the same translate/rotate scan geometry as the first generation, with the difference that the pencil beam is replaced by a fan beam and the single detector by multiple detectors [12]. In this way, a series of views can be acquired during each translation, thus reducing the scanning times.

3. ***Third Generation: Fan Beam, Rotating Detectors***

   In the third generation of CT scanners a fan beam of X-rays is rotated 360° around the isocenter without translation [10]. These rotate-only motion allows to drastically reduce the scanning time, which is now as little as 1 second.

4. ***Fourth Generation: Fan Beam, Fixed Detectors***

   With the fourth generation CT scanners, the X-ray source and the fan beam rotate about the isocenter, while the detector array remains stationary, with scan times similar to those of third-generation scanners [10].

5. ***Fifth Generation: Scanning Electron Beam***

   In the fifth-generation CT systems there is no mechanical motion involved. Thus, the detector array remains stationary, while a high-energy electron beam is electronically moved along a semicircular tungsten strip anode. X-rays are produced at the point where the electron beam hits the anode, resulting in a source of x-rays that rotates about the patient with no moving parts [10]. This configuration further reduces the scanning times, which are around 50 msec.

6. ***Spiral/Helical Scanning***

   With this technology, multiple images are acquired while the patient is translated through the gantry in a smooth continuous motion rather than stopping for each image [10] and the result is a helical profile instead of a circular one. By avoiding the time required to translate the patient table, the total scan time required to image the patient can be much shorter.

3.2.2 Positron emission tomography/computed tomography

Positron emission tomography (PET) is an imaging modality for obtaining in vivo cross-sectional images of positron-emitting isotopes that demonstrate biological function, physiology, or pathology. Unlike anatomical imaging techniques like CT, X-ray, and ultrasound, PET imaging provides “functional” information about the human body. In this technique, a chemical compound with the desired biological activity is labelled with a radioactive isotope that decays by emitting a positron. The emitted positron almost immediately combines with an electron and the two are mutually annihilated with the emission of two gamma rays. The two gamma ray photons travel in almost opposite directions, penetrate the surrounding tissue, and recorded outside the subject by a circular array of detectors. A mathematical algorithm applied by a computer rapidly reconstructs the spatial distribution of the radioactivity within the subject for a selected plane and displays the resulting image in the monitor.

During the 1990s there was interest in combining (fusing) different imaging modalities to obtain a more complete picture of disease. One of the most common combinations of imaging modalities is the PET/CT. Such a device provides a medical imaging department with the capability to acquire accurately aligned anatomic and functional images for a patient from a single scanning session.

The PET/CT scanner, by combining two established modalities such as CT and PET, is an evolution in imaging technology, integrating two existing technologies that have historically progressed along separate but parallel paths. The two modalities are complementary, with CT images lacking the functional specificity of PET and PET images lacking the anatomic detail seen on CT.

The improved PET imaging capability and the use of the CT provide the attenuation correction factors for PET in a few seconds instead of the many minutes required in PET-only scanners, resulted in greatly increased patient throughput compared with PET-only scanners.

4. Deep learning for automatic recognition of lung cancer: Theory

Deep learning (DL), that is a subfield of machine learning, which in turn is a field within artificial intelligence (AI), has become very popular in everyday life and can be used for many applications. In general, DL consists of massive multilayer networks of artificial neurons that can automatically discover useful features, that are needed for many tasks such as detection and classification, given large amounts of unlabeled or labeled data [13]. Nowadays, deep learning algorithms are commonly used for analyzing medical images and can be applied in many fields, such as radiology, oncology, neurology, musculoskeletal imaging, digital pathology, cardiology, and dentistry. [14]

4.1 Neural networks

Neural networks are a type of learning algorithm which forms the basis of most deep learning methods. Development of Neural Networks (NNs) dates back to the early 1940s and experienced an upsurge in popularity in the late 1980s. Much of the inspiration for the field of NNs came from the desire to produce artificial systems capable of sophisticated, perhaps intelligent, computations similar to those that the human brain routinely performs. For this reason, NNs are inspired by the neural network of our brain, that is composed by billions of interconnected neurons, which are cells that use biochemical reactions to receive, process and transmit information. Artificial neurons are arranged in layers and each artificial neuron is usually a simple processing unit which takes one or more inputs and produces an output. At each artificial neuron, every input has an associated weight which modifies the strength of each input. The artificial neuron simply adds together all the inputs and calculates an output to be passed on. Layers in between the input and output are often referred to as 'hidden' layers. When a neural network contains multiple hidden layers, it is typically considered a 'deep' neural network, hence the term 'deep learning' [15].

NNs learn from examples and exhibit some capability for generalization beyond the training data. Thus, there isn't the need of an explicit description of the problem, neither the need for a programmer. After sufficient training the NN is able to relate the problem data to the solutions, inputs to outputs, and it is then able to offer a viable solution to a brand-new problem.

4.1.1 Learning methods

- ***Supervised learning***

  In supervised training, both the inputs and the outputs are provided. The network then processes the inputs and compares its resulting outputs against the desired outputs, computing the error generated by the difference between the two outputs. Errors are then propagated back through the system, causing the system to adjust the weights which control the network in order to reduce this error. This process occurs over and over as the weights are continually tweaked.

- ***Unsupervised learning***
  In unsupervised training, the network is provided with inputs but not with desired outputs. The system itself must then decide what features it will use to group the input data. This is often referred to as self-organization or adaption.

The most common form of machine learning, deep or not, is supervised learning [16].

4.1.2 Supervised learning: Dataset split and truth labels for training, validating, and testing

Data and ground truth labels are the most important components in research applying deep learning or other machine learning methods. Indeed, careful collection of data and ground truth labels with which to train and test a model is mandatory for a successful deep learning project; however, obtaining high-quality labeled data can be costly and time-consuming.

Data are typically split into three sets: a training, a validation, and a test set, see *Figure 10.* A training set is used to train a network, where loss values are calculated via forward propagation and learnable parameters are updated via backpropagation. A validation set is used to evaluate the model during the training process. A test set is ideally used only once at the very end of the project in order to evaluate the performance of the final model [17].

For training data, augmentation is sometimes performed to increase the amount of data for training and to reduce the risk of the overfitting problem, i.e., when the system models the training data too well and thus fail in modeling other datasets. Even if the same patient is examined, images would not be identical between examinations due to slight differences in positioning. Therefore, use of rotated or parallel-shifted images would get the model robust for slight differences in patient positions. Mirrored images are also sometimes utilized.

Care should be paid to the data volume of input images. Use of input data with large data volume would be associated with some computation problems and computers with large memories would be necessary. Training of such large CNNs requires more calculations and longer times. To solve these problems, cropped images from the original image and/or resized images are utilized for much research [18].

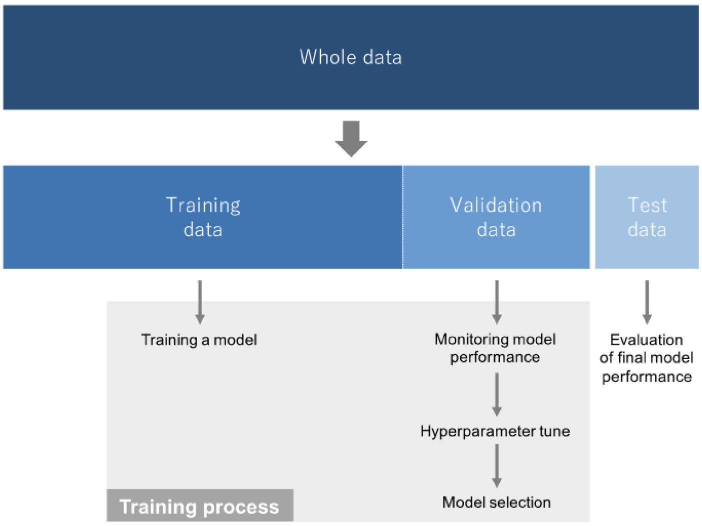

Figure 10: Division of the whole data into training, validation, and test data [17]

4.2 Convolutional neural network

The most successful type of models for image analysis to date are convolutional neural networks (CNNs). CNN is designed to learn spatial hierarchies of features automatically and adaptively through backpropagation by using multiple building blocks, such as convolution layers, pooling layers, and fully connected layers, see *Figure 11*. The first two, convolution and pooling layers, perform feature extraction, whereas the third, a fully connected layer, maps the extracted features into final output [17].

Let's analyze more in details the building blocks of the CNNs:

- ***Convolution layer***

  A convolution (Conv) layer is a component of the CNN architecture that performs feature extraction, which typically consists of a combination of linear and nonlinear operations, i.e., convolution operation and activation function [17].

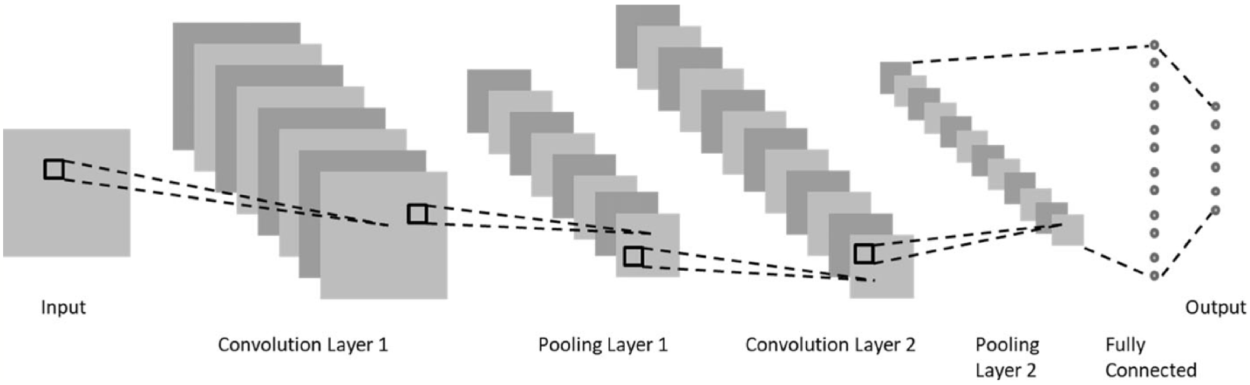

Figure 11: CNN with two convolution layer follower by a pooling layer and one fully connected layer [13].

Through the convolution, a small array of numbers, called a kernel, is applied across the input, which is an array of numbers, called a tensor. An element-wise product between each element of the kernel and the input tensor is calculated at each location of the tensor and summed to obtain the output value in the corresponding position of the output tensor, called a feature map, see *Figure 12*. This procedure is repeated applying multiple kernels to form an arbitrary number of feature maps. [17]

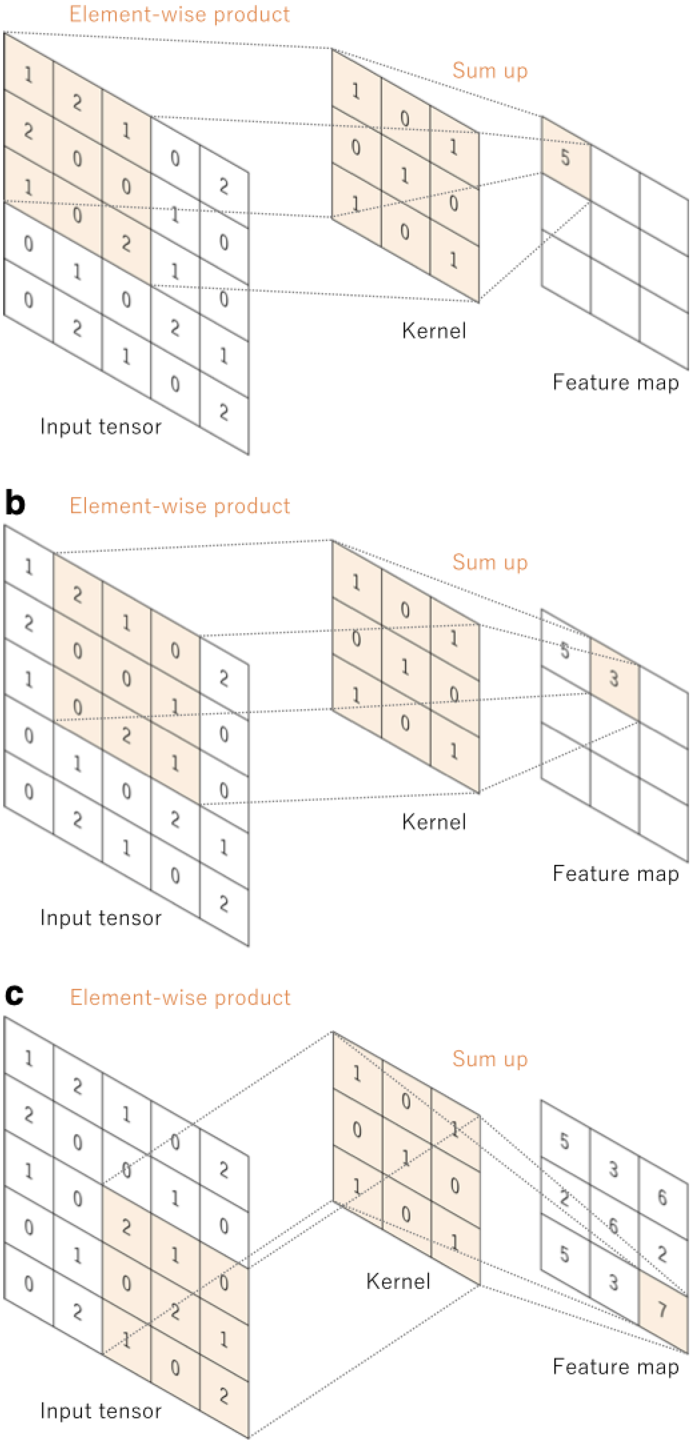

Figure 12: An example of convolution operation with kernel size of 3x3 [17].

The outputs of a linear operation such as convolution are then passed through a nonlinear activation function. The most common nonlinear activation function used presently is the rectified linear unit (ReLU), which simply computes the function: $f(x) = \max(0, x)$ [17].

- ***Pooling layer***

  The role of the pooling layer is to merge semantically similar features into one. A typical pooling unit computes the maximum of a local patch of units in one feature map. Neighboring pooling units take input from patches that are shifted by more than one row or column, thereby reducing the dimension of the representation, and creating an invariance to small shifts and distortions [16], as represented in *Figure 13*. The most popular form of pooling operation is max pooling, which extracts patches from the input feature maps, outputs the maximum value in each patch, and discards all the other values [17].

- ***Fully connected layer***

  The output feature maps of the final convolution or pooling layer is typically transformed into a one-dimensional (1D) array of numbers (or vector) and connected to one or more fully connected (FC) layers, in which every input is connected to every output by a learnable weight. The final fully connected layer typically has the same number of output nodes as the number of classes [17].

Training a network, *Figure 14*, is the process of finding kernels in convolution layers and weights in fully connected layers which minimize differences between output predictions and given ground truth labels on a training dataset. Backpropagation algorithm is the method commonly used for training convolutional neural networks where loss function and gradient descent optimization algorithm play essential roles.

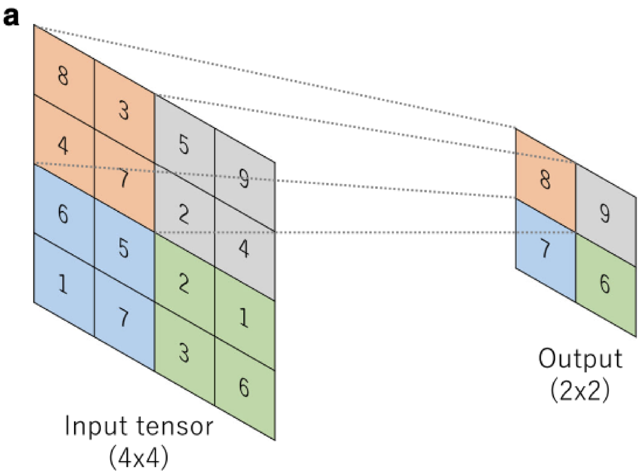

Figure 13: An example of max pooling operation with a filter size of 2x2 (17)

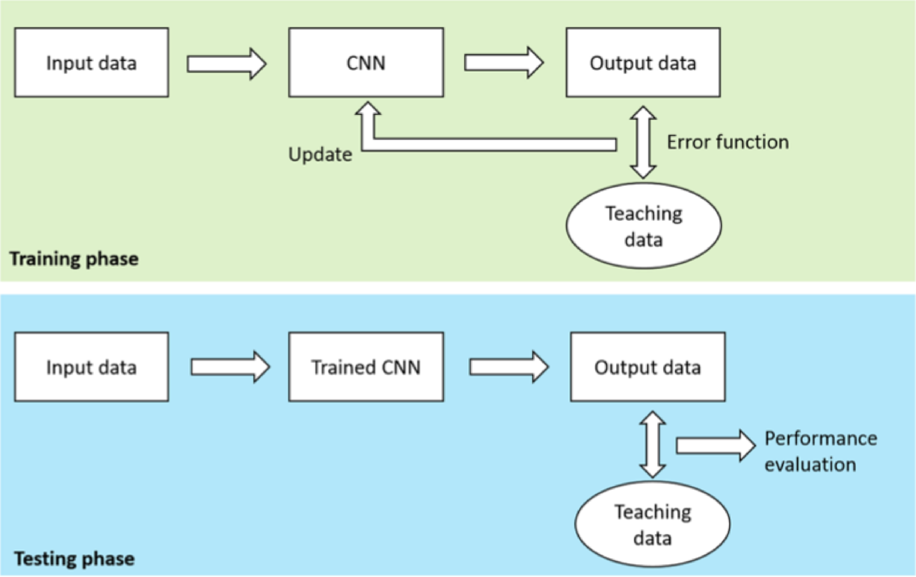

Figure 14: Overview of typical deep learning process [18]

A loss function, also referred to as a cost function, measures the compatibility between output predictions of the network through forward propagation and given ground truth labels.

Gradient descent, instead, is commonly used as an optimization algorithm that iteratively updates the learnable parameters, i.e., kernels and weights, of the network so as to minimize the loss [17].

4.2.1 Main applications of convolutional neural networks

CNNs, in particular those used for image analysis, may be applied for performing many tasks, such as:

- ***Image reconstruction***

  Tomographic reconstruction represents an important class of inverse problems in which externally measured data are linked to internal structures in a complicated way and processed to reconstruct internal features in cross sections or volumetrically [14].

- ***Image segmentation***

  Image segmentation is often an essential step within an overall image analysis pipeline and is used, for example, to identify regions of interest representing normal structures, diseases, or imaging artifacts [14]. The task of segmentation is typically defined as identifying the set of voxels which make up either the contour or the interior of the object(s) of interest [15]. Training data for the segmentation system consist of the medical images containing the organ or structure of interest and the segmentation result [17]. *Figure 15* reports an example of image segmentation.

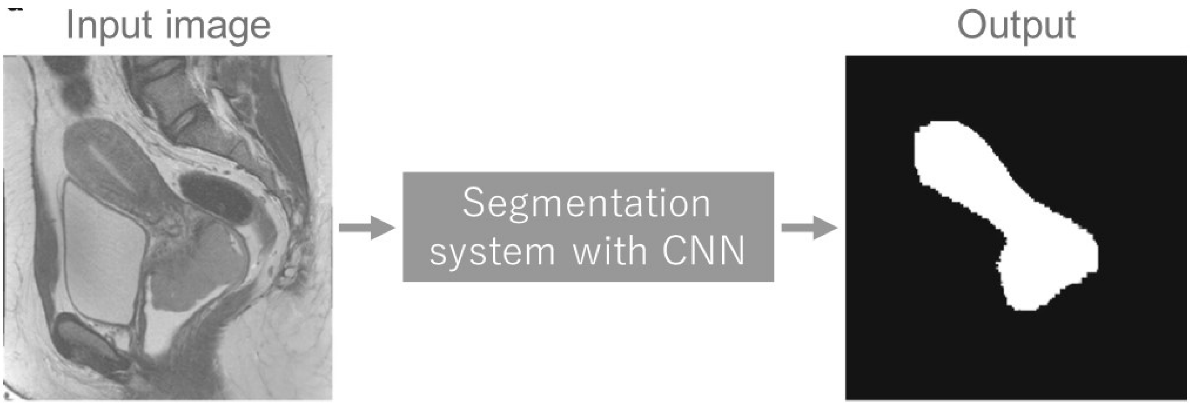

Figure 15: Segmentation system with CNN [17]

- ***Image registration***

  Image registration is the process of transforming different images into the same coordinate system with matched imaging contents. Depending on the medical purposes, the images to be registered may be acquired for the same subject using various modalities or in the same modality but from different subjects or the same subject at a different time. Registration may also be performed on images acquired over time for time series analysis or for longitudinal studies [14].

- ***Detection***

  Detection of organs or lesions in medical images is often an important application for CNNs. The detection of anatomical structures is closely related to that of segmentation of these structures. Deep-learning-based detection methods can generally be divided into classification-based and regression-based methods. The former methods identify structures of interest on a by-image (or image-patch) basis, while the latter provide more detailed information such as the coordinates of a lesion center [14].

- ***Classification***

  Image or exam classification was one of the first areas in which deep learning made a major contribution to medical image analysis. In exam classification, one typically has one or multiple images (an exam) as input with a single diagnostic variable as output (e.g., disease present or not). The application areas of these methods are very diverse, ranging from brain MRI to retinal imaging and digital pathology to lung computed tomography (CT).

4.2.2 Convolutional neural network for planar data classification

Two-dimensional image classification plays an important role in computer vision and can be applied in many fields, such as that of medical image classification. In particular, CNNs have demonstrated high performances on image classification. In recent years a variety of architectures have been proposed for classifying 2D images. Some of the most famous are:

- ***LeNet-5***

  LeCun et al. [19] proposed the first convolutional neural network called LeNet-5 in 1989. They developed a multi-layer artificial neural network which can classify handwritten numbers with the backpropagation algorithm. The handwritten digit-recognition application was chosen because it is a relatively simple machine vision task: the input consists of black or white pixels, the digits are usually well-separated from the background, and there are only ten output categories.

  Acquisition, binarization, location of the zip code, and preliminary segmentation were performed by Postal Service contractors. Then, the input data were resized to fit in a 16 x 16-pixel image. The network has four hidden layers respectively named HI, H2, H3, and H4. Layers HI and H3 are shared weights feature extractors, while H2 and H4 are averaging/subsampling layers.

  After 30 training passes the error rate on training set was 1.1%, while on the whole test set the error rate was 3.4%. However, due to the lack of large training data and computing power at that time, LeNet-5 cannot perform well on more complex problems, such as large-scale image and video classification.

- ***AlexNet***

  In 2012, Krizhevsky et al. [20] built a deep convolutional neural network called AlexNet and trained to classify the 1.2 million high-resolution images in the ImageNet database, which consists of over 15 million labeled high-resolution images in over 22,000 categories.

  The neural network, which has 60 million parameters and 650,000 neurons, consists of five convolutional layers, some of which are followed by max-pooling layers, and three fully connected layers with a final 1000-way SoftMax. The authors highlighted that removing any convolutional layer resulted in inferior performance.

  To reduce overfitting in the fully connected layers the authors employed two methods: data augmentation and “dropout”. Data augmentation is the easiest and most common method to reduce overfitting on image data by artificially enlarging the dataset using label-preserving transformations. The first form of data augmentation used in this CNN consists of generating image translations and horizontal reflections. The second form consists of altering the intensities of the RGB channels in training images. The recently introduced technique, called “dropout”, instead, consists of setting to zero the output of each hidden neuron with probability 0.5. The neurons which are “dropped out” in this way do not contribute to the forward pass and do not participate in back- propagation. The dropout has been used in the first two fully connected layers

Two numbers are usually reported: the top-1 accuracy rate, which compares the ground truth against the first predicted class, and the top-5 error rate, which compares the ground truth against the first 5 predicted classes: an image is deemed correctly classified if the ground truth is among the top-5, regardless of its rank in them. On the test data, this newly CNN achieved top-1 and top-5 error rates of 37.5% and 17.0% which is considerably better than the previous state-of-the-art at that time.

- ***VGGNet***

  In 2015 Simonyan et al. [21] proposed a very deep convolutional network (up to 19 weight layers) for large-scale image classification, which consists of running a model over several rescaled versions of a test image. It was demonstrated that the representation depth is beneficial for the classification accuracy, and that state-of-the-art performance on the ImageNet challenge dataset can be achieved using a conventional convolutional network architecture. The best single-network performance they achieved on the validation set was 24.8%/7.5% top-1/top-5 error.

  A peculiarity of the VGGNet proposed by Simonyan et al. is that the filter used in the convolutional layers were with a very small receptive field (i.e., 3 x 3). Moreover, spatial pooling is carried out by five max-pooling layers, which follow some of the convolutional layers. A stack of convolutional layers is followed by three fully connected layers. All hidden layers are equipped with the rectification (ReLU) non-linearity.

- ***GoogLeNet***

  In 2014 Szegedy et al. [22] proposed a deep convolutional neural network architecture codenamed Inception that achieves the new state of the art for classification and detection in the ImageNet Large-Scale Visual Recognition Challenge 2014 (ILSVRC14). The main idea of the Inception architecture is to consider how an optimal local sparse structure of a convolutional vision network can be approximated and covered by readily available dense components.

The main hallmark of this architecture is the improved utilization of the computing resources inside the network. By a carefully crafted design, we increased the depth and width of the network while keeping the computational budget constant. This newly architecture is called GoogLeNet, a 22 layers deep network, the quality of which is assessed in the context of classification and detection.

The final submission to the challenge obtains a top-5 error of 6.67% on both the validation and testing data, ranking the first among other participants. This is a 56.5% relative reduction compared to the SuperVision approach in 2012, and about 40% relative reduction compared to the previous year's best approach.

The obtained results yield solid evidence that approximating the expected optimal sparse structure by readily available dense building blocks is a viable method for improving neural networks for computer vision. The main advantage of this method is a significant quality gain at a modest increase of computational requirements compared to shallower and narrower architectures.

4.2.3 Convolutional neural network for volumetric data classification

Techniques for analyzing 3D shapes are becoming increasingly important due to the vast number of sensors that are capturing 3D data, as well as numerous computer graphics applications [23], even if they require a huge computational power. This high-cost approach forces to reduce the volume resolutions when applying 3D CNN on volumetric data [24].

In recent years a variety of deep architectures have been approached for classifying 3D shapes. These are:

- ***Multiview CNN (MVCNN)***

  The MVCNN architecture uses rendered images of the model from different views as input, as illustrated in *Figure 16*. Each image is fed into a CNN with shared weights. Then, a max-pooling layer across different views is used to perform an orderless aggregation of the individual representations followed by several non-linear layers for classification. [23]

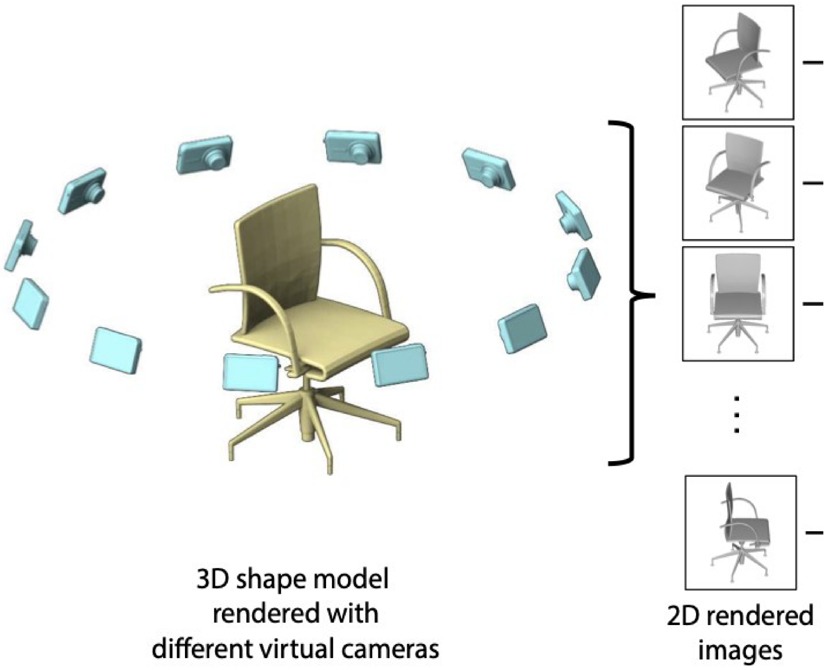

Figure 16: Multiple views obtained from a chair [25]

- ***Voxel-based CNN***

  This approach analyzes shapes represented as a 3D occupancy grid [23], as can be seen in *Figure 17*. An example of voxel-based CNN is the VoxNet, that uses convolution and pooling layers defined on 3D voxel grids [23]. Another one is NormalNet, a voxel-based CNN with normal vectors as input [26]. The performance of normal vector input should not be worse than voxel input, as the former contains more information (position and orientation of the model surface). Wang et al. [26], in fact, demonstrated that under the same network architecture, the normal vector input outperforms the voxel input consistently.

- ***Point-based CNN***

  This approach classifies shapes represented as collection of points [23]. The architecture applies a series of non-linear mappings individually to each input point and performs orderless aggregations using max-pooling operations. Despite its simplicity the model and its variants have been shown to be effective at shape classification and segmentation tasks [23].

- ***Octree CNN***

  An octree is a 3D grid structure with adaptive cell size, which allows for lossless reduction of memory consumption compared to a regular voxel grid [27]. A function defined on a voxel grid can be converted into a function defined on an octree starting from a single cell representing the entire space and recursively partitioning cells into eight octants. If every

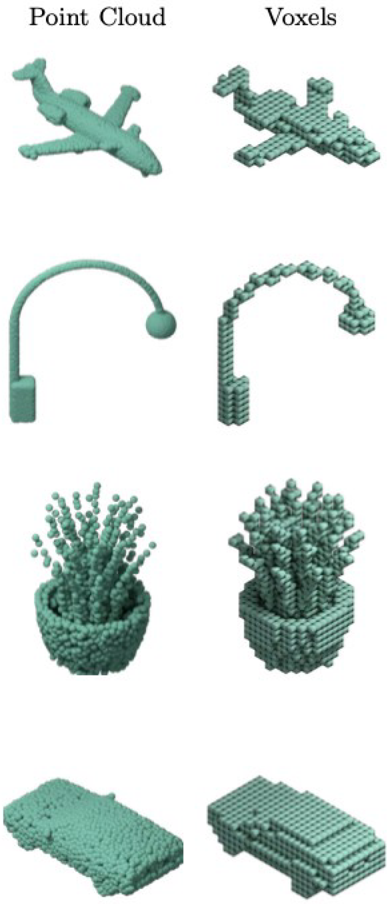

Figure 17: Example of objects represented with point clouds or voxels [23]

voxel within a cell has the same function value, this cell is not subdivided and becomes a leaf of the tree. Octree is becoming popular in CNN-based 3D shape analysis because of its compatibility in obtaining fine detail of a 3D model [24].

Su et al. [23] compared the first three CNNs and demonstrated that multiview methods provide better generalizability and outperform other methods. Moreover, point-based networks are more robust to point perturbations, while multiview and voxel-based networks can be fooled by imperceptible perturbations.

5. Deep learning for automatic recognition of lung cancer: Literature review

As mentioned earlier, computed tomography (CT) is the most effective method of detecting lung nodules due to its ability to extract three-dimensional (3D) images of the chest, resulting in greater resolution of nodules and tumor pathology. Given the complexity of lung cancer classification and the limitations of current practices, there is a need for innovative clinical data assessment tools to augment the biopsy and help better describe disease characteristics.

Some attempts have recently been made to use deep learning (DL) techniques and especially CNNs, since, unlike other feature learning methods that build data representation models in an unsupervised manner, CNNs learn features and train an ANN classifier at the same time, by minimizing the classification error [28].

Additionally, PET/CT is an imaging technique that provides both metabolic and anatomical information; therefore, it is also useful for the early detection of lung cancer [29]. In particular, PET/CT imaging using ^{18}F-fluorodeoxyglucose (FDG) is an established method for the staging of patients with lung cancer. Some FDG-related features that can be analyzed in order to establish the presence of lung cancer are the standard uptake value (SUV), i.e., the amount of tracer uptake inside the nodules, that has been demonstrated good quantitative repeatability in lung nodules, and the spatial distribution of FDG, which also contains important information since the uptake of FDG is not homogeneously distributed within the lesions [30].

In the following sections, some methods for classifying lung cancer from CT or from PET/CT images are reported, both for planar (two-dimensional) and volumetric (three-dimensional) data.

5.1 Methods for planar data classification

5.1.1 Computed tomography data

In this section, many different studies focused on planar CT data classification are reported. *Table 1* summarizes and compares the most important aspects of the studies.

5.1.1.1 Li et al. (2014)

In this study, Li et al. [31] proposed a customized CNN with shallow convolution layer to classify lung image patches with interstitial lung disease (ILD). Moreover, the authors designed a fully automatic neural-based machine learning framework to extract discriminative features from training samples and perform classification at the same time, instead of defining a set of features manually.

The proposed CNN is composed by a convolutional layer, a max pooling layer and three fully connected layers. In particular, the first layer is a convolutional layer with kernel size of 7 × 7 pixels and 16 output channels; the second layer is a max pooling layer with 2×2 kernel size; the following three layers are fully connected neural layers with 100-50-5 neurons in each layer.

The publicly available ILD database is used for evaluation. The database contains 113 sets of High-Resolution Computed Tomography (HRCT) images, with 2062 2D regions of interest (ROI) annotated indicting the ILD category. The authors chose to classify image patches of five ILD categories: normal (N), emphysema (E), ground glass (G), fibrosis (F) and micronodules (M). The CT slices were divided into half-overlapping image patches of 32 × 32 pixels.

The proposed method is capable of extracting discriminative features automatically without manual feature design and achieving good benchmark performance. In particular, in this study the precision obtained ranges from 50% for the detection of emphysema to 95% for the detection of micronodules.

5.1.1.2 Anthimopoulos et al. (2016)

In this work, Anthimopoulos et al. [28] proposed a deep CNN to classify CT image patches into 7 classes, including 6 different ILD (interstitial lung disease) patterns and healthy tissue. Their proposed network consists of 5 convolutional layers with 2 x 2 kernels and LeakyReLU activations, followed by just one average pooling, with size equal to the size of final feature maps and three dense layers. The LeakyReLU is a variant of ReLU, for activating every convolutional layer, which was proposed as a solution to the "dying ReLU" problem, i.e., the tendency of ReLU to keep a neuron constantly inactive as may happen after a large gradient update.

The dataset used for training and evaluating the proposed method was made using two databases of ILD CT scans from two different Swiss university hospitals:

- The first is the publicly available multimedia database of ILDs from the University Hospital of Geneva, which consists of 109 HRCT scans of different ILD cases with 512 x 512 pixels per slice. Manual annotations for 17 different lung patterns are also provided
- The second database was provided by the Bern University Hospital, "Inselspital", and consists of 26 HRCT scans of ILD cases with resolution 512 x 512.

After the preprocessing, the input of the network is a 32x32 image patch.

In this study, the authors used the Adam optimizer to minimize the categorical cross entropy, giving promising results with a relatively slow training (typically a few hours).

The proposed method was implemented using the Theano framework; the methods which do not involve convolutional networks were coded in python and MATLAB.

In this study, the authors reached an accuracy of 85.6%.

The same authors presented an improved method for training the proposed network by transferring knowledge from the similar domain of general texture classification in [32] by using the same architecture of their previous study.

5.1.1.3 Gao et al. (2018)

In this study, Gao et al. [33] proposed a method that classifies and labels ILD tags for holistic CT slices and can possibly be used to prescreen a large amount of radiology data. Their network contains multiple layers: first five layers are convolutional layers followed by three fully connected layers and the final SoftMax classification layer.

The database used in this work contains 120 HRCT scans with 512 × 512 pixels per axial slice, where 17 types of lung tissues are annotated on marked regions of interest (ROIs). Most existing classification methods evaluated on the ILD dataset first extract many image patches from ROIs and then only classify patches into five lung tissue classes: normal (N), emphysema (E), ground glass (G), fibrosis (F) and micronodules (M). Here, consolidation (CD), as a highly prevalent type of ILD, is also included within our classification scheme.

The database contains 2084 ROIs labelled with specific type of ILD disease, out of 120 patients. All patients are randomly split into two subsets at the patient level for training (100 patients) and testing (20 patients). All images containing the six types of diseases are selected, resulting 1689 images in total for training and testing. The input images were resized to 224 x 224 pixels.

For fair comparisons with previous work, Gao et al. conducted experiments under two different settings. One is patch-based classification. In this case, an overall accuracy of 87.9% is achieved. Another experiment shows the holistic image classification results. In this case, the overall accuracy is 68.6%.

5.1.1.4 Song et al. (2017)

In their work, Song et al. [34] compared three methods for the classification of benign and malign lung cancer using the convolution neural network (CNN), deep neural network (DNN), and stacked autoencoder (SAE).

The database used in this paper is the LIDC- IDRI, which contains 244,527 images of the 1010 cases. Each subject includes images from a clinical thoracic CT scan and an associated XML file that records the results of a two-phase image annotation process performed by four experienced thoracic radiologists, that classify the degree of malignancy of pulmonary nodules into five categories: (1) highly unlikely for cancer, (2) moderately unlikely for cancer, (3) indeterminate likelihood, (4) moderately suspicious for cancer and (5) highly suspicious for cancer.

Let’s now discuss the three methods separately, starting from the CNN. The architecture of the proposed CNN is composed by two convolutional layers with 5 x 5 kernels, two pooling layers with 2 x 2 kernels, two fully connected layers and a SoftMax classification layer. The results of the CNN architecture show an accuracy of 84.15%, a sensitivity of 83.96%, and a specificity of 84.32%.

The architecture of the DNN, instead, is composed by one input layer, four fully connected layers and a SoftMax classification layer. The results obtained with this architecture show an accuracy of 82.37%, a sensitivity of 80.66%, and a specificity of 83.9%.

A SAE neural network is a multilayer sparse autoencoder of a neural network and it is an unsupervised learning algorithm. The sparse autoencoder architecture for this study is composed by an input layer, three fully connected layers and a SoftMax classification layer. The results obtained with this neural network show an accuracy of 82.59%, a sensitivity of 83.96% and a specificity of 81.35%.

In conclusion, the result of this study show that the CNN architecture has the best precision.

5.1.2 Positron emission tomography/computed tomography data

In this section, many different studies focused on planar PET/CT data classification are reported. *Table 2* summarizes and compares the most important aspects of the studies.

5.1.2.1 Teramoto et al. (2016)

In this study, Teramoto et al. [29] aimed to develop an ensemble FP (false positive)-reduction method using a convolutional neural network (CNN). The outline of the overall scheme for the detection of pulmonary nodules presented is as follows: first, initial nodule candidates were identified separately on the PET and CT images using the algorithm specific to each image type; subsequently, candidate regions obtained from the two images were combined. FPs contained in the initial candidates were eliminated by an ensemble method using multistep classifiers on characteristic features obtained by a shape/metabolic analysis and a CNN.

With regard to the detection in CT images, the massive region was first enhanced using an active contour filter (ACF), which is a type of contrast-enhancement filter that has a deformable kernel shape. The active contour encloses the nodule without touching normal organs such as blood vessels and lung wall. The PET images were subsequently binarized using a pre-determined threshold to detect regions of increased uptake. Here, candidate regions other than the lungs were eliminated using the lung regions obtained by CT images.

Table 1: Main characteristics of the studies on planar data classification for CT images

	Dataset	Input shape (pixels)	Classes	Neural network	Composition of the neural network	Performance (%)
Li et al.	HRCT (92 scans, 16220 patches)	32 × 32	5 ILD categories (N E, G, F and M)	CNN	• 1 Conv • 1 max pooling • 3 FC	Precision from 50.0% to 95.00% for the 5 categories
Anthimopoulos et al.	2 databases: • HRCT (109 scans) • HRCT (26 scans)	32 x 32	7 classes, including 6 different ILD patterns and healthy tissue	CNN	• 5 Conv • LeakyReLU activation • 1 average pooling	Accuracy of 85.60%
Gao et al.	HRCT (120 scans)	224 x 224	6 ILD categories (N E, G, F, M and CD)	CNN	• 5 Conv • 3 FC • 1 SoftMax	Accuracy of 87.90% for the first setting Accuracy of 68.60% for the second
Song et al.	LIDC- IDRI (1010 scans, 244,527 images)	28 x 28	5 categories of degree of malignancy	CNN	• 2 Conv • 2 pooling • 2 FC • 1 SoftMax	Accuracy of 84.15% Sensitivity of 83.96% Specificity of 84.32%.
				DNN	• 1 input • 4 FC • 1 SoftMax	Accuracy of 82.37% Sensitivity of 80.66% Specificity of 83.90%.
				SAE	• 1 input • 3 FC • 1 SoftMax	Accuracy of 82.59% Sensitivity of 83.96% Specificity of 81.35%

Some FPs represent image features that are similar to nodules. In order to eliminate such FPs while maintaining the value of true positives (TPs), this study focused on a CNN, whose architecture consists of three convolution layers, three pooling layers, and two fully connected layers. The input to the first convolutional layer is (32×32) ×3 images; CT-axial, CT-sagittal, and PET-MIP (Maximum intensity projection PET) images are resized to 32 × 32 pixels.

Relating the dataset, Teramoto et al. considered a total of 104 Japanese men and women who underwent whole-body PET/CT during cancer screening programs from 2009 to 2012.

The number of detected nodules in the initial detection was 176, and its sensitivity exceeded 97%, so satisfactory performance was obtained. However, it was accompanied by a large number of FPs at initial detection.

5.1.2.2. Kirienko et al. (2018)

The present study by Kirienko et al. [35] aimed to develop an algorithm for the classification of lung cancer as T1-T2 or T3-T4 according to the TMN staging FDG positron emission tomography images. The stage of a cancer tells how big it is and whether it has spread. The TNM (where TNM stands for Tumor, Node, Metastasis) staging system is the most common way for doctors to stage non-small cell lung cancer, with stages 1 to 4.

With regard to the dataset, the authors screened all patients who underwent FDG-PET/CT between 01/01/2011 and 27/06/2017 for the purpose of staging a suspected lung lesion, within 60 days before biopsy or surgical procedure. The original CT and PET image size was $512 \times 512 \times N_{\text{slices}}$ and $128 \times 128 \times N_{\text{slices}}$, respectively, where N_{slices} is the number of slices in which the lesion appears. Consequently, the dataset consisted of 3D bounding boxes on both PET and CT images, cropped around the lesion center, identified by two nuclear medicine physicians with dimension $128 \times 128 \times N_{\text{slices}}$.

The algorithm was composed of two networks: a *feature extractor* and a *classifier*. The *feature extractor* was a CNN that took a CT-PET image patch of 128 × 128 pixels as input and performed classification (T1-T2 with label = 0 and T3-T4 with label = 1) according to the appearance of the image patch. Thus, the *feature extractor* aimed to extract the most relevant features from a single patch. The *classifier* took as input the mean of the second to last layer of features extracted from all slices of a single patient and aimed to perform a classification (T1-T2 vs. T3-T4) for that patient. The *SoftMax function* was applied to the last layer of both networks, in order to obtain the probability of being T1- T2 and T3-T4. The class having the highest probability was assigned to each patient.

The algorithm developed and tested in the present work achieved an accuracy of 69%, a recall of 70%, and a specificity of 67% in the test set, for the identification of T1-T2 and T3-T4 lung cancer, in the final model analysis. The AUC was 0.83, 0.73, and 0.68 in the training, validation, and test sets, respectively. The algorithm developed and tested in the present work achieved an accuracy of 87% and 69% in the training and validation sets, respectively.

In conclusion, the key result in the present preliminary investigation is the feasibility and promising performance of CNNs in assessing the T-parameter in lung cancer.

5.1.2.3 Zhang et al. (2019)

Zhang et al. [36] proposed a novel deep learning-based method using multiscale Mask Region–Based Convolutional Neural Network (Mask R-CNN) to address the issues for detecting lung tumor in PET imaging. Mask R-CNN, a deep neural network that can deduce instance segmentation and classification, is the latest and the most effective and beneficial in-depth learning model. The Mask R-CNN comes across as a network that is easy to implement and train.

In this proposed method, the authors firstly produced 3 models of Mask R-CNN, which is a state-of-the-art object detection and segmentation model for lung tumor candidate detection. All the 3 models were fine-tuned and trained with certain data sets using images from 3 different scales. Then, these 3 models of Mask R-CNN were integrated using weighted voting strategy to diminish false-positive outcomes.

Images with 3 different scales were used to produce 3 training data sets: PET images with resolution 512 x 512, 768 x 768, and 1024 x 1024, respectively. All the PET images used in this study were obtained from Changhai Hospital PET/CT Center, and the data were stored in the DICOM (Imaging and Communications in Medicine) format. The image pixels were 168 x 168, and the full-body PET image was of 274 slices, where the PET slice from 40^{th} to 120^{th} layers corresponded to the location of the thoracic cavity.

Recall of all 3 models was obtained to be 1, which affirmed that every single model was sensitive enough for detecting lung tumor with an effective detection of true positives. Precision values of the three models were 0.60, 0.53, and 0.59, which suggested that each single model could still produce several false positives. The precision and F score of the ensemble model was 0.90 and 0.95, which was 0.3 and 0.2 higher than that of the best-performing single model.

Table 2: Main characteristics of the studies on planar data classification for PET/CT images

	Dataset	**Input shape (pixels)**	**Classes**	**Neural network**	**Composition of the neural network**	**Performance (%)**
Teramoto et al.	Whole-body PET/CT (104 scans)	32 x 32	-	CNN	• 3 Conv • 3 pooling • 2 FC	Sensitivity of 97.00%
Kirienko et al.	FDG-PET/CT scans between 01/01/2011 and 27/06/2017	128 x 128	T1-T2 or T3-T4	CNN	-	Accuracy of 69.00%
Zhang et al.	PET/CT images from the Changhai Hospital	168 x 168	Cancer vs Non-cancer	Mask R-CNN	-	Precision of 90.00%

5.2 Methods for volumetric data classification

5.2.1 Computed tomography data

In this section, some methods for volumetric CT data classification are reported. *Table 3* illustrates some key aspects of the presented studies.

5.2.1.1 Jin et al. (2017)

Jin et al. [37] developed a 3D CNN architecture with segmented CT lung volumes as training and testing samples. The proposed model is constructed with eleven layers where there are five convolutions, three max pooling and three full connections layers. The input volume for the architecture of the CNN is of 128 x 128 x 20 pixels. The dataset used for this work comes from the Kaggle Data Science Bowl 2017 dataset, which includes 1,397 instances of high-risk patients' lung low-dose CT scans saved in DICOM standard, which is a standard format of medical imagery. For each patient's scan, it contains a series of axial 2D slices. The CT slices are reconstructed in size of 512 x 512 pixels while the number of slices in each chest cavity scan varies from 128 to 220 depending on the machine and patient. Each instance (patient scan) is labeled as 0 or 1, where 0 means the patient is not diagnosed as cancer, and 1 represents being diagnosed as cancer.

The new model proposed by Jin et al. extracts and projects 3D features to the hidden layers, which preserves the temporal relations between neighboring CT slices, achieving a prediction accuracy of 87.5%. The lowest accuracy of 75.0% was obtained without the pre-segmentation of lung volume.

5.2.1.2 Moradi et al. (2019)

Moradi et al. [38] presented a new method based on 3D convolutional neural networks (CNN) that can reduce the false positives rate while providing a high sensitivity in detecting lung cancer lesions using multi-level contextual CNNs and obtained 91.23% accuracy.

The architecture of their 3D CNN is composed by four convolutional layers, where two are with 5x5x3 kernels and two with 3x3x3 kernels, four max pooling layers and a SoftMax activation function.

The authors used the LUNA16 (Lung Nodule Analysis) dataset to train and evaluate their system. This dataset is made up based on the Lung Image Database Consortium (LIDC) dataset of CT images, with the exception that slices with a thickness of more than 2.5 mm were removed. The LIDC dataset contains 1018 CT image series. After applying this constraint, ultimately this dataset contains 888 CT image series.

5.2.1.3 Polat et al. (2019)

Polat et al. [39] compared two CNN-based models to diagnose lung cancer on lung CT images. These are Straight 3D-CNN with conventional SoftMax and hybrid 3D-CNN with Radial Basis Function (RBF)-based SVM. Straight 3D-CNN was the proposed CNN architecture which used SoftMax as its conventional classifier and additionally hybrid 3D-CNN was the second proposed CNN architecture which utilized Radial Basis Function (RBF)-based SVM as its classifier instead of SoftMax.

The main structures of both proposed 3D-CNN architectures, contain six convolutional layers, four max pooling layers, and two fully connected layers. The kernel size in all convolutional layers is considered 3 x 3 x 3 and 2 x 2 x 2 respectively. For non-linearity, the Relu activation function with less computation cost is applied to the output of each convolutional layer and fully connected layers. In this study, CT scan images of Data Science Bowl and Kaggle are used. This dataset consists of 2101 high-risk patients with CT scans of chest cavity which includes multiple 2D slices in a DICOM format. Moreover, each CT scan is labeled as with cancer or without cancer by pathology diagnosis which contains a different number of images. Each CT scan consists of slices by 512 × 512 sizes with a variable number of slices (z dimension) based on the resolution of the scanner machine. The input has been rescaled to a to a 227 × 227 × 227 volume. The proposed hybrid 3D-CNN with SVM achieved more satisfying results (91.81%, 88.53% and 91.91% for accuracy rate, sensitivity, and precision respectively) compared to straight 3D-CNN in the diagnosis of lung cancer (90.23%, 86.40% and 90.37% for accuracy rate, sensitivity, and precision respectively).

5.2.1.4 Ahmed et al. (2020)

Ahmed et al. [40] proposed a 3D CNN to detect early lung cancer by first preprocessing the raw images using thresholding technique and then using the Vanilla 3D CNN classifier to determine whether the image is cancerous or non-cancerous. The proposed CNN architecture mainly consists of the following layers: two convolution layers which follow two max-pooling layers and one fully connected layer with two SoftMax units.

The authors used LUNA16 datasets (CT scans with labeled nodules). The inputs are the image files that are in “DICOM” format. Actually, the images are of size (z × 512 × 512), where z is the number of slices in the CT scan and varies depending on the resolution of the scanner. Such large images cannot be fed directly into convolutional neural network architecture because of the limit on the computation power. To reduce the size of the input data, the authors have segmented the image and resized to 50 x 50 x 20. The experimental results show that the proposed method can achieve a detection accuracy of about 80%.

Table 3: Main characteristics of the studies on volumetric data classification for CT images

	Dataset	Input shape (pixels)	Classes	Neural network	Composition of the neural network	Performance (%)
Jin et al.	Low-dose CT (1,397 scans)	128 x 128 x 20	Cancer vs non-cancer	CNN	• 5 Conv • 3 max pooling • 3 FC	Accuracy of 87.50%
Moradi et al.	LUNA16 dataset (888 images)	-	-	CNN	• 4 Conv • 4 max pooling • 1 SoftMax	Accuracy of 91.23%
Polat et al.	Data Science Bowl and Kaggle dataset	227 x 227 x 227	Cancer vs non-cancer	Straight 3D CNN	• 6 Conv • 4 max pooling • 2 FC • 1 SoftMax	Accuracy of 90.23% Sensitivity of 86.40% Precision of 90.37%
				Hybrid 3D CNN	• 6 Conv • 4 max pooling • 2 FC • 1 RBF-based SVM classifier	Accuracy of 91.81% Sensitivity of 88.53% Precision of 91.91%
Ahmed et al.	LUNA16 datasets	50 x 50 x 20	Cancer vs non-cancer	CNN	• 2 Conv • 2 max pooling • 1 FC with two SoftMax units.	Accuracy of 80.00%

5.2.2 Positron emission tomography/computed tomography data

In this section, some methods for volumetric PET/CT data classification are reported. *Table 4* illustrates some key aspects of the presented studies.

5.2.2.1 Sibille et al. (2020)

In this study [41], Sibille et al. aimed to evaluate various configurations of deep CNNs for their ability to correctly localize and classify uptake patterns into foci suspicious and nonsuspicious for cancer from whole-body ^{18}F-FDG PET/CT images of patients with lung cancer and lymphoma.

The dataset they used was composed by 302 consecutive patients with lung cancer and 327 consecutive patients with lymphoma undergoing routine whole-body PET/CT at the University Hospital of Münster, Germany, from August 2011 to August 2013. Patients were randomly split into three independent subsets: validation (20% of all patients), training (60%), and test (20%).

PET volumes of interest were first segmented by using a fixed thresholding algorithm to create candidate regions. Subsequently, each volume of interest was evaluated by the CNN independently.

For the test set of 123 patients, the areas under the receiver operating characteristic curve (AUCs) of the CNN in determining hypermetabolic ^{18}F-FDG PET/CT foci that was suspicious for cancer versus nonsuspicious were as follows: CT alone, 0.78 (95% confidence interval [CI]: 0.72, 0.83); ^{18}F-FDG PET alone, 0.97 (95% CI: 0.97, 0.98); ^{18}F-FDG PET/CT, 0.98 (95% CI: 0.97, 0.99).

In conclusion, the study by Sibille et al. is promising in presenting a CNN that can detect, delineate, and classify lesions as suspicious or nonsuspicious for malignancy with high accuracy and although not stated, likely faster than most clinician.

5.2.2.2 Wang et al. (2017)

This study by Wang et al. [42] aimed to compare the performance of multiple machine learning methods for classifying mediastinal lymph node NSCLC from PET/CT images. The evaluated methods included both classical feature-based methods and the state-of-the-art deep learning approach. For the classical methods, the texture features were compared with the features used by human doctors for clinical diagnosis, such as tumor size, CT value, SUV (standardized uptake value), image contrast, and intensity standard deviation. The machine learning methods were also compared with human doctors, so as to evaluate the value of computerized methods for classifying mediastinal lymph node NSCLC from FDG PET/CT images.

^{18}F-FDG PET/CT images of 168 patients were retrieved from the Tumor Hospital of Harbin Medical University database within the period from June 2009 to September 2014. From the 168 patients, 1397 lymph nodes were confirmed cancerous by pathology, and the number of negative and positive samples were 1270 and 127, respectively. Based on the PET/CT images, diagnosis of lymph node metastasis status (positive or negative) was made by four doctors from our institute.

This study compared four mainstream classical machine learning methods and one deep learning method. The classical methods included random forest (RF), support vector machines (SVM), adaptive boosting (AdaBoost), and back-propagation artificial neural network (BP-ANN). The deep learning method was the CNN, which performs classification according to the appearance of the image patch; it learns the patterns of patch appearance from a large amount of training patches. The outputs of CNN are the scores for different classes, and the class with the highest score is deemed as the classification result. For their application, Wang et al. considered as the input of CNN a patch around the lymph node, whereas the outputs were two scores of being benign and malignant. This study used the well-known AlexNet architecture [20] implemented using the Keras library for Python. To avoid overfitting to our data, the number of AlexNet layers was reduced to five.

Table 4: Main characteristics of the studies on volumetric data classification for PET/CT images

	Dataset	Input shape (pixels)	Classes	Neural network	Composition of the neural network	Performance (%)
Sibille et al.	Whole-body PET/CT (302 scans with lung cancer and 327 scans with lymphoma)	-	Suspicious vs Nonsuspicious lesions	CNN	-	AUC of 98.00%
Wang et al.	^{18}F-FDG PET/CT (168 scans)	51 x 51	Lymph node metastasis status (positive or negative)	RF	-	Accuracy of 85.08% Sensitivity of 81.56% Specificity of 88.59%
				SVM	-	Accuracy of 82.73% Sensitivity of 77.08% Specificity of 88.37%
				AdaBoost	-	Accuracy of 85.05% Sensitivity of 85.65% Specificity of 84.45%
				BP-ANN	-	Accuracy of 80.51% Sensitivity of 75.65% Specificity of 85.38%
				CNN	AlexNet with 5 layers	Accuracy of 85.64% Sensitivity of 83.53% Specificity of 87.76%

For both training and testing stages, the inputs of CNN were six axial image patches cropped from the CT and PET SUV images. The patches were cropped around the lymph node center and resampled into 51 × 51 pixels of 1.0-mm size. To generate the patches, the center of each lymph node was specified by the doctor. This was the only step requiring user input.

The accuracy reached for the four classical methods was 82.73%, 85.08%, 85.05% and 80.51% for the SVM, RF, AdaBoost and BP-ANN respectively. For the CNN the accuracy reached was 85.64%. Relating the sensitivity, it was 77.08%, 81.56%, 85.65% and 75.65% for the SVM, RF, AdaBoost and BP-ANN respectively, while for the CNN it was 83.53%. Finally, the specificity was 88.37%, 88.59%, 84.45% and 85.38% for the SVM, RF, AdaBoost and BP-ANN respectively and for the CNN it was 87.76%.

Thus, it can be observed that in this study, the performance of CNN is not significantly different from the best classical methods, even if in many recent publications of medical image analysis, CNN was reported to outperform classical methods for imaging modalities other than PET/CT. The reason of this is probably related to the fact that CNN has not fully explored the functional nature of PET. In fact, before the image patches are input to CNN, the pixel intensities are normalized to a range of [−1, 1], thus the discriminative power of SUV is lost during the normalization.

6. Materials and methodology

This study aims to test the performance of an already-tested neural network, called Cloud-YLung [43], on PET scans. The mentioned neural network has been properly designed to classify two histological types of lung tumor directly from 3D CT whole-lung scans, particularly ADC and SCC, achieving a test accuracy of 75% and AUC of 84%.

In this work, instead, 3D PET whole-lung scans have been used to train the Cloud-YLung neural network, to compare the predictions obtained with those already achieved by [43] and evaluate which kind of radiological data is better suited to non-invasively classify ADC and SCC.

6.1 Dataset

The PET data used in this work consisted of 68 whole-body scans selected from three openly accessed datasets, notably TCGA-LUAD, TCGA-LUSC and NSCLC-Radiogenomics. All the scans were stored in DICOM format and, above them, 34 belonged to the class of ADC lung tumor and 34 to the class of SCC lung tumor. Each scan was composed by a variable number of slices from 91 to 551, with an original resolution of 192x192 pixels, 168x168 pixels or 128x128 pixels.

6.2 Proposed methodology

All the pre- and post-processing steps have been performed on the free version of Google Colab, whereas the training and the classification procedures have been done on the Pro version to greatly reduce the computational time. *Figure 18* shows the workflow of both pre- and post-processing steps.

6.2.1 Pre-processing

The first part of this works consisted of four preprocessing steps.

1 Firstly, all the scans have been converted from DICOM format to numpy format and a 1-mm resampling has been performed to make slices within scans spatially homogenous. After this first step, the scans were composed by a number of slices between 298 and 1802 with a resolution between 500x500 pixels and 700x700 pixels.

2 The second pre-processing step consisted of slice selection centering the lung volume. This step has been performed "manually", meaning that the initial and final slices to be kept were selected one by one by observing the scans, as well as the black parts to be removed around the body volume. This step is essential to enormously reduce the dimension of each scan and to remove all the useless black pixels from the images.

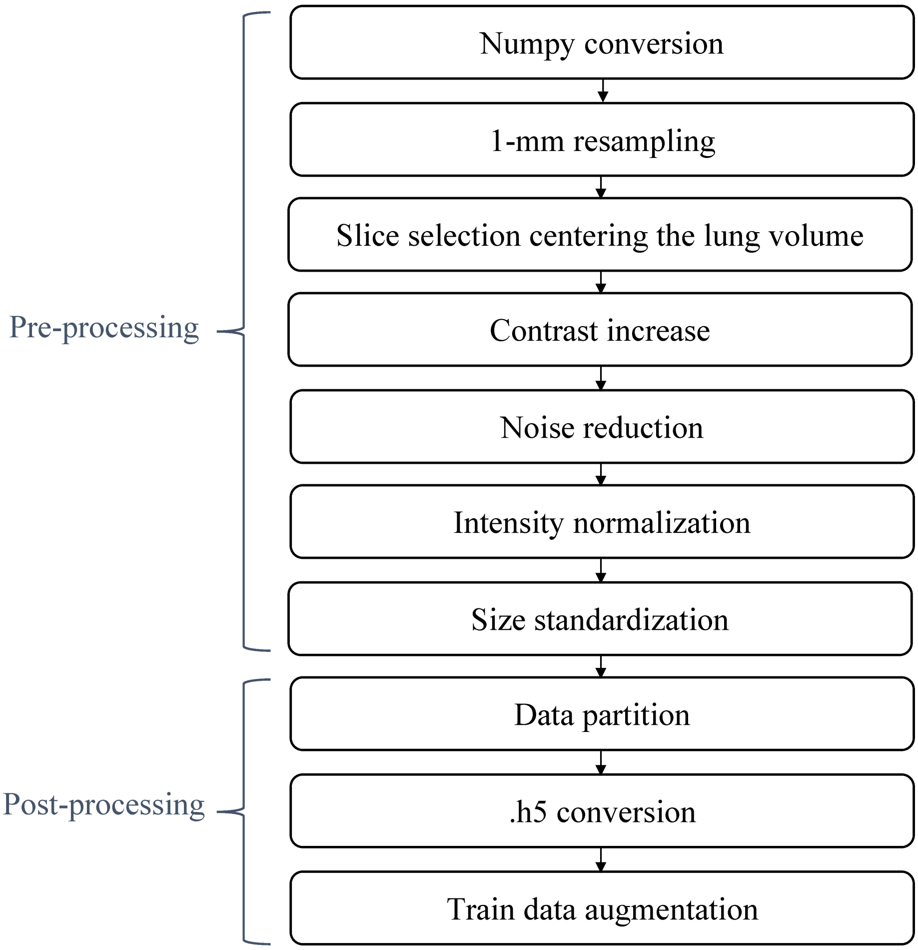

Figure 18: Workflow of pre- and post-processing steps

3 In the third step an enhancement of the scans has been performed by increasing the contrast, sharpening the images, and reducing the noise. Moreover, in this step all the images have been normalized so that, at the end, each pixel value lied between 0 and 1. The normalization has been simply performed by subtracting each pixel by the minimum value of all the pixels and dividing by the range of values. *Figure 19* reports a comparison between the original image, the image after the second and after the third pre-processing steps of the same subject.

4 The last step was the one needed to standardize the dimensions of the scans before entering the neural network. Specifically, two datasets of different sizes were created. The first one had dimensions of 250x200x250 pixels, according to [43]. The second one, instead, had dimensions of 180x270x430 pixels and it has been chosen according to the dimensions of the scans after the first three steps. In particular, a histogram of the distribution of each volume dimension was drawn to choose the values in which the greatest distribution resides, ad shown in *Figure 20*.

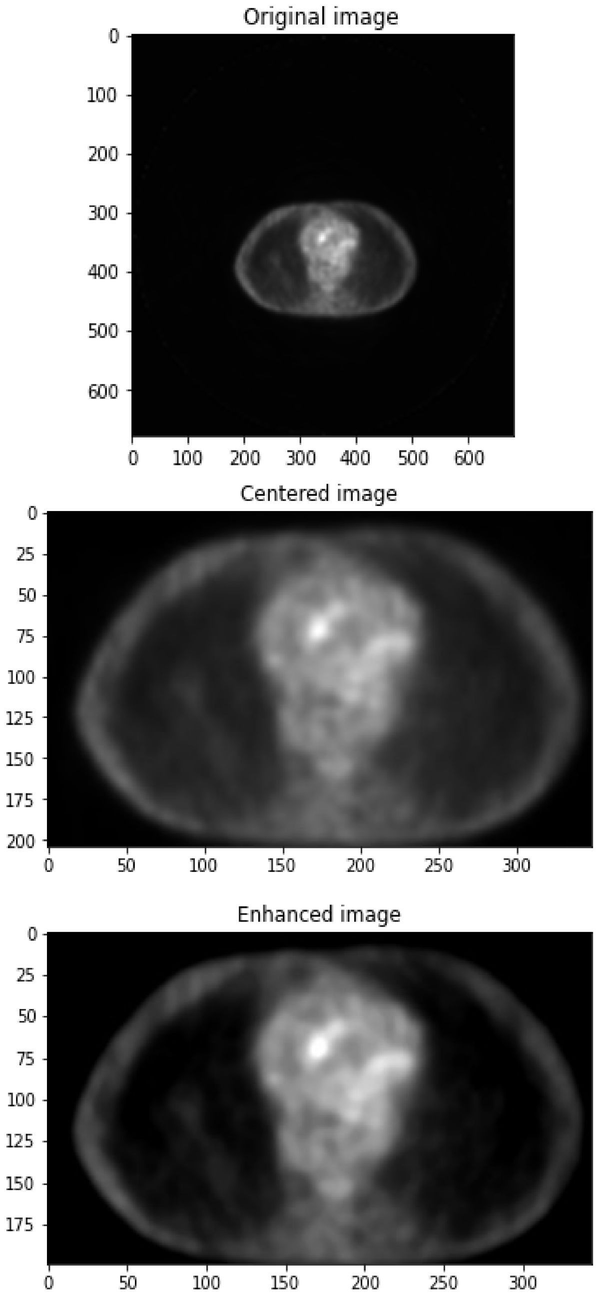

Figure 19: Comparison between the original image, the image without the black pixels surrounding the body volume and the enhanced image of the same patient.

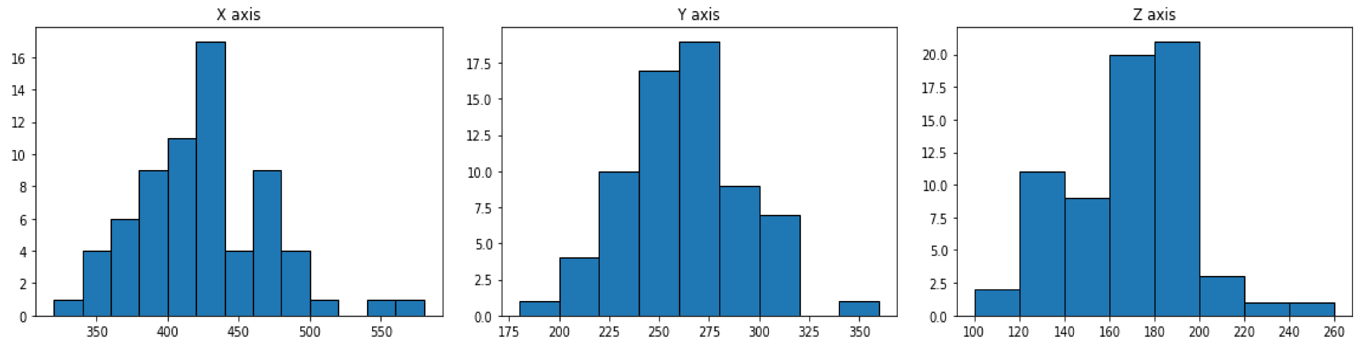

Figure 20: Histograms of the x, y, and z axes of all the 68 scans after the third pre-processing step

6.2.2 Post-processing

After performing the pre-processing steps, the final scans were divided into train and test set with a split ratio of 60:40 and 40% of the train set served as validation set, according to [43]. Each of these three sets was created so that the two classes, ADC and SCC, were balanced. In particular, the train set was composed by 24 scans, the validation set by 16 scans, and the test set by 28 scans, all equally distributed between ADC and SCC. Then, the post-processing steps simply consisted in compression of the three sets in .h5 format and augmentation of the train set, consisting in 15° left/right rotation and random zooming between 0.8 and 1.2. After the augmentation process, the number of scans in the train set increased from 24 to 72.

At this point, the scans were ready to train and test the Cloud-YLung neural network. For the first dataset, the neural network was trained using the same setups described in [43] in order to have a fair comparison with its results. For the second dataset, instead, a tuning of the parameters of the neural network has been accomplished to increase its performances. Particularly, the dropout rate was changed from 0.5 to 0.3 and the learning rate was reduced from 0.0001 to 0.00001.

6.3 Statistics

To evaluate the performances of the neural network on the test set, some statistical parameters have been considered. Particularly, the precision (1), which describes how close or dispersed the measurements are to each other.

$$Precision = \frac{TP}{TP + FP} \qquad (1)$$

The accuracy (2), which assesses whether a series of measurements are correct on average.

$$Accuracy = \frac{N^{\circ}\ of\ correct\ predictions}{N^{\circ}\ of\ total\ predictions} \qquad (2)$$

Then, the sensitivity, or true positive rate (3), and specificity, or true negative rate (4), which refer to the probability of a positive test, conditioned on truly being positive and to the probability of a negative test, conditioned on truly being negative, respectively.

$$Sensitivity = \frac{TP}{TP + FN} \qquad (3)$$

$$Specificity = \frac{TN}{TN + FP} \qquad (4)$$

Finally, also the F1-score (5), i.e., the harmonic mean of the precision and sensitivity, the Area Under the Curve (AUC) and the Receiver Operating Characteristics (ROC) curve were taken into account.

$$F1 - score = 2 * \frac{Precision * Sensitivity}{Precision + Sensitiviy} \qquad (5)$$

Especially, the AUC and ROC are two of the most important evaluation metrics for checking any classification model’s performance since ROC is a probability curve and AUC represents the degree or measure of separability, i.e., it tells how much the model is capable of distinguishing between classes. Thus, the higher is the AUC, the better the model is at distinguishing between the classes.

7. Results

Tables 5 and *6* report the performances of Cloud-YLung trained with PET scans in classifying ADC and SCC in terms of Precision, Specificity, Sensitivity, F1-score, Accuracy and AUC. In particular, *Table 5* shows the results obtained with the first dataset, that of size 250x200x250 pixels, while *Table 6* shows the results with the second dataset of size 180x270x430 pixels.

Figure 21, finally, shows the ROC curves obtained from the Cloud-YLung trained with PET scans of both sizes.

Table 5: Performance of Cloud-YLung in classifying ADC and SCC from PET scans with dimensions 250x200x250 pixels

	Precision (%)	**Specificity (%)**	**Sensitivity (%)**	**F-1 score (%)**	**Accuracy (%)**	**AUC (%)**
ADC	67	64	71	69		
SCC	69	71	64	67		
Average	68	68	68	68	68	64

Table 6: Performance of Cloud-YLung in classifying ADC and SCC from PET scans with dimensions 180x270x430 pixels

	Precision (%)	**Specificity (%)**	**Sensitivity (%)**	**F-1 score (%)**	**Accuracy (%)**	**AUC (%)**
ADC	73	79	57	64		
SCC	65	57	79	71		
Average	69	68	68	68	68	65

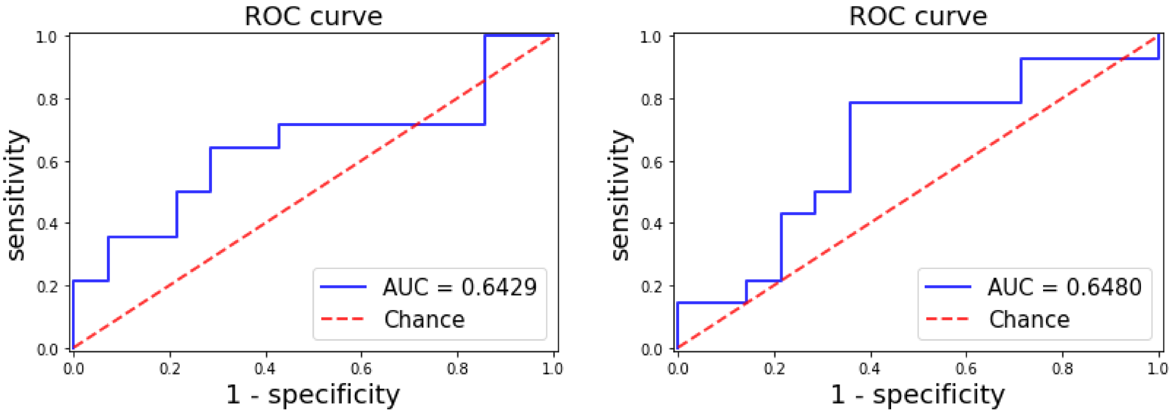

Figure 21: Left: ROC curve obtained after testing Cloud-YLung with PET 250x200x250 pixels; right: ROC curve obtained after testing Cloud-YLung with PET 180x270x430 pixels

8. Discussion of the results

In this study, two trainings have been done with the purpose of comparing the performances of Cloud-YLung in classifying two histological types of lung cancer from differently sized PET scans. In particular, the first experiment was performed using a pre-processed dataset with the same size as [43] and in this case the neural network was trained with the same parameter configuration used in [43] to have a fair comparison with its results. The second experiment, instead, was done to optimize the performance of Cloud-YLung as much as possible by choosing a size suitable for the PET scans and fine-tuning some parameters of the neural network, namely the learning rate and the dropout rate. Observing the results reported in *Tables 5* and *6*, it is possible to state that the performances of Cloud-YLung with the two datasets are almost comparable, since they show the same accuracy (68%), and the AUC of the second dataset is higher than that of the first dataset for only one point (65% versus 64%). This suggests that the tuning of the parameters performed in the second experiment does not improve so much the performances of this neural network and that probably those obtained are the maximum performances achievable with Cloud-YLung trained with PET scans. Presumably, using a neural network designed “ad hoc” for this task can increase the performances. However, it is worth to highlight how challenging this task is, as it is a matter of discriminating the histotype of lung cancer directly from a radiological datum and not from a histological one. Therefore, even if the accuracies reached in this study are not that high, they are nevertheless promising for the development of future studies focused on this task.

Table 7 illustrates a comparison between the results obtained training Cloud-YLung with CT scans [43] and that obtained training Cloud-YLung with PET scans of both sizes. From this comparison it is possible to note that the performances achieved in [43] are quite better than those obtained in this work, especially in terms of AUC, which in the case of CT scans is almost twenty points higher than those obtained in this work using PET scans (85% vs 64% and 65%). This result suggests that among the two types of radiological data, CT appears to be more informative than PET in terms of distinguishing between histological types of lung cancer. Based on this, it might be interesting to develop a future study focused on using PET / CT scans to train this neural network for the same task, understanding whether that type of radiological data can boost Cloud-YLung's performance.

Table 7: Performance of Cloud-YLung in classifying ADC and SCC from CT scans [43] and from PET scans

	Accuracy (%)	**AUC (%)**
Cloud-YLung CT	75	84
Cloud-YLung PET (250;200;250)	68	64
Cloud-YLung PET (180;270;430)	68	65

Comparing the results obtained in this work with those present in the literature is not so obvious, since almost all studies focused on neural networks for the classification of lung cancer are based on the recognition of scans containing the tumor with respect to healthy subjects' scans. In fact, focusing on the methods for volumetric data classification illustrated in Chapter 5.2, in the studies of Jin et al. [37], Polat et al. [39] and Ahmed et al. [40] the focus was on discriminating between the cancerous and non-cancerous scans. Moreover, in their study Sibille et al. [41] classified the PET/CT scans as suspicious or nonsuspicious, while Wang et al. [42] compared different approaches to classify the metastasis status as positive or negative. Although the accuracies achieved in the studies present in the literature are decidedly superior to the best achieved in this study, as it is can be seen in *Table 8*, this work is the only one that attempts to classify lung cancer according to its histological type from PET data, thus remarking how complex and underdeveloped this task is.

Table 8: Comparison between methods for volumetric data classification and Cloud-YLung trained with PET 180x270x430 pixels in terms of classification classes and accuracy

	Classes	**Accuracy (%)**
Jin et al.	Cancer vs non-cancer	87.50
Polat et al.	Cancer vs non-cancer	91.02[1]
Ahmed et al.	Cancer vs non-cancer	80.00
Sibille et al.	Suspicious vs nonsuspicious lesion	-
Wang et al.	Positive vs negative metastasis status	83.80[1]
Cloud-YLung PET (180;270;430)	ADC vs SCC	65.00

[1] *The accuracies reported for the studies of Polat et al. and Wang et al. are a mean of the accuracies reached in the single experiments*

Conclusion

This thesis investigated the performances of a neural network, namely Cloud-YLung, in classifying two histological types of lung cancer directly from PET scans. Two training experiments have been proposed: a first one in which the scans were cut with a standard dimension according to a previous experiment done with Cloud-YLung, and a second one in which the dimensions of the scans were accurately chosen, and the parameters of the neural network were carefully tuned to optimize its performances. The results obtained suggested that Cloud-YLung, between the two radiological data, CT scans resulted to be more informative in non-invasively classify ADC and SCC. However, making a new neural network suitably designed to work with PET scans as well as strengthening the pre-processing phase is likely to lead to increased performance.

Furthermore, in this work it was possible to realize that the classification of two or more histological types of lung cancer directly from radiological data is not such a developed field, neither using CT scans, much less using PET scans as input. For this reason, even if the accuracies achieved are not that high, they are still promising for further studies focused on this very demanding task.

Some future developments of this thesis could be the realization of a neural network carefully designed to work with PET scans in identifying the histological type of lung cancer or the training of Cloud-YLung with PET / CT scans to verify if this method of hybrid acquisition could increase the performances. These developments could lead to a future in which it will be possible to diagnose the histological type of lung cancer a patient is suffering from avoiding the need for biopsy, thus reducing the time required for diagnosis and the costs that the medical facility has to face.

Bibliography

[1] Novaes, Fabiola Trocoli, et al. "Lung cancer: histology, staging, treatment and survival." Jornal Brasileiro de Pneumologia 34 (2008): 595-600.

[2] Zheng, Min. "Classification and pathology of lung cancer." Surgical Oncology Clinics 25.3 (2016): 447-468.

[3] Nicholson, Andrew G., et al. "The 2021 WHO classification of lung tumors: impact of advances since 2015." Journal of Thoracic Oncology 17.3 (2022): 362-387.]

[4] Rueda, José-Ramón, et al. "Non-invasive interventions for improving well-being and quality of life in patients with lung cancer." Cochrane Database of Systematic Reviews 9 (2011).

[5] Cavalheri, Vinicius, et al. "Exercise training undertaken by people within 12 months of lung resection for non-small cell lung cancer." Cochrane Database of Systematic Reviews 6 (2019).

[6] Driessen, Elisabeth J., et al. "Effects of prehabilitation and rehabilitation including a home-based component on physical fitness, adherence, treatment tolerance, and recovery in patients with non-small cell lung cancer: a systematic review." Critical reviews in oncology/hematology 114 (2017): 63-76.

[7] Christensen, Jared D., and Betty C. Tong. "Computed Tomography Screening for Lung Cancer:: Where Are We Now?." North Carolina medical journal 74.5 (2013): 406.

[8] Ostrowski, Marcin, Tomasz Marjański, and Witold Rzyman. "Low-dose computed tomography screening reduces lung cancer mortality." Advances in medical sciences 63.2 (2018): 230-236.

[9] Huang, Kai-Lin, et al. "Effects of low-dose computed tomography on lung cancer screening: a systematic review, meta-analysis, and trial sequential analysis." BMC pulmonary medicine 19.1 (2019): 1-11.

[10] Bronzino, Joseph D., ed. Medical devices and systems. CRC press, 2006.

[11] Khandpur, Raghbir Singh. Handbook of biomedical instrumentation. McGraw-Hill Education, 1987.

[12] Cantatore, Angela, and Pavel Müller. "Introduction to computed tomography." Kgs. Lyngby: DTU Mechanical Engineering (2011).

[13] Sahiner, Berkman, et al. "Deep learning in medical imaging and radiation therapy." Medical physics 46.1 (2019): e1-e36.

[14] Drukker, Karen, et al. "Biomedical imaging and analysis through deep learning." Artificial Intelligence in Medicine. Academic Press, 2021. 49-74.

[15] Litjens, Geert, et al. "A survey on deep learning in medical image analysis." Medical image analysis 42 (2017): 60-88.

[16] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." nature 521.7553 (2015): 436-444.

[17] Yamashita, Rikiya, et al. "Convolutional neural networks: an overview and application in radiology." Insights into imaging 9.4 (2018): 611-629.

[18] Yasaka, Koichiro, et al. "Deep learning with convolutional neural network in radiology." Japanese journal of radiology 36.4 (2018): 257-272.

[19] LeCun, Yann, et al. "Handwritten digit recognition with a back-propagation network." Advances in neural information processing systems 2 (1989).

[20] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems25 (2012).

[21] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

[22] Szegedy, Christian, et al. "Going deeper with convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

[23] Su, Jong-Chyi, et al. "A deeper look at 3D shape classifiers." Proceedings of the European Conference on Computer Vision (ECCV) Workshops. 2018.

[24] Muzahid, A. A. M., et al. "3D Object classification using a volumetric deep neural network: An efficient Octree Guided Auxiliary Learning approach." IEEE Access 8 (2020): 23802-23816.

[25] Su, Hang, et al. "Multi-view convolutional neural networks for 3d shape recognition." Proceedings of the IEEE international conference on computer vision. 2015.

[26] Wang, Cheng, et al. "NormalNet: A voxel-based CNN for 3D object classification and retrieval." Neurocomputing 323 (2019): 139-147.

[27] Tatarchenko, Maxim, Alexey Dosovitskiy, and Thomas Brox. "Octree generating networks: Efficient convolutional architectures for high-resolution 3d outputs." Proceedings of the IEEE international conference on computer vision. 2017.

[28] Anthimopoulos, Marios, et al. "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network." IEEE transactions on medical imaging 35.5 (2016): 1207-1216.

[29] Teramoto, Atsushi, et al. "Automated detection of pulmonary nodules in PET/CT images: Ensemble false-positive reduction using a convolutional neural network technique." Medical physics43.6Part1 (2016): 2821-2827.

[30] Chen, Song, et al. "Diagnostic classification of solitary pulmonary nodules using dual time 18F-FDG PET/CT image texture features in granuloma-endemic regions." Scientific reports 7.1 (2017): 1-8.

[31] Li, Qing, et al. "Medical image classification with convolutional neural network." 2014 13th international conference on control automation robotics & vision (ICARCV). IEEE, 2014.

[32] Christodoulidis, Stergios, et al. "Multisource transfer learning with convolutional neural networks for lung pattern analysis." IEEE journal of biomedical and health informatics 21.1 (2016): 76-84.]

[33] Gao, Mingchen, et al. "Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks." Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization 6.1 (2018): 1-6.]

[34] Song, QingZeng, et al. "Using deep learning for classification of lung nodules on computed tomography images." Journal of healthcare engineering 2017 (2017).

[35] Kirienko, Margarita, et al. "Convolutional neural networks promising in lung cancer T-parameter assessment on baseline FDG-PET/CT." Contrast Media & Molecular Imaging 2018 (2018).

[36] Zhang, Rui, et al. "Multiscale mask R-CNN–based lung tumor detection using PET imaging." Molecular imaging 18 (2019): 1536012119863531.

[37] Jin, Taolin, et al. "Learning deep spatial lung features by 3D convolutional neural network for early cancer detection." 2017 international conference on digital image computing: techniques and applications (DICTA). IEEE, 2017.

[38] Moradi, Pouria, and Mansour Jamzad. "Detecting lung cancer lesions in CT images using 3D convolutional neural networks." 2019 4th International Conference on Pattern Recognition and Image Analysis (IPRIA). IEEE, 2019.

[39] Polat, Huseyin, and Homay Danaei Mehr. "Classification of pulmonary CT images by using hybrid 3D-deep convolutional neural network architecture." Applied Sciences 9.5 (2019): 940.]

[40] Ahmed, Tasnim, et al. "Lung cancer detection using CT image based on 3D convolutional neural network." Journal of Computer and Communications 8.03 (2020): 35.

[41] Sibille, Ludovic, et al. "18F-FDG PET/CT uptake classification in lymphoma and lung cancer by using deep convolutional neural networks." Radiology 294.2 (2020): 445-452.

[42] Wang, Hongkai, et al. "Comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from 18F-FDG PET/CT images." EJNMMI research 7.1 (2017): 1-11.

[43] Tomassini, Selene, et al. "Cloud-YLung for Non-Small Cell Lung Cancer Histology Classification from 3D Computed Tomography Whole-Lung Scans." *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2022.