

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Pre-training Autoencoder for Lung Cancer Characterisation

Francisco Carvalho Moreira da Silva

DISSERTAÇÃO



Mestrado Integrado em Engenharia Eletrotécnica e de Computadores

Supervisor: Tânia Maria Pereira Lopes, PhD

Co-Supervisor: Hélder Filipe Pinto de Oliveira, PhD

Co-Supervisor: António Manuel Trigueiros da Silva Cunha, PhD

July 31, 2020

Resumo

Dados estatísticos têm mostrado que o diagnóstico em fase avançada representa um dos mais importantes fatores responsáveis pela elevada taxa de mortalidade associada a pacientes de cancro do pulmão, e avanços tem sido mostrados na eficácia de um tratamento personalizado com base no perfil genético de cada paciente, sendo possível obter um maior controlo na resposta ao tratamento. A biópsia tem sido o método mais utilizado para obter a caracterização do tumor, e embora seja um método capaz de obter informação útil nesse sentido é uma abordagem que pode levantar inúmeros problemas clínicos e que pode não ser capaz de obter uma caracterização completa devido à heterogeneidade espacial do tumor.

A imagiologia médica oferece a possibilidade de obter um grande conjunto de informação útil, abrindo oportunidades para investigar a relação entre manifestações visuais presentes numa imagem médica e o perfil genético do tumor, usando uma abordagem não-invasiva. Especialistas têm trabalhado para fornecer anotações semânticas relativas a inúmeras patologias associadas a tecidos do pulmão. No entanto, tendo em conta que este é um processo lento e complexo, têm sido exploradas outras abordagens capazes de extrair informação relevante de forma direta, sendo o número de exemplos disponível para desenvolver estes modelos uma limitação comum.

Este trabalho investigou a capacidade de usar uma abordagem de *Transfer Learning* baseada em um *Convolutional Autoencoder* treinado para trabalhar como extrator de características de Tomografias Computadorizadas (TAC), com foco na caracterização do cancro do pulmão numa perspectiva local, onde a área analisada apenas engloba o nódulo, e numa perspectiva que abrange toda a secção do pulmão no plano axial. A malignidade do nódulo foi avaliada numa classificação binária, atingindo uma AUC média de 0.936, e usando a mesma área analisada foi explorada a existência de padrões relevantes para a previsão do estado de mutação do *EGFR* (AUC = 0.540). Alargando a área de análise para a secção axial do pulmão, foi implementado um modelo de segmentação do pulmão avaliado usando duas bases de dados distintas (DSC = 0.948). O estado de mutação do *EGFR* foi também explorado nesta análise, mostrando uma melhoria na deteção de manifestações relevantes deste gene (AUC = 0.645) comparativamente aos resultados obtidos usando apenas informação contida no nódulo.

Abstract

Statistics have shown that one of the main factors responsible for the high death rate related to lung cancer is the late diagnosis, and precision medicine practices have shown advances in the direction of an individualized treatment according to the genetic profile of each patient, providing a better control on tumor response. The biopsy has been the traditional method to obtain the tumor characterisation and despite being able to provide useful information, different clinical problems might be caused by this invasive procedure and the tumor heterogeneity may not be clearly captured.

Medical imaging offers valuable information with a full state perspective of the tumor, opening opportunities to explore the imaging manifestations associated with the tumor genotype in a non-invasive way. Radiology experts have been putting effort into providing qualitative annotations regarding different lesions; however, being such a complex and time-consuming process, other approaches have been explored to achieve a tumor characterisation in an end-to-end process, being often limited by the size of the available datasets.

This work investigated the ability to use a Transfer Learning approach based on a trained Convolutional Autoencoder to work as feature extractor of Computed Tomography (CT) images, focusing on lung cancer characterisation in a local nodule and in an axial lung slice perspectives. The nodule malignancy was assessed in a binary classification with a mean AUC of 0.936 achieved, and using the same ROI was attempted to predict the *EGFR* mutation status, evaluated by a mean AUC of 0.540. Extending the ROI to a lung axial slice, a 2D lung segmentation model was developed, achieving a mean Dice-Coefficient value of 0.948 in a multi-dataset evaluation, and a classification model achieved an AUC of 0.645 in the *EGFR* mutation status prediction. This work investigated the hypothesis that it might be possible to find relevant phenotype manifestations in lung structures outside the tumor region, achieving results that suggest the presence of useful biomarkers using a ROI that included an entire lung section, with a performance increase in comparison with the local nodule analysis.

Acknowledgements

The work accomplished in this dissertation would not be possible without the contribution of many persons.

I would like to express my gratitude to my advisor Tânia Pereira for constantly being available to provide the guidance and support needed, always allowing me to discuss my own thoughts on the work done.

I would also like to thank Professor Hélder P. Oliveira and Professor António Cunha for their eagerness to help me at any moment, motivating me with challenging and valuable advice.

Last but not least, a special acknowledgement to my family, girlfriend and closest friends, by unconditionally encouraging me throughout this journey.

Francisco Silva

This work is financed by the ERDF – European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme and by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia within project POCI-01-0145-FEDER-030263.

Contents

1	Introduction	1
1.1	Context	1
1.2	Motivations	1
1.3	Objectives	2
1.4	Contributions	3
1.5	Document Structure	3
2	Background	5
2.1	Lung Cancer	5
2.2	Precision Medicine	5
2.3	Medical Diagnostic Methods	5
2.3.1	Biopsy	5
2.3.2	CT Images	6
2.4	Imaging Features	6
2.5	Interstitial Lung Diseases	7
3	Literature Review	9
3.1	Classification of lung structures	9
3.2	Pre-trained neural networks used in biomedical images	13
3.3	Public Datasets	18
3.4	Summary	20
4	Data Description	23
4.1	Datasets	23
4.1.1	LIDC-IDRI	23
4.1.2	NSCLC-Radiogenomics	23
4.1.3	UHC of São João Dataset	24
4.1.4	Lung CT Segmentation Challenge 2017	24
4.1.5	ILD Dataset	25
4.1.6	Inclusion Criteria	25
4.2	Data Pre-processing	26
4.2.1	DICOM Objects to Image Data Conversion	26
4.2.2	Resampling	26
4.2.3	HU Normalization	26
4.3	Summary	27

5	Lung Nodule Characterisation	29
5.1	Unsupervised Feature Learning	29
5.1.1	Convolutional Autoencoder (CAE)	29
5.1.2	Proposed CAE	30
5.1.3	Materials and Methods	30
5.2	Nodule malignancy classification	32
5.2.1	Materials and Methods	32
5.2.2	Classifier	33
5.2.3	Model fine-tuning	34
5.2.4	Results	35
5.2.5	Discussion	37
5.3	EGFR Prediction	37
5.3.1	Materials and Methods	38
5.3.2	Results	39
5.3.3	Discussion	40
5.4	Summary	40
6	Lung Characterisation	43
6.1	Lung Segmentation	43
6.1.1	Materials and Methods	43
6.1.2	Segmentation Model	45
6.1.3	Results	45
6.1.4	Discussion	47
6.2	CAE for Lung Feature Learning	48
6.2.1	CAE architecture	48
6.2.2	Materials and Methods	49
6.3	EGFR Prediction	50
6.3.1	Materials and Methods	51
6.3.2	Results	52
6.3.3	Discussion	53
6.4	Summary	54
7	Conclusions	55
	References	57

List of Figures

2.1	Examples of patterns from three different subjects represented in CT images. . . .	7
3.1	DCAE architecture.	11
3.2	Comparison of the results for Emphysema.	12
3.3	Transfer Learning concept overview.	13
3.4	Model structure overview.	16
3.5	Block workflow.	17
3.6	Visual aspect of the most common lung tissue patterns.	19
3.7	Example patches from each of the 28 texture classes.	20
4.1	Lung segmentation challenge examples.	25
5.1	Proposed lung nodule CAE architecture.	30
5.2	Examples of lung nodule patch extraction.	31
5.3	Lung nodule reconstruction examples.	32
5.4	5-Fold Cross-validation overall scheme.	33
5.5	Nodule malignancy model.	34
5.6	Nodule malignancy classification fine-tuning strategy applied.	35
5.7	Averaged ROC curve for lung nodule malignancy classification using the Transfer Learning approach.	36
5.8	Averaged ROC curve for <i>EGFR</i> mutation status prediction with a nodule analysis.	40
6.1	Comparison on two distinct axial lung slices.	44
6.2	Lung segmentation architecture implemented.	45
6.3	Lung segmentation example comparison.	47
6.4	Lung segmentation example comparison.	49
6.5	Lung CAE data acquisition pipeline.	49
6.6	Lung axial slice CAE reconstructions.	50
6.7	Selection of the lung with nodule. Slices from NSCLC-Radiogenomics database patients.	52
6.8	Averaged ROC curve for <i>EGFR</i> mutation status prediction with a lung axial slice analysis.	53

List of Tables

3.1	Overview of published studies regarding the classification of lung tissue patterns.	10
3.2	ILD classifications using F-score evaluation metric.	13
3.3	Overview of collected publications regarding the use of pre-trained neural networks in medical imaging applications	15
3.4	Distribution of the lung tissue patterns.	18
3.5	Properties for the dataset version with rotated patches.	19
4.1	Number of patients from each dataset used considering each proposed task. . . .	27
5.1	Manual-search for CAE hyper-parameters	31
5.2	Manual search for malignancy classification hyper-parameters	34
5.3	Best hyper-parameters for lung nodule malignancy classification.	36
5.4	Lung nodule malignancy classification results.	36
5.5	Manual-search for <i>EGFR</i> classification hyper-parameters	38
5.6	Best hyper-parameters for lung nodule malignancy classification.	39
5.7	<i>EGFR</i> mutation status prediction results with nodule level analysis.	39
6.1	Manual-search for lung segmentation model hyper-parameters	44
6.2	Set of hyper-parameters that achieved highest performance.	46
6.3	Lung segmentation results.	46
6.4	Manual-search for Lung CAE hyper-parameters	50
6.5	Set of hyper-parameters values used in manual-search for <i>EGFR</i> mutation status assessment.	51
6.6	Best hyper-parameters found for lung slice <i>EGFR</i> mutation status classification. .	52
6.7	Lung with nodule: classification results for lung axial slice <i>EGFR</i> mutation status prediction.	53
6.8	Both lungs: classification results for lung axial slice <i>EGFR</i> mutation status prediction.	53

Abbreviations

2D	Two Dimensional
3D	Three Dimensional
ALK	Anaplastic Lymphoma Kinase
AUC	Area Under The Curve
CNN	Convolutional Neural Network
CT	Computed Tomography
ctDNA	Circulating Tumour DNA
DBT	Digital Breast Tomosynthesis
DCAE	Deep Convolutional Autoencoder
DL	Deep Learning
DNA	Deoxyribonucleic Acid
DSC	Sørensen–Dice Coefficient
EGFR	Epidermal Growth Factor Receptor
FC	Fully-Connected
FN	False-Negative
FP	False Positive
HRCT	High Resolution Computed Tomography
HU	Hounsfield Unit
HUG	University Hospitals of Geneva
ILD	Interstitial Lung Diseases
ILSVRC	ImageNet Large Scale Visual Recognition Competition
K-NN	K-Nearest Neighbours
KRAS	Kirsten Rat Sarcoma Viral Oncogene Homolog
KTD	Kyleberg Texture Dataset
LR	Logistic Regression
LTRC	Lung Tissue Research Consortium
MLP	Multi-Layer Perceptron
MRI	Magnetic Resonance Imaging
MT	Mutated Type
NB	Naive Bayes
NSCLC	Non-small cell lungcancer
OAR	Organs At Risk
pMCI	Progressive Mild Cognitive Impairment
RBF	Radius Basus Function
RF	Random Forest
RGB	Red, Green and Blue
ROC	Receiver Operating Characteristics
ROI	Region Of Interest
RTOG	Radiation Therapy Oncology Group
SCLC	Small cell lung cancer
sMCI	Stable Mild Cognitive Impairment
SVM	Support Vector Machines
TL	Transfer Learning
TKI	Tyrosine-kinase Inhibitor
WT	Wildtype

Chapter 1

Introduction

1.1 Context

Around the world, lung cancer is on the top of cancer-related mortality numbers [1, 2]. In the United States, 230,000 new cases were estimated in 2019, accounting for about 13% of all new cancer cases [1]. In Europe, 470,000 new cases and 388,000 deaths caused by lung cancer were estimated in 2018. Although below breast and colorectal cancers in the number of new estimated cases, lung cancer led in the number of estimated deaths [3]. Between 2000 and 2010, 9767 new lung cancer patients were registered in 12 hospitals from the north of Portugal. During this period, the incidence was significantly higher in males, about 80%; however, the number of female patients increased by 30%. According to this study, 77.8% of the patients were diagnosed with advanced-stage lung cancer [4].

Tobacco is undoubtedly considered the most significant risk factor, responsible for about 80% of lung cancer-related deaths. However, records show that from 2012 to 2016 there was a decrease in mortality rate of 4% and 3% per year in men and women, respectively, possibly associated with a decrease in tobacco use [1].

In clinical terms, lung cancer can be classified into three major categories: non-small cell lung cancer (NSCLC) which covers about 85% of all lung cancer cases [5], small cell lung cancer (SCLC), and lung carcinoid tumor. The 5-year survival rate after diagnosis in lung cancer patients is 19% (16% for men and 22% for women) and is higher in NSCLC (23%) than in SCLC (6%), since the tumor in this latter type is characterized by a quicker spread. Only 16% of cases are diagnosed in a local state, and in these cases the 5-year survival rate increases to 56% [1].

1.2 Motivations

Since a large percentage of the patients are diagnosed in advanced stage [1, 6], working on the ability to tailor the treatment regarding each patient's needs and tumor state is an important step to decrease lung cancer mortality rate.

Currently, the biopsy is the main method to characterise lung tumors, using a sample of extracted tumor tissue for molecular analysis. Although it provides essential information for characterisation, being an invasive procedure can lead to some associated side effects, as it can be painful and risky for the patient. In some cases, the tumor location may not allow the usage of this method [7]. Also, given the tumor spatial heterogeneity, the extracted sample analysis may not provide a complete and clear understanding of the tumor [8, 9]. With the growing need for a more realistic perspective of the tumor status to support decision-making on selecting each patient's treatment, it is important to find methods that respond to this problem in a non-invasive way. Medical images analysis can provide valuable information to contribute to a better tumor characterisation, being an active research point where different approaches have been investigated showing promise results [9–12].

In lung cancer, the most frequently mutated genes are Epidermal Growth Factor Receptor (*EGFR*) and Kirsten Rat Sarcoma Viral Oncogene Homolog (*KRAS*), constituting critical genomic biomarkers due to their importance in targeted treatment response [13]. About 50% of never-smokers patients with NSCLC exhibit mutant *EGFR*, while *KRAS* mutations are more frequent in NSCLC patients with smoking history [14, 15]. Being a research focal point, assessing these driver oncogenes might provide essential information for precision medicine practices.

Imaging qualitative features annotated by radiology experts in Computed Tomography (CT) scans have shown high correlation with some genes mutation status, and their ability to feed predictive models have shown promising results [9, 16]. Besides the majority of research made by far have focused on features related to the nodule in these type of prediction tasks, previous work considered an analysis where not only were used features related to the nodule but also from patterns outside the nodule, showing better results when both were used to develop the predictive model [9], supporting the idea that it might be possible to find important patterns related to lung cancer in structures outside the nodule region of interest (ROI) [9]. Since the visual assessment of these qualitative data is a time-consuming and highly subjective task [17], the idea of building models capable of learning these relevant patterns is raised. Deep learning techniques, namely Convolutional Neural Networks (CNN), are not only able to self-learn representative features from raw data, but also carry the advantage of preserving the original spatial representation by convolutional operations. However, these type of implementations need a large amount of data to perform well, which can be a problem in medical imaging related tasks, where publicly available data is often scarce. To overcome this problem, exploring Transfer Learning techniques might be an useful approach, given recent promising performances on classification tasks in this field [11, 18, 19].

1.3 Objectives

This dissertation aims to provide a lung cancer characterisation considering imaging phenotypes identified in CT images, in order to improve support on decision-making in the field of precision medicine. Thus, two distinct characterisations are intended to be conducted: (1) only considering

the local tumor information, and (2) extending the ROI to include lung patterns outside the tumor region.

Taking advantage of Transfer Learning techniques, this work's approach consists of pre-training a Deep Learning architecture - Convolutional Autoencoder (CAE) - on a similar dataset, being intended to learn generic features that enable the knowledge transfer into the target classification task. With this approach is expected to significantly reduce the necessary trainable parameters, allowing to develop classification models with a reduced amount of training data.

1.4 Contributions

This dissertation presents the following contributions:

- An investigation regarding the viability of using a Transfer Learning based approach by pre-training a Convolutional Autoencoder to develop a classification model;
- Lung nodule malignancy classification model in a Transfer Learning based approach;
- Assessment of *EGFR* mutation status using the same Transfer Learning technique, considering a local nodule and an axial lung slice analysis;
- Lung segmentation model implementation with a multi-dataset development.

1.5 Document Structure

This dissertation comprises 7 Chapters. The present Chapter presented an introduction considering statistical data related to lung cancer, explaining the urgency behind improving practices in order to decrease mortality rate, as well as exploring some general ideas about the investigation intended to be conducted. Chapter 2 details some important clinical concepts regarding current diagnostic methods and the relevance of some imaging manifestations in CT scans. Chapter 3 provides a detailed review on recent works regarding the classification of lung diseases and the use of Transfer Learning techniques in the biomedical imaging field. It is also detailed relevant publicly available databases found in the reviewed works. Chapter 4 explains the required steps in data preparation, characterising each dataset and for which task it was included, with a final summary mentioning the number of relevant samples included from each data collection. Chapter 5 and Chapter 6 details model implementations regarding a lung nodule and an axial lung slice characterisation. Each of these two last chapters includes an introduction explaining its content, followed by dataset and implementation details, and finishing with the presentation of the achieved results with discussion. Chapter 7 provides a final conclusion regarding the work done in this dissertation alongside some future work considerations.

Chapter 2

Background

2.1 Lung Cancer

In lung cancer cases, the main factor that highly contributes to the nearly 1.6 million deaths annually is the late diagnosis. Once diagnosed, a patient with a lung tumour that has spread to other organs and tissues has a 6% chance of a 5-year-survival [20], so treating a patient with different methods until a positive response to be noticed leads, in most cases, to minimal chances of survival. However, advances in genomics have driven to perceive that lung tumors can be sorted by their genetic causes and the treatment response can be different for each patient, rising opportunities to explore treatment strategies that rely on the individual's genetic profile [13].

2.2 Precision Medicine

Precision medicine consists of a genomic-based approach, at treatment level, that focus on targeted therapies development based on the tumor genotype characterisation. Thus, it is possible to predict the outcome for a specific treatment tailored for the individual. Standard procedures include mutational test of some target genes like *EGFR*, *KRAS*, and *ALK* [13].

One of the most well-known anti-cancer treatments are the Tyrosine-kinase inhibitors (TKIs), developed to prevent cell survival and uncontrolled proliferation processes, caused by mutated *EGFR* activation. However, since the co-occurrence of other mutated genes might change the treatment response by conferring resistance to the targeted gene, the treatment success depends on multiple factors that must be studied [13].

2.3 Medical Diagnostic Methods

2.3.1 Biopsy

As mentioned before, the biopsy is the currently employed diagnostic strategy to assess specific bio-markers status, supporting precision medicine by providing extracted tissue to molecular analysis. Although providing valuable information, a complete tumor characterisation often requires

several biopsies due to spatial heterogeneity, and multiple clinical risks might arise from these procedures.

The liquid biopsy has been used as an alternative and non-invasive approach, consisting of the analysis of circulating tumor DNA (ctDNA), which involves tests on body fluids samples (typically blood is used, but saliva and urine might be other possibilities) [21]. This approach overcomes the tumor heterogeneity problem, and as its study is increasing, it might replace tissue biopsy in the future. However, the main limitation relies on the fact that the number of ctDNA molecules present in a blood sample is much lower than the non-cancer-related DNA molecules, resulting in an abundance of ctDNA lower than the necessary for an accurate characterisation [22].

2.3.2 CT Images

Imaging information found in radiographic medical images have shown the potential to obtain a reliable tumor characterisation, offering a faster and less invasive approach compared with the traditional tissue biopsy.

The CT scan enables the tissue 3D reconstruction by using multiple X-ray projections from multiple angles. By passing through tissues with different density values, the absorbed radiation provide different contrast imaging of these tissues, changing the level of energy that reaches the detector. These density values are expressed using the Hounsfield (HU) scale values. As shown in Equation 2.1, the formula to compute the HU values depends on the linear attenuation coefficient of water μ_{water} , the linear attenuation coefficient of air μ_{air} , and the linear attenuation coefficient of the substance μ [23].

$$HU = 1000 \times \frac{\mu - \mu_{water}}{\mu_{water} - \mu_{air}} \quad (2.1)$$

The computed values are based on a scale where the radio density of water, at standard pressure and temperature conditions, is equal to 0 HU, and the radio density of air, at the same conditions, is equal to -1000 HU. Therefore, a higher HU value corresponds to a denser tissue region, resulting in a brighter region on the image [23].

2.4 Imaging Features

Progress in the computation ability to process medical images transformed these images into a large set of data that provide useful information to help in tumor characterisation.

Besides clinical features, which describe general information including patient's gender, age and smoking status, imaging features can be obtained by visual assessment, resulting in semantic annotations in a controlled vocabulary made by an expert; by a radiomic approach, which refers to extracting large amounts of quantitative features directly from the image or by taking advantage of Deep Learning (DL) methods. All of these strategies have shown to be able to successfully provide information to help in lung cancer diagnostic and prognostic.

2.5 Interstitial Lung Diseases

Interstitial lung diseases (ILD) refers to a group of more than 150 inflammations of lung tissues that largely affect the lung parenchyma, comprising the structures involved in gas transfer [24,25]. Although many of these ILD causes still remain unknown, the most common symptoms are related to breathing dysfunction [24], and these similarities in clinical manifestations often leads to an ambiguous diagnostic.

Giving the visual variability of these tissue patterns and their co-occurrence in different lung diseases, the identification of a disease might become a challenging and subjective task. The most typically studied tissue inflammations include emphysema, ground glass, fibrosis and micronodules, due to large variations between subjects, as presented in Figure 2.1.

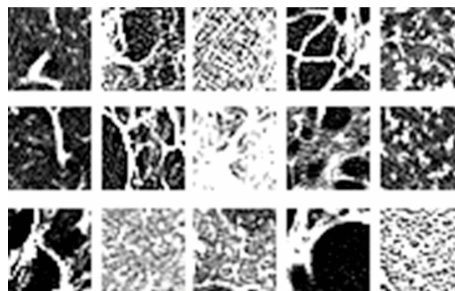


Figure 2.1: Examples of three different subjects. Patterns represented in CT images from left to right: healthy tissue, emphysema, ground glass, fibrosis, micronodules. From [26].

Chapter 3

Literature Review

The present Chapter provides a detailed description covering a selection of relevant studies in the classification of lung diseases field (Section 3.1), and in Section 3.2 is covered a collection of relevant investigations regarding Transfer Learning based approaches in different medical imaging applications.

3.1 Classification of lung structures

In lung cancer diagnosis, the visual identification of important chronic tissue inflammations has been a widely used procedure, due to their valuable correlation with lung cancer and impact on individualized treatment planning [9, 16]. However, even for experts, this task has shown to be complex, time-consuming due to the large quantity of radiological data to analyse, and often leads to high inter-observer variability [27].

Recent advances in image pattern recognition field allowed achieving promising results on the automatic classification of these features. In particular, Deep Learning techniques have proven to achieve successful results in this type of tasks. Table 3.1 describes some of the most relevant investigations found related to this research topic.

Table 3.1: Overview of published studies regarding the classification of lung tissue patterns.

Publication	Description	Datasets	Results	Classified Patterns
Huang et al. (2020) [28]	Classification of ILD patterns using a CNN and a deep convolutional autoencoder (DCAE) with different transfer learning approaches.	Source dataset: Kylberg Texture Dataset (KTD) [29]; Auxiliary dataset: patches extracted from non-labelled areas in target dataset; Target dataset: ILD database (University Hospitals of Geneva) [24].	Evaluation metric: F1-score CNN (two stage TL) = 0.9791. DCAE (two stage TL) = 0.9810.	Healthy, ground glass, emphysema, micronodules and fibrosis.
Joyseere et al. (2019) [30]	Classification of ILD patterns using the combination between Riesz and deep learning (DL) features by fine-tuning the InceptionV3 pre-trained model.	Training dataset: ILD dataset.	Best performance with late concatenation of both Riesz and DL features (AUC = 0.948).	Healthy, ground glass, emphysema, micronodules and fibrosis.
Negahdar et al. (2019) [31]	Propose an end-to-end pipeline based on CNN AlexNet architecture to detect and evaluate the proportion of emphysematous lung tissue.	Training dataset: ILD database; Test set: Lung Tissue Research Consortium (LTRC) [32].	Average evaluation of 0.96 considering Precision, Recall and F1-score regarding all classes.	Emphysema.
Gao et al. (2018) [33]	Investigate the impact of using the entire lung image as a holistic input in the classification of ILD imaging patterns using pre-trained AlexNet [34].	ILD database used for Training (100 patients) and testing (20 patients).	Overall accuracy of 87.9%, with Emphysema being perfectly classified (F-score = 1.00).	Healthy, ground glass, emphysema, micronodules and fibrosis.
Christodoulidis et al. (2017) [35]	Evaluate the impact of a multi-source Transfer Learning using 6 texture databases.	6 public texture benchmark databases (ALOT, DTD, FMD, KTB, KTH-TIPS-2b, UIUC); Test data: ILD database.	Ensemble approach achieved a F1-score = 0.8817.	Healthy, ground glass, emphysema, micronodules and fibrosis.

Huang et al. [28] proposed a CNN, which architecture consists of 6 convolutional layers followed by a fully-connected layer. Also, a Deep Convolutional Autoencoder (DCAE) was developed based on the proposed network, using a mirror deconvolution process as decoder for attempting the inputs reconstruction. The DCAE architecture is illustrated in Figure 3.1, where the convolutional part is equal to the proposed CNN architecture, as referred. It was also explored a different approach on Transfer Learning, consisting of an intermediate unsupervised fine-tuning using extracted non-labelled patches from the target dataset due to useful similarities with the target domain. The input patches size was 32×32 pixels, using 3 channels corresponding to 3 different CT attenuation channels used to highlight different patterns. This “two-stage transfer learning” approach was applied using the proposed CNN and the DCAE.

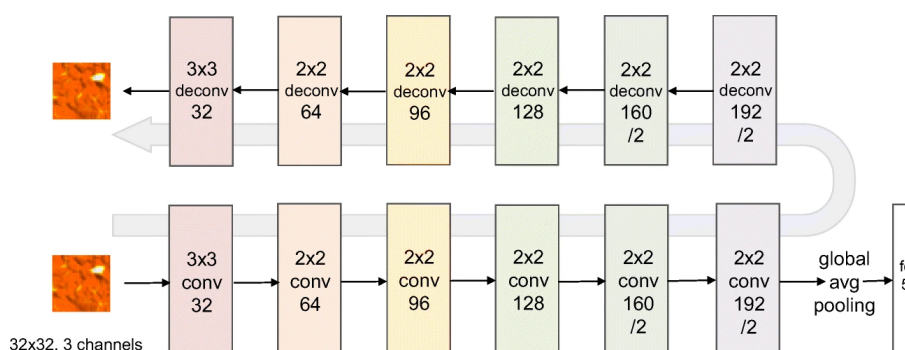


Figure 3.1: DCAE architecture. From [28].

The source and target databases used were the Kyleberg Texture Dataset (KTD) [29] and the ILD dataset, from the University Hospitals of Geneva [24], respectively. The ILD patterns classified were ground glass, emphysema, micronodules, fibrosis, and healthy tissue for comparison. The better classification performances were achieved using the two-stage transfer learning approach on both the CNN and the DCAE models proposed (0.9791 and 0.9810, respectively) using the average F1-score as evaluation metric.

Using the same target database, Joyseeree et al. [30] proposes an approach that complements the lack of invariance to local rotations in CNN models with multi-directional properties of Riesz wavelets. This feature-fusion approach intends to classify the same 5 tissue disorders: healthy, ground glass, emphysema, micronodules and fibrosis. Considering the lack of annotated data to train the model, the InceptionV3 architecture was the pre-trained model selected for learning the high-level pattern features using all annotated classes from the target dataset. Figure 3.2 show the results obtained for the emphysema pattern, and the results were very similar between all classes. The best average performance was achieved with the late fusion implementation. Given the fact that the DL feature vector dimension was approximately 4 times larger than the Riesz representation, the early-fusion approach did not show a significant impact in performance when compared with the use of DL features only.

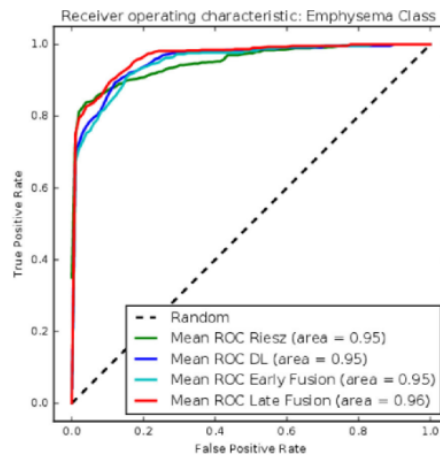


Figure 3.2: Comparison of the results for Emphysema using the receiver operating characteristic (ROC). Adapted from [30].

Negahdar et al. [31] evaluated the existence and proportion of emphysematous lung tissue, with co-occurrence of other lung diseases with an end-to-end pipeline based on CNNs. The proposed pipeline includes lung segmentation, patch generation, classification architecture, training and testing, and emphysema quantification. The patch generation process consists of dividing each slice into patches of 32x32 pixels, and the CNN architecture was based on AlexNet, replacing the softmax single-label with a multi-label loss function to allow class differentiation. The ILD database was employed for training, and for the test set was used the LTRC dataset (Lung Tissue Research Consortium) [32]. An average evaluation value of 0.96 was achieved considering precision, recall and F-score regarding all classes.

A holistic approach is presented by Gao et al. [33], where the proposed model receives as input an entire lung CT image. According to this study, the traditional image patch sizes are relatively small, raising the hypothesis that there are some details that may be overlooked in the patch-based representation. A CNN-based method was formulated to classify 6 lung tissue classes, adding the consolidation class to the previous studies classified patterns. The CNN architecture is identical to AlexNet's [34], containing 5 convolutional layers followed by 3 fully-connected layers, and the final softmax classification layer (changed to 6 classes in this specific application). Here, were selected 3 CT attenuation ranges, not only because it offers a better visual separation between the 6 classes, but also to accommodate the selected pre-trained network that uses RGB values from natural images. From the 120 patients contained in the ILD database, the random split into training (100 patients) and test (20 patients) sets was made with the concern that different slices, resulted from the data augmentation process, from the same patient would not appear in both sets. In Table 3.2 are represented the F-scores regarding all ILD patterns considered, noting that emphysema was perfectly classified.

Table 3.2: ILD classifications using F-score evaluation metric. Adapted from [33].

Pattern	Emphysema	Fibrosis	Ground glass	Normal tissue	Micronodules	Consolidation
F-score	1.00	0.80	0.75	0.40	0.56	0.50

Using 6 publicly available texture benchmark databases, Christodoulidis et al. [35] explored the performance impact of using a multi-source transfer learning approach for lung pattern classification. The CNN architecture used was the same as proposed in [36], using as input an image patch of 32x32 pixels. In this work, the knowledge is transferred from each source to a different CNN, which are aggregated and the fused knowledge is then used in a random initialized CNN with the same architecture. According to this study, this ensemble technique allows the system to capture different characteristics and usually achieves better performances than with each of the predictors alone. The test set used was provided by the Bern University Hospital, containing 26 CT scans of ILD cases with annotations for 7 different classes: healthy tissue, ground glass, reticulation, consolidation, micronodules, honeycombing, and a combination of ground glass and reticulation. The best performance (F-score = 0.8817) was achieved with the proposed multi-source approach.

3.2 Pre-trained neural networks used in biomedical images

Deep learning techniques have been demonstrating their value in several medical applications, mostly related to image recognition and classification tasks, including detection and classification of pulmonary nodes [37], diagnosis of skin cancer [38] and diagnosis of Alzheimer's disease [39]. However, the lack of publicly available data for most medical applications limits the model's ability to generalize on unseen data, due to training in smaller datasets. Transfer Learning refers to a technique that seeks to overcome this overfitting problem, using models pre-trained on a larger set of images and then fine-tuned using the target dataset, as represented in the TL concept overview in Figure 3.3.

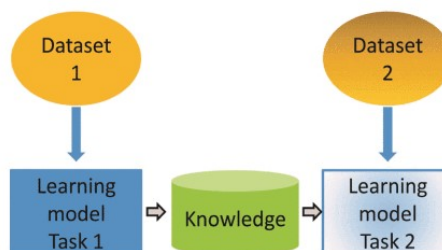


Figure 3.3: Transfer Learning concept overview. From [37].

The main problem that affects TL effectiveness lies in the similarities between the source and target domains. For image recognition problems, most of the methods based on Transfer Learning are pre-trained on ImageNet dataset [40]. The gap between the characteristics of those natural images and the medical images from the specific problem should be minimized [41].

To answer this problem, different neural networks' architectures have been explored using a variety of implementation strategies to investigate the ability of using pre-trained models in detection and classification problems. Table 3.3 includes a compilation of the most relevant works regarding the use of Transfer Learning based approaches in different medical imaging applications.

Table 3.3: Overview of collected publications regarding the use of pre-trained neural networks in medical imaging applications

Publication	Description	Datasets	Results	Model
Xiong et al. (2019) [11]	Comparison between different CNN implementation strategies in the CT assessment of <i>EGFR</i> mutation status.	Cohort of 1010 patients (balanced number of mutated and wildtype cases) from Shanghai Chest Hospital.	Best performance with fine-tuned 2D model using a multi-view perspective and fusion input size (AUC = 0.838).	ResNet-101.
Samala et al. (2019) [18]	Implementation of multi-stage Transfer Learning for the classification of masses in Digital Breast Tomosynthesis (DBT).	CNN pre-trained on ImageNet database; Mamography data (SFM and DM) and DBT data.	Best AUC value of 0.91 ± 0.03 achieved using the multi-stage approach, freezing the first convolutional layer (C1).	CNN architecture based: AlexNet.
Oh et al. (2019) [42]	Evaluate transfer learning techniques on pMCI vs sMCI classification task using Magnetic Resonance Imaging (MRI) data.	Data from the publicly available Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset.	Best performance achieved when using pre-trained weights from source AD/NC classification task to detect pMCI.	Proposed Architecture
Da Nóbrega et al. (2018) [43]	Evaluate different CNN architectures, as well as multiple classifiers for lung nodules malignancy classification task.	CNNs pre-trained on ImageNet database; LIDC/IDRI database used for fine-tuning all models.	Best performance achieved an AUC value of 93.19%.	CNN architecture: ResNet50; Classifier: SVM-RBF.
Tajbakhsh et al. (2016) [19]	Compared the performance between CNN trained from scratch and pre-trained and fine-tuned in 4 distinct medical imaging applications of specialties: radiology, cardiology and gastroenterology.	Data collected from non-specified databases for all experiments.	Considering the training dataset size, fine-tuning depth plays an important role in the performance achieved; in the worst case, the fine-tuned model performed as well as the CNN trained from scratch.	CNN architecture based: AlexNet.

Xiong et al. [11] proposed a model based on the ResNet-101 architecture to predict the *EGFR* mutation status in CT images. This work presents a comparison between different implementation strategies of the CNN model, considering dimension filters (2D/3D), input sizes (small/medium/large and fusion), slicing methods (transverse/multi-view plane) and training approach (transfer learning/training from scratch). The dataset used in all experiments include a cohort of 1010 patients from the Shanghai Chest Hospital, with a balance number of both mutated and wildtype classes: 510 and 500, respectively. The performance of each developed model was evaluated by the area under receiver operating characteristic (ROC) curve (AUC). The best performance was achieved by the 2D fine-tuned model with fusion input sizes and multi-view slicing, showing an AUC value of 0.838. Since there are no pre-trained 3D CNN available, all the 3D models were trained from scratch, noting the fact that these models performed better than all the 2D models when Transfer Learning was not used.

In a different medical field, Samala et al. [18] proposed a multi-stage Transfer Learning approach for the classification of malignant and benign masses in Digital Breast Tomosynthesis (DBT). The AlexNet architecture was adapted for a 2 class classification task, by adding to the last fully connected layer (FC), with 1000 outputs, two layers with 100 and 2 nodes, respectively, as shown in the structure overview represented in Figure 3.4. In the first-stage Transfer Learning approach, only the mammography data [44] was used to fine-tune the pre-trained AlexNet, adding the DBT data in the multi-stage strategy. Besides the freezing alternatives addressed, the effects of the sample size, regarding each one of the datasets used, were also studied. The best AUC value was of 0.91 ± 0.03 and was achieved in the model scheme with stage 1 training followed by stage 2 training using the DBT data. In both stages, the first convolutional layer, C1, was frozen.



Figure 3.4: Model structure overview. Adapted from [18].

Oh et al. [42] proposed a TL approach to identify progressive or stable mild cognitive impairment (pMCI and sMCI, respectively) on MRI data from the publicly available Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset [45]. A Convolutional Autoencoder (CAE) was developed for unsupervised pre-training with unlabelled data, transforming the input images into a lower dimension feature space on the encoding phase, and then decoding the compressed representation into the original image. Was developed another CAE, on which was applied an inception

module based on GoogLeNet inception version 2 [46], allowing the extraction of multilevel representations in a parallel way from the same input tensor. The encoding weights were used for knowledge transfer to build a classifier to differentiate between Alzheimer’s Disease (AD) and Normal Control (NC) classes, and the fine-tuned weights from this task were then applied to the learning model which aimed to distinguish between pMCI and sMCI, given the increased complexity and the lower amount of labelled data available for the latter task. Results showed that the learning models based on CAE and ICAE (Inception CAE) improved the performances by 5 – 12% and 7 – 14%, respectively, over already implemented baseline CNN-based models. Additionally, as described before, the pre-trained weights used on the MCI classification task were transferred from the AD/NC fine-tuned model as it was found to better improve pMCI complex patterns detection when compared with the weights from the initial ICAE.

To attempt the classification of lung nodules malignancy, Da Nóbrega et al. [43] explored several ImageNet pre-trained CNN, such as VGG-16-19 [47], MobileNet [48], Xception [49], InceptionV3 [50], ResNet50 [51], Inception-ResNet-V2 [52], DenseNet169, DeseNet201 [53], NASNetMobile and NASNetLarge [54] as deep features extraction from the Lung Image Database Consortium image collection dataset (LIDC-IDRI) [55]. This work also investigates the performance of 5 different classifiers, including Naive Bayes (NB) [56], Multilayer Perceptron (MLP), Support Vector Machines (SVM) [57], K-Nearest Neighbours (K-NN) [58], and Random Forest (RF) [59]. The model using the ResNet50 architecture with the SVM-RBF (Radius Basis Function kernel) classifier was the selected model, achieving an AUC value of ≈ 0.932 . Figure 3.5 shows the framework overview of this study.

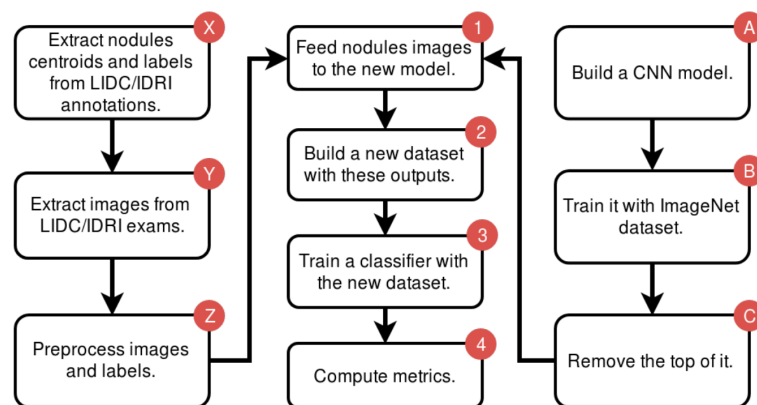


Figure 3.5: Block workflow. Blocks X, Y and Z refers to pre-processing the LIDC/IDRI data and blocks A, B and C represent the processes to build the feature extractor. From [43].

The central question “can the use of pre-trained deep CNNs with sufficient fine-tuning eliminate the need for training a deep CNN from scratch?” was addressed by Tajbakhsh et al. [19], by conducting experiments in 4 distinct medical imaging applications: poly detection in colonoscopy videos, colonoscopy frame classification, pulmonary embolism detection in CT, and intima-media boundary segmentation in ultrasonographic image. Using a model based on the AlexNet architecture, this work proposes a fine-tuning scheme, which consists of starting from the last layer and

then incrementally including more layers until the desired performance is achieved. The observations allowed to conclude that when the size of the training data was large, shallow fine-tuning was most often outperformed by CNNs trained from scratch, whereas using deeper fine-tuning the performance was identical or even superior; nevertheless, when training with a smaller dataset, a Transfer Learning approach always achieved best performances.

3.3 Public Datasets

Regarding the published works previously described, there are some publicly available datasets that should be detailed.

The ILD dataset [24], used in all presented works for classification of ILD patterns, contains 108 High Resolution CT (HRCT) scans with 512×512 pixels per axial slice, and 1946 regions of interest (ROI) were delineated and annotated for the specific tissue pattern. Table 3.4 represents the distribution of the lung tissue patterns included in the database. Although this database contains scans for 128 patients, only 108 have an annotated HRCT.

Table 3.4: Distribution of the lung tissue patterns. Adapted from [24].

Class	Annotated HRCTs
Healthy	7
Fibrosis	38
Ground glass	37
Micronodules	16
Consolidation	14
Reticulation	10
Emphysema	5
Bronchiectasis	8
Macronodules	7
Bronchial wall thickening	1
Cysts	3
Others	14
Total	108

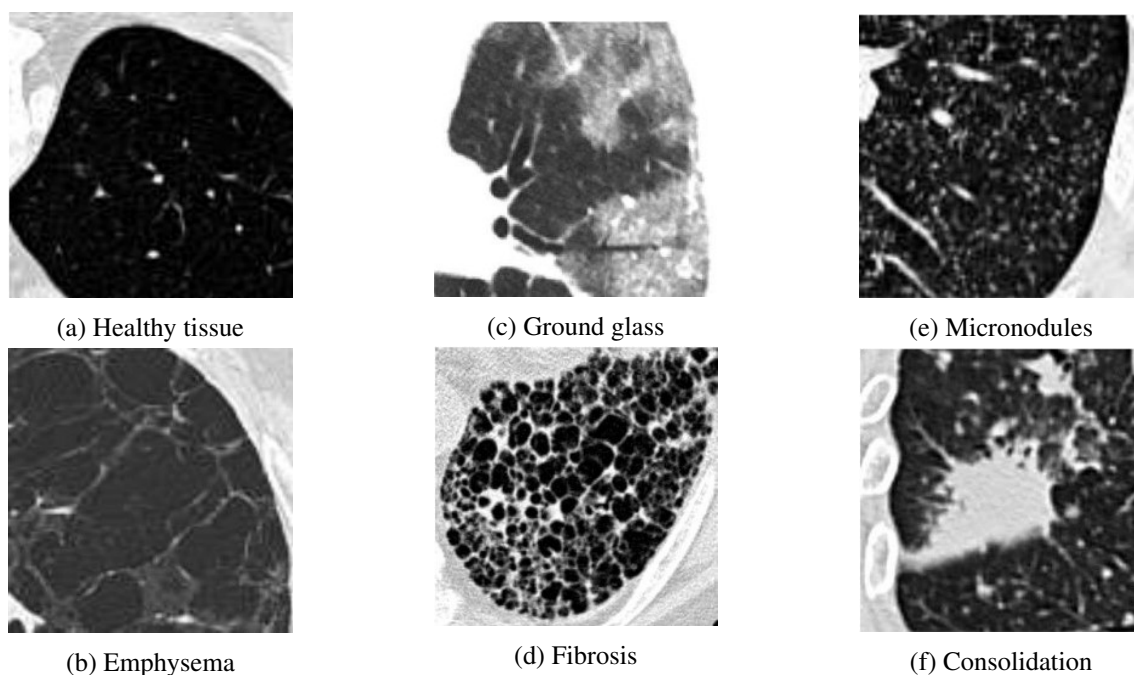


Figure 3.6: Visual aspect of the most common lung tissue patterns in HRCT of patients with ILDs. From [24].

From the public texture benchmark databases used by Christodoulidis et al. [35], the Kyleberg Texture Database (KTD) [29] should be detailed. This dataset includes two versions: without rotated texture patches, and with rotated texture patches. The properties of the latter version are depicted in Table 3.5. As one can see, the patches were rotated with θ degrees, where $\theta \in [0, 30, 60, \dots, 330]$, and then cropped to the same size of the non-rotated patches.

Table 3.5: Properties for the dataset version with rotated patches. Adapted from [29].

Number of texture classes	28
Number of rotations	12
Rotation increment	30 degrees
Number of samples/class	1,920
Total number of samples	53,760
Texture patch size	576×576 pixels

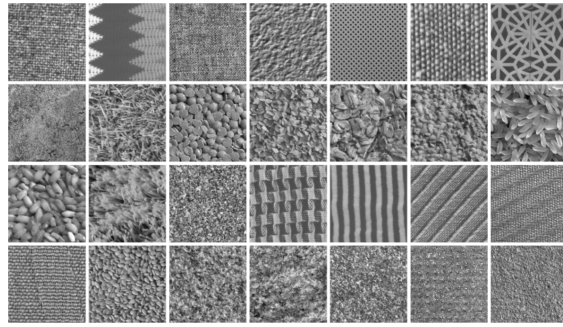


Figure 3.7: Example patches from each of the 28 texture classes. Several distinct objects are included, from rice and sugar to woven on chair. From [29].

Finally, the ImageNet dataset [60], on which the majority of the classification models described were pre-trained before the fine-tuning process, is a public large-scale collection of 2D natural images according to the WordNet [61] hierarchy. There are more than 80,000 nouns in WordNet and ImageNet offers an average of 1,000 images to illustrate each one.

Regarding classification tasks, there are other datasets with a large number of labeled natural images, for instance the CIFAR-10 and CIFAR-100 datasets [62], which contain 10 and 100 classes, respectively. The number of images included is the same, but the first dataset contains 6,000 images per class and the second only has 600.

3.4 Summary

The studies described in this Chapter were selected considering two main subjects: the classification of lung structures, and the use of pre-trained networks in medical imaging applications. Considering the analysis made and the wide variety of methodologies included, it is possible to identify some common conclusions based on the presented description:

- Deep learning techniques offer a successful approach in the classification of lung patterns, whether using CNNs trained from scratch as well as pre-trained networks;
- When using Transfer learning in medical imaging problems, considering different fine-tuning strategies might be an useful approach given the similarity gap between the source and target datasets;
- Given a medical imaging classification task, using a pre-trained model developed to perform a similar, less complex and with more data available task have shown to increase the performance of the more complex task.

In addition, looking at the whole process of Transfer Learning, it is possible to identify and describe two critical processes with a large impact on the resulting performance:

1. **Selection of a pre-trained model.** There are a variety of pre-trained models available, and one can evaluate which one would fit best for the specific problem to solve.

2. **Evaluation of fine-tuning options.** Considering aspects like the similarity between the data used to pre-train the model and the specific training dataset available, one can explore multiple fine-tuning approaches and find which one achieves the best performance.

The ImageNet Large Scale Visual Recognition Competition (ILSVRC) [63] consists of a classification challenge where different models are recognized due to promising performances. Since 2012, when the neural network “AlexNet” [34] achieved first place in the competition, CNN have been related to even more significant research as new models with more layers and innovative frameworks have emerged, including “VGG-Net” [47], “ResNet” [64] and “Inception” [65]. Direct comparison between all these models may not be the most realistic representation of each one’s performance, due to different sampling techniques. However, considering parameters like inference time, power and memory consumption is important to help to choose the most suitable model for a specific task [66].

As presented above, fine-tuning is a process that offers multiple implementation strategies. The gap between the source and target domains should be minimized, and one can choose how much the pre-trained weights should be adjusted in fine-tuning by freezing the layers that are not supposed to change during training with the target dataset. Here should be considered that the first layers of the network are related to general features learning, and as the network goes deeper, more complex and abstract representations are learned [67].

Chapter 4

Data Description

This chapter focuses on presenting a detailed description of all data collections used in this work, as well as the required steps to convert the medical images to the correct data type to enable model development. Thus, it includes a first section with a description of the datasets used in model implementation (Section 4.1), and a second Section, where are detailed the steps for converting from the original DICOM to an image array format, with a briefly explanation of some essential pre-processing steps, necessary to transform the data to improve model implementation as well as final results (Section 4.2).

4.1 Datasets

4.1.1 LIDC-IDRI

The Lung Image Database Consortium image collection (LIDC-IDRI) [55] is a lung cancer screening dataset which comprises thoracic CT scans for a total of 1010 patients, alongside with annotated lesions belonging to one of three classes: (1) nodule $\geq 3\text{mm}$, (2) nodule $< 3\text{mm}$ or (3) non-nodule $\geq 3\text{mm}$, made during a two-phase annotation process by four experienced radiologists. Despite including 7371 segmented nodules, only 2669 were labelled as $\geq 3\text{mm}$.

Retrospectively, was noticed that 2 scans of 8 different subjects were included in the dataset increasing the number of available scans to 1018. However, these extra data were retained in the database due to time already invested in the annotation process. Regarding data acquisition, slice thickness ranged from 0.6 to 5.0 mm, with X-ray current from 40 to 627 mA (mean: 222.1 mA) at 120-140 kVp [68].

4.1.2 NSCLC-Radiogenomics

The NSCLC-Radiogenomics dataset [69] is a publicly available collection with CT and PET/CT images for a cohort of 211 patients with NSCLC, being the only public dataset which comprises paired information on lung cancer-related gene mutation status for *EGFR* (mutant: 43, wildtype: 129), *KRAS* (mutant: 38, wildtype: 133) and *ALK* (translocated: 2, wildtype: 155). Additionally,

semantic tumor annotations are included in a controlled vocabulary as well as binary tumor masks, although not available for the entire set of subjects. This dataset includes CT scans obtained using different scanner models and scanning protocols, presenting variations in slice thickness from 0.625 to 3 mm (median: 1.5 mm) and X-ray tube current from 124 to 699 mA (mean 220 mA) at 80–140 kVp (mean 120 kVp) [69]. Tumor segmentation masks are stored as DICOM Segmentation Objects [70], representing 3D binary arrays where voxels belonging to the tumor ROI contain the value 1 and voxels outside are represented by the value 0.

4.1.3 UHC of São João Dataset

In collaboration with the University Hospital Center of São João (UHCSJ), a dataset of 141 patients with lung cancer was developed, including qualitative annotations for 18 semantic features by three radiologists and tumor mask coordinates for 26 patients. Additionally, this collection comprises paired information related to mutation status for an extensive list of genes. Regarding data acquisition protocol, slice thickness ranges from 3.0 to 5.0 mm, with pixel spacing in (x, y) directions of 0.455-0.808 mm.

4.1.4 Lung CT Segmentation Challenge 2017

The Lung CT Segmentation Challenge 2017 [71] is a data collection provided in association with a thoracic segmentation competition and the related conference session conducted at the AAPM 2017 Annual Meeting [72]. The competition provided a benchmark dataset for auto-segmentation algorithms of thoracic organs at risk (OAR): esophagus, heart, lungs and spinal cord, as shown in Figure 4.1. This data collection comprises 60 patients from 3 different institutions with different clinical practices. Thus, CT slice thickness took values of 1mm, 2.5mm and 3mm, depending on the institution, with number of slices ranging from 103 to 279.

The quality of the clinical contours provided by each institution was checked before made available and major inconsistencies were eliminated based on the Radiation Therapy Oncology Group (RTOG) 1106 contouring guidelines [73]. All necessary data was available for download in The Cancer Imaging Archive (TCIA) [74], enabling post-competition data access for thoracic OAR segmentation models development.

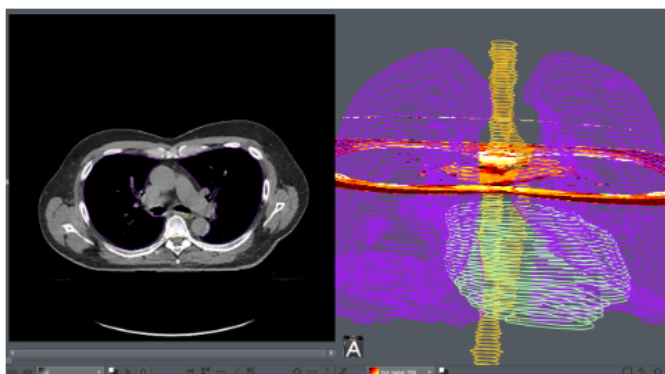


Figure 4.1: Example of CT slice with correspondent appended contour (left) and visualization of all contours for the OAR considered (right). Images obtained with Aliza Medical Imaging DICOM Viewer [75].

4.1.5 ILD Dataset

The ILD database [24] is a collection of lung pathologies built at the University Hospitals of Geneva (HUG), including examples for 13 of the most frequently diagnosed lung parenchyma disorders. Besides providing CT scans for 128 subjects who suffer from different lung diseases, this dataset also made available binary lung segmentation masks for a total of 112 cases. Regarding image acquisition protocol, CT scans present a slice thickness value ranging from 1mm to 2mm and a space between slices of 10-15mm. The axial plane is a 512×512 matrix and pixel spacing in (x, y) directions ranges from 0.4-1mm [24].

Considering the space between slices values previously mentioned, the 3D images available in this collection present a significantly lower number of slices along z axis when comparing with the rest of the databases used in this work.

4.1.6 Inclusion Criteria

Regarding different tasks, not all available patients suited the criteria to be used.

Considering the NSCLC-Radiogenomics dataset [69], despite providing information for a cohort of 211 patients, only 172 have a result for *EGFR* mutation test of either *Mutant* or *Wildtype*, and from those, 163 matched for CT modality. When the tumor segmentation mask was required, the number of useful patients decreased to 116. When multiple exams were available for the same patient, the selected CT scan was the one with *Slice Thickness* equal to 1mm, due to less detail compression.

From the LIDC-IDRI data collection [55], given the need for tumor annotations in this work and its availability only for the nodules labelled as ≥ 3 mm, only these were considered, resulting in 2669 nodules included.

Regarding the availability of lung binary masks, the entire set of CT scans provided by the Lung CT Segmentation Challenge [71] was used, and for the ILD dataset [24], only were included the 112 patients with provided lung mask.

When using the dataset from UHCSJ, only patients with available tumor mask annotated by at least one radiologist were considered, counting for a total of 26 cases.

4.2 Data Pre-processing

4.2.1 DICOM Objects to Image Data Conversion

After the selection of the useful data to be used on each experiment, it was necessary to convert the CT scan stored in DICOM format into an image array in order to perform the necessary processing. Using the *pydicom* package [76], designed to work with data elements of DICOM data, the *.dcm* files are stacked and converted to an image array with the *Pixel Array* attribute. Then, the pixel values were converted to Hounsfield Units (HU) using the attributes *Rescale Intercept* and *Rescale Slope* with tags (0028,1052) and (0028,1053) [77], respectively.

4.2.2 Resampling

When using multiple datasets with inconsistent scanning protocols for the same task, is of utmost importance to standardize these parameters for the entire input data. Having consecutive slices and adjacent pixels separated by different values, depending on the CT scan, a common practice to standardize the entire dataset consists of resampling all images so that the previous distances are equal to 1mm. For example, given an image with dimensions [397, 512, 512] and pixel spacing [0.625, 0.70, 0.70], the resize factor is computed by dividing the original spacing by the new spacing: [0.625/1.00, 0.70/1.00, 0.70/1.00]. The resampled image dimensions would be calculated using the computed resize factor: [0.625, 0.70, 0.70] \times [397, 512, 512] = [248, 358, 358], achieved by rounding. Then, the real resize factor is obtained by dividing the resampled image by the original image dimensions: [248/397, 358/512, 358/512] = [0.625, 0.699, 0.699]. The real new spacing is then computed as [0.625/0.625, 0.70/0.699, 0.70/0.699] = [1.00, 1.00, 1.00]. It is important to note that computations in this explanation rounded values to three decimal cases; however, in all results were held eight decimal cases. In this case, the real new spacing result is not exactly [1.00, 1.00, 1.00]. The resampled image is then obtained by interpolation using the exactly computed real new spacing value for the previously computed dimensions. This operation ensured that the whole set of data used to develop the model presented a similar representation, which is an important aspect in Deep Learning model development.

4.2.3 HU Normalization

When extracting the information of a CT scan, each voxel of the 3D reconstruction is represented by a value related to the correspondent tissue density and measured in the HU scale, as previously detailed in Section 2.3.2. Thus, all CT images were normalized by using the min-max normalization [78], where values under -1000 HU (HU_{min}), which corresponds to air's radiodensity value, were transformed into 0 and values above 400 HU (HU_{max}), representing hard tissues like bones,

were transformed into 1. A linear transformation was computed to map all values in the middle into the [0, 1] intended range as shown in Equation 4.1.

$$Image_{normalized} = \frac{Image_{original} - HU_{min}}{HU_{max} - HU_{min}} \quad (4.1)$$

4.3 Summary

Given the description of each dataset used in this work, Table 4.1 summarizes the number of considered samples for each task according to the inclusion criteria detailed.

Table 4.1: Number of patients from each dataset used considering each proposed task.

	Task	# Included Samples	Dataset
Nodule Analysis	Feature learning	2669	LIDC-IDRI [55]
	Malignancy classification	1095	LIDC-IDRI [55]
	<i>EGFR</i> prediction	116	NSCLC-Radiogenomics [69]
Lung Analysis	Lung segmentation	60	Lung CT Challenge [71]
		112	ILD Dataset [24]
	Feature learning	116	NSCLC-Radiogenomics [69]
	<i>EGFR</i> prediction	116	NSCLC-Radiogenomics [69]
		26	UHC São João

In nodule analysis, as only the nodules labelled as ≥ 3 mm were extracted from the LIDC-IDRI dataset [55], the entire set was included in the feature learning task. Using the provided malignancy annotation value by each radiology and averaging the given values for each nodule resulted in only 1095 included nodules, with 306 considered as malignant and 789 benign. Given the *EGFR* mutation status assessment in this analysis, only the 116 patients with tumor binary mask available were considered from the NSCLC-Radiogenomics collection [69].

When extending the analysed ROI to a lung axial slice, the whole set of 60 patients provided by the Lung CT Challenge [71] dataset was included, adding 112 more patients from the ILD dataset [24], being the ones with lung binary mask available. Considering the lung feature learning and *EGFR* mutation status prediction tasks, only patients with provided tumor mask were included due to the need of merging this mask, counting for 116 patients from the NSCLC-Radiogenomics dataset [69] and 26 from the UHC of São João dataset.

Chapter 5

Lung Nodule Characterisation

The present Chapter addresses the experiments conducted for the local nodule analysis. To employ the Transfer Learning technique under investigation in this work, a Convolutional Autoencoder is proposed in Section 5.1; to investigate the relevance of learned patterns in this ROI, in Section 5.2 a Transfer Learning based classification model intended to distinguish between benign and malignant nodules is implemented, and the same approach is applied in the *EGFR* mutation status assessment task in Section 5.3.

5.1 Unsupervised Feature Learning

Deep Learning architectures require a large number of trainable parameters, becoming imperative to collect a substantial amount of data in order to prevent the model to overfit the training data. However, in any medical field, obtaining more data to train a model is almost always a very complex task.

Given the ability of Transfer Learning techniques to reduce significantly the trainable weights, improving the model ability to generalize over unseen data, neural networks pre-trained on natural images have been widely explored and performances in medical imaging related tasks have proven the viability of this approach. However, similarities between natural and medical images are reduced, and in some cases, deeper fine-tuning approaches are necessary to reduce this gap between such different domains, increasing the amount of necessary data at the same time [18].

5.1.1 Convolutional Autoencoder (CAE)

One way to address this problem is to make advantage of the unsupervised self-learning abilities of Convolutional Autoencoders. In the encoding phase, a CAE transforms the input data into a lower dimensional structure while preserving the original spatial representation, extending the fully-connected AE; then a decoder is applied to reconstruct the compressed representation into the original data. Having the original input as target, the CAE learns the best features that enable input reconstruction while eliminating the need for labelled data.

An important advantage with the application of this dimensionality reduction technique lies in the idea that the network is capable of learning a good representation of the input data, providing an useful weight initialization if one intends to use the coding layers for a different task in the same domain.

This approach has been explored when attempting a classification task with a reduced number of labelled examples, consisting of a different manner of knowledge transfer by using the pre-trained parameters of the encoder layers [42, 79].

5.1.2 Proposed CAE

The proposed CAE architecture in this study is represented in Figure 5.1. As one can see, the encoding phase comprises four convolutional layers (C1 to C4) with 3×3 kernels and increasing number of filters as the network goes deeper. All the convolutional layers are followed by a Rectified Linear Unit (ReLU) activation function and a 2×2 strided max-pooling layer to reduce the output feature map by half. Giving an input tensor of size $C \times H \times W$, passing through the encoding layers results in a feature map of 256 filters with size $\frac{H}{8} \times \frac{W}{8}$. To reconstruct the original input, 3 max-unpooling layers were employed to double the input size before each of the first 3 convolution blocks (C5 to C7), from $(\frac{H}{8}, \frac{W}{8})$ to (H, W) . The last convolutional layer (C8) is followed by a sigmoid activation, ensuring that all output pixels belong to $[0, 1]$ range of values.

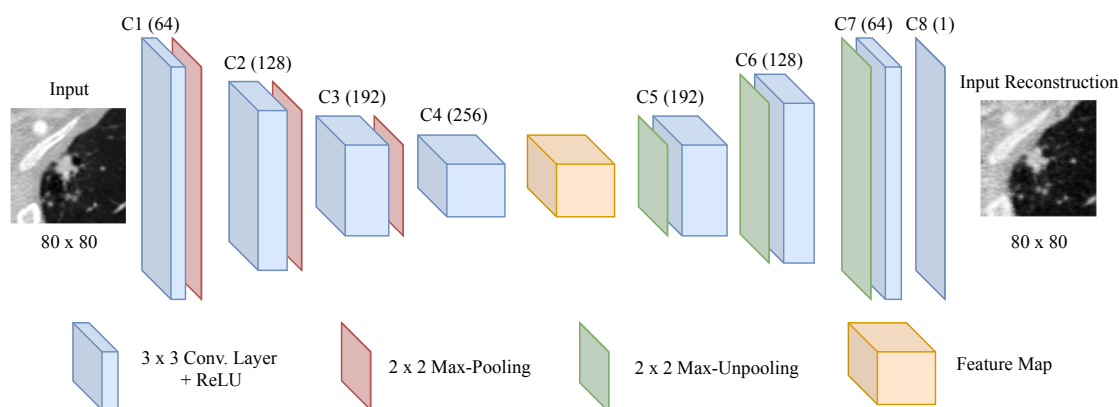


Figure 5.1: Proposed CAE architecture for lung nodule unsupervised feature learning by optimizing input reconstructions.

5.1.3 Materials and Methods

As the goal of this experiment was to extract nodule general features, an image of size $(80 \times 80 \times 80)$ voxels centered on the nodule was extracted for each of the 2669 examples included in this task. The provided nodule masks were averaged and a threshold = 0.5 was applied to obtain the final mask. Then, the mask axial slices were analysed to return the intended nodule cube. Given the fact that the proposed CAE was developed to receive 2D images as input, middle slices from the axial, coronal and sagittal planes were sampled to be used as input, as illustrated in Figure 5.2 below.

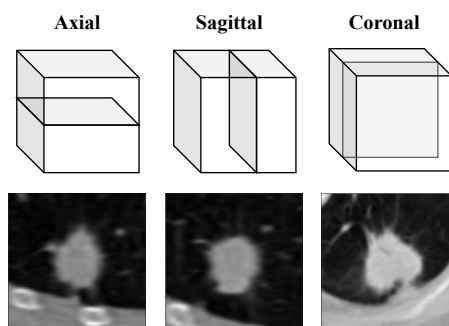


Figure 5.2: Lung nodule patch extraction example: middle slice from axial, sagittal and coronal planes from the same 3D nodule.

The experiments were conducted in order to optimize the reconstruction output. Even though a perfect reconstruction does not ensure the best set of learned features, the input/output similarities were the only optimization criteria used. A simple manual search was employed to find the best set of hyper-parameters, with values represented in Table 5.1.

Table 5.1: Set of hyper-parameters values used in CAE manual-search.

Hyper-parameter	Range values
Learning Rate	0.0001, 0.001, 0.01, 0.1
Optimizer	SGD, ¹ Adam
Momentum	0.1, 0.5, 0.9

¹ Stochastic Gradient Descent.

The entire set of nodules was divided into a training set (90%) and a testing set (10%), and the split was employed at nodule-level, i.e slices of the same nodule were only used for training or testing.

The best results were achieved when using mini-batches of 4 images with Stochastic Gradient Descent (SGD) as the optimizer, learning rate of 0.01 and momentum with value of 0.9. The Mean Squared Error (MSE) was used as loss function, representing the averaged error of each output pixel when compared to the same pixel in the input image. Figure 5.3 shows some examples collected from the test images used in this experiment.

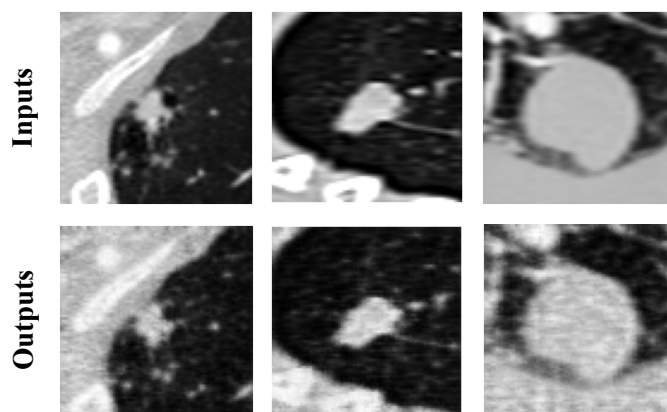


Figure 5.3: Lung nodule examples reconstructed by the proposed CAE. Collected from test images.

5.2 Nodule malignancy classification

To evaluate the self-learned features provided by the CAE, an experiment to classify lung nodules into malignant or benign was conducted. The major goal consisted of evaluating if the pre-trained weights provided by the feature extractor previously trained represented a good weight initialization and helped the model to converge while preventing overfitting, given the significantly reduction in the number of trainable weights.

5.2.1 Materials and Methods

Similarly to the experiments detailed before, the nodule malignancy classification model was developed using the examples annotated as $\geq 3mm$ from the LIDC-IDRI data collection [55]. The ground-truth malignancy value is available for each one of these nodules in the annotation XML file provided and it is an integer value ranging from 1 to 5 with the following designations related to the malignancy degree: (1) *Highly Unlikely*, (2) *Moderately Unlikely*, (3) *Indeterminate*, (4) *Moderately Suspicious* and (5) *Highly Suspicious*. To fit the binary classification of this study, each nodule's malignancy value was averaged over the provided annotations, and a mean malignancy value ≤ 2.0 was considered as benign, and ≥ 4.0 as malignant. Given these criteria, from the 7371 initially available nodules in this database, only 1095 were included in this task after considering the expert's annotations: 789 benign and 306 malignant.

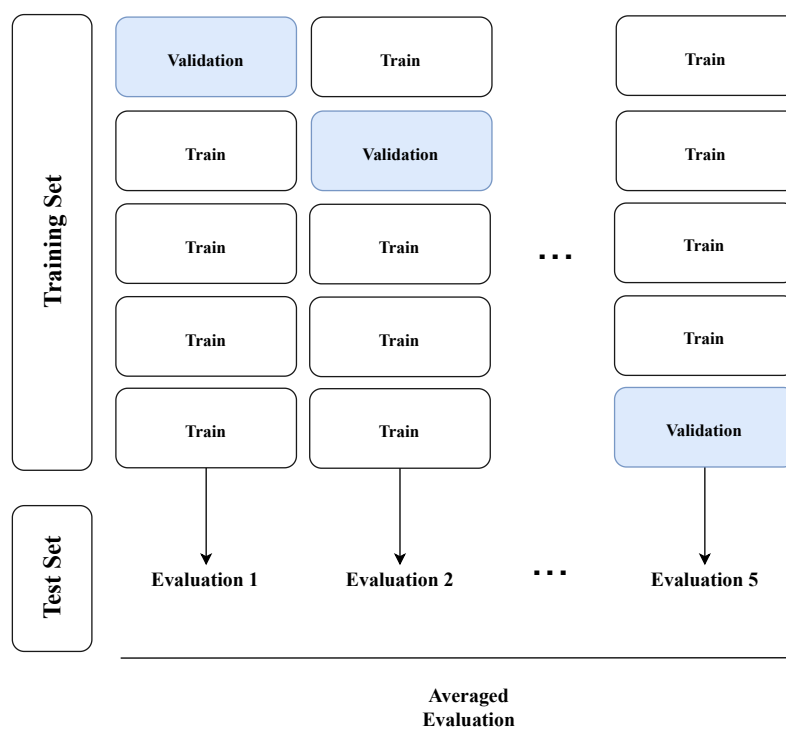


Figure 5.4: 5-Fold Cross-validation overall scheme for performance evaluation in lung nodule malignancy classification.

Data was randomly split into a training set (80%) and a test set (20%), with 5-fold cross-validation applied to the training data, as depicted in Figure 5.4. To achieve a more robust result, 5 random train/test splits were made and evaluation metrics were computed.

In the training phase, from the $(80 \times 80 \times 80)$ extracted cubes centered on the nodule, 3 slices correspondent to the axial, coronal and sagittal planes were sampled to increase the dataset size for the benign nodules, as shown in Figure 5.2. Additionally, to achieve a balanced number of malignant/benign samples, oversampling for malignant class consisted of extracting a total of 7 slices from cube symmetry planes, while only 3 were sampled for benign examples. To increase generalization, some data augmentation was also performed, consisting of horizontal and vertical flips, as well as random image rotations.

5.2.2 Classifier

To perform the intended classification, a Multi-layer Perceptron (MLP) was used as classifier. An MLP is an artificial neural network composed by an input layer where all the input values are received, an output layer with a number of neurons depending on the classification task in hands, and in between a variable number of hidden layers. Since it is a fully-connected neural network, each neuron is connected to all neurons of the following layer.

When approaching a classification task with Transfer Learning techniques, the developed model will consist of the feature extraction pre-trained layers and a classifier stacked on top completely trained with the new target data. Figure 5.5 illustrates the referred framework overview.

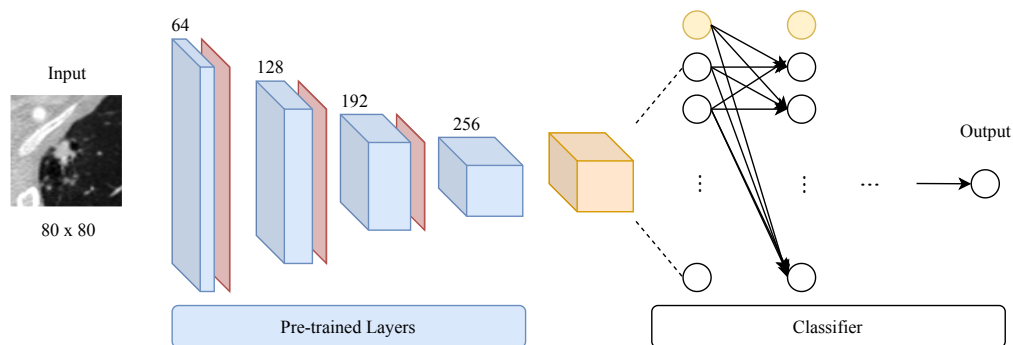


Figure 5.5: Model framework for lung nodule malignancy classification.

The binary output represented the probability of being malignant, with a dynamic decision threshold applied to evaluate this probability. The Binary Cross-Entropy (BCE) loss function was employed to be minimized. For the hyper-parameters manual-search employed, the range of values considered are presented in Table 5.2.

Table 5.2: Set of hyper-parameters values used in the manual search for malignancy classification model.

Hyper-parameter	Range values
Learning Rate	0.0001, 0.001, 0.01, 0.1
Batch-size	8, 16, 32, 64
Momentum	0.1, 0.5, 0.9
Weight decay	0.0001, 0.001, 0.01
Dropout	0.25, 0.5
Hidden Layers	1, 2, 3
Hidden Neurons	32, 64, 128, 256, 512
Optimizer	SGD ¹ , Adam

¹ Stochastic Gradient Descent.

5.2.3 Model fine-tuning

Considering basic observation on CNN behavior, the first convolutional layers of the proposed CAE learn generic features useful for many different tasks, but progressively, as the network goes deeper, more specific patterns are detected and most relevant information is learned regarding the target task [67]. Following this assumption, and given the target and source data similarities, as well as the size of the included training examples, the adopted fine-tuning strategy is illustrated in Figure 5.6.

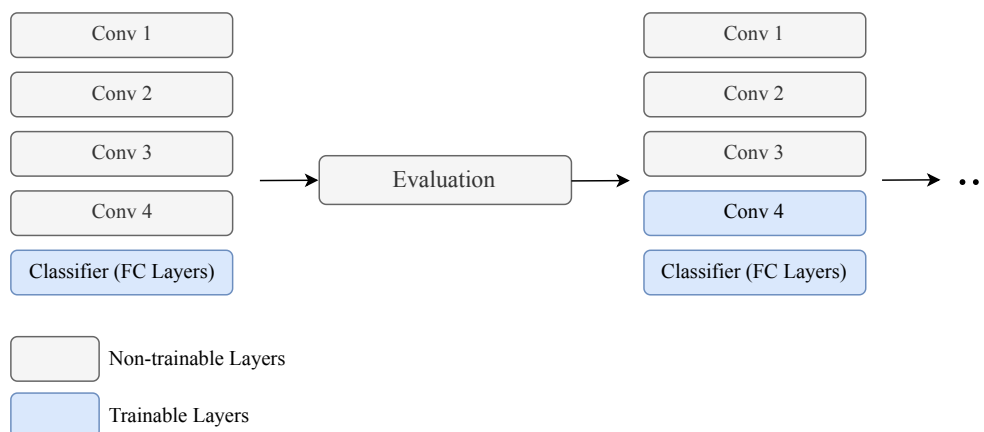


Figure 5.6: Fine-tuning layer-wise approach overview for lung nodule malignancy classification.

As represented above, in training first stage only the classifier FC layers were trained. Given the fact that the feature extractor layers were pre-trained with the same images, similarities and a good convergence were expected, but still far from a local minimum. Thus, to find the best classification performance, the convolutional layers were progressively unfrozen and retrained to detect more representative patterns related to lung nodule malignancy. The major advantage of this layer-wise fine-tuning approach relies on the fact that immediately unfreeze the entire model was not expected to be a viable option given the small amount of training data. Due to large similarities between source and target datasets, it was also considered an unnecessary approach.

5.2.4 Results

The hyper-parameters values that achieve the best performance on test set are presented in Table 5.3. A total number of 5 random train/test splits were employed to investigate data variance and for better performance robustness. In these experiments, the classifier was trained for 200 iterations until converge, then the last convolutional layer (C4) was retrained; finally, after the second visible convergence, gradients were updated for layers C4 and C3 for a final training. This strategy obtained the best results by preventing the model to overfit - which occurred when attempting to retrain layer C2, decreasing performance on validation set. Moreover, the same experiments were conducted with the same architecture completely trained from scratch, assessing the effect of the proposed Transfer Learning approach in this classification task.

For a better understanding of the model performance, different evaluation metrics were computed. Mean values for 5 random splits are presented in Table 5.4. By making advantage of cross-validation data, the decision threshold was tuned for F-score maximization by recall. This threshold optimization was employed by evaluating the precision-recall trade-off which, given the context, represents the cost of a missing malignant nodule (false negative) over a false suspicious of a benign tumor (false positive). Thus, as a missing malignant tumor should be a more penalized error, recall was maximized over precision in this classification task.

Table 5.3: Best hyper-parameters for lung nodule malignancy classification.

Hyper-parameter	Values
Learning Rate	0.001
Batch-size	8
Momentum	0.9
Weight decay	0.0001
Dropout	0.25
Hidden Layers	1
Hidden Neurons	64
Optimizer	SGD ¹

¹ Stochastic Gradient Descent

Table 5.4: Lung nodule malignancy classification results.

Training Method	Performance metrics (mean)			
	AUC	Precision	Recall	F-score
Transfer Learning	0.936	0.794	0.848	0.817
Trained from scratch	0.928	0.842	0.789	0.808

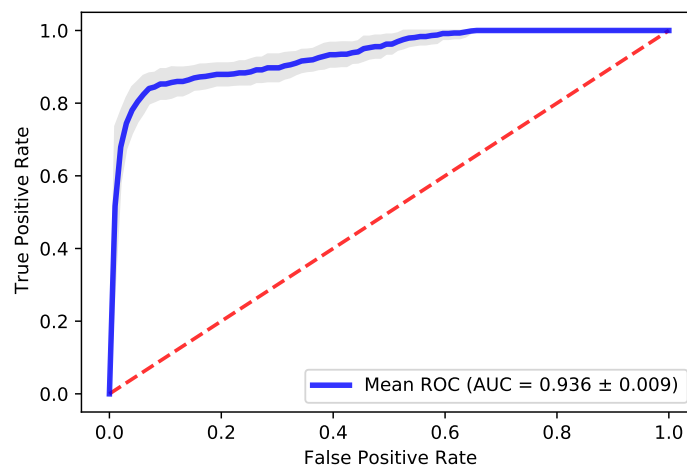


Figure 5.7: Averaged ROC curve for lung nodule malignancy classification using the Transfer Learning approach. The ROC curve is computed for each iteration, the arithmetic average is then calculated and represented by the blue line with a standard deviation represented by the gray shading area. The red dashed-line represents an at-chance classifier ROC curve.

5.2.5 Discussion

Analysing Table 5.4, the results achieved in this task suggest that a Transfer Learning approach based on the CAE unsupervised feature learning abilities is able to prevent the overfitting and perform well with the available data for this task. Given the lung nodule reconstruction results achieved by the proposed CAE (Figure 5.3), and despite a perfect input reconstruction was not achieved, it is possible to say that the patterns learned by the encoder convolutional layers helped the network to avoid the overfitting. When training the network from scratch, besides the overfitting occurrence early in training, it was possible to take the best model weights by maximizing validation AUC before the network starts to overfit the training data, achieving a very promising performance too as can be seen; however, the decision threshold tuning algorithm was not able to find a value that maximized f-score through recall, often resulting in models with more FN than FP predictions on test set, which was not desired in this task.

Considering a more conventional Transfer Learning strategy as using available pre-trained networks on ImageNet data collection as feature extractors, an important advantage provided by the proposed pipeline relies on the similarities between source and target datasets, which is often a problem when applying Transfer Learning in medical imaging tasks.

Several works have proposed CNN based models for lung nodule malignancy classification using data from LIDC-IDRI [55] dataset, either applying Transfer Learning techniques or not. Da Nóbrega et al. [43] proposed an investigation with multiple ImageNet pre-trained feature extractors and different classifiers, with a highest AUC of 0.932 achieved. Yan et al. [80] proposed a comparison between 2D and 3D implementations, achieving a mean AUC of 0.947 with 3D inputs. Kumar et al. [79] proposed to use a denoising autoencoder to extract deep features from 2D lung nodule slices, achieving a classification mean accuracy value of 75.01%, using a decision tree as classifier. Considering these attempts, the major contribution of this work relies on the investigation of the relevance of patterns learned by a Convolutional Autoencoder with any labelling necessary, making advantage of the computational resources lower consumption of a smaller architecture.

Finally, some improvements might be important to note to address different limitations in this work. Lung nodules are 3D elements, and with 2D or even 2.5D implementations, a large portion of useful information might be lost, which makes these perspectives sub-optimal ways of approaching this classification task. However, besides consuming more computational resources, a 3D approach does not allow a slice oversampling operation as employed in this study, which might be a problem given the lower amount of available training data.

5.3 EGFR Prediction

Following the proposed nodule characterisation task, a binary classification to distinguish between *EGFR* mutant or wildtype was conducted. More specifically, it was investigated if the same Transfer Learning technique previously applied for the nodule malignancy classification provided

a good result in this more complex task. Thus, two distinct experiments are presented regarding a small difference in the feature extraction layers' pre-training: besides transferring the pre-trained layers from the proposed CAE encoder, it was investigated if the encoder used in the nodule malignancy classification, after fine-tuning, showed better results. This evaluation was conducted with the idea that more complex patterns related to nodule malignancy could also be associated with *EGFR* mutation status.

5.3.1 Materials and Methods

5.3.1.1 Data

Using the NSCLC-Radiogenomics [69] data collection, 116 nodules of size $(80 \times 80 \times 80)$ were extracted to be used in this task, with 23 and 93 examples for mutant and wildtype classes, respectively.

Data were split into a training and test sets by patient level, ensuring that examples from the same patient would not appear in both sets, and 5-fold cross-validation was applied to the training examples, preventing the model to suffer from some possible bias from any specific input data. The model architecture for this task is the same as represented in Figure 5.5, and the best hyper-parameters were found by manual-search, with range values depicted in Table 5.5

Table 5.5: Set of hyper-parameters values used in the manual-search for *EGFR* classification model.

Hyper-parameter	Range values
Learning Rate	0.0001, 0.001, 0.01
Batch-size	8, 16, 32
Momentum	0.1, 0.5, 0.9
Weight decay	0.0001, 0.001, 0.01
Dropout	0.25, 0.5, 0.75
Hidden Layers	1, 2
Hidden Neurons	32, 64, 128
Optimizer	SGD ¹ , Adam

¹ Stochastic Gradient Descent

5.3.1.2 Overfitting

An important aspect that should be noted considering these experiments related to nodule characterisation lies in the number of features received by the classifier input layer vs number of training examples. As mentioned before, when a (80×80) nodule image passes through the encoder pre-trained layers, the extracted feature map will be a tensor with size $(256 \times 10 \times 10)$, which after a flattening operation is represented as a 1D tensor with 25600 neurons, resulting in more than

800,000 trainable parameters, only considering the first FC layer. Given the reduced number of training examples, this large amount of features is often a problem in any machine learning model, by increasing the chances of overfitting the training data. Since this work presents an approach based on an end-to-end DL structure, the dataset size was increased by oversampling the training examples with slices extracted from cube symmetry planes, improving class balance at the same time. Additionally, a dropout layer was appended before the first FC layer to force the network to learn more robust features by dropping out a neuron of the layer with a probability p at each iteration. These regularization techniques must be applied cautiously, too many of them might cause the model to underfit resulting in performance decrease.

5.3.2 Results

The best results were achieved with the set of hyper-parameters depicted in Table 5.6. Additionally, Table 5.7 shows the achieved performances, with mean values of AUC reported for the two Transfer Learning approaches followed in this classification task.

Table 5.6: Best hyper-parameters for lung nodule malignancy classification.

Hyper-parameter	Values
Learning Rate	0.1
Batch-size	32
Momentum	0.9
Weight decay	0.0001
Dropout	0.5
Hidden Layers	1
Hidden Neurons	32
Optimizer	SGD ¹

¹ Stochastic Gradient Descent

Using the feature extractor pre-trained to detect useful patterns related to nodule malignancy provided slightly better results, with a mean AUC of 0.540 ± 0.0505 averaged over a total of 20 train/test random splits, as represented in the mean ROC curve in Figure 5.8.

Table 5.7: EGFR mutation status prediction results considering 2 different Transfer Learning approaches.

Feature Extractor	AUC (mean \pm standard deviation)
Nodule CAE encoder	0.507 ± 0.055
Nodule malignancy encoder	0.540 ± 0.051

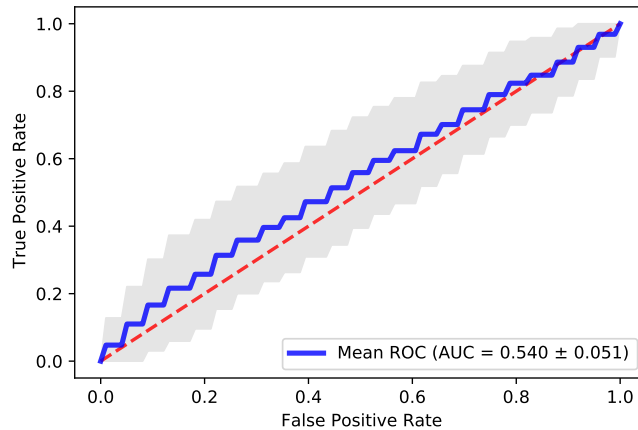


Figure 5.8: Averaged ROC curve for *EGFR* mutation status prediction when using the encoder transferred from nodule malignancy classification. The ROC curve is computed for each iteration, the arithmetic average is calculated and represented by the blue line, with standard deviation as represented by the gray shade area. The red dashed-line represents an at-chance classifier ROC curve.

5.3.3 Discussion

Performance results show that detecting relevant patterns related to *EGFR* mutation status is a very complex task. Besides all regularization techniques applied in these experiments, the proposed model was not able to generalize over unseen data.

Considering related works in this task, as far as it was concerned, no other work proposed a Deep Learning based approach using the same dataset used. Xiong et al. [11] was able to find useful features in lung nodules to classify *EGFR* mutation status, by proposing an investigation with different learning methods and input parameters, achieving a mean AUC of 0.838 with a 2D multi-view Transfer Learning approach. Wang et al. [81] also proposed a Deep Learning model to find deep features related to *EGFR* mutation status in 2D lung nodule slices, achieving a mean AUC of 0.81 in validation set. Considering these approaches, the proposed model attempted to perform this task with a smaller dataset and a different Transfer Learning technique.

Taking into account the achieved results, the biggest limitation of these experiments is the reduced size of the dataset used, which makes difficult to find useful, generalized and representative information to achieve unbiased predictions. Moreover, it might be interesting to explore more complex architectures for feature extraction, given the complexity and abstractness of the patterns that are intended to be captured.

5.4 Summary

In this Chapter, different experiments were conducted regarding a lung nodule characterisation. In Section 5.1, a CAE architecture was proposed with the intention to learn important represen-

tations of tumor-related patterns, and by using Transfer Learning strategies, in Section 5.2 was investigated if the pre-trained encoder layers represented a useful feature extractor by classifying lung nodules into benign or malignant. Moreover, an attempt to predict *EGFR* mutation status by evaluating two Transfer Learning approaches was also conducted (Section 5.3), where it was suggested that the available data did not allow the network to extract representative features, given the lack of heterogeneous information. All things considered, there are some conclusions possible to take:

- Nodule general features learned by the CAE encoding layers can be considered as a reliable weight initialization to apply Transfer Learning techniques in order to perform a malignancy classification task, being able to outperform a model trained from scratch under the same circumstances;
- Besides only supported by a small performance improvement, the hypothesis that patterns related to nodule malignancy might be useful in a more complex task as *EGFR* mutation status prediction was not refuted and it might make sense to continue to be considered;

Chapter 6

Lung Characterisation

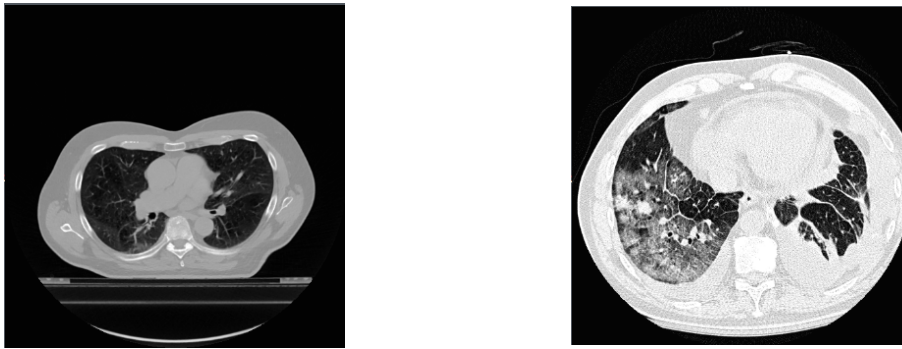
This Chapter presents experiments conducted for a lung characterisation using a holistic analysis. As it was only employed a 2D approach, the analysed ROI was extended to a lung section/slice and not the entire volumetric region. In Section 6.1 is detailed the the implementation of an adapted U-Net [82] to obtain a lung segmentation mask in a completely automatic manner. All data used to develop this model is detailed, as well as the performance evaluation given the available ground-truth for this task; similarly to Chapter 5, Section 6.2 addresses the same feature learning approach, where a CAE is implemented to learn relevant patterns inside the given lung section; finally, in Section 6.3 a classification task is attempted to predict *EGFR* mutation status.

6.1 Lung Segmentation

With the need of a holistic lung analysis, implementing a segmentation model which enabled an efficient lung representation was considered a pre-processing task of utmost importance, often included to perform other detection or classification tasks due to eliminating areas of non-relevant information. The complexity of this task is often related with the presence of abnormal tissues in lung parenchyma like large nodules, blood vessels and different tissue inflammations. Figure 6.1 illustrates two examples of different lung slices from two datasets used in this work where is possible to visualize some differences. In a general way, testing a lung segmentation model on these slices, chances of seeing a performance decrease for the image on the right are usually high. Thus, it is important to develop models able to perform the desired task given different circumstances.

6.1.1 Materials and Methods

For these experiments, the images used were collected from the Lung CT Segmentation Challenge 2017 [71] and the ILD Dataset [24]. It is important to note that, in the majority of cases, the tumor is excluded from the provided lung mask ground-truth in the challenge dataset. The ILD collection made available a total of 112 CT scans with correspondent lung binary mask. Given the multiple number of lung diseases found in the second database, these images represented a very



(a) Lung axial slice extracted from the Lung CT Segmentation Challenge [71].

(b) Lung axial slice extracted from the ILD dataset [24].

Figure 6.1: Comparison on two axial lung slices. From these, it is possible to expect that lung sections with multiple pathologies with imaging manifestation (Figure 6.1b) would make it more difficult to perform a lung segmentation task than with a more "clean" slice (Figure 6.1a).

useful dataset to test the model's performance. Pre-processing steps applied in these experiments included an image crop by body size, using the Otsu's method for histogram-based thresholding¹, followed by a resize to (256×256) by interpolation. These operations were considered an important step, reducing the computational consumption and eliminating background areas in the raw CT slice not relevant for this task.

The segmentation model implemented to perform the desired task was based on U-Net [82], a deep neural network especially designed for biomedical image segmentation. The robustness of the implemented model was an important factor in this task, so different datasets were used to performance evaluation. Experiments were conducted using the train/test split as provided for the challenge data collection, and test results were reported. Additionally, to evaluate the developed model robustness the ILD data was used to test. Given the 2D characteristic of these experiments, the ground-truth binary mask was used to locate the axial slice with the largest lung area to detect, and then a random slice from a range of 10 consecutive slices around the largest one was selected for each training iteration.

A simple manual-search was used to select the best set of hyper-parameters, with values ranging as shown in Table 6.1 below.

Table 6.1: Set of hyper-parameters values used in manual-search for lung segmentation model development.

Hyper-parameter	Range values
Learning Rate	0.0001, 0.001, 0.01
Optimizer	SGD ¹ , Adam
Batch-size	4, 8, 16

¹ Stochastic Gradient Descent

¹ Available at: <https://scipy-lectures.org/packages/scikit-image/index.html>.

6.1.2 Segmentation Model

The overall architecture of the implemented model to perform the lung segmentation task is shown in Figure 6.2. The architecture consists of a contraction, bottleneck and expansion phases. When contracting the input, four blocks of 3×3 convolutional layers followed by ReLU activation are employed, with each block's feature map being stored before passing through a 2×2 max-pooling layer with stride 2 for down-sampling. The number of filters duplicates every block so that more complex patterns can be learned by the network. The bottleneck feature map is achieved by two 3×3 convolutional layers with 256 filters each, and then is followed by a strided 2×2 transposed convolutional layer for up-sampling, starting the expansion phase. It consists of four new blocks composed by two 3×3 convolutional layers and one strided 2×2 transposed convolutional layer. The output is achieved by a fully-convolutional layer with only one 1×1 filter, followed by a sigmoid activation. Besides mirroring the contraction phase, the first convolutional layer of each expansion block uses a concatenation of the feature maps provided by the block before and the activations from the contraction block at the same level. This concatenation operation allows the expansion layers to build the desired mask using features not only learned from the bottleneck representation but also learned while contracting the input image [82].

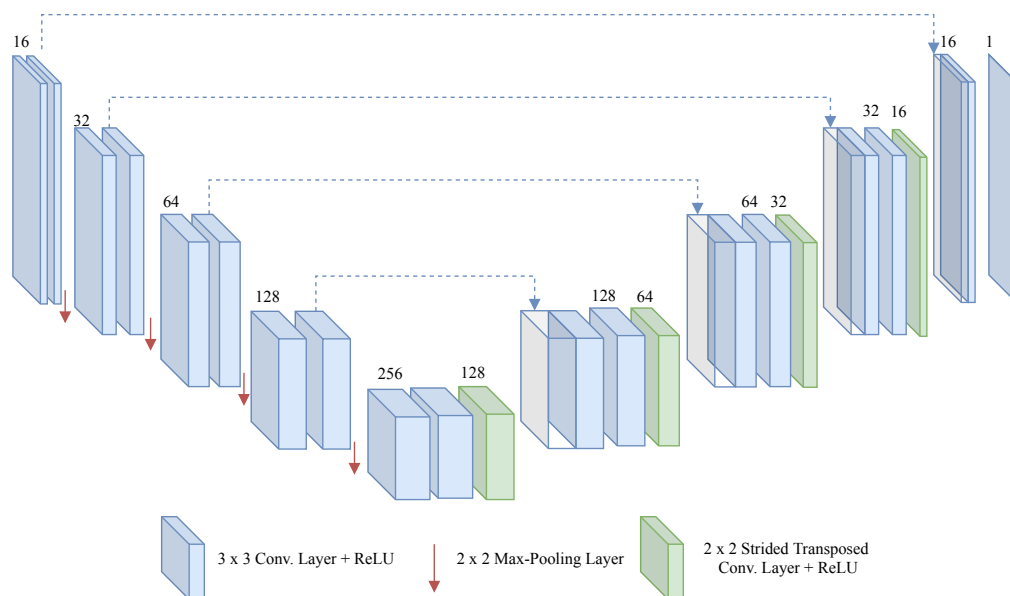


Figure 6.2: Architecture implemented based on U-Net [82], designed to perform the lung segmentation task. The network consists of five contracting and five expansion convolutional blocks. While expanding, each decoder block makes use of features learned in the encoder block with the same feature map ($H \times W$) size.

6.1.3 Results

The hyper-parameters set of values that achieve best performance are presented in Table 6.2. As previously mentioned, to investigate the model's robustness, a second dataset was used to test. Thus, results are presented in Table 6.3 regarding the experiments conducted with these datasets.

The loss function used to optimize was based on the Dice Coefficient (DSC) performance metric, which basically represents the similarity between the output and target images, by measuring the area of overlap over the area of union, being widely used in semantic segmentation tasks, as can be understood in Equation 6.1. To use this metric as loss function, it should measure the differences between output and target masks, so loss returns a value of $1 - dsc_metric$. Adam optimizer parameters were used as default ($\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$).

$$DSC = 2 \times \frac{output_mask \cap target_mask}{output_mask \cup target_mask} \quad (6.1)$$

Table 6.2: Set of hyper-parameters that achieved highest performance.

Hyper-parameter	Value
Learning Rate	0.001
Optimizer	Adam
Batch-size	4

To test the developed model, given the 2D approach in this segmentation task, experiments were employed changing the range of slices tested. Besides testing using the axial slice with the largest lung area to segment, a second experiment used a wider range of possible slices, giving an idea on the model’s performance when trying to segment slices with fewer lung area. Besides the Dice Coefficient metric, mean values of Hausdorff Distance (HD) were also reported for ten independent tests for each slice range configuration. The HD performance metric was employed to represent the model robustness in a less optimistic way, by measuring the longest distance that a point in one contour need to travel to reach its closest point in the other contour. Figure 6.3 shows an example of a poor segmentation achieved with an expected difficult example from the ILD database, representing the importance of using this dataset to test the model’s robustness.

Table 6.3: Lung segmentation results regarding both datasets used to test.

Test Dataset	Slice Range	Dice Coefficient	Hausdorff Distance (mm)
		(mean \pm standard deviation)	
Lung CT Segmentation Challenge 2017 [71]	<i>largest_slice</i>	0.959 \pm 0.061	4.45 \pm 1.65
	random slice in [<i>largest_slice</i> \pm 50]	0.940 \pm 0.065	4.54 \pm 1.72
ILD Dataset [24]	<i>largest_slice</i>	0.961 \pm 0.041	4.71 \pm 1.37
	random slice in [<i>largest_slice</i> \pm 5]	0.907 \pm 0.160	4.62 \pm 1.24

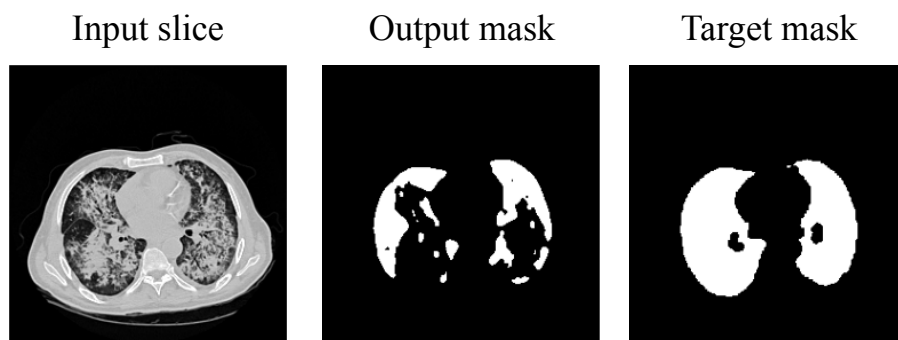


Figure 6.3: Comparison between the output lung mask achieved with the developed segmentation model and the ground-truth mask. Lung axial slice from the ILD dataset [24].

Changing the slice used to test the model was considered an important investigation given the 2D characterisation proposed in this study. Not having an ideally 3D lung segmentation, testing only in one slice for each CT scan would be a very optimistic analysis. Considering the number of slices in each CT scan in the ILD dataset, values ranged from 14 to 60, with mean 26.2. Given this reduced number of slices when comparing with the CT scans provided by the challenge dataset, the test slice range for each dataset was selected to include a similar proportion of the entire lung volume. In this case, an oversampling to standardize space between slices in these two datasets was not employed given the unrealistic results when creating the required pixels by interpolation.

To develop a more robust model, the same architecture was trained using a random train and test sets including examples from both datasets. Values of 0.948 ± 0.015 and 4.21 ± 1.29 were achieved for the DSC and HD metrics, respectively, over five independent tests using the $[largest_slice \pm 5]$ slice range configuration, showing an improvement in both metrics mean values, as well as a large decrease in performance variation results.

6.1.4 Discussion

Regarding the proposed lung segmentation task, it is possible to consider the achieved results as promising. Besides not being ideally, the explained way to evaluate the model performance on different slices with different lung sections within the same CT scan was the best found. The results achieved when testing in the ILD dataset examples were better than expected, which suggests that the developed model is able to perform well on different datasets, which is considered an extremely important factor. Moreover, an attempting to develop a more robust model trained on two very different datasets was conducted resulting in improvements on test performance.

Regarding related works in lung segmentation field, different 2D and 3D based implementations have been proposed. Skourt et al. [83] proposed an U-Net [82] implementation to segment 2D lung axial slices. Manually segmented images from LIDC [55] dataset were used, and an averaged Dice-Coefficient value of 0.950 was achieved. Despite some details not being specified, it should be considered the more similar approach in the literature to the implemented in this study. Considering submissions to the Lung CT Challenge 2017 [71, 72], several works employed

2.5D and 3D U-Net based approaches to lung segmentation, not being possible to obtain a direct comparison from this examples, due to different evaluation parameters.

With a simple architecture, the proposed model was able to perform well with the available input data, leaving space to some improvements as well: despite being computationally more expensive approaches, 2.5D or 3D implementations allow a complete lung analysis which is, without any question, an ideally approach; other architectures with more complex convolutional blocks in feature extraction, extending U-Net, have also been proposed and showed interesting performances in lung segmentation tasks [72, 84], which might be considered as another possible point of improvement in this work.

6.2 CAE for Lung Feature Learning

Following a similar approach to the lung nodule's analysis, a CAE architecture was proposed to make advantage of imaging patterns self-learning ability. After the lung segmentation task, it was possible to develop a CAE to reconstruct lung CT axial slices with only segmented lung sections as non-zero pixel values, allowing the CAE to focus on the lung areas where the potentially useful information might be present. Thus, this task intended to achieve a trained CAE with segmented lungs CT slices to be further applied in a Transfer Learning method to a classification task using the same ROI.

6.2.1 CAE architecture

The proposed CAE architecture for lung axial slice reconstructions is presented on Figure 6.4. Considering the architecture proposed to reconstruct lung nodules, the network implemented for this task is very similar: the encoding phase is composed by five convolutional layers (C1 to C5) with filters size of 3×3 , with a ReLU activation and a 2×2 max-pooling layer following each convolutional block. Thus, given an input with size $(C \times H \times W)$, a $(256 \times \frac{H}{16} \times \frac{W}{16})$ bottleneck represents the coded representation; the input reconstruction is achieved by a decoder with four blocks of a 2×2 strided max-unpooling layer for up-sampling followed by a 3×3 convolutional layer. The 3×3 output convolutional layer maps the 16 feature channels received into a final 2D image, which is followed by a sigmoid activation.

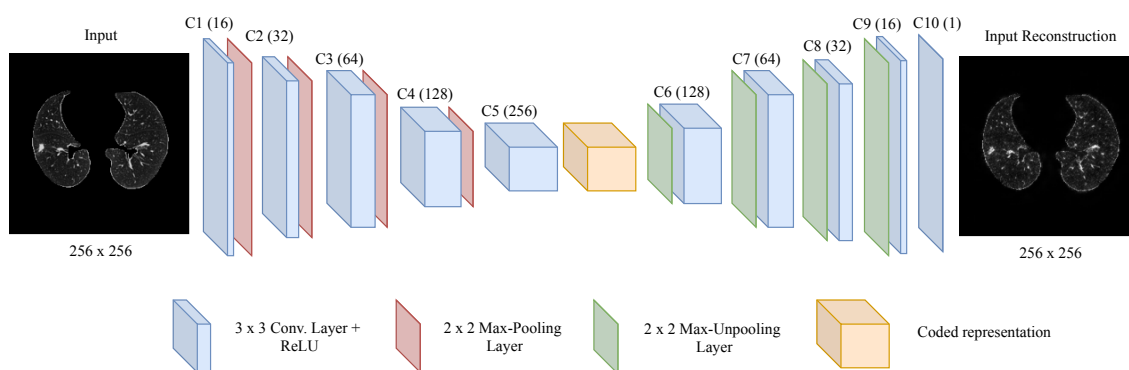


Figure 6.4: Comparison between the output lung mask achieved with the developed segmentation model and the ground-truth mask. Lung axial slice from the ILD dataset [24].

6.2.2 Materials and Methods

To perform this reconstruction task, the data used was collected from the NSCLC-Radiogenomics database [69]. Given the fact that the proposed model for lung segmentation was trained with a ground-truth where the tumors were often excluded, only the CT scans with available nodule mask annotation were included in this task, accounting for a total of 116 patients.

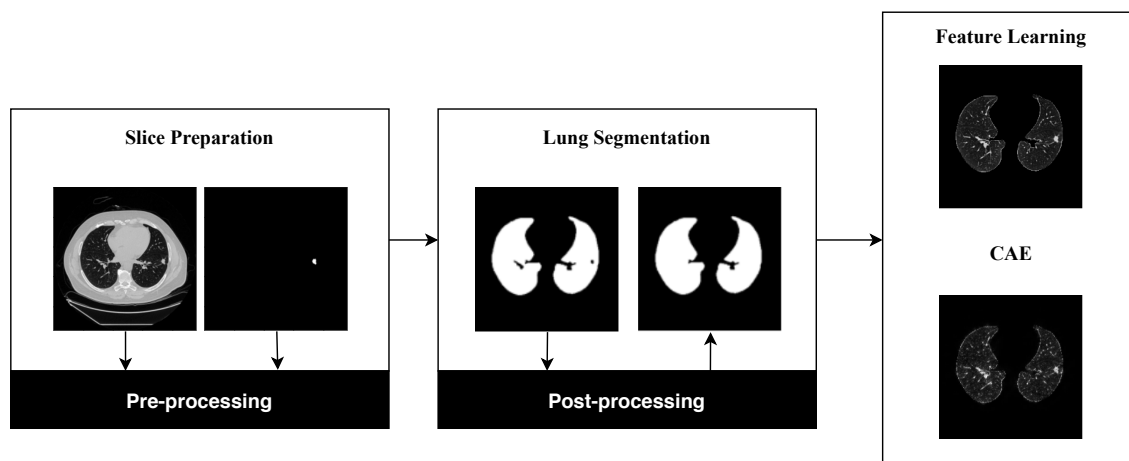


Figure 6.5: Lung CAE data acquisition pipeline. Before passing through the CAE, the represented framework transforms a raw CT slice into an image with segmented lung areas and background. This way, the CAE is not learning any patterns from unnecessary structures outside the lung.

As represented in the overall pipeline for this task in Figure 6.5, the CT scan and the correspondent nodule 3D binary mask are extracted for each patient. Then, given the importance of nodule information, the slice in the CT scan with the largest correspondent nodule mask is selected and before applying the developed lung segmentation model to this slice, an image crop by body size to eliminate unrequired pixel computation was employed, followed by a resize to (256×256) by interpolation. Following the application of lung segmentation model in the pre-processed image, some post-processing operations were employed to clean the output lung mask: (1) the nodule mask is merged given the probability of not being included in the lung segmentation output, (2)

false positive areas identified as lungs close to the image borders are excluded, and (3) some possible "holes" are filled inside the lung mask, to avoid losing important structures not detected as lung areas by the model. The final output is then used to feed the CAE to be reconstructed.

A simple manual-search was employed to find the set of hyper-parameters that allow the best possible image reconstruction. The searched values are depicted in Table 6.4 below:

Table 6.4: Set of hyper-parameters values used in Lung CAE manual-search.

Hyper-parameter	Range values
Learning Rate	0.001, 0.01, 0.1
Optimizer	SGD ¹ , Adam
Batch-size	1, 4, 8

¹ Stochastic Gradient Descent.

The best reconstruction results were achieved with a learning rate of 0.01, SGD as optimizer and mini-batches of one lung slice. The MSE was the cost function used to represent the distance between the input and output images. In these experiments, the model was forced to train for a longer number of iterations to achieve better image reconstructions, stopping the network training after 1000 iterations with no change in loss value. Figure 6.6 represents some results obtained when testing.

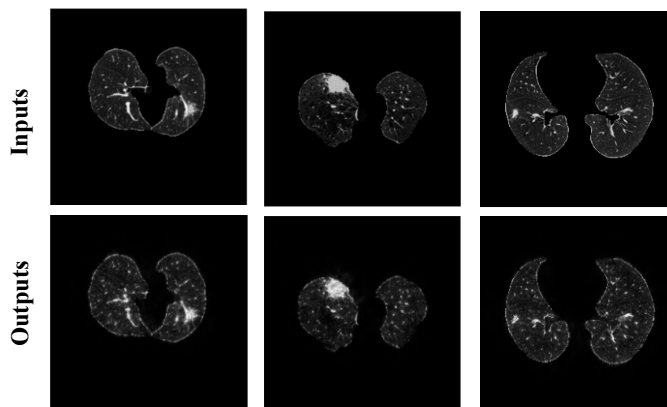


Figure 6.6: Lung axial slice CAE reconstruction examples from test set.

6.3 EGFR Prediction

To complete the proposed lung characterisation, a binary classification to assess the *EGFR* mutation status was conducted using information not only from the nodule, as in 5.3, but also considering a larger ROI including the entire lung section in a selected CT axial slice. The main motivation behind this investigation relied on the hypothesis that it might be possible to find relevant information related to *EGFR* mutation status outside the tumor ROI [9, 16].

6.3.1 Materials and Methods

To implement this task were used the same CT slices as the ones selected to develop the CAE: an axial slice for each case with annotated nodule mask from the NSCLC-Radiogenomics [69] collection, selected based on the correspondent nodule mask area. Having in mind the limitations of a 2D analysis, not having information on the location of relevant patterns to looking for, and assuming that not all slices would contain *EGFR* imaging manifestations, the detailed slice selection method provided confidence that useful patterns would be detected by including the nodule, where is proven that is possible to find *EGFR*-related information [11,81], not passing mislabelled slices to the network at the same time. All things considered, the number of slices used in these experiments were 23 and 93 for mutant and wildtype classes, respectively.

The model architecture proposed follows the same approach as previous classification tasks. The encoder layers from the pre-trained CAE were used as feature extractor, stacking a classifier on top to be completely trained for the new task. The manual-search employed to find the best set of hyper-parameters is presented in Table 6.5. The same fine-tuning layer-wise approach was implemented in these experiments to retrain the feature extractor layers.

The available images were split into training (85%) and test (15%) sets, with data augmentation applied to decrease overfitting. To improve class balance, two additional equidistant slices around the originally selected were extracted for mutant cases only in the training set. Similarly to the *EGFR* assessment experiments in the nodule analysis, a dropout layer was appended before the first FC layer of the classifier, to decrease the number of trainable neurons at each training iteration.

Table 6.5: Set of hyper-parameters values used in manual-search for *EGFR* mutation status assessment.

Hyper-parameter	Range values
Learning Rate	0.0001, 0.001, 0.01
Batch-size	8, 16, 32, 64
Momentum	0.1, 0.5, 0.9
Weight decay	0.0001, 0.001, 0.01
Dropout	0.25, 0.5
Hidden Layers	1, 2, 3
Hidden Neurons	32, 64, 128, 256

In an attempt to increase the available training examples, some cases from the UHC São João dataset were included. Even though limited by the availability of the annotated tumor mask, were included 6 mutant and 20 wildtype examples. These examples were used in two different experiments: (1) as hold-out test set, while training with the entire set of examples from NSCLC-Radiogenomics dataset, and (2) to create a hybrid collection, using examples from both datasets to train and test the model.

6.3.1.1 Lung with nodule selection

In an intermediate investigation, two different experiments were conducted in this analysis: only considering the lung with nodule in the analysed ROI, and considering both lungs. After all pre-processing steps explained above, to obtain an image with only the lung with nodule, one side (left or right) of the image was considered background. More specifically, depending on the nodule side of the image (left or right), the other side was filled with background pixels. By considering a static lung division in the middle of the image it was possible to obtain the desired representation in a fast way. Figure 6.7 illustrates the results for two examples from the NSCLC-Radiogenomics dataset.

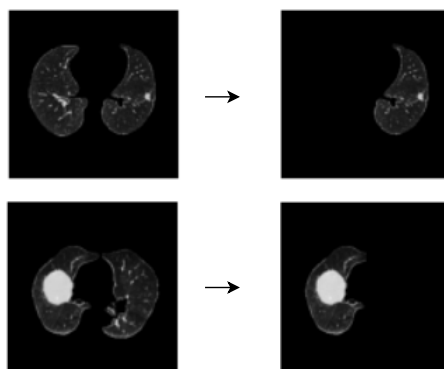


Figure 6.7: Selection of the lung with nodule. Slices from NSCLC-Radiogenomics database patients.

6.3.2 Results

The hyper-parameters that achieved best performance are depicted in Table 6.6. Given the small amount of training data, mean AUC values are averaged over a total of 20 random train/test splits, as presented in Tables 6.7 and 6.8, considering the lung with nodule and both lungs in the analysed ROI, respectively.

Table 6.6: Best hyper-parameters found for lung slice *EGFR* mutation status classification.

Hyper-parameter	Range values
Learning Rate	0.001
Batch-size	32
Momentum	0.9
Weight decay	0.0001
Dropout	0.5
Hidden Layers	1
Hidden Neurons	64

Table 6.7: Lung with nodule: classification results for lung axial slice *EGFR* mutation status prediction.

Train Dataset	Test Dataset	AUC (mean \pm standard deviation)
NSCLC-Radiogenomics	NSCLC-Radiogenomics	0.615 \pm 0.156

Table 6.8: Both lungs: classification results for lung axial slice *EGFR* mutation status prediction. Mean AUC values are depicted regarding the dataset used for each experiment.

Train Dataset	Test Dataset	AUC (mean \pm standard deviation)
NSCLC-Radiogenomics	NSCLC-Radiogenomics	0.645 \pm 0.145
NSCLC-Radiogenomics	UHC São João	0.557 \pm 0.066
NSCLC-Radiogenomics + UHC São João	NSCLC-Radiogenomics + UHC São João	0.556 \pm 0.142

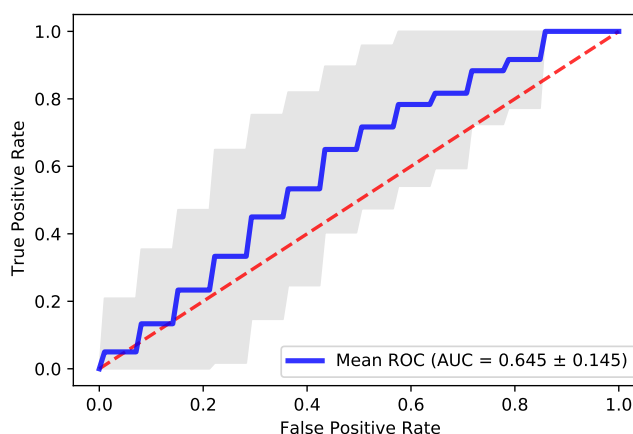


Figure 6.8: Averaged ROC curve for *EGFR* mutation status prediction when using only the examples from NSCLC-Radiogenomics dataset, where the network was able to perform better. The blue line represents the averaged ROC curve over each iteration, with standard deviation as represented by the gray shade area. The red dashed-line represents an at-chance classifier ROC curve.

6.3.3 Discussion

Considering the 2D lung analysis proposed in this investigation, results showed that makes sense to continue to study this holistic hypothesis. The experiments also showed a small performance increase when analysing both lungs, instead of only include the lung with tumor. To the best of gathered knowledge, no other Deep Learning based work attempted to assess *EGFR* mutation status extending the ROI to structures outside the tumor. Besides the scarce availability of data that could be included in this investigation, the proposed 2D analysis achieved a mean AUC of 0.645 ± 0.145 averaged over twenty random combinations for train and test patients, a value

that emphasizes the variance found when classifying unseen images. When joining data from the UHC São João database, the decrease in test performance also supports the lack of representative training data, which is a problem in such a complex classification task.

The performance achieved in previous work (mean AUC = 0.746) [9], where semantic features related to both the nodule and structures outside the nodule ROI were included, suggested that, with the available datasets, working with qualitative assessed information might increase chances of better *EGFR* mutation status predictions. With Deep Learning based approaches, the large set of deep features extracted from a larger ROI might not work well together with such a reduced dataset size, rising the idea that a hybrid approach where deep features combined with semantic *EGFR*-correlated information might gather the best of these two approaches.

6.4 Summary

This Chapter addressed different experiments regarding a lung characterisation. It was implemented a segmentation architecture in Section 6.1, with promising results achieved in a multi-dataset analysis, and in Section 6.2 a CAE was developed to learn imaging patterns related to lung structures. Given the same 2D input ROI, Transfer Learning techniques were applied to assess *EGFR* mutation status (Section 6.3).

Considering the proposed lung characterisation task, the achieved results make possible to take the following conclusions:

- Given a 2D analysis, a simple U-Net [82] based model implementation was capable to perform well on two different datasets, emphasizing the importance to work on a direction that allows robust model implementations able to perform under different circumstances;
- Considering a ROI including an entire lung section in axial view, it was possible to improve performance results over a tumor region analysis when predicting *EGFR* mutation status;
- The importance of more heterogeneous and representative data is a general issue when working in medical imaging field, and working with more abstract imaging manifestations highly increases this necessity to obtain a reliable generalization.

Chapter 7

Conclusions

Advances in precision medicine research have shown that treatment customization depending on each individual's necessities might be considered as a more efficient practice to decrease lung cancer mortality. Thus, the importance of a complete tumor characterisation is emphasized, and given its spatial and temporal heterogeneity, reliable non-invasive methods must be improved to obtain the necessary characterisation. Besides its complexity and wide range of active research points, a tumor characterisation by making advantage of relevant imaging patterns present in medical images has proven its reliability.

The work developed in this dissertation addressed an investigation in lung tumor characterisation by analysing not only the nodule region but also the entire lung section in a 2D perspective using Transfer Learning techniques in Computed Tomography images. Considering the local nodule analysis, the proposed characterisation resulted in a classification model with $AUC = 0.936$ when distinguishing between benign and malignant lung nodules, as well as an attempt to assess *EGFR* mutation status ($AUC = 0.540$). When extending the ROI to a lung axial perspective, a lung segmentation model was implemented achieving a robust performance in a multi-dataset evaluation, and the same input was used to predict the *EGFR* mutation status, achieving an improved classification performance ($AUC = 0.645$). All classification models were implemented with a feature extractor based on a trained Convolutional Autoencoder, investigating the relevance of learned features when trying to reconstruct the input image.

Besides not supported by a large performance improve, using the entire lung axial slice input provided better results in *EGFR* assessment, which emphasizes the idea behind this study's motivation that complex transformations related to lung cancer might be present in other lung structures, and not only in the nodule region. It is considered of utmost importance to continue to investigate this possibility, bringing a new perspective that might change the direction of this research topic. However, to develop models capable to work with larger images with more potentially relevant information, more representative data of the population affected by lung cancer are needed to enable such abstract and complex transformations detection. An important limitation in this work that should also be noted relies on the number of studied genes. An investigation in imaging phenotypes of a more extensive list of lung cancer related genes is necessary to ob-

tain a more complete characterisation, given the importance of other genes in targeted therapies development.

Thereby, in the future, it would be interesting to investigate the possibility to combine relevant qualitative information with abstract *EGFR* manifestations represented by extracted deep features, considering this holistic lung analysis. Providing tested and representative semantic information could help to reduce problems related to the scarce availability of data. Another aspect that should be considered relies on the analysed region of interest. Considering a 2.5D or 3D approach, investigating the possibility of using semantic features to assess the location of potentially useful information might be an interesting effort to extend the explored ROI by a more viable strategy than the entire lung volume analysis, given the computational resources optimization.

References

- [1] American Cancer Society, “Facts & Figures 2019,” Tech. Rep., 2019. [Online]. Available: <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2019/cancer-facts-and-figures-2019.pdf>
- [2] World Health Organisation, “Latest global cancer data: Cancer burden rises to 18.1 million new cases and 9.6 million cancer deaths in 2018,” *Int. Agency Res. Cancer*, 2018. [Online]. Available: <http://gco.iarc.fr/>,
- [3] J. Ferlay, M. Colombet, I. Soerjomataram, T. Dyba, G. Randi, M. Bettio, A. Gavin, O. Visser, and F. Bray, “Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018,” pp. 356–387, nov 2018. [Online]. Available: <https://doi.org/10.1016/j.ejca.2018.07.005>
- [4] V. Hespanhol, B. Parente, A. Araújo, J. Cunha, A. Fernandes, M. Figueiredo, R. Neveda, M. Soares, F. João, and H. Queiroga, “Lung cancer in Northern Portugal: A hospital-based study,” *Rev. Port. Pneumol. (English Ed.)*, vol. 19, no. 6, pp. 245–251, nov 2013. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S2173511513000699>
- [5] J. R. Molina, P. Yang, S. D. Cassivi, S. E. Schild, and A. A. Adjei, “Non-small cell lung cancer: Epidemiology, risk factors, treatment, and survivorship,” in *Mayo Clin. Proc.*, vol. 83, no. 5. Elsevier Ltd, 2008, pp. 584–594. [Online]. Available: <https://doi.org/10.4065/83.5.584>
- [6] T. Lu, X. Yang, Y. Huang, M. Zhao, M. Li, K. Ma, J. Yin, C. Zhan, and Q. Wang, “Trends in the incidence, treatment, and survival of patients with lung cancer in the last four decades,” *Cancer Manag. Res.*, vol. 11, pp. 943–953, 2019.
- [7] J. Marrugo-Ramírez, M. Mir, and J. Samitier, “Blood-based cancer biomarkers in liquid biopsy: A promising non-invasive alternative to tissue biopsy,” *Int. J. Mol. Sci.*, vol. 19, no. 10, p. 21, 2018.
- [8] P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R. G. Van Stiphout, P. Granton, C. M. Zegers, R. Gillies, R. Boellard, A. Dekker, and H. J. Aerts, “Radiomics: Extracting more information from medical images using advanced feature analysis,” *Eur. J. Cancer*, vol. 48, no. 4, pp. 441–446, 2012. [Online]. Available: <https://doi.org/10.1016/j.ejca.2011.11.036>
- [9] G. Pinheiro, T. Pereira, C. Dias, C. Freitas, V. Hespanhol, J. L. Costa, A. Cunha, and H. P. Oliveira, “Identifying relationships between imaging phenotypes and lung cancer-related mutation status: EGFR and KRAS,” p. 10, 2019. [Online]. Available: <https://doi.org/10.1101/794123>

- [10] Z. Bodalal, S. Trebeschi, T. D. L. Nguyen-Kim, W. Schats, and R. Beets-Tan, "Radiogenomics: bridging imaging and genomics," pp. 1960–1984, jun 2019.
- [11] J. Xiong, X. Li, L. Lu, L. H. Schwartz, X. Fu, J. Zhao, and B. Zhao, "Implementation Strategy of a CNN Model Affects the Performance of CT Assessment of EGFR Mutation Status in Lung Cancer Patients," *IEEE Access*, vol. 7, pp. 64 583–64 591, 2019. [Online]. Available: <https://ieeexplore.ieee.org/document/8713576/>
- [12] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "A survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, pp. 11–26, apr 2017.
- [13] W. Jiang, G. Cai, P. C. Hu, and Y. Wang, "Personalized medicine in non-small cell lung cancer: a review from a pharmacogenomics perspective," pp. 530–538, jul 2018.
- [14] S. E. Jorge, S. S. Kobayashi, and D. B. Costa, "Epidermal growth factor receptor (EGFR) mutations in lung cancer: Preclinical and clinical data," *Brazilian J. Med. Biol. Res.*, vol. 47, no. 11, pp. 929–939, nov 2014.
- [15] Z. Schrank, G. Chhabra, L. Lin, T. Iderzorig, C. Osude, N. Khan, A. Kuckovic, S. Singh, R. J. Miller, and N. Puri, "Current molecular-targeted therapies in NSCLC and their mechanism of resistance," *Cancers (Basel)*, vol. 10, no. 7, p. 17, 2018. [Online]. Available: <https://doi.org/10.3390/cancers10070224>
- [16] O. Gevaert, S. Echegaray, A. Khuong, C. D. Hoang, J. B. Shrager, K. C. Jensen, G. J. Berry, H. H. Guo, C. Lau, S. K. Plevritis, D. L. Rubin, S. Napel, and A. N. Leung, "Predictive radiogenomics modeling of EGFR mutation status in lung cancer," Tech. Rep., jan 2017.
- [17] M. M. W. Wille, L. H. Thomsen, A. Dirksen, J. Petersen, J. H. Pedersen, and S. B. Shaker, "Emphysema progression is visually detectable in low-dose CT in continuous but not in former smokers," *Eur. Radiol.*, vol. 24, no. 11, pp. 2692–2699, 2014.
- [18] R. K. Samala, H. P. Chan, L. Hadjiiski, M. A. Helvie, C. D. Richter, and K. H. Cha, "Breast cancer diagnosis in digital breast tomosynthesis: Effects of training sample size on multi-stage transfer learning using deep neural nets," *IEEE Trans. Med. Imaging*, vol. 38, no. 3, pp. 686–696, mar 2019.
- [19] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?" Tech. Rep. 5, 2016.
- [20] American Cancer Society, "Lung Cancer Survival Rates | 5-Year Survival Rates for Lung Cancer," 2019. [Online]. Available: <https://www.cancer.org/cancer/lung-cancer/detection-diagnosis-staging/survival-rates.html>
- [21] G. Siravegna, B. Mussolin, T. Venesio, S. Marsoni, J. Seoane, C. Dive, N. Papadopoulos, S. Kopetz, R. B. Corcoran, L. L. Siu, and A. Bardelli, "How liquid biopsies can change clinical practice in oncology," pp. 1580–1590, oct 2019.
- [22] Y. Bai and H. Zhao, "Liquid biopsy in tumors: opportunities and challenges," *Ann. Transl. Med.*, vol. 6, no. S1, pp. S89–S89, nov 2018.
- [23] A. Kalra, "Developing FE Human Models From Medical Images," in *Basic Finite Elem. Method as Appl. to Inj. Biomech.*, 2018, pp. 389–415.

- [24] A. Depeursinge, A. Vargas, A. Platon, A. Geissbuhler, P. A. Poletti, and H. Müller, “Building a reference multimedia database for interstitial lung diseases,” *Comput. Med. Imaging Graph.*, vol. 36, no. 3, pp. 227–238, apr 2012.
- [25] D. Bermejo-Peláez, S. Y. Ash, G. R. Washko, R. San José Estépar, and M. J. Ledesma-Carbayo, “Classification of Interstitial Lung Abnormality Patterns with an Ensemble of Deep Convolutional Neural Networks,” *Sci. Rep.*, vol. 10, no. 1, p. 15, dec 2020.
- [26] Q. Wang, Y. Zheng, G. Yang, W. Jin, X. Chen, and Y. Yin, “Multiscale Rotation-Invariant Convolutional Neural Networks for Lung Texture Classification,” *IEEE J. Biomed. Heal. Informatics*, vol. 22, no. 1, pp. 184–195, jan 2018.
- [27] I. Sluimer, A. Schilham, M. Prokop, and B. Van Ginneken, “Computer analysis of computed tomography scans of the lung: A survey,” *IEEE Trans. Med. Imaging*, vol. 25, no. 4, pp. 385–405, apr 2006.
- [28] S. Huang, F. Lee, R. Miao, Q. Si, C. Lu, and Q. Chen, “A deep convolutional neural network architecture for interstitial lung disease pattern classification,” *Med. Biol. Eng. Comput.*, jan 2020. [Online]. Available: <http://link.springer.com/10.1007/s11517-019-02111-w>
- [29] G. Kylberg, “Kylberg Texture Dataset v. 1.0,” Tech. Rep., 2011. [Online]. Available: <http://www.cb.uu.se/~gustaf/texture/>
- [30] R. Joyseeree, S. Otálora, H. Müller, and A. Depeursinge, “Fusing learned representations from Riesz Filters and Deep CNN for lung tissue classification,” *Med. Image Anal.*, vol. 56, pp. 172–183, aug 2019.
- [31] M. Negahdar, A. Coy, and D. Beymer, “An End-to-End Deep Learning Pipeline for Emphysema Quantification Using Multi-label Learning,” in *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* Institute of Electrical and Electronics Engineers (IEEE), oct 2019, pp. 929–932.
- [32] “Lung Tissue Research Consortium.” [Online]. Available: <https://ltrcpublic.com/>
- [33] M. Gao, U. Bagci, L. Lu, A. Wu, M. Buty, H. C. Shin, H. Roth, G. Z. Papadakis, A. Depeursinge, R. M. Summers, Z. Xu, and D. J. Mollura, “Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks,” *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.*, vol. 6, no. 1, pp. 1–6, jan 2018.
- [34] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Commun. ACM*, vol. 60, no. 6, pp. 84–90, jun 2017.
- [35] S. Christodoulidis, M. Anthimopoulos, L. Ebner, A. Christe, and S. Mougiakakou, “Multi-source Transfer Learning with Convolutional Neural Networks for Lung Pattern Analysis,” *IEEE J. Biomed. Heal. Informatics*, vol. 21, no. 1, pp. 76–84, jan 2017.
- [36] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, “Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network,” *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1207–1216, may 2016.
- [37] P. Monkam, S. Qi, H. Ma, W. Gao, Y. Yao, and W. Qian, “Detection and Classification of Pulmonary Nodules Using Convolutional Neural Networks: A Survey,” *IEEE Access*, vol. 7, pp. 78 075–78 091, 2019.

- [38] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, “Dermatologist-level classification of skin cancer with deep neural networks,” *Nature*, vol. 542, no. 7639, pp. 115–118, feb 2017.
- [39] S. Basheera and M. S. Sai Ram, “Convolution neural network–based Alzheimer’s disease classification using hybrid enhanced independent component analysis based segmented gray matter of T2 weighted magnetic resonance imaging with clinical valuation,” *Alzheimer’s Dement. Transl. Res. Clin. Interv.*, vol. 5, pp. 974–986, jan 2019.
- [40] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in *2009 IEEE Conf. Comput. Vis. pattern Recognit.* Institute of Electrical and Electronics Engineers (IEEE), mar 2010, pp. 248–255.
- [41] H. Azizpour, A. S. Razavian, J. Sullivan, A. Maki, and S. Carlsson, “Factors of Transferability for a Generic ConvNet Representation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 9, pp. 1790–1802, jun 2016. [Online]. Available: <http://arxiv.org/abs/1406.5774>
- [42] K. Oh, Y. C. Chung, K. W. Kim, W. S. Kim, and I. S. Oh, “Classification and Visualization of Alzheimer’s Disease using Volumetric Convolutional Neural Network and Transfer Learning,” *Sci. Rep.*, vol. 9, no. 1, pp. 1–16, dec 2019.
- [43] R. V. M. Da Nóbrega, S. A. Peixoto, S. P. P. Da Silva, and P. P. R. Filho, “Lung Nodule Classification via Deep Transfer Learning in CT Lung Images,” in *Proc. - IEEE Symp. Comput. Med. Syst.*, vol. 2018-June. Institute of Electrical and Electronics Engineers Inc., jul 2018, pp. 244–249.
- [44] “USF Digital Mammography Home Page.” [Online]. Available: <http://www.eng.usf.edu/cvprg/Mammography/Database.html>
- [45] Alzheimer’s Disease Neuroimaging Initiative, “ADNI | Alzheimer’s Disease Neuroimaging Initiative,” 2017. [Online]. Available: <http://adni.loni.usc.edu/>
- [46] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the Inception Architecture for Computer Vision,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* IEEE Computer Society, dec 2016, pp. 2818–2826.
- [47] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.* International Conference on Learning Representations, ICLR, 2015.
- [48] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” Tech. Rep., apr 2017. [Online]. Available: <http://arxiv.org/abs/1704.04861>
- [49] F. Chollet, “Xception: Deep Learning with Depthwise Separable Convolutions,” Tech. Rep., oct 2016. [Online]. Available: <http://arxiv.org/abs/1610.02357>
- [50] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the Inception Architecture for Computer Vision,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem. IEEE Computer Society, dec 2016, pp. 2818–2826.
- [51] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” Tech. Rep., 2015. [Online]. Available: <http://image-net.org/challenges/LSVRC/2015/>

- [52] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, “Inception-v4, inception-ResNet and the impact of residual connections on learning,” in *31st AAAI Conf. Artif. Intell. AAAI 2017*. AAAI press, 2017, pp. 4278–4284.
- [53] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” p. 9, aug 2018. [Online]. Available: <http://arxiv.org/abs/1608.06993>
- [54] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, “Learning Transferable Architectures for Scalable Image Recognition,” p. 14, jul 2018. [Online]. Available: <http://arxiv.org/abs/1707.07012>
- [55] L. Armato III, SG; McLennan, G; Bidaut, L; McNitt-Gray, MF; Meyer, CR; Reeves, AP; Zhao, B; Aberle, DR; Henschke, CI; Hoffman, Eric A; Kazerooni, EA; MacMahon, H; van Beek, EJ; Yankelevitz, D; Biancardi, AM; Bland, PH; Brown, MS; Engelmann, RM; Laderach, GE, “Data From LIDC-IDRI. The Cancer Imaging Archive.” 2015.
- [56] D. D. Lewis, “Naive(Bayes)at forty: The independence assumption in information retrieval,” in *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 1398. Springer Verlag, 1998, pp. 4–15.
- [57] C. Cortes and V. Vapnik, “Support-vector networks,” *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, sep 1995.
- [58] Z. Zhang, “Introduction to machine learning: K-nearest neighbors,” *Ann. Transl. Med.*, vol. 4, no. 11, jun 2016.
- [59] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, oct 2001.
- [60] “ImageNet.” [Online]. Available: <http://www.image-net.org/>
- [61] P. University, “WordNet | A Lexical Database for English,” 2010. [Online]. Available: <https://wordnet.princeton.edu/>
- [62] A. Krizhevsky, “CIFAR-10 and CIFAR-100 datasets,” 2009. [Online]. Available: <http://www.cs.toronto.edu/~kriz/cifar.html>
- [63] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, “ImageNet Large Scale Visual Recognition Challenge,” *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, dec 2015.
- [64] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem. IEEE Computer Society, dec 2016, pp. 770–778.
- [65] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. June-2015. IEEE Computer Society, oct 2015, pp. 1–9.
- [66] A. Canziani, A. Paszke, and E. Culurciello, “An Analysis of Deep Neural Network Models for Practical Applications,” p. 7, may 2016. [Online]. Available: <http://arxiv.org/abs/1605.07678>

- [67] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *Proc. 2017 Int. Conf. Eng. Technol. ICET 2017*, vol. 2018-Janua. Institute of Electrical and Electronics Engineers Inc., mar 2018, pp. 1–6.
- [68] S. G. Armato, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberle, C. I. Henschke, E. A. Hoffman, E. A. Kazerooni, H. MacMahon, E. J. Van Beek, D. Yankelevitz, A. M. Biancardi, P. H. Bland, M. S. Brown, R. M. Engelmann, G. E. Laderach, D. Max, R. C. Pais, D. P. Qing, R. Y. Roberts, A. R. Smith, A. Starkey, P. Batra, P. Caligiuri, A. Farooqi, G. W. Gladish, C. M. Jude, R. F. Munden, I. Petkovska, L. E. Quint, L. H. Schwartz, B. Sundaram, L. E. Dodd, C. Fenimore, D. Gur, N. Petrick, J. Freymann, J. Kirby, B. Hughes, A. Vande Casteele, S. Gupte, M. Sallam, M. D. Heath, M. H. Kuhn, E. Dharaiya, R. Burns, D. S. Fryd, M. Salganicoff, V. Anand, U. Shreter, S. Vastagh, B. Y. Croft, and L. P. Clarke, "The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans," *Med. Phys.*, vol. 38, no. 2, pp. 915–931, 2011. [Online]. Available: [/pmc/articles/PMC3041807/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3041807/](https://pubmed.ncbi.nlm.nih.gov/21876020/)
- [69] S. Bakr, O. Gevaert, S. Echegaray, K. Ayers, M. Zhou, M. Shafiq, H. Zheng, J. A. Benson, W. Zhang, A. N. Leung, M. Kadoch, C. D. Hoang, J. Shrager, A. Quon, D. L. Rubin, S. K. Plevritis, and S. Napel, "Data descriptor: A radiogenomic dataset of non-small cell lung cancer," *Sci. Data*, vol. 5, 2018.
- [70] C. E. Kahn, J. A. Carrino, M. J. Flynn, D. J. Peck, and S. C. Horii, "DICOM and Radiology: Past, Present, and Future," *J. Am. Coll. Radiol.*, vol. 4, no. 9, pp. 652–657, 2007.
- [71] M. Yang, Jinzhong; Sharp, Greg; Veeraraghavan, Harini ; van Elmpt, Wouter ; Dekker, Andre; Lustberg, Tim; Gooding, "Data from Lung CT Segmentation Challenge. The Cancer Imaging Archive." 2017.
- [72] J. Yang, H. Veeraraghavan, S. G. Armato, K. Farahani, J. S. Kirby, J. Kalpathy-Kramer, W. van Elmpt, A. Dekker, X. Han, X. Feng, P. Aljabar, B. Oliveira, B. van der Heyden, L. Zamdborg, D. Lam, M. Gooding, and G. C. Sharp, "Autosegmentation for thoracic radiation treatment planning: A grand challenge at AAPM 2017," *Med. Phys.*, vol. 45, no. 10, pp. 4568–4581, oct 2018.
- [73] F. M. Kong, T. Ritter, D. J. Quint, S. Senan, L. E. Gaspar, R. U. Komaki, C. W. Hurkmans, R. Timmerman, A. Bezjak, J. D. Bradley, B. Movsas, L. Marsh, P. Okunieff, H. Choy, and W. J. Curran, "Consideration of dose limits for organs at risk of thoracic radiotherapy: Atlas for lung, proximal bronchial tree, esophagus, spinal cord, ribs, and brachial plexus," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 81, no. 5, pp. 1442–1457, dec 2011. [Online]. Available: [/pmc/articles/PMC3933280/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC3933280/](https://pubmed.ncbi.nlm.nih.gov/21876020/)
- [74] F. Prior, K. Smith, A. Sharma, J. Kirby, L. Tarbox, K. Clark, W. Bennett, T. Nolan, and J. Freymann, "The public cancer radiology imaging collections of The Cancer Imaging Archive," *Sci. data*, vol. 4, p. 170124, sep 2017.
- [75] "Aliza Medical Imaging & DICOM Viewer." [Online]. Available: <https://sites.google.com/site/alizaviewer>
- [76] "pydicom · PyPI." [Online]. Available: <https://pypi.org/project/pydicom/>

- [77] “CT Image CIOD – DICOM Standard Browser.” [Online]. Available: <https://dicom.innolitics.com/ciods/ct-image>
- [78] S. Gopal, K. Patro, and K. Kumar Sahu, “Normalization: A Preprocessing Stage,” Tech. Rep., 2015. [Online]. Available: www.kiplinger.com,
- [79] D. Kumar, A. Wong, and D. A. Clausi, “Lung Nodule Classification Using Deep Features in CT Images,” in *Proc. -2015 12th Conf. Comput. Robot Vision, CRV 2015*. Institute of Electrical and Electronics Engineers Inc., jul 2015, pp. 133–138.
- [80] X. Yan, J. Pang, H. Qi, Y. Zhu, C. Bai, X. Geng, M. Liu, D. Terzopoulos, and X. Ding, “Classification of lung nodule malignancy risk on computed tomography images using convolutional neural network: A comparison between 2D and 3D strategies,” in *Lect. Notes Comput. Sci.*, vol. 10118 LNCS. Springer Verlag, 2017, pp. 91–101.
- [81] S. Wang, J. Shi, Z. Ye, D. Dong, D. Yu, M. Zhou, Y. Liu, O. Gevaert, K. Wang, Y. Zhu, H. Zhou, Z. Liu, and J. Tian, “Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning,” *Eur. Respir. J.*, vol. 53, no. 3, mar 2019.
- [82] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9351. Springer Verlag, may 2015, pp. 234–241.
- [83] B. Ait Skourt, A. El Hassani, and A. Majda, “Lung CT image segmentation using deep neural networks,” in *Procedia Comput. Sci.*, vol. 127. Elsevier B.V., jan 2018, pp. 109–113.
- [84] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, “Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation,” Tech. Rep., feb 2018. [Online]. Available: <http://arxiv.org/abs/1802.06955>