

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

# Information Fusion-Based Model for Lung Nodule Characterization

João António Maricato Malva

THESIS MANIFEST



Master in Informatics and Computation Engineering

Supervisor: Eduardo Rodrigues

October 10, 2024



# **Information Fusion-Based Model for Lung Nodule Characterization**

**João António Maricato Malva**

Master in Informatics and Computation Engineering

October 10, 2024

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Context . . . . .	1
1.2	Problem . . . . .	2
1.3	Hypothesis . . . . .	2
1.4	Motivation . . . . .	2
1.5	Research Questions . . . . .	3
	<b>References</b>	<b>4</b>

# Chapter 1

## Introduction

### 1.1 Context

Lung cancer is the leading cause of cancer-related deaths, often diagnosed in an advanced stage, resulting in a low 5-year survival rate of less than 10%, which occurs in 70% of cases, but if detected in an early stage, it is greater than 90% [1]. In 2022, lung cancer had the highest incidence and mortality rates of all cancers worldwide [2]. In particular, in upper-middle-income countries, there has been a significant increase in lung cancer-related deaths, with a rise of 442,000 deaths, more than 2.5 times the increase in deaths of the combined three other income groups [5].

Efforts to reduce lung cancer mortality by screening have been hampered by the aggressive and diverse nature of the disease [4], for example, CT scans help diagnose lung cancer more precisely and produce a reduction of 20% in mortality. Today, the classification of a pulmonary nodule is dependent on measuring the growth rate of that nodule from multiple CT scans and following it for approximately two years to avoid performing a biopsy, which is an invasive procedure for diagnosis. Another downside of slice-by-slice CT scans in lung cancer detection is that they are challenging for doctors, since the amount of data saved in this medical procedure is time-consuming, expensive, prone to reader bias, and requires a high degree of competency and concentration [6].

As medical data becomes more complex, there is a growing need for models that can effectively integrate and analyze these data to support clinical decision-making [3]. Computer-aided diagnosis (CAD) is increasingly being investigated as an alternative and complementary approach to conventional reading, as it avoids many of these issues. Automated nodule diagnostic systems can save both time and money while avoiding the risks of invasive surgical procedures. The non-invasive CAD system for lung nodule diagnosis is promising and has achieved very high accuracy measures from a single CT scan [6].

The combined gains in medical imaging and deep learning complement new approaches that are accurate and safer ways of recognizing diseases. Deep learning models can overcome projections that show how medical images have been analyzed to locate and determine the type of lung abnormalities that are a common cause of cancer.

## 1.2 Problem

This thesis addresses the need for a more accurate and reliable diagnostic tool. Existing diagnostic systems, largely based on deep learning models, have certain limitations when it comes to accuracy and generalization in regards towards medical image datasets. This sort of models are often based on deep features from neural networks, which can overshadow superficial features such as texture and shape that are important for accurately classifying nodules.

In addition, the lack of explainability of the model poses a challenge in clinical contexts, which can limit its reliability in making critical medical decisions. The inability to provide interpretable information hinders the adoption of these methods for the diagnosis of lung cancer.

## 1.3 Hypothesis

Feature extraction is critical for the characterization task, which involves both shallow features (texture and shape descriptors) and deep features learned by deep neural networks (DNNs). The state of the art demonstrates that performance through the application of information fusion techniques could be more efficient in deep learning models when applied in lung nodule characterization [8]. These advancements highlight the need for further research in deep learning and information fusion to prompt early detection and reduce mortality as well as to provide more effective treatment strategies for lung cancer.

We hypothesize that this information fusion-based model approach, with shallow and deep features, will result in a more accurate and reliable model for lung cancer characterization, making it better suited for early detection and precision diagnosis. We seek to overcome the current state-of-the-art limitations in automatic lung cancer diagnostics, offering a solution that not only improves prediction accuracy but also has the potential to assist in clinical decision-making and medical practice.

## 1.4 Motivation

Promoting the improvement of human life and health through early detection of diseases continues to be a concern throughout the world and is also the main objective of Goal 3 (Ensure healthy lives and promote well-being for all at all ages) of the Sustainable Development Goals (SDGs) of the United Nations [7]. Lung cancer is an enemy of public health care and the development of early and accurate diagnostic tools will help improve survival rates. This research aims to contribute to the goal of promoting health by harnessing modern technologies to address one of the greatest diagnostic issues in oncology today. Through the development of models that support more precise care, this dissertation aligns with the global imperative to promote well-being for all.

## 1.5 Research Questions

To bring clarity and precision to the hypothesis, we will break it down into three research questions. These questions will guide the investigation, helping us to understand the main points of our hypothesis.

1. Does fusing information from shallow and deep feature extractors bring any improvement (classification performance, generalization, reduction in the number of model parameters) compared to using only a deep feature extractor?
2. How does this approach compare with an approach that only uses a deep feature extractor when varying the dataset? (e.g. training on one set of data and testing on another, using different amounts of data, among others)
3. In what ways can information-fusion-based models contribute to improving the explainability of lung nodule malignancy predictions?

# References

- [1] Kirsten M. Bahmer, Andrew J. Schoenfeld, and Denise R. Aberle. Benefits and harms of ct screening for lung cancer: A systematic review. *JAMA*, 307(22):2418–2429, 2011. Accessed: 2024-10-03. Cited on page 1.
- [2] International Agency for Research on Cancer. Trachea, bronchus and lung cancer fact sheet. <https://gco.iarc.who.int/media/globocan/factsheets/cancers/15-trachea-bronchus-and-lung-fact-sheet.pdf>, 2024. Accessed: 2024-10-03. Cited on page 1.
- [3] Saeed Iqbal, Adnan N. Qureshi, Musaed Alhussein, Imran Arshad Choudhry, Khursheed Aurangzeb, and Tariq M. Khan. Fusion of textural and visual information for medical image modality retrieval using deep learning-based feature engineering. *IEEE Access*, 11:93238–93253, 2023. Cited on page 1.
- [4] National Lung Screening Trial Research Team, Denise R Aberle, Amanda M Adams, Christine D Berg, William C Black, Jonathan D Clapp, Richard M Fagerstrom, Ilana F Gareen, Constantine Gatsonis, Pamela M Marcus, and JoRean D Sicks. Reduced lung-cancer mortality with low-dose computed tomographic screening. *The New England journal of medicine*, 365(5):395—409, August 2011. Cited on page 1.
- [5] World Health Organization. The top 10 causes of death. <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death>, 2024. Accessed: 2024-10-03. Cited on page 1.
- [6] Ahmed Shaffie, Ahmed Soliman, Amr Eledkawy, Victor van Berkel, and Ayman El-Baz. Computer-assisted image processing system for early assessment of lung nodule malignancy. *Cancers*, 14(5), 2022. Cited on page 1.
- [7] United Nations. The 17 Goals: Sustainable Development Goals, 2015. Accessed: 2024-10-06. Cited on page 2.
- [8] Yutong Xie, Jianpeng Zhang, Yong Xia, Michael Fulham, and Yanning Zhang. Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest ct. *Information Fusion*, 42:102–110, 2018. Cited on page 2.