Early Lung Cancer Detection Using Artificial Intelligence

Moghioroş Eric
Croitoru Andreea Bianca

Prof. Drd. Dioşan Laura
Drd. Limboi Sergiu

May 2025

Contents

1. Introduction

Lung cancer remains one of the deadliest cancers worldwide, with high mortality largely due to late-stage diagnosis. Timely and accurate detection is therefore critical for improving patient outcomes. Advances in artificial intelligence, particularly deep learning, have enabled the development of automated systems capable of analyzing medical images with performance approaching that of radiologists.

This project presents a multi-stage deep learning pipeline based on 2D Convolutional Neural Networks (CNNs) for the comprehensive analysis of chest CT scans. The goal is to assist in the detection, classification, and localization of lung cancer from individual axial slices of CT data.

The proposed system is composed of the following sequential models:

1. **Cancer Detection Model**: A binary classification CNN trained to determine whether a given CT slice contains cancerous tissue. This serves as the entry point of the diagnostic pipeline.

2. **Cancer Type Classification Model**: For slices identified as cancer-positive, a second CNN performs multi-class classification to predict the type of cancer, such as Non-Small Cell Lung Cancer (NSCLC), Small Cell Lung Cancer (SCLC), or other relevant subtypes.

3. **Tumor Localization Model**: A regression-based CNN designed to output bounding box coordinates that localize the suspected tumor region within the slice, providing spatial interpretability and aiding visual diagnosis.

Each model in the pipeline is trained and validated independently to ensure optimal performance for its specific task. By combining classification and localization, the system provides both diagnostic decisions and visual evidence to support clinical workflows. This modular approach also facilitates future improvements and scalability across different datasets and imaging protocols.

Future sections of this document will detail the dataset preprocessing steps, model architectures, training methodologies, evaluation metrics, and performance results.