FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Information Fusion-Based Model for Lung Nodule Characterization

João António Maricato Malva

PDIS

Master in Informatics and Computation Engineering

Supervisor: Eduardo Rodrigues

January 17, 2025

Information Fusion-Based Model for Lung Nodule Characterization

João António Maricato Malva

Master in Informatics and Computation Engineering

January 17, 2025

Contents

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% [8]. In 2022, lung cancer had the highest incidence and mortality rates of all cancers worldwide [9]. 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 [24].

Efforts to reduce lung cancer mortality by screening have been hampered by the aggressive and diverse nature of the disease [22], 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 [1].

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 [12]. 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 [1].

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 [29]. 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 [23]. 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?

Chapter 2

Literature Review

The main goal of this literature review is to define a strategy for completing the state of research on information fusion-based models for lung nodule characterization, with a focus on identifying fusion techniques currently applied in this field. In particular, it has three main objectives: to gain knowledge by analyzing the different types of techniques that have been used in nodule characterization, to discover the specific methods that have been designed to automate nodule characterization and to evaluate the effectiveness of these methods based on the results obtained in the respective studies. However, in addition to achieving these objectives, the review will help to understand the current state of information fusion methods in this area and will lead to a better understanding of the various approaches.

Through recent studies analyses, we will study the most widely adopted techniques, as well as hybrid approaches that combine various strategies. In addition, methods used for automatic nodule characterization will also be analyzed, with a focus on deep learning architectures adapted to CT scan image analysis. This synthesis aims to establish a comprehensive basis for future studies in the characterization of pulmonary nodules and to guide the development of more effective, interpretable and applicable models in a clinical environment.

2.1 Eligibility Criteria

In order to achieve the relevance and rigor of the selected studies, the criteria for the systematic review were specified. These criteria seek to encompass the entire body of research that has been conducted between shallow and deep feature extractors and information fusion techniques in the characterization of CT scans, mainly related to lung nodules. To expand the scope, a more extensive view of applicable methodologies was included for medical conditions that use CT technologies if the studies presented relevant approaches.

In terms of eligibility, only studies published in the last five years (2019 - 2024) were considered, which ensures that the review reflects recent advances in fusion techniques in the field of medical imaging. Articles were limited to those published in English to maintain consistency and accessibility. In addition to the characterization of pulmonary nodules, studies aimed at diagnosing other medical conditions revealed by CT scans were also included, as long as they used methodologies that could be used within the scope of this study.

2.2 Search Strategy

In this review, a comprehensive search strategy was formulated to find relevant studies in various reliable databases. The search was conducted mainly in three academic databases: IEEE Xplore, PubMed and Google Scholar.

The search process employed a set of keywords and Boolean operators to develop comprehensive queries. The primary keywords included terms such as "lung nodule characterization," "information fusion," "shallow feature extraction," "deep feature extraction," and "CT scan analysis." Additionally, secondary terms were included to capture more specific methodologies and techniques, such as "texture features," "shape features," "convolutional neural networks (CNN)," and "medical image classification."

Furthermore, in addition to the database searches, reference chaining was used to expand the list of studies. Specifically, the references of key articles excavated in the initial phase were subjected to scrutiny. As an example, significant references were extracted from the influential article Fusion of Textural and Visual Information for Medical Image Modality Retrieval Using Deep Learning-Based Feature Engineering [12], which has been helpful in the study of fusion techniques for the analysis of pulmonary nodules. Reviewing the references cited in this document helped to identify other studies relevant to the objectives of the initial work, linked by common topics.

2.3 Screening Process

After the search drew out a set of studies that might be relevant, the title and abstract screening process followed. During this phase, each study's titles and abstracts were reviewed to weed out studies that are obviously not relevant. This first screening made it possible to discard the articles that didn't fit the core inclusion criteria, including those that are not associated with CT-based medical imaging or those focusing on work unrelated to the use of shallow or deep features extractors or information fusion methods.

This strategy was used to cut out irrelevant studies in a fast and easy way as well as to let potentially related papers through to the full-text analysis phase. At this stage, the objective was to screen only titles and abstracts. This enabled the easy inclusion of studies that were worth further investigation in subsequent stages.

2.4 Summary

This literature review lays the foundations for the study and research into lung nodule characterization models based on information fusion. With the implementation of a search strategy, this review presents a current selection of studies focused on the characterization of lung nodules and including other diseases based on CT scans. The structured procedure of using the main databases and linking references ensures that relevant studies are obtained, and the selection is further boosted by the use of the title and abstract screening process, which shortens the selection to studies that are actually dealing with research in terms of deep learning, shallow feature extraction and information fusion techniques.

Chapter 3

State of the Art

3.1 Lung Nodule Characterization

The detection of lung nodules is only one of the first steps when the objective is an assertive diagnosis of lung cancer. Once identified, the characteristics of these nodules must be analyzed to establish their risk of malignancy. This process, known as lung nodule characterization, is critical to reducing false positives and guiding more effective clinical decisions.

When discussing characterization, analyzing various features extracted from CT scans is important. These features include the size and volume of the nodules, as larger nodules may be more suspicious [6]. The shape and margins are also significant; benign nodules typically have more regular shapes, while malignant nodules often appear irregular, spiculated, or lobulated [1]. Other factors to consider are texture and density, as these can provide valuable insight into the nodule. [1] [6]

Recognizing that accurate diagnosis is crucial in preventing death from lung cancer, a traditional approach relies on visual assessment by radiologists. However, this method is often time-consuming and may lead to errors, especially when dealing with smaller nodules or those that have more subtle characteristics. The NLST has shown that low-dose CT screening is an effective way of fighting lung cancer, detecting tumors at early stages, and significantly reducing mortality. [22] On the other hand, it's also important to recognize that screening with low-dose CT can result in false-positive outcomes. There was a significantly higher rate of positive results in the low-dose CT group (24.2%) compared to the X-ray group (6.9%). False-positive results in lung cancer screening can lead to invasive and unnecessary diagnostic procedures, which entails risks for patients and additional costs for healthcare systems.

Here is where we found space for new technological advances like CAD systems and other techniques for automated lung characterization, that may help radiologists to interpret images more efficiently and consistently, reducing human error.

3.2 Datasets Overview

This section presents some of the most well-known and widely used datasets in lung nodule characterization tasks.

3.2.1 LIDC-IDRI

The Lung Image Database Consortium Image Collection (LIDC-IDRI) is a comprehensive collection of CT scans of the thorax, designed for the diagnosis of lung cancer and the detection of visualized lesions. This internationally accessible database serves as a valuable resource for the development of computer-aided detection (CAD) systems focused on lung cancer diagnosis and evaluation. Launched by the National Cancer Institute (NCI) and further developed by the Foundation for the National Institutes of Health (FNIH), with support from the Food and Drug Administration (FDA), this public-private partnership exemplifies the success of a consensus-based consortium.

The creation of this data registry involved collaboration among seven academic research centers and eight major medical imaging companies, resulting in a total of 1,018 cases. Each case includes clinical thoracic CT scan images for individual subjects, accompanied by an XML file that details the results of a two-phase image annotation process. In the first phase, four radiologists independently reviewed CT images and annotated lesions into one of three categories: "nodule >=3 mm," "nodule <3 mm," and "non-nodule >=3 mm." During the second phase, the radiologists reviewed their annotations alongside the anonymized annotations of their peers to reach a consensus. This process was designed to allow for the accurate tallying of lung nodules on a CT scan with minimal human intervention, without requiring forced agreement among the radiologists. [7]

3.2.2 LUNA16

The Lung Nodule Analysis 2016 dataset utilizes the publicly available LIDC/IDRI database mentioned earlier. Scans with a slice thickness greater than 2.5 mm were excluded from the dataset. In total, there are 888 CT scans included. The reference standard for this challenge consists of all nodules that are 3 mm or larger, which were accepted by at least 3 out of 4 radiologists. Annotations that are not part of the reference standard, such as non-nodules, nodules smaller than 3 mm, and nodules annotated by only 1 or 2 radiologists, are classified as irrelevant findings. [26]

3.2.3 NLST

The National Lung Screening Trial (NLST), previously mentioned, was a randomized controlled trial conducted by the Lung Screening Study Group (LSS) and the American College of Radiology Imaging Network (ACRIN). The purpose of the trial was to evaluate whether screening for lung cancer with low-dose helical computed tomography (CT) reduces mortality compared to screening with chest radiography in high-risk individuals. Approximately 54,000 participants were enrolled between August 2002 and April 2004. Data collection for the study has concluded, with the final information gathered by December 31, 2009. [22] [21]

3.2.4 Limitations Acknowledge

It is important to recognize that medical imaging datasets significantly contribute to developing and validating computer-aided detection (CAD) systems for lung cancer diagnosis. Nevertheless, even the most used datasets, such as LIDC-IDRI and LUNA16, have some limitations that can impact model performance and generalizability. These challenges include a lack of annotated data, subjectivity in labeling, inter-observer variability, and potential biases within the datasets. These factors complicate the creation of robust and broadly applicable models. Additionally, we acknowledge that the process of annotating medical images is both time-consuming and costly, often leading to datasets that are limited in size. This limitation is further exacerbated by patient data privacy regulations, which restrict access to larger datasets. [10]

3.3 Fusion Techniques

Exploring the state-of-the-art methodologies, we focus on feature fusion techniques that integrate texture, shape, and deep features. Fusion-based models [29] [16] that combine handcrafted features with deep learning representations have shown promise in reducing false positives and improving sensitivity, supporting their potential in clinical scenarios.

We aim to provide a comprehensive understanding of the advancements in nodule characterization. The focus is placed on the role of feature fusion techniques in enhancing diagnostic performance, particularly in terms of accuracy, sensitivity, and robustness against imaging artifacts.

3.3.1 Fusion of Radiomic Features

The following studies focus on the extraction and fusion of radiomic features to improve diagnostic accuracy, combining traditional image analysis approaches with advanced processing techniques.

Farag et al. [3] explored feature fusion by extracting texture descriptors (Gabor filters and Local Binary Patterns - LBP) and shape (signed distance transform fused with LBP). They showed that Gabor filters when implemented on a two-level cascaded framework with Support Vector Machines (SVM) classifiers, obtained the best performance: a mean area under the ROC curve (AUC) of 99% and an F1-score of 97.5%. This approach strongly encourages the premise that feature fusion, in particular with Gabor filters, could improve classification.

Shaffie et al. [2] proposed a framework to accurately diagnose lung nodules by integrating two types of features: appearance features from a seventh-order Gibbs random field model that captures spatial heterogeneities in nodules, and geometric features defining their shape. Then a deep autoencoder classifier uses these features to distinguish between malignant and benign nodules. Evaluated with data from the LIDC, which included 727 nodules from 467 patients, the system demonstrated potential for lung cancer detection, achieving a classification accuracy of 91.20%.

Building on this, the CAD system proposed by Shaffie et al. [1] uses an appearance features descriptor, a Histogram of oriented gradients, a Multi-view analytical Local Binary Pattern, and a Markov Gibbs Random Field, and shape features descriptor, a Multi-view Peripheral Sum Curvature Scale Space, Spherical Harmonics Expansion, and a group of some fundamental morphological features. Then a stacked auto-encoder followed by a soft-max classifier is applied to generate the initial malignancy probability. All of the resultant probabilities are fed to the last network that returns the diagnosis. When comparing with a previous study, they refer that the increase in the accuracy is small (from 93.97% to 94.73%), which is predictable, since the features used model the same nodule characteristics. However, the increase in system sensitivity from 90.48% to 93.97% represents a notable improvement, demonstrating that the new system, with additional features, is less affected by the segmentation process and image artifacts.

We establish that the fusion of radiomic features, particularly texture and shape descriptors, can significantly improve the performance of lung nodule classification systems. These results highlight the significance of incorporating more comprehensive features as a strategy to increase diagnostic accuracy.

3.3.2 Introduction on Deep Learning Approaches

The CNNs can automatically learn complex hierarchical features from raw image data, eliminating the need for complex manual feature engineering. Going deeper, both Halder et al. [11] and Gu et al. [10] recognize the paradigm shift to deep learning-based approaches for detecting and diagnosing pulmonary nodules.

Ali et al. [4] propose a Transferable Texture Convolutional Neural Network (CNN) for lung nodule classification, which architecture consists of three convolutional layers and an Energy Layer, omitting pooling layers to reduce trainable parameters and computational complexity. The EL preserves texture information and learns during both forward and backward propagation. The model was evaluated on the LIDC-IDRI and LUNGx Challenge datasets. The texture CNN achieved an accuracy of 96.69% ± 0.72% and an error rate of 3.30% ± 0.72% on LIDC-IDRI. Transfer learning improved accuracy on LUNGx from 86.14% to 90.91%.

The study by Ali et al. [5] evaluated the performance of Support Vector Machine and AdaBoostM2 algorithms using deep features from VGG-16, VGG-19, GoogLeNet, Inception-V3, ResNet-18, ResNet-50, ResNet-101 e InceptionResNet-V2 by identifying the optimal layers. Their results showed that SVM was more efficient for deep features as compared to AdaBoostM2. The proposed decision-level fusion technique demonstrates better results in terms of accuracy (90.46 ± 0.25%), recovery (90.10 ± 0.44%), and AUC (94.46 ± 0.11%). Although it was ranked second in terms of specificity (92.56 ± 0.18%), the deviation is notably lower compared to the Texture CNN approach [4]. Furthermore, the classification accuracy based on the simple average of the prediction scores is calculated at 89.10%, which highlights the robustness and effectiveness of the decision fusion technique compared to other methods.

3.3.3 Advanced Fusion-Based Approaches

This present section overviews various approaches, including convolutional neural networks, feature fusion methods, attention models, and multimodal learning techniques. These studies range from the application of deep learning methods for classification to the integration of textural and visual information and other more complex approaches. As previous, the primary goal is to enhance the ability to distinguish between benign and malignant nodules, enabling earlier and more accurate diagnosis.

Xie et al. [29] gave us an algorithm, Fuse-TSD, that takes texture, shape, and deep features to automatically classify lung nodules in chest CT images. It uses a texture descriptor based on the GLCM, a Fourier shape descriptor, and a DCNN to extract features. Then classifiers, AdaBoosted Back Propagation Neural Network (BPNN), are applied for each feature and the decision is made by the fusion of the respective results. Evaluated on the LIDC-IDRI dataset, Fuse-TSD achieved an AUC of 96.65% when nodules with a composite malignancy rate of 3 were discarded, 94.45% when they were considered benign, and 81.24% when they were considered malignant.

Saba et al. [25] propose a method for early-stage lung nodule detection, consisting of three main phases: nodule segmentation using Otsu's thresholding and morphological operations, feature extraction of geometric, texture, and deep learning features to select optimal features, and serial fusion of the optimal features for classifying nodules as malignant or benign. The study experiments with the LIDC-IDRI dataset, using Otsu's algorithm and morphological erosion for segmentation. Handcrafted geometric and texture features are combined with deep learning features extracted using a VGG-19 model. Feature optimization is performed using PCA, and the fused features are classified using multiple classifiers. Experimental results show the proposed method outperforms existing approaches, achieving an accuracy of 0.99 with fused features.

Muhammad Muzammil et al. [20] investigates different fusion approaches based on feature fusion and ensemble learning to classify lung nodules in CT scans. The authors propose two heterogeneous fusion techniques: fusion based on the average prediction score (AVG-Predict) and fusion based on majority voting (MAX-VOTE). The results showed that the MAX-VOTE technique, combining the predictions of twelve individual classifiers, achieved the highest accuracy in binary classification, with 95.59% ± 0.27%. While in multi classification, the SVM-FFCAT (Feature Fusion by Concatenation) method achieved superior performance, with an accuracy of 96.89%, an AUC of 99.21%, and a specificity of 97.70%. These results emphasizes that fusion features with ensemble learning can significantly enhance the performance of lung nodule classification.

The CAD system presented by Yuan, Wu e Dai [32] uses a multi-branch classification network with an effective attention mechanism (3D ECA-ResNet) to extract features from 3D images of nodules, adapting dynamically to improve the extraction of key information. Structured data, such as diameter and other radiological characteristics, are transformed into a feature vector. The experimental results show that the system achieves an accuracy of 94.89%, sensitivity of 94.91% and an F1-score of 94.65%, with a false positive rate of 5.55%. The study concludes that the combination of multimodal data increases the effectiveness of the CAD system, make it more likely to be able to assist doctors in diagnosing pulmonary nodules.

The study by Liu et al. [28] emphasizes the need to consider the temporal aspect in analyzing pulmonary nodules. It employs a Faster R-CNN to generate regions of interest and extract temporal and spatial features from lung nodule data. A 3D CNN fuses these features, and a time-modulated LSTM model (T-LSTM) analyzes trends and predicts the evolution and malignancy of lung lesions, achieving an area under the curve (AUC) of 92.71%, surpassing traditional methods like XGBoost and RNN.

Zhao et al. [30] proposed a lung nodule detection method that integrates multi-scale feature fusion. Candidate nodules are detected using a Faster R-CNN with multi-scale features, achieving a sensitivity of 98.6%, a 10% improvement over single-scale models. For false positive reduction, a 3D CNN based on multi-scale fusion achieved 90.5% sensitivity at 4 false positives per scan.

While studying the identification of COVID-19 cases, Mahmoud et al. [17], explored several different learning Deep-Learning Networks for thoracic image retrieval. It used two sets of data focused on the thorax: X-ray and CT scans. Pre-trained models were used, like ResNet-50, AlexNet, and GoogleNet, as feature extractors. Similarity between images was assessed using measures such as City Block and Cosine. ResNet-50 achieved the best accuracy, reaching 99% for positive COVID-19 cases and 98% for negative cases in chest X-rays.

Munoz et al. [19] used a predictive model, such as XGBoost, based on morphological characteristics extracted from CT scans, an approach called “3D-MORPHOMICS”. Its premise is that morphological changes can be quantified and used in the diagnostic process since irregularities in the nodules are indicators of malignancy. The classification model, using only 3D-morphomic features, achieved an AUC (Area Under the ROC Curve) of 96.4% on the NLST test set, and the combination with radiomic features resulted in even better performance, with an AUC of 97.8% on the NLST test set and 95.8% on the LIDC dataset.

Based on Hybrid Deep Learning models, Li et al. [14] proposed a CAD system that integrates deep learning techniques for feature extraction and feature fusion. The system uses VGG16 and VGG19 networks with a Convolutional Block Attention Module (CBAM) to extract relevant features. These features are reduced using Principal Component Analysis (PCA) and fused via Canonical Correlation Analysis (CCA) to create effective representations. The final analysis is performed using an optimized Multiple Kernel Learning Support Vector Machine - Improved Particle Swarm Optimization (MKL-SVM-IPSO). The proposed system achieved 99.56% accuracy, 99.3% sensitivity, and an F1-score of 99.65% on the LUNA16 dataset. These results demonstrate its competitiveness in reducing false positives and negatives in nodule detection.

Li et al. [13] evaluated the effectiveness of fusion models in predicting axillary lymph node (ALN) metastases in breast cancer, comparing traditional radiomics models, deep learning radiomics models, and fusion models using dynamic contrast-enhanced MRI (DCE-MRI) images. The imaging data were sourced from The Cancer Imaging Archive (TCIA) via the Duke-Breast-Cancer-MRI project. Handcrafted radiomic features and deep learning features were extracted from 3062 DCE-MRI images, with feature selection performed using mutual information algorithms and recursive feature elimination. The study found that the decision fusion model, integrating radiomic and deep learning features, achieved an AUC of 0.91, outperforming traditional and deep learning models. Adding clinical features to the decision fusion model further increased the AUC to 0.93. The findings demonstrate the efficacy of fusion models in predicting ALN metastases, with the decision fusion model showing significant potential to aid clinical decision-making in early-stage breast cancer treatment.

Alksas et al. [6] employ an approach that modifies the local ternary pattern (LTP) to use three levels instead of two and a new pattern identification algorithm to capture the heterogeneity and morphology of the nodule. Then the features were given as training data to a classification architecture based on hyper-tuned stacked generalization to classify nodules, achieving an overall accuracy of 96.17%, with 97.14% sensitivity and 95.33% specificity. On the other hand, the original LBP and other classification structures resulted in lower performance when compared to the proposed approach.

Liu et al. [16] present a novel method for classifying benign and malignant lung nodules by combining shallow visual features and deep learning features. The approach utilizes separate pipelines for feature extraction and classification. Shallow features, including texture and morphology, are extracted using statistical 3D data analysis and Haralick's texture model, while morphological features are derived from parameters such as size and shape. Support Vector Machines (SVMs) are employed to classify these extracted features. The deep learning branch uses neural architecture search to design a deep model with three sub-branches and integrates a Convolutional Block Attention Module (CBAM) for enhanced feature learning. The classification results from both shallow and deep models are fused using a weighted voting method.

Saeed Iqbal et al. [12], present with this study an innovative technique for classifying medical image modalities by combining visual and textural features. A pre-trained CNN extracts deep features, while manual methods like Zernike moments, Haralick features, and Global-Local Pyramid Pattern (GLPP) capture relevant textural and statistical attributes. These fused features are used to train ML classifiers such as SVM, KNN, and Decision Trees. The proposed approach outperformed standalone pre-trained CNNs, achieving 95.89% accuracy and 96.31% recall in modality classification.

3.4 Remarks

We all recognize that AI has shown potential for enhancing diagnostic, reducing false positives, and optimizing the management of pulmonary nodules. However, generalization, interpretability, and and clinical integration remain major obstacles. [15] It is essential that these tools are validated on larger datasets that are representative of the population to be applied and that they are integrated into clinical workflows. [31]

Standalone Deep learning approaches still miss on overcoming many challenges, for example, if the segmentation of the node is not accurate, the model may not be able to extract the features correctly, leading to inaccurate classification. [27] [10], allthough feature extraction show better accuracy and lower False Positive rates. Textural features, such as GLCM, are promising for differentiating nodules. The combination of feature extraction methods and neural networks optimizes diagnosis. [18]

Fusion-based techniques showed potential in the classification of pulmonary nodules, addressing the limitations of autonomous feature extraction methods. By making use of supplementary information from various feature domains, these approaches increase the accuracy and reliability of the diagnosis. The studies reviewed here highlight the promise of information fusion as a critical enabler of advanced Computer Aided Diagnosis (CAD) systems, leading the way for better clinical decision-making.

Chapter 4

Method Planning

4.1 Population and sampling

The lung nodule classification study will use a population of CT images from publicly available and clinically validated datasets, generally used for this purpose. We will fully trust in their assignment of labels (benign/malignant) to conduct experiments.

To retrieve insights into the characteristics and quality of the data we will submit it to an analysis and provide some visualization for easier interpretation. Computing and summarizing statistics such as mean, median, and standard deviation, for nodule size, patient demographics, and other clinical variables and distribution of classifications in the used datasets to evaluate balance and ascertain the need to perform sampling techniques. For each dataset selected, pre-processing will be carried out in order to improve the quality of the images: ensuring uniformity if necessary, such as applying resampling techniques, and removing, for example, noise and other irrelevant information.

In order to proceed with the experimentation phase, the data will be randomly divided into training and test sets, with the aim of carrying out evaluations afterward. This division will mitigate possible over-fitting effects.

4.2 Evaluation Strategy

Each combination will be evaluated using standardized cross-validation protocols to ensure robust and generalizable results. Detailed logs of the performance metrics for each run will be maintained to enable comprehensive comparative analysis.

4.3 Deep-Learning Selection

The first step in the research is to detect and manifest deep learning models that are the best of the best and specifically designed for the classification of pulmonary nodules in CT images. The primary target is to create a baseline for performance measurements that are to be taken into consideration. The retrieve models will be subject of a defined evaluation and ranked based on performance. We will use the best performing models to combine them with additional methods and features.

4.4 Shallow Extractors

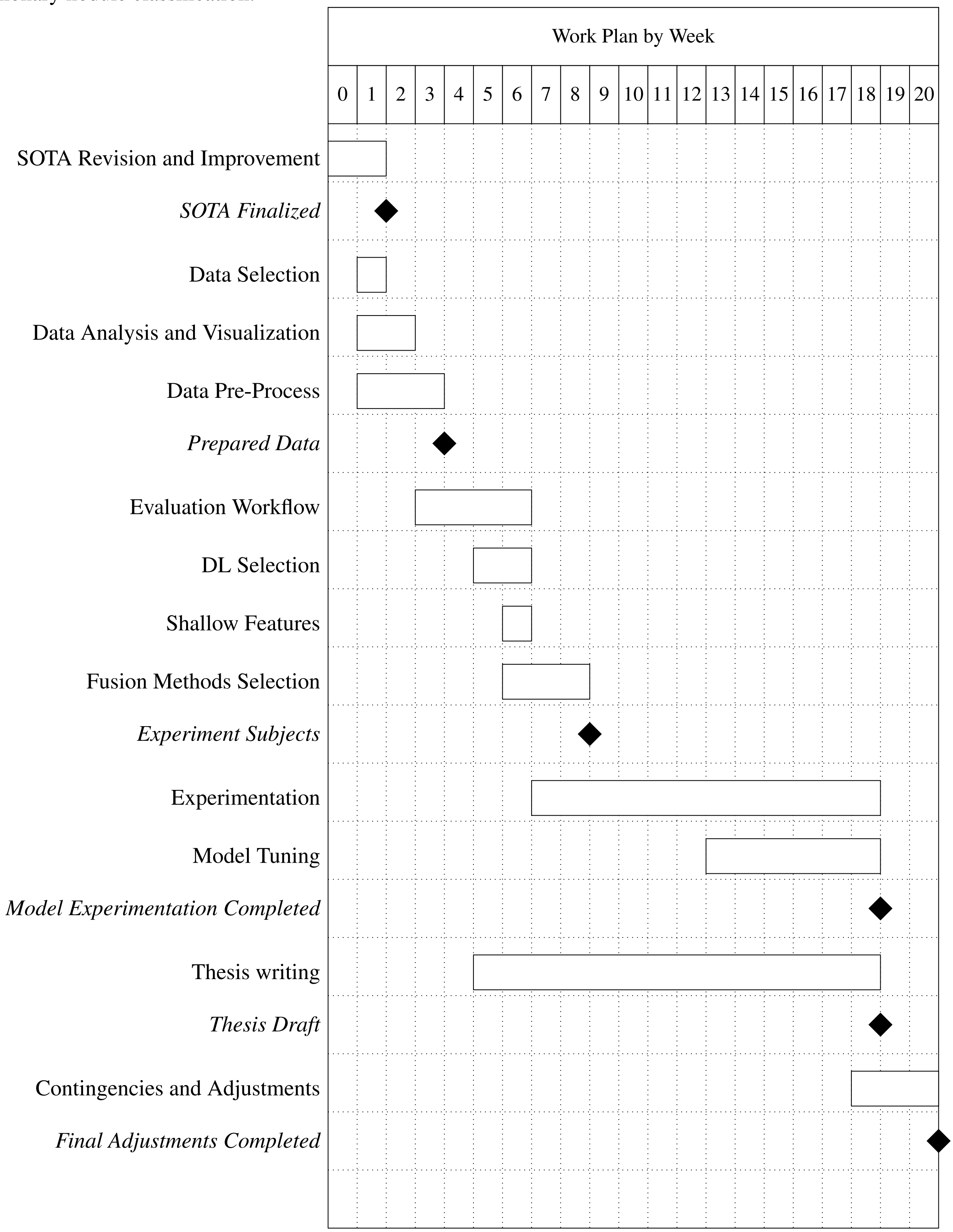
In parallel with deep learning, shallow feature extractors are employed to capture complementary visual and textural information about pulmonary nodules. This stage aims to identify and implement diverse extractors that contribute unique feature sets. We will select the most used extractors for texture, shape, moment-based, gradient-based, and hierarchical features used in the state of the art. The objective is to use them in the same dataset used for deep learning methods and ensuring compatibility with fusion techniques.

4.5 Fusion Methods

The integration of deep learning outputs with shallow feature sets is crucial for leveraging the strengths of both approaches. To achieve this, state-of-the-art fusion techniques will be examined and implemented. Fusion can be applied at two primary levels: decision-level and feature-level. Decision-level fusion combines predictions from multiple models or classifiers using techniques such as weighted voting, majority voting (MAX-VOTE), and prediction averaging, allowing the ensemble of predictions to refine the overall accuracy. On the other hand, feature-level fusion involves concatenating or averaging feature vectors from deep learning models and shallow extractors, often with dimensionality reduction techniques like Principal Component Analysis (PCA) to manage high-dimensional data. These techniques will be implemented in a modular framework, ensuring compatibility with various model configurations.

4.6 Experimentation

The experimentation phase is designed to systematically evaluate the performance of different combinations of deep learning models, shallow feature extractors, and fusion techniques in classifying pulmonary nodules. A flexible workflow will be developed to facilitate the testing of these combinations. This workflow begins by selecting one of the three top-performing deep learning models identified earlier, followed by choosing one to all five shallow feature extractors. The selected outputs are then combined using one of the chosen fusion methods. The pipeline processes the input CT images, extracts features, applies fusion, and trains the combined model. For each configuration, the model's performance will be assessed using metrics such as accuracy, sensitivity, specificity, F1-score, area under the curve (AUC), and false-positive rate (FPR). This systematic approach ensures that all potential interactions between deep learning models, shallow feature extractors, and fusion techniques are explored. The goal is to identify the configurations that deliver the highest classification accuracy, outperforming both standalone deep learning models and traditional classifiers. The experimental results will provide valuable insights into the synergies between deep and shallow methods, establishing a robust foundation for future advancements in pulmonary nodule classification.



References

[1] Shaffie A, Soliman A, Eledkawy A, van Berkel V, and El-Baz A. Computer-assisted image processing system for early assessment of lung nodule malignancy. *Cancers*, 14(5), 2 2022. Cited on pages 1, 7, and 10.

[2] Shaffie A, Soliman A, Fraiwan L, Ghazal M, Taher F, Dunlap N, Wang B, van Berkel V, Keynton R, Elmaghraby A, and El-Baz A. A generalized deep learning-based diagnostic system for early diagnosis of various types of pulmonary nodules. *Technology in cancer research & treatment*, 17:1533033818798800, 1 2018. Cited on page 9.

[3] Farag AA, Ali A, Elshazly S, and Farag AA. Feature fusion for lung nodule classification. *International journal of computer assisted radiology and surgery*, 12(10):1809–1818, 10 2017. Cited on page 9.

[4] Imdad Ali, Muhammad Muzammil, Ihsan Ul Haq, Muhammad Amir, and Suheel Abdullah. Efficient lung nodule classification using transferable texture convolutional neural network. *IEEE Access*, 8:175859–175870, 2020. Cited on page 10.

[5] Imdad Ali, Muhammad Muzammil, Ihsan Ul Haq, Muhammad Amir, and Suheel Abdullah. Deep feature selection and decision level fusion for lungs nodule classification. *IEEE Access*, 9:18962–18973, 2021. Cited on page 10.

[6] Ahmed Alksas, Ahmed Shaffie, Mohammed Ghazal, Fatma Taher, Adel Khelifi, Maha Yaghi, Ahmed Soliman, Eric VAN Bogaert, and Ayman El-Baz. A novel higher order appearance texture analysis to diagnose lung cancer based on a modified local ternary pattern. *Computer Methods and Programs in Biomedicine*, 240:107692, 2023. Cited on pages 7 and 13.

[7] Samuel G. Armato III, Geoffrey McLennan, Luc Bidaut, Michael F. McNitt-Gray, Charles R. Meyer, Anthony P. Reeves, Binsheng Zhao, Denise R. Aberle, Claudia I. Henschke, Eric A. Hoffman, Ella A. Kazerooni, Heber MacMahon, Edwin J. R. Van Beek, David Yankelevitz, Anna M. Biancardi, Paul H. Bland, Martha S. Brown, Robert M. Engelmann, Gregory E. Laderach, David Max, Rosalia C. Pais, Dan P. Y. Qing, Richard Y. Roberts, Andrew R. Smith, Adam Starkey, Prashant Batra, Philip Caligiuri, Aamer Farooqi, Gregory W. Gladish, Christopher M. Jude, Reginald F. Munden, Iva Petkovska, Leland E. Quint, Lawrence H. Schwartz, Balaji Sundaram, Lori E. Dodd, Christopher Fenimore, David Gur, Nicholas Petrick, John Freymann, Justin Kirby, Bethany Hughes, Arnaud Van Casteele, Sachin Gupte, Medhat Sallam, Michael D. Heath, Michael H. Kuhn, Ekta Dharaiya, Ryan Burns, Deborah S. Fryd, Marc Salganicoff, Vinay Anand, Uri Shreter, Gabor Vastagh, Bryan Y. Croft, and Laurence P. Clarke. Data from lidc-idri. Dataset, 2015. Cited on page 8.

[8] 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.

[9] 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.

[10] Yu Gu, Jingqian Chi, Jiaqi Liu, Lidong Yang, Baohua Zhang, Dahua Yu, Ying Zhao, and Xiaoqi Lu. A survey of computer-aided diagnosis of lung nodules from ct scans using deep learning. *Computers in Biology and Medicine*, 137:104806, 2021. Cited on pages 9, 10, and 14.

[11] Amitava Halder, Debangshu Dey, and Anup K. Sadhu. Lung nodule detection from feature engineering to deep learning in thoracic ct images: a comprehensive review. *Journal of Digital Imaging*, 33(3):655–677, 6 2020. Cited on page 10.

[12] 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 pages 1, 5, and 13.

[13] Xue Li, Lifeng Yang, and Xiong Jiao. Comparison of traditional radiomics, deep learning radiomics and fusion methods for axillary lymph node metastasis prediction in breast cancer. *Acad Radiol*, 30(7):1281–1287, November 2022. Cited on page 12.

[14] Yang Li, Hewei Zheng, Xiaoyu Huang, Jiayue Chang, Debiao Hou, and Huimin Lu. Research on lung nodule recognition algorithm based on deep feature fusion and mkl-svm-ipso. *Scientific Reports*, 12(1):17403, 10 2022. Cited on page 12.

[15] Jonathan A. Liu, Issac Y. Yang, and Emily B. Tsai. Artificial intelligence (ai) for lung nodules, from the ajr special series on ai applications. *American Journal of Roentgenology*, 219(5):703–712, 2022. PMID: 35544377. Cited on page 13.

[16] Shaojun Liu, Shujing Wang, Qixiang Wang, and Junhua Luo. Classification of benign and malignant pulmonary nodules based on mixed features. In *2023 42nd Chinese Control Conference (CCC)*, pages 8803–8808, July 2023. Cited on pages 9 and 13.

[17] Sawsan Mahmoud, Hanan Al-Jubouri, and Tawfeeq Abdulabbas. Chest radiographs images retrieval using deep learning networks. *Bulletin of Electrical Engineering and Informatics*, 11:1358–1369, 6 2022. Cited on page 12.

[18] P. Mathumetha, R. Sivakumar, and Ananthakrishna Chintanpalli. Feature extraction methods and deep learning models for detection of cancerous lung nodules at an early stage – recent trends and challenges. *Preprint*, 2024. This preprint research paper has not been peer-reviewed. Available at SSRN: https://ssrn.com/abstract=4756114. Cited on page 14.

[19] Elias Munoz, Pierre Baudot, Van-Khoa Le, Charles Voyton, Benjamin Renoust, Danny Francis, Vladimir Groza, Jean-Christophe Brisset, Ezequiel Geremia, Antoine Iannessi, Yan Liu, and Benoit Huet. 3d-morphomics, morphological features on ct scans for lung nodule malignancy diagnosis - cancer prevention through early detection. In Sharib Ali, Fons van der Sommen, Bartłomiej Władysław Papież, Maureen van Eijnatten, Yueming Jin, and Iris Kolenbrander, editors, *Cancer Prevention Through Early Detection*, pages 3–13, Cham, 2022. Springer Nature Switzerland. Cited on page 12.

[20] Muhammad Muzammil, Imdad Ali, Ihsan Ul Haq, Muhammad Amir, and Suheel Abdullah. Pulmonary nodule classification using feature and ensemble learning-based fusion techniques. *IEEE Access*, 9:113415–113427, 2021. Cited on page 11.

[21] National Lung Screening Trial Research Team. Data from the national lung screening trial (nlst). Dataset, 2013. Cited on page 9.

[22] 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, 8 2011. Cited on pages 1, 7, and 9.

[23] United Nations. The 17 goals: Sustainable development goals, 2015. Accessed: 2024-10-06. Cited on page 2.

[24] 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.

[25] Tanzila Saba, Ahmed Sameh, Fatima Khan, Shafqat Ali Shad, and Muhammad Sharif. Lung nodule detection based on ensemble of hand crafted and deep features. *Journal of Medical Systems*, 43(12):332, November 2019. Cited on page 11.

[26] Arnaud Arindra Adiyoso Setio, Andrea Traverso, Tom de Bel, Maria S. Berens, Casper van den Bogaard, Piergiorgio Cerello, Huanjun Chen, Qi Dou, Maria Enrica Fantacci, Bart Geurts, Rogier van der Gugten, Colin Jacobs, Aruch Jirapatnakul, Geert Litjens, Sergio Gomez Maldonado, Sacha Mucherino, Jesper Hjorth Pedersen, Tobias Penzkofer, Eduard T. Scholten, Michelle Stieren, Freek Tan, Sean Walsh, Xiaochen Wei, Bram van Ginneken, René van Klaveren, Jeroen van der Laak, Cornelia Schaefer-Prokop, Mathias Prokop, Fred Meijer, David Ardila, Xavier Ribo Farre, Panagiotis P. Ypsilantis, Luis M. Montuenga, Alejandro Sánchez-Salcedo, Jose Garcia Nogueira, Jian Pu, and Meindert Niemeijer. Luna16: A challenge for automatic nodule detection in low-dose ct scans, 2016. Data set available at https://luna16.grand-challenge.org/. Cited on page 8.

[27] Ahmed Shaffie, Ahmed Soliman, Hadil Abu Khalifeh, Mohammed Ghazal, Fatma Taher, Adel Elmaghraby, and Ayman El-Baz. A novel framework for accurate and non-invasive pulmonary nodule diagnosis by integrating texture and contour descriptors - 2021 ieee 18th international symposium on biomedical imaging (isbi). In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 1883–1886, 2021. Cited on page 14.

[28] Liu X, Wang M, and Aftab R. Study on the prediction method of long-term benign and malignant pulmonary lesions based on lstm. *Frontiers in bioengineering and biotechnology*, 10:791424, 2022. Cited on page 12.

[29] 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 pages 2, 9, and 11.

[30] Zhao Y, Wang Z, Liu X, Chen Q, Li C, Zhao H, and Wang Z. Pulmonary nodule detection based on multiscale feature fusion. *Computational and mathematical methods in medicine*, 2022:8903037, 2022. Cited on page 12.

[31] Wu YJ and Wu FZ. Ai-enhanced cad in low-dose ct: Balancing accuracy, efficiency, and overdiagnosis in lung cancer screening. *Thoracic cancer*, 11 2024. Cited on page 13.

[32] Haiying Yuan, Yanrui Wu, and Mengfan Dai. Multi-modal feature fusion-based multi-branch classification network for pulmonary nodule malignancy suspiciousness diagnosis. *Journal of Digital Imaging*, 36(2):617–626, 4 2023. Cited on page 11.