Lung Cancer Computer-Aided Diagnosis System (CADx) with 3D Deep Convolutional Neural Networks

Ioannis Marinakis

Supervisor: George Papadourakis, Professor

December 2023

Submitted in partial fulfillment of the requirements for the degree of Master of Science in Informatics Engineering



Hellenic Mediterranean University
School of Engineering
Department of Electrical and Computer Engineering

Abstract

This thesis presents the development and evaluation of a computer-aided diagnostics system for lung cancer screening. Utilizing advanced image analysis techniques, the system enhances the lung cancer screening process by exploring 3D CNNs for segmenting and classifying pulmonary nodules in low-dose CT images.

Early lung cancer detection is critical, and computer-aided diagnostics assist radiologists by automating image analyses. The study comprehensively investigates 3D CNNs for nodule segmentation and classification. Using a public low-dose CT dataset, the architecture learns intricate spatial features for accurate nodule segmentation. Classification capabilities are tuned to distinguish benign and malignant nodules.

Results show promise, enhancing lung cancer screening efficiency. Integration of 3D CNNs improves nodule segmentation, generating accurate 3D predictions. Classification performance advances accurate malignant nodule identification, aiding informed clinical decisions.

In conclusion, this thesis contributes to computer-aided diagnostics by presenting a novel lung cancer screening approach using 3D CNNs. Results highlight the developed system's viability as a valuable tool for radiologists, streamlining the detection and classification of pulmonary nodules in low-dose CT images.

Keywords: computer-aided diagnostics, lung cancer screening, 3D CNN, nodule segmentation, nodule classification, low-dose CT images.

Σύνοψη

Η παρούσα διπλωματική εργασία παρουσιάζει την ανάπτυξη και αξιολόγηση ενός διαγνωστικού συστήματος με τη βοήθεια υπολογιστή για τον προσυμπτωματικό έλεγχο του καρκίνου του πνεύμονα. Χρησιμοποιώντας προηγμένες τεχνικές ανάλυσης εικόνας, το σύστημα ενισχύει τη διαδικασία προσυμπτωματικού ελέγχου του καρκίνου του πνεύμονα εξερευνώντας τρισδιάστατα CNN για την κατάτμηση και ταξινόμηση των πνευμονικών όζων σε εικόνες CT χαμηλής δόσης.

Η πρώιμη ανίχνευση του καρκίνου του πνεύμονα είναι κρίσιμη και τα διαγνωστικά με τη βοήθεια υπολογιστή βοηθούν τους ακτινολόγους αυτοματοποιώντας τις αναλύσεις εικόνας. Η μελέτη διερευνά διεξοδικά τα 3D CNN για την κατάτμηση και την ταξινόμηση των οζιδίων. Χρησιμοποιώντας ένα δημόσιο σύνολο δεδομένων CT χαμηλής δόσης, η αρχιτεκτονική μαθαίνει περίπλοκα χωρικά χαρακτηριστικά για ακριβή τμηματοποίηση οζιδίων. Οι δυνατότητες ταξινόμησης είναι συντονισμένες ώστε να διακρίνουν καλοήθεις και κακοήθεις όζους.

Τα αποτελέσματα δείχνουν πολλά υποσχόμενα, ενισχύοντας την αποτελεσματικότητα του προσυμπτωματικού ελέγχου του καρκίνου του πνεύμονα. Η ενσωμάτωση τρισδιάστατων CNN βελτιώνει την κατάτμηση των οζιδίων, δημιουργώντας ακριβείς 3D προβλέψεις. Η απόδοση της ταξινόμησης προάγει την ακριβή αναγνώριση κακοήθων όζων, βοηθώντας στη λήψη τεκμηριωμένων κλινικών αποφάσεων.

Συμπερασματικά, αυτή η διατριβή συμβάλλει στη διαγνωστική με τη βοήθεια υπολογιστή παρουσιάζοντας μια νέα προσέγγιση προσυμπτωματικού ελέγχου του καρκίνου του πνεύμονα χρησιμοποιώντας τρισδιάστατα CNN. Τα αποτελέσματα υπογραμμίζουν τη βιωσιμότητα του ανεπτυγμένου συστήματος ως πολύτιμου εργαλείου για τους ακτινολόγους, βελτιστοποιώντας την ανίχνευση και την ταξινόμηση των πνευμονικών οζιδίων σε εικόνες CT χαμηλής δόσης.

Λέξεις-κλειδιά: διαγνωστικά με βοήθεια υπολογιστή, προσυμπτωματικός έλεγχος καρκίνου του πνεύμονα, 3D CNN, τμηματοποίηση οζιδίων, ταξινόμηση οζιδίων, εικόνες CT χαμηλής δόσης.

Contents

1 Introduction 1
1.1 Types of pulmonary nodules 2
1.2 Literature search and findings 3
1.3 CADx challenges 4
1.4 Databases 5
1.5 Preprocessing methods 6
1.6 Data augmentation 7
1.7 Motivation 7
1.8 Thesis objectives 9

2 Related works 10
2.1 Nodule segmentation works 10
2.1.1 Two-dimensional Convolutional Neural Networks (2D CNN) 11
2.1.2 Three-dimensional Convolutional Neural Networks (3D CNNs) 11
2.1.3 U-Net 12
2.1.4 Residual Networks 15
2.1.5 Generative Adversarial Networks (GANs) 16
2.1.6 Transformers 17
2.1.7 Other nodule segmentation works 18
2.2 Nodule classification works 19
2.2.1 Single-view 19
2.2.2 Multi-view 21
2.2.3 Three-dimensional classifiers 22
2.2.4 Auto Encoders 24
2.2.5 Multi-task Learning 25
2.2.6 Transformers 25
2.2.7 Other nodule classification works 26

3 Methods 29
3.1 Exploratory Data Analysis 29
3.1.1 Scan class 29
3.1.1.1 Slice spacing 29
3.1.1.2 Pixel spacing 30
3.1.2 Annotation class 31
3.1.3 Three-dimensional nodule samples 35
3.1.3.1 Nodule diameter 36
3.1.3.2 Nodule visualization example using PyLIDC 39
3.2 Hypotheses 40
3.2.1 Nodule semantic segmentation 40
3.2.2 Nodule malignancy classifiers 40

4 Modeling approach 41
4.1 Semantic segmentation architecture – 3D HighResNet 41
4.2 Classification architecture – 3D ResNet18 42

5 Implementation 44
5.1 Data preprocessing 44
5.1.1 Processed scan counts 47
5.1.2 Grouped nodule malignancy labels 47

5.2 Patch extraction and data augmentation 48
5.3 Dataset split 49
5.3.1 Nodule segmentation dataset 49
5.3.2 Nodule classification dataset 50
5.4 Training experiments 51
5.4.1 Nodule segmentation 51
5.4.2 Nodule classification 54
5.5 Evaluation Metrics 55
6 Experiment results 58
6.1 Segmentation performance 58
6.2 Classification performance 60
7 Experiment conclusions 64
7.1 Nodule segmentation 64
7.2 Nodule classification 66
8 CADx Inference 67
9 Limitations and future work 69
10 Discussion 71
References 75

List of Figures

Figure 1.1.1 Pulmonary Nodule Samples 3
Figure 2.1.1 DEHA-Net visual results comparison[28] 13
Figure 2.1.2 Block diagram of the proposed CSE-GAN [42] 17
Figure 2.2.1 The layout of the two CNN networks [69] 23
Figure 2.2.2 Proposed MIXCAPS [86] 26
Figure 3.1.1 LIDC-IDRI scans pixel spacing ranges and counts 31
Figure 3.1.2 Label distribution of nodule characteristics in LIDC-IDRI database 34
Figure 3.1.3 Three-dimensional (3D) nodule sample 35
Figure 3.1.4 Three-dimensional (3D) Ground-glass opacity (GGO) nodule sample . 35
Figure 3.1.5 Three-dimensional (3D) juxtapleural nodule sample 36
Figure 3.1.6 Three-dimensional (3D) solid nodule sample 36
Figure 3.1.7 Nodule diameter distribution 38
Figure 3.1.8 Nodule annotation 38
Figure 3.1.9 LIDC-IDRI sample annotation 39
Figure 4.1.1 3D HighResNet architecture 42
Figure 4.2.1 3D ResNet18 architecture 43
Figure 5.1.1 Preprocessing pipeline 45
Figure 5.1.2 Preprocessed data sample. 46
Figure 5.2.1 Data augmentation example 48
Figure 5.2.2 Patch extraction pipeline 49
Figure 5.4.1: Nodule representation in different views 55
Figure 6.1.1: Segmentation predictions 59
Figure 6.2.1 3D patch malignancy classification predictions 61
Figure 6.2.2: 3-view (masked) malignancy classification predictions 62
Figure 7.1.1 Segmentation underperforming cases 65
Figure 7.2.1 Classification underperforming cases 66
Figure 8.1 CADx pipeline 67
Figure 8.2 Diagnosis pipeline 68

List of Tables

Table 3.1 LIDC scans slice thicknesses range and counts 30
Table 3.2 LIDC scans pixel spacing ranges and counts 30
Table 3.3 LIDC-IDRI Nodule characteristics and annotations 32
Table 3.4 Nodule diameter distribution and counts 37
Table 5.1 Processed scan counts 47
Table 5.2 Grouped nodule malignancy labels 47
Table 5.3 CT scans split 49
Table 5.4 Classification nodule samples count 50
Table 5.5 Segmentation hyperparameters. 53
Table 5.6 Classification hyperparameters 54
Table 6.1 Nodule segmentation performance 58
Table 6.2 3D ResNet18 validation performance 60
Table 6.3 3D ResNet18 (masked input) validation performance 60
Table 6.4 3-view ResNet18 validation performance 63
Table 6.5 3-view ResNet18 (masked input) validation performance 63
Table 6.6 Classification testing performance 63

List of Abbreviations

2D	Two-dimensional
3D	Three-dimensional
AI	Artificial intelligence
AUC	Area under curve
BCE	Binary cross-entropy
CAD	Computer-aided diagnosis
CADe	Computer-aided detection
CADx	Computer-aided diagnosis
CAE	Convolutional Auto-Encoder
CapsNets	Capsule networks
CNNs	Convolutional neural networks
CT	Computed tomography
DL	Deep learning
DSC	Dice Similarity Coefficient
DTCNN	Deep Transfer Learning CNN
FPs/Scan	False positives per scan
GANs	Generative adversarial networks
GGO	Ground glass opacity
HD	Hausdorff distance
HU	Hounsfield units
IoU	Intersection over union
LDCT	Low-dose computed tomography
LR	Learning rate
MV-CNN	Multi-view convolutional neural networks
PPV	Positive predictive value
ResNet	Residual network
ROIs	Regions of interest
SGD	Stochastic gradient descent
SPNs	Solitary pulmonary nodules
VOI	Volume of interest

1 Introduction

Lung cancer is by far the leading cause of cancer death among both men and women, making up almost 25% of all cancer deaths [1]. Each year, more people die of lung cancer than colon, breast, and prostate cancers combined [2]. Lung cancer ranks as the most prevalent cancer among men and is the second most common among women [3]. The reported cancer stats from World Health Organization - WHO classify lung cancer as the deadliest one with 1.8 million deaths and 2.21 million new cases in 2020 [4]. Individuals who currently smoke or have smoked in the past are at an elevated risk of lung cancer due to the presence of more than sixty carcinogenic substances in tobacco smoke [5].

Early detection of lung cancer can increase the survival chance of the patients significantly. If lung cancer is diagnosed at an earlier stage, before it has spread, it is more likely to be successfully treated [5]. Typically, symptoms of lung cancer do not appear until the disease is already at an advanced stage. Even when lung cancer does cause symptoms, many people may mistake them for other problems, such as a virus infection or long-term effects from smoking. This may delay the diagnosis. Thus, delayed diagnosis contributes to the high mortality rates of lung cancer [6].

The risks of the invasive biopsy (Appendix A.1 Lung cancer screening methods) are many and the procedure can be very stressful and difficult for the patients [7]. During an invasive procedure, the patient may experience pain, discomfort, and loss of blood and there is also a risk of getting an infection or pneumonia. Screening tests can help detect cancer at an early stage before symptoms appear. When abnormal tissue or cancer is found early, it may be easier to treat or cure. By the time symptoms manifest, cancer may have advanced and metastasized, thereby rendering it more challenging to treat.

The low-dose computed tomography (LDCT) is defined by the National Cancer Institute at the National Institutes of Health (NIH) as a procedure that uses a computer linked to an X-ray machine that gives off a very low dose of radiation to make a series of detailed pictures of areas inside the body [8]. The pictures are taken from different angles and are used to create three-dimensional (3D) views of tissues and organs. LDCT is recommended as a screening test for adults who have a high risk of developing lung cancer based on their age and smoking history [9].

The screening process and protocols utilized in traditional methods are quite demanding in terms of time and expertise [9]. Reviewing the computed tomography (CT)

scan results of a patient to diagnose the condition requires significant time and experience from professionals. A deep-learning lung cancer computer-aided diagnosis system can help doctors, technicians, and patients by making the screening process much faster and even more accurate than the traditional method [10].

In the medical imaging field, computer-aided detection (CADe) or computer-aided diagnosis (CADx) is a computer-based system that helps doctors make decisions swiftly [11]. Medical imaging involves processing image data that medical practitioners and doctors have to evaluate and analyze for abnormalities quickly. Analysis of imaging in the medical field is a very crucial task because imaging is a basic modality to diagnose any disease at the earliest, but it is also important for the image acquisition process to not harm the human body. CAD is a technology that includes multiple elements like concepts of artificial intelligence (AI), computer vision, and medical image processing. The main application of the CAD system is finding abnormalities in the human body. Among all applications, tumor detection is particularly crucial, as unsuccessful screening can lead to cancer development [5], [12] .

1.1 Types of pulmonary nodules

Pulmonary nodules manifest in various forms, each exhibiting distinctive radiological features that contribute to their classification and clinical interpretation [13]. This thesis embarks on a thorough investigation of these diverse nodule types, aiming to provide a comprehensive understanding of their intricate characteristics.

The well-circumscribed nodule, for instance, presents with clearly defined borders and is often associated with benign pathologies. On the other hand, the juxta vascular nodule poses challenges due to its proximity to blood vessels, making accurate identification and differentiation crucial. Nodules with pleural tails extend linearly from the lung parenchyma to the pleura, while juxta pleural nodules are attached to the pleural surface, posing distinct challenges in assessment. The realm of sub-solid nodules further complicates matters, with lesions exhibiting both solid and ground-glass components, necessitating refined analytical approaches. Among sub-solid nodules, pure ground-glass nodules, characterized by their hazy appearance, present a particularly intriguing subset. **Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε.** depicts samples of various pulmonary nodules.

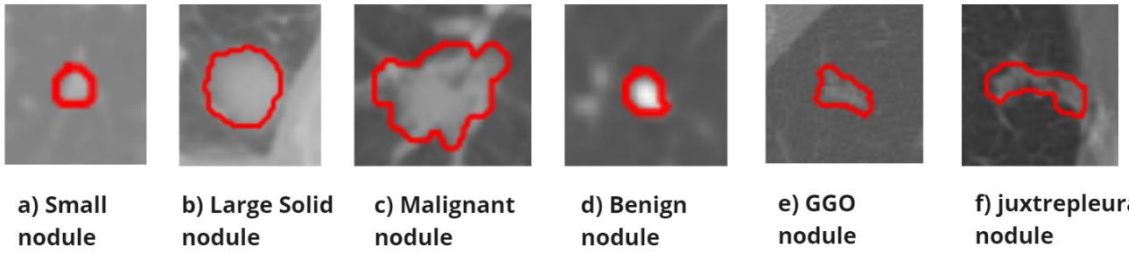


Figure 1.1.1 Pulmonary Nodule Samples

Amidst these variations, the overarching issue of subjectiveness in nodule interpretation emerges. This subjectivity underscores the importance of standardized reporting systems like LUNG-RADS, which provide a structured framework for radiologists to consistently classify and communicate their findings [14]. Moreover, with the growing prominence of lung cancer screening initiatives, an in-depth analysis of these diverse nodular entities becomes increasingly vital.

1.2 Literature search and findings

In the course of the literature review using the Publish and Perish tool [15] to query academic databases, our investigation was centered on the integration of deep learning techniques for lung cancer screening. Within this expansive domain, we identified and focused our research on three pivotal challenges: nodule detection, segmentation, and classification. The literature survey uncovered a diverse array of methodologies, ranging from established paradigms like two-dimensional (2D) convolutional neural networks (CNNs) and three-dimensional (3D) CNNs to more advanced techniques such as autoencoders and transformer models. Beyond these conventional approaches, our exploration extended to the examination of custom and innovative models, highlighting novel strategies that exhibit potential in augmenting the precision and efficacy of lung cancer detection. This thorough examination of the literature serves as a robust framework for our research, informing our endeavors to address these challenges and contribute substantially to the evolving landscape of deep learning applications in lung cancer screening.

We also extracted the performance metrics utilized in these studies, by collecting the quantitative benchmarks employed to assess the efficacy of the proposed models. This holistic analysis of datasets, preprocessing steps, data augmentation methods, model architectures, and performance metrics serves as a comprehensive foundation for our

research. A detailed literature search procedure is described in Appendix (A.2 Literature review procedure).

1.3 CADx challenges

Developing a Computer-Aided Diagnosis (CADx) system for lung nodules using deep learning algorithms presents a complex landscape, encompassing challenges that significantly shape the process. Our literature review underscored three primary hurdles as mentioned above, that underscore the intricacy of this task.

Nodule detection stands as a cornerstone challenge. The system must accurately locate potential nodule candidates within the intricate context of lung CT scans. This entails generating precise regions of interest (ROIs) in both 2D and 3D settings, requiring sophisticated algorithms that can navigate through extensive datasets while minimizing the risk of overlooking critical nodules.

Subsequently, nodule segmentation emerges as a crucial obstacle. Precisely outlining the boundaries of nodules, whether in 2D or 3D, is vital for downstream analysis. Segmentation not only supports accurate measurement but also facilitates comprehensive assessment of internal nodule properties, a prerequisite for reliable classification.

Nodule classification poses yet another significant challenge. Differentiating between benign and malignant nodules, while concurrently reducing false positives, demands advanced classification models proficient in identifying intricate patterns within nodule characteristics. This classification phase serves to inform clinicians, aiding in informed decision-making and prioritizing nodules that warrant closer evaluation.

The significance of CADx systems in the context of lung nodule screening becomes apparent as they augment and enrich the capabilities of radiologists, providing valuable assistance to doctors in the diagnostic process as well. During the screening process, CADx systems streamline tasks vulnerable to human subjectivity and errors. By precisely detecting nodules, accurately delineating their contours, and offering robust classification performance, CADx systems amplify the sensitivity and specificity of screenings. This, in turn, enables radiologists to direct their proficiency towards thorough and accurate evaluations, ultimately leading to early identification and intervention in potential malignancies. The synergy between CADx systems and medical professionals has the potential to reshape lung cancer screening and diagnosis, optimizing healthcare resources, mitigating false readings, and ultimately improving patient health outcomes.

1.4 Databases

The utilization of robust and diverse databases is pivotal in the development, training, and evaluation of methodologies aimed at enhancing the accuracy of pulmonary nodule detection and diagnosis. A review of the literature reveals that CT databases have consistently served as the cornerstone for these endeavors. Notably, two databases have prominently shaped the landscape of research in this domain. The Lung Image Database Consortium image collection (LIDC-IDRI) [16] stands as a monumental international resource, encompassing diagnostic and lung cancer screening thoracic CT scans enriched with meticulously marked-up annotated lesions. This publicly accessible database emerged from a collaborative initiative involving seven academic centers and eight medical imaging companies, yielding a comprehensive repository of 1018 cases. Each case comprises clinical thoracic CT scan images and XML files documenting the outcomes of a meticulous two-phase annotation process conducted by experienced thoracic radiologists. The LIDC-IDRI database's unique approach of independent and unblinded-read phases facilitated the identification of lung nodules without enforced consensus, fostering robustness in nodule classification.

LUNA16 [17], an essential derivative of LIDC-IDRI [16], offers a refined subset focused on enhancing algorithmic training and assessment. Comprising 888 CT scans with a slice thickness equal to or smaller than 2.5 mm, this database facilitates 10-fold cross-validation by splitting data into ten subsets. The LUNA16 challenge tasked participants with locating possible nodules and determining their probability of being a nodule — a binary classification endeavor that mirrored real-world diagnostic challenges. The creation of LUNA16 exemplifies the evolution of database design toward specialized training and evaluation, accommodating the intricacies of pulmonary nodule analysis.

Beyond LIDC-IDRI and LUNA16, other noteworthy databases contribute to the field's progression. The ELCAP Public Lung Image Database [18] furnishes a valuable resource for evaluating computer-aided detection systems. Boasting fifty low-dose whole-lung CT scans with radiologist-identified nodule locations, this database addresses the need for benchmarking performance while mimicking clinical scenarios. The SPIE-AAPM-NCI LungX [19] dataset, comprising seventy CT scans, was instrumental in the LUNGx Challenge [20], which sought to advance quantitative image analysis methods for classifying malignant and benign lung nodules. This collaborative effort underscores the intersection of multidisciplinary expertise in fostering comprehensive solutions.

Essentially, the integration of these varied databases, including LIDC-IDRI, LUNA16, ELCAP, and SPIE-AAPM-NCI LungX unveils a dynamic ecosystem that empowers researchers to refine methodologies, validate innovations, and ultimately push the boundaries of pulmonary nodule detection and classification. Through these databases, the field navigates toward a future where diagnostic accuracy and clinical impact converge to revolutionize lung cancer management.

1.5 Preprocessing methods

In the context of developing CADx systems for lung nodules using deep learning algorithms, a comprehensive set of data preprocessing techniques plays a pivotal role. These preprocessing steps intricately prepare the raw data within a dataset for subsequent training and analysis, greatly influencing the performance of the deep learning models. Among the several preprocessing techniques, some key procedures have emerged as critical factors in enhancing the accuracy and reliability of the CADx systems, as evidenced by the findings of previous studies.

An essential initial step involves the conversion of pixel values to Hounsfield units (HU). The Hounsfield scale, a standardized quantitative scale for radiodensity, allows for the transformation of raw pixel values into a scale that encapsulates the radiodensity of various tissues and materials. This facilitates accurate discrimination of different substances and aids in removing unnecessary elements, such as air, from the scans. Resampling for isotropy addresses the variance in pixel spacing among CT scans, ensuring uniform distances between slices. This enhances the compatibility of the dataset for CNNs and deep learning methods, promoting consistent and accurate analyses.

Lung segmentation, another preprocessing step, generates lung segmentation masks by isolating the lung parenchyma and removing extraneous artifacts, such as organs and bones. The goal is to establish a focused region of interest containing the inner lung area, while also considering the presence of tissue around the lungs for the potential detection of juxta-pleural nodules.

Normalization of HU values within a standardized range constitutes another crucial procedure, with common thresholds employed to eliminate unnecessary radiodensity variations. Zero centering, which involves subtracting the mean pixel value from all images to establish a mean of zero, enhances data consistency and facilitates convergence during

model training. Notably, the mean pixel value is calculated from all images within the dataset to ensure comprehensive normalization.

Patch extraction constitutes a fundamental step for training deep learning models. To this end, samples of nodules are extracted from nodule databases, often categorized based on experienced radiologists' annotations. These annotations guide the combination of benign and malignant classes, subsequently serving as labels for model training. The corresponding regions of interest, or nodule patches, are then cropped and can be extracted as 2D or 3D patches, incorporating volumetric information from axial, coronal, and sagittal views. These multi-view patches capture a comprehensive perspective of the nodule's structural characteristics, enriching the training process.

Collectively, these data preprocessing methods orchestrate a comprehensive transformation of raw lung CT data, rendering them amenable to robust and accurate training of deep learning models. By mitigating noise, standardizing radiodensity, and focusing on relevant features, these preprocessing steps pave the way for more precise and effective CADx systems.

1.6 Data augmentation

Data augmentation is a series of techniques that are commonly used in machine and deep learning applications to generate synthetic training data, effectively to bolster model performance. In numerous studies analyzed within our literature review, a recurrent focus has been on mitigating the dataset imbalance prevalent in datasets like LIDC-IDRI. This imbalance arises when certain classes are underrepresented or overrepresented, prompting researchers to adeptly employ a range of traditional data augmentation methods. These encompass diverse transformations such as rotations, translations, scaling, shear, flips, crops, and the replication of nodule samples. Furthermore, approaches have been devised to balance the representation of positive and negative nodule samples by adjusting their sampling rates. In tandem, modern data generation techniques, including the utilization of Generative Adversarial Networks (GANs), have gained prominence.

1.7 Motivation

The motivation driving the development of this lung cancer CADx system employing 3D deep neural networks to extract volumetric features stems from the pressing need for enhanced accuracy and efficiency in lung nodule diagnosis. Conventional methods often

grapple with the intricate 3D nature of lung CT scans, struggling to capture the full scope of nodule characteristics. By harnessing the power of 3D deep neural networks, this system aspires to revolutionize the analysis process by comprehensively examining volumetric data. This approach is anticipated to uncover nuanced spatial relationships and intricate patterns that may be obscured in traditional 2D analyses. The goal is to enable earlier and more accurate detection of lung nodules, ultimately leading to timely intervention and improved patient outcomes. Through the amalgamation of advanced deep learning techniques and volumetric insight, this CADx system seeks to address the inherent challenges of lung nodule diagnosis, enhancing diagnostic precision in the realm of lung cancer screening.

1.8 Thesis objectives

This thesis is designed with a set of comprehensive objectives aimed at advancing the field of lung cancer diagnosis through the development of an efficient and accurate CADx system. Firstly, it endeavors to conduct a comprehensive literature review focused on the primary challenges identified earlier, exploring studies that have demonstrated noteworthy experimental outcomes. This involves extracting vital information such as databases used, preprocessing strategies employed, data augmentation techniques applied, and proposed methodologies. Secondly, the research aims to identify publicly accessible lung cancer LDCT datasets, thereby expanding the availability of data crucial for robust training and evaluation. In pursuit of standardization, the thesis further seeks to establish a well-defined experimental protocol, outlining the procedures required to effectively train deep CNN architectures tailored for lung cancer screening. Finally, the culmination of this effort involves the deployment of the trained models and the subsequent execution of comprehensive tests to evaluate the CADx system's performance. This multifaceted approach aspires to contribute to the refinement and advancement of lung cancer diagnostics, culminating in a CADx system capable of aiding medical professionals in making accurate and timely assessments, thereby potentially enhancing patient outcomes.

2 Related works

From our comprehensive literature review, we detected 3 main challenges in developing a computer aided diagnostics system for lung cancer screening. These are pulmonary nodule detection, segmentation, and malignancy classification. Previous studies have explored these challenges extensively using many approaches (Appendix A.2-Literature review procedure). In this section, related works for the pulmonary nodule segmentation and malignancy classification tasks are summarized.

2.1 Nodule segmentation works

A nodule segmentation algorithm aims to segment a nodule region. In this segmentation task, a mask is calculated around the nodule and with that mask, the nodule is separated from other tissues. The segmentation of a nodule can be done in 2D or 3D space. CT scans contain rich volumetric information, so a 3D segmented nodule can provide useful information about the shape and the morphology of the nodule in a classification task. The preprocessing and augmentation of data constitute crucial stages that significantly impact the performance of a deep learning model. These steps aim to refine the raw data within a dataset, preparing it for the training process, and simultaneously generate synthetic samples to mitigate the risk of overfitting by altering the original data. Standard preprocessing techniques applied to lung CT databases include resampling for isotropy, conversion of pixel values to HU, lung parenchyma segmentation, normalization, resizing, cropping, and more. Depending on the deep learning model's input, nodule samples or patches are extracted based on radiologist annotations. These patches can be either 2D or 3D, considering the volumetric information in DICOM files. Synthetic data generation involves common methods such as rotations, flips, rescaling, contrast and brightness transformations, among others. The table (Table A.2.1 Nodule Segmentation - Preprocessing steps per work) reports the datasets, preprocessing steps and augmentation methods employed by prior studies in their works while table (Table A.2.2 Nodule Segmentation - Performance per work) reports the reported performance.

2.1.1 Two-dimensional Convolutional Neural Networks (2D CNN)

Singadkar et al. [21] presented a groundbreaking approach in lung nodule segmentation with their novel deep residual deconvolutional network. This end-to-end architecture incorporated multi-level contextual information, automatically learning nodule-sensitive features from 2D CT images to enhance segmentation performance. Despite achieving an impressive Dice Similarity Coefficient (DSC) score of 94.97% on the LIDC-IDRI dataset, it's crucial to note the lack of cross-validation and detailed preprocessing steps in their study.

Wang's et al. [22] Central Focused Convolutional Neural Networks (CF-CNN) stands out for its ability to effectively segment lung nodules from heterogeneous CT images. By capturing nodule-sensitive features from both 3D and 2D CT images simultaneously, Wang's model achieved average DSC scores of 82.15%±10.76 in the LIDC-IDRI dataset and 80.02%±11.09 in an independent private dataset. Notably, their approach excelled in juxta-pleural nodule segmentation, a challenging task, demonstrating a promising average DSC score difference of only 1.98% when compared with inter-radiologists' consistency on the LIDC-IDRI dataset.

Roy et al. [23] introduced a synergistic combination of deep learning and shape-driven level sets for accurate lung nodule segmentation. Their approach involved a deep fully convolutional network for coarse segmentation and shape-driven level sets for fine segmentation, achieving an impressive average DSC score of 93%±0.11 on isolated nodules in the LIDC-IDRI dataset and 90%±0.08 on pleura adhesion.

Khan et al. [24] introduced a deep learning framework using the VGG19 network for the joint segmentation and classification of lung nodules in CT images. The framework consists of two main components: VGG-SegNet for nodule segmentation and a pre-trained VGG19 for classification. For the segmentation task demonstrated impressive results were achieved a Dice similarity coefficient (DSC) of 0.91, which indicates a high degree of accuracy in the overlap between the predicted segmentation and the ground truth. The metrics of this work were evaluated using the LIDC-IDRI and Lung-PET-CT-Dx dataset.

2.1.2 Three-dimensional Convolutional Neural Networks (3D CNNs)

A Nested Three-Dimensional Fully Connected Convolutional Network for lung nodule segmentation has been introduced in [25]. Leveraging a single encoder, this model achieved promising results with a DSC of 0.845±0.007 and Intersection over Union (IoU) of

0.738±0.011 on a 128×128×64 input size. The housing unit structure and encoder-decoder connections through concatenation demonstrated the network's efficacy.

Tang et al. [26] proposed a unified model for pulmonary nodule detection, false positive reduction, and segmentation, exhibited substantial improvements in nodule detection accuracy by 10.27% compared to baseline models. The model achieved a state-of-the-art DSC of 83.10%, emphasizing its robust performance and contributions to enhancing pulmonary nodule detection and segmentation tasks. The shared underlying feature extraction backbone and end-to-end training showcased the effectiveness of their comprehensive approach.

Another study [27], employed a 2D-3D cascaded CNN for comprehensive lung nodule analysis, integrating segmentation, detection, and classification. The SquExUNet segmentation framework, combined with a 2D-3D cascaded CNN approach for detection, demonstrated sensitivity of 90% and a DSC of 80%. The study's holistic strategy showcased the potential for combining multiple tasks within a unified framework.

2.1.3 U-Net

Usman M. [28] proposed a semi-automated 3D segmentation method for lung nodules, employing a two-stage process. In the first stage, a 2D ROI containing the nodule was used for patch-wise exploration along the axial axis using an adaptive ROI algorithm. In the second stage, the volume of interest (VOI) was further explored along the coronal and sagittal axes using Residual U-Nets. The method achieved promising results with DSC scores of 85.29% ± 9.78 in axial, 84.76% ± 12.45 in coronal, and 83.58% ± 8.93 in sagittal views, averaging at 87.5% ± 10.58.

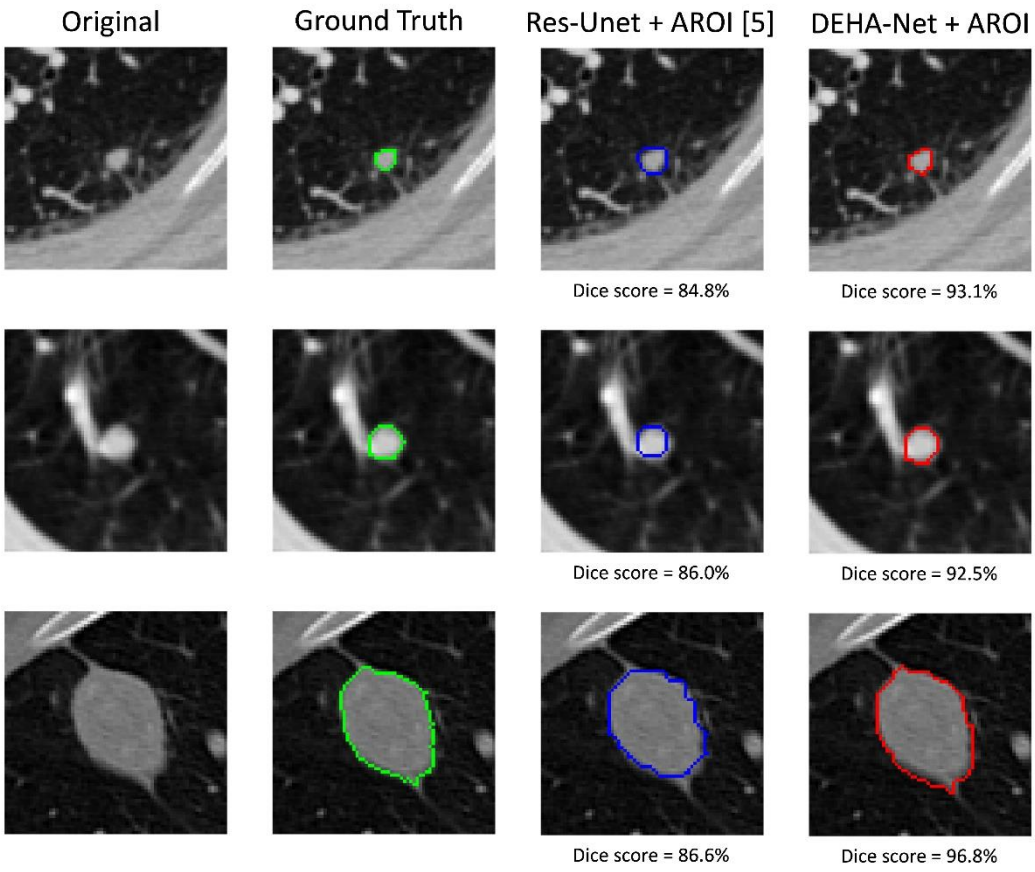


Figure 2.1.1 DEHA-Net visual results comparison[28]

Pezzano G. [29] took a unique approach by using a U-net based network to learn the context of nodules through two masks representing background and secondary-important elements in CT scans. The subtraction of these masks successfully extracted the nodule area, yielding an impressive IoU score of 76.6 ± 12.3, comparable to human performance. The masks generated by Pezzano's network closely resembled those produced by radiologists, as validated by the Kolmogorov-Smirnov test [30].

A novel approach introduced in a recent study [31] leveraging synthetic CT images with distinctive color patterns to represent evolving nodule features for detection and segmentation. Using a modified U-Net architecture, the method achieved a DSC of 93.14%, Recall of 91.76%, Precision of 93.3%, True Positive Rate of 2.3%, and a False Positive Rate of 0.21%. This innovative approach enhances the accuracy and efficiency of lung nodule segmentation by effectively incorporating inter-slice information through synthetic images.

Two studies, RAD-Unet [32] and SMR-Unet [33], focused on enhancing traditional U-Net models. RAD-UNet incorporated a residual network module and a pyramid pooling module, achieving a mean IoU of 87.76% and 88.13%, demonstrating improved lung nodule semantic segmentation performance. SMR-UNet introduced self-attention, multi-scale

features, and residual structures, achieving a DSC index of 0.9187 and an IoU of 0.8688, surpassing the traditional U-Net model by 4.22% and 4%, respectively.

Another study [34] introduced a novel deep learning architecture designed for lung cancer nodule detection with a focus on reducing false positives. It incorporates receptive regularization in both the convolution and deconvolution layers of the V-Net model. Nodule classification is carried out using a combination of SqueezeNet and Residual Network (ResNet), termed the nodule classification network (NCNet). Post-processing involves image enhancement on 2D slices through increased intensity using pseudo-color or fluorescence contrast. The RFR V-Net demonstrates impressive segmentation performance, achieving a DSC of 95.01% and an IoU of 0.83. The NCNet achieves a sensitivity of 98.38% and a low rate of false positives per scan (FPs/Scan) at 2.3 for 3D representations. The proposed approach, combining RFR V-Net and NCNet, exhibits substantial improvements over existing CAD systems for lung nodule detection.

A Coarse-to-Fine Lung Nodule Segmentation approach [35], utilizing image enhancement and a Dual-Branch U-Net (DB U-Net), demonstrated effective segmentation. The DB U-Net achieved DSC of 83.16% and 81.97% on the LIDC dataset and an additional dataset, showcasing its accuracy in lung nodule segmentation.

An accurate segmentation approach [36] introduced an improved U-Net convolutional network incorporating Batch Normalization (BN) for enhanced performance. Operating at a resolution of 32×32, the improved U-Net achieved a DSC of 0.8623, demonstrating its accuracy in segmenting various types of lung nodules.

Evo-GUNet3++ [37] a novel approach using evolutionary algorithms to optimize UNet-based architectures for efficient 3D lung cancer detection. Achieving a DSC of 0.972, sensitivity of 0.977, and positive predictive value (PPV) of 0.923, Evo-GUNet3++ demonstrated effectiveness in accurate semantic segmentation for lung cancer detection, outperforming baseline performances and showcasing the potential of evolutionary algorithms in UNet-based architecture optimization.

Additionally, AWEU-Net [38] was proposed for lung nodule detection and segmentation in CT images, utilizing a two-stage process. In the first stage, a fine-tuned Faster R-CNN model localizes the nodules. In the second stage, the segmented regions are refined using a U-Net architecture enhanced with position attention-aware weight excitation (PAWE) and channel attention-aware weight excitation (CAWE) blocks. The method achieved impressive results with Dice scores of 89.79% and 90.35%, and Intersection over

Union (IoU) metrics of 82.34% and 83.21% on the LUNA16 and LIDC-IDRI datasets, respectively.

2.1.4 Residual Networks

Liu et al. [39] proposed an innovative approach utilizing a residual block-based dual-path network to extract both local features and rich contextual information from lung nodules. The method takes advantage of multi-view and multi-scale features from CT images, showcasing effectiveness in segmenting small and juxtapleural nodules. The implemented CDP-ResNet achieved an impressive average DSC of 81.58%±11.05 on the LIDC dataset. Notably, the model's segmentation accuracy was reported to be slightly superior to that of human experts, emphasizing its potential for accurate nodule segmentation.

Building upon Liu's [39] work, Cao. et al. [40] proposed a model designed to simultaneously capture multi-view and multiscale features of different nodules in CT images. The model combines intensity and CNN features using a unique pooling method called the central intensity pooling layer (CIP). This layer extracts intensity features from the center voxel of the block, while a CNN is employed to obtain convolutional features from the same center voxel. To address overfitting, the authors introduced a weighted sampling strategy for training sample selection. The method achieved a remarkable DSC of 82.74%±10.19 on the LIDC-IDRI dataset, highlighting its potential for robust lung nodule segmentation.

Additionally, Yu et al. [41] proposed a lung nodule segmentation and recognition algorithm based on deep learning using 3D CT images. The method involves two main components: a 3D Res U-Net for segmentation and a 3D ResNet50 for classification. The 3D Res U-Net uses residual units in place of common convolutional layers and combines Dice loss and cross-entropy loss to improve accuracy and convergence. The 3D ResNet50 replaces 2D with 3D convolutional layers for better spatial feature extraction. The 3D Res U-Net achieved a Dice coefficient above 0.8 for nodules larger than 10 mm, while 3D ResNet50 achieved a classification accuracy of 87.3% and an AUC of 0.907 on the LIDC-IDRI dataset.

These studies collectively demonstrate the effectiveness of residual networks in lung nodule segmentation, showcasing advancements in leveraging multi-view and multiscale features for improved accuracy and performance.

2.1.5 Generative Adversarial Networks (GANs)

CSE-GAN, a 3D conditional generative adversarial network tailored specifically for lung nodule segmentation, has been introduced [42]. In this novel approach, the generator functions as a segmentation network, responsible for producing accurate segmentation mask images. Simultaneously, the discriminator operates as a classification network, distinguishing between authentic and generated segmented nodule masks. This adversarial setup encourages the generator to improve its segmentation capabilities continually.

A key innovation in CSE-GAN lies in the integration of a concurrent spatial and channel squeeze and excitation module within both the generator and discriminator. This module enhances segmentation performance by enabling the network to focus on critical spatial and channel-wise features during the segmentation process. This dual application of the squeeze and excitation module contributes to improving classification accuracy as well.

To validate the proposed model, an Indian lung nodule dataset collected from a local hospital is introduced. This dataset serves for both validation and generalizability testing, providing a diverse set of cases for the model to learn from. CSE-GAN demonstrates noteworthy results on two distinct datasets – the LUNA test set and the local dataset. Specifically, the model achieves DSC of 80.74% and 76.36% on these datasets, indicating a high level of accuracy in lung nodule segmentation. Additionally, sensitivities of 85.46% and 82.56% further underscore the effectiveness of CSE-GAN in accurately identifying and segmenting lung nodules.

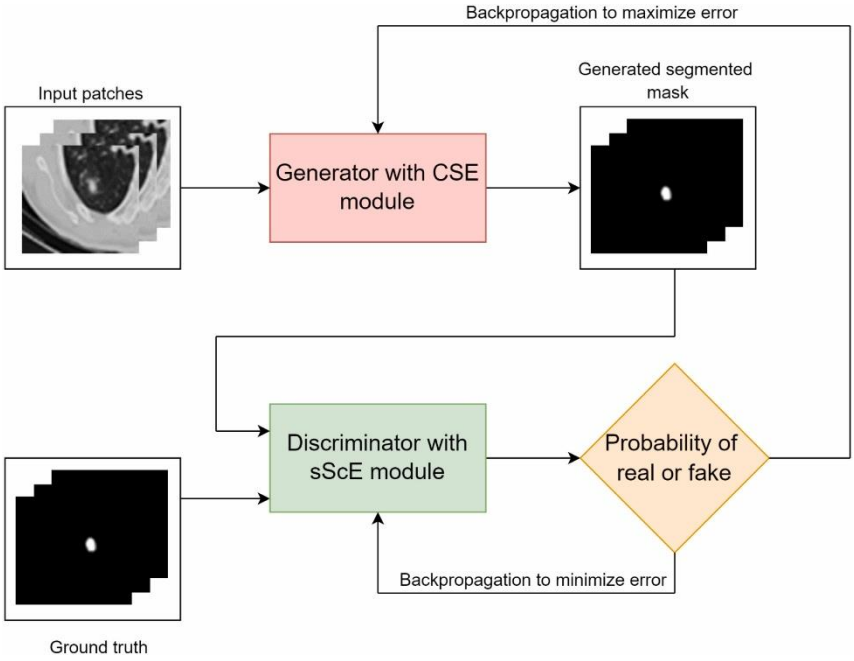


Figure 2.1.2 Block diagram of the proposed CSE-GAN [42]

2.1.6 Transformers

TPFR-Net is a U-shaped model designed for lung nodule segmentation, specifically addressing challenges related to global dependencies, computation efficiency, and feature loss during up-sampling [43]. TPFR-Net employs a unique encoder that integrates convolution and transformer modules for both local and global feature extraction. This allows for multiscale and channel attention convolutions to be incorporated effectively. To manage the computational demands of the transformer's self-attention layer, the authors ingeniously fuse it with the pooling layer. Additionally, a novel feature reorganization strategy, utilizing dual-attention mechanisms, is implemented to maximize feature retention during up-sampling. The loss function is enhanced by combining Binary Cross-Entropy (BCE), Dice, and Hausdorff Distance (HD Loss) to give due emphasis to segmentation boundaries. The deep supervision strategy further ensures the authenticity and reliability of image information post feature reorganization. Evaluated on datasets of varying dimensions (64×64, 96×96, and 128×128), TPFR-Net showcases remarkable results with a DSC of 91.84% and a sensitivity of 92.66%, highlighting its effectiveness in lung nodule segmentation. This model proves to be a promising advancement in addressing the challenges associated with global dependencies and feature loss during up-sampling in lung nodule segmentation tasks.

DPBET is introduced as a unique Cascade-Axial-Prune transformer (CAP-Trans) model for lung nodule segmentation [44]. DPBET utilizes a hybrid CNN-Transformer architecture to maximize the benefits of both convolutional and transformer blocks, allowing for effective capture of local details and global semantic representations. To enhance boundary information, the authors employ edge detection operators, constructing boundary enhancement datasets for the edge path to provide additional edge prior knowledge. The encoding method, termed "Down-Attention Sample (DASample)," is introduced in the edge path, leveraging channel and spatial attention to increase the perceptual field for multi-scale lung nodule information. Evaluated on a 64×64 dataset, the CAP-Trans model achieves impressive segmentation results with a DSC of 89.86% and an average sensitivity of 90.50%. This highlights the model's effectiveness in accurately delineating lung nodules and showcases its potential for advancing lung nodule segmentation tasks by incorporating a hybrid CNN-Transformer architecture and innovative boundary enhancement strategies.

The DEHA-Net [45] introduces a novel lung nodule segmentation framework incorporating a dual-encoder-based hard attention network and an adaptive ROI mechanism. The framework operates in two stages: firstly, a 2D ROI is generated along the axial axis using DEHA-Net, leveraging initial input from a radiologist or CADe system. An adaptive ROI algorithm then extends the generation of ROIs to surrounding slices, facilitating 3D mask reconstruction. In the second stage, ROIs are generated along sagittal and coronal views using the 3D mask obtained in the first stage, with DEHA-Net applied for segmentation along these views. The final 3D segmentation mask is produced by a consensus module. Notably, this pipeline avoids resizing, mitigating issues associated with input and output rescaling. The dual-encoder-based CNN achieves promising results with a DSC of 87.91%, sensitivity of 90.84%, and positive predictive value of 89.56%, demonstrating the effectiveness of DEHA-Net for accurate lung nodule segmentation.

2.1.7 Other nodule segmentation works

A novel Dual-Task Region-Boundary Aware Neural Network [46] for pulmonary nodule segmentation incorporates a hierarchical feature module for capturing multi-scale information, a boundary-guided module to model boundaries explicitly, and a feature aggregation module for fusing boundary and multi-scale features. The proposed region-boundary aware loss function enhances the relationship between regions and boundaries, resulting in improved segmentation performance. The method achieves superior segmentation results on the LIDC-IDRI and LUNA16 datasets, particularly excelling in small

and non-solid nodules. Utilizing a 3D U-Net architecture with a 64 × 64 × 32 input size, the proposed method demonstrates promising performance metrics, including a DSC of 82.48%, IoU of 70.86%, Sensitivity of 82.74%, Precision of 84.10%, and Average Surface Distance of 0.310 mm, highlighting its suitability for accurate pulmonary nodule segmentation.

DAS-Net [47] is a lung nodule segmentation method that leverages an adaptive dual-attention module and a novel 3D shadow mapping layer. The adaptive dual-attention module enhances the model's ability to perceive detailed information on the 3D nodule surface, improving segmentation accuracy. The 3D shadow mapping layer is introduced to construct the basic structure of the network, ensuring feature-rich extraction with a reduced number of parameters and computational efficiency. DAS-Net, with a dimension of 16×128×128, outperforms state-of-the-art methods in lung nodule segmentation tasks. The evaluation metrics demonstrate its superiority, with a DSC of 92.05% ± 3.08, sensitivity of 90.81% ± 6.35, and HD of 3.93 ± 1.87 mm, showcasing the effectiveness of DAS-Net in accurate 3D lung nodule segmentation with a compact parameter configuration.

2.2 Nodule classification works

A nodule malignancy classification algorithm aims to determine whether a detected nodule is benign or malignant. The classification of a nodule can be performed using 2D image slices or 3D volumetric data, allowing for a comprehensive assessment of the nodule's characteristics across multiple dimensions. Classification works follow very similar preprocessing steps to the segmentation works outlined in the previous section. Similarly, in Table A.2.3 Nodule Classification - Preprocessing steps per work, we demonstrate the datasets, preprocessing steps and data augmentation methods employed by prior studies while Table A.2.4 Nodule Classification - Performance per work reports the performance metrics.

2.2.1 Single-view

Da Silva et al. [48] employed a CNN alongside Particle Swarm Optimization (PSO) to optimize hyperparameters, achieving a remarkable accuracy of 97.62%, sensitivity of 92.20%, specificity of 98.21%, and an area under curve (AUC) of 0.955 on the LIDC-IDRI database. Bhandary et al. [49] explored a Modified AlexNet (MAN) architecture with

Ensemble-Feature-Technique (EFT), showcasing the model's superiority with a classification accuracy of 97.27%, sensitivity of 98.09%, specificity of 95.63%, and an AUC of 0.995.

Tran et al. [50] introduced a novel 15-layer 2D CNN, incorporating focal loss, resulting in a highly accurate model with 97.2% accuracy, 96.0% sensitivity, 97.3% specificity, and an AUC of 0.982 on the LUNA16 challenge database. Moreover, Al-Shabi et al. [51] proposed Gated-Dilated (GD) networks, achieving an accuracy of 92.57%, sensitivity of 92.67%, and AUC of 0.9514, outperforming traditional CNNs. Suresh et al. [52] developed a CNN with Generative Adversarial Networks (GAN), achieving a classification accuracy of 93.9%, specificity of 93%, sensitivity of 93.4%, and an AUC of 0.934, demonstrating the impact of GAN-generated synthetic samples.

Huang's et al. [53] novel approach combined Deep Transfer Learning CNN (DTCNN) and Extreme Learning Machine (ELM), achieving 94.57% accuracy, 93.69% sensitivity, and 95.15% specificity on LIDC-IDRI. Zhao et al. [54] explored modified CNN strategies with transfer learning, attaining an AUC of 0.94, sensitivity of 91%, and an overall accuracy of 88%. In addition, Ali et al. [55] introduced transferable texture CNN networks, achieving high performance on the LIDC-IDRI dataset and demonstrating the efficacy of transfer learning on smaller datasets. Naik et al. [56] combined Fractalnet and CNN, yielding a model with high sensitivity (97.52%) and AUC (0.98), addressing the overfitting problem with drop-path in the Fractalnet architecture.

Computational intelligence techniques were explored in another study [57], achieving improvements in accuracy, precision, sensitivity, and specificity using a U-net-based network. The ensemble framework, F-LSTM-CNN, proposed for benign-malignant classification, integrated nodule attributes and images, achieving an accuracy of 0.955, sensitivity of 1, specificity of 0.937, and an AUC of 0.995.

In another work [58] Suresh et al. proposed a NROI-based feature learning using DCNN which demonstrated outstanding performance with a classification accuracy of 97.8%, specificity of 97.2%, sensitivity of 97.1%, and an AUC of 0.9956. Residual-Transformer networks [59] are proposed in literature, showcasing improved accuracy (AUC of 0.9628) and overall effectiveness in classifying lung nodules. LDNNET [60], proposed for robust classification, exhibited excellent performance without extensive preprocessing, emphasizing its versatility and robustness. The reported metrics for a voxel resolution of 80×80 include sensitivity of 0.982072, accuracy of 0.988396, and specificity of 0.994584. A

Bilinear Convolutional Neural Network (BCNN) [61] introduced for lung nodule classification achieved an accuracy rate of 91.99% and an AUC rate of 95.9%, highlighting the efficacy of the proposed architecture.

Additionally, transfer learning techniques for the automatic characterization of Solitary Pulmonary Nodules (SPNs) [62] demonstrated peak accuracy of 94%, addressing the challenges of small datasets and showcasing the benefits of transfer learning and data augmentation.

These advancements collectively underscore the evolving landscape of deep learning in lung nodule classification, emphasizing improved model architectures, optimization strategies, and computational intelligence integration for enhanced clinical accuracy and reliability.

2.2.2 Multi-view

Nibali et al. [63] adopted a three-column ResNet architecture to assess the impact of curriculum learning, transfer learning, and varying network depth on malignancy classification. By using three 2D planar views instead of a full 3D volume, their model achieved a sensitivity of 91.07%, specificity of 88.64%, AUC of 0.9459, and an accuracy of 89.90%, showcasing computational efficiency compared to other multi-view approaches.

On a different front, Xie et al. [64] introduced a multi-view knowledge-based collaborative (MV-KBC) deep neural network model, utilizing nine views for nodule classification. Despite competitive performance metrics with an accuracy of 91.60%, specificity of 94%, and AUC of 95.70%, the model's drawback lay in its large size (nine sub-models) and significant computational requirements.

Kang et al. [65] proposed a 3D multi-view convolutional neural networks (MV-CNN) for lung nodule classification, exploring binary and ternary classifications. Achieving a lower error rate and higher performance metrics (Sensitivity: 95.60%, Specificity: 93.94%, AUC: 0.99) than similar works, this study highlighted the efficacy of 3D MV-CNN in diverse classification tasks.

Meanwhile, in another work Xie et al. [66] proposed the Transferable Multi-Model Ensemble (TMME) algorithm, leveraging pre-trained ResNet-50 models for nodule characterization. In experiments on the LIDC-IDRI database, TMME achieved a classification accuracy of 93.4%, sensitivity of 91.43%, and specificity of 94.09%, underscoring the effectiveness of transfer learning for accurate nodule classification.

A retrospective study [67] developed a deep learning (DL) algorithm for lung nodule classification, showcasing superior performance with an AUC of 0.93 and clinical relevance compared to existing models. Additionally, MV-CRecNet, a novel model architecture for pulmonary nodule identification, demonstrated superior performance in accuracy, sensitivity, specificity, and AUC compared to alternative model architectures. These studies collectively advance the landscape of lung nodule classification, introducing innovative approaches and demonstrating promising outcomes in accuracy and clinical applicability.

Another work [68] introduced a novel model architecture, MV-CrecNet which leverages Multi-view Convolutional Recurrent Neural Networks (MV-CNN). The comparative performance of MV-CRecNet is assessed alongside various other model architectures, employing 2D MV-CNN and 3D MV-CNN with different input sizes (20 x 20), (30 x 30), and (40 x 40). The reported metrics for a voxel resolution of (20 x 20) include an accuracy of 0.97, sensitivity of 0.98, specificity of 0.97, and an AUC of 0.99. These results underscore the effectiveness of the proposed MV-CRecNet in pulmonary nodule identification, showcasing superior performance in terms of accuracy, sensitivity, specificity, and AUC compared to alternative model architectures.

2.2.3 Three-dimensional classifiers

Several 3D classifiers for lung nodule classification have been developed over the past few years, Causey et al. [69] proposed a predictive model employing a deep learning CNN. Despite achieving high sensitivity of 94.80%, specificity of 94.30%, accuracy of 94.60%, and an AUC of 0.984, the model's major limitation lies in its reliance on manual specification of the nodule region of interest, making it less automated. Dey et al. [70] evaluated multiple 3D CNNs, introducing a 3D multi-output DenseNet (MoDenseNet) with transfer learning. The model achieved an accuracy of 90.40%, specificity of 90.33%, sensitivity of 90.47%, and AUC of 0.948, showcasing robust performance, especially with transfer learning on a challenging private dataset.

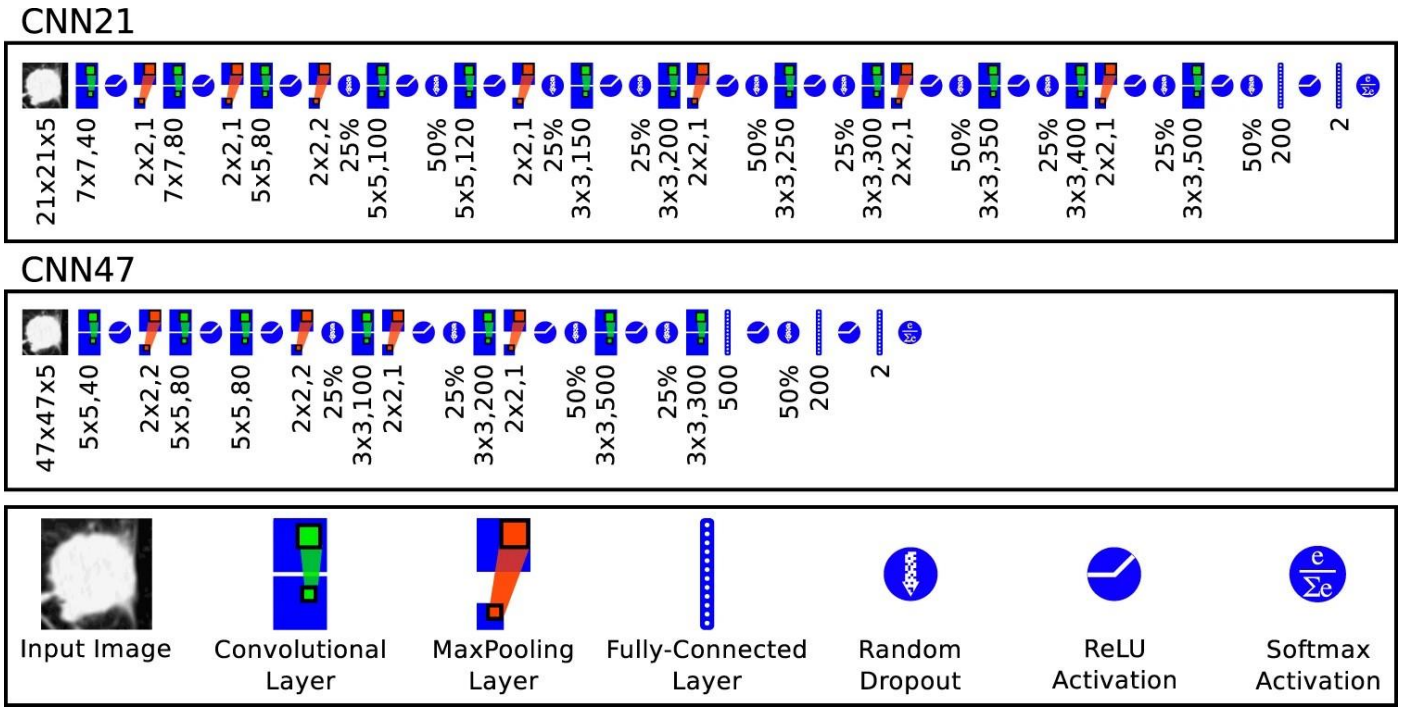


Figure 2.2.1 The layout of the two CNN networks [69]

Xu et al. [71] developed lightweight 3D CNN models, demonstrating that shallower networks outperformed deeper ones. Despite achieving a high accuracy of 92.65% and specificity of 95.87%, their sensitivity of 85.58% was relatively lower. Ren et al. [72] proposed a novel manifold regularized classification deep neural network (MRC-DNN), achieving a classification accuracy of 0.90 on the LIDC-IDRI validation set, with sensitivity and specificity of 0.81 and 0.95, respectively.

In addition, several other innovative approaches have been introduced. The Self-supervised Transfer Learning Framework driven by Visual Attention (STLF-VA) [73] demonstrated competitive performance with an accuracy of 92.36%, sensitivity of 91.62%, specificity of 93.08%, AUC of 97.17%, and precision of 92.99%. The ProCAN network [74] incorporating non-local network enhancement and curriculum learning, outperformed state-of-the-art methods with an AUC of 98.05% and an accuracy of 95.28%.

The Multi-View Coupled Self-Attention Network (MVCS) [75] addressed depth dimension relations and achieved notable performance metrics on a 64 × 64 × 32 voxel resolution, including an accuracy of 91.25%, sensitivity of 89.10%, specificity of 93.39%, precision of 91.59%, AUC of 91.25%, and F1-score of 90.19%. This study [76] focused on early detection of lung cancer utilized a 3D Convolutional Neural Network (CNN) with a Multiview-one-network strategy, achieving an AUC of 0.97, accuracy of 97.17%, F-score of 0.92, precision of 0.87, and recall of 0.94. These advancements collectively contribute to

the evolving landscape of 3D lung nodule classification, showcasing innovative approaches and promising outcomes in accuracy and clinical applicability.

Another study [77] proposed a lung nodule classification algorithm that integrates CT imagery with biomarker annotation and volumetric radiomic features. The algorithm employs a 3D CNN and a Random Forest. The performance is analyzed and compared across different combinations: using only imagery, only biomarkers, combined imagery and biomarkers, combined imagery and volumetric radiomic features, and the combination of imagery, biomarkers and volumetric features. The reported results for a voxel resolution of 32 pixels × 32 pixels × 16 slices include a Receiver Operating Characteristic - Area Under the Curve (ROC -AUC) of 0.8674. This comprehensive approach leverages multiple data modalities to enhance the classification of lung nodule malignancy suspicion levels. Jung H. [78], introduced a 3D DCNN with shortcut and dense layer connections achieved high sensitivity of 95.4% and a notably low false-positive rate.

2.2.4 Auto Encoders

Silva et al. [79] proposed a transfer learning approach utilizing a Convolutional Auto-Encoder (CAE) as a feature extractor, followed by a Multi-layer Perceptron classifier. Despite achieving a sensitivity of 78.9% and an AUC of 0.928 when trained from scratch, the model's performance improved to a sensitivity of 84.8% and an AUC of 0.936 on the LIDC-IDRI dataset. While these metrics are lower compared to other studies, the transfer learning approach showcases the potential of leveraging pre-trained autoencoder features for malignancy binary classification.

Another study [80] introduced an unsupervised feature extraction method employing multiple convolutional autoencoders for various 2.5-dimensional medical images, achieving high performance in detecting cerebral aneurysms and lung nodules with AUCs exceeding 0.96. This highlights the versatility and effectiveness of autoencoders in unsupervised feature extraction across different medical imaging applications.

In the CKAK pipeline [81], a method aimed at reliable lung nodule diagnosis, the authors effectively fused clinical and AI knowledge at both feature and decision-making levels. The proposed scale-aware feature extraction block (SAFE) integrates multi-scale contextual features using a lightweight Transformer, demonstrating superior accuracy in benign-malignant classification (92.82%), AUC (97.40%), sensitivity (87.69%), and specificity (95.38%) at a voxel resolution of 80 × 80 × 60. The CKAK pipeline showcases

the potential of combining clinical expertise with AI knowledge for robust lung nodule diagnosis, meeting clinical requirements for a reliable CAD system.

2.2.5 Multi-task Learning

Zhai. et al. [82] extracted nine 2D views from a candidate nodule cube from different projection angles. Then, for each view, they constructed a 2D Multi Task-CNN model, which consisted of a nodule classification branch and an image reconstruction branch. Finally, they obtained the classification results by the weighted fusion of the prediction results of the nine 2D MT-CNN models. The implementation of transfer learning improved their metrics slightly with an AUC of 0.9559 in the LIDC-IDRI dataset, and 0.973 in the LUNA16 challenge dataset, but the reported sensitivity performance of 87.74% in LIDC-DRI, and 84.00% in the LUNA16 challenge dataset is lower than other classification approaches. Authors state that there is no information communication between different views before fusion, thus rich contextual information and features are probably lost.

Zia et al. [83] addressed intra-class variation and inter-class similarity using a multi-deep model (MD model) for lung nodule classification. The model includes multi-scale dilated convolutional blocks, dual deep convolutional neural networks, and multi-task learning components. With reported performance metrics of almost 91% in sensitivity, specificity, and accuracy, this approach aims to overcome challenges related to diverse imaging modalities.

Another study [84] introduced a deep learning-based multi-task learning (MTL) model with attention modules for analyzing lung nodule attributes in CT images. This model, processing entire image volumes and simultaneously scoring multiple nodule attributes, achieved notable performance on the LIDC-IDRI dataset, with an accuracy of 94.7%, sensitivity of 96.2%, specificity of 82.9%, precision of 97.8%, and AUC of 95.9. The incorporation of attention modules enhances interpretability and clinical relevance.

2.2.6 Transformers

In the domain of Transformers, the Multi-Granularity Dilated Transformer (MGDFormer) [85] was introduced for lung nodule classification. This model aimed to learn pixel-wise global attention for robust long-range global representation and employs a Local Focus Scheme (LFS) to enhance focus on local discriminative features. With competitive performance metrics, including an AUC of 98.5%, accuracy of 96.1%, precision of 95.9%, sensitivity of 94.4%, and F1-Score of 95.2%, MGDFormer demonstrated effectiveness in addressing

challenges related to local and global information, making it a promising model for lung nodule classification.

2.2.7 Other nodule classification works

Capsule networks (CapsNets) were employed for the first time as individual experts within a mixture of experts (MoE) framework in the proposed MIXCAPS [86] model for lung nodule malignancy prediction. The study revealed how CapsNet itself can be considered an MoE framework, making MIXCAPS a hierarchical MoE technique. The output of the gating model was explored for potential correlations with hand-crafted nodule features, enhancing the interpretability of MIXCAPS. The analysis demonstrated that individual CapsNet experts specialize in different subsets of the dataset, and the gating model determines their contributions, showcasing how experts' activations change with different data subsets. MIXCAPS achieved robust generalizability, as illustrated through extension and evaluation on a separate dataset associated with a different prediction task. The reported metrics for an input size of 80×80×3 slices include sensitivity of 89.5%, specificity of 93.4%, accuracy of 90.7%, and an AUC of 0.956. These findings highlight the effectiveness of the proposed MIXCAPS model in lung nodule malignancy prediction and its potential for interpretability and generalization.

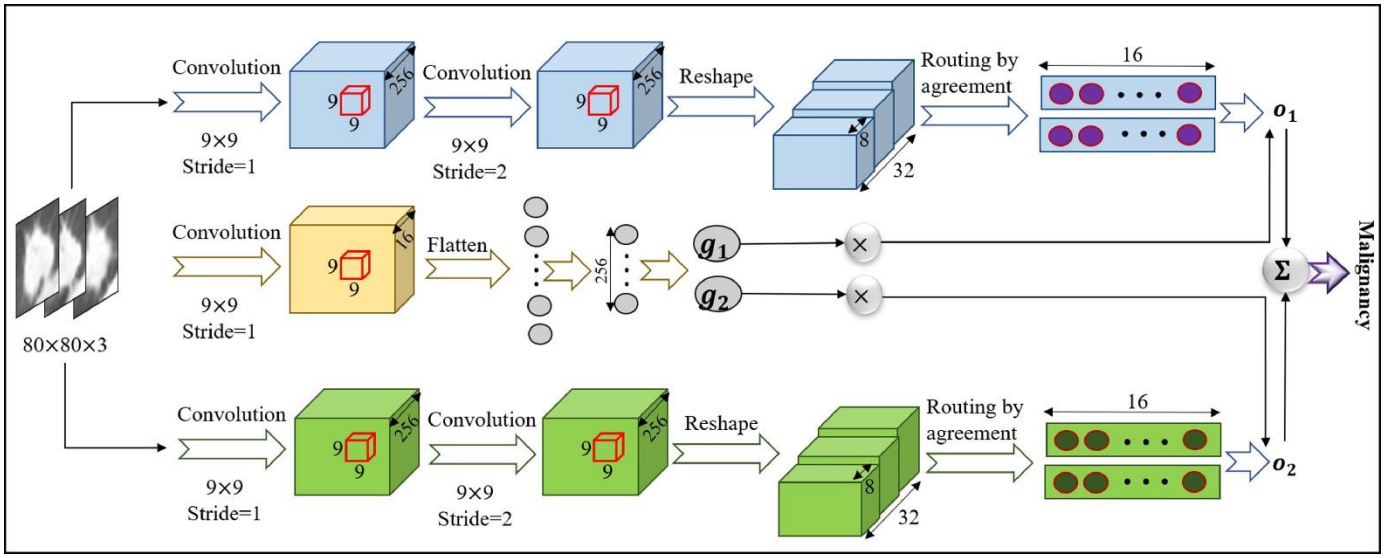


Figure 2.2.2 Proposed MIXCAPS [86]

DC-GAN [87] proved effective in generating realistic SPNs, leading to a notable increase in FF-VGG19's classification accuracy on the LIDC-IDRI dataset by +7% to 92.07% and on the CT dataset by 5% to 84.3%. The reported metrics, with a voxel resolution of

32×32, include an accuracy of 92.1%, sensitivity of 89.3%, specificity of 94.8%, and an AUC of 92.1%. These results underscore the efficacy of the proposed methodology, combining DC-GAN for realistic nodule generation and FF-VGG19 for improved classification accuracy in lung nodule malignancy assessment.

A groundbreaking study [88] introduced a novel approach utilizing neural architecture search (NAS) for automatic exploration of 3D network architectures. This method achieves an exceptional accuracy/speed trade-off, integrating the convolutional block attention module (CBAM) into networks to enhance the reasoning process. The use of A-Softmax loss during training fosters the learning of angularly discriminative representations. Notably, this study marks the first attempt to use NAS for pulmonary nodule classification. The model's reasoning process aligns with physicians' diagnosis, contributing to explainability, and the proposed ensemble strategy achieves high comparability with previous state-of-the-art methods while utilizing significantly fewer parameters. With reported metrics for a voxel resolution of 32×32×32, including an accuracy of 90.77%, sensitivity of 85.37%, specificity of 95.04%, and an F1 Score of 89.29, these results underscore the efficacy of the proposed 3D NAS method, CBAM module, A-Softmax loss, and ensemble strategy for achieving efficient, explainable, and discriminative representations in pulmonary nodule classification.

Another innovative study [89] adopted a multi-step approach for pulmonary nodule classification. The MIXUP method was employed to construct virtual training data for enhanced data diversity, followed by the use of the 3D dual-path network (3D DPN) to extract nodule features. The Gradient Boosting Machine (GBM) algorithm was then applied to differentiate pulmonary nodules using both deep features and raw nodule pixels. Spatial and contextual features are captured using the RAN and SE modules, with a novel multi-scale attention module proposed to capture multi-scale attentive features. This comprehensive approach, incorporating various techniques such as data augmentation, advanced network architectures, and attention mechanisms, yields robust pulmonary nodule classification with reported results, including an accuracy of 91.9%, sensitivity of 91.3%, a false positive rate of 8.0%, and an F1-score of 91.0%.

Additionally, another study [90] introduced a novel approach, 3D Axial-Attention, for lung nodule classification, aiming to improve efficiency compared to regular non-local networks. The 3D Axial-Attention operates on each axis independently, requiring less computing power. To address the invariant position problem, 3D positional encoding is added to shared embeddings. The proposed method outperforms state-of-the-art

approaches on the LIDC-IDRI dataset, achieving an AUC of 96.17, accuracy of 92.81, precision of 92.59, and sensitivity of 92.36. This highlights the effectiveness of the 3D Axial-Attention approach in enhancing lung nodule classification performance.

3 Methods

The objective of the present thesis is to develop deep learning models that can be incorporated into a lung cancer computer-aided diagnostics system and assist doctors with the screening. Our vision is to create a supporting system that processes the patient's CT scan, where a semantic segmentation model pre-labels suspicious nodules and another model classifies their malignancy. These predictions are presented to the doctors for their review and speed-up their diagnosis.

In this Chapter, we describe our observations from the Exploratory Data Analysis that was conducted in the LIDC-IDRI database, our hypotheses, and the experiment setting, which includes the data preparation, architectures used, and methods.

3.1 Exploratory Data Analysis

Before defining our experiment protocol, it is important to understand the nature of our data. This includes exploring our data to understand the structure and format. From the literature, most of the studies used the publicly available LIDC-IDRI database [16], or its Kaggle competition subset LUNA16 [17], while some studies used private data. For this thesis, we preferred to use only publicly available data and as a result, we decided to use data from the LIDC-IDRI database. To extract information and labels from the database we used the pylidc python package to load the CT scans and the corresponding metadata. This supplementary report contains the Exploratory Data Analysis results. In this section we represent some useful counts we retrieved from our data.

3.1.1 Scan class

This section contains information on the scan object. The pylidc.Scan class encompasses a subset of the DICOM attributes linked with the CT scans within the LIDC dataset. These attributes serve as valuable query parameters for accessing the dataset. It is noteworthy that we obtained the remaining metadata directly from the DICOM header.

3.1.1.1 Slice spacing

The slice_spacing property, represented as a floating-point value, is computed as the median of the differences between the coordinates of consecutive slices within a scan. These coordinates are referred to as the slice_zvals. Although the `slice_spacing` attribute generally aligns with the slice_thickness, it is important to note that this alignment is not a

strict rule and may vary. Additionally, it is crucial to understand that the `slice_spacing` does not necessarily indicate uniform spacing between all slices, although this is commonly observed. This property plays a pivotal role in characterizing the arrangement and relationships between the slices in a scan dataset.

Table 3.1 LIDC scans slice thicknesses range and counts

Slice thickness	Scans
≤ 1mm	97
1mm - 2mm	478
2mm - 2.5mm	322
2.5mm - 3mm	117
>3mm	4

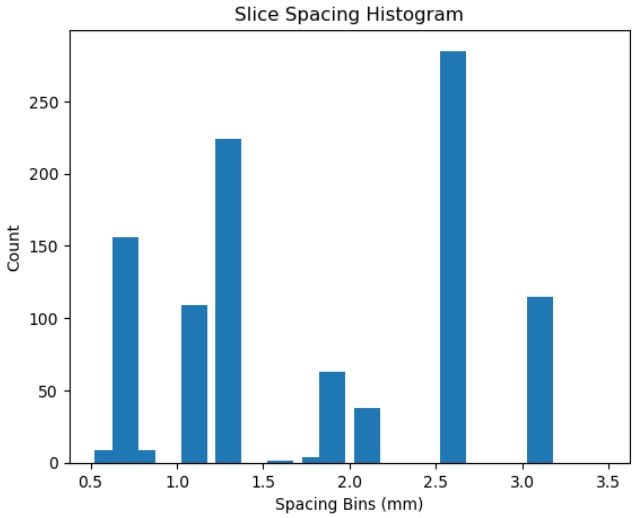


Figure 3.1 Slice spacing histogram

3.1.1.2 Pixel spacing

The pixel_spacing attribute, represented as a floating-point value, corresponds to a Dicom attribute (0028,0030) that typically consists of two values. In the context of the LIDC dataset, all scans share identical resolutions within the transverse plane. As a result, only one value is utilized to represent this attribute in the LIDC. This property is crucial for understanding the spatial relationship between pixels in medical images, aiding in accurate measurements and assessments during medical image analysis.

Table 3.2 LIDC scans pixel spacing ranges and counts

Slice pixel spacing	Scans
≤ 1mm	39
1mm - 2mm	412

2mm - 2.5mm	124
2.5mm - 3mm	322
>3mm	121

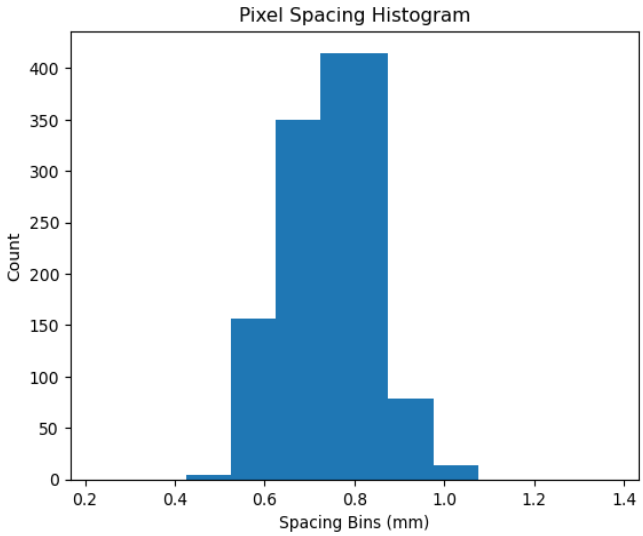


Figure 3.1.1 LIDC-IDRI scans pixel spacing ranges and counts

3.1.2 Annotation class

The Nodule model class serves as a comprehensive container for encapsulating vital information derived from the meticulous annotations of physicians regarding nodules of a size equal to or greater than 3mm, within the context of a specific medical scan. Each nodule instance encompasses a series of contours, wherein each contour corresponds to the precise outline drawn for the nodule across individual scan slices. This model class encompasses a spectrum of qualitative attributes that collectively offer a multidimensional representation of the nodule's characteristics is shown in Table 3.3.

Table 3.3 LIDC-IDRI Nodule characteristics and annotations

Attribute	Possible Values	Description
Subtlety	1, 2, 3, 4, 5	Difficulty of detection, ranging from 'Extremely Subtle' to 'Obvious'
Internal Structure	1, 2, 3, 4	Internal composition of the nodule, categorized as 'Soft Tissue', 'Fluid', ‘Air’
Calcification	1, 2, 3, 4, 5	Pattern of calcification, including 'Popcorn', 'Laminated', ‘Solid, ‘Non-central’, ‘Absent’
Sphericity	1, 2, 3, 4, 5	The three-dimensional shape of the nodule, like 'Linear', 'Ovoid/'Linear’, ‘Ovoid’, 'Ovoid/Round', ‘Round’
Margin	1, 2, 3, 4, 5	Description of nodule margin's well-defined nature, e.g., 'Poorly Defined', 'Sharp', etc.
Lobulation	1, 2, 3, 4, 5	Degree of lobulation, from 'No Lobulation' to 'Marked Lobulation'.
Spiculation	1, 2, 3, 4, 5	Extent of spiculation presence, ranging from 'No Spiculation' to 'Marked Spiculation'.
Texture	1, 2, 3, 4, 5	Radiographic solidity and internal texture, including 'Non-Solid/GGO', 'Solid', etc.
Malignancy	1, 2, 3, 4, 5	Subjective likelihood of malignancy, considering patient characteristics and labeled from 'Highly Unlikely' to 'Highly Suspicious'.

The subtlety attribute, represented by an integer within a predefined range, gauges the complexity of nodule detection. A higher value on this scale indicates an easier detection process, encompassing categories like 'Extremely Subtle' to 'Obvious', enabling clinicians to gauge the intricacy of identifying the nodule within the scan. The internal Structure attribute encapsulates the module's internal composition, employing categories such as 'Soft Tissue', 'Fluid', 'Fat', and 'Air', illuminating the diverse tissue constituents within the nodule.

Calcification patterns are captured through the calcification attribute, utilizing descriptors such as 'Popcorn', 'Laminated', 'Solid', and more, thereby providing insights into the presence and arrangement of calcified regions. The nodule's three-dimensional form is captured by the sphericity attribute, quantifying its roundness with values like 'Linear', 'Ovoid', and 'Round', enhancing the understanding of its morphological characteristics.

Details regarding the nodule's margin definition are encoded within the margin attribute, presenting categories ranging from 'Poorly Defined' to 'Sharp', delineating the clarity of its boundary. Lobulation, another crucial aspect, is quantified by the lobulation attribute, with degrees ranging from 'No Lobulation' to 'Marked Lobulation', providing an assessment of the nodule's surface irregularities.

Moreover, the presence of spiculation, or spike-like projections, is expressed through the spiculation attribute, capturing varying levels of spiculation extent. Texture, reflecting the internal solidity of the nodule in radiographic terms, is evaluated using labels such as 'Non-Solid', 'Solid/Mixed', and others. Lastly, the attribute malignancy provides a subjective evaluation of the potential malignancy of the nodule, considering factors like the patient's profile, with gradations like 'Highly Unlikely' to 'Highly Suspicious', facilitating an informed assessment of the nodule's clinical significance.

In summary, the Nodule model class amalgamates a rich array of attributes, each representing distinct facets of nodules' radiographic and morphological properties. This holistic framework equips medical professionals with a nuanced understanding of nodules, enabling more accurate diagnoses and facilitating informed clinical decisions. This class contains nodule-specific attributes. Each nodule has a unique ID. The pylidc library has a built-in function to group nodules based on the annotation coordinates in the scan. The following figure demonstrates the label counts for each attribute.

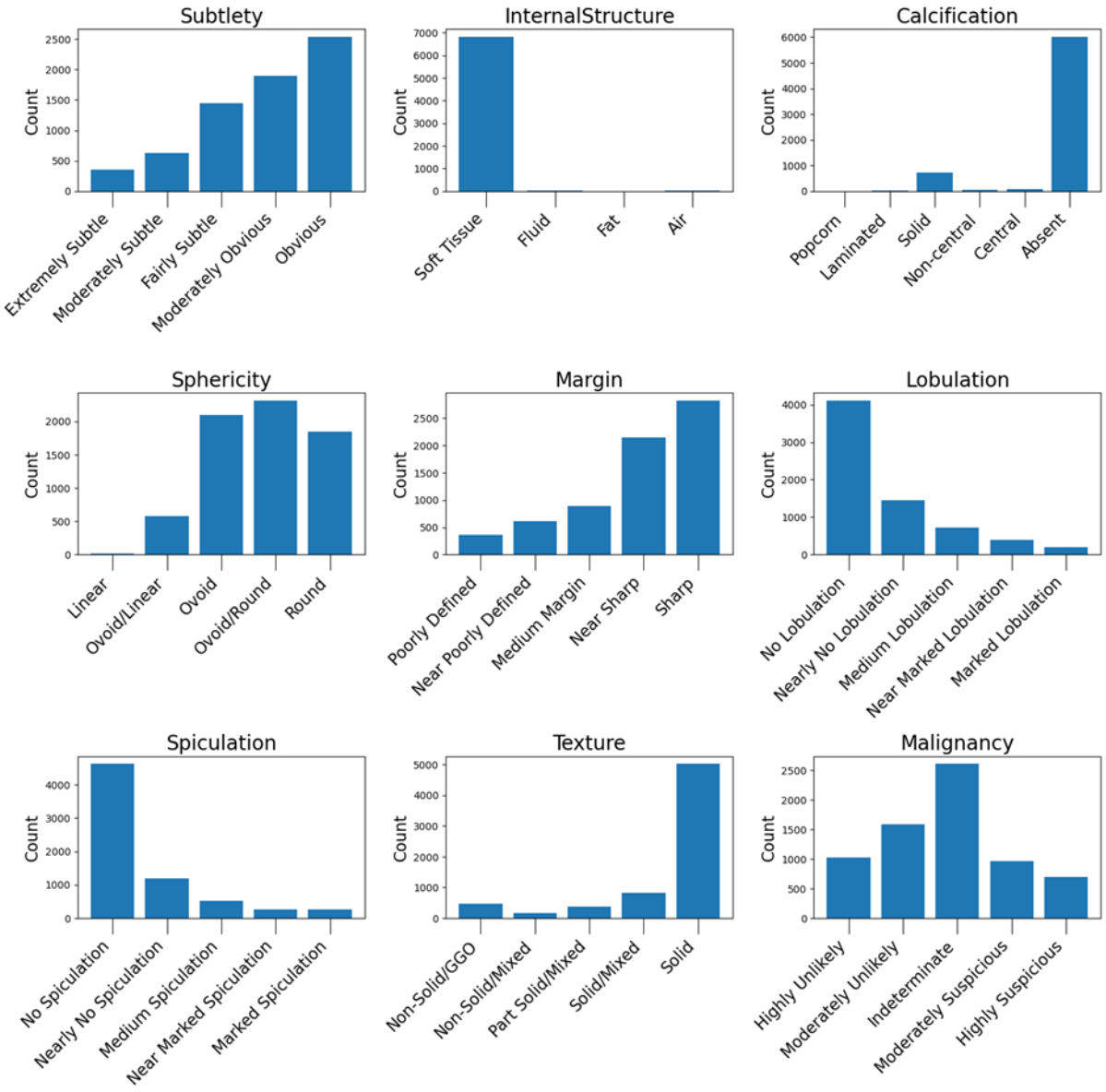


Figure 3.1.2 Label distribution of nodule characteristics in LIDC-IDRI database

3.1.3 Three-dimensional nodule samples

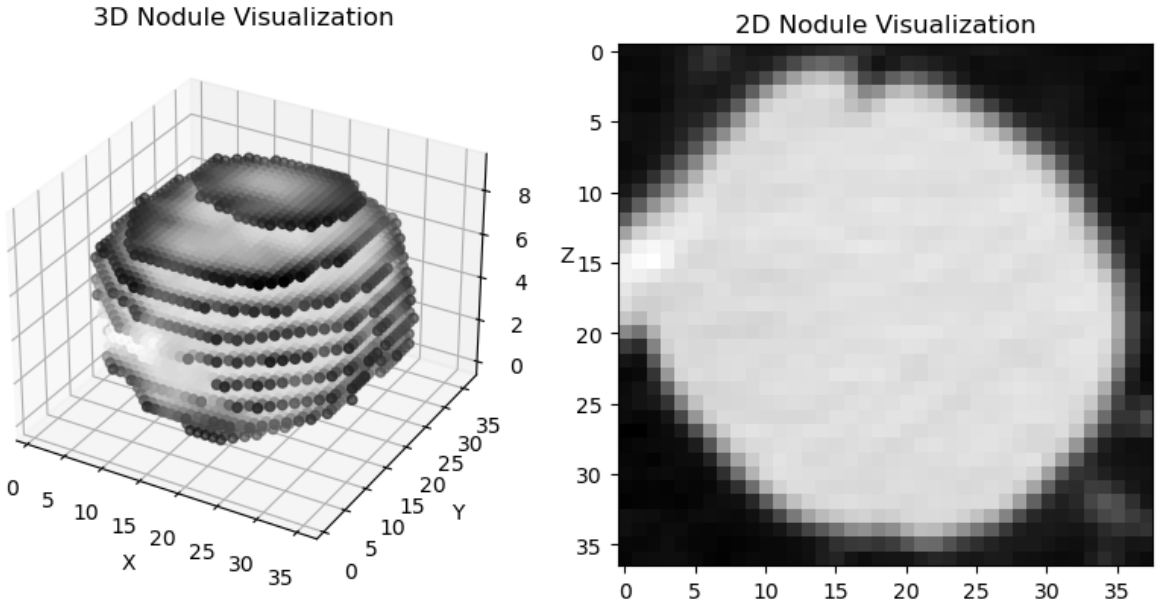


Figure 3.1.3 Three-dimensional (3D) nodule sample

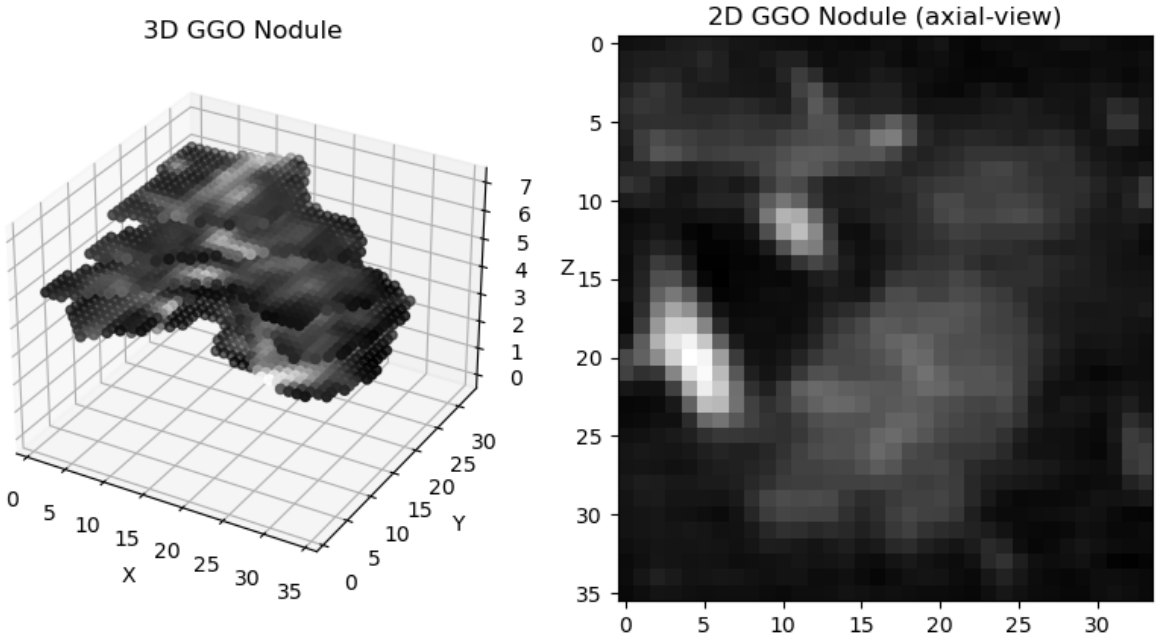


Figure 3.1.4 Three-dimensional (3D) Ground-glass opacity (GGO) nodule sample

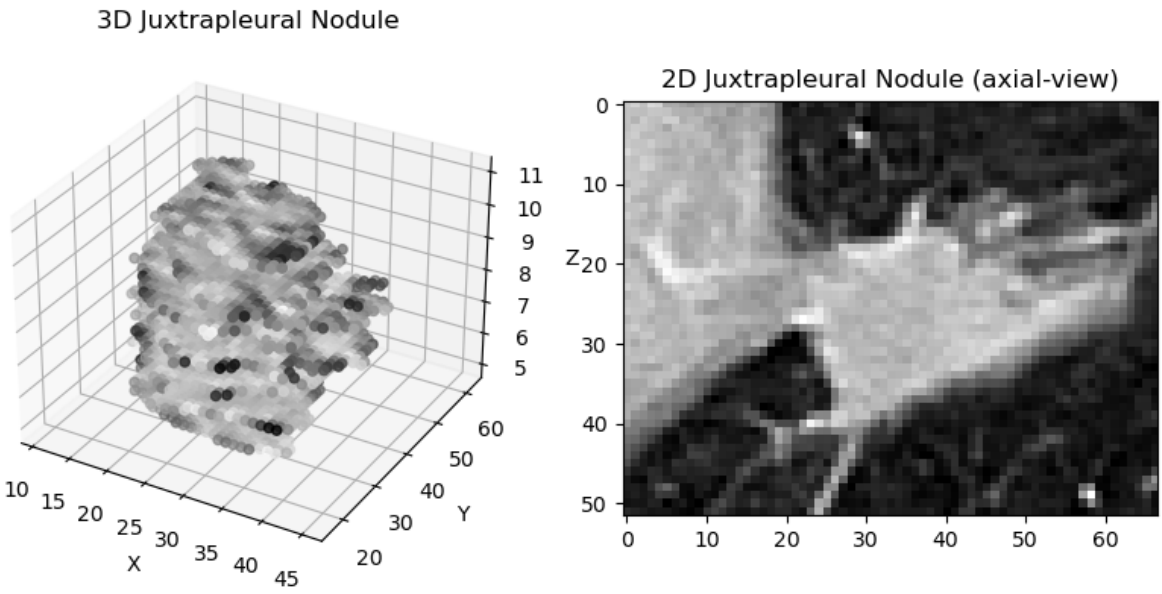


Figure 3.1.5 Three-dimensional (3D) juxtapleural nodule sample

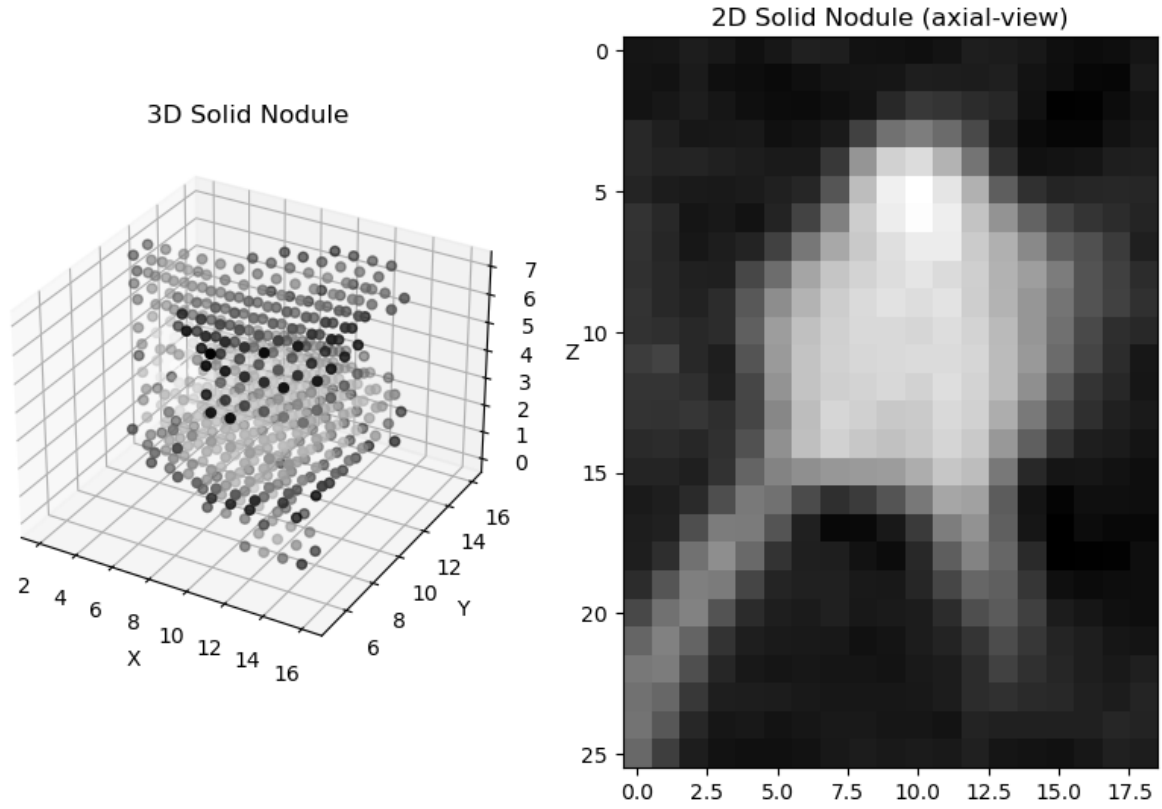


Figure 3.1.6 Three-dimensional (3D) solid nodule sample

3.1.3.1 Nodule diameter

The concept of "diameter" involves estimating the maximum axial plane diameter of a nodule through the utilization of its annotated contours. This estimation serves as a valuable metric

for assessing the size of nodules in medical imaging. However, it is important to note that the current estimation method does not incorporate situations in which the diameter extends beyond the nodule's boundary or traverses through internal cavities within the nodule. As a result, while this estimation provides a useful approximation of the nodule's size, it may not capture certain complex spatial configurations. This consideration underscores the need for further refinement and advanced techniques to comprehensively characterize nodules in medical image analysis.

The analysis of nodule diameters was organized into distinct groups based on their sizes. These groupings facilitate a comprehensive understanding of nodule characteristics. The diameters were categorized as follows: nodules with diameters less than or equal to 5 mm, those ranging from 5 mm to 10 mm, nodules spanning 10 mm to 15 mm, nodules within the 15 mm to 20 mm range, and nodules with diameters exceeding 20 mm. This categorization aids in gaining insights into the distribution and prevalence of nodules of varying sizes within the studied dataset. By counting the nodules falling into each diameter group, we acquire valuable information about the distribution of nodule sizes, which contributes to a more detailed and informative analysis of the dataset.

Table 3.4 Nodule diameter distribution and counts

Nodule diameter	**Nodule counts**
≤ 5 mm	349
5mm - 10mm	3.886
10mm - 15mm	1.325
15mm - 20mm	553
> 20mm	746

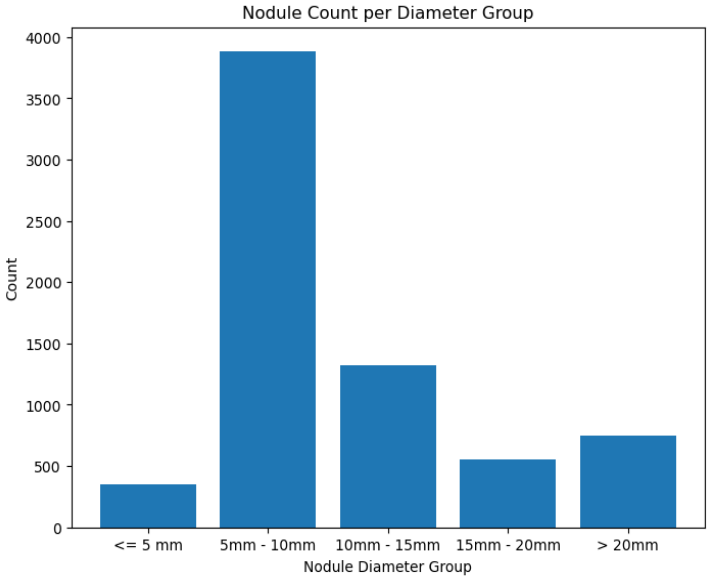


Figure 3.1.7 Nodule diameter distribution

After completing this stage, we gained more knowledge about our scan data and the labels. Using this knowledge, we will define the experiment protocol and design and implement our data preprocessing strategy.

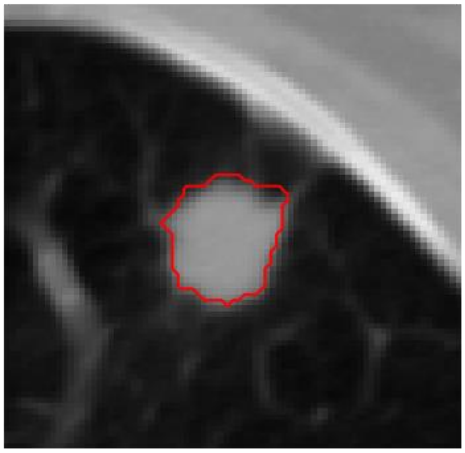


Figure 3.1.8 Nodule annotation

3.1.3.2 Nodule visualization example using PyLIDC

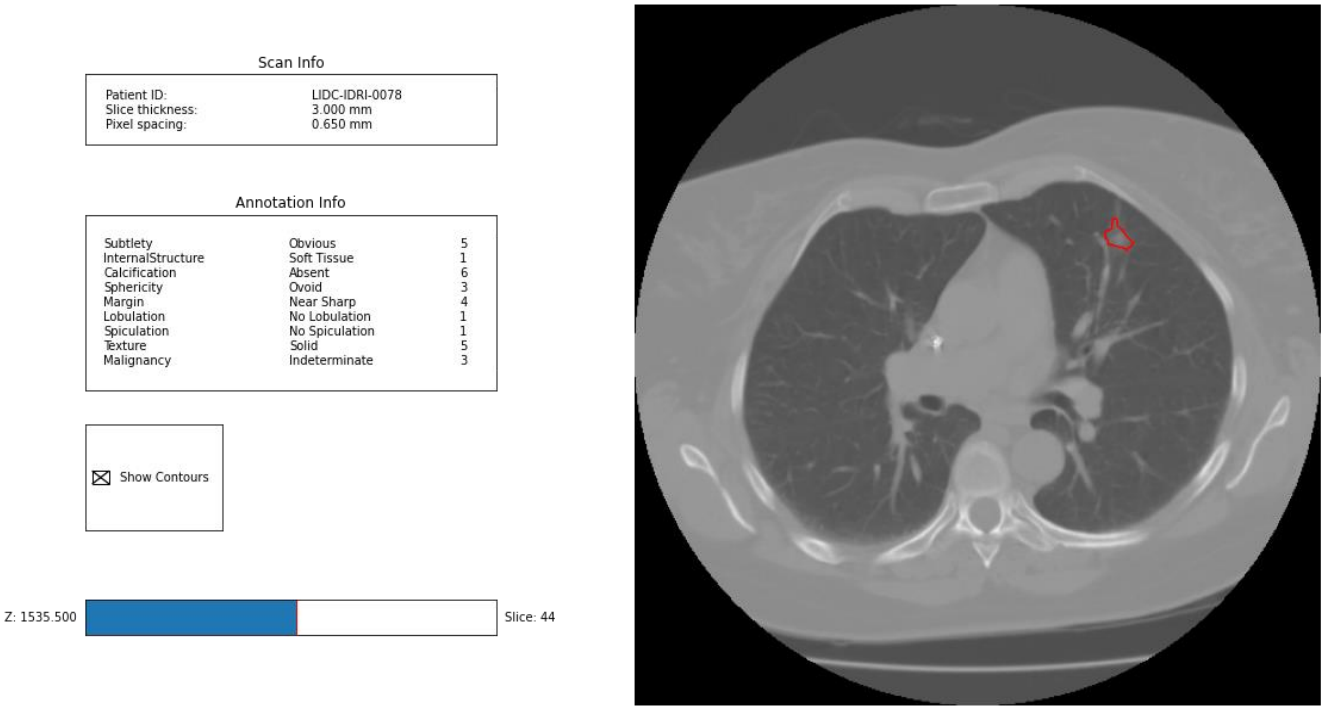


Figure 3.1.9 LIDC-IDRI sample annotation

3.2 Hypotheses

In our exploration of pulmonary nodule analysis within CT scans, we have outlined a series of hypotheses that will steer our research endeavors. These hypotheses form the guiding principles that will shape our experiments and propel us toward a deeper understanding of the subject. Our investigation focuses on two main areas: nodule segmentation, where we aim to precisely locate and define nodules, and nodule malignancy classification, where we strive to categorize nodules based as benign or malignant. Through these hypotheses, we aim to leverage advanced techniques to unveil valuable insights, contributing to the advancement of medical imaging analysis and its impact on patient well-being.

3.2.1 Nodule semantic segmentation

Usage of a semantic segmentation model to generate nodule candidates from a CT scan patch. This approach will provide us with pixel-level predictions for any nodules contained in the patch. Using this model output, we will be able to construct a mask over the nodule area.

- **Hypothesis 1**. Train a 3D ResNet (transfer learning) to segment pulmonary nodules in a 3D CT scan patch

3.2.2 Nodule malignancy classifiers

- **Hypothesis 1**. Fine-tune a 3D CNN to classify nodules as benign or malignant
- **Hypothesis 2**. Fine-tune a 3D CNN using masked nodules to classify nodules as benign or malignant

4 Modeling approach

Our modeling approach revolves around two key strategies: semantic segmentation and classification. For semantic segmentation, we employ advanced architectures like the 3D HighResNet. This model excels at extracting intricate details from three-dimensional data, making it a perfect fit for complex medical scans. It even learns from brain parcellation, enhancing its ability to accurately segment brain regions. This approach helps us precisely outline nodule structures within volumetric data. For the malignancy classification task, we turned to the 3D ResNet18. It is a powerful adaptation of a proven architecture, enabling us to classify volumetric medical scans efficiently. In essence, our approach involves utilizing advanced models to extract valuable insights from medical images using segmentation and classification techniques.

4.1 Semantic segmentation architecture – 3D HighResNet

Built on the foundation of residual connections, the 3D HighResNet model addresses gradient vanishing issues and supports deep network structures. It excels at learning complex three-dimensional features from raw data, enabling automatic extraction of anatomical and functional details present in brain scans. The architecture's emphasis on representation learning transforms brain parcellation into a data-driven problem rather than one reliant on hand-engineered features.

One notable aspect is the model's use of brain parcellation as a pretext task. This auxiliary training approach demonstrates the architecture's capacity to learn meaningful brain representations, ultimately enhancing its performance on the primary task of brain segmentation. The model's efficacy is validated through extensive experiments on diverse brain imaging datasets, showcasing its ability to accurately segment distinct brain regions.

In summary, the 3D HighResNet architecture presents an innovative solution for 3D brain parcellation. By leveraging the ResNet principles and emphasizing representation learning, it offers a compact and efficient means to address the complexities of volumetric neuroimaging data.

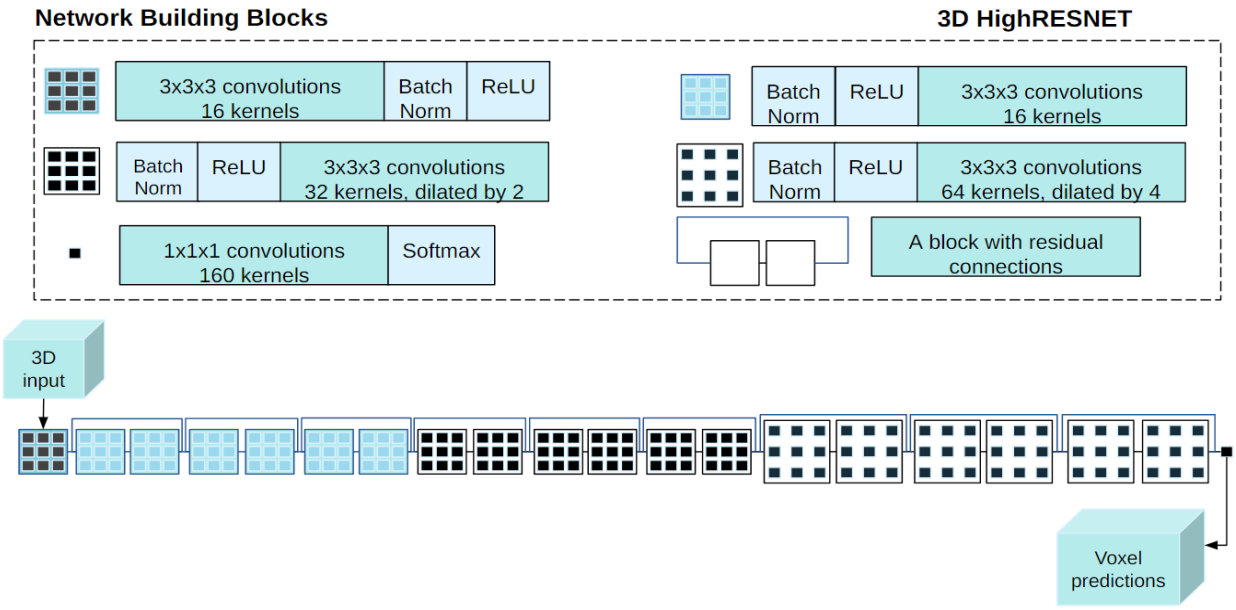


Figure 4.1.1 3D HighResNet architecture

4.2 Classification architecture – 3D ResNet18

The 3D ResNet18 architecture is an extension of the ResNet architecture that has been adapted to work with three-dimensional volumetric data. ResNet architectures are renowned for their ability to train very deep neural networks by mitigating the vanishing gradient problems through residual connections. The 3D ResNet18 architecture is effective for tasks such as medical image analysis due to its capacity to learn hierarchical features from volumetric data. It enables the network to capture both low-level local features and high-level global context, which is crucial for tasks like lung nodule detection and segmentation.

The 3D ResNet18 architecture follows the foundational principles of ResNet, introducing residual blocks that allow the network to efficiently learn and represent complex features. The architecture consists of a series of convolutional layers grouped into residual blocks of varying depths. Each residual block contains convolutional layers followed by batch normalization and ReLU activations, while the final layer in each block incorporates a skip connection that directly adds the input to the output. This skip connection allows gradients to flow freely, preventing degradation in the learning process even as the network becomes deeper.

In the 3D version, the convolutional layers within each residual block operate on three-dimensional data, accommodating the volumetric nature of medical images. The

architecture retains the fundamental ResNet building blocks, such as the basic block and bottleneck block, but adapts them to the 3D context.

In the case of 3D ResNet18, the architecture comprises multiple residual blocks with 3D convolutional layers. The overall depth of the network is significantly shallower compared to more advanced variants like ResNet50, making it computationally more efficient while still retaining the benefits of residual learning.

To tailor the 3D ResNet18 architecture to specific tasks, modifications can be made to accommodate task-specific input sizes, output classes, and loss functions. Data augmentation, regularization techniques, and transfer learning can also be incorporated to enhance generalization and improve performance.

In summary, the 3D ResNet18 architecture extends the ResNet concept to three-dimensional volumetric data, utilizing residual blocks to train deep neural networks effectively. By integrating skip connections and 3D convolutional layers, the architecture excels in capturing intricate features from medical images, making it a valuable asset for tasks demanding detailed analysis of volumetric data.

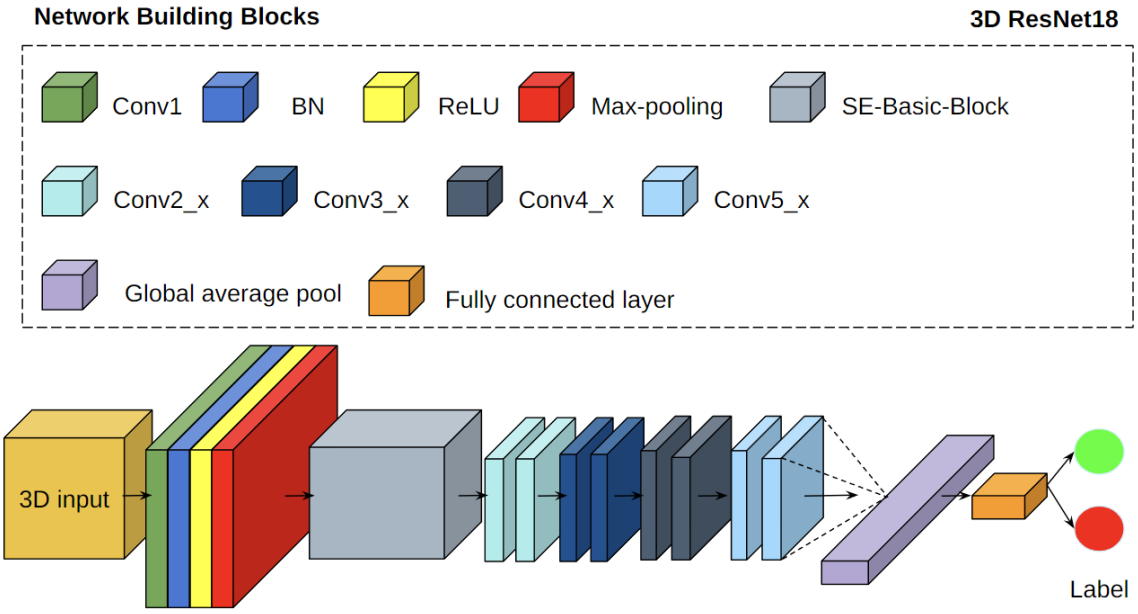


Figure 4.2.1 3D ResNet18 architecture

5 Implementation

5.1 Data preprocessing

Based on our hypotheses we will need volume patches for training the semantic segmentation models and nodule patches to train the classification models. We performed the following operations to prepare our data for training.

Scan level:

- Filtered out CT scan with over 2.5mm spacing between slices
- Converted CT scan values to Hounsfield units (HU)
- Resample scan and mask to uniform spacing of 1×1×1 mm using the pixel spacing and slice thickness
- Data normalization between -1 and 1
- Save the resampled volume and masks to Nifti images

Nodule level:

- Filtered out nodules with less than 3 annotations and with a diameter less than 3mm
- Nodule grouping. We used the built-in function of pylidc to group nodules based on their position
- Consensus nodule mask extraction. We combined all the masks that the doctors annotated and using 50% Consensus + extracted an average mask
- Nodule characteristics extraction. There were disagreements between some of the annotations generated by the doctors so we developed a method to extract the most common label per characteristic.
- Extract 3D and save nodule crops as Nifti images

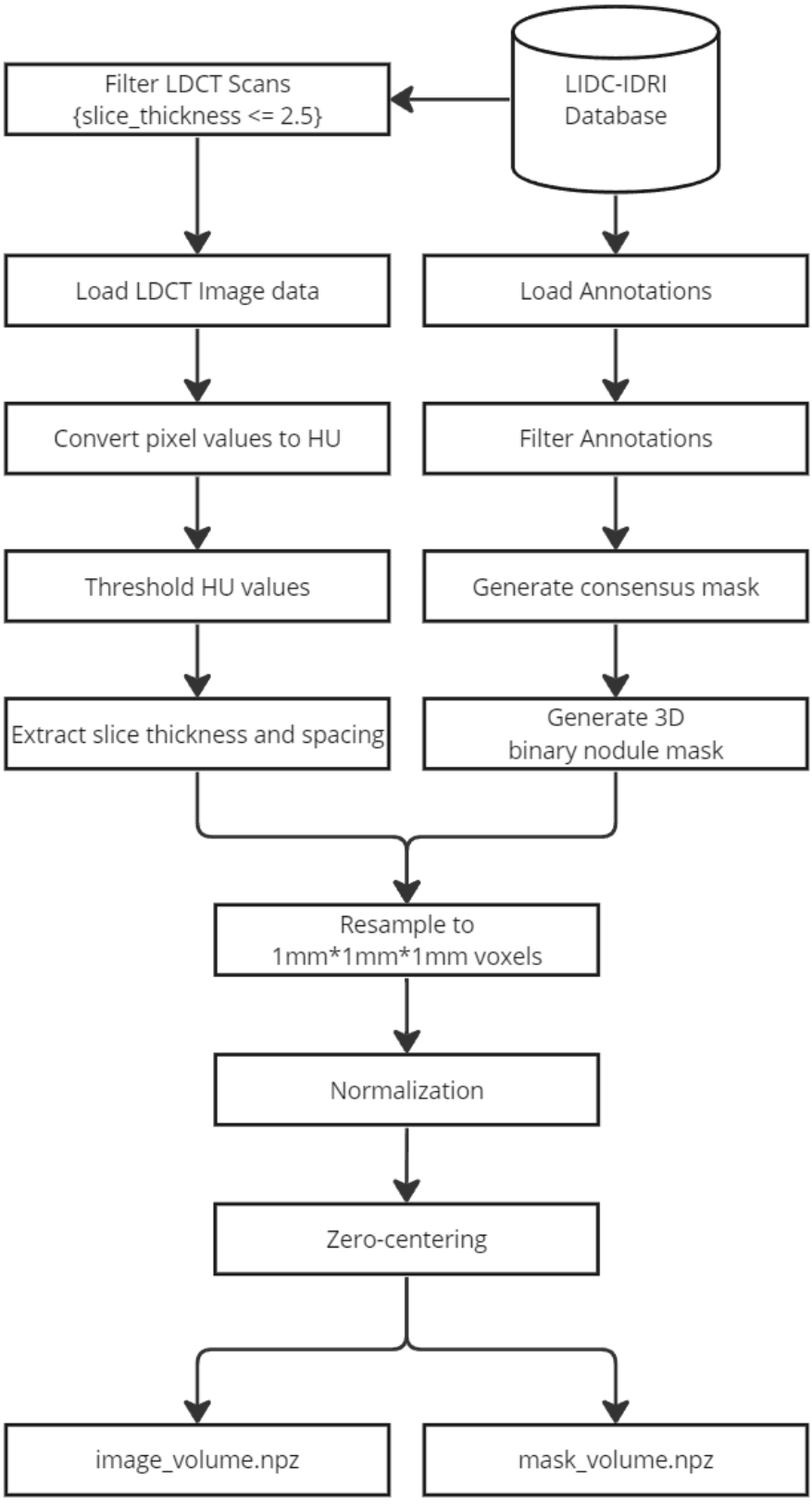


Figure 5.1.1 Preprocessing pipeline

The decision not to segment the lung parenchyma in our study was driven by the specific context of juxta-pleural nodules. Juxta-pleural nodules, owing to their proximity to the pleural lining, often exhibit intricate relationships with the surrounding lung tissue. Segmenting the lung parenchyma would potentially obscure or alter the spatial relationship between these nodules and the pleural boundary, which could have a direct impact on accurate nodule analysis and classification. By maintaining the contextual integrity of the juxta-pleural nodules in relation to their neighboring lung tissue, we aimed to ensure that the analysis accurately captured the unique characteristics of these nodules. This strategic choice aligns with our objective to provide precise insights into nodule attributes, acknowledging the significant implications that the lung parenchyma segmentation could have on the analysis of juxta-pleural nodules.

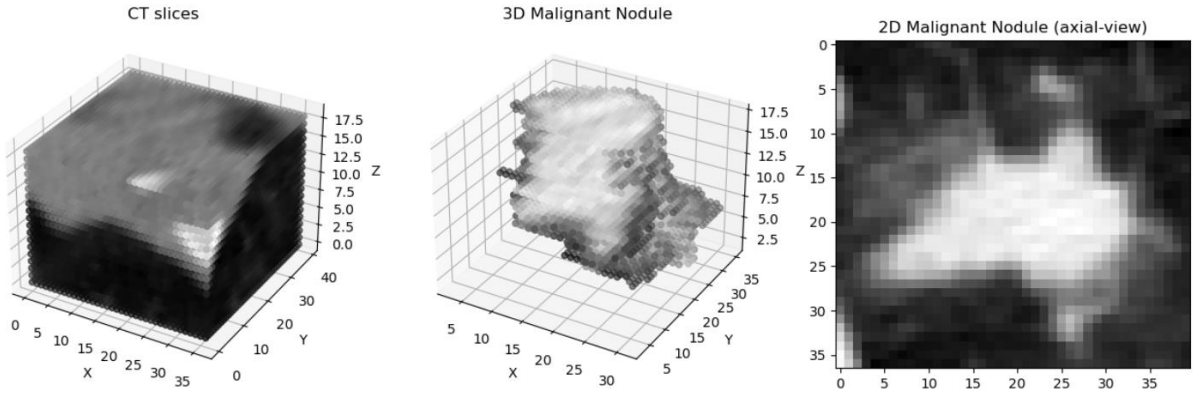


Figure 5.1.2 Preprocessed data sample.

5.1.1 Processed scan counts

With the dataset now preprocessed according to the specified criteria, we are ready to conduct our experiments. The counts provided here reflect the dataset after it has been filtered based on the earlier described parameters. This prepares us for a more targeted and effective experimental phase, utilizing data that aligns closely with our research objectives and criteria.

Table 5.1 Processed scan counts

Slice thickness	Scans count	Grouped Nodule counts
≤1mm	97	151
1mm-2mm	478	678
>2mm	322	363

5.1.2 Grouped nodule malignancy labels

Table 5.2 Grouped nodule malignancy labels

Malignancy Labels	Grouped Nodule counts
Highly Unlikely	155
Moderately Unlikely	222
Indeterminate	465
Moderately Suspicious	178
Highly Suspicious	172

5.2 Patch extraction and data augmentation

CT volumes are typically very large and it is hard to fit entire volumes to the memory while training. Instead of loading the entire scan, we decided to use a patch-based extraction pipeline.

To efficiently handle the loading and preprocessing of LDCT data for our patch-based analysis, we harnessed the capabilities of PyTorch IO. PyTorch IO is a specialized library tailored to the unique requirements of medical imaging data. It provided us with the tools necessary to seamlessly load and augment our CT images, ensuring that they were appropriately prepared for subsequent analysis

We developed a patch-based extraction method to extract 3D volumes and use them as training samples. These 3D patches are extracted online during training so in each epoch the patches are different and this is one of the methods we used to avoid overfitting on the training set. On top of that, we applied data augmentation techniques in the 3D space. More specifically we applied the following transformations:

- Random flip
- Random affine
- Random elastic deformation
- Random swap

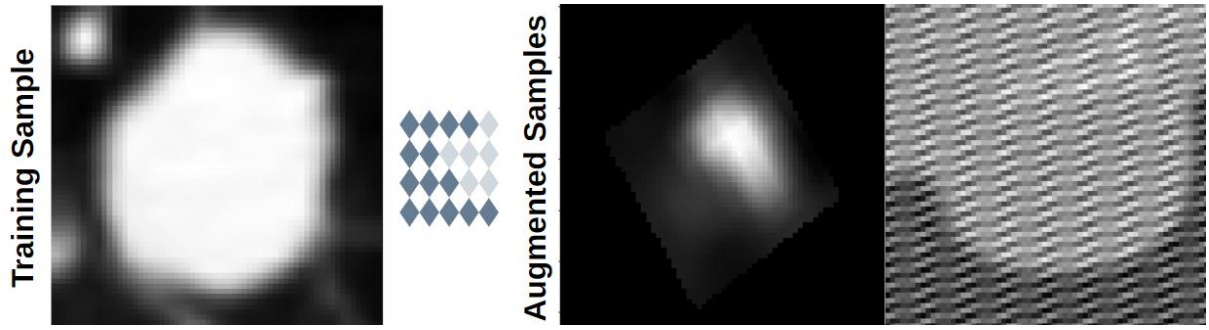


Figure 5.2.1 Data augmentation example

These transformations are randomly applied to the training data at the start of each epoch.

In our methodology, we implemented a 10:1 probability ratio for extracting a 3D patch containing a nodule sample, extracting four patches from each scan. This approach of probabilistic patch extraction allowed us to dynamically obtain 3D patches that are highly likely to contain nodule samples, while also maintaining a chance of extracting patches without nodules. To enhance the diversity of our training data, we applied transformations

to it in each epoch. However, for the validation and testing datasets, we only performed patch extraction without additional transformations. This strategy ensures a robust training process while maintaining the integrity of the validation and testing data for accurate model evaluation.

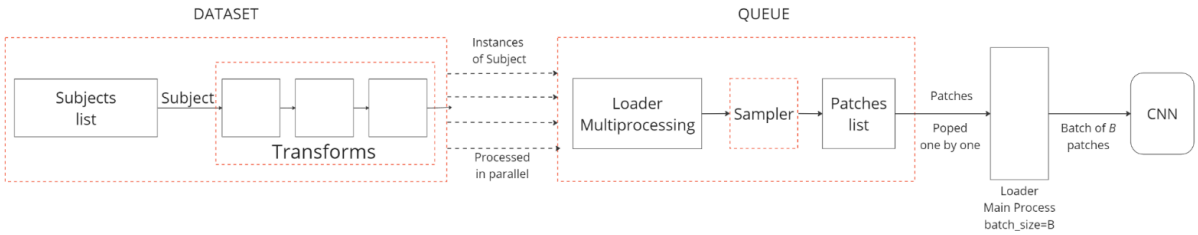


Figure 5.2.2 Patch extraction pipeline

5.3 Dataset split

In this thesis, a 5-fold cross-validation approach was employed to rigorously evaluate the performance and generalization capabilities of the developed models. By partitioning the dataset into five subsets and iteratively training and testing the model on different combinations of these subsets, we ensured a comprehensive assessment of its effectiveness. This approach not only minimized the potential for overfitting but also provided a robust estimation of the model's predictive power across diverse data instances.

5.3.1 Nodule segmentation dataset

An essential component of conducting experiments and evaluations in machine learning and deep learning research is the appropriate division of the dataset into training, validation, and test subsets. The data split strategy plays a crucial role in assessing the generalization performance of the models and ensuring their robustness. We applied a data split to our preprocessed dataset of 80% for training and 20% for validation. We further split the validation to extract a test using 10% of the validation set. We also generated a filtered dataset that contains only scans with at least one nodule. The following table represents the number of scans per set:

Table 5.3 CT scans split

Set	Number of scans	Scans with at least one nodule

Train	716	389
Validation	144	78
Testing	36	20

5.3.2 Nodule classification dataset

Given the limited number of samples in the 'Highly Unlikely' and 'Highly Suspicious' categories, we merged these with the 'Moderately Unlikely' and 'Moderately Likely' classes, respectively. This merging was done to create a new dataset tailored for our binary classification problem. Additionally, we chose not to include the 'Intermediate' class in our classification experiments, concentrating solely on the more clearly defined categories to ensure more distinct and reliable classification outcomes.

Table 5.4 Classification nodule samples count

Class Labels	**Sample Nodule counts**
Benign	377
Malignant	350

5.4 Training experiments

Our model architecture, a critical component of our project, was elegantly constructed using PyTorch Lightning. PyTorch Lightning is a high-level wrapper around PyTorch that abstracts away many of the intricacies of training deep learning models. We encapsulated our model definition within a lightning class, which streamlined our workflow. This class not only encapsulated the architecture but also incorporated the optimizer and loss functions.

One of the standout features of PyTorch Lightning is its ability to handle the complete training pipeline with ease in parameters setting. By leveraging its built-in training loop, we were able to focus on the higher-level aspects of our research, allowing PyTorch Lightning to efficiently orchestrate the core training and evaluation processes by managing the training, validation, and testing steps. The Lightning trainer executed our defined experiments and saved weight checkpoints based on the criteria we specified. This functionality was especially useful for maintaining a record of the model's progress throughout training. By automatically checkpointing the weights at specific intervals, we safeguarded our progress against unexpected interruptions.

Experiment tracking and metrics logging were essential for gaining insights into our model's performance and behavior. To accomplish this, we incorporated the wandb (Weights & Biases) library. wandb provided us with a robust framework for tracking experiments, logging metrics, and visualizing results. This streamlined our experimentation workflow, enabling us to make data-driven decisions and fine-tune our model based on real-time insights.

In conclusion, our project was empowered by a series of powerful tools and libraries. PyTorch IO facilitated the efficient handling of medical imaging data, PyTorch Lightning simplified our model architecture and training process, and wandb enhanced our experiment tracking and metric logging capabilities. Through the synergistic use of these libraries, we were able to conduct thorough analyses, make informed decisions, and achieve meaningful results.

5.4.1 Nodule segmentation

Hyperparameters

The first loss function we used for our binary semantic segmentation problem was BCEWithLogitsLoss combining a Sigmoid layer and the BCELoss in one single class. The function is described by the following formula (1):

$$BCEWithLogitsLoss(input, target) = pos_weight \cdot pos_loss + neg_loss \quad (1)$$

Where:

- input is the input tensor representing the logits from the model.
- target is the target tensor containing binary labels (0 or 1).
- pos_weight is an optional weight to balance positive and negative examples (default is 1).
- pos_loss is the binary cross-entropy loss for positive examples.
- neg_loss is the binary cross-entropy loss for negative examples.
- The binary cross-entropy loss for positive and negative examples is defined as:
- pos_loss = $-\log(\sigma(\text{input}))$
- neg_loss = $-\log(1-\sigma(\text{input}))$

Where σ is the sigmoid function, and log is the natural logarithm.

Putting it all together, the BCEWithLogitsLoss combines the positive and negative losses with an optional weight to form the final loss function used during the training of binary classification models.

The other utilized loss function for our binary semantic segmentation problem, was the Dice Loss, a well-suited metric for assessing the performance of models in scenarios where the task involves distinguishing between two classes. The Dice Loss, or Sørensen-Dice coefficient, is defined by the formula (2):

$$Dice\ Loss = 1 - \frac{2 \times Intersection(p, g)}{Union(p, g) + \epsilon} \quad (2)$$

Where:

- p denotes the predicted binary segmentation mask,
- g denotes the ground truth binary segmentation mask,
- Intersection(p,g) represents the intersection of the predicted and ground truth masks,
- Union(p,g) represents the union of the predicted and ground truth masks,

- ϵ is a small constant added to the denominator to prevent division by zero.

A lower Dice Loss indicates improved alignment between the predicted and ground truth masks. The Dice Loss is particularly advantageous for binary segmentation tasks, contributing to robust model performance even when faced with imbalanced datasets.

Throughout the training process, the model fine-tunes its parameters to minimize the Dice Loss, thereby enhancing its capability to accurately delineate the regions of interest in the binary segmentation images.

We began our training with an initial learning rate of 0.001. To dynamically adapt to the training progress, we incorporated a Learning Rate (LR) Scheduler, which effectively reduced the learning rate when the loss function reached a plateau. Throughout the training process, we employed the Stochastic Gradient Descent (SGD) [91] as our optimization algorithm during training. We performed 2 training rounds with different loss functions. First, we trained the model for 60 epochs using BCE as our loss function and then by using the model's checkpoint we continued the training for another 80 epochs using *Dice loss*.

Table 5.5 Segmentation hyperparameters.

Model	Learning rate (LR)	LR Scheduler	Epochs	Batch size	Loss function	Optimizer
3D HighResNet	0.001	StepLR	60	8	BCEWith LogitsLoss	SGD
3D HighResNet	0.0001	StepLR	80	8	Dice loss	SGD

5.4.2 Nodule classification

Hyperparameters

For the binary malignancy classifier, we used pre-trained weights and froze the first 10 layers of the network. Using this transfer learning technique, we will reduce the training time and reach model convergence in fewer epochs than training the network from scratch. We used CrossEntropyLoss as the loss function (3):

$$CrossEntropyLoss(input, target) = -\frac{1}{N}\sum_{i=1}^{N} \left(target_i \cdot log\ log\ (softmax(input)_i)\right) \quad (3)$$

Where:

- input is the input tensor representing the raw logits from the model.
- target is the target tensor containing class labels (as integers).
- *N* is the number of classes.
- softmax(input) applies the softmax function to the logits, converting them into probabilities.

The CrossEntropyLoss (3) computes the negative log likelihood of the predicted probabilities for the correct class. It implicitly applies the softmax function to the model's output and then calculates the log likelihood.

The initial learning rate was 0.0001 and we reduced it on the plateau of the loss function in the validation data. The optimizer we used was Adam.

Table 5.6 Classification hyperparameters

Model	Learning rate (LR)	LR Scheduler	Epochs	Batch size	Loss function	Optimizer
3D Resnet18	0.0001	StepLR	100	4	CrossEntropyLoss	Adam

We experimented with two different input shapes, and we performed the same experiment using masked nodules as inputs. More specifically, in the first set we used multiple slices from all three views (sagittal, coronal, transverse) while in the second

experiment we used a single slice per view from the center of the nodule. This resulted into three slices per nodule.

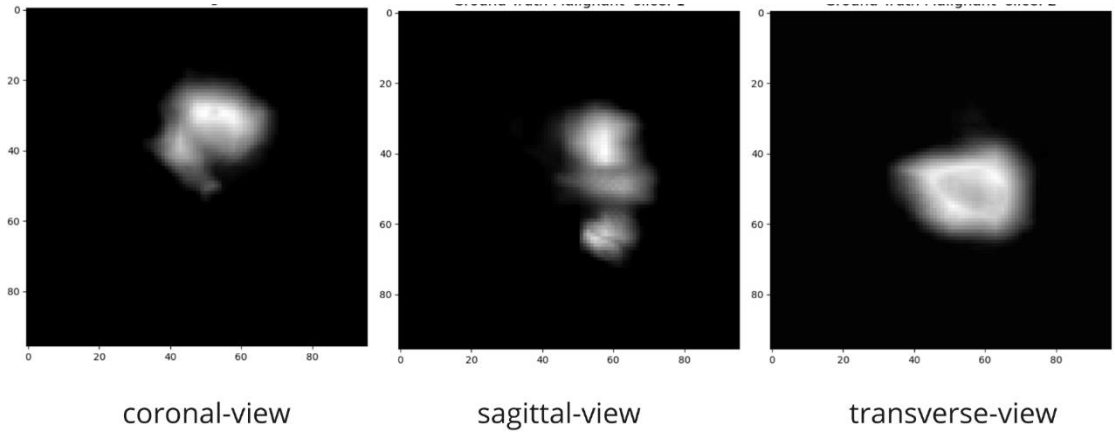


Figure 5.4.1: Nodule representation in different views

5.5 Evaluation Metrics

In order to assess the performance of our segmentation and classification models, it is essential to utilize robust evaluation metrics. These metrics provide quantitative insights into the accuracy and reliability of the predictions made by our models. This section outlines the evaluation metrics we employ to measure the effectiveness of our segmentation and classification approaches. The metrics chosen help us gain a comprehensive understanding of how well the models are performing in terms of both segmentation accuracy and classification precision. The following subsections delve into the specific evaluation metrics used for segmentation and classification, shedding light on their mathematical formulations and significance in gauging model performance.

For evaluating our segmentation performance, we will measure:

- **IoU:** Intersection over Union (IoU) is a commonly used evaluation metric for measuring the segmentation performance. It calculates the overlap between the predicted segmentation and the ground truth segmentation. The mathematical equation for IoU (4) is as follows:

$$Intersection\ over\ Union\,(IoU) = \frac{TP}{TP + FN + FP} \quad (4)$$

where TP (True Positive) represents the number of pixels correctly classified as positive, FP (False Positive) represents the number of pixels incorrectly classified as positive,

and FN (False Negative) represents the number of pixels incorrectly classified as negative.

- **DICE:** Another evaluation metric for segmentation is the DICE score, which measures the similarity between the predicted segmentation and the ground truth segmentation. The mathematical equation for the DICE score is as follows:

$$DICE = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (5)$$

The evaluation metrics for the classification model will be:

- **Precision**: Precision measures the proportion of correctly predicted positive instances out of the total instances predicted as positive. It is calculated as the ratio of true positives (TP) to the sum of true positives and false positives (FP).

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

- **Recall**: Recall, also known as sensitivity or true positive rate, measures the proportion of correctly predicted positive instances out of all actual positive instances. It is calculated as the ratio of true positives (TP) to the sum of true positives and false negatives (FN).

$$\text{Recall} = \frac{TP}{TP + FN} \quad (7)$$

- **Accuracy**: Accuracy measures the overall correctness of the model's predictions. It is calculated as the ratio of the sum of true positives and true negatives (TN) to the sum of true positives, true negatives, false positives, and false negatives.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

- **F1 score**: The F1 score is the harmonic mean of precision and recall. It provides a single metric that combines both precision and recall. It is calculated as:

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{9}$$

6 Experiment results

In this chapter, we present the experimental outcomes of our analysis, encompassing both nodules' semantic segmentation and characteristics classification. The semantic segmentation results demonstrate the effectiveness of our proposed methodology in accurately predicting the boundaries of nodules within CT scans. Furthermore, in the classification domain, our model showcases remarkable proficiency in categorizing nodules malignancy.

6.1 Segmentation performance

We conducted 3 training rounds with different loss functions (Nodule segmentation). In the first round we used the pre-trained weights of HighRes3DNet trained on an MRI brain dataset. For each next round we used the best checkpoint from the previous run. For this work we run the experiment for a single fold.

Table 6.1 Nodule segmentation performance

HighRes3DNet Performance						
Round	Set	Epochs	Loss function	Loss	IoU	DICE
1	Validation	80	Binary Cross Entropy	0.246	68.02%	79.3%
2	Validation	120	DICE_loss	0.229	71.00%	81.97%
3	Validation	60	Binary Cross Entropy	0.008	72.99%	82.88%
Evaluation	Testing	-	DICE_loss	0.222	**71.82%**	**82.65%**
Similar state-of-the-art work [46]					**70.86%**	**82.48%**

The following figure (Figure 6.1.1: Segmentation predictions) demonstrates the ground truth mask compared to the model predictions. More prediction results can be found at Appendix section A.3.1.

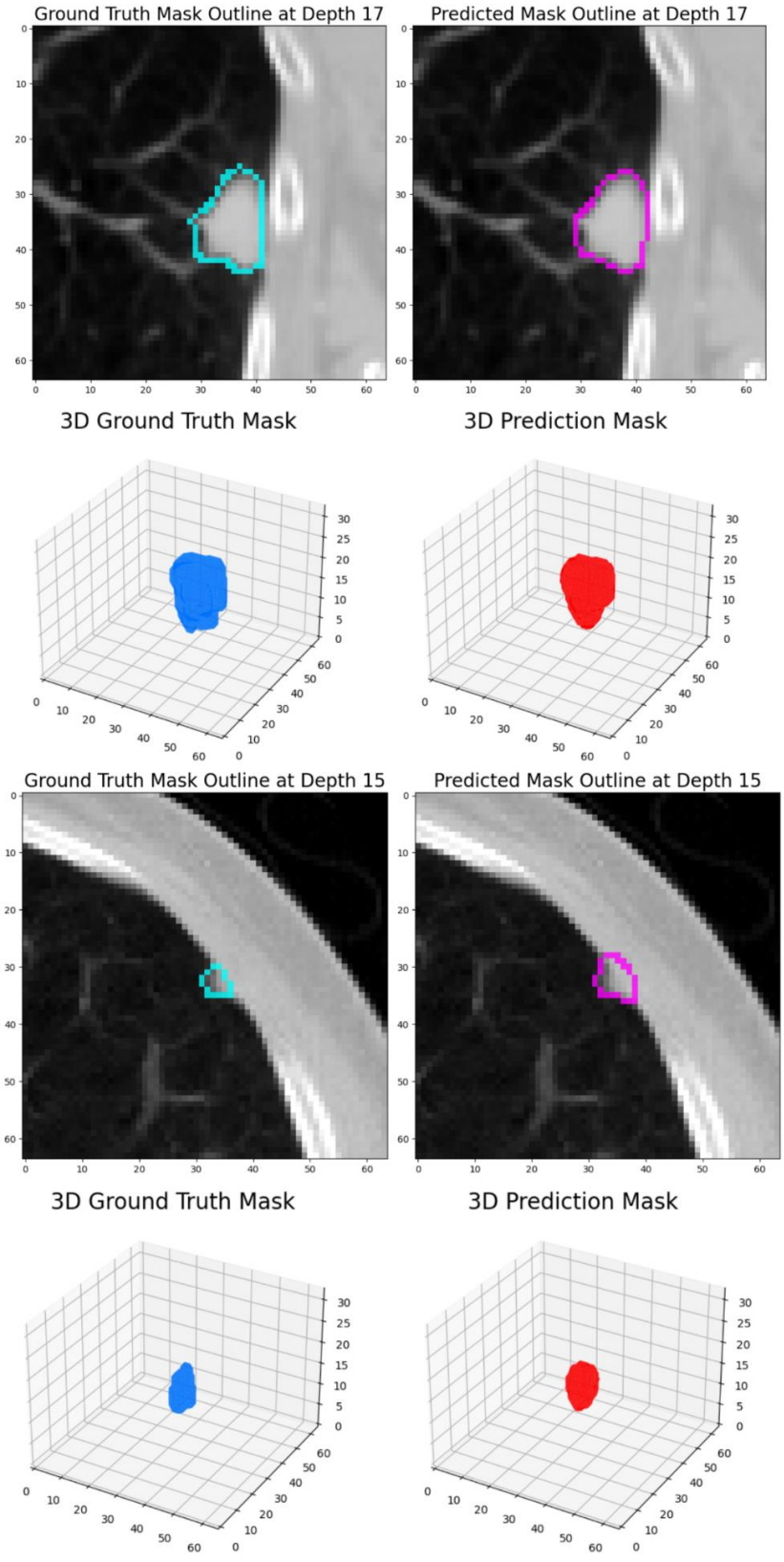


Figure 6.1.1: Segmentation predictions

6.2 Classification performance

We evaluated the performance of our malignancy classifier by calculating key metrics such as accuracy, precision, recall, and the F1-score. To enhance the reliability of our findings, we implemented a 5-fold cross-validation technique, which served to further validate our classification performance. Subsequently, we computed the average values of these metrics based on the data obtained from our 5-fold cross-validation process, providing a comprehensive assessment of the classifier's efficacy along different experiments (Nodule classification).

Table 6.2 3D ResNet18 validation performance

3D ResNet18[64, 64, 64] - Validation set				
Fold	**Precision**	**Recall**	**Accuracy**	**F1**
1	77.63%	80.98%	84.62%	79.27%
2	**90.19%**	82.19%	86.26%	86.00%
3	86.67%	**92.84%**	**89.31%**	**89.65%**
4	80.75%	89.25%	87.79%	84.79%
5	89.24%	84.89%	85.50%	87.01%
AVG	84.90%	86.03%	86.70%	85.34%

Table 6.3 3D ResNet18 (masked input) validation performance

3D ResNet18[64, 64, 64] (masked input) - Validation set				
Fold	**Precision**	**Recall**	**Accuracy**	**F1**
1	81.49%	**92.31%**	86.92%	86.56%
2	87.48%	87.07%	86.26%	87.27%
3	88.09%	82.70%	**90.84%**	85.31%
4	84.58%	86.51%	**90.84%**	85.53%
5	**88.98%**	89.41%	88.55%	**89.19%**
AVG	86.12%	87.60%	88.68%	86.78%

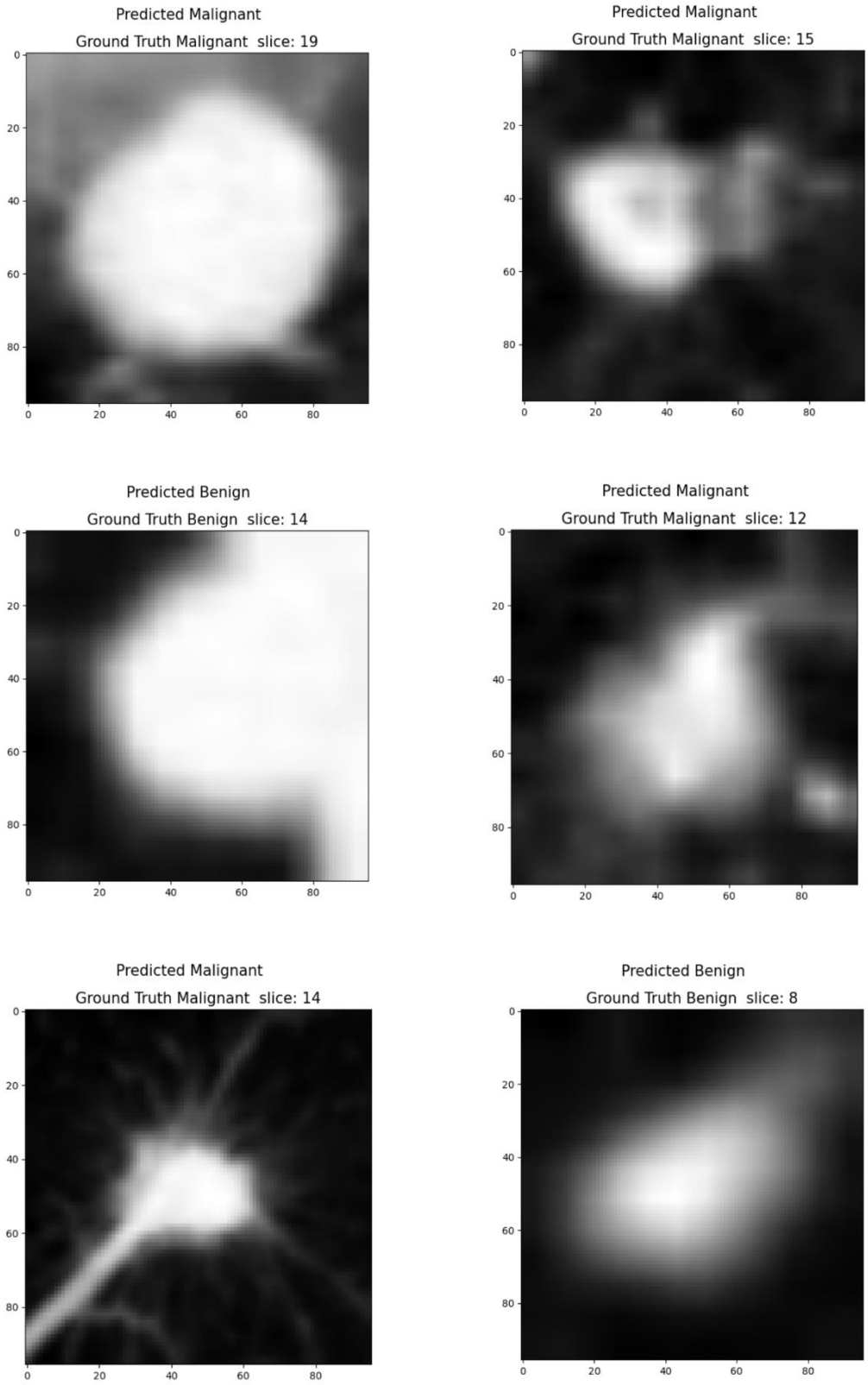


Figure 6.2.1 3D patch malignancy classification predictions

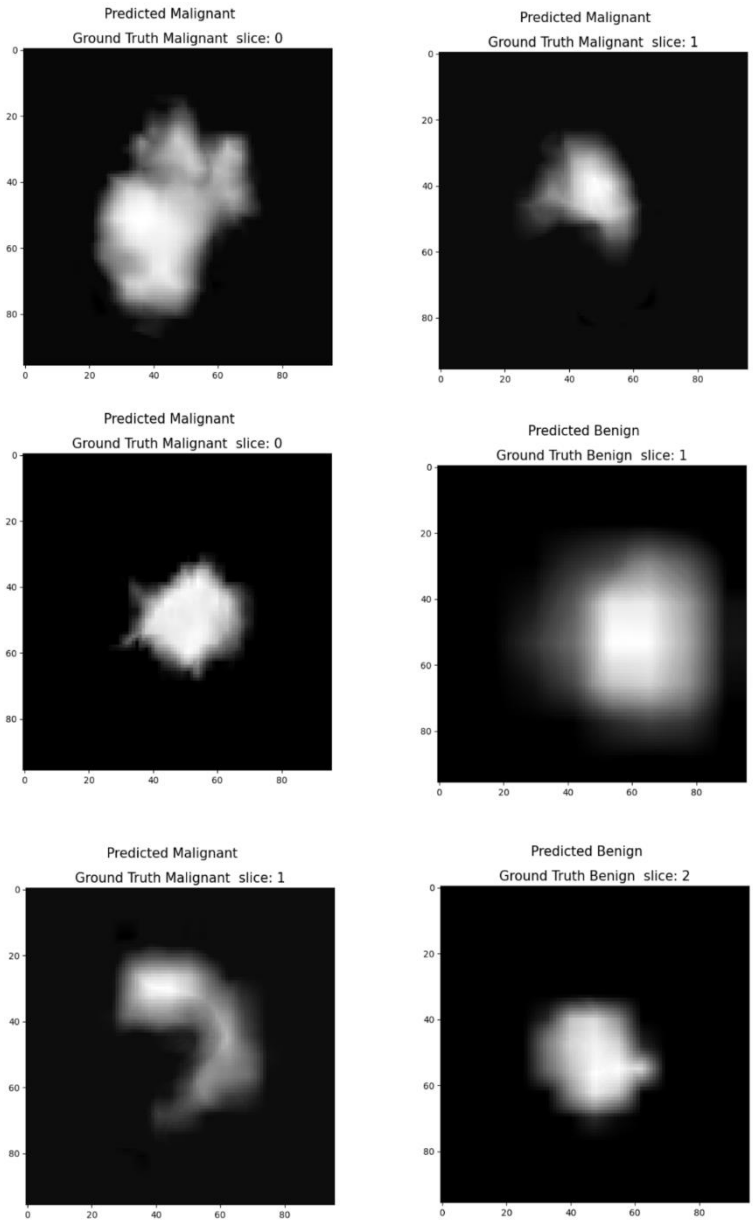


Figure 6.2.2: 3-view (masked) malignancy classification predictions

Table 6.4 3-view ResNet18 validation performance

3-view ResNet18 - Validation set (5-fold average)				
Input size	**Precision**	**Recall**	**Accuracy**	**F1**
[32, 32, 3]	**76.54%**	74.50%	**74.31%**	**75.50%**
[64, 64, 3]	74.62%	69.06%	73.94%	71.73%
[96, 96, 3]	72.42%	**76.19%**	**74.31%**	74.25%

Table 6.5 3-view ResNet18 (masked input) validation performance

3-view ResNet18(masked input) - Validation set (5-fold average)				
Input size	**Precision**	**Recall**	**Accuracy**	**F1**
[32, 32, 3]	**76.44**%	73.58%	74.49%	74.98%
[64, 64, 3]	76.08%	75.24%	**77.06**%	**75.65%**
[96, 96, 3]	72.20%	**77.09**%	70.27%	74.56%

Table 6.6 Classification testing performance

ResNet18 – Average Performance Summary - Testing set				
Experiment	**Precision**	**Recall**	**Accuracy**	**F1**
[64, 64, 64]	**83.78%**	71.41%	76.98%	**77.10%**
[64, 64, 64] (masked)	82.72%	70.51%	**77.78%**	76.12%
[32, 32, 3]	80.29%	67.92%	74.83%	70.27%
[32, 32, 3] (masked)	78.39%	71.41%	75.38%	74.73%
[64, 64, 3]	80.29%	71.35%	76.48%	75.55%
[64, 64, 3] (masked)	78.69%	69.18%	74.61%	73.62%
[96, 96, 3]	77.92%	**75.86%**	76.48%	76.87%
[96, 96, 3] (masked)	63.04%	70.28%	69.88%	66.46%

Our model's ability to precisely outline these regions holds significant promise for enhancing clinical diagnosis and treatment planning while the achieved classification accuracy underscores the potential utility of our approach in aiding radiologists to make informed decisions regarding nodule characteristics. Collectively, these results signify the successful fusion of advanced techniques in medical imaging analysis, with implications for advancing both nodule localization and characterization in clinical practice.

7 Experiment conclusions

Employing 5-fold cross-validation greatly improved the dependability of our results, adding to the credibility of the conclusions derived from our study. This method of validation, where the dataset was divided into five parts and each part was used as a test set while the others were used for training in turn, provided a more robust evaluation of our models. It helped in mitigating biases that could arise from a single train-test split and offered a more comprehensive assessment of the model's performance across different subsets of data.

Our segmentation model demonstrated performance levels that were on par with some of the current state-of-the-art 3D models. This comparison indicates that our approach, while possibly simpler or different in methodology, is effective and competitive in the field. Such findings underscore the potential of this model in practical applications and open up possibilities for their further development and refinement.

7.1 Nodule segmentation

The results we achieved were encouraging, with the model successfully segmenting the nodules in most of the scans. However, we noticed that the edge cases, where the model encountered difficulties, predominantly involved types of nodules that were underrepresented in our dataset, such as ground glass opacity (GGO), or occurred in scans that were particularly complex or the nodule size was ultra small. This observation highlights areas where our model needs further refinement to enhance its accuracy and reliability, especially in handling these less common or more challenging scenarios.

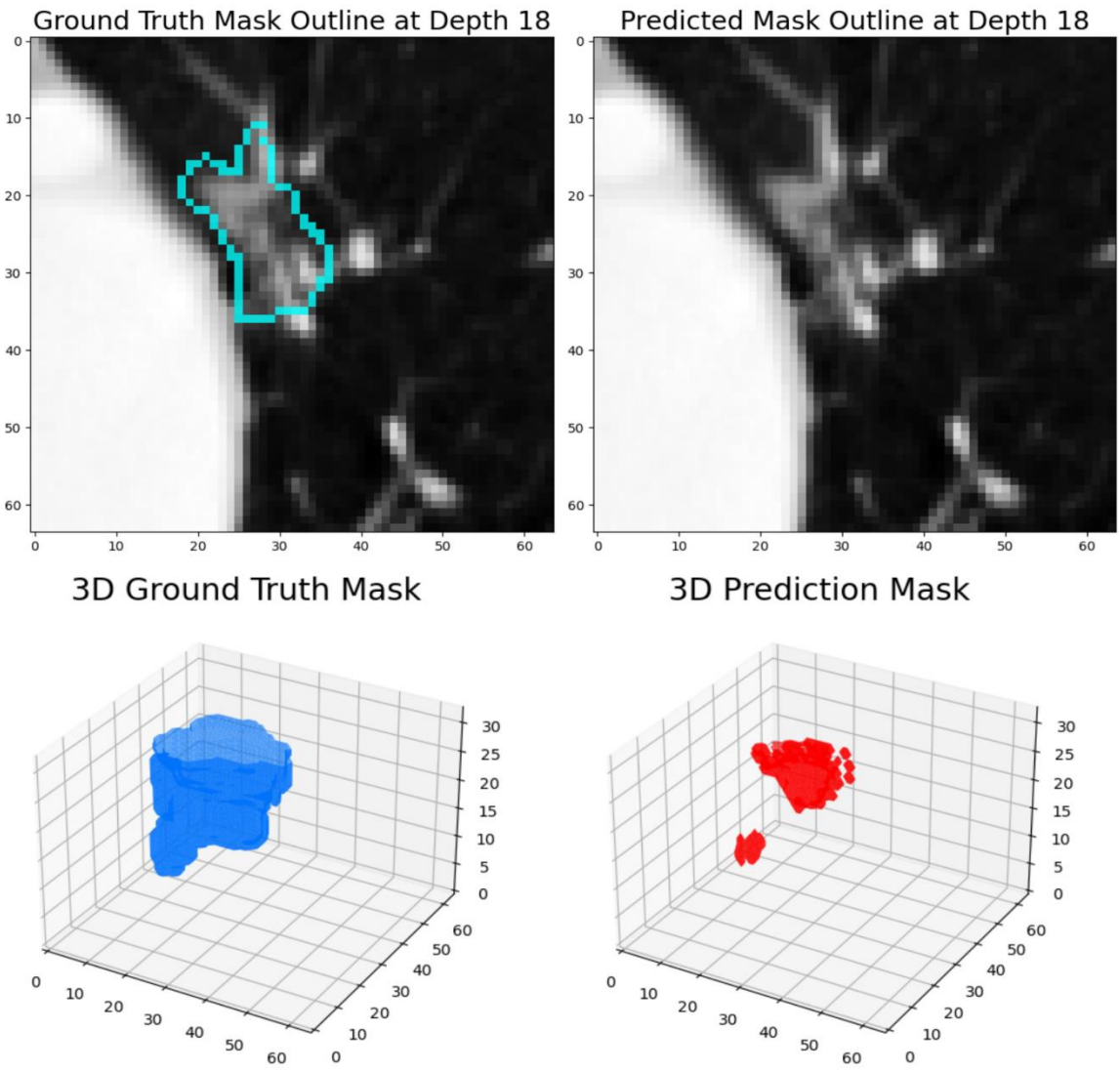


Figure 7.1.1 Segmentation underperforming cases

7.2 Nodule classification

We conducted experiments using a 3D ResNet18 classification network, designed to predict the malignancy of nodules with good accuracy. Our findings were quite positive, with the network making correct predictions in most cases. There were only a few instances of incorrect predictions or misclassifications, indicating the effectiveness of the ResNet18 architecture in this application. This suggests that with further refinement, the model could achieve even higher levels of accuracy in nodule malignancy prediction.

Given that our initial dataset was imbalanced and small in size, enhancing it with a more diverse range of data could lead to improved performance, especially in less common or 'edge case' scenarios. By incorporating a broader spectrum of data, we expect to address the limitations posed by the initial dataset's constraints, thereby refining our model's ability to handle a wider variety of cases with greater accuracy.

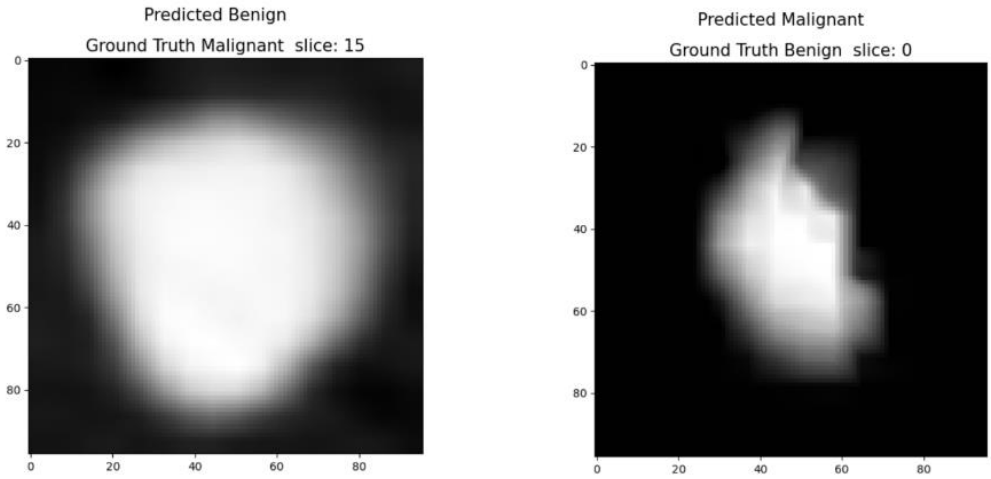


Figure 7.2.1 Classification underperforming cases

8 CADx Inference

During the inference phase of the computer-aided diagnosis (CADx) system, the primary goal is to seamlessly integrate the nodule segmentation predictions into the classification model, culminating in a comprehensive analysis of lung cancer characteristics. This critical step fuses the intricate details garnered from both the volumetric data and the nodule segmentation model, providing a holistic assessment for informed medical decision-making. Below, we outline the functionality of a comprehensive CADx system utilizing our models.

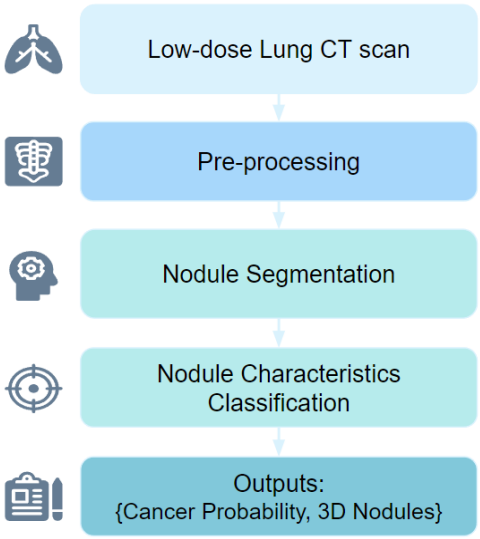


Figure 8.1 CADx pipeline

In the initial stage of the analysis, the CADx system begins by processing CT scan data and its accompanying metadata, such as pixel and slice spacings. This metadata is crucial for accurate interpretation of the scan. The system then standardizes the data by adjusting it to a consistent spacing of 1×1×1mm, which is essential for uniform analysis.

Following this, the data enters the 3D segmentation phase within the CADx system. In this phase, the volumetric image data is segmented into small, three-dimensional sections. The 3D segmentation model then methodically examines these sections, using its sophisticated neural network to identify and outline lung nodules. The accurate localization and definition of these nodules are key for the next stage of classification.

The process then advances to the classification of nodule characteristics, focusing specifically on malignancy. In our study, we concentrated on developing a malignancy classifier. However, the methodology we employed for this purpose can be adapted to train classifiers for a full range of nodule characteristics. This approach demonstrates the

versatility and potential scalability of our system in addressing various aspects of nodule analysis. By applying similar techniques, it is feasible to create classifiers that can assess other critical characteristics of nodules, thereby enhancing the comprehensiveness and accuracy of the CADx system in diagnosing and understanding lung cancer. Utilizing the information on nodule positions and outlines obtained from the segmentation phase, the system assesses each nodule's specific features such as size, shape, and texture. This analysis, powered by advanced machine learning algorithms and domain-specific knowledge, discerns various attributes of each nodule.

Ultimately, the analysis results from the nodule characteristic classifiers are synthesized. This integration of data results in a detailed and informative report on the nodules. This report, enriched by the comprehensive analysis conducted by the CADx system, provides valuable insights to medical professionals, aiding in the diagnosis and treatment of lung cancer.

In summary, the CADx system's inference process is a sophisticated journey from data standardization through segmentation to classification, culminating in a valuable diagnostic output. This intricate interplay of data processing and model analysis exemplifies the transformative impact of advanced technology on medical diagnostics, particularly in lung cancer screening.

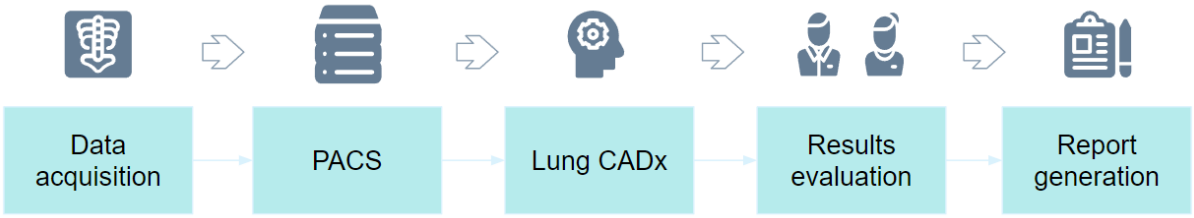


Figure 8.2 Diagnosis pipeline

9 Limitations and future work

The present study was limited by the data quality and quantity (see Discussion Chapter). In the context of future developments, we remain committed to improving our methodology. We are actively engaged in the exploration of alternative dimensions within our approach, seeking to unearth valuable insights that can propel our research forward. This exploration involves a thorough investigation into a variety of 3D architectures, each characterized by distinct capabilities and nuances. The goal of this ongoing exploration is to uncover the optimal configuration that effectively addresses our segmentation and classification tasks. Furthermore, our investigation extends to the influence of varying kernel sizes within our models. Understanding the impact of different kernel sizes is pivotal, as it directly affects the model's ability to capture features at different scales within the input data. Through systematic experimentation with varying sizes, we aspire to identify configurations that maximize the accuracy and robustness of our predictions.

An additional avenue of exploration is the integration of multiple classification heads into a unified network. This strategic consolidation not only streamlines the architecture but also enhances the model's comprehensive understanding of the data. By jointly processing diverse attributes and characteristics, we anticipate a significant enhancement in classification performance.

In summation, our forthcoming efforts will continue to encompass a multifaceted exploration involving diverse 3D architectures, kernel sizes, and the integration of classification heads for predicting all the nodule attributes. Through this methodical and forward-looking experimentation, we are dedicated to refining our approach, ultimately achieving an advanced grasp of medical images and yielding accurate segmentation and classification outcomes. This approach, underpinned by rigorous exploration, holds the promise of shaping the future direction of our research. One prospective enhancement involves the incorporation of attention mechanisms into existing models, specifically focusing on segmentation and classification tasks. Attention mechanisms can refine the model's focus on pertinent regions, thereby potentially improving accuracy and reducing false positives.

Another avenue for advancement involves the creation of nodule embeddings, with storage in a vector database, facilitating efficient retrieval and comparison. With the incorporation of a Large Vision Model (LVM) [92] in the context of lung cancer screening

using deep learning, the research trajectory gains an additional layer of complexity and potential. The introduction of LVM, akin to a Large Language Model (LLM) [93] but endowed with vision capabilities, opens up innovative possibilities for seamless integration of visual information into the analysis process. Moreover, the creation of nodule embeddings within the LVM framework, stored in a vector database, holds promise for efficient representation and retrieval of visual data. Expanding the LVM's capabilities to engage in real-time interactions with healthcare professionals, showcasing visually similar cases from historical data, could revolutionize the diagnostic decision-making process. As the research advances, the critical importance of interpretability and explainability remains, ensuring that the fusion of linguistic and visual information not only improves accuracy but also provides a transparent and understandable foundation for clinical application. Multi-modal models are the future of machine learning.

10 Discussion

In this thesis, we explored the potential of using 3D convolutions to extract rich features from our volumetric CT data. This approach proved to be much more computationally expensive than 2D approaches but our networks were able to extract and learn meaningful features from the volumetric data. Data quality and quantity are the most important factors for building a high-performance network but disagreements in labeling can confuse the network. An annotation guide developed by doctors and AI domain experts can improve the quality of the future dataset. Collecting and annotating the entire scan is a challenging and very time-consuming task. Model-assisted labeling can help with this task by pre-labeling the CT scans and the doctor will perform a quality check and correct any mislabeled annotations. This approach could help collect and annotate larger volumes of data and use them in an active learning process to improve the performance of the models. Model serving and privacy is a crucial part of a CADx system and can be explored in a future study. Data preprocessing strategy is also a critical aspect of building a robust model. With this thesis, we explored 3D patch extraction methods to generate training patches on the fly, generate synthetic samples by applying transformations, and fit the patches data to the memory. This thesis was focused on experiments using 3D convolutions to extract spatial features but more experiments can be made using different architectures and or datasets. Exploring the lung segmentation approach may improve the results. Patient info like age, gender, and smoking habits could be used as labels to develop models that assess a cancer risk score based on a patient's history and demographics. This will provide patient information along with model predictions to the doctors in an interactive medical report. This is highly important for monitoring the patient over time as follow-up scans can be easily compared and provide feedback on treatment progression.

In our literature research, we comprehensively gathered information from the datasets used, preprocessing procedures, data augmentation techniques, architectural designs, and reported performance metrics. Our analysis encompassed state-of-the-art deep learning approaches like 2D & 3D CNNs, autoencoders, and fast-growing and promising approaches, such as transformers. Furthermore, we assessed the credibility of each study by examining whether the authors presented lucid and comprehensive explanations of their methodologies and adhered to machine learning best practices.

Through these efforts, we provided a current and in-depth viewpoint on this dynamic and rapidly expanding field of study.

In this work we were limited by the constraints that the LIDC-IDRI database introduces. The dataset we used was limited in terms of the number of scans, and there was a significant imbalance in the distribution of nodule characteristics. For certain characteristics, such as texture and calcification, there were only a few examples per class. This scarcity of data for specific classes poses a challenge, as training classifiers on such a limited dataset can lead to overfitting. The models, in this case, might become overly tailored to the training data, particularly towards the classes with the most samples, resulting in a biased performance that does not generalize well to new, unseen data. Addressing this data imbalance is crucial for developing more robust and accurate models. Also, the very small sample size of some classes raised concerns about underfitting, wherein the model might fail to capture the underlying patterns due to the limited amount of training data available.

Labeling inconsistency and disagreements can be problematic while training a deep learning model. The mask labels produced by the doctors didn't always agree creating noise and confusing our model. Our solution to this was to extract the average mask. The same issues had to be faced while extracting the nodule characteristic. Furthermore, there were disagreements between some of the annotations generated by the doctors so we developed a method to extract the consensus label per characteristic.

By using a patch extraction method, we were able to efficiently extract and load 3D patches of our volumetric LDCT data. The queue approach helped utilize all the computing power of our machine by loading patches to the memory before they were requested by the data loader. This helped us manage our memory better but also dynamically create new training samples in each epoch of our experiments.

The decision not to perform lung parenchyma segmentation within our study was motivated by the specific context of juxta-pleural nodules. Given their proximity to the pleural lining, these nodules exhibit intricate relationships with the surrounding lung tissue. Segmentation of the lung parenchyma could potentially disrupt or modify these spatial relationships, impacting the accuracy of nodule analysis and subsequent classification. Our study's emphasis on accurately capturing the unique attributes of juxta-pleural nodules led us to preserve the contextual integrity of their proximity to the pleural boundary. By avoiding lung parenchyma segmentation, we aimed to ensure that our analysis provided a precise

representation of these nodules, acknowledging the potential implications that parenchyma segmentation could introduce to the analysis of juxta-pleural nodules. This strategic decision aligns with our overarching goal of offering detailed insights into nodule characteristics within their anatomical context.

We experimented with 3D versions of the popular architecture ResNet to extract contextual information from the LDCT images. This experiment requires significantly more computational power and memory but also more efficient data loading and management processes and adding more layers would increase these requirements exponentially. Our findings show that our models perform on par with other leading-edge studies in the field. With additional fine-tuning of our model's parameters, we have the potential to enhance our performance even further. This suggests that our approach not only holds its ground against current advanced systems but also has significant room for optimization and improvement.

The inference stage of the computer-aided diagnosis (CADx) system effectively combines nodule segmentation with a classification model to provide a thorough assessment of lung nodule malignancy. This process includes standardizing data spacing, breaking down volumetric data into 3D segments, and using a 3D segmentation model to detect nodules. The system then uses the identified nodule locations and outlines to classify nodule characteristics, examining aspects such as size, shape, texture, and surrounding features. The combined results offer a detailed view of each nodule, enhancing medical reports and providing doctors with crucial information for lung cancer diagnosis and treatment. The complex interaction between data and models in the CADx system highlights its potential to revolutionize medical diagnostics, particularly in lung cancer screening.

Importantly, the development of operational CADx systems that gather data and refine models over time is essential. These systems are not just vital tools for medical professionals; they also represent an opportunity for continual improvement. By consistently collecting data and progressively refining their models, these operational systems are designed to evolve, adapt, and steadily improve their diagnostic accuracy. This ongoing cycle of data gathering, model enhancement, and practical application could lead to more precise and reliable CADx systems, significantly benefiting patient care and advancing the medical diagnostics field.

The importance of interpretability and explainability in medical applications, particularly in the context of deep learning models for lung cancer screening, cannot be overstated. In healthcare, where decisions directly impact patient outcomes, understanding

the reasoning behind an algorithm's predictions is crucial for gaining the trust of healthcare professionals and ensuring the responsible integration of artificial intelligence into clinical practice. Interpretability refers to the ability to understand the internal mechanisms and decision-making processes of a model, while explainability involves providing clear and understandable reasons for a model's predictions. In the realm of medical applications, these aspects are paramount for several reasons. Firstly, interpretability and explainability contribute to the validation and assessment of the reliability of deep learning models. Healthcare professionals need to be confident in the accuracy and relevance of the predictions made by these models, especially when dealing with critical decisions such as disease diagnosis and treatment planning. Secondly, these aspects are essential for fostering collaboration between machine learning systems and human practitioners. Clinicians must comprehend how a model arrives at a particular diagnosis or recommendation to make informed decisions and provide the best possible care to their patients. This collaborative approach, where AI systems augment human expertise, can lead to more accurate and timely diagnosis.

Additionally, in the context of regulatory compliance and ethical considerations, interpretability and explainability are crucial. Regulatory bodies often require transparency in the functioning of medical AI systems, and patients have the right to understand the basis of the decisions affecting their health. Transparent models not only facilitate regulatory approval but also contribute to building public trust in the application of AI in healthcare. In the specific domain of lung cancer screening, where the consequences of a misdiagnosis can be severe, having interpretable and explainable models ensures that clinicians can confidently rely on AI systems as valuable tools, ultimately improving patient outcomes. Therefore, as the field progresses, the emphasis on interpretability and explainability remains integral to the responsible deployment of deep learning models in medical application.

References

[1] R. L. Siegel, K. D. Miller, and A. Jemal, “Cancer statistics, 2020.,” *CA Cancer J Clin*, vol. 70, no. 1, pp. 7–30, Jan. 2020, doi: 10.3322/caac.21590.

[2] The American Cancer Society medical and editorial content team, “Key Statistics for Lung Cancer,” American Cancer Society. Accessed: Sep. 21, 2023. [Online]. Available: https://www.cancer.org/cancer/types/lung-cancer/about/key-statistics.html

[3] Ferlay J, Ervik M, Lam F, Colombet M, Mery L, and et al. Piñeros M, “Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer.” Accessed: Sep. 21, 2023. [Online]. Available: https://gco.iarc.fr/today

[4] World Health Organization, “World Health Organization, ‘Cancer,’ World Health Organization, [Online],” World Health Organization. Accessed: Sep. 21, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cancer

[5] The American Cancer Society medical and editorial content team, “Lung Cancer Early Detection, Diagnosis, and Staging,” The American Cancer Society. Accessed: Sep. 21, 2023. [Online]. Available: https://www.cancer.org/content/dam/CRC/PDF/Public/8705.00.pdf

[6] P. M. Ellis and R. Vandermeer, “Delays in the diagnosis of lung cancer.,” *J Thorac Dis*, vol. 3, no. 3, pp. 183–8, Sep. 2011, doi: 10.3978/j.issn.2072-1439.2011.01.01.

[7] Johns Hopkins Medicine, “‘Lung Biopsy,’ Johns Hopkins Medicine, [Online].” Accessed: Sep. 21, 2023. [Online]. Available: https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/lung-biopsy

[8] C. Rampinelli, D. Origgi, and M. Bellomi, “Low-dose CT: technique, reading methods and image interpretation.,” *Cancer Imaging*, vol. 12, no. 3, pp. 548–56, Feb. 2013, doi: 10.1102/1470-7330.2012.0049.

[9] E. Tylski and M. Goyal, “Low Dose CT for Lung Cancer Screening: The Background, the Guidelines, and a Tailored Approach to Patient Care.,” *Mo Med*, vol. 116, no. 5, pp. 414–419, 2019.

[10] W. Hendrix *et al.*, “Deep learning for the detection of benign and malignant pulmonary nodules in non-screening chest CT scans,” *Communications Medicine*, vol. 3, no. 1, p. 156, Oct. 2023, doi: 10.1038/s43856-023-00388-5.

[11] M. Firmino, G. Angelo, H. Morais, M. R. Dantas, and R. Valentim, “Computer-aided detection (CADe) and diagnosis (CADx) system for lung cancer with likelihood of malignancy.,” *Biomed Eng Online*, vol. 15, no. 1, p. 2, Jan. 2016, doi: 10.1186/s12938-015-0120-7.

[12] J. Kim and K. H. Kim, “Role of chest radiographs in early lung cancer detection,” *Transl Lung Cancer Res*, vol. 9, no. 3, pp. 522–531, Jun. 2020, doi: 10.21037/tlcr.2020.04.02.

[13] Y. Weerakkody and J. Jones, “Pulmonary nodule,” in *Radiopaedia.org*, Radiopaedia.org, 2010. doi: 10.53347/rID-10187.

[14] B. Botz and M. Morgan, “Lung-RADS,” in *Radiopaedia.org*, Radiopaedia.org, 2014. doi: 10.53347/rID-32681.

[15] A. W. Harzing, “Publish or Perish.” 2007. Accessed: Oct. 11, 2023. [Online]. Available: https://harzing.com/resources/publish-or-perish

[16] S. G. Armato *et al.*, “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. 915–31, Feb. 2011, doi: 10.1118/1.3528204.

[17] A. A. A. Setio *et al.*, “Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the LUNA16 challenge,” Dec. 2016, doi: 10.1016/j.media.2017.06.015.
[18] ELCAP and VIA research groups, “ELCAP Public Lung Image Database.” Accessed: Oct. 18, 2023. [Online]. Available: https://www.via.cornell.edu/databases/lungdb.html
[19] S. G. Armato *et al.*, “Guest Editorial: LUNGx Challenge for computerized lung nodule classification: reflections and lessons learned,” *Journal of Medical Imaging*, vol. 2, no. 2, p. 020103, Jun. 2015, doi: 10.1117/1.JMI.2.2.020103.
[20] S. G. Armato *et al.*, “LUNGx Challenge for computerized lung nodule classification,” *Journal of Medical Imaging*, vol. 3, no. 4, p. 044506, Dec. 2016, doi: 10.1117/1.JMI.3.4.044506.
[21] G. Singadkar, A. Mahajan, M. Thakur, and S. Talbar, “Deep Deconvolutional Residual Network Based Automatic Lung Nodule Segmentation,” *J Digit Imaging*, vol. 33, no. 3, pp. 678–684, Jun. 2020, doi: 10.1007/s10278-019-00301-4.
[22] S. Wang *et al.*, “Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation,” *Med Image Anal*, vol. 40, pp. 172–183, Aug. 2017, doi: 10.1016/j.media.2017.06.014.
[23] R. Roy, T. Chakraborti, and A. S. Chowdhury, “A deep learning-shape driven level set synergism for pulmonary nodule segmentation,” *Pattern Recognit Lett*, vol. 123, pp. 31–38, May 2019, doi: 10.1016/j.patrec.2019.03.004.
[24] M. A. Khan *et al.*, “VGG19 Network Assisted Joint Segmentation and Classification of Lung Nodules in CT Images,” *Diagnostics*, vol. 11, no. 12, p. 2208, Nov. 2021, doi: 10.3390/diagnostics11122208.
[25] S. Kido *et al.*, “Segmentation of Lung Nodules on CT Images Using a Nested Three-Dimensional Fully Connected Convolutional Network,” *Front Artif Intell*, vol. 5, Feb. 2022, doi: 10.3389/frai.2022.782225.
[26] H. Tang, C. Zhang, and X. Xie, “NoduleNet: Decoupled False Positive Reduction for Pulmonary Nodule Detection and Segmentation,” 2019, pp. 266–274. doi: 10.1007/978-3-030-32226-7_30.
[27] P. Dutande, U. Baid, and S. Talbar, “LNCDS: A 2D-3D cascaded CNN approach for lung nodule classification, detection and segmentation,” *Biomed Signal Process Control*, vol. 67, p. 102527, May 2021, doi: 10.1016/j.bspc.2021.102527.
[28] M. Usman, B.-D. Lee, S.-S. Byon, S.-H. Kim, B. Lee, and Y.-G. Shin, “Volumetric lung nodule segmentation using adaptive ROI with multi-view residual learning,” *Sci Rep*, vol. 10, no. 1, p. 12839, Jul. 2020, doi: 10.1038/s41598-020-69817-y.
[29] G. Pezzano, V. Ribas Ripoll, and P. Radeva, “CoLe-CNN: Context-learning convolutional neural network with adaptive loss function for lung nodule segmentation,” *Comput Methods Programs Biomed*, vol. 198, p. 105792, Jan. 2021, doi: 10.1016/j.cmpb.2020.105792.
[30] “Kolmogorov–Smirnov Test,” in *Encyclopedia of Research Design*, 2455 Teller Road, Thousand Oaks California 91320 United States : SAGE Publications, Inc., 2010. doi: 10.4135/9781412961288.n204.
[31] M. H. Hesamian, W. Jia, X. He, Q. Wang, and P. J. Kennedy, “Synthetic CT images for semi-sequential detection and segmentation of lung nodules,” *Applied Intelligence*, vol. 51, no. 3, pp. 1616–1628, Mar. 2021, doi: 10.1007/s10489-020-01914-x.
[32] Z. Wu, X. Li, and J. Zuo, “RAD-UNet: Research on an improved lung nodule semantic segmentation algorithm based on deep learning,” *Front Oncol*, vol. 13, Mar. 2023, doi: 10.3389/fonc.2023.1084096.

[33] J. Hou, C. Yan, R. Li, Q. Huang, X. Fan, and F. Lin, “Lung Nodule Segmentation Algorithm With SMR-UNet,” *IEEE Access*, vol. 11, pp. 34319–34331, 2023, doi: 10.1109/ACCESS.2023.3264789.
[34] S. Dodia, A. Basava, and M. Padukudru Anand, “A novel receptive field- regularized <scp>V- net</scp> and nodule classification network for lung nodule detection,” *Int J Imaging Syst Technol*, vol. 32, no. 1, pp. 88–101, Jan. 2022, doi: 10.1002/ima.22636.
[35] Z. Wu, Q. Zhou, and F. Wang, “Coarse-to-Fine Lung Nodule Segmentation in CT Images With Image Enhancement and Dual-Branch Network,” *IEEE Access*, vol. 9, pp. 7255–7262, 2021, doi: 10.1109/ACCESS.2021.3049379.
[36] X. Zhang *et al.*, “Accurate segmentation for different types of lung nodules on CT images using improved U-Net convolutional network,” *Medicine*, vol. 100, no. 40, p. e27491, Oct. 2021, doi: 10.1097/MD.0000000000027491.
[37] P. Ardimento, L. Aversano, M. L. Bernardi, M. Cimitile, M. Iammarino, and C. Verdone, “Evo-GUNet3++: Using evolutionary algorithms to train UNet-based architectures for efficient 3D lung cancer detection,” *Appl Soft Comput*, vol. 144, p. 110465, Sep. 2023, doi: 10.1016/j.asoc.2023.110465.
[38] S. F. Banu, Md. M. K. Sarker, M. Abdel-Nasser, D. Puig, and H. A. Raswan, “AWEU-Net: An Attention-Aware Weight Excitation U-Net for Lung Nodule Segmentation,” *Applied Sciences*, vol. 11, no. 21, p. 10132, Oct. 2021, doi: 10.3390/app112110132.
[39] H. Liu *et al.*, “A cascaded dual-pathway residual network for lung nodule segmentation in CT images,” *Physica Medica*, vol. 63, pp. 112–121, Jul. 2019, doi: 10.1016/j.ejmp.2019.06.003.
[40] H. Cao *et al.*, “Dual-branch residual network for lung nodule segmentation,” *Appl Soft Comput*, vol. 86, p. 105934, Jan. 2020, doi: 10.1016/j.asoc.2019.105934.
[41] H. Yu, J. Li, L. Zhang, Y. Cao, X. Yu, and J. Sun, “Design of lung nodules segmentation and recognition algorithm based on deep learning,” *BMC Bioinformatics*, vol. 22, no. S5, p. 314, Nov. 2021, doi: 10.1186/s12859-021-04234-0.
[42] S. Tyagi and S. N. Talbar, “CSE-GAN: A 3D conditional generative adversarial network with concurrent squeeze-and-excitation blocks for lung nodule segmentation,” *Comput Biol Med*, vol. 147, p. 105781, Aug. 2022, doi: 10.1016/j.compbiomed.2022.105781.
[43] X. Li, A. Jiang, Y. Qiu, M. Li, X. Zhang, and S. Yan, “TPFR-Net: U-shaped model for lung nodule segmentation based on transformer pooling and dual-attention feature reorganization,” *Med Biol Eng Comput*, vol. 61, no. 8, pp. 1929–1946, Aug. 2023, doi: 10.1007/s11517-023-02852-9.
[44] S. Wang, A. Jiang, X. Li, Y. Qiu, M. Li, and F. Li, “DPBET: A dual-path lung nodules segmentation model based on boundary enhancement and hybrid transformer,” *Comput Biol Med*, vol. 151, p. 106330, Dec. 2022, doi: 10.1016/j.compbiomed.2022.106330.
[45] M. Usman and Y.-G. Shin, “DEHA-Net: A Dual-Encoder-Based Hard Attention Network with an Adaptive ROI Mechanism for Lung Nodule Segmentation,” *Sensors*, vol. 23, no. 4, p. 1989, Feb. 2023, doi: 10.3390/s23041989.
[46] J. Qiu, B. Li, R. Liao, H. Mo, and L. Tian, “A dual-task region-boundary aware neural network for accurate pulmonary nodule segmentation,” *J Vis Commun Image Represent*, vol. 96, p. 103909, Oct. 2023, doi: 10.1016/j.jvcir.2023.103909.

[47] S. Luo *et al.*, “DAS-Net: A lung nodule segmentation method based on adaptive dual-branch attention and shadow mapping,” *Applied Intelligence*, vol. 52, no. 13, pp. 15617–15631, Oct. 2022, doi: 10.1007/s10489-021-03038-2.
[48] G. L. F. da Silva, T. L. A. Valente, A. C. Silva, A. C. de Paiva, and M. Gattass, “Convolutional neural network-based PSO for lung nodule false positive reduction on CT images,” *Comput Methods Programs Biomed*, vol. 162, pp. 109–118, Aug. 2018, doi: 10.1016/j.cmpb.2018.05.006.
[49] A. Bhandary *et al.*, “Deep-learning framework to detect lung abnormality – A study with chest X-Ray and lung CT scan images,” *Pattern Recognit Lett*, vol. 129, pp. 271–278, Jan. 2020, doi: 10.1016/j.patrec.2019.11.013.
[50] G. S. Tran, T. P. Nghiem, V. T. Nguyen, C. M. Luong, and J.-C. Burie, “Improving Accuracy of Lung Nodule Classification Using Deep Learning with Focal Loss,” *J Healthc Eng*, vol. 2019, pp. 1–9, Feb. 2019, doi: 10.1155/2019/5156416.
[51] M. Al-Shabi, H. K. Lee, and M. Tan, “Gated-Dilated Networks for Lung Nodule Classification in CT Scans,” *IEEE Access*, vol. 7, pp. 178827–178838, 2019, doi: 10.1109/ACCESS.2019.2958663.
[52] S. Suresh and S. Mohan, “ROI-based feature learning for efficient true positive prediction using convolutional neural network for lung cancer diagnosis,” *Neural Comput Appl*, vol. 32, no. 20, pp. 15989–16009, Oct. 2020, doi: 10.1007/s00521-020-04787-w.
[53] X. Huang, Q. Lei, T. Xie, Y. Zhang, Z. Hu, and Q. Zhou, “Deep Transfer Convolutional Neural Network and Extreme Learning Machine for lung nodule diagnosis on CT images,” *Knowl Based Syst*, vol. 204, p. 106230, Sep. 2020, doi: 10.1016/j.knosys.2020.106230.
[54] X. Zhao *et al.*, “Deep CNN models for pulmonary nodule classification: Model modification, model integration, and transfer learning,” *J Xray Sci Technol*, vol. 27, no. 4, pp. 615–629, Sep. 2019, doi: 10.3233/XST-180490.
[55] I. Ali, M. Muzammil, I. U. Haq, A. A. Khaliq, and S. Abdullah, “Efficient Lung Nodule Classification Using Transferable Texture Convolutional Neural Network,” *IEEE Access*, vol. 8, pp. 175859–175870, 2020, doi: 10.1109/ACCESS.2020.3026080.
[56] A. Naik, D. R. Edla, and V. Kuppili, “A combination of FractalNet and CNN for Lung Nodule Classification,” in *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, IEEE, Jul. 2020, pp. 1–7. doi: 10.1109/ICCCNT49239.2020.9225365.
[57] J. Qiao, Y. Fan, M. Zhang, K. Fang, D. Li, and Z. Wang, “Ensemble framework based on attributes and deep features for benign-malignant classification of lung nodule,” *Biomed Signal Process Control*, vol. 79, p. 104217, Jan. 2023, doi: 10.1016/j.bspc.2022.104217.
[58] S. Suresh and S. Mohan, “NROI based feature learning for automated tumor stage classification of pulmonary lung nodules using deep convolutional neural networks,” *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 5, pp. 1706–1717, May 2022, doi: 10.1016/j.jksuci.2019.11.013.
[59] D. Liu, F. Liu, Y. Tie, L. Qi, and F. Wang, “Res-trans networks for lung nodule classification,” *Int J Comput Assist Radiol Surg*, vol. 17, no. 6, pp. 1059–1068, Jun. 2022, doi: 10.1007/s11548-022-02576-5.
[60] Y. Chen, Y. Wang, F. Hu, L. Feng, T. Zhou, and C. Zheng, “LDNNET: Towards Robust Classification of Lung Nodule and Cancer Using Lung Dense Neural Network,” *IEEE Access*, vol. 9, pp. 50301–50320, 2021, doi: 10.1109/ACCESS.2021.3068896.

[61] R. Mastouri, N. Khlifa, H. Neji, and S. Hantous-Zannad, “A bilinear convolutional neural network for lung nodules classification on CT images,” *Int J Comput Assist Radiol Surg*, vol. 16, no. 1, pp. 91–101, Jan. 2021, doi: 10.1007/s11548-020-02283-z.
[62] I. D. Apostolopoulos *et al.*, “Automatic classification of solitary pulmonary nodules in PET/CT imaging employing transfer learning techniques,” *Med Biol Eng Comput*, vol. 59, no. 6, pp. 1299–1310, Jun. 2021, doi: 10.1007/s11517-021-02378-y.
[63] A. Nibali, Z. He, and D. Wollersheim, “Pulmonary nodule classification with deep residual networks,” *Int J Comput Assist Radiol Surg*, vol. 12, no. 10, pp. 1799–1808, Oct. 2017, doi: 10.1007/s11548-017-1605-6.
[64] Y. Xie *et al.*, “Knowledge-based Collaborative Deep Learning for Benign-Malignant Lung Nodule Classification on Chest CT,” *IEEE Trans Med Imaging*, vol. 38, no. 4, pp. 991–1004, Apr. 2019, doi: 10.1109/TMI.2018.2876510.
[65] G. Kang, K. Liu, B. Hou, and N. Zhang, “3D multi-view convolutional neural networks for lung nodule classification,” *PLoS One*, vol. 12, no. 11, p. e0188290, Nov. 2017, doi: 10.1371/journal.pone.0188290.
[66] Y. Xie, Y. Xia, J. Zhang, D. D. Feng, M. Fulham, and W. Cai, “Transferable Multi-model Ensemble for Benign-Malignant Lung Nodule Classification on Chest CT,” 2017, pp. 656–664. doi: 10.1007/978-3-319-66179-7_75.
[67] K. V. Venkadesh *et al.*, “Deep Learning for Malignancy Risk Estimation of Pulmonary Nodules Detected at Low-Dose Screening CT,” *Radiology*, vol. 300, no. 2, pp. 438–447, Aug. 2021, doi: 10.1148/radiol.2021204433.
[68] M. M. Naeem Abid, T. Zia, M. Ghafoor, and D. Windridge, “Multi-view Convolutional Recurrent Neural Networks for Lung Cancer Nodule Identification,” *Neurocomputing*, vol. 453, pp. 299–311, Sep. 2021, doi: 10.1016/j.neucom.2020.06.144.
[69] J. L. Causey *et al.*, “Highly accurate model for prediction of lung nodule malignancy with CT scans,” *Sci Rep*, vol. 8, no. 1, p. 9286, Jun. 2018, doi: 10.1038/s41598-018-27569-w.
[70] R. Dey, Z. Lu, and Y. Hong, “Diagnostic classification of lung nodules using 3D neural networks,” in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, IEEE, Apr. 2018, pp. 774–778. doi: 10.1109/ISBI.2018.8363687.
[71] X. Xu *et al.*, “MSCS-DeepLN: Evaluating lung nodule malignancy using multi-scale cost-sensitive neural networks,” *Med Image Anal*, vol. 65, p. 101772, Oct. 2020, doi: 10.1016/j.media.2020.101772.
[72] Y. Ren *et al.*, “A manifold learning regularization approach to enhance 3D CT image-based lung nodule classification,” *Int J Comput Assist Radiol Surg*, vol. 15, no. 2, pp. 287–295, Feb. 2020, doi: 10.1007/s11548-019-02097-8.
[73] R. Wu, C. Liang, Y. Li, X. Shi, J. Zhang, and H. Huang, “Self-supervised transfer learning framework driven by visual attention for benign–malignant lung nodule classification on chest CT,” *Expert Syst Appl*, vol. 215, p. 119339, Apr. 2023, doi: 10.1016/j.eswa.2022.119339.
[74] M. Al-Shabi, K. Shak, and M. Tan, “ProCAN: Progressive growing channel attentive non-local network for lung nodule classification,” *Pattern Recognit*, vol. 122, p. 108309, Feb. 2022, doi: 10.1016/j.patcog.2021.108309.
[75] Q. Zhu, Y. Wang, X. Chu, X. Yang, and W. Zhong, “Multi-View Coupled Self-Attention Network for Pulmonary Nodules Classification,” 2022. [Online]. Available: https://github.com/ahukui/MVCS.
[76] E. S. Neal Joshua, D. Bhattacharyya, M. Chakkravarthy, and Y.-C. Byun, “3D CNN with Visual Insights for Early Detection of Lung Cancer Using Gradient-Weighted

Class Activation," *J Healthc Eng*, vol. 2021, pp. 1–11, Mar. 2021, doi: 10.1155/2021/6695518.
[77] K. Mehta, A. Jain, J. Mangalagiri, S. Menon, P. Nguyen, and D. R. Chapman, "Lung Nodule Classification Using Biomarkers, Volumetric Radiomics, and 3D CNNs," *J Digit Imaging*, vol. 34, no. 3, pp. 647–666, Jun. 2021, doi: 10.1007/s10278-020-00417-y.
[78] H. Jung, B. Kim, I. Lee, J. Lee, and J. Kang, "Classification of lung nodules in CT scans using three-dimensional deep convolutional neural networks with a checkpoint ensemble method," *BMC Med Imaging*, vol. 18, no. 1, p. 48, Dec. 2018, doi: 10.1186/s12880-018-0286-0.
[79] F. Silva *et al.*, "Pre-Training Autoencoder for Lung Nodule Malignancy Assessment Using CT Images," *Applied Sciences*, vol. 10, no. 21, p. 7837, Nov. 2020, doi: 10.3390/app10217837.
[80] M. Nemoto, K. Ushifusa, Y. Kimura, T. Nagaoka, T. Yamada, and T. Yoshikawa, "Unsupervised Feature Extraction for Various Computer-Aided Diagnosis Using Multiple Convolutional Autoencoders and 2.5-Dimensional Local Image Analysis," *Applied Sciences*, vol. 13, no. 14, p. 8330, Jul. 2023, doi: 10.3390/app13148330.
[81] D. Dai, Y. Sun, C. Dong, Q. Yan, Z. Li, and S. Xu, "Effectively fusing clinical knowledge and AI knowledge for reliable lung nodule diagnosis," *Expert Syst Appl*, vol. 230, p. 120634, Nov. 2023, doi: 10.1016/j.eswa.2023.120634.
[82] P. Zhai, Y. Tao, H. Chen, T. Cai, and J. Li, "Multi-Task Learning for Lung Nodule Classification on Chest CT," *IEEE Access*, vol. 8, pp. 180317–180327, 2020, doi: 10.1109/ACCESS.2020.3027812.
[83] M. B. Zia, Z. J. Juan, N. Xiao, J. Wang, A. Khan, and X. Zhou, "Classification of malignant and benign lung nodule and prediction of image label class using multi-deep model," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 3, pp. 35–41, 2020, doi: 10.14569/ijacsa.2020.0110305.
[84] X. Fu, L. Bi, A. Kumar, M. Fulham, and J. Kim, "An attention-enhanced cross-task network to analyse lung nodule attributes in CT images," *Pattern Recognit*, vol. 126, p. 108576, Jun. 2022, doi: 10.1016/j.patcog.2022.108576.
[85] K. Wu, B. Peng, and D. Zhai, "Multi-Granularity Dilated Transformer for Lung Nodule Classification via Local Focus Scheme," *Applied Sciences*, vol. 13, no. 1, p. 377, Dec. 2022, doi: 10.3390/app13010377.
[86] P. Afshar, F. Naderkhani, A. Oikonomou, M. J. Rafiee, A. Mohammadi, and K. N. Plataniotis, "MIXCAPS: A capsule network-based mixture of experts for lung nodule malignancy prediction," *Pattern Recognit*, vol. 116, p. 107942, Aug. 2021, doi: 10.1016/j.patcog.2021.107942.
[87] I. D. Apostolopoulos, N. D. Papathanasiou, and G. S. Panayiotakis, "Classification of lung nodule malignancy in computed tomography imaging utilising generative adversarial networks and semi-supervised transfer learning," *Biocybern Biomed Eng*, vol. 41, no. 4, pp. 1243–1257, Oct. 2021, doi: 10.1016/j.bbe.2021.08.006.
[88] H. Jiang, F. Shen, F. Gao, and W. Han, "Learning efficient, explainable and discriminative representations for pulmonary nodules classification," *Pattern Recognit*, vol. 113, p. 107825, May 2021, doi: 10.1016/j.patcog.2021.107825.
[89] K. Xia, J. Chi, Y. Gao, Y. Jiang, and C. Wu, "Adaptive Aggregated Attention Network for Pulmonary Nodule Classification," *Applied Sciences*, vol. 11, no. 2, p. 610, Jan. 2021, doi: 10.3390/app11020610.
[90] M. Al-Shabi, K. Shak, and M. Tan, "3D axial-attention for lung nodule classification," *Int J Comput Assist Radiol Surg*, vol. 16, no. 8, pp. 1319–1324, Aug. 2021, doi: 10.1007/s11548-021-02415-z.

[91] G. Turinici, “The convergence of the Stochastic Gradient Descent (SGD) : a self-contained proof,” Mar. 2021, doi: 10.5281/zenodo.4638694.
[92] J. Wang *et al.*, “Review of Large Vision Models and Visual Prompt Engineering,” Jul. 2023.
[93] Z. Wan *et al.*, “Efficient Large Language Models: A Survey,” Dec. 2023.
[94] Johns Hopkins Medicine, “‘Lung Biopsy,’ Johns Hopkins Medicine, [Online].” Accessed: Sep. 21, 2023. [Online]. Available: https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/lung-biopsy
[95] N. Mahmoud, R. Vashisht, D. K. Sanghavi, and S. Kalanjeri, *Bronchoscopy*. 2023.
[96] D. F. Sigmon and S. Fatima, *Fine Needle Aspiration*. 2023.
[97] M. Mehrotra, J. R. D’Cruz, and M. E. Arthur, *Video-Assisted Thoracoscopy*. 2023.
[98] P. A. McNally and M. E. Arthur, *Mediastinoscopy*. 2023.
[99] J. Kim and K. H. Kim, “Role of chest radiographs in early lung cancer detection,” *Transl Lung Cancer Res*, vol. 9, no. 3, pp. 522–531, Jun. 2020, doi: 10.21037/tlcr.2020.04.02.
[100] M. H. Winkler, H. R. Touw, P. M. van de Ven, J. Twisk, and P. R. Tuinman, “Diagnostic Accuracy of Chest Radiograph, and When Concomitantly Studied Lung Ultrasound, in Critically Ill Patients With Respiratory Symptoms: A Systematic Review and Meta-Analysis,” *Crit Care Med*, vol. 46, no. 7, pp. e707–e714, Jul. 2018, doi: 10.1097/CCM.0000000000003129.
[101] E. Tylski and M. Goyal, “Low Dose CT for Lung Cancer Screening: The Background, the Guidelines, and a Tailored Approach to Patient Care.,” *Mo Med*, vol. 116, no. 5, pp. 414–419, 2019.
[102] M. Vonder, M. D. Dorrius, and R. Vliegenthart, “Latest CT technologies in lung cancer screening: protocols and radiation dose reduction.,” *Transl Lung Cancer Res*, vol. 10, no. 2, pp. 1154–1164, Feb. 2021, doi: 10.21037/tlcr-20-808.
[103] K. H. Rubin *et al.*, “Developing and Validating a Lung Cancer Risk Prediction Model: A Nationwide Population-Based Study,” *Cancers (Basel)*, vol. 15, no. 2, p. 487, Jan. 2023, doi: 10.3390/cancers15020487.
[104] K.-H. Yu *et al.*, “Reproducible Machine Learning Methods for Lung Cancer Detection Using Computed Tomography Images: Algorithm Development and Validation.,” *J Med Internet Res*, vol. 22, no. 8, p. e16709, Aug. 2020, doi: 10.2196/16709.
[105] M. Cellina *et al.*, “Artificial Intelligence in Lung Cancer Screening: The Future Is Now,” *Cancers (Basel)*, vol. 15, no. 17, p. 4344, Aug. 2023, doi: 10.3390/cancers15174344.
[106] J. Zhang, Y. Xia, H. Cui, and Y. Zhang, “Pulmonary nodule detection in medical images: A survey,” *Biomed Signal Process Control*, vol. 43, pp. 138–147, May 2018, doi: 10.1016/j.bspc.2018.01.011.
[107] Y. Gu *et al.*, “A survey of computer-aided diagnosis of lung nodules from CT scans using deep learning,” *Comput Biol Med*, vol. 137, p. 104806, Oct. 2021, doi: 10.1016/j.compbiomed.2021.104806.
[108] M. A. Thanoon, M. A. Zulkifley, M. A. A. Mohd Zainuri, and S. R. Abdani, “A Review of Deep Learning Techniques for Lung Cancer Screening and Diagnosis Based on CT Images,” *Diagnostics*, vol. 13, no. 16, p. 2617, Aug. 2023, doi: 10.3390/diagnostics13162617.
[109] G. Zhang *et al.*, “Automatic nodule detection for lung cancer in CT images: A review,” *Comput Biol Med*, vol. 103, pp. 287–300, Dec. 2018, doi: 10.1016/j.compbiomed.2018.10.033.
[110] B. Liu *et al.*, “Evolving the pulmonary nodules diagnosis from classical approaches to deep learning-aided decision support: three decades’ development course and

future prospect," *J Cancer Res Clin Oncol*, vol. 146, no. 1, pp. 153–185, Jan. 2020, doi: 10.1007/s00432-019-03098-5.
[111] A. Halder, D. Dey, and A. K. Sadhu, "Lung Nodule Detection from Feature Engineering to Deep Learning in Thoracic CT Images: a Comprehensive Review," *J Digit Imaging*, vol. 33, no. 3, pp. 655–677, Jun. 2020, doi: 10.1007/s10278-020-00320-6.
[112] D. Gu, G. Liu, and Z. Xue, "On the performance of lung nodule detection, segmentation and classification," *Computerized Medical Imaging and Graphics*, vol. 89, p. 101886, Apr. 2021, doi: 10.1016/j.compmedimag.2021.101886.
[113] R. Li, C. Xiao, Y. Huang, H. Hassan, and B. Huang, "Deep Learning Applications in Computed Tomography Images for Pulmonary Nodule Detection and Diagnosis: A Review," *Diagnostics*, vol. 12, no. 2, p. 298, Jan. 2022, doi: 10.3390/diagnostics12020298.
[114] F. Silva *et al.*, "Towards Machine Learning-Aided Lung Cancer Clinical Routines: Approaches and Open Challenges," *J Pers Med*, vol. 12, no. 3, p. 480, Mar. 2022, doi: 10.3390/jpm12030480.

Appendix A

A.1 Lung cancer screening methods

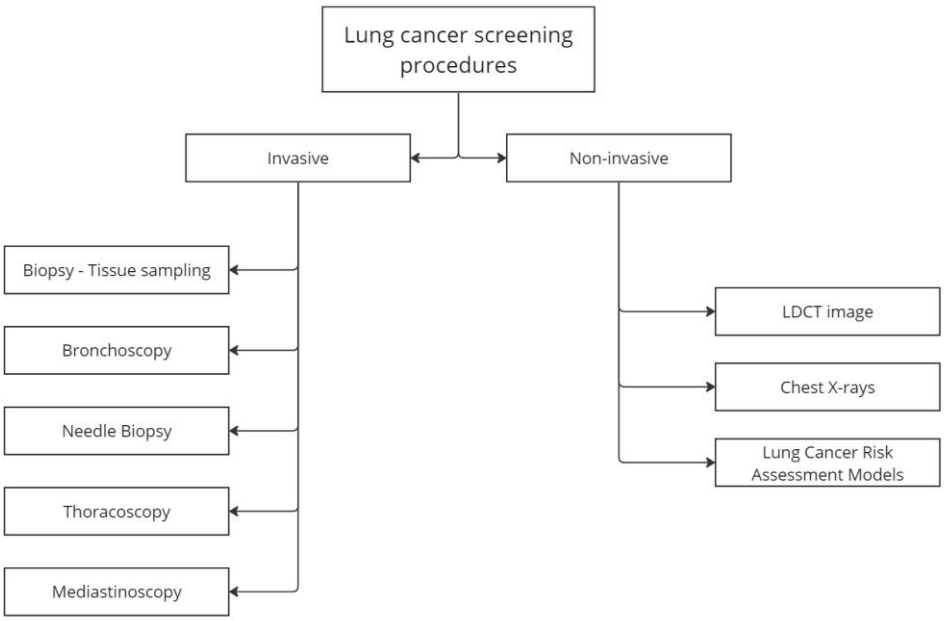


Figure A.1.1: Taxonomy of invasive and non-invasive lung cancer screening procedures

Invasive procedures concern physically entering the body, like using a scope to look inside the lungs or taking a piece of lung tissue for testing. If a suspicious nodule is detected through imaging, doctors may perform a biopsy to obtain tissue samples for a definitive diagnosis. This invasive procedure involves extracting a small piece of tissue for pathological examination [94]. Some methods for achieving that are the following:

- Bronchoscopy: A thin, flexible tube with a camera (bronchoscope) is inserted through the nose or mouth and into the airways to examine the lungs and collect tissue samples for biopsy [95].
- Needle Biopsy: A needle is used to extract a tissue sample from a suspicious lung nodule or lymph node for examination under a microscope. There are different types of needle biopsies, including transthoracic needle biopsy and endobronchial ultrasound-guided biopsy [96].

- Thoracoscopy or Video-Assisted Thoracoscopic Surgery (VATS): These minimally invasive surgical procedures involve making small incisions in the chest to access and biopsy lung tissue or remove a suspicious nodule [97].
- Mediastinoscopy: This procedure involves making a small incision in the neck and inserting a scope to examine and sample lymph nodes in the area between the lungs (mediastinum) [98].

These methods can give us a clear diagnosis and help plan treatment. However, these procedures pose numerous risks to patients, including pain, discomfort, potential blood loss, and an elevated risk of infection or pneumonia. Moreover, these procedures can be emotionally and physically taxing for the patient and require more recovery time [94].

These adverse effects underscore the urgent need for alternative, less invasive approaches to lung cancer screening, where deep learning technologies offer a promising avenue for improvement.

Nowadays, modern medical imaging techniques and tools employed by healthcare professionals have revolutionized patient screening, minimizing the need for invasive procedures and discomfort. Non-invasive procedures like low-dose CT or X-ray imaging, don't require physical entry into the body. These procedures are generally less uncomfortable and risky, but they might not always provide as much detailed information or accuracy in diagnosing lung cancer. Non-invasive procedures include:

- Chest X-rays: Historically, chest X-rays have been one of the primary tools for detecting lung abnormalities. They provide a two-dimensional image of the chest and can reveal the presence of lung nodules or other suspicious lesions. However, their sensitivity in detecting early-stage lung cancer is limited [99], [100].
- Low-dose Computed Tomography (LDCT) Scans: Computed Tomography scans have become a more advanced and widely adopted method for lung cancer screening. These scans use a series of X-rays to create detailed cross-sectional images of the chest. Low-dose CT (LDCT) scans, in particular, have gained prominence in recent years due to their ability to detect smaller nodules and early-stage cancers [101], [102].

- Lung Cancer Risk Assessment Models: Doctors often employ risk assessment models to identify individuals at a higher risk of developing lung cancer. These models take into account factors such as age, smoking history, and family history to stratify patients into different risk categories [103].

The choice between invasive and non-invasive procedures depends on the patient's situation and what the doctors need to find out. It's essential to consider both options carefully for lung cancer detection. The use of low-dose computed tomography (CT) scans for lung cancer screening has become increasingly popular due to their ability to detect pulmonary nodules at an early stage. However, interpreting these images requires expertise and time, which can lead to delays in diagnosis and treatment. They often lack in terms of sensitivity and objectivity required for optimal results [104]. These limitations underscore the urgency of exploring innovative, non-invasive, and more efficient approaches to lung cancer screening. To address this challenge, computer-aided diagnostic (CAD) systems have been developed to assist radiologists and other medical professionals in identifying and classifying pulmonary nodules.

These computer-assisted methods have significantly enhanced the capabilities of healthcare professionals in lung cancer screening and diagnosis. They not only improve the accuracy of detection but also streamline the workflow, leading to more efficient patient care and timely interventions when necessary. These methods can be discriminated in Figure A.1.2.

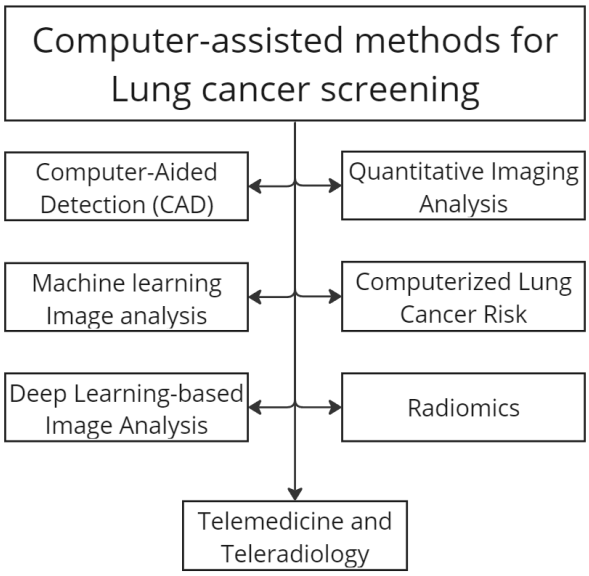


Figure A.1.2: Common computer assisted methods for lung cancer screening

Deep learning, a subset of artificial intelligence, has emerged as a promising technology to enhance the accuracy and efficiency of lung cancer screening [105]. Deep learning algorithms have shown great promise in improving the accuracy and efficiency of CAD systems, enabling them to automatically detect, segment, and classify pulmonary nodules in low-dose CT scans.

However, the choice of image modality plays a pivotal role. Researchers commonly utilize two primary modalities: 2D and 3D low-dose CT scans. The selection depends on the available data and the specific objectives of the screening program. Typically, a CT scan contains multiple slices and therefore both 2D and 3D option are available. Once the modality is determined, preprocessing steps are crucial to prepare the data for deep learning models. This includes candidate nodule generation to identify potential cancerous regions, resampling for uniformity, and lung masking to isolate the relevant structures. Finally, the model architecture is a critical component (Figure A.1.3). It involves selecting and configuring deep learning architectures, such as Convolutional Neural Networks (CNNs), 3D CNNs, Autoencoders and Deep Auto Encoders, to effectively analyze the pre-processed images and accurately detect lung cancer nodules or anomalies. Moreover, the utilization of synthetic data generation and data augmentation techniques holds significant importance due to the typically limited size of the datasets. The taxonomy presented to Figure A.1.3 will drive this

research and organize the presentation of the various deep learning methods used for lung cancer screening comprehensively.

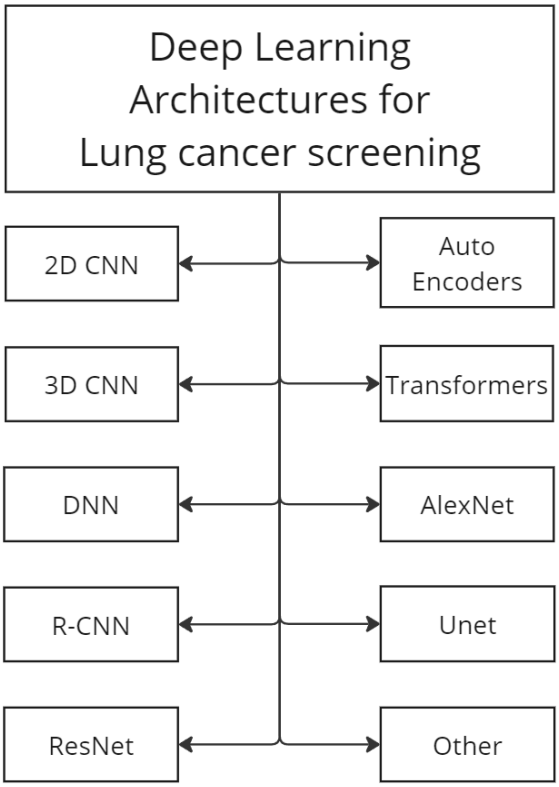


Figure A.1.3: Most common deep learning methods for lung cancer screening

A.2 Literature review procedure

From this literature review, we focused on investigating prior research efforts dedicated to the detection, segmentation, and classification of pulmonary nodules within low-dose CT scans, leveraging the capabilities of deep learning models.

Several review papers (**Σφάλμα! Το αρχείο προέλευσης της αναφοράς δεν βρέθηκε.**) have been published over the past decade on the subject of computer assisted lung cancer screening using machine learning. Older reviews focused on both traditional machine learning and deep learning methods [106], while more recent ones [107] have shifted their focus primarily the field of deep learning. Only few reviews [107] provide extensive details about the datasets, preprocessing methods and architecture details while advanced data augmentation methods that involve generative adversarial networks (GANs) for synthetic data generation are reported in this review [107]. This recent review [108] reports state-of-the-art deep learning methods but overlooks the current research trend of transformer techniques.

Table A.2.1: Literature review

Review	Publication Date	Date range	Datasets	Preprocessing	Methods
[109]	2018	2009-2018	yes	yes	Traditional ML, DL
[106]	2018	2006-2017	few (2)	briefly	Traditional ML, DL
[110]	2019	1990-2020	no	no	Traditional ML, DL
[111]	2020	2009-2018	yes	yes	Traditional ML, DL
[112]	2021	2005-2020	yes	no	Non-DL, DL
[107]	2021	2015-2020	yes	yes	DL, GAN
[113]	2022	2015-2021	yes	briefly	CNN
[114]	2022	2020-2021	no	no	Traditional ML, DL
[108]	2023	2018-2023	yes	no	DL
Ours	2024	2015-2023	yes	yes	CNN, DL, RNN, AE, GAN, Transformers

In our research, we comprehensively gathered information from the literature on the datasets used, preprocessing procedures, data augmentation techniques, architectural designs, and the reported performance metrics in the three tasks of interest - namely, pulmonary nodule detection, segmentation, and classification. Our analysis encompasses state-of-the-art deep learning approaches (CNN, autoencoders) and fast-growing and promising approaches, such as transformers. Furthermore, we assessed the credibility of each study by examining whether the authors presented lucid and comprehensive explanations of their methodologies and

adhered to machine learning best practices. Through these efforts, we provide a current and in-depth viewpoint on this dynamic and rapidly expanding field of study.

Given the vast number of studies published in recent years, our methodology aimed to ensure a rigorous and meticulous filtering process to isolate the most relevant and high-quality research. In our literature research, we employed the web scraping tool "Publish or Perish" [15] to enhance the efficiency of our literature search. This tool, driven by input parameters such as keywords and a specified year range, retrieves the most pertinent publications from Google Scholar. Upon initial investigation, it became evident that the field of lung nodule detection encompasses a multitude of diverse methodologies. In pursuit of a more comprehensive understanding of this subject, we opted to employ multiple keywords in our literature exploration. Table A.2.2: Query keywords and descriptions describes the queries used during our literature search:

Table A.2.2: Query keywords and descriptions

Query Keywords	Description
Lung nodule detection deep learning	Research on lung nodule detection using deep learning techniques.
Lung nodule convolutional neural networks	Studies involving convolutional neural networks (CNNs) for lung nodule detection.
Lung nodule segmentation	Research on the segmentation of lung nodules, often a critical step in detection.
Lung nodule transfer learning	Investigations into the use of transfer learning for lung nodule detection.
Lung nodule Generative Adversarial Networks synthetic data	Utilizing GANs to generate synthetic data for lung nodule detection.
Lung nodule convolutional auto-encoders	Studies involving convolutional auto-encoders for lung nodule analysis.

Our initial step involved sifting through a substantial population of 1,200 studies. During the first filtering phase, we applied a preliminary screening by examining the titles of these publications. This preliminary step allowed us to eliminate studies that were not directly aligned with the scope and objectives of our review. Through this process, we successfully narrowed down our selection to 560 studies that showed potential for inclusion in our analysis.

To further refine our selection, we conducted the second filtering phase. Here, we delved deeper into the studies by assessing their abstracts and methodologies. This phase was crucial in ensuring that the selected studies not only exhibited

relevance but also demonstrated a sound research design. By applying this stringent criterion, we were able to eliminate studies that lacked the necessary depth and rigor in their approach. This led us to a more focused set of 216 studies that provided a strong foundation for our review.

In the third and final filtering phase, we aimed to identify studies with exceptional performance metrics. We focused on publications that reported high-performance indicators such as accuracy, sensitivity, specificity, and area under the curve (AUC), all exceeding the threshold of approximately 90%. Additionally, we prioritized studies that exhibited a high DICE score, further emphasizing the quality and accuracy of their results. This approach allowed us to spotlight studies that showcased the most promising outcomes in terms of lung nodule detection, segmentation, and classification, resulting into 110 studies to analyze.

Throughout our filtering process, we also took into account the impact factor of each publication. Factors such as the publication year, the number of citations, and the reputation of the journal were considered as indicators of the study's influence within the field. This consideration provided us with an additional layer of insight into the significance and relevance of the studies. By employing this meticulous approach to study selection, we were able to curate a refined and high-quality collection of research articles that formed the basis of our review. In this thesis we report our finding for the segmentation and classification works.

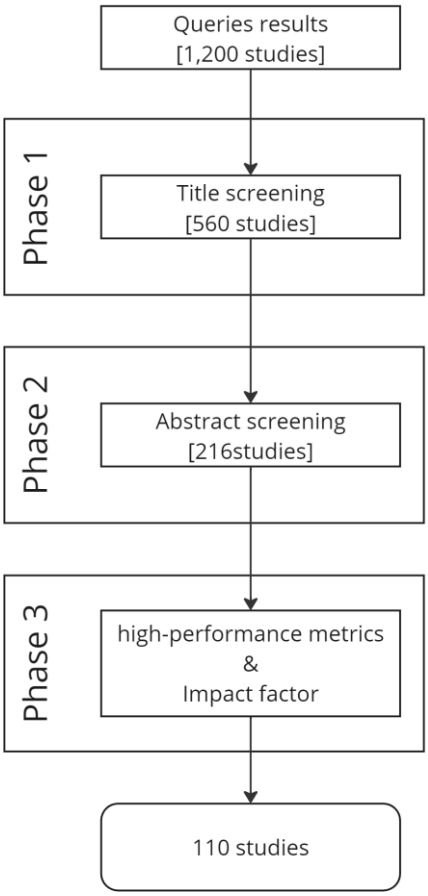


Figure A.2.1: Literature screening phases

A.2.1 Literature review tables

Table A.2.1 Nodule Segmentation - Preprocessing steps per work

Author	Year	Dataset	Preprocessing	Data augmentation	Deep Architecture
[22]	2017	LIDC-IDRI, private set	3D-PA, 2D-PA,	n/a	Central Focused Convolutional Neural Networks (CF-CNN)
[39]	2019	LIDC-IDRI	3D-PA	n/a	Cascaded Dual-Pathway Residual Network
[23]	2019	LIDC-IDRI	2D-PA, PSEG, Re	n/a	SegNet, deep fully convolutional network
[26]	2019	LIDC-IDRI	3D-PA	n/a	3D DCNN
[21]	2020	LIDC-IDRI	2D-PA	n/a	Deep residual deconvolutional network, TL
[28]	2020	LIDC-IDRI	2D-PA, No		Deep Residual U-Net
[40]	2020	LIDC-IDRI	3D-PA	n/a	DB-ResNet, CF-CNN
[29]	2020	LIDC-IDRI	2D-PA	Ro, Zoom, Padding	U-net
[27]	2021	LIDC-IDRI, LNDb, ILCID	TH, PSEG, Maximum intensity projection	Ro, Blur, No, Rand pixels to zero	2D CNN

[31]	2021	LIDC-IDRI	2D-PA, synthetic pseudo-color image.	Intensive augmentations	U-Net
[34]	2022	LUNA16	TH, No, 3D-Pa	Ro, Transpose, Affine Transform, Fl, Br, Contrast	V-net
[35]	2021	LIDC-IDRI, SHCH	TH, Contrast ench, PSEG, Lesion Localization With Region Growing	Fl, Ro, Cr, deformation	2D-3D U-net
[38]	2021	LIDC-IDRI, LUNA16	Cr, 2D-PA, Upscale	Ro, Fl, elastic transform	Faster R-CNN
[36]	2021	LIDC-IDRI	PSEG, 2D-PA	Ro, Fl, Sh, Zoom, Cr	U-net
[41]	2021	LIDC-IDRI	PSEG, TH, Re, 3D-PA	n/a	3D res U-net
[24]	2021	LIDC-IDRI	2D-PA, Re	n/a	VGG-SegNet
[25]	2022	hospital data	3D-PA	Ro, Mirroring	3D FCN
[42]	2022	LUNA16, ILND	TH, 3D-PA, Sampling for balance	patch-based augmentation	3D GAN
[47]	2022	LIDC-IDRI	TH, No, 3D-PA	Ro, Sc, Fl	3D Dual Attention Shadow Network (DAS-Net)
[44]	2022	LIDC-IDRI	2D-PA	Ro, Tr, Fl	Transformer
[45]	2023	LIDC-IDRI	grayscale thresholding, No, Re	n/a	Dual-encoder-based CNN
[32]	2023	LIDC-IDRI, AHAMU-LC	window selection, No	Fl	RAD - U-net

[33]	2023	LIDC-IDRI, private set	2D-PA	Cr, Sc, Br, Contrast, Sat, Random noise	SMR - U-net 2D
[43]	2023	LIDC-IDRI	2D-PA	Ro, random luminance, random gamma rays, Gaussian noise, hue/sat	U-shaped hybrid transformer
[46]	2023	LIDC-IDRI, LUNA16	TH, No, Re3D-PA	n/a	3D U-net based
[37]	2023	LIDC-IDRI	mask generation	n/a	GUNet3++

Abbreviations: Rs: Resampling, Rz: Resize, PA: Patch (2D or 3D), Cr: Crop, No: Normalization, Ro: Rotate, Sc: Scale, Tr: Translate, TH: Threshold (HU), Fl: Flip, Du: Duplicate

Table A.2.2 Nodule Segmentation - Performance per work

Author	Year	Dataset	Deep Architecture	Input size	Segmentation type	DSC (%)	IoU (%)	sensitivity/recall (%)
[22]	2017	LIDC-IDRI, private set	Central Focused Convolutional Neural Networks (CF-CNN)	572 × 572, 3×35×35	3D, 2D	82.15± 10.76, LIDC 80.02± 11.09 Private set	-	-
[39]	2019	LIDC-IDRI	Cascaded Dual-Pathway Residual Network	65 × 65 × 3	2D, 3D(mask)	81.58±11.05	-	-
[23]	2019	LIDC-IDRI	SegNet, deep fully convolutional network	128 × 128	2D	93±0.11	-	-
[26]	2019	LIDC-IDRI	3D DCNN	n/a	3D	83.10±8.85	71.85±10.48	-
[21]	2020	LIDC-IDRI	Deep residual deconvolutional network, TL	512×512	2D	94.97,	88.68	-
[28]	2020	LIDC-IDRI	Deep Residual U-Net	128×128	3D	87.5± 10.58	-	-
[40]	2020	LIDC-IDRI	DB-ResNet, CF-CNN	3 × 35 × 35	3D	82.74 ± 10.19	-	-
[29]	2020	LIDC-IDRI	U-net	64×64	2D	-	76.6± 12.3	-
[27]	2021	LIDC-IDRI, LNDb,	2D CNN	96×96	2D	80	-	

		ILCID						
[31]	2021	LIDC-IDRI	U-Net	256×256	2D	93.14	-	91.76
[34]	2022	LUNA16	V-net	96×96×16	3D	95,01	83	
[35]	2021	LIDC-IDRI, SHCH	2D-3D U-net	2D-3D (3-slices)	2D-3D	83.16/81.97	-	-
[38]	2021	LIDC-IDRI, LUNA16	Faster R-CNN	224×224	2D	89.79 /90.35	82.34 / 83.21	-
[36]	2021	LIDC-IDRI	U-net	32×32	2D	86.23	-	-
[41]	2021	LIDC-IDRI	3D res U-net	48 × 192 × 192	3D	80.5	-	80.5
[24]	2021	LIDC-IDRI	VGG-SegNet	224 × 224 × 3 channels	2D	90.49	82.64	-
[25]	2022	hospital data	3D FCN	128×128×64	3D	84.5	73.8	
[42]	2022	LUNA16, ILND	3D GAN	64 × 64 × 32	3D	80.74 / 76.36	-	85.46 / 82.56
[47]	2022	LIDC-IDRI	3D Dual Attention Shadow Network (DAS-Net)	16×128×128	3D	92.05	-	90.81
[44]	2022	LIDC-IDRI	Transformer	64×64	2D	89.86	-	90.50
[45]	2023	LIDC-IDRI	Dual-encoder-based CNN	512×512	2D	87.91	-	90.84
[32]	2023	LIDC-IDRI,	RAD - U-net	512×512	2D	-	87.76/88.13	-

		AHAMU-LC						
[33]	2023	LIDC-IDRI, private set	SMR - U-net 2D	128×128	2D	91.87	86.88	-
[43]	2023	LIDC-IDRI	U-shaped hybrid transformer	64×64, 96 × 96, 128×128	2D	91.84		92.66
[46]	2023	LIDC-IDRI, LUNA16	3D U-net based	64 × 64 × 32	3D	82.48	70.86	82.74
[37]	2023	LIDC-IDRI	GUNet3++	n/a	2D	97.2	-	97.7

Table A.2.3 Nodule Classification - Preprocessing steps per work

Author	Year	Dataset	Preprocessing	Data augmentation	Deep Architecture
[63]	2017	LIDC-IDRI	Rs, 2D-PA, No	Ro, Sc (only in the test set)	ResNet
[65]	2017	LIDC-IDRI	3D-PA, MV	Ro	3D MV-CNN + SoftMax
[66]	2017	LIDC-IDRI	2D-PA, Rs, Rz	Ro, Sh, Fl, Tr	ResNet-50, TL
[64]	2018	LIDC-IDRI	Rs, Rz, MV, Cr, 2D-Pa	Tr, Ro, Fl	MV-KBC
[69]	2018	LIDC-IDRI	3D-PA, QIF extraction	Ro, Sc, shifted up to 30%	CNN+ Random Forest
[70]	2018	LIDC-IDRI + Private set	Rs, 3D-PA, No	-	3D DenseNet, TL
[48]	2018	LIDC-IDRI	Rs,	-	CNN + PSO
[78]	2018	LUNA16	Re, 3D-PA	Tr, Ro	3D DCNN
[50]	2019	LUNA16	2D-PA	Ro, Fl	Novel 2D CNN
[51]	2019	LIDC-IDRI	No, Cr, 2D-PA	Ro, Sc, Gaussian Blurring	Novel 2D CNN
[54]	2019	LIDC-IDRI	Re, 2D-PA	Ro, Fl	CNN, TL
[49]	2020	LIDC-IDRI	Rs, Th	yes but no info	MAN (modified AlexNet), TL
[52]	2020	LIDC-IDRI	2D-PA	Tr, Ro, Sc, GAN*	CNN, TL
[58]	2020	LIDC-IDRI, private dataset (FAH-GMU)	Cr, 2D-PA	-	DTCNN, TL

[71]	2020	LIDC-IDRI, DeepLNDataset	No, Cr, 3D-PA	Tr, Fl	3D CNN
[55]	2020	LIDC-IDRI, LUNGx Challenge database	Re, No, 2D-PA	Tr, Ro, Fl	2D CNN, TL
[72]	2020	LIDC-IDRI	Cr, 3D-PA	Adjust sampling rate	MRC-DNN
[79]	2020	LIDC-IDRI	Re, No, Th, Cr, 3D-PA	Sampling different slices from the same nodule to achieve a better class balancing	CAE, TL
[82]	2020	LIDC-IDRI, LUNA16	Re, No, Cr, 3D-PA, 2D-PA	Tr, Ro, Fl	Multi-Task CNN
[56]	2020	LUNA16	Cr, 2D-PA	down sampling the negative samples, Ro	Fractalnet and CNN
[83]	2020	LIDC-IDRI	Re	Sc(zoom), FL, Ro	DCNN
[67]	2021	NLST, DLCST	3D-PA, 2D-PA (9 views)	n/a	2D CNN 9 views, 3D CNN
[60]	2021	LUNA16/Kaggle DSB 2017 dataset	raw data to png	Ro,H-Fl, clip, blurry	Dense Convolutional Network (DenseNet)
[68]	2021	LIDC-IDRI / ELCAP	2D-PA, 3D-PA	n/a	2D MV-CNN 3D MV-CNN
[86]	2021	LIDC	3D-PA, zero-padding	n/a	Capsule networks

					(CapsNets)
[88]	2021	LIDC-IDRI	3D-PA, padding, Cr	Fl	3D NAS method, CBAM module, A-Softmax loss, and ensemble strategy to learn efficient,
[87]	2021	LIDC-IDRI	2D-PA	n/a	Deep Convolutional Generative Adversarial Network (DC-GAN)/ FF-VGG19
[61]	2021	LUNA16	TH, 2D-PA, balancing samples	Ro, Fl	BCNN [VGG16, VGG19] combination with and without SVM
[76]	2021	LUNA16	n/a	n/a	3D CNN
[62]	2021	pet-ct private, LIDC-IDRI	n/a	Ro,Fl, Shift	2d cnn
[89]	2021	LIDC-IDRI	3D-PA, Re	Fl, Pad	3D DPN _ attention mech
[77]	2021	LIDC-IDRI	3D-PA	n/a	3D CNN + biomarkers
[90]	2021	LIDC-IDRI	Re, 3D-PA, No	Ro	3D attention

[74]	2022	LIDC-IDRI and LUNGx	Re, 3D-PA, TH, No	Ro	ProCAN
[58]	2022	LIDC-IDRI	2D-PA	Ro, Tr	DCNN
[59]	2022	LIDC-IDRI	Re, 2D-PA	Ro, overlays on the axial, coronal, and sagittal slices	Transformers
[84]	2022	LIDC-IDRI	interpolation, TH	Fl, Ro	CNN-based MTL model that incorporates multiple attention-based learning modules
[85]	2022	LIDC-IDRI	Re, Rz, No	Ro	Transformers
[75]	2022	LUNA16	3D-PA, No	Fl, Gaussian noise	3D ResNet + attention
[73]	2023	LIDC-IDRI / TC-LND Dataset/ CQUCH-LND	No, 3D-PA, Sc	Ro, Fl	STLF-VA
[81]	2023	LIDC-IDRI	Re, No	n/a	Transformer
[57]	2023	LIDC-IDRI	Re, 2D-PA, Re	Fl, Brightness, contrast, Sc	F-LSTM-CNN
[80]	2023	private	TH,3D-PA	n/a	CAE

Abbreviations: Rs: Resampling, Rz: Resize, PA: Patch (2D or 3D), Cr: Crop, No: Normalization, Ro: Rotate, Sc: Scale, Tr: Translate, TH: Threshold (HU), Fl: Flip, Du: Duplicate.

Table A.2.4 Nodule Classification - Performance per work

Author	Year	Dataset	Deep Architecture	Input size	Sensitivity/Recall (%)	Specificity (%)	Precision (%)	AUC (%)	Accuracy (%)
[63]	2017	LIDC-IDRI	ResNet	64×64	91.07	88.64	-	94.59	89.90
[65]	2017	LIDC-IDRI	3D MV-CNN + SoftMax	N [40, 50, 60] N×N×6 slices	95.60	93.94	-	99	-
[66]	2017	LIDC-IDRI	ResNet-50, TL	200×200	91.43	94.09	-	97.78	93.40
[64]	2018	LIDC-IDRI	MV-KBC	224×224	86.52	94		97.50	91.60
[69]	2018	LIDC-IDRI	CNN + Random Forest	N×N×S N[47,21,31], S[5, 3]]	94.80	94.30	-	98.4	94.60
[70]	2018	LIDC-IDRI + Private set	3D DenseNet, TL	N [50, 10] S [5,10] N×N×S	90.47	90.33	-	95.48	90.40
[48]	2018	LIDC-IDRI	CNN + PSO	28×28	92.20	98.64	-	95.5	97.62
[78]	2018	LUNA16	3D DCNN	64×64×64, 48×48×48	95.4/ 1 FP	-	-	0.910 (FROC)	-
[50]	2019	LUNA16	Novel 2D CNN	64×64	96.0	97.3	-	98.2	97.2
[51]	2019	LIDC-IDRI	Novel 2D CNN	32×32	92.67	-	-	95.14	92.57
[54]	2019	LIDC-IDRI	CNN, TL	53×53	91	-	-	94	88

[49]	2020	LIDC-IDRI	MAN (modified AlexNet) + SVM, TL	32× 32 × 32	-			95.70	91.60
[52]	2020	LIDC-IDRI	CNN	227×227	98.09	95.63	-	99.5	97.27
[58]	2020	LIDC-IDRI, private-set	DTCNN	52×52	93.4	93	-	93.4	93.9
[71]	2020	LIDC-IDRI, DeepLNDataset	3D CNN	64×64	93.69/ 100	95.15/ 100	-	94.9	94.57/ 100
[55]	2020	LIDC-IDRI, LUNGx Challenge database	CNN, TL	N [32, 48, 64] N×N×N	85.58	95.87	-	94	92.65
[72]	2020	LIDC-IDRI	MRC-DNN	64×64	97.19	-	-	99.1	96.69
[79]	2020	LIDC-IDRI	CAE, TL	32×32×32	81	95	-	-	90
[82]	2020	LIDC-IDRI, LUNA16	Multi Task CNN	80×80×80	84.8	-		93.6	-
[56]	2020	LUNA16	Fractalnet and CNN	64×64	87.74, 84.00	88.87, 96.80	-	0.955(LIDC), 0.973(LUNA)	-
[83]	2020	LIDC-IDRI	DCNN	50×50	97.52	86.76		98	94.06
[67]	2021	NLST, DLCST	2D CNN 9 views, 3D CNN	224×224	90.67	90.80		-	90.73
[60]	2021	LUNA16/Kaggle DSB 2017 dataset	Dense Convolutional Network (DenseNet)	64×64, 64×64×64	-	-	-	93	-

[68]	2021	LIDC-IDRI / ELCAP	2D MV-CNN 3D MV-CNN	80×80, 64×64 , 48×48 , 32×32 and 16×16	98.2	99.45	-		98.83
[86]	2021	LIDC	Capsule networks (CapsNets)	(20x20), (30x30) and (40x40)	98	97	-	99	97
[88]	2021	LIDC-IDRI	3D NAS method, CBAM module, A-Softmax loss, and ensemble strategy	80×80×3 slices	89.5	93.4	-	95.6	90.7
[87]	2021	LIDC-IDRI (SPN only)	Deep Convolutional Generative Adversarial Network (DC-GAN)/ FF-VGG19	32×32×32	85.37	95.04	-	-	90.77
[61]	2021	LUNA16	BCNN [VGG16, VGG19] combination with and without SVM	32×32	89.3	94.8	-	92.1	92.1
[76]	2021	LUNA16	3D CNN	50 × 50	-	-	-	95.9	91.99

[62]	2021	pet-ct private, LIDC-IDRI (SPN only)	2D CNN	64×64×240	94	-	87	97	97.17
[89]	2021	LIDC-IDRI	3D DPN _ attention mech	32×32	92.7	95.2	-	94	94
[77]	2021	LIDC-IDRI	3D CNN + biomarkers	32×32×32	91.3 (FP rate of 8.0%)	-	-	-	91.9
[90]	2021	LIDC-IDRI	3D attention	32×32×16 slices	-	-	-	86.74	-
[74]	2022	LIDC-IDRI and LUNGx	ProCAN	32×32×32	92.36	-	92.59	96.17	92.81
[58]	2022	LIDC-IDRI	DCNN	32×32×32	-	-	-	98.05	95.28
[59]	2022	LIDC-IDRI	Transformers	52×52	97.1	97.2		99.56	97.8
[84]	2022	LIDC-IDRI	CNN-based MTL model that incorporates multiple attention-based learning modules	32×32	-	-	-	96.28	92.92
[85]	2022	LIDC-IDRI	Transformers	64×64	96.2	82.9	97.8	95.9	94.7
[75]	2022	LUNA16	3D ResNet + attention	32×32	94.4		95.9	98.5	96.1

[73]	2023	LIDC-IDRI / TC-LND Dataset/ CQUCH-LND	STLF-VA	32×32×32	89.10	93.39	91.59	91.25	91.25
[81]	2023	LIDC-IDRI	Transformer	64×64×32	91.62	93.08	92.99	97.17	92.36
[57]	2023	LIDC-IDRI	F-LSTM-CNN	80×80×60	87.69	95.38	-	97.40	92.82
[80]	2023	private	CAE	224×224×3	100	93.7	-	99.5	95.5

A.3 Model predictions

This appendix section contains plots of our nodule segmentation model predictions

A.3.1 Nodule segmentation predictions

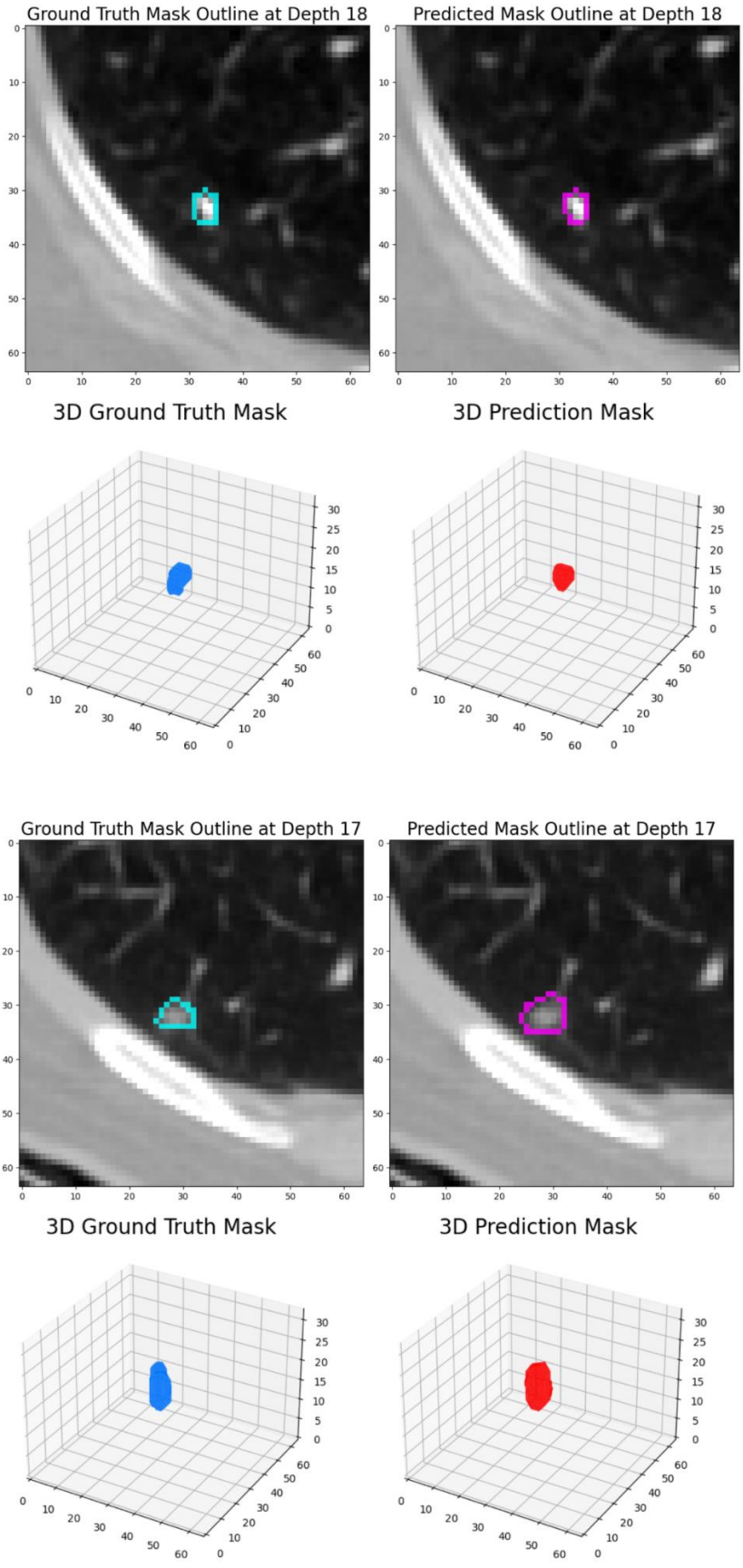


Figure A.3.1.1: Segmentation predictions i

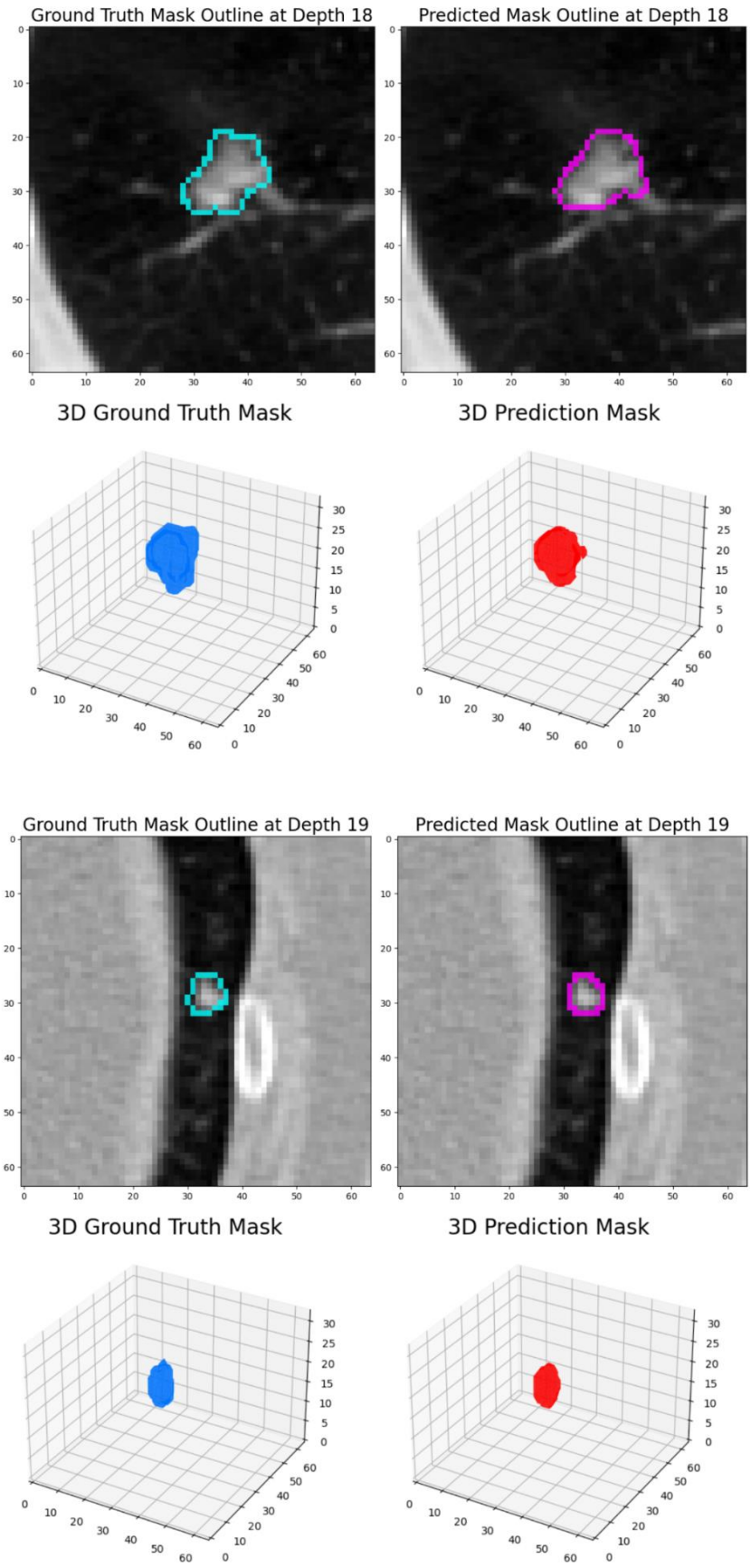


Figure A.3.1.2: Segmentation predictions ii

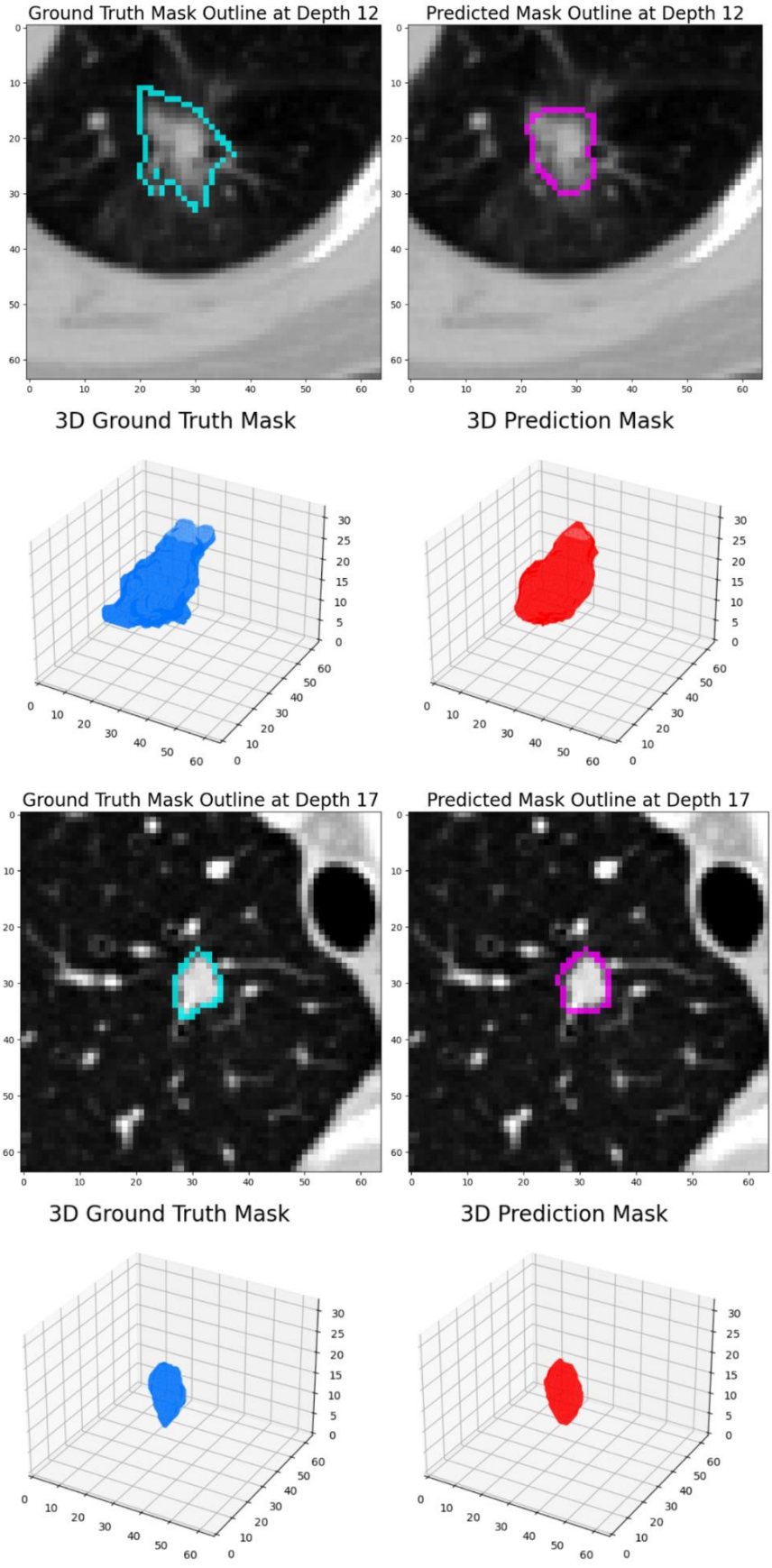


Figure A.3.1.3: Segmentation predictions iii

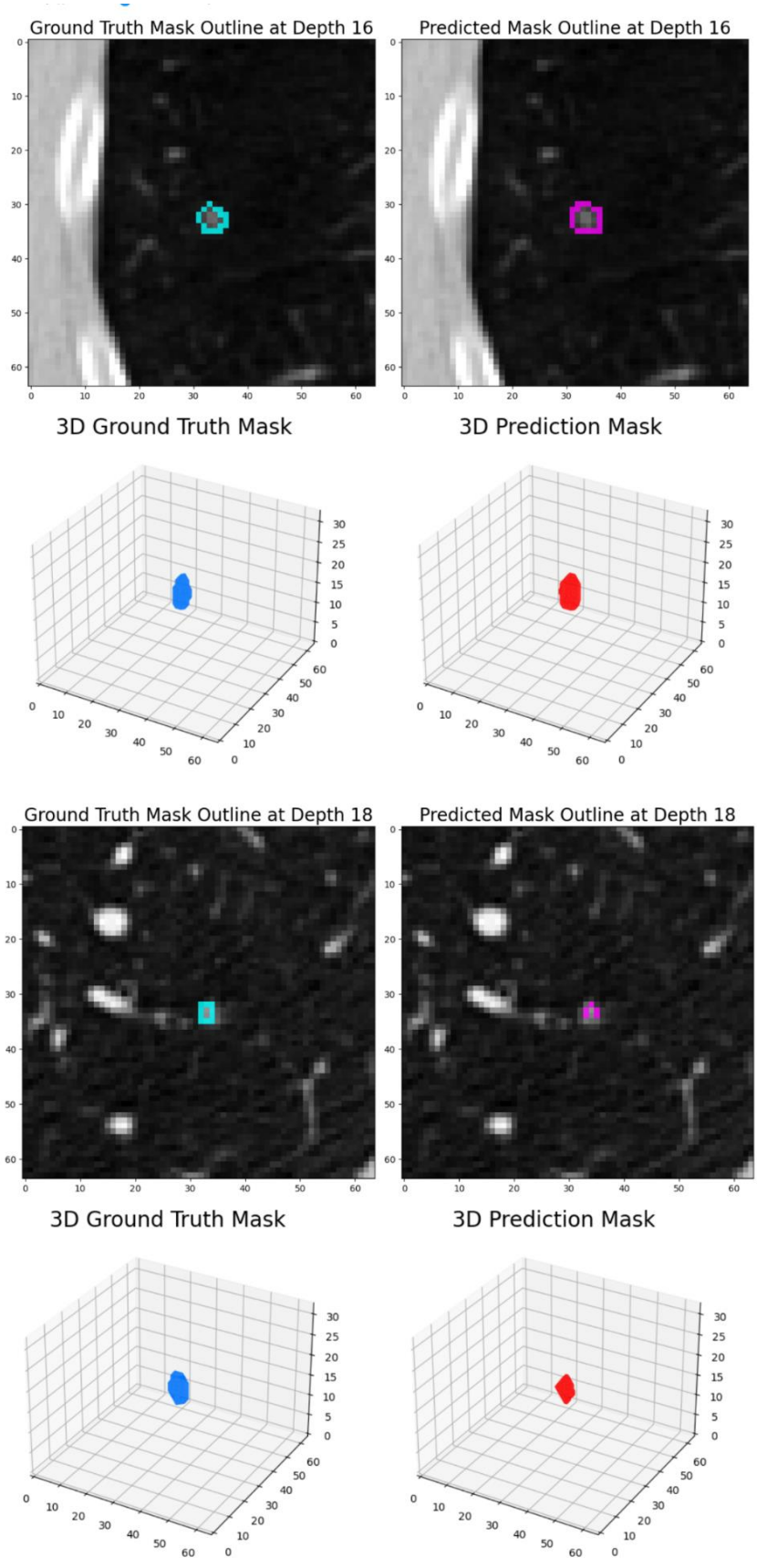


Figure A.3.1.4: Segmentation predictions iv

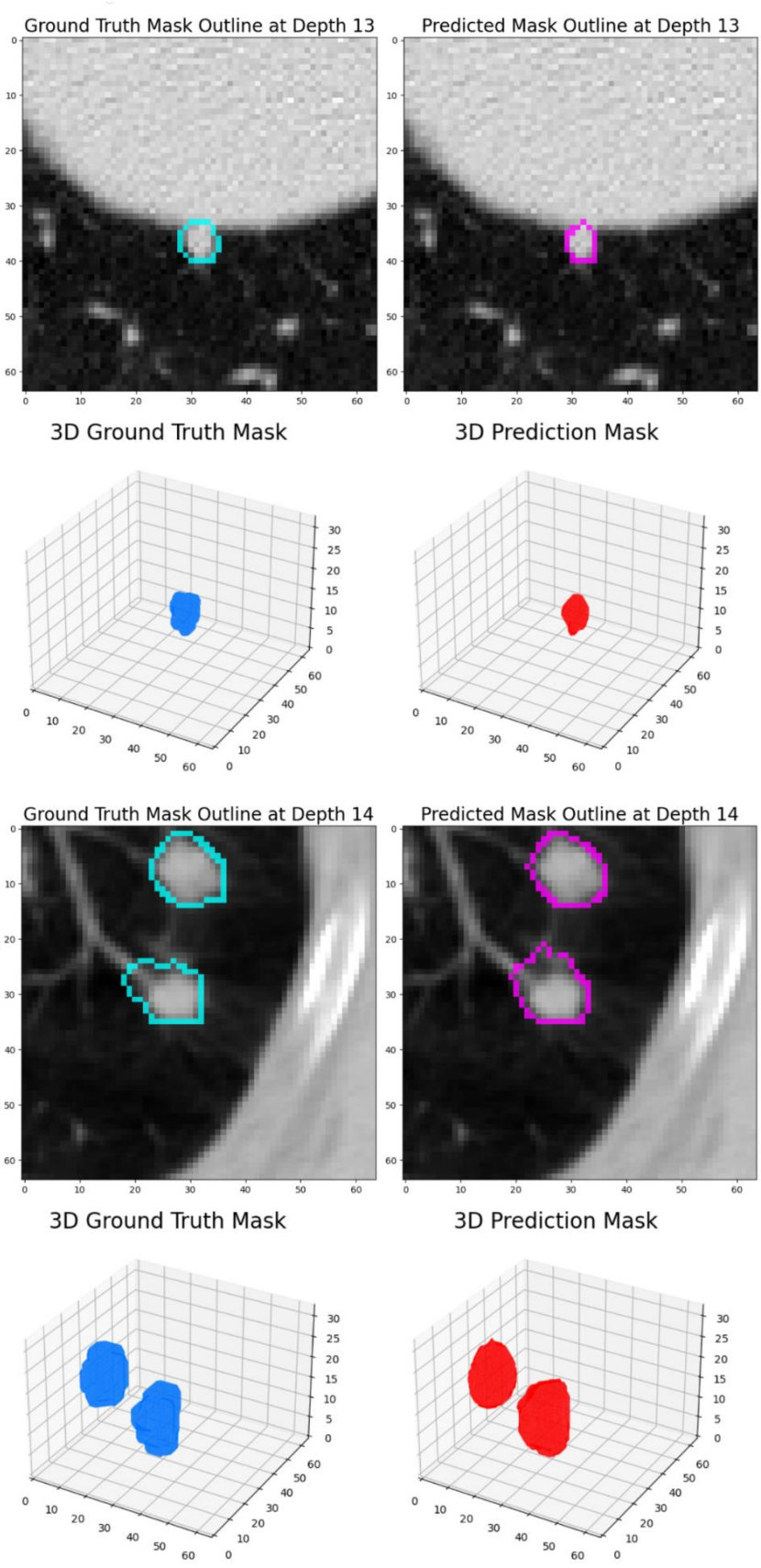


Figure A.3.1.5: Segmentation predictions v

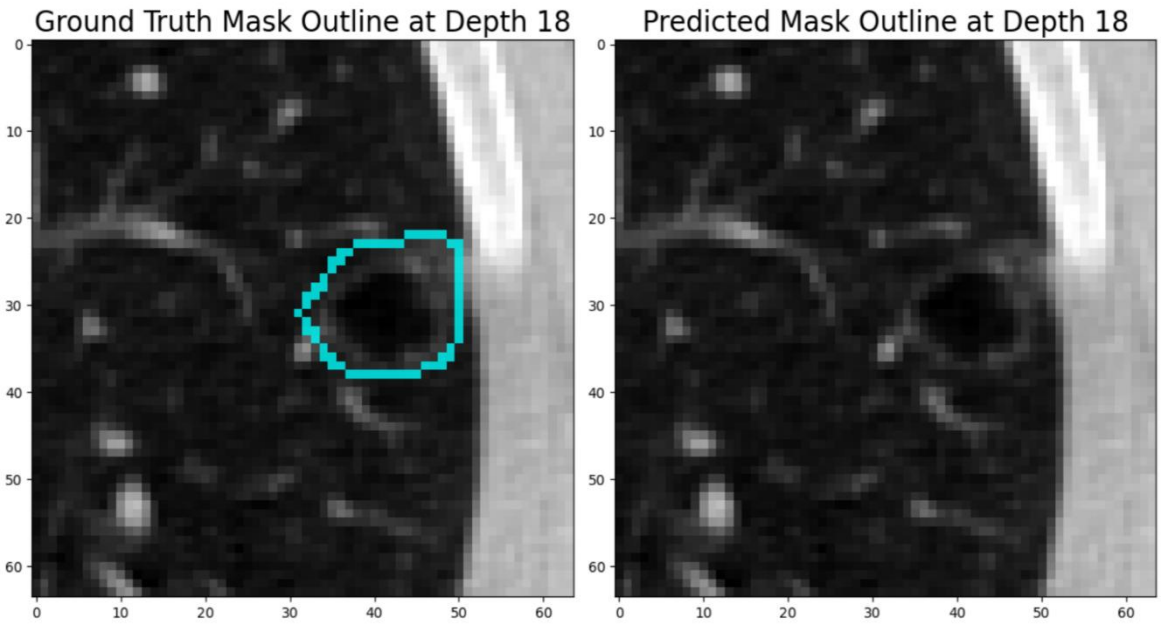


Figure A.3.1.6: False-negative prediction on Ground-glass opacity (GGO) nodule

A.3.2 Nodule classification confusion matrices

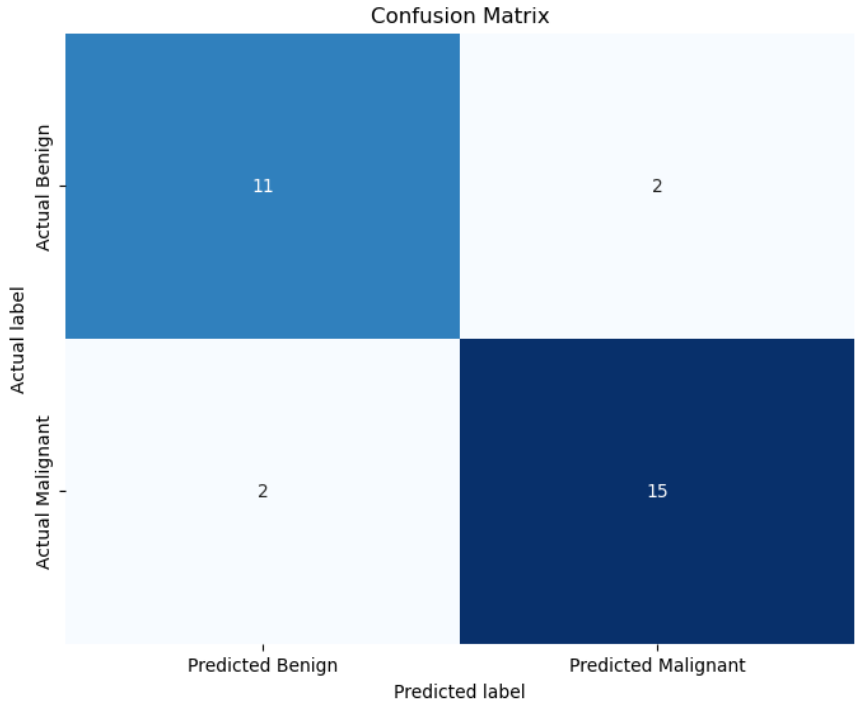


Figure A.3.2.1: Malignancy confusion matrix on test set– 3D patch input

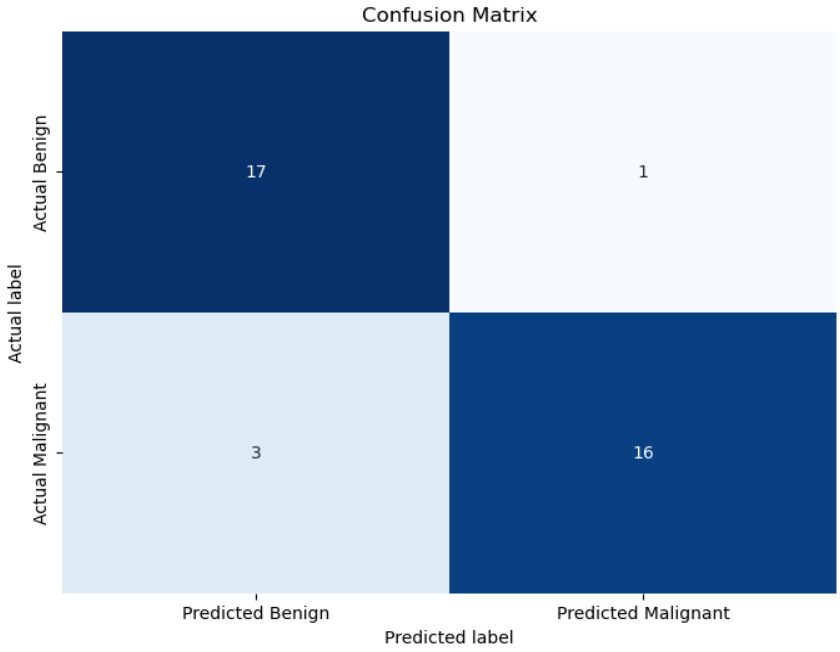


Figure A.3.2.2: Malignancy confusion matrix on test set – 3-view input

Appendix B Version Information

Version Number:	4.0
Date:	24 June 2024
Version Information:	Final version
Author:	Ioannis D. Marinakis

B.1 Copyright Information

© Ioannis D. Marinakis, 2024

All rights reserved. No part of this thesis may be reproduced, distributed, or transmitted in any form or by any means, including photocopying, recording, or other electronic or mechanical methods, without the prior written permission of the author, except in the case of brief quotations embodied in critical reviews and certain other non-commercial uses permitted by copyright law.

For permission requests, write to the author at:
mtp212@edu.hmu.gr
john@imaic.gr