
# YOLO PIPELINES FOR LUNG NODULE DETECTION

FRANCISZEK KACZMAREK

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN DATA SCIENCE & SOCIETY
AT THE SCHOOL OF HUMANITIES AND DIGITAL SCIENCES
OF TILBURG UNIVERSITY

STUDENT NUMBER

2134926

COMMITTEE

dr. Sharon Ong
prof. dr. Max Louwerse

LOCATION

Tilburg University
School of Humanities and Digital Sciences
Department of Cognitive Science &
Artificial Intelligence
Tilburg, The Netherlands

DATE

December 2th, 2024

WORD COUNT

8743

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my supervisor, Dr. Sharon Ong, for her guidance, support, and invaluable advice throughout this journey. Her willingness to take on the role of my supervisor, her encouragement, and her personalized approach have been crucial to the completion of this thesis. I feel truly fortunate to have had the opportunity to work under her mentorship. I am also grateful to Prof. Dr. Max Louwerse, whose insightful comments and suggestions have significantly enhanced the quality of this thesis.
I would also like to thank my girlfriend, Julia, for her endless patience and understanding during this challenging period. Her support and encouragement have kept me motivated and focused on my goals.
Finally, I am deeply grateful to my mother and my stepdad for their unwavering support and the opportunities they have provided, enabling me to complete my studies. Their belief in me and their generosity have been a constant source of strength throughout this journey.

# YOLO PIPELINES FOR LUNG NODULE DETECTION

FRANCISZEK KACZMAREK

**Abstract**

Lung cancer, as the leading cause of cancer-related deaths globally, underscores the urgent need for early and accurate detection of lung nodules. Lung nodules, small and well-defined masses of tissue within the lungs, are often the earliest indicators of malignancies. This thesis investigates applying state-of-the-art You Only Look Once (YOLO) models for detecting lung nodules in computed tomography scans, focusing on YOLOv5 and the recently introduced YOLOv10.

To support this investigation, a comprehensive pipeline was developed, which includes Hounsfield Unit windowing and bounding box adjustments around nodules to enhance detection accuracy. The study also evaluates the impact of YOLOv10's architectural advancements compared to YOLOv5. Using the LIDC-IDRI dataset, experiments revealed that YOLOv10-l outperformed YOLOv5-l, achieving a higher mAP@0.5 (0.61 vs. 0.54), faster inference times (6.8 ms vs. 8.6 ms), and improved precision (0.66 vs. 0.62) and recall (0.54 vs. 0.53), showcasing its suitability for clinical applications.

The findings demonstrate that optimizing windowing parameters and bounding box sizes significantly enhances detection accuracy, while error analysis highlighted challenges in detecting small nodules and nodules near lung boundaries. These insights underline the importance of preprocessing and model selection in addressing the unique complexities of lung nodule detection.

## 1 DATA SOURCE, ETHICS, CODE, AND TECHNOLOGY STATEMENT

The dataset in this thesis, the LIDC-IDRI dataset, was acquired from a publicly accessible online repository (Armato et al., 2015). This dataset is fully anonymized, and its use complies with ethical guidelines for de-identified medical imaging data research.

The implementation of this study employed openly available code frameworks. The models were developed using the YOLOv5 and YOLOv10 architectures made easily accessible through ultralitics (Jocher et al., 2023).

The Python libraries NumPy (Harris et al., 2020), Pandas (McKinney et al., 2010), Matplotlib (Hunter, 2007), SciPy (Virtanen et al., 2020), Scikit-Image (van der Walt et al., 2014), NiBabel (Brett et al., 2023), OpenCV (Itseez, 2015), PyTorch (Paszke et al., 2017), Torchvision (maintainers & contributors, 2016), and additional tools such as Albumentations (Buslaev et al., 2020), TQDM (da Costa-Luis, 2023) were utilized for various aspects of data processing, augmentation, visualization, and model implementation.

The specific versions of these libraries used in this work include NumPy (v1.23.5), Pandas (v1.1.4), Matplotlib (v3.3), SciPy (v1.4.1), Scikit-Image (v0.19.3), NiBabel (v5.1.0), OpenCV (v4.1.1 and v4.9.0.80), PyTorch (v1.8.0 and v2.0.1), Torchvision (v0.9.0 and v0.15.2), Albumentations (v1.0.3), TQDM (v4.66.3), and PyYAML (v6.0.1). The lung segmentation task used the `R231` model (Hofmanninger et al., 2020).

The code used in this thesis is publicly available on GitHub at https://github.com/franekaczmarek/YOLO-Pipelines-for-Lung-Nodule-Detection.

Writing assistance was provided through generative language models (ChatGPT by OpenAI) and spell-checking software (Grammarly). These tools were employed to improve language clarity and enhance grammar and spelling accuracy. All technical content, experimental design, and interpretation remain the author's intellectual contributions.

The research relied on Tilburg University's GPUs, ensuring computational efficiency for model training and evaluation.

# 2 INTRODUCTION

## 2.1 *Project Definition*

Lung nodules, which are small, clearly defined tissue masses in the lungs, frequently serve as the earliest signs of malignancies that may progress to lung cancer (Loverdos et al., 2019). The application of advanced diagnostic technologies, particularly Computerized Tomography (CT) scans, presents a valuable opportunity for the early detection of cancer-causing lung nodules, which could substantially improve patient outcomes and reduce mortality rates (Gautam et al., 2022).

Recent advancements in deep learning, especially Convolutional Neural Networks (CNNs), have significantly improved the analysis of CT images (Grover et al., 2024). CNNs are adept at identifying complex patterns within image data, making them invaluable for tasks like detection and segmentation, where accuracy is paramount (Grover et al., 2024). Object detection, a key task in computer vision, refers to the process of identifying and locating objects of interest within an image using bounding boxes, which correspond to rectangular shapes around the objects. Among these

advancements, models like You Only Look Once, which are specialized implementations of CNNs, have emerged as powerful tools for real-time object detection, including applications in medical imaging. Unlike traditional CNN approaches that process image regions sequentially, YOLO analyzes the entire image in a single forward pass, enabling faster and more efficient detection without compromising accuracy (Sapkota et al., 2024).

This study aims to develop a pipeline for detecting lung nodules from CT scans, aiming to expedite lung cancer detection and optimize CAD systems. The pipeline will leverage the YOLO framework for lung nodule detection. By comparing YOLOv5 (Jocher et al., 2022) and YOLOv10, the study will evaluate the effectiveness of YOLO-based detection techniques for clinical applications.

While conducting this study, the Lung Image Database Consortium image collection (LIDC-IDRI) (Armato et al., 2015) dataset will be used. It is a freely accessible dataset that contains described lesions from CT scans used in lung cancer diagnosis and screening.

## 2.2 *Societal Motivation*

In 2020, it was estimated that approximately 606,520 individuals would lose their lives to lung cancer in the United States alone (Siegel et al., 2020). Lung cancer remains the deadliest form of cancer, claiming more lives than breast, prostate, colorectal, and brain cancers combined (Siegel et al., 2020). These alarming statistics underscore the urgency of employing advanced technologies to refine Computer-Aided Diagnosis systems designed to support physicians in their diagnostic processes.

Medical imaging, mainly through CT scans, has become indispensable in modern clinical practice. CT scans provide detailed, three-dimensional visualizations of internal body structures, facilitating rapid and accurate diagnoses and informing treatment planning. Their capability to deliver high-resolution images swiftly enhances patient care by enabling timely medical interventions (Alanazi et al., 2024). However, interpreting these images remains mainly human-driven, which can introduce potential errors and delay analysis times (Alanazi et al., 2024).

Early detection of lung cancer is crucial, as it significantly improves overall survival rates and enhances progression-free survival and quality of life for patients (Saeed et al., 2024). Identifying lung cancer at an early stage allows for more effective treatment interventions, which can mitigate the risks of treatment-related severe adverse events commonly associated with late-stage diagnosis (Saeed et al., 2024). CT scans, in particular, serve as a powerful tool due to their ability to detect various internal

anomalies, including lung nodules, even when cancer is not the primary focus of the scan. This incidental detection capability enables healthcare providers to identify potentially malignant nodules during routine imaging, opening pathways for timely medical action. Such advancements in early diagnosis underscore the importance of integrating sophisticated detection technologies within CAD systems, ensuring that life-threatening conditions are managed as early as possible, thereby improving patient outcomes and overall survival.

Driven by the need to address these challenges, this study seeks to improve upon manual interpretation by constructing a pipeline designed to detect lung nodules. This objective will be pursued by leveraging the YOLO framework for early nodule detection.

## 2.3 *Scientific motivation*

Real-time object detection is a critical component across diverse applications, including autonomous driving, robotics, video surveillance, augmented reality, and healthcare. Within this domain, the YOLO framework has emerged as one of the top-performing algorithms, widely recognized for its impressive blend of speed and accuracy that enables rapid and precise object identification in images (Sapkota et al., 2024). Over time, the YOLO framework has undergone numerous enhancements, with each new version building upon the previous iterations to resolve challenges and elevate overall performance levels, which will be discussed in chapter 3.1.

A recent review of the YOLO architectures in computer vision (Sapkota et al., 2024) traces the development of the YOLO family from its initial release to its latest version, YOLOv10, highlighting the innovations and unique approaches introduced for image analysis. These advancements—focusing on improvements in speed, accuracy, and computational efficiency—have raised considerable expectations for CAD applications within the medical imaging field (Hendrix et al., 2023). Despite these developments, there has been limited exploration of YOLOv10's capabilities in specific areas of medical imaging, such as lung nodule detection, which this paper seeks to investigate in depth to assess its potential contributions to medical diagnosis.

This study focuses on advancing lung nodule detection by leveraging YOLOv10's capabilities and comparing its performance to the baseline approach, YOLOv5, which is a highly popular and frequently utilized object detection algorithm renowned for its exceptional detection speed and accuracy (Tong & Zhang, 2023). By evaluating YOLOv10's advanced detection features, such as its improved feature prioritization and enhanced backbone architecture with integrated attention mechanisms, this research aims to

assess its effectiveness in accurately identifying lung nodules. Comparing these two approaches will provide valuable insights into YOLOv10's potential for lung nodule detection, addressing a gap in the current literature and exploring its suitability for early lung cancer diagnosis.

## 2.4 *Research Question*

The interest in applying Deep Learning to automate processes in radiology has grown considerably, primarily due to its promising impact on medical imaging (Zhang & Qie, 2023). However, despite significant progress in DL methodologies, particularly in the YOLO family of models (Sapkota et al., 2024), their effectiveness for lung nodule detection still requires further exploration, especially with newer model iterations (X. Ma et al., 2024). This development, including its relevance and applications, will be further discussed in 3.1. The analysis of lung nodules presents unique challenges, including variations in nodule size, shape, and density and the inherent limitations of CT scan quality in routine clinical practice (Katase et al., 2022). These factors contribute to the complexity of accurately identifying nodules within lung tissue, underscoring the need to evaluate the performance of YOLO-based approaches in this critical application.

Consequently, the main research question formulated for this study is as follows:

**RQ: To what extent can YOLO models detect lung nodules?**

To delve deeper into these challenges, this thesis focuses on two critical aspects that may significantly influence the performance of DL models: the windowing technique applied during CT image preprocessing and the size of bounding boxes used for annotating lung nodules. The windowing technique affects the contrast and visibility of lung structures, which can impact the model's ability to detect nodules. Similarly, the bounding box size determines the context provided to the model for each nodule, influencing detection accuracy.

The windowing process in CT scans involves mapping the image's Hounsfield Unit (HU) values to a specific grayscale range to emphasize particular structures. By selecting appropriate windowing parameters, such as window width and window level, the visibility of lung tissues and nodules can be enhanced, facilitating better detection by Deep Learning models (Ain, 2021) as it will be further discussed in 3.3.

**SQ1: How does the choice of windowing parameters in CT image preprocessing affect the performance of YOLO models in detecting lung nodules?**

Bounding boxes play a crucial role in object detection models like YOLO by providing objects' spatial location and extent within an image. The size and scale of bounding boxes determine how much context around the lesion is retained and how the model interprets this information during training and inference. Adjusting these parameters can impact nodule detection accuracy, particularly when scaling or cropping affects the surrounding anatomical context.

**SQ2: How do the size and scaling of bounding boxes used for annotating lung nodules influence the accuracy of YOLO models in nodule detection?**

## 2.5 *Research Strategy*

This section outlines the methodological approach to address the **RQ** and associated subquestions posed in this study. The research strategy integrates model comparisons, systematic data handling to mitigate biases, and error analysis to identify potential areas for improvement in lung nodule detection.

### *Model Comparison*

The thesis evaluates the performance of two state-of-the-art object detection models: `YOLOv5` and `YOLOv10`. Specifically, three configurations of each model (small, medium, and large) are compared to assess the trade-offs between model complexity, computational efficiency, and detection accuracy. The focus is on understanding how architectural advancements in YOLOv10 influence its ability to detect lung nodules compared to YOLOv5.

### *Data Splitting to Avoid Biases*

To ensure unbiased evaluation and avoid data leakage, the dataset is split subject-wise. This method prevents images from the same patient from appearing in both the training and testing datasets, which could artificially inflate performance metrics. The data is divided into three subsets: training (70%), validation (20%), and testing (10%), with each patient assigned exclusively to one subset. This approach reflects real-world clinical scenarios where models encounter new patients during deployment.

*Error Analysis*

A detailed error analysis is conducted to gain insights into the model's limitations. This analysis focuses on identifying patterns in false positives and false negatives, such as:

- Whether smaller nodules are more prone to misclassification.
- The impact of nodule location, particularly for nodules near lung boundaries or in peripheral slices.

This analysis highlights areas where the model struggles and provides valuable information for future iterations and improvements.

## 2.6 *Summary of Findings*

The findings demonstrate the superiority of YOLOv10 over YOLOv5 in lung nodule detection. YOLOv10-l achieved the highest mAP@0.5 of 0.61, with improved precision 0.66 and recall 0.54, outperforming YOLOv5 across all configurations. Additionally, YOLOv10-l demonstrated faster inference times (6.8 ms) than YOLOv5-l (8.6 ms), making it more suitable for real-time applications.

Preprocessing techniques played a crucial role in enhancing detection performance. A windowing configuration with a center of $-600$ HU and a width of 1500 HU significantly improved the visibility of lung nodules. Bounding box adjustments, involving a 30% expansion and 5-pixel padding, further enhanced detection accuracy by incorporating contextual information around the nodules.

Error analysis highlighted specific challenges in lung nodule detection. Small nodules, nodules near lung boundaries, and peripheral slices were primarily responsible for false negatives, while false positives often resulted from the misclassification of blood vessels and image artifacts. These findings underscore the need for further refinements to address these limitations.

YOLOv10-l shows significant promise for accurate and efficient lung nodule detection. However, addressing detection challenges related to small nodules and boundary regions remains critical for further improving clinical applicability.

# 3 RELATED WORK

This section presents a structured review of existing research pertinent to lung nodule detection using advanced imaging and computational techniques. It begins with an overview of advancements in YOLO frameworks

- a family of real-time object detection models is discussed, highlighting their applicability to medical imaging tasks. The section then explores Hounsfield Units and windowing techniques, focusing on their significance in optimizing CT image analysis and enhancing the accuracy of detection tasks. Finally, different approaches to bounding box optimization are reviewed, emphasizing their impact on improving detection accuracy in AI models.

## 3.1 *YOLO Frameworks*

YOLO models were selected for this study due to their balance of speed and accuracy, making them well-suited for efficiently processing large volumes of CT scan data. Alternative frameworks, such as Faster R-CNN (Ren et al., 2016), Swin Transformers (Liu et al., 2021), and Vision Transformers (Dosovitskiy et al., 2021), offer unique strengths but were not included in this evaluation due to specific trade-offs.

Faster R-CNN is a highly accurate two-stage detection model, ideal for fine-grained localization tasks. However, its computational cost and slower processing times make it less practical for analyzing large-scale datasets (Ren et al., 2016). Swin Transformers leverage hierarchical attention mechanisms, excelling in capturing global context, but their complexity and resource requirements can hinder performance on high-resolution medical images (Yin et al., 2024). Vision Transformers provide state-of-the-art results in image analysis but often struggle with small-object detection, such as lung nodules, due to their coarse tokenization approach (Mkindu et al., 2023).

In contrast to these models, YOLO frameworks, or "You Only Look Once," revolutionized object detection by reinterpreting it as a regression task, allowing for real-time processing through a single pass over the image. Unlike conventional CNNs, which typically classify or segment objects after multiple processing stages, YOLO integrates object detection and recognition into a single, unified framework. This design enables the simultaneous identification of object boundaries and categories, achieving unparalleled speed and making YOLO a foundational framework in real-time object detection (Redmon et al., 2016). Building on this innovation, YOLOv2 introduced anchor boxes, enabling the model to handle objects of various sizes more effectively, thus improving detection performance (Redmon & Farhadi, 2016). This development established a benchmark for high-speed applications, making YOLOv2 a popular choice for tasks requiring efficient, real-time analysis, including emerging applications in medical imaging, such as lung nodule detection.

YOLOv3 further improved upon the YOLO framework by incorporating multi-scale detection, making it practical for detecting small objects through different layers in the network (Redmon & Farhadi, 2018). Building on this (Balaji et al., 2022), the MSF-YOLO model, an adaptation of YOLOv3, was developed to enhance the detection of small and fine tumors in MRI images. By integrating multi-scale feature fusion and optimizing anchor box sizes with K-means clustering, MSF-YOLO achieves greater contextual awareness and precision in identifying small tumors. Additionally, training was accelerated using weight parameters from the COCO (Lin et al., 2014) dataset, leading to substantial gains in detection accuracy. Experimental results showed that MSF-YOLO achieved an accuracy of 98.67% and 97.51% across two open datasets, outperforming traditional YOLOv3 and other models in mean Average Precision.

YOLOv4 then optimized the balance between speed and accuracy by implementing the CSPDarknet53 backbone alongside the Path Aggregation Network (PANet) for refined feature extraction and enhanced training via data augmentation and regularization techniques (Bochkovskiy et al., 2020). In a notable adaptation, Mei et al., 2021 tailored YOLOv4 for lung nodule detection by integrating DO-Conv layers for advanced feature extraction, CBAM attention modules to emphasize nodule locations, and a focal loss function to handle class imbalances in the data. The customized model, designed for the LIDC-IDRI dataset, achieved an Average Precision (AP) of 90.5% and processed each image in 0.04 seconds, demonstrating YOLOv4's applicability in high-performance medical detection tasks.

YOLOv5 brought modularity with multiple model sizes (e.g., YOLOv5n, YOLOv5s) tailored to different hardware requirements, broadening its use across varied applications (Hussain, 2024). Hendrix et al., 2023 developed a three-stage pipeline with YOLOv5 for lung nodule detection, involving lung localization to narrow the search area, detection of nodule candidates across CT slices, and a ResNet50-based false-positive reduction stage. This multi-step approach achieved 92% sensitivity at 1 false positive per scan using data from the LUng Nodule Analysis 2016 (LUNA16) dataset and additional hospital data, underscoring YOLOv5's versatility for layered detection workflows.

YOLOv6 introduced quantization and self-distillation techniques, which improved inference speed without sacrificing accuracy (Li et al., 2022). This made YOLOv6 particularly suitable for real-time medical applications where speed and efficiency are critical.

YOLOv7 further refined the architecture with the Extended Efficient Layer Aggregation Network for enhanced feature integration across layers, offering high precision and efficiency even on resource-constrained devices (C.-Y. Wang et al., 2022). The optimization of feature extraction without

added computational load allowed YOLOv7 to excel in high-resolution imaging tasks, making it applicable to detailed diagnostic tasks such as CT scans.

YOLOv8 introduced an anchor-free mechanism, advancing flexibility and accuracy across various detection scenarios (Yaseen, 2024). Z. Ma and Wu, 2024 adapted YOLOv8 for lung nodule localization by incorporating Receptive Fields Block (RFB) and Inner-IoU loss to fine-tune nodule positioning in CT scans. Tested on a dataset of 3009 images from Roboflow, these enhancements increased YOLOv8's precision by 3.4% and recall by 2%, demonstrating its efficacy for applications requiring precise localization.

YOLOv9 focused on a lightweight architecture optimized for mobile applications, maximizing detection with minimal computational demand (C.-Y. Wang et al., 2024). Chien et al., 2024 implemented YOLOv9 for detecting pediatric wrist fractures in X-rays, utilizing programmable gradient information (PGI) and a generalized efficient layer aggregation network (GELAN) for advanced feature extraction. In testing on the GRAZPEDWRI-DX dataset, YOLOv9-E achieved a mean average precision of 43.73%, surpassing YOLOv8+SA at 41.49%, highlighting YOLOv9's potential for precise imaging tasks requiring efficient resource usage.

YOLOv10 introduced refined CSPDarknet backbone with integrated convolutional attention mechanisms, a decoupled anchor-free detection head with dual-label assignment to eliminate Non-Maximum Suppression. (A. Wang et al., 2024). Despite these advancements, studies have yet to explore YOLOv10's specific application to lung nodule detection. Given its improved adaptability and robustness, YOLOv10 holds significant promise for medical imaging applications, particularly for lung nodule detection, where high accuracy and consistent performance across challenging imaging conditions are essential (Y. Wang, 2024). The lack of research in this area presents an opportunity to investigate YOLOv10's potential in detecting lung nodules, providing new insights into its effectiveness in clinical diagnostics.

## 3.2 *Identification of Differences Between the Baseline and Investigated YOLO Architecture*

YOLOv5 and YOLOv10 differ significantly in their architectural design and innovations (Hussain, 2024), which might impact their application in lung nodule detection. YOLOv5 employs a backbone based on CSPNet, incorporating multiple spatial pyramid pooling (SPP) blocks and a path aggregation network to optimize feature extraction and information flow (Jocher et al., 2022). In contrast, YOLOv10 introduces a refined CSPDarknet

backbone with convolutional attention mechanisms, enhancing precision and robustness, particularly for detecting small or occluded objects (A. Wang et al., 2024).

YOLOv10 eliminates non-maximum suppression by adopting a dual-label assignment system, balancing inference speed and accuracy, which is crucial for medical imaging tasks requiring real-time performance. Additionally, YOLOv10 incorporates advanced layers like partial self-attention (PSA) and spatially constrained (SC) layers, enabling improved feature refinement and higher adaptability to complex imaging conditions, such as varying nodule shapes and sizes (Sapkota et al., 2024).

While YOLOv5 has demonstrated strong performance in detection tasks, its anchor-based predictions may struggle with irregular or ambiguous nodule boundaries. In contrast, YOLOv10's anchor-free detection head and optimized training processes allow for better handling of these challenges, potentially leading to higher precision and recall in lung nodule detection. These advancements suggest that YOLOv10 could outperform YOLOv5 in both speed and accuracy, offering greater reliability in clinical diagnostics (Sapkota et al., 2024).

While YOLOv10 introduces advanced features that often enhance detection performance, studies indicate that newer architectures do not always guarantee better results across all tasks. For example, YOLOv10 excelled in tasks like kidney stone detection due to its superior computational efficiency and accuracy (Billah et al., 2024), yet it underperformed in steel surface defect detection compared to YOLOv9 (Huang et al., 2024). Similarly, a comparative study on knife safety detection highlights that YOLOv5, despite being an older model, delivered more balanced performance across multiple classes, achieving higher F1 scores in detecting specific objects, while YOLOv10 struggled with accuracy in certain classifications (Geetha & Hussain, 2024). These findings underscore the importance of aligning model selection with task requirements. Therefore, further investigation into YOLOv10's potential for lung nodule detection is necessary, as its advantages may depend heavily on the specific challenges and objectives of the application.

## 3.3 *Hounsfield units and windowing technique*

The invention of a method to reconstruct cross-sectional images from X-ray data laid the foundation for modern computed tomography by enabling the visualization of internal structures in the body with unprecedented detail (DenOtter & Schubert, 2023). HU values are crucial in medical imaging as they help differentiate between various types of tissues and fluids. For instance, fat typically has a negative HU value. In contrast, soft tissues

like muscle and organs have values closer to zero, and dense materials like bone can have values exceeding +1000 HU. This differentiation aids in diagnosing conditions and assessing the composition of lesions or tumors (DenOtter & Schubert, 2023).

The study conducted by Ain, 2021 showed that HU values improve medical imaging detection and segmentation by offering precise density data, enabling more accurate tissue differentiation during the segmentation process. In lung nodule detection, Hounsfield units allow for precise measurement and identification of nodules based on their density, which is often higher than that of the lung tissue itself (Youssif et al., 2011). This ability to differentiate tissues based on HU is essential in medical imaging and diagnosis, providing a reliable method for detecting abnormalities like lung nodules (Gould et al., 2015).

Abboud and Kadoury, 2023 study demonstrates that varying Hounsfield Unit window levels significantly improves the accuracy of deep learning models for liver lesion segmentation by optimizing contrast in the input data. By systematically adjusting HU levels during training and testing, the authors observed a notable enhancement in segmentation accuracy. They also found that varying HU levels allowed the model to generalize better across different contrasts and provided a practical way to estimate uncertainty, improving its reliability in clinical applications.

The study by Viriyavisuthisakul et al., 2023 thoroughly evaluated 45 different fixed Hounsfield Unit windowing strategies applied to non-contrast cranial CT images, focusing on optimizing HU values to improve diagnostic accuracy for ischemic stroke classification. By systematically adjusting HU window settings, the researchers aimed to enhance the ability of deep learning models to distinguish between hyperacute, acute, and normal brain conditions. Testing these strategies across 15 state-of-the-art models revealed that certain HU windows significantly boosted classification accuracy, underscoring the impact of appropriate HU value selection on model performance. This research highlights the critical role of HU windowing in medical imaging preprocessing, offering clinicians insights into optimizing HU settings to enhance diagnostic precision in stroke detection.

In a study by Karki et al., 2020, a distant-supervised method was proposed to automatically determine optimal window settings for preprocessing radiographic images, a crucial step for detecting abnormalities. Adaptive window settings were predicted by integrating a window estimator module with a DCNN-based lesion classifier, and the top four settings across the dataset were identified. Images were scaled based on these settings, and lesion classifiers were trained on each. Testing different combinations showed that using and combining predictions from the top

four window settings significantly improved the detection of intracranial hemorrhage in brain CT images.

## 3.4 *Bounding Boxes in Object Detection*

Bounding boxes play a pivotal role in object detection frameworks, serving as the foundational mechanism for annotating and localizing objects within an image. Their importance is amplified in medical imaging tasks, such as lung nodule detection, where accurate localization directly impacts diagnostic accuracy and subsequent clinical decisions. Recent research by Madan et al., 2023 highlights the impact of bounding box size and placement on the performance of anchor-based models, providing insights that are particularly relevant for single-class detection tasks like lung nodules.

The study by Madan et al., 2023 demonstrated that employing fixed-size bounding boxes significantly improved the performance of object detection models. In experiments using both real-world and synthetic datasets, fixed bounding boxes enhanced mean Average Precision by approximately 50% compared to randomly sized bounding boxes. An additional observation from the study was the improvement in object localization accuracy when fixed bounding boxes were used. For instance, YOLO Scaled v4 performed better on a synthetic dataset of circular shapes when using fixed bounding boxes compared to those adapted to object contours or randomly drawn.

A alternative approach was proposed by Zhong et al., 2020, focusing on dynamic optimization of anchor boxes. Their method enables anchor shapes to adapt automatically to data distribution and the learning process of the network by treating anchor shapes as trainable variables. Through backpropagation, the method minimizes localization errors, effectively learning optimal anchor shapes during training without adding computational overhead during inference. This approach improved detection accuracy by over 1% mean Average Precision on benchmark datasets such as Pascal VOC, MS COCO, and Brainwash. Furthermore, their method proved robust to variations in anchor shape initialization, greatly simplifying the design process.

## 3.5 *Literature Gap*

While the YOLO framework has seen extensive research and development in general object detection tasks, more studies are needed to evaluate its latest iterations, such as YOLOv10, in medical imaging contexts (Y. Wang, 2024). Specifically, the application of YOLOv10 for detecting lung nodules has not been systematically investigated, despite its potential to enhance

the detection of small and irregularly shaped objects like lung nodules. As Y. Wang (2024) emphasizes, "it is necessary to apply the YOLOv10 model in this lung nodule detection task for more exploration and research, which may further improve the clinical accuracy and reliability of deep learning methods in X-ray lung nodule detection tasks."

Additionally, while preprocessing techniques, such as windowing (Ain, 2021) and bounding box adjustments (Zhong et al., 2020), are critical for optimizing medical image analysis, their effects on YOLOv10's performance for lung nodule detection remain unexplored. This thesis addresses these gaps by evaluating YOLOv10 compared to YOLOv5, analyzing its performance on lung CT scans, and assessing the impact of domain-specific preprocessing.

By focusing on YOLOv10 and its potential improvements for lung nodule detection, this study advances the application of state-of-the-art object detection frameworks in medical imaging.

## 4 METHODOLOGY & EXPERIMENTAL SETUP

As depicted in Figure 1, the research process follows a structured pipeline to ensure a robust and unbiased evaluation of lung nodule detection models. The pipeline begins with loading and preprocessing CT scans, where images are normalized and windowed using Hounsfield Units to enhance nodule visibility. Preprocessed images are resized and formatted into YOLO-compatible labels with bounding box annotations, then split into training, validation, and test subsets to prevent data leakage. Data augmentation is applied to the training set to improve model generalization. YOLO models are trained with optimized hyperparameters to detect lung nodules, focusing on reducing false positives and negatives. Performance is evaluated using metrics like Mean Average Precision, Precision, and Recall, ensuring consistent and clinically relevant outcomes.

### 4.1 *Dataset description*

The dataset used in this study is the Lung Image Database Consortium image collection (LIDC-IDRI) (Armato et al., 2015), a comprehensive resource for lung nodule analysis. The dataset contains anonymized thoracic CT scans from multiple academic centers and medical imaging companies. Each case includes DICOM-format CT scans and corresponding annotations provided by four thoracic radiologists.

Annotations were created through a two-phase process. In the blinded-read phase, each radiologist independently marked nodules and categorized them based on size and significance. The subsequent unblinded-

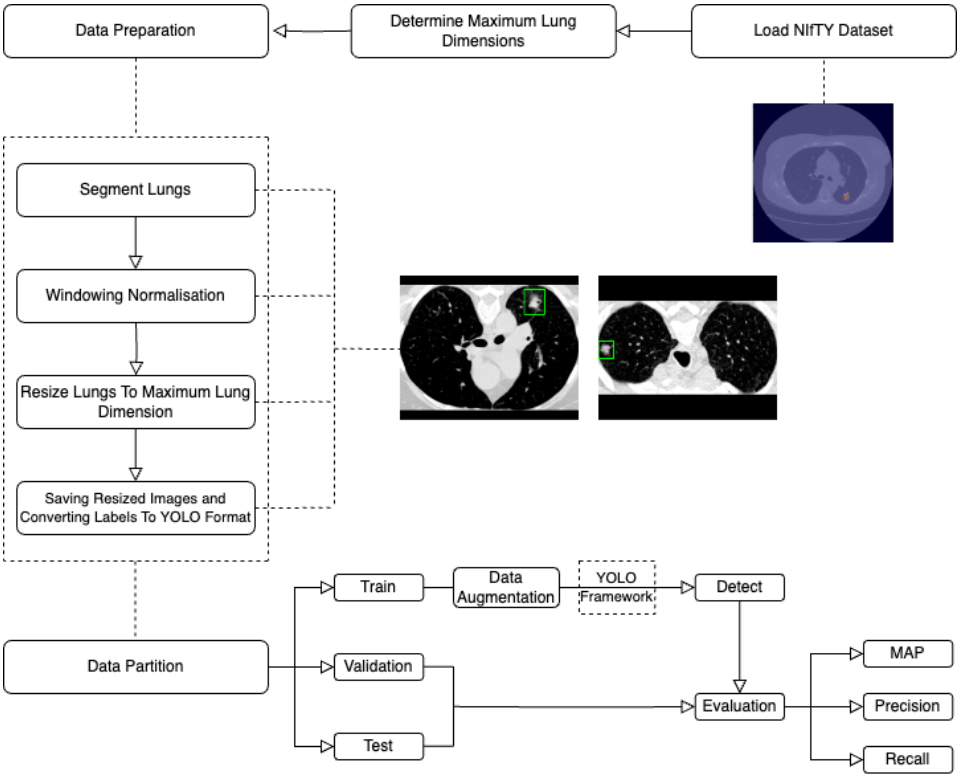


Figure 1: The diagram represents the workflow of the implemented methodology. The process begins with data loading and determining maximum lung dimensions for future steps. Next step is data preparation involving normalization and the application of specific windowing techniques to standardize the image data. The data is then resized to the maximum lung dimensions, ensuring consistency across samples, and reformatted to YOLO-compatible labels. Subsequently, the dataset is partitioned into training, validation, and test subsets. Data augmentation is performed on the training set. The YOLO framework is employed to train a detection model on the training data, while performance evaluation is conducted on the validation and test sets. Detection of lung nodules is carried out, and model performance is quantified using metrics such as the Mean Average Precision.

read phase allowed radiologists to refine their annotations after reviewing anonymized marks from their peers. This process ensured high-quality, non-consensus annotations of nodules.

All CT scans in this dataset are in DICOM format and are accompanied by XML annotation files. These files detail the location and size of each nodule, providing critical information for computer-assisted diagnosis and DL applications.

Table 1 summarizes key statistics of the LIDC-IDRI dataset.

Table 1: Summary of LIDC-IDRI Dataset

| | |
|---|---|
| Number of Patients | 1,018 |
| Total Nodules | 7,371 |
| Number of Nodules $\geq$ 3 mm | 2,669 |
| Number of Nodules $<$ 3 mm | 4,702 |

## 4.2 *Data preprocessing*

### 4.2.1 *Lung segmentation and cropping*

Lung segmentation was employed using the 'R231' model (Hofmanninger et al., 2020) to isolate lung regions within computed tomography images. This model is specifically designed for lung segmentation tasks and facilitates accurate delineation of lung boundaries in volumetric medical imaging data.

Initially, the maximum lung dimensions across all images in the dataset were determined. This process involved applying the segmentation model to each three-dimensional NIfTI image and calculating the maximum width and height of the lung regions detected. A standardized size was established for subsequent image processing steps by identifying these maximum dimensions.

With the maximum dimensions defined, each image underwent a series of preprocessing operations. The segmentation model was applied to extract the lung regions from the CT images. Slices that contained no lung tissue or exhibited only one lung were excluded to maintain consistency and focus on bilateral lung representations. This ensured that the dataset comprised images with complete and comparable anatomical structures.

For the remaining slices, the images were cropped to the bounding box encompassing the lung regions. This cropping was based on the segmentation masks generated, capturing the precise area of interest. The cropped images were then resized to match the previously determined maximum lung dimensions. To preserve the anatomical proportions, the

resizing maintained the original aspect ratio of the lung regions. If necessary, padding was added to align the images to the standardized size, ensuring uniformity across the dataset.

### 4.2.2 *Windowing and Normalization*

In the image preprocessing pipeline, windowing and normalization were applied to enhance the visibility of lung structures in the CT images. The windowing process adjusted the intensity values based on specified Hounsfield Unit parameters, followed by normalization to standardize the pixel intensity range.

The window boundaries were determined using the window center $C$ and window width $W$ as follows:

$$W_{\text{min}} = C - \frac{W}{2} \tag{1}$$

$$W_{\text{max}} = C + \frac{W}{2} \tag{2}$$

Each pixel intensity $I$ in the cropped image slice was then windowed using the following function:

$$I_{\text{windowed}} = \begin{cases} W_{\text{min}}, & \text{if } I < W_{\text{min}} \\ I, & \text{if } W_{\text{min}} \leq I \leq W_{\text{max}} \\ W_{\text{max}}, & \text{if } I > W_{\text{max}} \end{cases} \tag{3}$$

Following windowing, normalization was performed to scale the pixel intensities to an 8-bit grayscale range of 0 to 255:

$$I_{\text{normalized}} = \left( \frac{I_{\text{windowed}} - W_{\text{min}}}{W_{\text{max}} - W_{\text{min}}} \right) \times 255 \tag{4}$$

This linear transformation ensured that all images had consistent intensity levels, facilitating effective training and analysis in subsequent modeling tasks.

An evaluation was conducted to assess the impact of different window levels on the model's performance, measured by the Mean Average Precision. The results are summarized in Table 5.

### 4.2.3 *Bounding Box Adjustments*

After the windowing and normalization steps, the images were saved and bounding boxes were created to annotate regions containing pulmonary nodules. The process of creating bounding boxes involved several computational steps to accurately represent the location and size of the nodules within each image slice.

*Bounding Box Creation*

To identify the nodules in each label slice, connected component analysis was performed. Each connected component corresponds to a potential nodule. For each connected component $i$, the initial bounding box coordinates were determined based on the pixel positions:

$$x_{\text{min}}^{(i)} = \min\{x_k^{(i)}\} \tag{5}$$
$$x_{\text{max}}^{(i)} = \max\{x_k^{(i)}\} \tag{6}$$
$$y_{\text{min}}^{(i)} = \min\{y_k^{(i)}\} \tag{7}$$
$$y_{\text{max}}^{(i)} = \max\{y_k^{(i)}\} \tag{8}$$

where $(x_k^{(i)}, y_k^{(i)})$ are the coordinates of the pixels belonging to the $i$-th connected component.

To ensure that the bounding boxes fully encompassed the nodules and provided additional context, the width and height were increased by an appropriate scaling factor $e$ for each iteration. The expansion factor $e$ was adjusted as a parameter during experimentation to optimize detection performance. The adjusted dimensions were calculated as:

$$w^{(i)} = x_{\text{max}}^{(i)} - x_{\text{min}}^{(i)} \tag{9}$$
$$h^{(i)} = y_{\text{max}}^{(i)} - y_{\text{min}}^{(i)} \tag{10}$$
$$w_{\text{expanded}}^{(i)} = e \times w^{(i)} \tag{11}$$
$$h_{\text{expanded}}^{(i)} = e \times h^{(i)} \tag{12}$$

The center of each bounding box was determined to facilitate the expansion:

$$x_{\text{center}}^{(i)} = \frac{x_{\text{min}}^{(i)} + x_{\text{max}}^{(i)}}{2} \tag{13}$$
$$y_{\text{center}}^{(i)} = \frac{y_{\text{min}}^{(i)} + y_{\text{max}}^{(i)}}{2} \tag{14}$$

Using the expanded dimensions and the center coordinates, the new bounding box coordinates were calculated:

$$x_{\text{min,new}}^{(i)} = x_{\text{center}}^{(i)} - \frac{w_{\text{expanded}}^{(i)}}{2} - p \tag{15}$$

$$x_{\text{max,new}}^{(i)} = x_{\text{center}}^{(i)} + \frac{w_{\text{expanded}}^{(i)}}{2} + p \tag{16}$$

$$y_{\text{min,new}}^{(i)} = y_{\text{center}}^{(i)} - \frac{h_{\text{expanded}}^{(i)}}{2} - p \tag{17}$$

$$y_{\text{max,new}}^{(i)} = y_{\text{center}}^{(i)} + \frac{h_{\text{expanded}}^{(i)}}{2} + p \tag{18}$$

where $p$ represents the padding added to each side of the bounding box. The padding $p$ was also adjusted as a parameter in each iteration to find the optimal value for detection performance.

An evaluation was conducted to assess the impact of different expansion factors and padding values on the model's performance, measured by the Mean Average Precision. The results are summarized in Table 4.

*Bounding Box Merging*

To account for cases where bounding boxes were overlapping or in close proximity, a merging algorithm was applied. Bounding boxes were merged if the distance between them was less than a specified threshold $d_{\text{max}}$. The overlap condition was evaluated as:

$$\text{distance}(B^{(i)}, B^{(j)}) < d_{\text{max}} \tag{19}$$

where $\text{distance}(B^{(i)}, B^{(j)})$ represents the minimum distance between bounding boxes $B^{(i)}$ and $B^{(j)}$. Merged bounding boxes were recalculated to encompass all constituent boxes:

$$x_{\text{min,merged}} = \min\left(x_{\text{min,new}}^{(i)}, x_{\text{min,new}}^{(j)}\right) \tag{20}$$

$$x_{\text{max,merged}} = \max\left(x_{\text{max,new}}^{(i)}, x_{\text{max,new}}^{(j)}\right) \tag{21}$$

$$y_{\text{min,merged}} = \min\left(y_{\text{min,new}}^{(i)}, y_{\text{min,new}}^{(j)}\right) \tag{22}$$

$$y_{\text{max,merged}} = \max\left(y_{\text{max,new}}^{(i)}, y_{\text{max,new}}^{(j)}\right) \tag{23}$$

This ensured that closely situated nodules were annotated as a single detection target, which can be beneficial for certain object detection models.

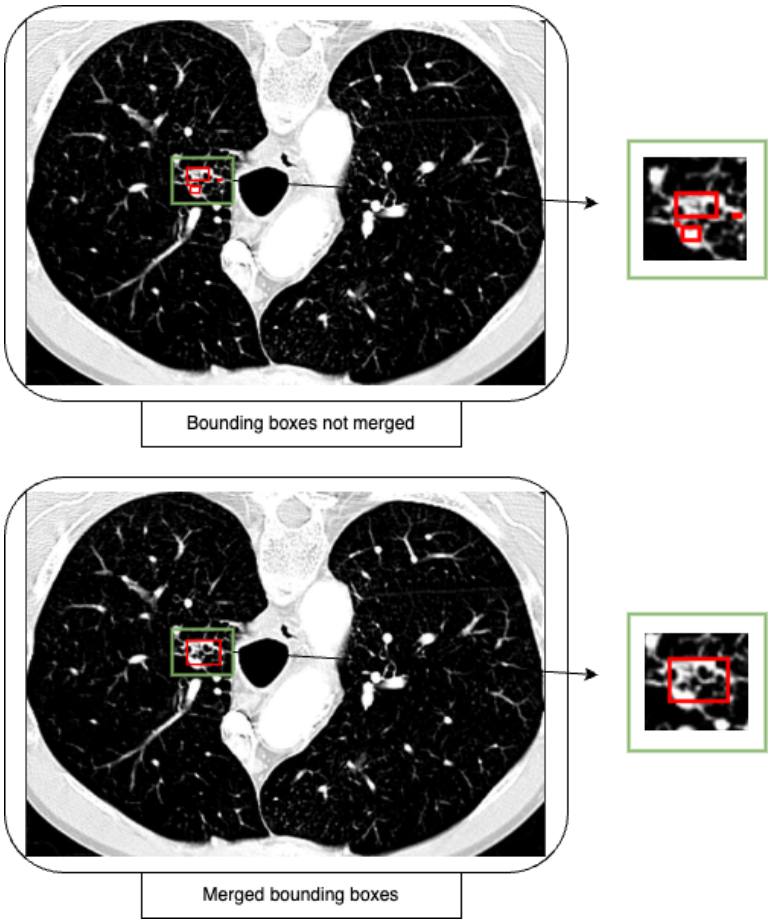


Figure 2: The process of merging bounding boxes is illustrated. The top panel shows bounding boxes before merging, where multiple boxes are present. The bottom panel demonstrates the result after applying the merging algorithm, where closely situated bounding boxes are combined into a single bounding box. This step is crucial for annotating closely located nodules as a unified detection target.

### 4.2.4 *Data Splitting and Augmentation*

The dataset was divided into training (70% of patients), validation (20% of patients), and testing (10% of patients) subsets using patient-wise splitting to prevent data leakage and ensure generalizability. This approach, crucial in medical imaging, ensures unbiased evaluation by segregating patients entirely.

Data augmentation balanced the training dataset by increasing nodule-containing images through geometric transformations such as rotation and zooming. Images were rotated by 19° increments and zoomed starting at 1.25, with bounding boxes updated accordingly. Slices without lung tissue

were excluded. Augmentation continued iteratively until a balanced class distribution was achieved, enhancing the model's detection capability.

## 4.3 *Models*

### 4.3.1 *YOLOv5*

As a baseline model, YOLOv5 was selected due to its balance between accuracy and inference speed. Implemented in PyTorch, YOLOv5 is a well-documented object detection framework offering multiple model variants (small, medium, large, and extra-large) to accommodate different computational requirements. In this study, three configurations of YOLOv5 were evaluated: `YOLOv5s`, `YOLOv5m`, and `YOLOv5l`, which contain 7.2M, 21.2M, and 46.5M parameters, respectively. These configurations were used to assess the trade-offs between detection performance and computational cost.

YOLOv5 incorporates Cross-Stage Partial (CSP) layers and a Path Aggregation Network (Hussain, 2024) for feature extraction and aggregation. The CSP layers enhance learning efficiency by partitioning feature maps for improved gradient flow, while PANet facilitates multi-scale feature fusion. These architectural features make YOLOv5 well-suited for detecting small objects, such as lung nodules, which present unique challenges due to their small size and irregular shapes.

The model was initialized with pre-trained weights from the COCO dataset (Lin et al., 2014) and fine-tuned on a domain-specific dataset of annotated CT lung scans. This dataset poses challenges due to the high spatial variability of lung nodules and their overlap with surrounding anatomical structures. Adapting YOLOv5 to this domain serves as a robust baseline for evaluating architectural advancements introduced in YOLOv10.

### 4.3.2 *YOLOv10*

The performance of YOLOv10 was evaluated as an enhanced object detection framework for detecting lung nodules in CT scans. YOLOv10 introduces several architectural innovations to improve small-object detection, computational efficiency, and scalability. These enhancements include a refined CSPDarknet backbone with integrated convolutional attention mechanisms, a decoupled anchor-free detection head with dual-label assignment to eliminate Non-Maximum Suppression (A. Wang et al., 2024).

Three configurations of YOLOv10 were evaluated: `YOLOv10-s`, `YOLOv10-m`, and `YOLOv10-l`, containing approximately 7.2, 15.4M, and 24.4M parame-

ters, respectively. These configurations align with their YOLOv5 counterparts, enabling a fair comparison of the two architectures.

Both YOLOv5 and YOLOv10 were trained using identical datasets and augmentation strategies to maintain consistency in evaluation. However, YOLOv10's improved feature prioritization and enhanced backbone architecture with integrated attention mechanisms are designed to address the limitations observed in YOLOv5 when detecting small, irregularly shaped objects. These advancements make YOLOv10 a promising alternative for lung nodule detection, a task with significant implications for early-stage lung cancer diagnosis.

### 4.3.3 *Evaluation Metrics*

The model's performance is evaluated using three key metrics: mean Average Precision, Precision, and Recall. These metrics are particularly relevant in medical image analysis, especially for tasks such as nodule detection, where accurate identification of pathological features is essential.

**Mean Average Precision (mAP)** provides a comprehensive assessment of model performance by combining Precision and Recall across various confidence thresholds. It is calculated as the area under the Precision-Recall curve, summarizing the model's effectiveness at different confidence levels. Specifically, mAP is the average of the Precision values at different recall thresholds and Intersection over Union (IoU) thresholds. The formula for mAP is as follows:

$$\text{mAP} = \frac{1}{Q} \sum_{q=1}^{Q} \text{AP}_q \tag{24}$$

where $Q$ is the number of queries or classes, and $\text{AP}_q$ represents the Average Precision for each class $q$, averaged across all relevant recall levels.

**Precision** measures the model's ability to avoid false alarms by determining the proportion of correct detections (true positives, *TP*) among all positive predictions. A true positive (*TP*) occurs when the model correctly identifies an object, and the predicted bounding box overlaps with the ground truth bounding box by more than a predefined threshold, commonly set at 0.5 for Intersection over Union (IoU). Conversely, a false positive (*FP*) occurs when the model predicts an object that does not correspond to any actual object in the dataset. High precision indicates that the model generates accurate predictions with minimal false alarms, which is crucial in ensuring the reliability of the detection system.

**Recall**, on the other hand, evaluates the model's ability to detect all actual instances of a given class, focusing on minimizing missed detections. A true positive (*TP*) for recall represents a correctly identified object, while

a false negative (*FN*) occurs when the model fails to detect an existing object in the dataset. High recall signifies that the model successfully identifies the majority of relevant instances, which is essential for tasks where missing an object can have significant consequences.

### 4.4 *Model Implementation*

This section describes the implementation of YOLO architectures optimized for lung nodule detection. The models used include YOLOv5s, YOLOv5m, YOLOv5l, and their counterparts in YOLOv10. These architectures were evaluated under various training configurations to determine their effectiveness.

The training process utilized key hyperparameters, including a batch size of 32, an input image size of 502 pixels, which corresponded to the maximum lung size detected in earlier steps, and a total of 100 training epochs. Early stopping was employed with a patience of 20 epochs to terminate training when performance stabilized, minimizing unnecessary computations and reducing overfitting.

The training configuration employed YOLO pre-trained weights trained on the COCO dataset. Built-in data augmentation techniques were applied, including combining multiple images, flipping, and adjusting contrast levels, to enhance the model's generalization capabilities. Despite efforts to improve accuracy through modifications such as learning rate adjustments, changes to the optimizer, alterations in early stopping criteria, adjustments to batch size, and the removal of YOLO's inherent data augmentation, these experiments did not yield better results and were excluded from further analysis.

Overall, the models demonstrated robust performance for lung nodule detection under the described configuration.

**Model Parameters**

| **Parameter** | **Setting** |
|---|---|
| Epochs | 100 |
| Batch Size | 32 |
| Image Size | 502 pixels |
| Early Stopping | Patience 20 epochs |
| Data Augmentation | Scaling, Flipping, Mosaic |
| Optimizer | SGD |

## 5 RESULTS

This section presents the outcomes of the experiments conducted to evaluate the effectiveness of YOLO models in detecting lung nodules from CT

images. The experiments were designed to address the main research question and the sub-questions posed in the study. The results are organized to reflect the impact of different preprocessing techniques, the comparison of model performances, error analysis, and the investigation of model generalization.

## 5.1 *Model Comparison*

Table 2 summarizes the performance metrics for YOLOv5 and YOLOv10 models across their small (s), medium (m), and large (l) configurations. The results clearly demonstrate the superiority of YOLOv10 models in detecting lung nodules, addressing the central research question: **To what extent can YOLO models detect lung nodules?**

Table 2: Performance Comparison of YOLOv5 and YOLOv10 Models

| Model | Params (M) | mAP@0.5 (Validation) | Precision (Validation) | Recall (Validation) | mAP@0.5 (Test) | Precision (Test) | Recall (Test) |
|---|---|---|---|---|---|---|---|
| YOLOv5s | 7.2 | 0.48 | 0.54 | 0.49 | 0.49 | 0.58 | 0.50 |
| YOLOv5m | 21.2 | 0.51 | 0.54 | 0.53 | 0.52 | 0.59 | 0.52 |
| YOLOv5l | 46.6 | 0.54 | 0.61 | 0.55 | 0.54 | 0.62 | 0.53 |
| YOLOv10-s | 7.2 | 0.54 | 0.62 | 0.49 | 0.56 | 0.66 | 0.52 |
| YOLOv10-m | 15.4 | 0.55 | 0.63 | 0.50 | 0.57 | 0.64 | 0.53 |
| YOLOv10-l | 24.4 | **0.59** | **0.65** | **0.54** | **0.61** | **0.66** | **0.54** |

The results reveal that YOLOv10 consistently outperformed YOLOv5 in lung nodule detection. YOLOv10-s achieved an mAP@0.5 of 0.56, surpassing YOLOv5-l's 0.54. This demonstrates that even the smallest YOLOv10 model exceeds the performance of the largest YOLOv5 model, highlighting the impact of YOLOv10's architectural improvements.

The highest-performing YOLOv10-l achieved an mAP@0.5 of 0.61, precision of 0.66, and recall of 0.54. This indicates that 66% of detected nodule candidates were true positives (34% were false positives), while only slightly more than half of all nodules were successfully detected. Studies from the literature, however, report mAP values exceeding 0.9, achieved through the use of advanced techniques and preprocessing (Hendrix et al., 2023; Mei et al., 2021). This comparison underscores that the overall performance of the YOLOv10 model remains relatively low compared to state-of-the-art methods.

It should also be noted that no post-processing steps were applied in this study. Post-processing techniques, used in related studies (Hendrix et al., 2023; Mei et al., 2021), are critical for refining the predictions and improving detection accuracy. Without these steps, the results of this study are not directly comparable to other approaches that employ such enhancements.

However, these findings confirm YOLOv10's superiority over YOLOv5 in this task. It is important to note that YOLOv10's performance does not

consistently surpass other YOLO versions across all detection tasks, as its effectiveness depends on the specific application and characteristics of the dataset (Huang et al., 2024). This variability underscores the importance of tailoring models to task-specific requirements, particularly in medical imaging, where precision and reliability are critical.

## 5.2 *Effect of Windowing Parameters on Model Performance*

To explore the sub-question **SQ1: How does the choice of windowing parameters in CT image preprocessing affect the performance of YOLO models in detecting lung nodules?**, several windowing configurations were evaluated during preprocessing. Windowing adjusts grayscale intensity values to enhance the contrast and visibility of specific tissues, making it particularly critical for emphasizing lung nodules within CT scans.

Three different windowing settings were tested, as shown in Table 3. Each model was trained using the YOLOv10l architecture with identical training parameters to ensure a fair comparison.

Table 3: Windowing Parameters Used in Experiments

| Window Center (HU) | Window Width (HU) | Description |
|---|---|---|
| no window center | no window width | CT scans not changed |
| −600 | 1500 | Lung Window |
| −350 | 1000 | Narrow Window |

The performance of the models was evaluated using the Mean Average Precision at IoU threshold 0.5. The results are presented in Table 4.

Table 4: Model Performance with Different Windowing Parameters

| Window Center (HU) | Window Width (HU) | mAP@0.5 | Precision | Recall |
|---|---|---|---|---|
| no window center | no window width | 0.55 | 0.59 | 0.51 |
| −600 | 1500 | **0.61** | **0.66** | **0.54** |
| −350 | 1000 | 0.57 | 0.63 | 0.49 |

As shown in Table 4, the windowing setting with a center of −600 HU and a width of 1500 HU yielded the highest mAP@0.5 of 0.61. This indicates that adjusting the windowing parameters to enhance contrast improves the model's ability to detect nodules.


Figure 3: Effect of windowing parameters on CT image visualization. Each panel represents a CT image processed with different windowing configurations: **Top:** No windowing applied; **Middle:** Window Center $-600$ HU, Window Width 1500 HU, emphasizing lung tissues; **Bottom:** Window Center $-350$ HU, Window Width 1000 HU. The red bounding box highlights the lung nodule, demonstrating how different windowing settings enhance or obscure the visibility of nodules and surrounding tissues.

## 5.3 *Impact of Bounding Box Size and Scaling on Detection Accuracy*

To address the sub-question **SQ2: How does the size and scaling of bounding boxes used for annotating lung nodules influence the accuracy of YOLO models in nodule detection?**, experiments were carried out to evaluate the effects of varying bounding box expansion factors and padding values on detection accuracy. The initial bounding boxes were systematically expanded and padded to incorporate additional contextual information around the nodules, aiming to enhance the model's ability to accurately differentiate nodules from surrounding anatomical structures.

Table 5 shows the mAP@0.5 scores for various combinations of expansion factors and padding.

Table 5: Effect of Bounding Box Expansion and Padding on Model Performance

| **Expansion Factor (%)** | **Padding (Pixels)** | **mAP@0.5** | **Precision** | **Recall** |
|---|---|---|---|---|
| 0 | 0 | 0.41 | 0.46 | 0.34 |
| 20 | 5 | 0.54 | 0.59 | 0.47 |
| 30 | 5 | **0.61** | **0.66** | **0.54** |
| 40 | 5 | 0.48 | 0.50 | 0.44 |

The combination of a 30% expansion factor and 5 pixels of padding achieved the highest mAP@0.5 of 0.61. Increasing the bounding box size up to this point provided sufficient contextual information without introducing excessive background noise.

These results suggest that appropriately scaling bounding boxes enhances the model's ability to detect nodules by including surrounding anatomical features that may aid in recognition.

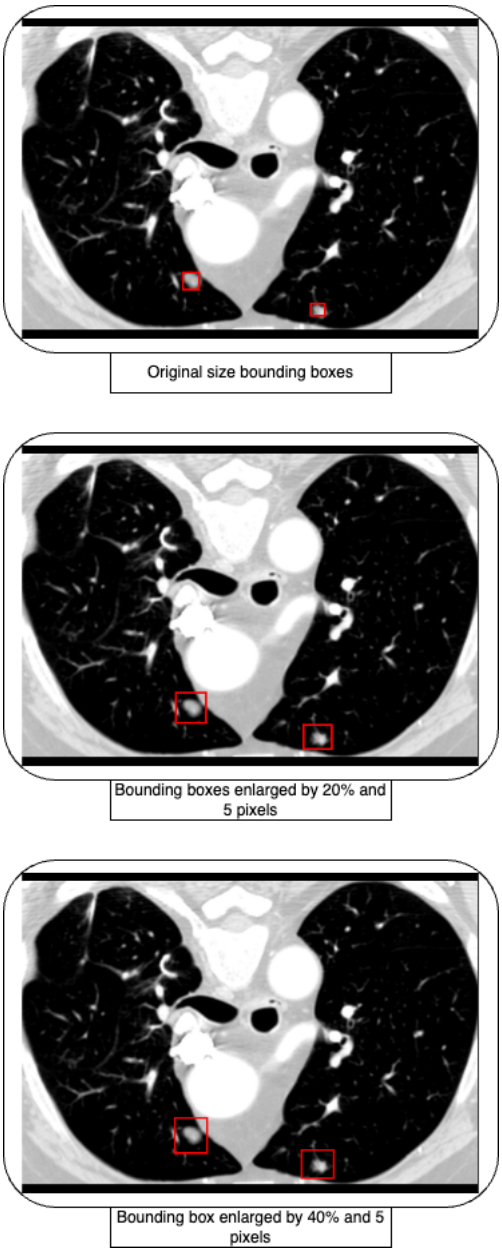


Figure 4: Visualization of bounding box expansion and padding effects on CT images. **Top:** Bounding box of normal size. **Middle:** Bounding box enlarged by 20% and 5 pixels. **Bottom:** Bounding box enlarged by 40% and 5 pixels. The red boxes highlight the expanded regions around the nodules, demonstrating how different configurations influence contextual information available to the model.

## 5.4 *Detailed Error Analysis*

An error analysis was conducted to identify and quantify the types of errors made by the models. This analysis focused on the YOLOv10-l model due to its superior performance.

### 5.4.1 *Error Types*

The errors were categorized into two main types:

**False Positives (FPs)** refer to non-nodule structures incorrectly identified as nodules. An analysis of the FPs revealed that the model often misclassified various structures or features that shared visual similarities with nodules. These misclassifications could involve anatomical elements or image irregularities, such as noise or artifacts, which contributed to the model's errors.

**False Negatives (FNs)** represent actual nodules that the model failed to detect. The analysis of these errors highlighted several challenges. Small nodules were often missed due to their subtle appearance in the scans. Additionally, nodules located near the lung boundaries posed detection difficulties, as their edges could blend with surrounding structures. Lastly, nodules appearing on peripheral slices of scans were more likely to be overlooked, as they were partially represented or less pronounced in those regions.

### 5.4.2 *Visualization of Errors*

To better understand the errors identified during model evaluation, visualizations were created for representative examples of both false positives and false negatives. These visualizations highlight key error patterns and provide insights into potential areas for improvement in the model. Below are examples corresponding to the different types of errors.

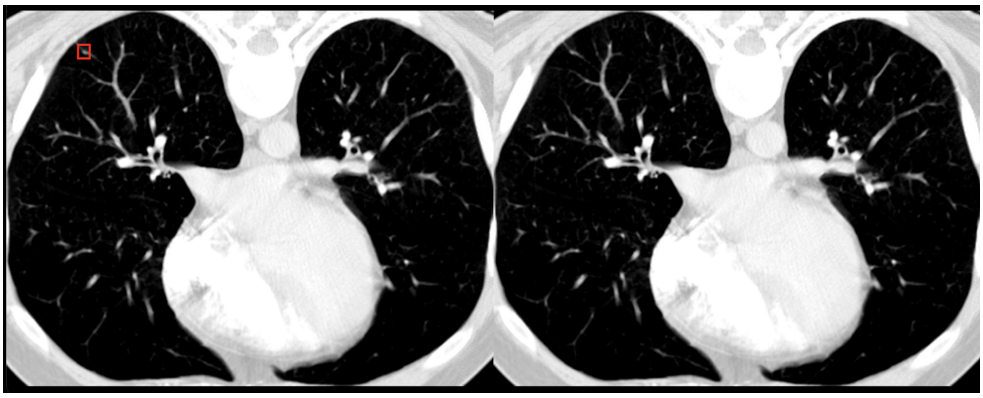

Figure 5: **False Negatives: Small Nodules.** This figure shows example of small nodule that were not detected by the model. The reduced visibility of these nodules due to their small size likely contributed to their misclassification. Notably, the small nodule visible in the upper left corner of the left scan was not detected by the model, as evidenced by its absence in the corresponding right scan.

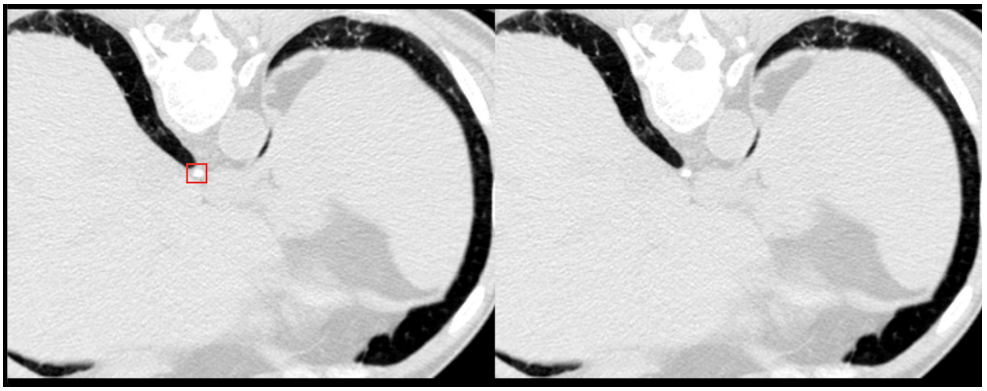

Figure 6: **False Negatives: Nodules Near Lung Boundaries.** This figure shows an example of a nodule located close to the lung boundaries that was missed by the model. Its location may complicate detection due to reduced contrast with surrounding structures. The nodule situated in the central region of the left scan is not detected by the model, as reflected in its absence in the corresponding right scan.

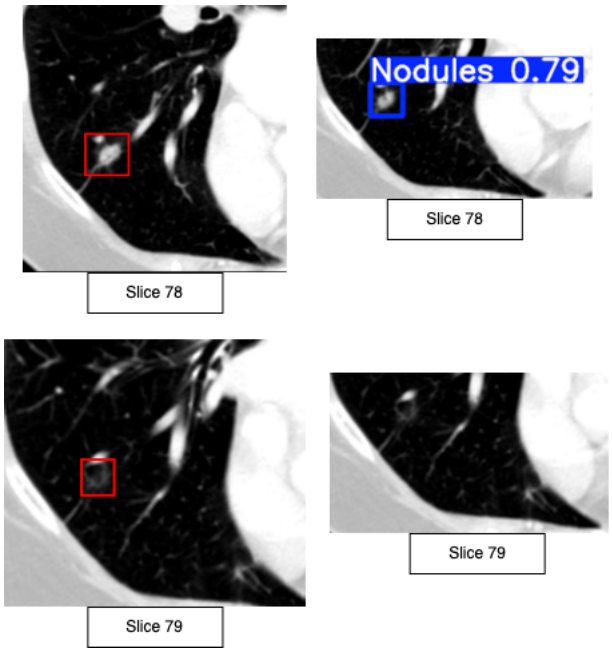


Figure 7: **False Negatives: Peripheral Slices of Nodules.** This visualization highlights errors occurring on the peripheral slices of nodules spanning multiple slices. The smaller portions of the nodule in these slices likely made detection more challenging. The top row displays slice 78 on the left and the corresponding model prediction on the right, while the bottom row shows the peripheral slice (in this case the last slice containing the nodule) on the left and the model's prediction on the right, where the nodule remains undetected.

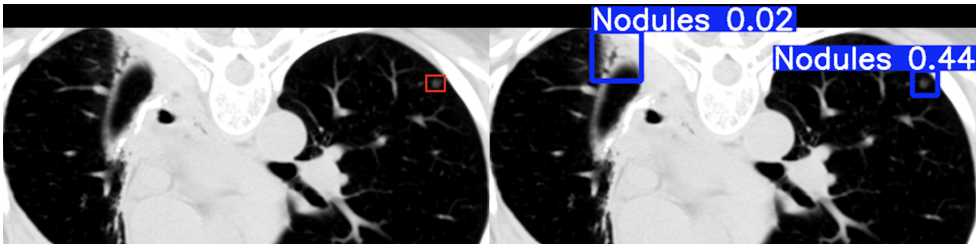


Figure 8: **False Positives: Misclassification of Other Structures.** Examples of non-nodule structures, such as blood vessels or image artifacts, that were incorrectly identified as nodules by the model. These errors highlight the need for improved differentiation of nodules from other anatomical structures. In the figure, the left scan shows a correctly identified nodule on the right side, while the right scan displays a false positive on the left side, where a structure that should not have been classified as a nodule was detected.

## 5.5 *Computational Efficiency*

The inference time per image was measured for both YOLOv5l and YOLOv10-l models to assess computational efficiency.

Table 6: Inference Time Comparison

| **Model** | **Inference Time (ms)** |
|---|---|
| YOLOv5l | 8.6 |
| YOLOv10-l | 6.8 |

As the YOLOv10-l model achieves a shorter inference time, this improvement enhances its efficiency, enabling faster analysis of medical images, which contributes to streamlined workflows and improved throughput in diagnostic settings.

## 5.6 *Visualization of Results*

To provide a comprehensive understanding of the model's performance, sample detection results are presented in Figure 9. The images show the original CT slices, the ground truth annotations, and the model's predictions.

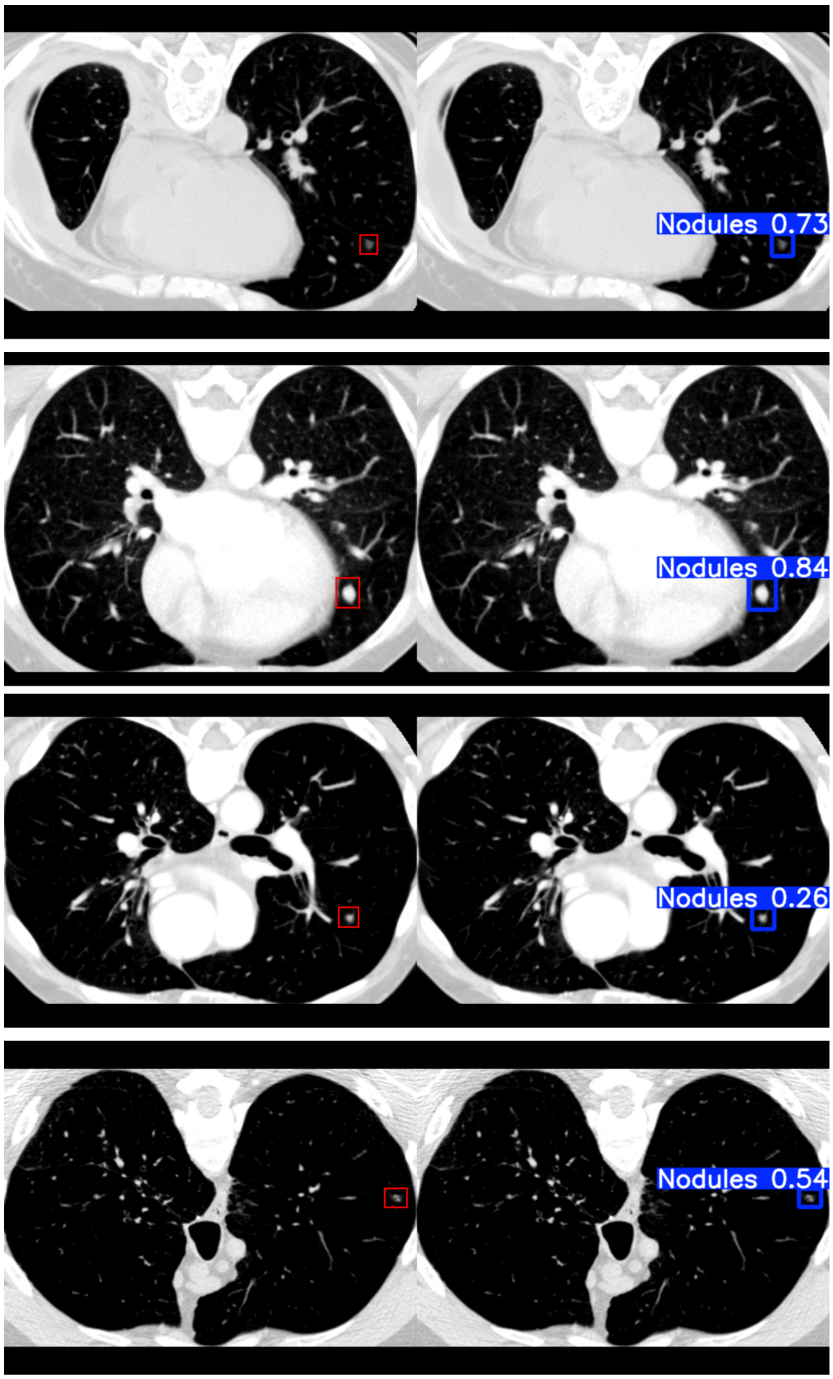


Figure 9: The figure illustrates examples of accurate detections by the model, with the ground truth displayed on the left and the corresponding predictions shown on the right.

# 6 DISCUSSION

This section critically evaluates the findings of this thesis, structured around the research questions presented earlier. It discusses the results, compares them with existing literature, and reflects on the strengths and limitations of the methods employed. Finally, it highlights implications for future research and practical applications.

## 6.1 *Results*

This thesis addressed the **RQ: To what extent can YOLO models detect lung nodules?** by evaluating the performance of YOLO-based deep learning frameworks for detecting lung nodules in CT scans. The findings confirm the effectiveness of YOLOv10 for this task, demonstrating its superiority over YOLOv5 and the impact of critical preprocessing strategies on detection accuracy.

The most notable result was that YOLOv10-l achieved a higher mAP@0.5 (0.61) compared to YOLOv5-l (0.54) while maintaining faster inference times (6.8 ms vs. 8.6 ms). However, these results fall short of state-of-the-art benchmarks reported in the literature, such as the Average Precision of 90.5% achieved by Mei et al., 2021 using a YOLOv4-based model or the 92% sensitivity at one false positive per scan reported by Hendrix et al., 2023 with a YOLOv5 pipeline. These discrepancies likely stem from suboptimal hyperparameter tuning or preprocessing limitations.

Despite these challenges, this study demonstrates that YOLOv10 significantly outperforms YOLOv5 in lung nodule detection, marking it a robust and promising candidate for CAD systems. This evidence supports further exploration of YOLOv10's potential in clinical diagnostics and highlights its relevance for improving automated detection workflows.

### 6.1.1 *Effect of Preprocessing on Performance*

Preprocessing played a key role in improving model performance and directly addressed **SQ1: How does the choice of windowing parameters in CT image preprocessing affect the performance of YOLO models in detecting lung nodules?** and **SQ2: How does the size and scaling of bounding boxes used for annotating lung nodules influence the accuracy of YOLO models in nodule detection?**

Optimized windowing parameters, specifically a center of $-600$ Hounsfield Units with a width of 1500 HU, improved model performance by 6% compared to unmodified CT scans by enhancing nodule visibility and contrast (**SQ1**). Adjusting bounding box configurations, particularly expanding

them by 30% with a 5-pixel padding, led to a 20% performance improvement by providing additional contextual information to better distinguish nodules from nearby structures (**SQ2**).

These results confirm that domain-specific preprocessing, including optimized windowing and bounding box adjustments, is critical for enhancing detection accuracy and addressing challenges in medical imaging data.

## 6.2 *Strengths and Limitations of the Methodology*

### 6.2.1 *Strengths*

This study incorporates several methodological strengths that enhance its reliability and impact. The comprehensive preprocessing pipeline, which included lung segmentation, optimized windowing, and tailored bounding boxes, ensured high-quality input data and contributed to robust model performance. By implementing patient-wise data splitting, the study successfully mitigated the risk of data leakage, ensuring that the evaluation results are representative of real-world scenarios and generalizable to unseen cases. Iterative augmentation strategies, such as rotation and zooming, were employed to balance the dataset, addressing the class imbalance inherent in medical imaging data and improving model generalization. Furthermore, the comparative framework, which directly evaluated YOLOv5 and YOLOv10 under identical conditions, provided a robust assessment of architectural advancements and their impact on performance.

### 6.2.2 *Limitations*

This study faces several limitations that highlight areas for improvement. YOLOv10 exhibited challenges in detecting nodules with very small diameters or those located near lung boundaries, which may be addressed by utilizing higher-resolution input images or exploring specialized loss functions to enhance detection accuracy. The use of bounding boxes for annotating nodules, while effective, may oversimplify complex nodule shapes, suggesting that incorporating advanced segmentation techniques could provide more precise representations. Additionally, the dataset used in this study was sourced from a single collection, potentially limiting the generalizability of the findings; thus, external validation using independent datasets, such as LUNA16 or NLST, is crucial to establish broader applicability. Furthermore, the study primarily relied on 2D slices, underutilizing the full 3D nature of CT data, which could offer richer spatial context. Employing an ensemble model that integrates data from all three axes—axial, sagittal, and coronal—and combines predictions using

majority voting may enhance model robustness and accuracy by leveraging multiple perspectives of the nodule.

Moreover, the study exclusively focused on comparing YOLO-based architectures and did not include other state-of-the-art deep learning frameworks, such as Faster R-CNN or Swin Transformers, which could provide valuable benchmarks. Future studies incorporating a broader range of architectures would offer a more comprehensive assessment of YOLOv10's performance relative to alternative approaches in lung nodule detection.

## 6.3 *Implications for Broader Society and Practical Applications*

### 6.3.1 *Societal Impact*

This research has direct implications for lung cancer diagnostics. By improving the detection of lung nodules, YOLOv10-based pipelines can enhance the accuracy and efficiency of Computer-Aided Diagnosis systems, reducing the burden on radiologists and enabling earlier interventions. Early detection is critical in lung cancer, where survival rates decrease significantly with disease progression (Saeed et al., 2024).

Furthermore, automated systems like the one proposed in this study could expand access to high-quality diagnostic tools in resource-limited settings. By deploying such systems in underserved regions, where radiologists are scarce, it is possible to reduce global disparities in cancer care.

### 6.3.2 *Clinical and Industrial Applications*

This study provides a foundation for integrating YOLOv10 into real-world CAD workflows with several potential applications. One key application is in triage systems, where the model can automate the initial review of CT scans, prioritizing cases with suspected nodules for further analysis, thereby streamlining radiologists' workflows. Additionally, YOLOv10 can serve as a valuable research tool, supporting large-scale epidemiological studies by enabling high-throughput nodule detection across extensive datasets. Furthermore, the model holds promise in training and educational settings, assisting radiologists in identifying nodules while enhancing their understanding of deep learning predictions, ultimately fostering confidence and trust in AI-based diagnostic tools.

## 6.4 *Recommendations for Future Research*

### 6.4.1 *Addressing Model Limitations*

Future research should focus on addressing the limitations identified in this study by exploring innovative techniques that enhance detection capabilities. One promising approach involves the integration of boundary refinement modules directly into the model architecture. These modules can dynamically adjust bounding boxes by analyzing edge information and ensuring precise localization, especially for nodules located near lung boundaries where contrast with surrounding structures is minimal.

Another avenue is the utilization of 3D context by incorporating volumetric analysis across sequential CT slices. By leveraging the spatial continuity of nodules across multiple slices, the model could better differentiate nodules from non-nodule structures such as blood vessels or image artifacts. This approach would provide richer context for detecting nodules that may appear less distinct in individual slices.

These strategies offer practical solutions to the current challenges in lung nodule detection and pave the way for further improvements in model accuracy and reliability.

### 6.4.2 *Expanding Dataset Diversity and Exploring Multi-Modality Approaches*

Validating the proposed pipeline on larger and more diverse datasets is essential for improving the generalizability of the results. Incorporating datasets with varying scanner types, patient demographics, and pathological conditions would ensure the robustness of the model across different scenarios. Additionally, integrating CT data with complementary imaging modalities, such as PET or MRI, could provide additional context, enhancing detection accuracy. Multi-modal learning frameworks could be explored to leverage the strengths of each modality and improve the overall performance of the pipeline.

### 6.4.3 *Advanced Techniques in Lung Nodule Detection*

Future advancements in lung nodule detection can leverage cutting-edge techniques to address existing challenges and improve model performance. Two promising directions include the integration of Explainable AI (XAI) and multi-modal learning frameworks.

Explainable AI (XAI) methods, such as Grad-CAM (Selvaraju et al., 2019) and SHAP (Lundberg & Lee, 2017), provide insights into the decision-making processes of YOLO-based frameworks by highlighting the regions or features that influence predictions. These methods foster trust among clinicians by ensuring that model predictions align with relevant anatomical

structures, such as lung nodules. Moreover, XAI can help identify biases or errors in the model, enabling targeted improvements and ensuring more reliable predictions.

Another approach involves combining CT data with complementary imaging modalities, such as PET or MRI, to provide additional context that can enhance detection accuracy. Multi-modal learning frameworks allow models to leverage diverse perspectives on nodule characteristics, improving performance on challenging cases. For example, integrating spatial information from 3D volumetric CT scans can help detect nodules that are less apparent in individual slices, thereby reducing false negatives.

### 6.4.4 *Benchmark Comparisons*

While this study focuses on evaluating YOLOv10 for lung nodule detection, future research should benchmark its performance against other state-of-the-art object detection frameworks. Models such as Faster R-CNN, Swin Transformers, and Vision Transformers have demonstrated strong capabilities in various image analysis tasks and could provide valuable points of comparison.

Benchmarking these models using standardized datasets, such as LUNA16 or NLST, would offer a comprehensive understanding of their relative strengths and weaknesses. Such comparisons can highlight scenarios where YOLOv10 excels or falls short, guiding further development and optimization. Incorporating benchmark comparisons into future studies ensures a balanced evaluation of available methods and advances the overall understanding of deep learning approaches in medical imaging.

## 7 CONCLUSIONS

This thesis investigated the application of YOLOv5 and YOLOv10 deep learning frameworks for detecting lung nodules from CT scans, with the overarching goal of improving Computer-Aided Diagnosis systems in lung cancer detection. The findings highlight the significant advancements introduced by YOLOv10 and the critical role of preprocessing and architectural design in enhancing model performance.

YOLOv10-l demonstrated superior performance compared to YOLOv5l, achieving a higher mAP@0.5 and faster inference times. The study further emphasized the importance of optimized preprocessing techniques, including windowing parameters and bounding box adjustments, in enhancing detection accuracy for challenging medical imaging tasks.

Despite these advancements, several limitations were identified. The model struggled with detecting small nodules and those near lung bound-

aries, reflecting challenges inherent in small-object detection. Additionally, the study's reliance on 2D data limited its ability to fully utilize the 3D spatial context of CT scans. Furthermore, the dataset's limited diversity highlight the need for external validation to ensure broader applicability of the findings.

The societal and clinical implications of this work are profound. By demonstrating the potential of YOLOv10 for lung nodule detection, this research paves the way for more accurate and efficient CAD systems, which can reduce diagnostic delays and improve early intervention in lung cancer cases. Moreover, the deployment of such systems in resource-limited settings could address disparities in global healthcare access.

Looking forward, several directions for future research have been identified. These include exploring ensemble models that leverage the 3D nature of CT data, validating the pipeline on diverse datasets, and integrating multi-modal imaging approaches. Additionally, developing explainable AI frameworks and benchmarking YOLOv10 against other state-of-the-art architectures will further enhance its clinical relevance and adoption.

In conclusion, this study underscores the transformative potential of deep learning in medical imaging, particularly in improving lung cancer diagnostics. While challenges remain, the insights gained from this research provide a strong foundation for advancing CAD systems, ultimately contributing to better patient outcomes and fostering innovation in healthcare technology.

## REFERENCES

Abboud, Z., & Kadoury, S. (2023). Impact of train- and test-time hounsfield unit window variation on ct segmentation of liver lesions. *Medical Imaging 2023: Image Processing*, *12464*, 124642E. https://doi.org/10.1117/12.2653974

Ain, K. (2021). Ct scan image segmentation based on hounsfield unit values using otsu thresholding method. *Jurnal Ilmiah Teknologi Informasi Asia*, *15*(1), 1–6. https://www.researchgate.net/publication/349911278_CT_scan_image_segmentation_based_on_hounsfield_unit_values_using_Otsu_thresholding_method

Alanazi, M., Almutairi, S. F. M., Alarjani, N., Alghaylan, M. Y. A., Aljawhari, M. S. M., & Alkhulaifi, A. A. S. (2024). Advancements in ai-driven diagnostic radiology: Enhancing accuracy and efficiency. *International Journal of Health Sciences (IJHS)*, *8*(S1), 737–749. https://doi.org/10.53730/ijhs.v8ns1.14928

Armato, S. G. I., McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., Zhao, B., Aberle, D. R., Henschke, C. I., Hoffman, E. A.,

Kazerooni, E. A., MacMahon, H., Van Beek, E. J. R., Yankelevitz, D. F., Biancardi, A. M., Bland, P. H., Brown, M. S., Engelmann, R. M., Laderach, G. E., ... Clarke, L. P. (2015). Data From LIDC-IDRI. https://doi.org/10.7937/K9/TCIA.2015.LO9QL9SX

Balaji, R., Prabaharan, G., Singh, A. R., Athisayamani, S., Sarveshwaran, V., & Daniya, S. (2022). Multi-scale features fusion with yolov3 for detecting small and fine tumors in mri images. *2022 6th International Conference on Electronics, Communication and Aerospace Technology*, 1545–1549. https://api.semanticscholar.org/CorpusID:255994048

Billah, M., Hossain, M. S., Ahamed, A., Al Rakib, A., & Haque, M. (2024). Real-time object detection in medical imaging using yolo models for kidney stone detection. *European Journal of Computer Science and Information Technology*, *12*, 54–65. https://doi.org/10.37745/ejcsit.2013/vol12n75465

Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. https://arxiv.org/abs/2004.10934

Brett, M., Markiewicz, C. J., Hanke, M., Côté, M.-A., Cipollini, B., McCarthy, P., Jarecka, D., Cheng, C. P., Halchenko, Y. O., Cottaar, M., Larson, E., Ghosh, S., Wassermann, D., Gerhard, S., Lee, G. R., Baratz, Z., Wang, H.-T., Kastman, E., Kaczmarzyk, J., & freec84. (2023). *Nipy/nibabel: 5.1.0* [Zenodo] (5.1.0). https://doi.org/10.5281/zenodo.7795644

Buslaev, A., et al. (2020). Albumentations: Fast and flexible image augmentations. *Information*, *11*(2), 125. https://doi.org/10.3390/info11020125

Chien, J. R., Chou, K., & Chiang, J. (2024). Yolov9-e: Lightweight object detection for pediatric wrist fractures in x-rays. *arXiv preprint arXiv:2403.11249*. https://arxiv.org/abs/2403.11249

da Costa-Luis, N. (2023). *Tqdm: A fast, extensible progress bar for python and cli*. https://github.com/tqdm/tqdm

DenOtter, T. D., & Schubert, J. (2023). Hounsfield unit. *StatPearls [Internet]*. https://europepmc.org/article/MED/31613501

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. https://arxiv.org/abs/2010.11929

Gautam, S., DeVera, R. E., & Patel, M. (2022). 3. abstract 6347a: Promotion of low-dose computed tomography for early-stage lung cancer detection. *Cancer Research*. https://doi.org/10.1158/1538-7445.am2022-6347a

Geetha, A. S., & Hussain, M. (2024). A comparative analysis of yolov5, yolov8, and yolov10 in kitchen safety. https://arxiv.org/abs/2407.20872

Gould, M. K., Donington, J., Lynch, C. W., Mazzone, P. J., Midthun, D. E., Naidich, D. P., & Wiener, R. S. (2015). Evaluation of patients with pulmonary nodules: When is it lung cancer? *Chest*, *148*(1), 308–318. https://doi.org/10.1378/chest.15-0677

Grover, V., Grover, V., & Nandal, M. (2024). Advancements in medical imaging. *Advances in healthcare information systems and administration book series*, 106–123. https://doi.org/10.4018/979-8-3693-5468-1.ch007

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hendrix, W., Hendrix, N., Scholten, E. T., Mourits, M., Trap-de Jong, J., Schalekamp, S., Korst, M., van Leuken, M., van Ginneken, B., Prokop, M., Rutten, M., & Jacobs, C. (2023). Deep learning for the detection of benign and malignant pulmonary nodules in non-screening chest ct scans. *Communications Medicine*, *3*(1), 156. https://doi.org/10.1038/s43856-023-00388-5

Hofmanninger, J., Prayer, F., Pan, J., et al. (2020). Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem. *European Radiology Experimental*, *4*(1), 50. https://doi.org/10.1186/s41747-020-00173-2

Huang, F., Zhang, Z., Zheng, L., Wen, H., Dai, M., & Zhai, C. (2024). A comparative study of yolov9 and yolov10 for steel surface defect detection. *2024 30th International Conference on Mechatronics and Machine Vision in Practice (M2VIP)*, 1–6. https://doi.org/10.1109/M2VIP62491.2024.10746104

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in science & engineering*, *9*(3), 90–95.

Hussain, M. (2024). Yolov5, yolov8 and yolov10: The go-to detectors for real-time vision. https://arxiv.org/abs/2407.02988

Itseez. (2015). Open source computer vision library.

Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., NanoCode012, Kwon, Y., Michael, K., TaoXie, Fang, J., imyhxy, Lorna, Yifu), Ê., Wong, C., V, A., Montes, D., Wang, Z., Fati, C., Nadar, J., Laughing, … Jain, M. (2022, November). *ultralytics/yolov5: v7.0 - YOLOv5 SOTA Realtime*

*Instance Segmentation* (Version v7.0). Zenodo. https://doi.org/10.5281/zenodo.7347926

Jocher, G., Qiu, J., & Chaurasia, A. (2023, January). *Ultralytics yolo* (Version 8.0.0) [Released under AGPL-3.0 license. Repository: https://github.com/ultralytics/ultr Ultralytics. https://ultralytics.com

Karki, M., Cho, J., Lee, E. J., Hahm, M. H., Yoon, S. Y., Kim, M. H., Ahn, J. Y., Son, J.-H., Park, S. H., Kim, K. H., & Park, S. J. (2020). Ct window trainable neural network for improving intracranial hemorrhage detection by combining multiple settings. *Artificial Intelligence in Medicine*, *106*, 101850. https://doi.org/10.1016/j.artmed.2020.101850

Katase, S., Ichinose, A., Hayashi, M., et al. (2022). Development and performance evaluation of a deep learning lung nodule detection system. *BMC Medical Imaging*, *22*, 203. https://doi.org/10.1186/s12880-022-00938-8

Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., Ke, Z., Li, Q., Cheng, M., Nie, W., Li, Y., Zhang, B., Liang, Y., Zhou, L., Xu, X., Chu, X., Wei, X., & Wei, X. (2022). Yolov6: A single-stage object detection framework for industrial applications. https://arxiv.org/abs/2209.02976

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. *European Conference on Computer Vision (ECCV)*, 740–755. https://doi.org/10.1007/978-3-319-10602-1_48

Lin, T., Maire, M., Belongie, S. J., Bourdev, L. D., Girshick, R. B., Hays, J., Perona, P., Ramanan, D., Doll'a r, P., & Zitnick, C. L. (2014). Microsoft COCO: common objects in context. *CoRR*, *abs/1405.0312*. http://arxiv.org/abs/1405.0312

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. https://arxiv.org/abs/2103.14030

Loverdos, K., Fotiadis, A., Kontogianni, C., Iliopoulou, M., & Gaga, M. (2019). Lung nodules: A comprehensive review on current approach and management. *Annals of thoracic medicine*, *14*(4), 226–238.

Lundberg, S., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. https://arxiv.org/abs/1705.07874

Ma, X., Jia, J., Yan, X., Wang, X., Song, Y., Zhu, G., & Gao, Y. (2024). Research on optimization of yolov8 pulmonary nodule detection algorithm based on deep learning, 782–788. https://doi.org/10.1109/ICECAI62591.2024.10675286

Ma, Z., & Wu, F. (2024). Research on pulmonary nodule detection method based on yolo algorithm. *2024 5th International Seminar on Artificial*

*Intelligence, Networking and Information Technology (AINIT)*, 635–639. https://doi.org/10.1109/AINIT61980.2024.10581847

Madan, M., Reich, C., & Hassenpflug, F. (2023). Drawing and¬†analysis of¬†bounding boxes for¬†object detection with¬†anchor-based models. In R. Gade, M. Felsberg, & J.-K. Kämäräinen (Eds.), *Image analysis* (pp. 359–373). Springer Nature Switzerland.

maintainers, T., & contributors. (2016). *Torchvision: Pytorch's computer vision library*. GitHub.

McKinney, W., et al. (2010). Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*, *445*, 51–56.

Mei, S., Jiang, H., & Ma, L. (2021). Yolo-lung: A practical detector based on imporved yolov4 for pulmonary nodule detection, 1–6. https://doi.org/10.1109/CISP-BMEI53629.2021.9624373

Mkindu, H., Wu, L., & Zhao, Y. (2023). 3d multi-scale vision transformer for lung nodule detection in chest ct images [Accessed on: December 1, 2024]. *Signal, Image and Video Processing*, *17*, 2473–2480. https://doi.org/10.1007/s11760-022-02464-0

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., & Lerer, A. (2017). Automatic differentiation in pytorch [NeurIPS Workshop]. *Proceedings of Neural Information Processing Systems*.

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779–788. https://doi.org/10.1109/CVPR.2016.91

Redmon, J., & Farhadi, A. (2016). Yolo9000: Better, faster, stronger. https://arxiv.org/abs/1612.08242

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. https://arxiv.org/abs/1804.02767

Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster r-cnn: Towards real-time object detection with region proposal networks. https://arxiv.org/abs/1506.01497

Saeed, A., Ahmad, J., Akhtar, H., Adeel, M., & Ahmad, A. (2024). The role of early diagnosis and intervention in improving outcomes for lung cancer. *Innovative Research in Applied, Biological and Chemical Sciences*, *2*, 131–137. https://doi.org/10.62497/IRABCS.2024.54

Sapkota, R., Qureshi, R., Calero, M. F., Badjugar, C., Nepal, U., Poulose, A., Zeno, P., Vaddevolu, U. B. P., Khan, S., Shoman, M., Yan, H., & Karkee, M. (2024). Yolov10 to its genesis: A decadal and comprehensive review of the you only look once (yolo) series. https://arxiv.org/abs/2406.19407

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2019). Grad-cam: Visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*, *128*(2), 336–359. https://doi.org/10.1007/s11263-019-01228-7

Siegel, R. L., Miller, K. D., & Jemal, A. (2020). Cancer statistics, 2020. *CA: A Cancer Journal for Clinicians*, *70*(1), 7–30. https://doi.org/https://doi.org/10.3322/caac.21590

Tong, B., & Zhang, M. (2023). Comparison of yolo series algorithms in mask detection. *2023 International Workshop on Intelligent Systems (IWIS)*, 1–5. https://doi.org/10.1109/IWIS58789.2023.10284631

van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., & Yu, T. (2014). Scikit-image: Image processing in python. *PeerJ*, *2*, e453. https://doi.org/10.7717/peerj.453

Viriyavisuthisakul, N., Wongsripuemtet, J., Sriton, C., Chotmongkol, R., Soontrapa, K., & Suthisisang, C. (2023). Evaluation of fixed hounsfield unit windowing in acute ischemic stroke detection on non-contrast computed tomography. *PLOS ONE*, *18*(1), e0274674. https://doi.org/10.1371/journal.pone.0274674

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024). Yolov10: Real-time end-to-end object detection. https://arxiv.org/abs/2405.14458

Wang, C.-Y., Bochkovskiy, A., & Liao, H.-Y. M. (2022). Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696*. https://arxiv.org/abs/2207.02696

Wang, C.-Y., Yeh, I.-H., & Liao, H.-Y. M. (2024). Yolov9: Learning what you want to learn using programmable gradient information. https://arxiv.org/abs/2402.13616

Wang, Y. (2024). Pulmonary nodule detection on chest x-rays: Exploring deep learning algorithms and their clinical impact. *2024 4th International Conference on Electronic Information Engineering and Computer Science (EIECS)*, 403–407. https://doi.org/10.1109/EIECS63941.2024.10800010

Yaseen, M. (2024). What is yolov8: An in-depth exploration of the internal features of the next-generation object detector. https://arxiv.org/abs/2408.15857

Yin, J., Chen, Y., Li, C., et al. (2024). Swin-transuper: Swin transformer-based upernet for medical image segmentation [Accessed on: December 1, 2024]. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-024-19009-x

Youssif, A. A. A., Hussein, S. A., & Ibrahim, A. S. (2011). Three-dimensional detection of pulmonary nodules in chest ct images. *Computer and Information Science*, *4*(5), 2. https://doi.org/10.5539/cis.v4n5p2

Zhang, H., & Qie, Y. (2023). Applying deep learning to medical imaging: A review. *Applied Sciences*, *13*(18). https://doi.org/10.3390/app131810521

Zhong, Y., Wang, J., Peng, J., & Zhang, L. (2020). Anchor box optimization for object detection. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*.