Arab American University

Faculty of Graduate Studies

Detection of Prostate Cancer Using Deep Learning Techniques and Positron Emission Tomography/Computed Tomography (PET/CT) Images

By

Ahmad Kathem Tawfeeq Mnaizel

Supervisor

Dr. Ahmad Ewais

This thesis was submitted in partial fulfillment of the requirements for the Master's degree in Data Science and Business Analytics

Feb/ 2026

I

Thesis Approval

Detection of Prostate Cancer Using Deep Learning Techniques and Positron Emission Tomography/Computed Tomography (PET/CT) Images

By

Ahmad Kathem Mnaizel

This thesis was defended successfully on 01/02/2026 and approved by:

Committee members	Signature
1. Dr. Ahmad Ewais: Supervisor	
2. Dr. Moh ammad Awad: Internal Examiner	....
3. Dr. Zain Salah: External Examiner	....

Declaration

I hereby declare that the work presented on this thesis is my own work and effort and has not been submitted anywhere for any degree in university, institution, or other colleges of higher education than Arab American University – Palestine (AAUP). Where other sources of information have been used, they have been acknowledged.

I confirm that I have read and understood the policies on plagiarism as outlined by Arab American University – Palestine (AAUP) and that this thesis is no plagiarism of any kind in this thesis.

Signed,

Student Name: Ahmad Kathem Tawfeeq Mnaizel
Student ID: 202112513

Signature: Ahmad Mnaizel
Date: 31.3.2026

Dedication

This work is dedicated to the memory of my beloved father, who passed away during my academic journey due to prostate cancer. May he rest in peace and may his memory continue to inspire me.

I also dedicate this work to my dear mother, who never spared her prayers, love, and care; she was my support and safe haven through the toughest moments. To my dear brothers and my lovely sister, for their constant encouragement and support, to my beloved wife and precious daughters, whose love, patience, and noble hearts were a beacon of hope and strength throughout this journey.

Finally, I dedicate this research to all my friends and everyone who supported me—through words, actions, prayers, or advice—helping me overcome challenges and continue this path. To all of you, I extend my deepest gratitude and heartfelt appreciation.

Acknowledgments

I acknowledge the support and assistance of my supervisor, Dr. Ahmad Ewais, I express my gratitude for his constant support and direction and encouragement throughout this journey. and other individuals who have contributed to this work.

Abstract

Prostate cancer represents one of the most frequent malignancies in men globally, and accurate lesion detection in medical imaging is very important for early diagnosis and treatment. Modern imaging modalities such as Positron Emission Tomography/Computed Tomography (PET/CT) imaging yield very valuable information for diagnostics. At the same time, the process of lesion identification in these images is very time-consuming and not very accurate. The proposed thesis investigates the potential of the use of object detection models based on deep learning for the automatic detection of lesions in the prostate.

Several state-of-the-art object detection models, such as YOLO-based object detection models and Roboflow OpenYOLO, were trained and compared based on the same data set and evaluation procedure. The Models were trained and evaluated using a local dataset collected from Augusta Victoria Hospital (AVH) in Jerusalem, comprising PET/CT images from 200 prostate cancer patients, totaling approximately 2,735 images focused on the prostate region, to my knowledge, PET/CT imaging has not been used in prior local studies, making this dataset a novel resource. Images were annotated with the help of expert radiologists. This dataset demonstrates the applicability of transformer-based detection models to real-world clinical data. Because of its high accuracy and robustness, RF-DETR was chosen as the final model for further testing. The chosen model was initialized with weights that were pretrained using COCO, and the usual COCO metrics, such as precision, recall, and Average Precision (mAP), were employed for evaluation.

The experimental results demonstrate that the fine-tuned RF-DETR has achieved a precision of 93.9%, recall of 92.6%, and mAP@50 of 93.4%. These measurements ensure high detection capability and accurate lesion location in a moderate overlap condition. Lower efficiency at higher intersection over union (IoU) thresholds, indicated by lower mAP@50-95 values, indicate the inherent challenge in precisely locating lesions at the boundary. The qualitative analysis has also substantiated these results with proper lesion location in both internal and external test sets, efficient processing of cases with no lesions, and successful detection of smaller lesions with challenging conditions. A detailed analysis involving true positive, false positive, and false-negative values has also given an insight into the experiment results.

In conclusion, this work has shown that a transformer-based models, especially RF-DETR, is an efficient tool for automatically detecting a prostate lesion from a medical image. Although there is still a problem in localization accuracy at a higher threshold of IoU, this study has a great potential for improvement and provides a good basis for further advancement by improving localization accuracy and multi-modal data integration.

Table of Contents

List of Tables

List of Figures

List of Abbreviation

AI	Artificial Inelegance
ANN	Artificial Neural Networks
EDA	Exploratory Data Analysis
WHO	World Health Organization
AUC	Area Under Curve
CNN	Convolutional Neural Networks
Conv	Convolution
ConvNet	Convolution Network
DenseNet121	Densely Connected Convolutional Networks
DL	Deep Learning
FC	Fully Connected
FN	False-Negative
FP	False-Positive
ML	Machine Learning
MRI	Magnetic resonance imaging
PC	Personal Computer
ReLU	Rectified Linear Unit
ResNet50	Residual learning Network
ROC	Receiver Operating Characteristic Curve
TN	True Negative
TNR	True Negative Rate
TP	True Positive
TPR	True Positive Rate
VGG-16	Visual Geometry Group-16 Mode
COCO	Common Objects in Context
PET	Positron Emission Tomography
CT	Computed Tomography

CHAPTER ONE

INTRODUCTION

1.1 BACKGROUND

Cancer is a disease in which some of the body's cells grow uncontrollably and spread to other parts of the body (Health, 2024). Cancer significantly contributes to global mortality, responsible for around 1 in 6 deaths, affecting almost every family (WHO, 2024). Prostate cancer (PCa) is one of the most common malignancies in men worldwide, affecting middle-aged men between 45 – 60 and it is the second most common cancer among men, main risk factor related to PCa are age, genetic nature, smoking, diet, physical activities, certain medications and work-related factors (Oskar Bergengren, 2023). Unfortunately, its biological traits often contribute to a delayed clinical diagnosis, although it is a common disease, when caught early enough, it is usually slow-growing and treatable.

The main challenge for healthcare industry is to understand the demand of this increasing group of men suffering from such life-threatening disease, in other words utilizing new technologies to improve prostate cancer detection.

1.1.1 Prostate Cancer Diagnosis Techniques:

Prostate Cancer traditional diagnosis methods include prostate-specific antigen (PSA) test, digital rectal examination (DRE), scans and biopsy (UK, 2024).

- Digital rectal examination (DRE), in which the doctor physically checks prostate gland for suspicious signs such as lumps or hard areas (Center, 2024).

- PSA is a protein produced by prostate cell either normal or cancerous ones, PSA test specify the amount of prostate specific antigen in the blood, it is a main indicator for diagnosis and screening of prostate cancer, also it was the first PCa biomarker approved by U.S. Food and Drug Administration (Elizondo-Riojas, 2018).
- Scans include magnetic resonance imaging (MRI) which use magnets and radio waves to get images with more details than CT images, MRI scan mainly focus on soft tissues in the body. Computed tomography (CT) images, in which x-rays are applied to monitor the potential presence of cancer, especially in lymph nodes or other organs. Bone scans are used to determine if the cancer has reached the bones by injecting low-level radioactive substances to identify damaged bone parts by cancer. Transrectal ultrasound (TRUS) uses ultrasound images to take images from specific areas targeted for biopsies by inserting probe into patient's rectum to get an image for the prostate. Positron emission tomography (PET) scan, which is a nuclear medicine image modality, in which a radioactive tracer liquid is used to monitor the metabolic processes in the body. By detecting the areas of increased tracer uptake, PET scans can identify and localize cancerous lesions (Schatten, 2018).
- Biopsy is a procedure in which little prostate samples are taken out and examined under a microscope, usually performed when the result of PSA or DRE is abnormal (Inform, 2024).

1.1.2 Treatment Options of PCa

Prognosis prediction is a key factor of clinical oncology, because the expected disease progression and chances of survival can influence treatment options (Khoa A. Tran, 2021). Treatment options of PCa include active surveillance, chemotherapy, radiation therapy, hormonal therapy, surgery, and cryotherapy. Treatment option choice depends on tumor nature, PSA level, tumor grade and stage and the possibility of cancer recurrence (Christian Bach, 2014).

1.1.3 PCa Traditional Diagnosis Techniques Limitations and issues

Classical prostate cancer (PCa) diagnosis techniques, including DRE, PSA testing, biopsies, and imaging scans have major limitations and side effects and could not be definitive enough. In case of using DRE while screening, it is limited because of its lack of reliability, sensitivity, and inability to accurately feel the entire prostate gland, in particular for tiny tumors (Dahm, 2013). As well in PSA tests sometimes the test recognize small cancers but couldn't distinguish between aggressive and non-aggressive tumors, which results in unnecessary biopsies and inaccurate diagnosis (Pierre-Jean Lamy, 2017), sometimes these methods generate false positives or false negatives which results in incorrect and inaccurate diagnosis and unnecessary medical interventions, in TRUS biopsy false-negative rate could reach 46% because of random location of the needle, and could reach 38% in tumor under-grading (Descotes, 2019). The accuracy of such techniques depends on the practitioner's knowledge, experience skills and the abilities which make it highly subjective. Also, invasive procedures are usually painful and uncomfortable, and can cause risky complications, while image scanning techniques might be costly and not always reliable (Nelly Gordillo, 2013). In addition to overdiagnosis, harms associated with overtreatment were false-positive PSA test results and complications of unnecessarily aggressive treatment (e.g., infection or bleeding from biopsies for subsequent diagnosis). It must be realized that there are consequences to increased anxiety as well additional examinations through biopsies and the extensive side-effects associated with treatments for various stages of prostate cancer. This issue becomes worse by inability to distinguish clinically significant neoplasms, which may be left untreated (Dahm, 2013). These limitations and drawbacks illustrate the need for more accurate, harmless, less costly and more definitive diagnoses techniques, such as those presented by the developments in artificial intelligence, to help and guide healthcare professionals to make more accurate diagnoses, consequently, offering better and effective treatments.

1.1.4 Artificial Intelligence in Healthcare

Artificial intelligence (AI) is defined as the technology that can perform, and in some cases can adapt to execute tasks without intervention from human, using computer knowledge to simulate intelligent behavior (Gopi Battineni, 2020).

Machine learning (ML) is recognized as a subset of AI techniques, ML inherently represents the "learning" quality which is part and parcel with what we associate human intelligence, it goes beyond patterns recognition to learning from data using computational algorithms. Machines trained on huge datasets with inputs and labels can then make decisions or provide recommendations on their own (Ramkumar, 2020).

Healthcare has encountered a massive developments and transformations due to the progress and advancements in technology, numerous information systems have been proposed and used in the healthcare industry, mainly decision support systems employed and helped healthcare practitioners in taking data-driven decisions resulting in enhancement of care quality (Malcolm Rozario, 2021).

Artificial Intelligence is transforming healthcare; main concepts that may be considered to highlight the impact of AI in healthcare are relieving workload, replacing tasks and augmenting clinical practices and medical knowledge (Ting, 2021). AI helps professionals by using electronic records and helps them with diagnosis and interpretation of images. It facilitates personalized treatment, processes automation and improve patient's outcomes and life expectations. Figure 1-1 depicts the different functionalities of utilizing AI in healthcare. For instance, AI constructs quickly and accurately evaluate medical tests and scan images, forecast disease outbreaks, and asses high-risk patient scores, which helps to detect and install instruments immediately (Candasamy, 2023). AI includes various technologies, including machine learning and deep learning which supports processing and interpret a huge amount of medical data, in diagnostics, AI algorithms analyze medical scan images like X-rays, MRI, Ultrasound, Positron Emission Tomography (PET) scan, etc. with same accuracy as human experts in the field, sometimes they often recognize situations early and more accurate than experts (Brown, 2018).

Particularly, AI and deep learning have demonstrated promising results in improving detection, diagnosis, and prognosis in prostate cancer. Machine learning models have been applied to clinical data, such as PSA levels, for early detection and risk assessment. In medical imaging, deep learning algorithms analyze MRI, CT, and PET/CT scans to automatically detect and classify prostate cancer lesions, often achieving accuracy comparable to expert radiologists. These AI applications facilitate personalized treatment planning, help predict disease progression, and support clinical decision-making, demonstrating the potential of AI to enhance prostate cancer care.

Machine learning (ML) models analyze patient's medical records, patient's behavior, histories and genetic information to predict possible diseases or patient's responses to treatments (Buzdin, 2019). Deep Learning (DL), a subset of ML inspired by artificial neural networks, has the advantage of extracting patterns and insights from unstructured medical data, also it has the ability to extract high-level features and recognize images. DL learns from massive amounts of data using neural networks with multiple layers, it shines in tasks like data analyses, Natural language processing (NLP) and image and speech recognition (Priya, 2021). DL is very effective in healthcare in many aspects such as disease diagnosis, disease prediction and detection, image analysis, drug supply, and smart monitoring and assistance systems (Bordoloi, 2022).

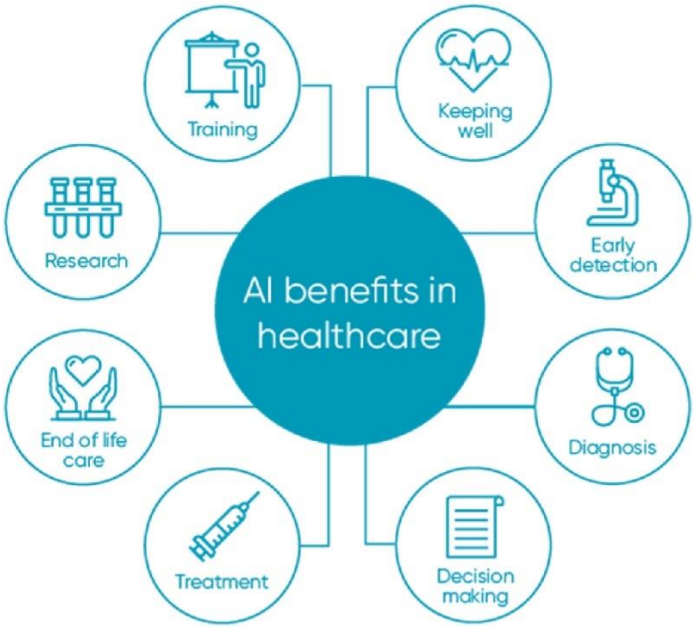


Figure 1-1 Applications of ML and DL in the healthcare system (Singh U. C., 2025).

As a result, customized treatment adapts the patient's genetic material and behavior to AI-generated diagnoses and treatment plans. Additionally, using AI-based risk profiles and AI personal assistants, virtual assistants and robotic surgery are readily available (Biancone, 2021). It also accelerates research in the medication discovery process and repurposes current inventions. AI is shifting hospitals procedures, involving scheduling and resource allocations, which improves efficiency and reduces expenses (Harfouche, 2023). As AI advances, it has the opportunity to revolutionize the healthcare industry, changing the focus from being reactive and defensive to diseases to being proactive and predictive in a way that will significantly benefit all parties in healthcare sector.

1.2 Problem Statement

Prostate cancer is one of the most common cancers affecting men globally, and early detection is a key factor in improving treatment results. In Palestine, challenges such as restricted access to advanced diagnostic tools, delays in results interpretation, and inconsistencies in radiological

assessment complicate early detection. While PET/CT imaging provides valuable insights, its accurate interpretation is time-consuming and requires highly qualified expertise and significant clinical experience.

Machine learning, in particular through the use of pre-trained models, has the potential to provide support to clinicians by improving the accuracy and reliability of prostate cancer detection. However, the effectiveness and relevance of these models on Palestinian patient data using PET/CT images remain insufficiently investigated. Therefore, there is a need to evaluate if current pre-trained models can be effectively implemented to enhance diagnostic decision-making in this context.

1.3 Motivation

The motivation behind this thesis came from the growing worldwide dependence on machine learning in medical imaging and the need to support healthcare systems that have limited resources. Utilizing pre-trained models can reduce development time and computational cost, enhancing the accessibility of AI-driven diagnosis solutions in regions such as Palestine

By applying machine learning to PET/CT images, this research seeks to help and support clinicians in achieving faster, more reliable, and more consistent prostate cancer diagnoses. The ultimate goal is to contribute to improved patient outcomes, support early detection plans, and promote the integration of modern AI methods in local healthcare processes so that it can be helpful for different departments in the ministry of health, hospitals, doctors, and radiologists.

1.4 Objective and Scope of the Study

This study evaluates the use of pre-trained machine learning models (YOLOv11, YOLOv12, Roboflow 3.0, and RF DETR) for prostate cancer diagnosis using PET/CT imaging data collected from patients in Palestine. The objectives of this study are:

- To collect a unique local PET/CT dataset for prostate cancer from local medical institutes.
- Preprocessing PET/CT images and preparing them for model input.
- Choosing appropriate pre-trained models and using them for this dataset.

- Evaluating model performance using common diagnostic metrics.
- Comparing results with existing approaches and relevant studies.
- Analyzing possibility of clinical integration, with focus on interpretation and computational performance.

1.5 CONTRIBUTION

This thesis offers several important contributions to prostate cancer diagnosis and the application of machine learning in medical imaging filed, especially within the Palestinian healthcare context:

Use of a Unique Local PET/CT Dataset:
The study uses a PET/CT dataset of Palestinian patients that was obtained locally, which has not been analyzed before using machine learning techniques or any other approach. It provides a valuable platform for developing AI-based diagnostic approaches related to local clinical characteristics, imaging protocols, and local population-specific patterns.

First Local Study uses PET/CT with Machine Learning for Prostate Cancer:
As far as we know, no other study elsewhere has explored using PET/CT imaging and pre-trained machine learning models for the diagnoses of prostate cancer in Palestine. This study does not only identify a gap in existing research but also creates a starting point for further research and investigation in this health domain.

Advancing Machine Learning Applications in Palestinian Healthcare management:
By researching on real local clinical data, the study indicates the importance and effectiveness of applying artificial intelligence approach within Palestine's healthcare diagnostic process. The findings highlight the possible role of AI-driven tools in ensuring the efficient use of resources and enhancing service delivery, even in healthcare environments which are financially and technically constrained.

Clinical Value for Early Detection:
The study shows insights into how machine learning might be used to enhance early detection accuracy for prostate cancer, so a chance of reducing diagnostic variability and making clinicians decision-making more informed.

Benchmarking and Performance Evaluation:

The study provides baseline performance measures for PET/CT based prostate cancer diagnosis by pre-trained models. These benchmarks can be instrumental in leading up the next pieces of research in the field, future models' development, and comparative studies within and outside the region.

Recommendations for Future Research and Clinical Integration:

The findings of the study address the opportunities and restrictions of employing AI for prostate cancer diagnosis in Palestine. It can be a source of practical advice for improving datasets, model implementations, and the general readiness, willingness and attitude of local healthcare institutions for AI implementation.

1.6 THESIS ORGANIZATION

This thesis is structured into five main chapters to provide a clear and logical flow from background concepts to experimental findings and conclusions.

Chapter 1 – Introduction

This chapter presents the overall background and motivation for the study, outlines the research problem, states the objectives, and highlights the significance and contributions of the work. It also provides an overview of the thesis structure.

Chapter 2 – Literature Review

This chapter reviews previous research related to prostate cancer diagnosis using machine learning. It begins with studies based on clinical biomarkers and traditional datasets, then transitions to imaging-based research using CT, MRI, PET, and PSMA-PET modalities. Special emphasis is placed on identifying research gaps, particularly the absence of local studies in Palestine and the lack of PET/CT-based machine learning studies in the region.

Chapter 3 – Methodology

This chapter describes the dataset used in the study, including details about the local PET/CT imaging dataset, preprocessing techniques, feature extraction, and model selection. It also explains the methodological workflow, including adaptation of pre-trained models, tuning

procedures, and evaluation metrics such as sensitivity, specificity, precision, recall, and accuracy.

Chapter 4 – Experiments and Results

This chapter presents the experimental findings obtained from applying the selected machine learning models to the PET/CT dataset. Results are analyzed and compared, highlighting model performance, interpretability, and clinical relevance. The chapter also discusses the implications of using machine learning for early prostate cancer detection in the local healthcare context.

Chapter 5 – Conclusions and Future Work

This chapter summarizes the main outcomes of the thesis, discusses the limitations of the work, and provides recommendations for future research. It also outlines potential directions for improving machine learning integration into local clinical practice and expanding prostate cancer diagnostic studies within Palestine.

CHAPTER TWO

Literature Review

2.1 BACKGROUND

Prostate cancer is one of the most common cancers which hit men around the world, and it is a major factor in cancer-related mortality (Bristow, 2021). Detecting the disease in its early stages increases the opportunity of effective treatments which eventually will result in mitigating the harm and guarantee better patients' outcomes (Fitzgerald, 2019). Classical diagnosis approaches often tend to have limitations, such as false positives, false negatives, and the intrusive nature of procedures (Tewari, 2023). These challenges have motivated the exploration of advanced technologies to improve diagnostic accuracy and eliminate the necessity for invasive tests. Recent years have witnessed the rise of ML machine learning (ML) as a powerful solution for the medical field challenges, with the potential to help in both analyzing large datasets and recognizing complicated patterns with high precision when making predictions. Machine learning techniques can predict prostate cancer with significantly better Area Under Curve (AUC), accuracy, and net clinical benefits compared to traditional methods (Chen S. a., 2022). Early detection, medical care optimization and mortality reduction are the intention of ML algorithms applicability in Prostate cancer diagnosis. The literature review highlights the state of research in this field along with consideration and key studies, methods and techniques for solving some relevant challenges and other challenges that remain unsolved.

2.2 LITERATURE REVIEW

The application of machine learning in prostate cancer diagnosis came from the clinical need to reduce dependency on invasive diagnostic procedures and to improve the diagnostic accuracy (Hassan, 2024). This motivation has pushed ongoing research efforts, first based on laboratory clinical biomarkers then progressing toward advanced medical imaging techniques supported by machine learning models (Ar, 2022). Each phase in this development addressed limitations of previous approaches and provided new opportunities for better accuracy, earlier detection, and non-invasive assessment.

2.2.1 Clinical Biomarker Based Approaches in Prostate Cancer Diagnosis

In the early days, prostate cancer detection was mainly dependent on biomarkers such as Prostate-Specific Antigen (PSA), a test which became a clinical routine in the 1980s (Bonfrer, 1990). While PSA testing transformed early diagnosis, its drawbacks, particularly false positives, emphasized an urgent need for advanced approaches (Schalken, 2016). In this context, machine learning began to show its potential (Nishiyama, 2019).

By the late 2010s, researchers started employing ML to clinical datasets, searching for patterns which conventional statistical methods could miss. This technique has proven potential in several medical domains, improving accuracy in diagnosis, predictive modeling, and medical decision-making (Sidey-Gibbons, 2019). An innovative approach innovative approach, integrating demographics, family history and transrectal ultrasound findings, and serum prostate-specific antigen (PSA) levels into a random forest model, was proposed in (Feng, 2016). The model was validated on 941 patients and achieved an accuracy of 83.10%, sensitivity of 65.64%, and specificity of 93.83%. Compared to individual diagnostic indicators, the algorithm significantly improved diagnostic performance, giving a useful tool to identify high-risk patients and supporting biopsy decisions. This illustrates the potential of machine learning in improving prostate cancer diagnosis. Building on limitations previously recognized with PSA testing as a solo diagnostic sign, current advances in machine learning have tried to overcome these limitations. For instance, the use of machine learning to improve prostate cancer risk assessment was investigated and outperformed standard PSA-based diagnostics. Using a dataset of 4,548 patients, a neural network model featuring PSA, free-PSA, age, and the FTR ratio had an AUC of 0.72, exceeding PSA alone. Whereas the model showed valuable diagnostic advances, the level of specificity (45.3% at 80% sensitivity) indicated possible constraints in practical application (Smith, 2020). Moreover, the observed AUC fluctuation with reduced sample sizes (AUC 0.64) emphasizes the need for reliable datasets. This research highlights the potential and difficulties associated with integrating machine learning into prostate cancer diagnosis, highlighting the need for deeper validation.

Another study (Nayan, 2023) employed machine learning techniques to address the challenge of predicting PSA readings for ongoing surveillance patients, with the intention of enhancing prostate cancer decision-making processes. The research used electronic medical records from 4,269 patients, including 1,134 on ongoing surveillance, and trained various machine learning models, such as Gradient Boosting Regression, to predict future PSA values using previous PSA measures

and clinical variables such as age, prostate volume, and BMI. The best-performing model, Gradient Boosting Regression, was trained on a diverse dataset and significantly decreased prediction error. The model's root means square error (RMSE) dropped and its accuracy improved when earlier PSA values were included. These findings indicate that machine learning could improve resource usage and tailor ongoing surveillance in prostate cancer treatment. Extending on the potential of machine learning beyond typical PSA screening, in (Chen S. a., 2022) retrospective research of 551 prostate biopsy patients generated five models utilizing four machine learning algorithms: logistic regression (LR), decision tree (DT), random forest (RF), and support vector machine (SVM). The models were tested based on parameters such as AUC, accuracy, and sensitivity. In the training set, RF, DT, and multivariate LR models maintained superior detection with AUCs of 1.0, 0.922, and 0.91, respectively. In the test set, multivariate LR outperformed traditional approaches (AUC = 0.918), indicating that machine learning models can deliver more accurate and generalizable predictions.

This momentum inspired the investigation of further biomarkers such as PCA3 and genetic signatures. In a landmark meta-analysis, the diagnostic effectiveness of PCA3 was evaluated thorough a comprehensive analysis of 65 studies with 8,139 cases and 14,116 controls (Jia, 2019). The analysis indicated a cumulative sensitivity of 0.68, specificity of 0.72, and AUC of 0.76, indicating that PCA3 is a fairly accurate, non-invasive biomarker for differentiating between prostate cancer patients and healthy people. Despite its potential, drawbacks such as inconsistent validation bias and various inclusion criteria among studies were identified, highlighting the need for a deeper study. When integrating PCA3 scores with machine learning models, the predictive performance of PCA3 improves significantly. For instance, a previous study (Passera, 2021) applying an XGBoost model found PCA3 as key predictor in a machine learning structure, with approximately four times the value of free-PSA and PSA. This combination of PCA3 and intelligent ML methods reveals PCA3's ability for enhancing diagnostic precision and prostate cancer treatments, increasing its role as a crucial biomarker for advanced diagnostic systems.

Previous findings indicate the potential and effectiveness of ML methods to improve biomarker-based prostate cancer prediction. This move from individual to multi-feature modeling was essential, but researchers immediately realized that even the most powerful biomarker models were unable to sufficiently tackle the complicated nature of prostate cancer.

Longitudinal biomarker datasets marked the next step forward. Longitudinal biomarker datasets have made significant contributions to health-care research by allowing for the tracking of biomarker changes over time, assisting in the analysis of disease progression, individual variance, and the development of individualized treatments (Schisterman, 2012). In order to enhance active surveillance (AS) techniques to monitor prostate cancer, researchers examined the predictive ability of prostate-specific antigen kinetics (PSAk) over time. In (Cooperberg, 2018) a linear mixed-effect model (LMEM) was created for analyzing serial PSA data and predict biopsy reorganization, which can be defined as an increase in cancer grade or volume. The study engaged 851 men from the multinational Canary Prostate Active Surveillance Study. Over a median follow-up of 3.7 years, 30% of individuals were reclassified. PSAk levels have been shown to be a significant predictor, with each 0.10 increase resulting in a 1.6-fold greater risk of reclassification ($p < 0.001$). Models involving PSAk beat those that used diagnostic PSA alone in terms of risk classification, with comparable results for PSA readings obtained every three or six months.

Time-series models enhance prostate cancer prediction and diagnosis by increasing long-term accuracy, such as dynamic clinical data, and offering extra therapeutic advantages. These models allow personalized treatment plans and adaptive clinical experiments, leading to enhanced patient outcomes.

However, the field had remained limited by the actually indirect character of biomarker data. To truly visualize the issue, researchers turned into imaging.

2.2.2 Imaging Based Machine Learning Approaches in Prostate Cancer Diagnosis

Imaging techniques such as MRI and PET scans provided something that biomarkers did not: a clear spatial view of the disease such as location and size. The process started with computed tomography (CT) images, which gave a baseline for understanding tumor anatomy. Computed tomography (CT) is a computerized X-ray imaging technology that produces comprehensive cross-sectional images, or slices, by revolving a narrow X-ray radiation across the human body. The segments produce more details than traditional X-rays and can be electronically merged to generate a 3D image, helping in the detection of structures, tumors, or variations (B.Burbridge, 2022). In the purpose of exploring further advancements and roles of CT scans in prostate cancer treatment field, a prior study (Jain, 2019) applying radiomics and machine learning showed potential in prostate cancer diagnosis and classification using CT imaging. CT-based radiomics attributes obtained from planning scans correctly categorized Gleason scores and risk groups, showing

AUCs of 0.90 for GS < 6 vs. GS ≥ 7 and 0.98 for GS 7(3 + 4) vs. GS 7(4 + 3). Risk group categorization proved to be accurate, with AUC values of 1.00 for low vs. intermediate risk and 0.96 for low vs. high risk. These results demonstrate the potential of radiomics for non-invasive insights into prostate cancer aggressiveness; however, further evidence is required. While these early efforts were inspiring, they were limited in that no consistent datasets and low-quality CT scans were available. The integral issue reveals some deficiencies of CT imaging in cancer diagnosis such as radiological characteristics of relatively low sensitivity and specificity compared to other imaging tools, biological issues of radiation exposure, the associated technicalities affecting image quality and the lack of molecular imaging targets (Jon, 2010) (Yu, 2023).

This was followed by the advent of multi-parametric MRI (mpMRI), which refers to a general and loose definition that includes various multi-contrast MRI techniques for radiotherapy simulation (Cai, 2023). Whereas CT provided a single imaging sequence, mpMRI combined multiple sequences, including diffusion-weighted imaging (DWI), T2-weighted imaging, and dynamic contrast-enhanced (DCE) imaging, thereby providing more informational and detailed data. Many machine learning techniques, mainly deep learning frameworks, have shown significant potential for enhancing mpMRI diagnostic accuracy. ML models, like convolutional neural networks (CNNs), may analyze complicated patterns in imaging data and detect minor signs of prostate cancer. These algorithms have the potential to enhance lesion recognition from images, feature categorization, and the prediction of clinically relevant prostate cancer (CSPCa). MRI images play a significant role in diagnosis and examination. (Singh S. K., 2024) presented a deep learning model for prostate cancer detection combining a 3D convolutional neural network (CNN) and Gleason grading of historical images. The model achieved cutting-edge performance by recognizing epithelial areas as well as Gleason scores at the same time. Using MRI slices, volumes, and segments aided by an endorectal coil, the model scored 87% accuracy, 85% specificity, and 89% sensitivity on the SPIE-AAPM-NCI Prostate dataset. These findings indicate the promise of deep learning for prostate cancer recognition and grading using MRI analysis. mpMRI achieves high diagnostic accuracy for detection, localization, and staging, impacting patient care and allows for MRI-targeted biopsy (Rosenkrantz, 2015). Building on these foundations as a base, specialists have widely understood machine learning's ability to improve the diagnostic abilities of MRI imaging for prostate cancer. Machine learning models can use complex algorithms to analyze the massive amount of data included in multiparametric MRI (mpMRI) images and find small patterns

indicating tumors that humans may miss. These improvements led to the creation of automated systems that can divide prostate areas, separate between malignant and benign tumors, and even predict the aggressiveness of identified cancers using imaging biomarkers (Sánchez, 2017) (Neto, 2020) (Hu, 2023).

The adoption of machine learning in the contextualization of MRI diagnosis has further stimulated patient-centered strategies, since models based on large and heterogeneous datasets are likely to flexibly conform to variations in anatomy and pathology. For instance, the use of deep learning methods such as convolutional neural networks (CNNs) has become popular in parsing prostate zonal structures and detecting PI-RADS-defined lesions (Cuocolo, 2019). Ensemble learning methods also improve prediction performance by combining predictions from multiple models resulting in reduced instances of false positives and improved diagnostic accuracy (Bertelli, 2022) (Al-Ani, 2023).

The Prostate Imaging-Reporting and Data System (PI-RADS), version 2.1 in particular, plays an important role in determining the ability of identifying clinically significant prostate cancer (csPCa) on a 5-grade scale. However, some PI-RADS categories (sch as PI-RADS 3 lesions) are of indeterminate nature sometimes, resulting in diagnostic issues (Alabousi, 2020). Recent studies have shown the utilization of machine learning in enhancing PI-RADS 3 lesions identification. For example, a radiomic analysis of apparent diffusion coefficient (ADC) maps in mpMRI has shown promise as a stand-alone approach for early and non-invasive prostate cancer diagnosis. As investigated by Gaudiano (2023), machine learning approach can be applied to PI-RADS 3 lesions – associated with diagnostic uncertainty – and demonstrated that the diagnostic reliability of the technology could be improved. This study included 133 patients with 155 PI-RADS 3 lesions, of which 84 were confirmed as prostate cancer (PCa) by prior biopsy. A support vector machine (SVM) classifier was trained and validated using linear and polynomial kernels after extracting radiomic features from ADC maps and selecting features using LASSO. The model exhibited an excellent PPV of 80%, specificity of 76%, and the sensitivity of 78%. These results, in addition to demonstrating the utility of radiomics in characterizing equivocal PI-RADS 3 lesions, also represent a step toward an image-based diagnostic pathway that may help avoid unnecessary invasive procedures in certain cases.

Furthermore, researchers have demonstrated the ability to connect imaging features with molecular and histological data using deep learning for radiomics, a procedure for extracting quantitative information from radiographs. Data on tumor heterogeneity is useful for treatment planning and may aid in the diagnosis of clinically relevant malignancies in a preclinical context (Papanikolaou, 2024) (Ebert, 2021). Concretely, the combination of mpMRI and machine learning may usher in a new era of PCa diagnosis and change the present clinical practice paradigm by enabling more accurate, data-driven clinical decision-making (Castillo T, 2020). Despite its promise, there are still challenges to overcome, such as the need for standard, high-quality, reproducible data, the difficulty of imaging variability, and the need for clinical confirmation prior to deployment.

In conclusion, the integration between machine learning and multiparametric MRI shows significant potential for enhancing prostate cancer diagnosis. By employing the power of advanced imaging and complex computational techniques, healthcare providers can improve diagnostic accuracy, adjust treatment plans, and eventually improve patient outcomes.

As MRI models matured, focus turned to positron emission tomography (PET), specifically when associated with CT. PET/CT scans, which merge metabolic and anatomical data, have made a massive leap in diagnostic accuracy. The announcement of prostate-specific membrane antigen (PSMA) - a highly specific marker for prostate cancer cells - PET imaging employment further improved the recognition of prostate cancer cells across the body (Lawhn-Heath, 2021).

PET scans have a strong history in molecular imaging, introduced back in the 1970s, and it was combined with CT images as PET/CT hybrid imaging which was rapidly spread in the early 2000s as global essential technique in oncological imaging, opening the path to the clinical practice molecular imaging based evaluation of tumors (Besson, 2021). PET is a nuclear medicine image modality that uses radiotracers to visualize and measure variations in metabolic processes, and other biological functions in the body. A radiotracer is, as a rule, a molecule tagged with a radioactive isotope injected into the patient’s body. This molecule concentrates on specific tissues or organs based on their biological features. A PET scanner captures the gamma rays issued by the radiotracer and produces a high resolution three-dimensional image of the tracer spread (Chiș, 2022) (Unterrainer, 2020).

PET based modality initially utilized radiotracers like fluorodeoxyglucose (FDG) to map glucose metabolism, providing insights into various cancers. However, FDG-PET had some limitations in

the diagnosis of prostate cancer due to the slow nature of the disease and its unique metabolic profile. Such constraints motivated the investigation of innovative radiotracers that are suitable to the biological features of prostate cancer (Sutherland, 2023) (Schuster, 2016).

In the recent few years, the foundation of PSMA-targeted radiotracers such as 68Ga-PSMA and 18F-PSMA has turned into evolution of prostate cancer imaging, achieving exceptional accuracy in detection and staging. These radiators attach specifically to PSMA, a transmembrane protein which is highly presented on prostate cancer cells compared to its present in normal tissues. This specific ability made the localization of cancerous lesions more accurate and useful processes, even those at micrometastatic levels, creating a new reference for diagnostic and therapeutic imaging. The application of machine learning (ML) into PSMA-based positron emission tomography (PET) and PET/CT imaging has enhanced its potential, allowing for more sophisticated and automated analyses of complex imaging data (Zhang, 2021).

PSMA-targeted radiotracers, such as 68Ga-PSMA and 18F-PSMA, have been widely validated and evaluated in clinical studies. These substances provide high sensitivity and specificity in detecting prostate cancer lesions, exceeding traditional imaging procedures like CT, MRI, and bone scans.

Thomsen (2018), Petersen (2018), Lan (2022) clearly showed the excellent diagnostic performance of 18F-PSMA and 68Ga-PSMA PET/CT scans in the detection of bone metastases, with significant sensitivity and specificity values as compared to traditional imaging methods such as MRI. These innovations presently offer a much more reliable means of identifying metastatic lesions, specifically in patients having prostate cancer; they are thus of crucial support for use in the context of early detection and accurate staging.

A study by Rowe (2020) emphasized the advantages of 18F-labeled PSMA tracers, for example, 18F-PSMA-1007 generally results in better spatial resolution because of its higher positron yield and lower positron energy. These characteristics reduce noise in the images and enhance contrast resolution, allowing for the detection of subtle lesions that may be undetected on other imaging techniques. Such improvements are invaluable in the early detection of metastatic disease and in the assessment of treatment response.

According to (Maisto, 2022), 18F-PSMA tracers are more widely available on the market due to their longer half-life since this affords delayed imaging protocols and more flexibility in clinical as well as research settings. Particularly important for smaller healthcare facilities or regions that do not have immediate radiotracer production capacities, this would therefore ensure much greater equity in the availability of this advanced imaging technology.

Findings by Choyke (2022), Calais (2024), Calais (2020) highlighted the transformation of treatment strategies by PSMA PET/CT imaging, which provides staging information on a very high level of detail, these imaging modalities very often change significantly therapeutic approaches-at least as regards radiotherapy planning. The increased precision of staging ensures treatment that is both targeted and efficient, thus reducing chances of either over treating the disease or undertreating it.

While Fendler (2022), Vjaters (2022) observed a higher rate of nonspecific bone uptake with 18F-PSMA-1007, the research highlighted that competent reporting practices could substantially reduce the chance of false positives. Improved diagnostic accuracy directly reduces the propensity of making errors and subsequently supports the credibility of imaging based on PSMA in prostate cancer management.

In conclusion, Positron Emission Tomography (PET) scanning is a vital component of prostate cancer management mainly in diagnosis and treatment as it provides accurate imagining of metabolic behavior and offer significant knowledge on the form and localization of the disease. PET imaging has evolved into the gold standard due to recent advancements in radiopharmaceuticals like 18F-PSMA-1007 and 68Ga-PSMA-11, which can cover the entire body at once and offer essential and comprehensive insights and details about the progression of the disease and the response to treatment. Furthermore, integration of machine learning into PET scans has strongly enhanced its diagnostic capabilities. Machine learning algorithms can evaluate massive amounts of PET data to uncover patterns and biomarkers that may indicate malignancy, predict disease progression, and optimize treatment regimens. This collaboration between PET scans and machine learning provides important promises for personalized treatment in prostate cancer, including more accurate projections, enhanced patient care, and the progress of precision oncology. PET imaging and machine learning are a solid pair in seeking optimal prostate cancer care.

Additionally, the integration of explainable AI (XAI) frameworks in PSMA-PET imaging may boost trust from clinicians by offering clear and understandable model results. As the imaging data increases and ML methods develop, the mix of PSMA-targeted radiotracers and machine learning has the ability to change prostate cancer diagnosis and treatment (Sarah Lindgren Belal, 2024) (Lawrentschuk, 2024).

Despite the revolutionary promise and hope of clarity that ML brings to PSMA-PET imaging in terms of precision oncology and improving prostate cancer outcomes, certain major difficulties remain. PET imaging is frequently quite expensive and expansive, leading to inequities in patient care. ML models require large, homogeneous datasets to train on, however inconsistencies in the dataset can be caused by improperly standardized images. Furthermore, the computing needs and small quantity of training datasets complicate the problem (Mooney, 2021) (Sarah Lindgren Belal, 2024). Efforts such as Image Biomarker Standardization Initiative (IBSI) are partnering to establish data agreements and maintain consistent patterns of radiomic feature extraction (O'Connor, 2020). Concerns related to the interpretability of ML algorithms also pose a barrier to clinical application (Tsui, 2023). Future studies are expected to focus on types of collaborative learning in which institutions together are in cooperation with the operation of training without compromising data privacy with ML models. The uplifting of these embarkations will be among the new perspectives for use of PET imaging in conjunction with ML in precision oncology.

The story of prostate cancer diagnostics does not end with imaging as it is only one aspect of the diagnosis of prostate cancer. Rather, it creates a multi-modal future by combining with various data sources. researchers are now integrating imaging, biomarkers, and even genomic data in order to create thorough models that take into consideration the entire range of disease complexity. According to a systematic review by Brinker (2022), combining these modalities, for example, improves classification performance in cancer subtype prediction and survival analysis. Even though these results are preliminary, they highlight the potential of multimodal techniques to advance precision oncology; however, additional validation is required to demonstrate their therapeutic relevance.

As we reflect on this journey, one theme is emerging, machine learning is a transformative power rather than just a tool. It has changed our understanding of the approach to diagnosing prostate cancer, from imaging to laboratory biomarkers and beyond. The journey is far from over, though.

There are still difficulties with data standardization, algorithm interpretability, and ethics. However, every advancement brings us one step closer to a time when prostate cancer is not only detected early but also accurately, individually, and fairly.

Prostate cancer in Palestine is a major health issue and is one of the most prevalent cancers among men, it has shown an increasing trend for years regarding incidence rates (Statistics, n.d.). Prostate cancer studies conducted in Palestine remain limited and mainly statistical, focusing on incidence and mortality statistics. Studies exploring innovative diagnostic or management applications for cancer are less common (Sharaf, 2006; Husseini, 2009; Alajerami, 2015; Abu-Rmeileh, 2016; Halahleh, 2022; AlWaheidi, 2023). It is noteworthy that no study has yet looked at the application of machine learning (ML) in prostate cancer management is interesting since it highlights larger challenges of implementing modern technology in environments with limited resources.

However, with the increasing availability of imaging modalities such as PSMA-PET and MRI, there is a lot of promises to use ML models for early identification and treatment planning. Local initiatives to gather high-quality datasets and establish partnerships between healthcare organizations and academic researchers could enable the development of tailored machine learning ML solutions. These advancements could address regional healthcare disparities, improve diagnosis accuracy and resolve regional healthcare inequities. By overcoming existing constraints, Palestine could incorporate machine learning (ML) and artificial intelligence (AI) into its healthcare system, leading to earlier detection, more accurate diagnoses, and personalized treatment plans that improve prostate cancer outcomes.

2.3 SUMMARY

Chapter 2 traced the history of diagnostic processes for prostate cancer, focusing on biomarkers and imaging. Molecular markers, including PSA and PCA3, can detect cancer but are non-spatial, while other methods, including MRI, PET scans, and PSMA-PET scans, are useful in understanding tumor localization, sizing, and

development. Machine learning algorithms continue to improve the potential of both biomarkers and imaging. Overall, there is an emerging concept within literature to combine both molecular and imaging for precise, non-invasive diagnosis, including literature on prostate cancer in Palestine.

CHAPTER THREE

METHODOLOGY

This chapter describes the detection methodology using PET/CT imaging local dataset. It includes dataset processing to prepare the dataset to use in object detection models training, then several pre-trained models are adjusted and refined to detect and locate prostate cancer lesions. Also, evaluation metrics and validation strategies used to evaluate models' performance. It also describes how this method provides a reliable framework for precise, non-invasive prostate cancer diagnosis in Palestine by combining machine learning techniques with local imaging data.

3.1 INTRODUCTION

This study utilizes an object detection approach to identify the exact anatomical prostate cancer spots in PET/CT images by generating bounding box predictions, using multiple state-of-the-art pre-trained models to perform and compare the detection performance for these models.

Object detection models such as YOLOv11, YOLOv12, Roboflow 3.0, and RF-DETR were implemented to achieve this goal. These models represent various generations and architectures of modern computer vision, which allow an extensive comparison of their capability to find the prostate cancer spots.

The methodology covers the whole process starting from dataset acquisition and preparation, preprocessing, selection of machine learning models, model training using Roboflow, models evaluation. The approach follows a defined workflow to ensure reproducibility, accuracy, and compatibility with clinical diagnostic procedures.

3.2 METHODOLOGICAL FRAMEWORK

The proposed methodology features multiple sequential phases, starting with dataset preparation and proceeding to model training, testing, and performance evaluation.

Figure 3-1 depicts the overall research workflow which includes the following steps:

- Collecting image data from a local medical institute: PET/CT dataset of prostate cancer patients acquired from Augusta Victoria Hospital (AVH) after proper ethical approval.
- Verifying and preparing PET/CT images: Images were analyzed for quality and eligibility, collecting only the prostate area-related cuts.
- Preprocessing images: The images were normalized to the [0,1] range (to ensure the model focuses on cancer patterns), resized, and enhancement included all the standard preprocessing steps.
- Dataset splitting: Dataset is divided into a training set, a validation set, and a test set to allow for appropriate model training and unbiased evaluation.
- Training suitable pre-trained models and model selection: Several pre-trained object detection models were trained initially based on their outstanding previous performance, then final RF-DETR was considered as the final model to proceed with.
- Applying transfer learning via Roboflow: Pre-trained model RF-DETR were fine-tuned using transfer learning on the local dataset of PET/CT to utilize prior knowledge while reducing training time.
- Pre-trained model training and parameter tuning: Mainly related to hyperparameter tuning for improving model accuracy and generalization, such as learning rate, batch size, number of epochs.
- Models' assessment using clinical metrics: Performance was measured considering standard metrics as precision, recall, F1-score, and mAP; all quantify the detection accuracy.
- Result interpretation and comparison: Analyzing model output to find the best performance with discussions on clinical relevance and implications of prostate cancer diagnosis.

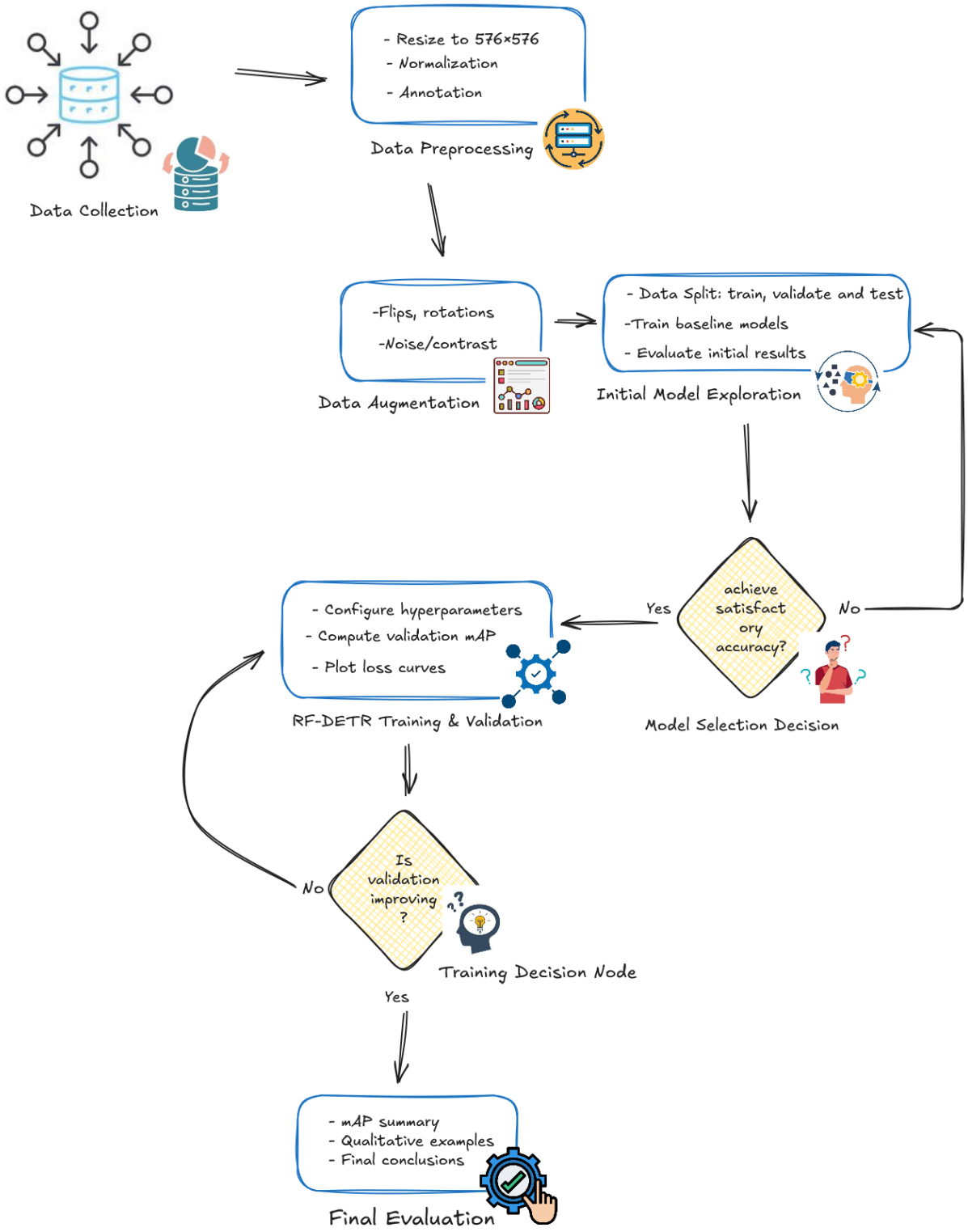


Figure 3-1 Proposed Methodology Workflow

3.3 DATASET DESCRIPTION

The dataset used in this study is composed of PET/CT images that were collected particularly for prostate cancer diagnosis. Every patient sample includes several image cuts, but only the ones representing the prostate area were chosen to ensure that the models focus on prostate clinically relevant features.

3.3.1 Dataset Source

The data were obtained from Augusta Victoria Hospital (AVH) in Jerusalem, a major medical center that provides PET/CT imaging services for cancer patients around Palestine. AVH's can be considered a reliable source for creating a representative and clinically significant and relevant dataset due to the diversity of patients' population and advanced imaging tools and facilities they have.

3.3.2 Dataset Structure and Annotation

Dataset consists of PET/CT images from 200 individual prostate cancer patients, totaling approximately 2,735 images were collected from the selected cuts from each case, with only the prostate area cuts were selected.

Each cut was manually annotated to highlight the cancer area within the prostate region to provide labels based on ground truth for supervised machine learning, Figure 3.2 shows example of annotated images.

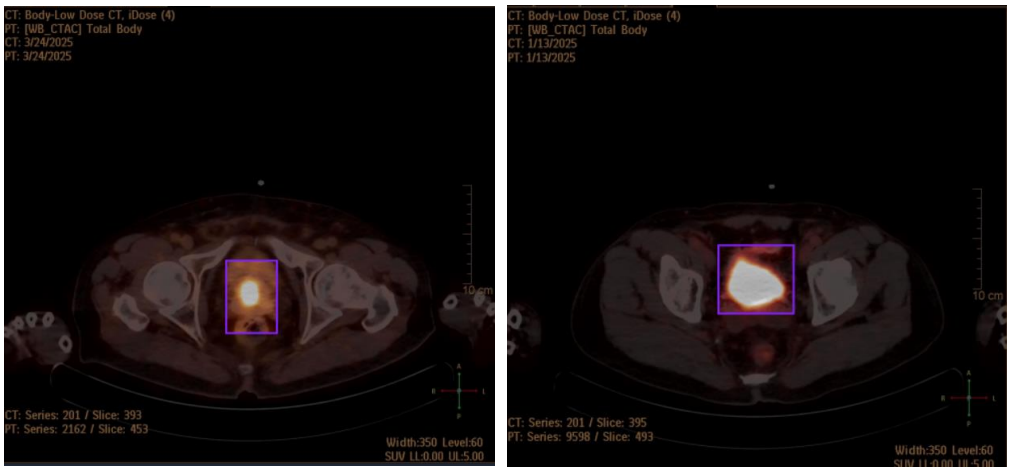


Figure 3-2 Annotated PET/CT cuts showing the highlighted prostate cancer region

In addition, in order to have a visual representation of the spatial distribution of the annotations over the dataset, the annotation heatmap shown in Figure 3.3 was generated. The color intensity

of each grid cell reflects the quantity of annotations in that area, making it evident where the majority of prostate areas are labeled. This makes it possible to evaluate the coverage of the dataset and aids in identifying the areas that are most important for model learning. As the heatmap shows, the annotations are highly concentrated almost in the central region of the images, which is the normal and expected location for prostate lesions. Overall, the coverage is sufficient for training detections models for prostate regions.

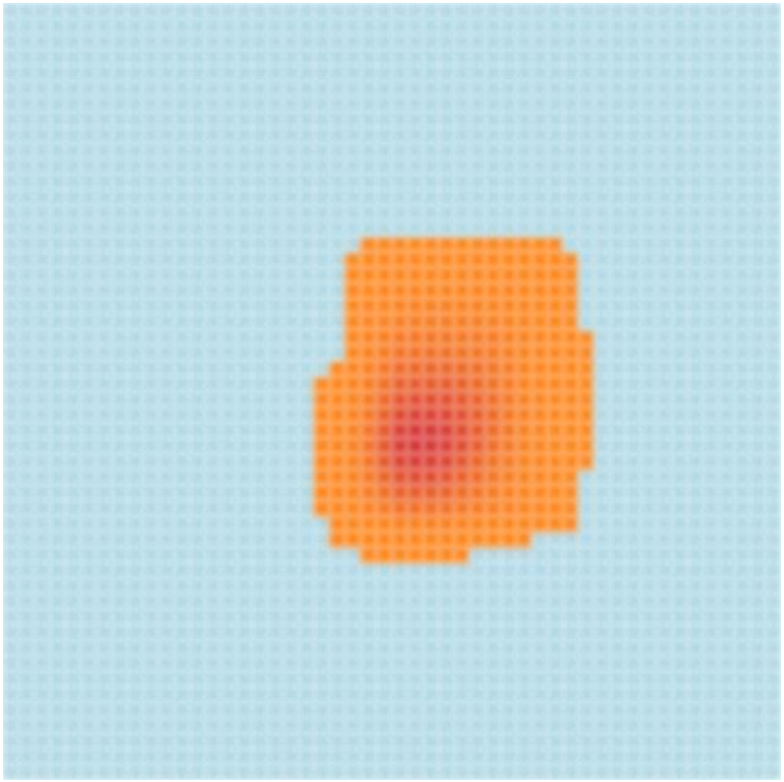


Figure 3-3 Annotation heatmap showing the density of prostate annotations in the dataset. Color gradients indicate the number of annotations per grid cell.

3.3.3 Dataset Samples

Figure 3.4 shows sample PET/CT images from the dataset. These samples reflect the variations in appearance, intensity, and anatomical structure present in the dataset.

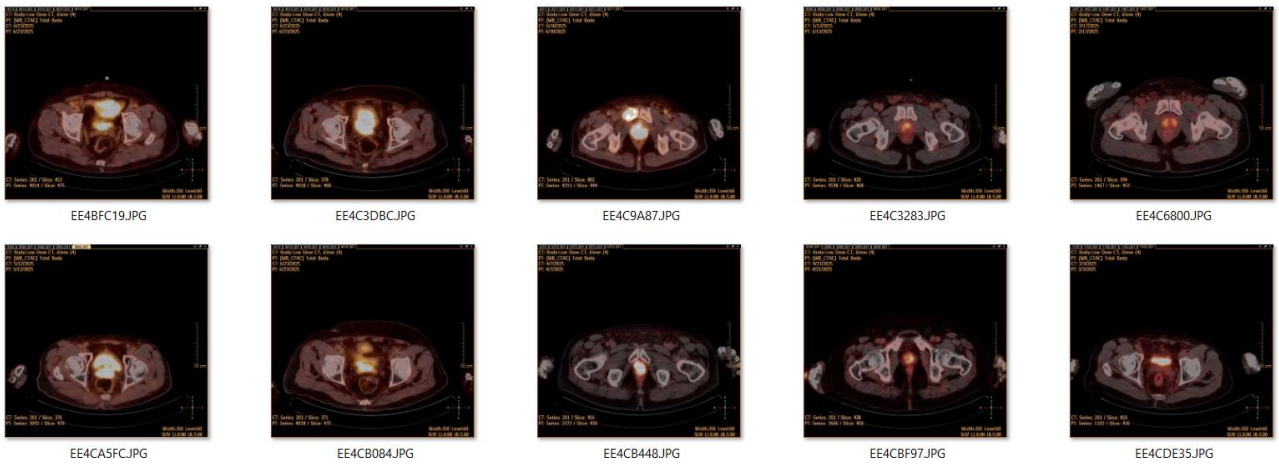


Figure 3-4 Sample PET/CT cuts from the dataset, showing variations in cancer appearance

3.3.4 Data Splitting

To efficiently train and evaluate the models, the dataset was split into training, validation, and test sets using the Roboflow platform:

- Training: 70% of the images were used to train the models.
- Validation: 15% of the images were used for hyperparameter tuning and early stopping.
- Testing: 15% of the images were used for final model evaluation.

The split was performed at the image level, with images randomly assigned to each subset, ensuring a balanced distribution of cases images across all sets while ensuring diversity of image features.

3.4 Preprocessing and Data Augmentation

3.4.1 Image Preprocessing

Initially all images were verified to ensure the quality and standardized in format. Preprocessing process involved resizing, normalizing, and selection of the prostate area cuts. This process was done to ensure that the input data is consistent and reliable for machine learning model training.

3.4.2 Data Augmentation

Data augmentation refers to the process of generating new, artificial data from existing data to increase both the diversity and the volume of a training dataset for machine learning models to

improve model training process and avoid overfitting and balance the data. Data augmentation was applied to increase dataset diversity and enhance model generalization, particularly with respect to the limited number of prostate cancer PET/CT cases. Augmentation makes the model more resistant to variability in patient anatomy, image orientation, and images collection circumstances.

These augmentation techniques were applied automatically to the training images using its built-in augmentation pipeline considering the following aspects:

- Horizontal and vertical flips: Introduced directional flips to generate geographical variations without losing the anatomy.
- Intensity adjustments: Adjusted brightness and contrast to simulate variations in imaging conditions across scans.
- Noise injection: Added a slight noise to increase resilience against scanner anomalies.

These augmentations techniques were only applied specifically to the training set to avoid affecting validation and test distributions. By extending the dataset size and providing controlled variation, augmentation mitigates the risk of overfitting and improves the reliability of the trained models when applied to unseen data.

3.5 Model Architecture

This study uses several advanced object detection models that are already trained on large datasets and made available through Roboflow. These pre-trained models come from different architectures and generations, meaning each one was designed using different computer vision techniques or improvements. This section discusses the models from a technical point of view focusing on their design concepts, training foundations, and anticipated performance characteristics.

3.5.1 YOLOv11

YOLOv11, an early version of the “You Only Look Once” family, with more robust multi-scale attention mechanism it enhances YOLOv10 ability to identify tiny and large objects, it focuses on real-time detection without much impact on accuracy. Its single-step process retrieves hierarchical features and predicts bounding boxes and class probabilities concurrently. YOLOv11 improves the detection of lesion through superior feature combination and stronger anchor box techniques.

It is well generalized to medical imaging when fine-tuned on prostate PET/CT images due to being pre-trained on large-scale datasets (Mao, 2025).

As shown in Figure 3.5 YOLOv11 architecture features backbone, neck and head which enable quick feature extraction, spatial enhancement, and reliable real-time detection.

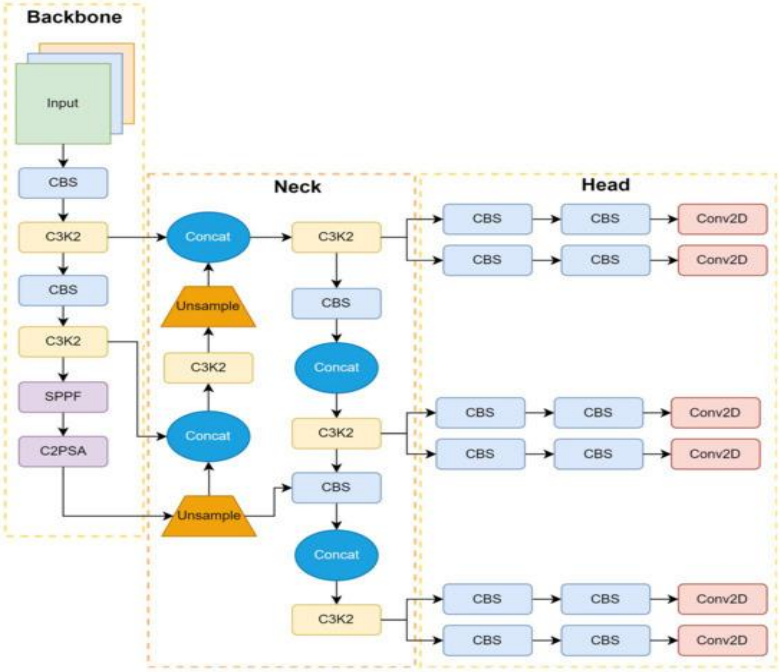


Figure 3-5 The architecture of YOLO-v11 (Mao, 2025).

Key advantages:

- High interpretation speed, appropriate for real-time detection.
- Reliable detection over different lesion sizes.
- Efficient treatment of unbalanced lesion distributions.

Limitations:

- Slightly decreased precision on complex or interconnected lesions compared to recent YOLO versions.
- May struggle when identifying tiny or small lesions in noisy images.

3.5.2 YOLOv12

YOLOv12, as the evolution of YOLO family, outperforms YOLOv11 by introducing deeper residual layers, flexible spatial concentration mechanisms, and optimized loss functions to speed,

improve accuracy and localization. It has been trained on massive, annotated datasets, it effectively employes transfer learning for domain-specific lesion recognition.

YOLOv12 introduce significant improvements to the backbone, feature aggregation and performance. Figure 3-6 shows baseline of this model which was YOLOv12n.

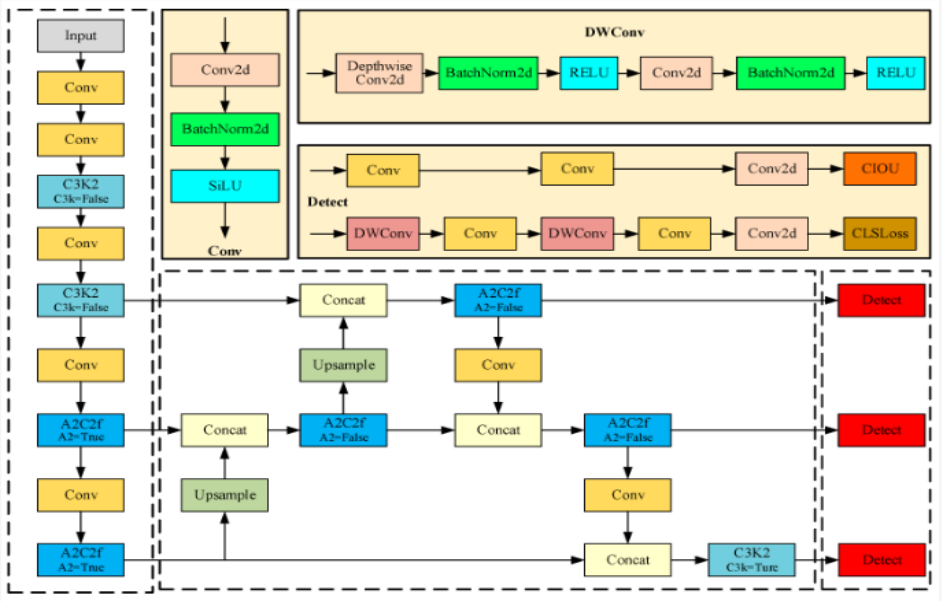


Figure 3-6 The structure diagram of YOLOv12 (Ji, 2025).

In order to enhance performance, YOLOv12 integrates lightweight attention-driven modules into each component while preserving the traditional backbone-neck-head layout of earlier YOLO models. The backbone focuses on collecting critical features from the input with a mix of Conv blocks, C3K2 (Cross-Stage Partial with Kernel-Split 2), and the Area Attention-enhanced A2C2f modules, offering better feature representation for real-time detection applications (Ji, 2025).

Key advantages:

- Higher detection accuracy than YOLOv11.
- Better identification of overlapping or complicated lesions
- Superior performance in noisy or low-contrast imaging areas

Limitations:

- Slightly slower interpretation speed because of increased complexity
- Training and fine-tuning require additional computational power.

3.5.3 Roboflow 3.0

Roboflow 3.0 offers a dynamic training environment which supports multiple backbone architectures, automatic data augmentation, and optimized end-to-end training pipelines. Here, it was used to train on annotated prostate imaging datasets. The platform's object detection system is built on Roboflow Train, supported and powered by OpenYOLO, which is a customized and optimized YOLO-based architecture. As shown in Figure 3-7, Roboflow 3.0 follows streamlined backbone–neck–head workflow, allowing the model to recognize visual features, utilize multi-scale information, and produces bounding box and classification predictions, enabling quick and accurate object detection that can be used for real-time medical imaging tasks (Gallagher, 2023).

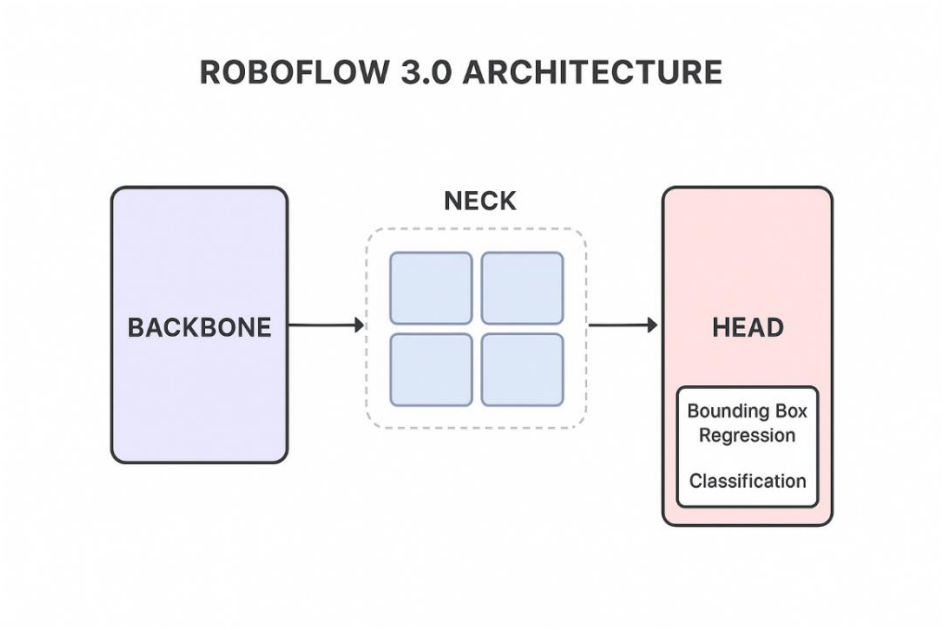


Figure 3-7 Roboflow 3.0 Object Detection Architecture (Chen W. a., 2024).

Key advantages:

- Flexible Adaptability to new datasets and architecture.
- Robust generalization due to strong augmentation techniques
- Simplifies model versioning and experimentation

Limitations:

- Performance depends on Training configuration and backbone chosen.
- Compared to specialized models it may require more manual tuning for ideal detection.

3.5.4 RF-DETR

RF-DETR is a transformer-based object detector that extends the DETR framework by including multi-scale deformable attention mechanisms, to enable efficient and effective object detection in images with diverse sizes and shapes. As illustrated in the architecture in Figure 3-8, the model first uses backbone network to extract image feature maps, then it feeds the extracted maps into a multi-scale encoder. The encoder first applies multi-scale deformable self-attention operations to the input features, which allows the network to selectively focus on relevant regions across different scales while decreasing computational complexity. These resulting multi-scale feature maps are fed to the decoder, which is implemented as a set object query that interacts with the encoded features through multi-scale deformable cross-attention and standard transformer self-attention. Such decoder design iteratively optimizes representations of objects over multiple layers, to produce accurate bounding box predictions for each detected object. This multi-scale deformable attention mechanism allows RF-DETR to handle high-resolution inputs efficiently and capture precise spatial information, making it particularly valuable for medical imaging tasks such as prostate cancer detection. RF-DETR model used in this research was initialized with COCO (Common Objects in Context)-pretrained weights. The COCO dataset is considered one of the most prominent large-scale benchmark datasets, comprising images with various context characteristics and annotations of different objects with bounding boxes. Although this model was not trained on the COCO dataset in this research, applying COCO weights facilitates successful transfer learning with enhanced low-level and mid-level visual representations.

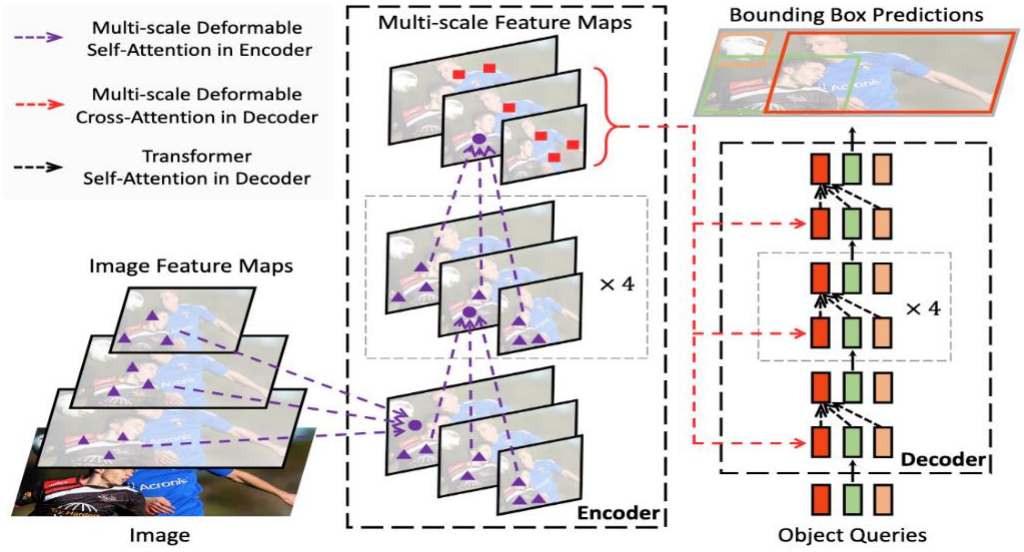


Figure 3-8 RF-DETR Architecture diagram for object detection (Sapkota, 2025)

Key advantages:

- Strong in detection of complex lesion patterns.
- Captures long-range relationships in images effectively.
- Robust to changes in lesion size and location.

Limitations:

- Slower conclusion speed due to transformer computations
- High computational requirements for training and fine-tuning

Altogether, these models provide complementary advantages in terms of speed, precision, and flexibility. Being included in this study allows for an extensive evaluation of how different architectural advancement and training strategies influence lesion detection performance in prostate PET/CT imaging.

3.6 Model Training and Fine-Tuning

The training process in this study followed a standard workflow to ensure consistency in all selected object detection models. At first, all PET/CT prostate dataset images were preprocessed, normalized, and resized according to the input requirements of each architecture, then, four pre-trained object detection models YOLOv11, YOLOv12, Roboflow 3.0, and RF-DETR were

evaluated using the annotated PET/CT prostate cancer dataset to test their initial detection performance.

After initial test, RF-DETR was selected to proceed with fine-tuning because of its superior capability to identify small lesions, handle global context through its transformer-based architecture, and reach better localization accuracy. The following sections will focus on RF-DETR, as it achieved an ideal trade-off between detection performance and computational efficiency.

For RF-DETR, fine-tuning was employed for the pre-trained weights using a learning rate schedule adjusted for convergence without overfitting. Data augmentation methods like rotation, flipping, and intensity scaling were tried to improve model generalization. The model was trained through several epochs with early stopping depending on validation loss, and checkpoints were saved to ensure repeatability and enable recovery in case of interruption. Combining these techniques produced a solid and reliable model, prepared for performance evaluation using standardized metrics.

The training and fine-tuning of the selected object detection models were employed with carefully chosen hyperparameters to guarantee best possible performance and repeatability. Table 3-1 provides a brief description of the main elements of the training environment of YOLOv11, YOLOv12, Roboflow 3.0 (OpenYOLO), and RF-DETR, it includes input size, batch size, learning rate, optimizer, number of epochs, data augmentation strategies, pretrained weights, and early stopping criteria. Although all models were initially evaluated on the dataset, full fine-tuning was only performed on RF-DETR, because it indicated superior preliminary performance, especially in detecting small and subtle prostate lesions. The experimental setups, which were customized to meet the unique needs of each architecture and the features of the PET/CT dataset, are clearly summarized in the table below.

Table 3-1 key components of the training setup for trained models

Component	YOLOv11	YOLOv12	Roboflow 3.0 (OpenYOLO)	RF-DETR
Input Size	640×640	640×640	640×640	576×576 (model-recommended)
Batch Size	Platform-managed	Platform-managed	Platform-managed	Platform-managed

Learning Rate	Platform default	Platform default	AutoLR	Platform-managed
Optimizer	Platform-managed	Platform-managed	AdamW (default)	AdamW (default)
Epochs	Platform-managed	Platform-managed	Platform-managed	Platform-managed (early stopping applied)
Augmentation	flip	flip	flip	flip
Pretrained Weights	COCO	COCO	Roboflow base	COCO-pretrained DETR weights

Roboflow's hosted train abstracts away low-level hyperparameter settings like the number of batches, the learning rate scheduler, and the number of epochs. As such, the values are recorded as platform-level defaults, while the architecture-level values for input resolution and model name are explicitly defined

After completing the training and fine-tuning process, and establishing the experimental configurations, the next step is to quantitatively evaluate RF-DETR's performance on the PET/CT dataset using standardized evaluation metrics.

3.7 Evaluation Metrics

After the fine-tuning of RF-DETR, the model's detection performance was measured quantitatively. This study focuses on a single-lesion object detection situation, in which both accurate spatial localization and reliable detection are clinically critical. COCO-style object detection metrics were used to evaluate RF-DETR model performance, which is consistent with those adopted by RF-DETR.

3.7.1 Intersection over Union (IoU)

Intersection over Union (IoU) measures the spatial overlap between the predicted bounding box (B_p) and the ground-truth bounding box (B_{gt}):

$$\text{IoU} = \frac{| B_p \cap B_{gt} |}{| B_p \cup B_{gt} |} \quad (1)$$

where $| B_p \cap B_{gt} |$ is the area of overlap and $| B_p \cup B_{gt} |$ is the total area covered by both boxes.

A prediction is considered a true positive (TP) if a prediction's IoU is greater than a certain threshold. IoU is an essential measure in object detection, especially in medical imaging, because it measures both the detection and localization quality of lesions.

Example of overlapping predicted and ground-truth boxes is shown in Figure 3-9 below.

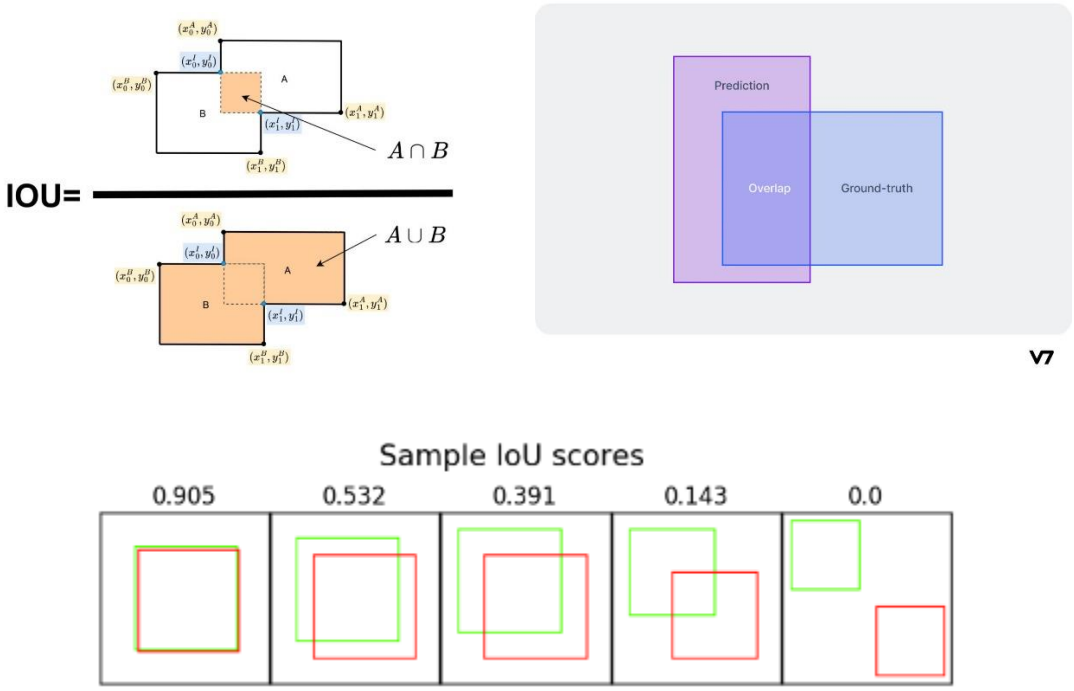


Figure 3-9 Example of overlapping predicted and ground-truth boxes.

3.7.2 Average Precision (AP) and Mean Average Precision (mAP)

Precision and Recall are defined as:

Precision: The percentage of correctly detected lesions among all predicted lesions.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{2}$$

Recall (Sensitivity): The percentage of actual lesions correctly detected by the model.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{3}$$

where $FP$is false positives and $FN$is false negatives.

Precision-recall curve for an example case of single lesion detection, Figure 3.10 illustrates the balance between precision and recall for different values of the confidence threshold; this is an indicator of how correctly the model identifies the lesion while suppressing false negatives and false positives.

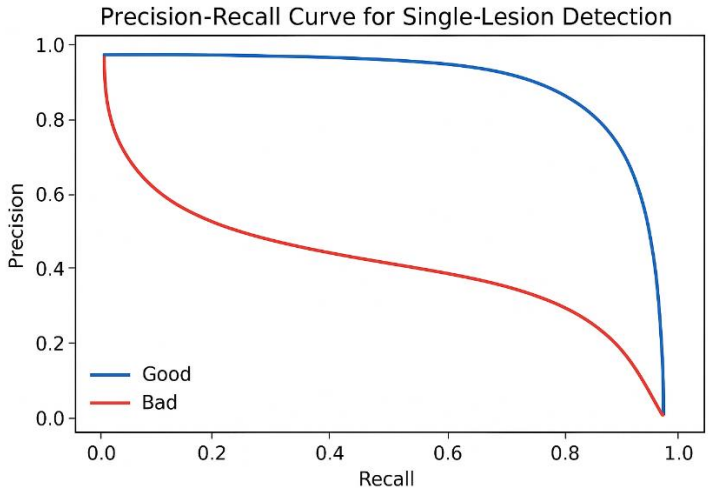


Figure 3-10 precision-recall curve for a single lesion detection example.

For prostate cancer detection with PET/CT images, recall (sensitivity) is the most important metric, since the primary objective here is trying to detect all true cancer lesions. Missing a cancer object (false negative) can have serious clinical consequences, whereas false positives mainly lead to additional diagnostic procedures, which are less critical.

The Average Precision (AP) summarizes the precision-recall curve at a certain IoU threshold. The mean Average Precision (mAP) aggregates AP across many IoU thresholds:

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i \tag{4}$$

where *N*is the number of IoU thresholds (e.g., 0.50 to 0.95 in increments of 0.05 for COCO-style evaluation).

In this study:

$$mAP@50{:}95 = \left(\frac{1}{|T|}\right)\sum_{\{t\in T\}} AP(t) \tag{5}$$

This metric restricts weakly aligned bounding boxes and prioritizes accurate lesion localization, which is essential for clinical diagnosis and treatment planning.

3.7.3 F1-Score and Specificity

The F1-score combines recall and precision to offer a balanced measure of detection performance:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{6}$$

Specificity measures how correctly the model can detect negative regions (i.e., avoid false positives):

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{7}$$

where $TN$is true negatives.

- Precision and specificity translate the model's ability to reduce false-positive detections.
- Recall evaluates sensitivity to actual lesions, reducing missed diagnoses.
- F1-score balances precision and recall, providing more insight into detection reliability in a single-lesion context.

Together, these evaluation metrics enable a comprehensive evaluation of RF-DETR's performance on the PET/CT dataset, covering both localization accuracy and detection reliability, which are essential for clinically relevance lesion detection.

3.8 Ethical Considerations

All patient data was completely de-identified before processing. The medical institution formally approved the dataset use, ensuring adherence to research ethics and data protection regulations.

3.9 SUMMARY

This chapter described the methodological pipeline of the research study, including dataset development, preprocessing, selection of pre-trained models, training procedures, and evaluation metrics. The following chapter presents the experimental results and evaluates the performance of the selected model.

CHAPTER FOUR

EXPERIMENTS AND RESULTS

In the earlier chapters, a number of object detection models were introduced as representatives of state-of-the-art approaches: such as YOLO11, YOLO12, Roboflow 3.0, and RF-DETR. These are some of the models used for the purpose of this study as baseline detectors, whereby quantitative performance comparison is enabled. Consequently, this chapter provides an extensive evaluation of the mentioned models by means of standard metrics of object detection. Special attention is paid to Average Precision and its variants.

This chapter describes the experimental setup, training procedures, and the outcome of the selected model for single-lesion detection. The purpose of these experiments is to evaluate the model's capacity in detecting lesions then localize them correctly. For this reason, the model's performance is measured using the evaluation metrics described in Chapter 3, while various qualitative and quantitative analyses are provided in order to demonstrate the effectiveness and reliability of the proposed approach.

4.1 EXPERIMENTAL SETUP

The experiments were performed on the Roboflow platform using pre-trained object detection models available. These models were fine-tuned with the selected PET/CT dataset, which enabled the use of solid feature representations that have been learned from large-scale datasets, which improves generalization and decreases training time.

Model training was performed in the Roboflow environment, with GPU acceleration utilized to optimize computational efficiency. Major parameters, including learning rate, batch size, number of epochs, and augmentation strategies were figured out through the preliminary experiments to ensure proper convergence. Data augmentation included image rotation, flipping, brightness adjustment, and scaling which helped to make the model more solid or less vulnerable to variations in lesion appearance.

Roboflow's cloud environment makes use of NVIDIA GPUs and enough RAM to manage common object detection pipelines, enabling models to train effectively on the given dataset (Peri, 2025).

4.2 Dataset and Splits

The PET/CT dataset was split into training, validation, and test sets to ensure objective and unbiased evaluation. The statistical distribution of the dataset across splits is summarized in figure 4-1.

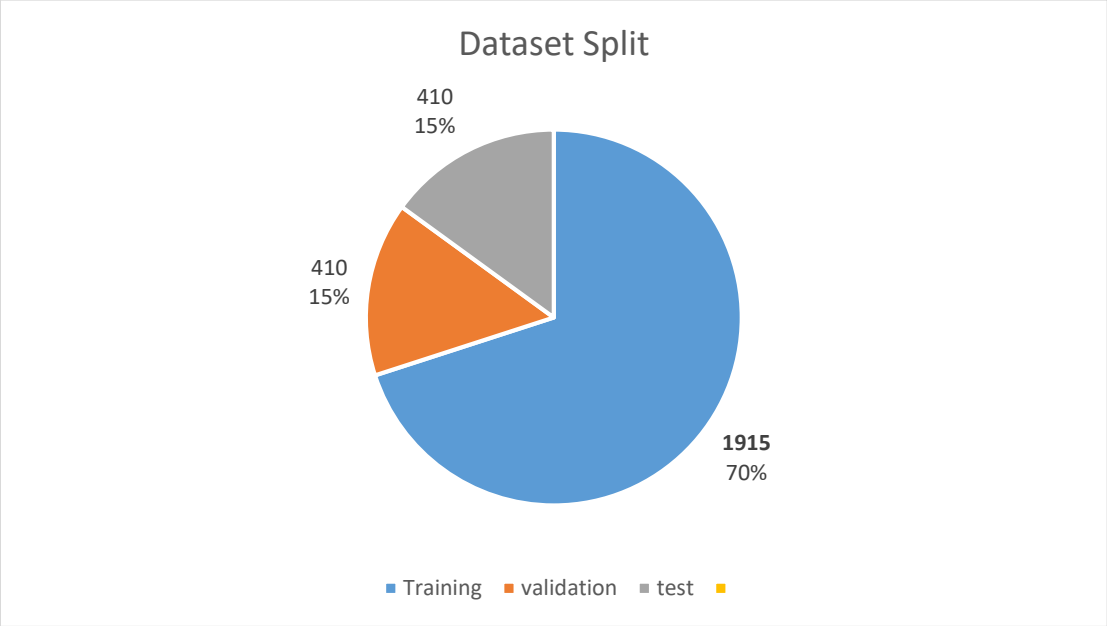
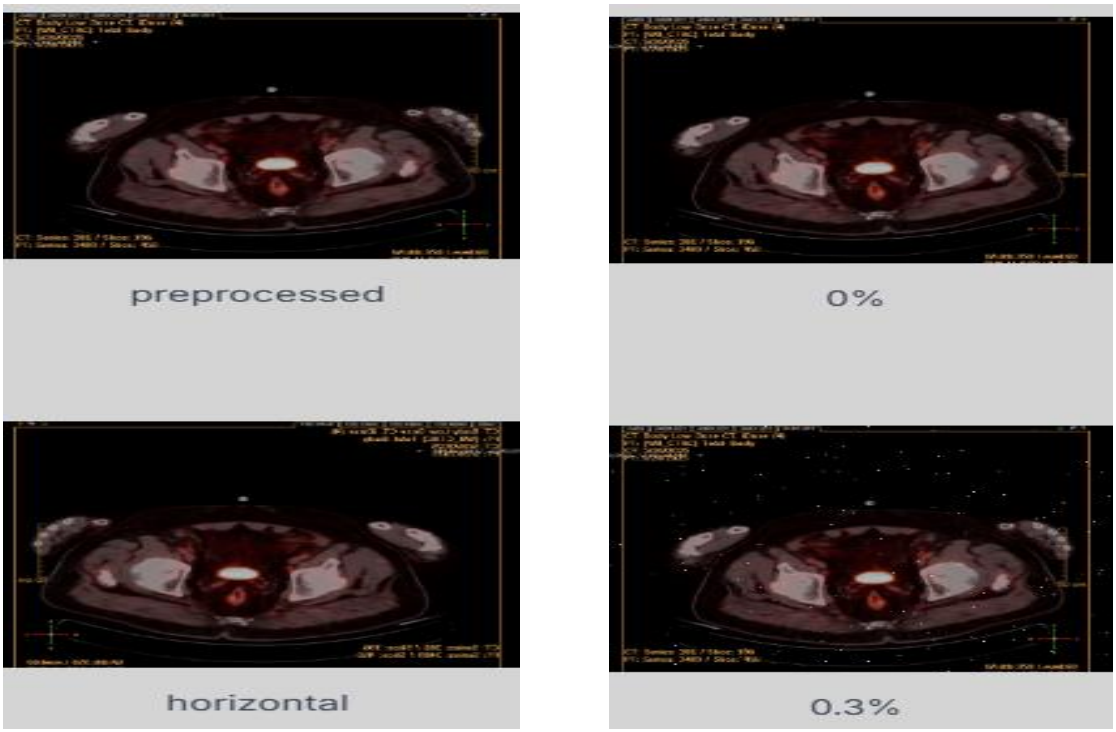


Figure 4-1 Dataset Splits

The images were initially resized to 576 × 576 pixels to meet model input requirements, then normalized for stabilizing the training process, and data augmentation to improve resilience and reduce overfitting by simulating changes in lesions appearances.

4.3 Training Details

In order to improve model generalization and robustness, multiple data augmentation techniques were implemented on the training dataset. These included horizontal flips, brightness and contrast adjustments, addition of noise and saturation changes. These augmentations increased the variety of training samples, helping the model to manage with variations in lesion appearance and images conditions (Figure 4-2 and Figure 4-3).



(a) Flip (b) Noise

Figure 4-2 Example of data augmentations applied during training, including horizontal flips, brightness and contrast adjustments, addition of noise and saturation changes

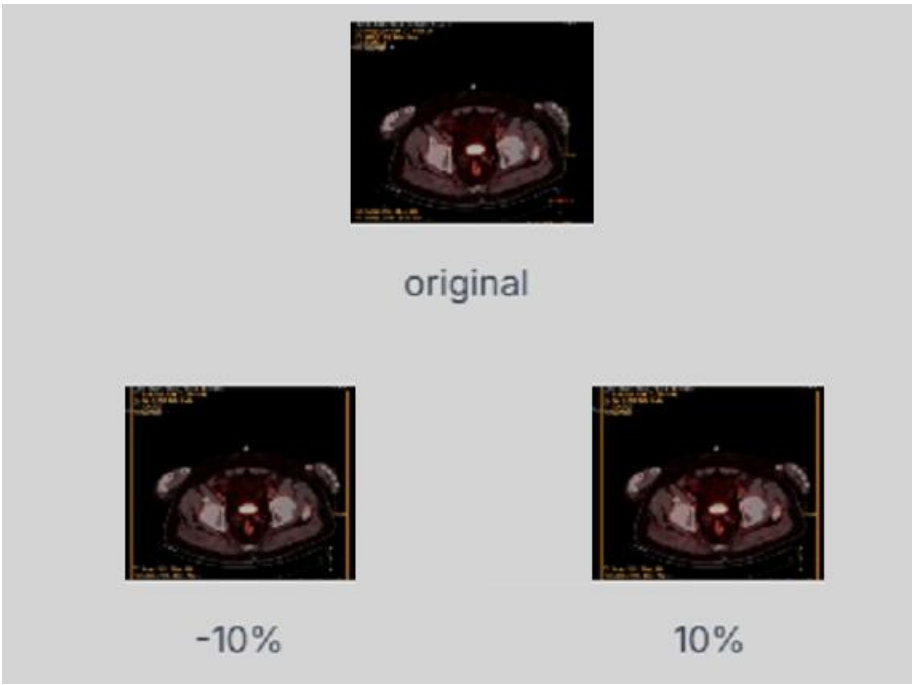


Figure 4-3 Example of saturation augmentation applied during training, showing random adjustments to the vibrancy of colors in the image

After applying these augmentation techniques, the model was trained and monitored with loss and detection performance metrics to evaluate learning progress and convergence.

The Roboflow platform's default loss functions and optimization techniques were used to refine the Roboflow pre-trained selected model. Validation performance was used to monitor the training, and early stopping employed to prevent overfitting. Also, model checkpoints were saved periodically.

Figure 4-4 shows Training and validation loss across epochs, with early stopping applied at the minimum of validation loss to prevent overfitting

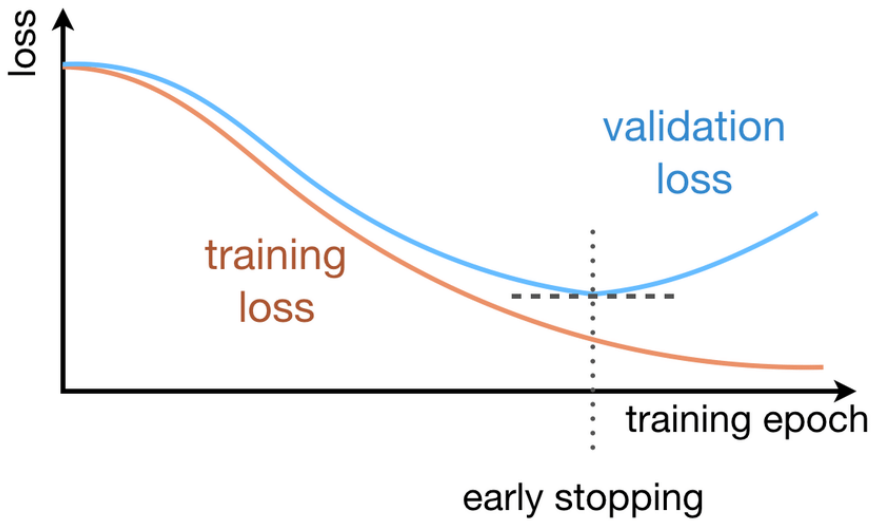


Figure 4-4 Training and validation loss across epochs, with early stopping applied at the minimum of validation loss to prevent overfitting

Figure 4.5 shows the training and validation loss curves, illustrating the model's convergence behavior over the training process. As shown in the figure, both mAP and mAP@50:95 metrics are significantly improved during the early epochs, which is a clear indication that the pre-trained network is effectively relearning and adapting to the lesion detection task. The subsequent stabilization of the curves implies that the training has converged, and no performance degradation can be observed. The mAP@50:95 values that are always lower than mAP indicate the increased challenge of achieving accurate localization at stricter IoU thresholds, which is a typical situation in medical lesion detection.

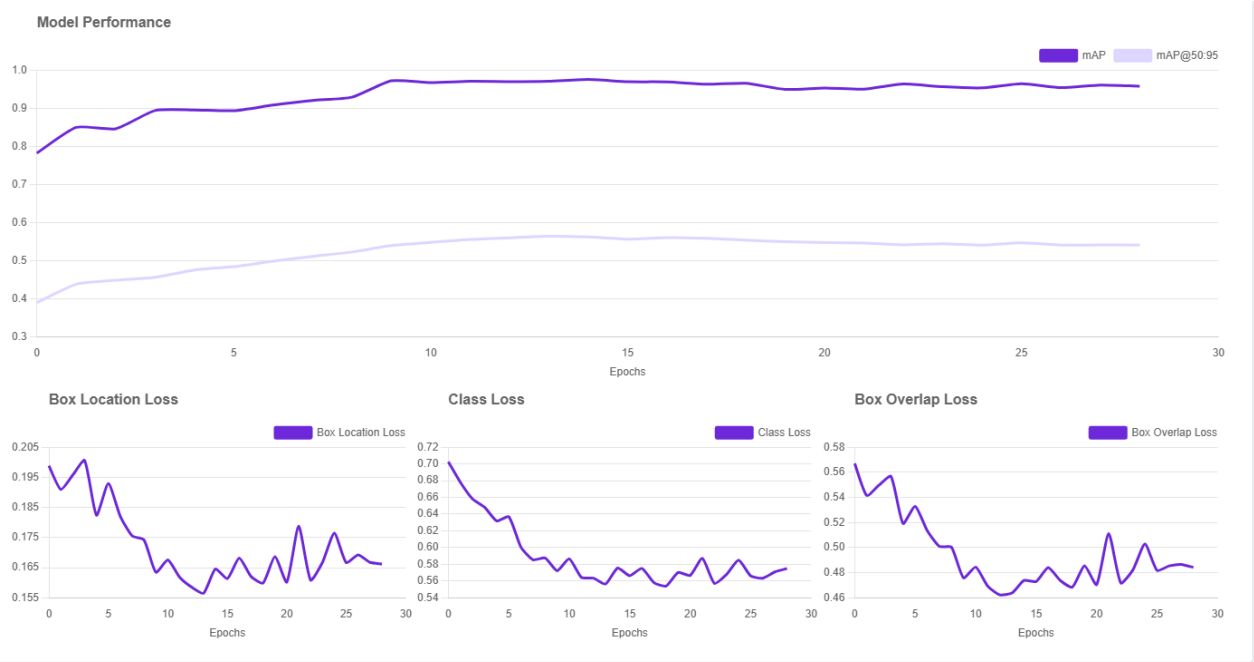


Figure 4-5 Training curves of the RF-DETR model illustrating the progression of mAP and mAP@50:95 across epochs, along with the corresponding box localization loss, classification loss, and box overlap loss during training.

The loss curves also offer further insights into the optimization process. The localization loss box gradually decreases, which indicates that there is better spatial matching between the predicted and ground truth bounding boxes. The classification loss box also remains focused on reducing and improving the class boundaries, then overlap loss box keeps reducing with time and improving overlap between bounding boxes.

The training curves apparent stability indicates successful convergence of the fine-tuned model. The following section provides a quantitative and qualitative assessment of the trained model on the reserved test dataset.

4.4 Computational Time Analysis

Apart from the analysis of the detection performance, the computational time cost was also investigated to assess the feasibility of the proposed model. The RF-DETR object detection model took approximately 8 hours to train on the Roboflow cloud-based training platform with a dataset of 2,735 PET/CT images.

Though this training time may seem relatively long, it is mainly due to the computational complexity of transformer-based object detection models, which are designed with multi-head attention mechanisms and deep feature extraction networks. It should be noted that the training time is offline and does not impact real-time clinical applications.

In real-world clinical applications, the inference time is more important than the training time. During the evaluation of the unseen PET/CT images, the trained model showed efficient prediction performance with an average inference time of 1-5 seconds per single image in the cloud-based environment. A slight delay can be related to network communication and cloud processing delays; however, this delay is still within the acceptable limits of clinical applications. Therefore, despite the computational complexity of the training process, the proposed transformer-based detection model is both feasible and applicable for incorporation into clinical diagnostic applications.

4.5 Experimental Results

4.5.1 Baseline Models Performance Comparison

Before choosing the suitable detection model for prostate PET/CT dataset, several of the modern object detection models, such as YOLO11, YOLO12, RF-DETR and Roboflow 3.0, have been evaluated. This is summarized in Table 4-1, which compares the performance of the models on the chosen dataset using standard metrics including mAP@50, mAP@75, mAP@50:95, precision, and recall. The results influenced the selection of the proposed RF-DETR model.

Table 4-1 Initial Models Performance

Model	Input Size	mAP@50	mAP@75	mAP@50:95	Precision	Recall	F1	Confidence
YOLO11	640*640	94.2%	54.4%	51.0%	97.6%	84.5%	90.6%	58%
YOLO12	640*640	89.6%	52.9%	51.3%	94.5%	88.7%	91.5%	58%

Roboflow 3.0	576*576	89.1%	47.9%	49.6%	92.8%	86.8%	89.7%	58%
RF-DETR	576*576	92.2%	45.7%	49.5%	92.5%	91.4%	92.0%	58%

Table 4.1 highlights that while all four models performed well in terms of detection performance, RF-DETR performed better than the other models in stricter IoU thresholds, supporting its selection for subsequent experiments.

4.5.2 Proposed Model Quantitative Evaluation

The RF-DETR model was evaluated on the test set with COCO-style metrics using precision, recall, F1-score, and mean Average Precision (mAP) to evaluate both detection and localization performance of lesions. Specifically, mAP@50 and mAP@50:95 were reported to measure detection accuracy at standard and stricter IoU thresholds increased respectively. Since the study focuses on single-object detection tasks, it was a priority to focus on high recall and strong precision to ensure reliable detection in clinical environments.

The quantitative performance of the fine-tuned model is summarized in Table 4.2.
The results correspond to the optimal confidence threshold calculated by figuring out the best model performance where the model is the most accurate across all classes. It optimizes F1 score, which balances precision and recall.

Table 4-2 Quantitative performance of the RF-DETR model

Metric	Value
Precision	93.9%
Recall	92.6%
F1-score	93.2%

mAP@50	93.4%

The Model showed strong detection ability which simultaneously achieves high recall and precision. Both of which are crucial for accurate cancer detection in clinical imaging.

In addition to evaluating the detection performance on single threshold, the performance was also evaluated based on different thresholds of IoU and object sizes.

Table 4.3 presents the Average Precision at IoU thresholds of 0.50 and 0.75, also the COCO-standard mAP@50–95 metric, which offer the mean AP across IoU thresholds from 0.50 to 0.95 in increments of 0.05.

Table 4-3 Average Precision at IoU thresholds of 0.50 and 0.75, also the COCO-standard mAP@50–95

Metric	Value
mAP@50–95	50.5%
mAP@75	49.6%
mAP@50	93.4%

Also Figure 4.6 shows the distribution of average precision for various IoU thresholds such as mAP@50, mAP@75, and mAP@50:95, also object size categories of small, medium and large.

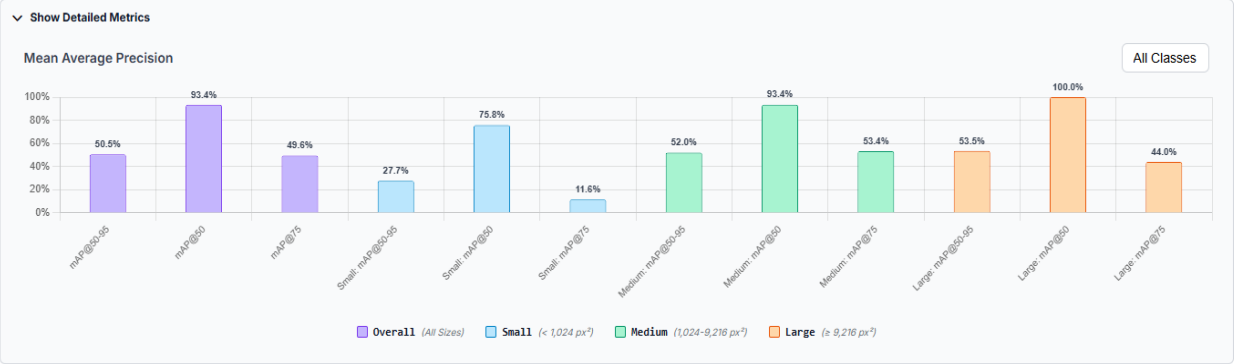


Figure 4-6 Average Precision at different IoU thresholds and object sizes.

The high AP@50 value shows excellent detection ability and ensures that the model is able to predict the lesion areas with high overlapping threshold constraints. With the increase of the IoU threshold to 0.75, the lowered AP value reveals the higher localization constraint requirement between the predicted and the correct bounding box values.

A lower mAP@50-95 score is what is to be expected in this task as this score combines performance over IoU thresholds that are increasingly strict. In the field of medical imaging, the boundaries of lesions are typically fuzzy and hard to precisely distinguish because of low contrast and the effects of the partial volume principle and inter-observer variability of annotations. A small deviation in localization can drastically influence the AP with IoU thresholds that are high and result in a lower average score. Nonetheless, the very high AP@50 indicates that this model performs well in terms of detecting clinically meaningful information.

A systematic evaluation of the impact of the confidence threshold on detection performance was conducted. The correlation between precision, recall, and F1-score and the confidence level is shown in Figure 4.8. The F1 score achieves its maximum value at 58%, which was determined to be the ideal confidence level.

Recall is strong at lower confidence levels, but there are more false positives as well. On the other hand, higher thresholds increase precision at the expense of decreased recall. Therefore, choosing the ideal threshold allows for a fair trade-off acceptable for the single-lesion detection task covered in the present study.

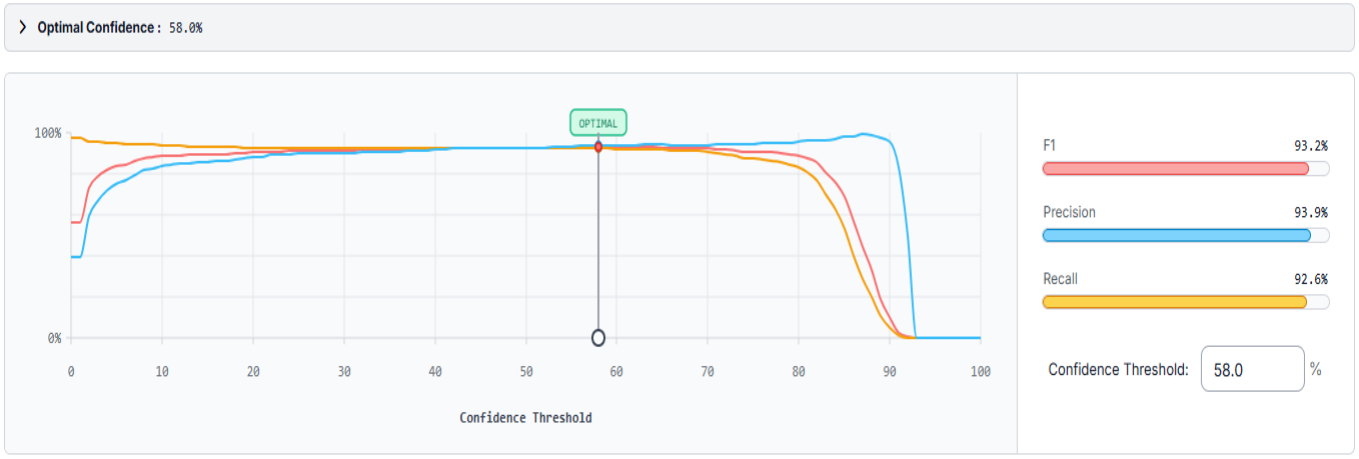


Figure 4-7 Precision, Recall, and F1-score as a function of the confidence threshold.

Performance analysis at the class level is shown in Figure 4.8. This figure summarizes the number of correct detections, the number of false positives, and the number of false negatives on the target class. This information helps in understanding the type of errors occurring in the detections. This ensures that most of the detections are true positives.

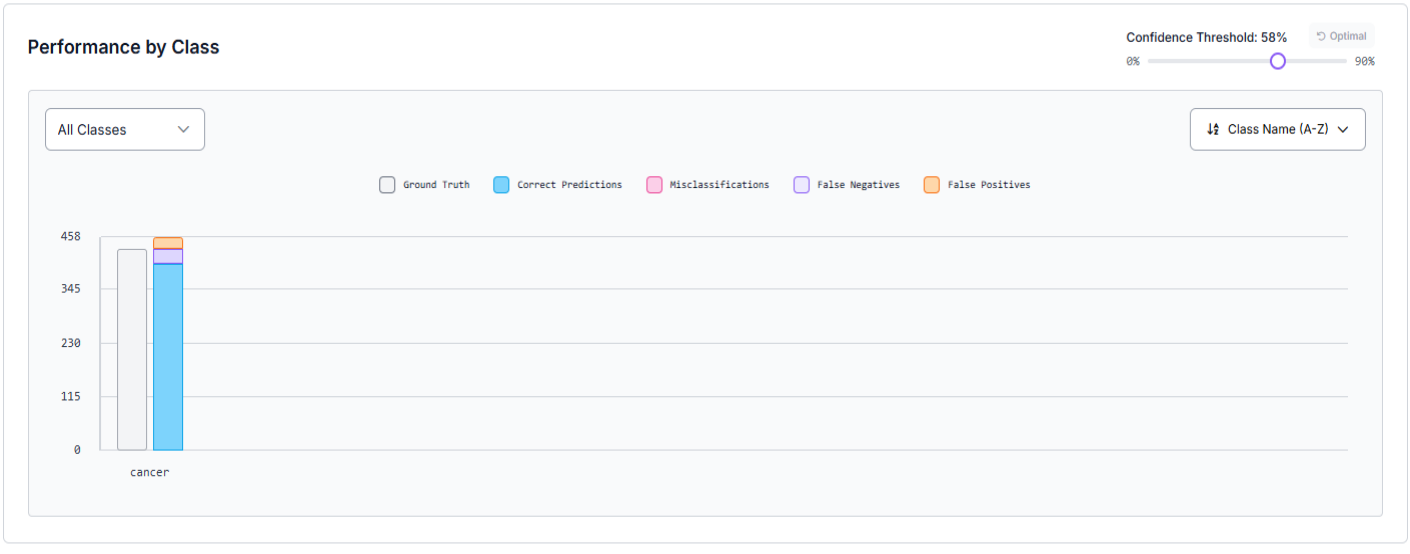


Figure 4-8 Performance by class at the optimal confidence threshold.

For more precise analysis of the classification behaviors, the confusion matrix presented below in Figure 4.9 has been used. This confusion matrix shows the relationship between the actual classifications and the predicted classifications made by the model. The matrix shows the number of true positives, false positives, and false negatives on the test set. Reliable lesion detection is confirmed by high true positive numbers.

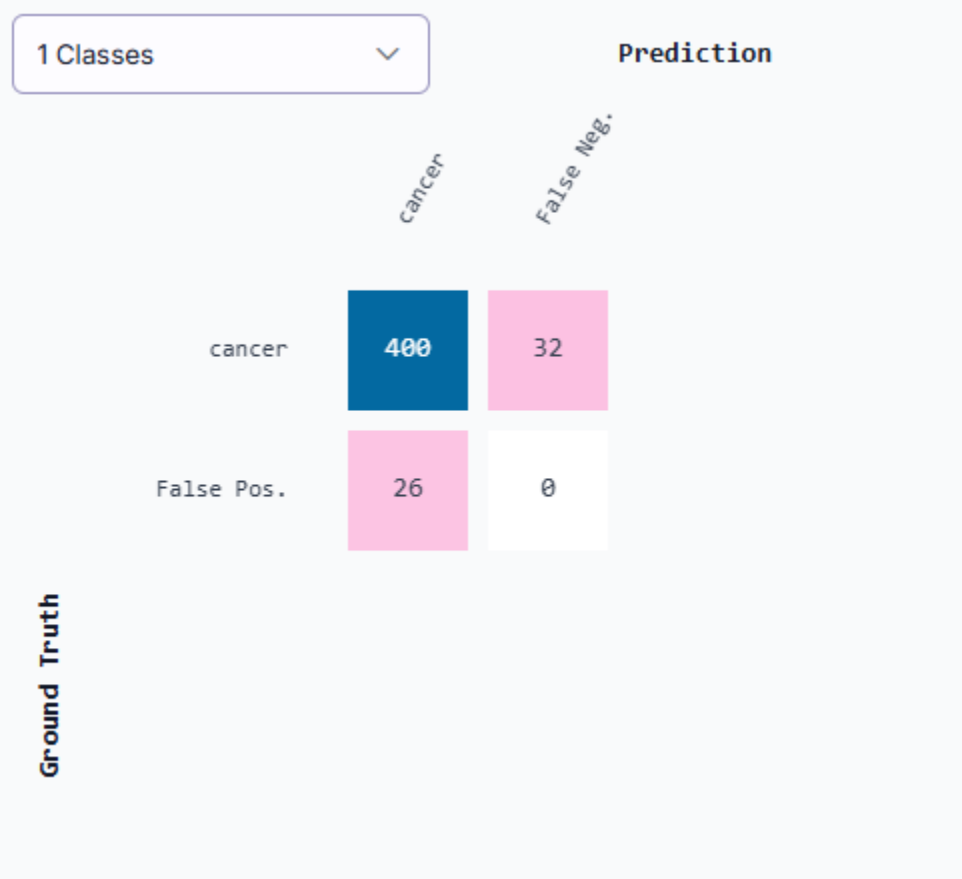


Figure 4-9 Confusion matrix of model predictions on the test set.

By this matrix, the representation again reinforces the quantitative values represented in Table 4.2. which confirms the model's reliability.

The qualitative evaluation takes into consideration true positives (TP), false positives (FP), and false negatives (FN). The concept of true negatives cannot be clearly framed in terms of object detection since the regions of the background are not individually enumerated.

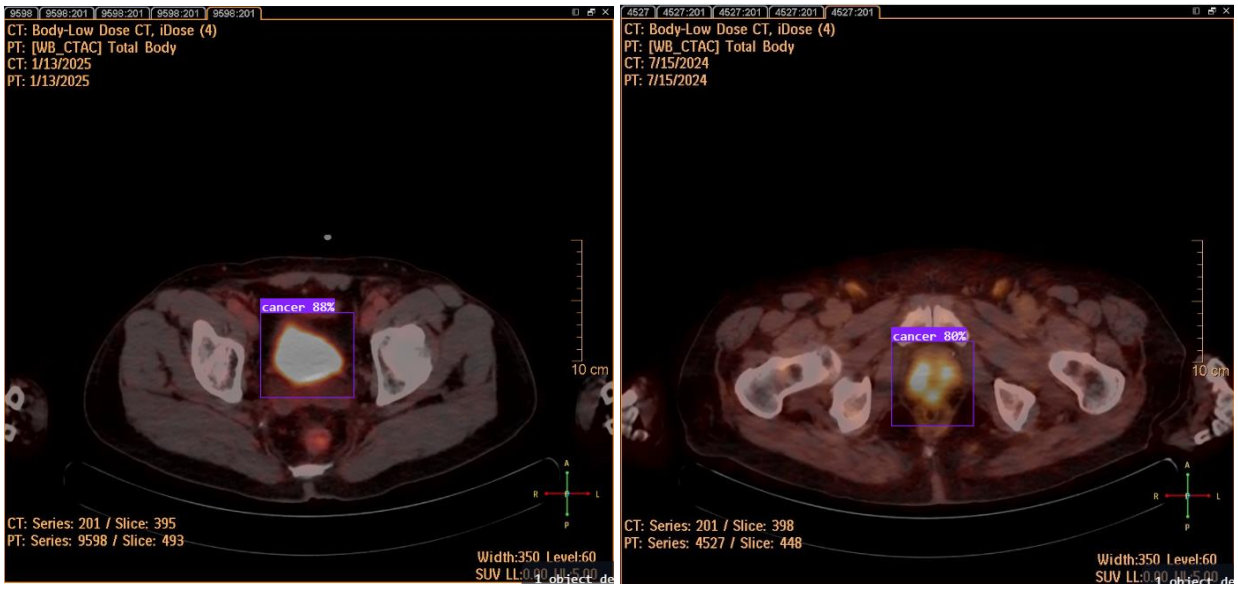
In general, the quantitative evaluation shows that the model can achieve high detection accuracy across various IoU thresholds, with strong precision and recall metrics. These results constitute a solid basis for further visual analysis in the qualitative evaluation section.

4.5.3 Proposed Model Qualitative Evaluation

Complementing the quantitative metrics presented in the previous Section 4.5.1, the representative qualitative results were analyzed in order to visually evaluate how the RF-DETR model performed during lesion detection on the test set. The evaluation includes general detection findings, generalization on external data, negative instances, small lesion difficulties, and explicit ground truth (GT) vs prediction comparisons are all included in the evaluation.

These examples offer insights into the model's localization accuracy, detection reliability, and it showed the model's behavior in both common and complex scenarios.

Figure 4.10 illustrates the general results of lesion detection, from the test dataset using model prediction. Predicting bounding box and confidence threshold displayed to show model's ability to detect lesions.



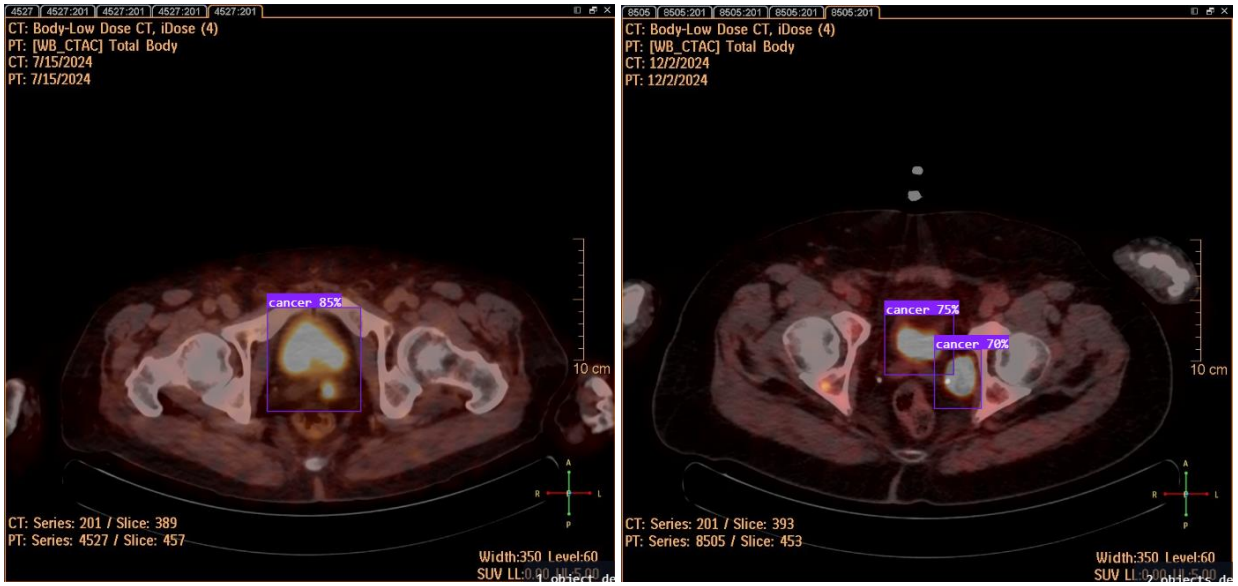


Figure 4-10 Example Detection Images

Detected lesions appear in representative images from the test set. Bounding boxes represent predicted lesion locations, indicating accurate localization with minimal false positives.

To further verify the model's generalization capacity, the model was also applied to an external image that was not included in the training or test sets (Figure 4.11).

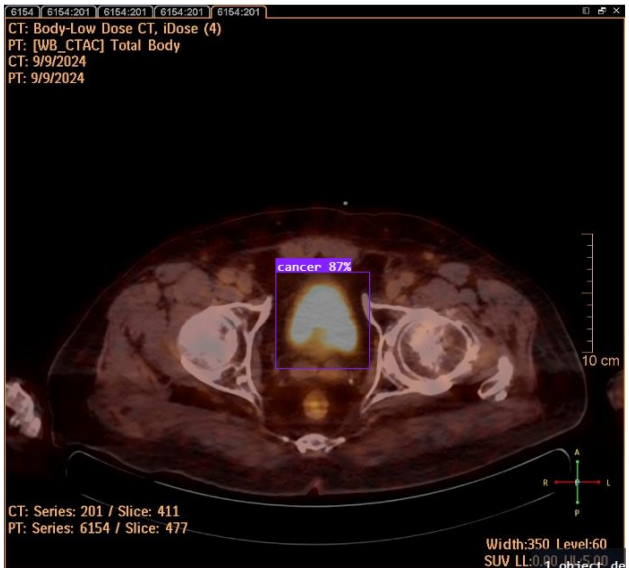


Figure 4-11 Detection on External Image

The successful detection on these external images represents the model's ability to generalize and remains effective on unseen imaging conditions, beyond the test dataset, which is critical for real-world clinical applications.

Negative test cases with no lesions were examined to further assess the model's resilience, and to verify the model does not produce false positives (Figure 4.12).

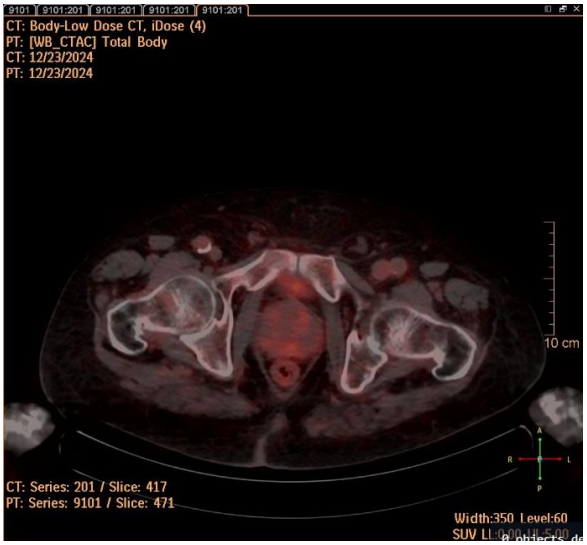


Figure 4-12 Example Negative Cases (No Lesions)

Images from the test set containing no lesions. The absence of predicted bounding boxes indicates the model's ability to decrease false positive detections.

Challenging cases with subtle or small lesions were examined to better understand the model's performance in complex cases (Figure 4.13). Some minor lesions are either completely ignored or have a lower localization accuracy. Under more stringent IoU criteria, these instances lead to lower results and are the main cause of false negatives.

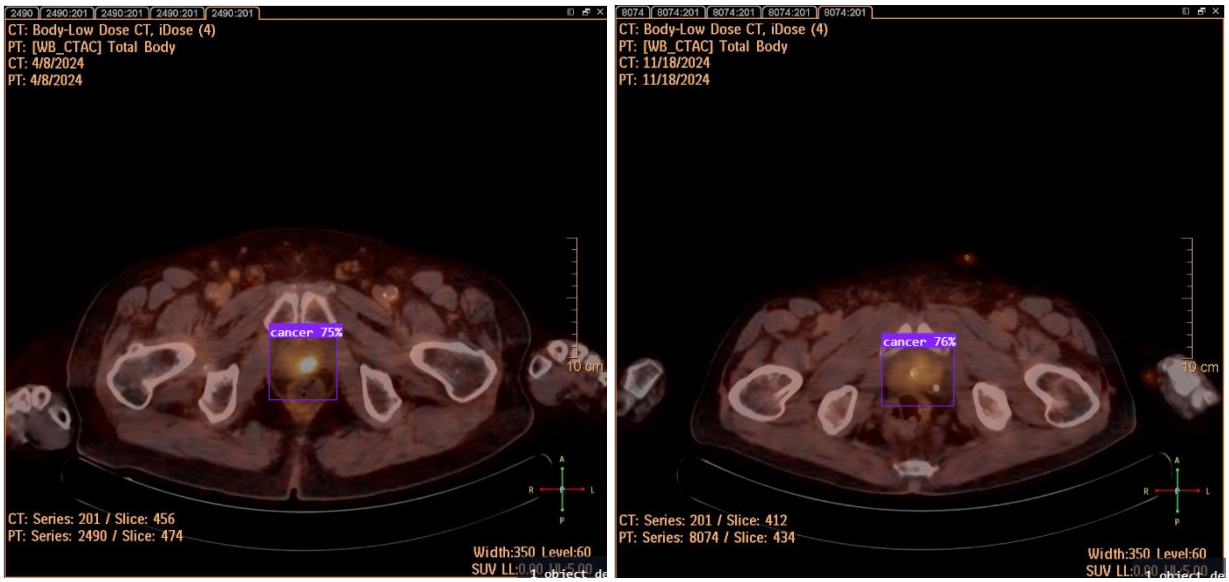


Figure 4-13 Examples of small Lesion Cases

"Images showing small or low-contrast lesions. Bounding boxes reflect the model's ability to identify small lesions that may be difficult to detect."

Following the visual detection findings on test and external images, also the localization performance was evaluated by analyzing the overlap between predicted and ground-truth bounding boxes using the Intersection over Union (IoU) metric.

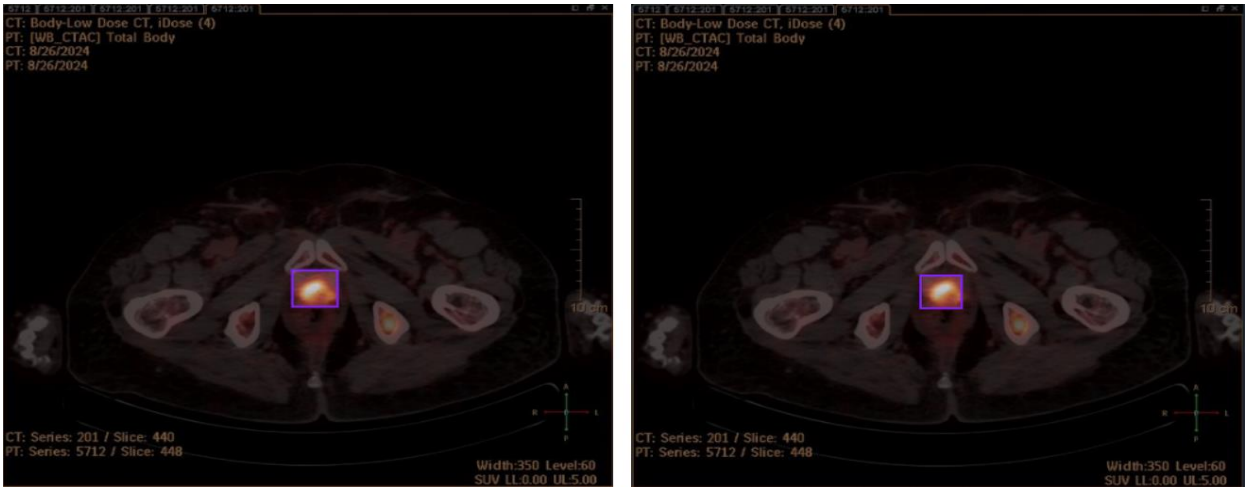
Figure 4.15–4.17: illustrates a direct comparison between ground truth (GT) annotations and the model prediction, which classifies detection results into True Positive (TP), False Positive (FP), and False Negative (FN). For each category, subfigure (a) shows the ground truth (GT) annotation, while subfigure (b) presents the model prediction.

TP: Correctly localized lesions with sufficient overlap with GT.

FP: Incorrect lesion localization on images or areas without GT labels.

FN: The lesions that are annotated and not found by the model.

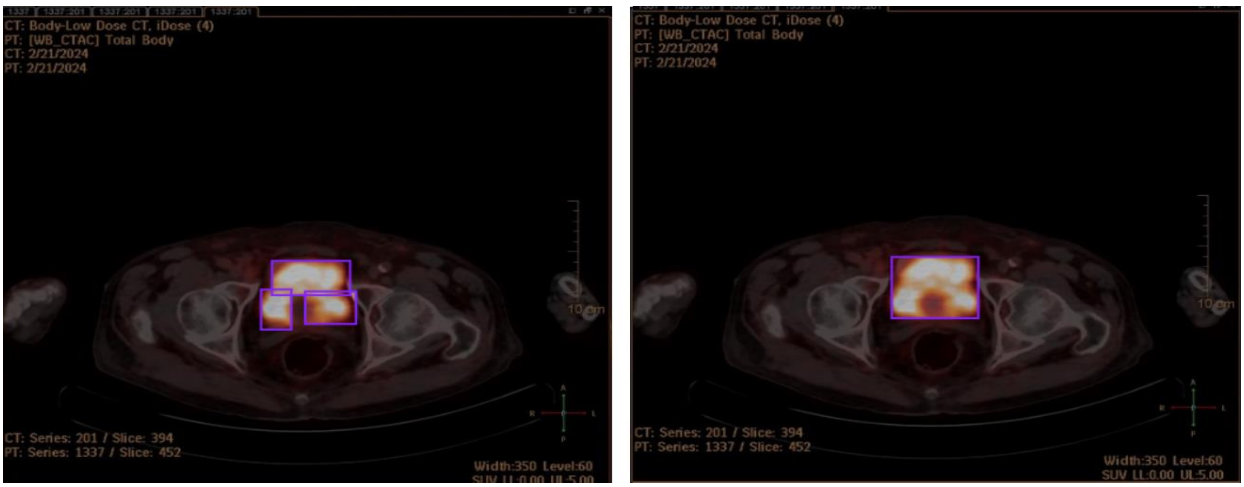
Figure 4.14 gives a few representative true positive results. The predicted bounding boxes very closely match the GT annotation, which shows the accurate localization of the lesions.



(a) Ground truth (b) Predicted

Figure 4-14 Examples of correctly detected lesions (true positives). (a) Ground truth annotations. (b) Predictions obtained from the model.

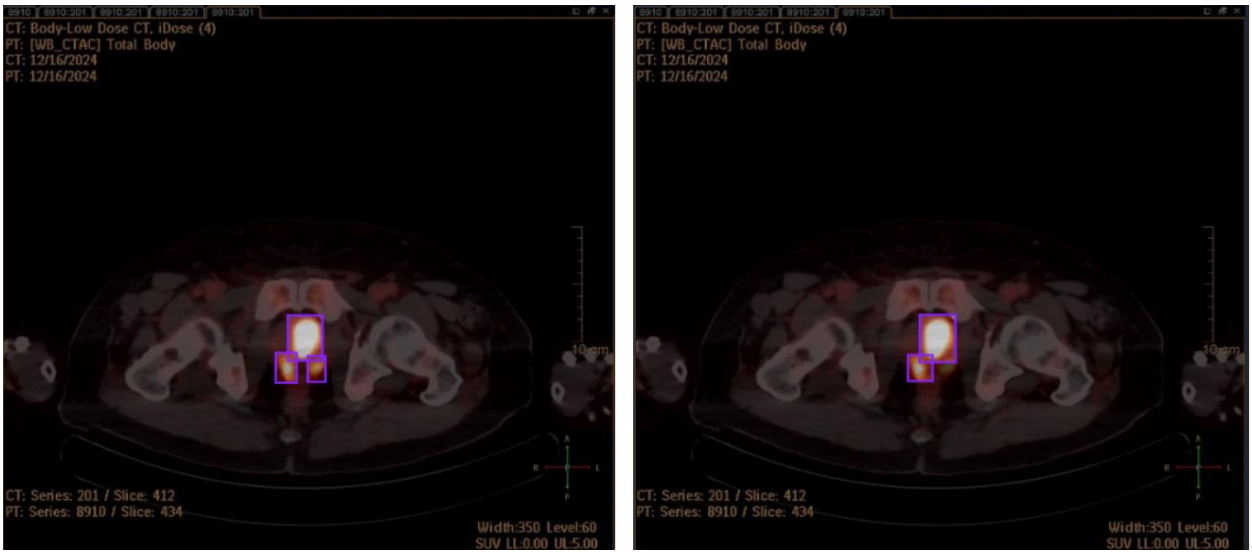
Figure 4.15: False positive detections with the model predicting lesions in regions that have no GT annotations.



(a) Ground truth (b) Predicted

Figure 4-15 Examples of false positive detections. (a) Ground truth images different annotated lesions. (b) Corresponding false positive predictions.

Figure 4.16 shows the false negatives, that is, the lesions annotated in GT which were missed by the model.



(a) Ground truth (b) Predicted

Figure 4-16 Examples of false negative detections. (a) Annotations of ground truth lesions. Corresponding model predictions in (b) showing missed detections

This structured form of TP/FP/FN enables a clear understanding of the behavior of the model in a single-class scenario of object detection. True positives ensure the effectiveness of localization, false positives indicate the similarity in ambiguity among the background regions, and false negatives show challenges related to small and low-contrast lesions.

This breakdown of TP, FP, and FN can be easily correlated with the components of the confusion matrix, which calculate precision, recall, and mean Average Precision. This gives a profound insight into how the matrix is affected by the failures in the object detection tasks. Collectively, these qualitative insights explicitly validate the findings, which present an in-depth analysis of the strengths and limitations of the approach. The correctness in lesion segmentation for positive, negative, and difficult cases explicitly validates the practicality of the approach in real-world setups.

It is a means of bridging between the upcoming analysis section and enables a connection between system model dynamics and vectors used for measurement.

4.6 Analysis and Discussion

The quantitative analysis showed excellent performance and confirmed that the fine-tuned RF-DETR model could achieve high levels of detection performance, with strong precision and recall scores that confirmed reliable lesion detection and low false positives. the model achieved mAP@50 = 93.4%, precision = 93.9%, and recall = 92.6%, confirming effective lesion detection with moderate levels of localization. The confidence threshold was set to 58%, which provides the balance between precision and recall for optimal overall lesion detection. Conversely, the mAP@50:95, which combines the results of the stricter levels of IoU, was lower (as illustrated in more detail in Section 4.5.1), since it is more difficult to accurately detect the positioning of the bounding box, particularly in smaller or less visible lesions.

These observations are further complemented by the qualitative analysis presented under Section 4.5.2. The examples of lesions detected from the test set are found to have good localization with few false positives, and successful lesion localization on external testing images indicates ability of the model to generalize to previously unseen data. The false positives are few, including only lesions on complicated backgrounds, and false negative cases mainly involved small or low-contrast lesions, as illustrated in Figures 4.15–4.17. IoU analysis indicates significant overlap between the actual and prediction bounding boxes, though slight boundary deviations in challenging instances. The absence of lesions in the negative examples shows that false detections have been effectively minimized.

For a better understanding of detection behavior, a vector analysis of the detected lesions is also included in Figure 4.17. In this figure, lesions are aggregated into clusters based on their F1 score (dark = poor detection, light = high F1), which reflects their general detection performance. Each cluster summarizes the distribution of true positives, false positives, and false negatives in the group, thus giving insight into the cases that are most contributing to the model errors. The approach shows the variation in performance for lesion types, sizes, and imaging conditions and visually links the qualitative behavior of the model to the quantitative metrics presented in Section 4.5.

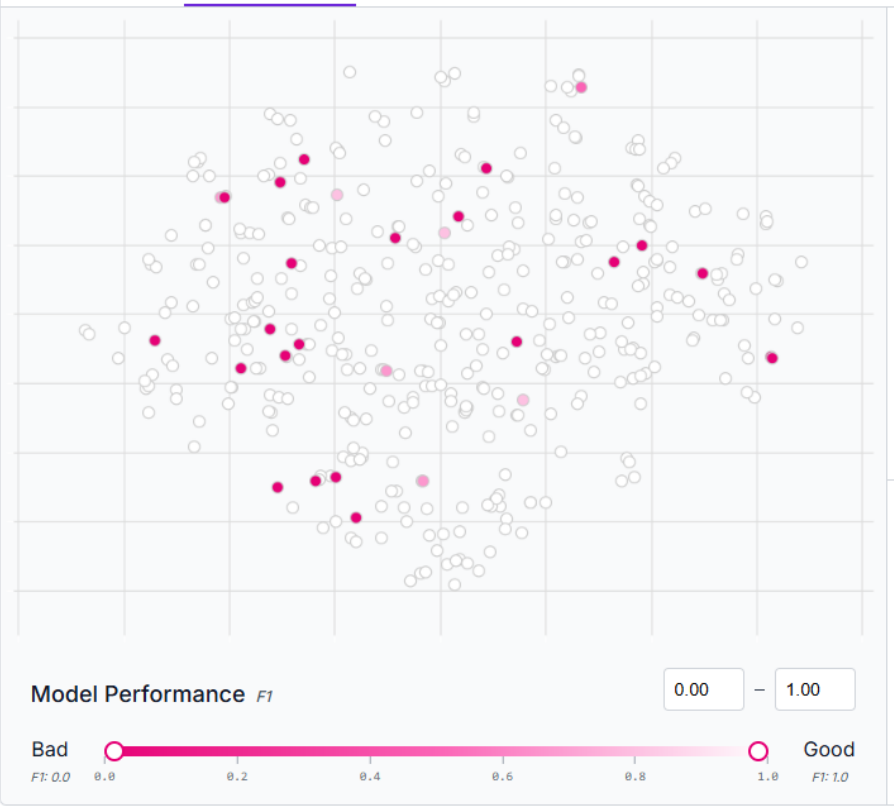


Figure 4-17 Clusters of lesions grouped by F1 score, highlighting detection performance variation across the test dataset and showing which cases contribute most to model errors.

The model indicates a strong trade-off between generalization capacity and detection performance. High TP rates and stable AP values indicate reliable lesion detection, whereas low FP rate indicates low false alarm probability. The slight decrease in mAP@[50:95] can be explained by false negative patterns, especially for small or subtle lesions, indicating that the main difficulty is localization precision. The TP/FP/FN vector distribution explains both precision-recall trade-offs and the variations in AP across different IoU thresholds.

Taken together both quantitative and qualitative aspects, it is easy to verify that these results demonstrate that the RF-DETR approach is effective for lesion detection, and that there is still scope for improvement with respect to precise localization. Some of these improvements may be achieved through more feature augmentation, and post-processing techniques.

The analysis concludes that the RF-DETR model is a useful instrument for automated lesion detection. It achieves high sensitivity (92.6%) and precision (93.9%), with mAP@50 = 93.4% and

optimal confidence at 58%, demonstrating low false positive rates and excellent generalization to external datasets. At higher IoU thresholds, spatial localization accuracy is still the primary constraint, especially for small or low-contrast lesions. However, the combination of qualitative examples and TP/FP/FN vector analysis offers a thorough understanding of detection behavior, establishing a solid foundation for further refinements and possible clinical deployment.

CHAPTER FIVE

CONCLUSIONS AND FUTURE WORK

The experiments and analysis results in Chapter 4 show that the fine-tuned RF-DETR model is able to effectively detect and localize lesions from medical images. The effectiveness in detecting lesions was reconfirmed by both quantitative metrics and qualitative evaluations on typical, minor, and negative cases. While the results emphasize the model's strengths, they also demonstrate few challenges in exact localization for small or low-contrast lesions and in generalization to external datasets. The following sections will summarize the important findings and draw conclusions, discuss limitations, and propose possibilities for future work.

5.1 CONCLUSIONS AND KEY FINDINGS

The study fine-tuned the RF-DETR model for single-lesion object detection using PET/CT medical imaging dataset.

The carefully fine-tuned RF-DETR model is effective in detecting and localizing lesions objects in medical images, making it a useful and promising tool for supporting clinical diagnosis and contributes in healthcare treatment management.

Quantitative evaluation demonstrated excellent detection accuracy with mAP@50 ≈ 95% and good localization performance (mAP@50:95 ≈ 55%).

Qualitative evaluation demonstrated reliable lesion localization, minimum false positives, and resilience in normal, minor, and challenging cases.

External image testing indicated the model's ability to generalize to unseen external images, but with slightly lower performance, indicating areas for improvement.

5.2 LIMITATIONS

The model performance heavily impacted by the quality and diversity of annotated datasets.

Small, subtle, or inconsistent lesions still can be challenging in some complex cases, which can reduce localization accuracy at stricter IoU thresholds.

Testing on external datasets showed some performance decrease, indicating that the model's generalization ability can be affected.

5.3 FUTURE WORK

Various approaches can be explored to improve the performance and applicability of the proposed lesion detection solution.

Future work may focus on using expanded datasets adding more diverse and higher-quality annotations, such as uncommon and unusual lesion types, this would help the model to better capture diversity and improve generalization to unseen clinical cases.

Additional performance improvements, especially in localization accuracy, can be accomplished by experimenting with architectural improvements like multi-scale feature enhancements, attention mechanisms, or specialized loss functions. These approaches might allow for more accurate bounding estimation, especially for small or low-contrast lesions.

Evaluating the model within real-world clinical operations could be an important next step. This involves evaluating computing efficiency, inference time, and usability, in addition to gathering feedback from specialists to assess the model’s practical value as guidance tool in clinics.

New research needs to evaluate how well the model performs on bigger sets of datasets gathered from several institutions, using different imaging machines and different imaging protocols. Such validation is required to determine how far this can be generalized and adopted in real-world clinical environments.

REFERENCES

Abu-Rmeileh, N. M. (2016). Cancer mortality in the West Bank, occupied Palestinian territory. *BMC public health}, 16*, 1-10.

Alabousi, A. W.-G. (2020). Impact of PI-RADS 3 lesions on the diagnostic accuracy of MRI for detecting prostate cancer and the prevalence of prostate cancer within each PI-RADS category: A systematic review and Meta-Analysis. *The British journal of radiology, 94*, 20191050. doi:10.1259/bjr.20191050

Alajerami, Y. a. (2015). Prostate Cancer Diagnostic and Evaluation in Gaza-Strip, Palestine. *Health, 7*(11), 1552.

Al-Ani, O. J. (2023). Ensemble Model for Prostate Cancer Detection Using MRI Images. *2023 16th International Conference on Developments in eSystems Engineering (DeSE)*, 492-497. doi:10.1109/DeSE60595.2023.10468866

AlWaheidi, S. (2023). Promoting cancer prevention and early diagnosis in the occupied Palestinian territory. *Journal of Cancer Policy, 35*, 100373. doi:https://doi.org/10.1016/j.jcpo.2022.100373

Ar, A. P.-Z.-R.-G.-M.-H.-D.-M.-B.-T. (2022). The promising role of new molecular biomarkers in prostate cancer: from coding and non-coding genes to artificial intelligence approaches. *Prostate Cancer and Prostatic Diseases, 25*(3), 431 - 443. doi:10.1038/s41391-022-00537-2

B.Burbridge, F. (2022). Computed Tomography (CT). *Comparative Veterinary Anatomy*, 47-56. doi:10.1515/iupac.87.0151

Bertelli, E. a. (2022). Machine and deep learning prediction of prostate cancer aggressiveness using multiparametric MRI. *Frontiers in oncology, 11*, 802964. doi:https://doi.org/10.3389/fonc.2021.802964

Besson, V. D. (2021). PET Molecular Imaging: A Holistic Review of Current Practice and Emerging Perspectives for Diagnosis, Therapeutic Evaluation and Prognosis in Clinical Oncology. *International Journal of Molecular Sciences, 22*(8), 4159. doi:10.3390/ijms22084159

Biancone, S. S. (2021). The role of artificial intelligence in healthcare: a structured literature review. *BMC Medical Informatics and Decision Making, 21*, 1-23. doi:10.1186/s12911-021-01488-9

Bonfrer, T. M. (1990). Prostate-specific antigen (PSA). A tissue-specific and sensitive tumor marker. *European journal of surgical oncology : the journal of the European Society of Surgical Oncology and the British Association of Surgical Oncology, 16* (1), 37-41.

Bordoloi, D. a. (2022). [Retracted] Deep Learning in Healthcare System for Quality of Service. *Journal of Healthcare Engineering, 2022*(1), 8169203. doi:10.1155/2022/8169203

Brinker, L. S.-P.-H. (2022). Integration of deep learning-based image analysis and genomic data in cancer pathology: A systematic review. *European Journal of Cancer, 160*, 80-91. doi:https://doi.org/10.1016/j.ejca.2021.10.007

Bristow, R. R. (2021). Prostate cancer. *Nature Reviews Disease Primers, 7*(1), 1-27. doi:10.1038/s41572-020-00243-0

Brown, D. M. (2018). Artificial Intelligence in Medical Practice: The Question to the Answer? *The American journal of medicine, 131*(2), 129-133. doi:10.1016/j.amjmed.2017.10.035

Buzdin, N. B. (2019). New Paradigm of Machine Learning (ML) in Personalized Oncology: Data Trimming for Squeezing More Biomarkers From Clinical Datasets. *Frontiers in Oncology, 9*, 658. doi:10.3389/fonc.2019.00658

Cai, T. L.-H. (2023). Multi-parametric MRI for radiotherapy simulation. *Medical physics, 50*(8), 5273--5293. doi:10.1002/mp.16256

Calais, I. S. (2020). Impact of 68Ga-PSMA-11 PET/CT on Staging and Management of Prostate Cancer Patients in Various Clinical Settings: A Prospective Single-Center Study. *The Journal of Nuclear Medicine, 61*(8), 1153 - 1160. doi:10.2967/jnumed.119.237602

Calais, J. N. (2024). Update from the PSMA-dRT trial: A randomized phase III trial of PSMA-PET prior to definitive radiation therapy for unfavorable intermediate-risk or high-risk prostate cancer. *Journal of Clinical Oncology, 42*(4), 291-291. doi:10.1200/jco.2024.42.4_suppl.291

Candasamy, S. B. (2023). Artificial Intelligence in Pharmaceutical and Healthcare Research. *Big Data Cogn. Comput, 7*(1), 10. doi:10.3390/bdcc7010010

Castillo T, J. M. (2020). Automated classification of significant prostate cancer on MRI: a systematic review on the performance of machine learning applications. *Cancers, 12*(6), 1606. doi:10.3390/cancers12061606

Center, U. H. (2024). *Digital rectal exam for prostate cancer*. Retrieved from hillman.upmc: https://hillman.upmc.com/cancer-care/prostate/screenings/digital-rectal-exam

Chen, S. a. (2022). Machine Learning-Based Models Enhance the Prediction of Prostate Cancer. *Frontiers in Oncology, 12*, 941349. doi:10.3389/fonc.2022.941349

Chen, W. a. (2024). A review of object detection: Datasets, performance evaluation, architecture, applications and current trends. *Multimedia Tools and Applications, 83*(24), 65603--65661.

Chiș, G. C.-C.-G. (2022). Radiopharmaceuticals for PET and SPECT Imaging: A Literature Review over the Last Decade. *International Journal of Molecular Sciences, 23*(9), 5023. doi:10.3390/ijms23095023

Choyke, E. M. (2022). The Impact of PSMA PET/CT Imaging in Prostate Cancer Radiation Treatment. *Seminars in nuclear medicine, 52*(2), 255--262. doi:10.1053/j.semnuclmed.2021.12.008

Christian Bach, S. P. (2014). The status of surgery in the management of high-risk prostate cancer. *Nature Reviews Urology*, 342–351.

Cooperberg, M. R. (2018). Refined analysis of prostate-specific antigen kinetics to predict prostate cancer active surveillance outcomes. *European urology, 74*(2), 211--217.

Cuocolo, R. a. (2019). Machine learning applications in prostate cancer magnetic resonance imaging. *European radiology experimental, 3*(1), 1-8. doi:https://doi.org/10.1186/s41747-019-0109-2

Dahm, D. I. (2013). Screening for prostate cancer. *Cochrane Database of Systematic Reviews*(1). doi:10.1002/14651858.CD004720.pub3

Descotes, J.-L. (2019). Diagnosis of prostate cancer. *Asian Journal of Urology, 6*(2), 129-136. Retrieved from https://doi.org/10.1016/j.ajur.2018.11.007.

Ebert, J. K. (2021). Radiomics for Identification and Prediction in Metastatic Prostate Cancer: A Review of Studies. *Frontiers in Oncology, 11*, 771787. doi:10.3389/fonc.2021.771787

Elizondo-Riojas, D. P.-I.-F. (2018). Prostate-specific antigen (PSA) as a possible biomarker in non-prostatic cancer: A review. *Cancer epidemiology, 54*, 48-55. doi:10.1016/j.canep.2018.03.009

Fendler, R. S. (2022). Unspecific 18F-PSMA-1007 Bone Uptake Evaluated Through PSMA-11 PET, Bone Scanning, and MRI Triple Validation in Patients with Biochemical Recurrence of Prostate Cancer. *The Journal of Nuclear Medicine, 64*(5), 738 - 743. doi:10.2967/jnumed.118.215434

Feng, L. X. (2016). Prostate cancer prediction using the random forest algorithm that takes into account transrectal ultrasound findings, age, and serum levels of prostate-specific antigen. *Asian Journal of Andrology, 19*, 586 - 590. doi:10.4103/1008-682X.186884

Fitzgerald, W. J. (2019). Early detection and therapeutics. *Molecular Oncology, 13*, 599 - 613. doi:10.1002/1878-0261.12458

Gallagher, J. (2023). Announcing roboflow train 3.0. *Roboflow Blog, 11*. Retrieved from https://blog.roboflow.com/roboflow-train-3-0/

Gaudiano, C. a. (2023). An apparent diffusion coefficient-based machine learning model can improve prostate cancer detection in the grey area of the prostate imaging reporting and data system category 3: a single-centre experience. *Cancers, 15*(13). doi:10.3390/cancers15133438

Gopi Battineni, G. G. (2020). Applications of Machine Learning Predictive Models in the Chronic Disease Diagnosis. *Journal of Personalized Medicine, 10*(2), 2. doi:10.3390/jpm10020021

Halahleh, K. a.-R. (2022). General oncology care in Palestine. In *Cancer in the Arab World* (pp. 195--213). Springer Singapore Singapor.

Harfouche, Y. A. (2023). Predicting Patient Length of Stay Using Artificial Intelligence to Assist Healthcare Professionals in Resource Planning and Scheduling Decisions. *Journal of Global Information Management (JGIM), 31*(1), 1-14. doi:10.4018/jgim.323059

Hassan, F. M.-S. (2024). Improved prostate cancer diagnosis using a modified ResNet50-based deep learning architecture. *BMC Medical Informatics and Decision Making, 24*(1), 23. doi:10.1186/s12911-024-02419-0

Health, N. I. (2024, 07 20). *What is cancer*. Retrieved from National Cancer Institute: https://www.cancer.gov/about-cancer/understanding/what-is-cancer#:~:text=Cancer%20is%20a%20disease%20caused,are%20also%20called%20genetic%20changes.

Hu, M. H. (2023). Research progress on deep learning in magnetic resonance imaging–based diagnosis and treatment of prostate cancer: a review on the current status and perspectives. *Frontiers in Oncology, 13*, 1189370. doi:10.3389/fonc.2023.1189370

Husseini, A. a.-R.-O. (2009). Cardiovascular diseases, diabetes mellitus, and cancer in the occupied Palestinian territory. *The lancet, 373*, 1041--1049.

Inform, N. (2024). *Biopsy*. Retrieved from NHS Inform: https://www.nhsinform.scot/tests-and-treatments/biopsies/biopsy/#:~:text=A%20biopsy%20is%20a%20medical,skin%2C%20organs%20and%20other%20structures.

Jain, S. O. (2019). Computed Tomography-based Radiomics for Risk Stratification in Prostate Cancer. *International Journal of Radiation Oncology*Biology*Physics, 105*(2), 448-456. doi:https://doi.org/10.1016/j.ijrobp.2019.06.2504

Ji, Y. a. (2025). Transmission Line Defect Detection Algorithm Based on Improved YOLOv12. *Electronics, 14*(12), 2432.

Jia, Z. Q. (2019). Diagnosis accuracy of PCA3 level in patients with prostate cancer: a systematic review with meta-analysis. *International Brazilian Journal of Urology : Official Journal of the Brazilian Society of Urology, 46*(5), 691 - 704. doi:10.1590/S1677-5538.IBJU.2019.0360

Jon, D. K. (2010). A drug-loaded aptamer-gold nanoparticle bioconjugate for combined CT imaging and therapy of prostate cancer. *ACS nano, 4* (7), 3689-96. doi:10.1021/nn901877h

Khoa A. Tran, O. K. (2021). Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Med, 13*, 152.

Lan, X. Z. (2022). PSMA-based 18F-DCFPyL PET: a better choice than multiparametric MRI for prostate cancer diagnosis? *American journal of nuclear medicine and molecular imaging, 12*(6), 195-200.

Lawhn-Heath, C. a. (2021). Prostate-specific membrane antigen PET in prostate cancer. *Radiology, 299*(2), 248 -260. Retrieved from https://doi.org/10.1148/radiol.2021202771

Lawrentschuk, J. L. (2024). A Systematic Review on Artificial Intelligence Evaluating Metastatic Prostatic Cancer and Lymph Nodes on PSMA PET Scans. *Cancers, 16*(3). doi:10.3390/cancers16030486

Maisto, C. a. (2022). Analysis of Pros and Cons in Using [68Ga]Ga-PSMA-11 and [18F]PSMA-1007: Production, Costs, and PET/CT Applications in Patients with Prostate Cancer. *Molecules, 27*(12). doi:10.3390/molecules27123862

Malcolm Rozario, A. A. (2021). Artificial Intelligence and Machine learning in the Healthcare Sector: A Review. *Malaysian Journal of Science and Advanced Technology, 1*(3), 89.

Mao, M. a. (2025). YOLO Object Detection for Real-Time Fabric Defect Inspection in the Textile Industry: A Review of YOLOv1 to YOLOv11. *Sensors (Basel, Switzerland), 25*(7), 2270.

Mooney, A. A. (2021). Current Challenges and Future Opportunities for XAI in Machine Learning-Based Clinical Decision Support Systems: A Systematic Review. *Applied Sciences, 11*(11), 5088. doi:10.3390/APP11115088

Nayan, A. A.-A. (2023). Machine learning to predict future PSA in patients with prostate cancer managed with active surveillance. *Journal of Clinical Oncology, 41*(16), e17098-e17098. doi:10.1200/jco.2023.41.16_suppl.e17098

Nelly Gordillo, E. M. (2013). State of the art survey on MRI brain tumor segmentation. *Magnetic Resonance Imaging, 31*(8), 1426–1438.

Neto, P. C. (2020). Deep Learning Based Analysis of Prostate Cancer from MP-MRI. *ArXiv, abs/2106.01835*.

Nishiyama, S. N. (2019). Machine learning methods can more efficiently predict prostate cancer compared with prostate-specific antigen density and prostate-specific antigen velocity. *Prostate International, 7*, 114 - 118. doi:10.1016/j.prnil.2019.01.001

O'Connor, I. F.-W.-F. (2020). Reliability and prognostic value of radiomic features are highly dependent on choice of feature extraction platform. *European Radiology, 30*(11), 6241 - 6250. doi:10.1007/s00330-020-06957-9

Oskar Bergengren, K. R. (2023). 2022 Update on Prostate Cancer Epidemiology and Risk Factors—A Systematic Review. *European Urology*, 191-206.

Papanikolaou, N. R. (2024). Deep Learning Features Can Improve Radiomics-Based Prostate Cancer Aggressiveness Prediction. *JCO clinical cancer informatics, 8,* e2300180 . doi:10.1200/CCI.23.00180

Passera, R. a. (2021). Machine learning techniques in prostate cancer diagnosis according to prostate-specific antigen levels and prostate cancer gene 3 score. *The Korean Journal of Urological Oncology, 19*(3), 164--173.

Peri, P. R. (2025). Roboflow100-VL: A Multi-Domain Object Detection Benchmark for Vision-Language Models. Retrieved from https://arxiv.org/abs/2505.20612

Petersen, H. Z.-O. (2018). Prospective comparison of 68Ga-PSMA PET/CT, 18F-sodium fluoride PET/CT and diffusion weighted-MRI at for the detection of bone metastases in biochemically recurrent prostate cancer. *European Journal of Nuclear Medicine and Molecular Imaging, 45*(11), 1884-1897. doi:10.1007/s00259-018-4058-4

Pierre-Jean Lamy, Y. A.-S.-C.-H. (2017). Prognostic Biomarkers Used for Localised Prostate Cancer Management: A Systematic Review. *European Association of Urology, 4*(6), 2.

Priya, S. Y. (2021). Deep Learning. In *Machine Learning — A Journey to Deep Learning* (pp. 423-488). doi:10.1142/9789811234064_0012

Ramkumar, J. M. (2020). Machine Learning and Artificial Intelligence: Definitions,Applications, and Future Directions. *Current reviews in musculoskeletal medicine, 13,* 69-76. doi:https://doi.org/10.1007/s12178-020-09600-8

Rosenkrantz, A. K. (2015). Implementation of Multi-parametric Prostate MRI in Clinical Practice. *Current Urology Reports, 16*, 1-8. doi:10.1007/s11934-015-0530-x

Rowe, R. W. (2020). 18F-Labeled, PSMA-Targeted Radiotracers: Leveraging the Advantages of Radiofluorination for Prostate Cancer Molecular Imaging. *Theranostics, 10*(1), 1 - 16. doi:10.7150/thno.37894

Sánchez, G. L. (2017). A survey on deep learning in medical image analysis. *Medical image analysis, 42*, 60--88. doi:10.1016/j.media.2017.07.005

Sapkota, R. a. (2025). RF-DETR Object Detection vs YOLOv12: A Study of Transformer-based and CNN-based Architectures for Single-Class and Multi-Class Greenfruit Detection in Complex Orchard Environments Under Label Ambiguity. *arXiv preprint arXiv:2504.13099*.

Sarah Lindgren Belal, S. F. (2024). Applications of Artificial Intelligence in PSMA PET/CT for Prostate Cancer Imaging. *Seminars in nuclear medicine, 54*(1), 141-149. doi:10.1053/j.semnuclmed.2023.06.001

Schalken, R. H. (2016). Blood-based and urinary prostate cancer biomarkers: a review and comparison of novel biomarkers for detection and treatment decisions. *Prostate Cancer and Prostatic Diseases, 20*(1), 12-19. doi:10.1038/pcan.2016.59

Schatten, H. (2018). Brief Overview of Prostate Cancer Statistics, Grading, Diagnosis and Treatment Strategies. In H. Schatten, *Advances in Experimental Medicine and Biology* (pp. 1-14).

Schisterman, P. A. (2012). Novel statistical methodology for analyzing longitudinal biomarker data. *Statistics in Medicine, 31*, 2457–2460. doi:10.1002/sim.5500

Schuster, D. M. (2016). PET tracers beyond FDG in prostate cancer. *Seminars in nuclear medicine, 46*(6), 507-521. doi:https://doi.org/10.1053/j.semnuclmed.2016.07.005

Sharaf, M. (2006). Epidemiology of prostate cancer in North-West-bank, Palestine 1998-2006.

Sidey-Gibbons, J. A.-G. (2019). Machine learning in medicine: a practical introduction. *BMC Medical Research Methodology, 19*, 1-18. doi:10.1186/s12874-019-0681-4

Singh, S. K. (2024). A novel deep learning-based technique for detecting prostate cancer in MRI images. *Multimedia Tools and Applications, 83*(5), 14173–14187. doi:10.1007/s11042-023-15793-0

Singh, U. C. (2025). The impact of artificial intelligence on remote healthcare: Enhancing patient engagement, connectivity, and overcoming challenges. *Intelligent Pharmacy, 3*(5), 323-329. doi:https://doi.org/10.1016/j.ipha.2024.12.003

Smith, M. P. (2020). PSA-based machine learning model improves prostate cancer risk stratification in a screening population. *World Journal of Urology, 39*, 1897--1902. doi:10.1007/s00345-020-03392-9

Statistics, P. C. (n.d.). Retrieved from Palestinian Central Bureau of Statistics: https://www.pcbs.gov.ps/default.aspx

Sutherland, D. E. (2023). Role of FDG PET/CT in Management of Patients with Prostate Cancer. *Seminars in nuclear medicine, 54*, 4-13. doi:https://doi.org/10.1053/j.semnuclmed.2023.06.005

Tewari, V. W. (2023). 'Stealth' Prostate Tumors. *Cancers, 15*(13), 3487. doi:10.3390/cancers15133487

Thomsen, E. D. (2018). 68Ga-PSMA-PET/CT in comparison with 18F-fluoride-PET/CT and whole-body MRI for the detection of bone metastases in patients with prostate cancer: a prospective diagnostic accuracy study. *European Radiology, 29*(3), 1221-1230. doi:10.1007/s00330-018-5682-x

Ting, Y. Y. (2021). The promise of artificial intelligence: a review of the opportunities and challenges of artificial intelligence in healthcare. *British medical bulletin, 139*(1), 4-15. doi:10.1093/bmb/ldab016

Trägårdh, S. L. (2023). Applications of Artificial Intelligence in PSMA PET/CT for Prostate Cancer Imaging. *Seminars in nuclear medicine, 54*(1), 141-149. doi:10.1053/j.semnuclmed.2023.06.001

Tsui, E. N.-L. (2023). Designing interpretable ML system to enhance trust in healthcare: A systematic review to proposed responsible clinician-AI-collaboration framework. *Inf. Fusion, 108*, 102412. doi:10.1016/j.inffus.2024.102412

UK, C. R. (2024, 01 15). *Tests for prostate cancer*. Retrieved from https://www.cancerresearchuk.org/about-cancer/prostate-cancer/getting-diagnosed/tests-for-prostate-cancer: https://www.cancerresearchuk.org/about-cancer/prostate-cancer/getting-diagnosed/tests-for-prostate-cancer

Unterrainer, M. a.-H.-S. (2020). Recent advances of PET imaging in clinical radiation oncology. *Radiation Oncology, 15*(1), 1--15. doi:10.1186/s13014-020-01519-1

Vjaters, L. S. (2022). Recurrent Prostate Cancer Diagnostics with 18F-PSMA-1007 PET/CT: A Systematic Review of the Current State. *Diagnostics*, *12*(12), 3176. doi:10.3390/diagnostics12123176

WHO. (2024, 07 21). *World Health Organization*. Retrieved from World Cancer Day 2024.

Yu, Y. H. (2023). Application of spectral CT in diagnosis, classification and prognostic monitoring of gastrointestinal cancers: progress, limitations and prospects. *Frontiers in Molecular Biosciences, 10*, 1284549. doi:10.3389/fmolb.2023.1284549

Zhang, Z. Y. (2021). Machine learning-based prediction of invisible intraprostatic prostate cancer lesions on 68 Ga-PSMA-11 PET/CT in patients with primary prostate cancer. *European Journal of Nuclear Medicine and Molecular Imaging, 49*(5), 1523 - 1534. doi:{10.1007/s00259-021-05631-6}

الملخص

يُعدّ سرطان البروستاتا أحد أكثر أنواع السرطان شيوعًا بين الرجال على مستوى العالم، ويُعدّ الكشف الدقيق عن الآفات في الصور الطبية أمرًا بالغ الأهمية للتشخيص المبكر والعلاج الفعّال. وتوفّر تقنيات التصوير الحديثة مثل التصوير المقطعي بالإصدار البوزيتروني/التصوير المقطعي المحوسب (PET/CT) معلومات قيّمة للغاية لأغراض التشخيص. ومع ذلك، فإن عملية تحديد الآفات في هذه الصور غالبًا ما تكون مستهلكة للوقت وتعتمد بدرجة كبيرة على الخبرة البشرية، مما قد يؤثر على دقة النتائج. تهدف هذه الرسالة إلى دراسة إمكانية استخدام نماذج كشف الأجسام المعتمدة على التعلم العميق للكشف التلقائي عن آفات البروستاتا في الصور الطبية.

في هذه الدراسة، تم تدريب ومقارنة عدة نماذج متقدمة في مجال كشف الأجسام، من بينها نماذج تعتمد على خوارزميات YOLO ونموذجRoboflow OpenYOLO ، وذلك باستخدام نفس مجموعة البيانات وإجراءات التقييم. وقد تم تدريب هذه النماذج وتقييمها باستخدام مجموعة بيانات محلية جُمعت من مستشفى أوغستا فيكتوريا (Augusta Victoria Hospital – AVH) في القدس، والتي تتكون من صور PET/CTلعدد 200 مريض مصاب بسرطان البروستاتا، بإجمالي يقارب 2735 صورة تركز على منطقة البروستاتا. وبحسب علم الباحث، لم تُستخدم صور PET/CT في دراسات محلية سابقة، مما يجعل هذه المجموعة من البيانات موردًا جديدًا ومهمًا للأبحاث. وقد تم إجراء عملية توصيف الصور (Annotation) بمساعدة أطباء مختصين في الأشعة. وتُظهر هذه البيانات إمكانية تطبيق نماذج الكشف المعتمدة على بنية المحولات (Transformer-based models) على بيانات سريرية واقعية. ونظرًا لدقته العالية ومتانته، تم اختيار نموذج RF-DETR كنموذج نهائي لإجراء الاختبارات المتقدمة. وقد تم تهيئة هذا النموذج باستخدام أوزان مدربة مسبقًا على مجموعة بياناتCOCO ، كما تم استخدام المقاييس القياسية الخاصة بـ COCO مثل الدقة(Precision) ، والاستدعاء(Recall) ، ومتوسط الدقة (mAP) في عملية التقييم.

أظهرت النتائج التجريبية أن نموذج RF-DETR بعد ضبطه (Fine-tuning) حقق دقة (Precision) بلغت 93.9%، واستدعاء (Recall) بنسبة 92.6%، ومتوسط دقة mAP@50 بنسبة 93.4%. وتشير هذه النتائج إلى قدرة عالية على اكتشاف الآفات وتحديد مواقعها بدقة في ظروف تداخل متوسطة. ومع ذلك، فقد أظهرت النتائج انخفاضًا نسبيًا في الأداء عند عتبات أعلى لمعامل التقاطع على الاتحاد(IoU) ، وهو ما

انعكس في انخفاض قيمmAP@50–95 ، مما يدل على التحديات الكامنة في تحديد الحدود الدقيقة للآفات. كما دعمت التحليلات النوعية هذه النتائج من خلال إظهار قدرة النموذج على تحديد مواقع الآفات بدقة في مجموعات الاختبار الداخلية والخارجية، والتعامل بكفاءة مع الحالات التي لا تحتوي على آفات، بالإضافة إلى الكشف الناجح عن الآفات الصغيرة في ظروف تصوير معقدة. وقد ساهم التحليل التفصيلي للحالات الصحيحة (True Positive)والخاطئة (False Positive) والحالات التي لم يتم اكتشافها (False Negative) في تقديم فهم أعمق لنتائج التجربة.

وفي الختام، تُظهر هذه الدراسة أن النماذج المعتمدة على بنية المحولات، وبخاصة نموذجRF-DETR ، تُعد أداة فعّالة للكشف التلقائي عن آفات البروستاتا في الصور الطبية. وعلى الرغم من استمرار وجود بعض التحديات المتعلقة بدقة تحديد المواقع عند عتبات IoU المرتفعة، فإن هذه الدراسة تبرز إمكانات كبيرة للتطوير المستقبلي، وتوفّر أساسًا علميًا جيدًا لمزيد من الأبحاث التي تهدف إلى تحسين دقة تحديد المواقع ودمج البيانات متعددة الأنماط.