UNIVERSITY OF CALIFORNIA

Los Angeles

Learning-Based Diagnostic Models Incorporating Imaging and Clinical Information

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Computer Science

by

Haoxin Zheng

2024

ABSTRACT OF THE DISSERTATION

Learning-Based Diagnostic Models Incorporating Imaging and Clinical Information

by

Haoxin Zheng
Doctor of Philosophy in Computer Science
University of California, Los Angeles, 2024
Professor Kyung Hyun Sung, Co-Chair
Professor Demetri Terzopoulos, Co-Chair

In recent years, learning-based computer-aid diagnostic (CAD) models have attracted significant attentions, and demonstrated their potential to make real-world impact. The prosperity of the models is currently being fueled by leveraging the rapid advancements in artificial intelligence (AI). They were shown to be capable of resolving various clinical tasks, like stratifying cancerous patients, predicting the prognosis results, and automatically allocating lesion locations. However, most models primarily rely on a single modality of information, such as imaging data or clinical data. In comparison, the real-world clinical specialists generally conclude and make diagnoses by incorporating imaging observations with anatomical knowledge and other clinical information, such as laboratory test results and demographic data. This suggests that despite the advancements in current learning-based CAD models, we posit that existing models can be further improved by integrating imaging data with broader clinical insights.

In this dissertation, we explore the potentials of incorporating imaging information together with clinical information into the learning-based CAD models, with focus on applications associated with magnetic resonance imaging (MRI). We begin by demonstrating how the integration of radiological imaging features with clinical information helps enhance the performance of traditional machine-learning-based CAD models, and showing its generalizability onto different diseases. We then delve into deep-learning-based

CAD models, and how it can be benefited by integrating both information on patient stratification. Finally, we discuss about the incorporation of domain-specific clinical knowledge into imaging-based deep learning network designs, using anatomical priors to reduce false positive detections and boost model efficacy on lesion allocation. Through extensive experimentation across diverse patient populations, our research not only validates the efficacy of integrating both imaging and clinical information in learning-based CAD model designs, but also highlights their potentials to facilitate the deployment of improved learning-based CAD models in real-world clinical settings in the foreseeable future.

The dissertation of Haoxin Zheng is approved.

Fabien Scalzo

Achuta Kadambi

Kai-Wei Chang

Demetri Terzopoulos, Committee Co-Chair

Kyung Hyun Sung, Committee Co-Chair

University of California, Los Angeles

2024

To my parents Shuying and Bowei,

to my grandparents Defang and Deheng,

and to my beloved Rujun,

for their love, understanding, support, sacrifices, and companionship.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First and foremost, I extend my deepest appreciation to my PhD advisors, Professors Kyung Hyun Sung and Fabien Scalzo. Their unwavering support and guidance were the cornerstones of my PhD journey, from its inception to its completion. Professor Scalzo opened the gateway to the fascinating world of medical imaging analysis, and Professor Sung, with his extraordinary mentorship, not only provided me with an unparalleled opportunity to explore and contribute to the field, but also guided and supported me throughout this journey and made my PhD experience profoundly enjoyable and rewarding. I am also profoundly grateful to my doctoral committee members: Professors Demetri Terzopoulos, Achuta Kadambi, and Kai-Wei Chang. Their remarkable expertise and insightful guidance have been pivotal in shaping my research endeavors and propelling me towards academic excellence. The journey towards this milestone would not have been attainable without their collective wisdom, steadfast encouragement, and unwavering support. For their invaluable contributions to my personal and professional development, I am forever indebted and thankful.

Besides my advisors and committee members, I wish to extend my gratitude towards all my collaborators, who have enriched my PhD journey in countless ways. My heartfelt respect and gratitude go to Dr. Qi Miao. As an exemplary collaborator, teacher, and highly skilled radiologist, Dr. Miao has played a crucial role in guiding me through the intricacies of my research projects. Her contributions have been significant in my development as a researcher, providing me with invaluable insights and support that have smoothed the path of my academic journey. Without her, the completion of most of my researches would not have been as smooth or successful. Moreover, I would like also thank Dr. Steven Raman for his support from the Radiological Sciences department. I am deeply thankful for the privilege of working alongside a group of dedicated professionals: Kai Zhao, Alex Ling Yu Hung, Kaifeng Pang, Ran Yan, Yongkai Liu, and Ruiming Cao. The wealth of knowledge and experience they brought to our collaboration has been invaluable to the success of my research projects. Their commitment and expertise have not only

enriched my work but have also fostered an environment of growth and learning.

Furthermore, I extend my heartfelt thanks to all the friends who have supported and assisted me throughout my enriching PhD journey in the United States. Special recognition is reserved for my dear friends Yaxuan Zhu and Jingchi Ma, who have been more than close friends; they have been integral to my journey. With patience and wisdom, they guided me on learn new things, navigated the complexities of PhD life, and significantly assisted my research and coursework. Additionally, I am profoundly grateful to Yihao Yang and Karen Kan, who warmly welcomed me upon my arrival in the United States, whose kindness and assistance were pivotal in easing my transition to a new country. Furthermore, my gratitude also extends to Mengshi Zhang, Xiaolin Li, and Yang Guo not only for their invaluable guidance that spanned the critical transition from my undergraduate to PhD life, but also for their sage advice regarding my career path, which have been tremendously beneficial. And of course, the unforgettable moments shared with my friends at UCLA will forever hold a special place in my heart. To Weinan Song, Xuan Hu, Fangyao Liu, Yu Wu, Yifei Xu, Yifeng Lan, Flora Fang, Jiangtao Chen, Hanzhi Xia, Linfang Wang, Wenyan Zhao, Xitong Zhou, Qianhui Gao, Yiming Zhou, Xiaofeng Gao, Ruiqi Gao, Yutan Gu, Qi Qu, Siting Liu, Qi Wu, Yu Zhao, Kiki Liu, Yuki Zhang, Lia Zhuo, Lin Du, Danlin Xie, Zihao Chen, Boyang Fu, Zixiang Chen, Meixi Lin, Ruoyu Peng, Zhonglin Zhang, Yanzhe Xu, Jinxi Zou, Tianyi Liu, Qi Zeng, Xudong Li, Tingyi Zhou, Jianhong He, Zhilai Shen, Guanqun Yang, Junheng Hao, Tianxiang Li, Joy Xu, Xinyu Ma, Mingfei Dong, Sai Ouyang, Fei Feng, Tengyu Liu, Xixie Zhou, Lingyu Zhan, Pei Han, Cheng Fang, Aoxuan Li, Yutong Lu, Shuangyu Li, Hengjie Yang, Danping Zhai, Hengjie Liu, Zhengyang Ming, Qing Dai, Shu-Fu Shih, Tianle Cao, Xinran Zhong, Jiahao Lin, Xinzhou Lin, Zhaohuan Zhang, Chang Gao, Le Zhang, Yannan Yu, Danfeng Guo, Yayun Du, Zhufeng Pan, Chong Zhang, Pan Lu, Jiyuan Wang, Yasi Zhang, and all who have been a part of my journey, I am deeply grateful. The camaraderie, laughter, and the spectrum of shared experiences — from engaging in badminton clubs and board games, to embarking on hiking adventures, fishing excursions, video gaming, and travels — have immeasurably enriched my life, rendering my time at UCLA genuinely unforgettable.

Each moment, whether filled with competition, exploration, or simple joy, has contributed to a tapestry of memories that I will cherish forever.

In closing, I would like to convey my deepest appreciation to my parents, Shuying Liu and Bowei Zheng, for their sacrifice and selfless support, and to my darling wife, Rujun Cai, for her companionship, heartwarming support, and trust. I also would like to say thank you to my grandparents, Defang Zhang and Deheng Zheng, for their love and kindness that warm my entire life and will never be forgotten.

VITA

2013–2017 B.S. in Electrical Engineering, Nankai University, China

2017–2019 M.S. in Electrical and Computer Engineering, University of California, Los Angeles, US

2019–2024 Graduate Student Researcher & Teaching Fellow in Computer Science, University of California, Los Angeles, US

PUBLICATIONS

Haoxin Zheng, Alex Ling-Yu Hung, Qi Miao, Weinan Song, Fabien Scalzo, Steven S. Raman, Kai Zhao, Kyunghyun Sung. AtPCa-Net: Anatomical-aware prostate cancer detection network on multi-parametric MRI. *Scientific Reports*, **14**(1), March 2024, 5740

Haoxin Zheng, Qi Miao, Alex Ling-Yu Hung, Kai Zhao, Fabien Scalzo, Steven S. Raman, Kyunghyun Sung. Integrative feature-enhanced network predicting prostate biopsy results from negative multi-parametric MRI. *Radiology: Imaging Cancer*. 2024. (Under Review)

Haoxin Zheng, Qi Miao, Alex Ling-Yu Hung, Kai Zhao, Steven S. Raman, Kyunghyun Sung. Anatomical-aware Siamese deep network for prostate cancer detection on multi-parametric MRI. *Annual Meeting of The International Society for Magnetic Resonance in Medicine (ISMRM)*, Toronto, Canada, June 2023.

Haoxin Zheng, Qi Miao, Yongkai Liu, Sohrab Afshari Mirak, Melina Hosseiny, Fabien Scalzo, Steven S. Raman, Kyunghyun Sung. Multiparametric MRI-based radiomics model to predict pelvic lymph node invasion for patients with prostate cancer. *European Radiology*, **32**(8), March 2022, 5688–5699

Haoxin Zheng, Qi Miao, Yongkai Liu, Steven S. Raman, Fabien Scalzo, Kyunghyun Sung. Integrative machine learning prediction of prostate biopsy results from negative multiparametric MRI. *Journal of Magnetic Resonance Imaging*, **55**(1), January 2022.

Haoxin Zheng, Qi Miao, Steven S. Raman, Fabien Scalzo, Kyunghyun Sung. Integrative radiomics models to predict biopsy results for negative prostate MRI. *IEEE International Symposium on Biomedical Imaging (ISBI)*, Online, April 2021, 877–881.

Alex Ling Yu Hung, **Haoxin Zheng**, Kai Zhao, Xiaoxi Du, Kaifeng Pang, Qi Miao, Steven S. Raman, Demetri Terzopoulos, Kyunghyun Sung. CSAM: A 2.5D cross-slice attention module for anisotropic volumetric medical image segmentation. *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, Waikoloa, Hawaii, January 2024, 5923–5932.

Weinan Song, **Haoxin Zheng**, Jiawei Yang, Chengwen Liang, Lei He. Oral-3Dv2: 3D Oral Reconstruction from Panoramic X-Ray Imaging with Implicit Neural Representation. *arXiv:2303.12123*, 2023.

Alex Ling Yu Hung, **Haoxin Zheng**, Qi Miao, Steven S. Raman, Demetri Terzopoulos, Kyunghyun Sung. CAT-Net: A cross-slice attention transformer model for prostate zonal segmentation in MRI. *IEEE Transactions on Medical Imaging (TMI)*, **42**(1), January 2023, 291–303.

Yongkai Liu, **Haoxin Zheng**, Zhengrong Liang, Qi Miao, Wayne G. Brisbane, Leonard S. Marks, Steven S. Raman, Robert E. Reiter, Guang Yang, Kyunghyun Sung. Textured-based deep learning in prostate cancer classification with 3T multiparametric MRI: comparison with PI-RADS-based classification. *Diagnostics*, **11**10, September 2021, 1785.

Alex Ling Yu Hung, Kai Zhao, **Haoxin Zheng**, Ran Yan, Steven S. Raman, Demetri Terzopoulos, Kyunghyun Sung. Med-cDiff: Conditional medical image generation with diffusion models. *Bioengineering*, **10**(10), October 2023, 1258:1–13.

Qi Miao, Xuechun Wang, Jingjing Cui, **Haoxin Zheng**, Yan Xie, Kexin Zhu, Ruimei Chai, Yuanxi Jiang, Dongli Feng, Xin Zhang, Feng Shi, Xiaodong Tan, Guoguang Fan, Keke Liang. Artificial intelligence to predict T4 stage of pancreatic ductal adenocarcinoma using CT imaging. *Computers in Biology and Medicine (CBM)*, March 2024, 108125.

Kai Zhao, Alex Ling Yu Hung, Kaifeng Pang, **Haoxin Zheng**, Kyunghyun Sung. PartDiff: Image super-resolution with partial diffusion models. *arXiv:2307.11926*, 2023.

Yongkai Liu, Qi Miao, Chuthaporn Surawech, **Haoxin Zheng**, Dan Nguyen, Guang Yang, Steven S. Raman, Kyunghyun Sung. Deep learning enables prostate MRI segmentation: A large cohort evaluation with inter-rater variability analysis. *Frontiers in Oncology*, **11**, December 2021.

CHAPTER 1

Introduction

1.1 Medical Imaging

Medical imaging is crucial in modern clinical practice, enabling the non-invasive acquisition of information about internal tissues and their functionalities. It encompasses various modalities, including magnetic resonance (MR) imaging, computed tomography (CT) imaging, ultrasound imaging, X-ray imaging, and positron emission tomography (PET) imaging, among others. In this dissertation, our primary focus is on MR imaging (MRI), given its widespread utilization in clinical practice. MRI is renowned for its ability to generate high-resolution three-dimensional anatomical images. Moreover, MRI, with its diverse sequence designs, can capture detailed characteristics of organ tissue from various perspectives, facilitating the multi-view interpretation of potential anatomical changes within the body. Furthermore, MRI's exceptional soft tissue contrast and absence of ionizing radiation make it a preferred imaging modality that minimizes the risk of radiation-induced harm. As a result, it now plays a crucial role in disease diagnosis and management. For instance, in the clinical diagnosis of prostate cancer, multi-parametric MRI (mpMRI) is the preferred imaging modality due to its ability to provide detailed anatomical information within the prostate gland, thus guiding diagnosis and treatment planning.

1.2 Clinical Information

While medical imaging can reflect detailed in-vivo anatomical changes, clinical information provides complementary insights into the patient's overall health status. Clinical information comprises routinely-collected data such as laboratory test results, demographic data, family histories, and details of previous treatments, among others. This wealth of information offers valuable insights into the patient's medical history, physiological condition, and potential risk factors. Furthermore, systematic anatomical prior knowledge about specific diseases and organs also forms an integral part of clinical information. Such knowledge, often derived from the medical literature, expert consensus, and clinical guidelines, significantly aids in diagnostic decision-making by providing context and guiding the interpretation of medical findings. In clinical practice, nurses may conduct blood pressure tests and evaluate weight and height before a patient's visit to clinical specialists. This preliminary assessment provides valuable reference points that assist the specialist in making informed decisions regarding the patient's care and treatment plan.

1.3 Computer-Aided Diagnostic Models

With the development and emergence of medical imaging, doctors are able to identify potential diseases in patients without resorting to invasive procedures. Typically, radiologists review images following clinical guidelines in order to provide diagnostic interpretations that can guide future treatment plans. Although medical imaging has proven to be greatly beneficial in clinical practice, the accuracy of diagnoses relies heavily on the expertise of radiologists, which necessitates specialized medical training and therefore leads to high costs. Furthermore, the repetitive nature of identifying abnormalities in medical images can be monotonous and exhausting for radiologists, potentially resulting in fatigue that compromises diagnostic accuracy and negatively affects treatment planning. The advent of artificial intelligence (AI), has led to the rise of learning-based computer-aided diagnostic (CAD) models, which can potentially mitigate these challenges. Learning-based CAD

models, composed of traditional machine learning models or deep learning networks, are increasingly popular not only for their ability to support radiologists in decision-making processes, potentially reducing costs, but also for their ability to maintain consistent decision-making quality regardless of workload volume.

Moreover, learning-based CAD models play a crucial role in aiding decision-making processes involving clinical information. Through the utilization of Nomograms, a set of disease-specific CAD models, doctors can comprehensively evaluate patients' overall health based on required clinical data. Nomograms analyze various clinical parameters and provide predictive assessments, empowering healthcare professionals to make well-informed decisions regarding patient care and treatment strategies. Overall, learning-based CAD models have demonstrated their potential in resolving various clinical tasks, such as stratifying cancer patients, predicting prognosis results, and automatically allocating lesion locations.

Despite the effectiveness of learning-based CAD models, many primarily rely either on radiological information extracted from medical images, or, on routinely-collected clinical data, thereby overlooking the potential significance of integrating both information for decision-making processes. In contrast, real-world clinical specialists rarely rely solely on a single modality for diagnosis. Instead, they incorporate observations from medical imaging along with anatomical knowledge and other clinical information. This comprehensive approach is essential, as disease-related factors such as laboratory test results, demographic data, and medical histories are critical for accurate diagnosis, but they may not be directly reflected in imaged features. Moreover, such comprehensive decision-making is crucial for procedures like biopsies or surgeries, which carry inherent risks and potential complications. For instance, the decision to proceed with prostate biopsies not only depends on the risk level of abnormalities observed on mpMRI, but also requires consideration of clinical factors such as prostate-specific antigen (PSA) level, PSA density (PSAD) level, age, and family history. Furthermore, the diagnosis of certain diseases may necessitate additional anatomical prior knowledge to help distinguish between cancerous and benign tissues, particularly when visual abnormalities in medical images

appear similar. These considerations underscore the potential for further improvements in current learning-based CAD models through the integration of imaging with broader clinical insights.

1.4 Research Objective

In this dissertation, we explore the potential of incorporating clinical information and imaging data in the design of learning-based CAD models for improving model performance, with a focus on MRI. The primary objective of this dissertation is twofold: firstly, to investigate how the integration of multi-modal information in learning-based CAD models can enhance model capacity and, secondly, to explore strategies for effectively incorporating anatomical priors into model designs to boost performance.

We commence by demonstrating how the integration of routinely-collected clinical information with radiological imaging features enhances the performance of traditional machine-learning-based CAD models. Subsequently, we broaden the scope and demonstrate the applicability of this multi-modal incorporation across different diseases and patient cohorts to demonstrate its effectiveness and generalizability. Furthermore, we delve into deep-learning-based CAD models and examine how the integration of radiological and clinical information can enhance deep learning model performance in diverse patient populations. Finally, we discuss the value of including domain-specific clinical prior knowledge into deep learning network designs, leveraging anatomical priors to reduce false positive detections and enhance model efficacy in lesion allocation.

1.5 Dissertation Outline and Contributions

The remainder of this dissertation is organized as follows:

- **Chapter 2** presents an integrative machine-learning-based radiomics model designed for predicting biopsy results for patients who have negative prostate mpMRI. The model is based on a quadratic support vector machine (SVM) model, to which

are inputted an integration of radiomics features, extracted representing shape, histogram, and low-level texture features of the patients' mpMRI images, and routinely-collected clinical information. Through experimentation, we demonstrated the feasibility of enhancing machine-learning-based CAD model performance by combining radiological information together with routinely-collected clinical information, compared with approaches using individual group of features in isolation. Moreover, the proposed integrative model exhibits superior performance compared with existing studies that rely solely on clinical information, which highlights its potential utility in addressing clinical challenges in practice. The work presented in this chapter has been published as (Zheng et al., 2022b).

- **Chapter 3** introduces a machine-learning-based CAD model for the prediction of pelvic lymph node invasion (LNI) for patients who have undergone radical prostatectomy (RP), utilizing radiomics features extracted from mpMRI, in conjunction with routinely-collected clinical information. In contrast to the study conducted in Chapter 2, which focused on predicting biopsy results for patients with negative prostate mpMRI, this research explores a different disease and patient cohort while adopting similar concepts and machine learning models. The experiment further illustrates the scalability and generalizability of integrating clinical information with radiological findings to enhance machine-learning-based CAD model capacity. Specifically, the model demonstrates improved performance when utilizing integrated features compared to using individual features alone. Furthermore, our proposed integrative model surpasses the predictive accuracy and reliability of existing Nomograms currently utilized in clinical practice, highlighting its potential to enhance diagnostic accuracy and patient care outcomes. The work presented in this chapter has been published as (Zheng et al., 2022a).

- **Chapter 4** explores the development of an integrative deep learning network for predicting prostate biopsy results for patients with negative prostate multi-parametric MRI, while incorporating routinely-collected clinical information. Building upon

the findings of Chapter 2, which showcased the enhanced prediction performance achieved through the integration of imaging information and clinical data, this chapter takes a further step. Here, we aim to devise a more flexible and adaptable feature extraction approach, diverging from the use of pre-defined radiomics features. Additionally, we seek to design a deep learning model specifically tailored to detecting nearly imperceptible lesion signals, thus addressing the prediction task more effectively. Our proposed integrative feature-enhanced network is presented and demonstrated to outperform methods proposed by existing studies, underscoring its efficacy and potential in improving diagnostic accuracy in clinical practice. The work presented in this chapter has been submitted for publication (Zheng et al., 2024b).

- **Chapter 5** delves into the development of an anatomical-aware deep-learning-based prostate cancer detection network based on mpMRI, attempting to incorporate the anatomical prior knowledge into the deep learning network architecture design. In this chapter, we recognize the value of anatomical knowledge in aiding radiologists' decision-making in clinical practice, and thus seek to determine if such knowledge can enhance the efficacy of deep learning models as well. We begin by incorporating symmetric-related clinical priors into the network design. This approach tries to leverage symmetric-related visual appearance differences to distinguish cancerous tissue from other prostate tissues with similar visual patterns on mpMRI, thus mitigating potential false positive predictions. Additionally, we explicitly integrate zonal appearance differences of prostate cancer on mpMRI into the loss design. By doing so, we can better guide the model in identifying lesions compared to existing models that overlook this property and treat all lesions identically regardless of their location. Extensive experimental comparisons demonstrate that each of the proposed anatomy-aware designs contributes to the improvement of the prostate cancer detection and patient-level classification performance, and the integration of both designs can achieve the best model performance on both tasks. The work presented in this chapter has been published as (Zheng et al., 2024a).

- **Chapter 6** concludes the dissertation, and discusses the challenges in current CAD models and the promising future work of integrating information of different modalities to arrive at better CAD model designs.

CHAPTER 2

Integrative Radiomics-Based Machine Learning Model Predicting Prostate Biopsy Results

2.1 Introduction

Multi-parametric MR Imaging (mpMRI) is now the preferred imaging technique for noninvasive diagnosis of prostate cancer (PCa). mpMRI is increasingly performed prior to prostate biopsy to maximize yield of clinically significant prostate cancer (csPCa) and minimize error Brizmohun Appayya et al. (2018). In the standardized 5 point Likert score based Prostate Imaging Reporting and Data System, version 2.1 (PI-RADS v2.1), intermediate and high suspicion MRI based lesions (PI-RADS≥3) typically undergo MRI-targeted biopsy with or without systematic biopsies with positive tissue diagnosis between 29.7-82.4% (Schoots et al., 2015; Sathianathen et al., 2018; Turkbey et al., 2019). However, when mpMRI findings are of low suspicion, mpMRI negative (PI-RADS 1 or 2), there is a lack of consensus of whether to proceed with a systematic biopsy which contributes to patient morbidity including pain, bleeding, urinary obstruction and erection dysfunction (Liss et al., 2017). Several strategies have been proposed in patients with negative mpMRI to predict low risk of csPCa including use of serum biomarkers such as prostate-specific antigen density (PSAD) levels less than either 0.10 ng/ml/ml or 0.15 ng/ml/ml (Hansen et al., 2018; Oishi et al., 2019; Distler et al., 2017; Otti et al., 2019; Pagniez et al., 2020). However, the current PSAD-based risk assessments are limited by negative predictive value (NPV) of 83.1% to 93.4% (Hansen et al., 2018; Oishi et al., 2019; Distler et al., 2017; Otti et al., 2019; Pagniez et al., 2020).

Radiomics is an emerging field in quantitative imaging that aims to associate radiomic

features with specific clinical endpoints (Gillies et al., 2016; Zwanenburg et al., 2020; Lambin et al., 2017). The radiomics features extracted from medical images can provide large-scale imaging information, and many studies have shown promising results on the PCa detection and aggressiveness assessment using radiomics features (Yan et al., 2019; Vignati et al., 2015; Gnep et al., 2017; Nketiah et al., 2017; Algohary et al., 2018; Starmans et al., 2019; Varghese et al., 2020; Zhang et al., 2020). The aim of the study is to construct and validate an integrative radiomics-based computer-aided diagnostic (CAD) model for predicting biopsy results in patients with negative MRI. Specifically, an integrative machine learning (iML) CAD model was proposed combining visually negative (PI-RADS 1 or 2) MRI-based radiomics features with routinely-collected clinical information to predict the prostate biopsy results. The efficacy of using the integrative multi-scale features was validated in comparisons with other machine learning approaches using either only clinical information or only radiomics features. In addition, the NPV and overall performance of the proposed iML approach was compared with pre-existing PSAD-based strategies to predict risks of csPCa in patients with negative mpMRI (Hansen et al., 2018; Oishi et al., 2019; Distler et al., 2017; Otti et al., 2019).

2.2 Materials and Methods

2.2.1 Study Population and MRI Data

The single arm observational study was performed in compliance with the United States Health Insurance Portability and Accountability Act (HIPAA) of 1996 and was approved by the institutional review board (IRB) with a waiver of the requirement for informed consent. The initial study cohort included all identified negative prostate 3 Tesla mpMRI cases by reviewing all clinical prostate MRI scans performed by a standard protocol via one of several 3 Tesla scanners: Siemens Magnetom Trio, Skyra, and Verio scanner (Siemens Medical Systems, Malvern, Pennsylvania, USA) from January 2016 to December 2018 at a single academic institution. All prostate mpMRI scans were acquired using a standardized imaging protocol following European Society of Urogenital Radiology

(ESUR) PI-RADS guidelines (Weinreb et al., 2016). Three genitourinary radiologists interpreted the mpMRI scans, as part of the clinical diagnostic procedure, where each had read 1,000-3000 prostate mpMRI scans with 10+ years of experience. The study cohort patients met the following inclusion criteria: 1) clinical suspicion of PCa, (elevated PSA level with respect to the current age and/or abnormal digital rectum exam results); 2) 3T-mpMRI with all lesions scored as PIRADS 1 or 2 (MR negative); 3) standardized 12-14 core systematic transrectal ultrasound-guided (TRUS) biopsy with or without magnetic resonance ultrasound fusion (MRUS) within six months after 3T-mpMRI study (Sonn et al., 2013); 4) serum prostate specific antigen (PSA) measured within six months prior to biopsy. All eligible cases were re-reviewed by another independent abdominal radiologist (X. X., 8 years of experience in clinical prostate MRI interpretation), and no discordant was observed. MRUS was used for a partial cohort to record and track biopsy site locations, and there was no difference between TRUS biopsy with and without MRUS. Patients with a known diagnosis of PCa, undergoing active surveillance, or PCa treatment (including surgery, focal therapy, radiation, or hormonal therapy), were excluded.

For patients with multiple mpMRI scans, we selected the mpMRI scan immediately preceding the first negative TRUS/MRUS biopsy. The detailed patient inclusion workflow is shown in Figure 2.1 .

In all, 330 men, median age 63 years (IQR: 58-67), with either systematic TRUS (n=87) or MRUS (n=243) biopsy were included in the final study cohort, for the primary analysis. A secondary analysis on a biopsy naïve cohort (n=227) was conducted to further evaluate the performance in a less cancer enriched population (Hansen et al., 2018; Oishi et al., 2019; Distler et al., 2017; Otti et al., 2019). Negative biopsy was defined as excluding csPCa (lack of primary or secondary Gleason Score (GS)$\geq$7) findings in each biopsy session (Matoso and Epstein, 2019). The following clinical information was evaluated: patient age, family history of PCa, prostate biopsy history, prostate volume, PSA, and PSAD. Other clinical information was incompletely available and thus not included in the study to avoid potential selection bias (Tripepi et al., 2010). All TRUS and MRUS biopsy cores were fixed in formalin, stained with hematoxylin and eosin (H&E)

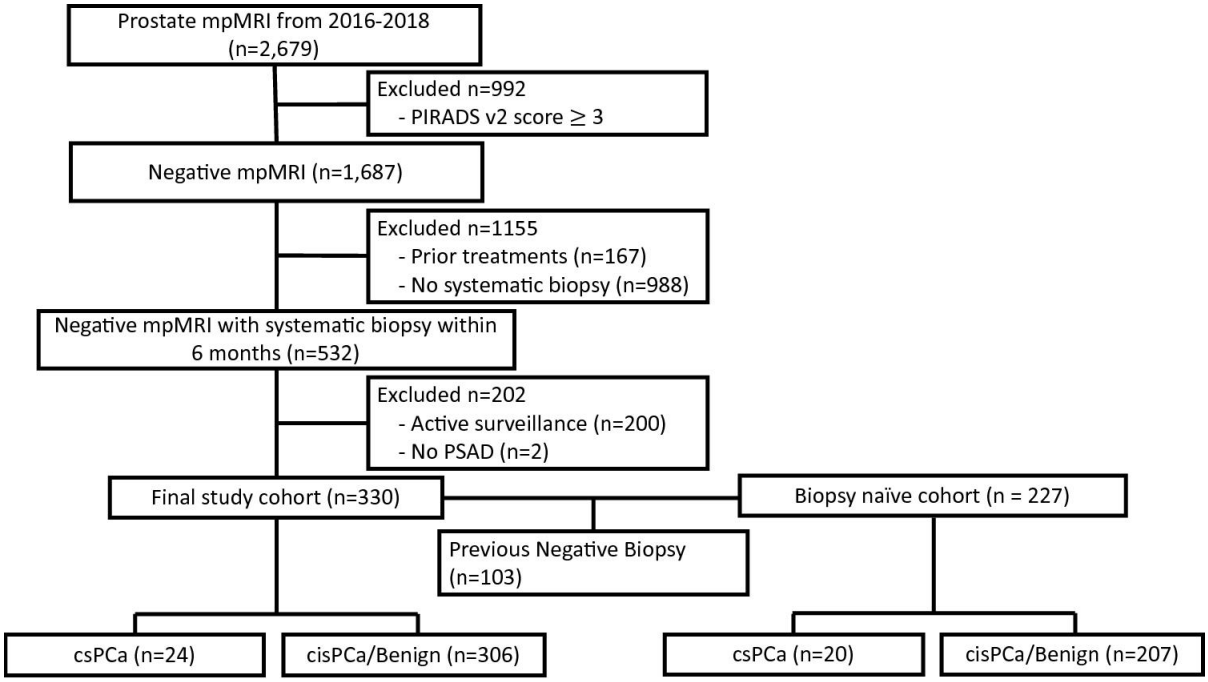


Figure 2.1: Patients cohorts selection pipeline. Finally, we generated two cohorts: a final study cohort (n=330) and a sub-cohort of biopsy naïve sub-cohort (n=227), which were used for further model construction, validation, and evaluation.

for histological evaluation performed by dedicated genitourinary pathologists as part of the routine clinical histopathological evaluation.

2.2.2 Integrative Machine Learning Model

The workflow for building our proposed iML model is shown in Figure 2.2. For a patient-basis prediction of positive or negative biopsy results, we used both apparent diffusion coefficient (ADC) maps and T2-weighted images (T2WI) from 3T mpMRI (Turkbey et al., 2019). The ADC maps were registered to T2WI through rigid spatial transformation using voxel size and real-world coordinates information for each patient (Yan et al., 2019; Cao et al., 2019a,c; Liu et al., 2013). After checking the quality of the registration, we found no observable discrepancies between T2WI and ADC. The whole prostate gland was manually segmented on T2WI slice-by-slice by the abdominal radiologist (X.X.; 8 years of experience in clinical prostate MRI interpretation) under the supervision of a senior genitourinary radiologist (Y.Y.; 20+ years of experience in clinical prostate MRI

interpretation) using OsiriX MD (ver. 11.0.3). We then applied N4 bias field correction to T2WI to compensate for the low-frequency intensity non-uniformities and applied z-score normalization to T2WI and ADC images (Tustison et al., 2010; Hectors et al., 2019).

Radiomics features were extracted from T2WI and ADC images after cropping the whole prostate, as shown in Figure 2.2. All the slices containing region of interest (ROI) of the whole prostate were used for feature extraction, and the mid-prostate slice was separately used to extract additional radiomics features. Among texture features, Gray-Level Cooccurrence Matrix (GLCM) and Gray-Level Run Length Matrix (GLRLM) were included using Pyradiomics package based on Python (Van Griethuysen et al., 2017). A total of 300 radiomics features were extracted for each patient, including 32 shape-based, 38 first-order, and 80 texture features from each of the T2WI and ADC images.

In order to pre-select important clinical features, significance levels, defined as $p<0.05$, were calculated for all routine clinical information between prostate biopsy positive group and negative group. Specifically, given the six initial clinical characteristics, Mann-Whitney U test was applied for continuous-valued features (i.e., age, PSA, PSAD, prostate volume) after checking the data normality using Kolmogorov-Smirnov test. Chi-Square test was applied for categorical features (i.e., family PCa history, prostate biopsy history). The detailed patients' clinical information can be seen in Table 2.1 We selected the clinical features that have a significant difference ($p<0.05$) between the biopsy positive and negative groups. Finally, we combined the pre-selected clinical features and all radiomics features and applied the Sequential Floating Forwarding Selection (SFFS) algorithm for integrative feature selection (Figure 2.2) (Zongker and Jain, 1996).

2.2.3 Model Comparison and Statistical Analysis

We used a quadratic-kernelized support vector machine (SVM) classifier with a class-balanced weight to train our proposed iML model. The model was validated by leave-one-out cross-validation (LOOCV) to reduce potential overfitting issues and also measure the variance of the evaluation results (Hawkins et al., 2003; Peng et al., 2013; Sutton

Table 2.1: Clinical information for 1) final study cohort and 2) biopsy naïve sub-cohort and the p-values that reflect the significance of the difference between biopsy positive and negative group within each cohort, respectively.

Feature Name	Cohort1				Cohort2			
	Overall	Biopsy Positive	Biopsy Negative	P-value	Overall	Biopsy Positive	Biopsy Negative	P-value
No. of Men (count {% of overall})	330{1}	24{7.3}	306{92.7}	-	227{100}	20{8.8}	207{91.2}	-
Prostate Volume (cc) (median {IQR})	55{39-73}	49{26-65}	55{40-77}	0.02	53{36-68}	42{26-57}	54{38-70}	0.01
Age (yr) (median {IQR})	63{58-67}	65{62-70}	62{58-67}	0.02	62{57-67}	65{63-68}	62{57-66}	0.02
PSA (ng/ml) (median {IQR})	6.3{4.6-8.9}	6.7{4.5-8.4}	6.3{4.6-9.0}	0.46	5.7{4.4-8.0}	6.3{4.5-8.1}	5.7{4.4-8.0}	0.32
PSAD (ng/ml/ml) (median {IQR})	0.11{0.08-0.16}	0.15{0.09-0.22}	0.11{0.09-0.15}	0.02	0.11{0.08-0.15}	0.15{0.09-0.22}	0.11{0.08-0.15}	0.02
PCa Family History (Yes/No: 1/0) (count {% of overall})	70{100}	7{10.0}	63{90.0}	0.46	42{100}	7{16.7}	35{83.3}	0.09
Prostate Biopsy History (Yes/No: 1/0) (count {% of overall})	103{100}	4{3.9}	99{96.1}	0.17	-	-	-	-

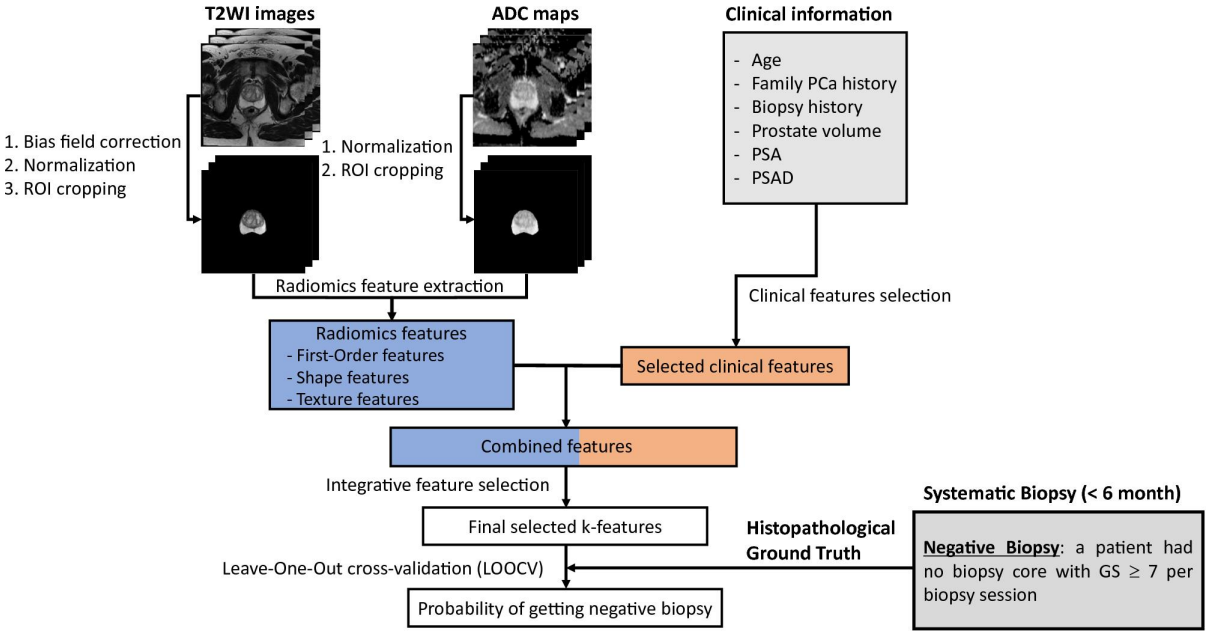


Figure 2.2: Workflow of building the integrative machine learning (iML) model for predicting negative biopsy results. The three inputs of the model are the patient's clinical information, T2WI, and ADC images. First, clinical features were selected from all clinical information, and radiomics features were extracted from the T2WI and ADC images that have been pre-processed and cropped based on ROI. Then, integrative feature selection was made based on the combination of the two categories of features. Finally, with the selected features, Leave-one-out cross-validation (LOOCV) was performed to evaluate the model's predictability.

et al., 2016; Pei et al., 2020) We first investigated the value of the iML approach by comparing the performance of iML with the models using only radiomics features or clinical features by DeLong test (DeLong et al., 1988). All models were using the same classifier, the quadratic-kernelized SVM with the class-balanced weight. We then compared the prediction performance of the proposed iML with two conventional PSAD-based strategies: 1) PSAD<0.10 ng/ml/ml as with low risks of having csPCa (Hansen et al., 2018) and 2) PSAD<0.15 ng/ml/ml as with low risks of having csPCa (Oishi et al., 2019; Distler et al., 2017; Otti et al., 2019). For each model, we identified the optimal cutoff point for the prediction of negative biopsy results by maximizing the Youden's index value (sensitivity+specificity-1) on ROC curves (Ruopp et al., 2008). NPV was calculated to measure the detection rate of true negative cases among all negative predictions, consistent with other studies (Hansen et al., 2018; Oishi et al., 2019; Distler et al., 2017; Otti et al., 2019). We further included sensitivity, specificity, and AUC in order to perform a more comprehensive evaluation to minimize the potential influence caused by data imbalance during model evaluation. Finally, all model comparisons were evaluated based on AUC with a 95% confidence interval (CI), and NPV, sensitivity, and specificity were calculated from ROC at the optimal cutoff point.

2.3 Results

The patient clinical characteristics in the final study cohort and the biopsy naïve sub-cohort are summarized in Table 2.1. Clinical information including age, prostate volume, and PSAD were selected during the procedure of clinical feature selection because of the significant difference ($p<0.05$) between biopsy positive and negative groups. Based on our inclusion criteria, 306 patients had negative biopsies and 24 patients had positive biopsies among the final study cohort (n=330). 207 patients had negative biopsies and 20 patients had positive biopsies among the biopsy naïve sub-cohort (n=227).

There were nine total integrative features comprised of six radiomics and three clinical features and are summarized in Table 2.2. Figure 2.3 shows representative examples of

Table 2.2: Nine selected features after integrative feature selection.

Selected Features	**Type**	**Imaging Sequence**
Gray Level Non-uniformity	GLRLM	ADC
Run Length Non-uniformity	GLRLM	ADC
Sum Squares	GLCM	T2WI
Least Axis Length	Shape	ADC/T2WI
Major Axis Length	Shape	ADC/T2WI
Minor Axis Length	Shape	ADC/T2WI
Age	Clinical Information	–
PSAD	Clinical Information	–
Prostate Volume	Clinical Information	–

3T mpMRI-based radiomics features, stratified as negative (top) and positive (bottom) biopsies. The six radiomics features consisted of three shape and three texture features. The shape features (Minor Axis_Length, Major Axis_Length and Least Axis_Length) descripted the shape and size information of the ROI region of prostate, and the texture features (Sum Squares, Gray Level Non-Uniformity and Run Length Non-Uniformity) descripted the texture information of the ROI region of prostate, on T2WI and ADC images. With negative MRI, the selected radiomics features show different visual patterns between two groups (A and B vs. C and D), as shown in the spider plots.

A and C in Figure 2.4 show the ROC comparisons between the proposed iML model and machine learning models with an individual feature group in two patient cohorts. The proposed iML approach achieved the highest AUC ($p<0.05$), compared with the models using an individual group of radiomics or clinical features, in both cohorts (Table 2.3). The AUC, and sensitivity, specificity and NPV that based on the optimum cutoff points of the iML approach were 0.798 (95% CI, 0.711-0.885), 83.3%, 75.2%, and 98.3% respectively in the final study cohort, which improved the AUC of [13.2%, 17.5%] compared with clinical-only and radiomics-only models ($p<0.05$), respectively. For the biopsy naïve cohort, the iML approach reached AUC, sensitivity, specificity and NPV of 0.749 (95% CI, 0.645-0.854), 85.0%, 72.0%, and 98.0% respectively. It thus improved the AUC of [10.3%,

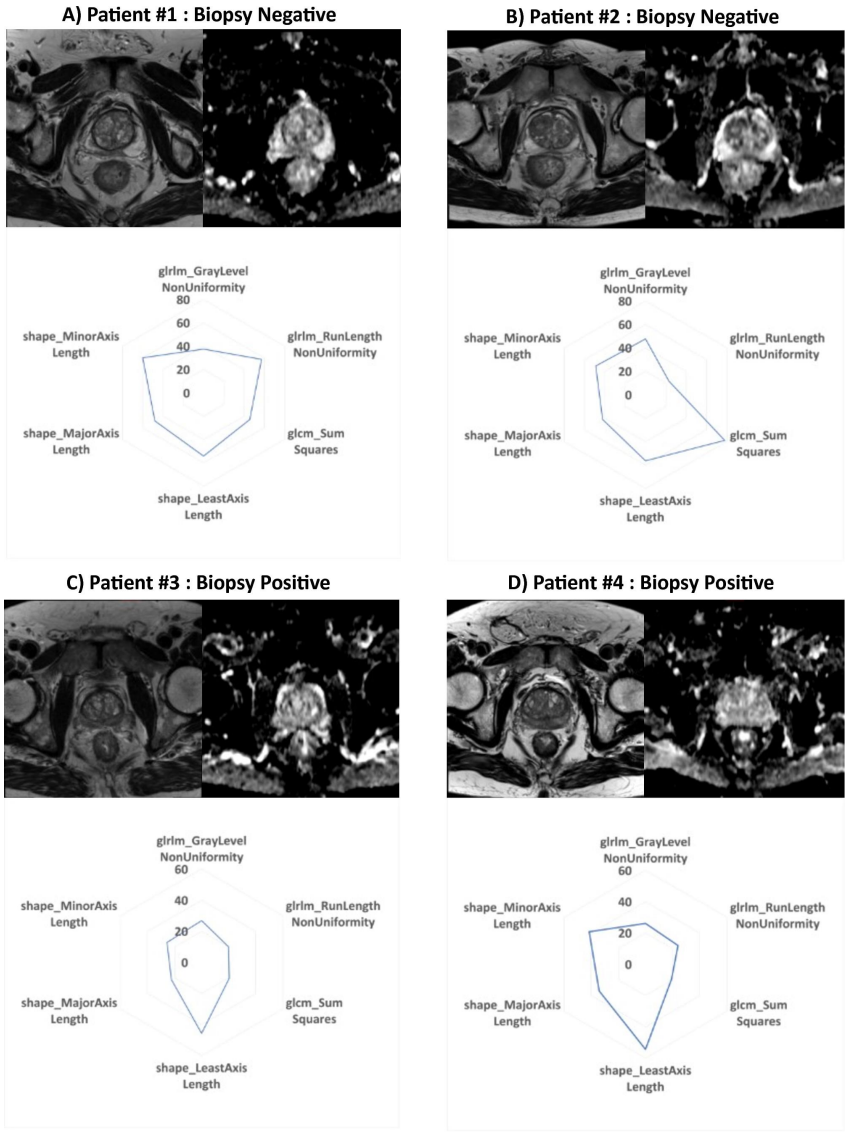


Figure 2.3: Visualizations of mpMRI (T2WI and ADC) images and the values of their corresponding radiomics features for patients with negative mpMRI; A) and B): patients with negative biopsies, and C) and D): patients with positive biopsies. Visualizations of the radiomics feature values are shown in spider plots, where the length of a feature's spoke is proportional to the value of that feature relative to that feature's maximum values across all patients. The numbers adjacent to each level of polygon represents the proportion value of the spoke at that level.

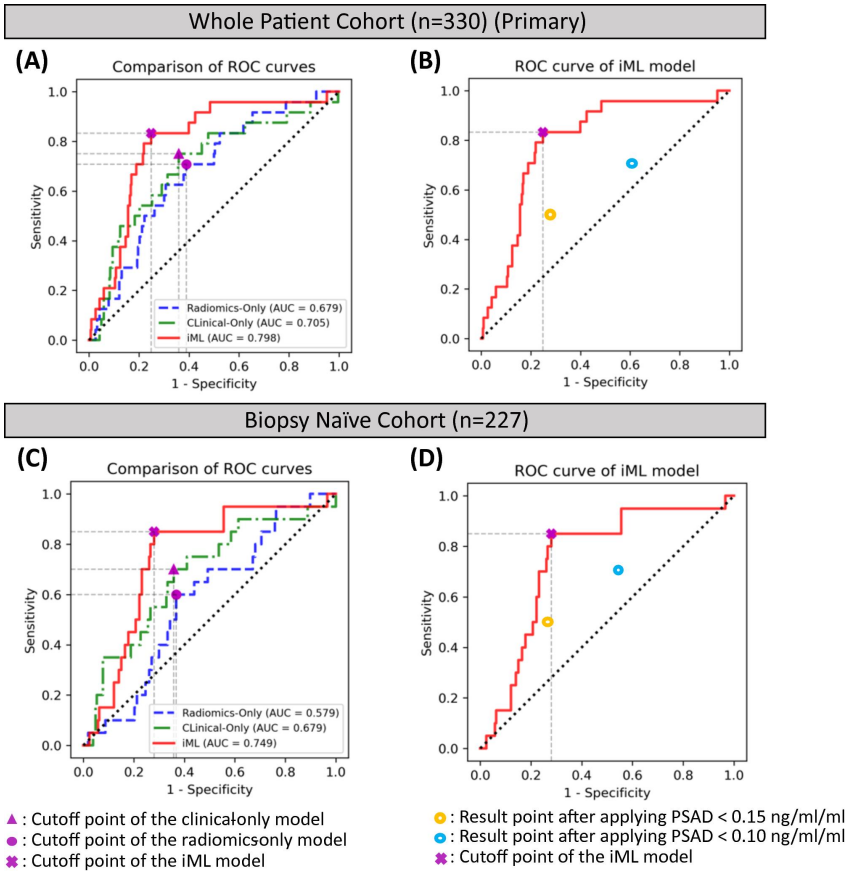


Figure 2.4: Comparisons between iML and machine learning approaches using individual feature groups for both patient cohorts. Red solid, blue dash, and green dot-dash curves are the ROC curves of the radiomics-only, clinical-only, and iML models. Horizontal and vertical gray dash lines of each optimal cutoff point aimed to visualize sensitivity and "1-specificity" value on each ROC curve.

29.4%], compared with clinical-only and radiomics-only models ($p<0.05$), respectively.

The comparison results between the PSAD-based risk prediction methods and the iML model conducted on the same study population are shown in Table 2.4, and B and D in Figure 2.4. The iML approach achieved higher specificity/sensitivity ($p<0.05$) while keeping the similar sensitivity/specificity to the results conducted by using thresholds of PSAD=0.10 ng/ml/ml and of PSAD=0.15 ng/ml/ml (B and D in Figure 2.4) for both cohorts (Hansen et al., 2018; Oishi et al., 2019; Distler et al., 2017; Otti et al., 2019). Moreover, for both final study cohort and biopsy naïve cohort, our proposed iML approach achieved NPVs of [98.3%, 98.0%], showed improvement compared with PSAD-based risk prediction methods, resulted in NPVs of [94.6%, 93.9%] for PSAD<0.10 ng/ml/ml and

Table 2.3: Comparisons of prediction performances of the proposed iML approach and the machine learning approaches that were clinical-only or radiomics-only for both final study cohort and biopsy naïve sub-cohort, respectively. P-values were calculated by DeLong test, for comparisons between AUCs of models using each individual feature group and the proposed iML model.

Method	**Final Study Cohort (N=330)**				
	AUC [%95 CI]	**Sensitivity(%)**	**Specificity(%)**	**NPV(%)**	**p-value**
Clinial-only	0.705 [0.589, 0.821]	75.0	64.1	97.0	0.011
Radiomics-only	0.679 [0.571, 0.787]	70.8	61.2	96.4	0.006
iML	**0.798 [0.711, 0.885]**	**83.3**	**75.2**	**98.3**	-
Method	**Biopsy Naïve Sub-cohort (N=227)**				
	AUC [%95 CI]	**Sensitivity(%)**	**Specificity(%)**	**NPV(%)**	**p-value**
Clinial-only	0.679 [0.553, 0.805]	70.0	64.3	95.7	<0.001
Radiomics-only	0.579 [0.464, 0.694]	60.0	63.3	94.2	0.046
iML	**0.749 [0.645, 0.854]**	**85.0**	**72.0**	**98.0**	-

NPVs of [94.9%, 93.9%] for PSAD<0.15 ng/ml/ml (Hansen et al., 2018; Oishi et al., 2019; Distler et al., 2017; Otti et al., 2019).

Comparisons of prediction performances using different approaches are shown in Table 2.4. Specifically, for the final study cohort, the iML approach improved the results conducted by using a threshold of PSAD=0.10 ng/ml/ml and a threshold of PSAD=0.15 ng/ml/ml on sensitivity of [17.7%, 66.6%], specificity of [88.5%, 2.4%] and NPV of [3.9%, 3.6%] (Hansen et al., 2018; Oishi et al., 2019; Distler et al., 2017; Otti et al., 2019). For the biopsy naïve cohort, the iML approach improved the results on sensitivity of [21.4%, 70.0%], specificity of [+60.4%, -3.2%] and NPV of [4.4%, 4.4%], respectively. Figure 2.5 visualized the prediction results using PSAD-based approaches and iML on both final study cohort and biopsy naïve cohort. The histograms show iML had the highest true positive ratio and the smallest true negative ratio among all methods (see Table 2.4 for statistical comparison between different methods).

Table 2.4: Comparisons of prediction performances of the proposed iML approach and approaches that using PSAD-based risk assessments for both final study cohort and biopsy naïve sub-cohort, respectively. P-values were calculated by Chi-square test, for comparisons of each measurement between PSAD-based prediction models and the proposed iML model.

Method	**Final Study Cohort (N=330)**					
	Sensitivity(%)	**p-value**	**Specificity(%)**	**p-value**	**NPV(%)**	**p-value**
PSAD<0.10	70.8	0.303	39.9	<0.001	94.6	0.048
PSAD<0.15	50.0	0.014	73.2	0.579	94.9	0.044
iML	**83.3**	-	**75.2**	-	**98.3**	-
Method	**Biopsy Naïve Sub-cohort (N=227)**					
	Sensitivity(%)	**p-value**	**Specificity(%)**	**p-value**	**NPV(%)**	**p-value**
PSAD<0.10	70.0	0.451	44.9	<0.001	93.9	0.171
PSAD<0.15	50.0	0.018	**74.4**	0.579	93.9	0.010
iML	**85.0**	-	72.0	-	**98.0**	-

2.4 Discussion

We proposed an integrative machine learning (iML) model as a potential triage test to obviate biopsy when 3T mpMRI was negative. Our findings showed that integrating both MRI and clinical information helped improve the prediction of the biopsy results ($p<0.05$), compared with the machine learning approaches conducted by individually using either MRI-based radiomics features or clinical features. Recent review studies reported that common strategies of using PIRADS<3 as a triage test to obviate biopsy resulted in NPVs with a range of 80.5% to 92.3%, and the PSAD-based assessment improved the NPVs to be in the range of 83.1% to 93.4% for predicting negative biopsy results among patients with negative MRI (Pagniez et al., 2020; Moldovan et al., 2017). In this study, the final study cohort had NPV of 92.7% and improved to 98.3% using the iML approach. This performance of the negative biopsy results was higher than other studies' NPVs, ranged from 89.0% to 89.9%, and the PSAD-based assessment with NPVs, ranged from 83.1 to 93.4% (Oishi et al., 2019; Distler et al., 2017; Pagniez et al., 2020). In the biopsy naïve

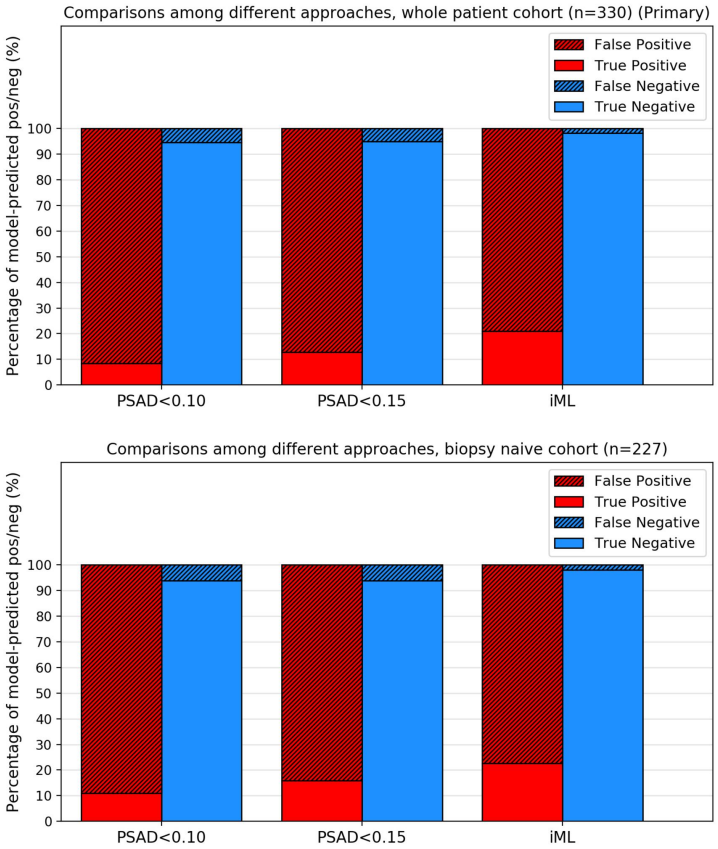


Figure 2.5: Histogram visualization of performance comparisons between our proposed iML approach and PSAD-based strategies from other studies, for both final study cohorts (left) and biopsy naïve cohort (right). Performances are measured by the percentage of true positive, false positive, true negative, and false negative in both cohorts. Red bars reveal prediction performance based on cases that are predicted as having biopsy positive, and blue bars reveal prediction performances based on cases that are predicted as having biopsy negative.

cohort, iML improved NPV from 91.2% to 98.0%, also higher than reported by other existing studies and the PSAD-based assessment strategies (Hansen et al., 2018; Otti et al., 2019; Pagniez et al., 2020). Furthermore, our results on both patient cohorts also achieved improvements in sensitivity/specificity with a small cost of specificity/sensitivity in comparison with reported specificities and sensitivities (Hansen et al., 2018; Oishi et al., 2019; Otti et al., 2019).

Prior studies had shown the MRI-based radiomics features had excellent performance for the prediction and aggressiveness assessment of PCa (Yan et al., 2019; Vignati et al.,

2015; Gnep et al., 2017; Nketiah et al., 2017; Algohary et al., 2018; Starmans et al., 2019; Varghese et al., 2020; Zhang et al., 2020). With the similar settings as the previous studies, our study also took the first-order, shape and texture features into consideration as the MRI-based radiomics features in order to comprehensively extracted the PCa-related information from T2WI and ADC images. The study showed the improved performances of using the integration of radiomics features and clinical information via machine learning when predicting patient-basis negative biopsy results, compared with the situation using individual feature groups only, or using the pre-existing PSAD-based methods.

2.4.1 Limitations

Our study includes a few limitations. The study included 330 patients with negative MRI who underwent systematic biopsies within six months. The study cohort was identified after investigating all in-house prostate 3T mpMRI scans for three years at a single academic institution (n=2,679). Although the size of the dataset was relatively small and contained imbalanced distribution between positive and negative biopsies, the data characteristic was similar to the previously investigated studies due to the study objectives (Hansen et al., 2018; Oishi et al., 2019; Distler et al., 2017; Otti et al., 2019). Moreover, the study did not conduct a separate study on the cohort of patients that have a prior negative biopsy (n=103) was conducted due to the limited number of positive biopsy cases (n=4). Additionally, we used LOOCV for the model evaluation due to the limited number of data with class imbalance (Hawkins et al., 2003), consistent with other studies when only limited data was available (Peng et al., 2013; Sutton et al., 2016; Pei et al., 2020). Our future works would include continuous collection of available data to evaluate our model with an external testing set. We believe this will further solidify our findings.

2.4.2 Conclusion

In conclusion, the negative biopsy results were highly predictable among patients with negative prostate MRI using the integrative machine learning (iML) model. The integra-

tion of MRI-based radiomics and clinical features improved the performance in predicting negative biopsy results. The proposed iML model outperformed the existing PSAD-based strategies with NPV of 98.3%, in the final study cohort, and NPV of 98.0%, in the biopsy naïve sub-cohort, respectively. It can thus be used to stratify patients who should obviate biopsies, potentially reducing the number of unnecessary biopsies for patients with negative prostate MRI.

CHAPTER 3

Integrative Radiomics-Based Machine Learning Model Predicting Pelvic Lymph Node Invasion

3.1 Introduction

The presence of lymph node invasion (LNI) is a poor prognostic marker for patients who have undergone radical prostatectomy (RP) for prostate cancer (Wilczak et al., 2018). An extended pelvic lymph node dissection (ePLND) is standard for staging of prostate cancer and a key decision after RP is to determine which patients should undergo ePLND (Chen et al., 2019). However, because of the morbidity of the procedure including lymphocele, lymphedema, and thromboembolic events, the American Urology Association (AUA) guidelines reserve ePLND for patients with a high risk of nodal involvement (Fossati et al., 2017). Current indications for ePLND are limited and vary considerably (Mottet et al., 2021; Schaeffer et al., 2021).

To select the most appropriate prostate cancer (PCa) patients for ePLND, several well-known nomograms have been developed to predict LNI utilizing clinical test results, demographic information and biopsy-related results (Yu et al., 2011; Venclovas et al., 2021; Soeterik et al., 2021; Roach et al., 1994; MSKCC; Briganti et al., 2012). Some recent studies showed the prostate specific membrane antigen (PSMA) positron emission tomography (PET) scans provided promising results in preoperative nodal staging for patients with PCa, and also could assist the prediction of LNI (Sprute et al., 2021; Cysouw et al., 2021; Zamboglou et al., 2019; Barbosa et al., 2018; Weinreb et al., 2016). However, the relatively high price and low availability in many areas and regions made the PSMA-PET not easily achieved in the current clinical practice. Recent studies show

that findings from multiparametric magnetic resonance imaging (mpMRI), which costs less and is more achievable, may help predict LNI, as described by the Prostate Imaging Reporting and Data System version 2 (PI-RADS v2) (Weinreb et al., 2016; Huang et al., 2020a; Hatano et al., 2020). However, the current assessment methods have limited prediction performance, where the area under the curve (AUC) ranged from 0.720 to 0.806 (Venclovas et al., 2021; Soeterik et al., 2021).

Radiomics, a quantitative imaging technique, provides quantitative imaging information within a given region of interest (ROI) (Lambin et al., 2017; Tomaszewski and Gillies, 2021; Zwanenburg et al., 2020; Gillies et al., 2016), and has shown promising results to help predict the aggressiveness of PCa (Cuocolo et al., 2021; Gugliandolo et al., 2021; Hectors et al., 2019; Yan et al., 2019; Zhang et al., 2020) and predict LNI in patients with colorectal cancer (Li et al., 2020; Huang et al., 2016). The aim of the study was to build a prediction model to predict LNI in patients with PCa via machine learning, integrating MRI-based radiomics features with routinely used clinical features including clinical test results, demographic information and biopsy-related results. We evaluated the benefits of integrating the radiomics features and clinical features, and the integrative radiomics model (IRM) was compared with commonly used nomograms (Yu et al., 2011; Roach et al., 1994; MSKCC; Briganti et al., 2012) when predicting LNI in patients with PCa. To our best knowledge, this is the first study trying to utilize mpMRI-based radiomics features to build the machine learning model to predict LNI for patients with PCa.

3.2 Material and Methods

3.2.1 Study Population and MRI Data

The single institutional retrospective study was approved by the institutional review board (IRB) and was conducted in compliance with the 1996 Health Insurance Portability and Accountability Act (HIPAA). The initial patient cohort consisted of all patients who had positive 3 Tesla (3T) mpMRI (defined as having any lesions with PI-RADS$\geq$3), from

January 2010 to December 2019 in the institution. The 3T mpMRI scans were performed on one of the five 3T scanners with standardized protocols (Prisma, Skyra, Trio, Vida and Verio (Siemens Healthcare)). All mpMRI scans were reviewed by two genitourinary radiologists following the PI-RADS v2.1 guideline (Turkbey et al., 2019).

The detailed flowchart of patient inclusion criteria is shown in Figure 3.1. The study cohort consisted of patients who underwent RP with ePLND within six months after mpMRI. Patients meeting one or more of the following criteria were excluded: 1) prior treatment for PCa (radiotherapy or focal ablation and anti-hormonal therapy); 2) poor mpMRI quality with strong artifact and/or extensive motion blur; 3) MR examination in other institutions 4) lack of formal preoperative biopsy report or review. All preoperative mpMRIs were interpreted by an abdominal imaging fellow and then reviewed by one of three board-certified attending genitourinary radiologists with 8-20 years of experience. The final study cohort comprised 244 patients, with median age of 62 years (IQR: 58-67) and was subdivided into training/validation set (mpMRI examination dates between 2010 and 2016; n =160), and testing set (mpMRI examination dates between 2017 and 2019; n=84). LNI positive patients were defined as those with LNI confirmed at the final histopathologic examination, and the rest were defined as LNI-negative patients (Amin et al., 2017).

For each patient, clinical features that positively related to the presence of LNI were initially included (Yu et al., 2011; Roach et al., 1994; MSKCC; Briganti et al., 2012; Huang et al., 2020a; Hatano et al., 2020). The clinical features consist of demographic information: age; clinical test results: serum level of prostate-specific antigen (PSA), PSA density (PSAD), digital rectum exam results (DRE); mpMRI observations: prostate volume, MR stage, maximum index lesion diameter, is PI-RADSv2.1 score>4, is index lesion located or involved in peripheral zone (PZ); biopsy-related results: number of positive prostate biopsy cores, percentage of positive prostate biopsy cores, the primary and secondary Gleason Score, and Gleason Grade Group (Yu et al., 2011; Roach et al., 1994; MSKCC; Briganti et al., 2012; Huang et al., 2020a; Hatano et al., 2020). Other clinical information that was partially available was not included to avoid selection bias

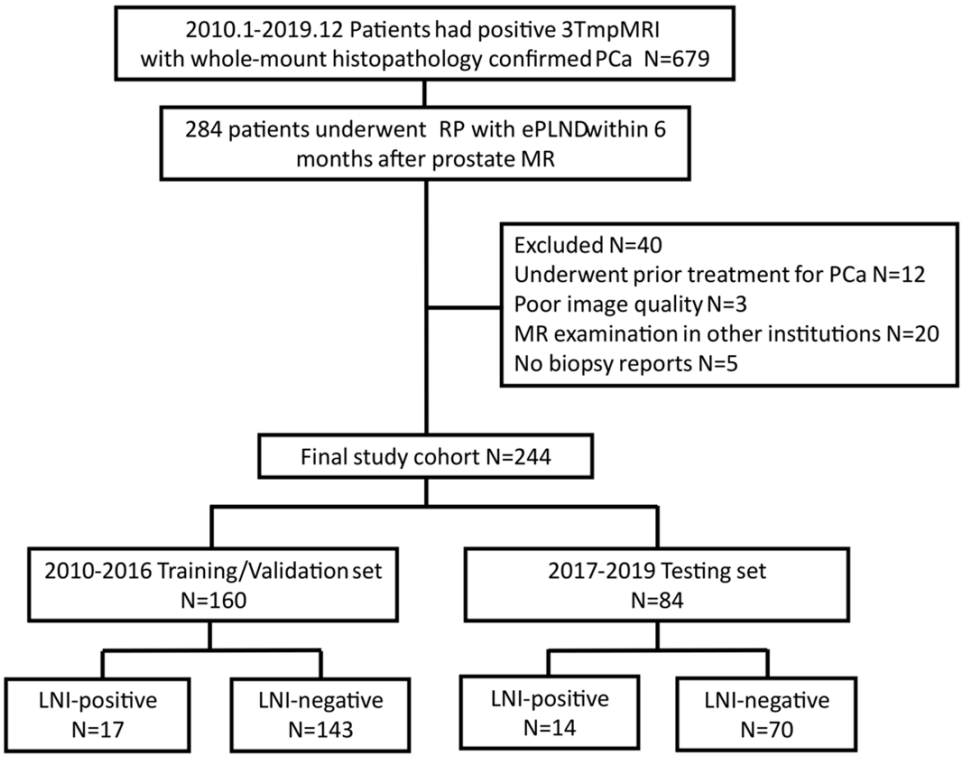


Figure 3.1: The inclusion workflow of the study population. Patients whose mpMRI examination dates between 2010 and 2016 were included into training/validation set (n=160), patients whose mpMRI examination dates between 2017 and 2019 were included into testing set (n=84).

(Tripepi et al., 2010). The detailed clinical feature information is described in Table 3.1.

3.2.2 Integrative Radiomics Model

The detailed workflow of the IRM construction is illustrated in Figure 3.2. We utilized both T2-weighted images (T2WI) and apparent diffusion coefficient (ADC) maps from mpMRI to build the patient-wise LNI prediction model. The ADC maps were registered to T2WI images using spatial transformation utilizing world coordinate system information, and no observable discrepancies were discovered after manually checking the registration results (Yan et al., 2019; Zheng et al., 2022b; Cao et al., 2019a). For each patient, the index lesion was considered as the one with the highest PI-RADSv2.1 score, or highest volume in cases where equivalent PI-RADSv2.1 scores were assigned to different lesions. The slice-by-slice manual segmentation of each index lesion was performed on T2WI

Table 3.1: Characteristics of clinical features for patients included in the study cohort.

Feature Name	**Training/Validation set (2010-2016)**			**Testing set (017-2019)**		
	Overall	**LNI-Positive**	**LNI-Negative**	**Overall**	**LNI-Positive**	**LNI-Negative**
No. of Men, (n {%})	160{100}	17{10.6}	143{89.4}	84{100}	14{16.7}	70{83.3}
Age (yr), (median {IQR})	62{58-67}	62{57.5-65}	62{58-67}	64{59-68}	63{60-68}	64{59-68}
PSA (ng/ml), (median {IQR})	6.5{4.9-10.0}	10.0{6.4-24.0}	6.1{4.5-9.0}	8.3{5.5-12.6}	10.3{8.0-18.3}	7.3{5.3-12.2}
PSAD (ng/ml/ml), (median {IQR})	0.19{0.13-0.29}	0.22{0.17-0.58}	0.18{0.12-0.27}	0.19{0.13-0.37}	0.25{0.15-0.40}	0.19{0.13-0.34}
DRE (Yes/No: 1/0), (n {%})	0: 116 {72.5} 1: 44 {27.5}	0: 9 {52.9} 1: 8 {47.1}	0: 107 {74.8} 1: 36 {25.2}	0: 67 {79.8} 1: 17 {20.2}	0: 9 {64.3} 1: 5 {35.7}	0: 58 {82.9} 1: 12 {17.1}
Prostate volume (cc), (median {IQR})	36.0{29.0-44.0}	40.0{37.0-45.0}	35.0{29.0-44.0}	37.3{29.0-48.2}	42.0{29.0-47.3}	36.3{30.0-49.0}
Maximum index lesion diameter (mm), (median {IQR})	1.6{1.1-1.2}	2.4{1.4-3.0}	1.5{1.1-2.0}	2.0{1.2-2.6}	2.1{1.2-2.9}	1.9{1.2-2.5}
MR stage (1: Organ-Confined; 2:Extraprostatic Extension; 3:Seminal Vesicle Invasion), (n {%})	1: 93 {58.1} 2: 56 {35.0} 3: 11 {6.9}	1: 5 {29.4} 2: 10 {58.8} 3: 2 {11.8}	1: 88 {61.5} 2: 46 {32.2} 3: 9 {6.3}	1: 40 {47.6} 2: 37 {44.0} 3: 7 {8.3}	1: 3 {21.4} 2: 8 {57.1} 3: 3 {21.4}	1: 37 {52.9} 2: 29 {41.4} 3: 4 {5.7}
Is index lesion located or involved in PZ (Yes/No: 1/0), (n {%})	0: 47 {29.4} 1: 113 {70.6}	0: 2 {11.8} 1: 15 {88.2}	0: 45 {31.5} 1: 98 {68.5}	0: 15 {17.9} 1: 69 {82.1}	0: 1 {7.1} 1: 13 {92.9}	0: 14 {20.0} 1: 56 {80.0}
Is PIRADSv2.1>4? (Yes/No: 1/0), (n {%})	0: 83 {51.9} 1: 77 {48.1}	0: 6 {35.3} 1: 11 {64.7}	0: 77 {53.8} 1: 66 {46.2}	0: 35 {41.7} 1: 49 {58.3}	0: 5 {35.7} 1: 9 {64.3}	0: 30 {42.9} 1: 40 {57.1}
Primary Gleason Score, (n {%})	3: 73 {45.6} 4: 83 {51.9} 5: 4 {2.5}	3: 3 {17.6} 4: 12 {70.6} 5: 2 {11.8}	3: 70 {49.0} 4: 71 {49.7} 5: 2 {1.4}	3: 31 {36.9} 4: 52 {61.9} 5: 1 {1.2}	3: 2 {14.3} 4: 12 {85.7} 5: 0 {0}	3: 29 {41.4} 4: 40 {57.1} 5: 1 {1.4}
Secondary Gleason Score, (n {%})	3: 63 {39.4} 4: 63 {39.4} 5: 34 {21.2}	3: 6 {35.3} 4: 5 {29.4} 5: 6 {35.3}	3: 57 {39.9} 4: 58 {40.6} 5: 28 {19.6}	3: 37 {44.0} 4: 39 {46.4} 5: 8 {9.5}	3: 5 {35.7} 4: 8 {57.1} 5: 1 {7.1}	3: 32 {45.7} 4: 31 {44.3} 5: 7 {10.0}
Gleason Grade Group, (n {%})	1: 19 {11.9} 2: 40 {25.0} 3: 43 {26.9} 4: 37 {23.1} 5: 21 {13.1}	1: 0 {0} 2: 2 {11.8} 3: 5 {29.4} 4: 5 {29.4} 5: 5 {29.4}	1: 19 {13.3} 2: 38 {26.6} 3: 38 {26.6} 4: 32 {22.4} 5: 16 {11.2}	1: 9 {10.7} 2: 21 {25.0} 3: 28 {33.3} 4: 18 {21.4} 5: 8 {9.5}	1: 0 {0} 2: 2 {14.3} 3: 5 {35.7} 4: 6 {42.9} 5: 1 {7.1}	1: 9 {12.9} 2: 19 {27.1} 3: 23 {32.9} 4: 12 {17.1} 5: 7 {10.0}
Number of positive biopsy cores, (median {IQR})	5{4-7}	6{4.5-9}	5{3-7}	5{3.5-8}	7.5{3-9}	5{4-7}
Percentage of positive biopsy cores, (median {IQR})	0.39{0.25-0.53}	0.50{0.33-0.75}	0.36{0.25-0.5}	0.48{0.29-0.67}	0.61{0.38-0.73}	0.40{0.29-0.67}

images by three clinical research fellows with more than three years of experience in segmenting prostate MRI. All segmentations were reviewed and validated by an abdominal radiologist (X.X., eight years of experience in prostate mpMRI interpretation) together with a board-certified abdominal radiologist (Y.Y., 20+ years of experience in prostate MRI interpretation) blinded to the histopathology reports of LNI. N4 bias field correction was applied to T2WI images to compensate for low-frequency intensity variations, and z-score normalization was then applied to both ADC and T2WI images for each patient (Hectors et al., 2019; Tustison et al., 2010).

Radiomics features were extracted from the region of interest (ROI) of the index lesions from both ADC and T2WI images, as shown in Figure 3.2. Shape features, first-order

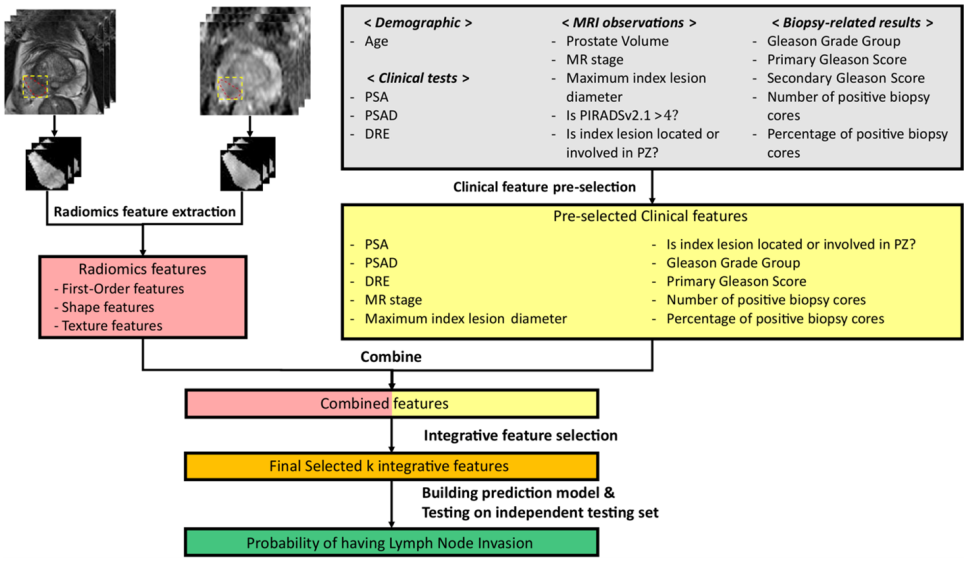


Figure 3.2: The Workflow of building the IRM for predicting LNI for patients with PCa. First, clinical features were pre-selected from initial clinical features, and radiomics features were extracted from ADC and T2WI images from the ROI of the index lesion for each patient. Then, integrative features were selected from the combination of clinical features and radiomics features. Finally, an SVM-based LNI prediction model was trained using the selected integrative features on the training/validation set and then tested on the internal independent testing set.

features and texture features, including Gray-level Run Length Matrix (GLRLM) features, Gray-level Co-occurrence Matrix (GLCM) features, Gray-level Size Zone Matrix (GLSZM) features, Neighboring Gray Tone Difference Matrix (NGTDM) features and Gray Level Dependence Matrix (GLDM) features, were extracted. We used the package PyRadiomics v3.0.1 in Python 3.6 for the radiomics feature extraction, following the guideline of Image Biomarker Standardization Initiative (IBSI), with slight optimization in discretizing and resampling of the input (Zwanenburg et al., 2020; Van Griethuysen et al., 2017). For each imaging modality, 16 shape features, 19 first-order features and 75 texture features were extracted and thus, 220 radiomics features were extracted for each patient.

We pre-selected the clinical features from the initially included demographic information, clinical tests results, MRI observations and biopsy-related results, via univariate logistic regression based on the training/validation set, thresholded by significance level $p<0.10$. Finally, the integrative features were selected from the integration of the pre-selected clinical features and the extracted radiomics features by the Sequential Forward

Floating Selection (SFFS) method, as shown in Figure 3.2 (Zongker and Jain, 1996).

3.2.3 Model Comparison and Statistical Analysis

A quadratic kernelized support vector machine (SVM) classifier was proposed to build the proposed IRM, implemented using package Scikit-learn v1.0.1 in Python 3.6. In this study, we applied 5-fold cross validation on the training/validation set to find the optimal hyper-parameters that resulted in the highest AUCs for the models. Then, the optimal hyper-parameters were fixed and the models were finally trained using the entire training/validation set and tested and evaluated in the internal independent testing set.

We first investigated the benefits of using the IRM approach compared with the approaches using radiomics features only or clinical features only by Delong test (DeLong et al., 1988), all using SVM classifier. Then, we compared the prediction performance of the proposed IRM approach with the pre-existing nomograms by the Delong test and Chi-square test. All comparisons were conducted in the testing set.

The cut-off points for machine learning models were chosen by maximizing Youden's index value (sensitivity+specificity-1) on the receiver operating characteristic (ROC) curve. The cut-off points for the pre-existing nomograms were chosen following suggestions from the National Comprehensive Cancer Network (NCCN) and the European Association of Urology (EAU) guidelines (Mottet et al., 2021; Schaeffer et al., 2021). Sensitivity, specificity, negative predictive value (NPV) and positive predictive value (PPV), that were calculated from the cut-off points and were compared via Chi-squared test or Fisher's exact test, were included in order to perform a more comprehensive evaluation to minimize the potential influence caused by data imbalance during model evaluation. The area under the curve (AUC) with 95% confidence interval (CI) calculated based on ROC curves were used to evaluate and compare the models' performance, and compared via the Delong test.

Table 3.2: Eleven selected features after integrative feature selection.

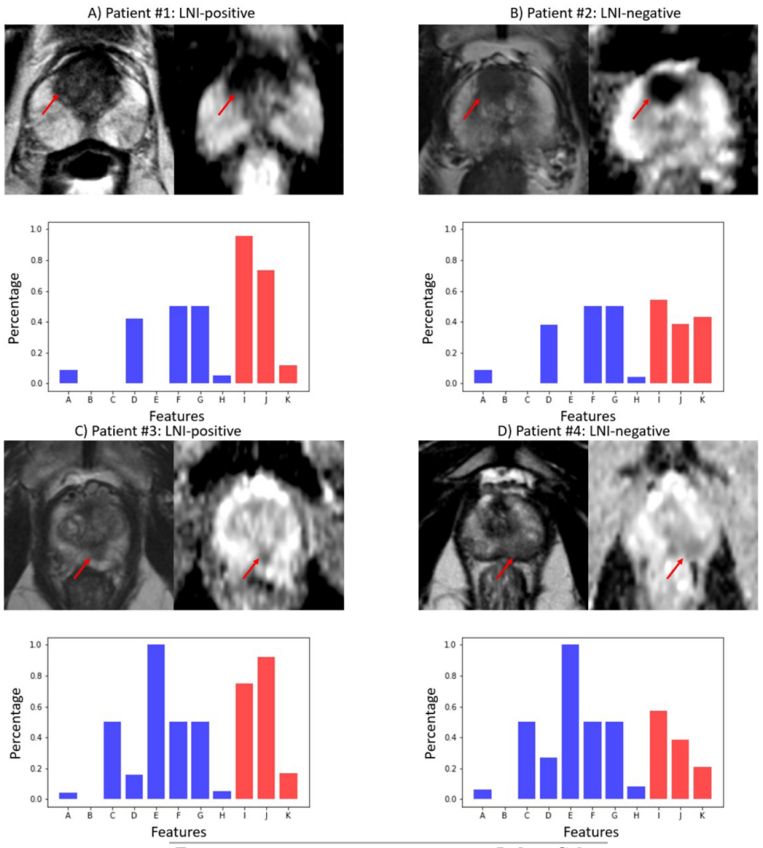
Feature Names	Feature Categories			
	Clinical Features			Radiomics Features
	Clinical test results	MRI observations	Biopsy-related results	
PSAD	×			
DRE	×			
MR Stage		×		
Maximum index lesion diameter		×		
Is index lesion located or involved in PZ		×		
The Primary Gleason Score			×	
Gleason Grade Group			×	
Percentage of positive biopsy cores			×	
Shape_Elongation				×
Shape_Maximum2DDiameterColumn				×
First-order_10Percentile_T2WI				×

3.3 Results

The final study cohort included 213 LNI-negative and 31 LNI-positive patients (n=244). The clinical features of the patients in the final study cohort are described in Table 3.1. The following ten features were pre-selected including PSA, PSAD, DRE, MR stage, maximum index lesion diameter, is index lesion located or involved in PZ, number of positive biopsy cores, percentage of positive biopsy cores, primary Gleason Score, and Gleason Grade Group, among the initial fourteen clinical features.

After integrative feature selection from the combination of radiomics features and the pre-selected features, a total of eleven features were finally selected by using the SFFS algorithm, as shown in Table 3.2. In Figure 3.3, examples of 3T mpMRI images, consisting of ADC and T2WI images from two LNI-positive patients (A and C) and two LNI-negative patients (B and D) are illustrated. With similar distributions of the clinical features (in blue), the radiomics features (in red) could show visually different patterns between the two groups (A vs. B and C vs. D), which can be observed through the histogram plots.

The model performances were all evaluated and compared based on the results conducted from the testing set. The ROC curves between the proposed IRM and the machine learning models using each individual feature group are shown in Figure 3.4(A).



Feature names	Index	Color
PSAD	A	Blue
DRE	B	Blue
MR stage	C	Blue
Maximum index lesion diameter	D	Blue
Is index lesion located or involved in PZ?	E	Blue
Primary Gleason Grade	F	Blue
Gleason Grade Group	G	Blue
Percentage of positive biopsy cores	H	Blue
Shape_Elongation	I	Red
Shape_Maximum2DDiameterColumn	J	Red
First-Order_10Percentile	K	Red

Figure 3.3: Visualizations of ADC and T2WI images and the values of the selected integrative features of each patient from patients that were predicted correctly using IRM at the optimal cutoff point. A) and C) are LNI-positive patients, B) and D) are LNI-negative patients. Visualizations of the values of the integrative features are shown in the histograms. The correspondence of features names and the indices on x-axis are shown in the table at the bottom, due to the limited space in the histogram. The height of a feature's bar, also known as the percentage number as shown in y-axis, is proportional to the value of that feature of the patient compared with the feature's maximum value across all patients. Blue bars are visualizations for clinical features, red bars are visualizations for radiomics features.

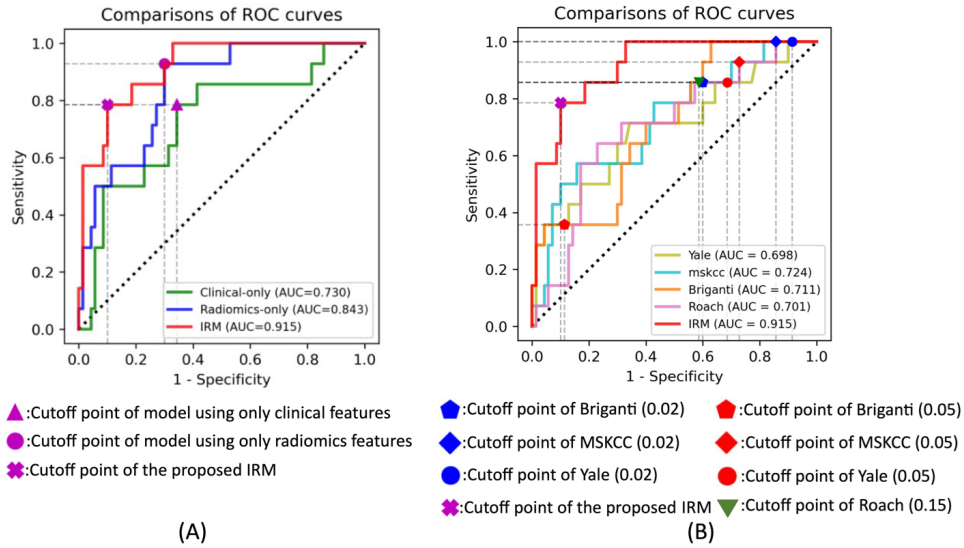


Figure 3.4: A): ROC curve comparisons between the proposed IRM and other machine learning approaches that only utilizing each individual feature group (clinical features or radiomics features). Green, blue and red curves are the ROC curves of the model that using radiomics features only, model that using clinical features only, and the IRM models, respectively. B): ROC curve comparisons between the proposed IRM and Yale, MSKCC, Briganti and Roach. Olive, cyan, orange, pink and red curves are the ROC curves of the Yale, MSKCC, Briganti, Roach and the IRM model.

The proposed IRM achieved an AUC, sensitivity, specificity, NPV, and PPV of 0.915 (95% CI: 0.846-0.984), 0.786, 0.90, of 0.955 and 0.611 respectively. Compared with the other two models using only clinical features and only radiomics features, the proposed IRM achieved the highest AUC ($p<0.05$), with a relative improvement in AUC of 25.3% and 8.5%, respectively. With the same sensitivity/specificity, the IRM achieved higher specificity/sensitivity (Figure 3.4(A)). Details are shown in Table 3.3.

Table 3.3: Comparisons of the prediction performance between the proposed IRM model and the other two machine learning approaches that using either only clinical features or radiomics features. The p values were calculated using Delong test for the comparisons of AUCs.

Methods	**AUC [95%CI]**	**Sensitivity(%)**	**Specificity(%)**	**NPV(%)**	**PPV(%)**	**P-value**
Clinical-only	0.730 [0.575-0.884]	78.6	65.7	93.9	31.4	0.006
Radiomics-Only	0.843 [0.744-0.942]	92.9	70.0	98.0	38.2	0.007
IRM	0.915 [0.846-0.984]	78.6	90.0	95.5	61.1	-

Table 3.4: Comparisons of the prediction performance between the proposed IRM model and the pre-existing Briganti, MSKCC, Yale and Roach. The p values for the comparisons of AUCs were calculated using Delong test. The p values for the comparisons of sensitivity, specificity, NPV and PPV were calculated using Chi-squared test or Fisher's exact test.

Methods	**AUC [95%CI]**	**Sensitivity(%)**	**Specificity(%)**	**NPV(%)**	**PPV(%)**
Briganti (0.02 cut-off)	0.711 [0.571-0.852]	85.7	40.0	93.3	22.2
Briganti (0.05 cut-off)	0.711 [0.571-0.852]	35.7	88.6	87.3	38.5
MSKCC (0.02 cut-off)	0.724 [0.571-0.878]	100.0	14.3	100.0	18.9
MSKCC (0.05 cut-off)	0.724 [0.571-0.878]	92.9	27.1	95.0	20.3
Yale (0.02 cut-off)	0.698 [0.528-0.868]	100.0	8.5	100.0	17.9
Yale (0.05 cut-off)	0.698 [0.528-0.868]	85.7	31.4	91.7	20.0
Roach (0.15 cut-off)	0.701 [0.560-0.842]	85.7	41.4	93.5	22.6
IRM	0.915 [0.846-0.984]	78.6	90.0	95.5	61.1

In Figure 3.4(B) and Table 3.4, we show performance results using IRM for detection of LNI with the cut-off points that maximized Youden index in comparison to common pre-existing nomograms including Briganti (Briganti et al., 2012), MSKCC (MSKCC), Yale (Yu et al., 2011) and Roach (Roach et al., 1994), with the guideline suggested cut-off points (risk probability of 0.02, 0.05 for Briganti, MSKCC and Yale and 0.15 for Roach) (Mottet et al., 2021; Schaeffer et al., 2021; Roach et al., 1994). The proposed IRM achieved the highest AUC of 0.915 (%95 CI: 0.846-0.984) ($p<0.05$) with relative improvement of AUC of 28.7%, 26.4%, 31.1%, 30.5% compared with Briganti, MSKCC, Yale and Roach, respectively. As shown in Table 3.4, the IRM achieved 120.2% higher sensitivity than Briganti's with the cut-off point as to the risk level of 0.05 ($p<0.05$), with no significant difference in specificity, NPV and PPV. For the rest nomograms and cutoff points, as shown in Table 3.4 following top-down order, the IRM achieved improvement in specificity of [125.0%, 529.4%, 232.1%, 958.8%, 186.6%, 141.5%], and in PPV of [175.2%, 223.2%, 201.0%, 241.3%, 205.5%, 170.4%] ($p<0.05$), with no significant difference in sensitivity and NPV.

In Table 3.5, we also compared the prediction results using different approaches under each risk group following the National Comprehensive Cancer Network (NCCN)

guidelines (Schaeffer et al., 2021). The NCCN risk groups were divided into Very Low (n=0), Low (n=2), Intermediate (n=25), High (n=47) and Very High (n=10) following the stratification methods given in the NCCN guidelines (Schaeffer et al., 2021). We showed the count of true positive, true negative, false positive and false negative instead of the sensitivity, specificity, NPV and PPV, because the prediction results under each NCCN risk group might result in zeros in some of the measurements and made the calculation of sensitivity, specificity, NPV and PPV invalid. In general, the proposed IRM achieved less false positives and more true negatives compared with the other pre-existing methods, while keeping similar number of true positives and false negatives under each NCCN risk group.

We also conducted a sub-analysis to investigate how the proposed IRM would perform when predicting LNI for patients with PCa without using biopsy-related results. If we excluded all the biopsy-related results from the integrative features, the IRM resulted in an AUC of 0.887 (%95 CI 0.808-0.965), and achieved better prediction AUC compared with results generated by Briganti, MSKCC, Yale and Roach ($p<0.05$).

3.4 Discussion

In this study, we proposed an integrative radiomics model (IRM) to predict LNI for patients with PCa, in order to avoid unnecessary ePLND. The results showed that the integration of clinical features and radiomics features from mpMRI helped improved the prediction of LNI in patients with PCa ($p<0.05$), compared with the approaches using only radiomics features or clinical features. Meanwhile, the IRM outperformed the pre-existing Briganti, MSKCC, Yale and Roach on the prediction of LNI (Yu et al., 2011; Roach et al., 1994; MSKCC; Briganti et al., 2012).

Several proposed nomograms have aimed to predict the risk of having LNI in patients with PCa, tried to avoid unnecessary ePLND using preoperative clinicopathological information (Yu et al., 2011; Roach et al., 1994; MSKCC; Briganti et al., 2012). Briganti, Yale and Roach were proposed to predict LNI using clinical test and biopsy-related results

Table 3.5: Comparisons of the prediction performance between the proposed IRM model and the pre-existing Briganti, MSKCC, Yale and Roach inside each NCCN risk group. LNI+ stands for LNI positive cases, LNI- stands for LNI negative cases. We showed the count of true positive, true negative, false positive and false negative instead of the sensitivity, specificity, NPV and PPV, because the prediction results under each NCCN risk group might result in zeros in some of the measurements and make the calculation of sensitivity, specificity, NPV and PPV invalid.

NCCN risk group (n=84)		**Briganti (0.02 cut-off)**	**Briganti (0.05 cut-off)**	**MSKCC (0.02 cut-off)**	**MSKCC (0.05 cut-off)**	**Yale (0.02 cut-off)**	**Yale (0.05 cut-off)**	**Roach (0.15 cut-off)**	**IRM**
Very Low (LNI +/-: 0/0)	TP	0	0	0	0	0	0	0	0
	TN	0	0	0	0	0	0	0	0
	FP	0	0	0	0	0	0	0	0
	FN	0	0	0	0	0	0	0	0
Low (LNI +/-: 0/2)	TP	0	0	0	0	0	0	0	0
	TN	2	2	2	2	1	2	2	2
	FP	0	0	0	0	1	0	0	0
	FN	0	0	0	0	0	0	0	0
Intermediate (LNI +/-: 2/23)	TP	1	0	2	1	2	1	1	1
	TN	12	23	3	8	3	9	12	22
	FP	11	0	20	15	20	14	11	1
	FN	1	2	0	1	0	1	1	1
High (LNI +/-: 9/38)	TP	8	4	9	9	9	8	8	7
	TN	13	32	5	9	2	10	15	35
	FP	25	6	33	29	36	28	23	3
	FN	1	5	0	0	0	1	1	2
Very High (LNI +/-: 3/7)	TP	3	1	3	3	3	3	3	2
	TN	1	5	0	0	0	1	0	4
	FP	6	2	7	7	7	6	7	3
	FN	0	2	0	0	0	0	0	1

(Yu et al., 2011; Roach et al., 1994; Briganti et al., 2012). MSKCC was also proposed considering additional demographic information (MSKCC). However, these nomograms were limited by the performance with an AUCs of 0.720 to 0.806 (Venclovas et al., 2021; Soeterik et al., 2021). Several studies have reported adding mpMRI imaging features and PI-RADS scores as additional nomogram variables to predict LNI in PCa patients and shown improved AUCs of 0.863-0.879 (Huang et al., 2020a; Hatano et al., 2020). In comparison, the proposed IRM improved the prediction AUC to 0.915, compared with the reported results, in the internal independent testing set.

In this study, following similar methodology to these prior studies (Cuocolo et al., 2021; Gugliandolo et al., 2021; Hectors et al., 2019; Yan et al., 2019; Zhang et al., 2020; Li et al., 2020; Huang et al., 2016; Zheng et al., 2022b), we incorporated first-order, shape, and texture features to build the IRM. Among radiomics features, three were finally included in the integrative features, including one first-order (intensity), and two shape features (shape and size), improved the predictability of LNI in patients with PCa. The inclusion of the shape feature Elongation, defined as the ratio between the largest and second largest principle components axes of the volume, indicated that the shape of the index lesion was correlated with LNI. The inclusion of the shape feature Maximum2DDiameterColumn, defined as the maximum length of the volume in the coronal plane direction, indicated that the size of the index lesion in a specific direction contributed to the prediction of LNI. In addition, the inclusion of first-order feature 10-Percentile, defined as the 10% high intensity value, indicated that the relative high intensity value on T2WI images also could be a useful feature correlated with the appearance of LNI. Our results showed the proposed IRM, integrating radiomics features and clinical features, outperformed the machine learning model using individual feature group only, or using the pre-existing Briganti, MSKCC, Yale and Roach (Yu et al., 2011; Roach et al., 1994; MSKCC; Briganti et al., 2012).

In the sub-analysis, the IRM's performance without using biopsy-related results was shown better than all four existing clinical nomograms (Briganti, MSKCC, Yale and Roach) (Yu et al., 2011; Roach et al., 1994; MSKCC; Briganti et al., 2012). This brought

us an insight that since the diagnosis of aggressiveness of PCa became increasingly accurate via machine learning methods without the need for prostate biopsies (Yan et al., 2019; Zhang et al., 2020; Gnep et al., 2017), the chance of having unnecessary prostate biopsies for patients with PCa could potentially be further reduced in the foreseeable future because the prediction of LNI in patients with PCa could also be accurate without using biopsy results.

Recently, PSMA-PET imaging has also shown great potential to help the assessment of the aggressiveness of PCa and also the prediction of LNI for patients with PCa (Sprute et al., 2021; Cysouw et al., 2021; Zamboglou et al., 2019; Barbosa et al., 2018). There were existing studies building PSMA-PET-based radiomics machine learning model for the prediction LNI for patients with PCa and achieved good AUCs ranged from 0.85 to 0.87 (Cysouw et al., 2021; Zamboglou et al., 2019). In general, a combination of PSMA-PET and mpMRI is not widely available, and thus it might be hard to achieve a fair comparison between the radiomics machine learning prediction models built based on the two image modalities on the same patient population. In the future, when both image modalities are available, a potential further improvement on the performance of prediction of LNI could be achieved by integrating information from both mpMRI-based radiomics features and PSMA-PET-based radiomics features.

The proposed IRM could also potentially be useful in the planning of radiotherapy for patients with PCa. Apart from surgery, the prediction and assessment of LNI is also essential for patients with PCa that will receive definitive radiotherapy. As the radiotherapy could potentially lead to radiation induced injury, unnecessary lymph node radiation should be tried to be avoided (Morris and Haboubi, 2015; De Meerleer et al., 2021). The proposed IRM could be used to predict the possibility of having LNI for the patients, and thus could potentially be helpful to stratify which patients are having lower risk of LNI, and thus assists to avoid the unnecessary lymph node radiation.

Our study has some limitations. One limitation is that the proposed IRM could only be applied to patients with PCa having positive mpMRI (exist lesions with PI-RADS$\geq$3), since radiomics features and some MRI observation features were calculated

and extracted from the index lesions region. We believe the superiority of our proposed IRM could be generalized to a more general population with PCa as long as the required imaging and clinical information are available, but the findings might be limited since the histopathological ground truth of LNI for patients with low suspicious are generally not available, and thus might be hard to test the results. In addition, after reviewing all patients with PCa that have went RP with ePLND from 2010-2019 in our institution (n=679), 244 eligible patients were finally included, with positive 3T mpMRI taken within six months before RP. Although the dataset was imbalanced, and the size of the dataset was relatively small, the characteristics were similar to existing studies conducted in a single institution (Soeterik et al., 2021). Moreover, inter-reader variability might exist in our study. The initial lesion annotations were generated as part of the clinical routine following the guidelines from PI-RADS (Weinreb et al., 2016). During our review and the quality control (QC) process, we found a small fraction of annotations were not sufficient enough, and thus slightly modified the annotations, blinded to the histopathology results of LNI. We used Dice coefficient and volume similarity to measure the possible inter-reader variability (Liechti et al., 2020; Fleiss et al., 1981), and it showed relative consistency with Dice coefficient of 0.89 and volume similarity of 0.95. The model evaluation might also be limited by the fact that the study is conducted single-institutional, and with MR machines from single vendor. Our future works would enlarge the dataset with multi-center collaborations to validate the generalizability onto other institutions' dataset, with different vendors and patient populations, also further solidify our findings with a prospective study design.

In conclusion, the proposed integrative radiomics model (IRM) could be potentially feasible in the prediction of presence of lymph node invasion (LNI) in patients with prostate cancer (PCa). The integration of mpMRI-based radiomics features with the clinical features helped improve the prediction performance, compared with the machine learning model using each individual feature group only. The proposed IRM presented superior performance with an AUC of 0.915, compared with pre-existing nomograms. The proposed model has the potential to be used to predict which patient with PCa may

have LNI and could thus possibly help reduce the number of unnecessary extended pelvic lymph node dissection (ePLND).

CHAPTER 4

Integrative Feature-Enhanced Network Predicting Prostate Biopsy Results

4.1 Introduction

According to the Prostate Imaging Reporting and Data System, version 2.1 (PI-RADS v2.1) (Turkbey et al., 2019), lesions that are classified with an intermediate or high level of suspicion (PI-RADS$\geq$3) typically undergo MRI-targeted biopsy with or without systematic biopsies (Schoots et al., 2015; Sathianathen et al., 2018). However, among patients with negative multi-parametric MRI (mpMRI) (PI-RADS$<$3), comprising 38% of the patient population (Tan et al., 2023), there exists a current lack of consensus regarding the circumstances under which systematic biopsies should be performed, which may lead to patient morbidity such as bleeding, pain, and sepsis. There are strategies proposed for predicting a low risk of clinically significant prostate cancer (csPCa) in patients with negative mpMRI, including the use of prostate-specific antigen density (PSAD) levels below thresholds of either 0.10 ng/ml/ml or 0.15 ng/ml/ml (Hansen et al., 2018; Oishi et al., 2019; Distler et al., 2017; Otti et al., 2019; Pagniez et al., 2020). However, current PSAD-based risk assessments are constrained by a negative predictive value (NPV) ranging from 83.1% to 93.4% (Hansen et al., 2018; Oishi et al., 2019; Distler et al., 2017; Otti et al., 2019; Pagniez et al., 2020)

There exists study proposed to predict the biopsy results for patient with negative mpMRI, via a radiomics-based integrative machine learning approach (Zheng et al., 2022b). The image information was represented by radiomics features, which was known to be effective in computer-aid diagnosis (Yan et al., 2019; Nketiah et al., 2017; Algohary

et al., 2018; Zhang et al., 2020). However, the fixed representation employed in radiomics modeling might constrain the potential for improved prediction performance, which we think could be improved by a more flexible setting, like deep learning approaches.

The Deep Learning (DL) models have demonstrated effectiveness in numerous computer-aid diagnosis tasks (Liu et al., 2021; Penzkofer et al., 2022; Hosseinzadeh et al., 2022; Cao et al., 2019a). However, general deep learning models might not perform well as expected when objects closely blend with the background and lack visual clarity (Ren et al., 2021; Fan et al., 2020), and thus is challenging to extract useful visual features between the cancerous and benign tissue from negative mpMRI images. Hence, there is potential to introduce additional design to enhance the model's predictive capabilities.

Studies have demonstrated the efficacy of texture features in aiding models to recognize objects when those objects blend with background signals (Ren et al., 2021). Moreover, studies shown that there exist statistical differences of image texture features of benign tissue within prostate region between the cancer-naïve patients and patients having PCa (Conlin et al., 2023; Nonn et al., 2009). Therefore, in addition to employing a DL backbone for predicting biopsy results, we also introduced the feature-enhanced (FE) block to enhance the texture feature representation, amplify the feature difference between the cancerous tissue and benign surroundings, and ultimately improve the model's predictive performance (Ren et al., 2021).

In this study, we aim to build and validate a feature-enhanced DL-based model for predicting prostate biopsy results in patients with negative mpMRI. Specifically, an integrative feature-enhanced network (IFN) was proposed to predict the biopsy results, combining image features, extracted using a DL network with the FE blocks, and routinely-collected clinical information. The effectiveness of the IFN integrating both categories of features was validated in comparisons with the DL-based approaches that including only clinical information or imaging features. Furthermore, the performance of the IFN was also compared with existing PSAD-based and radiomics-based strategy to predict the prostate biopsy results for patients with negative mpMRI.

4.2 Material and Methods

4.2.1 Study Population and MRI Data

The single institutional retrospective study was conducted in compliance with the 1996 Health Insurance Portability and Accountability Act (HIPAA), and approved by the institutional review board (IRB) with a waiver of the requirement for informed consent. The patient cohort consisted of cases with negative prostate 3 Tesla (3T) mpMRI results (PI-RADS<3), identified by reviewing clinical prostate MRI scans conducted under the standard protocol at a single academic institution between January 2016 and December 2020. The 3T mpMRI scans were performed on one of the three 3T scanners with standardized protocols (Trio, Skyra, and Verio (Siemens Healthcare)). All prostate mpMRI scans were conducted using a standardized imaging protocol in accordance with the European Society of Urogenital Radiology (ESUR) PI-RADS guidelines. The scans were interpreted following the PI-RADS v2.1 guideline by a panel of three genitourinary radiologists, each with 10+ years of diagnostic experience in prostate mpMRI.

In all, 508 patients with negative prostate mpMRI were included in the study. The patients included in the study cohort met the following inclusion criteria: (1) Clinical suspicion of PCa, indicated by an elevated Prostate Specific Antigen (PSA) level and/or abnormal digital rectum exam results; (2) 3T-mpMRI with all identified lesions scored as PIRADS 1 or 2 (MR negative); (3) Completion of a standardized systematic transrectal ultrasound-guided (TRUS) biopsy with or without magnetic resonance ultrasound fusion (MRUS), conducted within six months after the 3T mpMRI (Sonn et al., 2013); (4) PSA levels were measured within the six months preceding the biopsy. All cases were verified and re-reviewed by an abdominal radiologist with 6 years of diagnostic experience on prostate mpMRI and no discordant was observed. Patients were excluded if they were undergoing active surveillance, or were receiving PCa treatment (including surgery, focal therapy, radiation, or hormonal therapy). For patients had undergone multiple mpMRI scans, the scan that was conducted immediately prior to their first negative TRUS/MRUS biopsy was included. The Figure 4.1 presents a detailed flowchart outlining the patient

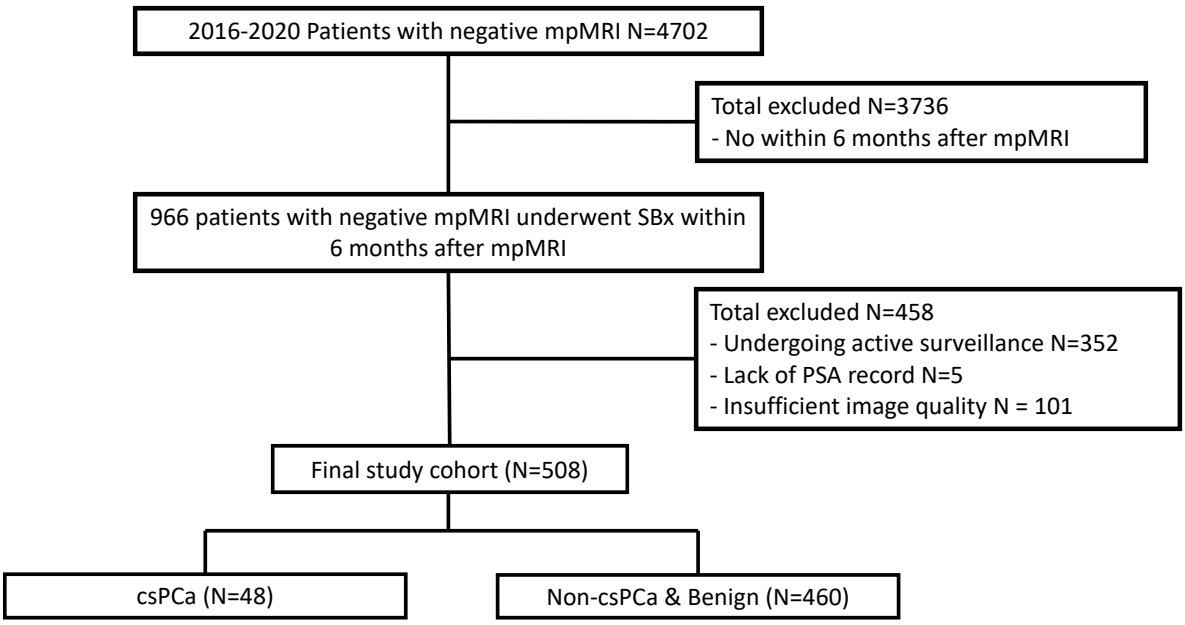


Figure 4.1: The overall pipeline of patient inclusion process. The study cohort consists of 508 patients with negative prostate mpMRI. 48 patients have positive prostate biopsies and 460 patients have negative prostate biopsies.

inclusion criteria for the study.

We defined negative prostate biopsy as excluding csPCa (Gleason Score (GS)<7) findings in all biopsy cores (Matoso and Epstein, 2019). We included clinical information that was reported to be positively co-related with the appearance of csPCa: family history of having csPCa, prostate biopsy history, prostate volume, patient age, and PSAD. All TRUS and MRUS biopsy cores were fixed in formalin and stained with hematoxylin and eosin (H&E) for histological analysis. All the evaluations were conducted by specialized genitourinary pathologists as a standard part of clinical histopathological assessment.

4.2.2 Integrative Feature-Enhanced Network

We defined negative prostate biopsy as excluding csPCa (Gleason Score (GS)<7) findings in all biopsy cores (Matoso and Epstein, 2019). We included clinical information that was reported to be positively co-related with the appearance of csPCa: family history of having csPCa, prostate biopsy history, prostate volume, patient age, and PSAD. All TRUS and MRUS biopsy cores were fixed in formalin and stained with hematoxylin and

eosin (H&E) for histological analysis. All the evaluations were conducted by specialized genitourinary pathologists as a standard part of clinical histopathological assessment.

As the zonal information shown to be important in both the clinical guideline and existing studies related to DL-based PCa diagnosis, we also integrated prostate zonal segmentation masks as for providing zonal information to help model predict the biopsy results given mpMRI images (Turkbey et al., 2019; Hosseinzadeh et al., 2022; Weinreb et al., 2016; De Vente et al., 2020). The prostate zonal segmentation mask of the transition zone (TZ) and the peripheral zone (PZ) were treated as part of the IFN's input, as shown in Figure 4.2. The zonal masks were generated using a separate automatic prostate zonal segmentation model, CAT-Net, to explicitly provide the PCa-related anatomical information (Hung et al., 2022). All segmentations were reviewed and validated by an abdominal radiologist and a radiological trainee (X.X. and Y.Y., eight and six years of experience in prostate mpMRI interpretation, respectively).

The proposed IFN was designed to integrate both clinical information and image features for predicting biopsy results. The image feature extraction module was built based on a 3D ResNet backbone (He et al., 2016). Compared with the original implementation of ResNet, we attached the proposed FE blocks after convolution layers within a residual block, which will be introduced in details in the next sub-section. The final extracted image feature vectors were first concatenated with the clinical feature vectors, which composed by age, prostate volume, PSAD, family history and biopsy history, and then went through the fully-connected (FC) layers for the final prediction of the prostate biopsy results.

4.2.3 Feature-Enhanced Block

The FE block was introduced to enhance the intermediate feature representation and amplify the texture feature difference between the cancer tissue and the benign surroundings (Ren et al., 2021). Details are shown in Figure 4.2. Assume $F_0 \in R^{C\times D\times H\times W}$ is the feature map generated from previous residual block, in which C, D, H and W

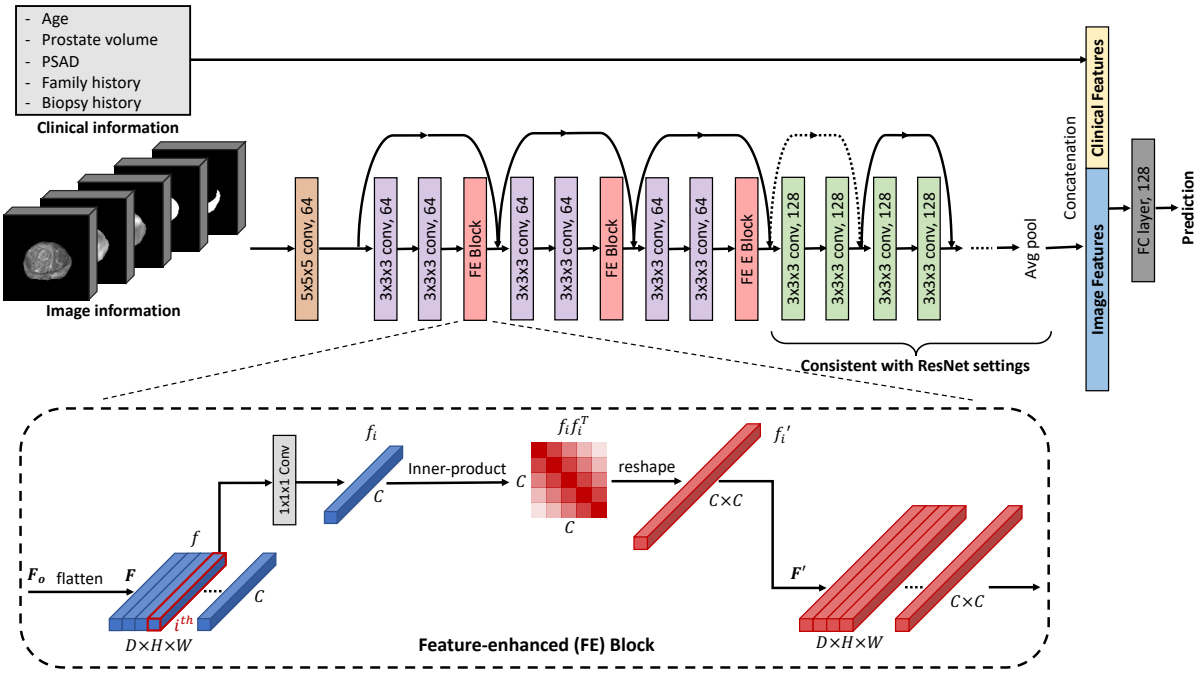


Figure 4.2: Visualization of the whole pipeline of the proposed integrative feature-enhanced network (IFN) for predicting prostate biopsy results. Clinical features consist of routinely collected clinical information: age, prostate volume, PSAD, family history and biopsy history. Image information consists of five components: 3D stacks of T2WI, ADC and high-B DWI images after being cropped based on ROI (prostate region), and also the masks of TZ and PZ of the prostate. Image features were extracted through the 3D ResNet with proposed feature-enhanced (FE) blocks added, and then concatenated with the clinical features. The integrated feature vectors were then fed into the fully-connected layer to predict the biopsy results.

represent for number of channels, depth, height and width of the feature map. Then, the feature map F_0 will first be flattened to $F \in R^{(D\times H\times W)\times C}$, and then, an inner-product operation is performed with respect to the channel dimension C. The operations are actually measuring the co-occurrence matrices of the features along the channel dimension at each spatial location, which were generally used to represent texture features (Ren et al., 2021; Haralick et al., 1973). For the i^{th} feature vector $f_i \in R^{C\times 1}$ on the feature map F, inner-product will be calculated as $f_i' = f_i f_i^T \in R^{(C\times C)\times 1}$, and we execute this inner-product operation for all $i = 1, 2, ...(D \times H \times W)$. The generated new feature map F' will be reshaped from $R^{(D\times H\times W)\times(C\times C)}$ to be $R^{(C\times C)\times D\times H\times W}$, and then will be treated as the new input feature map to the next residual block of ResNet for deeper feature extraction. The FE blocks were added into the three residual blocks at the first stage

of the ResNet, and the rest of architectures will remain the same setting as the original ResNet, since studies have shown that the texture features appeared to be learnt at the initial stage of a deep convolutional neural network (Zeiler and Fergus, 2014; Springenberg et al., 2014).

4.2.4 Statistical Analysis

We first investigated the proposed IFN approach compared with the approaches only image features or clinical information were available. Then, we also perform comparisons between the IFN model including or excluding the FE blocks. Finally, we showed the results of the comparisons between the proposed IFN's and other existing methods with same objectives: two conventional PSAD-based strategies: 1) PSAD<0.15 ng/ml/ml as with low risks of having csPCa and 2) PSAD<0.10 ng/ml/ml as with low risks of having csPCa, using in the clinical practice, and one existing radiomics-based approach with the same study objective (Zheng et al., 2022b).

The model performances were evaluated by receiver operating characteristic (ROC) analysis. For ROC analysis, we generate the mean and corresponding standard deviation after 1000 times bootstrapping using the results validated by 5-fold cross-validation (CV) (Cao et al., 2019a; Aamir et al., 2022). With known mean and variance, we were capable to perform a more comprehensive evaluation to minimize the potential influence caused by data imbalance during model evaluation. Sensitivity, specificity, and NPV were calculated from the cut-off points on the mean curves by maximizing the Youden's index value (sensitivity+specificity-1) (Ruopp et al., 2008), and were then compared via Chi-squared test. The area under the curve (AUC) with 95% confidence interval (CI) calculated based on ROC curves were used to evaluate and compare different models' performance, and compared via the Delong test (DeLong et al., 1988).

N4 bias field correction was applied to T2WI images to compensate for low-frequency intensity variations (Tustison et al., 2010). The intensity value of voxels in high-B DWI and T2WI images are linearly normalized to have a value in the range of [0, 1]. Since ADC

Table 4.1: Clinical information of the patients included in the study.

Clinical Information	**Overall**	**Biopsy Positive**	**Biopsy Negative**
No. of patients, n	508	48	460
Age (yr), median IQR	63 {58-67}	66 {60-71}	62 {57-67}
PSAD (ng/ml/ml), median IQR	0.12 {0.08-0.17}	0.13 {0.09-0.20}	0.11 {0.08-0.16}
Prostate Volume (cc), median IQR	53.0 {38.6-72.0}	47.5 {34.8-62.3}	54.5 {39.0-73.0}
Biopsy History, n (%)			
Present	133 {26.2}	8 {16.7}	125 {27.2}
Absent	375 {73.8}	40 {83.3}	335 {72.8}
Family History, n (%)			
Present	115 {22.6}	19 {39.6}	96 {20.9}
Absent	393 {77.4}	29 {60.4}	364 {79.1}

maps are quantitative and voxel intensities are consistent across patients, the intensity on ADC maps was first clipped by a patient-independent value and then normalized to be in the range of [0, 1] (Turkbey et al., 2019; Cao et al., 2019a; Vargas et al., 2011). The mpMRI images, composed of 3D stack of T2WI, ADC, and high-B DWI images, were first multiplied by the prostate masks, which constructed by combining the TZ and PZ segmentations, to get only prostate region as the region of interest (ROI). Then, the three 3D mpMRI image stacks were concatenated with the 3D masks of TZ and PZ. The final 3D stacks were 128x128 center-cropped on the HW dimension for efficient training purposes, and were treated as final inputs to the network. All the models were trained using Focal Loss with Adam optimizer and learning rate of 0.0001, on a single Nvidia RTX 3090 GPU.

4.3 Results

The patient clinical information, included age, PSAD, prostate volume, biopsy history and family history are summarized in Table 4.1. In total, the study cohort included 460 patients with negative prostate biopsy and 48 patients with positive prostate biopsy.

Figure 3 shows the ROC comparisons among the proposed IFN and the model using only clinical information, only imaging information and the case using IFN with both features but excluding FB blocks. Compared with the image-only and clinical-only

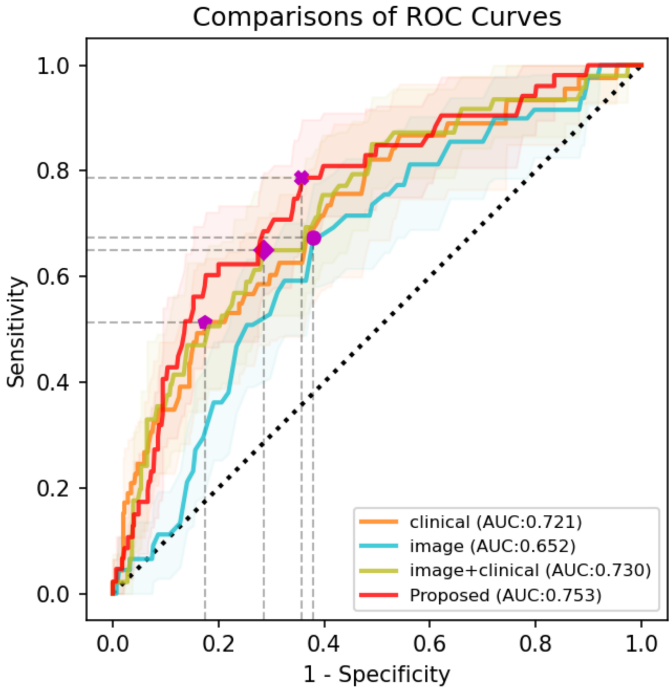


Figure 4.3: Comparisons between the proposed IFN, and approaches using only clinical information, only image features and the case using IFN with both image and clinical feature but without FB blocks. The solid curves with orange, cyan, gold and red color represent for the clinical-only, image-only, the IFN with no FB blocks and the IFN, respectively. Horizontal and vertical gray dash lines of each optimal cutoff point aimed to visualize sensitivity and "1-specificity" value on each ROC curve.

implementations, the DL network using both information achieves better AUC, shown in gold curve. One step further, by including additional FE blocks, the proposed IFN achieved the highest AUC ($p<0.05$), compared with all other approaches as shown in red curve, with relative improvements of 4.4%, 15.5% and 3.2%, respectively. The AUC, sensitivity, specificity and NPV that calculated from the cutoff point of the IFN were 0.753, 0.787, 0.642 and 0.967, as shown in Table 4.2(a). Table 4.2(b) showed that compared with other approaches, the IFN not only achieves the highest number of true positives, but also lowest number of false negatives. F1 score, which is generally used to evaluate model performance on dataset with imbalanced nature, like the patient cohort included in this study, was also calculated. The IFN achieves highest F1 score of 0.374 in the optimal situation, which shows improvements compared with other approaches with F1 scores of [0.327, 0.256, 0.338].

Table 4.2: Comparisons of prediction performances with different model components. 1) clinical-only, 2) image-only, 3) IFN without FE blocks, and 4) the proposed IFN. (a) showed the comparisons based on NPV, sensitivity, specificity and AUC. (b) showed comparisons based on number of the predicted True Positives, False Positives, True Negatives and False Negatives.

	NPV	**Sensitivity**	**Specificity**	**AUC**
Clinical	0.942	0.514	**0.825**	0.721
Image	0.948	0.674	0.620	0.652
Image+Clinical	0.951	0.650	0.715	0.730
Proposed	**0.967**	**0.787**	0.642	**0.753**

(a)

	True Positive	**False Positive**	**True Negative**	**False Negative**
Clinical	25	80	380	23
Image	32	175	285	16
Image+Clinical	31	131	329	17
Proposed	38	165	295	10

(b)

The comparison results between the proposed IFN, two conventional PSAD-based risk assessment methods, and a radiomics-based method from an existing study with same objective were shown in Table 4.3. The two PSAD-based approaches utilized the PSAD as the indicator, and treated patients with PSAD<0.10 ng/ml/ml or PSAD<0.15 ng/ml/ml as with low risks of having csPCa, respectively. The pre-existing study predicted the biopsy results for patients with negative mpMRI utilizing a machine learning model with integrated clinical information and radiomics features (Zheng et al., 2022b). Compare with all other approaches, the proposed IFN achieves the best NPV of 0.967 and sensitivity of 0.787, while other approaches have NPV of [0.923, 0.922, 0.953] and sensitivity of [0.688, 0.438, 0.665]. Compared with the pre-existing study (Zheng et al., 2022b), the IFN also achieves better AUC ($p<0.05$), with a relative improvement of 3.4%.

In Figure 4.4, we have included two representative examples to illustrate the effectiveness of the proposed IFN. In both cases, discerning clear visual cues for the diagnosis of csPCa on mpMRI images proved challenging, and the clinical information exhibited similar patterns. However, the two patients have different prostate biopsy outcomes. Imprecise diagnosis might result from the application of the criteria of using PSAD<0.15

Table 4.3: Comparisons of the prediction performance among different studies: 1) predict patients with PSAD<0.10 as low risk of having csPCa, 2) predict patients with PSAD<0.15 as low risk of having csPCa, 3) existing work utilizing radiomics-based machine learning method to predict the biopsy results for patients with negative prostate mpMRI from Zheng et al. (2022b) 4) the proposed IFN. (a) showed the comparisons based on NPV, sensitivity, specificity and AUC. (b) showed comparisons based on number of the predicted True Positives, False Positives, True Negatives and False Negatives.

	NPV	**Sensitivity**	**Specificity**	**AUC**
PSAD<0.10 ng/ml/ml	0.923	0.588	**0.393**	-
PSAD<0.15 ng/ml/ml	0.922	0.438	0.698	-
Radiomics-based model	0.953	0.665	**0.707**	0.728
Proposed	**0.967**	**0.787**	0.642	**0.753**

(a)

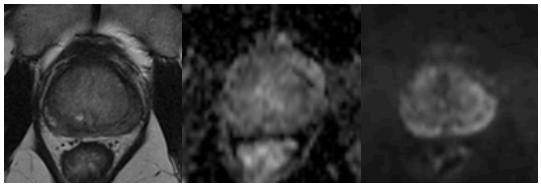
	True Positive	**False Positive**	**True Negative**	**False Negative**
PSAD<0.10 ng/ml/ml	33	279	181	15
PSAD<0.15 ng/ml/ml	21	139	321	27
Radiomics-based model	32	135	325	16
Proposed	38	165	295	10

(b)

ng/ml/ml as with low risks of having csPCa. In contrast, the IFN demonstrates its robust capability by accurately predicting the biopsy results for both example cases integrating both the imaging and clinical information.

4.4 Discussion

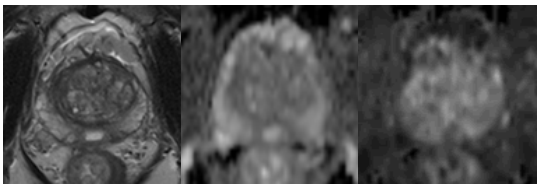
In this study, we proposed an integrative feature-enhanced network (IFN) designed to predict biopsy results for patients with negative mpMRI, with the aim of potentially avoiding invasive procedures when a patient's mpMRI shows negative findings. The results demonstrated that the integration of clinical information and mpMRI image features helped improved the performance of the DL model in predicting prostate biopsy results ($p<0.05$), when compared to the approaches utilizing only image or clinical information. In addition, the IFN's prediction performance was further improved with the inclusion of the design of feature enhance (FE) block, with enhanced intermediate feature representations.



Age	74
PSAD	0.12
Prostate Volume	49
Family History	No
Biopsy History	Yes

T2WI ADC DWI

(a) Patient without Biopsy-confirmed csPCa



Age	74
PSAD	0.15
Prostate Volume	44
Family History	No
Biopsy History	Yes

T2WI ADC DWI

(b) Patient with Biopsy-confirmed csPCa

Figure 4.4: mpMRI images and clinical information and of two representative examples of patients with negative mpMRI. Patient (a) has biopsy-confirmed csPCa, and patient (b) does not have biopsy-confirmed csPCa. The proposed IFN correctly predicted the biopsy results for both patients. Despite no clear visual signal of cancerous tissue, and two patients owned similar clinical information, the IFN was capable of distinguishing and making the correct prediction of the biopsy results for the two patients.

As reported in recent review studies, the strategy of utilizing PIRADS<3 as a decision criterion to avoid prostate biopsy led to a limited NPV ranging from 80.% to 92.3% (Tan et al., 2023; Pagniez et al., 2020; Moldovan et al., 2017), and the risk assessment methods involving PSAD levels improved the NPV to be in the range between 83.1% to 93.4% for predicting negative biopsy results among patients with negative mpMRI (Hansen et al., 2018; Oishi et al., 2019; Distler et al., 2017; Otti et al., 2019; Pagniez et al., 2020). In addition, experiments also show that the radiomics-based machine learning approach proposed by existing study reached NPV of 95.3%. In this study, the implementation of the IFN resulted in an improved NPV of 96.7%, which surpassed the PSAD-based strategies and the radiomics approach previously employed. In addition, the IFN also achieve better sensitivities compared with other approaches in our experiments.

Recent DL-based studies have been introduced to address various clinical tasks, like PCa detection and aggressiveness assessment using mpMRI (Liu et al., 2021; Penzkofer et al., 2022; Hosseinzadeh et al., 2022; Cao et al., 2019a). While deep learning models excel at extracting imaging features for decision-making when clear visual signals from potential

PCa lesions are evident on mpMRI images, they might not perform as expected in cases where the visual appearances of malignant tissue blended with benign tissue and is nearly imperceptible on mpMRI. In such instances, the effectiveness of general deep learning models may be limited without additional design considerations (Ren et al., 2021; Fan et al., 2020). To the best of our knowledge, we are not only the first study to investigate the feasibility of using DL model to predict biopsy results for patients with negative mpMRI, but also the first to introduce an additional design element, the FE block, into the DL network architecture, to enhance the intermediate feature representations, and thus improving the performance of the DL model. Our results conclusively demonstrated the feasibility of DL-based method, and also showed that the incorporation of the FE blocks design resulted in improved prediction outcomes when compared to the performance of the IFN lacking this design.

Our study is subject to several limitations. The initial screening involved an exhaustive review of all 3T prostate mpMRI scans (N=4,702) conducted over a five-year period at our institution, resulting in the inclusion of 508 patients with negative mpMRI who met the specified criteria. While this dataset is comprehensive, its size is relatively small. To address this limitation and enhance the robustness of our findings, we are actively seeking multi-institutional collaboration to expand the cohort size for future investigations. In addition, there exists data distribution imbalance between the patients with positive and negative biopsy results. However, the imbalance reflects the inherent nature of the study objectives, and is consistent with data distribution characteristics observed in existing studies (Hansen et al., 2018; Oishi et al., 2019; Distler et al., 2017; Otti et al., 2019). We conducted cross-validation to comprehensively evaluate model performance, and thus minimized the potential influence could be brought by the dataset size and the data imbalance, achieved by showing the reported mean and variance of measurements for different models.

In conclusion, the prostate biopsy results of patients with negative mpMRI can be effectively predicted using the proposed integrative feature-enhanced network (IFN). The combination of image features from mpMRI and routine-collected clinical information

helped improve the model's performance on predicting prostate biopsy results. In addition, the proposed feature-enhanced (FE) block help further improve the DL model's performance by enhancing the intermediate feature representations. Compared with the PSAD-based methods and the existing model with same objective, the proposed DL-based method, the IFN, achieved relatively superior performance with an NPV of 96.7%. These findings underscore the capabilities of DL-based models on predicting prostate biopsy results for patients with negative mpMRI. The IFN exhibits the potential to enable the identification of which patients with negative mpMRI could safely avoid the prostate biopsy, thereby contributing to reduce unnecessary biopsies.

CHAPTER 5

AtPCa-Net: Anatomical-Aware Prostate Cancer Detection Network on Multi-Parametric MRI

5.1 Introduction

Prostate cancer (PCa) is the second leading cancer-related cause of death and the most common cancer among men in the United States (Rawla, 2019). Multi-parametric MRI (mpMRI) is the preferred non-invasive imaging tool for PCa diagnosis before biopsies (Weinreb et al., 2016; Turkbey et al., 2019). According to the Prostate Imaging Reporting and Data System, version 2.1 (PI-RADS) (Weinreb et al., 2016; Turkbey et al., 2019), a combination of mpMRI findings is used for predicting the probability of clinically significant PCa, where the different combinations are used depending on the lesion location, in either transition zone (TZ) or peripheral zone (PZ) of the prostate. For example, following the PI-RADS, T2-weighted imaging (T2WI) is the primary imaging component for lesions in TZ with an additional assessment by diffusion-weighted imaging (DWI), while in the PZ, DWI/apparent diffusion coefficient (ADC) is the essential imaging component with an addition of dynamic contrast-enhanced (DCE) MRI (Weinreb et al., 2016; Turkbey et al., 2019). Therefore, given the significance of varying imaging appearances of PCa on mpMRI between TZ and PZ, there is considerable potential to improve PCa detection models further when this anatomical prior is thoughtfully incorporated.

With advances in deep learning, many studies proposed deep learning models for the detection of PCa using mpMRI. However, the different appearances of PCa lesions in TZ and PZ on different mpMRI components were generally not fully integrated into the model design. (Seetharaman et al., 2021; Hosseinzadeh et al., 2022; Vente et al., 2020; Yu

et al., 2020; Cao et al., 2019b, 2021; Saha et al., 2021). Overlooking this zonal-related anatomical prior, but treating all lesions equally regardless of the locations, could lead to potential suboptimal model performance (Weinreb et al., 2016; Turkbey et al., 2019). A design that can reflect both the zonal appearance differences and the commonality of them being PCa lesions is the key to improving the model's performance.

Hierarchical label and loss design embed structural information hierarchically among different classes into the loss function to better guide the model training (Verma et al., 2012; Bertinetto et al., 2020; Li et al., 2022). The design transforms the binary labeling to a more structured label space and is able to account for the distinct inter-class property differences while preserving shared properties among different classes (Verma et al., 2012; Bertinetto et al., 2020; Li et al., 2022). In this study, we propose an anatomical-aware hierarchical loss design, the Zonal Loss (ZL). The ZL can direct the model to learn both the unique and shared characteristics of PCa lesions across different prostate zones in accordance with clinical practice, thus enhancing the model's detection capabilities.

Furthermore, studies have shown that PCa, benign prostatic hyperplasia (BPH), and the central zone (CZ) of the prostate can occasionally present with visual similarities (Yu et al., 2014; Panebianco et al., 2015). This undesirable resemblance between PCa and other prostate tissues complicates the diagnosis process. In clinical practice, symmetric-related information as a reference is valuable for distinguishing BPH and CZ from PCa. Research indicates that BPH and CZ tend to be visually symmetric (Yu et al., 2014; Panebianco et al., 2015), while PCa is generally presented asymmetrically (Smith et al., 2003; Barentsz et al., 2012). Illustrations of PCa lesions and PCa-like visual patterns are shown in Figure 5.1. We can observe that BPH and CZ are shown to be similar to PCa lesions—low intensity on both ADC and T2WI images and high intensity on high-B DWI images.

The visual similarity of non-PCa prostate tissues not only complicates the diagnostic process but also leads to performance degradation in PCa detection models due to the generation of undesired false positive (FP) predictions—a common issue in existing deep-learning-based PCa detection models (Seetharaman et al., 2021; Hosseinzadeh et al., 2022;

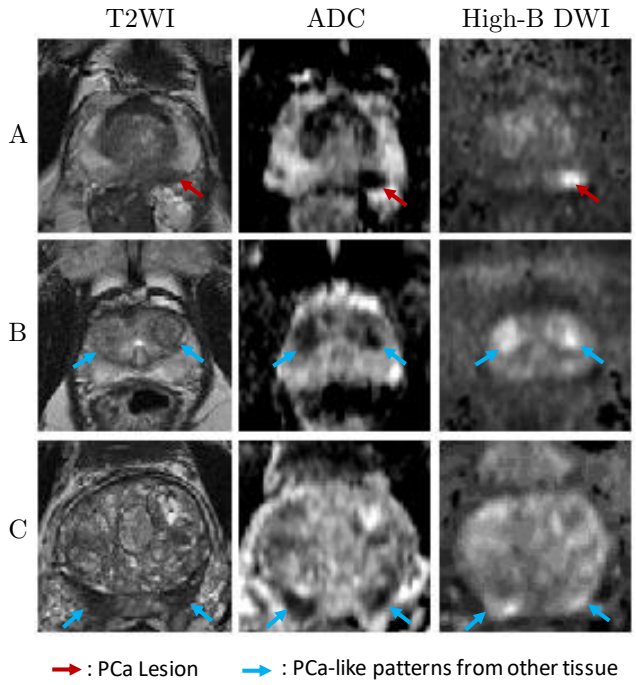


Figure 5.1: Examples of a PCa lesion and PCa-like patterns from other prostate tissues, like the BPH and the CZ, that can cause FP predictions from three mpMRI scans (A, B, and C). From the left to the right, the first column shows the T2WI images, the second column shows the ADC images, and the third column shows the high-B DWI images. Red arrows point to a PCa lesion, and blue arrows point to the PCa-like patterns in BPH (B) and CZ (C).

Yu et al., 2020; Cao et al., 2019b, 2021; Vente et al., 2020; Saha et al., 2021). By taking symmetric-related patterns into consideration, FP predictions may be reduced as PCa lesions can be further distinguished from BPH and CZ by their asymmetrical appearance differences. Therefore, integrating symmetry-related anatomical priors into the design of PCa detection models may be crucial in reducing potential FP predictions.

Existing study has shown that the human visual system recognizes symmetric patterns by comparing visual differences between the original image and the mirrored image after reflecting with respect to an imagined center-axis (Wagemans, 1997). Inspired by how symmetric patterns stimulate human visual perception, we propose a symmetric-aware network architecture that utilizes both original and mirrored mpMRI images for PCa detection. Simulating how the human vision system reacts to the symmetric patterns, the network can help distinguish the PCa lesions from other prostate tissue with visual similarities, like BPH and CZ, and thus reduce FP predictions (Yu et al., 2014; Panebianco

et al., 2015; Smith et al., 2003; Barentsz et al., 2012).

In this study, we introduced PCa-related anatomical priors into the deep learning framework design and developed an anatomical-aware deep learning network for PCa detection on mpMRI. The proposed network leverages symmetry-related information and PCa zonal appearance differences on mpMRI images to form a 3D anatomically-aware PCa detection network (AtPCa-Net), enhancing the accuracy of PCa lesion detection. Our main contributions include the following:

1) We exhibit that the introduction of the PCa-related anatomical priors into the DL network architecture design helps improve model performance. The extensive experiments demonstrate that either one of the anatomical-aware designs of the proposed AtPCa-Net can help improve the PCa detection and patient-level classification performance, and the integration of both designs can achieve the best model performance on both PCa detection and patient-level classification tasks.

2) We incorporate the symmetric-related clinical priors into the network architecture design to suppress potential FP predictions. By utilizing symmetric-related visual appearance differences, the design could help distinguish PCa from other prostate tissues shown similar visual patterns on mpMRI, thereby reducing possible FP predictions. To the best of our knowledge, our study is the first study to achieve FP reductions on PCa detection using anatomical-related clinical prior.

3) We integrate the zonal appearance differences of PCa on mpMRI explicitly into the loss design by proposing the Zonal Loss (ZL). Compared with existing models overlooking this property and treating all lesions equally regardless of their location, the ZL treats PCa in different prostate zones differently following the clinical guideline, and thus helps improve model performance.

5.2 Related Work

5.2.1 Prostate Cancer Detection

The deep-learning models for PCa detection and classification based on mpMRI have been widely investigated (Booven et al., 2021; Litjens et al., 2014; Bhattacharya et al., 2022; Sufyan et al., 2023). The models are generally built by convolutional neural networks (CNNs) for their outstanding performance on classification, segmentation, and detection tasks. Recent studies exhibit the feasibility of using CNNs for PCa detection using mpMRI (Cao et al., 2019b; Yu et al., 2020; Vente et al., 2020; Seetharaman et al., 2021; Cao et al., 2021; Hosseinzadeh et al., 2022; Li et al., 2023; Saha et al., 2021). Li et al. (2023) designed a multi-scale two-branch dilated-convolution-based deep learning network to segment both PCa lesions and prostate from mpMRI. Seetharaman et al. (2021) designed a PCa detection network to identify indolent and aggressive PCa separately with the help of two different encoder branches for T2WI and ADC images and the fusion of feature maps from multiple levels of the encoder branches. Cao et al. (2019b) introduced the idea of ordinal encoding for PCa with different severity doing a multi-class classification. The author also designed the mutual finding loss mimicking the process of how radiologists interpret the T2WI and ADC images to detect PCa. Cao et al. (2021) modified the FocalNet (Cao et al., 2019b) to have a stack of adjacent slides as input and did a comprehensive evaluation of the PCa detection performance between the radiologists and their proposed model.

There are some existing studies that tried to integrate clinical priors related to different diagnostic focuses for PCa in different zones into the PCa detection network design (Turkbey et al., 2019; Vente et al., 2020). Hosseinzadeh et al. (2022) utilized this zonal-related anatomical prior by stacking the prostate zonal segmentation masks together with mpMRI images as part of the input to the model and showed the detection performance improved. Vente et al. (2020) discovered the idea of both using ordinal encoding for PCa with different aggressiveness and also feeding prostate zonal masks into the PCa detection networks to let the model learn the anatomical relationship between

the prostate zones and PCa appearance. Duran et al. (2022) discussed the performance differences between using a prostate mask or using a PZ mask as part of the segmentation model input and observed the latter approach got better lesion segmentation performance.

Although they provide zonal information from the input, the cross-entropy (CE) loss with binary labeling they used explicitly treated all lesions identically regardless of the location (Weinreb et al., 2016; Turkbey et al., 2019). The ignorance of the lesion appearance differences in different prostate zones might lead to suboptimal PCa detection performance, which could be further improved.

As PCa detection models generally suffer from undesired FP predictions, some studies have introduced network designs aiming for better FP reduction ability (Yu et al., 2020; Saha et al., 2021). Yu et al. (2020) proposed a multi-scale patch-wise network together with a squeeze-and-excitation (SE) block (Hu et al., 2018). The design tried to reduce FP predictions by letting the model learn the FP patterns from the context information provided by the multi-scale patches automatically. To suppress FP predictions, Saha et al. (2021) first introduced an auxiliary network to classify if a given image patch contains PCa lesions or not and then multiplied the classification results with the detection probability map to conduct the final output. The experiments showed that the FP predictions could be suppressed by the patch-wise classification results. However, neither study has considered achieving FP reduction using anatomical-related clinical prior, which could also be capable of helping effectively reduce FP predictions.

5.2.2 Anatomical-Aware Design for Other Diseases

There are existing studies investigating how to incorporate anatomical-related clinical priors into the network architecture design for various tasks related to other diseases (Sun and Ji, 2021; Kamal et al., 2022; Ma et al., 2021). Sun and Ji (2021) introduced a weakly-supervised knowledge distillation model for breast mass segmentation, with auxiliary networks for reconstruction and aggressiveness classification. The anatomy property was designed to be learned from the encoder of the teacher model, an autoencoder network

reconstructing the input image, and then transferred to the student model, the desired breast mass segmentation network. Kamal et al. (2022) proposed a semi-supervised CNN for thoracic disease classification on chest Chest X-ray images. The anatomical information was brought by the prediction masks of lung and heart generated from the auxiliary segmentation network and then fed into the main classification network as an anatomy-informed reference for an attention module. Ma et al. (2021) proposed a dual-branch cascaded CNN for the segmentation of retinal layers and fluid from optical coherence tomography (OCT) images. The model first calculated a relative positional map based on the retinal layer boundaries and then fed them into the final segmentation network to inform the model of the anatomical relationships among different retinal layers. All the studies showed improvements in model performance when including anatomical-aware network architecture design. In this study, the anatomical-aware designs are not only composed of a symmetric-aware network architecture for FP prediction reduction but also shown through the design of the hierarchical loss, the ZL, considering the diagnostic differences of lesions on different prostate zones following clinical guideline (Weinreb et al., 2016; Turkbey et al., 2019).

5.3 Methods

5.3.1 Overview

We propose a 3D anatomical-aware PCa detection network (AtPCa-Net) to detect whole-mount histopathology (WMHP) confirmed clinically significant PCa (csPCa) utilizing the PCa-related anatomical priors. The proposed AtPCa-Net consists of two parts. First, a 3D symmetric-aware network takes the symmetric-related information into consideration to suppress FP predictions. Second, the ZL structurally integrates the PCa-related zonal differences into the label and loss design. The overall architecture of AtPCa-Net is illustrated in Figure 5.2. We adhered to the structure of nnU-Net as the backbone for AtPCa-Net because of its good performance on detection and segmentation of medical imaging tasks (Isensee et al., 2021).

5.3.2 Study Population and mpMRI Images

This retrospective study was carried out in compliance with the United States Health Insurance Portability and Accountability Act (HIPAA) of 1996 with approval from the institutional review board (IRB) of our institution with a waiver of the requirement for informed consent. All experiments conducted in this study adhered strictly to the relevant guidelines and regulations. The whole dataset consists of 652 patients. It is composed of two parts: 1) pre-operative mpMRI images from patients (N=220) who had confirmed PCa lesions (N=246) with whole-mount histopathology after radical prostatectomy (RP), and 2) mpMRI images from patients (N=432) who did not have indications of PCa lesions, confirmed by systematic biopsies followed by negative mpMRI (PI-RADS$\leq$2). We included mpMRI images with no indications of PCa lesions to balance the data distribution on model training and testing, as well as to perform patient-level classification evaluation. We used 5-fold cross-validation to validate and evaluate the model performance, in which each fold contains 130/131 patients assigned randomly from the entire dataset.

All mpMRI images are performed on Siemens 3T scanners with the standardized clinical prostate mpMRI protocol (Weinreb et al., 2016; Turkbey et al., 2019), including T2WI and DWI. We exclude the DCE-MRI images given the limited role of DCE-MRI (Kozlowski et al., 2018; Brancato et al., 2020; Weinreb et al., 2016; Turkbey et al., 2019). For T2WI, the repetition time (TR) and echo time (TE) are 3000-5900 ms and 101-109 ms, the field of view (FOV) of 20cm$\times$20cm with an in-plane resolution of 0.625 mm $\times$ 0.625 mm and through-plane resolution of 3 mm. For DWI, we use TR and TE of 4800-5300 ms and 60-81 ms, FOV of 26cm$\times$21cm with in-plane resolution of 1.625 mm $\times$ 1.625 mm and through-plane resolution of 3.6 mm. The ADC maps were calculated using linear least squares curve fitting of voxels in the four DWIs against their corresponding b values (0/100/400/800 s/mm2). We also denote the DWI images with b=1400 s/mm2 as high-B value DWI (high-B DWI).

5.3.3 Clinical Interpretation and Annotations

The mpMRI images were reviewed by three genitourinary (GU) radiologists (10+ years of clinical prostate MRI reading) as part of the standard clinical procedure following the clinical guideline (Weinreb et al., 2016; Turkbey et al., 2019). Lesion findings with PI-RADS score ≥ 3 are reported as MRI-positive findings, and the findings with PI-RADS score < 3 are interpreted as MRI-negative findings in this study.

The ground truth of the lesion annotations is confirmed by WMHP after RP matched to mpMRI prior to RP in this study. Blinded to all MRI-related information, the sliced WMHP specimens are examined and reported by three GU pathologists (with 14, 8, and 5 years of experience in clinical prostate histopathology interpretation) as part of the standard clinical procedure. Every PCa lesion was contoured and assigned a corresponding Gleason Score (GS) on WMHP. PCa lesions with GS$\geq$ 7 are defined as csPCa and are the detection targets of our proposed detection model in this study.

GU radiology research fellows, under the supervision of GU radiologists, retrospectively reviewed each mpMRI exam and contoured the region of interest (ROI) of MRI-visible lesions on T2WI images referring to the WMHP examination reports. MRI-positive findings are categorized as true positive if the radiological findings and the pathological findings are matched or false positive if no corresponding PCa lesion is found in histopathology reports. We defined the prospectively missed lesions that are retrospectively identified in the re-review procedure as false-negative (FN) lesions. The remaining PCa lesions that are MRI non-visible and also retrospectively unidentifiable on mpMRI are not included in the study as we cannot accurately contour them.

Compared to data consisting of biopsy-confirmed PCa, ground truth confirmed by WMHP offers additional insights into how the model would react to FN cases, which are generally harder to recognize in clinical practice. Understanding FN lesions is crucial, as overlooked or underestimated PCa can lead to insufficient treatment and undesired oncological outcomes (Borofsky et al., 2018; Lamb et al., 2015)

The prostate zonal segmentation of TZ and PZ are treated as part of the AtPCa-Net's

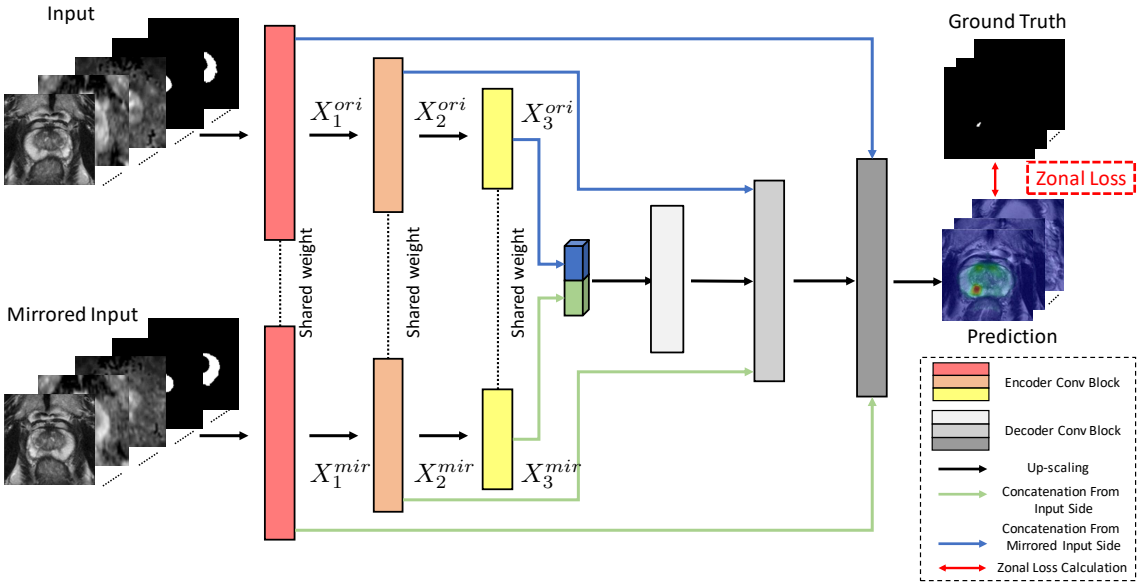


Figure 5.2: Architecture of the proposed AtPCa-Net. It combines the 3D symmetric-aware network and the proposed Zonal Loss (ZL). The network takes input stack images of T2WI, ADC, high-B DWI, TZ's mask, and PZ's mask images in an original way and a mirrored way. The weights of encoders at each level from the two sides of the figure are shared. At each level, the feature maps from two sides are concatenated first together and go through a bridge convolution block that consists of two consecutive 3D convolution blocks, then concatenated with the upscaled feature maps from the lower level, and finally upscaled to the upper-level decoder layers.

input, shown in Figure 5.2. The zonal masks are generated using a separate automatic prostate zonal segmentation model, CAT-Net (Hung et al., 2022), to explicitly provide the PCa-related anatomical information.

5.3.4 Preprocessing

The T2WI images underwent N4 bias filed correction to compensate for the low-frequency intensity non-uniformity (Tustison et al., 2010). The high-B DWI and ADC images are registered and resampled with respect to T2WI images using rigid spatial transformation while utilizing real-world coordinates information for each patient since the DWI and T2WI sequences are acquired temporally closed and only minimal patient motion are found (Cao et al., 2019b; Zheng et al., 2022b; Liu et al., 2013). After the registration, high-B DWI, ADC, and T2WI images are rotated with respect to the center line, generated

by connecting the volumetric center of the prostate and the TZ, to show the symmetric appearance. Then, high-B DWI, ADC, and T2WI images are center cropped with the size of 128x128 pixels from the original 320x320 pixels images as all prostates are allocated in the center of the acquired MR images following the clinical protocol (Weinreb et al., 2016; Turkbey et al., 2019). The intensity value of voxels in high-B DWI and T2WI images are linearly normalized to have a value in the range of [0, 1]. As the values of ADC maps are quantitative, the voxel intensities are consistent across patients (Cao et al., 2019b; Weinreb et al., 2016; Turkbey et al., 2019; Vargas et al., 2011). Therefore, the intensity on ADC maps is first clipped by a patient-independent value and then normalized to be in the range of [0, 1] (Cao et al., 2019b).

5.3.5 Symmetric-Aware Network Architecture

We first introduce the proposed symmetric-aware network architecture design that is capable of taking symmetric information into consideration explicitly. The detailed network architecture can be seen in Figure 5.2. In this study, we implement a UNet-like backbone structure since the UNet-like structures have shown great performances in medical-imaging-related segmentation and detection tasks (Isensee et al., 2021). The inputs are the 3D volumetric stack of images with a dimension of $[N \times C_{in} \times D_{in} \times H_{in} \times W_{in}]$, where N is the batch size, C_{in} is the number of channels, D_{in} is through-plane resolution, $[H_{in}, W_{in}]$ are in-plane resolution of the input mpMRI images. Different categories of images (T2WI, ADC, high-B DWI, binary mask of TZ, or binary mask of PZ) correspond to different channels of the input, and each imaging modality has the same volumetric size of $[D_0, H_0, W_0]$. The network takes two inputs: one is the original 3D stack of images, and the other is the 3D stack of images mirrored across the vertical axis.

The weights of all the convolution blocks (ConvBlock) in each encoder layer of the network are shared by both the original and mirrored paths. The design of shared-weight encoders has proven to be useful when visual comparisons are applied in downstream tasks in the DL network architecture designs (Chicco, 2021), similar as how the symmetric-

related anatomical priors were used to distinguish between benign and cancerous prostate tissue in the clinical practice. By sharing weights on the two encoders, features extracted from both pathways will maintain symmetric information. This, in turn, assists the network in learning how to utilize these symmetric features, resembling the human visual system, and thus enhances the model's decision-making process. In each level i other than the bottom level, the extracted feature maps $X_i^{ori} \in [N \times C_i \times D_i \times H_i \times W_i]$ and $X_i^{mir} \in [N \times C_i \times D_i \times H_i \times W_i]$ from both sides will be first concatenated channel-wisely and format a combined feature map $X_i^{cat} \in [N \times 2C_i \times D_i \times H_i \times W_i]$. Then X_i^{cat} will pass through a bridge block (BridgeBlock), composed of two 3D convolution blocks, and finally concatenated together with the upscaled feature map X_{i+1}^{up} from level $i+1$. The final concatenation will then be used to do further feature extraction at that level. Detail representations can be seen in the sub-figure of Decoder Blocks in Figure 5.2.

The output of the network is the detection probability map of where suspicious csPCa is allocated. The difference between the probability map and the ground truth mask is measured by the proposed ZL, which will be introduced in the following sub-section.

5.3.6 Zonal Loss

Current labeling strategies in PCa detection models generally inadequately account for the significance of PCa's zonal appearance differences, but using CE loss treating all PCa lesions identically regardless of their location (Cao et al., 2019b; Yu et al., 2020; Vente et al., 2020; Seetharaman et al., 2021; Cao et al., 2021; Hosseinzadeh et al., 2022; Saha et al., 2021; Weinreb et al., 2016; Turkbey et al., 2019). We propose an anatomical-aware hierarchical label and loss design, the ZL, to guide the model to learn the different appearances of PCa lesions in different zones with anatomy-informed constraints.

We denote the set of voxels of the PZ region as $\mathbb{P}$, the TZ region as $\mathbb{T}$, and the csPCa lesion region as $\mathbb{L}$ for a given prostate mpMRI image. Given an input image $\boldsymbol{\mathcal{I}} \in \boldsymbol{R}^{C \times D \times H \times W}$ and the corresponding binary mask $\boldsymbol{\mathcal{M}} \in \{0,1\}^{D \times H \times W}$, where C, D, H, W are the channel number, depth, height and width of the input image $\boldsymbol{\mathcal{I}}$, for any voxel

$v \in \boldsymbol{\mathcal{I}}$, the corresponding label voxel m on $\boldsymbol{\mathcal{M}}$ in binary CE loss design is given as

$$m = \begin{cases} 1 & \text{if } v \in \mathbb{L}, \\ 0 & \text{otherwise.} \end{cases} \tag{5.1}$$

One of the key points of the hierarchical label and loss design is the multi-level labeling design with respect to the number of properties each class holds - lower-level classes hold fewer properties and constraints, higher-level classes hold more properties and more constraints correspondingly (Verma et al., 2012; Bertinetto et al., 2020; Li et al., 2022). According to PI-RADS (Weinreb et al., 2016; Turkbey et al., 2019), PCa lesions in PZ mostly require visual information related to DWI, while lesions in TZ require a combined evaluation of both T2WI and DWI for accurate diagnosis. We adopt this clinical interpretation process using the hierarchical label and loss design by treating the class of TZ lesions as requiring additional information from T2WI images compared with the class of lesions in PZ for improved PCa detection on mpMRI. Hence, the ZL design is adept at acknowledging the distinct zonal appearance of PCa while preserving the anatomical congruence between lesions in the TZ and PZ.

We design a hierarchical labeling with ground truth mask $\boldsymbol{\mathcal{M}} \in \{0,1\}^{2\times D\times H\times W}$ for a given image $\boldsymbol{\mathcal{I}} \in \boldsymbol{R}^{C\times D\times H\times W}$. For any voxel $v \in \boldsymbol{\mathcal{I}}$, the corresponding label vector $\boldsymbol{m} = [m_0, m_1] \in \{0,1\}^2$ in $\boldsymbol{\mathcal{M}}$ in our loss design is given as

$$\boldsymbol{m} = \begin{cases} [1,1], & \text{if } v \in \mathbb{L} \cap \mathbb{T} \\ [1,0], & \text{if } v \in \mathbb{L} \cap \mathbb{P} \\ [0,0], & \text{otherwise.} \end{cases} \tag{5.2}$$

This label design aims to adopt the clinical prior knowledge to the detection of csPCa lesions. Abnormalities should be observed on image sequences related to DWI in common for both the lesions in TZ and PZ (m_0), and additional abnormal observations from T2WI are needed for lesions in TZ in order to make more accurate diagnoses (m_1) (Weinreb et al., 2016; Turkbey et al., 2019).

We denote the probability vector $\boldsymbol{p} = [p_0, p_1] \in [0,1]^2$ in the output probability map $\boldsymbol{\mathcal{P}} \in [0,1]^{2\times D\times H\times W}$ for the corresponding voxel $v \in \boldsymbol{\mathcal{I}}$, where

$$p_0 = \begin{cases} p_0 & \text{if } m_0 = 1 \\ \max(p_0, p_1) & \text{if } m_0 = 0 \end{cases} \tag{5.3}$$

and

$$p_1 = \begin{cases} \min(p_0, p_1) & \text{if } m_1 = 1 \\ p_1 & \text{if } m_1 = 0. \end{cases} \tag{5.4}$$

The modified CE loss can be written as

$$\mathcal{L}(\boldsymbol{\mathcal{P}}, \boldsymbol{\mathcal{M}}) = \sum_v -\boldsymbol{m} \log \boldsymbol{p} - (\mathbf{1} - \boldsymbol{m}) \log (\mathbf{1} - \boldsymbol{p}), \tag{5.5}$$

where $\mathbf{1} = [1, 1]$. It is designed to suppress prediction vector patterns that should not exist. Based on our labeling design, label vector pattern $\hat{\boldsymbol{m}} = [0, 1]$ is not defined, since solely abnormalities found on T2WI have limited contribution to the diagnosis of suspicious PCa (Weinreb et al., 2016; Turkbey et al., 2019). Therefore, any output probability vectors with $p_1 > p_0$ should be penalized in order to teach the model not to conduct such predictions. However, the original CE loss with binary labeling only computes the loss of each class independently but ignores this inter-class relationship. We intentionally add this constraint onto the original CE loss, which is shown in 5.3 and 5.4. In 5.3, $p_0 = \max(p_0, p_1)$ when $m_0 = 0$, and in 5.4, $p_1 = \min(p_0, p_1)$ when $m_1 = 1$ both indicate that p_0 should be greater than p_1 in any prediction outputs, and any patterns disobey this rule should be penalized. This modification could further help the model converge to a better solution.

In addition, we adopt Focal Loss onto the modified CE loss in (5.5) to account for the imbalance in the number of voxels between the csPCa lesions and background (Lin et al., 2020). This would reduce the relative weight for well-classified voxels and emphasize focus on hard ones like lesion voxels (Lin et al., 2020; Cao et al., 2019b). The final ZL form

follows:

$$\mathcal{L}^{ZL}(\boldsymbol{\mathcal{P}}, \boldsymbol{\mathcal{M}}) = \sum_{v} -\boldsymbol{m}(\mathbf{1}-\boldsymbol{p})^{\gamma}\log \boldsymbol{p} - (\mathbf{1}-\boldsymbol{m})\boldsymbol{p}^{\gamma}\log(\mathbf{1}-\boldsymbol{p}), \tag{5.6}$$

where $\mathbf{1} = [1, 1]$, and $\boldsymbol{p} \in [0, 1]^2$ is defined in 5.3 and 5.4.

5.3.7 Implementation Details

In each of the three levels of the network architecture, the channel number is [64, 128, 256] for each level of the convolutional layers of the encoders, and [256, 128, 64] for each level of the convolutional layers in the decoders, correspondingly (Isensee et al., 2021).

Each level of the convolutional layers comprises four consecutive ConvBlocks, and each ConvBlocks consists of a $3 \times 3 \times 3$ 3D convolution kernel, following by a LeakyReLU activation function and an instance normalization, following the settings of the nnU-Net (Isensee et al., 2021).

Each training procedure takes 60 epochs, with early-stopping strategy applied when no loss degradation for 30 accumulate epochs was found to avoid potential overfitting issues. Adam optimizer (Kingma and Ba, 2015) was adopted with the loss function of the Focal Loss (Lin et al., 2020) by default, the ZL when specifically stated. All models are trained on an Nvidia RTX3090 GPU.

5.4 Results

5.4.1 Quantitative Results

For csPCa lesion detection, we evaluate the overall csPCa detection performance of the AtPCa-Net using the free-response receiver operating characteristic (FROC) analysis (Litjens et al., 2014). The FROC curve helps analyze the relationship between model detection sensitivity and the level of FP predictions per patient. In the experiment, we consider the local maxima on the output probability map as the csPCa detection points.

The csPCa detection point is defined as a true positive (TP) when the point is within 5 mm of any csPCa ground truth ROIs to account for a potential mismatch between the whole-mount specimen and mpMRI of the corresponding ROI (Litjens et al., 2014; Cao et al., 2019b).

We also evaluate the per-patient level classification performance of the proposed AtPCa-Net by defining patients with csPCa as positive cases, and patients without csPCa as negative cases. For each patient, we treat the highest value on the output probability map as its probability of having csPCa. The evaluation of the per-patient level classification performance is done by using Receiver Operating Characteristic (ROC) analysis. In both ROC and FROC analysis, we evaluate the model performances by 5-fold cross-validation after 1000 times bootstrapping.

ROCs were compared with DeLong Test (ER Delong, 1988), and the sensitivity results at each number of FP predictions per patient were compared using Chi-squared Test, in accordance with 95% confidence interval (95% CI), correspondingly.

We performed comparisons between our proposed model and other popular 3D image segmentation models, including SEResUNet (Hu et al., 2018), ResidualUNet (Bhalerao and Thakur, 2020), VNet (Chen et al., 2021), AttentionUNet (Oktay et al., 2018), VoxResNet (Chen et al., 2018), nnUNet (Isensee et al., 2021), and UNETR (Hatamizadeh et al., 2022) for csPCa detection and patient-level classification. Figure 5.3 visualizes the comparison of csPCa detection performance among different models via different FROC curves. Figure 5.4 visualizes the comparison of patient-level classification performance among different models via ROC curves. Table 5.1 shows the comparisons of the patient level classification AUCs in the format of mean, and the comparisons of csPCa detection performance via showing the sensitivity results against 0.5/1/1.5/2/2.5 FP predictions per patient. Our proposed AtPCa-Net outperforms all other models on all the FROC measurements on 0.5/1/1.5/2/2.5 FP predictions per patient with higher mean sensitivities (p¡0.05). The AtPCa-Net also outperforms all other models on the patient-level classification AUCs (p¡0.05).

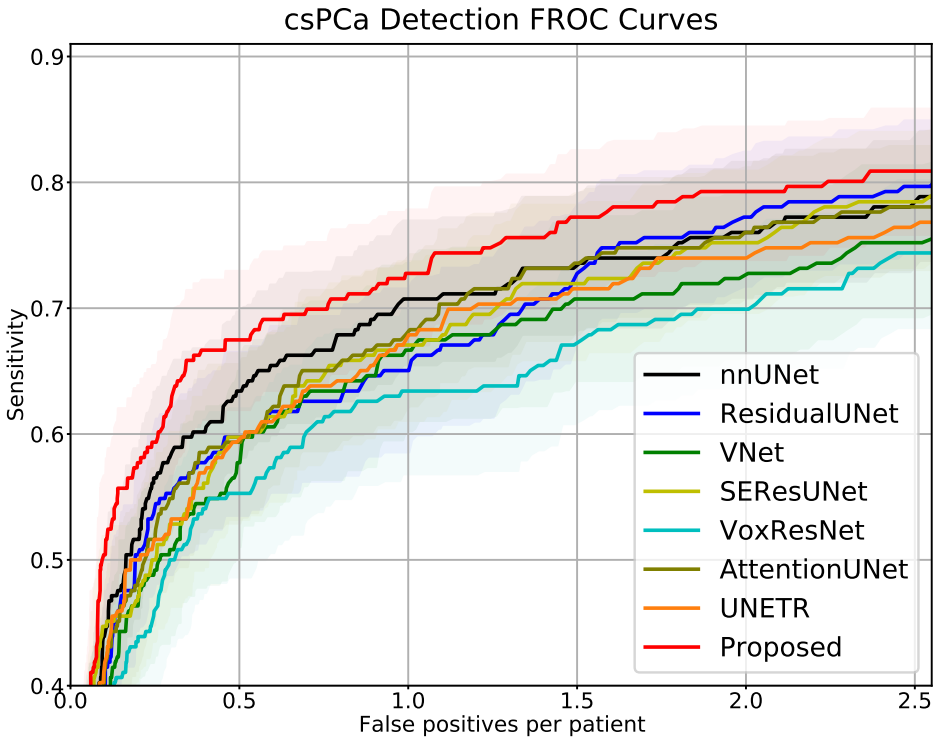


Figure 5.3: csPCa detection performance comparison via FROC curves, which measured by the detection sensitivity (y-axis) against number of false-positive predictions per patient (x-axis). Solid lines are the mean FROC curves, and shadow areas represent the corresponding 95% confidence interval.

Compared with some of the existing studies proposing PCa detection using PCa biopsy results (Liu et al., 2013; Hosseinzadeh et al., 2022; Seetharaman et al., 2021), our study uses results confirmed by WMHP, which results in additional FN csPCa annotations as the prospectively missed csPCa lesions were retrospectively annotated. In order to discuss the possible performance discrepancies caused by the dataset's differences between existing studies and ours, we also perform ROC and FROC analysis to the results using the dataset after excluding FN lesions, shown in Table 5.5. The proposed AtPCa-Net outperforms all other baseline models (Hu et al., 2018; Oktay et al., 2018; Chen et al., 2018; Bhalerao and Thakur, 2020; Chen et al., 2021; Hatamizadeh et al., 2022) on both ROC and FROC measurements when using the dataset excluding all FN lesions ($p<0.05$), which keeps consistent to its performance when including FN lesions, as shown in Table 5.3.

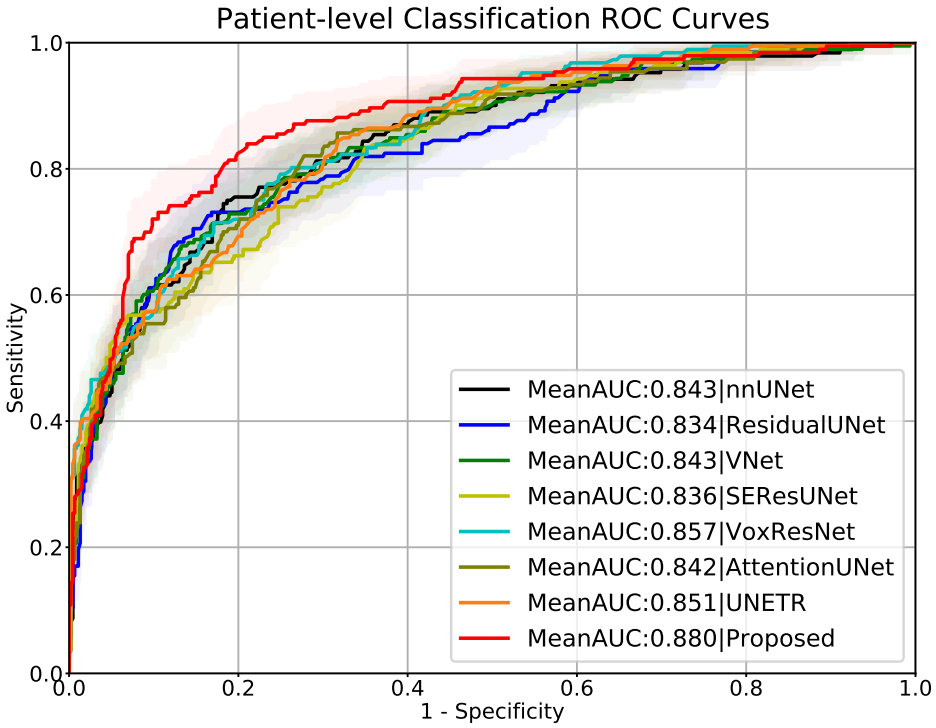


Figure 5.4: Patient-level classification performance comparisons via ROC curves, which are measured by the sensitivity (y-axis) against the false-positive rate (x-axis). Solid lines are the mean ROC curves, and shadow areas represent the corresponding 95% confidence interval.

Table 5.1: Patient-level classification and csPCa detection performance comparisons among different models.

Models	Patient Classification	csPCa Detection Sensitivity (95% CI)				
	AUC (%95 CI)	0.5 FP/Patient	1 FP/Patient	1.5 FP/Patient	2 FP/Patient	2.5 FP/Patient
nnUNet	0.843 (0.803,0.883)	0.634 (0.571,0.697)	0.707 (0.651,0.763)	0.736 (0.684,0.788)	0.760 (0.707,0.813)	0.789 (0.735,0.853)
ResidualUNet	0.834 (0.794,0.874)	0.595 (0.534,0.656)	0.650 (0.592,0.708)	0.727 (0.674,0.780)	0.772 (0.724,0.820)	0.797 (0.745,0.849)
VNet	0.843 (0.801,0.885)	0.594 (0.529,0.659)	0.667 (0.604,0.730)	0.707 (0.644,0.770)	0.728 (0.667,0.789)	0.752 (0.696,0.808)
SEResUNet	0.836 (0.797,0.875)	0.598 (0.534,0.662)	0.671 (0.609,0.733)	0.720 (0.656,0.784)	0.752 (0.692,0.812)	0.785 (0.728,0.842)
VoxResNet	0.857 (0.824,0.890)	0.553 (0.489,0.617)	0.634 (0.572,0.696)	0.675 (0.615,0.735)	0.699 (0.643,0.755)	0.744 (0.683,0.805)
AttentionUNet	0.842 (0.804,0.880)	0.594 (0.532,0.656)	0.683 (0.624,0.742)	0.736 (0.677,0.795)	0.760 (0.708,0.812)	0.781 (0.732,0.830)
UNETR	0.851 (0.815,0.887)	0.598 (0.537,0.659)	0.679 (0.619,0.739)	0.715 (0.659,0.771)	0.740 (0.685,0.795)	0.769 (0.716,0.822)
Proposed	**0.880 (0.846,0.914)**	**0.675 (0.620,0.730)**	**0.728 (0.674,0.782)**	**0.772 (0.716,0.828)**	**0.793 (0.741,0.845)**	**0.809 (0.757,0.861)**

5.4.2 Qualitative Results

We qualitatively evaluate the model performance by showing representative examples of csPCa detection performance comparisons in Figure 5.5. Figure 5.5 A and B correspond to two patients with csPCa, and Figure 5.5 C and D correspond to two patients without csPCa. Overall, the proposed AtPCa-Net conducted fewer FP predictions with the same TP predictions on all cases compared with other models (Hu et al., 2018; Oktay et al., 2018; Chen et al., 2018; Bhalerao and Thakur, 2020; Chen et al., 2021; Hatamizadeh et al., 2022).

We can also observe its ability to suppress symmetric FP predictions with its better ability to distinguish symmetric abnormal patterns of csPCa from other normal prostate tissue, like BPH and CZ, compared with other models (Hu et al., 2018; Oktay et al., 2018; Chen et al., 2018; Bhalerao and Thakur, 2020; Chen et al., 2021; Hatamizadeh et al., 2022). Patients B and C are representative examples of patients who have BPH, and Patient D is a representative example of patients whose CZ's appearance could mislead the model's prediction. The BPH, pointed by green arrows for Patient B and C on the MR images, and the CZ region, pointed by the yellow arrows for Patient D, show visually similar appearances as the csPCa on mpMRI images but with symmetric patterns. We can observe that for all patients, other models that do not take the symmetric-related anatomical information into consideration misidentify the BPH and the CZ as csPCa and result in FP predictions. Our models can correctly detect the csPCa with fewer FP predictions with the help of the symmetric-related anatomical-aware architecture design.

5.4.3 Backbone Network Extension

To show the generalizable potential of the proposed anatomical-aware design, we also try to transplant the architecture onto another UNet-like backbone network. We implement the nnUNet++ (Hung et al., 2022), a UNet++ (Zhou et al., 2018) variant, with the proposed ZL and symmetric-aware architecture. Similar to the nnUNet-based approach, the weights of each non-decoder block are shared. In nnUNet++, the feature maps from

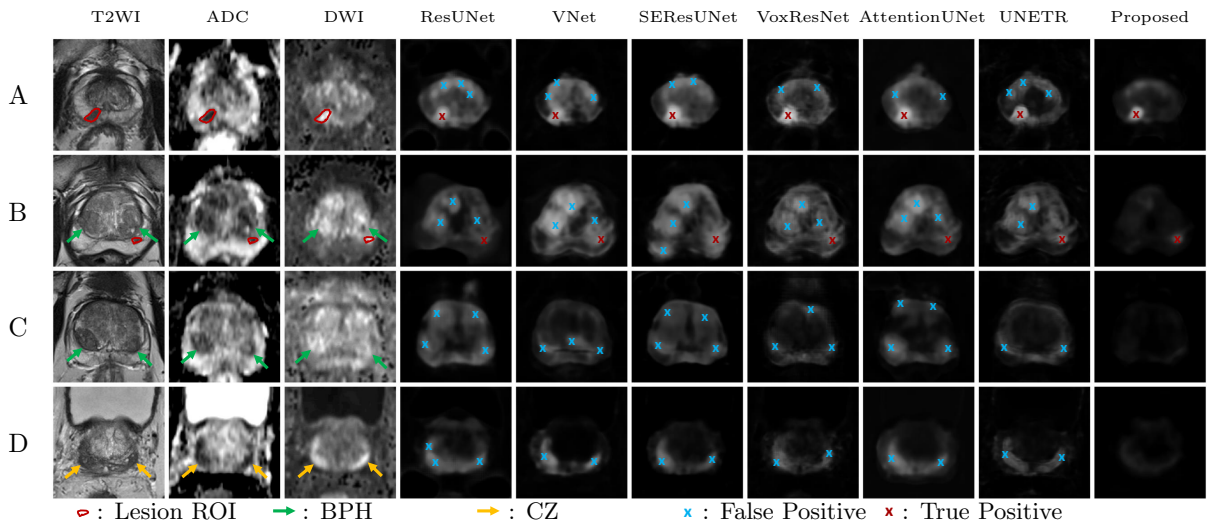


Figure 5.5: Visualizations of csPCa detection results among different models ResUNet (Bhalerao and Thakur, 2020), VNet (Chen et al., 2021), SEResUNet (Hu et al., 2018), VoxResNet(Chen et al., 2018), AttentionUNet (Oktay et al., 2018), UNETR (Hatamizadeh et al., 2022) on two patients with csPCa (A and B) and two without csPCa (C and D). From left to right, the first column shows the T2WI images, the second column shows the ADC images, the third column shows the high-B DWI images and all other columns show the detection probability maps generated by different models for each patient A, B, C and D correspondingly. Red contours shown on T2WI, ADC, and high-B DWI images indicate the contour of csPCa lesions. Yellow arrows point to the regions of CZ, and Green arrows point to the regions of BPH on the MR images. Blue crossings represent FP predictions, and red crossings represent TP predictions.

both sides of the network merge after each skip connection that ends at the decoder blocks on each level, similar to the implementation with nnUNet as the backbone network.

Table 5.2 shows the performance of using the two backbone networks on the original dataset, and Table 5.3 shows the performance comparisons when excluding all FN lesions. The AtPCa-Net(nnUNet) and AtPCa-Net(nnUNet++) represents for AtPCa-Net using nnUNet and nnUNet++ as a backbone network, respectively. From both the Table 5.2 and the Table 5.3, we can observe that both nnUNet-based AtPCa-Net and nnUNet++-based AtPCa-Net achieved better detection and classification performance than when only using the nnUNet or nnUNet++, respectively (p¡0.05). This indicates the generalizable potential of applying the proposed anatomical-aware design with different backbone networks. From Table 5.2, we see the nnUNet-based AtPCa-Net performs better on patient-level classification and also achieves higher sensitivities at 0.5/1 FP predictions

Table 5.2: Model performance comparisons using different backbone networks. A: nnUNet, B:nnUNet++, C: AtPCa-Net(nnUNet), D:AtPCa-Net(nnUNet++).

Models	Patient Classification	csPCa Detection Sensitivity (95% CI)				
	AUC (95% CI)	0.5 FP/Patient	1 FP/Patient	1.5 FP/Patient	2 FP/Patient	2.5 FP/Patient
A	0.843 (0.803,0.883)	0.634 (0.571,0.697)	0.707 (0.651,0.763)	0.736 (0.684,0.788)	0.760 (0.707,0.813)	0.789 (0.735,0.843)
B	0.867 (0.832,0.902)	0.626 (0.554,0.698)	0.695 (0.632,0.758)	0.756 (0.698,0.814)	0.781 (0.724,0.838)	0.817 (0.769,0.865)
C	**0.880 (0.846,0.914)**	**0.675 (0.620,0.730)**	**0.728 (0.674,0.782)**	**0.772 (0.716,0.828)**	0.793 (0.741,0.845)	0.809 (0.757,0.861)
D	0.877 (0.842,0.912)	0.662 (0.601,0.723)	0.719 (0.662,0.776)	**0.772 (0.719,0.825)**	**0.809 (0.757,0.861)**	**0.821 (0.771,0.871)**

Table 5.3: Model performance comparisons using different backbone networks after excluding FN lesions. A: nnUNet, B:nnUNet++, C: AtPCa-Net(nnUNet), D:AtPCa-Net(nnUNet++)

Models	Patient Classification	csPCa Detection Sensitivity (95%CI)				
	AUC (95%CI)	0.5 FP/Patient	1 FP/Patient	1.5 FP/Patient	2 FP/Patient	2.5 FP/Patient
A	0.873 (0.810,0.936)	0.783 (0.720,0.846)	0.862 (0.811,0.913)	0.889 (0.845,0.933)	0.915 (0.880,0.950)	0.921 (0.885,0.957)
B	0.876 (0.839,0.913)	0.762 (0.697,0.827)	0.836 (0.783,0.889)	0.884 (0.836,0.932)	0.905 (0.863,0.947)	0.921 (0.879,0.963)
C	**0.898 (0.866,0.930)**	**0.825 (0.767,0.883)**	**0.873 (0.822,0.924)**	**0.900 (0.856,0.944)**	**0.931 (0.890,0.972)**	**0.931 (0.891,0.971)**
D	**0.898 (0.866,0.930)**	0.820 (0.765,0.875)	0.852 (0.800,0.904)	**0.900 (0.855,0.945)**	0.921 (0.882,0.960)	0.926 (0.890,0.962)

per patient. When the rate of FP predictions per patient raises to 2/2.5 FP predictions per patient, the nnUNet++-based AtPCa-Net achieves higher sensitivities. In Table 5.3, nnUNet-based AtPCa-Net outperforms the nnUNet++-based AtPCa-Net on all situations, except similar on patient-level classification AUC and sensitivity at 1.5 FP predictions per patient. We select nnUNet as the backbone as it outperforms the nnUNet++-based AtPCa-Net in the majority of situations and represents the nnUNet-based AtPCa-Net as AtPCa-Net if without any other descriptions in this paper.

5.4.4 Ablation Study

We conduct ablation studies to discover the importance of each component of the proposed AtPCa-Net, shown in Table 5.4. We can observe that either modifying the Focal Loss to the ZL or modifying the network architecture to be in the symmetric-aware architecture improves the performance on both per-patient level classification and csPCa detection. When all the components are included, which formats the proposed AtPCa-Net, it outperforms all other situations when only partial components are included, showing the superiority of our proposed method and the usefulness of integrating all the mentioned prostate anatomical-related prior into the model.

Table 5.4: Ablation study of effects of including/excluding components of the AtPCa-Net

Components		Patient Classification	csPCa Detection Sensitivity (95%CI)				
Sym-aware	Zonal Loss	AUC (95%CI)	0.5 FP/Patient	1 FP/Patient	1.5 FP/Patient	2 FP/Patient	2.5 FP/Patient
		0.843 (0.803,0.883)	0.634 (0.571,0.697)	0.707 (0.651,0.763)	0.736 (0.684,0.788)	0.760 (0.707,0.813)	0.789 (0.735,0.843)
✓		0.878 (0.843,0.913)	0.650 (0.586,0.714)	0.707 (0.643,0.771)	0.756 (0.693,0.819)	0.785 (0.729,0.841)	0.789 (0.731,0.847)
	✓	0.854 (0.817,0.891)	0.642 (0.580,0.704)	0.707 (0.651,0.663)	0.744 (0.690,0.798)	0.781 (0.729,0.833)	0.793 (0.740,0.846)
✓	✓	**0.880 (0.846,0.914)**	**0.675 (0.620,0.730)**	**0.728 (0.674,0.782)**	**0.772 (0.716,0.828)**	**0.793 (0.741,0.845)**	**0.809 (0.757,0.861)**

Table 5.5: Patient-level classification and csPCa detection performance comparisons of different models without FN lesions.

Models	Patient Classification	csPCa Detection Sensitivity (95%CI)				
	AUC (95%CI)	0.5 FP/Patient	1 FP/Patient	1.5 FP/Patient	2 FP/Patient	2.5 FP/Patient
nnUNet	0.873 (0.810,0.936)	0.783 (0.720,0.846)	0.862 (0.811,0.913)	0.889 (0.845,0.933)	0.915 (0.880,0.950)	0.921 (0.885,0.957)
ResidualUNet	0.844 (0.803,0.885)	0.689 (0.610,0.768)	0.762 (0.686,0.838)	0.834 (0.769,0.899)	0.887 (0.835,0.939)	0.901 (0.853,0.949)
VNet	0.861 (0.823,0.899)	0.698 (0.628,0.768)	0.799 (0.736,0.862)	0.841 (0.785,0.897)	0.862 (0.812,0.912)	0.884 (0.840,0.928)
SEResUNet	0.860 (0.822,0.898)	0.725 (0.658,0.792)	0.799 (0.742,0.856)	0.852 (0.803,0.901)	0.884 (0.838,0.930)	0.900 (0.857,0.943)
VoxResNet	0.874 (0.841,0.907)	0.683 (0.616,0.750)	0.773 (0.713,0.833)	0.815 (0.762,0.868)	0.836 (0.784,0.888)	0.884 (0.836,0.932)
AttentionUNet	0.857 (0.819,0.895)	0.735 (0.666,0.804)	0.820 (0.767,0.873)	0.857 (0.806,0.908)	0.884 (0.832,0.936)	0.910 (0.869,0.951)
UNETR	0.873 (0.836,0.910)	0.735 (0.668,0.802)	0.831 (0.774,0.888)	0.862 (0.813,0.911)	0.889 (0.841,0.937)	0.905 (0.862,0.948)
Proposed	**0.898 (0.866,0.930)**	**0.825 (0.767,0.883)**	**0.873 (0.822,0.924)**	**0.900 (0.856,0.944)**	**0.931 (0.890,0.972)**	**0.931 (0.891,0.971)**

5.5 Discussion

Our study demonstrated that the anatomical-aware designs, specifically the symmetric-aware architecture and the ZL, of the AtPCa-Net can help improve the csPCa detection performance and also patient-level classification results. We attribute the improvements in our model not only to the zonal-related knowledge learned under the guidance of the anatomically-aware ZL but also to the ability to reduce FP, which is a direct result of our symmetric-aware network architecture design.

The proposed anatomical-aware designs in AtPCa-Net help improve model performance on both csPCa detection and patient-level classification. The ZL shows its effectiveness by taking the lesion appearance differences on mpMRI images in different prostate zones into consideration. There are several approaches (Hosseinzadeh et al., 2022; Vente et al., 2020) trying to utilize the zonal information by stacking the zonal mask as part of input together with the CE loss function and have shown improvement in model performance. However, PCa lesions located in different prostate regions are treated identically by using the CE loss, ignoring the essential anatomical information related to PCa's zonal appearance differences. By using the ZL, an additional anatomical-aware constraint is added, and the zonal masks are further utilized. In addition, the symmetric-aware architecture of the AtPCa-Net helps reduce the FP predictions that are related closely

to other normal prostate tissue with similar visual appearances as the PCa lesions on mpMRI, like BPH and CZ. The symmetric nature of the proposed network design helps distinguish the differences between the asymmetric patterns of PCa and the symmetric patterns of other normal prostate tissue. We can see that the integration of both the anatomical-aware designs, the ZL and the symmetric-aware architecture, helps improve the model performance more compared with the situations when including each individual design only, with 4.1%/2.1%/3.6%/3.3%/2.0% sensitivity per 0.5/1/1.5/2/2.5 FP/Patient and 3.7% AUC improvements.

In this study, all patients with csPCa are confirmed by the WMHP results. Different from some of the existing studies using results confirmed by prostate biopsies (Liu et al., 2013; Hosseinzadeh et al., 2022; Seetharaman et al., 2021), our WMHP dataset has the retrospective annotations for MR-visible FN lesions that are prospectively missed. The model performance regarding the FN lesions is important as the missing lesions or underestimation of the PCa's volume and significance could result in inadequate therapy and consequently undesired oncologic outcomes (Borofsky et al., 2018; Lamb et al., 2015). Although both Table 5.1 and Table 5.5 show the consistent superiority of the proposed AtPCa-Net compared with other models, it also reveals that all model performances dropped on both ROC and FROC measurements when including FN lesions compared with the situation when all FN lesions are excluded.

The results highlight the challenges in automatically identifying FN csPCa lesions via deep-learning models, which aligns with the observations from existing study about the difficulty to identify FN lesions in clinical practice (Borofsky et al., 2018). FN lesions are typically tiny and sometimes might be affected by the spatial resolution of the MRI imaging, making it hard to be detected (Borofsky et al., 2018). Future studies could potentially be conducted regarding how to build an effective automatic csPCa detection model focusing on issues related to FN csPCa lesions, in conjunction with advancements MR technology to enhance the resolution of mpMRI.

We also evaluated the proposed AtPCa-Net on patients cohorts grouped by different prostate-specific antigen density (PSAD) level. The PSAD level is one of the clinical

Table 5.6: Patient-level classification and csPCa detection performance comparisons using the proposed AtPCa-Net on patients with prostate-specific antigen density (PSAD) level above 0.15 ng/ml/ml and below 0.15 ng/ml/ml. A represents for patients with PSAD≥0.15 ng/ml/ml, and B represents for patients with PSAD≤0.15 ng/ml/ml

Settings	Patient Classification	csPCa Detection Sensitivity (95% CI)				
	AUC (95% CI)	0.5 FP/Patient	1 FP/Patient	1.5 FP/Patient	2 FP/Patient	2.5 FP/Patient
A	0.923 (0.869, 0.966)	0.780 (0.683, 0.881)	0.840 (0.750, 0.936)	0.900 (0.820, 0.976)	0.900 (0.820, 0.976)	0.900 (0.820, 0.976)
B	0.841 (0.738, 0.925)	0.674 (0.519, 0.825)	0.744 (0.587, 0.881)	0.814 (0.667, 0.950)	0.837 (0.719, 0.950)	0.884 (0.776, 0.977)

Table 5.7: Patient-level classification and csPCa detection performance comparisons using the proposed AtPCa-Net on patients with prostate-specific antigen density (PSAD) level above 0.20 ng/ml/ml and below 0.20 ng/ml/ml. A represents for patients with PSAD≥0.20 ng/ml/ml, and B represents for patients with PSAD≤0.20 ng/ml/ml

Settings	Patient Classification	csPCa Detection Sensitivity (95% CI)				
	AUC (95% CI)	0.5 FP/Patient	1 FP/Patient	1.5 FP/Patient	2 FP/Patient	2.5 FP/Patient
A	0.924 (0.862, 0.970)	0.800 (0.667, 0.926)	0.833 (0.706, 0.957)	0.900 (0.780, 1.000)	0.900 (0.780, 1.000)	0.900 (0.780, 1.000)
B	0.868 (0.785, 0.934)	0.698 (0.574, 0.820)	0.778 (0.660, 0.885)	0.825 (0.714, 0.927)	0.841 (0.732, 0.941)	0.873 (0.787, 0.952)

factors indicating the level of potential risk of patients having PCa (Yusim et al., 2020; Washino et al., 2017). The results can be found in Table 5.6 and Table 5.7. In all, we showed the csPCa detection and patient-level classification performances of the proposed AtPCa-Net on patient cohorts grouped by cut-off PSAD level of 0.15 ng/ml/ml and 0.20 ng/ml/ml, which used as recommended thresholds for evaluating the risk of patients having PCa in existing studies (Yusim et al., 2020; Washino et al., 2017). The results exhibited the proposed AtPCa-Net performed better in patient cohort with higher PSAD compared with the cohort with lower PSAD in both cut-off settings. Further improvement could be made to improve the model performance when integrating the clinical information with the DL model design. For example, the DL model may be able to capture the risk for the patient having csPCa by the imported PSAD level, and then learn to enhance the prediction efficacy accordingly. Collecting potential related clinical and demographic information and discovering how to effectively integrating them with the DL model designs could be our future research directions.

Several limitations exist in the study. The model evaluations might be affected by the fact that the WMHP dataset was collected from a single institution and with MR machines from a single vendor. In the future, we will expand the WMHP dataset with

multi-center collaborations and multi-vendor data, improving the diversity of the dataset with multiple clinical settings and patient demographics, and validate the proposed model's generalizability and further solidify our findings. In addition, the real-world diagnosis of PCa generally integrates radiological findings together with clinical test results and demographic information (Weinreb et al., 2016; Turkbey et al., 2019). However, just like other existing studies (Seetharaman et al., 2021; Hosseinzadeh et al., 2022; Yu et al., 2020; Cao et al., 2019b, 2021; Vente et al., 2020; Saha et al., 2021; Li et al., 2023), our study is limited on only utilizing information from mpMRI images. Potential performance improvements could be achieved if including clinical test results and demographic information in the csPCa detection model design, since they have shown the ability to improve model performance compared with using imaging information only in other computer-aided disease diagnosis studies (Huang et al., 2020b; Reda et al., 2018; Zheng et al., 2022a; Dong et al., 2023; Liu et al., 2023). In addition, due to the limit role played by the DCE imaging in the clinical practice, we excluded the DCE imaging from the model design, like other existing studies with the same research objectives (Huang et al., 2020b; Reda et al., 2018; Zheng et al., 2022a; Dong et al., 2023; Liu et al., 2023). As the DCE imaging can provide microvascular structure information, it could also potentially contribute to improved PCa diagnosis by providing imaging information from another perspectives (Turkbey et al., 2019). The integration of clinical information and radiological findings, and the inclusion of the DCE imaging could be our future research directions.

We have shown that by taking PCa-specific anatomical priors into consideration, the PCa detection model improves its performance on both csPCa detection and patient-level classification. We believe the advantage comes from the key ideas of fusing the anatomical-related clinical priors into the loss function and network architecture design, which can better guide model training.

The achievement could potentially influence the future designs of the DL-based PCa detection models on how the anatomical priors could help enhance the performance when integrating with DL model designs. We hypothesize that integrating the disease-specific anatomical-related knowledge into the model design could also potentially improve the

model performance for other diseases, which could be a future research direction.

5.6 Conclusions

We have demonstrated that by integrating anatomical priors into the deep learning network architecture design, the model efficacy is enhanced on both clinically significant prostate cancer (csPCa) detection and patient-level classification. Adopted from the clinical interpretation, the anatomical priors are carefully achieved by a symmetric-aware architecture design and the Zonal Loss (ZL), which format the proposed 3D anatomical-aware prostate cancer detection network (AtPCa-Net). Our experiments show that the model performance improves when either symmetric-related anatomical priors or zonal appearance differences of PCa are considered, with the best results achieved when the model incorporates both information. The proposed AtPCa-Net shows superior performance to other baseline models in both csPCa detection and patient-level classification, and shows the potential to further reduce the number of unnecessary biopsies may be caused by using current DL models. Our approach also reveals the potential flexibility of the anatomical-aware designs as they can improve the model performance with different backbone networks. How to generalize the anatomical-aware design idea to other specific diseases and how to integrate the design with clinical test results and demographic information could be our future research directions.

CHAPTER 6

Conclusions and Future Work

6.1 Research Summary

In this dissertation, we have explored the potential of integrating imaging information with clinical data into learning-based CAD models, focusing primarily on applications related to magnetic resonance imaging. In summary, our four primary contributions are as follows:

6.1.1 Integrative Radiomics-Based Machine Learning Model Predicting Prostate Biopsy Results

We have introduced an integrative radiomics-based machine learning model designed to predict biopsy results in patients with negative prostate mpMRI. Our aim was to address the current clinical challenge of unclear criteria for determining which patients with negative prostate mpMRI should undergo prostate biopsy. Compared with methods utilizing only radiological findings or clinical data, our proposed learning-based CAD model integrates both imaging information, through radiomics modeling, and clinical information, resulting in superior prediction performance. Our results demonstrate that the enhanced performance of the learning-based CAD model, which is superior to existing methods in clinical practice, is attributable to the integration of both clinical data and imaging information. This underscores the potential of a broader utilization of the proposed integration approach in addressing clinical challenges in practice.

6.1.2 Integrative Radiomics-Based Machine Learning Model Predicting Pelvic Lymph Node Invasion

We explored the potential of applying a novel integrative CAD model to address another critical clinical problem: predicting pelvic lymph node invasion for patients undergoing radical prostatectomy. In this study, we adopted similar concepts and machine learning models, aiming to investigate the generalizability and feasibility of integrating imaging and clinical information to enhance the performance of learning-based CAD models. Compared with methods utilizing a single modality of information, our proposed integrative learning-based CAD model, which incorporates both modalities of information, still achieves the best prediction performance. Additionally, when compared with existing routinely-used clinical nomograms, the proposed model also demonstrates superior performance. Our results underscore the effectiveness of including both imaging and clinical information in the design of learning-based CAD models.

6.1.3 Integrative Feature-Enhanced Network Predicting Prostate Biopsy Results

We explored the potential of incorporating clinical data and imaging information within the framework of deep-learning-based CAD models, inspired by the successful implementation observed with traditional machine-learning-based CAD models. Taking on a similar clinical problem as discussed in 6.1.1, but with a more comprehensive dataset, we developed more flexible feature extraction methods, specifically targeted to nearly imperceptible object signals, compared with approaches uilizing pre-defined radiomics features. Furthermore, we adopted more capable deep learning networks compared to traditional machine learning models. In comparison, our proposed integrative feature-enhanced network demonstrates superior performance when incorporating both imaging and clinical features, compared to situations considering only each individual group of features. Additionally, our proposed model also outperforms the radiomics-based machine learning model discussed in 6.1.1, highlighting the potential superiority of the proposed deep-learning-based CAD models.

6.1.4 AtPCa-Net: Anatomical-Aware Prostate Cancer Detection Network on Multi-Parametric MRI

We discussed how to incorporate clinical anatomical knowledge into the deep learning architecture design for prostate cancer detection task, after recognizing the value of anatomical knowledge in aiding radiologists' decision-making. We proposed a Siamese-like architecture, incorporating a clinical prior into the network design that leverages symmetric-related visual appearance differences to distinguish cancerous tissue from other prostate tissues with similar visual patterns. We then explicitly integrated zonal appearance differences of prostate cancer on mpMRI into the loss design, treating lesions in different locations differently as indicated by the clinical guideline, thus better guiding model training. Our extensive experimental comparisons demonstrate that each of the proposed anatomical-aware designs contributes to the improvement of the prostate cancer detection and patient-level classification performance, and the integration of both designs can achieve the best model performance on both tasks.

6.2 Conclusions and Future Work

Our explorations have demonstrated the significant enhancement in model efficacy achieved through the integration of multi-modal information, particularly imaging and clinical insights, into learning-based CAD models. Our approaches have proven to effectively improve the performance of learning-based CAD models compared to utilizing single modality information alone. By leveraging the complementary nature of imaging and clinical data, our findings underscore the importance of holistic model development approaches in medical imaging research. These advancements hold promise for enhancing diagnostic accuracy and may ultimately help improve patient care outcomes in real-world clinical practice.

There are several promising avenues for future research in the field. Firstly, while data collection remains crucial in the medical imaging domain, attention should be given

not only to gathering more data and establishing multi-center collaborations, but also to developing powerful CAD models with limited data access; for instance, better pre-training or more advanced zero-shot/few-shot learning approaches. Moreover, there is ongoing discussion on how more effectively to utilize clinical data. In addition to feeding raw clinical numbers into deep learning networks alongside extracted imaging features, the integration of clinical data with imaging features using mechanisms such as cross-attention may prove beneficial as well. Additionally, our research has primarily focused on implementations related to MRI images. We believe there remains untapped potential for exploring how best to incorporate other medical imaging modalities. Furthermore, the recent advancements in large language models present an opportunity to extract clinical observations from medical records and integrate them into CAD model designs alongside clinical data and medical images, thus creating more comprehensive multi-modality learning-based CAD models. Additionally, most current medical image analysis or CAD model research efforts are retrospective in nature. As the environment for CAD model development matures, it will become increasingly important to conduct prospective studies. These studies offer insights into how CAD models perform in live clinical environments, aiding in the assessment of real-world effectiveness, usability, and impact on clinical decision-making processes. These promising opportunities for further exploration can be the focus of future research endeavors.

Aamir, S., Rahim, A., Aamir, Z., Abbasi, S. F., Khan, M. S., Alhaisoni, M., Khan, M. A., Khan, K., and Ahmad, J. (2022). Predicting breast cancer leveraging supervised machine learning techniques. *Computational and Mathematical Methods in Medicine*, 2022. 46

Algohary, A., Viswanath, S., Shiradkar, R., Ghose, S., Pahwa, S., Moses, D., Jambor, I., Shnier, R., Böhm, M., Haynes, A.-M., et al. (2018). Radiomic features on mri enable risk categorization of prostate cancer patients on active surveillance: Preliminary findings. *Journal of Magnetic Resonance Imaging*, 48(3):818–828. 9, 21, 40

Amin, M. B., Greene, F. L., Edge, S. B., Compton, C. C., Gershenwald, J. E., Brookland, R. K., Meyer, L., Gress, D. M., Byrd, D. R., and Winchester, D. P. (2017). The eighth edition ajcc cancer staging manual: continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA: a cancer journal for clinicians*, 67(2):93–99. 25

Barbosa, F. d. G., Queiroz, M. A., Nunes, R. F., Marin, J. F. G., Buchpiguel, C. A., and Cerri, G. G. (2018). Clinical perspectives of psma pet/mri for prostate cancer. *Clinics*, 73. 23, 37

Barentsz, J. O., Richenberg, J., Clements, R., Choyke, P., Verma, S., Villeirs, G., Rouviere, O., Logager, V., Fütterer, J. J., and Radiology, E. S. o. U. (2012). ESUR prostate MR guidelines 2012. *European Radiologuy*, 22(4):746–757. 55, 57

Bertinetto, L., Mueller, R., Tertikas, K., Samangooei, S., and Lord, N. A. (2020). Making better mistakes: Leveraging class hierarchies with deep networks. In *Conference on Computer Vision and Pattern Recognition (CVPR)*. 55, 66

Bhalerao, M. and Thakur, S. (2020). Brain tumor segmentation based on 3d residual u-net. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 218–225. 69, 70, 72, 73

Bhattacharya, I., Khandwala, Y. S., Vesal, S., Shao, W., Yang, Q., Soerensen, S. J., Fan, R. E., Ghanouni, P., Kunder, C. A., Brooks, J. D., Hu, Y., Rusu, M., and Sonn, G. A. (2022). A review of artificial intelligence in prostate cancer detection on imaging. *Therapeutic Advances in Urology*, 14. 58

Booven, D. J. V., Kuchakulla, M., Pai, R., Frech, F. S., Ramasahayam, R., Reddy, P., Parmar, M., Ramasamy, R., and Arora, H. (2021). A Systematic Review of Artificial Intelligence in Prostate Cancer. *Research and Reports in Urology*, 13:31–39. 58

Borofsky, S., George, A. K., Gaur, S., Bernardo, M., Greer, M. D., Mertan, F. V., Taffel, M., Moreno, V., Merino, M. J., Wood, B. J., Pinto, P. A., Choyke, P. L., and Turkbey, B. (2018). What are we missing? false-negative cancers at multiparametric mr imaging of the prostate. *Radiology*, 286(1):186–195. 62, 76

Brancato, V., Di Costanzo, G., Basso, L., Tramontano, L., Puglia, M., Ragozzino, A., and Cavaliere, C. (2020). Assessment of dce utility for pca diagnosis using pi-rads v2.1: Effects on diagnostic accuracy and reproducibility. *Diagnostics*, 10(3). 61

Briganti, A., Larcher, A., Abdollah, F., Capitanio, U., Gallina, A., Suardi, N., Bianchi, M., Sun, M., Freschi, M., Salonia, A., et al. (2012). Updated nomogram predicting lymph node invasion in patients with prostate cancer undergoing extended pelvic lymph node dissection: the essential importance of percentage of positive cores. *European urology*, 61(3):480–487. 23, 24, 25, 33, 34, 36

Brizmohun Appayya, M., Adshead, J., Ahmed, H. U., Allen, C., Bainbridge, A., Barrett, T., Giganti, F., Graham, J., Haslam, P., Johnston, E. W., et al. (2018). National implementation of multi-parametric magnetic resonance imaging for prostate cancer detection–recommendations from a uk consensus meeting. *BJU international*, 122(1):13–25. 8

Cao, R., Bajgiran, A. M., Mirak, S. A., Shakeri, S., Zhong, X., Enzmann, D., Raman, S., and Sung, K. (2019a). Joint prostate cancer detection and gleason score prediction in mp-mri via focalnet. *IEEE transactions on medical imaging*, 38(11):2496–2506. 11, 26, 41, 46, 47, 51

Cao, R., Mohammadian Bajgiran, A., Afshari Mirak, S., Shakeri, S., Zhong, X., Enzmann, D., Raman, S., and Sung, K. (2019b). Joint prostate cancer detection and gleason score prediction in mp-mri via focalnet. *IEEE Transactions on Medical Imaging (TMI)*, 38(11):2496–2506. 55, 56, 58, 63, 64, 65, 67, 69, 78

Cao, R., Zhong, X., Afshari, S., Felker, E., Suvannarerg, V., Tubtawee, T., Vangala, S., Scalzo, F., Raman, S., and Sung, K. (2021). Performance of Deep Learning and Genitourinary Radiologists in Detection of Prostate Cancer Using 3-T Multiparametric Magnetic Resonance Imaging. *Journal of Magnetic Resonance Imaging*, 54(2):474–483. 55, 56, 58, 65, 78

Cao, R., Zhong, X., Shakeri, S., Bajgiran, A. M., Mirak, S. A., Enzmann, D., Raman, S. S., and Sung, K. (2019c). Prostate cancer detection and segmentation in multi-parametric mri via cnn and conditional random field. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 1900–1904. IEEE. 11

Chen, C., Wang, Y., Niu, J., Liu, X., Li, Q., and Gong, X. (2021). Domain knowledge powered deep learning for breast cancer diagnosis based on contrast-enhanced ultrasound videos. *IEEE Transactions on Medical Imaging (TMI)*, 40(9):2439–2451. 69, 70, 72, 73

Chen, H., Dou, Q., Yu, L., Qin, J., and Heng, P.-A. (2018). Voxresnet: Deep voxelwise residual networks for brain segmentation from 3d mr images. *NeuroImage*, 170:446–455. 69, 70, 72, 73

Chen, J., Wang, Z., Zhao, J., Zhu, S., Sun, G., Liu, J., Zhang, H., Zhang, X., Shen, P., Shi, M., et al. (2019). Pelvic lymph node dissection and its extent on survival benefit

in prostate cancer patients with a risk of lymph node invasion¿ 5%: a propensity score matching analysis from seer database. *Scientific Reports*, 9(1):17985. 23

Chicco, D. (2021). *Siamese Neural Networks: An Overview*, pages 73–94. Springer US, New York, NY. 64

Conlin, C. C., Karunamuni, R., Hussain, T. S., Zhong, A. Y., Kallis, K., Do, D. D., Lui, A. J., Mani, G., Ollison, C., Domingo, M. R., Shabaik, A., Kane, C., Bagrodia, A., Mckay, R. R., Kuperman, J. M., Rakow-Penner, R., Hahn, M. E., Dale, A. M., and Seibert, T. M. (2023). Background prostate tissue is quantitatively abnormal on t2-weighted mri in patients with clinically significant prostate cancer. In *Anuual Meeting of The International Society for Magnetic Resonance in Medicine (ISMRM)*. 41

Cuocolo, R., Stanzione, A., Faletti, R., Gatti, M., Calleris, G., Fornari, A., Gentile, F., Motta, A., Dell'Aversana, S., Creta, M., et al. (2021). Mri index lesion radiomics and machine learning for detection of extraprostatic extension of disease: a multicenter study. *European radiology*, pages 1–9. 24, 36

Cysouw, M. C., Jansen, B. H., van de Brug, T., Oprea-Lager, D. E., Pfaehler, E., de Vries, B. M., van Moorselaar, R. J., Hoekstra, O. S., Vis, A. N., and Boellaard, R. (2021). Machine learning-based analysis of [18 f] dcfpyl pet radiomics for risk stratification in primary prostate cancer. *European journal of nuclear medicine and molecular imaging*, 48:340–349. 23, 37

De Meerleer, G., Berghen, C., Briganti, A., Vulsteke, C., Murray, J., Joniau, S., Leliveld, A. M., Cozzarini, C., Decaestecker, K., Rans, K., et al. (2021). Elective nodal radiotherapy in prostate cancer. *The Lancet Oncology*, 22(8):e348–e357. 37

De Vente, C., Vos, P., Hosseinzadeh, M., Pluim, J., and Veta, M. (2020). Deep learning regression for prostate cancer detection and grading in bi-parametric mri. *IEEE Transactions on Biomedical Engineering*, 68(2):374–383. 44

DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, pages 837–845. 14, 29, 46

Distler, F. A., Radtke, J. P., Bonekamp, D., Kesch, C., Schlemmer, H.-P., Wieczorek, K., Kirchner, M., Pahernik, S., Hohenfellner, M., and Hadaschik, B. A. (2017). The value of psa density in combination with pi-rads™ for the accuracy of prostate cancer prediction. *The Journal of urology*, 198(3):575–582. 8, 9, 10, 14, 17, 18, 19, 21, 40, 51, 52

Dong, C., Dai, D., Zhang, Y., Zhang, C., Li, Z., and Xu, S. (2023). Learning from dermoscopic images in association with clinical metadata for skin lesion segmentation and classification. *Computers in Biology and Medicine*, 152:106321. 78

Duran, A., Dussert, G., Rouvière, O., Jaouen, T., Jodoin, P.-M., and Lartizien, C. (2022). Prostattention-net: A deep attention model for prostate cancer segmentation by aggressiveness in mri scans. *Medical Image Analysis*, 77:102347. 59

ER Delong, DM Delong, D. C.-P. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 44. 69

Fan, D.-P., Ji, G.-P., Sun, G., Cheng, M.-M., Shen, J., and Shao, L. (2020). Camouflaged object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2777–2787. 41, 52

Fleiss, J. L., Levin, B., and Paik, M. C. (1981). *Statistical methods for rates and proportions*, volume 1981218. New York: Wiley. 38

Fossati, N., Willemse, P.-P. M., Van den Broeck, T., van den Bergh, R. C., Yuan, C. Y., Briers, E., Bellmunt, J., Bolla, M., Cornford, P., De Santis, M., et al. (2017). The benefits and harms of different extents of lymph node dissection during radical prostatectomy for prostate cancer: a systematic review. *European urology*, 72(1):84–109. 23

Gillies, R. J., Kinahan, P. E., and Hricak, H. (2016). Radiomics: images are more than pictures, they are data. *Radiology*, 278(2):563–577. 9, 24

Gnep, K., Fargeas, A., Gutiérrez-Carvajal, R. E., Commandeur, F., Mathieu, R., Ospina, J. D., Rolland, Y., Rohou, T., Vincendeau, S., Hatt, M., et al. (2017). Haralick textural features on t2-weighted mri are associated with biochemical recurrence following radiotherapy for peripheral zone prostate cancer. *Journal of Magnetic Resonance Imaging*, 45(1):103–117. 9, 21, 37

Gugliandolo, S. G., Pepa, M., Isaksson, L. J., Marvaso, G., Raimondi, S., Botta, F., Gandini, S., Ciardo, D., Volpe, S., Riva, G., et al. (2021). Mri-based radiomics signature for localized prostate cancer: a new clinical tool for cancer aggressiveness prediction? sub-study of prospective phase ii trial on ultra-hypofractionated radiotherapy (airc ig-13218). *European Radiology*, 31:716–728. 24, 36

Hansen, N. L., Barrett, T., Kesch, C., Pepdjonovic, L., Bonekamp, D., O'Sullivan, R., Distler, F., Warren, A., Samel, C., Hadaschik, B., et al. (2018). Multicentre evaluation of magnetic resonance imaging supported transperineal prostate biopsy in biopsy-naïve men with suspicion of prostate cancer. *BJU international*, 122(1):40–49. 8, 9, 10, 14, 17, 18, 20, 21, 40, 51, 52

Haralick, R. M., Shanmugam, K., and Dinstein, I. H. (1973). Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, pages 610–621. 45

Hatamizadeh, A., Yang, D., Roth, H. R., and Xu, D. (2022). Unetr: Transformers for 3d medical image segmentation. *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1748–1758. 69, 70, 72, 73

Hatano, K., Tanaka, J., Nakai, Y., Nakayama, M., Kakimoto, K.-i., Nakanishi, K., and Nishimura, K. (2020). Utility of index lesion volume assessed by multiparametric mri combined with gleason grade for assessment of lymph node involvement in patients with high-risk prostate cancer. *Japanese Journal of Clinical Oncology*, 50(3):333–337. 24, 25, 36

Hawkins, D. M., Basak, S. C., and Mills, D. (2003). Assessing model fit by cross-validation. *Journal of chemical information and computer sciences*, 43(2):579–586. 12, 21

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778. 44

Hectors, S. J., Cherny, M., Yadav, K. K., Beksaç, A. T., Thulasidass, H., Lewis, S., Davicioni, E., Wang, P., Tewari, A. K., and Taouli, B. (2019). Radiomics features measured with multiparametric magnetic resonance imaging predict prostate cancer aggressiveness. *The Journal of urology*, 202(3):498–505. 12, 24, 27, 36

Hosseinzadeh, M., Saha, A., Brand, P., Slootweg, I., de Rooij, M., and Huisman, H. (2022). Deep learning–assisted prostate cancer detection on bi-parametric mri: minimum training data size requirements and effect of prior knowledge. *European Radiology*, pages 1–11. 41, 44, 51, 54, 55, 58, 65, 70, 75, 76, 78

Hu, J., Shen, L., and Sun, G. (2018). Squeeze-and-excitation networks. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7132–7141. 59, 69, 70, 72, 73

Huang, C., Song, G., Wang, H., Lin, Z., Wang, H., Ji, G., Zhang, S., Guo, Y., Li, J., Bao, Z., et al. (2020a). Preoperative pi-rads version 2 scores helps improve accuracy of clinical nomograms for predicting pelvic lymph node metastasis at radical prostatectomy. *Prostate Cancer and Prostatic Diseases*, 23(1):116–126. 24, 25, 36

Huang, S.-C., Pareek, A., Seyyedi, S., Banerjee, I., and Lungren, M. P. (2020b). Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. *NPJ digital medicine*, 3(1):136. 78

Huang, Y.-q., Liang, C.-h., He, L., Tian, J., Liang, C.-s., Chen, X., Ma, Z.-l., and Liu, Z.-y. (2016). Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *Journal of clinical oncology*, 34(18):2157–2164. 24, 36

Hung, A. L. Y., Zheng, H., Miao, Q., Raman, S. S., Terzopoulos, D., and Sung, K. (2022). Cat-net: A cross-slice attention transformer model for prostate zonal segmentation in mri. *IEEE transactions on medical imaging*, 42(1):291–303. 44, 63, 72

Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P. F., Kohl, S., Wasserthal, J., Koehler, G., Norajitra, T., Wirkert, S., et al. (2021). nnu-net: a self-configuring

method for deep learning-based biomedical image segmentation. *Nature Methods*. 60, 64, 68, 69

Kamal, U., Zunaed, M., Nizam, N. B., and Hasan, T. (2022). Anatomy-XNet: An Anatomy Aware Convolutional Neural Network for Thoracic Disease Classification in Chest X-Rays. *IEEE Journal of Biomedical and Health Informatics (JBHI)*, 26(11):5518–5528. 59, 60

Kingma, D. P. and Ba, J. (2015). Adam: A method for stochastic optimization. *CoRR*. 68

Kozlowski, P., Chang, S., Jones, E., and Goldenberg, S. (2018). Assessment of the need for dce mri in the detection of dominant lesions in the whole gland: Correlation between histology and mri of prostate cancer. *NMR in Biomedicine*, 31(3). 61

Lamb, B. W., Tan, W. S., Rehman, A., Nessa, A., Cohen, D., O'Neil, J., Green, J. S., and Hines, J. E. (2015). Is Prebiopsy MRI Good Enough to Avoid Prostate Biopsy? A Cohort Study Over a 1-Year Period. *Clinical Genitourinary Cancer*, 13(6):512–517. 62, 76

Lambin, P., Leijenaar, R. T., Deist, T. M., Peerlings, J., De Jong, E. E., Van Timmeren, J., Sanduleanu, S., Larue, R. T., Even, A. J., Jochems, A., et al. (2017). Radiomics: the bridge between medical imaging and personalized medicine. *Nature reviews Clinical oncology*, 14(12):749–762. 9, 24

Li, L., Zhou, T., Wang, W., Li, J., and Yang, Y. (2022). Deep hierarchical semantic segmentation. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1236–1247. 55, 66

Li, M., Zhang, J., Dan, Y., Yao, Y., Dai, W., Cai, G., Yang, G., and Tong, T. (2020). A clinical-radiomics nomogram for the preoperative prediction of lymph node metastasis in colorectal cancer. *Journal of translational medicine*, 18(1):1–10. 24, 36

Li, Y., Wu, Y., Huang, M., Zhang, Y., and Bai, Z. (2023). Attention-guided multi-scale learning network for automatic prostate and tumor segmentation on mri. *Computers in Biology and Medicine*, 165:107374. 58, 78

Liechti, M. R., Muehlematter, U. J., Schneider, A. F., Eberli, D., Rupp, N. J., Hötker, A. M., Donati, O. F., and Becker, A. S. (2020). Manual prostate cancer segmentation in mri: interreader agreement and volumetric correlation with transperineal template core needle biopsy. *European radiology*, 30:4806–4815. 38

Lin, T., Goyal, P., Girshick, R., He, K., and Dollar, P. (2020). Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 42(02):318–327. 67, 68

Liss, M. A., Ehdaie, B., Loeb, S., Meng, M. V., Raman, J. D., Spears, V., and Stroup, S. P. (2017). An update of the american urological association white paper on the

prevention and treatment of the more common complications related to prostate biopsy. *The Journal of urology*, 198(2):329–334. 8

Litjens, G., Debats, O., Barentsz, J., Karssemeijer, N., and Huisman, H. (2014). Computer-aided detection of prostate cancer in mri. *IEEE Transactions on Medical Imaging (TMI)*, 33(5):1083–1092. 58, 68, 69

Liu, P., Wang, S., Turkbey, B., Grant, K., Pinto, P., Choyke, P., Wood, B. J., and Summers, R. M. (2013). A prostate cancer computer-aided diagnosis system using multimodal magnetic resonance imaging and targeted biopsy labels. In *Medical Imaging 2013: Computer-Aided Diagnosis*, volume 8670, pages 369–374. SPIE. 11, 63, 70, 76

Liu, Y., Yu, Y., Ouyang, J., Jiang, B., Yang, G., Ostmeier, S., Wintermark, M., Michel, P., Liebeskind, D. S., Lansberg, M. G., Albers, G. W., and Zaharchuk, G. (2023). Functional outcome prediction in acute ischemic stroke using a fused imaging and clinical deep learning model. *Stroke*, 24:2316. 78

Liu, Y., Zheng, H., Liang, Z., Miao, Q., Brisbane, W. G., Marks, L. S., Raman, S. S., Reiter, R. E., Yang, G., and Sung, K. (2021). Textured-based deep learning in prostate cancer classification with 3t multiparametric mri: comparison with pi-rads-based classification. *Diagnostics*, 11(10):1785. 41, 51

Ma, D., Lu, D., Chen, S., Heisler, M., Dabiri, S., Lee, S., Lee, H., Ding, G. W., Sarunic, M. V., and Beg, M. F. (2021). LF-UNet – A novel anatomical-aware dual-branch cascaded deep neural network for segmentation of retinal layers and fluid from optical coherence tomography images. *Computerized Medical Imaging and Graphics*, 94:101988. 59, 60

Matoso, A. and Epstein, J. I. (2019). Defining clinically significant prostate cancer on the basis of pathological findings. *Histopathology*, 74(1):135–145. 10, 43

Moldovan, P. C., Van den Broeck, T., Sylvester, R., Marconi, L., Bellmunt, J., van den Bergh, R. C., Bolla, M., Briers, E., Cumberbatch, M. G., Fossati, N., et al. (2017). What is the negative predictive value of multiparametric magnetic resonance imaging in excluding prostate cancer at biopsy? a systematic review and meta-analysis from the european association of urology prostate cancer guidelines panel. *European urology*, 72(2):250–266. 19, 51

Morris, K. A. and Haboubi, N. Y. (2015). Pelvic radiation therapy: Between delight and disaster. *World journal of gastrointestinal surgery*, 7(11):279. 37

Mottet, N., van den Bergh, R. C., Briers, E., Van den Broeck, T., Cumberbatch, M. G., De Santis, M., Fanti, S., Fossati, N., Gandaglia, G., Gillessen, S., et al. (2021). Eau-eanm-estro-esur-siog guidelines on prostate cancer—2020 update. part 1: screening, diagnosis, and local treatment with curative intent. *European urology*, 79(2):243–262. 23, 29, 33

MSKCC. Memorial Sloan Kettering Cancer Center Dynamic, Prostate Cancer Nomogram: Coefficients. www.mskcc.org/nomograms/prostate/pre-op/coefficients. Available via the MSKCC website. 23, 24, 25, 33, 34, 36

Nketiah, G., Elschot, M., Kim, E., Teruel, J. R., Scheenen, T. W., Bathen, T. F., and Selnæs, K. M. (2017). T2-weighted mri-derived textural features reflect prostate cancer aggressiveness: preliminary results. *European radiology*, 27:3050–3059. 9, 21, 40

Nonn, L., Ananthanarayanan, V., and Gann, P. H. (2009). Evidence for field cancerization of the prostate. *The Prostate*, 69(13):1470–1479. 41

Oishi, M., Shin, T., Ohe, C., Nassiri, N., Palmer, S. L., Aron, M., Ashrafi, A. N., Cacciamani, G. E., Chen, F., Duddalwar, V., et al. (2019). Which patients with negative magnetic resonance imaging can safely avoid biopsy for prostate cancer? *The Journal of urology*, 201(2):268–277. 8, 9, 10, 14, 17, 18, 19, 20, 21, 40, 51, 52

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B., and Rueckert, D. (2018). Attention u-net: Learning where to look for the pancreas. In *Medical Imaging with Deep Learning*. 69, 70, 72, 73

Otti, V. C., Miller, C., Powell, R. J., Thomas, R. M., and McGrath, J. S. (2019). The diagnostic accuracy of multiparametric magnetic resonance imaging before biopsy in the detection of prostate cancer. *BJU international*, 123(1):82–90. 8, 9, 10, 14, 17, 18, 20, 21, 40, 51, 52

Pagniez, M., Kasivisvanathan, V., Puech, P., Drumez, E., Villers, A., and Olivier, J. (2020). Predictive factors of missed clinically significant prostate cancers in men with negative magnetic resonance imaging: a systematic review and meta-analysis. *The Journal of urology*, 204(1):24–32. 8, 19, 20, 40, 51

Panebianco, V., Barchetti, F., Barentsz, J., Ciardi, A., Cornud, F., Futterer, J., and Villeirs, G. (2015). Pitfalls in Interpreting mp-MRI of the Prostate: A Pictorial Review with Pathologic Correlation. *Insights into Imaging*, 6(6):611–630. 55, 56

Pei, C., Sun, Y., Zhu, J., Wang, X., Zhang, Y., Zhang, S., Yao, Z., and Lu, Q. (2020). Ensemble learning for early-response prediction of antidepressant treatment in major depressive disorder. 14, 21

Peng, Y., Jiang, Y., Yang, C., Brown, J. B., Antic, T., Sethi, I., Schmid-Tannwald, C., Giger, M. L., Eggener, S. E., and Oto, A. (2013). Quantitative analysis of multiparametric prostate mr images: differentiation between prostate cancer and normal tissue and correlation with gleason score—a computer-aided diagnosis development study. *Radiology*, 267(3):787–796. 12, 21

Penzkofer, T., Padhani, A. R., Turkbey, B., and Ahmed, H. U. (2022). Assessing the clinical performance of artificial intelligence software for prostate cancer detection on mri. *European Radiology*, 32(4):2221–2223. 41, 51

Rawla, P. (2019). Epidemiology of Prostate Cancer. *World Journal of Oncology*, 10(2):63–89. 54

Reda, I., Khalil, A., Elmogy, M., Abou El-Fetouh, A., Shalaby, A., Abou El-Ghar, M., Elmaghraby, A., Ghazal, M., and El-Baz, A. (2018). Deep learning role in early diagnosis of prostate cancer. *Technology in cancer research & treatment*, 17:1533034618775530. 78

Ren, J., Hu, X., Zhu, L., Xu, X., Xu, Y., Wang, W., Deng, Z., and Heng, P.-A. (2021). Deep texture-aware features for camouflaged object detection. *IEEE Transactions on Circuits and Systems for Video Technology*. 41, 44, 45, 52

Roach, M. r., Marquez, C., Yuo, H. S., Narayan, P., Coleman, L., Nseyo, U. O., Navvab, Z., and Carroll, P. R. (1994). Predicting the risk of lymph node involvement using the pre-treatment prostate specific antigen and gleason score in men with clinically localized prostate cancer. *International journal of radiation oncology, biology, physics*. 23, 24, 25, 33, 34, 36

Ruopp, M. D., Perkins, N. J., Whitcomb, B. W., and Schisterman, E. F. (2008). Youden index and optimal cut-point estimated from observations affected by a lower limit of detection. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 50(3):419–430. 14, 46

Saha, A., Hosseinzadeh, M., and Huisman, H. (2021). End-to-end prostate cancer detection in bpMRI via 3D CNNs: Effects of attention mechanisms, clinical priori and decoupled false positive reduction. *Medical Image Analysis*, 73:102155. 55, 56, 58, 59, 65, 78

Sathianathen, N. J., Konety, B. R., Soubra, A., Metzger, G. J., Spilseth, B., Murugan, P., Weight, C. J., Ordonez, M. A., and Warlick, C. A. (2018). Which scores need a core? an evaluation of mr-targeted biopsy yield by pirads score across different biopsy indications. *Prostate cancer and prostatic diseases*, 21(4):573–578. 8, 40

Schaeffer, E., Srinivas, S., Antonarakis, E. S., Armstrong, A. J., Bekelman, J. E., Cheng, H., D'Amico, A. V., Davis, B. J., Desai, N., Dorff, T., et al. (2021). Nccn guidelines insights: Prostate cancer, version 1.2021: Featured updates to the nccn guidelines. *Journal of the National Comprehensive Cancer Network*, 19(2):134–143. 23, 29, 33, 34

Schoots, I. G., Roobol, M. J., Nieboer, D., Bangma, C. H., Steyerberg, E. W., and Hunink, M. M. (2015). Magnetic resonance imaging–targeted biopsy may enhance the diagnostic accuracy of significant prostate cancer detection compared to standard transrectal ultrasound-guided biopsy: a systematic review and meta-analysis. *European urology*, 68(3):438–450. 8, 40

Seetharaman, A., Bhattacharya, I., Chen, L. C., Kunder, C. A., Shao, W., Soerensen, S. J. C., Wang, J. B., Teslovich, N. C., Fan, R. E., Ghanouni, P., Brooks, J. D., Too, K. J., Sonn, G. A., and Rusu, M. (2021). Automated detection of aggressive and indolent

prostate cancer on magnetic resonance imaging. *Medical Physics*, 48(6):2960–2972. 54, 55, 58, 65, 70, 76, 78

Smith, R., Mettlin, C., and H., E. (2003). Cancer screening and early detection. *Holland-Frei Cancer Medicine*, 31. 55, 57

Soeterik, T. F., Hueting, T. A., Israel, B., van Melick, H. H., Dijksman, L. M., Stomps, S., Biesma, D. H., Koffijberg, H., Sedelaar, M., Witjes, J. A., et al. (2021). External validation of the memorial sloan kettering cancer centre and briganti nomograms for the prediction of lymph node involvement of prostate cancer using clinical stage assessed by magnetic resonance imaging. *BJU international*, 128(2):236–243. 23, 24, 36, 38

Sonn, G. A., Natarajan, S., Margolis, D. J., MacAiran, M., Lieu, P., Huang, J., Dorey, F. J., and Marks, L. S. (2013). Targeted biopsy in the detection of prostate cancer using an office based magnetic resonance ultrasound fusion device. *The Journal of urology*, 189(1):86–92. 10, 42

Springenberg, J. T., Dosovitskiy, A., Brox, T., and Riedmiller, M. (2014). Striving for simplicity: The all convolutional net. *arXiv preprint arXiv:1412.6806*. 46

Sprute, K., Kramer, V., Koerber, S. A., Meneses, M., Fernandez, R., Soza-Ried, C., Eiber, M., Weber, W. A., Rauscher, I., Rahbar, K., et al. (2021). Diagnostic accuracy of 18f-psma-1007 pet/ct imaging for lymph node staging of prostate carcinoma in primary and biochemical recurrence. *Journal of Nuclear Medicine*, 62(2):208–213. 23, 37

Starmans, M. P., Niessen, W. J., Schoots, I., Klein, S., Veenland, J. F., et al. (2019). Classification of prostate cancer: High grade versus low grade using a radiomics approach. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 1319–1322. IEEE. 9, 21

Sufyan, M., Shokat, Z., and Ashfaq, U. A. (2023). Artificial intelligence in cancer diagnosis and therapy: Current status and future perspective. *Computers in Biology and Medicine*, 165:107356. 58

Sun, Y. and Ji, Y. (2021). AAWS-Net: Anatomy-aware weakly-supervised learning network for breast mass segmentation. *PLoS ONE*, 16(8):e0256830. 59

Sutton, E. J., Dashevsky, B. Z., Oh, J. H., Veeraraghavan, H., Apte, A. P., Thakur, S. B., Morris, E. A., and Deasy, J. O. (2016). Breast cancer molecular subtype classifier that incorporates mri features. *Journal of Magnetic Resonance Imaging*, 44(1):122–129. 12, 21

Tan, N., Pollock, J. R., Margolis, D. J., Padhani, A. R., Tempany, C., Woo, S., and Gorin, M. A. (2023). Management of patients with a negative multiparametric prostate mri examination: Ajr expert panel narrative review. *American Journal of Roentgenology*. 40, 51

Tomaszewski, M. R. and Gillies, R. J. (2021). The biological meaning of radiomic features. *Radiology*, 298(3):505–516. 24

Tripepi, G., Jager, K. J., Dekker, F. W., and Zoccali, C. (2010). Selection bias and information bias in clinical research. *Nephron Clinical Practice*, 115(2):c94–c99. 10, 26

Turkbey, B., Rosenkrantz, A. B., Haider, M. A., Padhani, A. R., Villeirs, G., Macura, K. J., Tempany, C. M., Choyke, P. L., Cornud, F., Margolis, D. J., et al. (2019). Prostate imaging reporting and data system version 2.1: 2019 update of prostate imaging reporting and data system version 2. *European urology*, 76(3):340–351. 8, 11, 25, 40, 44, 47, 54, 55, 58, 59, 60, 61, 62, 64, 65, 66, 67, 78

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., and Gee, J. C. (2010). N4itk: improved n3 bias correction. *IEEE transactions on medical imaging*, 29(6):1310–1320. 12, 27, 46, 63

Van Griethuysen, J. J., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R. G., Fillion-Robin, J.-C., Pieper, S., and Aerts, H. J. (2017). Computational radiomics system to decode the radiographic phenotype. *Cancer research*, 77(21):e104–e107. 12, 28

Vargas, H. A., Akin, O., Franiel, T., Mazaheri, Y., Zheng, J., Moskowitz, C., Udo, K., Eastham, J., and Hricak, H. (2011). Diffusion-weighted endorectal mr imaging at 3 t for prostate cancer: tumor detection and assessment of aggressiveness. *Radiology*, 259(3):775–784. 47, 64

Varghese, B., Chen, F., Hwang, D., Palmer, S. L., De Castro Abreu, A. L., Ukimura, O., Aron, M., Aron, M., Gill, I., Duddalwar, V., et al. (2020). Objective risk stratification of prostate cancer using machine learning and radiomics applied to multiparametric magnetic resonance images. In *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, pages 1–10. 9, 21

Venclovas, Z., Muilwijk, T., Matjosaitis, A. J., Jievaltas, M., Joniau, S., and Milonas, D. (2021). Head-to-head comparison of two nomograms predicting probability of lymph node invasion in prostate cancer and the therapeutic impact of higher nomogram threshold. *Journal of Clinical Medicine*, 10(5):999. 23, 24, 36

Vente, C. d., Vos, P., Hosseinzadeh, M., Pluim, J., and Veta, M. (2020). Deep Learning Regression for Prostate Cancer Detection and Grading in Bi-Parametric MRI. *IEEE Transactions on Biomedical Engineering (TBME)*, 68(2):374–383. 54, 56, 58, 65, 75, 78

Verma, N., Mahajan, D., Sellamanickam, S., and Nair, V. (2012). Learning hierarchical similarity metrics. In *Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2280–2287. 55, 66

Vignati, A., Mazzetti, S., Giannini, V., Russo, F., Bollito, E., Porpiglia, F., Stasi, M., and Regge, D. (2015). Texture features on t2-weighted magnetic resonance imaging: new potential biomarkers for prostate cancer aggressiveness. *Physics in Medicine & Biology*, 60(7):2685. 9, 20

Wagemans, J. (1997). Characteristics and models of human symmetry detection. *Trends in Cognitive Sciences*, 1(9):346–352. 56

Washino, S., Okochi, T., Saito, K., Konishi, T., Hirai, M., Kobayashi, Y., and Miyagawa, T. (2017). Combination of prostate imaging reporting and data system (pi-rads) score and prostate-specific antigen (psa) density predicts biopsy outcome in prostate biopsy naïve patients. *BJU international*, 119(2):225–233. 77

Weinreb, J. C., Barentsz, J. O., Choyke, P. L., Cornud, F., Haider, M. A., Macura, K. J., Margolis, D., Schnall, M. D., Shtern, F., Tempany, C. M., et al. (2016). Pi-rads prostate imaging–reporting and data system: 2015, version 2. *European urology*, 69(1):16–40. 10, 23, 24, 38, 44, 54, 55, 59, 60, 61, 62, 64, 65, 66, 67, 78

Wilczak, W., Wittmer, C., Clauditz, T., Minner, S., Steurer, S., Büscheck, F., Krech, T., Lennartz, M., Harms, L., Leleu, D., et al. (2018). Marked prognostic impact of minimal lymphatic tumor spread in prostate cancer. *European urology*, 74(3):376–386. 23

Yan, C., Peng, Y., and Li, X. (2019). Radiomics analysis for prostate cancer classification in multiparametric magnetic resonance images. In *BIBE 2019; The Third International Conference on Biological Information and Biomedical Engineering*, pages 1–4. VDE. 9, 11, 20, 24, 26, 36, 37, 40

Yu, J., Fulcher, A. S., Turner, M. A., Cockrell, C. H., Cote, E. P., and Wallace, T. J. (2014). Prostate cancer and its mimics at multiparametric prostate MRI. *The British Journal of Radiology*, 87(1037):20130659. 55, 56

Yu, J. B., Makarov, D. V., and Gross, C. (2011). A new formula for prostate cancer lymph node risk. *International journal of radiation oncology, biology, physics*. 23, 24, 25, 33, 34, 36

Yu, X., Lou, B., Shi, B., Winkel, D., Arrahmane, N., Diallo, M., Meng, T., Busch, H. v., Grimm, R., Kiefer, B., Comaniciu, D., IEEE International Symposium on Biomedical Imaging (ISBI)Kamen, A., Huisman, H., Rosenkrantz, A., Penzkofer, T., Shabunin, I., Choi, M. H., Yang, Q., and Szolar, D. (2020). False Positive Reduction Using Multiscale Contextual Features for Prostate Cancer Detection in Multi-Parametric MRI Scans. *IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1355–1359. 54, 56, 58, 59, 65, 78

Yusim, I., Krenawi, M., Mazor, E., Novack, V., and Mabjeesh, N. J. (2020). The use of prostate specific antigen density to predict clinically significant prostate cancer. *Scientific reports*, 10(1):20015. 77

Zamboglou, C., Carles, M., Fechter, T., Kiefer, S., Reichel, K., Fassbender, T. F., Bronsert, P., Koeber, G., Schilling, O., Ruf, J., et al. (2019). Radiomic features from psma pet for non-invasive intraprostatic tumor discrimination and characterization in patients with intermediate-and high-risk prostate cancer-a comparison study with histology reference. *Theranostics*, 9(9):2595. 23, 37

Zeiler, M. D. and Fergus, R. (2014). Visualizing and understanding convolutional networks. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13*, pages 818–833. Springer. 46

Zhang, G.-m.-y., Han, Y.-q., Wei, J.-w., Qi, Y.-f., Gu, D.-s., Lei, J., Yan, W.-g., Xiao, Y., Xue, H.-d., Feng, F., et al. (2020). Radiomics based on mri as a biomarker to guide therapy by predicting upgrading of prostate cancer from biopsy to radical prostatectomy. *Journal of Magnetic Resonance Imaging*, 52(4):1239–1248. 9, 21, 24, 36, 37, 41

Zheng, H., Hung, A. L. Y., Miao, Q., Song, W., Scalzo, F., Raman, S. S., Zhao, K., and Sung, K. (2024a). Atpca-net: anatomical-aware prostate cancer detection network on multi-parametric mri. *Scientific Reports*, 14(1):5740. 6

Zheng, H., Miao, Q., Hung, A. L. Y., Zhao, K., Scalzo, F., Raman, S. S., and Sung, K. (2024b). Integrative feature-enhanced network predicting prostate biopsy results from negative multi-parametric mri. *Radiology: Imaging Cancer*. Under review. 6

Zheng, H., Miao, Q., Liu, Y., Mirak, S. A., Hosseiny, M., Scalzo, F., Raman, S. S., and Sung, K. (2022a). Multiparametric mri-based radiomics model to predict pelvic lymph node invasion for patients with prostate cancer. *European radiology*, 32(8):5688–5699. 5, 78

Zheng, H., Miao, Q., Liu, Y., Raman, S. S., Scalzo, F., and Sung, K. (2022b). Integrative machine learning prediction of prostate biopsy results from negative multiparametric mri. *Journal of Magnetic Resonance Imaging*, 55(1):100–110. 5, 26, 36, 40, 46, 49, 50, 63

Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., and Liang, J. (2018). Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pages 3–11. 72

Zongker, D. and Jain, A. (1996). Algorithms for feature selection: An evaluation. In *Proceedings of 13th international conference on pattern recognition*, volume 2, pages 18–22. IEEE. 12, 29

Zwanenburg, A., Vallières, M., Abdalah, M. A., Aerts, H. J., Andrearczyk, V., Apte, A., Ashrafinia, S., Bakas, S., Beukinga, R. J., Boellaard, R., et al. (2020). The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*, 295(2):328–338. 9, 24, 28