PREOPERATIVE PREDICTION OF LATERAL CERVICAL LYMPH NODE METASTASIS IN PAPILLARY THYROID CARCINOMA USING CONVENTIONAL ULTRASOUND, SONAZOID CONTRAST ENHANCED ULTRASOUND, FINE NEEDLE ASPIRATION THYROGLOBULIN AND RADIOMICS

LIU YAN

UNIVERSITI KEBANGSAAN MALAYSIA

UNIVERSITI KEBANGSAAN MALAYSIA The National University of Malaysia	UKM-SPKPPP-PT(PdP)-05-AK04-BO07	No. Semakan: 02	Tarikh Kuat Kuasa: 27/01/2026
	BORANG DAN SENARAI SEMAK PENYERAHAN TESIS/DISERTASI SELEPAS PEMBETULAN ***FORM AND CHECKLIST OF AMENDED THESIS/DISSERTATION SUBMISSION***		

F. KELULUSAN DEKAN/PENGARAH / *APPROVAL FROM DEAN/DIRECTOR*

Tandatangan: *Signature*	Tarikh: *Date* 8/9/2026
Nama dan Cap Rasmi: *Name and Official Stamp*	PROFESOR DATO' DR MARINA MAT BAKI MD (UKM), MS ORL-HNS (UKM), PhD (UCL) Dekan Fakulti Perubatan Pakar Perunding Kanan ORL-HNS Universiti Kebangsaan Malaysia

G. PERAKUAN TESIS/ DISERTASI SARJANA / DOKTOR FALSAFAH
(CERTIFICATION OF MASTER'S / DOCTORAL THESIS / DISSERTATION)

Nama penuh pengarang (*Author's Full Name*)	LIU YAN		
No. Pendaftaran Pelajar (Student's Registration No.)	P142742	Sesi Akademik (*Academic Session*)	Semester2, Session2024/2025
Tajuk Tesis / Disertasi (*Thesis / Dissertation Title*)	PREOPERATIVE PREDICTION OF LATERAL CERVICAL LYMPH NODE METASTASIS IN PAPILLARY THYROID CARCINOMA USING CONVENTIONAL ULTRASOUND, SONAZOID CONTRAST ENHANCED ULTRASOUND, FINE NEEDLE ASPIRATION THYROGLOBULIN, AND RADIOMICS		

Semua tesis/disertasi dan penerbitan berkaitan hasil penyelidikan pelajar adalah tertakluk kepada Dasar Harta Intelek Universiti Kebangsaan Malaysia.
All theses/ dissertations and publications relating to the research work of a student are subject to the Intellectual Property Policy of Universiti Kebangsaan Malaysia.

Saya mengaku bahawa tahap akses tesis/disertasi ini sebagai (*Tandakan ✓ dalam kotak bagi maklumat yang berkaitan; tandakan satu (1) kotak di bawah sahaja):
I hereby declare the level of access for this thesis/dissertation as follows (tick ✓ the relevant box below; tick only one option).

Pilihan Tahap Akses *Access Level Selection*		Tafsiran Tahap Akses Tesis/Disertasi *Definition of Thesis/Dissertation Access Level*
☐	**RAHSIA** ***(CONFIDENTIAL)***	Mengandungi maklumat rahsia sepertimana yang diperuntukan bawah Akta Rahsia Rasmi 1972. *Contains classified information as stipulated under the Official Secrets Act 1972.*
☐	**TERHAD** ***(RESTRICTED)***	Mengandungi maklumat yang hanya boleh diakses oleh pihak yang mempunyai kebenaran atau keperluan tertentu untuk mengetahuinya yang ditentukan oleh organisasi/ badan di mana penyelidikan dijalankan. *Contains information that may only be accessed by parties with specific authorization*

UNIVERSITI KEBANGSAAN MALAYSIA *The National University of Malaysia*	UKM-SPKPPP-PT(PdP)-05-AK04-BO07	No. Semakan: 02	Tarikh Kuat Kuasa: 27/01/2026
	BORANG DAN SENARAI SEMAK PENYERAHAN TESIS/DISERTASI SELEPAS PEMBETULAN ***FORM AND CHECKLIST OF AMENDED THESIS/DISSERTATION SUBMISSION***		

		or a defined need to know, as determined by the organization or body where the research is conducted.
√	**AKSES TERBUKA** ***(OPEN ACCESS)***	Tesis/disertasi versi akhir dalam format PDF yang boleh diakses secara percuma secara dalam talian dan bebas dari isu hak cipta dan pelesenan *The final version of the thesis/dissertation in PDF format that is freely accessible online and free from copyright and licensing issues.* Kementerian Pendidikan Tinggi melalui surat bertarikh 27 Mac 2025 memutuskan bahawa kebenaran akses tesis / disertasi dalam borang serahan akses terbuka ditentukan oleh penyelia (bukan pelajar). *The Ministry of Higher Education, through a letter dated 27 March 2025, has decided that permission for thesis / dissertation access in the open access submission form shall be determined by the supervisor (not the student).*
	EMBARGO	Tempoh masa di mana akses kepada tesis/disertasi disekat (ditutup) selama dua (2) tahun atas sebab-sebab tertentu sebagai contoh kerahsiaan atau hak cipta. Tesis/disertasi akan berstatus akses terbuka selepas tamat tempoh *embargo*. *The period during which access to the thesis/dissertation is restricted for two (2) years due to specific reasons, such as confidentiality or copyright. The thesis/dissertation will be granted open-access status upon the expiry of the embargo period.*

PENGESAHAN PELAJAR ***(STUDENT VERIFICATION)***	**PENGESAHAN PENYELIA** ***(SUPERVISOR VERIFICATION)***
LIU YAN **TANDATANGAN PELAJAR** ***(STUDENT'S SIGNATURE)***	**TANDATANGAN PENYELIA / PENGERUSI JK SISWAZAH** ***(SUPERVISOR'S / CHAIRPERSON SUPERVISION COMMITTEE SIGNATURE)***
Nama Pelajar *(Student name)*: LIU YAN No. Pendaftaran Pelajar *(Student's Registration No.)*: P142742	Nama Penyelia / Pengerusi JK Siswazah : *(Supervisor's / Chairperson Supervision Committee Name)*: HANANI ABDUL MANAN
Tarikh: *(Date)* 19 August 2026	Tarikh: *(Date)* 25 Ogos 2026

PREOPERATIVE PREDICTION OF LATERAL CERVICAL LYMPH NODE METASTASIS IN PAPILLARY THYROID CARCINOMA USING CONVENTIONAL ULTRASOUND, SONAZOID CONTRAST ENHANCED ULTRASOUND, FINE NEEDLE ASPIRATION THYROGLOBULIN AND RADIOMICS

LIU YAN

THESIS SUBMITTED IN FULFILMENT FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF MEDICINE
UNIVERSITI KEBANGSAAN MALAYSIA
KUALA LUMPUR

2026

PRAOPERATIF RAMALAN METASTASIS NODUS LIMFA SERVIKAL LATERAL DALAM KARSINOMA PAPILARI TIROID MENGGUNAKAN ULTRABUNYI KONVENSIONAL, ULTRABUNYI BERKONTRAS SONAZOID, TIROGLOBULIN ASPIRASI JARUM HALUS DAN RADIOMIK

LIU YAN

TESIS YANG DIKEMUKAKAN UNTUK MEMPEROLEH IJAZAH
DOKTOR FALSAFAH

FAKULTI PERUBATAN
UNIVERSITI KEBANGSAAN MALAYSIA
KUALA LUMPUR

2026

DECLARATION

I hereby declare that the work in this thesis is my own except for quotations and summaries which have been duly acknowledged.

01 September 2026

LIUYAN
P142742

ACKNOWLEDGEMENT

I would like to express my sincere gratitude to my main supervisor, Assoc. Prof. Dr. Hanani Abdul Manan, for her invaluable guidance, patience, encouragement, and continuous support throughout this research. Her insightful advice and academic mentorship have been essential to the completion of this thesis.

I would also like to extend my heartfelt appreciation to my co-supervisors, Assoc. Prof Dr. Noorazrul Yahya, Professor Dr. Hamzaini Abdul Hamid, and Dr. Chai Jia Ning, for their constructive suggestions, professional support, and valuable comments that greatly contributed to this study.

I am also sincerely thankful to the lecturers and staff of the Faculty of Medicine, Universiti Kebangsaan Malaysia, for their assistance and academic support throughout my doctoral study.

My heartfelt appreciation is extended to all colleagues, medical staff, and study participants who contributed directly or indirectly to this research. Their cooperation and support were indispensable to the successful completion of this study.

This work was supported by research grants from China, including the Research Project of Pan Zhihua Medical Research Center (PYYZ-2022-36), the Science and Technology Project of Pan Zhihua (2024ZD-S-35), and the Health Commission of Sichuan Province Medical Science and Technology Program (24WSXT001), as well as by the Faculty of Medicine Fundamental Grants (GFFP), Universiti Kebangsaan Malaysia (FF-2025-002).

Finally, I would like to express my deepest gratitude to my family for their understanding, patience, and unwavering support throughout this journey.

ABSTRAK

Pengenalpastian metastasis nodus limfa servikal lateral (LLNM) secara praoperatif dengan tepat adalah penting bagi pembuatan keputusan pembedahan dan penilaian prognosis dalam kalangan pesakit karsinoma tiroid papilari (PTC). Oleh sebab diseksi leher lateral profilaksis tidak disyorkan secara rutin berikutan risiko komplikasi pembedahan dan kesan kosmetik, pengesahan LLNM yang boleh dipercayai sebelum pembedahan masih merupakan satu cabaran klinikal yang penting. Kajian ini bertujuan untuk menilai secara menyeluruh prestasi diagnostik ultrasonografi konvensional, ultrasonografi dipertingkat kontras menggunakan Sonazoid (CEUS), tiroglobulin aspirasi jarum halus (FNA-Tg), dan radiomik berasaskan ultrasonografi dalam meramalkan LLNM secara praoperatif pada pesakit PTC. Kohort analisis akhir terdiri daripada 178 pesakit. Daripada jumlah ini, 44 pesakit telah menjalani pemeriksaan berkaitan sebagai sebahagian daripada penjagaan klinikal rutin antara 1 Februari dan 13 Jun 2024 dan hanya dikenal pasti serta dimasukkan secara retrospektif selepas kelulusan etika tempatan diperoleh pada 14 Jun 2024. Seramai 134 pesakit lagi dimasukkan dari Julai 2024 hingga November 2025 di bawah protokol kajian yang telah diluluskan. Setiap pesakit menyumbang satu nodus limfa servikal lateral yang menunjukkan abnormaliti pengimejan paling mencurigakan atau paling ketara untuk dianalisis. Semua nodus limfa yang dipilih menjalani ultrasonografi konvensional, penilaian CEUS Sonazoid semasa fasa arteri, vena, dan pascavaskular, serta pengukuran FNA-Tg. Pendekatan diagnostik kualitatif dan kuantitatif dibandingkan menggunakan analisis ciri pengendalian penerima dan ujian DeLong berpasangan. Gabungan multimodal yang melibatkan ultrasonografi konvensional, CEUS Sonazoid, dan FNA-Tg turut dinilai. Ciri-ciri radiomik diekstrak daripada imej ultrasonografi skala kelabu, pengimejan aliran Doppler warna, dan imej daripada ketiga-tiga fasa CEUS. Sebelas algoritma pembelajaran mesin dinilai menggunakan luas di bawah lengkung ciri pengendalian penerima, analisis lengkung keputusan, lengkung kalibrasi, dan skor ramalan pada peringkat sampel. Pendekatan kuantitatif secara konsisten menunjukkan prestasi diagnostik yang lebih baik berbanding penilaian kualitatif bagi ultrasonografi konvensional, CEUS Sonazoid berbilang fasa, dan FNA-Tg. CEUS Sonazoid mencapai nilai AUC sebanyak 0.979 dan menunjukkan prestasi yang secara signifikan lebih baik berbanding ultrasonografi konvensional ($p = 0.0057$). Dalam kalangan strategi multimodal, gabungan ultrasonografi konvensional, CEUS Sonazoid, dan FNA-Tg menghasilkan nilai AUC tertinggi yang diperhatikan, iaitu 0.993. Prestasi radiomik berbeza mengikut modaliti pengimejan. Model perceptron berbilang lapisan berasaskan CEUS fasa pascavaskular mencapai nilai AUC tertinggi dalam kohort ujian, iaitu 0.882. Dapatan ini menunjukkan potensi nilai suatu laluan diagnostik yang disesuaikan mengikut risiko, dengan ultrasonografi konvensional digunakan untuk pemetaan anatomi awal, CEUS Sonazoid untuk pencirian fungsi, FNA-Tg untuk pengesahan biokimia, dan radiomik untuk stratifikasi risiko kuantitatif tambahan. Melalui integrasi maklumat struktur, fungsi, biokimia, dan kuantitatif, pendekatan multimodal ini berpotensi meningkatkan penilaian risiko praoperatif serta menyokong pengurusan yang lebih diperibadikan bagi pesakit PTC.

ABSTRACT

Accurate preoperative identification of lateral cervical lymph node metastasis (LLNM) is crucial for surgical decision-making and prognostic assessment in patients with papillary thyroid carcinoma (PTC). Because prophylactic lateral neck dissection is not routinely recommended owing to the risks of surgical complications and cosmetic impairment, reliable preoperative confirmation of LLNM remains an important clinical challenge. This study aimed to comprehensively evaluate the diagnostic performance of conventional ultrasound, Sonazoid contrast-enhanced ultrasound (CEUS), fine-needle aspiration thyroglobulin (FNA-Tg), and ultrasound-based radiomics for the preoperative prediction of LLNM in PTC. The final analytical cohort comprised 178 patients. Of these, 44 had undergone the relevant examinations as part of routine clinical care between 1 February and 13 June 2024 and were retrospectively included only after local ethical approval was obtained on 14 June 2024. A further 134 patients were enrolled from July 2024 to November 2025 under the approved study protocol. Each patient contributed one lateral cervical lymph node showing the most suspicious or pronounced imaging abnormalities for analysis. All selected lymph nodes underwent conventional ultrasound, Sonazoid CEUS assessment during the arterial, venous, and post-vascular phases, and FNA-Tg measurement. Qualitative and quantitative diagnostic approaches were compared using receiver operating characteristic analysis and paired DeLong tests. Multimodal combinations of conventional ultrasound, Sonazoid CEUS, and FNA-Tg were also evaluated. Radiomics features were extracted from grayscale ultrasound, colour Doppler flow imaging, and images from the three CEUS phases. Eleven machine-learning algorithms were assessed using the area under the receiver operating characteristic curve, decision curve analysis, calibration curves, and sample-level prediction scores. Quantitative approaches generally showed higher diagnostic performance than the corresponding qualitative assessments for conventional ultrasound, multiphase Sonazoid CEUS, and FNA-Tg. Sonazoid CEUS achieved an AUC of 0.979 and significantly outperformed conventional ultrasound ($p = 0.0057$). Among the multimodal strategies, the combination of conventional ultrasound, Sonazoid CEUS, and FNA-Tg yielded the highest observed AUC of 0.993, although it was not statistically superior to the high-performing two-modality combinations. Radiomics performance varied across imaging modalities. The multilayer perceptron model based on post-vascular-phase CEUS achieved the highest observed test-cohort AUC of 0.882 among the radiomics models evaluated. These findings support a risk-adapted diagnostic approach in which conventional ultrasound provides initial anatomical assessment, Sonazoid CEUS adds functional information, FNA-Tg provides biochemical confirmation, and radiomics offers additional quantitative risk stratification. External multicentre validation is required before the radiomics models or integrated pathway can be adopted routinely.

CONTENTS

CHAPTER II LITERATURE REVIEW

CHAPTER IV RESULTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

LIST OF ABBREVIATIONS

ATA	American Thyroid Association
AUC	Area Under the Curve
BRAF	B-Raf proto-oncogene, serine/threonine kinase
CDFI	Color Doppler flow imaging
CEUS	Contrast-enhanced ultrasound
CI	Confidence interval
CLNM	Cervical lymph node metastasis
CT	Computed tomography
DCA	Decision curve analysis
ETE	Extrathyroidal extension
FN	False negative
FNAC	Fine-needle aspiration cytology
FNA-Tg	Fine-needle aspiration thyroglobulin
FP	False positive
GLCM	Grey-level co-occurrence matrix
GLDM	Grey-level dependence matrix
GLRLM	Grey-level run-length matrix
ICC	Intraclass correlation coefficient

IQR	Interquartile range
KNN	K-Nearest Neighbor
LASSO	Least Absolute Shrinkage and Selection Operator
LCLN	Lateral cervical lymph node
LLNM	Lateral cervical lymph node metastasis
LR	Logistic Regression
MI	Mechanical index
MLP	Multilayer Perceptron
MRI	Magnetic resonance imaging
NPV	Negative predictive value
PET	Positron emission tomography
PPV	Positive predictive value
PTC	Papillary thyroid carcinoma
PTMC	Papillary thyroid microcarcinoma
RF	Random Forest
ROC	Receiver Operator Characteristic
ROI	Region of interest
SD	Standard deviation
SE	Standard error

Tg	Serum thyroglobulin
TgAb	Serum thyroglobulin antibody
TN	True negative
TP	True positive
TPO-Ab	Serum thyroid peroxidase antibody
TSH	Serum thyroid-stimulating hormone
UKM	Universiti Kebangsaan Malaysia
US	Conventional ultrasound
YI	Youden Index

CHAPTER I

INTRODUCTION

1.1 RESEARCH BACKGROUND

Papillary thyroid carcinoma (PTC) is the most common histological subtype of differentiated thyroid cancer. Although its overall prognosis is generally favorable, cervical lymph node metastasis is frequent and influences staging, recurrence-risk stratification, surgical planning, and postoperative surveillance. Because prophylactic lateral neck dissection is not routinely recommended, reliable preoperative confirmation of lateral cervical lymph node metastasis (LLNM) is essential for selecting patients who may benefit from therapeutic dissection while avoiding unnecessary surgery and related morbidity (Haugen et al. 2016; Ringel et al. 2025).

Preoperative assessment remains difficult because metastatic involvement may be focal, partial, or microscopic and therefore may not produce obvious morphological changes. Conversely, benign reactive lymph nodes can show cortical thickening, altered vascularity, cystic change, or other features that overlap with metastatic disease. This biological and imaging overlap may lead either to missed metastases and incomplete initial surgery or to false-positive diagnoses and avoidable lateral neck dissection (Stack et al. 2012).

Conventional ultrasound (US) is the first-line imaging method because it is non-invasive, accessible, inexpensive, and suitable for real-time assessment. It evaluates nodal size, shape, echogenicity, hilum status, calcification, cystic degeneration, and vascular pattern. However, these findings mainly reflect macroscopic morphology, are operator-dependent, and may be insufficient for indeterminate or

small-volume disease; consequently, reported diagnostic sensitivity remains inconsistent (Bergeron et al. 2018).

Fine-needle aspiration thyroglobulin (FNA-Tg) provides complementary biochemical evidence of metastatic thyroid tissue and is particularly useful in cystic or cytologically equivocal nodes, although cut-off selection, assay variability, and sampling-related limitations remain unresolved (Yu et al. 2025).

Contrast-enhanced ultrasound (CEUS) adds functional information by depicting microvascular perfusion. Most previous cervical lymph node studies used SonoVue, a blood-pool agent whose enhancement is confined to a relatively short vascular phase. Sonazoid differs because its perfluorobutane microbubbles are retained by phagocytic cells of the reticuloendothelial system, enabling both vascular-phase imaging and a prolonged post-vascular phase (Yoo et al. 2026). This longer observation window may permit assessment of multiple neck compartments and may reveal tissue-related differences between benign and metastatic lymph nodes that are not apparent on conventional ultrasound or early vascular-phase imaging (Wei et al. 2022; Liu et al. 2026).

Radiomics provides a quantitative means of extracting intensity, texture, shape, and spatial-heterogeneity features from medical images. Its value, however, depends on the biological relevance and reproducibility of the source image rather than on machine learning alone. This issue is particularly important in ultrasound, where acquisition settings, operator technique, image quality, and segmentation can introduce variability. A phase-specific comparison is therefore needed to determine whether the post-vascular phase of Sonazoid offers a more stable and biologically informative substrate for radiomics than grayscale, Doppler, or early vascular-phase images (Lambin et al. 2012; Gillies et al. 2016).

Existing studies have usually assessed conventional ultrasound, FNA-Tg, CEUS, or radiomics separately, and many radiomics investigations have focused on central rather than lateral cervical lymph nodes (Wei et al. 2022; Chung et al. 2022). As a result, the incremental value of integrating anatomical, functional, biochemical, and quantitative information has not been established in a single cohort. A clinically useful

framework should therefore determine not only which modality performs best, but also how the modalities complement one another and whether radiomics adds value beyond conventional multimodal assessment.

Accordingly, the present study systematically evaluated conventional ultrasound, multiphase Sonazoid CEUS, FNA-Tg, and ultrasound-based radiomics for the preoperative prediction of LLNM in PTC. It compared qualitative and quantitative diagnostic strategies, assessed single and combined modalities, and examined the influence of imaging phase and machine learning classifier on radiomics performance, intending to develop a biologically grounded and clinically interpretable multimodal framework.

1.2 PROBLEM STATEMENT

Accurate preoperative assessment of lateral cervical lymph node metastasis (LLNM) in papillary thyroid carcinoma (PTC) remains a clinically important problem because the decision to perform therapeutic lateral neck dissection depends on reliable evidence of metastatic disease. Conventional ultrasound is recommended as the first-line examination, but its diagnostic performance is variable because metastatic deposits may be small or partial and may preserve apparently benign nodal morphology, whereas reactive nodes may show cortical thickening, altered vascularity, or other suspicious changes. Previous studies and guidelines have therefore emphasised that no single conventional ultrasound sign is sufficiently reliable in all patients, and diagnostic performance is affected by nodal size, operator experience, feature definition, and the reference standard used (Haugen et al. 2016; Mulla et al. 2012; Chung et al. 2022). This limitation has direct clinical consequences: false-negative assessment may leave metastatic lateral disease untreated at the initial operation, whereas false-positive assessment may expose patients to unnecessary lateral neck dissection and associated neurological, vascular, lymphatic, and cosmetic morbidity.

Adjunctive methods improve different dimensions of this uncertainty but remain incomplete when used alone. Published CEUS evidence has reported high pooled diagnostic performance for differentiating benign and malignant cervical lymph nodes, including pooled sensitivity and specificity of approximately 95% and 87%,

respectively, although most studies used the blood-pool agent SonoVue and heterogeneous diagnostic criteria (Yu et al. 2023). In addition, CEUS combined with FNA-Tg has been reported to achieve sensitivity of 97.3% and a negative predictive value of 93.0%, illustrating the potential value of combining functional imaging with biochemical confirmation (Liu et al. 2022). Nevertheless, FNA-Tg remains affected by assay and cut-off variability, conventional blood-pool CEUS provides only a short vascular observation window, and evidence for intravenous Sonazoid post-vascular imaging of lateral cervical lymph nodes remains comparatively limited. Radiomics adds quantitative information beyond visual interpretation, but published thyroid radiomics studies are predominantly retrospective, frequently single-centre, and often focus on central rather than lateral nodal disease, with substantial heterogeneity in image acquisition, segmentation, feature selection, and validation (Gillies et al. 2016; Liu et al. 2024).

The central research gap is therefore the lack of a unified and clinically interpretable evaluation that directly compares and integrates structural ultrasound information, Sonazoid vascular- and post-vascular-phase information, FNA-Tg biochemical evidence, and radiomics-based quantitative information in the same lateral-node cohort under a consistent reference standard. Existing studies do not adequately establish the incremental contribution of each component, the most informative Sonazoid phase for quantitative modelling, or whether radiomics provides clinically meaningful information beyond conventional multimodal assessment. To address these gaps, this study compared qualitative and quantitative strategies, evaluated conventional ultrasound, multiphase Sonazoid CEUS, and FNA-Tg alone and in combination, and developed modality-specific radiomics models for preoperative LLNM prediction.

1.3 OBJECTIVES OF THE RESEARCH

1.3.1 General Objective

The general objective of the present study was to develop and evaluate a multimodal and radiomics-based diagnostic framework for the preoperative prediction of lateral cervical lymph node metastasis in papillary thyroid carcinoma.

1.3.2 Specific Objectives

a. To evaluate and compare the diagnostic performance of conventional ultrasound, Sonazoid contrast-enhanced ultrasound, and fine-needle aspiration thyroglobulin in the preoperative assessment of lateral cervical lymph node metastasis in papillary thyroid carcinoma.

b. To assess whether quantitative approaches, including radiomics analysis, provide superior diagnostic accuracy compared with conventional qualitative assessment.

c. To investigate the incremental diagnostic value of multimodal integration by combining imaging modalities and biochemical markers.

d. To develop radiomics-based predictive models using multiphase ultrasound data and evaluate the performance, stability, and clinical utility of the developed models, including comparison of machine learning classifiers and assessment of their potential role in supporting surgical planning and personalized management.

1.4 RESEARCH QUESTIONS

a. What are the diagnostic performances of conventional ultrasound, Sonazoid contrast-enhanced ultrasound, and fine-needle aspiration thyroglobulin in the preoperative assessment of lateral cervical lymph node metastasis in papillary thyroid carcinoma?

b. Do quantitative approaches provide superior diagnostic accuracy compared to conventional qualitative assessment for suspicious lateral cervical lymph nodes?

c. Does multimodal integration of ultrasound modalities and fine-needle aspiration thyroglobulin improve the accuracy of preoperative prediction, and can effective radiomics-based models be developed using multiphase ultrasound data?

d. How do different machine learning classifiers and imaging modalities compare in terms of predictive performance and stability, and can the resulting models provide clinically meaningful support for surgical planning and personalized management?

1.5 RESEARCH HYPOTHESES

The present study was based on the following hypotheses:

1.5.1 General Hypothesis

The general hypothesis of the present study was that a multimodal and radiomics-based diagnostic framework could improve the preoperative prediction of lateral cervical lymph node metastasis in papillary thyroid carcinoma, and would demonstrate superior diagnostic performance compared with single-modality assessment.

1.5.2 Specific Hypotheses

a. It was hypothesized that conventional ultrasound, Sonazoid contrast-enhanced ultrasound, and fine-needle aspiration thyroglobulin would show significantly different diagnostic performance in the preoperative assessment of lateral cervical lymph node metastasis in papillary thyroid carcinoma.

b. It was hypothesized that quantitative approaches, including radiomics analysis, would achieve superior diagnostic accuracy compared with conventional qualitative assessment.

c. It was hypothesized that multimodal integration of imaging modalities and biochemical markers would provide incremental diagnostic value beyond single-modality assessment, and that radiomics-based predictive models derived from multiphase ultrasound data would effectively predict lateral cervical lymph node metastasis.

d. It was hypothesized that the developed models would demonstrate good performance, stability, and clinical utility, and that different machine learning classifiers would show different levels of predictive performance in supporting surgical planning and personalized management.

1.6 SIGNIFICANCE OF THE STUDY

The principal original contribution of this study is the development and direct comparative evaluation of a unified preoperative framework that integrates conventional ultrasound, multiphase Sonazoid contrast-enhanced ultrasound, FNA-Tg, and modality-specific radiomics for lateral cervical lymph node metastasis. In contrast to previous studies that generally evaluated these components separately, the present study assessed their individual, complementary, and incremental diagnostic value within the same patient cohort and against a common reference standard.

The methodological contribution lies in demonstrating that radiomics performance depends not only on classifier selection but also on the biological relevance and temporal stability of the source image. By comparing grayscale ultrasound, color Doppler flow imaging (CDFI), and arterial-, venous-, and post-vascular-phase Sonazoid images, the study provides a phase-specific evaluation of the imaging substrate used for quantitative modelling. This approach helps clarify why post-vascular Sonazoid imaging may provide more informative and reproducible features than conventional or early vascular-phase images.

The clinical contribution is the definition of a risk-based diagnostic pathway rather than the promotion of a single superior test. Conventional ultrasound provides the initial anatomical assessment; Sonazoid CEUS adds phase-specific functional and tissue-related information; FNA-Tg supplies biologically specific biochemical evidence; and radiomics offers an additional quantitative layer. Their integration may improve the identification of patients who require therapeutic lateral neck dissection while reducing missed metastases and unnecessary surgery.

The study therefore contributes original knowledge at three levels: it establishes a unified multimodal comparison for preoperative LLNM prediction, provides evidence for biologically informed selection of radiomics imaging substrates, and proposes a clinically interpretable framework in which radiomics supports rather than replaces multimodal clinical assessment. These findings provide a foundation for prospective multicentre validation, standardized acquisition and segmentation protocols, and future integration with clinical and molecular variables.

1.7 SCOPE OF THE STUDY

The present study focused on the preoperative assessment of lateral cervical lymph node metastasis in patients with papillary thyroid carcinoma. Its scope was defined in terms of study population, diagnostic modalities, analytical methods, and intended clinical application.

First, the study population was limited to patients with papillary thyroid carcinoma who underwent preoperative evaluation for suspicious lateral cervical lymph nodes. The study specifically addressed metastatic involvement of the lateral cervical compartment and did not primarily focus on central compartment lymph node metastasis or other thyroid malignancies (Mulla et al. 2012). Therefore, the findings are most directly applicable to the diagnostic assessment and management of lateral cervical lymph nodes in PTC.

Second, the study was confined to selected diagnostic modalities relevant to preoperative evaluation. These included conventional ultrasound, Sonazoid contrast-enhanced ultrasound, fine-needle aspiration thyroglobulin, and radiomics-based modelling derived from ultrasound images. Within the radiomics component, the study examined features extracted from multiple ultrasound modalities, including two-dimensional ultrasound, CDFI, and phase-specific Sonazoid-enhanced ultrasound images. Other imaging modalities, such as computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), or molecular imaging techniques, were beyond the scope of the present investigation.

Third, the study focused primarily on diagnostic prediction rather than therapeutic intervention. Its main aim was to evaluate the ability of multimodal and radiomics-based approaches to predict lateral cervical lymph node metastasis before surgery. Although the study considered the potential clinical implications of improved prediction for surgical planning and patient management, it did not directly assess treatment outcomes, postoperative complications, long-term survival, or recurrence through a prospective interventional design.

Fourth, the radiomics analysis was limited to handcrafted feature extraction and machine learning-based predictive modelling derived from the available ultrasound data. While different classifiers were compared, the study did not extend to fully deep learning-based end-to-end image analysis or automated clinical deployment systems. In addition, although efforts were made to assess model stability and generalizability, validation was primarily performed within the available study dataset and test cohort, rather than across large-scale external multicenter populations (Gillies et al. 2016).

Fifth, the study was conducted within a defined methodological framework that emphasized multimodal integration, quantitative analysis, model validation, and clinical interpretability. Broader issues such as health economics, regulatory implementation, and large-scale workflow integration were considered only in terms of future implications rather than as primary study objectives.

In summary, the present study was limited to the multimodal preoperative prediction of lateral cervical lymph node metastasis in papillary thyroid carcinoma using conventional ultrasound, Sonazoid contrast-enhanced ultrasound, FNA-Tg, and radiomics-based modelling. Its scope was diagnostic and methodological, with particular emphasis on improving preoperative risk stratification and supporting clinically relevant decision-making.

1.8 OPERATIONAL DEFINITIONS OF KEY TERMS

For the present study, the following key terms were operationally defined:

a. Papillary thyroid carcinoma (PTC)

A histologically confirmed malignant epithelial tumor arising from thyroid follicular cells and classified as papillary thyroid carcinoma.

b. Lateral cervical lymph node metastasis (LLNM)

Metastatic involvement of lymph nodes located in the lateral cervical compartment in patients with papillary thyroid carcinoma, confirmed by pathological examination and used as the primary outcome variable in the study.

c. Fine-needle aspiration cytology (FNAC)

A diagnostic procedure in which cells are aspirated from suspicious lymph nodes under ultrasound guidance for cytological examination.

d. Fine-needle aspiration thyroglobulin (FNA-Tg)

The thyroglobulin concentration measured in needle-washout fluid obtained after ultrasound-guided fine-needle aspiration of a suspicious lymph node, used as a biochemical indicator of metastatic thyroid tissue.

e. Conventional ultrasound (US)

Routine grayscale ultrasound examination used to assess lymph node morphology and structural characteristics such as size, shape, echogenicity, hilum status, calcification, cystic change, and CDFI was an ultrasound technique used to evaluate blood flow distribution and vascular patterns within lymph nodes.

f. Contrast-enhanced ultrasound (CEUS)

An ultrasound imaging method using intravenously administered microbubble contrast agents to assess nodal perfusion characteristics and enhancement patterns.

g. Sonazoid contrast-enhanced ultrasound (Sonazoid CEUS)

A CEUS technique using perfluorobutane microbubbles as the contrast agent, enabling assessment of both vascular-phase and post-vascular-phase enhancement characteristics.

h. Arterial phase

The early phase after contrast administration, during which the microbubbles circulate predominantly within the arterial system, and early nodal perfusion is observed.

i. Venous phase

The subsequent vascular phase, after arterial enhancement, during which contrast washout and delayed perfusion characteristics can be assessed.

j. Post-vascular phase

The latter phase, after Sonazoid administration, in which persistent tissue-related contrast retention can be observed beyond the purely vascular circulation phase.

k. Radiomics

A quantitative imaging analysis approach involving the extraction of high-dimensional features from medical images to characterize image intensity, texture, shape, and spatial heterogeneity.

l. Radiomics feature

A quantitative descriptor extracted from medical images that numerically represents a specific imaging characteristic, such as texture, intensity distribution, shape, or spatial relationship; includes gray-level size zone matrix (GLSZM), first-order features, gray-level dependence matrix (GLDM), shape features, gray-level co-occurrence matrix (GLCM), neighbouring gray tone difference matrix (NGTDM), and gray-level run length matrix (GLRLM), etc.

m. Region of interest (ROI)

The manually or semi-automatically delineated area of a lymph node on an image from which radiomics features are extracted.

n. Machine learning classifier

A computational algorithm used to learn the relationship between radiomics features and lymph node metastatic status for predictive modelling, such as Logistic Regression (LR), Naive Bayes, support vector machine (SVM),k-Nearest Neighbor (KNN), Random Forest (RF), ExtraTrees, Extreme Gradient Boosting (XGBoost), LightGBM, gradient boosting, AdaBoost, multilayer perceptron (MLP).

o. Least Absolute Shrinkage and Selection Operator (LASSO)

A regularization and variable selection method used in regression analysis, in which an L1 penalty is applied to the absolute values of regression coefficients to shrink less important coefficients towards zero and, in some cases, reduce them exactly to zero, thereby selecting the most relevant features and reducing overfitting in radiomics-based modelling.

p. Decision curve analysis (DCA)

A method used to evaluate the clinical usefulness of a diagnostic model or predictive model by quantifying its net benefit across a range of threshold probabilities, thereby assessing whether the model provides meaningful support for clinical decision-making in radiomics-based modelling.

q. Calibration Curve

Calibration curves were used to evaluate the agreement between the predicted probabilities and the observed frequencies of lateral cervical lymph node metastasis

r. Sample-level prediction scores

Quantitative output values generated by a predictive model for each sample, representing the estimated probability, confidence, or likelihood that the sample belongs to a particular class or outcome category in radiomics-based modelling.

s. Multimodal diagnostic framework

An integrated diagnostic approach combining conventional ultrasound, Sonazoid contrast-enhanced ultrasound, fine-needle aspiration thyroglobulin, and radiomics-based modelling for the prediction of lateral cervical lymph node metastasis.

1.9 ORGANIZATION OF THE THESIS

This thesis is organized into six chapters, each addressing a specific component of the research process and contributing to the overall investigation of multimodal and radiomics-based prediction of lateral cervical lymph node metastasis in papillary thyroid carcinoma.

Chapter One presents the general introduction to the study. It includes the research background, problem statement, general and specific objectives, research questions, research hypotheses, significance of the study, scope of the study, and operational definitions of key terms.

Chapter Two reviews the relevant literature related to papillary thyroid carcinoma, lateral cervical lymph node metastasis, conventional ultrasound, contrast-enhanced ultrasound, fine-needle aspiration thyroglobulin, radiomics, and multimodal diagnostic approaches. It also identifies current knowledge gaps and provides the theoretical basis for the present study.

Chapter Three describes the research methodology. It outlines the study design, study setting, study population, inclusion and exclusion criteria, data collection procedures, imaging protocols, radiomics workflow, machine learning modelling, outcome measures, and statistical analysis methods.

Chapter Four presents the results of the study. It reports the diagnostic performance of conventional ultrasound, Sonazoid contrast-enhanced ultrasound, and FNA-Tg, as well as the results of multimodal integration and radiomics-based modelling. Validation findings and comparative model performance are also described.

Chapter Five discusses the findings in relation to previous studies and the broader clinical and methodological context. It interprets the diagnostic value of the evaluated modalities, the significance of multimodal integration, the role of radiomics and machine learning classifiers, and the clinical implications of the findings.

Chapter Six provides the overall conclusion and recommendations. It summarizes the main findings of the study, highlights their clinical and research implications, limitation of the study and proposes recommendations for practice and future investigation.

CHAPTER II

LITERATURE REVIEW

2.1 INTRODUCTION OF PAPILLARY THYROID CARCINOMA AND LATERAL CERVICAL LYMPH NODE METASTASIS

Papillary thyroid carcinoma (PTC) is the predominant differentiated thyroid malignancy. Although disease-specific survival is generally excellent, cervical lymphatic spread is common and materially affects staging, recurrence risk, surgical planning, and surveillance. Lateral cervical lymph node metastasis (LLNM) is especially important because it may require therapeutic neck dissection, a procedure associated with greater technical complexity and morbidity than central compartment surgery (Haugen et al. 2016; Gu et al. 2025).

PTC demonstrates marked biological heterogeneity. Some tumours remain indolent for many years, whereas others show multifocality, extrathyroidal extension, aggressive variants, or early nodal dissemination. This heterogeneity explains why nodal status cannot be inferred reliably from the behaviour of the primary tumour alone and why direct assessment of the lateral neck is required (Haugen et al. 2016; Stack et al. 2012).

LLNM is clinically distinct from central compartment metastasis. Lateral disease is more likely to alter the operative field, extend surgical duration, and expose patients to injury of the spinal accessory nerve, internal jugular vein, cervical plexus, and thoracic duct. The diagnostic threshold for recommending surgery must therefore balance oncological completeness against avoidable morbidity.

The relevance of LLNM extends beyond a binary metastatic status. Nodal burden, size of the largest metastatic focus, number of involved levels, extranodal extension, and the relationship between nodal disease and the primary tumour all influence recurrence risk and the intensity of subsequent management. Consequently, a clinically useful preoperative assessment should not merely detect abnormal nodes; it should also characterize the probability, distribution, and biological plausibility of metastatic involvement (Filetti et al. 2019).

A further difficulty is that the reference standard itself may vary across studies. Histopathology after compartment dissection offers the strongest confirmation but is available only in operated patients, whereas cytology, biochemical analysis, and longitudinal imaging follow-up are often used in non-operated cohorts. This heterogeneity partly explains the variability in reported diagnostic performance and highlights the importance of clearly defining node-level and patient-level outcomes.

2.1.1 Epidemiology and Clinical Profile of Papillary Thyroid Carcinoma

PTC accounts for approximately 80–85% of thyroid malignancies. Its rising incidence is attributed largely to improved detection through high-resolution imaging, although genuine epidemiological changes may also contribute. Despite an indolent course in many patients, cervical lymph node metastasis is reported in a substantial proportion of cases and is associated with locoregional recurrence, reoperation, and intensified follow-up (Ruiz et al. 2020).

PTC should therefore not be considered uniformly low risk. Tumour size, histological subtype, multifocality, extrathyroidal extension, molecular profile, and nodal distribution contribute to biological heterogeneity and should be integrated into risk-adapted management(Yilmaz et al. 2023).

The apparent increase in PTC incidence has been driven partly by the detection of small tumours through high-resolution ultrasound and cross-sectional imaging. Nevertheless, the clinical burden is not determined by incidence alone. The distribution and volume of cervical nodal disease influence recurrence patterns, the need for

adjuvant treatment, and the intensity of postoperative surveillance (Haugen et al. 2016; Filetti et al. 2019).

Age, tumour size, upper-pole location, multifocality, extrathyroidal extension, and selected molecular alterations have been associated with cervical nodal spread. However, none of these variables is sufficiently accurate to replace direct lymph node evaluation. They are better regarded as contextual risk factors that modify pre-test probability.

2.1.2 Patterns and Characteristics of Lateral Cervical Lymph Node Metastasis

PTC most commonly spreads through the cervical lymphatic system. Central compartment nodes are often involved first, followed by levels II–V; however, skip metastasis may occur, particularly from upper-pole tumours that drain along the superior thyroid vessels (Yilmaz et al. 2023).

LLNM is associated with larger tumours, multifocality, extrathyroidal extension, greater recurrence risk, and more complex surgery. Although its effect on overall survival may be limited in younger patients, it has important consequences for disease-free survival and the need for reoperative treatment (Dong et al. 2024; Haugen et al. 2016). The lateral neck contains several nodal levels with different drainage pathways and anatomical constraints. Levels III and IV are commonly involved, while level II disease may occur in upper-pole tumours. Level V involvement is less frequent but clinically relevant because posterior triangle dissection may increase shoulder morbidity (Haugen et al. 2016; Stack et al. 2012; Kessler et al. 2003).

Metastatic deposits may replace the entire node or involve only a focal cortical region. Partial replacement explains why some metastatic nodes retain a fatty hilum or oval contour. Conversely, reactive nodes may become enlarged and hypervascular. The biological spectrum of nodal involvement therefore produces considerable overlap in imaging appearance (Stack et al. 2012). Recent evidence also emphasizes the prognostic importance of nodal burden rather than nodal positivity alone. The number and size of metastatic nodes, extranodal extension, and involvement of multiple lateral levels may

provide more clinically meaningful information than a simple N1b category. These factors influence recurrence risk and the intensity of postoperative surveillance. Nevertheless, they are generally available only after surgery; preoperative imaging must therefore identify not only whether metastasis is present, but also its distribution and probable extent (Haugen et al. 2016). Skip metastasis illustrates why central and lateral compartments cannot be assessed sequentially in a rigid manner. Upper-pole tumours may drain directly to the lateral neck, and small primary tumours may coexist with clinically meaningful lateral disease. A systematic level-by-level lateral neck examination is therefore required even when the central compartment appears negative (Haugen et al. 2016; Kessler et al. 2003).

The distribution of LLNM is not random. Levels III and IV are frequently involved because of drainage along the internal jugular chain, while level II involvement is particularly relevant in upper-pole tumours. Level V metastasis is less common but may be overlooked because of its posterior location and the technical limitations of routine ultrasound. Multi-level disease and extranodal extension generally indicate a greater tumour burden and may justify more extensive surgical planning (Dong et al. 2024). Skip metastasis challenges the assumption that a normal central compartment excludes lateral disease. Direct drainage from the superior thyroid pole toward the lateral chain may permit lateral spread without demonstrable central involvement. This phenomenon reinforces the need for systematic lateral neck examination even in patients with small primary tumours or apparently negative central nodes (Di Filippo et al. 2026).

2.1.3 Surgical and Diagnostic Management of Lateral Cervical Lymph Node Metastasis for Papillary Thyroid Carcinoma

Prophylactic lateral neck dissection is not routinely recommended because clinically significant LLNM is not universal and surgery may cause shoulder dysfunction, chyle leakage, nerve injury, vascular injury, and cosmetic morbidity. Current guidelines therefore favour therapeutic dissection when metastasis is confirmed preoperatively (Haugen et al. 2016; Thigpen and Shah 2008).

This policy makes diagnostic accuracy central to management. False-negative assessment may lead to incomplete initial surgery and later reoperation, whereas false-positive assessment may expose patients to unnecessary dissection. Conventional ultrasound remains the first-line method, but its limitations justify complementary functional, biochemical, and quantitative approaches. Preoperative confirmation is especially important because lateral neck dissection is usually planned as a compartment-oriented procedure rather than removal of a single visible node. A false-positive result may therefore lead to a much larger operation than the diagnostic target itself would suggest (Haugen et al. 2016; Stack et al. 2012).

Diagnostic evaluation is typically stepwise. Ultrasound identifies and maps suspicious nodes; fine-needle aspiration provides cytological and biochemical confirmation; cross-sectional imaging may be added when disease is extensive, deep, or anatomically complex. Newer techniques such as CEUS and radiomics are intended to improve confidence before invasive confirmation or surgery (Haugen et al. 2016). The consequences of diagnostic error are asymmetric. Missing a true metastatic node may result in incomplete initial treatment, persistent disease, and a technically more difficult reoperation. Conversely, falsely diagnosing metastasis may lead to an unnecessary compartment dissection with risks of accessory nerve dysfunction, chyle leak, sensory disturbance, and visible scarring (Zhang et al. 2023). The preferred diagnostic pathway must therefore achieve high sensitivity without sacrificing clinically meaningful specificity.

From a decision-theory perspective, the value of a diagnostic test depends not only on its AUC but also on the probability threshold at which the result changes management. A test may be statistically accurate yet clinically unhelpful if it does not alter the decision to observe, aspirate, or operate. This distinction is important when evaluating advanced imaging and radiomics models.

2.1.4 Methodological Heterogeneity, Selection Bias, and Reference Standards

Interpretation of the diagnostic literature requires attention to heterogeneity that extends beyond the imaging technique itself. Studies differ in whether they recruit consecutive

patients, only surgically treated patients, or only lymph nodes already considered suspicious on conventional ultrasound. Cohorts enriched with overt metastatic disease may yield higher sensitivity and area under the curve than routine clinical populations containing small, partially replaced, or inflammatory nodes. Conversely, studies confined to indeterminate nodes may underestimate the apparent performance of highly specific signs because unequivocal benign and malignant appearances have been excluded. Diagnostic estimates should therefore be interpreted in relation to disease spectrum, inclusion criteria, nodal compartment, and whether analysis is performed at the patient or node level (Haugen et al. 2016; Stack et al. 2012; Mulla & Knoefel 2012).

Selection and verification bias are particularly relevant. Nodes with strongly suspicious imaging findings are more likely to undergo aspiration or surgery, whereas benign-appearing nodes are often verified only by follow-up. When histopathology is available predominantly in high-risk cases, specificity may be overestimated, and occult false-negative disease may be underrepresented. Mixed reference standards—histopathology, cytology, FNA-Tg, and longitudinal imaging—improve feasibility but are not methodologically equivalent. Histopathology after compartment dissection provides the strongest confirmation, while follow-up depends on duration, imaging quality, and the assumption that stability excludes microscopic metastasis (Stack et al. 2012; Haugen et al. 2016).

Reference-standard heterogeneity also affects comparisons among modalities. A test evaluated against node-by-node pathology addresses a different question from a patient-level analysis in which one positive node determines metastatic status. Similarly, surgical series may include larger or more aggressive nodes than surveillance cohorts. Robust literature synthesis therefore requires consideration of spectrum effects, partial verification, incorporation bias, and the independence of the index test from the reference standard (Stack et al. 2012; Mulla and Knoefel 2012). These issues explain part of the wide variation in reported diagnostic performance and support cautious interpretation of apparently superior AUC values across separate studies.

2.2 CONVENTIONAL ULTRASOUND FOR THE ASSESSMENT OF LATERAL CERVICAL LYMPH NODES

Conventional ultrasound is the first-line imaging method for lateral cervical lymph node assessment because it is accessible, radiation-free, inexpensive, and capable of real-time morphological and vascular evaluation. It also permits immediate ultrasound-guided aspiration when indicated (Haugen et al. 2016; Kessler et al. 2003).

Its principal limitation is that benign reactive and metastatic nodes may share similar appearances. Diagnostic performance therefore depends on combined interpretation of grayscale and Doppler findings rather than any single feature (Sohn et al. 2010; Mulla and Knoefel 2012; Filetti et al. 2019).

Ultrasound examination should be systematic and level-based. The operator evaluates nodal distribution, short-axis diameter, shape, echogenicity, hilum, cortical pattern, calcification, cystic change, and vascularity while comparing suspicious nodes with adjacent benign-appearing nodes. This contextual comparison is often more informative than a single absolute threshold (Filetti et al. 2019). Although CT and MRI may provide complementary anatomical information, ultrasound remains superior for superficial high-resolution assessment and image-guided sampling. Its main disadvantage is that image acquisition and interpretation occur simultaneously, making the final diagnosis dependent on the examiner's technique and experience (Kessler et al. 2003; Haugen et al. 2016).

A comprehensive ultrasound examination should include transverse and longitudinal scanning from the submandibular region to the supraclavicular fossa, with documentation by cervical level. Suspicious nodes should be assessed in relation to the carotid sheath, sternocleidomastoid muscle, clavicle, and adjacent nerves and vessels because these relationships influence both sampling feasibility and operative planning (Ahuja and Ying 2005; Filetti et al. 2019).

Conventional ultrasound also functions as the gateway to subsequent tests. It identifies the node selected for CEUS, defines the region of interest for quantitative analysis, and guides FNA-C and FNA-Tg. Its value therefore persists even when an

adjunctive modality ultimately provides the decisive evidence (Haugen et al. 2016; Chung et al. 2001).

2.2.1 Grayscale Ultrasound for Lateral Cervical Lymph Node Assessment

Grayscale assessment includes nodal size, shape, long-to-short axis ratio, echogenicity, cortical thickness, hilar status, margins, calcification, and cystic change. Round shape, loss of the fatty hilum, heterogeneous echotexture, punctate echogenic foci, and cystic degeneration increase suspicion for metastasis (Ahuja and Ying 2005; Filetti et al. 2019). No single criterion is sufficiently reliable. Reactive nodes may be enlarged or cortically thickened, while small metastatic deposits may preserve an oval shape and echogenic hilum. Size thresholds also vary by nodal level and perform poorly when used alone(Leenhardt et al. 2013). Grayscale findings should therefore be interpreted as a pattern rather than as isolated binary signs.

Certain findings have relatively high specificity. Punctate echogenic foci compatible with microcalcifications, cystic degeneration, and focal hyperechogenicity are strongly suggestive of metastatic PTC. However, these features are not sensitive because they occur only in a subset of metastatic nodes (Ahuja and Ying 2005; Kessler et al. 2003; Filetti et al. 2019). Loss of the echogenic hilum and a round configuration are common suspicious features but are less specific. The hilum may be absent in small normal nodes, and reactive nodes may become round during inflammation (Cai et al. 2025; Di Filippo et al. 2026). Similarly, a small short-axis diameter does not exclude metastasis, especially when tumour deposits are focal.

These observations support a pattern-based approach in which multiple features are considered together. A weighted score may improve consistency by assigning greater importance to highly specific findings while reducing reliance on size alone (Kessler et al. 2003; Filetti et al. 2019).

The diagnostic significance of individual grayscale features differs. Microcalcifications and cystic degeneration have high specificity for metastatic PTC but relatively low sensitivity. Loss of the fatty hilum and round shape are more common but less specific because they may also occur in reactive or small normal nodes.

Hyperechogenicity can reflect intranodal deposits of thyroid tissue, but subtle echogenic change is prone to observer variation (Stack et al. 2012; Ferreira et al. 2013).

The long-to-short-axis ratio is often used as a surrogate for shape, yet its threshold varies among studies (Ardakani et al. 2018). A binary cut-off may oversimplify a continuous morphological transition from elongated benign nodes to rounder metastatic nodes. Quantitative scoring systems can preserve more information by assigning graded weights to multiple features rather than treating each sign as independently positive or negative.

2.2.2 Color Doppler Flow Imaging in Lateral Cervical Lymph Nodes

Color Doppler flow imaging (CDFI) complements grayscale ultrasound by depicting vascular distribution. Benign or reactive nodes usually show hilar flow, whereas metastatic nodes more often demonstrate peripheral, mixed, or disorganized vascularity (Ahuja and Ying 2005; Kim et al. 2017). However, inflammatory hypervascularity may mimic malignancy, while necrotic metastases may show little detectable flow. Doppler findings are also affected by gain, pulse repetition frequency, transducer pressure, patient position, and operator experience, which limits reproducibility (Yuan et al. 2006; Chammas 2019).

Peripheral vascularity is thought to arise from tumour-induced neovascularization, whereas preserved hilar flow reflects normal nodal architecture. Mixed vascularity may indicate partial replacement or coexistence of residual hilar vessels and peripheral tumour vessels(Ahuja and Ying 2005).

Spectral Doppler indices have also been explored, but their clinical use remains limited because measurements are technically sensitive and show substantial overlap between benign and malignant nodes (Ahuja and Ying 2005; Chammas 2019). CDFI therefore contributes supportive rather than definitive evidence.

Vascular distribution is biologically linked to nodal architecture. Hilar vessels are usually preserved in benign lymph nodes, whereas peripheral or mixed vascularity

may reflect tumour-driven angiogenesis and destruction of the normal vascular tree. Nevertheless, these patterns are influenced by inflammation, necrosis, and technical settings, which explains their limited standalone specificity (Kim et al. 2017).

Power Doppler and microvascular flow techniques may improve sensitivity for low-velocity vessels, but they remain sensitive to motion and transducer pressure. Thus, Doppler assessment should be standardized and interpreted as complementary evidence rather than a definitive marker of metastatic status (Zhao et al. 2021).

2.2.3 Combined Interpretation of Morphological and Vascular Features in Lateral Cervical Lymph Nodes

Clinical ultrasound diagnosis is based on the combined interpretation of morphology and vascularity. This approach improves confidence compared with isolated criteria, but indeterminate combinations remain common, such as preserved hilar architecture with peripheral flow or suspicious morphology with minimal vascularity. (Ahuja and Ying 2005; Mulla and Knoefel 2012)

Weighted or quantitative scoring may reduce subjectivity, yet conventional ultrasound still reflects only macroscopic features and cannot directly characterize the biochemical or microstructural changes associated with early metastasis. This limitation provides the rationale for adjunctive FNA-Tg, CEUS, and radiomics (Ahuja and Ying 2005; Haugen et al. 2016).

Combined interpretation is clinically useful because malignant transformation is rarely expressed through a single feature. For example, a round node with absent hilum and peripheral vascularity is more suspicious than a node showing only one of these findings.

Nevertheless, combinations remain qualitative unless formal scoring rules are applied. Different observers may assign different importance to the same findings, resulting in inconsistent classifications. Quantitative scoring systems aim to translate this multidimensional assessment into a reproducible estimate of metastasis risk (Bournaud et al. 2010).

2.2.4 Limitations of Conventional Ultrasound in Lateral Cervical Lymph Nodes

Ultrasound performance is constrained by anatomical complexity, variable nodal depth and orientation, and partial shielding by muscles or vessels. Image quality and interpretation also depend heavily on examiner experience, scanning technique, equipment settings, and diagnostic thresholds.

Published sensitivity and specificity vary substantially across studies, reflecting differences in populations, protocols, and reference standards. Interobserver agreement is often only moderate for cortical thickness, hilum status, and vascular distribution (Mulla and Knoefel 2012). These limitations reduce standardization and may affect decisions regarding lateral neck dissection. Anatomical accessibility differs across nodal levels. Deep level II nodes may be partially obscured by the mandible or sternocleidomastoid muscle, while low level IV nodes may be difficult to visualize near the clavicle. Obesity, short neck configuration, postoperative change, and patient motion further reduce image quality (Chammas 2019).

Operator dependence affects both acquisition and interpretation. A poorly optimized image may conceal subtle calcification or vascularity, while an experienced examiner may identify suspicious focal changes that are not captured in a single stored image. These factors complicate retrospective review and multicentre standardization (Mulla et al. 2012). Another limitation is verification bias. Nodes with strongly suspicious ultrasound findings are more likely to undergo aspiration or surgery, while benign-appearing nodes are often followed without pathological confirmation. Studies that do not account for this process may overestimate specificity or underestimate false-negative disease (Kim et al. 2017).

Reproducibility is also affected by the use of static representative images. Real-time examination allows dynamic comparison, compression, multiplanar assessment, and targeted Doppler interrogation, whereas retrospective review is restricted to selected frames. The diagnostic performance of stored images may therefore differ from that of an experienced examiner performing the complete examination.

2.2.5 Failure Patterns of Conventional Ultrasound in the Lateral Cervical Compartment

Two recurrent failure patterns are clinically important. First, small-volume metastases may retain near-normal architecture and therefore produce false-negative findings. Second, reactive lymphadenopathy may cause cortical thickening, altered echogenicity, and increased vascularity, resulting in false-positive interpretation (Cai et al. 2025).

These errors arise because conventional ultrasound depicts downstream macroscopic changes rather than tumour-specific biology. Higher image resolution alone cannot fully resolve this limitation; complementary techniques are required when findings remain equivocal (Murata 2018).

False-negative findings are most likely when metastatic deposits are microscopic, focal, or located within an otherwise preserved node. They may also occur in deeply situated nodes or when necrosis eliminates detectable vascularity (Mulla et al. 2012).

False-positive findings frequently arise in reactive nodes associated with infection, autoimmune disease, or granulomatous inflammation. Such nodes may show cortical thickening, increased flow, and heterogeneous echogenicity. Recognizing these failure patterns is essential when selecting nodes for aspiration(Stack et al. 2012).

2.2.6 Summary of the Role of Conventional Ultrasound in Lateral Cervical Lymph Nodes

Conventional ultrasound is indispensable for initial detection, anatomical mapping, and image-guided sampling. Its main strengths are accessibility, real-time assessment, and high spatial resolution for superficial structures.

Nevertheless, overlap between benign and metastatic appearances, operator dependence, and limited sensitivity for early or focal disease prevent ultrasound from serving as a definitive standalone test. Its most appropriate role is as the foundation of a broader multimodal diagnostic pathway.

Taken together, the evidence supports a two-level interpretation of conventional ultrasound. At the first level, highly specific findings such as cystic change, punctate echogenic foci, or focal hyperechogenicity may justify direct sampling. At the second level, nodes with less specific combinations require structured assessment using multiple morphological and vascular features. This distinction is important because a binary qualitative label may obscure the degree of diagnostic confidence(Murata 2018).

Quantitative or weighted scoring may serve as a pragmatic intermediate step between visual interpretation and complex machine-learning analysis. Such scores remain clinically interpretable, can be applied in real time, and provide a reproducible comparator for evaluating whether CEUS, FNA-Tg, or radiomics offers true incremental value.

2.3 FINE-NEEDLE ASPIRATION THYROGLOBULIN (FNA-TG) IN LATERAL CERVICAL LYMPH NODE METASTASIS

Fine-needle aspiration thyroglobulin (FNA-Tg) measures thyroglobulin in needle washout fluid from a suspicious lymph node. It complements cytology, particularly when aspirates are cystic, necrotic, or paucicellular, and provides a thyroid-specific biochemical marker of metastatic tissue (Jeon et al. 2013).

FNA-Tg is especially useful in the lateral neck, where contamination from the thyroid bed is less likely than in the central compartment. Its value, limitations, and role in clinical decision-making are reviewed below (Pacini et al. 1992; Wang et al. 2020).

FNA-Tg is usually performed during ultrasound-guided aspiration by rinsing the needle with a defined volume of saline and measuring thyroglobulin in the washout. The result is interpreted together with cytology, serum thyroglobulin, anti-thyroglobulin antibodies, and imaging findings (Wang et al. 2020). The method is particularly valuable when cytology is compromised by cystic fluid, haemorrhage, fibrosis, or low cellularity. In these situations, a biochemical marker may remain positive even when malignant cells are not recovered.

FNA-Tg occupies a distinctive position between imaging and pathology. It does not provide direct histological proof, but it offers tissue-specific biochemical evidence that is often more sensitive than cytology in cystic or paucicellular nodes. Its interpretation is strongest when the sampled node has been accurately localized and when the result is concordant with imaging and clinical context.

Pre-analytical factors are central to reliability. Needle gauge, number of passes, washout volume, saline composition, sample transport, assay platform, and the interval between aspiration and analysis may all influence measured concentrations. Transparent reporting of these details is therefore necessary for cross-study comparison and clinical reproducibility (Boux de Casson et al. 2017).

2.3.1 Biological Rationale of FNA-Tg for Lateral Cervical Lymph Nodes

Thyroglobulin is produced by normal thyroid follicular cells and most differentiated thyroid carcinoma cells. Metastatic PTC cells may retain this secretory function after lymphatic spread, allowing thyroglobulin to accumulate in solid or cystic nodal tissue.

Detection of thyroglobulin in lymph node washout fluid therefore provides evidence of thyroid-derived tissue even when cytology is non-diagnostic. The anatomical distance of lateral nodes from the thyroid reduces contamination and generally improves specificity (Kim et al. 2021).

The diagnostic principle depends on tissue specificity rather than tumour morphology. A metastatic node containing differentiated thyroid cells can release large amounts of thyroglobulin into the aspirate, producing concentrations far above background serum values (Deng et al. 2022).

However, poorly differentiated tumour components may produce less thyroglobulin, and extensive necrosis or sampling outside the metastatic focus may lower the measured value. Thus, a negative result does not absolutely exclude metastasis when imaging suspicion remains high.

2.3.2 Diagnostic Performance of FNA-Tg in Lateral Cervical Lymph Node Metastasis

Most studies report high sensitivity and specificity for FNA-Tg, frequently exceeding the performance of cytology alone in cystic or indeterminate nodes. Meta-analytic estimates are approximately 94% sensitivity and 92% specificity, although performance varies with assay method, threshold, and patient population (Jeon et al. 2015; Wang et al. 2020).

Its principal advantage is objectivity: a quantitative biochemical result is less dependent on visual interpretation than ultrasound. Combining FNA-Tg with cytology or imaging improves sensitivity and negative predictive value and reduces missed metastases. Diagnostic performance improves when FNA-Tg is combined with cytology because the two tests fail for different reasons. Cytology provides direct cellular confirmation, whereas FNA-Tg is sensitive to thyroid-derived protein even in acellular fluid (Yu et al. 2025; Wang et al. 2022).

The greatest incremental value is observed in cystic nodes, previously treated patients, and nodes with non-diagnostic cytology. In clearly benign-appearing nodes, routine aspiration is unnecessary; in unequivocally malignant nodes, FNA-Tg may provide confirmation before compartment dissection.

The apparent diagnostic superiority of FNA-Tg over cytology is most pronounced in cystic metastases, where malignant epithelial cells may be sparse or absent despite abundant thyroglobulin in the fluid (Shin et al. 2015). Conversely, cytology remains valuable in non-thyroid malignancies and in poorly differentiated lesions that may produce little thyroglobulin (Su et al. 2024). The two tests should therefore be viewed as complementary rather than competing procedures.

Reported performance also depends on whether analysis is performed at the patient level or node level. Patient-level classification may appear favorable when one positive node determines the outcome, whereas node-level analysis is more demanding and more relevant to selecting individual targets for surgery or follow-up.

2.3.3 Controversies Surrounding FNA-Tg Cut-off Values in Lateral Cervical Lymph Nodes

No universal FNA-Tg cut-off has been established. Reported thresholds range widely because of differences in assay calibration, dilution protocols, serum thyroglobulin, anti-thyroglobulin antibodies, and sample handling.

Thresholds may also depend on nodal location and clinical context. Lower values may be more meaningful in the lateral neck because contamination is less likely, but partial nodal involvement and sampling error can produce borderline results. Quantitative values should therefore be interpreted together with imaging and cytology rather than as an isolated binary test. (Kim et al. 2012; Jeon et al. 2015; Mikosiński et al. 2023)

Some studies use a fixed concentration threshold, whereas others compare washout thyroglobulin with serum thyroglobulin or use receiver operating characteristic analysis to derive an optimal cut-off. These approaches are not directly interchangeable. Assay-specific reference ranges, washout volume, and laboratory platform should therefore be reported. A locally validated threshold may be more reliable than adoption of a universal value from another institution. Continuous quantitative interpretation may also preserve more information than simple positive-negative classification (Mikosiński et al. 2023, Sakamoto et al. 2024).

Three broad interpretive strategies have been used: a fixed absolute threshold, comparison with the concurrent serum thyroglobulin concentration, and a study-specific ROC-derived cut-off. Fixed thresholds are simple but may not transfer across assays; serum-referenced ratios may reduce background effects but introduce additional biological variability; ROC-derived thresholds optimize a particular dataset but require external validation (Jeon et al. 2015; Liu et al. 2021). For this reason, continuous quantitative analysis may be more informative than dichotomization. A very high FNA-Tg value conveys stronger evidence than a borderline elevation, and integration with ultrasound or CEUS findings can generate a clinically interpretable probability rather than a rigid positive-negative label.

Three principal threshold strategies recur in the FNA-Tg literature. The first uses a fixed absolute concentration, which is simple and readily applicable but highly dependent on assay calibration, washout volume, and laboratory platform. The second interprets washout Tg relative to the concurrent serum Tg concentration, which may reduce the influence of circulating background but introduces additional biological variability. The third derives an optimal threshold from receiver operating characteristic analysis within the study cohort; this approach maximizes apparent discrimination locally but may be optimistic and requires validation before transfer to another population (Jeon et al. 2013; Duval et al. 2017).

These strategies alter the balance between sensitivity and specificity. A low absolute cut-off may detect small-volume or cystic metastases but can increase false-positive results through blood contamination or residual thyroid-derived Tg. A higher threshold improves specificity but may miss focal metastatic deposits, poorly differentiated components, or samples obtained outside the involved portion of a heterogeneous node. The lateral compartment provides a favorable context because contamination from the thyroid bed is less likely than in central nodes, yet anatomical separation does not eliminate assay and sampling variability (Kim and Bae 2021).

Accordingly, FNA-Tg is better understood as graded biochemical evidence than as a universally transferable binary result. A markedly elevated value in a cystic lateral node with suspicious imaging is substantially more persuasive than a borderline elevation in a small node with otherwise benign morphology. Continuous interpretation, or integration of the measured concentration with cytology and imaging, preserves more information than dichotomisation and may be more clinically meaningful. Nevertheless, any locally derived threshold should be reported with the assay method, dilution protocol, serum Tg and anti-thyroglobulin antibody status, and should be externally validated before widespread adoption (Wang et al. 2020).

2.3.4 Clinical Decision-Making Value of FNA-Tg in Lateral Cervical Lymph Nodes

FNA-Tg can provide decisive preoperative confirmation when imaging is equivocal or cytology is inadequate. A clearly elevated result supports therapeutic lateral neck

dissection, while a low result may justify repeat sampling, additional imaging, or surveillance when other high-risk features are absent (Haugen et al. 2016).

Its limitations are equally important. FNA-Tg is invasive, node-specific, and susceptible to sampling error; it cannot evaluate all nodes simultaneously and provides no information on architecture, perfusion, or whole-node heterogeneity. It is best used as a confirmatory biochemical component of an integrated diagnostic strategy (Jeon et al. 2015). Discordant results require careful interpretation. A suspicious node with negative cytology but markedly elevated FNA-Tg is likely metastatic, whereas a mildly elevated FNA-Tg result in a node close to the thyroid may require repeat sampling or correlation with other findings (Shin et al. 2015). Because aspiration samples only a small part of one node, FNA-Tg does not solve the problem of multiple nodal assessment. Functional imaging and quantitative image analysis may help identify which node should be sampled and whether additional suspicious nodes are present (Wang et al. 2020).

Evidence synthesis suggests that FNA-Tg has its greatest clinical value in three circumstances: cystic or paucicellular nodes, discordance between imaging and cytology, and small nodes in which cytological sampling is technically difficult. In these settings, thyroglobulin provides a tumor-lineage signal that is complementary to morphology rather than redundant with it (Zhu et al. 2020).

However, the clinical meaning of FNA-Tg depends on how the result changes management. A positive value is most useful when it converts an indeterminate node into a surgically actionable diagnosis. Conversely, a negative result should not automatically override highly suspicious imaging because partial replacement, poorly differentiated components, or sampling outside the tumor focus may produce false-negative measurements. This argues for continuous quantitative analysis and multimodal interpretation rather than exclusive reliance on a single threshold.

2.4 CONVENTIONAL SONOVUE CONTRAST-ENHANCED ULTRASOUND IN THE EVALUATION OF LATERAL CERVICAL LYMPH NODES

SonoVue contrast-enhanced ultrasound (CEUS) depicts microvascular perfusion in real time and can reveal vascular heterogeneity not visible on grayscale or Doppler imaging.

It is radiation-free, has high temporal resolution, and can be performed during the same ultrasound session (Yuan et al. 2006).

Because SonoVue is a pure intravascular blood-pool agent, its information is confined to the arterial and venous phases. This biological constraint limits its ability to distinguish inflammatory hyperperfusion from tumour-related neovascularity. CEUS is particularly attractive for nodes with equivocal Doppler findings because microbubbles can depict low-volume microvascular flow that is not visible on conventional CDFI. Dynamic observation also allows assessment of enhancement sequence and regional perfusion defects (Zhan et al. 2019).

However, perfusion is not tumour-specific. Both inflammation and malignancy can increase vascularity, while necrosis can reduce enhancement. The clinical value of CEUS therefore depends on integration with morphology and the biological properties of the contrast agent. CEUS improves visualization of microvascular flow by suppressing background tissue signals and selectively displaying nonlinear microbubble responses. This allows the examiner to observe enhancement direction, distribution, intensity, perfusion defects, and wash-out in real time. These features may reveal functional abnormalities before gross morphological changes become evident (Wei et al. 2014).

Nevertheless, early-phase enhancement primarily reflects haemodynamics. Reactive hyperplasia and metastatic angiogenesis can both produce rapid or heterogeneous enhancement, whereas necrotic metastases may appear hypoenhancing. The diagnostic ceiling of blood-pool CEUS is therefore biological rather than merely technical (Wilson and Burns 2010; Zhan et al. 2019; Yu et al. 2023).

2.4.1 Principles of Conventional SonoVue Contrast-Enhanced Ultrasound

SonoVue consists of sulphur hexafluoride microbubbles enclosed by a phospholipid shell. At a low mechanical index, microbubble oscillation generates nonlinear signals that enhance visualization of capillary perfusion (Bonini et al. 2007).

Assessment is based on wash-in, enhancement distribution, peak intensity, and wash-out. Quantitative time-intensity analysis can improve objectivity, but results depend on acquisition protocol, motion control, and region-of-interest selection. (Wilson and Burns 2010; Zhang et al. 2023)

Qualitative analysis describes enhancement direction, homogeneity, intensity, and wash-out. Quantitative analysis uses time-intensity curves to measure parameters such as arrival time, time to peak, peak intensity, wash-in slope, and wash-out rate (Zhang et al. 2023).

Quantitative analysis may reduce observer variability, but reliable measurement requires stable imaging, consistent machine settings, avoidance of transducer pressure, and reproducible placement of the region of interest. These requirements limit routine application.

2.4.2 Diagnostic Performance of Conventional SonoVue CEUS for Lateral Cervical Lymph Nodes

Reactive nodes often enhance rapidly and homogeneously from the hilum, whereas metastatic nodes may show peripheral or heterogeneous enhancement, perfusion defects, and early wash-out. These patterns are useful but not specific because inflammation, necrosis, and small tumor burden may produce overlapping appearances (Hong et al. 2017).

Reported sensitivity is generally 70–90% and specificity 65–85%. SonoVue CEUS may improve confidence in indeterminate nodes, but it rarely provides definitive discrimination when used alone. (Zhan et al. 2019; Yu et al. 2023). Homogeneous hilar enhancement is generally associated with benign nodes, while heterogeneous or peripheral enhancement raises suspicion for metastasis. Non-enhancing areas may correspond to necrosis or cystic degeneration but are not exclusive to malignancy.

Variation in diagnostic criteria is a major reason for inconsistent performance across studies. Some investigators emphasize enhancement direction, while others focus on homogeneity or quantitative parameters. Standardized definitions are therefore necessary for meaningful comparison (Zhan et al. 2019; Yu et al. 2023).

Differences in reported accuracy arise from inconsistent definitions of enhancement patterns. Terms such as heterogeneous enhancement, peripheral enhancement, centripetal filling, and perfusion defect are not always operationalized in the same way. Variation in contrast dose, mechanical index, recording duration, and reference standard further limits pooled interpretation(Hong et al. 2017; Chen et al. 2022).

Quantitative time-intensity analysis can reduce visual subjectivity, but it introduces its own sources of error, including respiratory and swallowing motion, unstable probe position, non-representative ROI placement, and differences in curve-fitting software. The reliability of quantitative CEUS therefore depends on strict acquisition and analysis protocols.

2.4.3 Biologically Driven Failure Patterns of Conventional SonoVue CEUS in Lateral Cervical Lymph Nodes

Inflammatory angiogenesis can produce intense or heterogeneous enhancement that mimics metastasis, whereas micrometastases may not sufficiently alter perfusion to be detected. Technical factors, transducer pressure, and physiological variation further affect enhancement (Hong et al. 2017).

These failures reflect the blood-pool nature of SonoVue: it depicts perfusion but not tissue-specific retention. Consequently, CEUS findings should be interpreted with structural imaging and biochemical or pathological confirmation (Yu et al. 2023).

2.4.4 Summary of Conventional SonoVue CEUS for Lateral Cervical Lymph Nodes

SonoVue CEUS adds functional perfusion information to conventional ultrasound and can improve sensitivity in equivocal nodes. Its main limitations are overlap between inflammatory and malignant enhancement, subjective interpretation, brief imaging duration, and absence of a tissue-specific delayed phase (Chen et al. 2022).

These constraints explain the interest in Sonazoid, which provides both vascular-phase imaging and a post-vascular phase related to reticuloendothelial uptake (Xue et al. 2023).

The principal contribution of SonoVue CEUS is improved visualization of microvascular perfusion. Its principal limitation is equally clear: perfusion is a shared biological response in malignancy and inflammation. This explains why CEUS frequently improves sensitivity and lesion conspicuity without consistently achieving tumour-specific discrimination (Wei et al. 2014).

From a research perspective, SonoVue studies provide an important benchmark. They demonstrate that enhancement can be quantified and that phase-specific assessment is feasible, but they also show that a purely intravascular agent cannot address all sources of diagnostic overlap (Radzina et al. 2021). This limitation provides the biological rationale for investigating Sonazoid post-vascular imaging.

2.5 SONAZOID CONTRAST-ENHANCED ULTRASOUND FOR LATERAL CERVICAL LYMPH NODES

Sonazoid is a second-generation perfluorobutane microbubble agent that combines real-time vascular imaging with a prolonged post-vascular phase. Its stability and reticuloendothelial uptake provide a longer observation window and potentially more biologically informative assessment of lymphatic tissue than conventional blood-pool agents. (Zhang et al. 2023)

Sonazoid has been used extensively for liver imaging, where the post-vascular phase improves lesion detection by demonstrating defects in normal reticuloendothelial uptake. Translation of this principle to lymph nodes is biologically plausible because normal lymphoid tissue also contains macrophages.

The ability to revisit a node several minutes after injection may reduce the time pressure associated with short vascular-phase imaging. It also permits assessment of multiple nodal levels using a single contrast dose. Sonazoid differs from conventional blood-pool agents because its phospholipid-stabilized perfluorobutane microbubbles remain sufficiently stable for delayed imaging. After the vascular phase, uptake by phagocytic cells creates a post-vascular phase that may reflect the integrity of macrophage-containing tissue. This property introduces a tissue-related dimension that is absent from purely intravascular agents (Barr et al. 2020).

In lymph nodes, the biological interpretation of post-vascular enhancement is plausible but more complex than in the liver. Normal nodal sinuses and macrophages may retain microbubbles, while metastatic replacement, necrosis, fibrosis, or inflammation may alter retention. Accordingly, post-vascular defects should be interpreted as evidence of disrupted nodal tissue rather than as a tumour-specific sign in isolation.

2.5.1 Introduction of Sonazoid Contrast-Enhanced Ultrasound Agent

Sonazoid microbubbles contain perfluorobutane gas stabilized by a phospholipid shell. Their resistance to acoustic pressure permits prolonged enhancement and repeated scanning at a low mechanical index. After the vascular phase, microbubbles are taken up by phagocytic cells of the reticuloendothelial system, creating a post-vascular phase (Yoo et al. 2026).

Although this property is established in hepatic imaging, emerging evidence indicates that it may also be relevant to lymph nodes because normal lymphoid tissue contains macrophages and sinusoidal structures capable of contrast uptake (Zhang et al. 2023).

2.5.2 Micro-perfusion Imaging in the Vascular Phase and Macrophage-Mediated Uptake in the Post-Vascular Phase of Sonazoid

During the early vascular phase, Sonazoid depicts arterial inflow, microvascular distribution, and wash-out in a manner broadly comparable to other blood-pool agents. These findings reflect perfusion but may overlap between inflammatory and malignant nodes (Wei et al. 2022).

The post-vascular phase provides additional information. Preserved lymphoid tissue may retain contrast through macrophage-mediated uptake, whereas metastatic replacement or architectural destruction may reduce or alter retention. This phase therefore has the potential to reflect tissue integrity rather than perfusion alone (Machado et al. 2025). Vascular phase abnormalities may indicate altered angiogenesis, but their specificity remains limited. The post-vascular phase may provide

complementary evidence by demonstrating the degree to which normal macrophage-rich architecture is preserved or replaced.

A metastatic node may show complete or partial post-vascular defects depending on tumor burden. Focal defects may correspond to partial replacement, whereas homogeneous retention may support benignity. These interpretations require pathological correlation and standardized timing.

The multiphase nature of Sonazoid permits sequential assessment of distinct biological processes. Arterial-phase findings mainly reflect inflow and early microvascular distribution; venous-phase findings reflect continued perfusion and wash-out; and post-vascular findings may reflect phagocytic uptake and residual tissue architecture. Direct comparison of these phases can clarify which phase provides the greatest diagnostic discrimination (Barr et al. 2020; Zhang et al. 2023).

The post-vascular phase may also be advantageous methodologically. The longer observation window reduces dependence on second-by-second timing, allows repeated scanning of multiple nodal levels, and may yield more stable images for segmentation and radiomics. These potential advantages require empirical validation rather than assumption.

2.5.3 Comparison Between Sonazoid and Conventional SonoVue CEUS Agents

Both SonoVue and Sonazoid permit real-time microvascular imaging, but their biological behaviour differs (**Table 2.1**). SonoVue remains intravascular and is cleared rapidly, whereas Sonazoid is more stable and supports delayed post-vascular imaging.

The practical advantage of Sonazoid is not simply longer enhancement. It offers a second diagnostic dimension related to tissue retention, enabling assessment of multiple neck compartments after a single injection and providing a potentially more reproducible substrate for quantitative analysis (Li et al. 2021).

2.5.4 Advantages and Performance of Sonazoid CEUS for Lateral Cervical Lymph Nodes

Early studies indicate that Sonazoid CEUS can improve characterization of cervical lymph nodes, particularly when post-vascular enhancement is evaluated. Persistent homogeneous enhancement may favour benign lymphoid tissue, whereas focal or diffuse defects may indicate metastatic replacement (Liu et al. 2026).

Sonazoid may be especially useful for multiple or equivocal lateral nodes because of its prolonged imaging window. However, published evidence remains limited, and differences in administration route, phase timing, and diagnostic criteria restrict direct comparison across studies.

The prolonged post-vascular phase may improve lesion conspicuity and reduce variation caused by the exact timing of image acquisition. It may also facilitate side-by-side comparison among multiple nodes in the same patient (Liu et al. 2026). Importantly, Sonazoid should not be viewed as a replacement for conventional ultrasound or FNA-Tg. Its strongest role may be in resolving indeterminate nodes, selecting aspiration targets, and contributing phase-specific information to an integrated model.

A clinically important application is the assessment of small nodes that are morphologically indeterminate (Liu et al. 2026). In such nodes, a focal post-vascular defect may reveal partial metastatic replacement that is not apparent on grayscale imaging. Conversely, homogeneous delayed retention may support benignity when early vascular enhancement is nonspecific.

The strongest evidence for Sonazoid is likely to emerge from integrated interpretation rather than a single qualitative sign (Xiao et al. 2023). Phase-specific scores, quantitative measurements, and combination with FNA-Tg may improve robustness by linking tissue-related imaging with biochemical confirmation.

2.5.5 Limitations of Sonazoid CEUS for Lateral Cervical Lymph Nodes

Sonazoid CEUS remains operator-dependent and requires standardized acquisition, timing, and interpretation. Inflammation, fibrosis, necrosis, and incomplete metastatic replacement may affect post-vascular patterns and cause overlap between benign and

malignant nodes. Evidence is also derived mainly from single-centre studies with limited sample sizes. The technique should therefore be considered promising but not yet definitive, and its incremental value over conventional ultrasound and FNA-Tg requires systematic evaluation (Sasaki et al. 2012).

Interpretation of post-vascular enhancement is still evolving. Macrophage distribution may be altered by inflammation, fibrosis, previous biopsy, or treatment, potentially producing false-positive or false-negative findings. Availability, cost, regulatory approval, and operator familiarity also differ among countries and institutions. These practical factors may influence external applicability even when diagnostic performance is favorable.

The post-vascular phase should not be interpreted as a tumor-specific equivalent of histopathology. Its biological rationale is based on microbubble retention by phagocytic cells and preservation of macrophage-containing lymphoid tissue; however, nodal macrophage distribution is heterogeneous and can be altered by reactive hyperplasia, inflammation, fibrosis, necrosis, previous aspiration, or partial tumor replacement. A focal or diffuse defect is therefore plausibly associated with disrupted nodal architecture, but it cannot independently distinguish every metastatic node from every benign inflammatory node (Wilson & Burns 2010; Barr et al. 2020; Zhang et al. 2023).

This limitation is important when extrapolating from hepatic Kupffer-phase imaging. The liver has a comparatively well-characterised reticuloendothelial background, whereas lymph nodes contain variable sinuses, follicles, germinal centers, stromal elements, and macrophage populations. The strength of post-vascular Sonazoid assessment lies in providing a tissue-related dimension beyond transient perfusion, not in creating an inherently pathognomonic sign. Interpretation should therefore remain integrated with grayscale morphology, vascular-phase enhancement, clinical context, and—where required—cytological or biochemical confirmation (Wei et al. 2022; Chung et al. 2022).

Methodologically, the longer post-vascular window may offer more stable imaging, repeated interrogation, and improved opportunity to assess multiple nodes after a single injection. These characteristics may reduce sensitivity to exact second-by-second timing and may create a more reproducible substrate for segmentation and quantitative analysis. Nevertheless, stability of acquisition does not automatically guarantee diagnostic validity; direct pathological correlation, interobserver assessment, standardised phase timing, and multicenter reproducibility remain necessary before delayed-phase criteria can be generalised (Chen et al. 2021).

2.5.6 Summary of Sonazoid CEUS for Lateral Cervical Lymph Nodes

Sonazoid extends CEUS beyond perfusion by providing a post-vascular phase that may reflect macrophage distribution and nodal microstructural integrity. This biological distinction is the principal reason for its potential advantage in LLNM assessment.

Current evidence supports further evaluation of phase-specific qualitative and quantitative criteria, particularly in combination with conventional ultrasound, FNA-Tg, and radiomics.

The available literature indicates that the diagnostic value of Sonazoid should be evaluated at two complementary levels. Vascular-phase imaging assesses micro-perfusion and tumor-related angiogenesis, whereas post-vascular imaging may reflect macrophage distribution and preservation of lymphoid architecture. Combining these phases may therefore characterize both blood flow and tissue integrity (Liu et al. 2026).

Nevertheless, the evidence base remains smaller than that for conventional ultrasound and FNA-Tg. Studies differ in route of administration, phase definitions, qualitative signs, and quantitative parameters. Direct within-cohort comparison of arterial, venous, and post-vascular phases is therefore necessary before the post-vascular phase can be positioned as the preferred diagnostic window.

2.6 RADIOMICS IN THE EVALUATION OF CERVICAL LYMPH NODE METASTASIS

Radiomics converts medical images into quantitative descriptors of intensity, texture, shape, and spatial heterogeneity. It may identify subtle patterns not consistently recognized by visual assessment and can support machine-learning classification (Lambin et al. 2012). Its clinical value depends on the quality and biological relevance of the source image, robust segmentation, reproducible feature extraction, appropriate feature selection, and independent validation. (Lambin et al. 2017)

Radiomics aims to quantify image heterogeneity that may reflect microscopic tissue organization. First-order features describe intensity distribution; texture matrices describe spatial relationships; and shape features characterize geometry. Wavelet or filtered features may reveal patterns at different spatial scales (Lambin et al. 2012; Gillies et al. 2016; Lambin et al. 2017). Because hundreds of features can be extracted from a modest sample, radiomics is vulnerable to overfitting. Reproducibility testing, dimensionality reduction, penalized selection, cross-validation, and an independent test cohort are essential components of credible model development.

Radiomics provides a formal framework for analysing image heterogeneity beyond visual description. First-order features summarize intensity distributions; shape features quantify geometry; and texture matrices describe spatial relationships among pixel or voxel values. Filtered and wavelet-derived features can emphasize patterns at different spatial frequencies, potentially capturing subtle microstructural variation (Coroller et al. 2017).

The large number of candidate features creates a high-dimensional problem in which apparent model performance can arise from chance correlations. Reliable studies therefore require reproducibility screening, removal of redundant features, penalized feature selection, internal validation, and an independent test cohort. Model complexity should be proportionate to the available sample size (Liu et al. 2019; Tran et al. 2019).

2.6.1 Conceptual Foundations of Radiomics in Cervical Lymph Node Imaging

Radiomics does not create new biological information; it reorganizes image signals into quantitative variables. Features may reflect heterogeneity associated with tumor infiltration, necrosis, vascular remodeling, and disruption of normal nodal architecture. (Lambin et al. 2012)

Ultrasound radiomics is attractive because ultrasound is already central to thyroid cancer care, but it is particularly vulnerable to variation in equipment, settings, compression, operator technique, and image selection. Model performance must therefore be interpreted in relation to the imaging substrate. Biological interpretability is strengthened when features are linked to plausible tissue characteristics, such as heterogeneity from necrosis, disruption of nodal architecture, or uneven contrast retention (Scheckenbach et al. 2017).

Nevertheless, radiomics features remain mathematical surrogates and should not be equated directly with histopathology. Model performance may also reflect technical artefacts. Differences in gain, depth, focal zone, compression, and image preprocessing can alter feature values. Standardization is therefore a prerequisite for generalization.

2.6.2 Radiomics Workflow in Cervical Lymph Node Imaging

A typical workflow includes image acquisition and standardization, region-of-interest segmentation, feature extraction, reproducibility testing, dimensionality reduction, model training, and validation(Rabasco Meneghetti et al. 2022). Common feature families include first-order, shape, GLCM, GLRLM, GLSZM, GLDM, and NGTDM descriptors. To reduce overfitting, feature selection should be confined to the training data and combined with cross-validation. Performance evaluation should extend beyond AUC to calibration, confusion matrices, decision-curve analysis, and independent testing(Gillies et al. 2016).

Segmentation is a major source of variability. Manual delineation is clinically intuitive but time-consuming and observer-dependent; semi-automatic methods may improve efficiency but still require correction. Inter- and intra-observer reproducibility should be quantified before modelling (Scheckenbach et al. 2017).

Feature selection should be performed only within the training cohort to avoid information leakage. The final model should then be evaluated in untouched test data using discrimination, calibration, and clinical utility metrics (Zhang et al. 2020).

Image standardization should precede feature extraction. Differences in pixel spacing, grey-level discretization, image depth, gain, dynamic range, and preprocessing can materially alter feature values. Reproducible radiomics therefore depends on predefined acquisition settings and transparent reporting of normalization and resampling procedures (Wen et al. 2022).

Segmentation uncertainty is another major source of variation. Manual ROI delineation may exclude surrounding tissue and focus on the visible lesion, but it is time-consuming and observer-dependent. Inter- and intra-observer ICC analysis can identify robust features, while sensitivity analysis can determine whether model performance changes with alternative contours (Zhang et al. 2020; Zhang et al. 2023).

Data leakage must be avoided throughout the pipeline. Feature scaling, feature selection, hyperparameter tuning, and class balancing should be performed within the training data only. Applying these steps before the train-test split can produce optimistically biased performance that will not generalize.

2.6.3 Diagnostic Performance of Radiomics for Central Lymph Node Metastasis

Most ultrasound radiomics studies in PTC have focused on predicting central lymph node metastasis from primary tumor images. Many reports show promising discrimination, but results are heterogeneous because of differences in cohorts, segmentation methods, feature pipelines, and validation strategies (Lambin et al. 2017; Liu et al. 2024). These studies establish methodological feasibility but cannot be

directly extrapolated to lateral nodes, which differ in anatomy, prevalence, biological behavior, and clinical consequences.

Published models for central nodal metastasis frequently analyze the primary thyroid tumor rather than the lymph node itself. Such models estimate a patient-level risk of occult central disease and may be useful for surgical planning, but they answer a different question from direct classification of a suspicious lateral node (Zhou et al. 2020; Shi et al. 2022). This distinction is important because patient-level prediction cannot determine which specific lateral node is metastatic or guide node-targeted aspiration.

2.6.4 Diagnostic Performance of Radiomics for Lateral Lymph Node Metastasis

Evidence specifically addressing radiomics of lateral cervical lymph nodes is limited. Available studies suggest that grayscale and Doppler features may contribute to classification, but performance is affected by small sample sizes, class imbalance, and inadequate external validation (Park et al. 2020).

Direct analysis of the lymph node, rather than prediction from the primary tumour alone, may provide more lesion-specific information (Ardakani et al. 2018). However, the most informative ultrasound modality and phase for this purpose remain uncertain. Direct nodal radiomics offers the potential to classify an individual suspicious node (Zhu et al. 2023). It can therefore be linked more readily to cytology, FNA-Tg, pathology, or follow-up as a node-specific reference standard.

However, sample sizes are often limited because lateral nodes meeting aspiration or surgical criteria are less common than primary thyroid lesions (Tong et al. 2022). This increases the importance of conservative feature selection and transparent validation.

Direct node-based modelling is clinically attractive because each prediction can be linked to a specific node and its pathological or follow-up outcome (Zhang et al. 2020; Feng et al. 2025). However, multiple nodes from the same patient introduce

statistical dependence. Randomly splitting nodes rather than patients may place correlated observations in both training and test cohorts and inflate performance. Patient-level partitioning is therefore preferable. Class imbalance is another concern. If metastatic and non-metastatic nodes are unequally represented, accuracy alone may be misleading. Sensitivity, specificity, precision, recall, F1 score, AUC, calibration, and decision-curve analysis provide a more complete evaluation (Tong et al. 2022).

2.6.5 Differential Performance of Ultrasound-Based Radiomics Models in Central and Lateral Cervical Compartments

Central and lateral nodal disease should not be treated as interchangeable modelling targets. Central nodes are smaller, deeper, and more difficult to image, whereas lateral nodes are more accessible but frequently affected by reactive changes (Wang et al. 2024; Han et al. 2025).

These anatomical and biological differences influence segmentation quality, feature stability, disease prevalence, and classifier performance (Zou et al. 2021; Tian et al. 2024). Compartment-specific model development and validation are therefore required.

2.6.6 Contrast-Enhanced Ultrasound-Based Radiomics in Lateral Lymph Node Metastasis

CEUS-based radiomics incorporates quantitative information from enhancement intensity and spatial heterogeneity. It may capture perfusion patterns more comprehensively than qualitative CEUS interpretation, but conventional blood-pool CEUS remains limited by short acquisition windows and perfusion overlap (Huang et al. 2022). Phase selection is crucial. Early vascular images may be sensitive to timing and motion, whereas delayed images may provide more stable tissue-related information if the contrast agent supports a post-vascular phase (Liu et al. 2026).

CEUS radiomics may capture enhancement heterogeneity, perfusion defects, and spatial variation that are difficult to summarize visually. Static phase-specific images are easier to analyze than complete cine loops, but they may lose temporal information (Chen et al. 2020; Xue et al. 2023). Future approaches may combine static radiomics with time-intensity parameters or deep temporal modelling. Direct

comparison of standardized images from distinct phases provides a practical way to determine how the biological content of the imaging substrate influences model performance.

CEUS radiomics may quantify spatial heterogeneity in perfusion and contrast retention that is difficult to characterize visually. Yet the biological meaning of a feature depends on the selected phase. A texture feature extracted from an arterial image may reflect vascular distribution, while the same mathematical feature extracted from a post-vascular image may reflect tissue retention and architectural integrity(Lentge et al. 2025).

This phase dependence means that classifier comparison alone is insufficient. A modest classifier applied to a biologically informative image may outperform a sophisticated algorithm trained on a less relevant substrate. Comparative radiomics should therefore evaluate both the imaging modality and the modelling method.

2.6.7 Potential Applications of Sonazoid CEUS Radiomics in Lateral Cervical Lymph Node Metastasis

Sonazoid post-vascular images may be a particularly relevant radiomics substrate because they provide a longer acquisition window and may reflect macrophage distribution and structural replacement within lymph nodes. These characteristics could improve feature stability and biological interpretability (Lu et al. 2025). However, this hypothesis has not been sufficiently tested in lateral cervical nodes. Comparative modelling across grayscale, Doppler, arterial, venous, and post-vascular images is needed to determine whether the imaging substrate influences predictive performance more strongly than classifier selection.

Sonazoid post-vascular radiomics is conceptually attractive because the delayed phase offers a relatively stable acquisition window and may encode spatial patterns of preserved and replaced lymphoid tissue. Compared with rapidly changing vascular-phase images, this may reduce timing-related variability and improve the reproducibility of texture features (Xiao et al. 2023). However, an apparently superior model may reflect differences in acquisition stability rather than superior biological

information. To distinguish these explanations, radiomics models should be developed under the same segmentation, preprocessing, feature-selection, and validation pipeline across all imaging modalities and phases (Jiang et al. 2023). The imaging substrate should therefore be treated as an experimental variable rather than a fixed input.

2.6.8 Methodological Quality, Validation, and Reporting Considerations

The apparent performance of a radiomics model is influenced by study design as much as by the classifier itself. Retrospective single-center cohorts, small test sets, class imbalance, and feature selection performed before data splitting can produce optimistic estimates (Gillies et al. 2016). Transparent separation of training and test data, control of information leakage, and reporting of confidence intervals are therefore essential.

Reproducibility should be assessed at the levels of image acquisition, segmentation, feature extraction, and model calibration (Jiang et al. 2023). Discrimination alone is insufficient for clinical translation: calibration, decision-curve analysis, confusion matrices, and sample-level prediction distributions provide additional evidence about reliability and potential net benefit. External multicenter validation remains the definitive test of generalizability.

Discrimination is only one dimension of model quality. Calibration describes agreement between predicted and observed risk, while decision-curve analysis evaluates net clinical benefit across threshold probabilities (Fu et al. 2024). A model with a high AUC but poor calibration may rank patients correctly yet provide unreliable absolute probabilities.

External validation is essential because ultrasound features are sensitive to equipment, operator technique, and patient population. Temporal validation within the same institution is useful but does not fully test transportability. Multicenter validation and prospective workflow evaluation are required before radiomics can influence surgical decisions(Lu et al. 2025).

Transparent reporting should include cohort assembly, missing data handling, image acquisition, segmentation, feature definitions, preprocessing, feature selection, model tuning, validation strategy, and confidence intervals. Adherence to radiomics quality and prediction-model reporting principles improves reproducibility and allows meaningful comparison across studies.

2.6.9 Validation, Generalizability, and Clinical Utility of Radiomics Models

High apparent discrimination in a training cohort does not by itself establish the clinical reliability of a radiomics model. Feature selection, hyperparameter tuning, and classifier comparison performed repeatedly on a small dataset can produce optimistic performance even when cross-validation is used. Internal train-test splitting is preferable to reporting resubstitution performance, but it does not reproduce the variation in scanners, operators, acquisition settings, patient spectrum, and disease prevalence encountered across institutions. External validation and prospective evaluation therefore remain central requirements for judging generalizability (Lambin et al. 2012; Lambin et al. 2017).

Reproducibility also depends on the stability of the imaging substrate and segmentation. Ultrasound radiomics is particularly vulnerable to gain, depth, focus, dynamic range, compression, transducer pressure, Doppler settings, and selection of a representative frame. Manual delineation introduces observer variability, while automated segmentation may fail when nodal margins are indistinct. (Lambin et al. 2012; Gillies et al. 2016) Reporting should therefore include acquisition standardization, segmentation reproducibility, feature robustness, handling of class imbalance, prevention of data leakage, calibration, confidence intervals, and transparent separation of training and test procedures.

Clinical utility is distinct from statistical discrimination. An AUC describes ranking ability across thresholds but does not indicate whether a model improves decisions to observe, repeat imaging, perform aspiration, or undertake lateral neck dissection. Calibration is required to determine whether predicted probabilities correspond to observed risk, and decision curve analysis can evaluate whether using the

model produces greater net benefit than treating all or no patients across clinically relevant thresholds (Vickers & Elkin 2006). A radiomics model is most likely to be useful when it adds reproducible information to standard clinical assessment and changes management in indeterminate cases, rather than when it merely achieves a numerically higher AUC.

The biological source image should also be considered when comparing algorithms. Grayscale ultrasound, CDFI, arterial-phase CEUS, venous-phase CEUS, and post-vascular Sonazoid images encode different structural and functional processes. A more complex classifier cannot compensate for a poorly standardized or biologically weak substrate (Agyekum et al. 2022). Direct within-cohort comparisons using the same segmentation, feature extraction, selection, and validation framework are therefore more informative than comparisons of headline performance across unrelated studies (Liu et al. 2024; Zhang et al. 2023).

2.7 MULTIMODAL INTEGRATION IN THE PREOPERATIVE ASSESSMENT OF LATERAL CERVICAL LYMPH NODE METASTASIS FOR PAPILLARY THYROID CARCINOMA

No single modality captures all relevant dimensions of LLNM. Conventional ultrasound provides anatomy, CEUS provides perfusion and tissue-related enhancement, FNA-Tg provides biochemical specificity, and radiomics provides quantitative pattern analysis (Wang et al. 2022; Liu et al. 2024). Multimodal integration is therefore biologically and clinically rational, but its incremental value must be demonstrated against single-modality assessment using a common reference standard.

A clinically meaningful integrated pathway may begin with conventional ultrasound, proceed to CEUS for functional characterization, and use FNA-Tg for targeted biochemical confirmation. Radiomics may provide an additional quantitative risk estimate when conventional interpretation remains uncertain (Hong et al. 2017). Integration can be performed through simple diagnostic rules, statistical combinations, or machine-learning models. Simple combinations are easier to interpret and implement, whereas complex models may capture interactions but require stronger validation.

Multimodal integration is justified when each component contributes non-redundant information. Conventional ultrasound describes anatomy and morphology; CEUS characterizes perfusion and, with Sonazoid, delayed tissue retention; FNA-Tg provides thyroid-specific biochemical evidence; and radiomics quantifies image heterogeneity. Their combination may reduce uncertainty created by the limitations of any single modality.

Integration can occur at several levels. Clinicians may use sequential decision rules, statistical models may combine selected predictors, and machine-learning approaches may fuse high-dimensional features. The optimal strategy should balance diagnostic gain, interpretability, procedural burden, cost, and feasibility.

Conventional ultrasound contributes structural and anatomical information; Sonazoid CEUS contributes vascular and post-vascular functional information; FNA-Tg provides thyroid-specific biochemical evidence; and radiomics adds quantitative descriptors of image heterogeneity. These components were evaluated both individually and in combination in relation to the final reference diagnosis of lateral cervical lymph node metastasis. Figure 2.1 illustrates the conceptual framework.

2.7.1 Limitations of Isolated Diagnostic Modalities

Conventional ultrasound is non-invasive but subjective; FNA-Tg is specific but invasive and node-specific; CEUS adds functional information but may show overlapping patterns (Chen et al. 2020; Wang et al. 2022), and radiomics is quantitative but sensitive to acquisition and validation quality.

These complementary limitations explain why isolated tests may produce discordant or indeterminate results and why combination strategies may reduce diagnostic uncertainty.

2.7.2 Theoretical Basis for Multimodal Integration

Multimodal assessment combines structural, vascular, tissue-retention, biochemical, and quantitative information. When modalities measure different biological processes,

their errors are less likely to be identical, creating the potential for incremental diagnostic gain.

Integration should nevertheless remain clinically interpretable. Complex models are not inherently superior unless they improve discrimination, calibration, net benefit, or decision-making compared with simpler approaches (Vickers and Elkin 2006). Anatomical, functional, biochemical, and quantitative modalities are complementary because they measure different aspects of disease. A node may appear morphologically benign yet show abnormal contrast retention or elevated FNA-Tg, while another may look suspicious on ultrasound but prove reactive biochemically (Wang et al. 2022). The most useful combination is not necessarily the one containing the greatest number of variables. Incremental performance, feasibility, invasiveness, cost, and influence on management should all be considered.

Complementarity is most valuable in discordant cases. A morphologically suspicious node with preserved post-vascular enhancement and low FNA-Tg may represent reactive disease, whereas a nearly normal-appearing node with a focal delayed defect and markedly elevated FNA-Tg may represent early metastasis. Integrated interpretation can resolve such conflicts more effectively than repeated reliance on the same type of information(Ding et al. 2023).

The incremental value of adding a modality should be demonstrated rather than presumed. Improvements in AUC, sensitivity, specificity, reclassification, calibration, and net benefit should be weighed against additional contrast administration, invasive sampling, analysis time, and resource requirements.

2.7.3 Role of Radiomics as an Integrative Analytical Layer

Radiomics can translate heterogeneous image patterns into quantitative variables and may serve as an analytical layer within a multimodal framework. Its role is to support, rather than replace, clinical imaging and biochemical interpretation (Lambin et al. 2017).

Models should be developed with strict separation of training and test data, reproducibility assessment, transparent feature selection, and clinically meaningful validation (Agyekum et al. 2022).

2.7.4 Clinical Implications of an Integrated Strategy

An integrated strategy may improve preoperative risk stratification, identify nodes that require aspiration or surgery, and reduce unnecessary lateral neck dissection. It may also improve communication among radiologists, surgeons, endocrinologists, and pathologists (Stack et al. 2012).

The practical objective is not maximal test complexity, but a stepwise pathway in which non-invasive imaging guides targeted biochemical confirmation and quantitative analysis is used when it adds clinically relevant information (Vickers and Elkin 2006). An integrated strategy may reduce unnecessary aspiration by identifying clearly benign nodes and may improve targeting by prioritizing nodes with discordant or high-risk features. It may also help determine whether surgery should include lateral compartment dissection (Wang et al. 2022).

For clinical translation, the output should be understandable to clinicians and linked to a management action. A highly accurate model without a clear role in decision-making has limited practical value.

A stepwise pathway may reserve invasive sampling for nodes that remain indeterminate after conventional and contrast-enhanced ultrasound. This approach could reduce unnecessary aspirations while ensuring that high-risk nodes receive biochemical or cytological confirmation. It may also assist in selecting the most representative node when multiple nodes are present.

For surgical planning, the integrated result should be translated into a clear management category, such as low probability requiring surveillance, intermediate probability requiring aspiration or repeat assessment, and high probability supporting therapeutic dissection when confirmed. Such categories may be more clinically useful than an isolated numerical probability.

The principal justification for multimodal integration is complementarity, but adding tests does not automatically improve clinical care. Incremental value should be distinguished from simple association: a new modality should improve discrimination, calibration, or decision-making beyond information already available from conventional ultrasound and FNA-based assessment. The possibility of correlated information is especially relevant when grayscale, Doppler, CEUS, and radiomics features are derived from the same lymph node. Apparent gains may be small, unstable, or accompanied by greater complexity and reduced interpretability (Su et al. 2024; Wang et al. 2022).

A clinically credible integrated pathway should therefore be selective rather than indiscriminate. Conventional ultrasound remains the whole-neck screening and mapping method; Sonazoid CEUS may refine functional and tissue-related characterization of equivocal nodes; FNA-C and FNA-Tg provide targeted confirmation; and radiomics may support quantitative stratification when image quality and model validity are sufficient. The added burden of contrast administration, aspiration, segmentation, and computation should be justified by fewer indeterminate classifications, more appropriate sampling, or better-informed surgical decisions (Su et al. 2024).

Evaluation should extend beyond AUC to calibration, net benefit, feasibility, reproducibility, and the consequences of false-positive and false-negative classifications. Because unnecessary lateral neck dissection carries substantial morbidity, a model with improved sensitivity but reduced specificity may not be clinically preferable at the thresholds used for surgery. Conversely, a highly specific but insensitive strategy may fail to prevent reoperation. The preferred approach depends on the intended decision and should be assessed at clinically relevant probability thresholds rather than by a single summary metric.

2.7.5 Current Evidence and Future Research Directions in Multimodal Integration

Published studies generally support combinations of ultrasound, CEUS, cytology, and FNA-Tg, but direct comparisons are limited by heterogeneous thresholds, reference

standards, and study designs. Evidence integrating Sonazoid post-vascular imaging with radiomics is particularly sparse.

Future research should emphasize prospective multicenter validation, standardized acquisition and segmentation, external testing, calibration, decision-curve analysis, and transparent reporting. These requirements are essential before integrated models can be adopted in routine practice.

The three complementary reviews consistently support multimodal diagnosis but also demonstrate substantial heterogeneity in the published evidence. Conventional ultrasound, CEUS, FNA-Tg, and radiomics are frequently discussed as individually promising, yet few studies compare them directly in the same patients using the same reference standard. This prevents reliable estimation of incremental rather than isolated diagnostic value (Wang et al. 2022; Liu et al. 2024).

A clinically useful framework should also distinguish between complementary and substitutive tests. Conventional ultrasound and CEUS are largely complementary because they depict different imaging dimensions. FNA-Tg is confirmatory and invasive, while radiomics is an analytical layer derived from imaging. Their roles should therefore be organized in a stepwise pathway instead of being treated as interchangeable predictors.

2.7.6 Proposed Stepwise Integrated Diagnostic Pathway

On the basis of the reviewed evidence, a rational clinical pathway begins with systematic conventional ultrasound for detection, mapping, and initial risk classification. Sonazoid CEUS can then be used to characterize indeterminate nodes by combining vascular-phase perfusion with post-vascular tissue-retention information. FNA-C and FNA-Tg provide targeted pathological and biochemical confirmation when the result is expected to alter surgery.

Radiomics should be positioned as an adjunctive quantitative layer, particularly for nodes that remain equivocal after conventional interpretation. Its output

should be interpreted alongside imaging and biochemical findings rather than used as an autonomous decision. This stepwise structure preserves clinical interpretability while allowing each modality to contribute information at the point where it has the greatest value (Lambin et al. 2017).

The proposed pathway begins with systematic conventional ultrasound and quantitative morphological scoring. Nodes with unequivocally benign features may undergo surveillance, whereas suspicious or indeterminate nodes proceed to Sonazoid CEUS for phase-specific assessment. FNA-C and FNA-Tg are then targeted to nodes in which confirmation would change management.

Radiomics can be positioned as an adjunctive analytical layer, particularly for indeterminate nodes or research settings. Its output should be interpreted alongside imaging and biochemical findings rather than used as an autonomous decision-maker (Park et al. 2021). This preserves clinical interpretability and reduces the risk of algorithm-driven overtreatment.

The pathway is intentionally modular. Institutions without access to Sonazoid or radiomics can still apply the underlying principle of combining structurally, functionally, and biologically distinct information. Future prospective studies should determine the most cost-effective sequence and the threshold at which each additional test provides meaningful clinical benefit.

2.8 SUMMARY AND CRITICAL SYNTHESIS OF THE LITERATURE

Conventional ultrasound provides accessible structural mapping but remains limited by overlap in benign and malignant nodal appearances and by observer dependence. FNA-Tg contributes thyroid-specific biochemical evidence but is invasive, node-specific, and sensitive to sampling procedures, assay characteristics, and threshold selection. Blood-pool CEUS adds microvascular information, although inflammatory and malignant lymph nodes may demonstrate overlapping perfusion responses. Sonazoid further extends the imaging window by introducing post-vascular tissue-related information, but the biological interpretation of this phase in cervical lymph nodes remains plausible

rather than pathognomonic. Radiomics provides an additional quantitative analytical layer, yet its credibility depends on acquisition standardization, feature robustness, appropriate validation, calibration, and demonstrable clinical benefit (Haugen et al. 2016; Wang et al. 2022; Lambin et al. 2017).

The apparent differences among published studies are attributable not only to true biological variation but also to substantial methodological heterogeneity. Patient selection, prevalence of cystic disease, nodal size and compartment, unit of analysis, reference standard, operator expertise, assay platform, CEUS phase definition, segmentation strategy, feature-selection method, and validation design can all influence reported diagnostic performance. Consequently, isolated estimates of sensitivity, specificity, or AUC should not be interpreted as universally transferable rankings of diagnostic methods (Wang et al. 2022; Liu et al. 2024).

Taken together, the literature suggests that the most rational direction is not replacement of one diagnostic modality by another, but structured integration of complementary information. Structural, vascular, tissue-retention, biochemical, and quantitative evidence address different biological dimensions and exhibit different failure modes. Their integration is most defensible when each component provides non-redundant and clinically actionable information, when discordant findings are explicitly considered, and when resulting probability estimates are sufficiently calibrated to support decisions regarding observation, aspiration, or surgery.

Accurate preoperative identification of LLNM remains clinically important because each available method has a diagnostic ceiling. Conventional ultrasound is accessible but operator-dependent; FNA-Tg is biochemically specific but invasive and subject to sampling and threshold variability; SonoVue CEUS depicts microvascular perfusion but does not provide the prolonged tissue-related imaging window available with Sonazoid; and radiomics is promising but highly dependent on image quality, segmentation reproducibility, feature stability, and validation. Sonazoid introduces a post-vascular phase that may provide additional biological information through macrophage-related uptake and tissue retention. However, evidence regarding its application in lateral cervical lymph nodes remains limited, and the comparative

diagnostic value of arterial, venous, and post-vascular phases has not been clearly established. It also remains uncertain whether post-vascular Sonazoid images provide a superior biological substrate for radiomics analysis.

Current evidence is further fragmented because few investigations directly compare qualitative and quantitative assessment of conventional ultrasound, multiphase Sonazoid CEUS, and FNA-Tg within the same node-level cohort and under the same reference standard. Similarly, relatively few studies have evaluated single and combined diagnostic strategies while simultaneously developing modality-specific radiomics models within a unified analytical framework. A common-cohort design would reduce between-study confounding and allow a clearer assessment of the independent and incremental contribution of each diagnostic component (Mulla et al. 2012).

Methodological inconsistency remains particularly important when interpreting the existing literature. Studies differ in the definitions of suspicious ultrasound features, FNA-Tg cut-off values, CEUS protocols, contrast agents, reference standards, and radiomics pipelines. These differences limit direct comparison and external generalizability (Jiang and Hsiao 2020; Chen et al. 2022; Lentge et al. 2025). Therefore, apparently superior performance in one study may reflect differences in study design, disease spectrum, or analytical workflow rather than a true intrinsic advantage of a particular modality.

The evidence for Sonazoid in cervical lymph-node assessment is less mature than that for conventional ultrasound or FNA-Tg. In particular, the biological and diagnostic contribution of the post-vascular phase has not been sufficiently compared with arterial and venous phases within the same patient cohort. This represents an important knowledge gap because the post-vascular phase may reflect tissue characteristics that differ fundamentally from early perfusion-based information.

Radiomics introduces a related but distinct area of uncertainty. Model performance varies according to image type, segmentation, feature selection, dimensionality reduction, classifier choice, and validation strategy. Few studies have directly compared radiomics derived from grayscale ultrasound, CDFI, arterial-phase

CEUS, venous-phase CEUS, and post-vascular Sonazoid images while maintaining a consistent analytical workflow (Lambin et al. 2017). As a result, the influence of the underlying imaging substrate on radiomics performance remains less well understood than the influence of classifier selection itself.

Four principal gaps therefore remain in the current literature. First, few studies directly compare qualitative and quantitative conventional ultrasound, multiphase Sonazoid CEUS, and FNA-Tg within the same node-level cohort. Second, the relative diagnostic contributions of the arterial, venous, and post-vascular phases of Sonazoid CEUS have not been systematically established. Third, the effect of biological imaging substrate on radiomics performance has received less attention than model or classifier selection. Fourth, evidence remains limited regarding whether integrated multimodal assessment provides clinically meaningful incremental benefit beyond the best-performing single modality.

These gaps justify a unified investigation that compares qualitative and quantitative assessment, evaluates individual and combined diagnostic strategies, and tests modality-specific radiomics models under a consistent reference standard (Liu et al. 2019; Zhang et al. 2023). Such a design can clarify not only which approach demonstrates the highest diagnostic performance, but also why different modalities may contribute complementary information and whether these differences translate into statistically robust, well-calibrated, and clinically meaningful decision support.

The literature therefore supports a multimodal framework based on a common cohort and reference standard, with direct comparison of qualitative and quantitative approaches, individual and combined modalities, and radiomics models derived from multiple ultrasound substrates. Such a framework enables diagnostic discrimination, calibration, clinical utility, and the biological rationale underlying modality-specific differences to be evaluated under comparable conditions. More importantly, it shifts the focus from identifying a single “best” diagnostic test towards understanding how complementary structural, vascular, tissue-related, biochemical, and quantitative information can be integrated to reduce uncertainty in the preoperative assessment of lateral cervical lymph nodes in PTC.

The landscape tables consolidate and synthesize the principal evidence reviewed in this chapter. **Table 2.2** provides a comparative overview of modality-specific diagnostic contributions, methodological limitations, clinical or research relevance in the preoperative assessment of lateral cervical lymph nodes. **Table 2.3** summarizes selected key studies that collectively provide the evidence base for the multimodal diagnostic framework evaluated in this study.

Overall, the literature establishes a clear rationale for investigating conventional ultrasound, multiphase Sonazoid CEUS, FNA-Tg, and ultrasound-based radiomics within a unified analytical framework. The unresolved methodological and diagnostic questions identified in this chapter directly inform the design of the present study. The following chapter therefore describes the study population, reference standard, imaging and FNA-Tg procedures, radiomics workflow, model development, and statistical methods used to address these questions systematically.

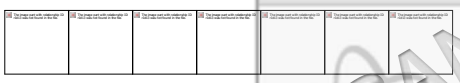


Table 2.1 Comparative technical characteristics of SonoVue and Sonazoid contrast-enhanced ultrasound

Attribute	SonoVue	Sonazoid
Composition	Sulfur hexafluoride (SF_6)	Perfluorobutane (C_4F_{10})
Microbubble structure	Lipid shell, gas core	Phospholipid shell, gas core
Imaging duration	2-4 minutes	Up to 30 minutes or longer
Macrophage uptake	Absent	Present (Post-vascular/Kupffer phase)
Applicable imaging phases	Vascular phase only	Vascular + post-vascular phases
Clinical use	Widely applied in liver and thyroid imaging	Primarily liver; expanding indications

Table 2.2 Comparative characteristics of diagnostic approaches for preoperative LLNM assessment

Approach	Main diagnostic information	Key strengths	Main limitations	Clinical / Research relevance
Conventional US	Nodal morphology and architecture	Accessible, real-time, high spatial resolution	Operator-dependent; benign–malignant overlap	First-line mapping and target selection
CDFI	Macroscopic vascular pattern	Adds hilar/peripheral flow information	Low-flow sensitivity; setting dependent	Supportive vascular assessment
FNA-C	Cellular morphology	Direct cytological confirmation	Sampling error; low yield in cystic nodes	Pathological confirmation of selected nodes
FNA-Tg	Thyroid-specific biochemical signal	Sensitive in cystic or paucicellular metastases	Invasive; threshold and assay variability	Adjunct to cytology and imaging
SonoVue CEUS	Microvascular perfusion	Real-time enhancement assessment	Short blood-pool window; inflammatory overlap	Problem-solving for indeterminate nodes
Sonazoid vascular phases	Arterial and venous perfusion	Phase-specific microvascular information	Timing and interpretation remain operator-dependent	Comparison of arterial and venous diagnostic value
Sonazoid post-vascular phase	Tissue-related contrast retention	Prolonged window; potential tissue information	Limited nodal evidence; possible inflammatory confounding	Evaluate incremental value and radiomics potential
Radiomics	Quantitative image heterogeneity	Objective features beyond visual assessment	Overfitting; acquisition and segmentation variability	Risk prediction, validation, and calibration
Multimodal integration	Combined structural, vascular, biochemical, and quantitative information	May reduce diagnostic uncertainty	Greater complexity; requires validation	Integrated preoperative decision support

Table 2.3 Selected key studies informing the multimodal diagnostic framework

Study	Focus	Design	Key contribution	Limitation	Relevance
Haugen et al. (2016)	Guideline / PTC management	ATA guideline	Therapeutic lateral neck dissection is recommended when nodal metastasis is confirmed; supports the need for accurate preoperative assessment.	Guideline evidence depends on heterogeneous source studies.	Defines the clinical decision context and consequences of diagnostic error.
Lim et al. (2011)	Clinicopathological predictors of LLNM	Retrospective surgical cohort	Identified factors associated with lateral cervical nodal metastasis in PTC.	Retrospective and centre-specific.	Supports risk stratification but also shows that clinical variables cannot replace direct nodal assessment.
Mulla et al. (2012)	Conventional ultrasound	Systematic review / diagnostic evidence	Conventional US is first-line, but sensitivity varies and depends on feature definition and operator experience.	Heterogeneous criteria and reference standards.	Justifies quantitative scoring and adjunctive modalities.
Jeon et al. (2012)	FNA-Tg	Diagnostic cohort	FNA-Tg improves diagnosis of metastatic cervical nodes, especially when cytology is inadequate.	Assay and cut-off variability.	Provides biochemical rationale for inclusion of FNA-Tg.
Al-Hilli et al. (2016)	FNA-Tg	Retrospective diagnostic study	Needle-washout thyroglobulin improves detection of PTC nodal metastasis.	Single-center and assay-specific.	Supports FNA-Tg as a confirmatory biochemical test.
Duval et al. (2017)	FNA-Tg analytical factors	Diagnostic laboratory study	Serum TSH and anti-thyroglobulin antibodies may influence interpretation of washout Tg.	Platform-specific and heterogeneous sampling.	Highlights need for contextual and quantitative interpretation.

…continuation

Liu et al. (2020)	FNA-Tg	Meta-analysis	Reported high pooled sensitivity and specificity for cervical nodal metastasis.	Threshold and protocol heterogeneity.	Supports high diagnostic value but identifies a standardization gap.
Tong et al. (2020)	Ultrasound radiomics for LLNM	Radiomics nomogram study	An ultrasound-based radiomic nomogram showed potential for predicting lateral cervical nodal metastasis.	Retrospective; limited external validation.	Directly relevant to node-specific lateral-neck radiomics.

to be continued…

Yu et al. (2020)	Transfer-learning radiomics	Multicenter imaging study	Transfer-learning radiomics improved prediction of lymph node metastasis in PTC.	Model interpretability and protocol variation.	Demonstrates the value of quantitative imaging and the need for robust validation.
Cao et al. (2021)	Radiomics review	Narrative review	Summarized applications and limitations of radiomics in differentiated thyroid cancer.	Broad scope; not lateral-node specific.	Frames reproducibility, standardization, and clinical translation issues.
Jiang et al. (2023)	CEUS clinical-radiomics nomogram	Retrospective modelling study	A CEUS-based clinical-radiomics model improved preoperative prediction of cervical nodal metastasis.	Potential overfitting and lack of broad external validation.	Supports combining CEUS-derived quantitative information with clinical factors.
Xiao et al. (2023)	Perfluorobutane CEUS for small LLNM	Diagnostic study of small lateral nodes	Perfluorobutane CEUS improved diagnostic confidence for small lateral cervical nodal metastases.	Protocol and contrast-agent experience may limit generalizability.	Strongest direct evidence for Sonazoid/perfluorobutane in small lateral nodes.
Liu et al. (2023)	FNA-Tg	Large retrospective diagnostic study	Confirmed high diagnostic efficacy of FNA-Tg and evaluated influencing factors.	Retrospective and laboratory-specific.	Supports quantitative FNA-Tg and analysis of impact factors.
Liu et al. (2024)	Clinical multimodal ultrasound radiomics	Retrospective multimodal model	Clinical and multimodal ultrasound radiomics improved prediction over single inputs.	Potential center and acquisition dependence.	Supports multimodal integration and comparison of imaging substrates.
Zhang et al. (2024)	Ultrasound radiomics	Systematic review and meta-analysis	Ultrasound radiomics showed promising pooled performance for nodal metastasis.	Substantial heterogeneity and publication bias risk.	Supports promise but underscores need for rigorous validation.
Liu et al. (2024)	FNA-Tg in PTC	Clinical diagnostic study	Evaluated FNA-Tg for cervical nodal metastasis and its clinical application.	Cut-off and population dependence.	Reinforces FNA-Tg as complementary rather than standalone.

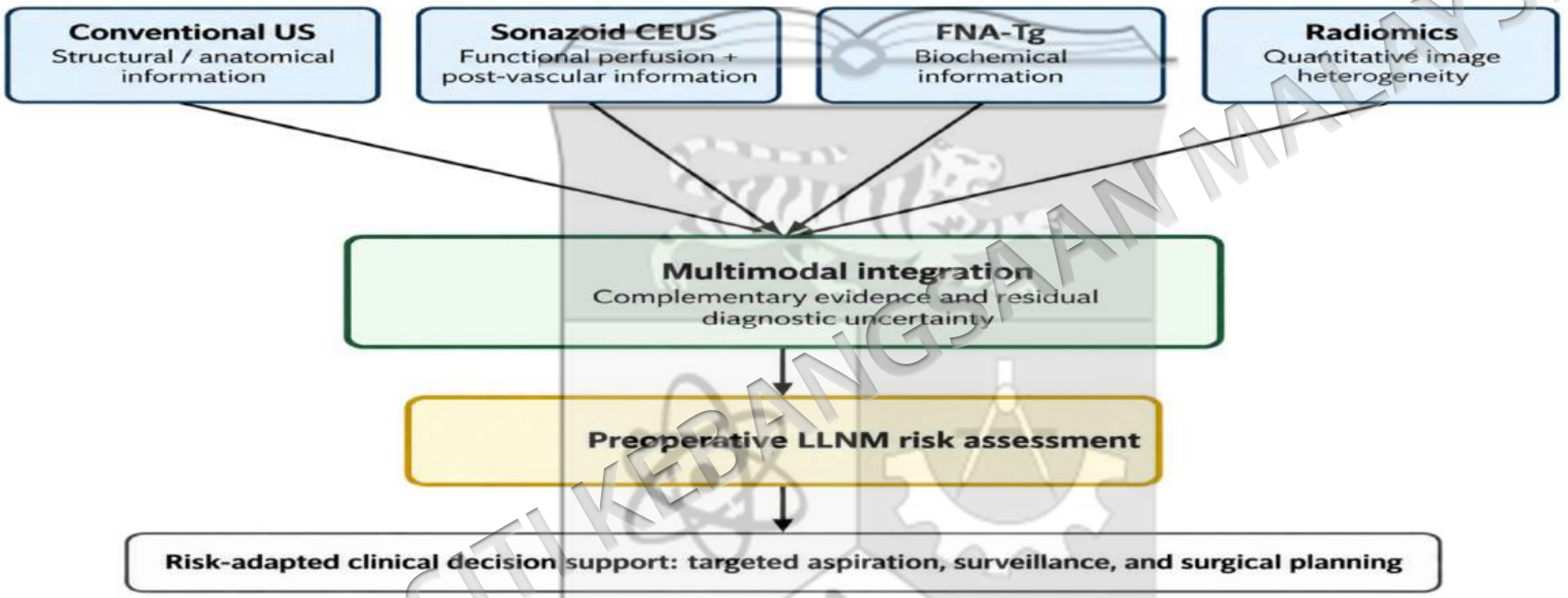


Figure 2.1 Conceptual framework linking structural, functional, biochemical, and quantitative information for preoperative LLNM assessment

CHAPTER III

METHODOLOGY

3.1 DESIGN AND LOCATION OF THE STUDY

This cross-sectional diagnostic study was conducted in the Department of Ultrasound, Affiliated Hospital of Pan Zhihua University, Sichuan, China. The clinical source period extended from 1 February 2024 to November 2025. Between 1 February and 13 June 2024, 44 patients who later fulfilled the final eligibility criteria underwent the relevant examinations as part of routine clinical care; no study-specific recruitment took place before local ethical approval. After approval on 14 June 2024, their existing records and stored images were screened against the prespecified eligibility, completeness, image-quality, and outcome-verification criteria and were included only when all requirements were met. From July 2024 to November 2025, a further 134 eligible patients were enrolled under the approved protocol. The final analytical cohort therefore comprised 178 patients (44 retrospectively included routine-care cases and 134 patients enrolled during the formal study period). One index lateral cervical lymph node per patient was analysed.

3.1.1 Preliminary Methodological Feasibility Assessment

The preliminary methodological assessment was performed after local ethical approval had been obtained. The 44 pre-existing routine-care cases were retrospectively reviewed to confirm the consistency of image selection, phase-specific recording, FNA-C/FNA-Tg documentation, target-node matching, ROI delineation, observer procedures, data-collection forms, and radiomics preprocessing. No research database entry or research analysis of these cases was undertaken before approval. These checks did not change the study objectives, eligibility criteria, reference standard, or primary outcome, and the same procedures were subsequently applied throughout the formal study period.

3.1.2 Sample Size Justification

An a priori sample-size calculation was performed using G*Power version 3.1.9.7. A two-tailed point-biserial correlation model under the t-test family was specified, with an assumed effect size of $|\rho| = 0.30$, a two-sided significance level (α) of 0.05, and statistical power ($1-\beta$) of 0.95. The calculation yielded a minimum required sample size of 134 participants (df = 132; actual power = 0.951). This target was used for recruitment during the formal study period from July 2024 to November 2025. Separately, 44 patients who had undergone the relevant examinations as part of routine clinical care between 1 February and 13 June 2024 were screened for research eligibility only after local ethical approval was obtained on 14 June 2024. They met the same eligibility, imaging-quality, and outcome-verification criteria as patients enrolled during the formal study period. The final analytical cohort therefore comprised 178 patients: 44 retrospectively included pre-existing clinical cases and 134 patients enrolled during the planned study period.

3.2 PATIENTS

The study protocol was approved by the Research Ethics Committee of the Affiliated Hospital of Pan Zhihua University, China, on 14 June 2024 (Appendix A), and by the Medical Research and Ethics Committee, Faculty of Medicine, Universiti Kebangsaan Malaysia, on 12 December 2024 (Appendix B). Between 1 February and 13 June 2024, 44 patients who later met the final eligibility criteria underwent the relevant examinations as part of routine clinical care. No study-specific recruitment, research database construction, radiomics processing, or statistical analysis of these cases was undertaken before local ethical approval. After 14 June 2024, their existing clinical records and stored images were screened for research eligibility and included only when the prespecified criteria were satisfied. From July 2024 to November 2025, 134 patients were enrolled under the approved study protocol. Participants undergoing study-specific procedures after approval received verbal and written information on the study purpose, procedures, potential risks, confidentiality, voluntary participation, and the right to withdraw without affecting clinical care; written informed consent was obtained where required (Appendix C). All data were de-identified before analysis and were accessible only to authorised members of the research team.

3.2.1 Inclusion criteria:

a. Age 18 years or older and a diagnosis of primary PTC.

b. Availability of preoperative conventional ultrasound and Sonazoid CEUS images, including arterial, venous, and post-vascular phases, of the target lateral cervical lymph node.

c. Availability of preoperative FNA-C and FNA-Tg results from the same target lymph node.

d. Availability of serum thyroglobulin measured within 14 days of FNA-Tg sampling and sufficient clinical information to establish the final reference diagnosis.

3.2.2 Exclusion criteria:

a. Known allergy to eggs, egg products, or any component of Sonazoid.

b. Pregnancy, age below 18 years, or inability to provide informed consent where consent was required.

c. Lateral cervical lymphadenopathy caused by another malignant tumour, lymphoma, infection, or other non-PTC disease confirmed by pathology or clinical assessment.

d. Severe cardiopulmonary, hepatic, or renal dysfunction; clinically significant coagulopathy; or any condition precluding CEUS or needle aspiration.

e. Incomplete imaging, laboratory, cytological, surgical, pathological, or follow-up information; images of insufficient quality for interpretation or radiomics analysis; or withdrawal from the study.

3.3 ETHICS APPROVAL & INFORMED CONSENT

The study protocol was approved by the Research Ethics Committee of the Affiliated Hospital of Pan Zhihua University, China, on 14 June 2024 (Appendix A), and by the Medical Research and Ethics Committee, Faculty of Medicine, Universiti Kebangsaan Malaysia, on 12 December 2024 (Appendix B). Between 1 February and 13 June 2024, 44 patients who later fulfilled the final eligibility criteria underwent the relevant examinations as part of routine clinical care and were not prospectively recruited for research. No study-specific recruitment, research database construction, radiomics processing, or statistical analysis of these cases was undertaken before local ethical approval. After approval on 14 June 2024, their pre-existing clinical records and stored images were formally screened for research eligibility and included only when the prespecified criteria were satisfied. From July 2024 to November 2025, a further 134 patients were enrolled under the approved study protocol. Participants undergoing study-specific procedures after approval received verbal and written information regarding the study purpose, procedures, potential risks, confidentiality safeguards, voluntary participation, and the right to withdraw without affecting clinical care; written informed consent was obtained where required (Appendix C). All data were de-identified before analysis and were accessible only to authorised members of the research team.

3.4 LATERAL CERVICAL LYMPH NODES CLINICAL EVALUATION FLOW

Before surgery, all patients with PTC underwent systematic ultrasound examination of the central and lateral cervical compartments. When a suspicious lateral cervical lymph node was identified, the same target node underwent Sonazoid CEUS and ultrasound-guided FNA for FNA-C and FNA-Tg. The target lymph node was documented by cervical level, side, size, and image landmarks to ensure correspondence across imaging, aspiration, surgery, pathology, and follow-up.

Patients with one or more strongly positive findings—malignant FNA-C, malignant Sonazoid CEUS features, or a positive FNA-Tg result according to the prespecified criteria—were considered clinically suspicious and were evaluated for

lateral cervical lymph node dissection during thyroid surgery. Immediately before surgery, the target node was localized by skin marking and ultrasound-guided injection of diluted methylene blue. The labelled node was removed separately and submitted for histopathological examination to permit node-to-node correlation.

The primary reference standard was postoperative histopathology of the corresponding target lymph node. For patients who did not undergo lateral neck dissection because all three preoperative tests were negative, scheduled ultrasound surveillance at 1, 3, and 6 months formed a secondary clinical reference standard. Follow-up was used only to classify the outcome of non-resected nodes and was not treated as a preoperative predictor. A node was classified as non-metastatic only when it remained stable without new suspicious features throughout at least 6 months of surveillance; any interval suspicious change prompted repeat CEUS, FNA-C, and FNA-Tg, when indicated, surgery or biopsy. This approach is illustrated in **Figure 3.1**.

3.5 STUDY PROCEDURE

For Observer Roles and Blinding. Conventional ultrasound and Sonazoid CEUS examinations were interpreted independently of the final reference diagnosis. Image reviewers were not provided with postoperative histopathology or follow-up outcomes at the time of qualitative scoring. The physician performing FNA was aware of the ultrasound appearance for procedural safety but did not have access to the final histopathological diagnosis. Cytopathologists and laboratory personnel interpreted FNA-C and measured FNA-Tg without access to radiomics results. For radiomics, the primary segmenter was blinded to the outcome during ROI delineation, and the second observer independently segmented the reproducibility subset without viewing the first observer's contours. Dataset labels were revealed only after segmentation and feature extraction had been completed. Where complete blinding was not feasible because clinical procedures occurred sequentially in routine care, this was acknowledged as a potential source of review bias and was mitigated by using prespecified criteria and the independent reference standard.

3.5.1 Conventional Ultrasound Assessment

a. Image Acquisition

i. Conventional ultrasound and CEUS examinations were performed using a Sequoia system (Siemens Healthineers), LOGIQ E11 system (GE Healthcare), or EPIQ Elite system (Philips Healthcare) with a 6–15 MHz linear-array transducer. Equipment settings were optimized for superficial cervical imaging and kept as consistent as clinically feasible within each examination.

ii. Patients were examined in the supine position with the neck mildly extended and rotated away from the side being assessed.

iii. Lateral cervical lymph nodes were systematically assessed at levels II–V according to standard anatomical landmarks.

iv. The index suspicious lymph node was examined in longitudinal and transverse planes, and representative still images and cine loops were stored in Digital Imaging and Communications in Medicine format.

v. Recorded features included short-axis diameter, long-to-short-axis ratio, shape, margin, cortical morphology, internal echogenicity, fatty hilum, calcification, cystic degeneration, vascular distribution, and relationship to adjacent structures.

b. Suspicious Conventional Ultrasound Features

Lateral cervical lymph nodes were considered suspicious if one or more of the following features were present (Luo et al. 2022):

i. Feature ① (Size/Shape): Short diameter >0.5 cm, long diameter/short diameter (L/S) ratio <1.5, and a rounded shape.

ii. Feature ② (Margin/Cortex): Ill-defined or irregular margin, centripetal thickening of the cortex, or decreased echogenicity levels in the cortex and medulla.

iii. Feature ③ (Lymphatic Hilum): Absence or eccentricity of the lymphatic hilum.

iv. Feature ④ (Vascularization): Peripheral vascularization or loss of regular blood flow signals.

v. Feature ⑤ (echogenicity): Presence of diffuse or focal hyperechogenicity.

vi. Feature ⑥ (Cystic Appearance): Presence of cystic components within the node.

vii. Feature ⑦ (Calcifications): Presence of any type of calcification.

c. Qualitative Conventional Ultrasound Diagnosis

Qualitative metastasis was diagnosed if the lymph node exhibited any of the following pathognomonic features: hyper-echogenicity (⑤), cystic appearance(⑥), or calcifications (⑦).

d. Conventional Ultrasound Quantitative Scoring System

The quantitative score in **Table 3.1** was an investigator-defined ordinal scoring system developed for the present study and was not a previously validated clinical scoring system. The component variables were selected from recognised suspicious sonographic features of metastatic cervical lymph nodes in thyroid cancer, including increased short-axis diameter or rounded configuration, irregular margin/cortical abnormality, loss or eccentricity of the fatty hilum, abnormal peripheral or disorganised vascularity, hyperechogenicity, cystic change, and calcification (Ahuja & Ying 2005; Haugen et al. 2016; Luo et al. 2022; Chung et al. 2022). The numerical weights were not copied from a published validated coefficient set and were not derived from a fitted

regression model. Instead, they were assigned ordinally to represent increasing diagnostic suspicion: common but relatively nonspecific morphologic abnormalities were assigned 1 point; loss of the hilum and abnormal vascularity were assigned 2 points; and hyperechogenicity, cystic change, and calcification received higher scores because these findings are more characteristic of metastatic PTC nodes. Within the latter categories, increasing extent or greater metastatic specificity was represented by progressively higher study-defined scores (3-5points). Thus, the literature supported selection and relative diagnostic importance of the suspicious features, whereas the exact numerical weighting remained investigator-defined. The total score was obtained by summing all features present in the index node and was evaluated as an index test in the present cohort; its optimal cut-off and diagnostic performance were determined empirically by ROC analysis rather than assumed a priori.

Table 3.1 Quantitative scoring criteria for conventional ultrasound features

Feature Category	Specific Ultrasound Sign	Score
Morphology	Short diameter > 0.5 cm	1
	L/S ratio < 1.5	1
Structure	Irregular or ill-defined margin	1
	Cortical thickening / decreased echogenicity/hilum eccentricity	1
	Absence of lymphatic hilum	2
Vascularity	Peripheral vascularization / irregular flow / abundant flow	2
Echogenicity	Focal hyper-echogenicity	4
	Diffuse hyper-echogenicity	5
Cystic Changes	Cystic appearance < 0.3 cm	3
	Cystic appearance 0.3–0.5 cm	4
	Cystic appearance > 0.5 cm	5
Calcifications	Calcification with comet tail / linear calcification	3
	Coarse calcification (> 0.1 cm)	4
	Fine punctate calcification	5

3.5.2 Sonazoid CEUS Assessment

a. Sonazoid Administration and Imaging Protocol

After written consent for CEUS had been obtained, Sonazoid (perfluorobutane microbubbles; Daiichi Sankyo, Tokyo, Japan) was prepared according to the

manufacturer's instructions. A 1.0 mL bolus diluted with 2 mL saline was administered through a 20-gauge antecubital intravenous cannula and followed by a 5 mL saline flush. The largest reproducible section of the target lymph node was selected. Low-mechanical-index contrast imaging (approximately 0.2) was used, with the focal zone positioned immediately below the node. Patients were instructed to breathe quietly and avoid swallowing. Continuous vascular-phase cine loops were stored for 60 seconds after injection, and additional images were acquired through the venous phase. Post-vascular-phase imaging was performed at 10 minutes or later and stored for at least 10 seconds. Gain, depth, focus, and dynamic range were maintained as consistently as possible within each examination.

b. Sonazoid CEUS Phase Definitions

CEUS evaluation was divided into three phases:

i. Arterial phase

ii. (10–30 s): assessment of initial contrast inflow.

iii. Venous phase (30–120 s): evaluation of washout and perfusion homogeneity.

iv. Post-vascular phase (≥10 min): assessment of contrast retention related to macrophage uptake.

c. Qualitative Sonazoid CEUS Assessment

Two ultrasound physicians with experience in thyroid and cervical lymph-node imaging independently reviewed the stored conventional ultrasound and Sonazoid CEUS images using prespecified criteria. Reviewers were blinded to histopathological and follow-up outcomes. Disagreements were resolved by consensus after the independent assessments had been recorded. Inter-observer agreement for categorical features was quantified using Cohen's kappa. Suspicious Sonazoid CEUS features included centripetal or chaotic arterial enhancement, heterogeneous enhancement or perfusion

defects during the vascular phase, and sparse, defective, hypo-enhancing, ring-like, or absent enhancement in the post-vascular phase (Xiao et al. 2023; Yu et al. 2023).

i. Sonazoid CEUS in the arterial phase

Enhancement modes manifested as chaotic, centripetal, circular enhancement with no enhancement inside indicate metastasis. Centrifugal and overall (integral) enhancement modes indicate benign.

ii. Sonazoid CEUS in the venous phase

Enhancement modes manifest as heterogeneous, and no enhancement mode indicates metastasis. Homogeneous enhancement mode indicates benign.

iii. Sonazoid CEUS in the post-vascular phase

Enhancement modes manifest as hyper-enhancement with sparse microbubbles, hyper-enhancement with defect> 0.1 cm, circular enhancement with internal non-enhancement, homogeneous hypo-enhancement, heterogeneous hypo-enhancement, and non-enhancement in the post-vascular phase indicate metastasis. Homogeneous hyper-enhancement indicates benign.

d. Quantitative Sonazoid CEUS Scoring

An investigator-defined ordinal score was assigned to each enhancement pattern to represent increasing suspicion of metastasis; the higher the score, the greater the suspicion. The arterial-, venous-, and post-vascular-phase scores were analyzed separately and in combination. The scoring system was not treated as an externally validated instrument; observed in our clinical practice, its cut-off values and diagnostic performance were evaluated in this cohort, and observer agreement for the underlying categorical patterns was assessed before model comparison.

i. Sonazoid CEUS in the arterial phase

Centrifugal (+1), overall (integral +2), chaotic (+3), no enhancement (+4), circular enhancement (+4), centripetal (+5).

ii. Sonazoid CEUS in the venous phase

Homogeneous (+1), heterogeneous with single defect 0.1- 0.2 cm (+2), heterogeneous with single defect 0.2- 0.3 cm (+3), no enhancement (+4), extremely heterogeneous with defect>0.3cm (+5), disordered with more than 3 defects and every defect > 0.1cm (+5).

iii. Sonazoid CEUS in the post-vascular phase

Homogeneous hyper-enhancement (+1), heterogeneous hyper-enhancement with defect 0.1-0.3cm(+2), homogeneous hypo-enhancement(+3), heterogeneous hypo-enhancement with defect 0.1-0.3cm (+4), circular enhancement no enhancement inside (+4), no enhancement (+4), marked filling with single defect > 0.3cm(+5), marked filling with multiple defect > 0.1cm (+5)， marked filling only sparse microbubbles(+5).

3.5.3 FNA-C and FNA-Tg Examination

All index lymph nodes underwent ultrasound-guided FNA for cytology and washout thyroglobulin irrespective of the qualitative Sonazoid CEUS result, thereby reducing partial verification bias between the index tests.

a. FNA-C Acquisition

i. Participants were informed about the puncture procedure and provided written consent. Anticoagulant or antiplatelet medication was managed according to institutional clinical protocols and the patient's thrombotic risk rather than discontinued automatically.

ii. Relevant laboratory tests, including complete blood count and coagulation indices, were reviewed before puncture when clinically indicated.

iii. Ultrasound-guided FNA was performed by a senior ultrasound physician experienced in cervical intervention using an in-plane needle approach.

The patient was positioned to provide a short and safe route to the target node. The puncture path was planned to avoid major vessels, nerves, and other critical structures.

iv. After skin antisepsis and local anaesthesia with 2% lidocaine, a 22-gauge needle attached to a 5 mL syringe was advanced into the solid component of the node. Multiple short back-and-forth passes were performed in different directions under real-time ultrasound guidance. At least three passes were obtained when technically feasible. Aspirated material was expelled onto glass slides, smeared, immediately fixed in 95% ethanol, labelled by nodal level and side, and sent for haematoxylin and eosin staining and cytological examination.

v. All target locations and corresponding specimens were recorded to maintain imaging–cytology–pathology correspondence.

vi. FNA-C was classified as positive when malignant epithelial cells compatible with metastatic PTC were identified. A specimen containing lymphocytes without atypical epithelial cells was classified as negative; non-diagnostic samples were repeated when clinically feasible and were excluded if no definitive classification could be obtained.

b. FNA-Tg Acquisition

After the cytology pass, a separate needle pass was performed in the same target lymph node. The needle and syringe were rinsed with 1.0 mL normal saline in an Eppendorf tube. For predominantly cystic nodes, one drop of the aspirated cyst fluid was added to the saline washout. The sample was centrifuged for 5 minutes, and the supernatant was analyzed for thyroglobulin by electrochemiluminescence immunoassay using a COBAS e601 analyzer (Roche Diagnostics). Laboratory personnel were blinded to the imaging and final reference results.

c. FNA-Tg Qualitative Assessment

The reference standard of FNA-Tg (Jiang and Hsiao 2020; Liu et al. 2022; Su et al. 2024):

i. Positive: FNA-Tg >10 ng/mL, or FNA-Tg of 1–10 ng/mL with an FNA-Tg/serum Tg ratio >1.

ii. Negative: FNA-Tg <1 ng/mL, or FNA-Tg of 1–10 ng/mL with an FNA-Tg/serum Tg ratio ≤1, in the absence of another positive criterion.

d. FNA-Tg Quantitative Assessment

FNA-Tg was additionally analyzed as a continuous variable. ROC analysis was used to estimate the AUC and 95% confidence interval and to determine the cohort-specific optimal cut-off by the maximum Youden Index. Sensitivity, specificity, PPV, NPV, and accuracy were calculated at that cut-off.

3.5.4 Lateral Cervical Lymph Node Surgery

A target lateral cervical lymph node was considered highly suspicious when at least one of the following criteria was present:

a. FNA-C result (+).

b. Sonazoid CEUS vascular phase or post-vascular phase showing a malignant image.

c. FNA-Tg > 10 ng/mL or FNA-Tg 1–10 ng/ml, FNA-Tg/serum Tg > 1.

Patients meeting the surgical criteria underwent lateral cervical lymph node dissection according to clinical indications. Before surgery, the target node was

documented by level and size, marked on the skin, and injected under ultrasound guidance with 0.1 mL of 1:10 diluted methylene blue within approximately 15 minutes before the operation. The labelled node was removed first when feasible, submitted separately, and matched with the preoperative imaging and aspiration records. Paraffin-embedded histopathology of the corresponding node served as the primary reference standard.

3.5.5 Lateral Cervical Lymph Node Follow-up

Patients without lateral neck dissection entered follow-up only when all of the following low-suspicion criteria were met:

a. FNA-C result (-)

b. No signs of malignancy in vascular and post-vascular phases of sonazoid CEUS.

c. FNA-Tg<1ng/ml or FNA-Tg is 1~10 ng/ml, FNA- Tg/serum Tg<1.

These patients underwent routine cervical ultrasonography at approximately 1, 3, and 6 months after surgery. Follow-up was used solely to establish the outcome of non-resected nodes and did not contribute predictor information to the preoperative models. New growth or suspicious morphologic change prompted repeat conventional ultrasound, Sonazoid CEUS, FNA-C, and FNA-Tg. Nodes that remained stable and showed no suspicious features throughout at least 6 months were classified as non-metastatic for the final analysis. This secondary reference standard and its potential for misclassification were acknowledged as a study limitation.

3.5.6 Radiomics Analysis of Conventional Ultrasound and Sonazoid CEUS

The radiomics workflow comprised image selection, preprocessing, ROI segmentation, reproducibility assessment, feature extraction, feature reduction, model development, internal validation, and clinical utility assessment (**Figure 3.2**).

a. Image Selection and Preprocessing

For each patient, one representative image of the index lymph node was selected from each available modality: grayscale ultrasound, color Doppler flow imaging (CDFI), arterial-phase CEUS, venous-phase CEUS, and post-vascular-phase CEUS. Images were exported in DICOM or lossless format and anonymized. To reduce non-biological variation, images were checked for consistent orientation and complete lesion display. Intensities were normalized within each modality before feature extraction, grayscale values were discretized using a fixed bin-width strategy, and images were resampled to a common in-plane pixel spacing when required by the extraction pipeline. No smoothing or filtering beyond the predefined PyRadiomics transforms was applied unless explicitly specified in the feature configuration. The same preprocessing parameters were applied to all patients within a modality and were fitted without using outcome information.

b. ROI Segmentation and Reproducibility Assessment

The index lymph node was manually delineated in ITK-SNAP on each selected image. The primary observer, an ultrasound physician with more than 3 years of thyroid-imaging experience, segmented all 178 cases while blinded to the outcome. A second ultrasound physician independently re-segmented a simple random sample of 40 cases selected from the full cohort. Random selection was stratified by outcome to preserve representation of metastatic and non-metastatic nodes. The 40-case subset was used only for segmentation reproducibility analysis and was not a separate model-development dataset. Intraclass correlation coefficients (ICCs) were calculated from features extracted from the two independent segmentations. After an interval of at least 2 weeks, the primary observer repeated segmentation of the same subset to assess intra-observer reproducibility. Features with ICC ≥ 0.80 for both intra- and inter-observer comparisons were retained for subsequent analysis. The observers did not view each other's contours, imaging interpretations, or final histopathological outcomes during segmentation.

c. Radiomics Feature Extraction

Radiomics features were extracted using PyRadiomics in Python (Tong et al. 2022; Lu et al. 2023; Mu et al. 2023; Liu et al. 2024) on the OneKey AI platform. Feature families included first-order features, shape features, gray-level co-occurrence matrix (GLCM), gray-level run length matrix (GLRLM), gray-level size zone matrix (GLSZM), gray-level dependence matrix (GLDM), and neighbouring gray tone difference matrix (NGTDM) features(**Figure 3.3**). Three-dimensional shape features were not used for single two-dimensional images. The extraction settings were fixed before outcome modelling and applied identically to the training and test cohorts.

d. Feature Selection

All data-dependent preprocessing and feature selection were performed after the 8:2 training–test split and used the training cohort only. Within each training fold, features with poor reproducibility were removed, near-zero-variance features were excluded, missing values were imputed using training-fold estimates, and continuous features were standardized using parameters derived from the training fold. Highly correlated features were reduced before least absolute shrinkage and selection operator (LASSO) regression. The LASSO penalty was selected by 10-fold cross-validation within the training cohort. The resulting transformations and selected feature set were then applied unchanged to the held-out test cohort. This nested sequence prevented information leakage from the test cohort into feature selection.

e. Classifier Selection and Model Construction

Eleven supervised machine-learning classifiers were evaluated: logistic regression, naive Bayes, support vector machine, k-nearest neighbour, random forest, ExtraTrees, extreme gradient boosting, LightGBM, gradient boosting, AdaBoost, and multilayer perceptron. These methods were selected because the radiomics analysis was based on handcrafted quantitative features organised as tabular data. The selected classifiers included linear, probabilistic, distance-based, kernel-based, tree-based, ensemble, boosting, and shallow neural-network methods, allowing different modelling

assumptions and levels of non-linearity to be compared within the same analytical framework.

CNNs and Vision Transformers were not included because the study was designed as a handcrafted-radiomics analysis rather than an end-to-end image-learning study. With a single-centre cohort of 178 patients and no independent external validation cohort, direct training of high-capacity image-based deep-learning models would have increased the risk of overfitting and produced results that would be difficult to evaluate reliably. Conventional machine-learning classifiers were therefore considered more appropriate for the available data structure and sample size, while preserving direct assessment of the contribution of the selected radiomics features.

Each classifier was developed separately for grayscale ultrasound, CDFI, arterial-phase CEUS, venous-phase CEUS, and post-vascular-phase CEUS using the same preprocessing, feature-selection, training, and validation framework.

f. Training, Hyperparameter Tuning, and Internal Validation

The full cohort of 178 patients—not the 40-case reproducibility subset—was used for model development. Patients were randomly divided into training and test cohorts at an 8:2 ratio using stratified random sampling to preserve the proportion of metastatic and non-metastatic cases. The random seed was fixed before analysis to allow reproducibility. Hyperparameters were tuned exclusively within the training cohort by grid search with stratified cross-validation. Class weights or resampling were considered only when imbalance materially affected a classifier; any such procedure was applied within training folds only. The final model for each classifier was refitted on the complete training cohort using the selected hyperparameters and evaluated once on the untouched test cohort. Overfitting was limited through reproducibility filtering, correlation reduction, penalized selection, constrained hyperparameter search, cross-validation, and strict separation of training and test data.

g. Model Performance, Internal Validation, and Clinical Utility

Discrimination was evaluated using receiver operating characteristic (ROC) analysis,

and the area under the ROC curve (AUC) with its 95% confidence interval was reported for each radiomics classifier and imaging modality. At the prespecified or training-derived classification threshold, sensitivity, specificity, accuracy, positive predictive value (PPV), negative predictive value (NPV), precision, recall, and F1-score were calculated. For comparisons between radiomics models evaluated in the same patients, correlated ROC curves were compared using the non-parametric DeLong test. These comparisons were performed in the held-out test cohort for prespecified clinically relevant pairs, including the best-performing models derived from grayscale ultrasound, CDFI, arterial-phase CEUS, venous-phase CEUS, and post-vascular-phase CEUS. The null hypothesis was that the two correlated AUCs were equal; a two-sided P value <0.05 indicated a statistically significant difference. Where multiple exploratory pairwise comparisons were presented, the results were interpreted cautiously because of the limited test-cohort size and multiplicity of testing.

Model calibration was assessed by plotting calibration curves that compared predicted probabilities of lateral cervical lymph node metastasis with the corresponding observed event proportions. The 45-degree reference line represented perfect agreement between predicted and observed risks. Curves above the reference line indicated underestimation of metastatic risk, whereas curves below the line indicated overestimation. Calibration was examined primarily in the held-out test cohort and was additionally displayed for the training cohort where relevant. The Brier score, calculated as the mean squared difference between predicted probabilities and observed outcomes, was used as an overall measure of probabilistic accuracy, with lower values indicating better calibration and overall prediction. Calibration intercept and calibration slope were estimated where sample size permitted, ideal calibration corresponded to an intercept of 0 and a slope of 1. Because the test cohort was relatively small, calibration findings were interpreted as exploratory internal-validation evidence rather than definitive proof of transportability.

Decision curve analysis was performed in the held-out test cohort across clinically plausible threshold probabilities, with the net benefit of each model compared with the default treat-all and treat-none strategies. The displayed threshold range was selected to represent probabilities at which clinicians might consider additional biopsy

or lateral neck surgery.

Sample-level prediction distributions were plotted to demonstrate separation and overlap between metastatic and non-metastatic cases. These plots were interpreted as visualizations of discrimination and probability distribution rather than as direct evidence of model stability. Model stability was assessed through cross-validation variability in the training cohort and performance consistency in the held-out test cohort. Sample-level prediction distributions were plotted to show separation and overlap between metastatic and non-metastatic cases. These plots were interpreted as visualizations of discrimination and probability distribution rather than as evidence of model stability. Model stability was evaluated through cross-validation variability in the training cohort and performance consistency in the held-out test cohort.

All radiomics analyses were performed in Python using PyRadiomics on the OneKey AI platform.

h. External Validation

External validation was not performed because the study was conducted at a single center and no independent cohort with the same Sonazoid multiphase protocol, FNA-Tg workflow, image storage format, and node-level reference standard was available during the study period. The held-out test cohort therefore provided internal validation only and did not establish transportability to other institutions, scanners, operators, or patient populations. To reduce optimism, the test cohort remained untouched during preprocessing, feature selection, and hyperparameter tuning. The absence of external validation was treated as an important limitation, and multicenter temporal and geographic validation was recommended before clinical deployment.

3.6 STATISTICAL ANALYSIS

Statistical analyses were performed using SPSS Statistics version 25.0 (IBM Corp., Armonk, NY, USA), MedCalc version 11.4 (MedCalc Software, Ostend, Belgium), and Python (on the OneKey AI platform) for radiomics modelling and validation.

All tests were two-sided, and P<0.05 was considered statistically significant. Continuous variables were assessed for distributional form using graphical methods and an appropriate normality test. Normally distributed variables were summarized as mean ± standard deviation and compared with the independent-samples t-test; non-normally distributed variables were reported as median and interquartile range. Categorical variables were reported as counts and percentages and compared using the chi-square test or Fisher's exact test, as appropriate.

Inter-observer agreement for categorical ultrasound and CEUS features was quantified using Cohen's kappa; kappa values were interpreted as ≤0.20 slight, 0.21–0.40 fair, 0.41–0.60 moderate, 0.61–0.80 substantial, and 0.81–1.00 almost perfect agreement.

Paired binary diagnostic outcomes obtained from the same patients were compared using the McNemar test. McNemar testing was not used to compare sensitivity, specificity, PPV, NPV, or accuracy as isolated unpaired quantities; instead, it was applied to paired discordant classifications at a common reference standard.

Diagnostic performance was evaluated by ROC analysis. AUCs with 95% confidence intervals were reported and interpreted descriptively as 0.50 no discrimination, 0.60–0.69 poor, 0.70–0.79 acceptable, 0.80–0.89 good, and ≥0.90 excellent discrimination. The optimal diagnostic cut-off for continuous or ordinal index tests was selected by maximizing the Youden Index (sensitivity + specificity − 1), which identifies the threshold with the greatest vertical distance from the chance line when sensitivity and specificity are weighted equally. Correlated AUCs from the same participants were compared using the DeLong test.

For radiomics, intra- and inter-observer reproducibility of extracted features was quantified using intraclass correlation coefficients (ICCs). ICC values <0.50, 0.50–0.75, 0.75–0.90, and ≥0.90 were interpreted as poor, moderate, good, and excellent reliability, respectively. Features meeting the prespecified reproducibility threshold

were retained for subsequent analysis. All data-dependent preprocessing, feature selection, and hyperparameter tuning were confined to the training data. Model performance was summarized separately for training cross-validation and the held-out test cohort to distinguish optimization from independent internal evaluation. Sensitivity, specificity, PPV, NPV, accuracy, precision, recall, F1-score, and AUC with 95% confidence intervals were reported where appropriate.

For radiomics model comparison, AUCs obtained from predictions on the same held-out test patients were treated as correlated. The non-parametric DeLong test was used to compare prespecified pairs of ROC curves, particularly the best-performing model from each imaging modality. The DeLong test statistic was based on the covariance of the paired AUC estimates, and two-sided P values were reported. Statistical significance was defined as $P<0.05$. Comparisons not formally tested were described as numerical differences only and were not interpreted as evidence of statistical superiority. Exploratory multiple comparisons were interpreted cautiously in view of the limited number of test-cohort observations.

Decision curve analysis quantified net benefit across the stated threshold-probability range in the held-out test cohort. Sample-level prediction-score distributions were also generated for individual test-cohort patients to illustrate the separation and overlap of predicted probabilities between metastatic and non-metastatic groups.

Calibration of the radiomics models was evaluated using calibration curves, in which predicted probabilities were plotted against observed proportions of metastasis. A 45-degree line represented perfect calibration. Departure above this line indicated that the model underestimated risk, whereas departure below the line indicated overestimation. Calibration was assessed primarily in the held-out test cohort. The Brier score was calculated as the mean squared error between predicted probabilities and binary outcomes; values closer to 0 indicated better overall probabilistic accuracy. Calibration intercept and slope were estimated where feasible, with ideal values of 0 and 1, respectively. Given the small test cohort, visual calibration patterns and associated estimates were interpreted cautiously.

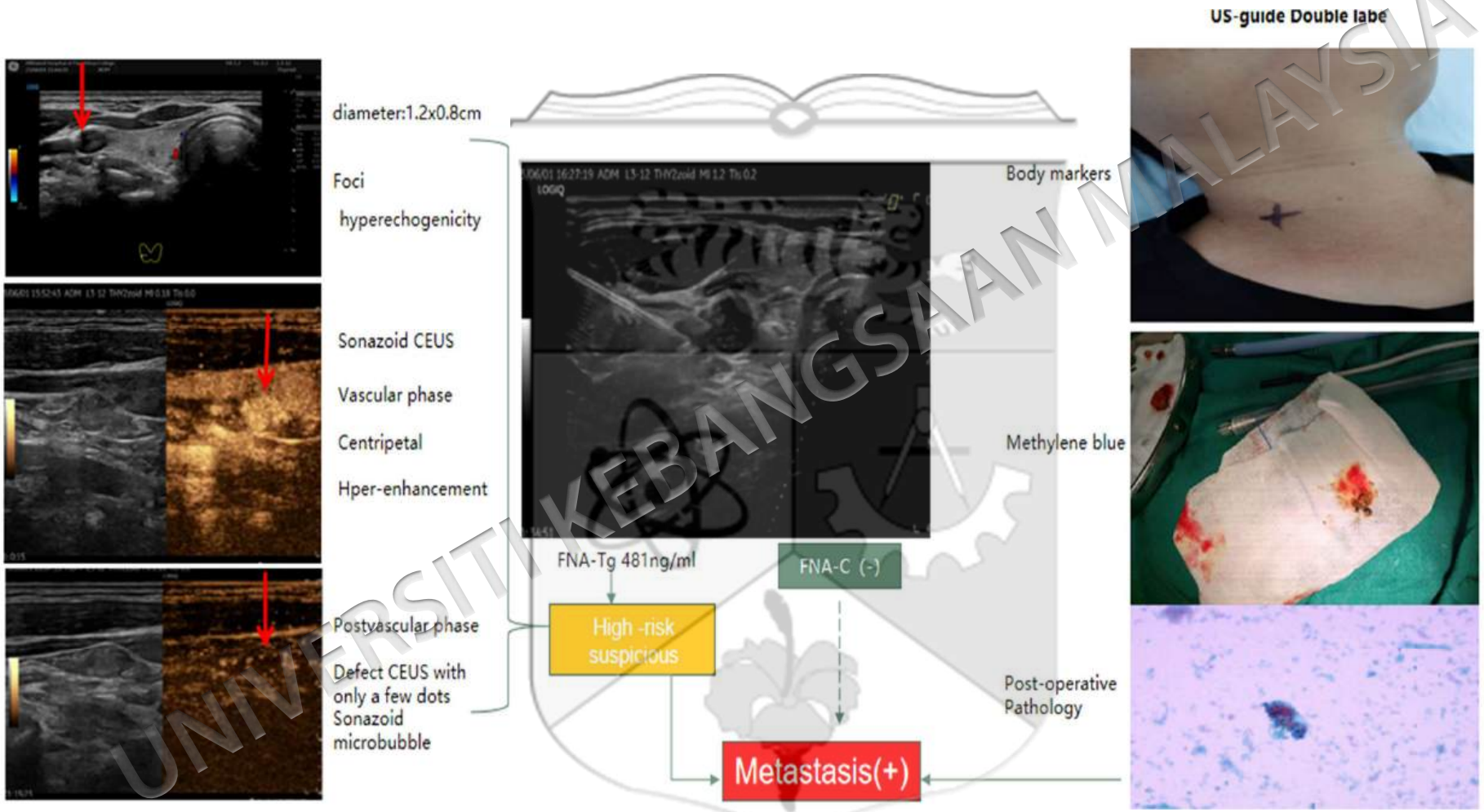


Figure 3.1 Flowchart of clinical evaluation of lateral cervical lymph nodes

A 22-year-old woman with a 0.3 cm papillary thyroid carcinoma in the right thyroid lobe underwent conventional ultrasound, Sonazoid contrast-enhanced ultrasound, FNA-C, and FNA-Tg evaluation of a level III lymph node in the right lateral neck (as indicated by the arrow). Conventional ultrasound showed a moderately hyperechoic area within the lymph node. The FNA-Tg value was 481 ng/mL. Sonazoid contrast-enhanced ultrasound and FNA-Tg both suggested metastatic papillary thyroid carcinoma, whereas FNA-C showed lymph node cells without definite tumor cells. Following dual localization by surface marking and ultrasound-guided methylene blue injection, the patient underwent total thyroidectomy, bilateral central lymph node dissection, and right lateral cervical lymph node dissection. Postoperative paraffin pathology confirmed lymph node metastasis.

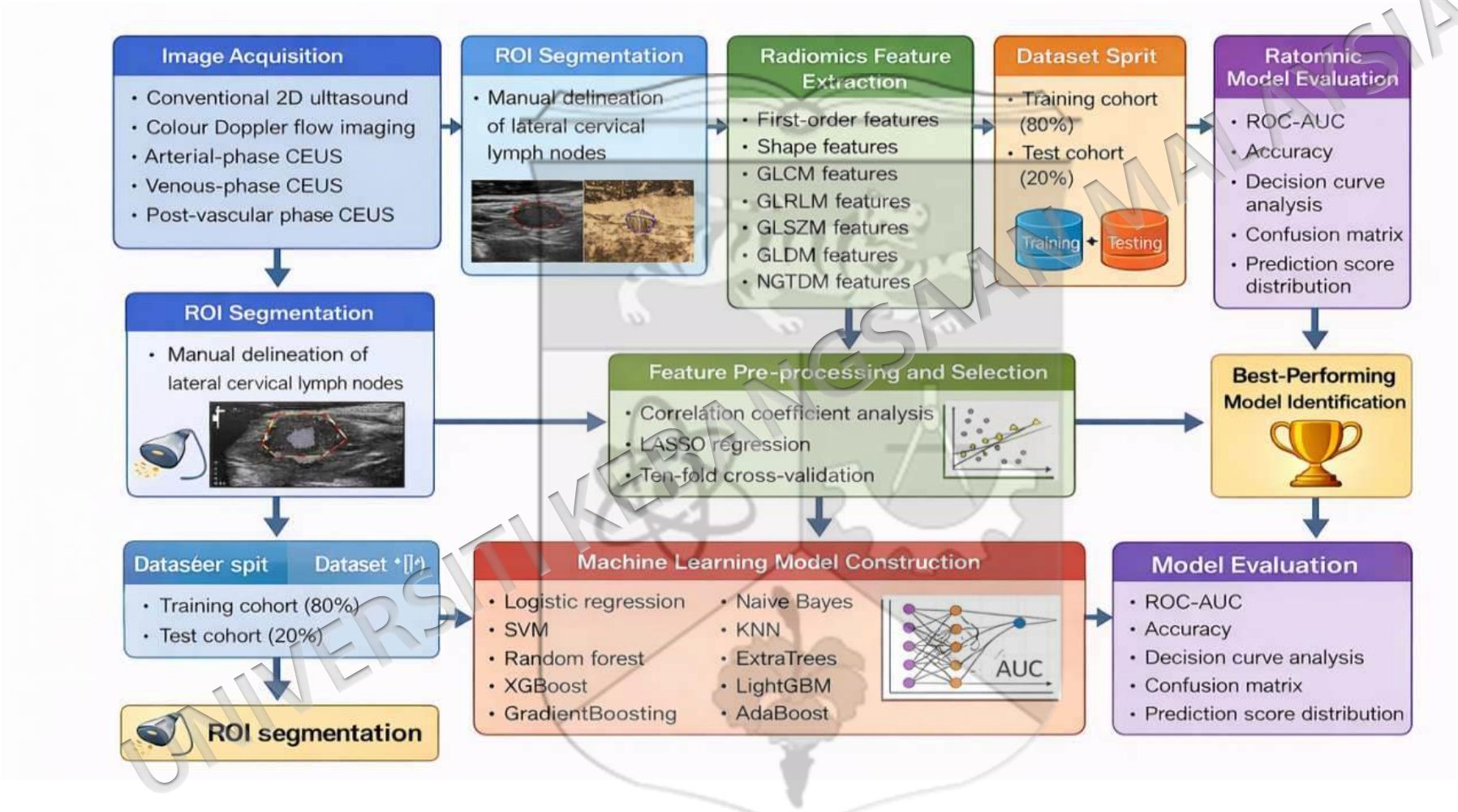
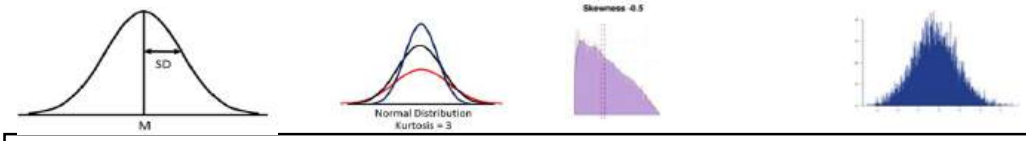


Figure 3.2 Radiomics workflow

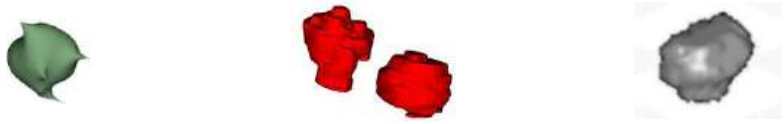
First Order Statistics (19 features):

Extracts the voxel-intensities within the region of interest in an image.



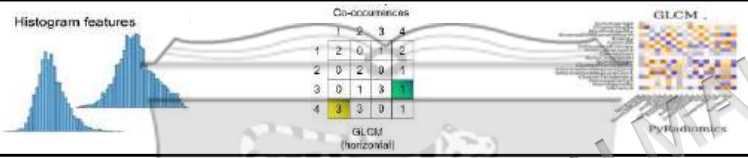
Shape-based (3D) (16 features) +Shape-based (2D) (10 features):

Describes the dimension shape/size of a volume or region of interest.



Gray Level Co-occurrence Matrix (24 features):

Describes the quantitative view of texture in an image using image histograms.



Gray Level Run Length Matrix (16 features):

Describes the distribution of pixel-intensities in the image run-length features.

Gray Level Size Zone Matrix (16 features):

Describes the number of related voxels which shares the similar gray-level intensities.

Neighbouring Gray Tone Difference Matrix (5 features):

Quantifies the difference between a gray-value and the average gray-value ofits neighbors.

Gray Level Dependence Matrix (14 features):

Describes the gray-level dependencies in an image from its central voxel.

Figure 3.3 Radiomics Features

CHAPTER IV

RESULTS

4.1 DEMOGRAPHIC AND CLINICAL CHARACTERISTICS OF THE STUDY POPULATION

4.1.1 Overall Demographic Characteristics

A total of 178 patients were included in the final analytical cohort at the Affiliated Hospital of Pan Zhihua University, China. Of these, 44 were retrospectively included from patients who had undergone the relevant examinations as part of routine clinical care between 1 February and 13 June 2024, with research inclusion occurring only after local ethical approval. The remaining 134 patients were enrolled from July 2024 to November 2025 under the approved study protocol. The same eligibility criteria, imaging requirements, reference standard, and data-quality procedures were applied to both groups. The mean age was 45.60 ± 13.30 years (range, 18–83 years). All participants were adults, and age was analysed as a continuous variable.

4.1.2 Final Grouping of Patients

Based on the findings of fine-needle aspiration cytology (FNA-C), 55 patients were diagnosed with metastatic lateral cervical lymph nodes, whereas 123 patients were diagnosed without metastasis.

Among the 178 patients, 140 highly suspicious patients underwent surgery and postoperative pathological examination. Histopathological examination confirmed lateral cervical lymph node metastasis in 82 patients, while 58 patients were confirmed to have no metastasis. The remaining 38 patients did not undergo surgery and were evaluated through follow-up. During follow-up, 1 patient was found to have lateral cervical lymph node metastasis, whereas 37 patients showed no evidence of metastasis.

Based on the final pathological and follow-up findings, 83 patients were included in the metastasis group, and 95 patients were included in the non-metastasis group. The process of patient inclusion and final grouping is illustrated in **Figure 4.1**.

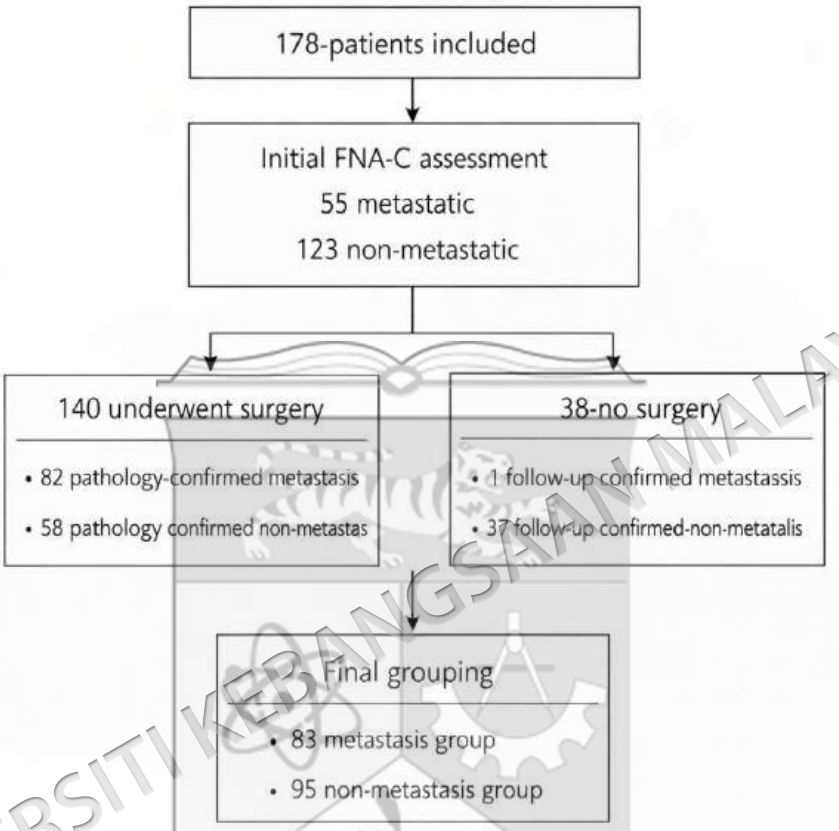


Figure 4.1 Flowchart of patient recruitment and final grouping of the study

4.1.3 Comparison of Age and Sex Between the Two Groups

In the non-metastasis group, 76 patients were female, and 19 were male. In the metastasis group, 54 patients were female, and 29 were male. The mean age of the metastasis group was 40.65 ± 14.13 years, whereas the mean age of the non-metastasis group was 49.92 ± 10.88 years. The mean age of patients in the metastasis group was significantly lower than that in the non-metastasis group ($p = 0.014$). The demographic and clinical characteristics of the patients are presented in **Table 4.1**.

Table 4.1 Demographic characteristics of the study participants

Characteristic	Category	Total (n=178)	Metastasis (n=83)	Non-Metastasis (n=95)	P-value
Sex	Female	130	54	76	-
	Male	48	29	19	-
Age (Years)	Mean ± SD	45.60 ±13.30	40.65 ± 14.13	49.92 ±10.88	0.014

4.1.4 Distribution of Lateral Cervical Lymph Nodes

A total of 109 lateral cervical lymph nodes were located on the left side, whereas 69 were located on the right side. According to cervical nodal level, 1 lymph node was located in level I, 14 in level II, 72 in level III, 89 in level IV, and 2 in level V. There was no statistically significant difference in the distribution of cervical nodal levels between the metastasis and non-metastasis groups ($p = 0.880$). The distribution of lateral cervical lymph nodes is shown in **Table 4.2**.

Table 4.2 Regional distribution of lateral cervical lymph nodes

Group	I	II	III	IV	V	Total	P-value
Non-Metastasis	1	7	40	46	1	95	
Metastasis	0	7	32	43	1	83	0.88
Total	1	14	72	89	2	178	

4.1.5 Comparison of Lymph Node Dimensions Between the Two Groups

The long diameter of the lateral cervical lymph nodes ranged from 0.40 cm to 2.84 cm, whereas the short diameter ranged from 0.24 cm to 1.60 cm. There was no statistically significant difference in the long diameter between the metastasis and non-metastasis groups ($p = 0.517$). Similarly, no statistically significant difference was observed in the long-to-short diameter ratio between the two groups ($p = 0.598$). However, the short diameter was significantly greater in the metastasis group than in the non-metastasis group ($p < 0.001$). The comparison of lymph node dimensions is presented in **Table 4.3**.

Table 4.3 Comparison of lymph node diameters by metastasis status

Parameter	Non-Metastasis (n=95)	Metastasis (n=83)	P-value
Long diameter (cm)	1.19 ± 0.45	1.30± 0.48	0.517
Short diameter (cm)	0.55 ± 0.13	0.64 ± 0.23	<0.001
L/S Ratio	2.22 ± 0.08	2.16 ±0.09	0.598

4.1.6 Laboratory Findings of the Study Population

The laboratory findings of the study population are presented in **Table 4.4**. Serum thyroglobulin (Tg) levels were within the normal range in 99 patients (55.6%), decreased in 62 patients (34.8%), and increased in 17 patients (9.6%). Serum thyroglobulin antibody (TgAb) levels were within the normal range in 151 patients (84.8%) and increased in 27 patients (15.2%). Serum thyroid-stimulating hormone (TSH) levels were within the normal range in 142 patients (79.8%), decreased in 29 patients (16.3%), and increased in 7 patients (3.9%). Serum thyroid peroxidase antibody (TPO-Ab) levels were within the normal range in 143 patients (80.3%) and increased in 35 patients (19.7%).

Table 4.4 Thyroid function indicators of the study participants(N=178)

Status	Serum Tg	Serum TgAb	Serum TSH	Serum TPO-Ab
Normal (n%)	99 (55.6%)	151 (84.8%)	142 (79.8%)	143 (80.3%)
Decrease (n%)	62 (34.8%)	-	29 (16.3%)	-
Increase (n%)	17 (9.6%)	27 (15.2%)	7 (3.9%)	35 (19.7%)

4.1.7 Comparison of Laboratory Parameters Between the Two Groups

As shown in **Table 4.5**, the mean serum Tg level in the metastasis group was significantly higher than that in the non-metastasis group ($p < 0.001$). The mean serum TgAb level was also significantly higher in the metastasis group than in the non-metastasis group ($p = 0.002$). In addition, the mean serum TPO-Ab level was significantly higher in the metastasis group than in the non-metastasis group ($p = 0.003$). Although the mean serum TSH level was higher in the metastasis group than in the non-metastasis group, the difference was not statistically significant ($p = 0.663$).

Table 4.5 Thyroid function values by metastasis status

Parameter	Non-Metastasis (n=95)	Metastasis (n=83)	P-value
Serum Tg (unit)	10.44 ± 2.33	24.39 ± 3.67	<0.001
Serum TgAb (unit)	7.35 ± 2.38	17.14 ± 6.04	0.002
Serum TSH (unit)	2.06 ± 0.23	2.30 ± 0.21	0.663
Serum TPO-Ab (unit)	45.41 ± 18.93	98.98 ± 31.18	0.003

4.2 CONVENTIONAL ULTRASOUND FINDINGS

4.2.1 Distribution of Conventional Ultrasound Manifestations

The evaluation of 178 lateral cervical lymph nodes focused on seven suspicious features identified on conventional ultrasound, as described in the methodology. The detailed distribution and frequency of these manifestations are shown in **Table 4.6**.

Among the 178 lateral cervical lymph nodes examined, the most frequent combination of suspicious features was ①②, which was observed in 37 lymph nodes (20.8%). This combination was found exclusively in the non-metastasis group. The second most frequent combinations were ①②④ and ①②⑦, each occurring in 13 lymph nodes (7.3%). Feature ② alone was observed in 12 lymph nodes (6.7%), while feature combination ②⑦ was identified in 11 lymph nodes (6.2%). Several other combinations, including ⑤⑥, ④⑤, ②④⑤, ④⑤⑦, ④⑤⑥⑦, ①③⑤⑥⑦, and ①③④⑤⑥, were each observed in only 1 lymph node (0.6%).

Lymph nodes demonstrating feature ⑤ (hyperechogenicity), feature ⑥ (cystic appearance), or feature ⑦ (calcification) were observed only in the metastasis group.

Table 4.6 Conventional ultrasound features of lateral cervical lymph **nodes**

	Conventional ultrasound characteristic																																															
features	①②	①②③	①②③④⑤	①②③④⑤⑥	①②③④⑥	①②③④⑥⑦	①②③④⑦	①②③⑥	①②④	①②④⑤	①②④⑦	①②⑤	①②⑥	①②⑥⑦	①②⑦	①③	①③④⑤	①③④⑤⑥	①③④⑤⑥⑦	①③④⑤⑦	①③④⑦	①③⑤	①③⑤⑥⑦	①③⑤⑦	①③⑥	①③⑦	①④⑤	①④⑤⑥⑦	①④⑤⑦	①④⑦	①⑤⑦	①⑦	②	②④	②④⑤	②④⑦	②⑤	②⑥	②⑦	③④⑤⑥⑦	③④⑤⑦	③⑤	③⑥⑦	④⑤	④⑤⑥⑦	④⑤⑦	⑤⑥	Total
Frequency	37	3	2	1	1	2	2	4	13	2	7	3	1	1	13	4	4	1	2	5	3	2	1	2	2	3	2	1	3	1	1	1	12	11	1	1	1	1	11	2	1	1	2	1	1	1	1	178
Percent	20.8	1.7	1.1	0.6	0.6	1.1	1.1	2.2	7.3	1.1	3.9	1.7	0.6	0.6	7.3	2.2	2.2	0.6	1.1	2.8	1.7	1.1	0.6	1.1	1.1	1.7	1.1	0.6	1.7	0.6	0.6	0.6	6.7	6.2	0.6	0.6	0.6	0.6	6.2	1.1	0.6	0.6	1.1	0.6	0.6	0.6	0.6	100
Non-metastasis	Yes	Yes	-	-	-	-	-	-	Yes	-	-	-	-	-	-	Yes	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Yes	Yes	-	-	-	-	-	-	-	-	-	-	-	-	-	80
Metastasis	-	-	Yes	Yes	Yes	Yes	Yes	Yes	-	Yes	Yes	Yes	Yes	Yes	Yes	-	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	-	-	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	98

Note: ① lymph nodes with a short diameter >0.5 cm, lymph nodes with a long diameter / short diameter (L / S) ratio<1.5, rounded shape. ②ill-defined margin or irregular margin， centripetal thickening of the cortex of lymph nodes, or decreased echogenicity levels in the cortex and medulla ③absence of lymphatic hilum or eccentricity.④ peripheral vascularization， or blood flow signals, were losing their regularity.⑤ diffuse or focal hyper-echogenicity. ⑥ cystic appearance. ⑦ calcifications; Have ⑤or ⑥or ⑦ conventional ultrasound qualitative diagnoses metastasis

4.2.2 Qualitative Diagnoses of Conventional Ultrasound

For qualitative diagnoses, lateral cervical lymph nodes were classified as metastatic when they demonstrated feature ⑤, ⑥, or ⑦ on conventional ultrasound. Based on this criterion, 98 lymph nodes were diagnosed as metastatic and 80 as non-metastatic. When compared with the final diagnoses, qualitative conventional ultrasound showed moderate agreement, with a kappa value of 0.676. The association between qualitative conventional ultrasound diagnoses and the final diagnoses was statistically significant ($p < 0.001$). These findings are presented in **Table 4.7**.

Table 4.7 Qualitative diagnostic efficacy of conventional ultrasound

Diagnoses Method	Non-metastasis (Final)	Metastasis (Final)	Total	P	Kappa
Ultrasound Qualitative (+)	22	76	98		
Ultrasound Qualitative (-)	73	7	80	<0.001	0.676
Total	95	83	178		

4.2.3 Quantitative Scoring of Conventional Ultrasound

To objectively quantify the risk of metastasis, a weighted scoring system was applied to the suspicious signs observed during conventional ultrasound examination, as described in the methodology. Scores were assigned according to morphological, structural, echogenicity-related, and vascular characteristics, ranging from +1 for basic size-related criteria to +5 for highly specific features such as fine punctate calcification or diffuse hyper-echogenicity.

The mean conventional ultrasound score was 9.22 ± 3.78 in the metastasis group and 3.39 ± 2.23 in the non-metastasis group. The score was significantly higher in the metastasis group than in the non-metastasis group ($p < 0.001$). The comparison of conventional ultrasound scores between the two groups is presented in **Table 4.8**.

Table 4.8 Quantitative scores of conventional ultrasound

Group	Mean Score	Standard Deviation	P-value
Non-metastasis (n=95)	3.39	2.23	< 0.001
Metastasis (n=83)	9.22	3.78	

4.2.4 Diagnostic Performance of Quantitative and Qualitative Conventional Ultrasound

When the cut-off value for the quantitative conventional ultrasound score was set at 5.5, the sensitivity and specificity were 0.880 and 0.968, respectively. The Youden index was 0.848, and the area under the receiver operating characteristic curve (ROC-AUC) was 0.914 (95% CI: 0.863–0.951). By comparison, qualitative conventional ultrasound showed a sensitivity of 0.924, a specificity of 0.661, a Youden index of 0.585, and an AUC of 0.842. The diagnostic performance of qualitative and quantitative conventional ultrasound is presented in **Table 4.9**.

Table 4.9 Qualitative and quantitative diagnostic performance of conventional ultrasound

Method	Cut-off	Sensitivity	Specificity	YI	AUC	95% CI	SE	P
Qualitative	N/A	0.924	0.661	0.585	0.842	0.780–0.892	0.0266	0.0001
Quantitative	5.5	0.88	0.968	0.848	0.914	0.863–0.951	0.0224	

4.2.5 Comparison of Quantitative and Qualitative Diagnostic Efficacy of Conventional Ultrasound

The quantitative diagnostic efficacy of conventional ultrasound was higher than that of qualitative conventional ultrasound for the assessment of lateral cervical lymph node metastasis, as indicated by the higher AUC value (0.914 vs. 0.842). The difference between the two methods was statistically significant according to the DeLong test ($p < 0.001$), and the corresponding standard error was 0.0224. This standard error reflects the precision of the AUC estimate but should not, by itself, be interpreted as evidence of model robustness or external validity. The comparison of quantitative and qualitative diagnostic efficacy is illustrated in **Figure 4.2**.

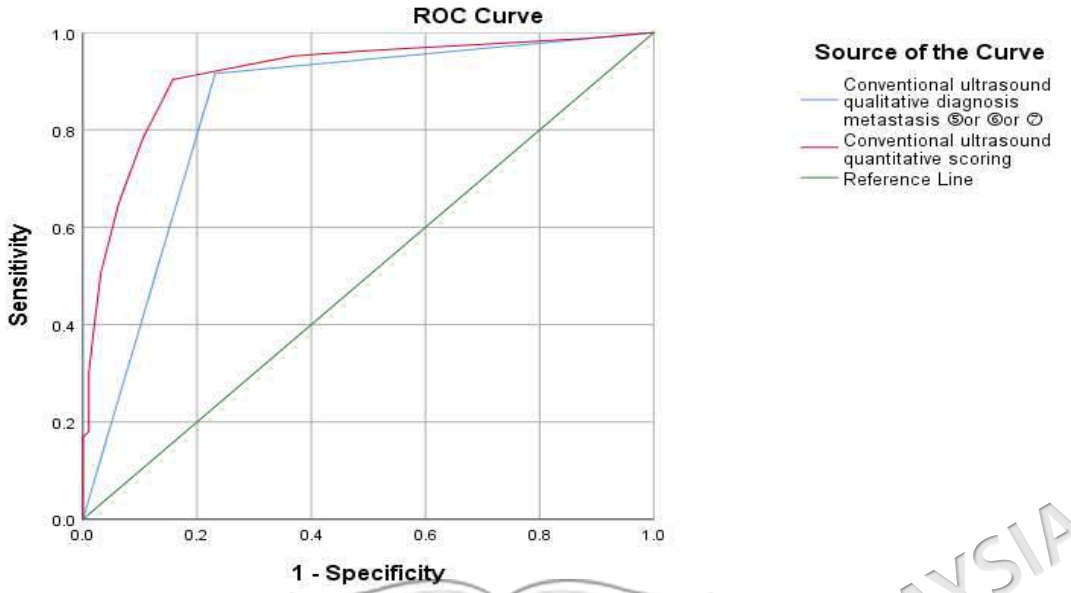


Figure 4.2 Comparison of quantitative and qualitative diagnostic efficacy of conventional ultrasound

4.3 FA-C AND FNA-TG FINDINGS

4.3.1 Diagnostic Performance of FNA-C

Fine-needle aspiration cytology (FNA-C) classified 55 lateral cervical lymph nodes as metastatic and 123 as non-metastatic. Among the 178 patients included in the study, 140 with highly suspicious findings underwent surgery and postoperative pathological examination, which confirmed 82 metastatic and 58 non-metastatic lateral cervical lymph nodes. The remaining 38 patients did not undergo surgery and were evaluated by follow-up, which identified 1 additional metastatic lymph node and 37 non-metastatic lymph nodes. When compared with the final diagnoses, FNA-C showed moderate agreement, with a kappa value of 0.562. The association between FNA-C diagnoses and the final diagnoses was statistically significant ($p < 0.001$). The diagnostic performance of FNA-C is presented in **Table 4.10**.

Table 4.10 Comparison of FNA-C and final diagnoses

FNA-C Result	Non-metastasis (Final)	Metastasis (Final)	Total	P-value	Kappa
Non-metastasis	90	33	123		
Metastasis	5	50	55	< 0.001	0.562
Total	95	83	178		

4.3.2 Distribution and Qualitative Classification of FNA-Tg

The FNA-Tg level ranged from 0 to 79,005 ng/ml. Among the 178 lateral cervical lymph nodes, 65 had an FNA-Tg value of < 1.0 ng/ml, 37 had an FNA-Tg value of 1.0–10.0 ng/ml, and 76 had an FNA-Tg value of > 10.0 ng/ml. According to the predefined qualitative criteria, lymph nodes were classified as non-metastatic when FNA-Tg < 1.0 ng/ml or when FNA-Tg was 1.0–10.0 ng/ml with an FNA-Tg/serum Tg ratio ≤ 1. Lymph nodes were classified as metastatic when FNA-Tg > 10.0 ng/ml or when FNA-Tg was 1.0–10.0 ng/ml with an FNA-Tg/serum Tg ratio > 1. The detailed distribution and classification of FNA-Tg are presented in **Table 4.11**.

Table 4.11 Classification of FNA-Tg levels and qualitative criteria

FNA-Tg Value	FNA-Tg/Serum-Tg Ratio	Frequency	Percentage (%)	Qualitative Diagnoses
< 1.0 ng/ml	-	65	36.5	Non-metastasis
1.0–10.0 ng/ml	≤ 1	25	14.1	Non-metastasis
	> 1	12	6.7	Metastasis
> 10.0 ng/ml	-	76	42.7	Metastasis
Total	-	178	100	-

4.3.3 Qualitative Diagnostic Performance of FNA-Tg

Using the above qualitative criteria, FNA-Tg showed good agreement with the final diagnoses, with a kappa value of 0.831. The association between qualitative FNA-Tg diagnoses and the final diagnoses was statistically significant ($p < 0.001$). The qualitative diagnostic performance of FNA-Tg is presented in **Table 4.12**.

Table 4.12 Qualitative diagnostic performance of FNA-Tg

FNA-Tg Qualitative Diagnoses	Non-metastasis (Final)	Metastasis (Final)	Total	P-value	Kappa
Non-metastasis	85	5	90		
Metastasis	10	78	88	< 0.001	0.831
Total	95	83	178		

4.3.4 Quantitative Analysis of FNA-Tg

The mean FNA-Tg value in the metastasis group was 3675.10 ± 1179.62 ng/ml, whereas the mean value in the non-metastasis group was 1.58 ± 0.28 ng/ml. The mean FNA-Tg value was significantly higher in the metastasis group than in the non-metastasis group ($p < 0.001$). The comparison of the FNA-Tg value between the two groups is presented in **Table 4.13**.

Table 4.13 Quantitative FNA-Tg values by metastasis status

Group	n	Mean FNA-Tg (ng/ml)	P-value
Metastasis	83	3675.10 ± 1179.62	< 0.001
Non-Metastasis	95	1.58 ± 0.28	

4.3.5 Diagnostic Performance of Quantitative and Qualitative FNA-Tg

When the cut-off value of FNA-Tg was set at 9.10 ng/ml, the sensitivity and specificity of quantitative FNA-Tg for the diagnosis of lateral cervical lymph node metastasis were 0.880 and 0.968, respectively. The Youden index was 0.848, and the ROC-AUC was 0.953 (95% CI: 0.917–0.990). By comparison, qualitative FNA-Tg showed a sensitivity of 0.940, a specificity of 0.895, a Youden index of 0.835, and an AUC of 0.917. The diagnostic performance of qualitative and quantitative FNA-Tg is presented in **Table 4.14**.

Table 4.14 Qualitative and quantitative diagnostic performance of FNA-Tg

Method	Cut-off	Sensitivity	Specificity	AUC	95% CI	SE	P-value
Qualitative	NA	0.94	0.895	0.917	0.871–0.964	0.0206	0.0456
Quantitative	9.1	0.88	0.968	0.953	0.917–0.990	0.0185	

4.3.6 Comparison of Quantitative and Qualitative Diagnostic Efficacy of FNA-Tg

The quantitative diagnostic efficacy of FNA-Tg was higher than the qualitative diagnostic efficacy, as reflected by the higher AUC value (0.953 vs. 0.917). The difference between the two methods was statistically significant according to the DeLong test ($p = 0.0456$); the standard error of the quantitative AUC estimate was

0.0185. This value reflects estimation precision and does not independently establish robustness or generalizability. The comparison of quantitative and qualitative diagnostic efficacy is illustrated in **Figure 4.3**.

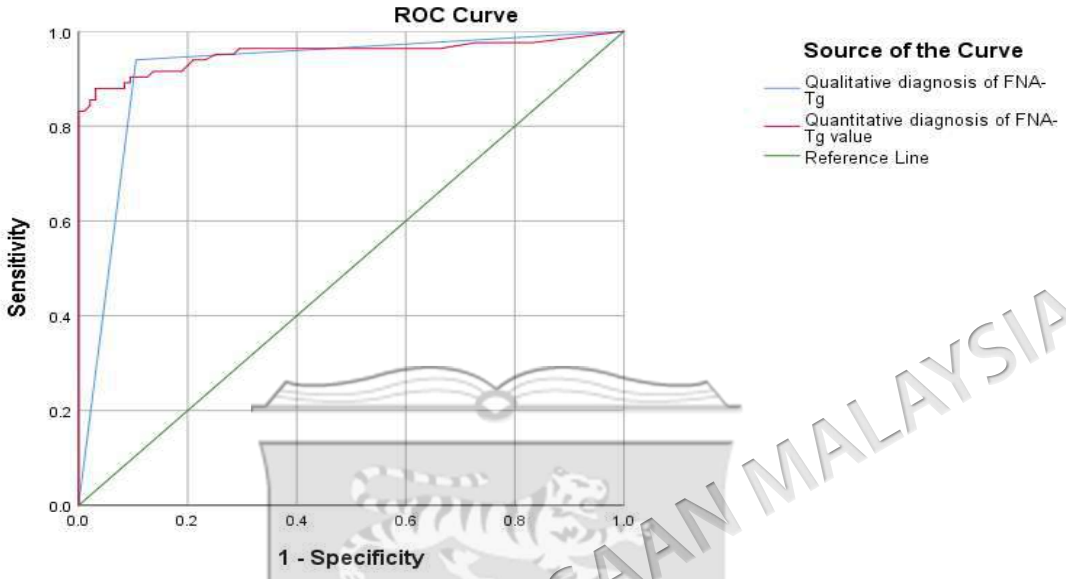


Figure 4.3 Comparison of quantitative and qualitative diagnostic efficacy of FNA-Tg

4.4 SONAZOID CONTRAST-ENHANCED ULTRASOUND (CEUS) FINDINGS

4.4.1 Arterial-Phase of Sonazoid CEUS

In the arterial phase of Sonazoid contrast-enhanced ultrasound (CEUS), six enhancement patterns were identified, namely centrifugal enhancement, overall (integral)enhancement, chaotic enhancement, centripetal enhancement, no enhancement, and circular enhancement with internal non-enhancement. The distribution of enhancement patterns in the arterial phase across 178 lateral cervical lymph nodes is presented in **Table 4.15**.

For qualitative diagnoses, centrifugal enhancement and overall enhancement were classified as non-metastatic patterns, whereas chaotic enhancement, centripetal enhancement, no enhancement, and circular enhancement with no enhancement inside were classified as metastatic patterns. The qualitative diagnostic results of arterial phase Sonazoid CEUS for lateral cervical lymph nodes are shown in **Table 4.16**.

A quantitative scoring system was also applied to the arterial phase enhancement patterns, as described in the methodology, with the following scores assigned: centrifugal enhancement (+1), overall enhancement (+2), chaotic enhancement (+3), no enhancement (+4), circular enhancement with no enhancement inside (+4), and centripetal enhancement (+5). The quantitative score distribution of arterial phase Sonazoid CEUS for lateral cervical lymph nodes was presented in **Table 4.17**.

The mean arterial phase quantitative score was 4.37 ± 0.129 in the metastasis group and 1.13 ± 0.043 in the non-metastasis group. The arterial phase quantitative score was significantly higher in the metastasis group than in the non-metastasis group ($p < 0.001$).

When the cut-off value of the arterial phase quantitative score was set at 2.5, the sensitivity and specificity of quantitative arterial phase Sonazoid CEUS for the diagnosis of lateral cervical lymph node metastasis were 0.892 and 0.989, respectively. The Youden index was 0.881, and the ROC-AUC was 0.969 (95% CI: 0.940–0.998). By comparison, qualitative arterial phase Sonazoid CEUS showed an AUC of 0.941. The diagnostic performance of qualitative and quantitative arterial-phase Sonazoid CEUS for lateral cervical lymph node metastasis is presented in **Table 4.18**.

Quantitative score arterial-phase Sonazoid CEUS demonstrated higher diagnostic efficacy than qualitative assessment for lateral cervical lymph node metastasis, as indicated by the higher AUC value (0.969 vs. 0.941). The difference between the two AUC values was statistically significant according to the DeLong test ($p = 0.0227$); the standard error of the quantitative AUC estimate was 0.0133. This value reflects the precision of the AUC estimate and should not be interpreted as proof of external robustness. The comparison of diagnostic efficacy between the two methods of qualitative and quantitative arterial phase Sonazoid CEUS for lateral cervical lymph node metastasis is illustrated in **Figure 4.4**.

Table 4.15 Arterial-phase enhancement patterns of Sonazoid CEUS

Enhancement pattern	Frequency	Percentage (%)	Non-metastasis	Metastasis
Centrifugal enhancement	89	50	85	4
Overall(integral)enhancement	14	7.9	9	5
Chaotic enhancement	8	4.5	0	8
Centripetal enhancement	61	34.3	0	61
Non-enhancement	2	1.1	1	1
Circular enhancement with internal non-enhancement	4	2.2	0	4
Total	178	100	95	83

Table 4.16 Qualitative diagnoses of arterial-phase Sonazoid CEUS

Arterial phase qualitative diagnoses	Non-metastasis (Final)	Metastasis (Final)	Total	P
Non-metastasis	94	9	103	
Metastasis	1	74	75	<0.001
Total	95	83	178	

Table 4.17 Quantitative scores of arterial phase Sonazoid CEUS

Arterial phase quantitative score	Non-metastasis (Final)	Metastasis (Final)	Total
1	85	4	89
2	9	5	14
3	0	8	8
4	1	5	6
5	0	61	61
Total	95	83	178

Table 4.18 Qualitative and quantitative diagnostic performance of arterial-phase Sonazoid CEUS

Method	Cut-off value	Sensitivity	Specificity	YI	AUC	95% CI		SE	P
Qualitative	NA	0.892	0.989	0.881	0.941	0.899	0.982	0.0180	0.0227
Quantitative	2.5	0.892	0.989	0.881	0.969	0.940	0.998	0.0133	

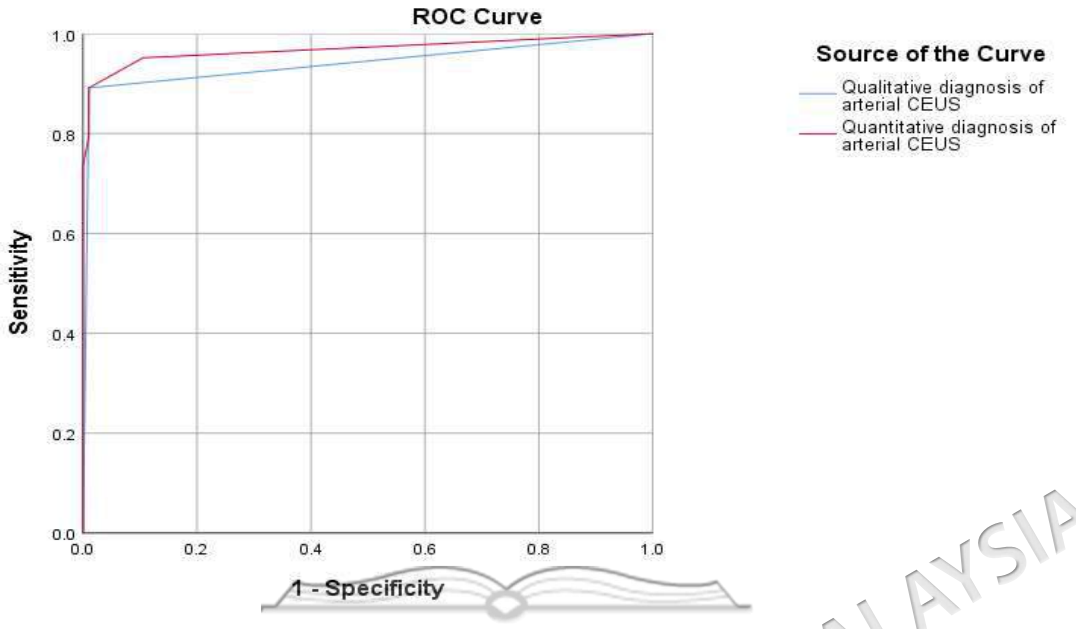


Figure 4.4 Comparison of quantitative and qualitative diagnostic efficacy of arterial-phase Sonazoid CEUS

4.4.2 Venous-Phase of Sonazoid CEUS

In the venous phase of Sonazoid CEUS, three enhancement classifications were identified, namely homogeneous enhancement, heterogeneous enhancement, and no enhancement. The distribution of these venous phase enhancement classifications of lateral cervical lymph nodes is shown in **Table 4.19**.

For qualitative diagnoses, homogeneous enhancement was classified as non-metastatic, whereas heterogeneous enhancement and no enhancement were classified as metastatic. The qualitative diagnostic results of venous phase Sonazoid CEUS for lateral cervical lymph nodes in PTC are presented in **Table 4.20.**

A quantitative scoring system was further applied to the venous phase enhancement as described in the methodology, homogeneous (+1), heterogeneous with single defect 0.1- 0.2cm (+2), heterogeneous with single defect 0.2-0.3cm (+3), no enhancement (+4), extremely heterogeneous with defect>0.3cm (+5), disordered with more than 3 defect and every defect > 0.1cm (+5). The quantitative score distribution for the venous phase Sonazoid CEUS is presented in **Table 4.21**.

The mean venous phase score was 3.06 ± 0.163 in the metastasis group and 1.16 ± 0.062 in the non-metastasis group. The score was significantly higher in the metastasis group than in the non-metastasis group ($p < 0.001$). At a cut-off value of 1.5, the sensitivity and specificity of quantitative venous phase Sonazoid CEUS for the diagnoses of lateral cervical lymph node metastasis were 0.880 and 0.916, respectively. The Youden index was 0.796, and the ROC-AUC was 0.904 (95% CI: 0.855–0.954). The diagnostic performance of qualitative and quantitative venous phase Sonazoid CEUS is presented in **Table 4.22**.

Qualitative venous phase Sonazoid CEUS showed an AUC of 0.898. Although the quantitative venous phase method showed a numerically higher AUC than the qualitative method (0.904 vs. 0.898), but no statistically significant according to the DeLong test; the p-value was 0.3416. The comparison of diagnostic efficacy between the two methods of qualitative and quantitative venous phase Sonazoid CEUS for lateral cervical lymph node metastasis is illustrated in **Figure 4.5**.

Table 4.19 Venous-phase enhancement patterns of Sonazoid CEUS

Venous phase classification	Frequency	Percentage (%)	Non-metastasis (Final)	Metastasis (Final)
Homogeneous	97	54.5	87	10
Heterogeneous	78	43.8	7	71
Non-enhancement	3	1.7	1	2
Total	178	100.0	95	83

Table 4.20 Qualitative diagnoses of venous-phase Sonazoid CEUS

Venous phase qualitative diagnoses	Non-metastasis (Final)	Metastasis (Final)	Total	P
Non-metastasis	87	10	97	
Metastasis	8	73	81	<0.001
Total	95	83	178	

Table 4.21 Quantitative scores of venous-phase Sonazoid CEUS

Venous phase quantitative score	Non-metastasis (Final)	Metastasis (Final)	Total
5	1	27	28
4	1	2	3
3	2	13	15
2	4	31	35
1	87	10	97
Total	95	83	178

Table 4.22 Qualitative and quantitative diagnostic performance of venous-phase Sonazoid CEUS

Method	Cut-off value	Sensitivity	Specificity	YI	AUC	95% CI		SE	P
Qualitative	NA	0.880	0.916	0.796	0.898	0.846	0.950	0.023	0.3416
Quantitative	1.5	0.880	0.916	0.796	0.904	0.855	0.954	0.228	

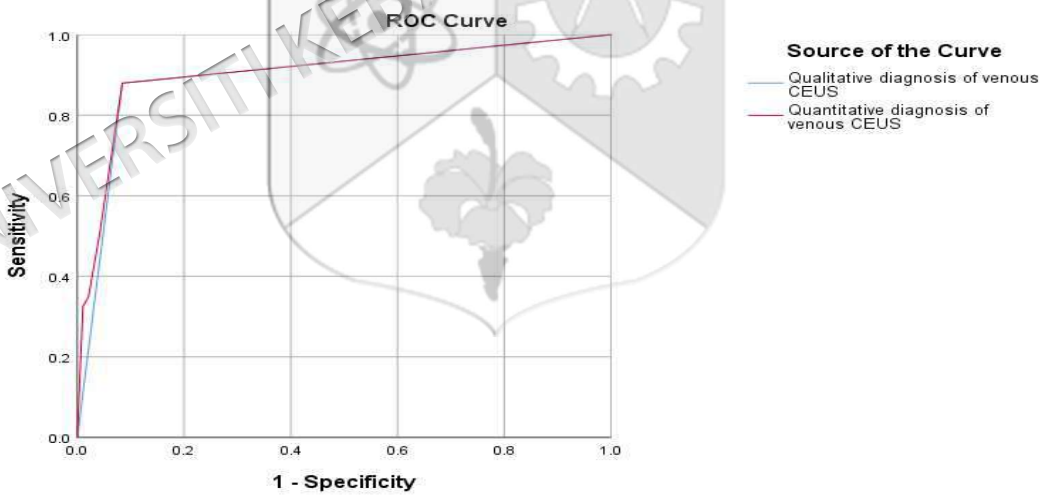


Figure 4.5 Comparison of quantitative and qualitative diagnostic efficacy of venous-phase CEUS

4.4.3 Post-Vascular-Phase of Sonazoid CEUS

In the post-vascular phase of Sonazoid CEUS, seven enhancement patterns were identified: homogeneous hyper-enhancement, heterogeneous hyper-enhancement with

a defect of 0.1- 0.3 cm, homogeneous hypo-enhancement, heterogeneous hypo-enhancement with a defect of 0.1- 0.3 cm, circular enhancement with internal non-enhancement, non-enhancement, and marked filling defect. Marked filling defect was defined as a single defect> 0.3 cm, multiple defects>0.1cm, or sparse microbubbles. The distribution of post-vascular phase enhancement patterns of lateral cervical lymph nodes is presented in **Table 4.23**.

For qualitative diagnoses, homogeneous hyper-enhancement was classified as non-metastatic, whereas heterogeneous hyper-enhancement with a defect of 0.1- 0.3 cm, homogeneous hypo-enhancement, heterogeneous hypo-enhancement, circular enhancement with internal non-enhancement, non-enhancement, and marked filling defect were classified as metastatic patterns. The qualitative diagnostic results of post-vascular phase Sonazoid CEUS for lateral cervical lymph nodes are shown in **Table 4.24**.

A quantitative scoring system was applied to the post-vascular phase enhancement as described in the methodology, homogeneous hyper-enhancement (+1), heterogeneous hyper-enhancement with single defect 0.1-0.3cm(+2), homogeneous hypo-enhancement(+3), heterogeneous hypo-enhancement with single defect 0.1-0.3cm (+4), circular enhancement no enhancement inside (+4), no enhancement (+4), marked filling with single defect > 0.3cm(+5)， marked filling with multiple defect > 0.1cm (+5)， marked filling only sparse microbubbles(+5). The quantitative score distribution for the post-vascular phase of lateral cervical lymph nodes is presented in **Table 4.25**.

The mean post-vascular phase score was 4.37 ± 0.121 in the metastasis group and 1.43 ± 0.097 in the non-metastasis group. The score was higher in the metastasis group than in the non-metastasis group ($p = 0.044$). At a cut-off value of 3.5, the sensitivity and specificity of quantitative post-vascular phase Sonazoid CEUS for diagnosing lateral cervical lymph node metastasis were 0.843 and 0.937, respectively. The Youden index was 0.780, and the ROC-AUC was 0.951 (95% CI: 0.908–0.978). The diagnostic performance of qualitative and quantitative post-vascular phase Sonazoid CEUS of lateral cervical lymph nodes is presented in **Table 4.26**.

Qualitative post-vascular phase Sonazoid CEUS showed an AUC of 0.901. Quantitative post-vascular phase Sonazoid CEUS demonstrated higher diagnostic efficacy than qualitative assessment, as reflected by the higher AUC value (0.951 vs. 0.901). The difference between the two AUC values was statistically significant according to the DeLong test (p = 0.00026); the standard error of the quantitative AUC estimate was 0.015. This value describes the precision of the estimate and does not independently demonstrate model robustness. The comparison of diagnostic efficacy between the qualitative and quantitative post-vascular phase Sonazoid CEUS for lateral cervical lymph nodes is illustrated in **Figure 4.6**.

Table 4.23 Post-vascular-phase enhancement patterns of Sonazoid CEUS

Post-vascular-phase classification	**Frequency**	**Percentage (%)**
Homogeneous hyper-enhancement	87	48.9
Heterogeneous hyper-enhancement (single defect 0.1- 0.3 cm)	12	6.8
Homogeneous hypo-enhancement	3	1.7
Heterogeneous hypo-enhancement (single defect 0.1 - 0.3cm)	7	3.9
Circular enhancement with internal non-enhancement	2	1.1
Non-enhancement	7	3.9
Marked filling defect (single defect > 0.3cm or multiple > 0.1cm or sparse microbubbles)	60	33.7
Total	178	100.0

Table 4.24 Qualitative diagnoses of post-vascular-phase Sonazoid CEUS

Post-vascular-phase qualitative diagnoses	**Non-metastasis (Final)**	**Metastasis (Final)**	**Total**	**P**
Non-metastasis	82	5	87	
Metastasis	13	78	91	<0.001
Total	95	83	178	

Table 4.25 Quantitative scores of post-vascular-phase Sonazoid CEUS

Post-vascular-phase quantitative score	Non-metastasis (Final)	Metastasis (Final)	Total
1	72	1	73
2	14	11	25
3	3	1	4
4	3	13	16
5	3	57	60
Total	95	83	178

Table 4.26 Qualitative and quantitative diagnostic performance of post-vascular-phase CEUS

Method	Cut-off value	Sensitivity	Specificity	YI	AUC	95% CI		SE	P
Qualitative	NA	0.940	0.863	0.803	0.901	0.848	0.941	0.022	0.000
Quantitative	3.5	0.843	0.937	0.780	0.951	0.908	0.978	0.015	26

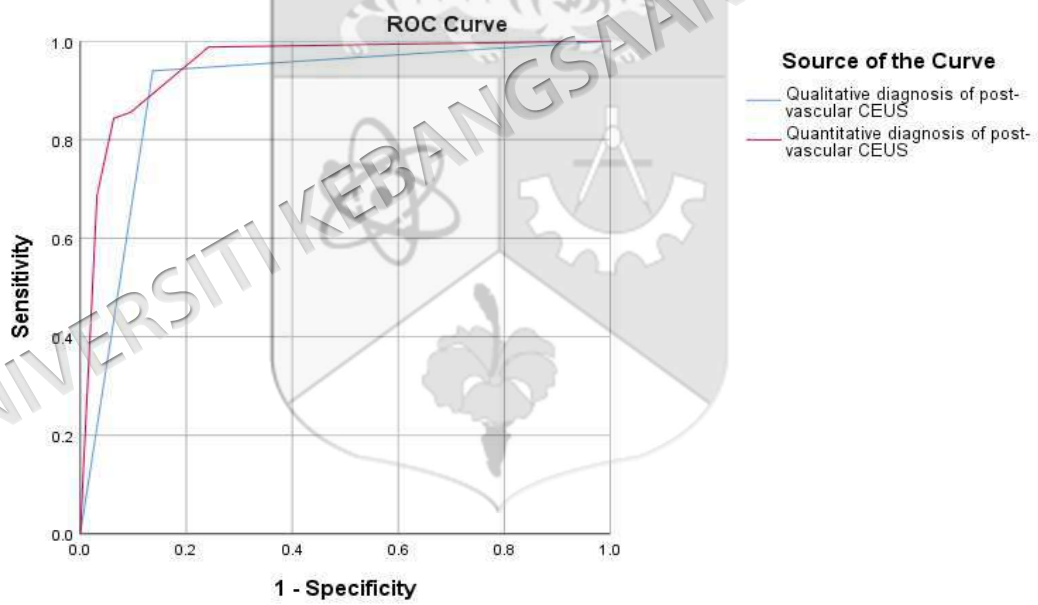


Figure 4.6 Comparison of quantitative and qualitative diagnostic efficacy of post-vascular-phase Sonazoid CEUS

4.4.4 Overall Phase Sonazoid CEUS Assessment

The arterial, venous, and post-vascular phase findings of Sonazoid CEUS were further combined to evaluate lateral cervical lymph node metastasis. In the qualitative overall phase Sonazoid CEUS assessment, the sensitivity, specificity, and Youden index were

0.988, 0.832, and 0.820, respectively. The ROC-AUC was 0.910 (95% CI: 0.862–0.957). In the quantitative overall phase Sonazoid CEUS assessment, the cut-off value was 6.5, and the sensitivity, specificity, and Youden index were 0.976, 0.947, and 0.923, respectively. The ROC-AUC was 0.978 (95% CI: 0.958–0.999); the diagnostic performance of the overall phase Sonazoid CEUS for lateral cervical lymph nodes is presented in **Table 4.27**.

The quantitative overall phase Sonazoid CEUS assessment demonstrated higher diagnostic efficacy than the qualitative overall phase Sonazoid CEUS assessment, as indicated by the higher AUC value (0.978 vs. 0.910). The AUC difference between the two methods was statistically significant according to the DeLong test (p =0.0001), and the standard error of the quantitative AUC estimate was 0.0103. This value reflects estimation precision rather than external robustness or transportability. The comparison of diagnostic efficacy between the qualitative and quantitative overall phase Sonazoid CEUS for lateral cervical lymph nodes is illustrated in **Figure 4.7**.

Table 4.27 Diagnostic performance of the overall phase Sonazoid CEUS

Method	Cut-off value	Sensitivity	Specificity	YI	AUC	95% CI	SE	P
Qualitative	NA	0.988	0.832	0.820	0.910	0.862-0.957	0.0202	0.000
Quantitative	6.5	0.976	0.947	0.923	0.978	0.958-0.999	0.0103	1

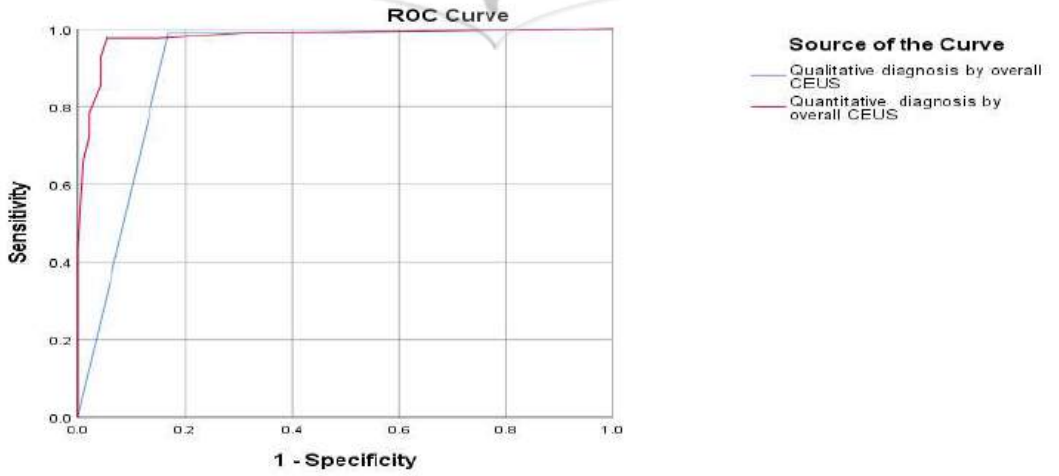


Figure 4.7 Comparison of quantitative and qualitative diagnostic efficacy of the overall phase Sonazoid CEUS

4.5 COMPARISON BETWEEN CONVENTIONAL ULTRASOUND AND OVERALL SONAZOID CEUS

4.5.1 Comparison of Qualitative Diagnostic Performance Between Conventional Ultrasound and Overall Sonazoid CEUS

The qualitative diagnostic efficacy of overall Sonazoid contrast-enhanced ultrasound (CEUS) was higher than that of conventional ultrasound for the diagnosis of lateral cervical lymph node metastasis. The area under the receiver operating characteristic curve (AUC) was 0.910 for qualitative overall Sonazoid CEUS and 0.842 for qualitative conventional ultrasound. The difference between the two AUC values was statistically significant according to the DeLong test (p = 0.0214). The comparison of qualitative diagnostic performance between the conventional ultrasound and overall Sonazoid for lateral cervical lymph nodes was presented in **Table 4.28** and illustrated in **Figure 4.8**.

Table 4.28 Qualitative diagnostic performance of conventional ultrasound and Sonazoid CEUS

Qualitative diagnoses	Sensitivity	Specificity	YI	AUC	95% CI		SE	P
Conventional ultrasound	0.924	0.661	0.585	0.842	0.780	0.892	0.0266	0.0214
Overall Sonazoid CEUS	0.988	0.832	0.820	0.910	0.858	0.947	0.0202	

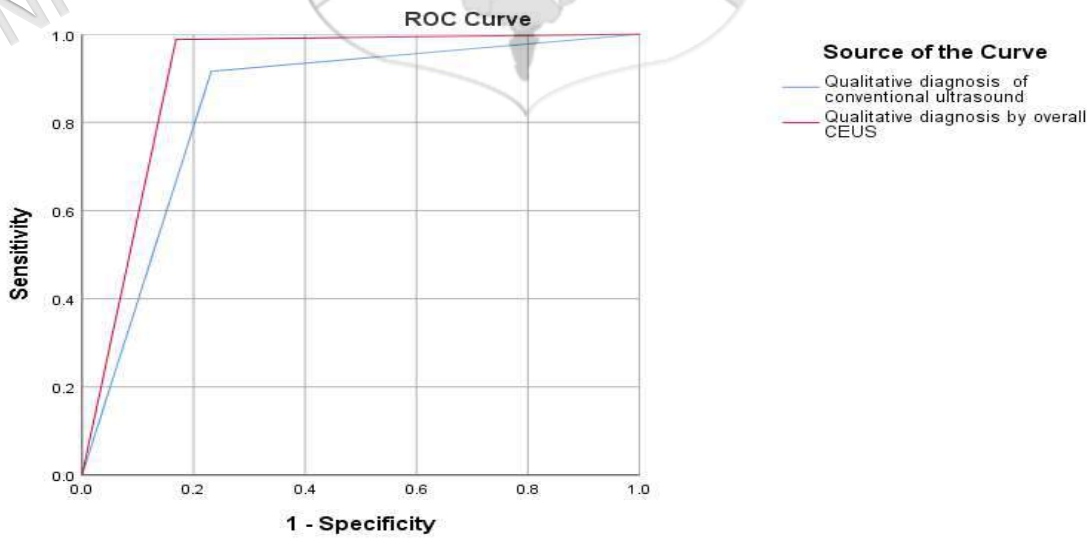


Figure 4.8 Comparison of qualitative diagnostic efficacy of conventional ultrasound and overall Sonazoid CEUS

4.5.2 Comparison of Quantitative Diagnostic Performance Between Conventional Ultrasound and Overall Sonazoid CEUS

The quantitative diagnostic efficacy of overall Sonazoid CEUS was also higher than that of conventional ultrasound for the diagnosis of lateral cervical lymph node metastasis. The AUC was 0.978 for quantitative overall Sonazoid CEUS and 0.914 for quantitative conventional ultrasound. The difference between the two AUC values was statistically significant according to the DeLong test (p = 0.0069). The comparison of quantitative diagnostic performance between the two methods for lateral cervical lymph nodes was presented in **Table 4.29** and illustrated in **Figure 4.9**.

Table 4.29 Quantitative diagnostic performance of conventional ultrasound and Sonazoid CEUS

Quantitative diagnoses	Cut-off	Sensitivity	Specificity	YI	AUC	95% CI	SE	P
Conventional ultrasound	5.5	0.880	0.968	0.848	0.914	0.863-0.951	0.0224	0.0069
Overall Sonazoid CEUS	6.5	0.976	0.947	0.923	0.978	0.945-0.994	0.0103	

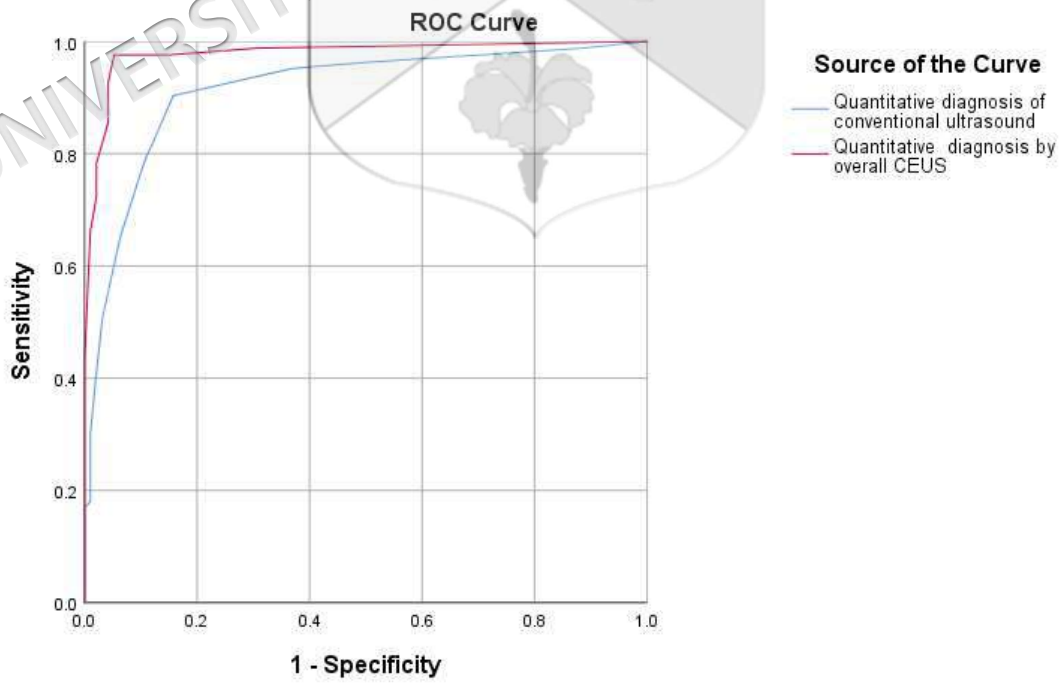


Figure 4.9 Comparison of quantitative diagnostic efficacy of conventional ultrasound and overall Sonazoid CEUS

4.5.3 Overall Comparison Between Integrated Conventional Ultrasound and Integrated Sonazoid CEUS

When the qualitative and quantitative findings were integrated, Sonazoid CEUS demonstrated higher diagnostic efficacy than conventional ultrasound for the diagnosis of lateral cervical lymph node metastasis. The integrated conventional ultrasound approach showed a sensitivity of 0.904, a specificity of 0.842, and a Youden index of 0.746. The ROC-AUC was 0.914, with a 95% confidence interval (95% CI) of 0.870–0.958.

By comparison, the integrated Sonazoid CEUS approach showed a sensitivity of 0.976, a specificity of 0.947, and a Youden index of 0.923. The ROC-AUC was 0.979, with a 95% CI of 0.959–1.000. These findings indicate that integrated Sonazoid CEUS achieved better overall diagnostic performance than integrated conventional ultrasound. The difference between the two AUC values was statistically significant according to the DeLong test (p = 0.0057). The overall comparison between the two integrated approaches is presented in **Table 4.30** and illustrated in **Figure 4.1**0.

Table 4.30 Integrated diagnostic performance of conventional ultrasound versus Sonazoid CEUS

Qualitative +Quantitative	Sensitivity	Specificity	YI	AUC	95% CI	SE	P
US	0.904	0.842	0.746	0.914	0.870-0.958	0.0101	0.0057
CEUS	0.976	0.947	0.923	0.979	0.959-1.000	0.0224	

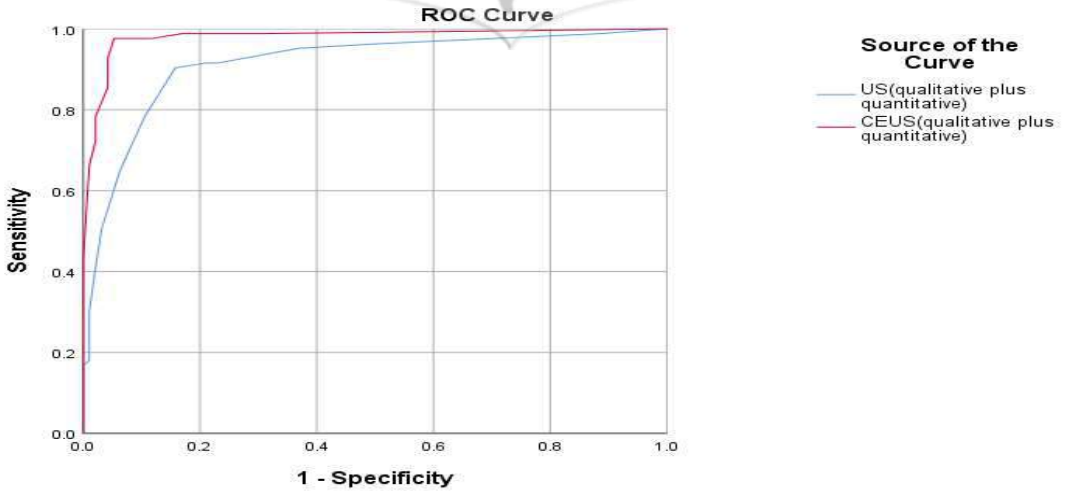


Figure 4.10 ROC comparison between integrated conventional ultrasound and integrated Sonazoid CEUS

4.6 DIAGNOSTIC PERFORMANCE OF CONVENTIONAL ULTRASOUND, OVERALL SONAZOID CEUS, AND FNA-TG USED ALONE OR IN COMBINATION

The integrated diagnostic performance of conventional ultrasound (US), overall Sonazoid CEUS, and FNA-Tg, based on both qualitative and quantitative assessments and used alone or in combination for the diagnosis of lateral cervical lymph node metastasis, is summarized in **Table 4.31**.

Among the single-modality approaches, overall Sonazoid CEUS showed the highest diagnostic efficacy, with a sensitivity of 0.976, a specificity of 0.947, a Youden index of 0.923, and an AUC of 0.979 (95% CI: 0.959–1.000). Integrated FNA-Tg showed a sensitivity of 0.904, a specificity of 0.958, a Youden index of 0.862, and an AUC of 0.959 (95% CI: 0.924–0.995), whereas conventional ultrasound showed a sensitivity of 0.904, a specificity of 0.842, a Youden index of 0.746, and an AUC of 0.914 (95% CI: 0.870–0.958).

Among the combined diagnostic approaches, US + FNA-Tg showed a sensitivity of 0.928, a specificity of 1.000, a Youden index of 0.928, and an AUC of 0.983 (95% CI: 0.961–1.000). The CEUS + FNA-Tg combination showed a sensitivity of 0.976, a specificity of 0.989, a Youden index of 0.965, and an AUC of 0.988 (95% CI: 0.966–1.000). The US+CEUS showed a sensitivity of 0.976, a specificity of 0.979, a Youden index of 0.955, and an Area Under the Curve of the Receiver Operating Characteristic (AUC) of 0.989 (95% CI: 0.976-1.000). The highest observed diagnostic performance was achieved by the combination of US + CEUS + FNA-Tg, which showed a sensitivity of 0.988, a specificity of 0.989, a Youden index of 0.977, and a ROC-AUC of 0.993 (95% CI: 0.981–1.000).

These findings show that the combined approaches had higher observed discrimination than the single-modality assessments. The US + CEUS + FNA-Tg model had the highest observed AUC; however, statistical superiority over the two-modality combinations was not demonstrated. The ROC curve comparison of the single and combined diagnostic approaches is illustrated in **Figure 4.11**.

Numerically, the combined diagnostic approaches demonstrated higher discriminatory performance than the single-modality assessments. The four highest-performing combinations were US + FNA-Tg, with an AUC of 0.983 (95% CI: 0.961–1.000); CEUS + FNA-Tg, with an AUC of 0.988 (95% CI: 0.966–1.000); US + CEUS, with an AUC of 0.989 (95% CI: 0.976–1.000); and US + CEUS + FNA-Tg, with an AUC of 0.993 (95% CI: 0.981–1.000). Although the three-modality model achieved the highest observed AUC, its 95% confidence interval substantially overlapped with those of the three two-modality models.

Pairwise DeLong tests further showed that the differences between US + CEUS + FNA-Tg and US + CEUS, CEUS + FNA-Tg, and US + FNA-Tg were not statistically significant, with p-values of 0.2849, 0.2769, and 0.2717, respectively. Therefore, the three-modality model should be described as having the highest observed AUC rather than being statistically superior to the two-modality models. The results of the selected pairwise ROC comparisons and their statistical and clinical interpretations are summarized in **Table 4.32**.

Table 4.31 Alone or combination of conventional ultrasound, CEUS, and FNA-Tg

Diagnostic approach	Sensitivity	Specificity	YI	AUC	95% CI
US	0.904	0.842	0.746	0.914	0.870-0.958
FNA-Tg	0.904	0.958	0.862	0.959	0.924-0.995
CEUS	0.976	0.947	0.923	0.979	0.959-1.000
US+FNA-Tg	0.928	1.000	0.928	0.983	0.961-1.000
CEUS+FNA-Tg	0.976	0.989	0.965	0.988	0.966-1.000
US+CEUS	0.976	0.979	0.955	0.989	0.976-1.000
US+CEUS+FNA-Tg	0.988	0.989	0.977	0.993	0.981-1.000

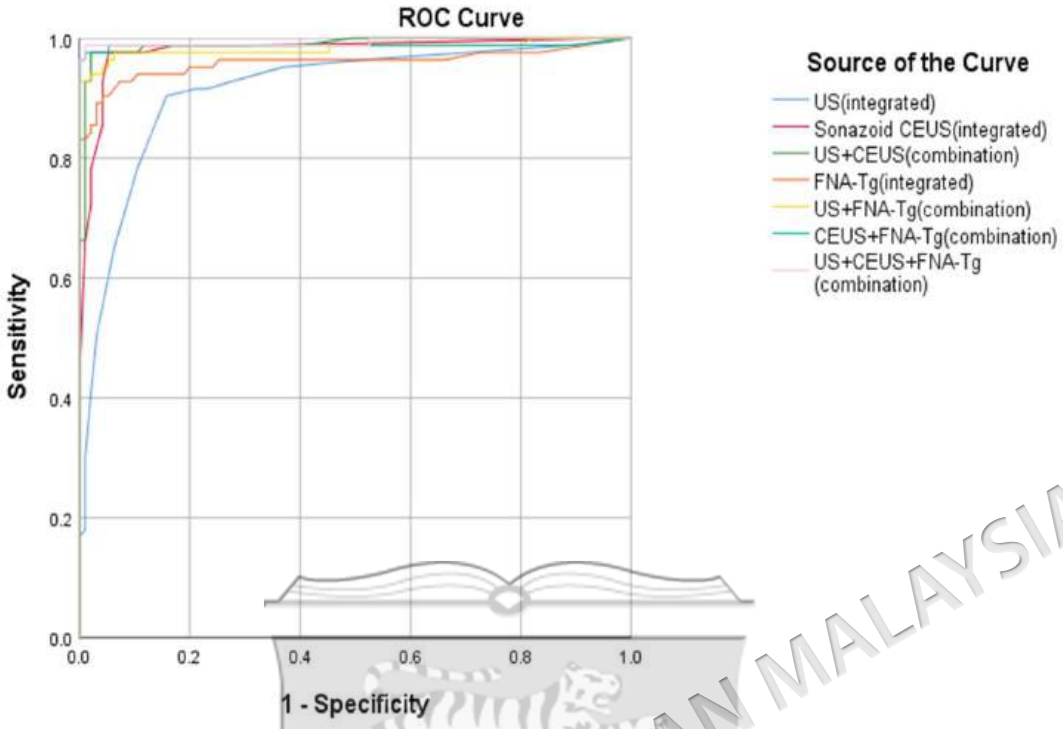


Figure 4.11 ROC comparison of single and combined diagnostic modalities using conventional ultrasound, Sonazoid CEUS, and FNA-Tg

Table 4.32 Statistical comparison and interpretation of selected ROC curves

Comparison	AUCs compared	DeLong p-value	Statistical conclusion
US + CEUS + FNA-Tg **vs** US + CEUS	0.993 vs 0.989	0.2849	Not significant
US + CEUS + FNA-Tg **vs** CEUS + FNA-Tg	0.993 vs 0.988	0.2769	Not significant
US + CEUS + FNA-Tg **vs** US + FNA-Tg	0.993 vs 0.983	0.2717	Not significant

4.7 RADIOMICS ANALYSIS OF LATERAL CERVICAL LYMPH NODES

The study cohort was randomly divided into a training cohort and a test cohort at a ratio of 8:2, comprising 142 and 36 lymph nodes, respectively. Radiomics features were extracted using an in-house analysis program implemented in Pyradiomics on the OneKey AI platform.

For each imaging modality, a total of 107 radiomics features were extracted, including 18 first-order features, 14 shape features, 24 gray-level co-occurrence matrix (GLCM) features, 16 gray-level run length matrix (GLRLM) features, 16 gray-level size zone matrix (GLSZM) features, 14 gray-level dependence matrix (GLDM) features, and 5 neighbouring gray tone difference matrix (NGTDM) features. Among all extracted features, GLCM features accounted for the largest proportion (22.4%), whereas NGTDM features accounted for the smallest proportion (4.7%).

4.7.1 Radiomics of Conventional Grayscale Ultrasound

The distributions of the extracted radiomics features and their corresponding p-values are presented in **Figure 4.12**. To assess redundancy and interrelationships among the features, correlation analysis was subsequently performed and visualized using a clustered correlation matrix, as shown in **Figure 4.13**.

Following correlation coefficient-based screening, 36 radiomics features were retained. These comprised 4 first-order features, 4 grey-level co-occurrence matrix (GLCM) features, 4 grey-level dependence matrix (GLDM) features, 9 grey-level run-length matrix (GLRLM) features, 8 grey-level size-zone matrix (GLSZM) features, 2 neighbouring grey-tone difference matrix (NGTDM) features, and 5 shape features.

To further reduce dimensionality and identify the most informative predictors, least absolute shrinkage and selection operator (LASSO) regression was applied with ten-fold cross-validation. Using the minimum-error criterion, 6 radiomics features were ultimately selected at the optimal λ value of 0.0295. The LASSO coefficient profiles and the corresponding cross-validation mean squared error are shown in **Figure 4.14** and **Figure 4.15**, respectively. Among the retained features, original_gldm_LargeDependenceHighGrayLevelEmphasis had the largest positive coefficient, indicating that it contributed most strongly to the calculated radiomics score.

Based on the 6 selected features and their corresponding LASSO coefficients, the radiomics score was calculated for each lymph node according to **Equation4.1**.

Radiomics score

$$=0.4662921348314609 + 0.039035X1 + 0.046980X_2 + 0.012973X_3 + 0.094064X_4 + 0.039005X_5 - 0.016734X_6 \quad (0.1)$$

X_1=original_firstorder_RobustMeanAbsoluteDeviation,

X_2=original_firstorder_RootMeanSquared,

X_3=original_firstorder_TotalEnergy,

X_4=original_gldm_LargeDependenceHighGrayLevelEmphasis,

X_5=original_shape_MinorAxisLength,

X_6=original_shape_SurfaceVolumeRatio.

The 6 selected radiomics features were then used to construct 11 machine-learning models based on conventional grayscale two-dimensional ultrasound images. These models included logistic regression (LR), Naive Bayes, support vector machine (SVM), k-nearest neighbour (KNN), random forest (RF), ExtraTrees, extreme gradient boosting (XGBoost), LightGBM, GradientBoosting, AdaBoost, and multilayer perceptron (MLP).

Model performance was evaluated sequentially using cross-validation and independent training- and test-cohort analyses. The cross-validation AUC values of the 11 models are presented in **Figure 4.16**, while their detailed diagnostic performance in the training and test cohorts is summarized in **Table 4.33**.The corresponding ROC curves are presented in **Figure 4.17**.

Among the grayscale-ultrasound radiomics models, ExtraTrees achieved the highest observed AUC in the test cohort, with an AUC of 0.830. However, this ranking should be interpreted descriptively because pairwise statistical comparisons among the classifiers were not available. The ExtraTrees model achieved an AUC of 0.927 in the training cohort and 0.830 in the test cohort, representing an absolute reduction of 0.097. This decline suggests a degree of performance optimism and possible overfitting, although the model retained acceptable discriminatory ability in the independent test cohort.

The training- and test-cohort ROC curves of the ExtraTrees model are shown in **Figure 4.18**. To further assess its potential clinical value, decision curve analysis was performed in the test cohort. The results indicated that the ExtraTrees radiomics model provided a greater net clinical benefit than the treat-all and treat-none strategies across the evaluated threshold-probability range, as shown in **Figure 4.19**. Finally, the sample-level predicted probabilities generated by the ExtraTrees model are presented in **Figure 4.20**.

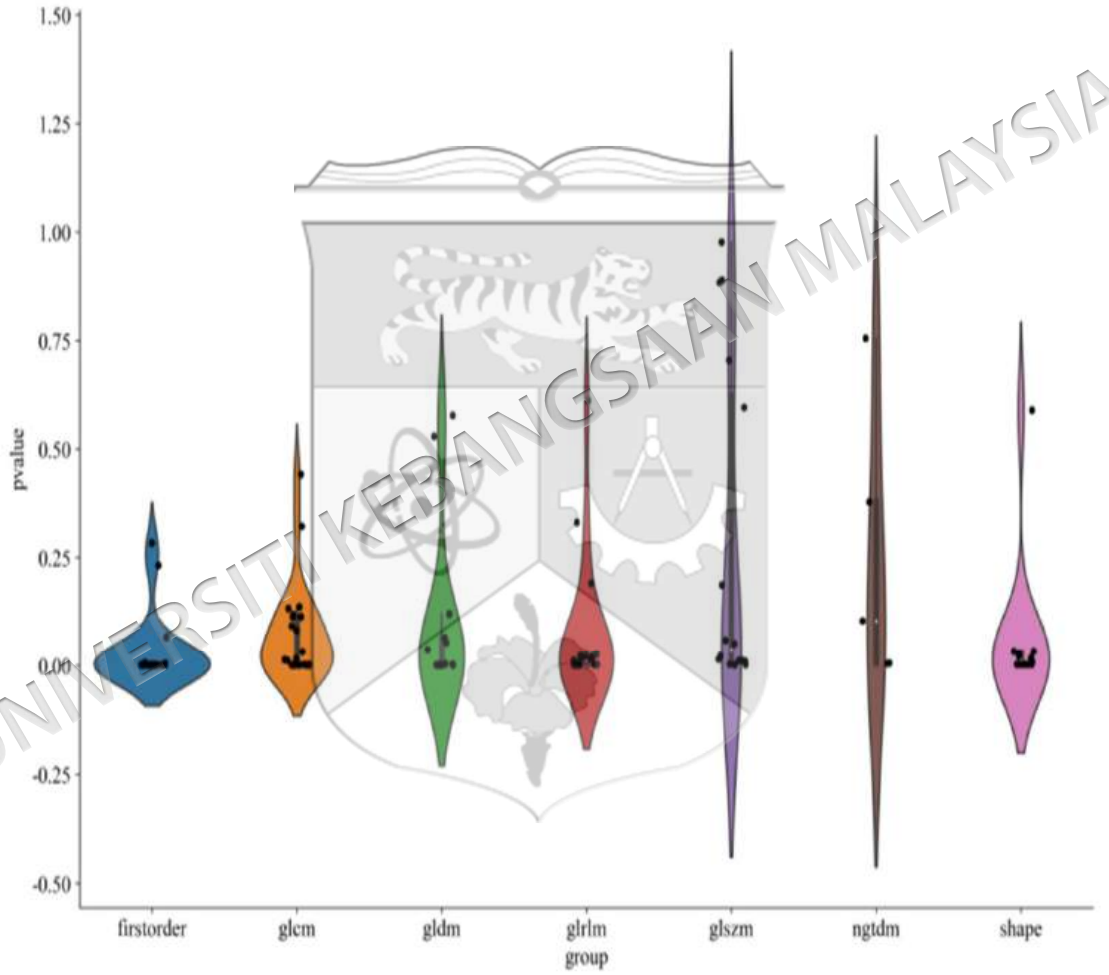


Figure 4.12 Distribution of radiomics features and corresponding p-values derived from grayscale ultrasound images

The figure illustrates the distributions of the extracted radiomics features in metastatic and non-metastatic lateral cervical lymph nodes, together with the corresponding p-values; it provides an overview of the statistical significance of individual features in differentiating metastatic from non-metastatic lateral cervical lymph nodes.

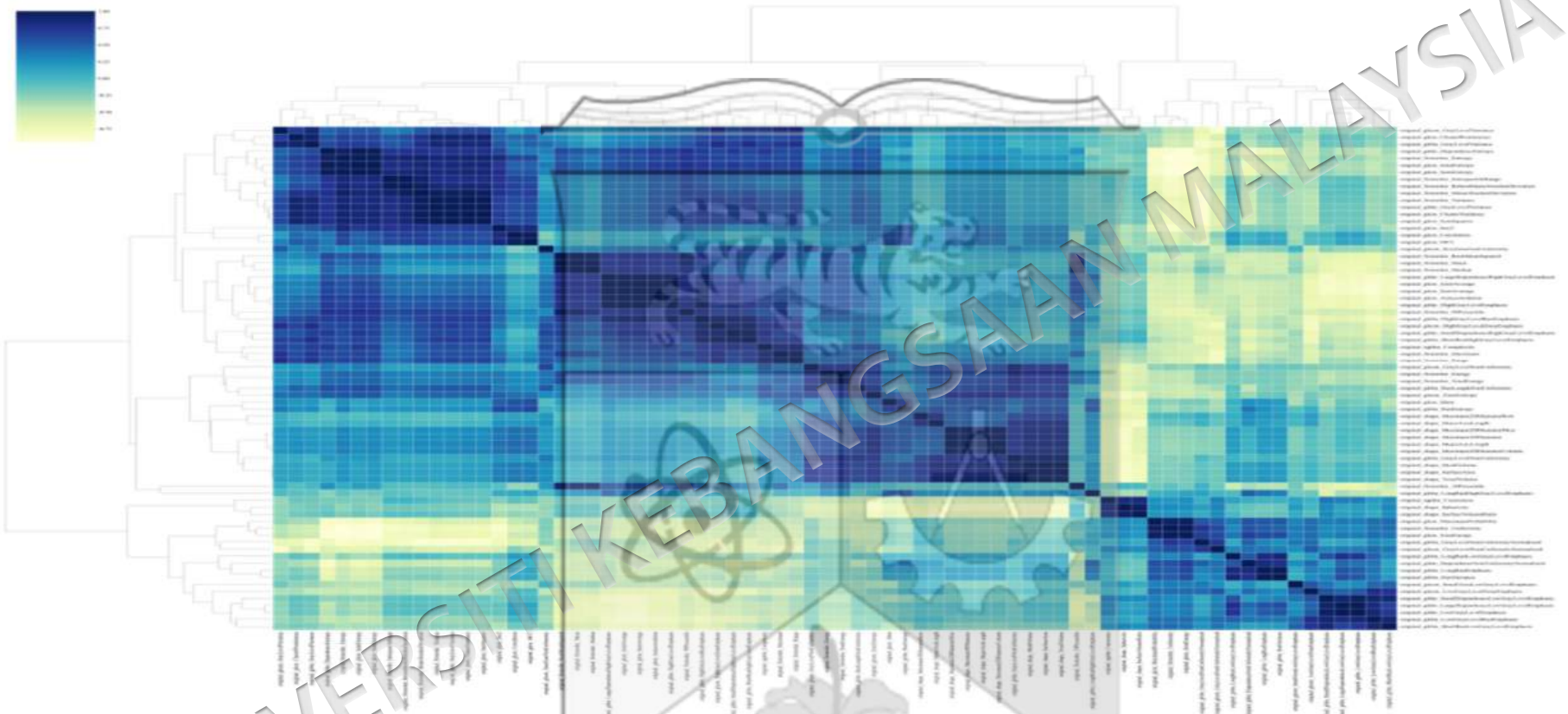


Figure 4.13 Clustered correlation matrix of radiomics features extracted from grayscale ultrasound

The figure illustrates the correlation structure among radiomics features extracted from conventional grayscale ultrasound images. Highly correlated features are grouped by hierarchical clustering, thereby facilitating the identification of redundant or highly correlated variables. Darker colors indicate stronger correlations.

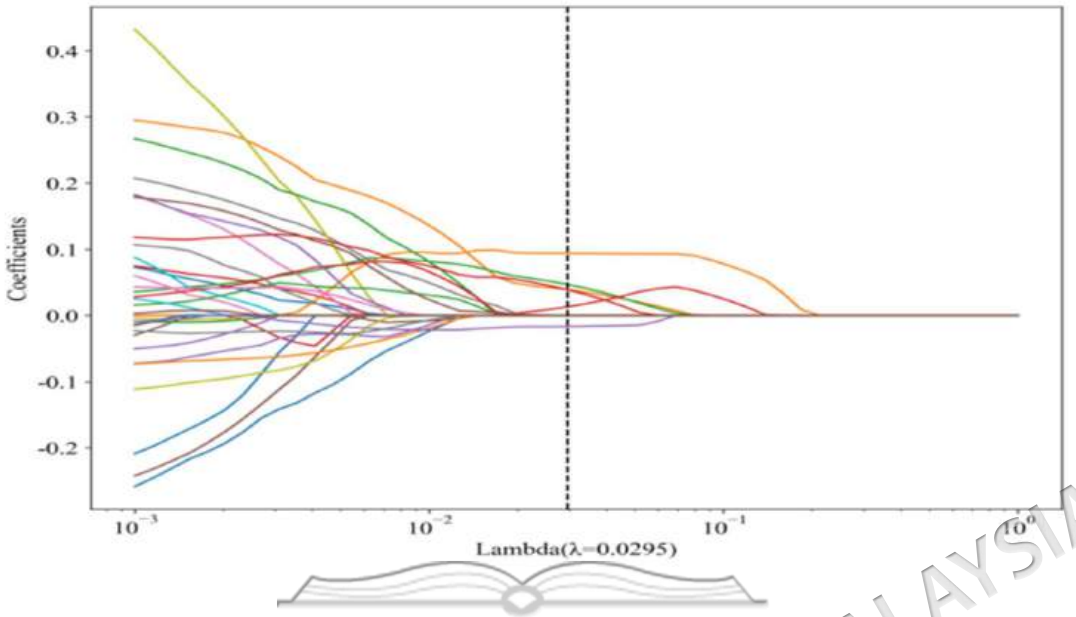


Figure 4.14 LASSO coefficient profiles of radiomics features derived from grayscale ultrasound images

The figure illustrates the coefficient trajectories of the candidate radiomics features during least absolute shrinkage and selection operator (LASSO) regression as the regularization parameter (λ) changed. As λ increased, the feature coefficients were progressively shrunk towards zero, enabling selection of the most informative features. Different colored lines represent the retained 36 radiomics features.

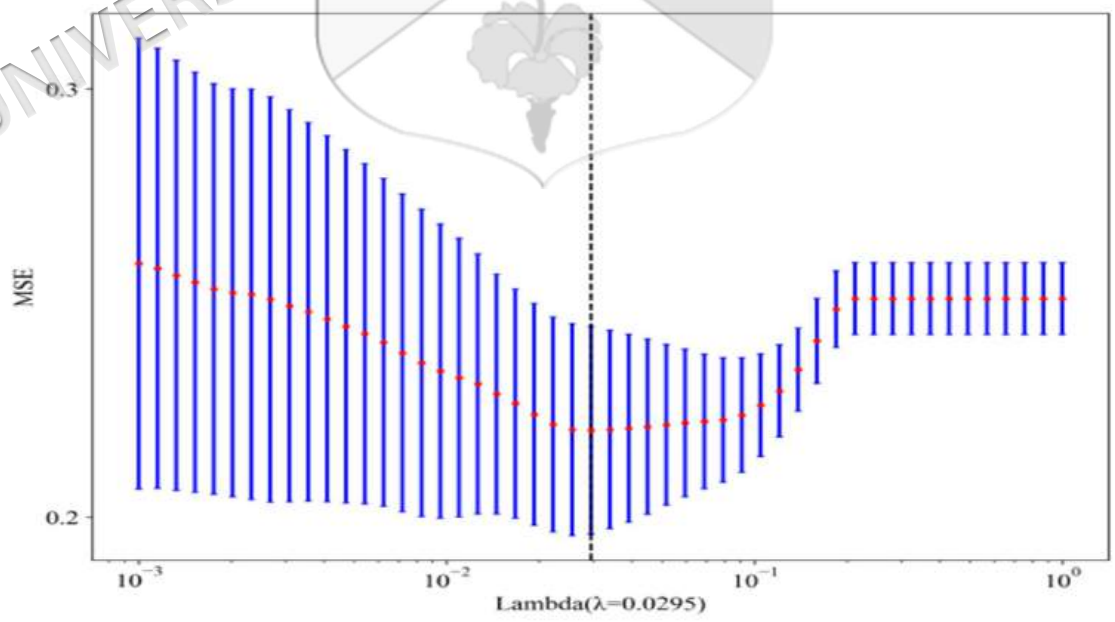


Figure 4.15 Mean squared error of LASSO regression with ten-fold cross-validation for radiomics features derived from grayscale ultrasound images

The figure illustrates the cross-validated mean squared error across different values of the regularization parameter (λ) during LASSO regression, with an optimal λ value of 0.0295. The optimal λ value was determined using ten-fold cross-validation according to the minimum-error criterion.

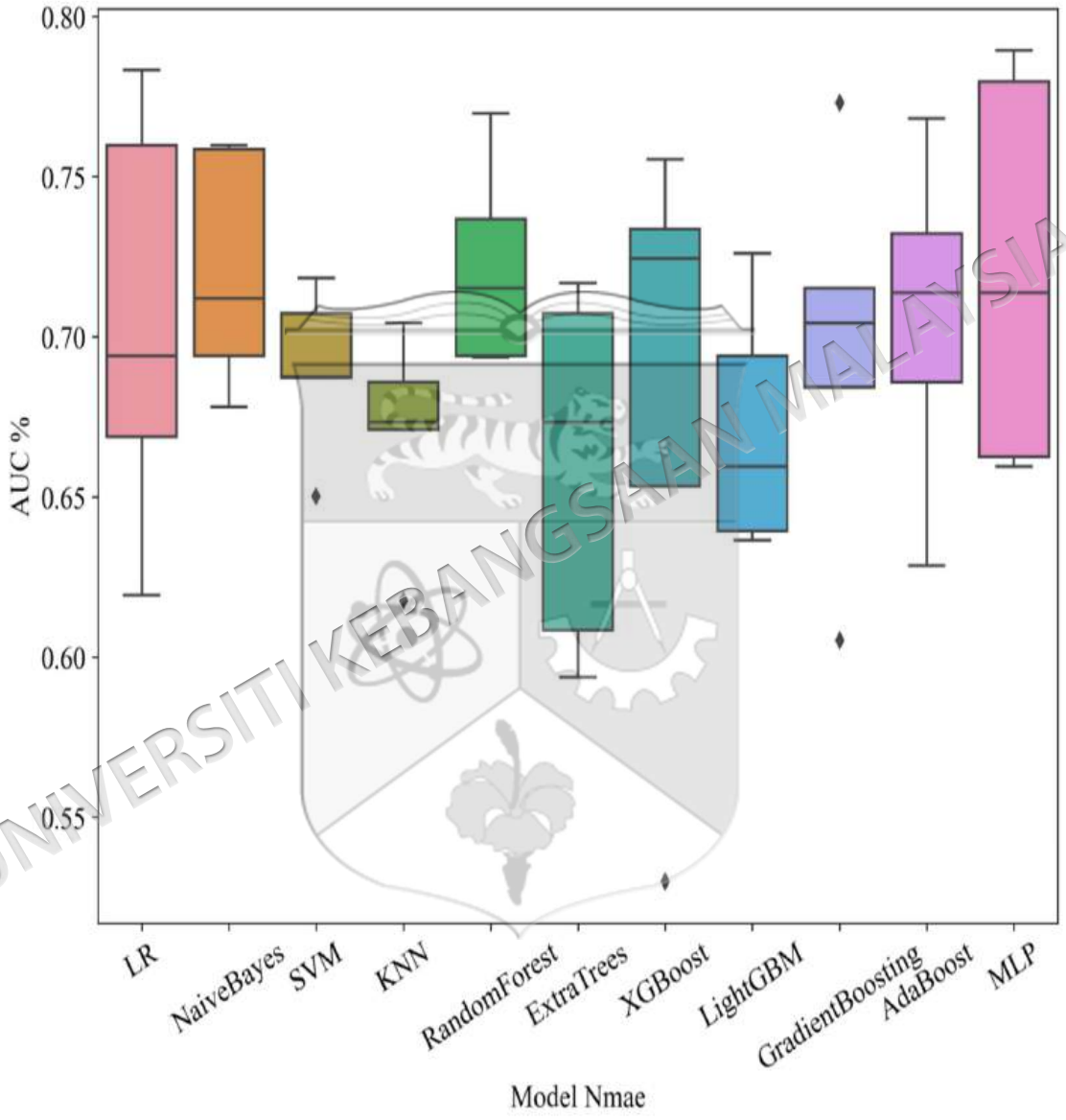


Figure 4.16 Cross-validation AUCs of machine learning models based on grayscale ultrasound radiomics

The figure compares the cross-validation area under the receiver operating characteristic curve (AUC) values of 11 machine learning models constructed using radiomics features derived from conventional grayscale ultrasound images. It provides an overall comparison of model discrimination performance during cross-validation..

Table 4.33 Performance of conventional grayscale ultrasound radiomics-based machine learning models

The table summarizes the diagnostic performance of 11 machine learning models constructed using radiomics features derived from conventional grayscale ultrasound images, including their performance in the training cohort and test cohort

model_name	Accuracy	AUC	95% CI	Sensitivity	Specificity	PPV	NPV	Precision	Recall	F1	Threshold	Task
LR	0.718	0.741	0.6584 - 0.8241	0.591	0.829	0.750	0.700	0.750	0.591	0.661	0.507	label-train
LR	0.778	0.774	0.6180 - 0.9300	0.588	0.947	0.909	0.720	0.909	0.588	0.714	0.518	label-test
NaiveBayes	0.704	0.741	0.6600 - 0.8229	0.439	0.934	0.853	0.657	0.853	0.439	0.580	0.654	label-train
NaiveBayes	0.750	0.759	0.5968 - 0.9202	0.647	0.842	0.786	0.727	0.786	0.647	0.710	0.136	label-test
SVM	0.796	0.836	0.7701 - 0.9026	0.682	0.895	0.849	0.764	0.849	0.682	0.756	0.500	label-train
SVM	0.694	0.718	0.5483 - 0.8882	0.882	0.526	0.625	0.833	0.625	0.882	0.732	0.276	label-test
KNN	0.739	0.836	0.7737 - 0.8975	0.924	0.579	0.656	0.898	0.656	0.924	0.767	0.400	label-train
KNN	0.694	0.704	0.5308 - 0.8779	0.765	0.632	0.650	0.750	0.650	0.765	0.703	0.400	label-test
RandomForest	0.880	0.953	0.9235 - 0.9820	0.848	0.908	0.889	0.873	0.889	0.848	0.868	0.417	label-train
RandomForest	0.722	0.712	0.5373 - 0.8868	0.471	0.947	0.889	0.667	0.889	0.471	0.615	0.505	label-test
ExtraTrees	0.873	0.927	0.8856 - 0.9684	0.788	0.947	0.929	0.837	0.929	0.788	0.852	0.430	label-train
ExtraTrees	0.778	0.830	0.6971 - 0.9624	0.647	0.895	0.846	0.739	0.846	0.647	0.733	0.415	label-test
XGBoost	0.859	0.943	0.9103 - 0.9767	0.758	0.947	0.926	0.818	0.926	0.758	0.833	0.478	label-train
XGBoost	0.694	0.686	0.5059 - 0.8656	0.353	1.000	1.000	0.633	1.000	0.353	0.522	0.560	label-test
LightGBM	0.817	0.896	0.8468 - 0.9451	0.742	0.882	0.845	0.798	0.845	0.742	0.790	0.515	label-train

to be continued…

…continuation

LightGBM	0.750	0.681	0.4897 - 0.8725	0.529	0.947	0.900	0.692	0.900	0.529	0.667	0.510	label-test
GradientBoosting	0.887	0.933	0.8893 - 0.9762	0.833	0.934	0.917	0.866	0.917	0.833	0.873	0.458	label-train
GradientBoosting	0.667	0.715	0.5459 - 0.8845	0.824	0.526	0.609	0.769	0.609	0.824	0.700	0.329	label-test
AdaBoost	0.824	0.896	0.8474 - 0.9455	0.803	0.842	0.815	0.831	0.815	0.803	0.809	0.502	label-train
AdaBoost	0.611	0.686	0.5109 - 0.8606	1.000	0.263	0.548	1.000	0.548	1.000	0.708	0.029	label-test
MLP	0.718	0.762	0.6823 - 0.8412	0.652	0.776	0.717	0.720	0.717	0.652	0.683	0.446	label-train
MLP	0.750	0.789	0.6424 - 0.9365	0.588	0.895	0.833	0.708	0.833	0.588	0.690	0.473	label-test

AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value; F1=2 x (precision x recall)/(precision + recall); LR, logistic regression; SVM, support vector machine; KNN, K-nearest neighbors; XGBoost, extreme gradient boosting; Light GBM, light gradient boosting machine; MLP, multi-layer perceptron.

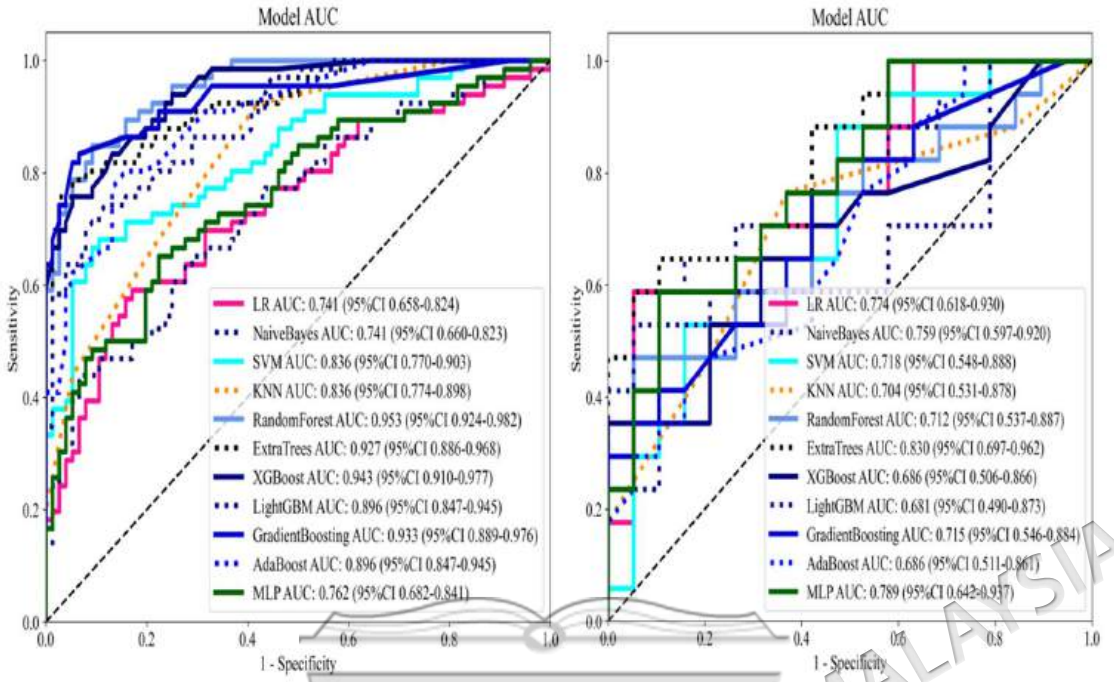


Figure 4.17 ROC curves of 11 machine learning models based on grayscale ultrasound radiomics

The figure illustrates the receiver operating characteristic (ROC) curves of the machine learning models in the training cohort and test cohort. (left) Training cohort. (right) Test cohort.

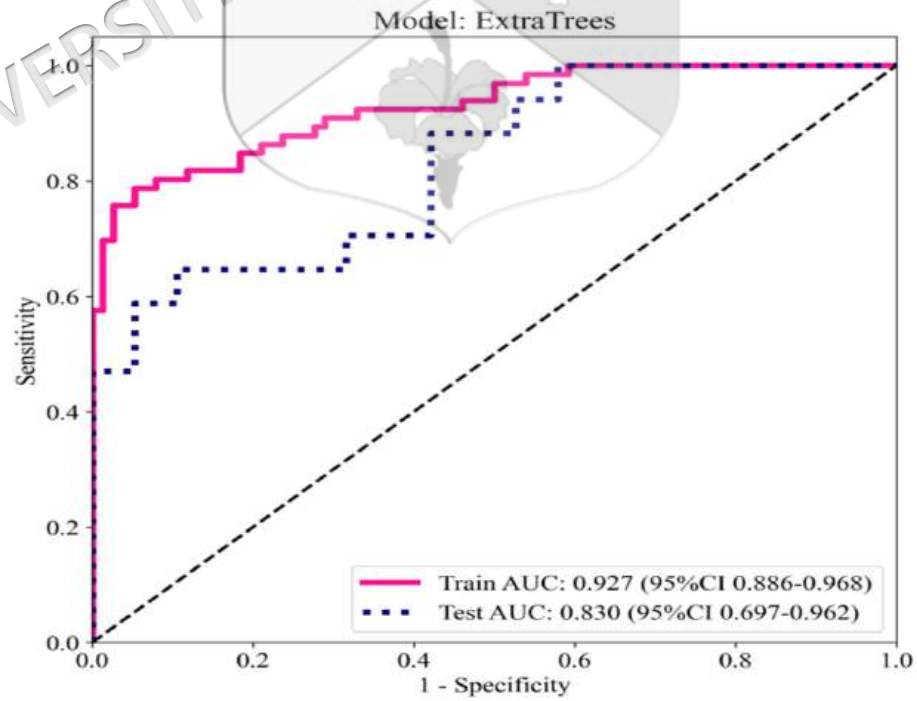


Figure 4.18 ROC curves of the ExtraTrees radiomics model based on grayscale ultrasound

The figure shows the receiver operating characteristic curves of the ExtraTrees model in the training and test cohorts. The AUC was 0.927 in the training cohort and 0.830 in the test cohort.

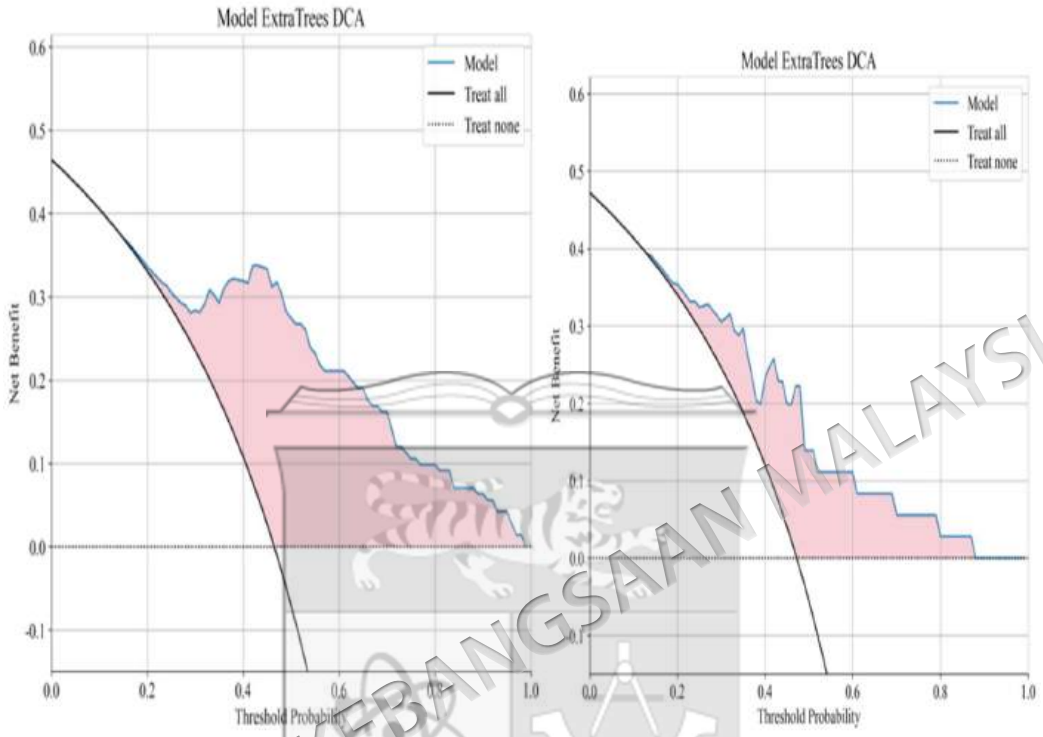


Figure 4.19 Decision curve analysis of the ExtraTrees radiomics model based on grayscale ultrasound

The left panel represents the training cohort and the right panel represents the test cohort. Net benefit is plotted against threshold probability and compared with the treat-all and treat-none strategies. The ExtraTrees model has potential clinical utility over threshold ranges in which its curve lies above both reference strategies. Interpretation should focus on clinically plausible thresholds and should not rely on isolated fluctuations at the extremes.

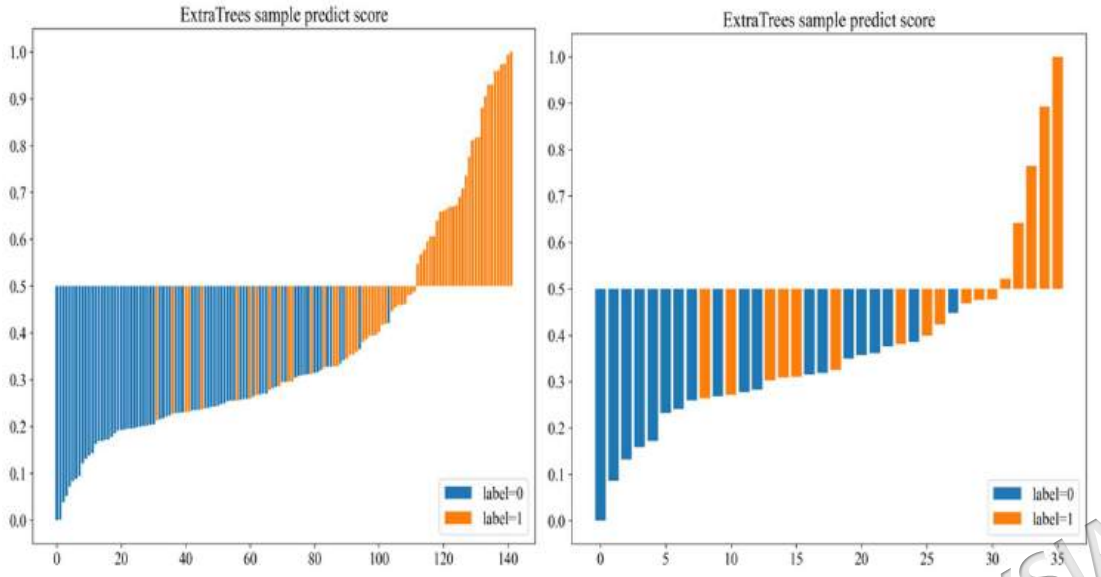


Figure 4.20 Distribution of sample-level prediction scores of the ExtraTrees radiomics model based on grayscale ultrasound

The figure shows the distribution of prediction scores generated by the ExtraTrees radiomics model for individual samples in the cohort. (left) Training cohort. (right) Test cohort.

4.7.2 Radiomics of Color Doppler Flow Imaging

The distributions of the extracted radiomics features and their corresponding p-values are presented in **Figure 4.21**. To assess the relationships and potential redundancy among these features, correlation analysis was subsequently performed and visualized using a clustered correlation matrix, as shown in **Figure 4.22**.

Following correlation coefficient-based feature screening, 31 radiomics features were retained. These comprised 5 first-order features, 3 grey-level dependence matrix (GLDM) features, 5 grey-level run-length matrix (GLRLM) features, 10 grey-level size-zone matrix (GLSZM) features, 4 neighbouring grey-tone difference matrix (NGTDM) features, and 4 shape features. To further reduce dimensionality and identify the most informative predictors, least absolute shrinkage and selection operator (LASSO) regression was performed with ten-fold cross-validation. Using the minimum-error criterion, 9 radiomics features were ultimately selected at the optimal λ value of 0.0193. The LASSO coefficient profiles and the corresponding cross-validation

mean squared error are presented in **Figure 4.23 and Figure 4.24**, respectively. Among the selected features, original_firstorder_RobustMeanAbsoluteDeviation had the largest positive coefficient, indicating the greatest contribution to the calculated radiomics score.

Based on the 9 selected features and their corresponding LASSO coefficients, the radiomics score for each lymph node was calculated according to **Equation 4.2**.

Radiomicsscore

$$=0.4662921348314604+0.087041X_1-0.047663X_2+0.058990X_3+0.065381X_4$$
$$+0.001175X_5+0.015429X_6+0.071173X_7+0.057219X_8-0.021894X_9$$

X_1=original_firstorder_RobustMeanAbsoluteDeviation,

X_2=original_firstorder_Skewness,

X_3=original_gldm_LargeDependenceHighGrayLevelEmphasis,

X_4=original_glszm_GrayLevelNonUniformityNormalized,

X_5=original_glszm_SizeZoneNonUniformityNormalized,

X_6=original_glszm_ZonePercentage,

X_7=original_ngtdm_Complexity,

X_8=original_shape_MinorAxisLength,

X_9=original_shape_SurfaceVolumeRatio. (0.2)

The 9 selected radiomics features were then used to construct and evaluate 11 machine-learning models based on color Doppler flow imaging (CDFI) images. The cross-validation AUC values of these models are presented in **Figure 4.**25, while their detailed diagnostic performance in the training and test cohorts is summarized in **Table** 4.34 The corresponding ROC curves are shown in **Figure 4.26**.

Among the evaluated CDFI radiomics models, AdaBoost achieved the highest observed AUC in the test cohort, with an AUC of 0.824. However, this ranking should be interpreted descriptively because pairwise statistical comparisons among the classifiers were not available. The AdaBoost model achieved an AUC of 0.899 in the training cohort and 0.824 in the test cohort, representing an absolute reduction of 0.075.

This relatively modest decline suggests limited performance optimism compared with models showing larger training-to-test differences, although independent external validation remains necessary to confirm the model's generalizability.

The ROC curves of the AdaBoost model in the training and test cohorts are presented in **Figure 4.27**. To further evaluate its potential clinical utility, decision curve analysis was performed. The results indicated that the AdaBoost radiomics model provided a greater net clinical benefit than the treat-all and treat-none strategies across the evaluated threshold-probability range, as shown in **Figure 4.28**. Finally, the sample-level predicted probabilities generated by the AdaBoost model are presented in **Figure 4.29**.

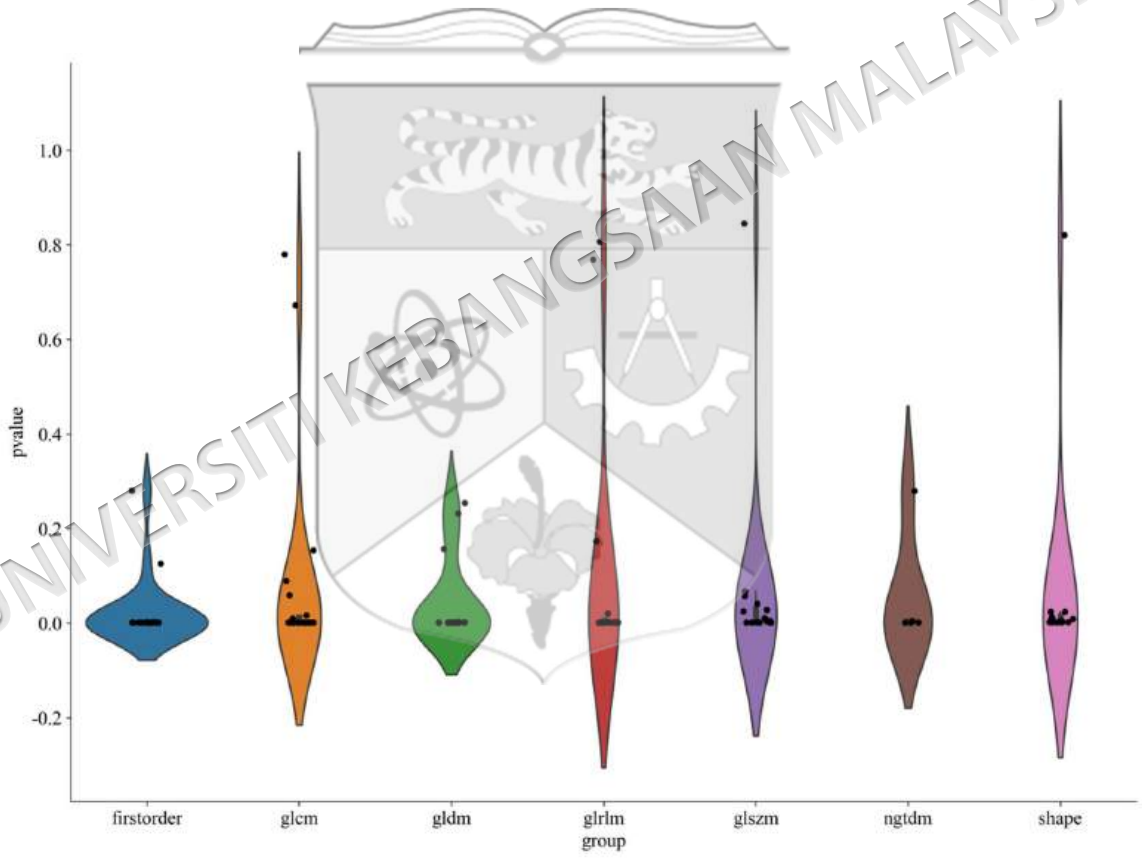


Figure 4.21 Distribution of radiomics features and corresponding p-values derived from colour Doppler flow imaging images

The figure illustrates the distributions of radiomics features extracted from colour Doppler flow imaging images in metastatic and non-metastatic lateral cervical lymph nodes, together with the corresponding p-values indicating the statistical significance of between-group differences.

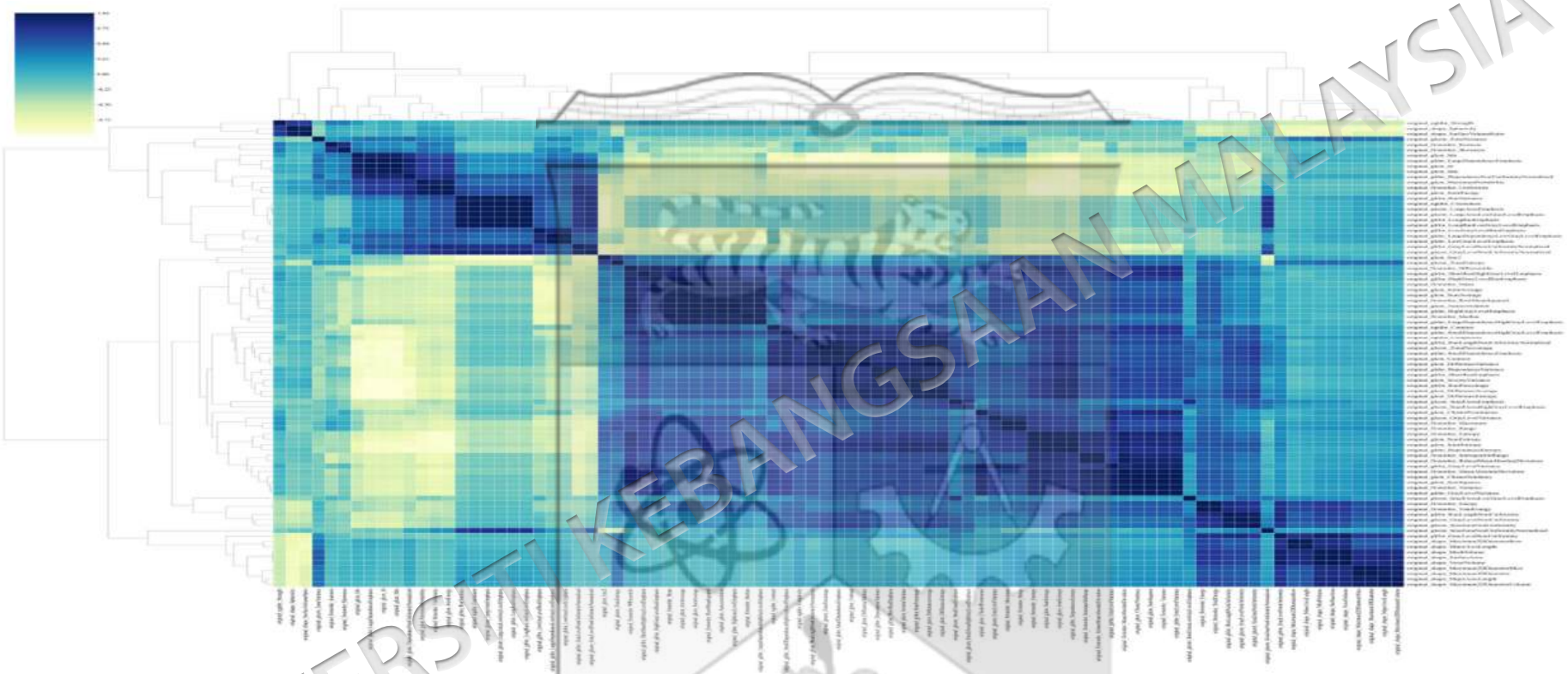


Figure 4.22 Clustered correlation matrix of radiomics features extracted from CDFI

The figure illustrates the correlation structure among radiomics features extracted from color Doppler flow imaging ultrasound images. Highly correlated features are grouped by hierarchical clustering. Darker colors indicate stronger correlations

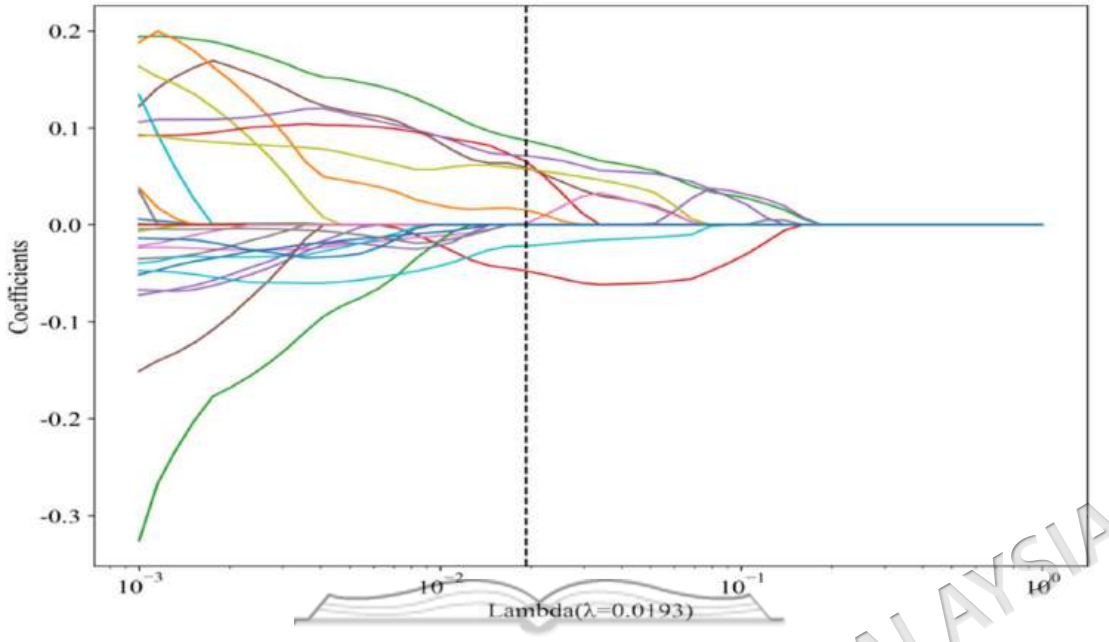


Figure 4.23 LASSO coefficient profiles of radiomics features derived from colour Doppler flow imaging images

The figure illustrates the coefficient trajectories of the candidate radiomics features during least absolute shrinkage and selection operator (LASSO) regression as the regularization parameter (λ) changed. As λ increased, the feature coefficients were progressively shrunk towards zero, enabling selection of the most informative features. Different colored lines represent the retained 31 radiomics features.

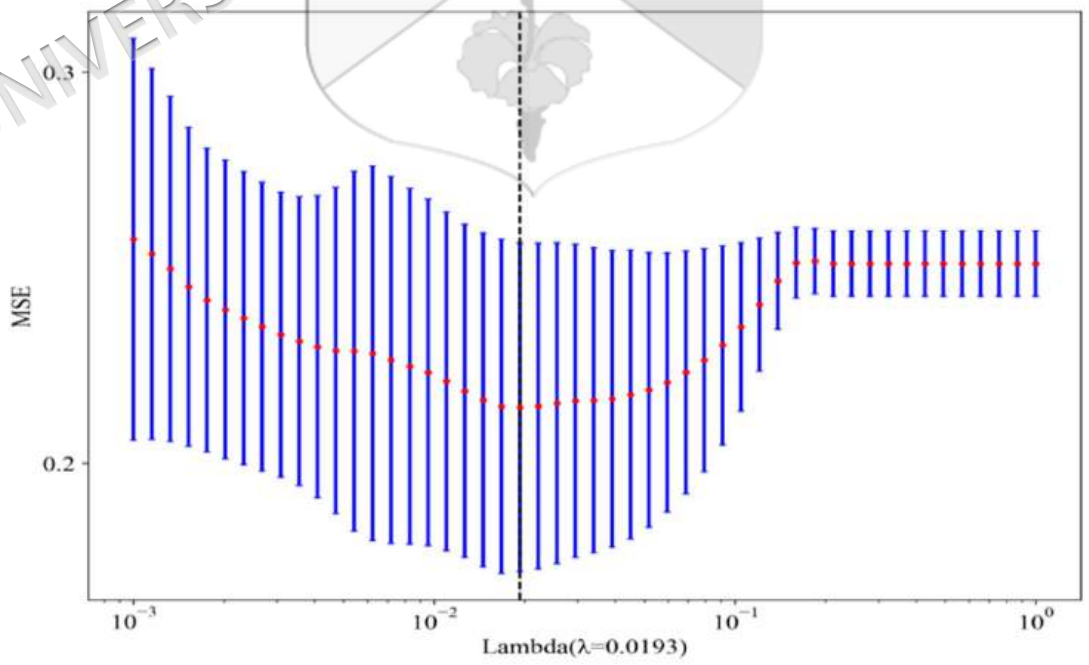


Figure 4.24 Mean squared error of LASSO regression with ten-fold cross-validation for radiomics features derived from color Doppler flow imaging images

The figure illustrates the cross-validated mean squared error across different values of the regularization parameter (λ) during LASSO regression, with an optimal λ value of 0.0193. The optimal λ value was determined using ten-fold cross-validation according to the minimum-error criterion.

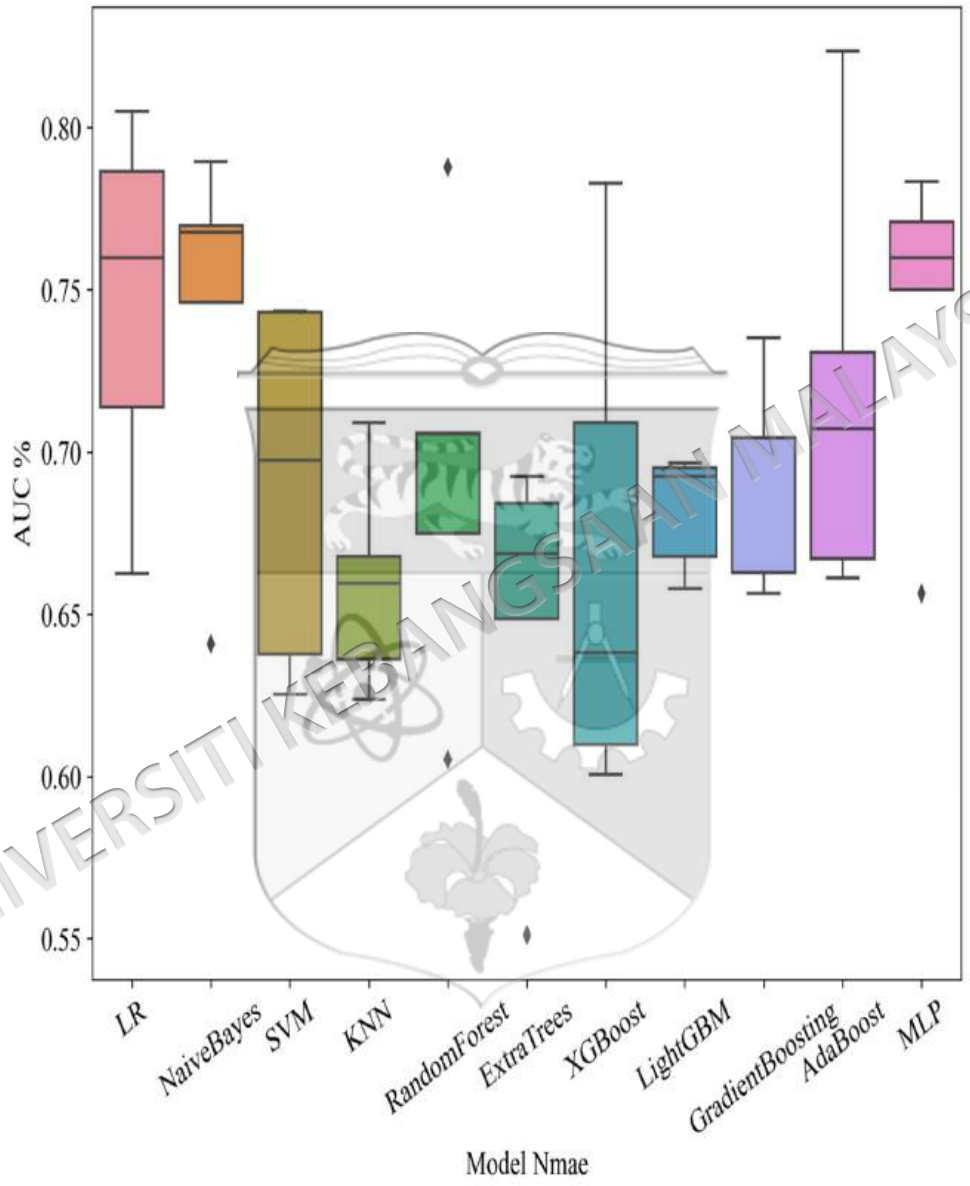


Figure 4.25 Cross-validation AUC values of machine learning models based on CDFI radiomics

The figure compares the cross-validation area under the receiver operating characteristic curve (AUC) values of 11 machine learning models constructed using radiomics features derived from color Doppler flow imaging ultrasound images. It provides an overall comparison of model discrimination performance during cross-validation.

Table 4.34 Performance of CDFI radiomics-based machine learning models

The table summarizes the diagnostic performance of 11 machine learning models constructed using radiomics features derived from color Doppler flow imaging ultrasound images, including their performance in the training cohort and test cohort

model_name	Accuracy	AUC	95% CI	Sensitivity	Specificity	PPV	NPV	Precision	Recall	F1	Threshold	Task
LR	0.746	0.801	0.7290 - 0.8735	0.712	0.776	0.734	0.756	0.734	0.712	0.723	0.463	label-train
LR	0.750	0.789	0.6416 - 0.9373	0.882	0.632	0.682	0.857	0.682	0.882	0.769	0.263	label-test
NaiveBayes	0.725	0.754	0.6743 - 0.8336	0.515	0.908	0.829	0.683	0.829	0.515	0.636	0.449	label-train
NaiveBayes	0.667	0.746	0.5861 - 0.9062	1.000	0.368	0.586	1.000	0.586	1.000	0.739	0.010	label-test
SVM	0.796	0.867	0.8077 - 0.9259	0.667	0.908	0.863	0.758	0.863	0.667	0.752	0.572	label-train
SVM	0.722	0.638	0.4379 - 0.8376	0.588	0.842	0.769	0.696	0.769	0.588	0.667	0.409	label-test
KNN	0.768	0.859	0.8015 - 0.9164	0.879	0.671	0.699	0.864	0.699	0.879	0.779	0.400	label-train
KNN	0.611	0.636	0.4566 - 0.8158	0.412	0.789	0.636	0.600	0.636	0.412	0.500	0.600	label-test
RandomForest	0.915	0.970	0.9482 - 0.9924	0.909	0.921	0.909	0.921	0.909	0.909	0.909	0.453	label-train
RandomForest	0.667	0.703	0.5305 - 0.8751	0.412	0.895	0.778	0.630	0.778	0.412	0.538	0.554	label-test
ExtraTrees	0.880	0.954	0.9240 - 0.9847	0.909	0.855	0.845	0.915	0.845	0.909	0.876	0.403	label-train
ExtraTrees	0.722	0.759	0.5996 - 0.9174	0.941	0.526	0.640	0.909	0.640	0.941	0.762	0.326	label-test
XGBoost	0.915	0.950	0.9111 - 0.9886	0.909	0.921	0.909	0.921	0.909	0.909	0.909	0.408	label-train
XGBoost	0.694	0.692	0.5142 - 0.8697	0.588	0.789	0.714	0.682	0.714	0.588	0.645	0.406	label-test

to be continued…

… continuation

LightGBM	0.873	0.926	0.8862 - 0.9667	0.848	0.895	0.875	0.872	0.875	0.848	0.862	0.506	label-train
LightGBM	0.694	0.731	0.5632 - 0.8981	0.824	0.579	0.636	0.786	0.636	0.824	0.718	0.222	label-test
GradientBoosting	0.915	0.958	0.9224 - 0.9926	0.909	0.921	0.909	0.921	0.909	0.909	0.909	0.447	label-train
GradientBoosting	0.667	0.704	0.5323 - 0.8764	0.882	0.474	0.600	0.818	0.600	0.882	0.714	0.296	label-test
AdaBoost	0.803	0.899	0.8514 - 0.9465	0.818	0.789	0.771	0.833	0.771	0.818	0.794	0.485	label-train
AdaBoost	0.778	0.824	0.6881 - 0.9590	0.647	0.895	0.846	0.739	0.846	0.647	0.733	0.495	label-test
MLP	0.732	0.804	0.7322 - 0.8767	0.773	0.697	0.689	0.779	0.689	0.773	0.729	0.422	label-train
MLP	0.722	0.783	0.6350 - 0.9315	1.000	0.474	0.630	1.000	0.630	1.000	0.773	0.337	label-test

AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value; F1=2 x (precision x recall)/(precision + recall); LR, logistic regression; SVM, support vector machine; KNN, K-nearest neighbors; XGBoost, extreme gradient boosting; Light GBM, light gradient boosting machine; MLP, multi-layer perceptron.

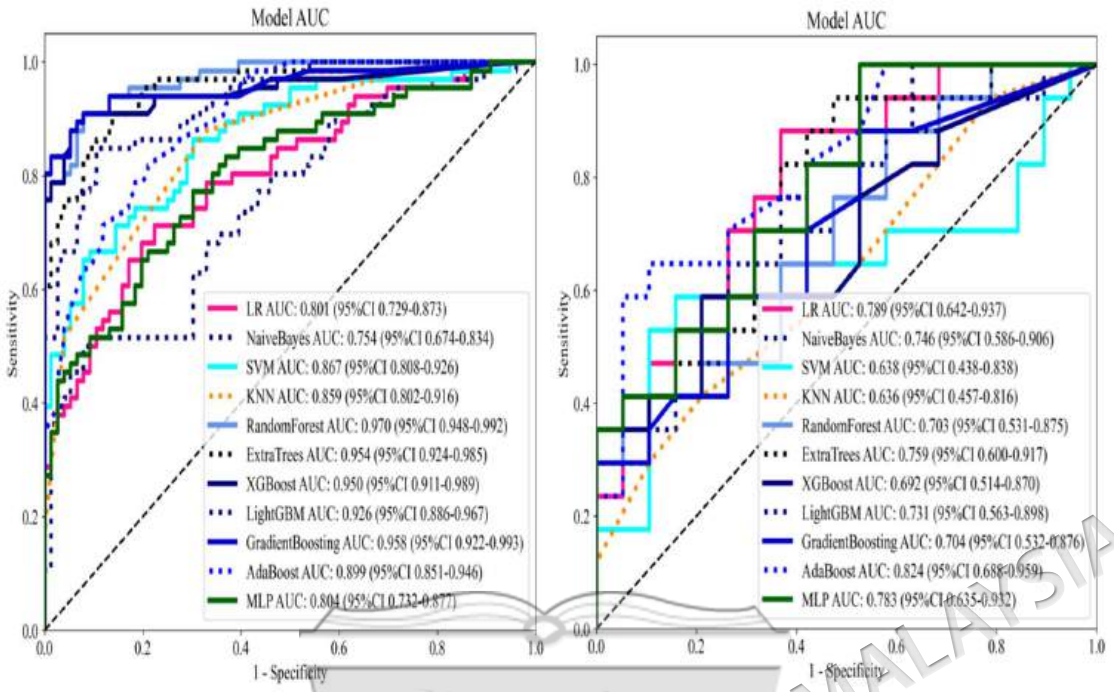


Figure 4.26 ROC curves of 11 machine-learning models based on CDFI radiomics

The figure illustrates the receiver operating characteristic (ROC) curves of the 11 machine-learning models in the training and test cohorts. (Left) Training cohort. (Right) Test cohort.

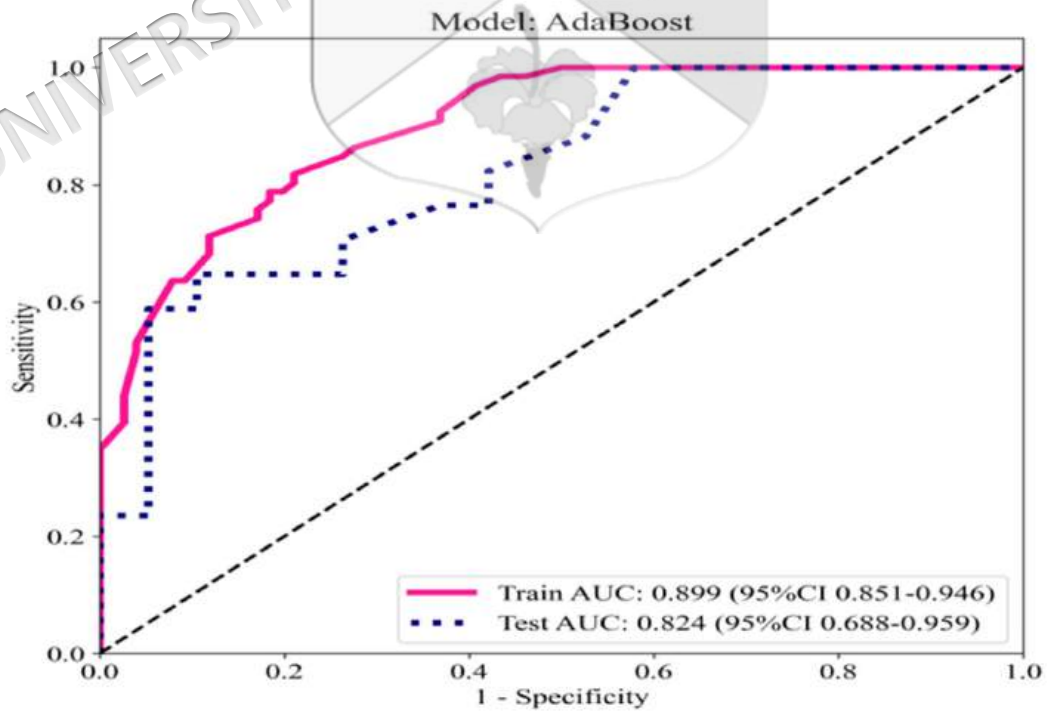


Figure 4.27 ROC curves of the AdaBoost radiomics model based on CDFI

The figure shows the receiver operating characteristic curves of the AdaBoost model in the training and test cohorts. The AUC was 0.899 in the training cohort and 0.824 in the test cohort.

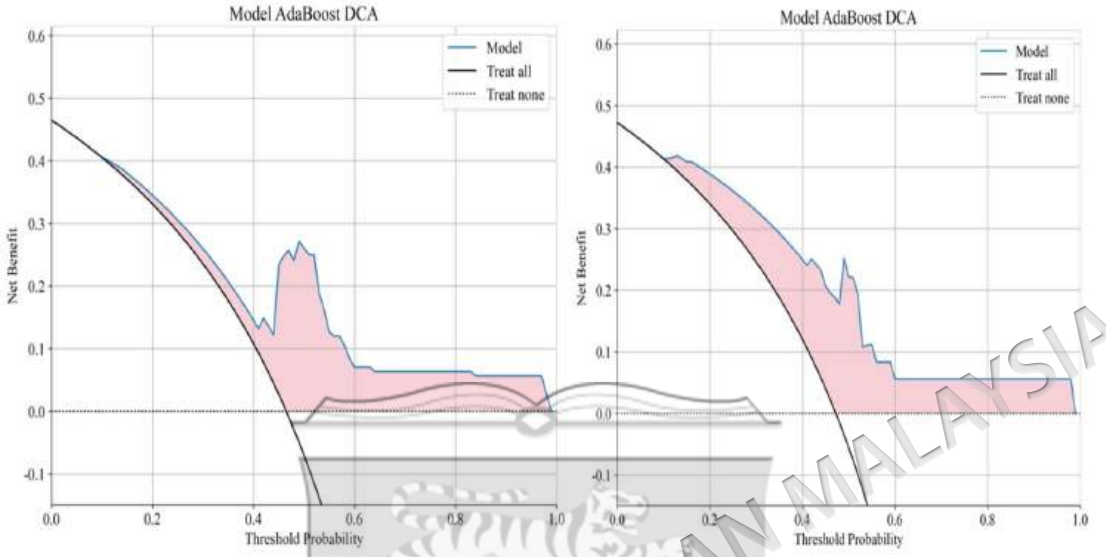


Figure 4.28 Decision curve analysis of the AdaBoost radiomics model based on CDFI

The left panel represents the training cohort and the right panel represents the test cohort. The curves show the net benefit of the CDFI AdaBoost model across threshold probabilities relative to the treat-all and treat-none strategies. Clinical utility is supported only within ranges where the model curve remains above both reference lines; curve crossing and instability at high thresholds should be interpreted cautiously.

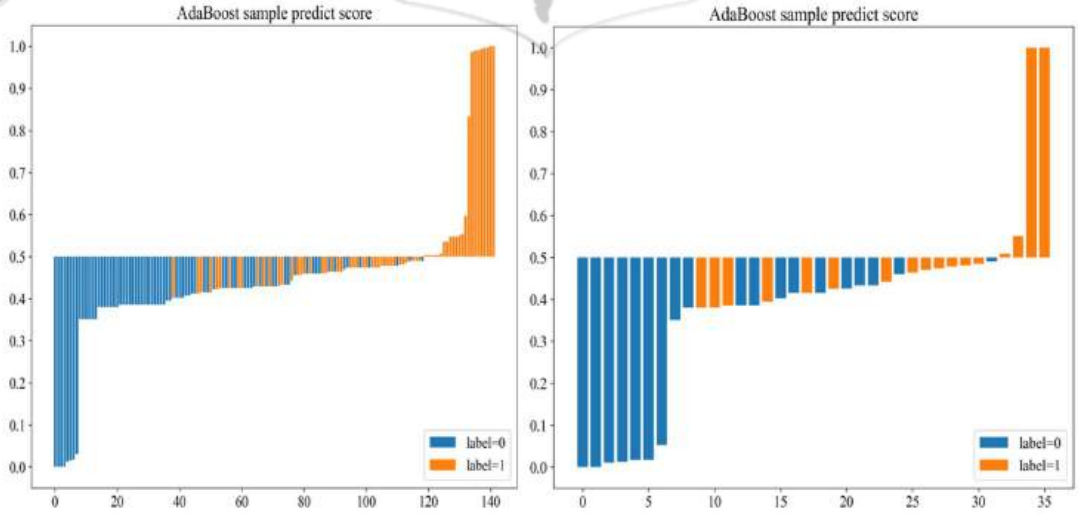


Figure 4.29 Distribution of sample-level prediction scores of the AdaBoost radiomics model based on CDFI

The figure shows the distribution of prediction scores generated by the AdaBoost radiomics model for individual samples in the training and test cohorts. (left) Training cohort. (right) Test cohort.

4.7.3 Radiomics of Arterial-Phase Sonazoid Contrast-Enhanced Ultrasound

The distributions of the extracted radiomics features and their corresponding p-values are presented in **Figure 4.30**. To evaluate inter-feature relationships and identify potential redundancy, correlation analysis was subsequently performed and visualized using a clustered correlation matrix, as shown in **Figure 4.31.**

Following correlation coefficient-based feature screening, 15 radiomics features were retained. These comprised 4 first-order features, 2 grey-level co-occurrence matrix (GLCM) features, 4 grey-level run-length matrix (GLRLM) features, 4 grey-level size-zone matrix (GLSZM) features, and 1 neighbouring grey-tone difference matrix (NGTDM) feature.

To further reduce dimensionality and identify the most informative predictors, least absolute shrinkage and selection operator (LASSO) regression was performed with ten-fold cross-validation.7 radiomics features were ultimately selected at the optimal λ value of 0.0146. The LASSO coefficient profiles and the corresponding cross-validation mean squared error are presented in **Figure 4.32** and **Figure 4.33**, respectively. Among the selected features, original_glszm_GrayLevelNonUniformityNormalized had the largest positive coefficient, indicating the greatest contribution to the calculated radiomics score.

Based on the 7 selected features and their corresponding LASSO coefficients, the radiomics score for each lymph node was calculated according to **Equation 4.3**.

Radiomics score

$$=0.4662921348314607+0.057806X_1+0.008531X_2+0.073431X_3+0.036814X_4$$
$$+0.018299X_5 + 0.106895X_6 + 0.044394X_7$$

X_1=original_firstorder_Maximum,

X_2=original_firstorder_TotalEnergy,

X_3=original_glrlm_GrayLevelVariance,

X_4=original_glrlm_RunEntropy,

X_5=original_glrlm_ShortRunHighGrayLevelEmphasis,

X_6=original_glszm_GrayLevelNonUniformityNormalized,

X_7=original_ngtdm_Complexity. (0.3)

The 7 selected radiomics features were then used to construct and evaluate 11 machine-learning models based on arterial-phase contrast-enhanced ultrasound images. The cross-validation AUC values of these models are presented in **Figure 4.34**, while their detailed diagnostic performance in the training and test cohorts is summarized in **Table 4.35**. The corresponding ROC curves are shown in **Figure 4.35**.

Among the evaluated arterial-phase CEUS radiomics models, Naive Bayes achieved the highest observed AUC in the test cohort, with an AUC of 0.822. However, this ranking should be interpreted descriptively because pairwise statistical comparisons among the classifiers were not available. The model achieved an AUC of 0.664 in the training cohort and 0.822 in the test cohort. The higher test-cohort AUC should not be interpreted as evidence of superior generalizability, as this pattern may reflect sampling variability associated with the relatively small test cohort and differences in sample distribution.

The ROC curves of the Naive Bayes model in the training and test cohorts are presented in **Figure 4.36**. To further assess its potential clinical utility, decision curve analysis was performed. The results indicated that the Naive Bayes radiomics model provided a greater net clinical benefit than the treat-all and treat-none strategies across the evaluated threshold-probability range, as shown in **Figure 4.37**. Finally, the sample-level predicted probabilities generated by the model are presented in **Figure 4.38**.

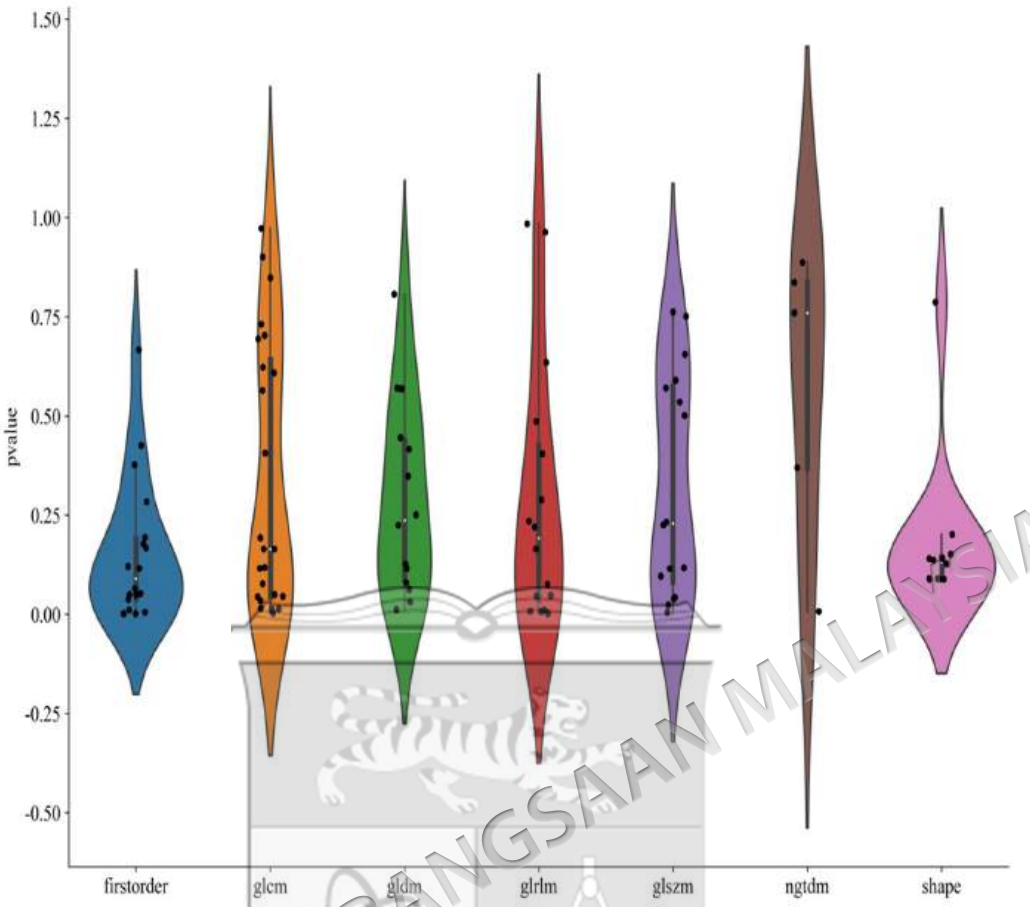


Figure 4.30 Radiomics features distribution and corresponding p-values for arterial-phase Sonazoid CEUS

This figure presents the distribution of radiomics features extracted from arterial-phase contrast-enhanced ultrasound images together with their corresponding p-values. The results provide an overview of the statistical significance of individual features in differentiating metastatic from non-metastatic lateral cervical lymph nodes.

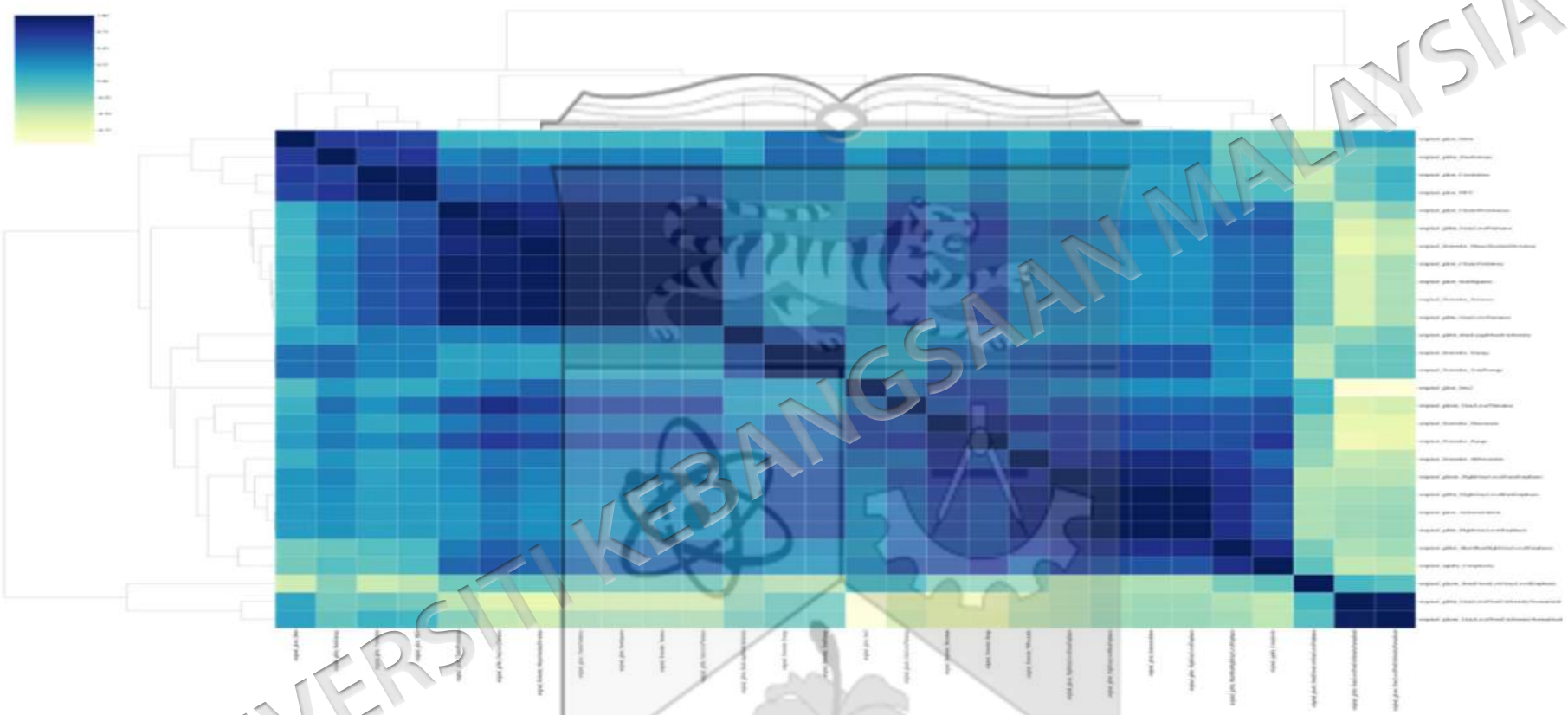


Figure 4.31 Clustered correlation matrix of radiomics features extracted from arterial-phase Sonazoid CEUS

This figure illustrates the correlation structure among radiomics features extracted from arterial-phase contrast-enhanced ultrasound images. Features with similar correlation patterns are grouped by hierarchical clustering, thereby facilitating the identification of redundant or highly correlated variables before feature selection. Darker colors indicate stronger correlations.

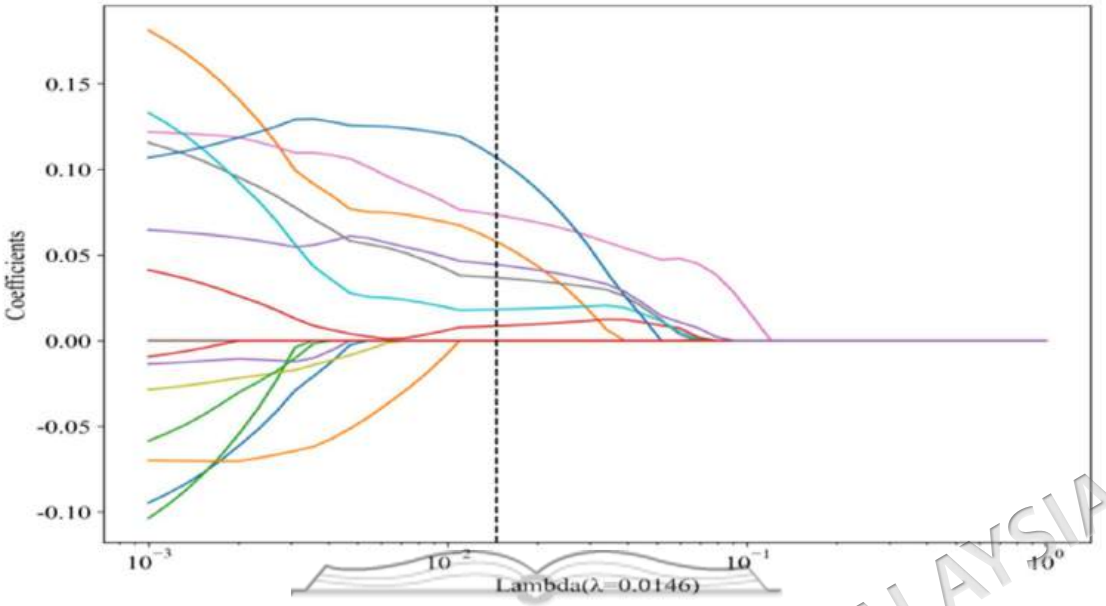


Figure 4.32 LASSO coefficient profiles of radiomics features from arterial-phase Sonazoid CEUS

This figure shows the coefficient trajectories of radiomics features during least absolute shrinkage and selection operator (LASSO) regression as the regularization parameter varied. Changes in coefficient magnitude reflect the shrinkage process used to identify the most informative radiomics features. Different colored lines represent the retained 15 radiomics features.

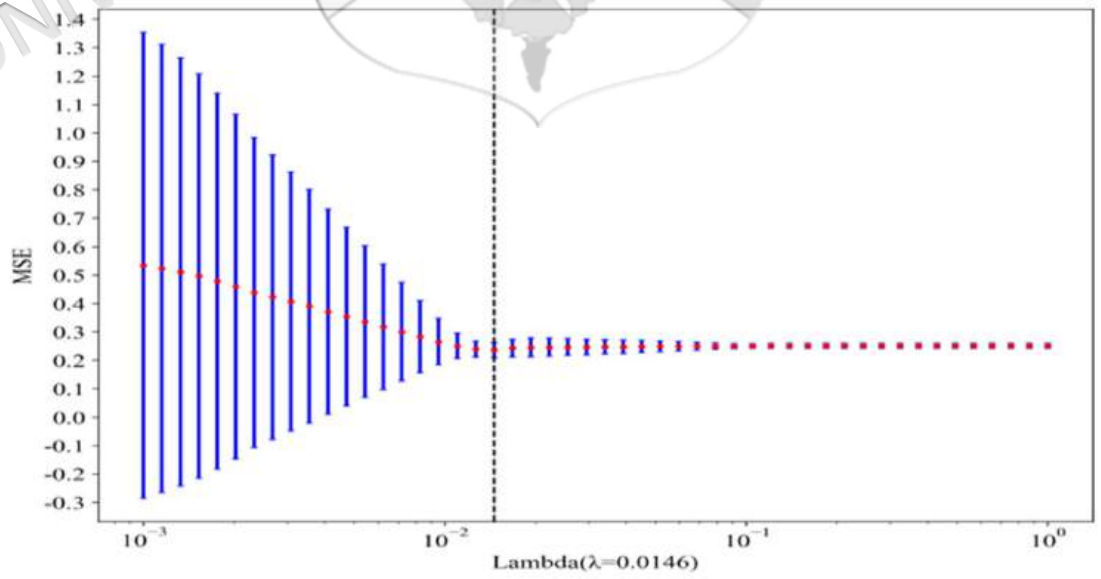


Figure 4.33 Mean squared error of LASSO regression with ten-fold cross-validation for arterial-phase Sonazoid CEUS

The figure shows the mean squared error across different values of the regularization parameter in LASSO regression, with an optimal λ value of 0.0146. The optimal λ value was selected using ten-fold cross-validation.

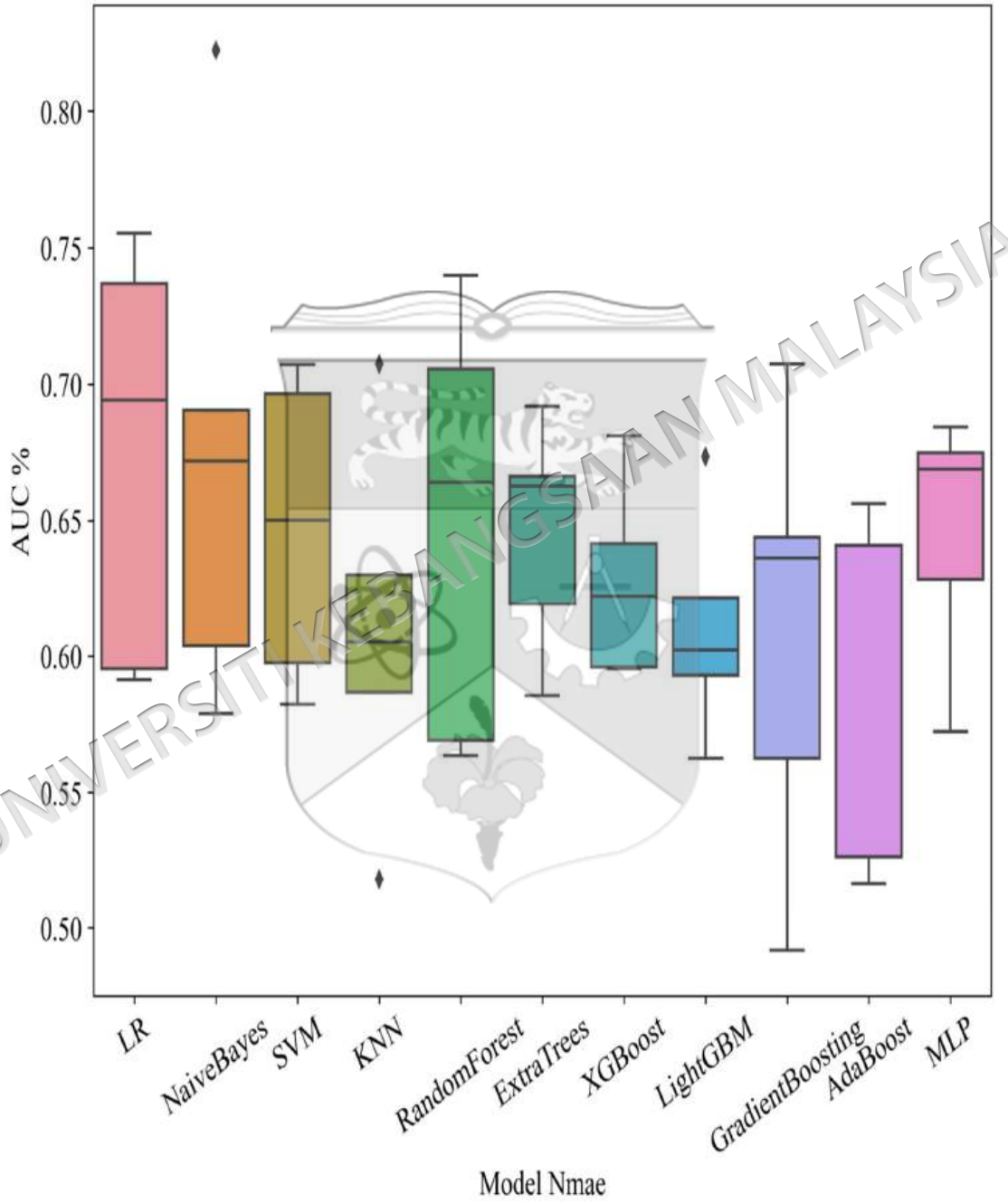


Figure 4.34 Cross-validation AUC values of machine learning models based on arterial-phase Sonazoid CEUS radiomics

This figure compares the cross-validation area under the receiver operating characteristic curve (AUC) values of 11 machine learning models constructed using radiomics features derived from arterial-phase contrast-enhanced ultrasound images. It provides an overall comparison of model discrimination performance during cross-validation.

Table 4.35 Performance of arterial-phase Sonazoid CEUS radiomics-based machine learning models

This table summarizes the diagnostic performance of 11 machine learning models constructed using radiomics features derived from arterial-phase contrast-enhanced ultrasound images. The results include model performance in both the training and test cohorts and provide a basis for identifying the best-performing model.

model_name	Accuracy	AUC	95% CI	Sensitivity	Specificity	PPV	NPV	Precision	Recall	F1	Threshold	Task
LR	0.699	0.719	0.6349 - 0.8034	0.746	0.658	0.658	0.746	0.658	0.746	0.699	0.444	label-train
LR	0.771	0.720	0.5370 - 0.9038	0.500	1.000	1.000	0.704	1.000	0.500	0.667	0.587	label-test
NaiveBayes	0.622	0.664	0.5751 - 0.7521	0.687	0.566	0.582	0.672	0.582	0.687	0.630	0.078	label-train
NaiveBayes	0.800	0.822	0.6807 - 0.9641	0.562	1.000	1.000	0.731	1.000	0.562	0.720	0.391	label-test
SVM	0.762	0.786	0.7101 - 0.8626	0.836	0.697	0.709	0.828	0.709	0.836	0.767	0.436	label-train
SVM	0.743	0.707	0.5231 - 0.8913	0.500	0.947	0.889	0.692	0.889	0.500	0.640	0.562	label-test
KNN	0.762	0.844	0.7840 - 0.9039	0.746	0.776	0.746	0.776	0.746	0.746	0.746	0.600	label-train
KNN	0.657	0.605	0.4134 - 0.7971	0.312	0.947	0.833	0.621	0.833	0.312	0.455	0.800	label-test
RandomForest	0.867	0.955	0.9266 - 0.9827	0.955	0.789	0.800	0.952	0.800	0.955	0.871	0.446	label-train
RandomForest	0.743	0.711	0.5256 - 0.8954	0.437	1.000	1.000	0.679	1.000	0.437	0.609	0.643	label-test
ExtraTrees	0.825	0.886	0.8337 - 0.9381	0.821	0.829	0.809	0.840	0.809	0.821	0.815	0.473	label-train
ExtraTrees	0.714	0.704	0.5161 - 0.8918	0.562	0.842	0.750	0.696	0.750	0.562	0.643	0.490	label-test
XGBoost	0.874	0.939	0.9037 - 0.9742	0.851	0.895	0.877	0.872	0.877	0.851	0.864	0.542	label-train
XGBoost	0.686	0.607	0.4009 - 0.8129	0.625	0.737	0.667	0.700	0.667	0.625	0.645	0.536	label-test

to be continued…

… continuation

LightGBM	0.839	0.882	0.8275 - 0.9369	0.910	0.776	0.782	0.908	0.782	0.910	0.841	0.460	label-train
LightGBM	0.714	0.656	0.4574 - 0.8551	0.500	0.895	0.800	0.680	0.800	0.500	0.615	0.582	label-test
GradientBoosting	0.853	0.937	0.9007 - 0.9725	0.866	0.842	0.829	0.877	0.829	0.866	0.847	0.539	label-train
GradientBoosting	0.629	0.562	0.3559 - 0.7691	0.562	0.684	0.600	0.650	0.600	0.562	0.581	0.539	label-test
AdaBoost	0.804	0.888	0.8357 - 0.9398	0.970	0.658	0.714	0.962	0.714	0.970	0.823	0.483	label-train
AdaBoost	0.686	0.656	0.4593 - 0.8532	0.687	0.684	0.647	0.722	0.647	0.687	0.667	0.498	label-test
MLP	0.706	0.713	0.6282 - 0.7980	0.716	0.697	0.676	0.736	0.676	0.716	0.696	0.471	label-train
MLP	0.743	0.684	0.4892 - 0.8792	0.437	1.000	1.000	0.679	1.000	0.437	0.609	0.558	label-test

AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value; F1=2 x (precision x recall)/ (precision + recall); LR, logistic regression; SVM, support vector machine; KNN, K-nearest neighbors; XG Boost, extreme gradient boosting; Light GBM, light gradient boosting machine; MLP, multi-layer perceptron

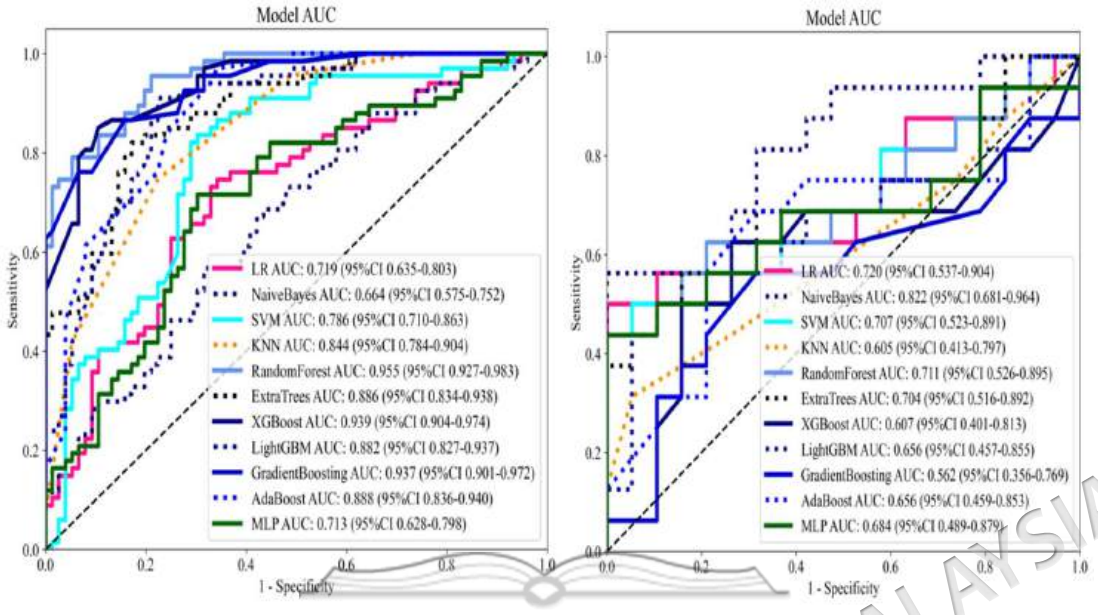


Figure 4.35 ROC curves of 11 machine learning models based on arterial-phase Sonazoid CEUS

The figure illustrates the receiver operating characteristic (ROC) curves of the 11 machine learning models in the training cohort and test cohort. (left) Training cohort. (right) Test cohort.

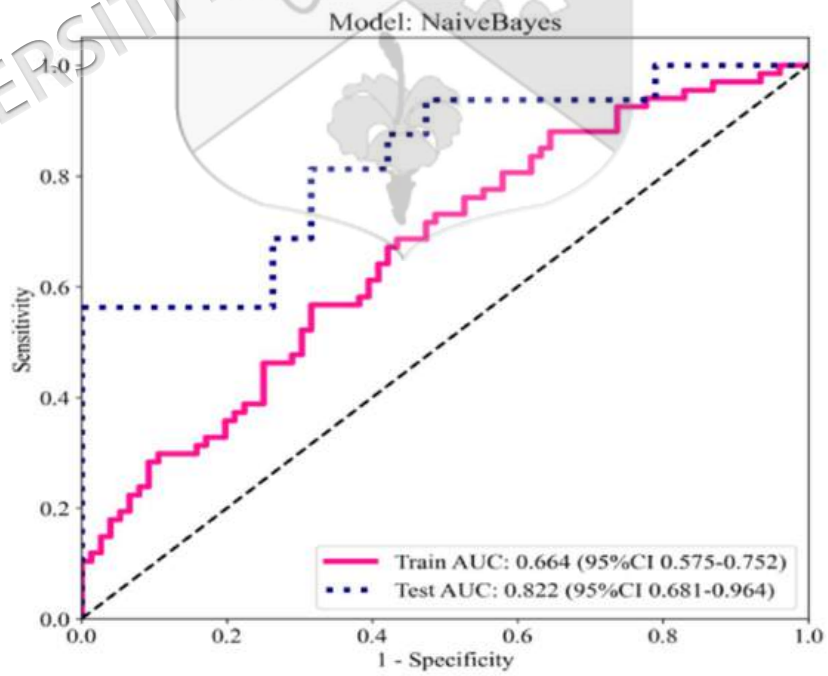


Figure 4.36 ROC curves of the NaiveBayes radiomics model based on arterial-phase Sonazoid CEUS

This figure shows that the AUC of the NaiveBayes radiomics model in the training and test cohorts was 0.664 and 0.822.

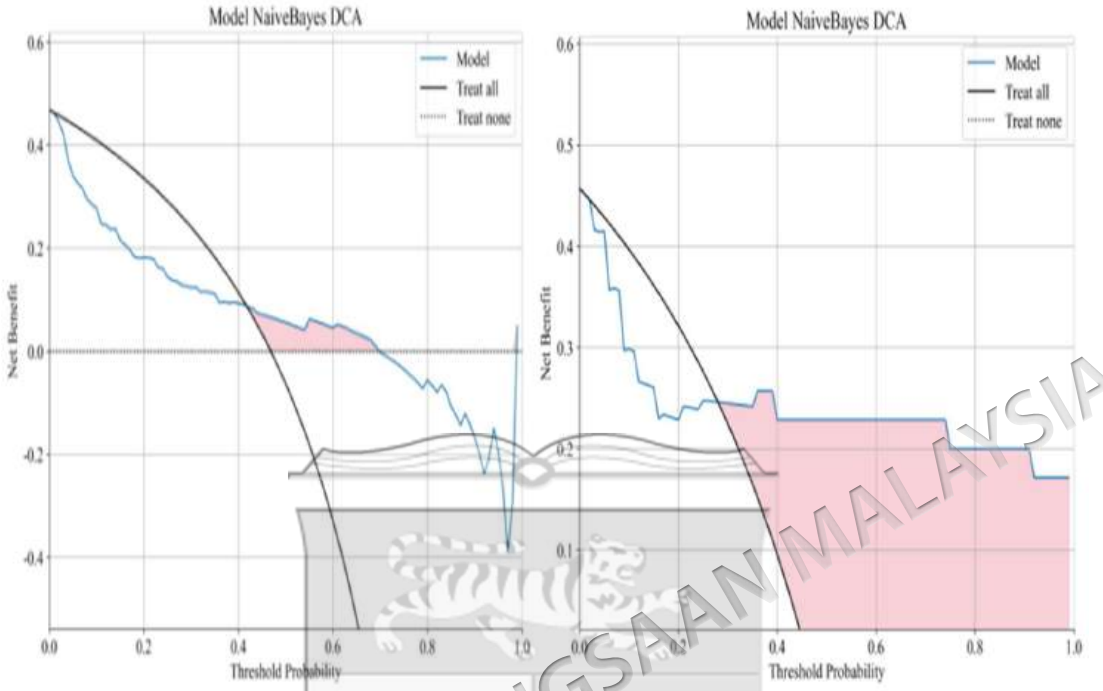


Figure 4.37 Decision curve analysis of the Naive Bayes radiomics model based on arterial-phase Sonazoid CEUS

The left panel represents the training cohort and the right panel represents the test cohort. The curves compare the net benefit of the arterial-phase CEUS Naive Bayes model with the treat-all and treat-none strategies across threshold probabilities. A positive and higher net benefit than both reference strategies indicate potential clinical usefulness within that threshold range. Differences between cohorts may reflect sampling variability, particularly in the small test cohort.

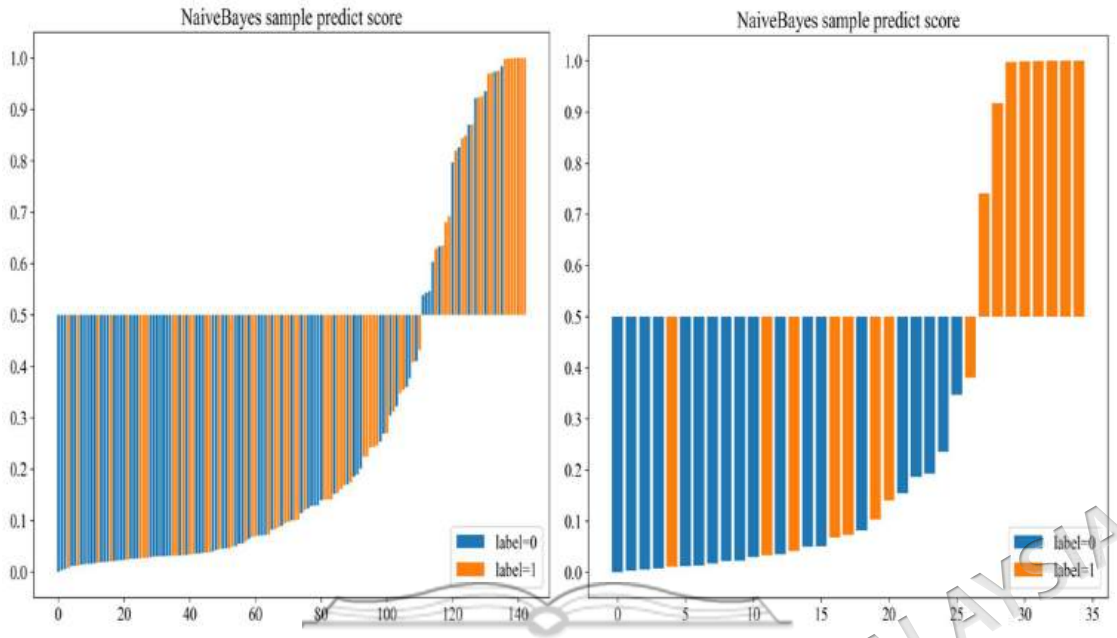


Figure 4.38 Distribution of sample-level prediction scores of the NaiveBayes radiomics model based on arterial-phase Sonazoid CEUS

The figure shows the distribution of prediction scores generated by the NaiveBayes radiomics model for individual samples in the cohort. (left) Training cohort. (right) Test cohort.

4.7.4 Radiomics of Venous-Phase Sonazoid Contrast-Enhanced Ultrasound

The distributions of the extracted radiomics features and their corresponding p-values are presented in **Figure 4.39**. To assess inter-feature relationships and identify potential redundancy, correlation analysis was subsequently performed and visualized using a clustered correlation matrix, as shown in **Figure 4.40**.

Following correlation coefficient-based feature screening, 23 radiomics features were retained. These comprised 3 first-order features, 3 grey-level co-occurrence matrix (GLCM) features, 1 grey-level dependence matrix (GLDM) feature, 4 grey-level run-length matrix (GLRLM) features, 7 grey-level size-zone matrix (GLSZM) features, 1 neighbouring grey-tone difference matrix (NGTDM) feature, and 4 shape features.

To further reduce dimensionality and identify the most informative predictors, least absolute shrinkage and selection operator (LASSO) regression was performed with ten-fold cross-validation. Using the minimum-error criterion, 5 radiomics features were ultimately selected, with an optimal λ value of 0.0518. The LASSO coefficient profiles and the corresponding cross-validation mean squared error are presented in **Figure 4.41 and Figure 4.42**, respectively. Among the selected features, original_glrlm_GrayLevelVariance had the largest positive coefficient, indicating the greatest contribution to the calculated radiomics score.

Based on the 5 selected features and their corresponding LASSO coefficients, the radiomics score for each lymph node was calculated according to **Equation 4.4**.

Radiomics score

$$=0.4662921348314607+0.019927X_1+0.067493X_2+0.001606X_3+0.015258X_4+0.021372X_5$$

X_1=original_glcm_ClusterProminence,

X_2=original_glrlm_GrayLevelVariance,

X_3=original_glszm_SizeZoneNonUniformity,

X_4=original_shape_Maximum2DDiameterColumn,

X_5=original_shape_VoxelVolume. (0.4)

The 5 selected radiomics features were then used to construct and evaluate 11 machine-learning models based on venous-phase contrast-enhanced ultrasound images. The cross-validation AUC values of these models are presented in **Figure 4.43**, while their detailed diagnostic performance in the training and test cohorts is summarized in **Table 4.36**. The corresponding ROC curves are shown in **Figure 4.44**.

Among the evaluated venous-phase CEUS radiomics models, Random Forest achieved the highest observed AUC in the test cohort, with an AUC of 0.701. However, this ranking should be interpreted descriptively because pairwise statistical comparisons among the classifiers were not available. The model achieved an AUC of 0.936 in the training cohort and 0.701 in the test cohort, representing an absolute reduction of 0.235.

This marked decline indicates substantial performance optimism and suggests possible overfitting, with only modest discriminatory ability retained in the test cohort.

The ROC curves of the Random Forest model in the training and test cohorts are presented in **Figure 4.45**. To further assess its potential clinical utility, decision curve analysis was performed. The results indicated that the Random Forest radiomics model provided a greater net clinical benefit than the treat-all and treat-none strategies across the evaluated threshold-probability range, as shown in **Figure 4.46**. Finally, the sample-level predicted probabilities generated by the model are presented in **Figure 4.47**.

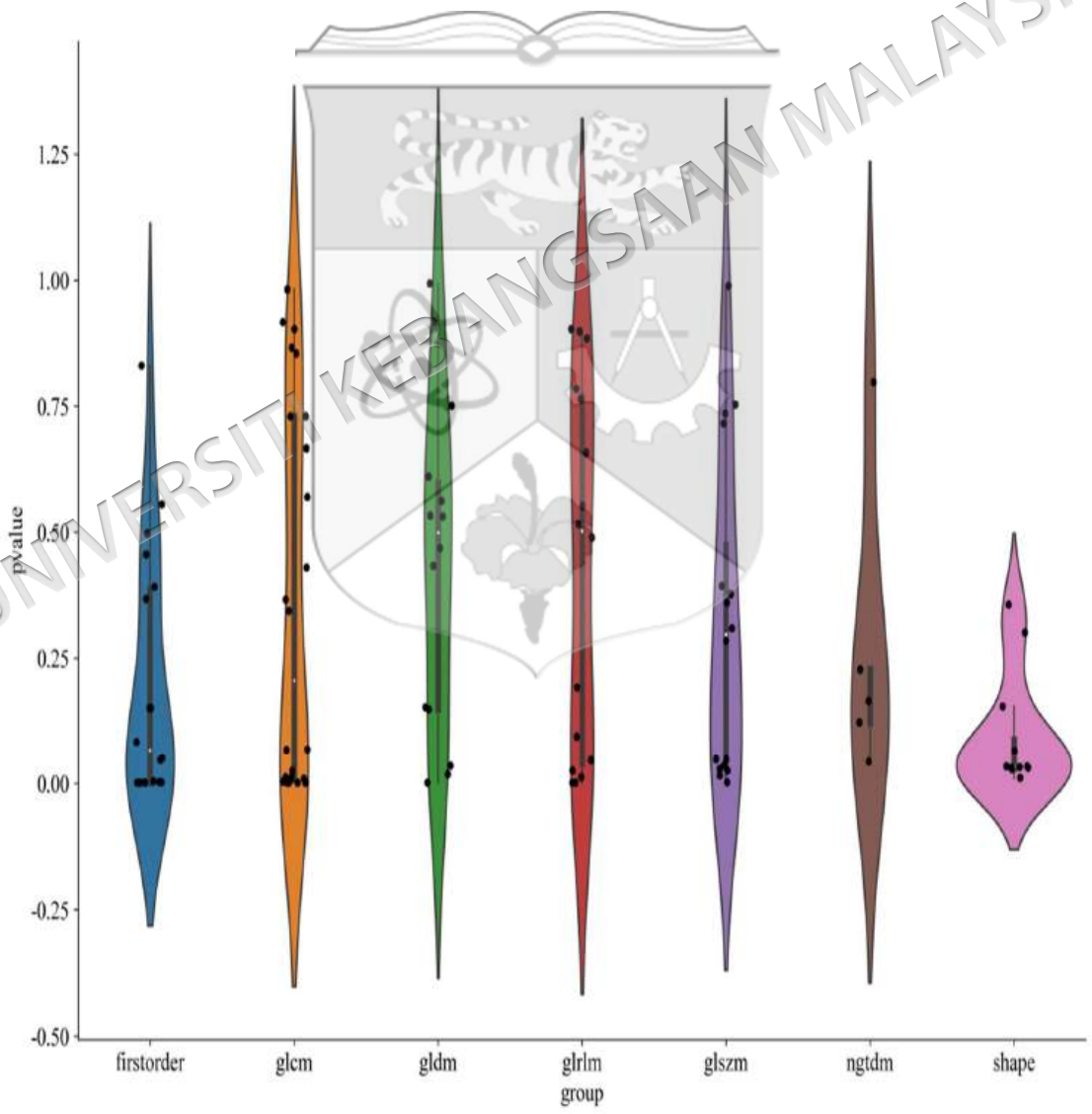


Figure 4.39 Radiomics features distribution and corresponding p-values for venous-phase Sonazoid CEUS

This figure presents the distribution of radiomics features extracted from venous-phase contrast-enhanced ultrasound images together with their corresponding p-values. The results provide an overview of the statistical significance of individual features in differentiating metastatic from non-metastatic lateral cervical lymph nodes.

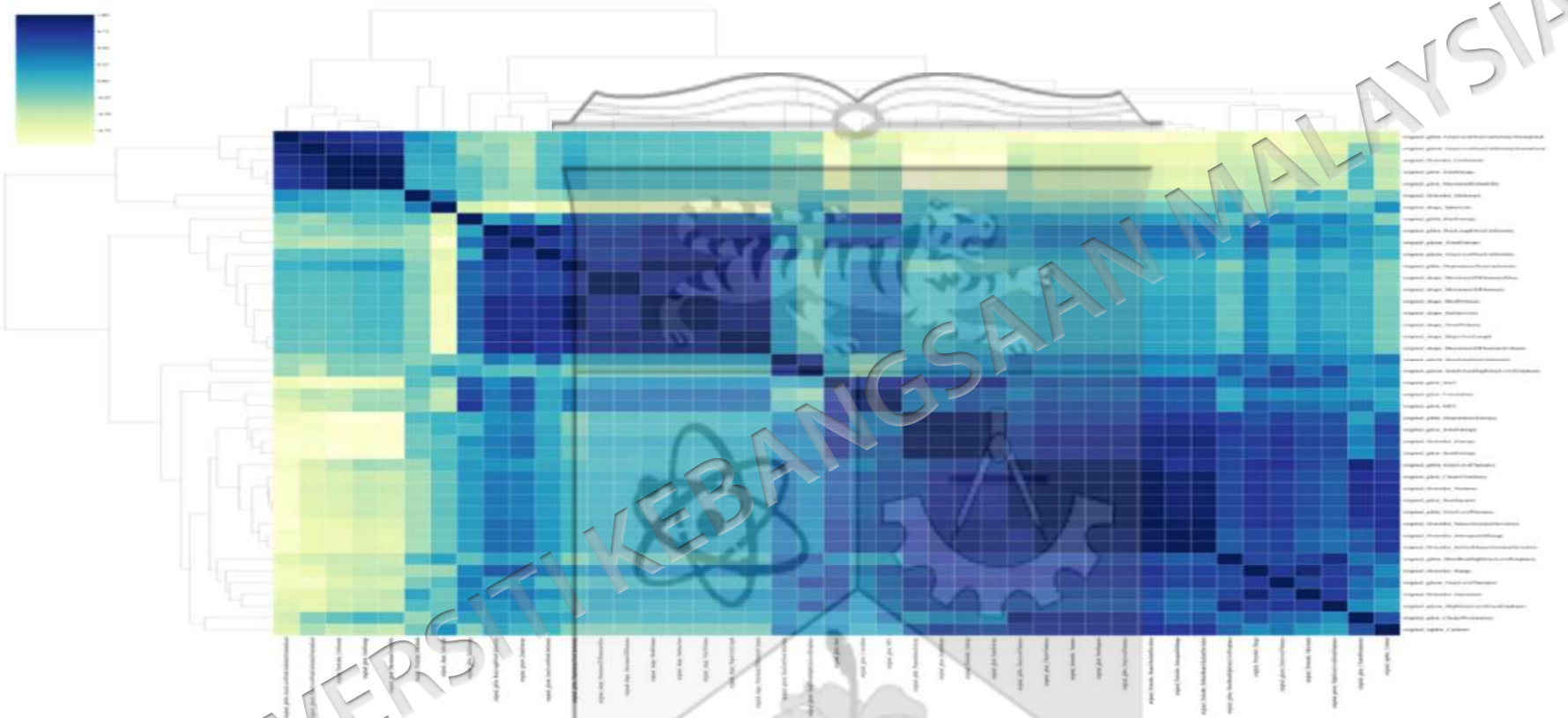


Figure 4.40 Clustered correlation matrix of radiomics features extracted from venous-phase Sonazoid CEUS

This figure illustrates the correlation structure among radiomics features extracted from venous-phase contrast-enhanced ultrasound images. Features with similar correlation patterns are grouped by hierarchical clustering, thereby facilitating the identification of redundant or highly correlated variables before feature selection. Darker colors indicate stronger correlations

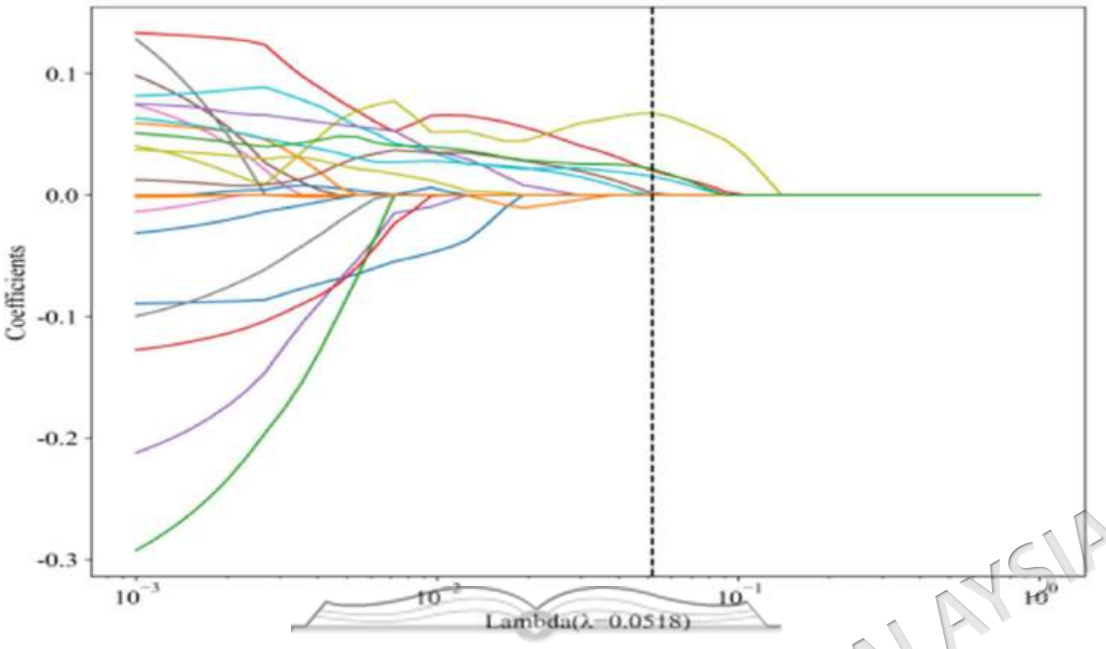


Figure 4.41 LASSO coefficient profiles of radiomics features from venous-phase Sonazoid CEUS

This figure shows the coefficient trajectories of radiomics features during least absolute shrinkage and selection operator (LASSO) regression as the regularization parameter varied. Changes in coefficient magnitude reflect the shrinkage process used to identify the most informative radiomics features. Different colored lines represent the retained 23 radiomics features.

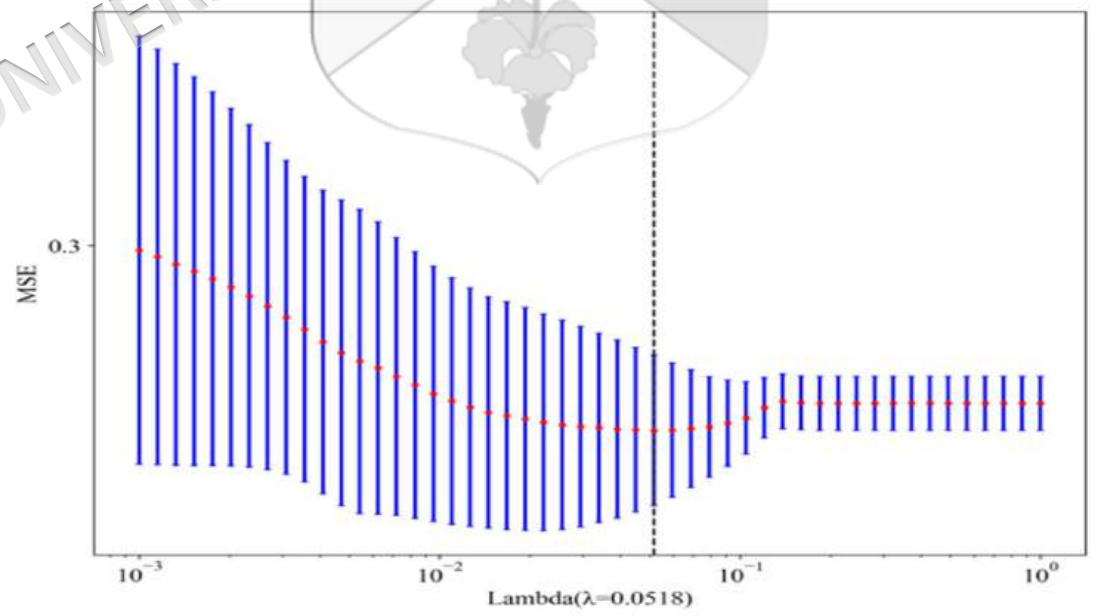


Figure 4.42 Mean squared error of LASSO regression with ten-fold cross-validation for venous-phase Sonazoid CEUS

This figure presents the mean squared error across different values of the regularization parameter in LASSO regression, with an optimal λ value of 0.0518. Ten-fold cross-validation was used to identify the optimal feature subset.

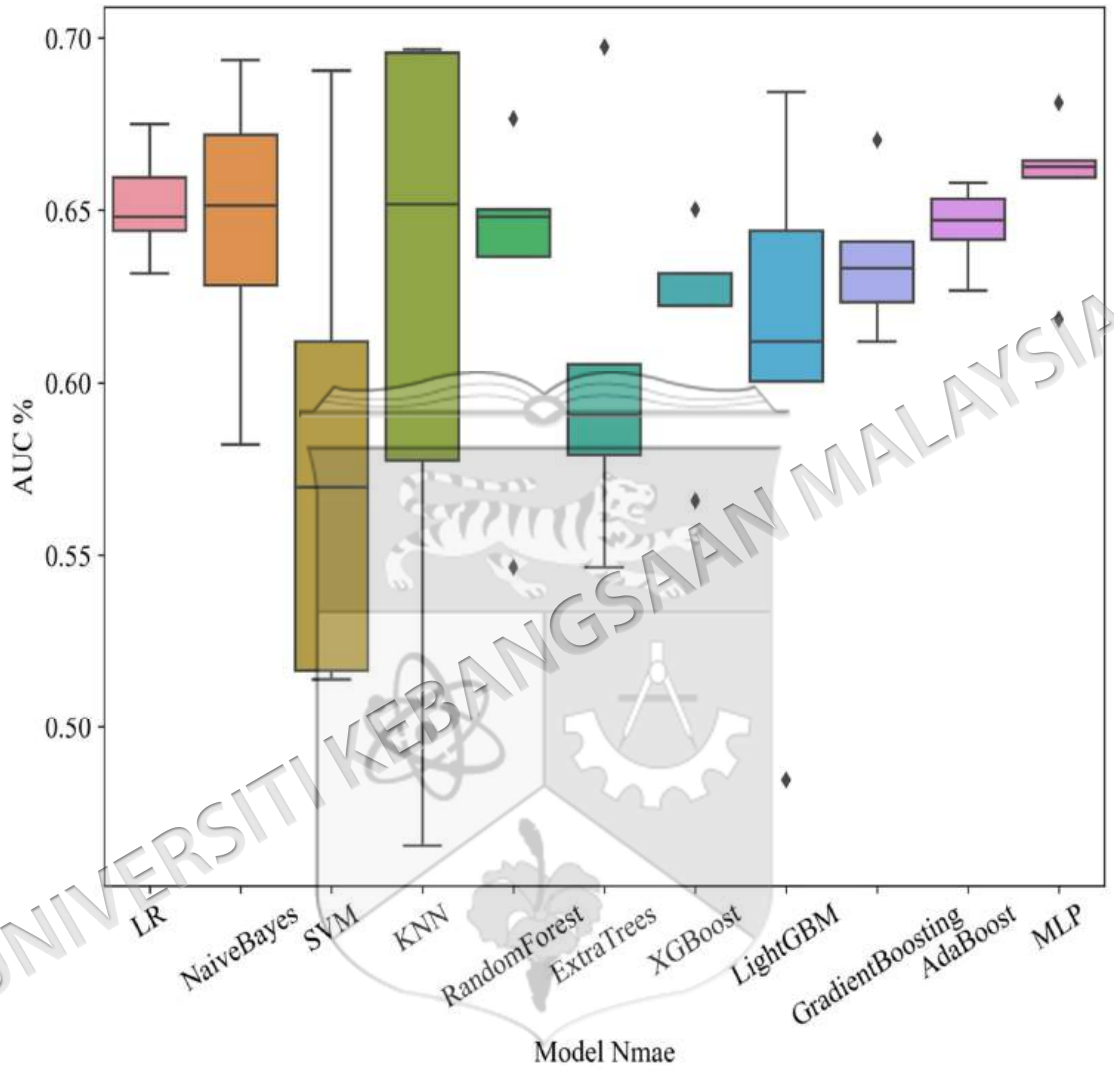


Figure 4.43 Cross-validation AUC values of machine learning models based on venous-phase Sonazoid CEUS radiomics

This figure compares the cross-validation area under the receiver operating characteristic curve (AUC) values of 11 machine learning models constructed using radiomics features derived from venous-phase contrast-enhanced ultrasound images. It provides an overall comparison of model discrimination performance during cross-validation.

Table 4.36 Performance of venous-phase Sonazoid CEUS radiomics-based machine learning models

This table summarizes the diagnostic performance of 11 machine learning models constructed using radiomics features derived from venous-phase contrast-enhanced ultrasound images. The results include model performance in both the training and test cohorts and provide a basis for identifying the best-performing model.

model_name	Accuracy	AUC	95% CI	Sensitivity	Specificity	PPV	NPV	Precision	Recall	F1	Threshold	Task
LR	0.671	0.694	0.6079 - 0.7798	0.463	0.855	0.738	0.644	0.738	0.463	0.569	0.556	label-train
LR	0.629	0.625	0.4322 - 0.8178	0.750	0.526	0.571	0.714	0.571	0.750	0.649	0.377	label-test
NaiveBayes	0.650	0.672	0.5843 - 0.7605	0.448	0.829	0.698	0.630	0.698	0.448	0.545	0.511	label-train
NaiveBayes	0.714	0.651	0.4609 - 0.8417	0.437	0.947	0.875	0.667	0.875	0.437	0.583	0.524	label-test
SVM	0.706	0.718	0.6307 - 0.8057	0.552	0.842	0.755	0.681	0.755	0.552	0.638	0.514	label-train
SVM	0.629	0.516	0.3078 - 0.7251	0.250	0.947	0.800	0.600	0.800	0.250	0.381	0.580	label-test
KNN	0.706	0.764	0.6905 - 0.8380	0.597	0.803	0.727	0.693	0.727	0.597	0.656	0.600	label-train
KNN	0.657	0.696	0.5332 - 0.8582	0.937	0.421	0.577	0.889	0.577	0.937	0.714	0.400	label-test
RandomForest	0.867	0.936	0.8982 - 0.9738	0.851	0.882	0.864	0.870	0.864	0.851	0.857	0.505	label-train
RandomForest	0.714	0.701	0.5164 - 0.8849	0.750	0.684	0.667	0.765	0.667	0.750	0.706	0.484	label-test
ExtraTrees	0.818	0.885	0.8315 - 0.9388	0.851	0.789	0.781	0.857	0.781	0.851	0.814	0.438	label-train
ExtraTrees	0.686	0.632	0.4395 - 0.8237	0.375	0.947	0.857	0.643	0.857	0.375	0.522	0.526	label-test
XGBoost	0.853	0.903	0.8543 - 0.9527	0.896	0.816	0.811	0.899	0.811	0.896	0.851	0.476	label-train
XGBoost	0.686	0.538	0.3286 - 0.7471	0.875	0.526	0.609	0.833	0.609	0.875	0.718	0.415	label-test

to be continued…

… continuation

LightGBM	0.748	0.832	0.7664 - 0.8968	0.791	0.711	0.707	0.794	0.707	0.791	0.746	0.441	label-train
LightGBM	0.600	0.562	0.3640 - 0.7610	0.562	0.632	0.562	0.632	0.562	0.562	0.562	0.463	label-test
GradientBoosting	0.832	0.921	0.8801 - 0.9612	0.925	0.750	0.765	0.919	0.765	0.925	0.838	0.445	label-train
GradientBoosting	0.657	0.612	0.4147 - 0.8090	0.875	0.474	0.583	0.818	0.583	0.875	0.700	0.422	label-test
AdaBoost	0.741	0.822	0.7581 - 0.8867	0.806	0.684	0.692	0.800	0.692	0.806	0.745	0.500	label-train
AdaBoost	0.657	0.627	0.4364 - 0.8169	0.750	0.579	0.600	0.733	0.600	0.750	0.667	0.500	label-test
MLP	0.650	0.703	0.6181 - 0.7880	0.716	0.592	0.608	0.703	0.608	0.716	0.658	0.446	label-train
MLP	0.686	0.618	0.4225 - 0.8143	0.437	0.895	0.778	0.654	0.778	0.437	0.560	0.514	label-test

AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value; F1=2 x (precision x recall)/ (precision + recall); LR, logistic regression; SVM, support vector machine; KNN, K-nearest neighbors; XGBoost, extreme gradient boosting; LightGBM, light gradient boosting machine; MLP, multi-layer perceptron.

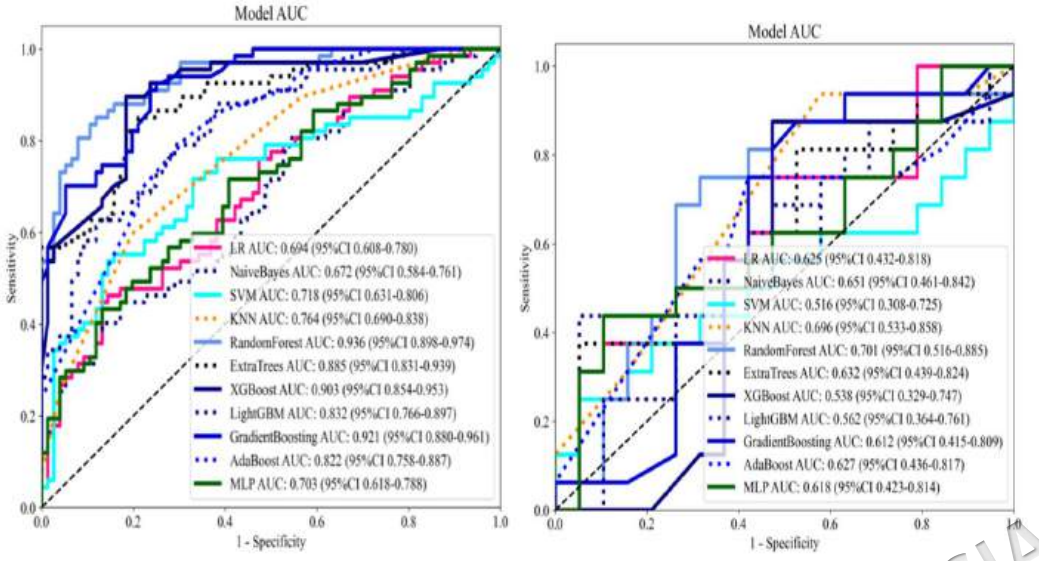


Figure 4.44 ROC curves of 11 machine learning models based on venous-phase Sonazoid CEUS

This figure illustrates the receiver operating characteristic (ROC) curves of the machine learning models in the training and the test cohorts. The curves provide a visual comparison of the diagnostic performance of the candidate models across the two datasets. (left) Training cohort. (right) Test cohort.

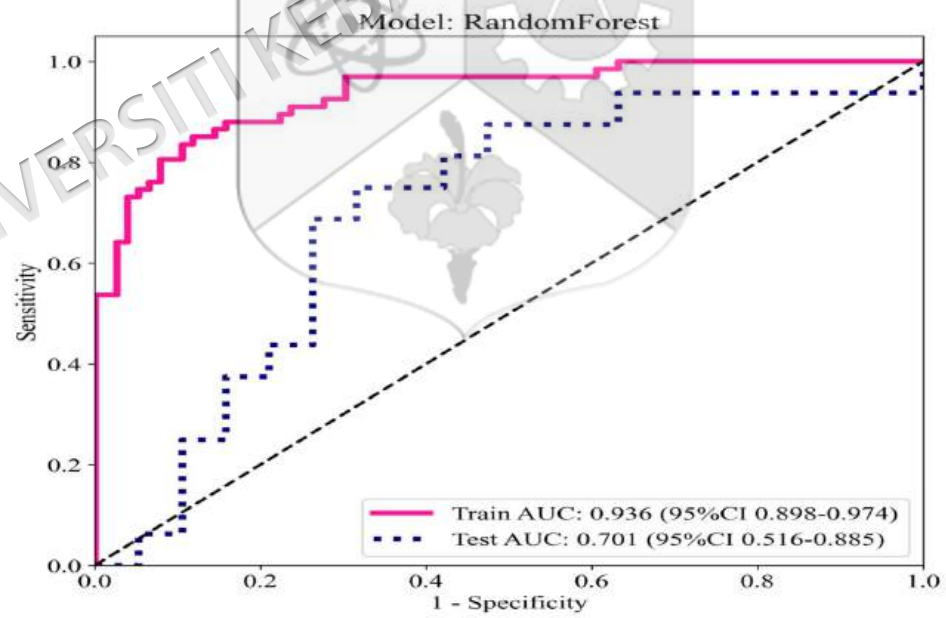


Figure 4.45 ROC curves of the Random Forest radiomics model based on venous-phase Sonazoid CEUS

This figure shows the receiver operating characteristic curves of the Random Forest radiomics model in the training and test cohorts. The AUC was 0.936 in the training cohort and 0.701 in the test cohort.

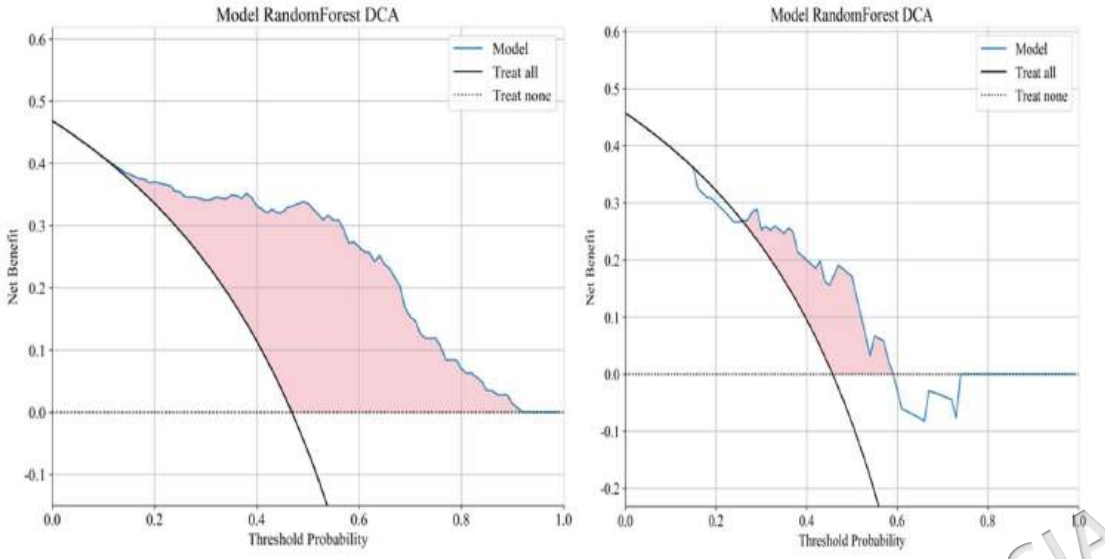


Figure 4.46 Decision curve analysis of the Random Forest radiomics model based on venous-phase Sonazoid CEUS

The left panel represents the training cohort and the right panel represents the test cohort. Net benefit of the venous-phase CEUS Random Forest model is compared with the treat-all and treat-none strategies across threshold probabilities. The marked training-to-test reduction in discrimination should be considered when interpreting the DCA; any apparent test-cohort benefit is exploratory and requires confirmation in a larger external cohort.

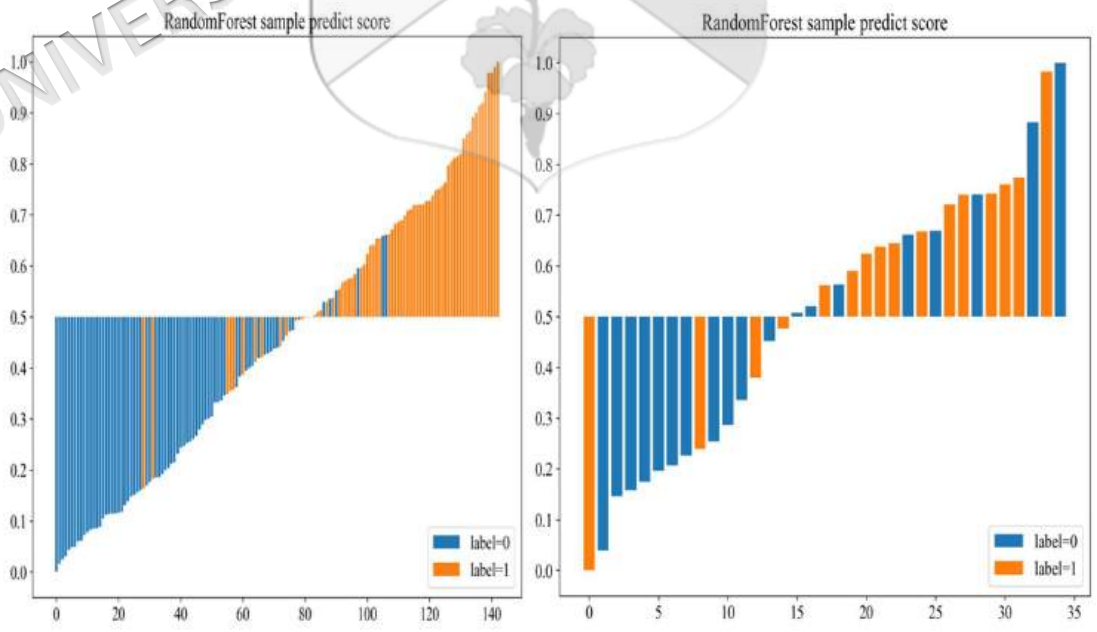


Figure 4.47 Distribution of sample-level prediction scores of the Random Forest radiomics model based on venous-phase Sonazoid CEUS

This figure presents the distribution of prediction scores generated by the Random Forest radiomics model for individual samples in the training and test cohorts. The score distribution provides additional insight into the separation between metastatic and non-metastatic cases. (left) Training cohort. (right) Test cohort.

4.7.5 Radiomics of Post-Vascular Phase Contrast-Enhanced Ultrasound

The distributions of the extracted radiomics features and their corresponding p-values are presented in **Figure 4.48**. To assess inter-feature relationships and identify potential redundancy, correlation analysis was subsequently performed and visualized using a clustered correlation matrix, as shown in **Figure 4.49**.

Following correlation coefficient-based feature screening, 30 radiomics features were retained. These comprised 6 first-order features, 3 grey-level co-occurrence matrix (GLCM) features, 5 grey-level dependence matrix (GLDM) features, 7 grey-level run-length matrix (GLRLM) features, 5 grey-level size-zone matrix (GLSZM) features, 3 neighbouring grey-tone difference matrix (NGTDM) features, and 1 shape feature.

To further reduce dimensionality and identify the most informative predictors, least absolute shrinkage and selection operator (LASSO) regression was performed with ten-fold cross-validation. 8 radiomics features were ultimately selected, with an optimal λ value of 0.0518. The LASSO coefficient profiles and the corresponding cross-validation mean squared error are presented in **Figure 4.50 and Figure 4.51**, respectively; the minimum-error criterion λ value was 0.0518. Among the selected features, original_gldm_SmallDependenceLowGrayLevelEmphasis had the largest positive coefficient, indicating the greatest contribution to the calculated radiomics score.

Based on the 8 selected features and their corresponding LASSO coefficients, the radiomics score for each lymph node was calculated according to **Equation 4.5**.

Radiomics score

$$=0.46629213483146065-0.036909X_1+0.010307X_2+0.009439X_3$$
$$+0.071532X_4+0.013551X_5+0.091566X_6+0.001690X_7+0.007599X_8$$

X_1=original_firstorder_10Percentile,

X_2=original_firstorder_Kurtosis,

X_3=original_glcm_ClusterProminence,

X_4=original_glcm_ClusterShade,

X_5=original_gldm_GrayLevelNonUniformity,

X_6=original_gldm_SmallDependenceLowGrayLevelEmphasise,

X_7=original_glszm_LargeAreaLowGrayLevelEmphasis,

X_8=original_glszm_SizeZoneNonUniformity. (0.5)

The 8 selected radiomics features were then used to construct and evaluate 11 machine-learning models based on post-vascular-phase contrast-enhanced ultrasound images. The cross-validation AUC values of these models are presented in **Figure 4.52**, while their detailed diagnostic performance in the training and test cohorts is summarized in **Table 4.37**. The corresponding ROC curves are shown in **Figure 4.53**.

Among the evaluated post-vascular-phase CEUS radiomics models, the multilayer perceptron (MLP) achieved the highest observed AUC in the test cohort, with an AUC of 0.882. However, this ranking should be interpreted descriptively because pairwise statistical comparisons among the classifiers were not available. The model achieved an AUC of 0.821 in the training cohort and 0.882 in the test cohort. The higher test-cohort AUC should not be interpreted as evidence of improved performance in unseen populations, as the difference may reflect sampling variability associated with the relatively small held-out test cohort.

The ROC curves of the MLP model in the training and test cohorts are presented in **Figure 4.54**. To further assess its potential clinical utility, decision curve analysis was performed. The results indicated that the MLP radiomics model provided a greater net clinical benefit than the treat-all and treat-none strategies across the

evaluated threshold-probability range, as shown in **Figure 4.55**. Finally, the sample-level predicted probabilities generated by the MLP model are presented in **Figure 4.56**.

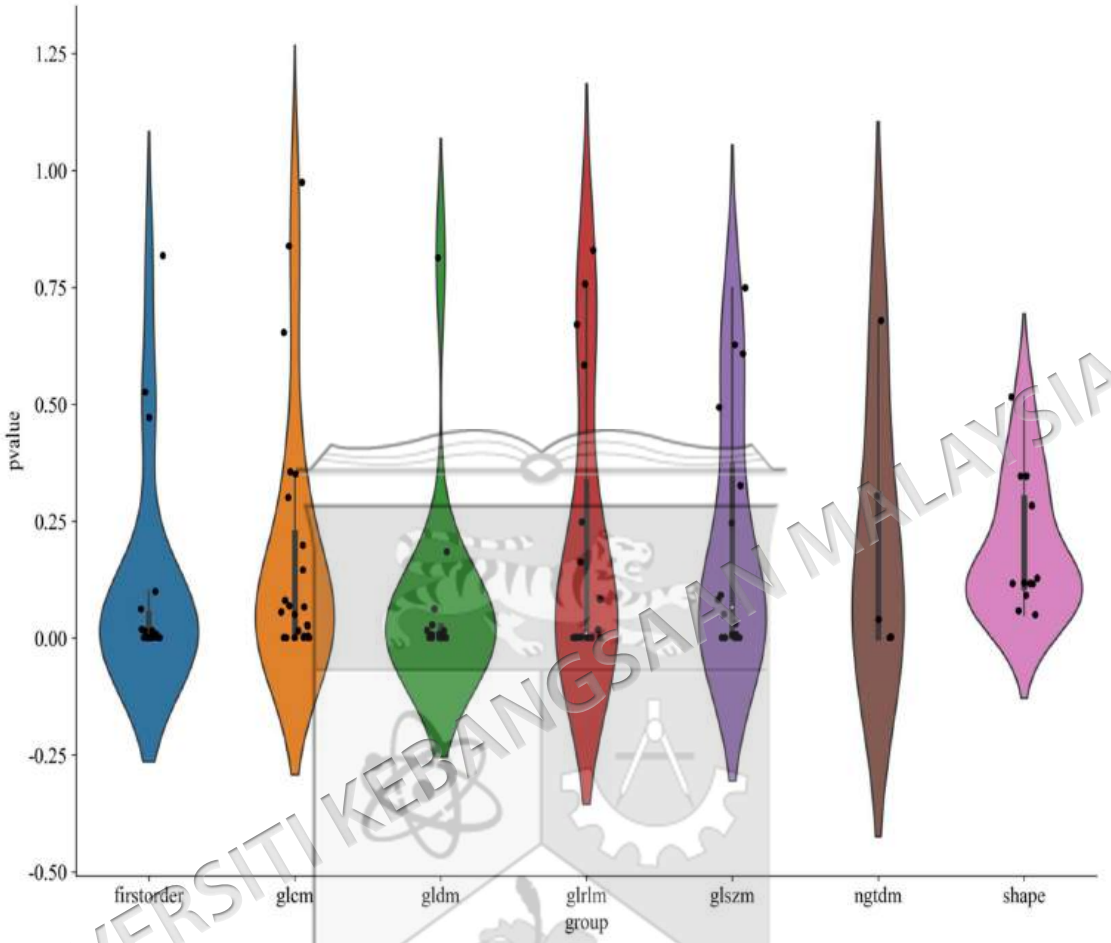


Figure 4.48 Radiomics features distribution and corresponding p-values for post-vascular phase Sonazoid CEUS

This figure presents the distribution of radiomics features extracted from post-vascular phase contrast-enhanced ultrasound images together with their corresponding p-values. The results provide an overview of the statistical significance of individual features in differentiating metastatic from non-metastatic lateral cervical lymph nodes.

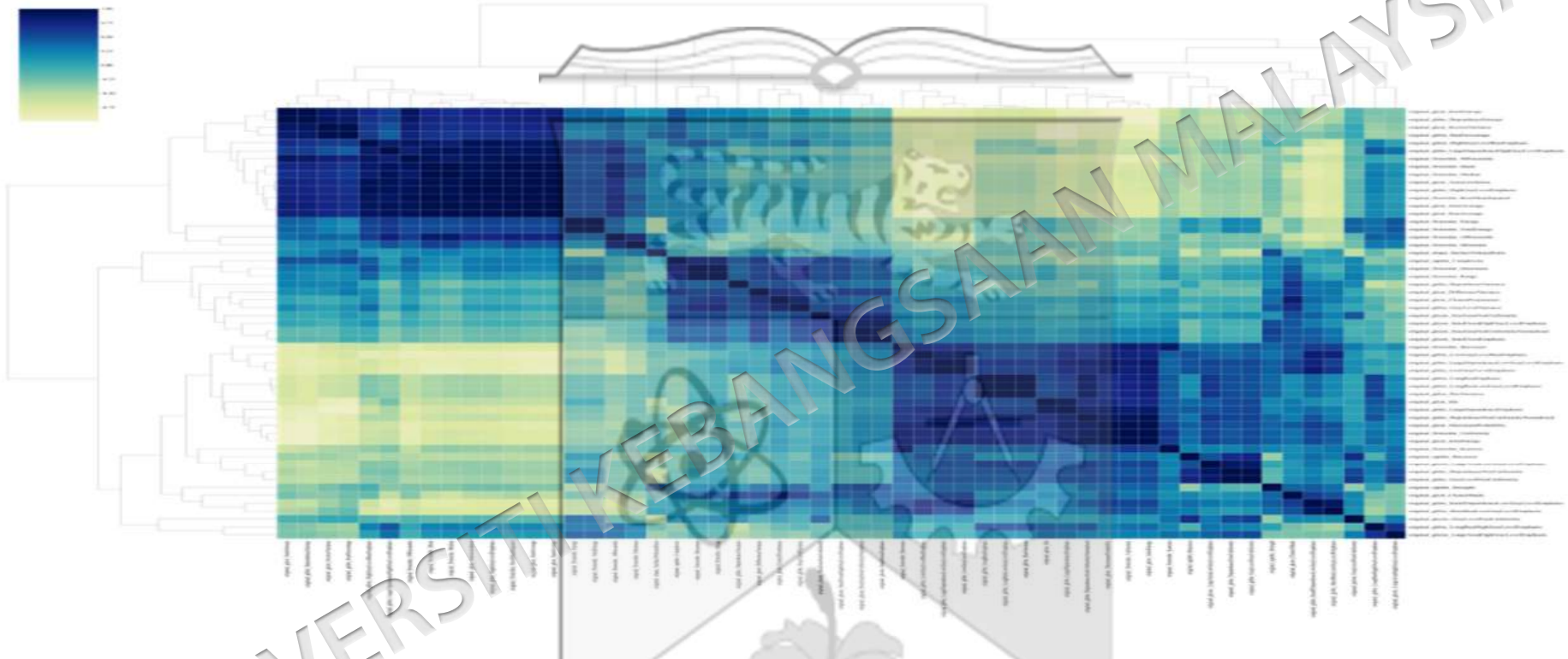


Figure 4.49 Clustered correlation matrix of radiomics features extracted from post-vascular phase Sonazoid CEUS

This figure illustrates the correlation structure among radiomics features extracted from post-vascular phase contrast-enhanced ultrasound images. Features with similar correlation patterns are grouped by hierarchical clustering, thereby facilitating the identification of redundant or highly correlated variables. Darker colors indicate stronger correlations.

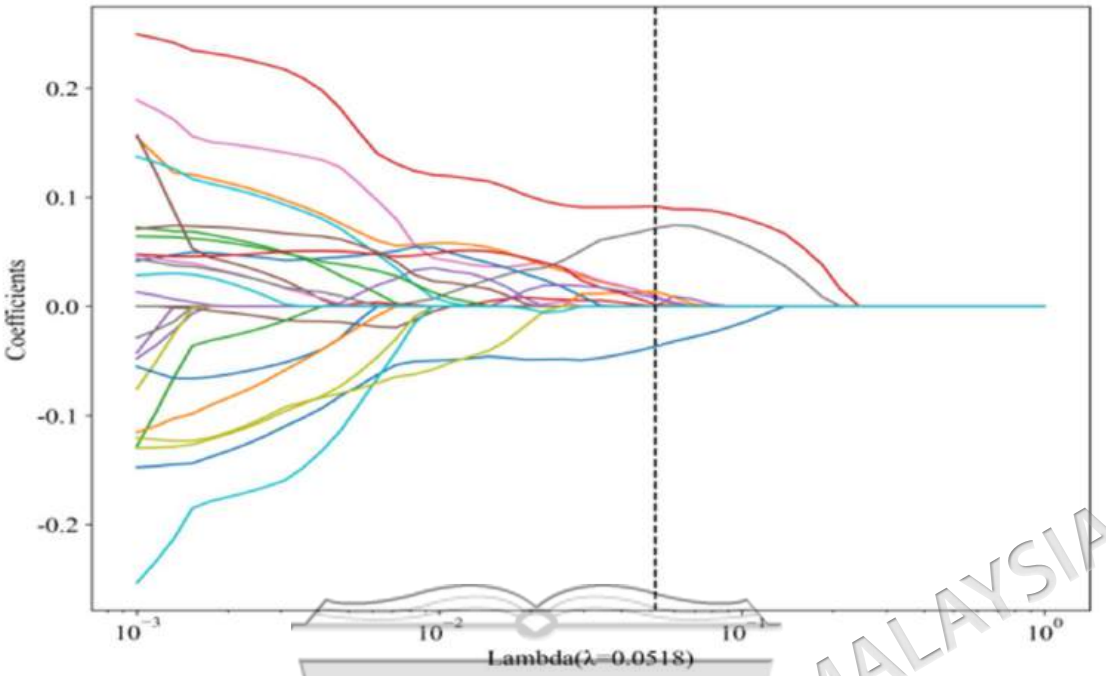


Figure 4.50 LASSO coefficient profiles of radiomics features from post-vascular phase Sonazoid CEUS

This figure shows the coefficient trajectories of radiomics features during least absolute shrinkage and selection operator (LASSO) regression as the regularization parameter varied. Changes in coefficient magnitude reflect the shrinkage process used to identify the most informative radiomics features. Different colored lines represent the retained 30 radiomics features.

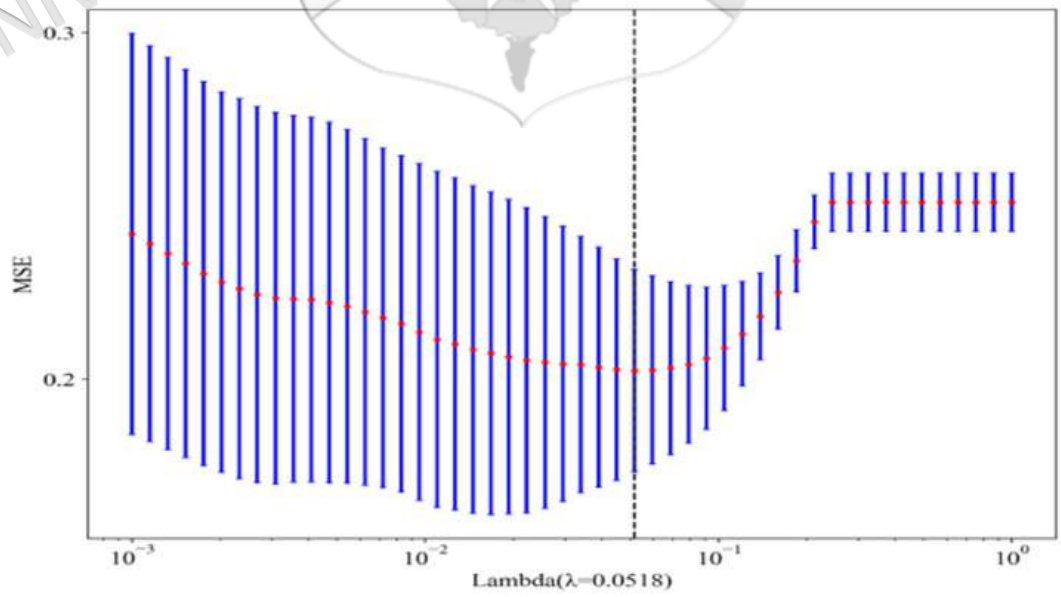


Figure 4.51 Mean squared error of LASSO regression with ten-fold cross-validation for post-vascular phase Sonazoid CEUS

This figure presents the mean squared error across different values of the regularization parameter in LASSO regression, with an optimal λ value of 0.0518. Ten-fold cross-validation was used to identify the optimal feature subset.

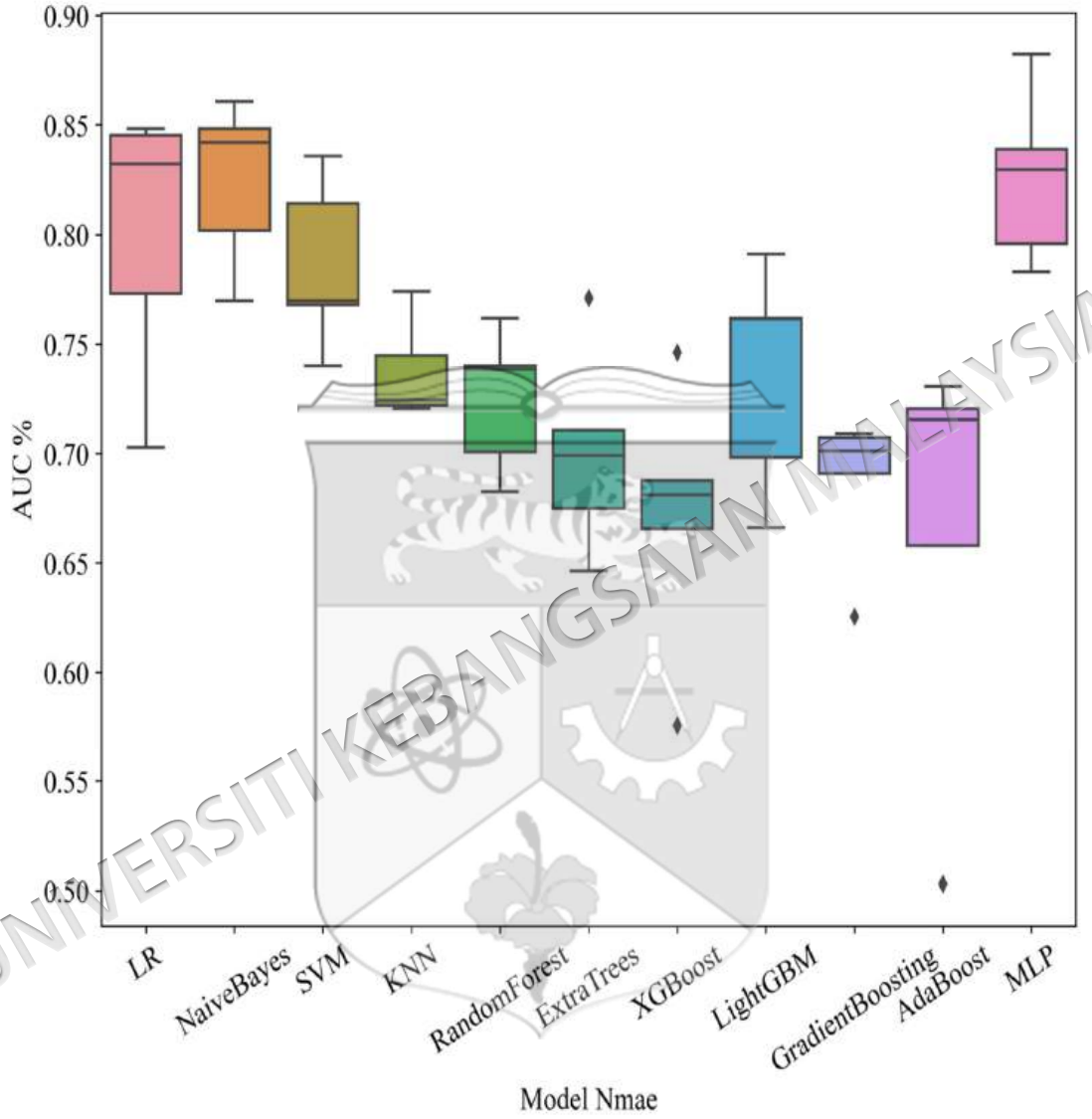


Figure 4.52 Cross-validation AUC values of machine learning models based on post-vascular phase Sonazoid CEUS

This figure compares the cross-validation area under the receiver operating characteristic curve (AUC) values of 11 machine learning models constructed using radiomics features derived from post-vascular phase contrast-enhanced ultrasound images. It provides an overall comparison of model discrimination performance during cross-validation.

Table 4.37 Performance of post-vascular-phase Sonazoid CEUS radiomics-based machine learning models

This table summarizes the diagnostic performance of 11 machine learning models constructed using radiomics features derived from post-vascular phase CEUS images. The results include model performance in both the training and test cohorts and provide a basis for identifying the best-performing model.

Model_Name	Accuracy	AUC	95% CI	Sensitivity	Specificity	PPV	NPV	Precision	Recall	F1	Threshold	Task
LR	0.754	0.820	0.7519 - 0.8876	0.879	0.645	0.682	0.860	0.682	0.879	0.768	0.344	label-train
LR	0.806	0.830	0.6875 - 0.9719	0.941	0.684	0.727	0.929	0.727	0.941	0.821	0.393	label-test
NaiveBayes	0.739	0.812	0.7431 - 0.8805	0.864	0.632	0.671	0.842	0.671	0.864	0.755	0.003	label-train
NaiveBayes	0.806	0.848	0.7194 - 0.9772	0.882	0.737	0.750	0.875	0.750	0.882	0.811	0.010	label-test
SVM	0.782	0.858	0.7989 - 0.9176	0.803	0.763	0.746	0.817	0.746	0.803	0.774	0.471	label-train
SVM	0.778	0.814	0.6721 - 0.9564	0.941	0.632	0.696	0.923	0.696	0.941	0.800	0.387	label-test
KNN	0.796	0.862	0.8050 - 0.9194	0.758	0.829	0.794	0.797	0.794	0.758	0.775	0.600	label-train
KNN	0.667	0.745	0.5850 - 0.9041	0.882	0.474	0.600	0.818	0.600	0.882	0.714	0.400	label-test
RandomForest	0.866	0.952	0.9217 - 0.9814	0.939	0.803	0.805	0.938	0.805	0.939	0.867	0.444	label-train
RandomForest	0.750	0.765	0.6073 - 0.9221	0.882	0.632	0.682	0.857	0.682	0.882	0.769	0.449	label-test
ExtraTrees	0.817	0.902	0.8554 - 0.9492	0.742	0.882	0.845	0.798	0.845	0.742	0.790	0.497	label-train
ExtraTrees	0.750	0.814	0.6759 - 0.9525	0.941	0.579	0.667	0.917	0.667	0.941	0.780	0.417	label-test
XGBoost	0.908	0.956	0.9262 - 0.9861	0.909	0.908	0.896	0.920	0.896	0.909	0.902	0.514	label-train
XGBoost	0.722	0.697	0.5191 - 0.8741	0.941	0.526	0.640	0.909	0.640	0.941	0.762	0.278	label-test

to be continued…

… continuation

LightGBM	0.838	0.910	0.8633 - 0.9563	0.803	0.868	0.841	0.835	0.841	0.803	0.822	0.540	label-train
LightGBM	0.722	0.745	0.5817 - 0.9074	0.941	0.526	0.640	0.909	0.640	0.941	0.762	0.291	label-test
GradientBoosting	0.901	0.957	0.9279 - 0.9860	0.894	0.908	0.894	0.908	0.894	0.894	0.894	0.523	label-train
GradientBoosting	0.694	0.701	0.5246 - 0.8779	0.824	0.579	0.636	0.786	0.636	0.824	0.718	0.474	label-test
AdaBoost	0.845	0.913	0.8687 - 0.9578	0.924	0.776	0.782	0.922	0.782	0.924	0.847	0.488	label-train
AdaBoost	0.722	0.731	0.5613 - 0.9000	0.941	0.526	0.640	0.909	0.640	0.941	0.762	0.454	label-test
MLP	0.754	0.821	0.7540 - 0.8879	0.652	0.842	0.782	0.736	0.782	0.652	0.711	0.567	label-train
MLP	0.833	0.882	0.7748 - 0.9899	0.824	0.842	0.824	0.842	0.824	0.824	0.824	0.532	label-test

AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value; F1=2 x (precision x recall)/ (precision + recall); LR, logistic regression; SVM, support vector machine; KNN, K-nearest neighbors; XGBoost, extreme gradient boosting; Light GBM, light gradient boosting machine; MLP, multi-layer perceptron.

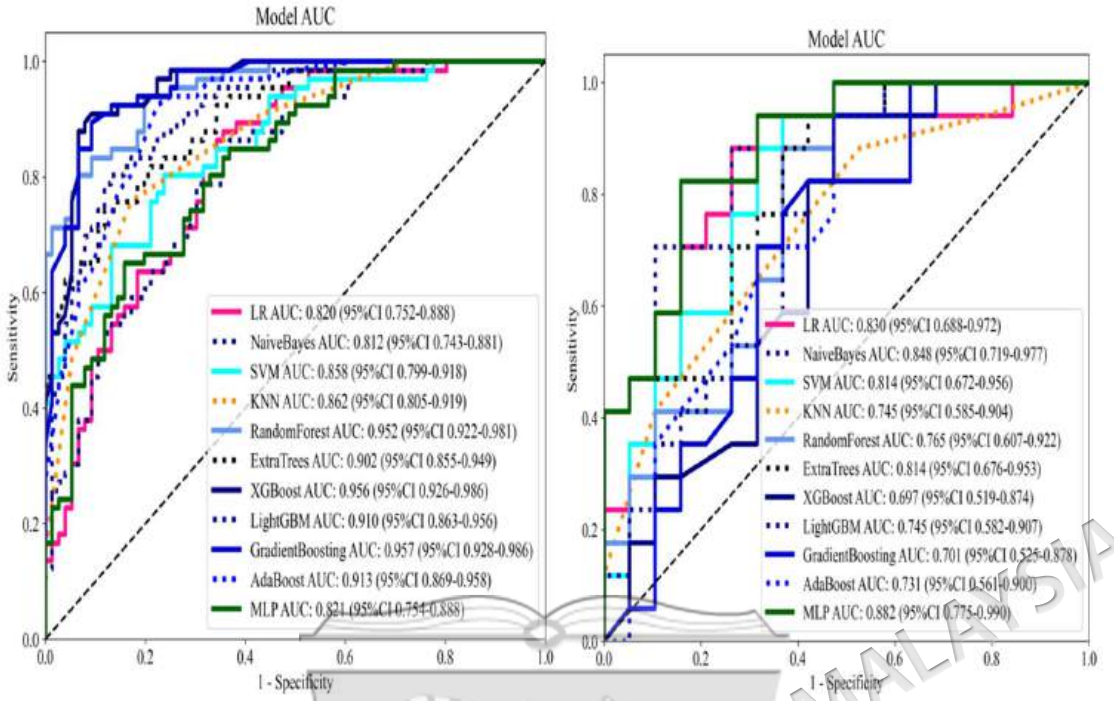


Figure 4.53 ROC curves of 11 machine learning models based on post-vascular phase Sonazoid CEUS

This figure illustrates the receiver operating characteristic (ROC) curves of the machine learning models in the training cohort and the test cohort. The curves provide a visual comparison of the diagnostic performance of the candidate models across the two datasets. (left) Training cohort. (right) Test cohort.

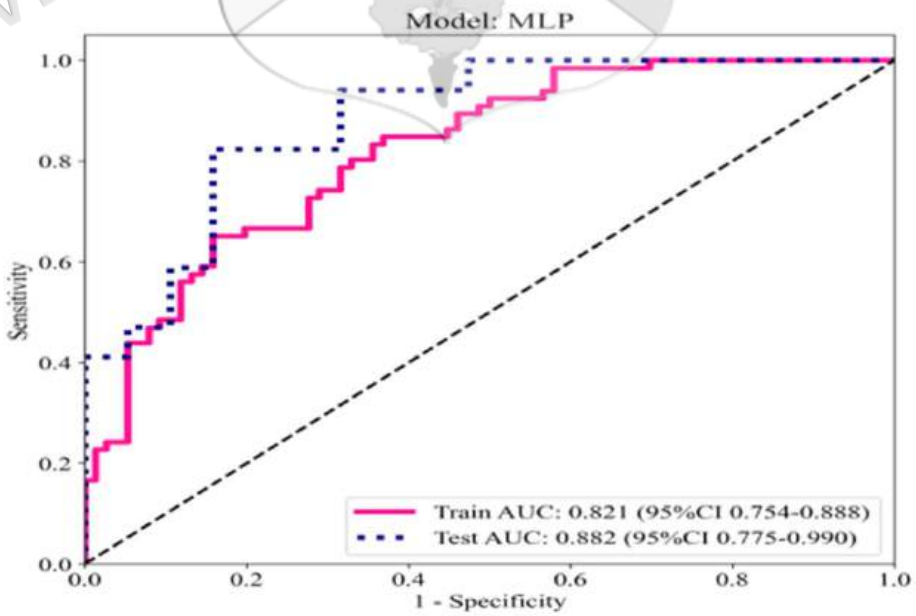


Figure 4.54 ROC curves of the MLP radiomics model based on post-vascular phase Sonazoid CEUS

This figure shows the receiver operating characteristic curves of the MLP radiomics model in the training cohort and the test cohort. The AUC was 0.821 in the training cohort and 0.882 in the test cohort.

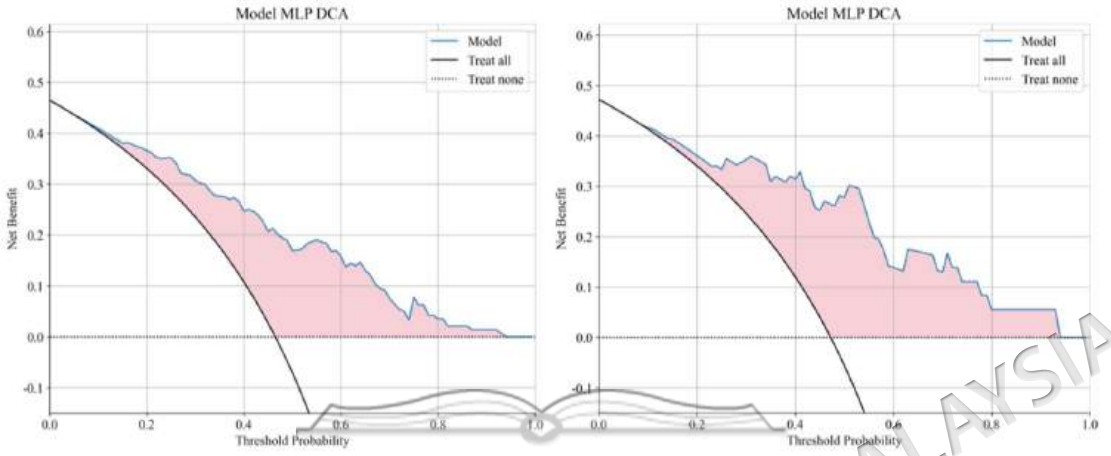


Figure 4.55 Decision curve analysis of the MLP radiomics model based on post-vascular-phase Sonazoid CEUS

The left panel represents the training cohort and the right panel represents the test cohort. The curves show the net benefit of the post-vascular-phase CEUS MLP model across threshold probabilities relative to the treat-all and treat-none strategies. The model demonstrates potential clinical utility over threshold ranges where its curve is above both reference strategies, but it should not be considered uniformly superior because net benefit varies with the selected threshold and the test cohort is small.

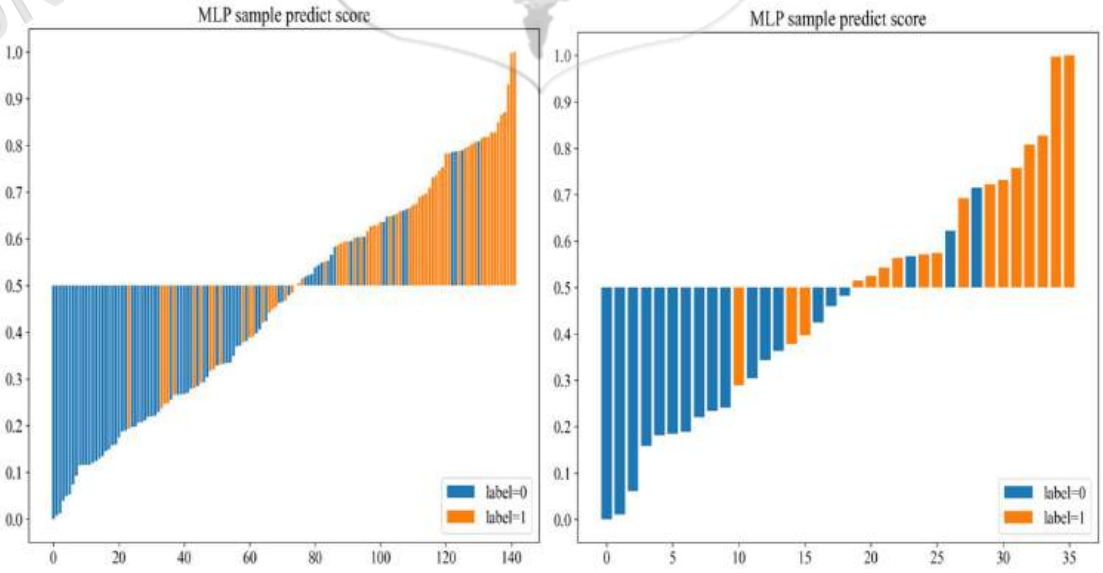


Figure 4.56 Distribution of sample-level prediction scores of the MLP radiomics model based on post-vascular phase Sonazoid CEUS

This figure presents the distribution of prediction scores generated by the MLP radiomics model for individual samples in the training and test cohorts. The score distribution provides additional insight into the separation between metastatic and non-metastatic cases. (left) Training cohort. (right) Test cohort.

4.7.6 Summary of Radiomics Findings

Overall, radiomics performance varied by imaging modality and classifier. In the independent test cohort, the AUCs of the five selected modality-specific models ranged from 0.701 to 0.882. The post-vascular-phase CEUS MLP model had the highest observed test-cohort AUC, followed by the conventional grayscale-ultrasound Extra Trees, CDFI AdaBoost, arterial-phase CEUS Naive Bayes, and venous-phase CEUS Random Forest models. These rankings should be interpreted together with the confidence intervals and paired DeLong comparisons, rather than based on the numerical AUC values alone.

The post-vascular-phase CEUS MLP model achieved the highest observed test-cohort AUC of 0.882 (95% CI: 0.775–0.990). The corresponding values were 0.830 (95% CI: 0.697–0.962) for the grayscale-ultrasound Extra Trees model, 0.824 (95% CI: 0.688–0.959) for the CDFI AdaBoost model, 0.822 (95% CI: 0.681–0.964) for the arterial-phase CEUS Naive Bayes model, and 0.701 (95% CI: 0.516–0.885) for the venous-phase CEUS Random Forest model. The broad and overlapping confidence intervals indicate considerable uncertainty in the test-cohort estimates. Pairwise DeLong tests were therefore performed separately in the training and test cohorts to compare the AUCs of the 5 selected models (**Figure 4.57**). The heatmaps of pairwise DeLong tests demonstrated varying differences in ROC-AUC among the 5 modality-specific models. The comparison between the grayscale-ultrasound Extra Trees and CDFI AdaBoost models was not statistically significant in either the training cohort ($p = 0.655$) or the test cohort ($p = 0.720$). Some other pairwise comparisons showed lower p-values; however, their statistical significance should be determined according to the exact p-values, with $p < 0.05$ considered statistically significant. The matrices should be interpreted cautiously because multiple pairwise tests were performed and the test cohort was small ($n = 36$); therefore, the results are exploratory and should be considered together with AUC differences and their 95% confidence intervals.

Decision curve analysis was used to assess the clinical consequences of applying the five selected radiomics models across a range of threshold probabilities. In both cohorts, the model curves showed varying net benefit relative to the treat-all and treat-none strategies, and no single model was uniformly dominant across the entire threshold range. In the test cohort, the post-vascular-phase CEUS MLP model showed comparatively favorable net benefit over portions of the clinically relevant threshold range, consistent with its higher observed discrimination; however, the curves crossed and became unstable at some higher thresholds. Therefore, DCA supports potential clinical utility but does not establish universal superiority of one modality. The comparative DCA curves for all 5 selected models are presented in **Figure 4.58**. DCA curves indicate that the preferred model may vary according to the chosen clinical threshold; apparent differences at extreme thresholds should be interpreted cautiously because few observations contribute to those regions.

Calibration curves for the best-performing model from each imaging modality are shown in **Figure 4.59**. In the training cohort, all models generally showed increasing observed event proportions with increasing predicted probabilities, although deviations from the 45-degree reference line were present. In the test cohort, the curves were more irregular, with modality-specific fluctuations across probability intervals. Overall, none of the models showed perfect agreement with the reference line across the full probability range. Calibration performance was therefore assessed descriptively by visual inspection; visual differences should be interpreted cautiously because the test cohort was small (n = 36) and numerical calibration indices were not available.

Table 4.38 Best-performing radiomics models in the test cohort

Imaging modality	Selected model	Training AUC (95% CI)	Test AUC (95% CI)	Train–test AUC	Interpretation
Grayscale-US	Extra Trees	0.927 (0.886–0.968)	0.830 (0.697–0.962)	−0.097	Acceptable test discrimination; possible overfitting.
CDFI	AdaBoost	0.899 (0.851–0.946)	0.824 (0.688–0.959)	−0.075	Moderate generalization gap; wide test CI.

to be continued…

… continuation

Arterial-phase CEUS	Naive Bayes	0.664 (0.575–0.752)	0.822 (0.681–0.964)	+0.158	Higher test AUC likely reflects sampling variability.
Venous-phase CEUS	Random Forest	0.936 (0.898–0.974)	0.701 (0.516–0.885)	−0.235	Weak test discrimination and limited generalizability .
Post-vascular CEUS	MLP	0.821 (0.754–0.888)	0.882 (0.775–0.990)	+0.061	Highest observed test AUC; superiority not statistically established.

AUC, area under the receiver operating characteristic curve; CI, confidence interval; MLP, multilayer perceptron

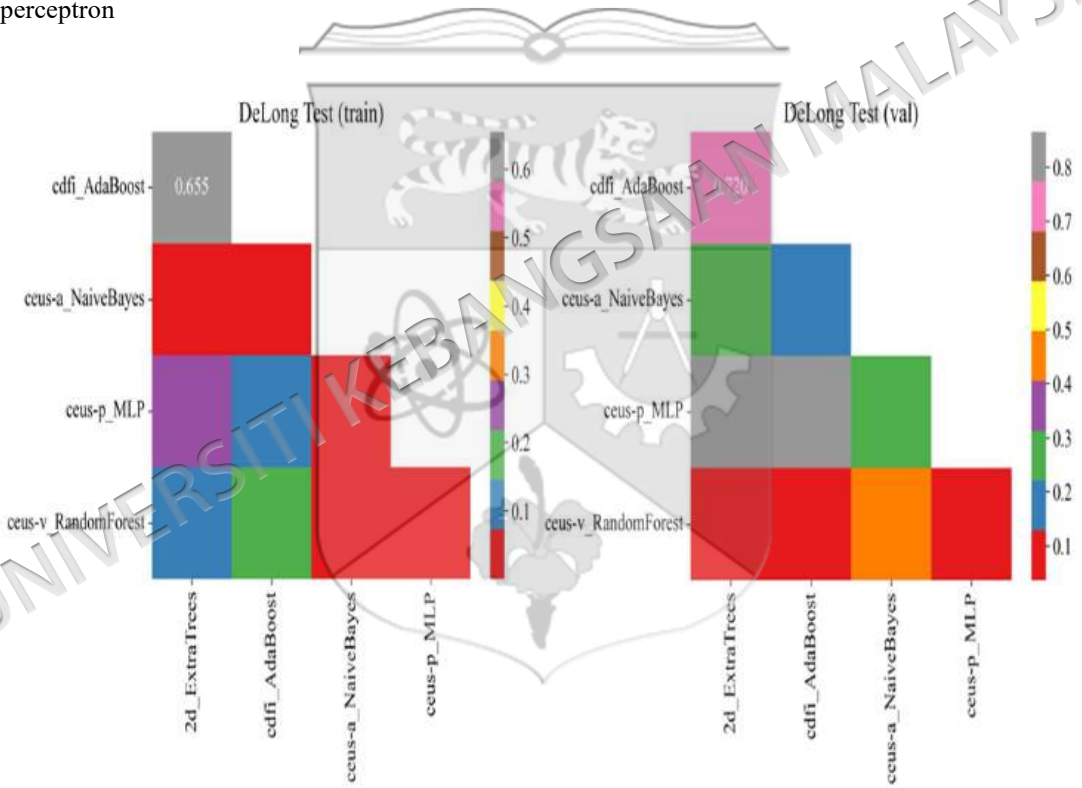


Figure 4.57 Pairwise DeLong comparisons of the AUCs of the five selected radiomics models in the training and test cohorts

Heatmaps of pairwise DeLong test p-values for the best-performing models from the five imaging modalities are shown for the training cohort (left) and test cohort (right). Each cell represents a pairwise comparison of ROC-AUCs, with $p < 0.05$ indicating a statistically significant difference.

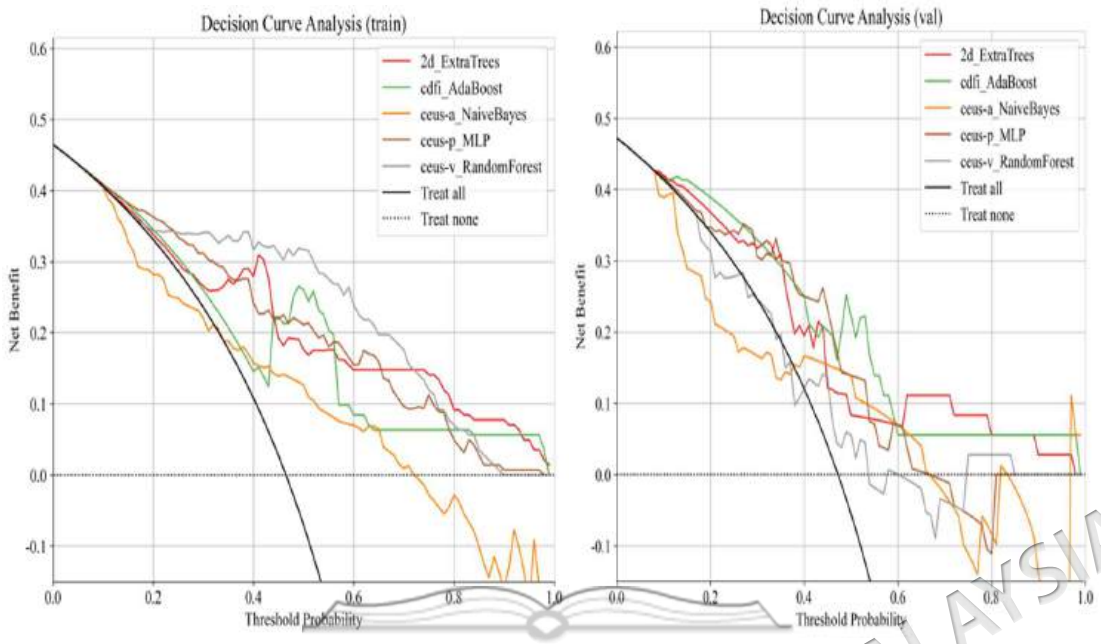


Figure 4.58 Decision curve analysis of the five selected radiomics models in the training and test cohorts

Decision curves for the best-performing models from the five imaging modalities are shown for the training cohort (left) and test cohort (right). The curves compare net benefit across a range of threshold probabilities, with the treat-all and treat-none strategies included as references.

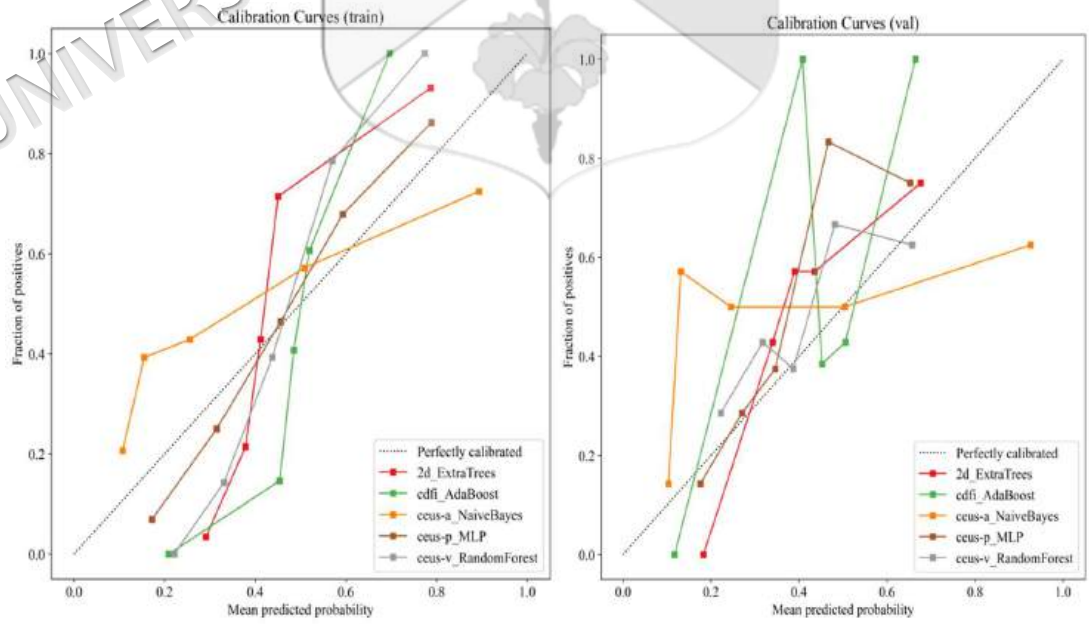


Figure 4.59 Calibration curves of the best-performing radiomics models in the training and test cohorts

The dotted diagonal line represents perfect calibration. Curves above and below the line indicate underestimation and overestimation of metastatic risk, respectively. Calibration curves are shown for the training cohort (left) and test cohort (right).

CHAPTER V

DISCUSSION

5.1 OVERVIEW AND PRINCIPAL FINDINGS

5.1.1 Purpose and Overall Interpretation of the Findings

This chapter will interpret the principal findings of the study in the context of existing evidence, with particular attention to their clinical and methodological implications. As the detailed numerical results have already been presented in Chapter IV, only the key findings necessary for the subsequent discussion are reiterated here. The discussion is organised around the complementary roles of conventional ultrasound, Sonazoid contrast-enhanced ultrasound (CEUS), fine-needle aspiration thyroglobulin (FNA-Tg), multimodal integration, and radiomics in the preoperative assessment of lateral cervical lymph node metastasis (LLNM) in papillary thyroid carcinoma (PTC).

The central finding is that diagnostic performance improved as the assessment moved from subjective morphology to structured quantitative analysis and then to multimodal integration. Conventional ultrasound remained valuable for initial detection and anatomical localisation, but its specificity was limited when interpretation depended mainly on visual pattern recognition. Quantitative scoring improved discrimination by combining several suspicious features. Sonazoid CEUS added functional and microenvironmental information, with the post-vascular phase providing the most distinctive contribution. FNA-Tg supplied biochemical evidence of thyroid tissue within the sampled node. When these sources were integrated, the highest observed diagnostic performance was achieved. The radiomics results extended this interpretation. The performance of machine-learning models depended more on the biological and technical quality of the source images than on classifier complexity alone.

Post-vascular-phase Sonazoid CEUS produced the strongest radiomics model in the test cohort, whereas venous-phase images produced the weakest. This pattern supports a biologically informed view of radiomics: algorithms can only extract clinically useful signatures when the underlying image contains stable and disease-relevant information.

5.1.2 Clinical and Biological Context of the Baseline Findings

Patients in the metastasis group were younger than those in the non-metastasis group. This finding is consistent with the recognised paradox in differentiated thyroid cancer whereby younger patients may show more frequent nodal dissemination while retaining favourable disease-specific survival. Age should therefore not be interpreted as a direct marker of poor prognosis; rather, in the present cohort, it functioned as a contextual risk indicator for lymphatic spread. Its clinical value is likely to be greatest when combined with imaging and biochemical evidence rather than used as an isolated selection criterion.

The short-axis diameter, but not the long-axis diameter or long-to-short-axis ratio, differed significantly between groups. This suggests that metastatic infiltration may be reflected more reliably by nodal thickening than by overall length. Lateral cervical nodes are frequently elongated even when benign, and their shape can be altered by anatomical constraints. Short-axis enlargement may therefore be a more direct expression of cortical replacement and internal tumour burden. Nevertheless, size alone cannot distinguish early metastasis from reactive hyperplasia, which explains why morphometric variables should remain part of a composite assessment.

Higher serum Tg, TgAb, and TPO-Ab levels were observed in the metastasis group. These associations should be interpreted cautiously because serum markers are influenced by the primary tumour, background thyroid tissue, autoimmune thyroid disease, and assay characteristics. They do not establish a causal immunological mechanism for LLNM. Their value in this study is primarily contextual: they indicate that metastatic status was accompanied by a different biochemical profile, but they were not sufficiently specific to replace node-directed tests such as FNA-Tg.

5.1.3 Hierarchy of Diagnostic Evidence

Taken together, the findings support a hierarchy of diagnostic evidence. Conventional ultrasound identifies and localises suspicious nodes; structured ultrasound scoring reduces subjectivity; Sonazoid CEUS characterises perfusion and post-vascular retention; FNA-Tg confirms thyroid-derived biochemical material; and radiomics quantifies image heterogeneity beyond visual perception. This hierarchy does not imply that every patient requires every test. It provides a rational escalation pathway in which additional modalities are used when the expected clinical benefit exceeds the burden, cost, and invasiveness of further investigation.

The observed AUC of 0.993 for the three-modality combination was the highest among the evaluated strategies. Pairwise DeLong testing in the latest analysis showed that its differences from US + CEUS, CEUS + FNA-Tg, and US + FNA-Tg were not statistically significant. Accordingly, the three-modality model should be described as having the highest observed discrimination rather than being statistically superior to the high-performing two-modality alternatives. This distinction is clinically important because a small numerical gain in AUC may not justify an additional invasive or resource-intensive procedure when it is unlikely to alter management.

5.2 CONVENTIONAL ULTRASOUND: ESSENTIAL FIRST-LINE ASSESSMENT WITH A DIAGNOSTIC CEILING

5.2.1 Interpretation of Qualitative Ultrasound Findings

Qualitative conventional ultrasound achieved high sensitivity but substantially lower specificity. This profile is clinically understandable. Sonographers are generally trained to avoid missing metastatic disease, and suspicious features such as loss of the hilum, cortical abnormality, peripheral vascularity, hyperechogenicity, cystic change, and calcification appropriately trigger further investigation. The cost of this sensitive strategy is false-positive classification, particularly when reactive nodes demonstrate cortical thickening or altered vascularity.

The occurrence of hyperechogenicity, cystic change, and calcification only in metastatic nodes in this cohort supports their high specificity, but their absence cannot exclude metastasis. Early or partial nodal involvement may not produce these classic signs, and metastatic foci may be missed when only a limited portion of the node is affected. Thus, qualitative ultrasound is most useful as a screening and triage method rather than a definitive rule-in or rule-out test.

These findings broadly agree with previous studies and guideline-based practice, which recognise conventional ultrasound as the first-line examination for cervical nodal staging while also acknowledging substantial operator dependence and overlap between benign and malignant appearances (Ahuja and Ying 2005; Sohn et al. 2010; Haugen et al. 2016). The present study adds compartment-specific evidence by focusing on lateral cervical nodes, where anatomical depth, heterogeneous metastatic replacement, and reactive changes make purely visual diagnosis especially difficult.

5.2.2 Added Value of Quantitative Ultrasound Scoring

Quantitative scoring improved the AUC from 0.842 to 0.914 and markedly increased specificity, although sensitivity decreased slightly. The important implication is not simply that a numerical score performed better, but that structured integration changed the balance of diagnostic errors. Qualitative assessment favoured sensitivity and produced more false positives; quantitative scoring imposed a reproducible threshold and reduced unnecessary classification of reactive nodes as metastatic.

The improvement probably arose from aggregation. No single ultrasound sign is sufficiently reliable across all lymph nodes, but several moderately informative features can form a stronger composite indicator when weighted together. The scoring system therefore transformed ultrasound from an impression-based assessment into a more standardised risk estimate. This is consistent with the broader movement toward structured reporting and quantitative decision support in medical imaging (Lambin et al. 2012; Gillies et al. 2016).

Nevertheless, quantitative scoring cannot overcome limitations inherent to grayscale and Doppler imaging. It still depends on acquisition quality, machine settings, operator technique, and the visibility of macroscopic changes. Reactive hyperplasia and early metastatic infiltration may continue to overlap. The score should therefore be viewed as a refinement of conventional ultrasound rather than a substitute for functional or biochemical assessment.

This advantage likely arises because quantitative scoring combines several moderate predictors into a stronger composite indicator. Individual ultrasound features such as nodal size, shape, margin irregularity, hilum loss, abnormal echogenicity, and peripheral vascularity are each informative but not completely reliable in isolation (Ahuja and Ying 2005; Sohn et al. 2010; Tian et al. 2022). When these findings are weighted and aggregated, their combined diagnostic value becomes more stable and reproducible. In the present study, this cumulative approach appeared to provide a more balanced assessment of lymph node status than qualitative pattern recognition alone.

The benefit of quantitative analysis is particularly relevant in the lateral cervical compartment, where metastatic involvement is often partial, heterogeneous, and anatomically complex. Early metastatic infiltration may affect only limited cortical regions and may not yet produce overt architectural distortion. Under such circumstances, a structured scoring framework may be more sensitive to the cumulative presence of subtle abnormalities than an unstructured visual assessment. This is clinically important because lateral cervical lymph nodes often display overlapping benign and malignant features, and purely morphology-based interpretation may therefore result in diagnostic inconsistency (Sohn et al. 2010; Filetti et al. 2019).

In addition, the incorporation of color Doppler flow imaging into a quantitative framework provides incremental value by standardizing the assessment of vascular patterns. Abnormal peripheral vascularity and loss of normal hilar flow have long been regarded as suspicious signs of nodal metastasis in thyroid cancer(Ahuja and Ying 2005; Sohn et al. 2010). Expressing these Doppler findings within a scoring system reduces reliance on subjective visual grading and allows vascular information to contribute more systematically to overall risk estimation.

However, despite these advantages, quantitative conventional ultrasound remains constrained by the intrinsic limitations of grayscale and Doppler imaging. Grayscale ultrasound primarily reflects macroscopic structural characteristics, whereas Doppler techniques depict only flow signals above a certain velocity and vessel calibre threshold. Consequently, early metastatic infiltration may not yet produce sufficient architectural disruption or detectable vascular alteration to be reliably captured, even within a quantitative framework. In addition, Doppler signals remain susceptible to technical factors such as insonation angle, gain settings, motion artefact, and machine optimization, which may still introduce variability into the final score (Ahuja and Ying 2005; Haugen et al. 2016).

Another important limitation is the biological overlap between benign and malignant lymph nodes. Reactive lymph nodes may exhibit cortical thickening or increased vascularity due to inflammatory hyperplasia, thereby mimicking metastatic involvement. Quantitative scoring can reduce subjectivity, but it cannot eliminate the overlap inherent to conventional ultrasound signs. This likely explains why diagnostic performance, although improved, still did not reach the level achieved by Sonazoid CEUS, FNA-Tg, or multimodal integration in the present study.

5.2.3 Clinical Positioning of Conventional Ultrasound

Clinically, conventional ultrasound should remain the entry point to the diagnostic pathway because it is non-invasive, widely available, real-time, and suitable for mapping multiple cervical levels. Its most appropriate roles are localisation, initial risk stratification, selection of a target for aspiration, and guidance of subsequent procedures. A node with overtly malignant morphology may proceed directly to confirmatory sampling, whereas an indeterminate node may benefit from quantitative scoring and Sonazoid CEUS before an invasive test is performed.

The present findings therefore support neither abandonment of qualitative expertise nor exclusive reliance on a numerical score. Expert pattern recognition remains necessary for acquisition, localisation, and detection of uncommon appearances, while quantitative scoring provides consistency and reduces threshold

variation. The most defensible approach is complementary: qualitative examination identifies the relevant biological pattern, and quantitative assessment standardises its diagnostic weight.

Despite continuous technical refinement, quantitative scoring, and the integration of qualitative and quantitative assessment, conventional ultrasound ultimately reaches a diagnostic ceiling in the evaluation of lateral cervical lymph node metastasis. This limitation is not primarily due to operator inexperience or methodological weakness, but rather to the intrinsic constraints of morphology-based and macroscopic vascular imaging. Conventional ultrasound is fundamentally limited in its ability to characterize the microstructural, microvascular, and biochemical alterations associated with early or heterogeneous metastatic disease (Lambin et al. 2012; Gillies et al. 2016).

Conventional ultrasound primarily evaluates structural characteristics such as nodal size, shape, echogenicity, and gross architectural change (Ahuja and Ying 2005; Kim et al. 2015). However, metastatic involvement in papillary thyroid carcinoma is often biologically subtle, particularly in early or partial disease. Micro-metastases or focal cortical infiltration may not produce sufficiently obvious morphological abnormalities to be detected reliably by conventional ultrasound (Kessler et al. 2003; Surov et al. 2016). In addition, there is substantial biological overlap between benign and malignant lymph nodes. Reactive lymph nodes may demonstrate enlargement, cortical thickening, and increased vascularity in response to inflammation, thereby mimicking metastatic disease (Filetti et al. 2019; Shin et al. 2016). Even when quantitative thresholds are applied, distinguishing reactive hyperplasia from early metastatic infiltration remains difficult because similar biological processes may generate similar sonographic appearances (Kim et al. 2017).

Another important reason for this diagnostic ceiling is that conventional ultrasound cannot adequately depict microenvironmental changes central to tumor metastasis, including microvascular perfusion, extracellular matrix alteration, immune infiltration, and tumor –stroma interaction. These processes are highly relevant to metastatic progression but lie beyond the detection capability of grayscale and Doppler

imaging (Chammas 2019). Furthermore, technical factors such as nodal depth, surrounding anatomy, patient habitus, and operator dependency may affect image quality and reduce diagnostic consistency(Ahuja and Ying 2005). Although quantitative analysis improves objectivity, it cannot fully overcome the inherent limitations of signal acquisition and tissue contrast.

Taken together, these findings indicate that conventional ultrasound reaches a performance plateau in lateral cervical lymph node assessment even after quantitative enhancement. This plateau highlights the need for diagnostic escalation beyond ultrasound alone. Importantly, recognition of this diagnostic ceiling does not diminish the clinical value of conventional ultrasound. Rather, it clarifies its appropriate role within a broader multimodal strategy and allows its strengths and limitations to be interpreted more rationally.

Within the multimodal diagnostic framework proposed in this thesis, conventional ultrasound remains foundational and indispensable. Rather than functioning as a definitive diagnostic tool for lateral cervical lymph node metastasis, it should be regarded as the entry point and structural basis of the preoperative evaluation pathway (Ahuja and Ying 2005; Haugen et al. 2016). In clinical practice, conventional ultrasound provides the first opportunity to assess the lateral cervical compartment systematically. Its real-time capability allows rapid evaluation of multiple nodal levels, identification of suspicious lymph nodes, and guidance for targeted fine-needle aspiration (Ahuja and Ying 2005; Lee et al. 2012). These practical advantages ensure its continued importance in the routine staging of papillary thyroid carcinoma.

The present study, therefore, supports repositioning conventional ultrasound as a screening and stratification modality rather than the final determinant of metastatic status. This view is consistent with contemporary risk-adapted management strategies (Haugen et al. 2016). Ultrasound findings guide further diagnostic escalation by identifying lymph nodes that require additional functional, biochemical, or quantitative assessment. In this way, ultrasound determines not only which lymph nodes should undergo further investigation, but also how complementary diagnostic resources should be deployed. Lymph nodes with clearly benign sonographic features may be followed

conservatively, whereas suspicious or indeterminate nodes may benefit from further evaluation using contrast-enhanced ultrasound, fine-needle aspiration thyroglobulin (FNA-Tg), or advanced quantitative and radiomics-based methods (Park et al. 2019; Song et al. 2015).

5.3 FNA-TG: STRONG BIOCHEMICAL CONFIRMATION WITH IMPORTANT CONTEXTUAL LIMITATIONS

5.3.1 Interpretation of the Diagnostic Performance of FNA-Tg

FNA-Tg showed excellent diagnostic performance, with quantitative analysis achieving an AUC of 0.953. Its biological rationale is strong: thyroglobulin detected in needle washout fluid provides evidence of thyroid-derived material within a sampled lymph node. This is particularly valuable when cytology is non-diagnostic because of scant cellularity, cystic degeneration, or sampling of a non-representative area.

The difference between quantitative and qualitative interpretation was statistically significant but modest. Qualitative classification retained higher sensitivity, whereas the quantitative cut-off improved specificity. This resembles the pattern observed with ultrasound and indicates that dichotomous clinical rules and optimized numerical thresholds serve different purposes. A sensitive qualitative rule is useful for avoiding missed metastasis; a quantitative threshold is useful when the clinical priority is to reduce false-positive diagnoses and unnecessary neck dissection.

The present cut-off of 9.10 ng/mL should not be treated as universally transferable. Published thresholds vary widely because of differences in assay platforms, washout volume, residual thyroid tissue, serum Tg concentration, anti-thyroglobulin antibodies, node composition, and patient selection. The study therefore provides a cohort-specific operating threshold rather than a universal biological boundary. External calibration is required before direct adoption in another institution.

This continuous-variable approach is particularly important in the lateral cervical compartment, where metastatic involvement is often focal, partial, and heterogeneous. Quantitative interpretation is therefore better aligned with the biological

continuum of nodal metastasis than binary categorization. Reviews and meta-analyses have similarly concluded that FNA-Tg has high overall diagnostic accuracy, while also emphasizing that numerical values and optimal cut-offs should be interpreted within the assay and clinical context (Al-Hilli et al. 2017; Khadra et al. 2019).

Another important strength of quantitative interpretation is its compatibility with advanced predictive modelling. Because the variable is continuous, it can be incorporated directly into multivariate analyses, nomograms, and multimodal diagnostic frameworks together with imaging-based quantitative parameters. This makes quantitative FNA-Tg particularly suitable for the objective and measurement-based diagnostic strategy emphasized in the present study. Broader imaging and biomarker research has similarly shown the methodological advantages of retaining continuous variables rather than collapsing them into binary groups when building predictive models (Jeon et al. 2015; Qiang et al. 2016).

Nevertheless, quantitative interpretation is also influenced by factors such as assay variation, thyroid gland status, and cut-off selection. For this reason, it should still be interpreted within the broader clinical and imaging context rather than in isolation. Published studies have also noted the importance of preanalytical standardization and awareness of factors such as serum TgAb and gland status when interpreting FNA-Tg results (Bournaud et al. 2010; Jeon et al. 2013). Even so, the findings of the present study indicate that quantitative FNA-Tg provides strong diagnostic discrimination and represents a more refined analytical approach than threshold-based categorization alone.

Despite the well-established diagnostic value of fine-needle aspiration thyroglobulin (FNA-Tg), the optimal cut-off value for defining a positive FNA-Tg result remains one of the most controversial issues in clinical practice, particularly in the evaluation of lateral cervical lymph nodes. Wide variation in proposed thresholds across studies has limited standardization and complicated interpretation of FNA-Tg results (Qiang et al. 2016; Wang et al. 2022).

One major contributor to this variability is assay-dependent differences in thyroglobulin measurement. Different immunoassays exhibit varying analytical sensitivity, calibration standards, and susceptibility to interference, resulting in non-

comparable absolute values across institutions (Jeon et al. 2013). As a result, cut-off values derived in one laboratory setting may not be directly applicable elsewhere.

Another important factor is patient-related heterogeneity, particularly with respect to thyroid status and prior treatment. In patients who have undergone thyroidectomy, baseline thyroglobulin levels in lymph node aspirates are expected to be negligible, allowing lower cut-off values to retain high specificity. Conversely, in preoperative patients with intact thyroid tissue, background thyroglobulin contamination may elevate aspirate levels, necessitating higher thresholds to avoid false-positive interpretation (Duval et al. 2017).

The presence of anti-thyroglobulin antibodies further complicates cut-off determination. Although FNA-Tg is less affected by antibodies than serum thyroglobulin, interference cannot be completely excluded and may contribute to underestimation in selected cases (Shin et al. 2015). This issue reinforces the need for cautious interpretation rather than rigid reliance on a single numerical threshold.

Several authors have therefore advocated context-dependent or relative cut-off strategies, such as comparing aspirate thyroglobulin levels with serum thyroglobulin or using institution-specific thresholds derived from local validation cohorts (Bournaud et al. 2010; Onoue et al. 2021; Yu et al. 2025). While these approaches improve adaptability, they reduce comparability across studies and limit guideline-level standardization.

The findings of the present study provide further evidence that cut-off selection should be interpreted within a specific clinical and analytical context. In this cohort, when the cut-off value of FNA-Tg was set at 9.10 ng/mL, the sensitivity and specificity for diagnosing lateral cervical lymph node metastasis were 0.880 and 0.968, respectively. The Youden index was 0.848, and the ROC-AUC reached 0.953 (95% CI: 0.917–0.990), indicating excellent diagnostic performance. This result suggests that, in the setting of lateral cervical lymph node evaluation, a threshold within the higher range reported in previous studies may offer an appropriate balance between avoiding false-positive overdiagnoses and maintaining high diagnostic accuracy.

At the same time, the present study does not imply that 9.10 ng/mL should be regarded as a universal threshold applicable to all institutions and patient populations. The optimal operating point for FNA-Tg is likely to remain influenced by assay platform, specimen handling, patient thyroid status, and study population characteristics. Therefore, the value identified in this study should be interpreted as a locally validated threshold with strong relevance to the present cohort, rather than as an absolute standard.

Importantly, controversies regarding FNA-Tg cut-off values become most relevant in equivocal cases. Extremely high FNA-Tg values are usually strongly suggestive of metastatic involvement, whereas very low values reliably exclude disease. The diagnostic uncertainty lies mainly in intermediate ranges, where numerical interpretation may be influenced by biological heterogeneity, blood contamination, and assay-related variability (Kuo et al. 2025). In such situations, FNA-Tg should not be interpreted in isolation, but rather in conjunction with ultrasound findings and overall clinical context (Haugen et al. 2016).

5.3.2 Discrepancies with Cytology and Other Evidence

FNA-C showed only moderate agreement with the final reference standard, whereas FNA-Tg performed substantially better. This discrepancy is plausibly explained by sampling and cellularity. Cytology requires representative malignant cells, but metastatic material may be focal, cystic, necrotic, or diluted (Bakuła-Zalewska et al. 2021). FNA-Tg can remain positive even when the aspirate contains few diagnostic cells because the marker is measured in the fluid phase. The two tests therefore interrogate different aspects of the sample and should be regarded as complementary rather than interchangeable.

False-positive FNA-Tg results remain possible, especially when the needle traverses thyroid tissue, when blood contamination is substantial, or when benign thyroid inclusions are present. False-negative results may occur in poorly differentiated tumor tissue, very small metastatic deposits, inadequate aspiration, or assay interference. These limitations reinforce the need to interpret FNA-Tg together with ultrasound localization, CEUS behavior, cytology, and the clinical context.

Qualitative FNA-Tg is particularly useful in lateral cervical lymph nodes with cystic change, necrosis, or low cellularity, where cytological specimens may be inadequate or non-diagnostic. In such cases, the presence of thyroglobulin in aspirated washout fluid provides biochemical evidence of metastatic thyroid tissue even when malignant cells are not clearly identified cytologically. This diagnostic advantage has been repeatedly described in studies evaluating FNA-Tg in suspicious cervical lymph nodes and in reviews summarizing its added value over cytology alone (Zhang et al. 2022; Junsong et al. 2025).

However, qualitative interpretation also has limitations. Its threshold-based nature reduces a continuous biochemical signal to a binary classification and therefore may obscure biologically meaningful variation. This limitation is especially relevant in lymph nodes with intermediate FNA-Tg values, where interpretation may depend on the FNA-Tg/serum Tg ratio and patient-specific biochemical context. Published evidence has shown that lack of standardization in assay methods, preanalytical handling, and cut-off selection remains an important source of heterogeneity in qualitative FNA-Tg interpretation (Xu et al. 2019; Kim et al. 2021).

5.3.3 Clinical Role of FNA-Tg

FNA-Tg is most valuable when the result is expected to change management. It is particularly helpful for nodes with suspicious or indeterminate imaging, discordant ultrasound and cytology, or cystic morphology. Routine aspiration of every sonographically visible lateral cervical node would be neither practical nor necessary. A risk-adapted strategy can reserve FNA-Tg for nodes in which biochemical confirmation is likely to influence the decision for lateral neck dissection.

The excellent performance of FNA-Tg in this study should therefore be interpreted as evidence for targeted biochemical confirmation, not as support for replacing imaging. FNA-Tg cannot map the lateral neck, characterise perfusion, identify the optimal target, or assess intranodal heterogeneity. Those functions remain dependent on imaging, particularly ultrasound and CEUS.

5.4 SONAZOID CEUS: FUNCTIONAL AND MICROENVIRONMENTAL CHARACTERISATION

5.4.1 Why Sonazoid Added Diagnostic Information

Sonazoid CEUS improved discrimination beyond conventional ultrasound because it depicts biological processes that are not visible on grayscale morphology alone. The arterial and venous phases reflect vascular arrival, distribution, and washout, while the post-vascular phase is additionally influenced by microbubble retention and macrophage-related uptake. Metastatic replacement can disrupt normal nodal microarchitecture and reticuloendothelial activity, producing enhancement patterns that differ from those of reactive nodes.

This mechanism distinguishes Sonazoid from purely blood-pool agents and provides a plausible explanation for the strong performance of post-vascular imaging. The post-vascular phase is less dependent on the exact timing of the initial bolus and may provide a more stable representation of tissue-level integrity. In contrast, arterial- and venous-phase appearances are more vulnerable to injection technique, circulatory variation, and transient inflammatory hyperaemia.

5.4.2 Sonazoid CEUS in Lateral Cervical Lymph Nodes

Sonazoid contrast-enhanced ultrasound offers a distinct biological and diagnostic advantage over conventional blood-pool agents because of its unique micro-perfusion behavior and post-vascular phase characteristics. Unlike agents such as SonoVue, which remain confined to the intravascular space, Sonazoid can provide not only vascular-phase information but also delayed tissue-related imaging associated with reticuloendothelial uptake, particularly by macrophages. This feature allows Sonazoid to reflect not only blood flow but also tissue integrity and microenvironmental change (Zhang et al. 2023).

This mechanism is particularly relevant to lateral cervical lymph node metastasis. Metastatic infiltration disrupts normal nodal architecture, alters vascular organization, replaces native lymphoid tissue, and modifies the immune

microenvironment. These changes influence not only early perfusion patterns but also delayed contrast distribution and retention. Benign nodes are therefore more likely to demonstrate preserved and relatively regular enhancement behavior, whereas metastatic nodes more often show heterogeneous enhancement, focal defects, irregular retention, or loss of normal contrast distribution (Wei et al. 2021; Wei et al. 2022).

An additional advantage of Sonazoid is its longer and more stable imaging window. Because assessment is not restricted to a brief intravascular phase, Sonazoid allows observation of both early vascular enhancement and delayed post-vascular behavior. This dual-phase capability is especially valuable in lateral cervical lymph nodes, where diagnoses often depend on subtle differences in tissue organization and microenvironment rather than on marked morphological abnormalities alone (Wakisaka et al. 2019).

For the present study, the use of Sonazoid was therefore conceptually central. The aim was not merely to visualize transient perfusion, but to determine whether a contrast agent capable of depicting both microvascular flow and post-vascular tissue behavior could improve the detection and characterization of lateral cervical lymph node metastasis. In this setting, Sonazoid provides a biologically richer imaging signal than conventional blood-pool CEUS and aligns more closely with the need for multimodal, functionally informed assessment in papillary thyroid carcinoma (Xiao et al. 2023).

The findings of the present study support this rationale. Sonazoid-enhanced imaging provided diagnostically meaningful information for lateral cervical lymph node assessment, and the post-vascular phase showed particular value in reflecting metastasis-related tissue alteration. These observations suggest that the advantage of Sonazoid lies not simply in being another contrast agent, but in offering a different imaging perspective that integrates perfusion assessment with tissue integrity and microenvironmental change. This distinction is also important for subsequent quantitative and radiomics analysis, because Sonazoid provides richer phase-specific information and a more suitable basis for objective feature extraction than purely transient blood-pool enhancement.

In summary, Sonazoid CEUS is particularly well suited to the evaluation of lateral cervical lymph node metastasis because it combines vascular-phase perfusion imaging with post-vascular assessment of tissue-level contrast behavior. This dual biological basis distinguishes it from conventional blood-pool agents and provides a strong mechanistic rationale for its use in the present study. On this basis, the following sections further examine the qualitative, quantitative, and radiomics applications of Sonazoid CEUS in lateral cervical lymph node assessment (Sidhu et al. 2018; Wei et al. 2021).

5.4.3 Interpretation Across CEUS Phases

Arterial-phase quantitative assessment showed excellent discrimination. Its strength lies in identifying abnormal perfusion architecture, including centripetal, chaotic, absent, or ring-like enhancement patterns. However, arterial enhancement can be influenced by benign hypervascularity and by technical timing, so it should not be interpreted independently of morphology and later-phase behavior.

The venous phase was less discriminative, particularly in the radiomics analysis. This may reflect its transitional nature: by this stage, contrast distribution can become more homogeneous and differences between reactive and metastatic nodes may narrow. Washout dynamics can also vary with inflammation, fibrosis, necrosis, and vascular shunting. The relatively weak venous-phase performance is therefore biologically plausible rather than merely an algorithmic failure.

Post-vascular-phase assessment provided the most consistent advantage. The persistence or defect of enhancement during this phase likely captures loss of normal nodal tissue and altered macrophage-rich microenvironments. This phase also produced the best radiomics model, strengthening the interpretation that its diagnostic value is embedded in both visually recognizable patterns and high-dimensional texture information.

Despite its strong performance, arterial-phase imaging still has limitations. Because it reflects an early hemodynamic state, it remains susceptible to physiological and technical variation, including injection technique, circulatory status, and timing of

acquisition. In addition, reactive or inflammatory lymph nodes may also demonstrate increased arterial perfusion, creating overlap with metastatic nodes. Arterial-phase imaging, therefore, mainly reflects perfusion behavior rather than stable tissue characteristics and cannot fully capture delayed tissue retention or broader microenvironmental change. These limitations support multiphase interpretation, particularly in a Sonazoid-based framework where later phases may provide additional tissue-level information (Sidhu et al. 2018; Wei et al. 2021).

In summary, the present study demonstrates that arterial-phase Sonazoid CEUS has excellent diagnostic value for lateral cervical lymph node metastasis, especially when enhancement patterns are interpreted quantitatively. However, because arterial-phase findings remain influenced by hemodynamic variability and provide only limited information on stable tissue characteristics, they should be interpreted within a multiphase and multimodal framework rather than in isolation (Sidhu et al. 2018; Chen et al. 2020).

At a cut-off value of 1.5, quantitative venous-phase analysis yielded a sensitivity of 0.880, specificity of 0.916, a Youden index of 0.796, and an AUC of 0.904 (95% CI: 0.855–0.954). Qualitative assessment showed a similar AUC of 0.898, and the difference between the two methods was not statistically significant ($p = 0.3416$). These findings suggest that, although venous-phase imaging can distinguish metastatic from non-metastatic nodes to some extent, its incremental value is limited, and quantitative scoring does not provide the same clear advantage as in the arterial phase. Similar variability in CEUS performance has been reported in pooled analyses of papillary thyroid carcinoma lymph node metastasis (Yu et al. 2023).

This limited discrimination is biologically plausible. Venous-phase enhancement remains influenced by systemic and local hemodynamic factors, and reactive or inflammatory lymph nodes may also show delayed or heterogeneous contrast behavior, creating overlap with metastatic nodes (Sidhu et al. 2018; Yu et al. 2023). In addition, the venous phase does not fully exploit the unique biological advantage of Sonazoid, because microbubbles are still largely confined to the vascular compartment, and tissue-retentive information has not yet become dominant. As a result, venous-

phase imaging provides only limited insight into tissue integrity and microenvironmental alteration, which are more directly relevant to metastatic infiltration (Xiao et al. 2023).

Quantitative analysis further improved discrimination. At a cut-off value of 3.5, quantitative post-vascular phase Sonazoid CEUS achieved a sensitivity of 0.843, specificity of 0.937, a Youden index of 0.780, and an ROC-AUC of 0.951 (95% CI: 0.908–0.978). By comparison, qualitative assessment showed a sensitivity of 0.940, specificity of 0.863, and an AUC of 0.901. The significantly higher AUC of the quantitative method ($p = 0.00026$) indicates that structured scoring of post-vascular contrast retention provides more effective discrimination than visual pattern recognition alone.

The biological basis of this advantage is closely related to the mechanism of Sonazoid. In later phases, Sonazoid is associated with reticuloendothelial uptake and tissue-related contrast behavior, so enhancement patterns are more closely linked to macrophage distribution, tissue integrity, and preservation of nodal architecture than to blood flow alone (Sidhu et al. 2018; Zhang et al. 2023). Benign or reactive lymph nodes are therefore more likely to retain relatively homogeneous enhancement, whereas metastatic lymph nodes, in which tumor infiltration disrupts native lymphoid tissue and microenvironmental structure, more often show reduced, heterogeneous, or absent enhancement. This principle is consistent with studies of Sonazoid-based cervical lymph node imaging in papillary thyroid carcinoma, which found that lymphatic or tissue-related contrast behavior can improve identification of metastatic nodes (Wei et al. 2022; Chung et al. 2022).

An additional strength of the post-vascular phase is its greater temporal stability. Because imaging is no longer confined to the brief early vascular window, assessment is less dependent on precise timing and immediate operator response. This improves reproducibility and allows more deliberate evaluation of intranodal heterogeneity, including focal filling defects and partial metastatic replacement. Such features are particularly relevant in lateral cervical lymph nodes, where metastatic involvement is often incomplete and spatially heterogeneous.

From the perspective of this thesis, the post-vascular phase also provides a more suitable substrate for quantitative analysis and radiomics. Because contrast distribution is more stable and biologically meaningful than in the earlier phases, extracted features are more likely to reflect genuine tissue alteration rather than transient hemodynamic fluctuation. This helps explain why post-vascular phase Sonazoid CEUS showed strong diagnostic performance in the present study and why it is conceptually central to advanced quantitative modelling.

5.4.4 Comparison with Conventional Ultrasound and Prior Evidence

Overall Sonazoid CEUS outperformed conventional ultrasound in both qualitative and quantitative analyses. The improvement was greatest in specificity, which is clinically important because the major limitation of conventional ultrasound in this cohort was false-positive classification. Sonazoid CEUS therefore appears to refine rather than duplicate morphological assessment: it helps distinguish nodes that look suspicious because of reactive structural change from those showing functional and microenvironmental disruption compatible with metastasis.

Direct comparison with earlier CEUS studies requires caution because many investigations used blood-pool contrast agents, different phase definitions, different qualitative criteria, or mixed nodal populations. The present study contributes specific evidence for Sonazoid and for lateral cervical lymph nodes in PTC. Its findings should not be extrapolated automatically to other contrast agents, primary tumors, or nodal compartments.

5.4.5 Practical Constraints

Despite its diagnostic value, Sonazoid CEUS requires additional time, contrast administration, trained operators, appropriate equipment, and consistent acquisition protocols. Interpretation can still be affected by motion, small node size, depth, probe pressure, and selection of the imaging plane. In institutions without Sonazoid availability or CEUS expertise, quantitative conventional ultrasound and targeted FNA-Tg may remain the more feasible pathway.

Accordingly, Sonazoid CEUS is best positioned as a problem-solving and risk-refining test for indeterminate or clinically consequential nodes, rather than an obligatory examination for every patient. Its use should be guided by whether the additional information is likely to alter sampling, surgical planning, or surveillance.

5.5 MULTIMODAL INTEGRATION AND DISCREPANT FINDINGS

5.5.1 Why Integration Outperformed Single-Modality Assessment

The superior performance of multimodal combinations is explained by complementarity rather than simple accumulation of tests. Conventional ultrasound provides anatomy and morphology; Sonazoid CEUS provides perfusion and post-vascular tissue behavior; and FNA-Tg provides biochemical confirmation. Because each modality interrogates a different aspect of nodal pathology, concordant results reduce uncertainty more effectively than repeated measurements of the same biological domain.

The three-modality combination achieved the highest observed AUC, sensitivity, and specificity. However, several two-modality combinations also performed extremely well, and their confidence intervals overlapped. The clinical implication is that diagnostic escalation should be selective. When two non-invasive modalities are strongly concordant, the incremental benefit of aspiration may be small; when imaging is discordant or when surgery carries major consequences, FNA-Tg may provide decisive confirmation.

5.5.2 Diagnostic Synergy of Conventional Ultrasound, Sonazoid CEUS, and FNA-Tg

The present study strongly supports the concept that multimodal integration provides greater diagnostic accuracy than any individual modality. This improvement is not simply additive, but reflects genuine complementarity among conventional ultrasound, Sonazoid CEUS, and FNA-Tg, each of which interrogates a different aspect of lateral cervical lymph node pathology.

All combined approaches outperformed the corresponding single-modality strategies. The combination of US + FNA-Tg achieved a sensitivity of 0.928, a specificity of 1.000, a Youden index of 0.928, and an AUC of 0.983, indicating strong complementarity between morphology-based imaging and biochemical confirmation. The combination of CEUS + FNA-Tg yielded a sensitivity of 0.976, a specificity of 0.989, a Youden index of 0.965, and an AUC of 0.988, suggesting that functional contrast-enhanced imaging and biochemical evidence of thyroid tissue dissemination reinforce one another effectively. Similarly, US + CEUS achieved a sensitivity of 0.976, a specificity of 0.979, a Youden index of 0.955, and an AUC of 0.989, indicating that integration of morphology-based and contrast-enhanced functional imaging alone already provides near-optimal discrimination.

The full combination of US + CEUS + FNA-Tg produced the highest observed point estimates, with a sensitivity of 0.988, a specificity of 0.989, a Youden index of 0.977, and an AUC of 0.993. Nevertheless, Table 4.32 of Chapter IV shows that its AUC was not significantly higher than those of US + CEUS, CEUS + FNA-Tg, or US + FNA-Tg. The result should therefore be interpreted as evidence of excellent discrimination and biological complementarity, but not as proof that routine use of all three modalities is statistically superior to every two-modality strategy.

The synergy observed is not merely statistical, but conceptual. Conventional ultrasound identifies suspicious nodes anatomically, Sonazoid CEUS characterizes perfusion and tissue-level contrast behavior, and FNA-Tg confirms thyroid-origin material biochemically. Together, these modalities generate a multidimensional diagnostic profile that aligns more closely with the biological reality of metastatic lymph node involvement than any isolated test. This interpretation is consistent with previous evidence showing that CEUS combined with conventional ultrasound improves preoperative nodal evaluation in papillary thyroid carcinoma, while FNA-Tg adds high diagnostic accuracy as an adjunctive test (Zhang et al. 2020; Liu et al. 2022; Yu et al. 2023).

This synergy is particularly valuable in clinically challenging scenarios. In lymph nodes with equivocal morphological features, Sonazoid CEUS may reveal

suspicious functional patterns that justify aspiration and biochemical confirmation. Conversely, elevated FNA-Tg values in morphologically subtle nodes can be reinforced by CEUS findings, thereby increasing diagnostic confidence and reducing both false-negative and false-positive classification. Overall, the present study demonstrates that triple-modality integration of conventional ultrasound, Sonazoid CEUS, and FNA-Tg represents the most diagnostically robust strategy for preoperative evaluation of lateral cervical lymph node metastasis.

5.5.3 Statistical Interpretation of the Comparative ROC Analysis

Tables 4.31 and 4.32 of Chapter IV add an important inferential layer to the descriptive AUC rankings by distinguishing statistically supported improvements from merely higher observed point estimates. Because the ROC curves were derived from the same participants, the DeLong test appropriately accounted for the correlation between paired AUC estimates. This distinction is interpretaion: a numerically larger AUC does not by itself establish superiority, particularly when performance is already close to the upper boundary and confidence intervals overlap.

The quantitative approaches significantly outperformed their corresponding qualitative assessments for conventional ultrasound, FNA-Tg, and overall CEUS. The gain for conventional ultrasound (AUC 0.914 versus 0.842; $p < 0.001$) was driven mainly by the marked improvement in specificity obtained through structured scoring. This suggests that the principal contribution of quantification was not simply to detect more abnormal nodes, but to reduce false-positive classification caused by subjective overcalling of reactive or morphologically equivocal nodes. In clinical terms, this improvement is relevant because false-positive lateral nodal diagnoses may expose patients to unnecessary lateral neck dissection and its associated morbidity.

The improvement for quantitative FNA-Tg was smaller (AUC 0.953 versus 0.917; $p = 0.0456$). Although statistically significant, the p-value was close to the conventional threshold, and the absolute difference in AUC was modest. This finding should therefore be interpreted as refinement rather than transformation of diagnostic performance. Qualitative FNA-Tg criteria already performed strongly, and quantitative

thresholding added precision primarily in borderline values. The result also reinforces the need to judge statistical significance together with confidence intervals, assay variability, threshold transportability, and the clinical consequences of reclassifying a small number of cases.

The largest absolute gain from quantification was observed for overall Sonazoid CEUS (AUC 0.978 versus 0.910; p = 0.0001). This supports the view that multiphase enhancement contains diagnostically relevant information that cannot be fully captured by visual pattern labels alone. Quantitative integration can preserve graded differences in perfusion, enhancement distribution, washout, and post-vascular retention that would otherwise be compressed into broad qualitative categories. The magnitude of this improvement is biologically plausible because metastatic nodal involvement is heterogeneous and may produce partial rather than categorical disruption of vascular and reticuloendothelial behavior.

Table 4.30 of Chapter IV also demonstrates that integrated CEUS significantly outperformed integrated conventional ultrasound (AUC 0.979 versus 0.914; p = 0.0057). This comparison provides direct evidence that the benefit of CEUS was not merely due to adding another scoring system. Rather, contrast-enhanced imaging contributed a distinct functional domain beyond the morphology and Doppler vascularity available from conventional ultrasound. The result supports diagnostic escalation to Sonazoid CEUS when conventional ultrasound leaves clinically important uncertainty, especially for nodes lacking highly specific structural signs.

By contrast, the three-modality combination did not significantly outperform any of the high-performing two-modality combinations. The AUC gains over US + CEUS, CEUS + FNA-Tg, and US + FNA-Tg were only 0.004, 0.005, and 0.010, respectively, with DeLong p-values of 0.2849, 0.2769, and 0.2717. These findings illustrate a ceiling effect: once two biologically complementary modalities have already resolved most diagnostic uncertainty, the third modality may improve only a small number of classifications. The absence of statistical significance does not prove equivalence, but it does mean that superiority cannot be claimed from the present sample.

This ceiling effect has direct implications for clinical implementation. A three-modality pathway may still be justified when the cost of a false-negative or false-positive decision is particularly high, when two modalities are discordant, or when the additional test is likely to change the planned extent of surgery. Conversely, routine application of all three modalities to every suspicious node may yield limited additional discrimination while increasing procedure time, invasiveness, cost, and patient burden. The results therefore support a selective, stepwise strategy in which testing is escalated according to residual uncertainty rather than a uniform maximal-testing approach.

Finally, very high AUC values should not be interpreted as evidence that the combined diagnostic strategies are automatically ready for probability-based clinical decisions. Discrimination describes the ability to rank patients, whereas calibration describes the agreement between predicted and observed risk. Therefore, the multimodal models can be described as having excellent discrimination, while their absolute-risk calibration still requires prospective and external evaluation before probability thresholds are used to determine surgery.

5.5.4 Interpretation of Discordance

Discordant findings are not necessarily errors. A node may have suspicious morphology but benign post-vascular retention because structural change precedes extensive microenvironmental disruption. Conversely, a morphologically subtle node may show abnormal CEUS behavior because functional alteration occurs before obvious enlargement or architectural distortion. FNA-Tg may be positive when cytology is negative because biochemical material is detected despite inadequate cellular sampling.

A clinically useful multimodal framework should therefore treat discordance as information requiring explanation. The pattern of disagreement can suggest the most appropriate next step: repeat targeted aspiration, review of image acquisition, short-interval surveillance, or surgical consideration based on the totality of evidence. A rigid majority-vote approach would ignore the different biological meanings of the tests.

5.5.5 Multimodal Diagnoses as a Foundation for Risk-Based Stratification and Quantitative Modelling

The implications of multimodal diagnoses extend beyond improvement in diagnostic accuracy alone. More fundamentally, the present findings support a transition from binary diagnostic thinking toward probability-based risk stratification. Traditional workflows for lateral cervical lymph node metastasis have largely categorized nodes as either metastatic or non-metastatic. Although operationally convenient, this dichotomous framework inadequately reflects the biological continuum of metastatic progression and provides only limited support for nuanced clinical decision-making.

In reality, metastatic involvement of lateral cervical lymph nodes is rarely an all-or-none phenomenon. Metastasis may evolve gradually from microscopic tumor foci to partial infiltration and eventually extensive nodal replacement, accompanied by dynamic changes in vascularity, stromal interaction, and microenvironmental integrity (Murata 2018). Within this context, multimodal diagnoses offer an important conceptual advance. By integrating structural, functional, and biochemical information, it allows clinicians to estimate the probability of metastatic involvement rather than depending on a single threshold or isolated criterion. The progressive improvement observed from single-modality to dual-modality and finally triple-modality combinations in the present study strongly supports this model.

At the same time, the increasing complexity of multimodal data also exposes the limitations of purely qualitative integration. Even when multiple modalities are available, interpretation may still depend heavily on expert judgement to weigh structural, functional, and biochemical findings. This reliance on subjective synthesis may reduce reproducibility, particularly in borderline cases. For this reason, multimodal diagnoses naturally lead toward quantitative modelling approaches capable of integrating heterogeneous data more systematically.

Radiomics represents a logical extension of this framework. By transforming imaging data into high-dimensional quantitative features, radiomics can capture intranodal heterogeneity, spatial organization, and intensity variation beyond visual assessment. Ultrasound-based radiomics studies in papillary thyroid carcinoma have

shown that quantitative imaging models can predict both central and lateral cervical lymph node metastasis, supporting the view that multimodal diagnoses are a practical bridge toward more objective and reproducible risk modelling (Li et al. 2024).

In summary, the present study supports a coherent shift from isolated single-modality testing toward integrated, probability-based multimodal assessment of lateral cervical lymph node metastasis. This approach better reflects the biological reality of metastatic evolution, improves flexibility in equivocal cases, and provides the practical and conceptual foundation for subsequent quantitative modelling and radiomics-based prediction.

5.6 RADIOMICS: PROMISING QUANTITATIVE SUPPORT, NOT YET A REPLACEMENT FOR MULTIMODAL DIAGNOSIS

5.6.1 Modality-Dependent Model Performance

Radiomics performance varied substantially by source image. The best test-cohort model was the multilayer perceptron based on post-vascular-phase Sonazoid CEUS, with an AUC of 0.882. This exceeded the best models from grayscale ultrasound, color Doppler flow imaging, arterial-phase CEUS, and venous-phase CEUS. The ranking supports the interpretation that biologically stable and information-rich images provide a stronger substrate for quantitative modelling.

The relatively good performance of grayscale and color Doppler radiomics indicates that conventional images contain useful heterogeneity beyond visual assessment. Nevertheless, their test performance remained below that of the best post-vascular model, probably because grayscale morphology and Doppler flow are more sensitive to acquisition angle, gain, compression, vessel detectability, and operator technique.

The weaker venous-phase model is an important negative finding. It demonstrates that adding a contrast-enhanced phase does not automatically improve radiomics. When the biological signal is transient, overlapping, or technically variable, high-dimensional feature extraction may amplify noise rather than reveal disease-

specific structure. Reporting this limitation strengthens rather than weakens the study because it prevents indiscriminate selection of imaging phases.

5.6.2 Why Radiomics Performance Fundamentally Depends on Imaging Biology

Although radiomics can extract large numbers of quantitative features from medical images, its diagnostic performance is fundamentally determined by the biological information encoded in the source images. Radiomics does not create new biological signals; it only quantifies and reorganizes information that is already present in the imaging data (Lambin et al. 2017). As a result, the success of any radiomics model depends not only on analytical technique, but also on the biological relevance, stability, and reproducibility of the imaging modality from which the features are derived (Gillies et al. 2016).

This principle is highly relevant in lateral cervical lymph node assessment. Metastatic involvement in papillary thyroid carcinoma is associated with complex biological changes, including partial tumor infiltration, vascular remodeling, stromal alteration, and disruption of normal lymphoid and immune architecture (Scheckenbach et al. 2017). Imaging modalities that incompletely reflect these processes will inevitably limit the discriminative potential of radiomics, regardless of how sophisticated the downstream modelling may be.

Conventional ultrasound provides a useful example. Its images are primarily based on macroscopic structural contrast and therefore encode mainly morphological information. Although radiomics can extract texture and intensity features from grayscale ultrasound, these features remain derived from an imaging substrate that may not change substantially in early or partial metastatic disease. Consequently, radiomics models based solely on conventional ultrasound may show only moderate performance when the biological changes are subtle (Coroller et al. 2017).

A similar limitation applies to early-phase contrast-enhanced ultrasound. Arterial and venous phase images mainly reflect transient hemodynamic phenomena. Features extracted from these phases may therefore be influenced by injection timing,

systemic circulation, and physiological variability, which can reduce feature stability and weaken model generalizability. In such cases, the limitation lies less in the radiomics algorithm than in the instability of the underlying biological signal (Theek et al. 2018; Jiang et al. 2020).

By contrast, imaging modalities that better encode tissue-level biology provide a more suitable substrate for radiomics. In the present study, the post-vascular phase of Sonazoid contrast-enhanced ultrasound was particularly advantageous because it reflects tissue-related contrast retention rather than hemodynamic behavior alone. As discussed in previous sections, this phase is more closely linked to lymphoid tissue integrity, macrophage-associated retention, and microenvironmental disruption caused by metastatic infiltration. Radiomics analysis of such images is therefore more likely to capture biologically meaningful heterogeneity rather than transient physiological variation.

This explains an important finding of the present thesis: radiomics performance differs substantially across ultrasound modalities and phases because the biological informativeness of the source images differs. Increasing feature dimensionality or model complexity cannot compensate for weak biological input. On the contrary, when the source images are biologically limited, more complex models may simply amplify noise and increase the risk of overfitting (Lambin et al. 2017).

This observation has broader implications for radiomics research. It suggests that the critical first step in building clinically useful radiomics models is not algorithm selection alone, but biologically informed image selection. Radiomics should be preferentially applied to imaging phases and modalities that are stable, reproducible, and closely linked to disease biology. In the context of the present study, this principle strongly supports the use of Sonazoid post-vascular phase imaging as a more suitable radiomics substrate than conventional ultrasound alone or transient early-phase CEUS.

In summary, radiomics performance is fundamentally dependent on imaging biology. Its clinical value is determined by the quality and biological relevance of the source images from which quantitative features are extracted. Recognizing this

dependency is essential for interpreting radiomics results appropriately and for developing robust, biologically meaningful models for lateral cervical lymph node metastasis.

From a biological perspective, this pattern is understandable. The arterial phase primarily reflects the arrival and early intranodal distribution of microbubbles and therefore captures imaging correlates of tumor -associated angiogenesis, perfusion asymmetry, and early vascular disorganization. Such alterations may generate heterogeneous enhancement patterns that are difficult to quantify through visual assessment alone but may be partially detectable by radiomics analysis. At the same time, arterial-phase enhancement is intrinsically influenced by transient hemodynamic conditions, injection timing, and systemic circulatory dynamics. Reactive or inflammatory lymph nodes may also exhibit increased arterial perfusion or heterogeneous enhancement, thereby creating overlap with metastatic nodes. These factors likely explain why arterial-phase radiomics was informative, yet insufficiently stable for stronger standalone performance (Wakisaka et al. 2019; Guo et al. 2021; Zhang et al. 2023).

Several factors may account for this limitation. First, arterial-phase imaging represents a rapidly changing physiological state, and radiomics features extracted from a single frame or narrow temporal window may not consistently represent the same biological substrate across patients. This temporal instability is likely to reduce feature reproducibility and model robustness. Second, arterial-phase CEUS remains influenced by technical factors such as injection rate, precise acquisition timing, and subtle differences in systemic hemodynamics. Third, ultrasound radiomics in general remains vulnerable to variability in machine settings, image selection, and lesion segmentation. These issues are consistent with broader radiomics methodology, in which model performance depends strongly on the biological and technical robustness of the source images rather than on algorithmic complexity alone (Theek et al. 2018; Zhou et al. 2021).

From a biological standpoint, these findings are understandable. The venous phase represents a transitional imaging stage in which the observed phenotype is

influenced by a mixture of residual perfusion, washout behavior, vascular permeability, and circulation-related effects rather than by a single dominant pathological characteristic. As a result, venous-phase heterogeneity may arise not only from metastatic infiltration, but also from benign inflammatory hyperaemia, reactive vascular change, or timing variation in image acquisition. Such overlap is likely to weaken the specificity of extracted radiomics features and reduce the ability of venous-phase models to distinguish metastatic from non-metastatic lymph nodes with high confidence (Zhao et al. 2024).

Several broader considerations also arise from these findings. One important limitation is temporal instability. Because the venous phase still represents a dynamically changing physiological state, radiomics features extracted from a single frame or a narrow temporal window may not consistently represent the same biological substrate across patients. In addition, ultrasound radiomics remains vulnerable to variation in acquisition conditions, including machine settings, imaging angle, and segmentation location. Such technical variability may further weaken the robustness of venous-phase models. This interpretation is consistent with general radiomics methodology, which emphasizes that reproducibility of the source image strongly influences downstream model performance (Lambin et al. 2017).

The post-vascular phase of Sonazoid CEUS differs fundamentally from the arterial and venous phases in the type of information it encodes. Whereas early-phase imaging mainly reflects transient hemodynamic behavior and intravascular contrast kinetics, the post-vascular phase provides relatively stable tissue-level information. More specifically, Sonazoid produces a post-vascular phase because perfluorobutane microbubbles can be taken up by phagocytic cells in reticuloendothelial tissue, resulting in prolonged enhancement beyond the purely intravascular period. In lymph node imaging, this delayed enhancement pattern has been associated with preservation or disruption of nodal structure and with metastatic involvement. From a biological standpoint, this distinction is highly important because it implies that post-vascular phase imaging is more closely linked to tissue integrity and microenvironmental alteration than to short-lived circulatory events alone (Xiao et al. 2023; Zhao et al. 2024).

From the perspective of radiomics methodology, such biological stability is of central importance. Radiomics studies and methodological reviews consistently emphasize that stable source images are more likely to yield reproducible features, lower noise levels, and models with stronger generalizability. Compared with arterial- or venous-phase CEUS, the longer and more temporally stable acquisition window of the post-vascular phase is therefore more likely to generate coherent feature distributions across patients. This characteristic is particularly advantageous in ultrasound-based radiomics, where feature reproducibility can otherwise be compromised by hemodynamic fluctuation, timing inconsistency, and acquisition variability (Lambin et al. 2012; Jiang et al. 2023).

An additional important point is that post-vascular-phase radiomics outperformed all other ultrasound-based radiomics sources evaluated in the present study. Its best test AUC of 0.882 exceeded that of conventional grayscale ultrasound radiomics (0.830), CDFI radiomics (0.824), arterial-phase CEUS radiomics (0.822), and venous-phase CEUS radiomics (0.696). This ranking is highly informative. It suggests that the effectiveness of radiomics depends heavily on the biological relevance and temporal stability of the source images rather than on computational modelling alone. Put differently, the superiority of post-vascular-phase radiomics does not simply indicate that more features or more complex algorithms were used more effectively; rather, it indicates that this imaging phase contains more robust and more pathophysiologically informative content for predicting metastatic status.

From a broader methodological perspective, the present findings reinforce the principle that phase selection is crucial in CEUS-based radiomics. Not all contrast-enhanced images are equally suitable for quantitative modelling, even when obtained using the same contrast agent. The mere presence of contrast enhancement does not guarantee that the resulting images will provide a stable or biologically specific basis for radiomics analysis. Rather, the phase that most faithfully reflects durable tissue-level properties is likely to yield the most robust and clinically meaningful quantitative features. In this regard, the superiority of post-vascular-phase Sonazoid CEUS supports prioritizing biologically stable imaging phases for ultrasound-based radiomics research (Lambin et al. 2017; Wei et al. 2023).

5.6.3 Feature Interpretation and Classifier Differences

The selected radiomics features represented first-order intensity, texture, dependence, run-length, size-zone, neighbouring-tone, and shape characteristics. These features can be interpreted as quantitative surrogates of heterogeneity, tissue organization, focal defects, and nodal architecture. However, individual feature names should not be equated directly with specific histopathological processes without correlative tissue analysis. Their biological interpretation remains plausible but inferential.

Different classifiers performed best for different modalities. This does not establish that one algorithm is universally superior. It suggests that classifier performance depends on the structure of the feature space, sample size, noise, and nonlinear relationships. The best model is therefore the model that generalizes most consistently for a given image substrate, not necessarily the most complex algorithm.

A major challenge in radiomics research is ensuring that extracted features and predictive models remain biologically interpretable rather than functioning as purely computational abstractions. In lateral cervical lymph node evaluation, this issue is particularly important because acceptance of radiomics in clinical practice depends not only on predictive performance, but also on whether the extracted imaging patterns can be plausibly linked to known pathological processes. In the present study, the biological relevance of radiomics features appeared closely related to the known microstructural and microenvironmental alterations associated with metastatic lymph node involvement in papillary thyroid carcinoma.

Radiomics features are typically grouped into first-order statistics, shape descriptors, and higher-order texture features. First-order features describe the overall distribution of pixel intensities, whereas texture features derived from matrices such as GLCM, GLRLM, GLDM, and GLSZM quantify spatial heterogeneity, local dependence, and structural complexity within an image. In the context of lateral cervical lymph nodes, such features are biologically plausible because metastatic infiltration is often partial and spatially heterogeneous rather than uniformly distributed. Tumor deposits, fibrosis, necrosis, altered vascularity, and disruption of normal lymphoid

architecture may produce subtle intranodal heterogeneity that is difficult to recognize visually but can be quantified computationally (Lambin et al. 2012, 2017).

This biological interpretability was reflected in the present study. In conventional grayscale ultrasound radiomics, the retained features included first-order, GLDM, and shape descriptors, suggesting that both internal signal heterogeneity and structural configuration contributed to metastatic discrimination. In CDFI radiomics, selected features were related not only to intensity and texture, but also to vascular signal complexity, consistent with the known vascular abnormalities of metastatic lymph nodes. In arterial- and venous-phase CEUS radiomics, retained features often reflected intensity non-uniformity, run-length heterogeneity, and complexity, which may correspond to perfusion asymmetry and contrast distribution irregularity. Most notably, post-vascular-phase Sonazoid CEUS radiomics showed high importance of features related to gray-level dependence and heterogeneity, which is biologically consistent with tissue-level disruption, uneven contrast retention, and alteration of nodal microenvironment in metastatic nodes.

These observations reinforce the view that radiomics features should not be understood as arbitrary mathematical constructs. Rather, they act as quantitative surrogates of biologically meaningful heterogeneity. Importantly, this does not require each feature to correspond to a single pathological change. The biological value of radiomics lies in the fact that combinations of features capture multiscale patterns produced by multiple interacting processes, including tumor infiltration, vascular remodeling, immune-cell redistribution, and necrosis. This multivariate nature is particularly appropriate for lateral cervical lymph node metastasis, which is itself biologically heterogeneous and rarely expressed as a single uniform imaging phenotype.

The present findings also show that classifier performance must be interpreted in relation to the biological quality of the extracted features rather than as an independent indicator of algorithmic superiority. Different classifiers performed best in different imaging modalities: ExtraTrees in grayscale ultrasound radiomics, AdaBoost in CDFI radiomics, NaïveBayes in arterial-phase CEUS radiomics, Random Forest in venous-phase CEUS radiomics, and MLP in post-vascular-phase CEUS radiomics. This

pattern suggests that classifier success depended strongly on the statistical structure and biological informativeness of the input data. Ensemble methods such as ExtraTrees and AdaBoost appeared to provide robust baseline performance in heterogeneous but only moderately informative image sources, whereas the multilayer perceptron showed the greatest advantage when applied to post-vascular-phase radiomics, where the extracted features appeared more stable and biologically meaningful.

This observation is important because it indicates that the best-performing model is not necessarily the most complex one, but the one whose inductive bias is most compatible with the structure of the radiomics feature space. When source images are biologically rich and relatively stable, more flexible classifiers may successfully capture complex nonlinear patterns. Conversely, when source images are biologically limited or noisy, greater algorithmic complexity may not improve performance and may even increase instability. Thus, the present study supports a central principle of radiomics-based modelling: classifier selection should be guided not by presumed algorithmic superiority alone, but by compatibility with the biological and statistical properties of the source imaging data (Lambin et al. 2012; Gillies et al. 2016).

Taken together, these findings suggest that feature interpretability and classifier performance are closely linked. The extracted radiomics features derive their predictive value from underlying biological heterogeneity, while the relative success of different classifiers depends on how effectively they can exploit these biologically grounded quantitative patterns. In this sense, the present results support a biologically informed approach to radiomics, in which both feature extraction and classifier selection are interpreted within the context of imaging biology rather than as purely technical exercises.

5.6.4 Internal Validation and Risk of Overfitting

The use of feature reduction, ten-fold cross-validation, and a held-out test cohort provided an important internal check against overfitting. The separation between training and test performance was nevertheless evident for some models, indicating that optimism remained possible. This is expected in a high-dimensional study with a modest sample size, especially when multiple modalities and classifiers are compared.

The test cohort should not be described as external validation because it was drawn from the same institution, population, equipment environment, and acquisition protocol as the training cohort. It provides internal hold-out validation. True external validation requires independent patients from another time period, institution, operator group, or equipment platform. This distinction directly addresses the reviewers' concern regarding generalizability.

Decision curve analysis added clinically relevant information by estimating net benefit across threshold probabilities. In the latest comparative analysis, no single radiomics model was uniformly dominant across the full threshold range, and the curves crossed in both cohorts. The post-vascular-phase CEUS MLP model showed comparatively favorable net benefit over portions of the test-cohort threshold range, but instability at higher thresholds limits claims of universal superiority. These findings indicate that the preferred model may depend on the clinical action threshold and that apparent advantages at extreme thresholds should be interpreted cautiously because they are supported by few observations.

A critical requirement for the clinical translation of radiomics-based models is robust validation and demonstrated generalizability. High apparent performance in a training cohort does not necessarily imply reliable performance in independent datasets, particularly in radiomics studies characterized by high-dimensional features and relatively limited sample sizes (Gillies et al. 2016; Lambin et al. 2017). For this reason, the present study emphasized not only model discrimination but also stability across validation layers and indicators of potential clinical reliability.

In the present work, radiomics models were evaluated using both internal cross-validation and an independent test cohort. This approach is important because cross-validation provides an estimate of how model performance varies across repeated data partitions, whereas an independent test cohort offers a more stringent approximation of real-world application. Models that maintained broadly consistent performance across these validation layers were therefore considered more reliable than those exhibiting substantial performance degradation between training and testing. This distinction was particularly relevant in the present study, where some models showed

strong apparent discrimination in the training cohort but noticeably weaker performance in the test cohort, suggesting that not all radiomics signatures were equally generalizable.

The contrast between different imaging substrates was especially informative. Conventional grayscale ultrasound and CDFI radiomics showed moderate predictive performance but also some reduction in test-cohort discrimination, indicating limited robustness. Arterial-phase CEUS radiomics demonstrated measurable value but an unstable relationship between training and test performance, suggesting sensitivity to sample distribution and reduced model stability. Venous-phase CEUS radiomics performed less well overall and showed limited specificity, further weakening its suitability as a standalone substrate. By contrast, post-vascular-phase Sonazoid CEUS radiomics showed the most favorable balance between discrimination and consistency, with the MLP model achieving relatively stable performance across the training and test cohorts. This pattern strongly supports the view that model generalizability depends not only on classifier design, but also on the biological stability of the source image.

Beyond AUC and accuracy, clinical reliability also depends on the pattern of classification errors. In the context of lateral cervical lymph node evaluation, false-negative predictions may contribute to undertreatment and persistent disease, whereas false-positive predictions may lead to unnecessary lateral neck dissection and avoidable morbidity. For this reason, confusion matrix analysis is particularly relevant. It allows model performance to be interpreted not simply as a statistical summary, but in terms of clinically meaningful misclassification behavior. In the present study, some models achieved high sensitivity at the expense of specificity, whereas others showed the opposite pattern. This reinforces the importance of evaluating radiomics models in relation to intended clinical use rather than relying on discrimination metrics alone.

Decision curve analysis (DCA) provided an additional layer of clinically relevant validation. Unlike purely statistical measures, DCA assesses whether a model offers net clinical benefit across a range of threshold probabilities relative to treat-all or treat-none strategies (Vickers and Elkin 2006). In the present study, the inclusion of DCA demonstrated that several radiomics models had potential practical utility despite differences in AUC. This is important because models with similar discrimination may

differ substantially in their usefulness for decision-making. In preoperative assessment of lateral cervical lymph node metastasis, where management decisions must balance oncologic adequacy against procedural morbidity, net clinical benefit is a particularly meaningful criterion.

At the same time, the present findings indicate that even the best-performing radiomics models should be interpreted cautiously in translational terms. Although post-vascular-phase radiomics showed the strongest performance among all radiomics approaches, it did not exceed the best multimodal conventional diagnostic combination, which incorporated conventional ultrasound, Sonazoid CEUS, and FNA-Tg. This comparison has important clinical implications. It suggests that radiomics, even when derived from the most biologically informative imaging substrate, is currently best regarded as an adjunctive quantitative layer rather than a replacement for multimodal diagnostic integration. Its strength lies in refining risk estimation, improving objectivity, and quantifying heterogeneity, but not in superseding the combined use of morphological, functional, and biochemical evidence.

Another important dimension of clinical reliability is calibration, namely the agreement between predicted probabilities and observed outcomes. The latest analysis included calibration curves for the best-performing model from each imaging modality. In the training cohort, observed event proportions generally increased with predicted risk, but all curves showed some departure from the 45-degree reference line. The test-cohort curves were more irregular and showed modality-specific underestimation and overestimation across probability intervals. Consequently, none of the models demonstrated perfect calibration across the full risk range. Because the test cohort contained only 36 cases and numerical indices such as calibration intercept, calibration slope, and Brier score were unavailable, the calibration findings should be considered descriptive rather than definitive.

Overall, the validation framework used in the present study—including cross-validation, independent test-cohort assessment, confusion matrix analysis, and decision curve analysis—provides a stronger basis for judging model credibility than reliance on training performance alone. The results suggest that generalizability and clinical

reliability vary substantially across imaging phases and radiomics substrates, and that these differences are closely linked to biological image stability. This reinforces the broader conclusion of the present thesis: radiomics models are clinically most promising when built upon imaging phases that encode stable and biologically relevant tissue information.

a. Comparative Generalizability of the Best-Performing Radiomics Models

Table 4.38 of Chapter IV is particularly informative because it places test-cohort discrimination beside training performance and therefore exposes the degree of model stability that would be hidden by reporting only the best AUC. The table shows that the ranking of models cannot be interpreted solely from their test AUCs; the train-test gap, confidence interval width, and direction of performance change are equally important indicators of generalizability.

The ExtraTrees model based on conventional grayscale ultrasound achieved a training AUC of 0.927 and a test AUC of 0.830, corresponding to a reduction of 0.097. This represents acceptable but clearly diminished test discrimination. The gap is compatible with moderate overfitting, which is plausible because tree ensembles can model complex interactions among radiomics features but may also learn sample-specific partitioning patterns. Nevertheless, the test AUC remained clinically meaningful, suggesting that conventional grayscale images contain reproducible structural information, albeit insufficient to support highly accurate standalone prediction.

The CDFI AdaBoost model showed a somewhat smaller train-test decline, from 0.899 to 0.824, but the test confidence interval was wide. Its performance indicates that vascular texture contributes complementary information, while also highlighting the instability of Doppler-derived radiomics. CDFI is sensitive to gain, pulse repetition frequency, wall-filter settings, probe pressure, and low-flow detectability. Thus, a moderate generalization gap may arise not only from statistical overfitting but also from variation in the physical acquisition of the vascular signal.

The arterial-phase Naive Bayes model displayed the opposite pattern, with a lower training AUC of 0.664 but a higher test AUC of 0.822. Such an increase should not be interpreted as evidence that the model performs better in unseen populations. In a small held-out cohort, differences in case mix, class balance, lesion difficulty, and random partitioning can produce an apparently favorable test estimate. The wide confidence interval further supports sampling variability as the most likely explanation. This finding underlines why a single split should be supplemented by repeated resampling, nested cross-validation, or external validation.

The venous-phase Random Forest model showed the largest deterioration, from a training AUC of 0.936 to a test AUC of 0.701, a gap of 0.235. This combination of excellent apparent training performance and weak test discrimination is a classic warning sign of overfitting and limited transportability. It is also consistent with the biological and temporal ambiguity of the venous phase, in which residual perfusion, washout, vascular permeability, and inflammatory hyperaemia may overlap. The model may therefore have captured unstable or cohort-specific feature interactions rather than a robust metastatic phenotype.

The post-vascular-phase MLP model produced the highest observed test AUC of 0.882, compared with a training AUC of 0.821. The positive train-test difference of 0.061, like that of the arterial-phase model, is most plausibly explained by sampling variability rather than a true improvement in external performance. Even so, its ranking is biologically coherent because the post-vascular phase offers a longer, more stable acquisition window and reflects tissue retention and reticuloendothelial disruption rather than transient circulation alone. The result therefore identifies post-vascular imaging as the most promising radiomics substrate in this dataset, while not establishing statistical superiority over the other models.

The overlapping and relatively wide test-cohort confidence intervals are critical to this interpretation. Although the point estimate for the post-vascular-phase MLP model was the highest, pairwise DeLong comparisons among the five selected radiomics models in both the training and test cohorts. These comparisons did not support a simple conclusion that the numerically highest test AUC represented

definitive superiority. For example, the grayscale-ultrasound ExtraTrees and CDFI AdaBoost models were not significantly different in either cohort. Given the small test sample, overlapping confidence intervals, and multiple pairwise tests, the appropriate conclusion remains that the post-vascular-phase MLP model had the highest observed internal test AUC, not that it was conclusively superior to all other models.

Decision curve analysis suggested potential net benefit for selected models over portions of the evaluated threshold range, but the benefit was model- and threshold-dependent. Calibration curves further showed deviations from perfect agreement, especially in the small test cohort. Thus, discrimination, calibration, and net benefit should be interpreted together: a model may rank patients effectively yet misestimate absolute risk, and a model with favorable net benefit at one threshold may not remain preferable at another. These findings support cautious clinical positioning of radiomics as an adjunct rather than a stand-alone determinant of lateral neck dissection.

Collectively, Table 4.38 of Chapter IV indicates that model performance was determined by the interaction between imaging biology, acquisition stability, classifier behavior, and sample variation. It also demonstrates why the development process should not select a model solely because it has the highest training AUC or even the highest single test AUC. A clinically credible model should show stable performance across resampling procedures, acceptable calibration, consistent net benefit, robustness to segmentation and equipment variation, and successful validation in an independent population.

b. Pairwise DeLong Comparisons Across the Selected Radiomics Models

The pairwise DeLong heatmaps in Figure 4.57 of Chapter IV provide an inferential complement to the descriptive ranking of the five modality-specific models. Although the post-vascular-phase CEUS MLP model achieved the highest observed test AUC, broad and overlapping confidence intervals indicate that the numerical ranking alone cannot establish superiority. The grayscale-ultrasound ExtraTrees and CDFI AdaBoost models, for example, were not significantly different in either the training cohort ($p = 0.655$) or the test cohort ($p = 0.720$). This supports the interpretation that several

modalities may provide broadly comparable discrimination within the precision permitted by the available sample.

The matrices also contain lower p-values for some comparisons, but these should be interpreted in the context of multiple pairwise testing, wide confidence intervals, and a test cohort of only 36 cases. Without adjustment for multiplicity, isolated p-values below 0.05 may overstate the strength of evidence. Moreover, statistical significance does not necessarily imply a clinically important difference. A small AUC difference may have little effect on the number of patients reclassified or on surgical decisions, whereas a larger difference may still be uncertain when confidence intervals are wide. The DeLong results therefore reinforce the need to integrate effect size, precision, calibration, and decision-curve findings rather than selecting a model solely from its rank order.

c. Interpretation of Comparative Decision Curve Analysis

Figure 4.58 of Chapter IV compares the net benefit of the five selected models against treat-all and treat-none strategies in the training and test cohorts. The absence of a uniformly dominant curve is clinically informative. It indicates that the relative value of a model changes with the threshold probability at which a clinician would recommend further invasive assessment or lateral neck dissection. At lower thresholds, avoiding missed metastasis may be prioritized; at higher thresholds, avoiding unnecessary treatment becomes more important. Consequently, the same model need not be optimal across all decision contexts.

The comparatively favorable test-cohort net benefit of the post-vascular-phase CEUS MLP model over portions of the threshold range is consistent with its higher observed AUC, but the crossed and unstable curves caution against describing it as universally superior. At extreme thresholds, few patients contribute to the calculation, making net-benefit estimates highly sensitive to individual classification errors. DCA should therefore be viewed as evidence of potential clinical usefulness under selected conditions, not proof that implementation will improve patient outcomes. Prospective impact studies are needed to determine whether model-assisted decisions actually reduce unnecessary aspiration or surgery without increasing missed metastasis.

d. Interpretation of Calibration Curves

The calibration curves in Figure 4.59 of Chapter IV address a question that cannot be answered by AUC alone: whether predicted probabilities correspond to the observed frequency of metastasis. Curves above the diagonal indicate underestimation of risk, whereas curves below the diagonal indicate overestimation. In the training cohort, the five models generally preserved an increasing relationship between predicted probability and observed metastasis, but departures from the ideal diagonal remained. In the test cohort, the curves were more irregular, indicating that the absolute probabilities were less stable when applied to unseen cases.

This pattern is compatible with the small test sample, model-selection variability, and possible overfitting. It also explains why a model with acceptable discrimination may still be unsuitable for direct threshold-based treatment decisions. For example, systematic overestimation could increase unnecessary aspiration or neck dissection, whereas underestimation could delay confirmation of metastatic disease. Because calibration was assessed visually and numerical measures such as calibration intercept, slope, and Brier score were not reported, no model should be described as well calibrated. Future validation should quantify these indices, apply bootstrap correction where appropriate, and perform recalibration before fixed probability thresholds are transferred to another institution.

5.6.5 Clinical Positioning of Radiomics

The best radiomics model did not surpass the highest-performing multimodal diagnostic combination. Radiomics should therefore be positioned as an adjunctive decision-support method rather than a replacement for expert imaging and targeted biochemical testing. Its potential advantages are objectivity, repeatability, and the ability to quantify subtle heterogeneity; its current disadvantages include segmentation burden, dependence on acquisition standardization, limited interpretability, and uncertain transportability.

A realistic translational role would be automated or semi-automated risk estimation integrated into the ultrasound workstation after a node has been detected and segmented. The output could support prioritization for CEUS or aspiration, highlight cases requiring expert review, or provide a second opinion in centers with limited subspecialty experience. Such use requires prospective workflow studies rather than retrospective performance comparisons alone.

The integration of radiomics into a multimodal diagnostic framework represents the logical culmination of the methodological strategy adopted in the present study. Rather than functioning as an isolated computational technique, radiomics is more appropriately understood as a quantitative component within a structured diagnostic pathway that integrates morphological imaging, functional assessment, biochemical confirmation, and predictive modelling. In this context, radiomics does not replace established diagnostic approaches; instead, it serves as a quantitative augmentation layer that enhances the interpretation and standardization of complementary diagnostic information (Lambin et al. 2012, 2017).

A central finding of the present study was that radiomics performance was strongly modality-dependent. Among the five evaluated imaging modalities, the best overall diagnostic performance was achieved by radiomics models derived from post-vascular-phase Sonazoid contrast-enhanced ultrasound, followed by models based on conventional grayscale ultrasound, color Doppler flow imaging, arterial-phase CEUS, and venous-phase CEUS. This ranking reinforces a key principle of the present thesis: the predictive value of radiomics is constrained not only by computational modelling strategy, but also by the biological relevance and stability of the source images. This interpretation is consistent with radiomics methodology literature showing that image quality, feature robustness, and modelling strategy act together to determine predictive performance (Lambin et al. 2017).

At the same time, even the best-performing radiomics model did not exceed the diagnostic performance of the strongest multimodal conventional combination identified in the present study. The combined model integrating conventional ultrasound, Sonazoid CEUS, and FNA-Tg achieved substantially higher diagnostic performance

than any single radiomics model alone. This finding indicates that radiomics should presently be regarded not as a replacement for multimodal diagnostic evaluation, but as an adjunctive quantitative tool capable of refining risk stratification, reducing subjectivity, and quantifying subtle intranodal heterogeneity. This positioning is also supported by recent multimodal ultrasound radiomics studies in papillary thyroid carcinoma, which suggest that radiomics is most valuable when incorporated into broader integrated prediction frameworks rather than used in isolation (Dai et al. 2023).

From a translational perspective, the most promising role of radiomics lies in its incorporation into multimodal prediction frameworks. By converting imaging heterogeneity into numerical features, radiomics provides a bridge between visual imaging phenotype and objective risk estimation. When integrated with conventional ultrasound findings, CEUS characteristics, biochemical markers such as FNA-Tg, and relevant clinical variables, radiomics may contribute to a more individualized and biologically informed assessment of lateral cervical lymph node metastasis. However, meaningful clinical translation will require further methodological refinement, including standardization of image acquisition and preprocessing, robust feature selection, and validation across independent external cohorts (Lambin et al. 2017; Dai et al. 2023).

Another important translational issue is clinical reliability. High apparent model performance in a training cohort is insufficient if it does not generalize across different datasets or provide meaningful decision-level benefit. For this reason, model evaluation must extend beyond discrimination alone to include independent testing, clinically interpretable error patterns, and decision-analytic assessment. Decision curve analysis is particularly relevant in this setting because it evaluates whether a prediction model provides net clinical benefit compared with treat-all or treat-none strategies, which is directly applicable to preoperative management of lateral cervical lymph nodes (Vickers and Elkin 2006; Wei et al. 2023). In the present study, the incorporation of cross-validation, independent test-cohort evaluation, confusion matrix analysis, and decision curve analysis therefore strengthened the translational credibility of the proposed radiomics models (Steyerberg et al. 2010; Xue et al. 2023).

5.7 TRANSPORTABILITY AND CLINICAL INTERPRETATION

The present findings were derived from a single-center cohort using locally defined acquisition protocols, Sonazoid phase selection, FNA-Tg procedures, and analytical workflows. Accordingly, their transportability cannot be assumed from internal performance alone. Differences in ultrasound systems, operator experience, contrast administration, laboratory assays, patient spectrum, and reference standards may alter both conventional diagnostic thresholds and radiomics performance. This limitation is particularly relevant to the FNA-Tg cut-off and to models based on post-vascular-phase images, because both depend on technical standardization.

This external-validity boundary does not negate the observed diagnostic value; rather, it defines how the results should be interpreted. Conventional ultrasound, Sonazoid CEUS, FNA-Tg, and radiomics showed complementary strengths within the study setting, but the present data support clinical plausibility and internal discrimination rather than universal implementation. The discussion therefore positions the framework as a carefully qualified interpretation of the evidence, while consolidating the limitations and proposed clinical pathway.

CHAPTER VI

CONCLUSION AND RECOMMENDATIONS

6.1 SUMMARY OF THE STUDY

This study investigated a multimodal and quantitatively oriented framework for the preoperative prediction of lateral cervical lymph node metastasis (LLNM) in papillary thyroid carcinoma (PTC). A total of 178 patients were evaluated using conventional ultrasound, Sonazoid contrast-enhanced ultrasound (CEUS), fine-needle aspiration cytology (FNA-C), fine-needle aspiration thyroglobulin (FNA-Tg), and radiomics derived from conventional grayscale ultrasound, color Doppler flow imaging (CDFI), and the arterial, venous, and post-vascular phases of Sonazoid CEUS. The study examined the diagnostic value of each method, compared qualitative and quantitative interpretation, assessed the incremental value of multimodal integration, and evaluated the discrimination, calibration, and potential clinical utility of modality-specific radiomics models.

The findings demonstrated a consistent progression from subjective morphology towards structured quantification and multimodal integration. Conventional ultrasound remained the essential first-line examination for cervical mapping, anatomical localization, and selection of lymph nodes for further assessment. Quantitative scoring improved discrimination compared with qualitative interpretation, although morphology-based imaging retained an inherent diagnostic ceiling because reactive and metastatic lymph nodes may share overlapping features.

FNA-Tg provided strong biochemical confirmation of thyroid-derived material within the sampled lymph node. Quantitative FNA-Tg achieved excellent discrimination at the cohort-derived cut-off of 9.10 ng/mL, but this value should be

interpreted as locally derived rather than universally transferable because assay platform, washout technique, serum thyroglobulin, anti-thyroglobulin antibodies, thyroid status, and patient spectrum may alter the optimal threshold.

Sonazoid CEUS added functional and tissue-level information beyond conventional ultrasound. Quantitative interpretation generally outperformed qualitative assessment, and integrated Sonazoid CEUS significantly outperformed integrated conventional ultrasound. Among the CEUS phases, the post-vascular phase provided the most biologically informative and temporally stable imaging window, supporting its value for both visual interpretation and radiomics analysis.

Multimodal integration produced the highest observed diagnostic performance. The combination of conventional ultrasound, Sonazoid CEUS, and FNA-Tg achieved an AUC of 0.993. However, pairwise DeLong testing showed that this three-modality combination did not significantly outperform US + CEUS, CEUS + FNA-Tg, or US + FNA-Tg ($p = 0.2849$, $p = 0.2769$, and $p = 0.2717$, respectively). It should therefore be described as having the highest observed discrimination rather than as being statistically superior to all high-performing two-modality alternatives.

The radiomics analysis further showed that model performance depended strongly on the biological and technical quality of the source images. Among the five modality-specific best-performing models, the post-vascular-phase Sonazoid CEUS multilayer perceptron model (MLP) achieved the highest observed test-cohort AUC of 0.882, followed by grayscale-ultrasound ExtraTrees, CDFI AdaBoost, arterial-phase CEUS Naive Bayes, and venous-phase CEUS Random Forest. Pairwise DeLong comparisons showed that the numerical ranking did not establish universal statistical superiority, particularly given overlapping confidence intervals, multiple comparisons, and the small test cohort of 36 cases.

Decision curve analysis indicated that several radiomics models provided potential net clinical benefit over selected threshold ranges, but no model was uniformly dominant across all thresholds. Calibration curves showed departures from the ideal 45-degree line, with greater irregularity in the test cohort and modality-specific

underestimation and overestimation of risk. Because calibration was assessed visually and calibration intercept, calibration slope, and Brier score were not available, none of the models should be described as well calibrated. These findings position radiomics as a promising adjunctive risk-refinement tool rather than a validated replacement for multimodal clinical assessment.

6.2 OVERALL CONCLUSIONS

First, conventional ultrasound remains indispensable as the entry point for preoperative lateral neck assessment. Quantitative scoring improves objectivity and specificity, but does not eliminate the limitations of morphology- and Doppler-based evaluation.

Second, Sonazoid CEUS provides clinically relevant information beyond conventional ultrasound. Its post-vascular phase is particularly valuable because it reflects tissue-level contrast behavior and provides the most promising source images for ultrasound radiomics.

Third, FNA-Tg is a highly effective biochemical adjunct, especially when cytology is non-diagnostic or imaging findings are equivocal. Its numerical interpretation should remain assay- and context-specific.

Fourth, the most accurate observed assessment was achieved through multimodal integration of structural, functional, and biochemical information. Nevertheless, the lack of significant DeLong differences between the three-modality model and the best two-modality combinations supports selective diagnostic escalation rather than routine maximal testing for every suspicious lymph node.

Fifth, radiomics adds quantitative information beyond visual interpretation, but its apparent performance must be judged using discrimination, calibration, decision-curve findings, error patterns, and transportability together. Post-vascular-phase Sonazoid CEUS is the most promising radiomics substrate in the present study, but the models require external validation and recalibration before probability-based clinical decisions.

Overall, LLNM in PTC should be approached as a multidimensional diagnostic problem. The most defensible current strategy is a biologically informed and risk-adapted framework in which conventional ultrasound provides structural mapping, Sonazoid CEUS supplies functional and post-vascular characterisation, FNA-Tg offers targeted biochemical confirmation, and radiomics provides optional quantitative refinement.

6.3 CONTRIBUTIONS OF THE STUDY

The principal clinical contribution of this study is the establishment of a stepwise, compartment-specific diagnostic framework for lateral cervical lymph nodes in PTC. Rather than proposing that every patient undergo every test, the framework assigns complementary roles to conventional ultrasound, Sonazoid CEUS, FNA-Tg, and radiomics according to residual diagnostic uncertainty and the consequences of an incorrect decision.

Methodologically, the study demonstrates the value of structured quantitative interpretation across several diagnostic domains. It also shows that radiomics performance depends not only on classifier selection, but more fundamentally on the biological relevance, temporal stability, and reproducibility of the source images.

Scientifically, the study identifies the post-vascular phase of Sonazoid CEUS as the most informative phase for both conventional assessment and radiomics. The concordance between post-vascular visual performance and post-vascular radiomics performance provides biologically plausible internal consistency.

A further contribution is the balanced statistical interpretation of high-performing models. By incorporating pairwise DeLong comparisons, DCA, and calibration curves, the study distinguishes numerical ranking from statistical superiority, discrimination from absolute-risk accuracy, and potential net benefit from demonstrated clinical impact.

6.4 PROPOSED CLINICAL INTEGRATION

The proposed framework can be incorporated into current practice as a staged pathway. Conventional ultrasound should remain the first step for systematic cervical mapping and initial risk stratification. Structured scoring may be used to reduce interpretative variability while preserving expert visual assessment.

Sonazoid CEUS should be considered for suspicious or indeterminate nodes when additional functional information could alter the decision to perform aspiration or lateral neck dissection. Interpretation should include arterial and venous behavior but should place particular emphasis on post-vascular-phase findings.

When pathological confirmation is required for a management-changing decision, targeted FNA-C and FNA-Tg should be obtained together where feasible. FNA-Tg should be interpreted using locally validated laboratory procedures rather than by automatic transfer of the 9.10 ng/mL cut-off to other institutions.

Concordant high-risk findings across ultrasound, Sonazoid CEUS, and FNA-Tg provide the strongest basis for multidisciplinary consideration of therapeutic lateral neck dissection. When findings are discordant, the clinical team should review image quality, sampling adequacy, nodal level, primary-tumor status, and the relative consequences of false-negative and false-positive decisions.

Radiomics should currently be limited to decision support in research-capable or specialist centers. Its output may assist prioritization, second review, or risk refinement, but fixed probability thresholds should not be used until external validation and recalibration confirm reliable performance.

6.5 LIMITATIONS OF THE STUDY

The principal limitation is a single-center design. Referral patterns, operator expertise, equipment, contrast protocols, laboratory assays, and local clinical practice may have influenced the observed performance and limit direct generalization to other institutions.

External validation was not performed. Although the radiomics models were evaluated in a held-out test cohort, this internal split cannot reproduce interinstitutional differences in scanners, acquisition settings, segmentation, patient spectrum, and reference standards. The test cohort contained only 36 cases, which also contributed to wide confidence intervals, unstable decision curves, and irregular calibration patterns.

The final reference classification combined surgical histopathology with follow-up for patients who did not undergo surgery. This approach reflects ethical clinical practice, but the two pathways do not provide identical certainty and may introduce verification or classification bias.

The quantitative ultrasound and CEUS scoring systems were developed and evaluated within the same institutional environment. Their thresholds require interobserver, interequipment, and external reproducibility assessment. Similarly, the FNA-Tg threshold is dependent on assay and sampling procedures and requires local verification.

The radiomics workflow relied on manual or semi-manual segmentation and a specific preprocessing pipeline. Scanner effects, segmentation variability, feature-selection choices, and comparison of multiple classifiers increase the risk of overfitting and optimistic model selection.

Calibration evaluation was descriptive. Numerical measures such as calibration intercept, calibration slope, and Brier score were not available, and no recalibration analysis was performed. Therefore, the predicted probabilities should not yet be used as absolute risk estimates for treatment thresholds.

Finally, the study evaluated diagnostic performance and potential net benefit rather than actual changes in patient outcomes. Prospective evidence is required to determine whether the proposed framework reduces unnecessary aspiration or surgery, prevents missed metastasis, improves surgical planning, and is cost-effective.

6.6 RECOMMENDATIONS FOR FUTURE RESEARCH

The highest research priority is a prospective multicenter validation study using prespecified ultrasound, Sonazoid CEUS, FNA-Tg, segmentation, and analytical protocols. Such a study should include an independent external test cohort and represent different ultrasound systems, operators, laboratories, and patient populations.

Future model evaluation should report discrimination together with calibration intercept, calibration slope, Brier score, confidence intervals, and decision curve analysis. Bootstrap correction and formal recalibration should be performed before probability thresholds are transferred to another center.

Radiomics research should prioritize post-vascular-phase Sonazoid CEUS while testing feature reproducibility, scanner harmonization, test-retest stability, automated segmentation, and transparent model explanation. Comparisons with simpler clinical models are necessary to establish genuine incremental value.

Future implementation studies should evaluate the entire staged diagnostic pathway rather than accuracy alone. Outcomes should include biopsy selection, extent of surgery, unnecessary lateral neck dissection, reoperation, complications, surveillance burden, cost-effectiveness, clinician confidence, and patient-reported outcomes.

The three-modality strategy should be studied selectively in patients with discordant or high-consequence findings rather than assumed to be necessary for all patients. Research should identify the clinical situations in which the third modality changes management sufficiently to justify its additional invasiveness, cost, or resource use.

6.7 FINAL CONCLUSION

Accurate preoperative assessment of LLNM is essential for balancing oncological adequacy against the morbidity of unnecessary lateral neck dissection. This study demonstrates that no single modality fully captures the structural, functional, biochemical, and quantitative complexity of metastatic lymph nodes.

Conventional ultrasound should remain the structural foundation of assessment; Sonazoid CEUS, particularly the post-vascular phase, should provide functional and tissue-level refinement; and FNA-Tg should be used selectively for biochemical confirmation when the result is likely to change management. The three-modality combination achieved the highest observed discrimination, but it was not statistically superior to the best two-modality combinations and should therefore support selective rather than routine maximal testing.

Radiomics offers meaningful quantitative potential, especially when derived from post-vascular-phase Sonazoid CEUS, but current evidence supports its use as an adjunctive decision-support layer. The absence of external validation, imperfect and descriptively assessed calibration, and threshold-dependent DCA findings preclude immediate autonomous clinical use.

The conclusion of this doctoral research is that a stepwise, biologically informed, and risk-adapted multimodal framework provides the strongest current basis for preoperative LLNM assessment in PTC. Prospective multicenter validation and outcome-based implementation studies are required before the framework and its radiomics components can be translated into routine probability-based clinical decision-making.

REFERENCES

Agyekum, E. A., Y. Z. Ren, X. Wang, S. S. Cranston, Y. G. Wang, J. Wang, D. Akortia, F. J. Xu, L. Gomashie, Q. Zhang, D. Zhang & X. Qian. 2022. Evaluation of Cervical Lymph Node Metastasis in Papillary Thyroid Carcinoma Using Clinical-Ultrasound Radiomic Machine Learning-Based Model. Cancers 14(21).

Ahuja, A. T. & T. C. Ying. 2005. Sonographic evaluation of cervical lymph nodes. AJR Am J Roentgenol 184(5): 1691-1699.

Al-Hilli, Z., V. Strajina, T. J. McKenzie, G. B. Thompson, D. R. Farley, M. R. Castro, A. Algeciras-Schimnich & M. L. Richards. 2017. Thyroglobulin Measurement in Fine-Needle Aspiration Improves the Diagnosis of Cervical Lymph Node Metastases in Papillary Thyroid Carcinoma. Ann Surg Oncol 24(3): 739-744.

Ardakani, A. A., A. Rasekhi, A. Mohammadi, E. Motevalian & B. K. Najafabad. 2018. Differentiation between metastatic and tumour-free cervical lymph nodes in patients with papillary thyroid carcinoma by grey-scale sonographic texture analysis. Polish Journal of Radiology 83: e37-e46.

Bakuła-Zalewska, E., A. Żyłka, J. Długosińska, E. Musiał, J. Gałczyński & M. Dedecjus. 2021. Thyroglobulin measurements in washouts of fine-needle aspiration biopsy in the monitoring of patients with differentiated thyroid carcinoma. Endokrynol Pol 72(6): 601-608.

Barr, R. G., P. Huang, Y. Luo, X. Xie, R. Zheng, K. Yan, X. Jing, Y. Luo, H. Xu, X. Fei & J. M. Lee. 2020. Contrast-enhanced ultrasound imaging of the liver: a review of the clinical evidence for SonoVue and Sonazoid. Abdom Radiol (NY) 45(11): 3779-3788.

Bergeron, S. R., N. P. Papanikolopoulos & A. Truskinovsky. 2018. Applicability of color and texture features in the universal detection of lymph node metastases by computer vision. Laboratory Investigation 98: 585.

Bonini, G., Pezzotta, G., Morzenti, C., Agazzi, R. & Nani, R. 2007. Contrast-enhanced ultrasound with SonoVue in the evaluation of postoperative complications in pediatric liver transplant recipients. J Ultrasound 10(2): 99-106.

Bournaud, C., A. Charrié, C. Nozières, K. Chikh, V. Lapras, M.-L. Denier, C. Paulin, M. Decaussin-Petrucci, J.-L. Peix, J.-C. Lifante, C. Cornu, C. Giraud, J. Orgiazzi & F. Borson-Chazot. 2010. Thyroglobulin measurement in fine-needle aspirates of lymph nodes in patients with differentiated thyroid cancer: a simple definition of the threshold value, with emphasis on potential pitfalls of the method. Clin Chem Lab Med 48(8): 1171-1177.

Boux de Casson, F., V. Moal, A.-S. Gauchez, M.-P. Moineau, C. Sault, M.-H. Schlageter & C. Massart. 2017. Thyroglobulin assay in fluids from lymph node fine needle-aspiration washout: influence of pre-analytical conditions. Ann Biol Clin (Paris) 75(2): 173-180.

Cai, S., C. Qin, M. Liu, Y. Liu, H. Tang, W. Xu, M. Yin, Q. Ji, T. Liao & Y. Wang. 2025. Decoding lymph node skip metastasis: impact on long-term outcomes in stage T1 papillary thyroid carcinoma. Eur J Endocrinol 192(3): 318-326.

Chammas, M. C. 2019. Evaluation of cervical lymph nodes (duplex scan and color Doppler). Ultrasound in Medicine and Biology 45: S72.

Chang, L. C., Y. Q. Zhang, J. L. Zhu, L. F. Hu, X. Q. Wang, H. Z. Zhang, Q. Gu, X. Y. Chen, S. Zhang, M. Gao & X. Wei. 2023. An integrated nomogram combining deep learning, clinical characteristics and ultrasound features for predicting central lymph node metastasis in papillary thyroid cancer: A multicenter study. Frontiers in Endocrinology 14.

Chen, L., B. Dong, L. Jiang, J. Zhang, L. Chen, T. Li, Y. Shao & X. Sun. 2022. Microbubble contrast agent SonoVue: An efficient medium for the preoperative lymphatic mapping of thyroid carcinoma. Frontiers in Bioengineering and Biotechnology 10: 1077145.

Chen, L., L. Chen, J. Liu, B. Wang & H. Zhang. 2020. Value of Qualitative and Quantitative Contrast-Enhanced Ultrasound Analysis in Preoperative Diagnosis of Cervical Lymph Node Metastasis from Papillary Thyroid Carcinoma. J Ultrasound Med 39(1): 73-81.

Chen, L., L. Yuan, W. Mingxiao, D. Weide, S. Jinghai & W. Yong. 2025. Kupffer Phase Radiomics Signature in Sonazoid Contrast-Enhanced Ultrasound Predicts Immunohistochemistry Marker Expression in Hepatocellular Carcinoma. Cancer Med 14(19).

Chen, L., L. Yuan, X. Jingyong, S. Jinghai, W. Mingxiao & C. Jian. 2021. Contrast-Enhanced Intraoperative Ultrasonography with Kupffer Phase May Change Treatment Strategy of Metastatic Liver Tumors - A Single-Centre Prospective Study. Ther Clin Risk Manag 17(0), 789-796.

Chen, Y., Y. Y. Wang, Z. H. Cai & M. A. Jiang. 2021. Predictions for Central Lymph Node Metastasis of Papillary Thyroid Carcinoma via CNN-Based Fusion Modeling of Ultrasound Images. Traitement Du Signal 38(3): 629-638.

Chung, S. R., J. H. Baek, Y. H. Rho, Y. J. Choi, T.-Y. Sung, D. E. Song, T. Y. Kim & J. H. Lee. 2022. Sonographic Diagnosis of Cervical Lymph Node Metastasis in Patients with Thyroid Cancer and Comparison of European and Korean Guidelines for Stratifying the Risk of Malignant Lymph Node. Korean J Radiol 23(11): 1102-1111.

Chung, S. Y., E. K. Kim, J. H. Kim, K. K. Oh, D. J. Kim, Y. H. Lee, H. J. An & J. S. Kim. 2001. Sonographic findings of metastatic disease to the thyroid. Yonsei Medical Journal 42(4): 411-417.

Coroller, T. P., V. Agrawal, E. Huynh, V. Narayan, S. W. Lee, R. H. Mak & H. J. W. L. Aerts. 2017. Radiomic-Based Pathological Response Prediction from Primary Tumors and Lymph Nodes in NSCLC. J Thorac Oncol 12(3): 467-476.

Dai, Q., Y. Tao, D. Liu, C. Zhao, D. Sui, J. Xu, T. Shi, X. Leng & M. Lu. 2023. Ultrasound radiomics models based on multimodal imaging feature fusion of papillary thyroid carcinoma for predicting central lymph node metastasis. Front Oncol 13: 1261080.

Deng, P., X. Han, X. Wei & L. Chang. 2022. Automatic classification of thyroid nodules in ultrasound images using a multi-task attention network guided by clinical knowledge. Computers in Biology and Medicine 150.

Di Filippo, G., G. L. Canu, G. Gobbo, F. Medas, L. Rossi, F. Cappellacci, P. Papini, M. Paternoster, A. Chorti, I. Pliakos, M. Moysidis, G. Lazzari, E. Morelli, D. Serbusca, A. Ruzzenente, T. Papavramidis, G. Materazzi & P. G. Calò. 2026. Skip metastases in papillary thyroid carcinoma: evidence from a multicenter European retrospective study. Front Endocrinol (Lausanne) 17: 1712563.

Ding, X. H., Y. T. Liu, J. J. Zhao, R. Wang, C. F. Li, Q. Y. Luo & C. T. Shen. 2023. A novel wavelet-transform-based convolution classification network for cervical lymph node metastasis of papillary thyroid carcinoma in ultrasound images. Computerized Medical Imaging and Graphics 109.

Dong, S., J. Pan, X.-L. Du, X.-J. Xie, Q. Xia & Y.-J. Wu. 2024. Prediction of bilateral thyroid carcinoma and lateral cervical lymph node metastasis in PTC patients with suspicious thyroid nodules. Endocrine 85(2): 803-810.

Duval, M. A. S., Zanella, A. B., Cristo, A. P., Faccin, C. S., Graudenz, M. S. & Maia, A. L. 2017. Impact of serum TSH and anti-thyroglobulin antibody levels on lymph node fine-needle aspiration thyroglobulin measurements in differentiated thyroid cancer patients. Eur Thyroid J 6(6): 292-297.

Feng, J.-W., Y.-X. Yang, R.-J. Qin, S.-Q. Liu, A.-C. Qin & Y. Jiang. 2025. Application and validation of the machine learning-based multimodal radiomics model for preoperative prediction of lateral lymph node metastasis in papillary thyroid carcinoma. Frontiers in Endocrinology 16: 1618902.

Ferreira, R. C., L. L. Cunha, P. S. Matos, R. L. Adam, F. Soares, J. Vassallo & L. S. Ward. 2013. Chromatin changes in papillary thyroid carcinomas may predict patient outcome. Cellular Oncology 36(3): 259-264.

Filetti, S., C. Durante, D. Hartl, S. Leboulleux, L. D. Locati, K. Newbold, M. G. Papotti, A. Berruti & E. G. C. E. a. clinicalguidelines@esmo.org. 2019. Thyroid cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-updagger. Ann Oncol 30(12): 1856-1883.

Fu, Y., L.-G. Cui, J.-Y. Ma, M. Fang, Y.-X. Lin & N. Li. 2024. Development of a novel contrast-enhanced ultrasound-based nomogram for superficial lymphadenopathy differentiation: Postvascular phase value. Ultrasound in Medicine and Biology 50(6): 852-859.

Gillies, R. J., P. E. Kinahan & H. Hricak. 2016. Radiomics: Images Are More than Pictures, They Are Data. Radiology 278(2): 563-577.

Gu, Y., M. Yu, J. Deng & Y. Lai. 2025. Preoperative circulating tumor cells level is associated with lymph node metastasis in patients with unifocal papillary thyroid carcinoma. World J Surg Oncol 23(1): 47.

Guo, S. Y., P. Zhou, Y. Zhang, L. Q. Jiang & Y. F. Zhao. 2021. Exploring the Value of Radiomics Features Based on B-Mode and Contrast-Enhanced Ultrasound in Discriminating the Nature of Thyroid Nodules. Frontiers in Oncology 11.

Han, L., H. Chun-Jie, W. Min, L. Ke-Feng, L. Ying, D. Pei, S. Li-Tao & T. Jing-Lan. 2025. High-risk habitat radiomics model based on ultrasound images for predicting lateral neck lymph node metastasis in differentiated thyroid cancer. BMC Med Imaging 25(1).

Han, N., J. R. Fan, D. W. Chen & Y. P. Wang. 2024. APPLICATION OF ARTIFICIAL INTELLIGENCE IN THE DIAGNOSIS OF THYROID CANCER WITH ENHANCED COMPUTED TOMOGRAPHY. Journal of Mechanics in Medicine and Biology 24(02).

Haugen, B. R., E. K. Alexander, K. C. Bible, G. M. Doherty, S. J. Mandel, Y. E. Nikiforov, F. Pacini, G. W. Randolph, A. M. Sawka, M. Schlumberger, K. G. Schuff, S. I. Sherman, J. A. Sosa, D. L. Steward, R. M. Tuttle & L. Wartofsky. 2016. 2015 American Thyroid Association Management Guidelines for Adult Patients with Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. Thyroid 26(1): 1-133.

Hong, Y.-R., Z.-Y. Luo, G.-Q. Mo, P. Wang, Q. Ye & P.-T. Huang. 2017. Role of Contrast-Enhanced Ultrasound in the Pre-operative Diagnosis of Cervical Lymph Node Metastasis in Patients with Papillary Thyroid Carcinoma. Ultrasound Med Biol 43(11): 2567-2575.

Huang, Y., Y. Mao, L. Xu, J. Wen & G. Chen. 2022. Exploring risk factors for cervical lymph node metastasis in papillary thyroid microcarcinoma: construction of a novel population-based predictive model. BMC Endocrine Disorders 22(1).

Jeon, M. J., J. W. Park, J. M. Han, J. H. Yim, D. E. Song, G. Gong, T. Y. Kim, J. H. Baek, J. H. Lee, Y. K. Shong & W. B. Kim. 2013. Serum antithyroglobulin antibodies interfere with thyroglobulin detection in fine-needle aspirates of metastatic neck nodes in papillary thyroid carcinoma. J Clin Endocrinol Metab 98(1): 153-160.

Jeon, M. J., W. G. Kim, E. K. Jang, Y. M. Choi, Y.-M. Lee, T.-Y. Sung, J. H. Yoon, K.-W. Chung, S. J. Hong, J. H. Baek, J. H. Lee, T. Y. Kim, Y. K. Shong & W. B. Kim. 2015. Thyroglobulin level in fine-needle aspirates for preoperative diagnosis of cervical lymph node metastasis in patients with papillary thyroid carcinoma: two different cutoff values according to serum thyroglobulin level. Thyroid 25(4): 410-416.

Jiang, H. J. & P. J. Hsiao. 2020. Clinical application of the ultrasound-guided fine needle aspiration for thyroglobulin measurement to diagnose lymph node metastasis from differentiated thyroid carcinoma-literature review. Kaohsiung J Med Sci 36(4): 236-243.

Jiang, L., Z. Zhang, S. Guo, Y. Zhao & P. Zhou. 2023. Clinical-Radiomics Nomogram Based on Contrast-Enhanced Ultrasound for Preoperative Prediction of Cervical Lymph Node Metastasis in Papillary Thyroid Carcinoma. Cancers (Basel) 15(5).

Jiang, M., C. Li, S. Tang, W. Lv, A. Yi, B. Wang, S. Yu, X. Cui & C. F. Dietrich. 2020. Nomogram Based on Shear-Wave Elastography Radiomics Can Improve Preoperative Cervical Lymph Node Staging for Papillary Thyroid Carcinoma. Thyroid 30(6): 885-897.

Jun, Z., Z. Zhiyu, H. Xin, H. Xingyue, Z. Yao, H. Yugang, D. Qing, Z. Qing & Z. Qing. 2025. Super-resolution ultrasound quantifying microvascular alterations for early detection of metastatic cervical lymph nodes: a prospective diagnostic study. Sci Rep 16(1).

Junsong, L., W. Rui, X. Chongwen, Z. Ruimin, W. Shiyang, Z. Qian, L. Honghui, Y. Xiaobao, G. Rui, B. Yanxia & Z. Shaoqiang. 2025. FNA-Tg improves the diagnostic efficacy of FNAC for PTC lateral cervical LN metastasis. Sci Rep 15(1).

Ke, L., Z. Hongyan, J. Yuxin, L. Ping, X. Hui-Xiong, D. Lianfang, C. Yi-Hong, X. Xiaoyan, L. YuKun, L. Young Joon, L. Jae Young, H. Bing, L. Baoming, W. Yi, L. Ying, K. Christina, C. Kun, W. Wenping & L. Ja-Der. 2021. Prospective assessment of diagnostic efficacy and safety of Sonazoid(TM) and SonoVue(®) ultrasound contrast agents in patients with focal liver lesions. Abdom Radiol (NY) 46(10).

Kessler, A., Y. Rappaport, A. Blank, S. Marmor, J. Weiss & M. Graif. 2003. Cystic appearance of cervical lymph nodes is characteristic of metastatic papillary thyroid carcinoma. J Clin Ultrasound 31(1): 21-25.

Khadra, H., H. Mohamed, Z. Al-Qurayshi, A. Sholl, M. Killackey & E. Kandil. 2019. Superior detection of metastatic cystic lymphadenopathy in patients with papillary thyroid cancer by utilization of thyroglobulin washout. Head Neck 41(1): 225-229.

Kim, D. W., S. J. Jeon & C. G. Kim. 2012. Usefulness of thyroglobulin measurement in needle washouts of fine-needle aspiration biopsy for the diagnosis of cervical lymph node metastases from papillary thyroid cancer before thyroidectomy. Endocrine 42(2): 399-403.

Kim, K., J. S. Bae & J. S. Kim. 2021. Measurement of thyroglobulin level in lateral neck lymph node fine needle aspiration washout fluid in papillary thyroid cancer. Gland Surg 10(9): 2686-2694.

Kim, S. Y., E. K. Kim, H. J. Moon, J. H. Yoon & J. Y. Kwak. 2015. Application of texture analysis in the differential diagnosis of benign and malignant thyroid nodules: Comparison with gray-scale ultrasound and elastography. American Journal of Roentgenology 205(3): W343-W351.

Kim, S. Y., E. Lee, S. J. Nam, E. K. Kim, H. J. Moon, J. H. Yoon, K. H. Han & J. Y. Kwak. 2017. Ultrasound texture analysis: Association with lymph node metastasis of papillary thyroid microcarcinoma. PLoS ONE 12(4).

Kuo, P.-C., W.-C. Chen, W.-C. Lin, S.-Y. Chen, Y.-H. Chen, Y.-C. Yeh & C.-K. Chen. 2025. Clinical implications of lymph node thyroglobulin in papillary thyroid carcinoma metastases: Independent from thyroglobulin antibody interference. International Journal of Molecular Sciences 26(11): 5340.

Lambin, P., E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R. G. P. M. van Stiphout, P. Granton, C. M. L. Zegers, R. Gillies, R. Boellard, A. Dekker & H. J. W. L. Aerts. 2012. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer 48(4): 441-446.

Lambin, P., R. T. H. Leijenaar, T. M. Deist, J. Peerlings, E. E. C. de Jong, J. van Timmeren, S. Sanduleanu, R. Larue, A. J. G. Even, A. Jochems, Y. van Wijk, H. Woodruff, J. van Soest, T. Lustberg, E. Roelofs, W. van Elmpt, A. Dekker, F. M. Mottaghy, J. E. Wildberger & S. Walsh. 2017. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 14(12): 749-762.

Lee, C. Y., S. K. Snyder, T. C. Lairmore, S. C. Dupont & D. C. Jupiter. 2012. Utility of surgeon-performed ultrasound assessment of the lateral neck for metastatic papillary thyroid cancer. J Oncol 2012: 973124.

Leenhardt, L., M. F. Erdogan, L. Hegedus, S. J. Mandel, R. Paschke, T. Rago & G. Russ. 2013. 2013 European Thyroid Association guidelines for cervical ultrasound scan and ultrasound-guided techniques in the postoperative management of patients with thyroid cancer. Eur Thyroid J 2(3): 147-159.

Lentge, F., Korn, P., Gellrich, N.-C., Neuhaus, M.-T., Bettag, S. A., Linderkamp, M. L. & Jehn, P. 2025. Comparison of ultrasonographic (SonoVue) and radiographic (Imeron) contrast agents for intralesional application in the head and neck area: An in vitro study. Ultrasound.

Li, J., F. T. Xia, X. Q. Wang, Y. Jin, J. Yan, X. Wei & Q. Zhao. 2023. Multiclassifier Radiomics Analysis of Ultrasound for Prediction of Extrathyroidal Extension in Papillary Thyroid Carcinoma in Children. International Journal of Medical Sciences 20(2): 278-286.

Li, M. H., L. Liu, L. Feng, L. J. Zheng, Q. M. Xu, Y. J. Zhang, F. R. Zhang & L. N. Feng. 2024. Prediction of cervical lymph node metastasis in solitary papillary thyroid carcinoma based on ultrasound radiomics analysis. Front Oncol 14: 1291767.

Li, Q. L., T. Ma, Z. J. Wang, L. Huang, W. Liu, M. Chen, T. Sang, X. G. Ren, J. Tong, C. L. Cao, J. Dong & J. Li. 2022. The value of contrast-enhanced ultrasound for the diagnosis of metastatic cervical lymph nodes of papillary thyroid carcinoma: A systematic review and meta-analysis. J Clin Ultrasound 50(1): 60-69.

Li, T., J. Sheng, W. Li, X. Zhang, H. Yu, X. Chen, J. Zhang, Q. Cai, Y. Shi & Z. Liu. 2015. A new computational model for human thyroid cancer enhances the preoperative diagnostic efficacy. Oncotarget 6(29): 28463-28477.

Li, Y., W. Sun, X. Liao, M. Zhang, F. Xie, D. Chen, Y. Zhang & Y. Luo. 2021. A Thyroid Ultrasound Image-based Artificial Intelligence Model for Diagnosis of Central Compartment Lymph Node Metastasis in Papillary Thyroid Carcinoma. Acta Academiae Medicinae Sinicae 43(6): 911-916.

Liu, C., S. Yang, T. Xue, Q. Zhang, Y. Zhang, Y. Zhao, G. Yin, X. Yan, P. Liang & L. Liu. 2024. The application of a clinical-multimodal ultrasound radiomics model for predicting cervical lymph node metastasis of thyroid papillary carcinoma. Front Oncol 14: 1507953.

Liu, N., L. Lan, W. Lin, Z. Zhong, Z. Wu, Z. Du, Y. Chen, L. Tang & S. Wu. 2026. A stepwise diagnostic strategy combining a simplified US-based Node-RADS with postvascular phase perfluorobutane-CEUS for indeterminate cervical lymph nodes: A dual-center retrospective study. Ultrasound in Medicine and Biology 52(8): 1502-1511.

Liu, N., L. Tang, Y. Chen, Y. Wang, W. Huang, Z. Du, Y. Shen, Z. Wu, T. He, G. Su, W. Xie & Y. Chen. 2022. A Combination of Contrast-Enhanced Ultrasound and Thyroglobulin Level in Fine-Needle Aspirates Improves Diagnostic Accuracy for Metastatic Lymph Nodes of Papillary Thyroid Carcinoma. J Ultrasound Med 41(10): 2431-2443.

Liu, Q., Y. Li, Y. Hao, W. Fan, J. Liu, T. Li & L. Liu. 2024. Multi-modal ultrasound multistage classification of PTC cervical lymph node metastasis via DualSwinThyroid. Front Oncol 14: 1349388.

Liu, T., S. Zhou, J. Yu, Y. Guo, Y. Wang, J. Zhou & C. Chang. 2019. Prediction of Lymph Node Metastasis in Patients with Papillary Thyroid Carcinoma: A Radiomics Method Based on Preoperative Ultrasound Images. Technol Cancer Res Treat 18: 1533033819831713.

Liu, Z. Q., X. W. Zhang, X. H. Zhao, Q. Q. Guo, Z. J. Li, M. H. Wei, L. J. Niu & C. M. An. 2024. Combining radiomics with thyroid imaging reporting and data system to predict lateral cervical lymph node metastases in medullary thyroid cancer. Bmc Medical Imaging 24(1).

Liu, R.-B., D.-L. Zhou, B.-H. Xu, X.-H. Yang, Q. Li, X. Zhou, T. Tao, Z.-L. Yu & Y. Li. 2021. Comparison of the diagnostic performances of US-guided fine needle aspiration cytology and thyroglobulin measurement for lymph node metastases in patients with differentiated thyroid carcinoma: a meta-analysis. Eur Radiol 31(5): 2903-2914.

Lu, J., X. Wu, W. Wang, Z. Chen, C. Jin, D. Zhao, K. Zhang & F. Dong. 2025. Conventional ultrasound combined with contrast-enhanced ultrasound predicts lateral lymph node metastasis in papillary thyroid carcinoma. Clin Hemorheol Microcirc 89(1): 83-94.

Lu, W. J., L. Mao, J. Li, L. Y. OuYang, J. Y. Chen, S. Y. Chen, Y. Y. Lin, Y. W. Wu, S. N. Chen, S. D. Qiu & F. Chen. 2023. Three-dimensional ultrasound-based radiomics nomogram for the prediction of extrathyroidal extension features in papillary thyroid cancer. Front Oncol 13: 1046951.

Lu, X., Y. Dang, B. Kondowe, H. Zhang, J. Shang, W. Wang & X. Wang. 2025. Combining Thyroglobulin Levels in Lymph Node Wash-out Fluid with TI-RADS to Predict Lymph Node Metastasis in Papillary Thyroid Carcinoma. Malawi Med J 36(5): 298-302.

Lu, Z., Z. Wei & Z. WeiWei. 2017. Role of ultrasound in the assessment of percutaneous laser ablation of cervical metastatic lymph nodes from thyroid carcinoma. Acta Radiol 59(4).

Luo, Z. Y., Y. R. Hong, C. X. Yan, Y. Wang, Q. Ye & P. Huang. 2022. Utility of quantitative contrast-enhanced ultrasound for the prediction of lymph node metastasis in patients with papillary thyroid carcinoma. Clin Hemorheol Microcirc 80(1): 37-48.

Machado, P., T. Vu, N. Rosenblum, A. Shafer, F. Forsberg & J.-B. Liu. 2025. Initial evaluation of lymphosonography to identify sentinel lymph nodes in cervical and vulvar cancer patients. Journal of Clinical Ultrasound 53(9): 2159-2164.

Mikosiński, P., E. Wołowiec-Korecka, L. Pomorski, A. Mikosińska, K. Kaczka & S. Mikosiński. 2023. Cut-off value for thyroglobulin washout concentration in the detection of cervical lymph node metastases in patients after thyroidectomy due to differentiated thyroid cancer. Biomedicines 11(9): 2433.

Mu, J., Y. Cao, X. Zhong, W. Diao & Z. Jia. 2023. Prediction of Lymph Node Metastasis in Differentiated Thyroid Cancer Based on Radiomics Models: A Metaanalysis and Systematic Review. European Journal of Nuclear Medicine and Molecular Imaging 50: S546-S547.

Mulla, M. G., W. T. Knoefel, J. Gilbert, A. McGregor & K. M. Schulte. 2012. Lateral cervical lymph node metastases in papillary thyroid cancer: a systematic review of imaging-guided and prophylactic removal of the lateral compartment. Clin Endocrinol (Oxf) 77(1): 126-131.

Murata, M. 2018. Inflammation and cancer. Environ Health Prev Med 23(1): 50.

Onoue, K., N. Fujima, V. C. Andreu-Arasa, B. N. Setty & O. Sakai. 2021. Cystic cervical lymph nodes of papillary thyroid carcinoma, tuberculosis and human papillomavirus positive oropharyngeal squamous cell carcinoma: utility of deep learning in their differentiation on CT. American Journal of Otolaryngology - Head and Neck Medicine and Surgery 42(5).

Pacini, F., L. Fugazzola, F. Lippi, C. Ceccarelli, R. Centoni, P. Miccoli, R. Elisei & A. Pinchera. 1992. Detection of thyroglobulin in fine needle aspirates of nonthyroidal neck masses: a clue to the diagnosis of metastatic differentiated thyroid cancer. J Clin Endocrinol Metab 74(6): 1401-1404.

Park, V. Y., E. Lee, H. S. Lee, H. J. Kim, J. Yoon, J. Son, K. Song, H. J. Moon, J. H. Yoon, G. R. Kim & J. Y. Kwak. 2021. Combining radiomics with ultrasound-based risk stratification systems for thyroid nodules: an approach for improving performance. European Radiology 31(4): 2405-2413.

Park, V. Y., K. Han, E. Lee, E. K. Kim, H. J. Moon, J. H. Yoon & J. Y. Kwak. 2019. Association Between Radiomics Signature and Disease-Free Survival in Conventional Papillary Thyroid Carcinoma. Scientific reports 9(1): 4501.

Park, V. Y., K. Han, H. J. Kim, E. Lee, J. H. Youk, E. K. Kim, H. J. Moon, J. H. Yoon & J. Y. Kwak. 2020. Radiomics signature for prediction of lateral lymph node metastasis in conventional papillary thyroid carcinoma. PLoS One 15(1): e0227315.

Qiang, L., Z. Bowen, L. Jianghong, S. Jinduo, X. Haishan, X. Lilong, G. Li, Z. Murui & Z. Jiang. 2016. [The value of thyroglobulin measurement in fine-needle aspiration for diagnosis of suspicious lymph nodes in patients with thyroid carcinoma after thyroidectomy]. Zhonghua Er Bi Yan Hou Tou Jing Wai Ke Za Zhi 51(5).

Rabasco Meneghetti, A., Zwanenburg, A., Linge, A., Lohaus, F., Grosser, M., Baretton, G. B., Kalinauskaite, G., Tinhofer, I., Guberina, M., Stuschke, M., Balermpas, P., von der Grün, J., Ganswindt, U., Belka, C., Peeken, J. C., Combs, S. E., Böke, S., Zips, D., Troost, E. G. C., Krause, M., Baumann, M. & Löck, S. 2022. Integrated radiogenomics analyses allow for subtype classification and improved outcome prognosis of patients with locally advanced HNSCC. Sci Rep 12: 16755.

Radzina, M., M. Ratniece, D. S. Putrins, L. Saule & V. Cantisani. 2021. Performance of Contrast-Enhanced Ultrasound in Thyroid Nodules: Review of Current State and Future Perspectives. Cancers (Basel) 13(21).

Ringel, M. D., J. A. Sosa, Z. Baloch, L. Bischoff, G. Bloom, G. A. Brent, P. L. Brock, R. Chou, R. R. Flavell, W. Goldner, E. G. Grubbs, M. Haymart, S. M. Larson, A. M. Leung, J. Osborne, J. A. Ridge, B. Robinson, D. L. Steward, R. P. Tufano & L. J. Wirth. 2025. 2025 American Thyroid Association Management Guidelines for Adult Patients with Differentiated Thyroid Cancer. Thyroid 35(8): 841-985.

Rui, D., Z. Ying, K. Jiedong, L. Jingting, S. Chengqiu, Z. Daqi, F. Yantao, Z. Le, F. Qingfeng, L. Fang, D. Gianlorenzo, L. Nan & S. Hui. 2024. A novel risk stratification model based on tumor size and multifocality to predict recurrence in pediatric PTC: comparison with adult PTC. Front Endocrinol (Lausanne) 14(0).

Ruiz, E. M. L., T. Niu, M. Zerfaoui, M. Kunnimalaiyaan, P. L. Friedlander, A. B. Abdel-Mageed & E. Kandil. 2020. A novel gene panel for prediction of lymph-node metastasis and recurrence in patients with thyroid cancer. Surgery (United States) 167(1): 73-79.

Sakamoto, K., H. Ozawa, Y. Sato, M. Nakaishi, A. Sakanushi, T. Matsunobu, K. Okubo & S. Shinden. 2024. Cutoff value of thyroglobulin in needle aspirates for screening neck masses of thyroid carcinoma. Endocrine-Related Cancer 31(12): e240067.

Sasaki, N., N. Kudo, K. Nakamura, S. Y. Lim, M. Murakami, W. R. Kumara, Y. Tamura, H. Ohta, M. Yamasaki & M. Takiguchi. 2012. Activation of microbubbles by short-pulsed ultrasound enhances the cytotoxic effect of cis-diamminedichloroplatinum (II) in a canine thyroid adenocarcinoma cell line in vitro. Ultrasound Med Biol 38(1): 109-118.

Scheckenbach, K., Colter, L. & Wagenmann, M. 2017. Radiomics in head and neck cancer: Extracting valuable information from data beyond recognition. ORL J Otorhinolaryngol Relat Spec 79(1-2): 65-71.

Shi, Y., Y. Zou, J. Liu, Y. Wang, Y. Chen, F. Sun, Z. Yang, G. Cui, X. Zhu, X. Cui & F. Liu. 2022. Ultrasound-based radiomics XGBoost model to assess the risk of central cervical lymph node metastasis in patients with papillary thyroid carcinoma: Individual application of SHAP. Front Oncol 12: 897596.

Shin, H. J., H. S. Lee, E.-K. Kim, H. J. Moon, J. H. Lee & J. Y. Kwak. 2015. A Study on Serum Antithyroglobulin Antibodies Interference in Thyroglobulin Measurement in Fine-Needle Aspiration for Diagnosing Lymph Node Metastasis in Postoperative Patients. PLoS One 10(6): e0131096.

Shin, Y. G., J. Yoo, H. J. Kwon, J. H. Hong, H. S. Lee, J. H. Yoon, E. K. Kim, H. J. Moon, K. Han & J. Y. Kwak. 2016. Histogram and gray level co-occurrence matrix on gray-scale ultrasound images for diagnosing lymphocytic thyroiditis. Computers in Biology and Medicine 75: 257-266.

Sidhu, P. S., V. Cantisani, C. F. Dietrich et al. 2018. The EFSUMB Guidelines and Recommendations for the Clinical Practice of Contrast-Enhanced Ultrasound (CEUS) in Non-Hepatic Applications: Update 2017 (Long Version). Ultraschall Med 39(2): e2-e44.

Sohn, Y. M., J. Y. Kwak, E. K. Kim, H. J. Moon, S. J. Kim & M. J. Kim. 2010. Diagnostic approach for evaluation of lymph node metastasis from thyroid cancer using ultrasound and fine-needle aspiration biopsy. AJR Am J Roentgenol 194(1): 38-43.

Song, G. S., F. Z. Xue & C. Q. Zhang. 2015. A Model Using Texture Features to Differentiate the Nature of Thyroid Nodules on Sonography. Journal of Ultrasound in Medicine 34(10): 1753-1760.

Stack, B. C. Jr, R. L. Ferris, D. Goldenberg, M. Haymart, A. Shaha, S. Sheth, J. A. Sosa & R. P. Tufano. 2012. American Thyroid Association consensus review and statement regarding the anatomy, terminology, and rationale for lateral neck dissection in differentiated thyroid cancer. Thyroid 22(5): 501-508.

Steyerberg, E. W., A. J. Vickers, N. R. Cook, T. Gerds, M. Gonen, N. Obuchowski, M. J. Pencina & M. W. Kattan. 2010. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 21(1): 128-138.

Su, X., L. Shang, C. Yue & B. Ma. 2024. Ultrasound-guided fine needle aspiration thyroglobulin in the diagnosis of lymph node metastasis of differentiated papillary thyroid carcinoma and its influencing factors. Front Endocrinol (Lausanne) 15: 1304832.

Surov, A., A. Machens, H. J. Holzhausen, R. P. Spielmann & H. Dralle. 2016. Radiological features of metastases to the thyroid. Acta Radiologica 57(4): 444-450.

Theek, B., T. Opacic, Z. Magnuska, T. Lammers & F. Kiessling. 2018. Radiomic analysis of contrast-enhanced ultrasound data. Scientific Reports 8.

Thigpen, J., S. Shah & Ieee. 2008. Multispectral Microscopy for Cell Differentiation in Thyroid Cytology. IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems, Seoul, SOUTH KOREA.

Tian, D., X. Li & Z. Jia. 2024. Analysis of Risk Factors and Risk Prediction for Cervical Lymph Node Metastasis in Thyroid Papillary Carcinoma. Cancer Manag Res 16: 1571-1585.

Tian, X., L. Chang, L. Jing-Jing, H. Yan-Hong, S. Yan-Ping, Z. Xiu-Xiu, Z. Yan-Jing, Z. Yu-Fang & L. Li-Ping. 2022. Analysis of the Relevance of the Ultrasonographic Features of Papillary Thyroid Carcinoma and Cervical Lymph Node Metastasis on Conventional and Contrast-Enhanced Ultrasonography. Front Oncol 11(0).

Tong, Y. Y., J. W. Zhang, Y. Wei, J. H. Yu, W. W. Zhan, H. S. Xia, S. C. Zhou, Y. Y. Wang & C. Chang. 2022. Ultrasound-based radiomics analysis for preoperative prediction of central and lateral cervical lymph node metastasis in papillary thyroid carcinoma: a multi-institutional study. Bmc Medical Imaging 22(1).

Tran, W. T., H. Suraweera, K. Quaioit, D. Cardenas, K. X. Leong, I. Karam, I. Poon, D. Jang, L. Sannachi, M. Gangeh, S. Tabbarah, A. Lagree, A. Sadeghi-Naini & G. J. Czarnota. 2019. Predictive quantitative ultrasound radiomic markers associated with treatment response in head and neck cancer. Future Sci OA 6(1): FSO433.

Vickers, A. J. & E. B. Elkin. 2006. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 26(6): 565-574.

Wakisaka, N., Endo, K., Kitazawa, T., Shimode, Y., Kato, K., Moriyama-Kita, M., Koda, W., Ikeda, H., Ishikawa, K., Ueno, T., Nakanishi, Y., Kondo, S., Sugimoto, H., Yoshimura, K., Tsuji, H., Kawashiri, S., Omoto, K. & Yoshizaki, T. 2019. Detection of sentinel lymph node using contrast-enhanced agent, Sonazoid™, and evaluation of its metastasis with superb microvascular imaging in oral and oropharyngeal cancers: A preliminary clinical study. Acta Otolaryngol 139(1): 94-99.

Wang, J., X. Jiang, G. Xiao, W. Zhou & Y. Hu. 2020. Excellent diagnostic performance of FNA-Tg in detecting lymph nodes metastases from papillary thyroid cancer. Future Oncology 16(33): 2735-2746.

Wang, Y., H. L. Tan, S. L. Duan, N. Li, L. Ai & S. Chang. 2024. Predicting central cervical lymph node metastasis in papillary thyroid microcarcinoma using deep learning. PeerJ 12: e16952.

Wang, Y., S. Zhang, M. Zhang, G. Zhang, Z. Chen, X. Wang, Z. Yang, Z. Yu, H. Ma, Z. Wang & L. Sang. 2024. Prediction of lateral lymph node metastasis with short diameter less than 8 mm in papillary thyroid carcinoma based on radiomics. Cancer Imaging 24(1): 155.

Wang, Y., Y. Duan, H. Li, K. Yue, J. Liu, Q. Lai, M. Zhou, B. Ye, Y. Wu, J. Zhu, P. Chen, C. Jing, Y. Wu & X. Wang. 2022. Detection of thyroglobulin in fine-needle aspiration for diagnosis of metastatic lateral cervical lymph nodes in papillary thyroid carcinoma: A retrospective study. Frontiers in Oncology 12: 909723.

Wei, T. J., W. Wei, Q. Ma, Z. B. Shen, K. B. Lu & X. M. Zhu. 2023. Development of a Clinical-Radiomics Nomogram That Used Contrast-Enhanced Ultrasound Images to Anticipate the Occurrence of Preoperative Cervical Lymph Node Metastasis in Papillary Thyroid Carcinoma Patients. International Journal of General Medicine 16: 3921-3932.

Wei, X., Y. Li, S. Zhang & M. Gao. 2014. Prediction of thyroid extracapsular extension with cervical lymph node metastases (ECE-LN) by CEUS and BRAF expression in papillary thyroid carcinoma. Tumour Biol 35(9): 8559-8564.

Wei, Y., M. A. Yu, Y. Niu, Y. Hao, J. X. Di, Z. L. Zhao, X. J. Cao, L. L. Peng & Y. Li. 2021. Combination of Lymphatic and Intravenous Contrast-Enhanced Ultrasound for Evaluation of Cervical Lymph Node Metastasis from Papillary Thyroid Carcinoma: A Preliminary Study. Ultrasound Med Biol 47(2): 252-260.

Wei, Y., Y. Niu, Z. L. Zhao, X. J. Cao, L. L. Peng, Y. Li & M. A. Yu. 2022. Effectiveness of Lymphatic Contrast Enhanced Ultrasound in the diagnosis of Cervical Lymph node metastasis from papillary thyroid carcinoma. Sci Rep 12(1): 578.

Wen, Q., Z. X. Wang, A. Traverso, Y. J. Liu, R. F. Xu, Y. Feng & L. X. Qian. 2022. A radiomics nomogram for the ultrasound-based evaluation of central cervical lymph node metastasis in papillary thyroid carcinoma. Frontiers in Endocrinology 13.

Wilson, S. R. & P. N. Burns. 2010. Microbubble-enhanced US in body imaging: what role? Radiology 257(1): 24-39.

Xiao, L., J. Zhou, W. Tan, Y. Liu, H. Zheng, G. Wang, W. Zheng, X. Pei, A. Yang & L. Liu. 2023. Contrast-enhanced US with Perfluorobutane to Diagnose Small Lateral Cervical Lymph Node Metastases of Papillary Thyroid Carcinoma. Radiology 307(4): e221465.

Xu, Y., D. Wu, W. Wu, J. Jiang, C. Xi, N. Yu, Y. Wang & X. Xu. 2019. Diagnostic value of cytology, thyroglobulin, and combination of them in fine-needle aspiration of metastatic lymph nodes in patients with differentiated thyroid cancer: A systematic review and network meta-analysis. Medicine (Baltimore) 98(45): e17859.

Xue, J., S. Y. Li, N. A. Qu, G. Y. Wang, H. Z. A. Chen, Z. H. Wu & X. L. Cao. 2023. Value of clinical features combined with multimodal ultrasound in predicting lymph node metastasis in cervical central area of papillary thyroid carcinoma. Journal of Clinical Ultrasound 51(5): 908-918.

Yilmaz, E., Ismaila, N., Dabney, R., Saba, N. F. & Mell, L. K. 2023. Immunotherapy and biomarker testing in recurrent and metastatic head and neck cancers: ASCO guideline Q and A. JCO Oncol Pract 19(4): 194-196.

Ying, Y., R. Jiliang, S. Yiqian & T. Xiaofeng. 2019. MRI-based radiomic signature as predictive marker for patients with head and neck squamous cell carcinoma. Eur J Radiol 117(0).

Yoo, J., J. M. Lee, H.-J. Kang & J. S. Bae. 2026. Low-mechanical-index B-mode phase-inversion harmonic imaging significantly improves hepatic lesion conspicuity during Sonazoid Kupffer phase: a prospective comparative study. Ultrasonography 45(2): 163-173.

Yu, C.-Y., E. H.-L. Chu, C.-H. Lin & Y.-C. Chen. 2025. Diagnostic Efficiency of Thyroglobulin in Lymph Node Fine-needle Aspiration Washout: A Systematic Review and Meta-analysis. J Clin Endocrinol Metab 110(12): 3569-3587.

Yu, Y., L. L. Shi, H. W. Zhang & Q. Wang. 2023. Performance of contrast-enhanced ultrasound for lymph node metastasis in papillary thyroid carcinoma: a meta-analysis. Endocr Connect 12(2): e220341.

Yuan, W. H., H. J. Chiou, Y. H. Chou, H. C. Hsu, C. M. Tiu, C. Y. Cheng & C. H. Lee. 2006. Gray-scale and color Doppler ultrasonographic manifestations of papillary thyroid carcinoma: analysis of 51 cases. Clinical Imaging 30(6): 394-401.

Zhan, J., X. Diao, Y. Chen, W. Wang & H. Ding. 2019. Predicting cervical lymph node metastasis in patients with papillary thyroid cancer (PTC) - Why contrast-enhanced ultrasound (CEUS) was performed before thyroidectomy. Clin Hemorheol Microcirc 72(1): 61-73.

Zhang, L.-Y., Y. Chen & Y.-Z. Ao. 2022. Value of thyroglobulin combined with ultrasound-guided fine-needle aspiration cytology for diagnosis of lymph node metastasis of thyroid carcinoma. World J Clin Cases 10(2): 492-501.

Zhang, Y., S. Yu, L. Zhang & L. Kang. 2020. Radiomics Based on CECT in Differentiating Kimura Disease rom Lymph Node Metastases in Head and Neck: A Non-Invasive and Reliable Method. Front Oncol 10: 1121.

Zhang, H., S. Hu, X. Wang, J. He, W. Liu, C. Yu, Z. Sun, Y. Ge & S. Duan. 2020. Prediction of Cervical Lymph Node Metastasis Using MRI Radiomics Approach in Papillary Thyroid Carcinoma: A Feasibility Study. Technol Cancer Res Treat 19: 1533033820969451.

Zhang, M., Y. Zhang, H. Wei, L. Yang, R. Liu, B. Zhang & S. Lyu. 2023. Ultrasound radiomics nomogram for predicting large-number cervical lymph node metastasis in papillary thyroid carcinoma. Frontiers in Oncology 13.

Zhang, Q., X. Liang, Y. Zhang, H. Nie & Z. Chen. 2023. A review of contrast-enhanced ultrasound using SonoVue® and Sonazoid™ in non-hepatic organs. Eur J Radiol 167: 111060.

Zhao, H.-N., H. Yin, M.-H. Li, H.-Q. Zhang, Y.-S. He, H.-H. Luo, B.-Y. Ma, L. Ma, D.-Q. Liu & Y.-L. Peng. 2024. Contrast-enhanced ultrasound image sequences based on radiomics analysis for diagnosis of metastatic cervical lymph nodes from thyroid cancer. Gland Surgery 13(8): 1437-1447.

Zhao, Z. J., C. M. Yang, Q. Wang, H. W. Zhang, L. L. Shi & Z. W. Zhang. 2021. A deep learning-based method for detecting and classifying the ultrasound images of suspicious thyroid nodules. Medical Physics 48(12): 7959-7970.

Zhou, S. C., T. T. Liu, J. Zhou, Y. X. Huang, Y. Guo, J. H. Yu, Y. Y. Wang & C. Chang. 2020. An Ultrasound Radiomics Nomogram for Preoperative Prediction of Central Neck Lymph Node Metastasis in Papillary Thyroid Carcinoma. Front Oncol 10: 1591.

Zhou, Y., X. Xu, G. Su, X. Tao, Y. Ge, Y. Si, M. Shen & F. Wu. 2021. Radiomics based on arterial-venous mixed images derived from dual-energy CT data in diagnosis of lymph nodes metastasis of papillary thyroid cancer. Chinese Journal of Radiology (China) 55(7): 703-709.

Zhu, J., L. Chang, D. Li, B. Yue, X. Wei, D. Li & X. Wei. 2023. Nomogram for preoperative estimation risk of lateral cervical lymph node metastasis in papillary thyroid carcinoma: A multicenter study. Cancer Imaging 23: 55.

Zhu, X.-H., J.-N. Zhou, Y.-Y. Qian, K. Yang, Q.-L. Wen, Q.-H. Zhang, L. Xia, M.-H. Ge & C.-X. Sun. 2020. Diagnostic values of thyroglobulin in lymph node fine-needle aspiration washout: a systematic review and meta-analysis diagnostic values of FNA-Tg. Endocr J 67(2): 113-123.

Zou, Y., Y. Shi, J. H. Liu, G. H. Cui, Z. Yang, M. L. Liu & F. Sun. 2021. A Comparative Analysis of Six Machine Learning Models Based on Ultrasound to Distinguish the Possibility of Central Cervical Lymph Node Metastasis in Patients With Papillary Thyroid Carcinoma. Frontiers in Oncology 11.

APPENDIX A

ETHICAL APPROVAL FROM THE AFFILIATED HOSPITAL OF PAN ZHIHUA UNIVERSITY, CHINA

PCWEC-C-003

攀枝花学院附属医院伦理审查委员会

科研项目伦理审查批件

批件号：（研）伦审第 2024-02-001 号

项目名称	基于 Sonazoid 微球超声造影微循环灌注联合 FNA-Tg 构建 Radiomics 模型预测甲状腺乳头状癌颈侧区淋巴结转移				
项目编号					
项目类别	临床试验	经费来源	自筹	起止时间	2024.6-2026.12
申请科室	超声科	科室负责人	王瑜	电话	13320712097
项目负责人信息	姓名	刘燕		性别	女
	学历	硕士		职称	副主任医师
	联系电话	18096307906		电子邮箱	304176240@qq.com
	主要研究方向	腹部与浅表器官超声造影及超声介入			
委员会应到人数	17	委员会实到人数	9	委员会同意人数	9
审查方式	☐快速审查 ☑会议审查 ☐紧急情况受试者审查 ☐应急审查				

单位审批意见：经伦理审查，审查结果为批准。具体意见如下：

课题实施中严格按照国家有关法律、法规、及研究方案开展研究。

攀枝花学院附属医院伦理委员会

（副）主任委员签字：

日期：2024 年 6 月 14 日

批件有效期：☐六个月 ☑一年

地址：攀枝花市东区炳草岗桃源街 27 号

邮编：617000 电话：0812-2213396

V2.0/2022-07-12

标准操作规程/Standard Operating Procedures

1

PCWEC-C-003

Ethics Review Committee of Affiliated Hospital of Panzhihua University

Research project ethics review approval

Approval Number: (Research) No. 2024-02-001

Project Title	Contrast-enhanced ultrasound microperfusion of Sonazoid microsphere combined with FNA-Tg to construct radiomics to predict lateral cervical lymph node metastasis in papillary thyroid carcinoma clinical application study				
Project number					
Project types	clinical study	financial resources	self funding	starting and ending time	2024.6-2026.12
Application department	Ultrasound department	Head of department	Wang Yu	telephone	13320712097
Principal Investigator	name	LiuYan		gender	female
	education background	PhD		professional title	associate chief physician
	telephone	18096307906		E-mail	304176240@qq.com
	main direction of studying	Contrast-enhanced ultrasound and ultrasound intervention in abdominal and superficial organs			
The number of members of the committee should be present	17	The actual number of committee members	9	Committee consent	9
method of review	☐Expedited review ☑Conference review ☐Emergency subject review ☐Emergency review				

Approval comments of the unit: After the ethical review, the review result is approved. Specific suggestions are as follows:

In the implementation of the project, the research is carried out in strict accordance with the relevant national laws, regulations and research programs.

Signature of the (Vice) Chairman:

Date: June 14, 2024

Validity period of approval: ☐Six months ☑One year

Address: No. 27 Taoyuan Street, Bingcaogang, Panzhihua City, sichuan , China

postcodes: 617000 telephone: 0812-2213396

APPENDIX B

ETHICAL APPROVAL FROM UNIVERSITI KEBANGSAAN MALAYSIA (UKM)

UNIVERSITI KEBANGSAAN MALAYSIA
The National University of Malaysia

SEKRETARIAT ETIKA PENYELIDIKAN UKM • UKM RESEARCH ETHICS SECRETARIAT

Reference: UKM PPI/111/8
Date : 24 December 2024

Associate Profesor Dr. Hanani Abdul Manan
Department of Radiology
Faculty of Medicine UKM

Dear Professor/Datuk/Dato'/Datin/Sir/Madam,

ETHICAL APPROVAL TO CONDUCT RESEARCH IN THE NATIONAL UNIVERSITY OF MALAYSIA

Research title	:	Contrast-Enhanced Ultrasound Microperfusion Of Sonazoid Microsphere Combined With FNA-Tg To Construct Radiomics To Predict Lateral Cervical Lymph Node Metastasis In Papillary Thyroid Carcinoma Clinical Application Study
Ethics Ref. No.	:	JEP-2024-867
Approval period	:	12 December 2024 – 11 September 2027
Study Site	:	1. Malaysia 2. China
Sample Size	:	134 Subjects

With reference to the above.

2. The Research Ethics Committee, The National University of Malaysia (RECUKM) has provided ethical approval for above study. Please be reminded permission from the Deputy Dean of research of the faculty or Director of Institute/Centre and all relevant heads of departments / units where the study will be carried out must be obtained prior to the study. You are required to follow and comply with their decision and all other relevant regulations.

3. Please note that the investigator's responsibility is to ensure that:

i. All Adverse Events should be reported to RECUKM as soon as possible.
ii. Progress report should be submitted every **6 Months.**
iii. All changes / amendments to the study documents / study sites / study team must be notified to the RECUKM. All changes must be approved by RECUKM before continuation of the study.
iv. Application for renewal of the approval has to be submitted to RECUKM within one month (1 month) prior to the expiry of ethical approval.
v. Final Report should be provided to RECUKM when the project is complete.
vi. Please take note that all records and data are to be kept strictly CONFIDENTIAL and can only be used for the purpose of this study. All precautions are to be taken to maintain data confidentiality.

Required forms can be obtained from the The Research Ethics Committee, The National University of Malaysia (RECUKM) website: **https://www.ukm.my/jepukm**

Thank you.

Sekretariat Etika Penyelidikan Universiti Kebangsaan Malaysia
Tingkat 1, Blok Klinikal, Hospital Canselor Tuanku Muhriz,Pusat Perubatan UKM, Jalan Yaacob Latif, Bandar Tun Razak, 56000 Cheras Kuala Lumpur.
Telefon: +603-9145 5046 / 5048
Email: sepukm@ukm.edu.my Web: https://www.ukm.my/jepukm/

ILMU, MUTU DAN BUDI

UNIVERSITI KEBANGSAAN MALAYSIA
The National University of Malaysia

SEKRETARIAT ETIKA PENYELIDIKAN UKM • UKM RESEARCH ETHICS SECRETARIAT

Yours sincerely,

PROFESOR DR MOHD SHAHRIR MOHAMED SAID
Chairman
Research Ethics Committee
The National University Of Malaysia

c.c.
- **Director**
  Hospital Canselor Tuanku Muhriz
- **Director**
  Hospital Pakar Kanak-Kanak UKM (HPKK UKM)
- **Deputy Dean (Research & Innovation)**
  Faculty of Medicine, UKM
- **Head of Department Radiology**
  Faculty of Medicine UKM
- **Professor Dr. Hamzaini Abdul Hamid**
- **Liu Yan (P142742 – PhD Candidate)**
  Department of Radiology
  Faculty of Medicine UKM
- **Associate Professor Dr. Noorazrul Azmie Yahya**
  Center for Diagnostic
  Therapeutic and Investigative Studies (CODTIS)
  Faculty of Health Sciences, UKM
- **Dr. Chai Jia Ning**
  Department of Radiology
  Hospital Pakar Kanak-Kanak UKM (HPKK UKM)
- Approval letter File 2024

UNIVERSITI KEBANGSAAN MALAYSIA
The National University of Malaysia

SEKRETARIAT ETIKA PENYELIDIKAN UKM • UKM RESEARCH ETHICS SECRETARIAT

NAME OF ETHICS COMMITTEE/IRB: Research Ethics Committee, The National University of Malaysia	ETHICS COMMITTEE/IRB REF NO : JEP-2024-867

PROTOCOL TITLE:
Contrast-Enhanced Ultrasound Microperfusion Of Sonazoid Microsphere Combined With FNA-Tg To Construct Radiomics To Predict Lateral Cervical Lymph Node Metastasis In Papillary Thyroid Carcinoma Clinical Application Study

PRINCIPAL INVESTIGATOR:
Associate Profesor Dr. Hanani Abdul Manan
Department of Radiology
Faculty of Medicine UKM

The following items [X] have been received and reviewed in connection with the above study to be conducted by the above investigator.

Documents

- [X] Research Application Form (UKM-JEP-BO01)
- [X] Review Form (UKM-JEP-BO02)
- [X] Non-Disclosure Agreement (UKM-JEP-BO03)
- [X] Conflict Of Interest Form
- [X] Research Proposal / Protocol
- [ ] Participant Information Sheet: [ ] Malay [ ] English [ ] Others :
- [X] Participant Consent Form: [ ] Malay [X] English [ ] Others :
- [ ] Questionnaire: [ ] Malay [ ] English [ ] Others :
- [X] Curriculum Vitae of Researchers: [X] Principal [X] Co-researcher [X] Student
- [X] Good Clinical Practice Certificate (GCP)
- [ ] Project Agreement

The Research Ethics Committee, The National University of Malaysia operates in accordance to the International Conference of Harmonization Good Clinical Practice Guidelines.

Comments (if any): _____

Date of Approval: 12 December 2024

PROFESSOR DR. **MOHD SHAHRIR MOHAMED SAID**
Chairman
Research Ethics Committee
The National University of Malaysia

APPENDIX C

RESEARCH INFORMED CONSENT FORM

攀枝花学院附属医院
Affiliated Hospital of Panzhihua University

UNIVERSITI KEBANGSAAN MALAYSIA
The National University of Malaysia

Informed Consent Form

We will carry out a study "Contrast-enhanced ultrasound microperfusion of Sonazoid microsphere combined with FNA-Tg to construct radiomics to predict lateral cervical lymph node metastasis in papillary thyroid carcinoma clinical application study". Your specific situation meets the eligibility criteria for this study, so we invite you to participate. This informed consent will introduce this study's purpose, steps, benefits, and risks to you. Please read it carefully before deciding whether to participate. When the researcher explains and discusses the informed consent with you, you can ask questions at any time and ask him/her to explain what you do not understand. You can discuss this with family, friends, and your doctor before deciding. If you are currently participating in another clinical study, be sure to inform your study physician or investigator.The project leader of the study is Yan Liu, deputy chief physician, at Affiliated Hospital of Pan Zhihua University.

Normal lymph nodes contain macrophages in the cortex and medulla, which are uniformly enhanced after phagocytosis of Sonazoid microspheres in the post-vascular phase. In metastatic lymph nodes, macrophages are destroyed by thyroid papillary carcinoma (PTC) cells and lose their phagocytosis of Sonazoid microspheres, resulting in changes in perfusion morphology in the post-vascular phase. In this study, we will combine FNA-Tg value, conventional ultrasonography, Sonazoid CEUS image feature, and Sonazoid CEUS radiomics signature to predict the metastasis of lateral cervical lymph nodes in PTC patients. It is of great significance to provide evidence-based medical evidence for whether to perform cervical lymph node dissection in PTC patients during the operation, avoid unnecessary sweeping damage and the omission of

early lymph node metastasis, promote the improvement of preoperative diagnosis and treatment level of papillary thyroid cancer, and develop individualized thyroidectomy, cervical lymph node dissection, postoperative treatment and follow-up strategies for patients.

We plan to collect 100-150 patients with papillary thyroid carcinoma from June 2024 to December 2026 in the Affiliated Hospital of Pan Zhihua University. 135-180 suspected lymph nodes with Sonazoid microspheres enhanced ultrasonography, and FNA-Tg biopsy will be performed, and a Radiomics model will be constructed. Subjects who met the following criteria: Inclusion criteria (1) All patients underwent thyroid surgery for the first time in our hospital, had thyroid nodule FNA before surgery, cytopathological diagnosis of papillary thyroid carcinoma on at least one side, suspicious ultrasonic signs of lymph nodes in the lateral cervical region, and lymph node dissection in the lateral cervical region; (2) After thyroid PTC follow-up, suspicious lymph nodes were found by ultrasound, and lymph nodes were dissected in the lateral cervical region; (3) The ultrasonographic characteristics of 1 or more suspicious lymph nodes were consistent with the following: ① the short diameter of lymph nodes was >0.5cm; ② Lymph node short diameter/long diameter (S/L) ratio ≥0.5; ③ The boundary is unclear or fuzzy ④ the lymphatic hilus is eccentric or missing; ⑤ Peripheral or diffuse blood flow signals; ⑥ internal echo is heterogenous, accompanied by cystic changes or clumpy hyper-echo; ⑦ Multiple foci calcification or cluster calcification. (4) The patient's medical records are complete. All the above conditions were included in the study. Exclusion criteria: ① lymph node metastases were confirmed by other malignant tumors pathologically; ② Tuberculosis of lymph nodes; ③ Other serious complications. Meeting any of these criteria was not included in the study.

During the study period, there may be a small amount of bleeding during the puncture of the lateral neck lymph node, which can be stopped with the pressure of the doctor. If you have any other discomfort, please inform the doctor in time. During the study period, if you are pregnant or think that you may become pregnant, you should immediately tell the study doctor.

Ultrasound-guided fine needle puncture of the cervical lymph node costs 177 RMB/time, sonazoid contrast agent 512RMB/time, and the fine needle 278RMB/time, which is paid by the research group, and other expenses should be borne by themselves. Participation in this study may change your surgical approach for radical papillary thyroid carcinoma. Depending on the results, you may choose whether to have a lateral neck lymph node dissection during radical papillary thyroid cancer surgery. You may choose not to participate in the study, which will have no adverse effect on your access to conventional treatment. Even after you agree to participate, you can change your mind at any time and tell the investigator to withdraw from the study, and this will not affect your access to normal medical care.

If you decide to participate in the study, your data in the study will be kept confidential, and information that can identify you will not be disclosed to anyone other than a member of the study unless you have given your permission.

If you have any questions about the study, please contact the doctor at +86 18096307906. If you have any questions related to the rights and interests of the subjects, you can contact the Ethics Committee of Pan Zhihua University Affiliated Hospital at +86 0812-2213396.

The researcher states:

I have explained to the subjects (and legal representatives) the research background, purpose, steps, risks, and benefits of (Contrast-enhanced ultrasound microperfusion of Sonazoid microsphere combined with FNA-Tg to construct radiomics to predict lateral cervical lymph node metastasis in papillary thyroid carcinoma clinical application study). Give him/her enough time to read the informed consent form, discuss it with others, and answer any questions he/she has about the study; I told the subject how to contact him for questions related to the study; I informed the subject (or legal representative) that he or she may withdraw from the study at any time during the study period without any reason.

Investigator Signature: Date:

The subject states:

The researcher explained to me the research background, purpose, steps, risks, and benefits of Contrast-enhanced ultrasound microperfusion of Sonazoid microsphere combined with FNA-Tg to construct radiomics to predict lateral cervical lymph node metastasis in papillary thyroid carcinoma clinical application study. I had enough time and opportunity to ask questions, and I was satisfied with the answers given by the researcher. I know who to contact when I have questions or want further information. I read this informed consent form and decided to participate in the study. I knew that I could withdraw from the study at any time during the study period without any reason. I was told that I would get a copy of this informed consent form, with my signature and the investigator's.

Subject's signature: Date:

Signature of Legal Representative [if applicable]: Date:

Relationship with subjects:

APPENDIX D

PROOF OF PUBLICATION 1

（Copy of the First Page of the Publication）

Original Investigation

Accuracy of Radiomics in the Identification of Extrathyroidal Extension and BRAF^{V600E} Mutations in Papillary Thyroid Carcinoma: A Systematic Review and Meta-analysis

Yan Liu, Ling Xiang, Fang-Yue Liu, Noorazrul Yahya, Jia-Ning Chai, Hamzaini Abdul Hamid, Qiang Lu, Hanani Abdul Manan

Rationale and Objectives: Extrathyroidal extension (ETE) and BRAF^{V600E} mutation in papillary thyroid cancer (PTC) increase mortality and recurrence risk. Preoperative identification presents considerable challenges. Although radiomics has emerged as a potential tool for identifying ETE and BRAF^{V600E} mutation, systematic evidence supporting its effectiveness remains insufficient. Therefore, this paper aims to determine the effectiveness of radiomics in detecting ETE and BRAF^{V600E} mutations in PTC.

Materials and Methods: PubMed, Web of Science, Cochrane, and Embase databases were searched until May 7th, 2024. The Radiomics Quality Score tool assessed bias risk. Subgroup analyses based on radiomics and clinical characteristics were conducted.

Results: Our systematic review included 19 studies, encompassing 5337 PTC cases. Among these, 12 articles focused on ETE and seven articles focused on BRAF^{V600E} mutations. For the identification of ETE in the validation set, the summarized machine learning (ML) models demonstrated 0.80c-index (95%CI: 0.77–0.83), 0.77 sensitivity (95%CI: 0.72–0.81), and 0.78 specificity (95%CI: 0.73–0.82). Radiomics based on ultrasound demonstrated 0.82c-index (95%CI: 0.78–0.86), 0.77 sensitivity (95%CI: 0.68–0.84), and 0.84 specificity (95%CI: 0.75–0.91). For the identification of BRAF^{V600E} mutations in the validation set, the summarized ML models showed 0.80c-index (95%CI: 0.72–0.87), 0.76 sensitivity (95%CI: 0.67–0.84), and 0.88 specificity (95%CI: 0.77–0.94). ML models based on ultrasound-guided radiomics had 0.81c-index (95%CI: 0.74–0.89), 0.79 sensitivity (95%CI: 0.71–0.86), and 0.87 specificity (95%CI: 0.74–0.94).

Conclusion: Radiomics in identifying ETE and BRAF^{V600E} mutation have high c-index, sensitivity, and specificity, especially images from ultrasound, demonstrating the potential for diagnosing ETE and BRAF^{V600E} mutations in PTC.

Key Words: Papillary thyroid cancer; Radiomics; Extrathyroidal extension; BRAFV600E mutation.

Abbreviations: **CI** confidence intervals, **ETE** extrathyroidal extension, **LNM** lymph node metastasis, **ML** machine learning, **PTC** papillary thyroid cancer, **RQS** Radiomics Quality Score, **TNM** tumor node metastasis

Acad Radiol 2025; 32:1385–1397

From the Department of Radiology and Intervention, Hospital Pakar Kanak-Kanak (UKM Specialist Children's Hospital), Universiti Kebangsaan Malaysia, Jalan Yaacob Latif, Bandar Tun Razak, 56000, Kuala Lumpur, Malaysia (Y.L., F.Y.L., J.N.C., H.A.H., H.A.M.); Department of Ultrasound, Affiliated Hospital of Pan Zhihua University, Panzhihua, 61700, Sichuan Province, China (Y.L., L.X.); Tianfu Jincheng Laboratory, City of Future Medicine, Chengdu 641400, China (Y.L., Q.L.); Diagnostic Imaging & Radiotherapy Program, Faculty of Health Sciences, School of Diagnostic & Applied Health Sciences, University Kebangsaan Malaysia, Kuala Lumpur 50300, Malaysia (N.Y.); Department of Ultrasound, West China Hospital, Sichuan University, Chengdu 610041, China (Q.L.); Makmal Pemprosesan Imej Kefungsian (Functional Image Processing Laboratory), Department of Radiology, Universiti Kebangsaan Malaysia, Jalan Yaacob Latif, Bandar Tun Razak, Kuala Lumpur 56000, Malaysia (H.A.M.). Received September 2, 2024; revised November 4, 2024; accepted November 4, 2024. Address correspondence to: H.A.M. e-mail: hanani@ukm.edu.my

https://doi.org/10.1016/j.acra.2024.11.014

APPENDIX E

PROOF OF PUBLICATION 2

（Copy of the First Page of the Publication）

ORIGINAL ARTICLE

Ultrasound Combined With FNA-Tg Predicts the Lateral Cervical Lymph Node Metastasis in Papillary Thyroid Carcinoma

Yan Liu[1,2] | Ling Xiang[2] | Yi Wang[3] | Noorazrul Azmie Yahya[4] | Jing-kun Yin[2] | Wei Li[2] | Hamzaini Bin Abdul Hamid[1] | Jia-ning Chai[1] | Hanani Abdul Manan[1]

[1]Department of Radiology, University Kebangsaan Malaysia, Cheras, Kuala Lumpur, Malaysia | [2]Department of Ultrasound, Affiliated Hospital of Panzhihua University, Panzhihua, China | [3]Department of Thyroid Gland Breast Surgery, Affiliated Hospital of Panzhihua University, Panzhihua, China | [4]Diagnostic Imaging & Radiotherapy Program, School of Diagnostic & Applied Health Sciences, Faculty of Health Sciences, National University of Malaysia, Kuala Lumpur, Malaysia

Correspondence: Yan Liu (liuyan_panzhihua@163.com) | Hanani Abdul Manan (hanani@ukm.edu.my)

Received: 21 December 2024 | **Revised:** 16 May 2025 | **Accepted:** 23 May 2025

Funding: This work was supported by the Research Project of Panzhihua Medical Research Center (PYYZ-2022-36), the Health Commission of Sichuan Province Medical Science and Technology Program (24WSXT001) and the Faculty of Medicine Fundamental Grants (GFFP) (FF-2025-002).

Keywords: eluent | lymph node metastasis | papillary thyroid carcinoma | thyroglobulin | ultrasound

ABSTRACT

Objective: The present study analyzed typical ultrasound manifestations and fine-needle aspiration thyroglobulin (FNA-Tg) levels to investigate their association with cervical lymph node metastasis in patients with papillary thyroid carcinoma (PTC).

Methods: The data of 139 PTC patients with ultrasonically suspected cervical lymph node metastasis treated in our hospital from December 2022 to November 2023 were retrospectively analyzed. All included patients underwent ultrasound examination of cervical lymph nodes, fine needle aspiration cytology (FNA-C) examination, and ultrasound-guided lymph node aspiration eluent thyroglobulin (FNA-Tg). Typical ultrasound signs for diagnosing cervical lymph node metastasis (US-M) and ultrasound-guided FNA-Tg for diagnosing cervical lymph node metastasis were compared and analyzed.

Results: Results indicate that 71 patients were diagnosed with cervical lymph node metastasis through surgery and subsequently included in the metastatic group; the remaining 68 patients were included in the nonmetastatic group. The FNA-Tg value in the metastatic group was higher than that in the nonmetastatic group; the difference was significant ($p < 0.001$). The AUC values for diagnosing cervical lymph node metastasis in PTC patients using US-M, FNA-Tg, and US-M+FNA-Tg were 0.854, 0.927, and 0.952. When the cut-off value of FNA-Tg was 229.1 ng/mL, the sensitivity and specificity for diagnosing cervical lymph node metastasis in PTC patients were 84.5% and 89.5%.

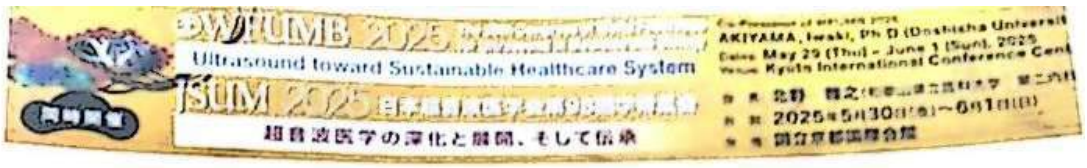
Conclusions: Ultrasound-guided FNA-Tg level is closely related to cervical lymph node metastasis in patients with PTC. The combination of ultrasound examination and FNA-Tg testing significantly enhances the accuracy of predicting lateral cervical lymph node metastasis in patients with PTC.

1 | Introduction

Thyroid cancer is the most common malignant tumor in the endocrine system of the head and neck, and among this type of cancer, papillary thyroid carcinoma (PTC) accounts for the largest proportion [1], accounting for over 80%. In patients with PTC, the metastasis rate of cervical lymph nodes is high (up to 20%–50%), and cervical lymph node metastasis is the main

APPENDIX F

PROOF OF ORAL PRESENTATION 1



Certificate of Presentation

Name: Yan LIU

This is to certify that you had an oral presentation in the 20th congress of World Federation for Ultrasonic in Medicine and Biology (WFUMB 2025).

Abstract submission number: 01928
Abstract title: ACCURACY OF RADIOMICS IN THE IDENTIFICATION OF EXTRATHYROIDAL EXTENSION AND BRAFV600EMUTATIONS IN PAPILLARY THYROID CARCINOMA
Presentation number: O-36-2
Oral Session: O-36 "Head and Neck (Thyroid) 4"
Date & Time: May 30 (Fri), 2025 4:30:00 PM - 5:15:00 PM
Place: Room 18 (Room K, 2F, Kyoto International Conference Center)

Sincerely,

President of WFUMB Congress 2025
CONDOUS, George

President of WFUMB Congress 2025
AKIYAMA, Iwaki

APPENDIX G

PROOF OF ORAL PRESENTATION 2