frontiers | Research Topics

Workflow optimisation for radiological imaging

Edited by
Jie-Zhi Cheng, Minjeong Kim and Xin Yang

Published in
Frontiers in Medicine

FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-8325-5588-0
DOI 10.3389/978-2-8325-5588-0

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Workflow optimisation for radiological imaging

Topic editors

Jie-Zhi Cheng — Shanghai United Imaging Intelligence, Co., Ltd., China

Minjeong Kim — University of North Carolina at Greensboro, United States

Xin Yang — Huazhong University of Science and Technology, China

Citation

Cheng, J.-Z., Kim, M., Yang, X., eds. (2024). *Workflow optimisation for radiological imaging*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-5588-0

Topic Editor Jie-Zhi Cheng is employed by Shanghai United Imaging Intelligence Co., Ltd. Other Topic Editors declare no conflicts of interest with the Research Topic.

Table of contents

frontiers | Frontiers in Medicine

TYPE Editorial
PUBLISHED 10 October 2024
DOI 10.3389/fmed.2024.1487230

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Giorgio Treglia,
Ente Ospedaliero Cantonale (EOC), Switzerland

*CORRESPONDENCE
Jie-Zhi Cheng
jzcheng@ntu.edu.tw

RECEIVED 27 August 2024
ACCEPTED 30 August 2024
PUBLISHED 10 October 2024

CITATION
Cheng J-Z, Kim M and Yang X (2024) Editorial: Workflow optimisation for radiological imaging. *Front. Med.* 11:1487230. doi: 10.3389/fmed.2024.1487230

COPYRIGHT
© 2024 Cheng, Kim and Yang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Editorial: Workflow optimisation for radiological imaging

Jie-Zhi Cheng[1]*, Minjeong Kim[2] and Xin Yang[3]

[1]United Imaging Intelligence Co., Ltd., Shanghai, China, [2]Department of Computer Science, University of North Carolina at Greensboro, Greensboro, NC, United States, [3]School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China

KEYWORDS

artificial intelligence, radiological imaging, computer-aided diagnosis, cancer staging, metastasis prediction, quantitative analysis, prognosis prediction

Editorial on the Research Topic
Workflow optimisation for radiological imaging

Deep learning techniques have effectively addressed numerous technical challenges in medical image analysis, leading to the flourishing advancement of artificial intelligence (AI) applications in medicine. The scope of these applications may include the improvement of the image diagnostic workup, imaging workflow optimisation, burden alleviation of repetitive medical tasks, and comprehensive analysis for treatment planning. In recent years, numerous commercial AI applications have been introduced into clinical practice, demonstrating clear benefits. At the same time, there has been a proliferation of promising clinical validation studies of AI in the literature. Many of these important studies have demonstrated the efficacy of AI, supported by large datasets from multiple institutions and countries. To advance research on AI applications in radiology, we organised this Research Topic to explore new AI and machine learning methods for improving medical workflows, procedures, and diagnostic processes related to radiological imaging. We received numerous high-quality manuscripts, and after a rigorous review process, 16 of them were ultimately published. The published articles in this Research Topic can be categorised into several research topics of computer-aided diagnosis, cancer staging and metastasis prediction, quantitative analysis, and prognosis prediction. Each research topic is further summarised below.

Computer-aided diagnosis

Li et al. utilised deep learning and radiomic techniques to diagnose Parkinson's disease based on T1-weighted magnetic resonance imaging. The VB-net demarcated the brain into 109 regions, from which a total of 2,264 radiomic features were extracted. The support vector machine was further applied to identify Parkinson's disease with high accuracy. Yang et al. focused on the early and rapid diagnosis of mild cognitive impairment, which is crucial for improving the prognosis of Alzheimer's disease. The structures of the bilateral hippocampus and parahippocampal gyrus were segmented from MR images for the Gaussian process models for the final identification of MCI. Wei et al. developed a novel prediction model that integrated multi-modal radiomic features derived from diffusion-weighted imaging and apparent diffusion coefficient images, along with multi-clinical features for the diagnosis of acute ischaemic stroke. Lin et al. proposed a deep learning-based diagnostic system to achieve early detection of esophageal cancer from non-contrast chest computed tomography (CT) images.

For differential diagnosis, Feng et al. extensively compared the performance of different machine learning approaches for the staging of testicular lesions in MR images. Lu et al. developed a deep learning radiomic model on multimodal images from ultrasonography, mammography, and MRI, which is further augmented with transfer learning, for malignancy identification of breast lesions. Cui et al. developed a CT-based radiomic model to distinguish between laryngeal squamous cell carcinoma and squamous cell hyperplasia. The applications of CAD are quite extensive, ranging from brain diseases to cancer problems in various parts of the body.

Cancer staging, metastasis prediction

Wu et al. manifested the early stages of cervical cancer with radiomic features from the T2-weighted images and apparent diffusion coefficient maps. Tian et al. developed a segmentation-classification paradigm for rectal cancer (RC) lesions for the assessment of different T-stages.

For metastasis prediction, Ma et al. developed a clinical-radiomic nomogram, which leveraged deep learning techniques, preoperative MR images, and clinical characteristics, to achieve accurate prediction of lymph node metastasis in RC. Fu et al. developed multitask prediction models for High-grade serous ovarian cancer with radiomic profiling on preoperative CT scans. Specifically, the radiomic features were derived from manual ROIs, and then classification models were constructed with the feature selection scheme to predict the assessments of R0 resection, lymph node invasion, and distant metastasis status. Cancer staging and metastasis prediction are very useful for treatment planning.

Quantitative analysis

Sun et al. employed ResUNet to delineate the testis from the MR scans for the measurement of testicular volume, which is essential for evaluating function and pathology. Jiang et al. utilised radiomic analysis on dual-energy CT images of the pancreas to establish a quantitative biomarker for type 2 diabetes mellitus. Deep learning algorithms were applied to segment the pancreas and were sequenced with the later classifiers of random forest, support vector machine and logistic regression. Meng et al. combined the clinical features, CT imaging signs, and radiomic features for the differentiation of Mycoplasma pneumonia in adults and children. The radiomic features are well-correlated with CT imaging signs, which provided a representation of different CT imaging signs as potential quantitative biomarkers. Jia et al. found that peri-plaque pericoronary adipose tissue (PCAT) was more valuable than proximal PCAT for the evaluation of coronary atherosclerosis. Spectral parameters of peri-plaque PCAT could further improve diagnostic accuracy. The quantitative parameters of peri-plaque PCAT attenuation were shown to be potential biomarkers of plaque vulnerability and the haemodynamic characteristics of coronary atherosclerosis. The quantitative measurement can replace tedious manual drawing, while the biomarkers will be very beneficial for disease diagnosis and treatment decisions.

Prognosis prediction

Xue et al. predicted the prognosis of early acute pancreatitis with a clinical radiomic model on CT scans. Prediction of disease prognosis is very helpful in the determination of treatment approaches.

AI applications in medicine are booming, and can potentially benefit various aspects of clinical practice, such as simplifying medical procedures, improving efficiency, increasing diagnostic accuracy, optimising workflow, etc. In this Research Topic, we have published articles on a wide range of applications like CAD, cancer staging and metastasis prediction, quantitative analysis, and prognosis prediction that used radiomics and CNN techniques to attain specific goals. In particular, the studies by Lu et al., Meng et al., and Xue et al. further applied the multimodal data for better performance. The integration of multimodal data may reflect the real practice of clinical diagnosis and may be the future research trend of medical AI. The recent progress of large model techniques may further resonate and expedite this research trend with a high capability of context and long-range reasoning.

In conclusion, We hope that the articles published in this Research Topic could serve as a reference for the promotion of more advanced AI studies in the medical field.

Author contributions

J-ZC: Writing – original draft, Writing – review & editing. MK: Writing – original draft, Writing – review & editing. XY: Writing – original draft, Writing – review & editing.

Conflict of interest

J-ZC was employed by United Imaging Intelligence. Co. Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 19 September 2023
DOI 10.3389/fmed.2023.1277535

Check for updates

OPEN ACCESS

EDITED BY
Jie-Zhi Cheng,
Shanghai United Imaging Intelligence, Co., Ltd., China

REVIEWED BY
Ning Mao,
Yantai Yuhuangding Hospital, China
He Deng,
Wuhan University of Technology, China

*CORRESPONDENCE
Liming Xia
✉ xialiming2017@outlook.com

†These authors have contributed equally to this work and share first authorship

RECEIVED 14 August 2023
ACCEPTED 06 September 2023
PUBLISHED 19 September 2023

CITATION
Sun K, Fan C, Feng Z, Min X, Wang Y, Sun Z, Li Y, Cai W, Yin X, Zhang P, Liu Q and Xia L (2023) Magnetic resonance imaging based deep-learning model: a rapid, high-performance, automated tool for testicular volume measurements.
Front. Med. 10:1277535.
doi: 10.3389/fmed.2023.1277535

COPYRIGHT
© 2023 Sun, Fan, Feng, Min, Wang, Sun, Li, Cai, Yin, Zhang, Liu and Xia. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Magnetic resonance imaging based deep-learning model: a rapid, high-performance, automated tool for testicular volume measurements

Kailun Sun[1†], Chanyuan Fan[2†], Zhaoyan Feng[2†], Xiangde Min[2], Yu Wang[3], Ziyan Sun[2], Yan Li[2], Wei Cai[2], Xi Yin[4], Peipei Zhang[2], Qiuyu Liu[2] and Liming Xia[2*]

[1]Department of Urology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [2]Department of Radiology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, Hubei, China, [3]Department of Research and Development, Infervision Medical Technology Co., Ltd., Beijing, China, [4]Department of CT & MRI, The First Affiliated Hospital, College of Medicine, Shihezi University, Shihezi, China

Background: Testicular volume (TV) is an essential parameter for monitoring testicular functions and pathologies. Nevertheless, current measurement tools, including orchidometers and ultrasonography, encounter challenges in obtaining accurate and personalized TV measurements.

Purpose: Based on magnetic resonance imaging (MRI), this study aimed to establish a deep learning model and evaluate its efficacy in segmenting the testes and measuring TV.

Materials and methods: The study cohort consisted of retrospectively collected patient data ($N = 200$) and a prospectively collected dataset comprising 10 healthy volunteers. The retrospective dataset was divided into training and independent validation sets, with an 8:2 random distribution. Each of the 10 healthy volunteers underwent 5 scans (forming the testing dataset) to evaluate the measurement reproducibility. A ResUNet algorithm was applied to segment the testes. Volume of each testis was calculated by multiplying the voxel volume by the number of voxels. Manually determined masks by experts were used as ground truth to assess the performance of the deep learning model.

Results: The deep learning model achieved a mean Dice score of 0.926 ± 0.034 (0.921 ± 0.026 for the left testis and 0.926 ± 0.034 for the right testis) in the validation cohort and a mean Dice score of 0.922 ± 0.02 (0.931 ± 0.019 for the left testis and 0.932 ± 0.022 for the right testis) in the testing cohort. There was strong correlation between the manual and automated TV (R^2 ranging from 0.974 to 0.987 in the validation cohort; R^2 ranging from 0.936 to 0.973 in the testing cohort). The volume differences between the manual and automated measurements were 0.838 ± 0.991 (0.209 ± 0.665 for LTV and 0.630 ± 0.728 for RTV) in the validation cohort and 0.815 ± 0.824 (0.303 ± 0.664 for LTV and 0.511 ± 0.444 for RTV) in the testing cohort. Additionally, the deep-learning model exhibited excellent reproducibility (intraclass correlation >0.9) in determining TV.

Conclusion: The MRI-based deep learning model is an accurate and reliable tool for measuring TV.

KEYWORDS

testicular volume, magnetic resonance imaging, deep-learning, T2-weighted imaging, ResUNet

Introduction

The testis is an important organ for male spermatogenesis and testosterone synthesis (1–3). As the seminiferous tubules account for approximately 80–90% of the testicular mass, the testicular volume (TV) reflects sperm and hormonal status (1, 2, 4–7). In clinical practice, the TV is an essential parameter for monitoring testicular functions and pathologies (2). An increased TV is the earliest sign of pubertal gonadotropin elevation; thus, TV measurements are used to monitor testicular development and pubertal status. Normal spermatogenesis occurs only when the total TV is normal or approximately normal, and the amount of TV loss is associated with the degree of spermatogenesis disorder (5). The TV has been proven to be related to semen profiles, and TV measurements are key components of male infertility evaluations (1, 6, 8). Therefore, accurate and individualized TV measurements may improve the diagnosis and treatment of patients with various disorders that affect testicular growth and fertility (2, 4, 7).

Several methods are used to assess TV, including calipers, different types of orchidometers, and ultrasonography (US) (4–6, 8–13). Clinical methods, such as calipers and orchidometers, are subjective in nature and tend to overestimate the true TV due to potential interference from the adjacent soft tissue, such as the epididymis, scrotal skin, and subcutaneous tissues, particularly in the case of small testes and hydrocele (5, 6, 11–14). Formula-derived US is generally used as the standard method for determining TV nowadays. TV is usually calculated as length (L) × width (W) × height (H) × constant (C), where C is a correction factor (often recommended as 0.71 or 0.52), and the length, width, and height are the sizes of the testicular axes determined by the sonographers (2, 15). However, the formula-derived measurement is recognized as a rough estimate of the TV, because the testis is an elastic and compressible organ with a shape that is neither uniform nor necessarily ellipsoid (5, 11). TV measurements obtained *via* US have been proven to vary from study to study, formula to formula, and examiner to examiner (2, 9, 10, 15); thus, establishing normative TV values and cutoffs for distinguishing pathological conditions has proven challenging, limiting the standardized use of TV in clinical practice (2, 5, 11, 15). Therefore, efforts are still needed to develop methods that are accurate, convenient, and individualized.

Recently, deep learning models have demonstrated great potential in attaining highly accurate volume measurements (16). These models first automatically segment the targets using deep learning algorithms, and then calculate the volumes of the targets by multiplying the voxel size by the voxel number (17). Measurement accuracy relies more on the precision of automatic segmentation results than on the match between the shape of the targets and the formula employed (18). Highly accurate auto-segmentation and volume estimation using deep learning models have been reported in several organs and pathologies, such as brain tumors, the liver, the kidney, the spleen, and the inner ear (16–23). However, the performance of deep learning models in estimating testicular volume has not been reported previously.

Therefore, in this study, a deep learning model, specifically, a ResUNet algorithm, is used to automatically segment the testes on T2-weighted imaging (T2WI) and calculate testicular volume. Masks manually defined by experts served as the reference standard for evaluating the performance of the deep learning model. A subset of subjects was scanned multiple times to evaluate the repeatability of the segmentation results. Our findings demonstrate that the T2WI-based deep learning model is an accurate and reliable tool for TV measurement.

Materials and methods

Our institutional review board approved this study, and the requirement for informed consent was waived.

Study population

The study population consisted of a retrospective dataset and a prospective dataset. For the collection of the retrospective data, we searched the electronic database of our institution from February 2014 to September 2021 for males who underwent magnetic resonance imaging (MRI) of the scrotum for any reason, such as scrotal pain and infertility. The inclusion criteria were defined as follows: (1) Both testes exhibited anatomically intact morphology, (2) no visible intratesticular lesions were present, and (3) patients underwent 3.0 T MRI scans of the scrotum, with available T2WI included in the MRI protocol. The exclusion criteria were defined as follows: (1) undescended testes, (2) testis was too small to observe in three image slices, (3) the quality of the MR images was poor, (4) patients underwent treatment, such as orchiectomy, partial orchiectomy, testis-sparing surgery, radiotherapy, or chemotherapy, due to testicular diseases, and (5) patients underwent androgen deprivation therapy due to prostate cancer. Finally, a total of 200 consecutive patients (400 testes) were enrolled in the retrospective dataset. This dataset was divided into training and independent validation cohorts according to a random distribution of 8:2. The training cohort was employed to train the network, while the validation cohort was used to evaluate the segmentation performance of the network.

A prospectively collected dataset comprising of ten healthy volunteers was used as the testing cohort. Each volunteer was scanned 5 times. The subjects were repositioned (removed from the scanner and asked to sit up and move on the bed) and reregistered on the scanner console between scans in each session; thus, all scans were treated as separate measurements. In addition, we attempted to vary the acquisition geometry between each scan while still acquiring full testes coverage. The testing data were used to assess the reproducibility of the MRI-based measurements.

MRI acquisition

All images were acquired using a 3 T MAGNETOM Skyra (Siemens Healthcare, Erlangen, Germany) and an anterior 18-element body matrix coil combined with a posterior 32-channel spine coil. Multiple sequences were scanned, but only the T2-weighted turbo spin–echo sequences were used in this study. The transverse T2WI were acquired using the following parameters: 3 mm slice thickness, 0 mm slice gap, 6,500 ms repetition time, 104 ms echo time, 180 × 180 in field of view and 384 × 320 acquisition matrix.

Notably, the T2WI parameters were consistent with the standardized technical requirements for scrotal imaging recommended by the Scrotal and Penile Imaging Working Group of the European Society of Urogenital Radiology (24, 25). The acquisition time of the transverse T2WI was approximately 180 s.

Manual segmentation

The manual segmentation results were used as the ground truth. Manual segmentation was performed using ITK SNAP software (version 3.4.0; www.itksnap.org). Three-dimensional binary masks of the entire testes were generated by tracing the testicular boundaries slice-by-slice on the transverse T2WI. The non-testicular parenchyma area, including the epididymis and mediastinum, was excluded from the manual segmentation. Manual segmentation was carried out by two radiologists (observer 1, with 10 years of experience in interpreting MRI of scrotum, and observer 2, with 5 years of experience in interpreting MRI of scrotum) in a blind manner. For the manual segmentation of the retrospective dataset, the images were collectively analyzed by the two observers, and discrepancies were resolved through discussion until a consensus was reached.

For the initial segmentation of the prospective dataset, which served as the ground truth, readers 1 and 2 collectively segmented the images [region of interest (ROI) A] for all 5 repeated acquisitions. Then, 1 month after the initial segmentation, readers 1 (ROI B) and 2 (ROI C) independently segmented all 5 repeated acquisitions to assess the inter- and intra-observer variability of the manual segmentation. The volume of each testis was computed by multiplying the voxel volume by the number of voxels in each testis mask. Subsequently, the total testicular volume (TTV) was calculated by summing the volumes of both testes.

Automated segmentation using ResUNet

All images were preprocessed, including resampling, normalization, cropping, and padding, to generate homogeneous MRI volumes. First, all volumes were resampled to the same voxel size of 0.46875 mm × 0.46875 mm × 1 mm. Subsequently, the intensities of each volume were normalized to the range [−1, 1]. The architecture of the model is based on the ResUNet algorithm (7, 26–28). Briefly, the model has encoding, bridge, and decoding parts. The encoding part encodes the input image into compact representations, while the decoding part recovers the representations for pixel-wise categorization. The bridge part connects the encoding and decoding paths. The ResUNet algorithm was implemented in Python 3.9.7 using PyTorch version 1.8.0. The network uses a Tversky loss function. The model was trained with a batch size of 1 over 200 epochs using the Adam optimizer. We set the initial learning rate to 0.0001 and trained the network for 600 iterations, reducing the learning rate to 80% of the current value every 20 iterations. The ResUNet model was trained using RTX 2080Ti GPUs (NVIDIA).

Statistical analysis

The baseline demographics are reported in the form of mean ± standard deviation (SD). The accuracy of the deep learning model was assessed by comparing the automated segmentation results with the manual segmentation results. The reliability of the manual segmentation results and the reproducibility of the deep learning model were evaluated using the testing dataset. Voxel-based similarity metrics (e.g., Dice score) and surface-based similarity metrics (e.g., Hausdorff distance) were employed to evaluate the overlap between masks. In addition, volume differences, including actual volume difference and percentage volume difference, were computed. The mean coefficient of variation (CoV; defined as SD/mean) and the intraclass correlation coefficient (ICC) were used to assess repeatability. Bland–Altman and regression analyses were conducted to evaluate the correlation between manual TV and automated TV.

Results

Patients

The final training dataset included MRI scans of 160 cases from 160 patients (aged 9–74 years; mean age 34.713 ± 14.542 years). In the training cohort, the average left testicular volume (LTV) was 12.539 ± 2.625 mL (1.471–34.628 mL), the average right testicular volume (RTV) was 13.579 ± 4.366 mL (1.824–36.601 mL), and the average total testicular volume (TTV) was 26.333 ± 8.357 mL (3.295–71.229 mL). The validation dataset included MRI scans of 40 cases from 40 patients (aged 11–70 years; mean age 33.4 ± 13.388 years). In the validation dataset, the average LTV was 12.351 ± 4.133 mL (2.356–21.373 mL), the average RTV was 12.672 ± 4.821 mL (1.539–23.126 mL), and the average TTV was 25.023 ± 8.676 mL (4.629–43.276 mL). The prospective testing dataset included MRI scans of 50 cases from 10 healthy volunteers (aged 13–30 years; mean age 19.7 ± 5.33 years). In the testing dataset, the average LTV was 12.539 ± 2.625 mL (8.162–16.072 mL), the average RTV was 13.549 ± 2.505 mL (8.187–16.945 mL), and the average TTV was 26.089 ± 5.052 mL (16.833–32.354 mL). The characteristics of the enrolled patients are provided in Table 1. The distributions of the TV in the training, validation and testing datasets are shown in Figure 1.

Reliability of the manual segmentation results

The healthy volunteers in the testing dataset were utilized to analyze the reproducibility of the manual segmentation results, as healthy testes have more consistent morphologies and therefore provide better performance for repeatability evaluation. First, based on masks manually determined by different experts, interobserver

TABLE 1 Characteristics of the enrolled patients.

Patients	Number of patients	Number of datasets	Mean Age (years)	LTV (mL)	RTV (mL)	TTV (mL)
Training Cohort	160	160	34.713 ± 14.542	12.753 ± 4.342	13.579 ± 4.366	26.333 ± 8.357
Validation Cohort	40	40	33.400 ± 13.388	12.351 ± 4.133	12.672 ± 4.821	25.023 ± 8.676
Testing Cohort	10	50	19.700 ± 5.330	12.539 ± 2.625	13.549 ± 2.505	26.089 ± 5.052

LTV, left testicular volume; RTV, right testicular volume; TTV, total testicular volume.
All values are quoted as mean ± SD.

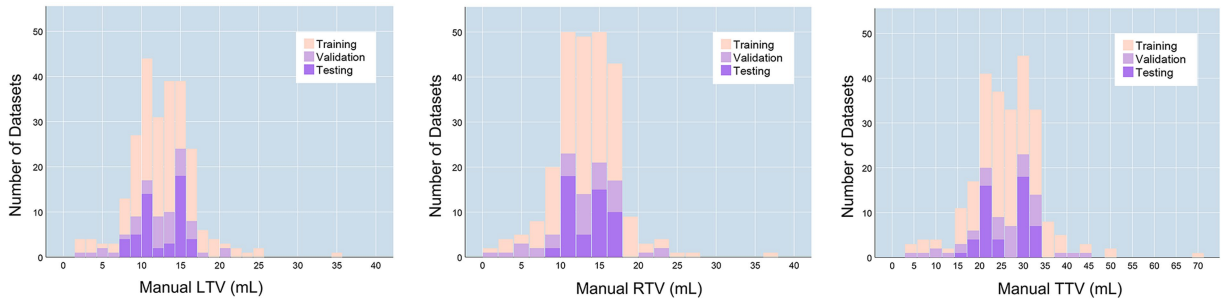


FIGURE 1
TV distributions. LTV, left testicular volume; RTV, right testicular volume; TTV, total testicular volume.

variability of the manual segmentation results was evaluated, as shown in Supplementary Table S1. The overlap between different manual masks was analyzed using similarity metrics, including the Dice score, Jaccard index, and Hausdorff distance. The actual volume difference was calculated. Next, based on the 5 repeated scans in the testing dataset, the intra-observer variability of the manual segmentation results was assessed. As shown in Table 2, the intra-observer repeatability of the manual TV was excellent (ICC > 0.9), regardless of the experiments of the observers or whether the manual segmentations were performed independently by one radiologist or collectively by two radiologists.

Accuracy of the deep learning model

As shown in Table 3, there was excellent similarity between the automatic and manual segmentations, with a mean Dice score of 0.922 ± 0.02 (0.921 ± 0.026 for the left testis and 0.926 ± 0.034 for the right testis) in the validation cohort and a mean Dice score of 0.931 ± 0.018 (0.931 ± 0.019 for the left testis and 0.932 ± 0.022 for the right testis) in the testing cohort. Linear regression analysis indicated a strong positive correlation (R2 ranging from 0.974 to 0.987, $p < 0.001$ for the validation cohort; R2 ranging from 0.936 to 0.973, $p < 0.001$ for the testing cohort) between the manual TV and automated TV (Figure 2, Supplementary Figure S1). For TTV, the bias (mean) and precision (SD) of the automated measurements were 0.838 and 0.991 in the validation cohort and 0.815 and 0.824 in the testing cohort. For LTV, the bias and precision of the automated measurements were 0.209 and 0.665 in the validation cohort and 0.303 and 0.664 in the testing cohort. For RTV, the bias and precision of the automated measurements were 0.630 and 0.728 in the validation cohort and 0.511 and 0.824 in the testing cohort. In terms of volume error, the actual volume differences between manual measurements and automated measurements were 0.209 ± 0.665 for LTV, 0.630 ± 0.728 for RTV, and 0.838 ± 0.991 for TTV in the validation cohort. In the testing cohort, the percentage volume differences between manual measurements and automated measurements were 0.303 ± 0.664 for LTV, 0.511 ± 0.444 for RTV, and 0.815 ± 0.824 for TTV. The percentage volume differences between manual measurements and automated measurements were 2.192 ± 6.129% for LTV, 4.654 ± 7.355% for RTV, and 3.711 ± 4.983% for TTV in the validation cohort. In the testing cohort, the percentage volume differences between manual measurements and automated measurements were 2.621 ± 5.580% for LTV, 3.909 ± 3.856% for RTV, and 3.266 ± 3.668% for TTV. Figure 3 illustrates an example of manual segmentation alongside the corresponding automated segmentation generated by the deep learning model.

Repeatability of the deep learning model

Based on the 5 repeated scans in the testing dataset, the repeatability of the MR-based automated measurements was evaluated (Table 4). Across the 5 different measurements, the automated method demonstrated excellent repeatability, with ICCs of 0.973 for LTV, 0.970 for RTV, and 0.982 for TTV. The mean CoV across the 5 different measurements were 2.964% ± 1.873% for LTV, 2.556% ± 1.690% for RTV, and 2.156% ± 1.352% for TTV, which were similar to the CoV of the manual methods ($p = 0.961$, $p = 0.118$, and $p = 0.343$, respectively).

Discussion

In this study, utilizing retrospectively collected patient data and prospectively collected data from healthy volunteers, we developed a

deep learning model to automatically segment the testes and measure TV. The deep learning model achieved accurate segmentation and provided reliable TV measurements. For the first time, we report that the MR-based deep learning model holds promise as a valuable tool for TV measurements.

As an essential parameter for monitoring testicular functions and pathologies, TV measurements have long been a subject of research focus (1, 2, 29). Over the past decades, efforts have been made to improve the accuracy of TV measurements, and formula-derived US measurements are generally used as the standard method for TV determination (1, 6, 10). However, the testis is an elastic and compressible organ whose elasticity varies across different developmental stages and pathological conditions. Moreover, the testis does not always conform to a strictly ellipsoidal shape. Consequently, precise and individualized measurements cannot be achieved through formula-derived approaches (5, 15). Recently, deep learning models have been reported to obtain highly accurate volume measurements of various organs and tissues, including the lungs, liver, kidney, spleen, and brain tumors (16, 18–21). For example, Daniel AJ et al. enrolled 30 healthy volunteers and 30 chronic disease patients, reporting that their deep learning model allowed for accurate segmentation and volume measurements of the kidney, yielding a mean Dice score of 0.93 ± 0.01 and a mean volume difference of 1.2 ± 16.2 mL (20). Modanwal G et al. demonstrated that a deep learning model enabled accurate segmentation of the liver and spleen in non-contrast computed tomography images, achieving a Dice coefficient of 0.95 in an independent validation cohort (16). In this study, utilizing retrospectively collected patient data ($N = 200$, comprising the training and independent validation cohorts) and prospective data from healthy volunteers ($N = 50$, serving as the testing cohort), we found that the ResUNet deep learning model enabled accurate TV measurements. This was reflected in mean Dice scores of 0.926 ± 0.034 and 0.922 ± 0.02, respectively in the validation and testing cohorts, along with volume differences of 0.838 ± 0.991 and 0.815 ± 0.824, respectively in the validation and testing cohorts. The possible reason might be as follows. On one hand, the testis exhibits relatively uniform characteristics in T2-weighted MR images, and the ResUNet model has previously demonstrated exceptional performance in automatically segmenting organs and tissues with repetitive structures (22–24, 26–28, 30). On the other hand, MRI, especially T2WI, provides excellent soft tissue contrast, facilitating the clear delineation of the tunica albuginea and tunica vaginalis that enclose the testes. Consequently, the testes can be accurately differentiated from the surrounding tissue in T2WI.

Another point of concern in automated volume measurement is its repeatability. Longitudinal follow-up of TV may be necessary in certain clinical settings, such as closely monitoring changes in pubertal status, tracking testicular involvement in pathological processes, and assessing the impact of chemotherapeutic or hormonal agents on the testes. TV measurements must exhibit high reproducibility to be valuable in longitudinal studies (4, 5, 31). In this study, we obtained 5 scans for each volunteer in the testing cohort to investigate the reproducibility of the MR-based measurement. Our results showed that MR-based deep learning model have small variations and excellent reproducibility; Thus is a reliable tool for TV measurements. In addition, our results also suggest that the MR-based manual measurements showed excellent inter- and intra-observer repeatability,

TABLE 2 Intra-observer repeatability of the manual measurements.

Observer	Testis	CoV (%)	ICC
Intra ROI A	Left	2.931 ± 1.291	0.971
	Right	3.829 ± 2.792	0.946
	Total	2.487 ± 1.193	0.981
Intra ROI B	Left	2.685 ± 0.965	0.977
	Right	3.991 ± 2.192	0.949
	Total	2.842 ± 0.813	0.978
Intra ROI C	Left	2.797 ± 0.842	0.976
	Right	4.051 ± 2.833	0.946
	Total	2.753 ± 1.420	0.979

CoV, coefficient of variation.
All CoV values are represented as the mean ± SD. ROI A was segmented collectively by readers 1 and 2. ROI B was independently segmented by reader 1, and ROI C was independently segmented by reader 2.

TABLE 3 The accuracy of the deep learning model.

Datasets	testis	Dice score	Jaccard index	Hausdorff distance (95th percentage)	Actual volume difference (mL)	Percentage volume difference (%)
Training	Left	0.918 ± 0.044	0.852 ± 0.064	1.412 ± 0.756	0.388 ± 0.761	2.639 ± 7.475
	Right	0.926 ± 0.034	0.864 ± 0.051	1.886 ± 7.471	0.578 ± 0.816	4.337 ± 8.170
	Total	0.923 ± 0.029	0.859 ± 0.046	1.926 ± 7.272	0.967 ± 1.231	3.627 ± 6.153
Validation	Left	0.921 ± 0.026	0.854 ± 0.043	1.364 ± 0.687	0.209 ± 0.665	2.192 ± 6.129
	Right	0.921 ± 0.027	0.856 ± 0.046	1.389 ± 0.747	0.630 ± 0.728	4.654 ± 7.355
	Total	0.922 ± 0.02	0.856 ± 0.033	1.386 ± 0.482	0.838 ± 0.991	3.711 ± 4.983
Testing	Left	0.931 ± 0.019	0.871 ± 0.033	1.182 ± 0.426	0.303 ± 0.664	2.621 ± 5.580
	Right	0.932 ± 0.022	0.873 ± 0.037	1.222 ± 0.515	0.511 ± 0.444	3.909 ± 3.856
	Total	0.931 ± 0.018	0.872 ± 0.030	1.183 ± 0.424	0.815 ± 0.824	3.266 ± 3.668

All values are represented as the mean ± SD.

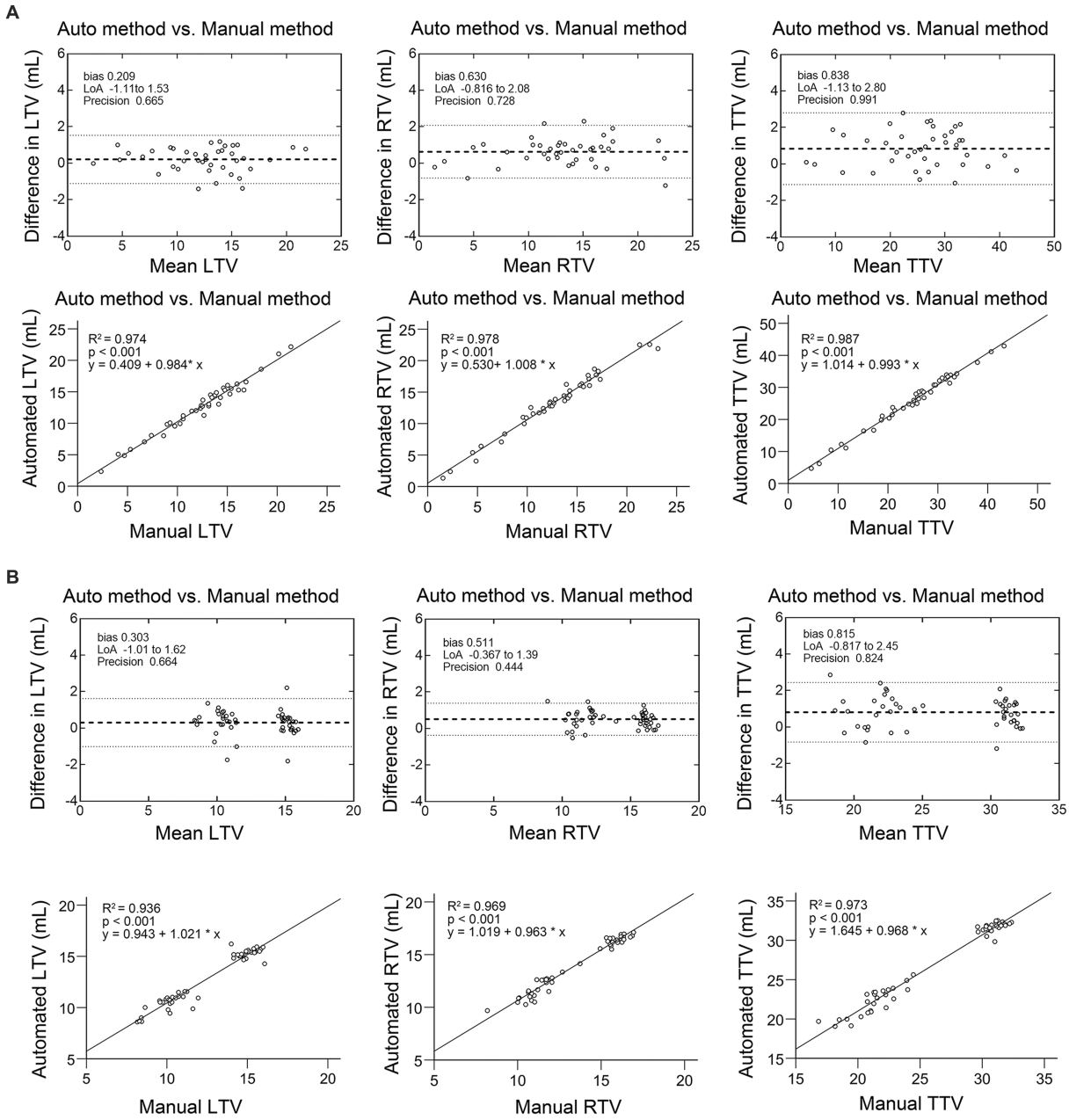


FIGURE 2
Scatter plot and Bland–Altman graph showing the difference between automated TV and manual TV. **(A)** Validation dataset. **(B)** Testing dataset. In the Bland–Altman graph, the solid lines show the actual mean difference (bias), and the dotted lines show 95% limits of agreements (LoAs).

regardless of the experiments of the observer or whether the manual segmentations were performed independently by one radiologist or collectively by two radiologists. These results demonstrate the reliability and rationality of the proposed MR-based measurement approach. One possible reason for these findings is that the testes could be well discriminated from the surrounding tissue in the T2WI.

Although MRI provides richer morphological and functional information and is less dependent on operator experience, US remains the first choice for diagnostic imaging of the scrotum (15, 24, 29). MRI is recommended as a valuable alternative diagnostic tool for investigating scrotal pathology (24, 25). The main reason is that US is faster, more easily accessible, and more convenient, whereas multiplane and multimodal imaging are needed for scrotal MRI (24). However, in this study, the deep learning model was trained on only transverse T2WI, which takes only about 180 s to obtain the images. Therefore, the MRI-based deep-learning model proposed in this study is low time consuming, reliable and individualized.

This study has several limitations. First, there is a lack of data on US-derived measurements to conduct a comparison between US-derived measurements and MRI-based measurements.

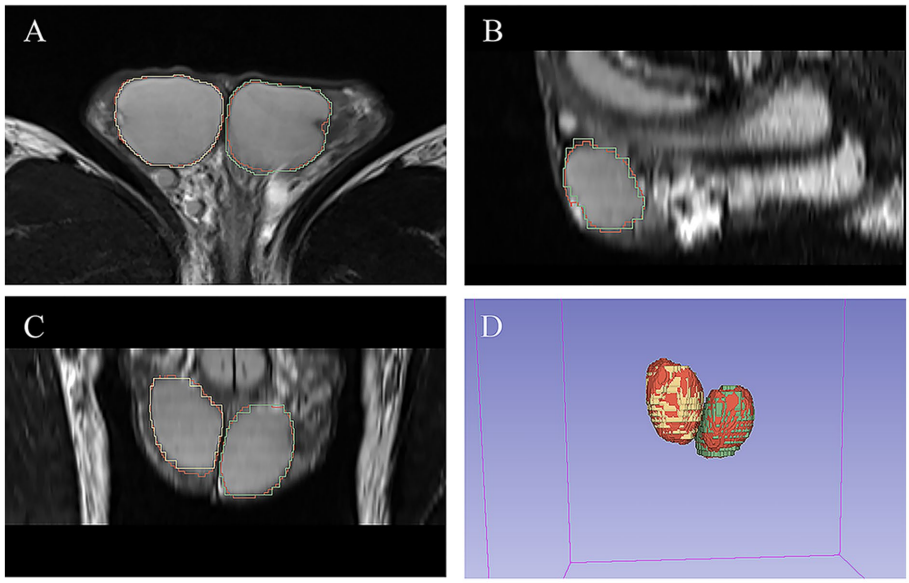


FIGURE 3
Example of manual and automated segmentation of the testes. **(A)** Axial view. **(B)** sagittal view. **(C)** Coronal view. **(D)** 3D volume. The manual mask generated by experts for left testis is shown in blue, the manual mask for right testis is shown in green, and the automatically generated mask is shown in red.

TABLE 4 **Comparison of the repeatability between the manual and automated measurements.**

Testis	ICC		CoV (%)		
	Manual	Auto	Manual	Auto	*p**
Left	0.971	0.973	2.931 ± 1.291	2.964 ± 1.873	0.961
Right	0.946	0.967	3.829 ± 2.792	2.779 ± 1.853	0.118
Total	0.981	0.984	2.487 ± 1.193	2.047 ± 1.319	0.343

CoV, coefficient of variation.
All CoV values are represented as the mean ± SD.
*Paired student *t* test.

Second, the retrospective data served as training and validation cohorts containing heterogeneous patient populations, including infertility, hydrocele, scrotal pain, etc. A deep learning model trained with heterogeneous patient data can be clinically significant since the TV is typically used to assess patients with a variety of disorders that may affect testicular growth and fertility, such as infertility and varicocele. Third, this work was a single-center study. Multicenter studies are needed to validate our findings. Notably, the scan parameters used in this study were consistent with the standardized scrotal MRI technical requirements recommended by the Scrotal and Penile Imaging Working Group of the European Society of Urogenital Radiology, suggesting the universality of the proposed deep learning model.

Conclusion

In conclusion, the proposed MRI-based deep learning model is an accurate and reliable tool for the segmentation and volume measurement of the testes.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by the Ethical Committee of Tongji Hospital of Tongji Medical College of Huazhong University of Science and Technology. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

KS: Conceptualization, Methodology, Project administration, Writing – original draft, Writing – review & editing. CF: Conceptualization, Methodology, Project administration, Writing – original draft, Writing – review & editing. ZF: Conceptualization, Methodology, Project administration, Writing – review & editing. XM: Data curation, Methodology, Software, Writing – review & editing. YW: Formal analysis, Methodology, Software, Writing – review & editing. ZS: Data curation, Formal analysis, Writing – review & editing. YL: Data curation, Formal analysis, Writing – original draft. WC: Data curation, Formal analysis, Writing – review & editing. XY: Data curation, Formal analysis, Writing – review & editing. PZ: Data curation, Formal analysis, Writing – review & editing. QL: Data curation, Formal analysis, Writing – review &

editing. LX: Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing.

Funding

This work was supported by the National Natural Science Foundation of China [grant numbers 81873889]. The author(s) declare financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

YW is employed by Department of Research and Development, Infervision Medical Technology Co., Ltd., Beijing, China.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2023.1277535/full#supplementary-material

SUPPLEMENTARY FIGURE S1
Scatter plot and Bland–Altman graph showing the difference between automated TV and manual TV in the training dataset. In the Bland–Altman graph, solid lines represent the actual mean difference (bias), while dotted lines represent the 95% limits of agreement (LoAs).

References

1. Bahk JY, Jung JH, Jin LM, Min SK. Cut-off value of testes volume in young adults and correlation among testes volume, body mass index, hormonal level, and seminal profiles. *Urology*. (2010) 75:1318–23. doi: 10.1016/j.urology.2009.12.007

2. Lotti F, Frizza F, Balercia G, Barbonetti A, Behre HM, Calogero AE, et al. The European academy of andrology (EAA) ultrasound study on healthy, fertile men: scrotal ultrasound reference ranges and associations with clinical, seminal, and biochemical characteristics. *Andrology*. (2021) 9:559–76. doi: 10.1111/andr.12951

3. Tang Fui MN, Hoermann R, Wittert G, Grossmann M. Testicular volume and clinical correlates of hypothalamic-pituitary-testicular function: a cross-sectional study in obese men. *Asian J Androl*. (2020) 22:354–9. doi: 10.4103/aja.aja_96_19

4. Cai D, Wu S, Li Y, Chen Q. Validity of measurements of testicular volume obtained by a built-in software of ultrasound systems: with formula recommended by updated guidelines as reference. *J ultrasonography*. (2020) 20:e181–4. doi: 10.15557/JoU.2020.0030

5. Paltiel HJ, Diamond DA, Di Canzio J, Zurakowski D, Borer JG, Atala A. Testicular volume: comparison of orchidometer and US measurements in dogs. *Radiology*. (2002) 222:114–9. doi: 10.1148/radiol.2221001385

6. Lenz S, Giwercman A, Elsborg A, Cohr KH, Jelnes JE, Carlsen E, et al. Ultrasonic testicular texture and size in 444 men from the general population: correlation to semen quality. *Eur Urol*. (1993) 24:231–8. doi: 10.1159/000474300

7. Nguyen Hoai B, Hoang L, Tran D, Nguyen Cao T, Doan Tien L, Sansone A, et al. Ultrasonic testicular size of 24,440 adult Vietnamese men and the correlation with age and hormonal profiles. *Andrologia*. (2022) 54:e14333. doi: 10.1111/and.14333

8. Diamond DA, Paltiel HJ, Dicanzio J, Zurakowski D, Bauer SB, Atala A, et al. Comparative assessment of pediatric testicular volume: orchidometer versus ultrasound. *J Urol*. (2000) 164:1111–4. doi: 10.1016/S0022-5347(05)67264-3

9. Hsieh ML, Huang ST, Huang HC, Chen Y, Hsu YC. The reliability of ultrasonographic measurements for testicular volume assessment: comparison of three common formulas with true testicular volume. *Asian J Androl*. (2009) 11:261–5. doi: 10.1038/aja.2008.48

10. Pilatz A, Rusz A, Wagenlehner F, Weidner W, Altinkilic B. Reference values for testicular volume, epididymal head size and peak systolic velocity of the testicular artery in adult males measured by ultrasonography. *Ultraschall Med*. (2013) 34:349–54.

11. Taskinen S, Taavitsainen M, Wikström S. Measurement of testicular volume: comparison of 3 different methods. *J Urol*. (1996) 155:930–3. doi: 10.1016/S0022-5347(01)66349-3

12. Sakamoto H, Saito K, Ogawa Y, Yoshida H. Testicular volume measurements using Prader orchidometer versus ultrasonography in patients with infertility. *Urology*. (2007) 69:158–62. doi: 10.1016/j.urology.2006.09.013

13. Sakamoto H, Saito K, Oohta M, Inoue K, Ogawa Y, Yoshida H. Testicular volume measurement: comparison of ultrasonography, orchidometry, and water displacement. *Urology*. (2007) 69:152–7. doi: 10.1016/j.urology.2006.09.012

14. Mbaeri TU, Orakwe JC, Nwofor AM, Oranusi KC, Mbonu OO. Accuracy of Prader orchidometer in measuring testicular volume. *Niger J Clin Pract*. (2013) 16:348–51. doi: 10.4103/1119-3077.113460

15. members of the ESUR-SPIWG WGFreeman S, Bertolotto M, Richenberg J, Belfield J, Dogra V, et al. Ultrasound evaluation of varicoceles: guidelines and recommendations of the European Society of Urogenital Radiology Scrotal and Penile Imaging Working Group (ESUR-SPIWG) for detection, classification, and grading. *Eur Radiol*. (2020) 30:11–25. doi: 10.1007/s00330-019-06280-y,

16. Modanwal G, al-Kindi S, Walker J, Dhamdhere R, Yuan L, Ji M, et al. Deep-learning-based hepatic fat assessment (DeHFt) on non-contrast chest CT and its association with disease severity in COVID-19 infections: a multi-site retrospective study. *EBioMedicine*. (2022) 85:104315. doi: 10.1016/j.ebiom.2022.104315

17. Perez AA, Noe-Kim V, Lubner MG, Graffy PM, Garrett JW, Elton DC, et al. Deep learning CT-based quantitative visualization tool for liver volume estimation: defining Normal and hepatomegaly. *Radiology*. (2022) 302:336–42. doi: 10.1148/radiol.2021210531

18. Cancian P, Cortese N, Donadon M, di Maio M, Soldani C, Marchesi F, et al. Development of a deep-learning pipeline to recognize and characterize macrophages in Colo-rectal liver metastasis. *Cancers*. (2021) 13:13133313. doi: 10.3390/cancers13133313

19. Li S, Liu J, Song Z. Brain tumor segmentation based on region of interest-aided localization and segmentation U-net. *Int J Mach Learn Cybern*. (2022) 13:2435–45. doi: 10.1007/s13042-022-01536-4

20. Daniel AJ, Buchanan CE, Allcock T, Scerri D, Cox EF, Prestwich BL, et al. Automated renal segmentation in healthy and chronic kidney disease subjects using a convolutional neural network. *Magn Reson Med*. (2021) 86:1125–36. doi: 10.1002/mrm.28768

21. Chao H, Shan H, Homayounieh F, Singh R, Khera RD, Guo H, et al. Deep learning predicts cardiovascular disease risks from lung cancer screening low dose computed tomography. *Nat Commun*. (2021) 12:2963. doi: 10.1038/s41467-021-23235-4

22. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O, (2016). 3D U-net: learning dense volumetric segmentation from sparse annotation. Medical image computing and computer-assisted intervention – MICCAI 2016: 19th international conference, Athens, Greece, October 17–21, 2016, proceedings, part II; Cham: Springer International Publishing.

23. Li D, Chu X, Cui Y, Zhao J, Zhang K, Yang X. Improved U-net based on contour prediction for efficient segmentation of rectal cancer. *Comput Methods Prog Biomed*. (2022) 213:106493. doi: 10.1016/j.cmpb.2021.106493

24. Tsili AC, Bertolotto M, Turgut AT, Dogra V, Freeman S, Rocher L, et al. MRI of the scrotum: recommendations of the ESUR scrotal and penile imaging working group. *Eur Radiol*. (2018) 28:31–43. doi: 10.1007/s00330-017-4944-3

25. Tsili AC, Argyropoulou MI, Dolciami M, Ercolani G, Catalano C, Manganaro L. When to ask for an MRI of the scrotum. *Andrology*. (2021) 9:1395–409. doi: 10.1111/andr.13032

26. Li J, Liu K, Hu Y, Zhang H, Heidari AA, Chen H, et al. Eres-UNet++: liver CT image segmentation based on high-efficiency channel attention and res-UNet+. *Comput Biol Med*. (2022) 158:106501. doi: 10.1016/j.compbiomed.2022.106501

27. Xu Z, Yu F, Zhang B, Zhang Q. Intelligent diagnosis of left ventricular hypertrophy using transthoracic echocardiography videos. *Comput Methods Prog Biomed*. (2022) 226:107182. doi: 10.1016/j.cmpb.2022.107182

28. Zhang Z, Liu Q, Wang Y. Road extraction by deep residual U-net. *IEEE Geosci Remote Sens Lett*. (2018) 15:749–53. doi: 10.1109/LGRS.2018.2802944

29. Members of the ESUR-SPIWG WGBertolotto M, Freeman S, Richenberg J, Belfield J, Dogra V, et al. Ultrasound evaluation of varicoceles: systematic literature review and rationale of the ESUR-SPIWG guidelines and recommendations. *J Ultrasound*. (2020) 23:487–507. doi: 10.1007/s40477-020-00509-z

30. Topff L, Groot Lipman KBW, Guffens F, Wittenberg R, Bartels-Rutten A, van Veenendaal G, et al. Is the generalizability of a developed artificial intelligence algorithm for COVID-19 on chest CT sufficient for clinical use? Results from the international consortium for COVID-19 imaging AI (ICOVAI). *Eur Radiol*. (2023) 33:4249–58. doi: 10.1007/s00330-022-09303-3

31. Oehme NHB, Roelants M, Bruserud IS, Eide GE, Bjerknes R, Rosendahl K, et al. Ultrasound-based measurements of testicular volume in 6- to 16-year-old boys - intra- and interobserver agreement and comparison with Prader orchidometry. *Pediatr Radiol*. (2018) 48:1771–8. doi: 10.1007/s00247-018-4195-8

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 29 November 2023
DOI 10.3389/fmed.2023.1326324

Check for updates

OPEN ACCESS

EDITED BY
Jie-Zhi Cheng,
Shanghai United Imaging Intelligence, Co., Ltd., China

REVIEWED BY
Yi Xiao,
Shanghai Changzheng Hospital, China
Zhang Shi,
Fudan University, China
Yiqun Sun,
Fudan University Shanghai Cancer Center, China

*CORRESPONDENCE
Yuan Yuan
✉ yuanyuan19870108@163.com
Fu Shen
✉ ssff_53@163.com

†These authors have contributed equally to this work

RECEIVED 23 October 2023
ACCEPTED 14 November 2023
PUBLISHED 29 November 2023

CITATION
Tian C, Ma X, Lu H, Wang Q, Shao C, Yuan Y and Shen F (2023) Deep learning models for preoperative T-stage assessment in rectal cancer using MRI: exploring the impact of rectal filling.
Front. Med. 10:1326324.
doi: 10.3389/fmed.2023.1326324

COPYRIGHT
© 2023 Tian, Ma, Lu, Wang, Shao, Yuan and Shen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Deep learning models for preoperative T-stage assessment in rectal cancer using MRI: exploring the impact of rectal filling

Chang Tian[1], Xiaolu Ma[2], Haidi Lu[2], Qian Wang[3], Chengwei Shao[2], Yuan Yuan[2*†] and Fu Shen[2*†]

[1]School of Information Science and Technology and School of Biomedical Engineering, ShanghaiTech University, Shanghai, China, [2]Department of Radiology, Changhai Hospital, The Navy Medical University, Shanghai, China, [3]School of Biomedical Engineering, ShanghaiTech University, Shanghai, China

Background: The objective of this study was twofold: firstly, to develop a convolutional neural network (CNN) for automatic segmentation of rectal cancer (RC) lesions, and secondly, to construct classification models to differentiate between different T-stages of RC. Additionally, it was attempted to investigate the potential benefits of rectal filling in improving the performance of deep learning (DL) models.

Methods: A retrospective study was conducted, including 317 consecutive patients with RC who underwent MRI scans. The datasets were randomly divided into a training set (n=265) and a test set (n=52). Initially, an automatic segmentation model based on T2-weighted imaging (T2WI) was constructed using nn-UNet. The performance of the model was evaluated using the dice similarity coefficient (DSC), the 95th percentile Hausdorff distance (HD95), and the average surface distance (ASD). Subsequently, three types of DL-models were constructed: Model 1 trained on the total training dataset, Model 2 trained on the rectal-filling dataset, and Model 3 trained on the non-filling dataset. The diagnostic values were evaluated and compared using receiver operating characteristic (ROC) curve analysis, confusion matrix, net reclassification index (NRI), and decision curve analysis (DCA).

Results: The automatic segmentation showed excellent performance. The rectal-filling dataset exhibited superior results in terms of DSC and ASD ($p = 0.006$ and 0.017). The DL-models demonstrated significantly superior classification performance to the subjective evaluation in predicting T-stages for all test datasets (all $p < 0.05$). Among the models, Model 1 showcased the highest overall performance, with an area under the curve (AUC) of 0.958 and an accuracy of 0.962 in the filling test dataset.

Conclusion: This study highlighted the utility of DL-based automatic segmentation and classification models for preoperative T-stage assessment of RC on T2WI, particularly in the rectal-filling dataset. Compared with subjective evaluation, the models exhibited superior performance, suggesting their noticeable potential for enhancing clinical diagnosis and treatment practices.

KEYWORDS

rectal cancer, T staging, MRI, deep learning, rectal filling

Background

Colorectal cancer (CRC) stands as the second most prevalent contributor to cancer-related mortality in the United States. Projections for the year 2023 indicate that approximately 153,020 individuals will receive a diagnosis of CRC, and regrettably, 52,550 individuals will succumb to the disease. This includes a concerning subset of 19,550 cases and 3,750 deaths among individuals below the age of 50 years old (1). Rectal cancer (RC) is a subset of CRC, a disease that poses a grave risk to people's lives. Rectal magnetic resonance imaging (MRI) has witnessed widespread utilization in the comprehensive assessment of RC, assuming a vital role in treatment planning for patients by facilitating accurate preoperative tumor staging. Within clinical practice, high-resolution T2-weighted imaging (HR-T2WI) has gained unanimous acceptance as the optimal approach for preoperative staging of RC (2). The precise preoperative differentiation between T1-2 and T3-4 stages in RC holds immense significance for clinicians in guiding individualized treatment strategies. The ability to discern which patients should undergo total mesorectal excision (TME) or receive neoadjuvant treatment while minimizing the risks of both over-treatment and under-treatment has become paramount (3, 4). However, the traditional approach to MRI staging relies heavily on the expertise and subjective evaluation of radiologists, leading to diminished repeatability and accuracy rates. This reliance poses significant challenges in achieving an accurate preoperative T-stage diagnosis for RC (5). Furthermore, a contentious issue surrounds the use of rectal distension during rectal MRI, specifically regarding whether the rectal lumen should be filled with fluid or gel (2–4). While the primary objective of rectal filling is to optimize lesion visualization and improve T-stage assessment on MRI, the question of its routine application remains unresolved due to a lack of robust evidence demonstrating substantial improvements in lesion conspicuity (3–5).

In recent years, radiomics has emerged as a potential method for addressing diverse clinical challenges, surpassing traditional methods in several studies. By leveraging high-throughput analysis to extract a multitude of quantitative features from medical images, radiomics approaches have demonstrated promising potential in the field of digestive tumors (6–17). However, the predominant methodologies in this domain typically involve manually determining the volume of the entire primary tumor. This process is not only arduous and time-consuming but also heavily reliant on the operator's expertise, demanding a high level of proficiency (9, 16, 17).

Previous study yielded an initial finding indicating the development of two distinct radiomics models utilizing rectal HR-T2WI, both with and without rectal filling. These models were devised to evaluate the T staging of RC. Notably, our results demonstrated the superior performance of the radiomics model incorporating rectal filling in effectively distinguishing between T1-2 and T3 stages. This promising outcome suggests that the utilization of this model could offer valuable support in clinical decision-making when evaluating T-stage in RC patients (6).

Meanwhile, the deep learning (DL)-based method, as a novel technology, could significantly improve lesion automatic localization and segmentation, tumor diagnosis, staging, and prognosis prediction to facilitate treatment strategy, and could even greatly help radiologists work more efficiently and reduce their burden (18–20). Despite the considerable significance of T staging in RC, there exists a notable research gap regarding the validation and comparative analysis of MRI-based DL approaches specifically tailored for T staging evaluation, taking into account the presence or absence of rectal filling.

Therefore, in this study, we initially attempted to construct a convolutional neural network for the automatic localization and segmentation of RC lesions. Subsequently, we developed DL networks for the assessment of RC T-staging. Of utmost importance was our exploration of whether rectal filling could prove beneficial in guiding clinical decision-making for RC T stage evaluation.

Methods

Participants

This study followed the Declaration of Helsinki and had approval from the Ethics Committees of Changhai Hospital. Written informed consent was waived from all patients.

This retrospective trial enrolled a total of 492 consecutive patients with RC who underwent radical resection at Changhai Hospital between January 2017 and May 2023. The study's inclusion criteria comprised the following: (1) confirmation of rectal adenocarcinoma through postoperative pathological examination; (2) presence of a single lesion; (3) baseline rectal magnetic resonance (MR) examination conducted within 14 days prior to surgical resection. Exclusion criteria included: (1) receipt of any local or systemic treatment prior to surgical resection, such as neoadjuvant chemoradiotherapy ($n = 108$); (2) concurrent diagnosis of other malignancies ($n = 7$); (3) poor image quality ($n = 25$); (4) synchronous distant metastasis ($n = 23$); (5) palliative resection ($n = 7$); (6) history of previous pelvic surgery ($n = 5$). Consequently, a total of 317 cases were included in the final analysis, as depicted in Figure 1.

Clinicopathologic data

Patients' demographic and clinicopathological data were retrospectively extracted from the clinicopathological databases. These data encompassed various factors, including sex, age, body mass index (BMI), histological differentiation, pathological T-stage, pathological N-stage, carcinoembryonic antigen (CEA) levels (with <5 ng/mL considered as negative), and carbohydrate antigen 19-9 (CA19-9) levels (with <37 U/mL considered as negative). These parameters were recorded concurrently with the baseline MRI examinations. Employing the criteria set forth by the National Comprehensive Cancer Network (NCCN) and American Joint Committee on Cancer (AJCC) staging system (21), the patients involved in the study were meticulously stratified into distinct cohorts, each characterized by their respective pathological T stages.

Abbreviations: RC, Rectal cancer; TME, Total mesorectal excision; DL, Deep learning; T2WI, T2-weighted imaging; VOI, Volume of interest; ROC, Receiver operating characteristic; AUC, Area under the ROC curve; MRF, Mesorectal fascia; EMVI, Extramural vascular invasion; LNM, Lymph node metastasis; DSC, Dice similarity coefficient; HD95, 95th percentile Hausdorff distance; ASD, Average surface distance.

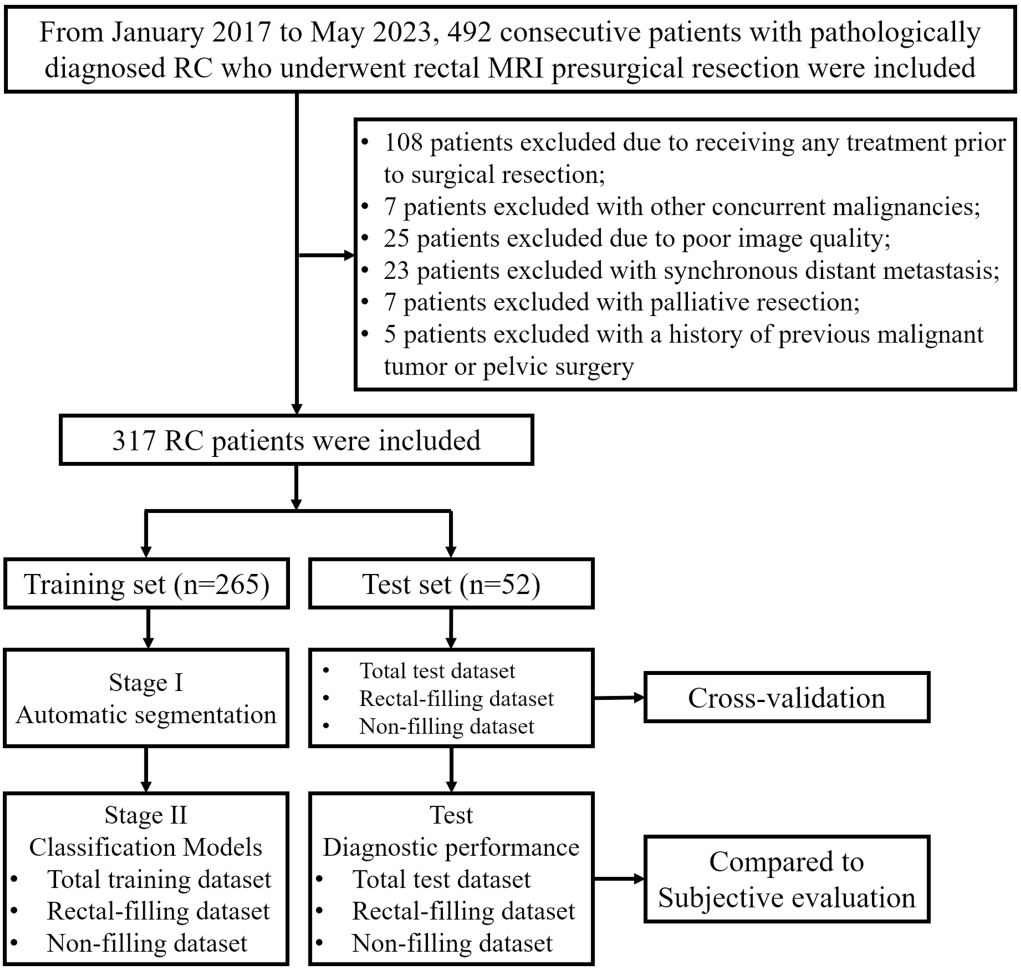


FIGURE 1
Flowchart of the study.

Specifically, the T1-2 group encompassed individuals with tumors confined solely to the submucosal and muscularis propria layers. In contrast, the T3-4 group comprised patients with tumors that exhibited invasive growth beyond the confines of the muscularis propria.

Image acquisition and analysis

Prior to the study, baseline rectal MRI scans were performed using 3.0 T MR systems, including the Siemens 3.0 T MAGNETOM Skyra MRI System, GE 3.0 T Discovery MR 750w, and Signa HDX System, coupled with a specialized phased array coil for enhanced imaging sensitivity. To ensure optimal image quality, intestinal cleaning was meticulously carried out through the administration of a 20 mL glycerin enema. Considering the possibility of contraindications, the administration of raceanisodamine hydrochloride, a commonly used agent, was deliberately omitted. As part of the routine imaging procedure, oblique axial HR-T2WI was conducted with careful consideration of the orientation perpendicular to the long axis of the rectum, encompassing the region of interest (ROI). Notably, detailed information regarding the parameters employed for HR-T2WI, which played a pivotal role in the subsequent analysis, can be found in Supplementary Table S1.

Within the filling group, patients underwent a baseline MRI with rectal filling, involving the administration of warm ultrasound (US) transmission gel to achieve rectal distention. Prior to acquiring the oblique axial HR-T2W images, the volume of gel used for rectal filling was tailored based on the endoscopic evaluation of tumor location. Specifically, 60–80 mL of gel was administered for lesions situated in the lower and middle rectum, while 80–100 mL was utilized for lesions in the upper rectum (22). Conversely, in the non-filling group, rectal distention using US gel was omitted during the baseline MRI procedure.

Subjective evaluation and ROI delineation were performed by 3 radiologists with systematic training, including FS, HL, and YY with 15, 11, and 13 years of experience in MR diagnosis, respectively, who were blinded to pathological data. A subjective classification task was assigned to the experts, requiring them to categorize each lesion as either T1-2 or T3-4 based on the established TNM staging system. Interobserver agreement for MR T-staging among the three radiologists was calculated. To facilitate accurate lesion segmentation, the ROI encompassing the entire rectal lesion was manually delineated in a meticulous slice-by-slice manner on the T2WI using

ITK-SNAP 4.0.0 software[1]. The delineated borders, representing the ground truth (GT), were meticulously determined by consensus among the experts. In cases of any discrepancies or differences of opinion, a thorough discussion ensued until a consensus was reached, requiring the agreement of at least two experts.

Dataset and pre-processing

A dataset comprising 317 MRI scans of RC and their corresponding T-stage labels was extracted and subsequently divided into a training set ($n = 265$) and a test set ($n = 52$) using a random allocation in a ratio of approximately 5:1. For the segmentation task, we utilized the preprocessing pipeline of nn-UNet (23–25), which could select the suitable data fingerprint automatically. We adopted the data preprocessing strategy through data fingerprint information, including resampling strategy, cropping area size, gray value distribution, etc. information, thus forming a so-called "configuration plan." While for the classification task, to ensure consistency, all images underwent preprocessing too from the "configuration plan," including resampling to a target spacing of [0.36, 0.36, 0.36] mm. Additionally, the size of each imaging scan was adjusted by cropping or padding to achieve a uniform dimension of $384 \times 384 \times 64$.

DL model construction

The U-Net architecture (23), introduced in 2015 as an Encoder–Decoder model (24), made a significant impact in the field of medical segmentation, generating widespread enthusiasm. Subsequent studies have primarily concentrated on maximizing the potential of U-Net and enhancing its performance through various modifications. Presently, UNet-like Encoder–Decoder architectures remain robust and highly regarded in the field. One notable variant, nn-UNet (25), exemplifies the remarkable qualities of U-Net as a self-configuring approach and pipeline for DL-based biomedical image segmentation, consistently delivering exceptional results.

Taking inspiration from these advancements, we incorporated a powerful and highly acclaimed network architecture, slightly modified by nn-UNet to be tailored for our rectal cancer data, as the backbone in Stage I of our study, where we trained it on a total of 265 cases. To adapt this network to our specific dataset and enable automatic segmentation of RC (Figure 2A), we rebuilt the training pipeline. We employed a larger dropout rate and more data augmentation strategy to prevent overfitting. In order to enhance the performance and generalizability of our model, we randomly divided the dataset with 5-fold cross-validation, implemented group normalization instead of batch normalization, and introduced larger convolution kernels. To evaluate the accuracy of the segmentation results, we calculated the dice similarity coefficient (DSC), the 95th percentile Hausdorff distance (HD95), and the average surface distance (ASD) between the automatically segmented images and GT images (26–28).

In contrast to the segmentation sub-task, the classification sub-task focuses on feature extraction after the convolution stage without the need to restore the features to their original size, which is one of the differences between the classification and segmentation tasks. Thus, in Stage II of our study, we designed an appropriate encoder as the backbone, which is the encoder of the 3D UNet, to extract features after convolution. For the output layer, we incorporated a multi-layer perception network to classify the T stages of RC. The flowchart of the classification task is depicted in Figure 2B. To facilitate T-stage classification in Stage II, we introduced a simple and easily manageable padding-cropping strategy. This involved utilizing the segmentation results obtained from Stage I and treating them as the input for the classification task, following the process outlined in Figure 2B. To construct our DL models, we divided the dataset into three categories based on the rectal filling status: model 1 trained on the complete training set of 265 cases, model 2 trained exclusively on rectal-filling cases, and model 3 trained using the non-filling dataset. The construction details of these models are provided in Supplementary material.

Statistical analysis

To perform the statistical analysis, we employed two software tools: MedCalc (version 19.8, MedCalc Software, Mariakerke, Belgium) and the R package (version 4.1.3, Vienna, Austria). Normality testing of all continuous variables was conducted using the Kolmogorov–Smirnov test to assess their distribution. Categorical data were compared using either the Pearson Chi-square test or the Fisher's exact test, depending on the expected cell counts. For continuous variables, presented as mean ± standard deviation, comparisons were made using either the Student's *t*-test for normally distributed data or the Kruskal–Wallis H test for variables with non-normal distributions. To comprehensively evaluate the diagnostic performance of the T-staging classification models, we employed rigorous statistical techniques. Receiver operating characteristic (ROC) curve analysis and the confusion matrix were utilized to assess the models' discriminative abilities in the independent test datasets. Essential performance measures, such as sensitivity, specificity, accuracy, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (PLR), and negative likelihood ratio (NLR), were determined to provide a comprehensive understanding of the models' diagnostic values. Furthermore, to compare the classification models with subjective evaluation, we conducted net reclassification index (NRI) analysis. To gauge the clinical significance of the models, decision curve analysis (DCA) was performed, allowing us to calculate the net benefit. Statistical significance was established at a two-sided *p*-value less than 0.05, indicating strong evidence for significance.

Results

Patients' characteristics

A comprehensive overview of patient demographic characteristics can be found in Table 1. After thorough evaluation, a total of 317 patients were included in the final analysis. Among them, 158 out of 317 cases (49.8%) underwent rectal filling, while the remaining 159 cases (50.2%) were in the non-filling group. Notably, there were no significant differences observed between these two cohorts of patient demographic characteristics ($p > 0.05$). The subsequent examination of the 52 test cases revealed an equal distribution of 26 cases each in

1 http://www.itksnap.org/

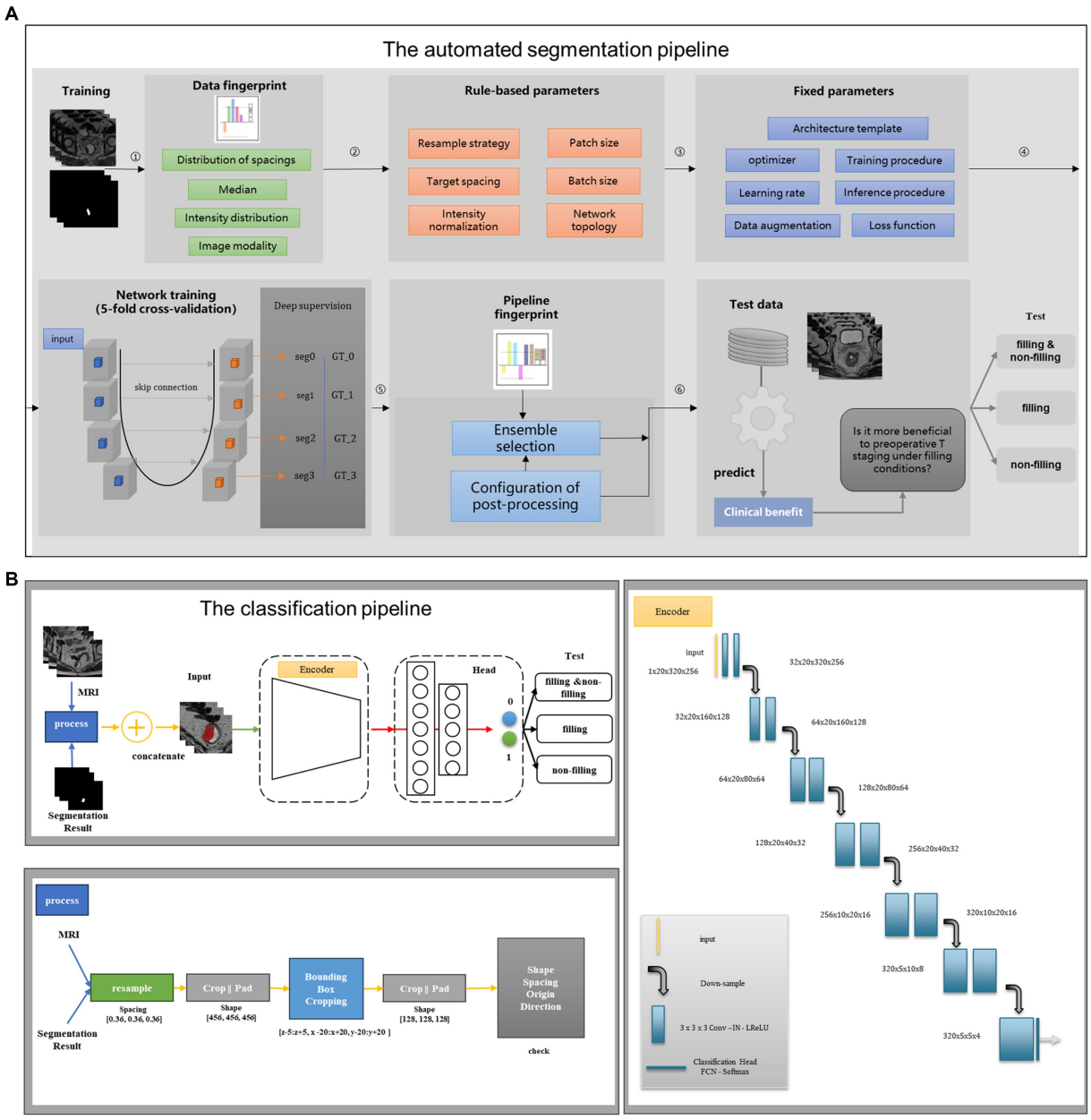


FIGURE 2
Structure diagram of the deep learning model. **(A)** The automated segmentation pipeline; **(B)** the classification pipeline. The first stage is to construct a segmentation flow chart to segment rectal cancer. The second stage is to build a classification flow chart and use the segmentation results of the first stage to do specific staging tasks.

both the filling and non-filling groups. Importantly, no statistically significant differences in T stage were detected between the filling and non-filling groups (T1-2/T3-4: 14/12 vs. 10/16, $p = 0.404$). Moreover, it is noteworthy that none of the cases exhibited positive circumferential resection margin (CRM) involvement.

Automatic segmentation results

In Stage I, we developed a segmentation pipeline utilizing DL models, which can be succinctly referred to as nn-UNet. These automatic segmentation models exhibited exceptional performance in the test datasets, as illustrated in Figure 3. For the overall test dataset, the median values of DSC, HD95, and ASD were 0.835, 2.236 mm, and 0.647 mm, respectively. In the rectal-filling cases, the median values were 0.862 for DSC, 2.118 mm for HD95, and 0.584 mm for ASD. Conversely, in the non-filling cases, the median values were 0.807 for DSC, 3.000 mm for HD95, and 0.879 mm for ASD. Notably, the DSC and ASD values were higher in the rectal-filling dataset ($p = 0.006$ and $p = 0.017$, respectively). The detailed results of the automatic segmentation are presented in Table 2 and Supplementary Figure S1.

TABLE 1 Pathological characteristics of patients.

Variables		Total	Rectal filling	Non-rectal filling	*p* value
		N = 317	*N* = 158	*N* = 159	
Gender					0.326
	Male	206 (65.0%)	98 (62.0%)	108 (67.9%)	
	Female	111 (35.0%)	60 (38.0%)	51 (32.1%)	
Age (years)		59.4 ± 12.0	58.3 ± 9.8	59.8 ± 10.3	0.185
BMI (kg/m^2)		23.9 ± 3.4	23.8 ± 3.0	24.2 ± 3.3	0.260
Tumor location					0.161
	Upper	45 (14.2%)	20 (12.7%)	25 (15.7%)	
	Middle	190 (59.9%)	103 (65.2%)	87 (54.7%)	
	Lower	82 (25.9%)	35 (22.1%)	47 (29.6%)	
Differentiation					0.977
	High-Moderate	250 (78.9%)	124 (78.5%)	126 (79.2%)	
	Poor	67 (21.1%)	34 (21.5%)	33 (20.8%)	
T stage					0.078
	T1-2	180 (56.8%)	98 (62.0%)	82 (51.6%)	
	T3-4	137 (43.2%)	60 (38.0%)	77 (48.4%)	
N stage					0.971
	N0	222 (70.0%)	110 (69.6%)	112 (70.4%)	
	N1-2	95 (30.0%)	48 (30.4%)	47 (29.6%)	
Tumor deposit					0.816
	Negative	242 (76.3%)	122 (77.2%)	120 (75.5%)	
	Positive	75 (23.7%)	36 (22.8%)	39 (24.5%)	
Lymphovascular invasion					0.184
	Negative	190 (59.9%)	101 (63.9%)	89 (56.0%)	
	Positive	127 (40.1%)	57 (36.1%)	70 (44.0%)	
Perineural invasion					0.197
	Negative	152 (47.9%)	82 (51.9%)	70 (44.0%)	
	Positive	165 (52.1%)	76 (48.1%)	89 (56.0%)	
Tumor budding					0.669
	Negative	204 (64.4%)	104 (65.8%)	100 (56.6%)	
	Positive	113 (35.6%)	54 (34.2%)	59 (43.4%)	
CEA*					0.061
	Negative	177 (55.8%)	97 (61.4%)	80 (50.3%)	
	Positive	140 (44.2%)	61 (38.6%)	79 (49.7%)	
CA19-9*					0.593
	Negative	254 (80.1%)	129 (81.6%)	125 (78.6%)	
	Positive	63 (19.9%)	29 (18.4%)	34 (21.4%)	

*Preoperative blood samples. BMI, body mass index; CEA, carcinoembryonic antigen; CA19-9, carbohydrate antigen 19-9.

Classification performance

In Stage II, concentrating on T-staging classification, the subjective evaluation yielded area under the curve (AUC) values of 0.735, 0.810, and 0.713 for the total, filling, and non-filling test datasets, respectively. The corresponding accuracies were 0.731, 0.808, and 0.692. Interobserver agreement for MR T-staging among the three radiologists is presented in Supplementary Table S2. Notably, the DL models outperformed the subjective evaluation across all test datasets. Model 1 achieved AUC values of 0.902, 0.958, and 0.900 for the total, filling, and non-filling test datasets, respectively. The accuracies were 0.846, 0.962, and 0.808, respectively. Model 2, trained on rectal-filling cases, exhibited an AUC of 0.946 and an accuracy of 0.885 in the filling test dataset. Model 3, trained exclusively on non-filling cases,

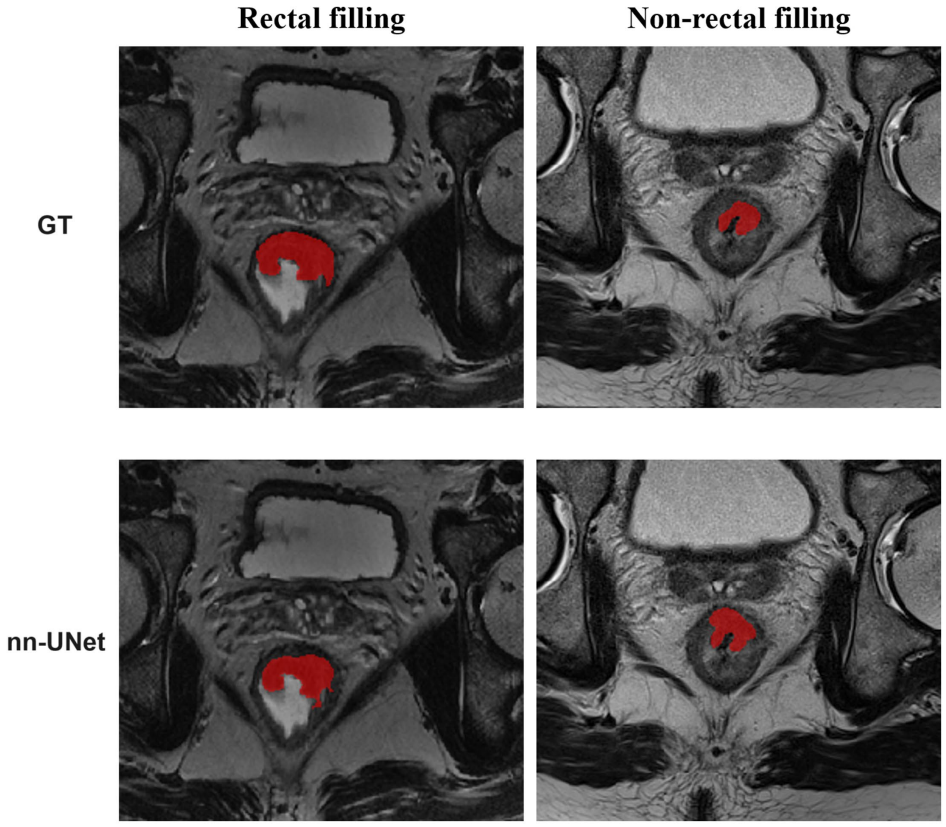


FIGURE 3
Representative diagram of automatic segmentation.

TABLE 2 Automatic segmentation results.

Test data sets	Total	Rectal filling	Non-rectal filling	*p* value*
DSC	0.835 (0.779, 0.871)	0.862 (0.817, 0.879)	0.807 (0.719, 0.850)	0.006
HD95 (mm)	2.236 (2.000, 3.742)	2.118 (1.732, 2.450)	3.000 (2.000, 5.385)	0.127
ASD (mm)	0.647 (0.516, 1.034)	0.584 (0.440, 0.737)	0.879 (0.616, 1.332)	0.017

Median (IQR).
*Kruskal–Wallis rank sum test.
DSC, dice similarity coefficient; HD95, the 95th percentile Hausdorff distance; ASD, average surface distance.

demonstrated an AUC of 0.863 and an accuracy of 0.885 in the non-filling test dataset. Comprehensive ROC analyses are presented in Table 3, and the corresponding curves are depicted in Figure 4.

Model comparison and clinical utility

Compared to subjective evaluation for RC T-staging, NRIs of DL-models were 0.167 to 0.310, demonstrating an improved clinical utility in all datasets (Table 3).

Considering the influence of rectal filling or non-filling, the confusion matrix highlighted the superior classification performance of Model 1 in the rectal-filling dataset compared to Model 1 in the non-filling dataset. Likewise, the performance of Model 2 in the filling dataset outperformed that of Model 3 in the non-filling dataset (Figure 5). The net clinical advantage of Model 1 over Model 2 in the rectal-filling dataset and Model 1 over Model 3 in the non-filling dataset is illustrated by the DCA chart (Figure 6). Overall, Model 1 in the rectal-filling dataset demonstrated notably improved diagnostic performance.

Discussion

In this study, we developed an advanced and automated segmentation model based on nn-UNet to achieve precise segmentation of rectal adenocarcinomas from T2W images, particularly in the rectal-filling dataset. Subsequently, we constructed DL-based classification models that exhibited significantly improved performance in T-staging classification compared to subjective evaluation for RC cases. Notably, Model 1, trained on the total training dataset, demonstrated higher AUC and accuracy in the rectal-filling cohort. To the best of our knowledge, this is the first investigation into the impact of rectal filling on DL models, highlighting its influence on classification performance.

TABLE 3 ROC curve analysis and comparison in the test dataset.

	Total		Filling			Non-filling		
	Model 1	SE	Model 1	Model 2	SE	Model 1	Model 3	SE
AUC	0.902	0.735	0.958	0.946	0.810	0.900	0.863	0.713
95% CI	0.818–0.985	0.594–0.848	0.874–0.999	0.865–0.999	0.653–0.966	0.784–0.999	0.706–0.999	0.533–0.892
Sensitivity	0.750	0.679	1.000	1.000	0.833	0.688	1.000	0.625
Specificity	0.958	0.792	0.929	0.786	0.786	1.000	0.700	0.800
Accuracy	0.846	0.731	0.962	0.885	0.808	0.808	0.885	0.692
PLR	18.000	3.257	14.000	4.667	3.889	Inf.	3.333	3.125
NLR	0.261	0.406	0.000	0.000	0.212	0.313	0.000	0.469
PPV	0.955	0.792	0.923	0.800	0.769	1.000	0.842	0.833
NPV	0.767	0.679	1.000	1.000	0.846	0.667	1.000	0.571
NRI*	0.238		0.310	0.167		0.263	0.275	

SE, subjective evaluation; NPV, negative predictive value; PPV, positive predictive value; NLR, negative likelihood ratio; PLR, positive likelihood ratio.
*NRI, net reclassification index, compared to SE.
Model 1: trained on total training datasets; Model 2: trained on rectal-filling datasets; Model 3: trained on non-filling datasets.

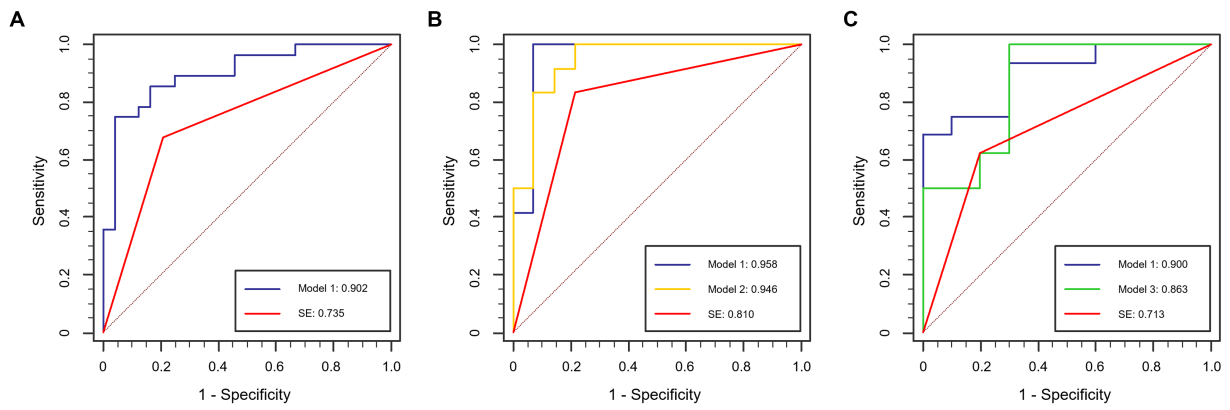


FIGURE 4
ROC curves in the test dataset. **(A)** Total test dataset; **(B)** rectal-filling test dataset; **(C)** non-filling test dataset.

In the current landscape of medical practice, rectal MRI has been widely endorsed as the optimal approach for preoperative T-staging in RC. However, its diagnostic accuracy is compromised by the connective tissue hyperplastic response in the surrounding rectal mesenteric fat, leading to indistinct tumor boundaries. This limitation of traditional MRI techniques in distinguishing between T2 and T3 stages has been well-documented in previous studies (3–5), and our own results align with these observations. Our study conducted a comprehensive ROC analysis, revealing that the subjective discrimination of preoperative T-stage by radiologists was significantly inferior to the proposed DL model. The accuracies of radiologists' assessments ranged from 69.2 to 80.8%. Additionally, the net reclassification index (NRI) analysis demonstrated improved classification performance achieved by employing DL approaches, while decision curve analysis (DCA) highlighted the favorable clinical usefulness of these models. These findings can be attributed to the inherent challenges faced by radiologists in accurately interpreting irregular tumor shapes and blurred boundaries. Therefore, the accurate identification and precise segmentation of lesions serve as crucial prerequisites for future research endeavors aimed at advancing preoperative evaluation and staging methodologies in RC.

The routine utilization of manual or semi-manual segmentation methods is often plagued by inherent challenges, including their arduous and time-consuming nature, as well as their heavy reliance on operator expertise (16, 17). In recent years, several studies have explored the application of 2D convolutional neural network (CNN) models for the discrimination of T2 and T3 stages using 2D MR images (29, 30). However, these approaches introduce an additional burden on radiologists, as they require manual selection of a representative slice (2D) from each MR volume (3D). This manual selection step adds complexity and potential subjectivity to the process. Hou et al. conducted a study where they developed a DL model using 3D T2W images, achieving an impressive AUC value of 0.869 (31). However, it is important to note that the segmentation process in their research was carried out manually, which may introduce subjectivity and potential variability. In a separate study by Wei et al., a multi-parametric MR image fusion model was employed, achieving an AUC of 0.854. This approach involved the manual determination

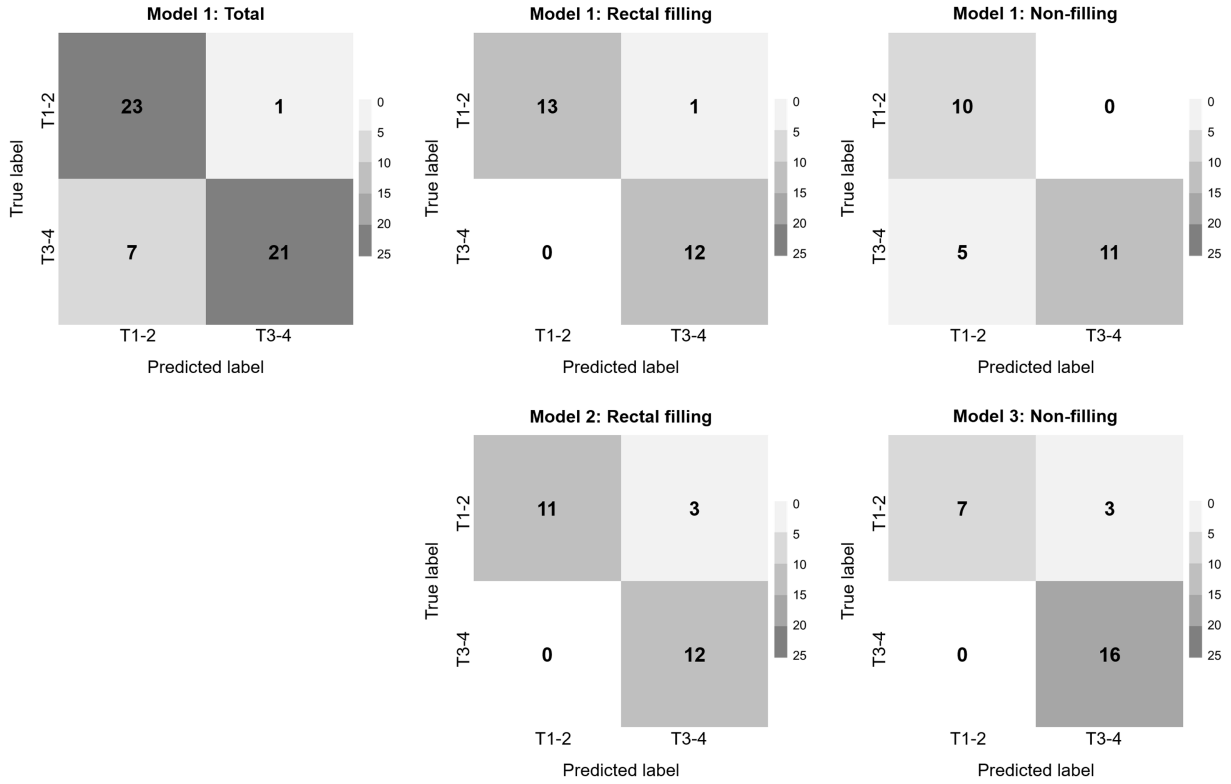


FIGURE 5
Confusion matrices.

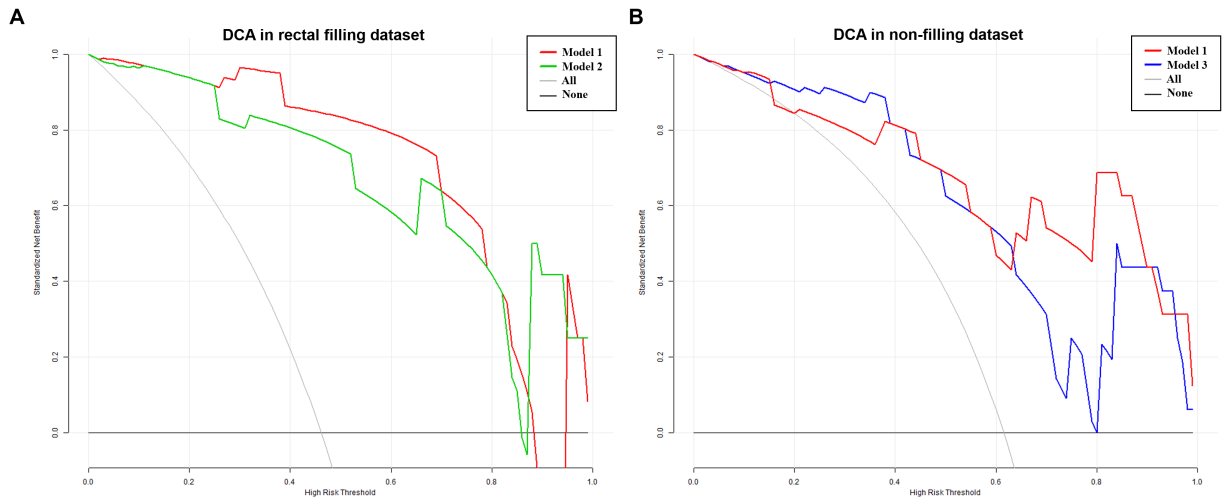


FIGURE 6
DCA in filling and non-filling test datasets. Results for the rectal-filling dataset. **(A)** The net benefit analysis showed that for *p* probability thresholds ranging from 0.25 to 0.88 in the test dataset, Model 1 provided greater benefits compared to Model 2 assessment. Moreover, Model 1 exhibited larger net benefits when compared to all/no intervention methods. **(B)** Results for the non-filling dataset. The net benefit analysis demonstrated that for P probability thresholds ranging from 0.42 to 0.91 in the test dataset, Model 1 yielded additional benefits compared to Model 3 assessment. Furthermore, Model 1 showed larger net benefits when compared to all/no intervention methods.

of the location and size of a 3D bounding box containing the tumor (32).

In Stage I, we successfully developed an automatic segmentation model for rectal adenocarcinomas using a 3D nn-UNet architecture. As a standardized and dataset-agnostic framework, nnU-Net was proposed as a robust and powerful tool for medical image segmentation. The results demonstrated impressive performance, with median values of 0.807–0.862 for DSC, 2.118–3.000 for HD95, and

0.584–0.879 for ASD in the test dataset. To enhance the robustness and generalization of the model while avoiding overfitting, we employed a data augmentation strategy along with 5-fold cross-validation. Furthermore, two experienced radiologists carefully examined the visualizations of the segmentation results, and no noticeable segmentation errors were detected. The implementation of this method bears the potential to serve as a viable replacement for the prevailing manual segmentation method, which is notorious for its time-consuming nature and lack of reproducibility. Subsequently, we conducted additional evaluations on the total test dataset, as well as the rectal-filling and non-filling datasets within the test set. Our findings revealed that the DSC and ASD values were significantly better in the rectal-filling datasets compared to both the total datasets and the non-filling datasets ($p = 0.006$ and $p = 0.017$, respectively). These results suggest that the model exhibited a tendency towards better performance and metrics in rectal-filling cases.

In the Stage II of this study, we introduced a 3D CNN to classify RC lesions as T1-2 or T3-4 stages on HR-T2WI. For the classification models, we utilized widely-used UNet-like Encoder–Decoder architectures. We recognized that directly inputting the original images into the models would make it challenging to distinguish between T1-2 and T3-4 stages, as the models might concentrate on areas other than the cancer of interest. To address this concern, we devised a novel approach using the information from the automated segmentation results obtained in Stage I. We combined the original MRI of RC with its corresponding segmentation result, incorporating them as the complete input. This approach involved employing a region-of-interest cropping strategy, as mentioned earlier. Our initial experiments demonstrated the effectiveness and correctness of this approach compared to solely using the original MRI of RC with a center-cropping strategy. We believe that the center-cropping strategy may not accurately select the cancerous region, as the cancer might not always be located in the center of every image. Although expanding the cropping area could be considered, it would introduce redundant information that is not helpful. Therefore, our cropping strategy, as described above, represents a promising approach for precisely selecting the cancerous region in each original image.

A distinctive feature of our study was its groundbreaking exploration and validation of DL models for preoperative T staging in RC, with a particular focus on the influence of rectal filling. To the best of our knowledge, this was the first endeavor to address this specific aspect, shedding new light on the application of DL in this context.

Our study encompassed the evaluation of automatic segmentation models for rectal adenocarcinomas across three distinct datasets: the total dataset, the rectal-filling dataset, and the non-filling dataset. Following this, three DL models were trained using these datasets to explore their performance. Through a comprehensive analysis involving segmentation results, ROC evaluation, confusion metrics, and DCA, a noteworthy finding emerged: Model 1 exhibited superior performance specifically in the rectal-filling dataset. These results underscore the additional benefits conferred by the use of rectal contrast material in DL models. Previous research has already demonstrated the advantages of rectal filling, including improved lesion visualization and enhanced evaluation of tumor penetration on MRI (2). Furthermore, our previous study has corroborated the value of rectal-filling in accurately delineating rectal lesions and distinguishing them from normal rectal tissue, thus facilitating precise segmentation (6). This likely explains the higher performance observed in the DL model trained on the rectal-filling dataset compared to the non-filling dataset.

Despite the notable contributions of our study, several limitations warrant consideration. Firstly, our dataset consisted solely of HR-T2WI of RC, lacking the inclusion of other imaging modalities. Moreover, being a retrospective single-center study, potential selection biases may have influenced our findings. Thus, for further validation and to enhance the generalizability of our results, larger datasets and multi-center studies incorporating diverse imaging modalities are necessary. Secondly, it is crucial to acknowledge that factors, such as extramural vascular invasion (EMVI), lymph node metastasis (LNM), and mesorectal fascia (MRF) significantly influence the prognosis and survival of RC patients (2–4, 33). While the impact of rectal luminal distention on DL-models pertaining to MRF, EMVI, and LNM remains a topic of debate, further investigation is essential to elucidate these associations comprehensively. Thirdly, an important consideration is the generalizability of our findings to lesions that have undergone neoadjuvant treatment, as this aspect remains elusive and requires further clarification. Finally, the current investigation primarily employed CNN-based models, which are known to perform well with small datasets. However, we did not explore the use of transformer-based models, which are better suited for larger datasets. Therefore, future research should encompass the incorporation of transformer-based models to leverage their potential in handling larger datasets effectively.

Conclusion

Leveraging high-resolution rectal MR imaging, we developed a DL-based segmentation model to automatically extract the region of RC. Subsequently, we constructed DL-based classification models to explore an innovative approach for preoperative T-staging of RC using DL networks. Through a comprehensive comparison, we observed that the DL models exhibited superior predictive capabilities compared to subjective evaluation, particularly in distinguishing between T1-2 and T3-4 stages in the test dataset with rectal-filling. These findings strongly indicate that the DL model, augmented by rectal-filling, holds significant potential as an optimal tool for guiding clinical practice in the preoperative T-staging of RC patients.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by the Shanghai Changhai Hospital, Naval Medical University. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

CT: Data curation, Formal analysis, Writing – original draft. XM: Data curation, Methodology, Software, Writing – review & editing. HL: Data curation, Writing – review & editing. QW: Data curation, Writing – review & editing. CS: Conceptualization, Writing – review & editing. YY: Conceptualization, Formal analysis, Investigation, Methodology, Software, Supervision, Writing – review & editing. FS: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by the Guhai project of Changhai hospital (GH145-09) and the "Yi Yuan Xin Xing" young medical talents funding project of Shanghai (project's number: N/A).

Acknowledgments

We thank Cheng Li and Xiaolin Meng for providing the DL algorithm support in the Research and Advanced Algorithm Department of HSW BU, Shanghai United Imaging Healthcare Co., Ltd., Shanghai, China.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2023.1326324/full#supplementary-material

References

1. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics. *CA Cancer J Clin*. (2023) 73:17–48. doi: 10.3322/caac.21763

2. Gollub MJ, Lall C, Lalwani N, Rosenthal MH. Current controversy, confusion, and imprecision in the use and interpretation of rectal MRI. *Abdom Radiol*. (2019) 44:3549–58. doi: 10.1007/s00261-019-01996-3

3. Horvat N, Carlos Tavares Rocha C, Clemente Oliveira B, Petkovska I, Gollub MJ. MRI of rectal cancer: tumor staging, imaging techniques, and management. *Radiographics*. (2019) 39:367–87. doi: 10.1148/rg.2019180114

4. Nougaret S, Reinhold C, Mikhael HW, Rouanet P, Bibeau F, Brown G. The use of MR imaging in treatment planning for patients with rectal carcinoma: have you checked the "DISTANCE"? *Radiology*. (2013) 268:330–44. doi: 10.1148/radiol.13121361

5. Maas M, Lambregts DMJ, Lahaye MJ, Beets GL, Backes W, Vliegen RFA, et al. T-staging of rectal cancer: accuracy of 3.0 tesla MRI compared with 1.5 tesla. *Abdom Imaging*. (2012) 37:475–81. doi: 10.1007/s00261-011-9770-5

6. Yuan Y, Lu H, Ma X, Chen F, Zhang S, Xia Y, et al. Is rectal filling optimal for MRI-based radiomics in preoperative T staging of rectal cancer? *Abdominal Radiol*. (2022) 47:1741–9. doi: 10.1007/s00261-022-03477-6

7. Li M, Zhang J, Dan Y, Yao Y, Dai W, Cai G, et al. A clinical-radiomics nomogram for the preoperative prediction of lymph node metastasis in colorectal cancer. *J Transl Med*. (2020) 18:46. doi: 10.1186/s12967-020-02215-0

8. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J Clin Oncol*. (2016) 34:2157–64. doi: 10.1200/JCO.2015.65.9128

9. Ma X, Shen F, Jia Y, Xia Y, Li Q, Lu J. MRI-based radiomics of rectal cancer: preoperative assessment of the pathological features. *BMC Med Imaging*. (2019) 19:86. doi: 10.1186/s12880-019-0392-7

10. Liu Z, Wang Y, Shen F, Zhang Z, Gong J, Fu C, et al. Radiomics based on readout-segmented echo-planar imaging (RS-EPI) diffusion-weighted imaging (DWI) for prognostic risk stratification of patients with rectal cancer: a two-center, machine learning study using the framework of predictive, preventive, and personalized medicine [J]. *EPMA J*. (2022) 13:633–47. doi: 10.1007/s13167-022-00303-3

11. Jing G, Chen Y, Ma X, Li Z, Lu H, Xia Y, et al. Predicting mismatch-repair status in rectal cancer using multiparametric mri-based radiomics models: a preliminary study. *Biomed Res Int*. (2022) 2022:1–11. doi: 10.1155/2022/6623574

12. Li Z, Li S, Zang S, Ma X, Chen F, Xia Y, et al. Predicting treatment response to neoadjuvant chemoradiotherapy in rectal mucinous adenocarcinoma using an MRI-based radiomics nomogram. *Front Oncol*. (2021) 11:671636. doi: 10.3389/fonc.2021.671636

13. Liu M, Ma X, Shen F, Xia Y, Jia Y, Lu J. MRI-based radiomics nomogram to predict synchronous liver metastasis in primary rectal cancer patients. *Cancer Med*. (2020) 9:5155–63. doi: 10.1002/cam4.3185

14. Dong D, Fang MJ, Tang L, Shan XH, Gao JB, Giganti F, et al. Deep learning radiomic nomogram can predict the number of lymph node metastasis in locally advanced gastric cancer: an international multicenter study. *Ann Oncol*. (2020) 31:912–20. doi: 10.1016/j.annonc.2020.04.003

15. Wang Y, Liu W, Yu Y, Liu JJ, Xue HD, Qi YF, et al. CT radiomics nomogram for the preoperative prediction of lymph node metastasis in gastric cancer. *Eur Radiol*. (2020) 30:976–86. doi: 10.1007/s00330-019-06398-z

16. Chen Q, Zhang L, Liu S, You J, Chen L, Jin Z, et al. Radiomics in precision medicine for gastric cancer: opportunities and challenges. *Eur Radiol*. (2022) 32:5852–68. doi: 10.1007/s00330-022-08704-8

17. Qin Y, Deng Y, Jiang H, Hu N, Song B. Artificial intelligence in the imaging of gastric cancer: current applications and future direction. *Front Oncol*. (2021) 11:631686. doi: 10.3389/fonc.2021.631686

18. Sahiner B, Pezeshk A, Hadjiiski LM, Wang X, Drukker K, Cha KH, et al. Deep learning in medical imaging and radiation therapy. *Med Phys*. (2019) 46:e1–e36. doi: 10.1002/mp.13264

19. Chen C, Qin C, Qiu H, Tarroni G, Duan J, Bai W, et al. Deep learning for cardiac image segmentation: a review. *Front Cardiovasc Med*. (2020) 7:25. doi: 10.3389/fcvm.2020.00025

20. Bandyk MG, Gopireddy DR, Lall C, Balaji KC, Dolz J. MRI and CT bladder segmentation from classical to deep learning based approaches: current limitations and lessons. *Comput Biol Med*. (2021) 134:104472. doi: 10.1016/j.compbiomed.2021.104472

21. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, et al. The eighth edition AJCC cancer staging manual: continuing to build a bridge from a population-based to a more personalized approach to cancer staging. *CA Cancer J Clin*. (2017) 67:93–9. doi: 10.3322/caac.21388

22. Kaur H, Choi H, You YN, Rauch GM, Jensen CT, Hou P, et al. MR imaging for preoperative evaluation of primary rectal cancer: practical considerations. *Radiographics*. (2012) 32:389–409. doi: 10.1148/rg.322115122

23. Ronneberger Olaf, Fischer Philipp, Brox Thomas. U-net: convolutional networks for biomedical image segmentation medical image computing and computer-assisted intervention-MICCAI, (2015): 18th international conference, Munich, Germany, October 5–9, 2015, Springer International Publishing

24. Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell*. (2017) 39:2481–95. doi: 10.1109/TPAMI.2016.2644615

25. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. (2021) 18:203–11. doi: 10.1038/s41592-020-01008-z

26. Krithika Alias AnbuDevi M, Suganthi K. Review of semantic segmentation of medical images using modified architectures of UNET. *Diagnostics*. (2022) 12:64. doi: 10.3390/diagnostics12123064

27. Taha AA, Hanbury A. An efficient algorithm for calculating the exact Hausdorff distance. *IEEE Trans Pattern Anal Mach Intell*. (2015) 37:2153–63. doi: 10.1109/TPAMI.2015.2408351

28. Yeghiazaryan V, Voiculescu I. Family of boundary overlap metrics for the evaluation of medical image segmentation. *J Med Imaging*. (2018) 5:015006. doi: 10.1117/1.JMI.5.1.015006

29. Wu QY, Liu SL, Sun P, Li Y, Liu GW, Liu SS, et al. Establishment and clinical application value of an automatic diagnosis platform for rectal cancer T-staging based on a deep neural network. *Chin Med J*. (2021) 134:821–8. doi: 10.1097/CM9.0000000000001401

30. Jian J, Xiong F, Xia W, Zhang R, Gu J, Wu X, et al. Fully convolutional networks (FCNs)-based segmentation method for colorectal tumors on T2-weighted magnetic resonance images. *Australas Phys Eng Sci Med*. (2018) 41:393–401. doi: 10.1007/s13246-018-0636-9

31. Hou M, Zhou L, Sun JH. Deep-learning-based 3D super-resolution MRI radiomics model: superior predictive performance in preoperative T-staging of rectal cancer [J]. *Eur Radiol*. (2023) 33:1–10. doi: 10.1007/s00330-022-08952-8

32. Wei Y, Wang H, Chen Z, et al. Deep learning-based multiparametric MRI model for preoperative T-stage in rectal cancer [published online ahead of print, 2023 Jun 27]. *Magn Reson Imaging*. (2023) 2:28856. doi: 10.1002/jmri.28856

33. Gollub MJ, Arya S, Beets-Tan RG, et al. Use of magnetic resonance imaging in rectal cancer patients: Society of Abdominal Radiology (SAR) rectal cancer disease-focused panel (DFP) recommendations 2017. *Abdom Radiol*. (2018) 43:2893–902. doi: 10.1007/s00261-018-1642-9

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 29 November 2023
DOI 10.3389/fmed.2023.1289295

Check for updates

OPEN ACCESS

EDITED BY
Jie-Zhi Cheng,
Shanghai United Imaging Intelligence, Co., Ltd., China

REVIEWED BY
Yihua Song,
Nanjing University of Chinese Medicine, China
Xiaoguang Tu,
Civil Aviation Flight University of China, China

*CORRESPONDENCE
Xigang Xiao
✉ xxgct_417@126.com
Yulin Jia
✉ lin80078558@163.com

†These authors have contributed equally to this work and share first authorship

RECEIVED 05 September 2023
ACCEPTED 27 October 2023
PUBLISHED 29 November 2023

CITATION
Xue M, Lin S, Xie D, Wang H, Gao Q, Zou L, Xiao X and Jia Y (2023) The value of CT-based radiomics in predicting the prognosis of acute pancreatitis.
Front. Med. 10:1289295.
doi: 10.3389/fmed.2023.1289295

COPYRIGHT
© 2023 Xue, Lin, Xie, Wang, Gao, Zou, Xiao and Jia. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

The value of CT-based radiomics in predicting the prognosis of acute pancreatitis

Ming Xue[1†], Shuai Lin[2†], Dexuan Xie[1], Hongzhen Wang[1], Qi Gao[1], Lei Zou[1], Xigang Xiao[1*] and Yulin Jia[1*]

[1]Department of Radiology, First Affiliated Hospital of Harbin Medical University, Harbin, China, [2]Department of Magnetic Resonance, First Affiliated Hospital of Harbin Medical University, Harbin, China

Purpose: Early judgment of the progress of acute pancreatitis (AP) and timely intervention are crucial to the prognosis of patients. The purpose of this study was to investigate the application value of CT-based radiomics of pancreatic parenchyma in predicting the prognosis of early AP.

Materials and methods: This retrospective study enrolled 137 patients diagnosed with AP (95 cases in the progressive group and 42 cases in the non-progressive group) who underwent CT scans. Patients were randomly divided into a training set ($n = 95$) and a validation set ($n = 42$) in a ratio of 7: 3. The region of interest (ROI) was outlined along the inner edge of the pancreatic parenchyma manually, and the Modified CT Severity Index (MCTSI) was assessed. After resampling and normalizing the CT image, a total of 2,264 radiomics features were extracted from the ROI. The radiomics features were downscaled and filtered using minimum redundancy maximum correlation (mRMR) and the least absolute shrinkage and selection operator algorithm (LASSO) regression, in turn, and the more optimal subset of radiomics features was selected. In addition, the radiomics score (rad-score) was calculated for each patient by the LASSO method. Clinical data were also analyzed to predict the prognosis of AP. Three prediction models, including clinical model, radiomics model, and combined clinical–radiomics model, are constructed. The effectiveness of each model was evaluated using receiver operating characteristic (ROC) curve analysis. The DeLong test was employed to compare the differences between the ROC curves. The decision curve analysis (DCA) is used to assess the net benefit of the model.

Results: The mRMR algorithm and LASSO regression were used to select 13 radiomics features with high values. The rad-score of each texture feature was calculated to fuse MCTSI to establish the radiomics model, and both the clinical model and clinical–radiomics model were established. The clinical–radiomics model showed the best performance, the AUC and 95% confidence interval, accuracy, sensitivity, and specificity of the clinical–radiomics model in the training set were 0.984 (0.964–1.000), 0.947, 0.955, and 0.931, respectively. In the validation set, they were 0.942 (0.870–1.000), 0.929, 0.966, and 0.846, respectively. The Delong test showed that the predictive efficacy of the clinical–radiomics model was higher than that of the clinical model ($Z = 2.767$, $p = 0.005$) and the radiomics model ($Z = 2.033$, $p = 0.042$) in the validation set. Decision curve analysis demonstrated higher net clinical benefit for the clinical–radiomics model.

Conclusion: The pancreatic parenchymal CT clinical–radiomics model has high diagnostic efficacy in predicting the progression of early AP patients, which is significantly better than the clinical or radiomics model. The combined model

can help identify and determine the progression trend of patients with AP and improve the prognosis and survival of patients as early as possible.

KEYWORDS

acute pancreatitis, computed tomography, modified CT severity index, radiomics, prognosis

Introduction

Acute pancreatitis (AP) is an inflammatory disease of the exocrine pancreas with a complex and variable clinical course. It is a common acute abdomen of the digestive system. The pathogenesis is due to the abnormal activation of pancreatic enzymes that destroy the pancreas itself and surrounding tissues and organs, with the local inflammatory infiltration of the pancreas as the main feature. The etiology of AP is diverse, mainly biliary tract diseases, alcoholism, hypertriglyceridemia, and less commonly drugs, endoscopic retrograde cholangiopancreatography, postoperative period, metabolic factors, infections, heredity, autoimmune disorders, and trauma (1). According to the pathology, AP is classified into interstitial edematous pancreatitis (EEP) and acute necrotizing pancreatitis (ANP) (2).

Patients can feel severe abdominal pain and often trigger systemic inflammatory response syndrome (SIRS) and multiple organ dysfunction syndrome (MODS) and other severe complications. Then, this triggers pancreatic necrosis and persistent organ failure, which in severe cases can even lead to death with a mortality rate of 1–5% (3). Numerous studies have shown that controlling inflammation within 72 h of onset is crucial for reducing the incidence of complications and mortality in patients (4, 5).

Therefore, early prediction of the development of acute pancreatitis and taking reasonable measures timely are essential for the prognosis of patients. The occurrence of complications can be used as a valid indicator for prognostic assessment of pancreatitis (6). Complications include both immediate and long-term complications. Immediate complications include bleeding, pancreatic leakage, and gastrointestinal perforation (7). However, long-term complications are symptoms that lead to long-term weakness, disease recurrence, and endocrine and exocrine pancreatic insufficiency in some cases.

Clinicians diagnose AP mainly by observing patients' symptoms, signs, and changes in laboratory indicators. However, early clinical symptoms and signs of AP are not specific. Serum amylase and lipase levels do not fully reflect the severity and progression of AP (8, 9). CRP > 150 mg/L suggests a complex course of acute pancreatitis. It has a sensitivity of 85%, but it is not specific. Procalcitonin (PCT) is also considered as a marker to evaluate the prognosis of acute pancreatitis. It is more responsive in the acute phase and can respond to bacterial and/or fungal infections or sepsis (10). Several scoring systems are now clinically available to assess the severity and prognosis of AP, such as the Bedside Index for Severity in Acute Pancreatitis (BISAP) and MCTSI (11). BISAP can identify patients early with a high risk of complications and death, including five indicators: BUN, impaired mental status, SIRS, age, and pleural effusion (12). The MCTSI has a high value in predicting severe acute pancreatitis, pancreatic necrosis, and organ failure (13). They have similar predictive efficacy for AP severity (14, 15). Except for the MCTSI system, all clinical scoring systems are based on clinical information and laboratory data. Although they reflect the pathologic and physiologic status of the patient, they may overlook both the pathoanatomical changes and local complications of AP. Imaging diagnosis is the basis for accurate clinical treatment, and CT is the main method to assess AP complications (16). In the early stage of AP, some patients' pancreatic parenchymal changes are not significant on CT plain scan. Radiomics is the more novel technology of the moment. It applies high-throughput computation to rapidly extract features from tomograms and quantify them. It converts digital medical images into high-dimensional data with the aim of revealing biomedical images that reflect underlying pathophysiological information that cannot be observed by the naked eye through quantitative image analysis (17). A large number of studies have confirmed that a single radiomics model has great value in predicting the severity and recurrence of acute pancreatitis (18, 19). However, there is currently no research exploring the value of clinical–radiomics models in predicting the prognosis of AP patients.

This study is based on pancreatic CT plain scan images for radiomics analysis and fusion of clinical data. Exploring the prediction of complications in pancreatitis based on clinical–radiomics models to determine the prognosis of early AP.

Materials and methods

Patients

A total of 137 patients from April 2021 to November 2022 at the First Affiliated Hospital of Harbin Medical University were retrospectively collected, and imaging data and relevant clinical laboratory data of patients with a clinical diagnosis of AP were enrolled. Ethics committee approval was granted for this retrospective study, and the requirement for written informed consent was waived. The diagnostic criteria of AP were based on the 2012 revised Atlanta Classification of AP (2).

Inclusion criteria: ① Patients with the first onset of pancreatitis; ② the time interval between the onset and the examination was not more than 1 week; ③ good CT imaging quality.

Exclusion criteria: ① Patients with incomplete images, poor quality, or incomplete patient case information data; ② autoimmune pancreatitis, trauma, or recurrent pancreatitis and AP due to pancreatic tumor; ③ difficulty in outlining pancreatic parenchymal ROI.

Patients were divided into progressive and non-progressive groups according to the presence or absence of new local or systemic complications or exacerbation of complications. Patients were randomly divided into a training set and a validation set in a ratio of 7:3.

TABLE 1 Bedside index of severity of acute pancreatitis (BISAP).

Risk factors	Scoring criteria	Scores
BUN	>25 mg/dL	1
	≤25 mg/dL	0
Impaired mental status	2	1
	1	0
SIRS	≥2 Criteria	1
	<2 Criteria	0
Age	>60 years	1
	≤60 years	0
Pleural effusion	Presence	1
	Absence	0

TABLE 2 MCTSI scores.

Index		MCTSI
Pancreatic inflammation	Normal pancreas	0
	Intrinsic pancreatic abnormalities with or without inflammatory changes in peripancreatic fat	2
	Pancreatic or peripancreatic fluid collection or peripancreatic fat necrosis	4
Pancreatic necrosis	None	0
	30% or less	2
	More than 30%	4
Extrapancreatic complications	one or more of pleural effusion, ascites, vascular complications, parenchymal complications, and/or gastrointestinal involvement	2

Clinical information

The medical records were reviewed to collect baseline clinical and imaging information, including sex, age, BISAP, CPR, PCT, and MCTSI. The BISAP and MCTSI criteria are shown in Tables 1, 2.

BISAP is an abbreviation for five indicators. They are as follows: B: BUN; I: impairment; S: SIRS; A: age; P: pleural effusion. Note: SIRS has two or more of the following signs: ① temperature > 38°C or < 36°C; ② heart rate > 90 beats/min; ③ respiration >20 breaths/min or $PaCO_2$ < 32 mmHg; ④ white blood cell count $>12.0\times10^9$ /L or $<4.0\times10^9$/L or infantile cells >10%.

CT image acquisition and image analysis

Siemens 64-row spiral CT was used to scan the mid-abdomen of the AP patient, with the patient in the supine position and hands raised flat over the head. Scanning with advanced head. Image acquisition parameters: tube voltage of 120 kV, tube current of 220 mA, layer thickness of 5 mm, interlayer of 5 mm, DFOV of 32 cm × 32 cm, rotation speed of 0.28 s/turn, pitch of 1.7, image layer thickness of 5 mm, matrix 512 × 512, and reconstructed thin layer of 1-mm image.

The CT plain scan images of all patients were exported in DICOM format from the image archiving and communication system PACS workstation, and all of them were uploaded to the uAI Research Portal (uAI Research Portal version: 20220915, Shanghai United Imaging Intelligence, Co., Ltd.). Moreover, the radiomics feature extraction, feature selection, and machine learning models building were established on the uAI Research Portal (version: 20220915, Shanghai United Imaging Intelligence, Co., Ltd.), which was integrated with PyRadiomics (version 2.2.0), Scikit-Learn (version 1.2.0), and so on. The CT images were analyzed by two radiologists with 10 and 20 years of clinical experience blinded to outline the region of interest (ROI) along the edge of the pancreatic parenchyma by manual segmentation (Figure 1), and MCTSI scoring was performed, and in case of disagreement, the decision was made after discussion between the two physicians.

Image preprocessing and feature extraction

To ensure repeatability, gray intensity normalization and resampling were performed to eliminate the heterogeneity between different scanners before feature extraction. Images were resampled to 1 mm × 1 mm × 1 mm voxels using the B-Spline interpolation method and normalized by subtracting the window level (WL: 30) and dividing by the window width (WW: 300). The extractable feature groups include first-order features, shape, and texture features (gray-level co-occurrence matrix [GLCM], gray-level travel length matrix [GLRLM], gray-level size zone matrix [GLSZM], gray-level dependence matrix [GLDM], and neighborhood gray-tone difference matrix [NGTDM]). The filter transforms include 14 filters, such as the Laplace-Gaussian filter, wavelet analysis, and local binary mode transform. A total of 2,264 radiomic features were finally extracted from the pancreatic parenchyma.

Intraclass correction coefficients (ICCs) were used to calculate intra-observer and inter-observer reproducibility (20). Two radiologists, A and B, manually delineate ROIs for all patient images. Moreover, 2 weeks later, radiologist B performed a second manual delineation to select image features with inter-observer and intra-observer ICCs >0.75.

Filtering and establishment of radiomics labels

To avoid overfitting, it is necessary to reduce the dimensionality of the image data before the establishment of the radiomics label. The obtained radiomic features were reduced and filtered by using mRMR algorithm and LASSO regression to obtain the optimal subset of features, and the rad-score of each patient was calculated for each patient based on feature weights.

Construction of the radiomics model

The obtained rad-score was fused with the MCTSI score to create a radiomics model. The baseline data were used to build a clinical model. A joint clinical–radiomics model was established by combining

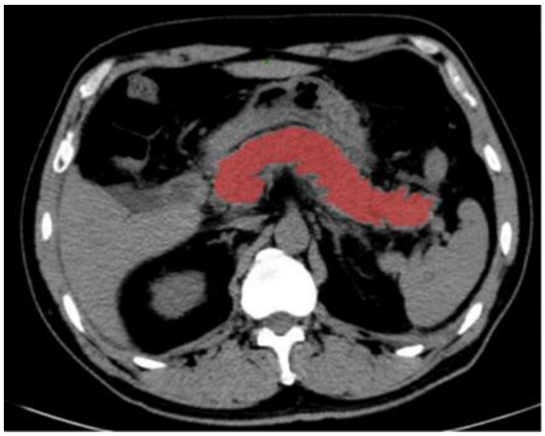


FIGURE 1
Manually delineate the ROI of pancreatic parenchyma with the uAI Research Portal.

the two. The AUC values of the three models were analyzed using ROC curves. Compare the performance of different models in the training and validation sets. The performance of the same model is also compared in the training and validation sets. The DeLong test was used to compare the statistical differences among the three models. Evaluate clinical–radiomics characteristics by univariate and multivariate logistic regression analyses. The net clinical benefit of the models was compared by DCA.

Statistical analysis

Statistical analysis was performed using SPSS (version 25.0, IBM) and R statistical software (Version 4.1.0). Clinical data for the training set and validation set were analyzed according to the variable type. Continuous variables that obeyed normal distribution are presented as mean ± standard deviation and analyzed by *t*-test for comparison of differences between groups; continuous variables that did not obey normal distribution are presented as median (interquartile range, IQR), and Wilcoxon rank sum test was used for comparison of differences between groups. Categorical variables were presented as frequency, and differences between groups were compared using the chi-square test or Fisher's exact test for comparison of differences between groups. $p < 0.05$ was considered statistically different. The AUC and 95% confidence interval, accuracy, sensitivity, and specificity were used to evaluate the performance of the models, and the DeLong test was used to compare the differences between the ROC curves of the three models. DCA is used to compare net clinical benefit.

Results

Clinical characteristics and MCTSI scores

A total of 137 AP patients were enrolled in this study, which consisted of 84 male and 53 female cases with a mean age of 43 (35, 53) years. Patients were divided into progressive ($n = 95$) and non-progressive ($n = 42$) groups. There were 95 cases in the training set (66 cases in the progressive group and 29 cases in the non-progressive group) and 42 cases in the validation set (29 cases in the progressive group and 13 cases in the non-progressive group).

In both the training set and validation set, the BISAP ($Z = -5.19$; -3.55, $P < 0.001$) and PCT ($Z = -3.92$; -2.2, $P < 0.001$; $p = 0.028$) in the progression group were higher than in the non-progression group, while the gender differences were not statistically significant ($Z = 2.27$; 0.14, $p = 0.707$; 0.132). In the training set, the age in the progressive group was younger than the non-progressive group ($Z = 2.11$, $p = 0.035$), and the CRP was higher than the non-progressive group ($Z = -3.05$, $p = 0.002$) (Table 3).

In both the training set and validation set, the MCTSI scores in the progression group were higher than in the non-progression group ($Z = -5.3$; -3.34, $P < 0.001$).

Radiomics analysis

The meaningful texture features were obtained by extracting the texture and filtering the transform by the uAI Research Portal. Selection of the best parameters for binomial bias through *Z*-score, ICC, mRMR algorithm, and LASSO regression. Figure 2A shows that λ increases the variation of binomial deviation of the model on the training set samples. The value with the smallest binomial deviation is selected as the best parameter value. Figure 2B shows the variation of the coefficients of the variables in the model. Thirteen texture features with large values were selected from the CT plain scan images, and their corresponding coefficients are shown in Figure 3. Calculate the rad-score for each patient based on the following formula:

Radscore = 1.0661 + 1.0718 × recursiveGaussian_glcm_ClusterShade+0.4923 × additivegaussiannoise_glrlm_RunEntropy+0.3503 × boxsigmaimage_glszm_LowGrayLevelZoneEmphasis+0.2422 × normalize_glszm_ZoneEntropy-0.0386 × log_ngtdm_log_sigma_2_0_mm_3D_Strength-0.0424 × wavelet_firstorder_wavelet_LLL_Kurtosis-0.048 × mean_gldm_DependenceVariance-0.1576 × log_glszm_log_sigma_2_0_mm_3D_SizeZoneNonUniformity-0.1904 × specklenoise_glszm_SmallAreaEmphasis-0.1940 × normalize_glszm_LargeAreaLowGrayLevelEmphasis-0.3478 × wavelet_gldm_wavelet_HHL_DependenceVariance-0.4221 × log_glcm_log_sigma_0_5_mm_3D_Correlation-0.5333 × boxmean_glszm_SmallAreaHighGrayLevelEmphasis.

The rad-score fused MCTSI is derived for each patient by calculating the coefficients of each texture feature to build a radiomics model.

Based on the texture features of plain scan images, the optimal features are selected. By fusing MCTSI scores, establish clinical models, radiomics models, and clinical–radiomics models, respectively. According to the ROC curve (Table 4), the AUCs in the training set were 0.911, 0.933, and 0.984, respectively, and the AUCs in the validation set were 0.857, 0.897, and 0.942, respectively. In the training set, the clinical–radiomics model has the best predictive performance. Its predictive performance is higher than the clinical model ($Z = 2.767$, $p = 0.005$) and the radiomics model ($Z = 2.033$, $p = 0.042$) (Figure 4). There was no statistically significant difference in AUC between the training and validation sets for clinical models ($D = 0.644$, $p = 0.522$),

TABLE 3 Intergroup comparison of clinical indicators and MCTSI Scores in the training set and validation set for patients in the progressive and non-progressive groups.

Project	Training set ($n = 95$)		χ^2/z	P	Validation set ($n = 42$)		χ^2/z	P
	Non-progressive group ($n = 29$)	Progressive group ($n = 66$)			Non-progressive group ($n = 13$)	Progressive group ($n = 29$)		
Gender			2.27	0.1322			0.14	0.7066
Male	21 (72.41)	37 (56.06)			7 (53.85)	19 (65.52)		
Female	8 (27.59)	29 (43.94)			6 (46.15)	10 (34.48)		
Age	47 (40.00,54.00)	39 (32.00,50.75)	2.11	0.0348	45.62 ± 16.83	43.52 ± 15.39	0.40	0.6935
BISAP	1 (1,1)	2 (2,3)	−5.19	<0.0001	1 (0,1)	2 (1,3)	−3.55	0.0004
CRP	141 (94.4,191.0)	201 (141.8,289.6)	−3.05	0.0023	185 (34,250)	220.58 (161,392)	−1.80	0.0725
PCT	0.79 (0.46,1.50)	1.93 (0.91,4.24)	−3.92	0.0001	1.1 (0.25,1.60)	1.93 (0.87,8.15)	−2.20	0.0275
MCTSI	2 (0,2)	4 (2,5.5)	−5.3	<0.0001	2 (0,2)	4 (2,6)	−3.34	0.0008

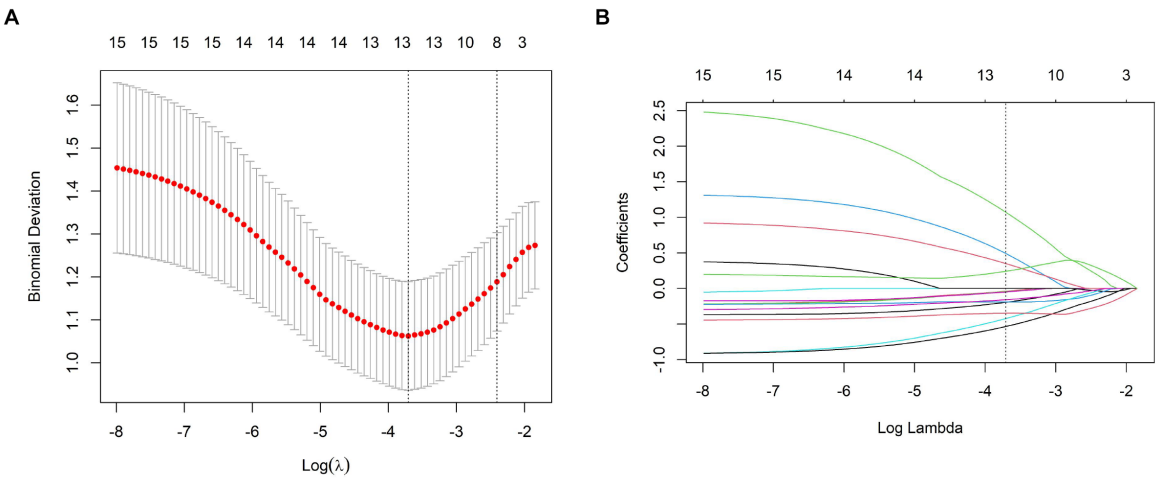


FIGURE 2
(A) The binomial deviation of the LASSO model in the training set samples with λ of the binomial deviation, and the binomial deviation with the smallest λ value as the best λ value. **(B)** Change of the coefficients of the variables in the LASSO model with λ.

radiomics models ($D = 0.823$, $p = 0.414$), and clinical–radiomics models ($D = 1.108$, $p = 0.274$). The results of DCA indicate that the clinical–radiomics model has a higher clinical net benefit (Figure 5).

Assessment of clinical–radiomics features

The results of univariate and multivariate logistic regression analysis of clinical–radiomics features predicting the progression of acute pancreatitis are shown in Table 5. The results of univariate logistic regression analysis showed that BISAP, CRP, PCT, MCTSI, and rad-score were significant factors in differentiating the progression of acute pancreatitis. The results of multivariate logistic regression analysis showed that BISAP (OR = 6.187; 95% *CI:* 2.259–27.047; $p = 0.003$), PCT (OR = 1.923; 95% *CI*: 1.040–5.568; $p = 0.124$), and rad-score (OR = 3.841; 95% *CI*: 1.578–13.481; $p = 0.013$) were independent predictors of progression of acute pancreatitis, and two factors, MCTSI (OR = 1.281; 95% *CI:* 0.690–2.617; $p = 0.455$) and CRP (OR = 1.001; 95% *CI*: 0.992–1.010; $p = 0.872$), were not independent factors.

Discussion

The initial diagnosis of AP patients accounts for approximately 0.3% of the total number of emergency department patients. The pathogenesis is due to the abnormal activation of pancreatic enzymes. The pancreatic enzyme causes damage to the pancreas itself and surrounding organs, which, in turn, triggers related inflammatory and immune responses. In severe cases, it can lead to organ dysfunction and even death (21). There is evidence that mild and acute pancreatitis is associated with hyperperfusion within the first few hours of symptom onset, while moderate and severe pancreatitis is accompanied by complications, and pancreatic parenchyma has progressive tissue ischemia and decreased perfusion. It has been suggested that damage to the pancreatic parenchyma from microcirculatory disorders is present at an early stage, and it is difficult to reflect such changes by conventional CT plain scan (22). In this study, we established and validated a clinical–radiomics model based on CT plain scan images to predict the progression of AP in a non-invasive and quantitative way.

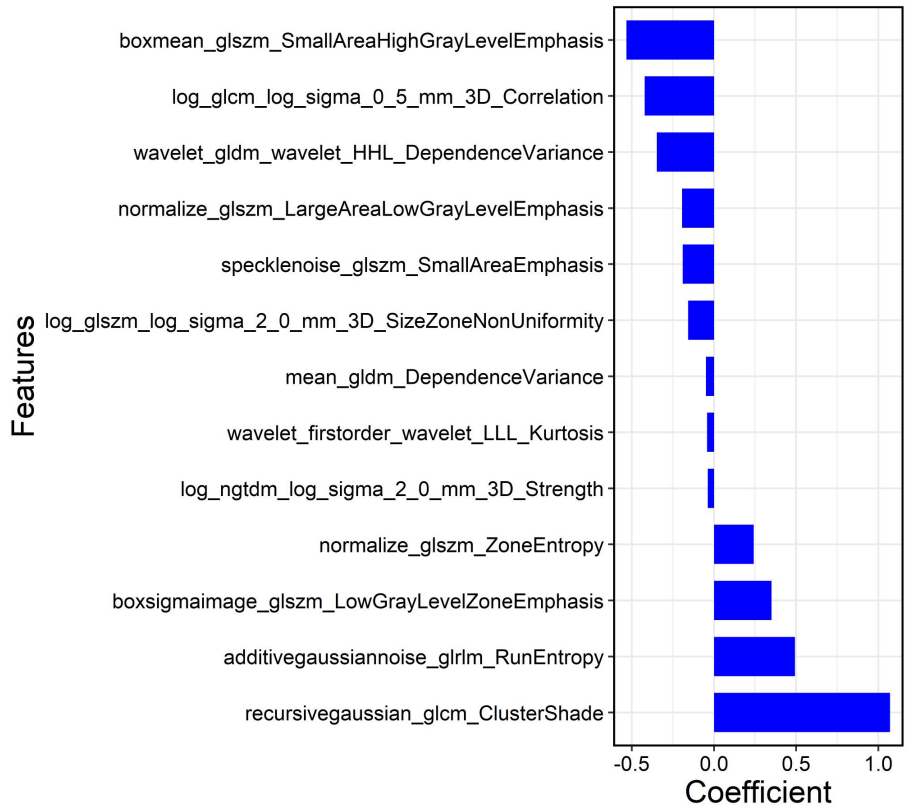


FIGURE 3
Coefficients of the 13 texture features selected by LASSO.

TABLE 4 Clinical characteristics model, radiomics label, and radiomics prediction model ROC result.

Model	Training set (n = 95)						Validation set (n = 42)					
	AUC	ACC	Sensitivity	Specificity	PPV	NPV	AUC	ACC	Sensitivity	Specificity	PPV	NPV
Clinical features	0.911 (0.853–0.969)	0.853 (0.765–0.917)	0.864	0.828	0.919	0.727	0.857 (0.741–0.973)	0.81 (0.659–0.914)	0.793	0.846	0.920	0.647
Radiomics	0.933 (0.875–0.990)	0.895 (0.815–0.948)	0.909	0.862	0.938	0.806	0.897 (0.803–0.990)	0.81 (0.659–0.914)	0.759	0.923	0.957	0.632
Clinical–radiomics	0.984 (0.964–1.000)	0.947 (0.881–0.983)	0.955	0.931	0.969	0.900	0.942 (0.87–1.000)	0.929 (0.805–0.985)	0.966	0.846	0.933	0.917

Majidi et al. (23) found CRP and PCT in the SAP group were higher than in the MAP group ($p < 0.05$). With the progressing pancreatitis, PCT and CRP also increased. BISAP can early identify patients with a high risk of complications and death using five simple indicators (12). Singh et al. (22) found that the BISAP can effectively predict the severity of AP. They found that gender was balanced by comparing clinical baseline data. Khanna et al. (10) found that the age of AP patients is between 21 and 50 years old. By comparing the clinical characteristics of the training set and the validation set, we found that the BISAP, CRP, and PCT in the progressive group were higher than those in the non-progressive group, and the age in the progressive group was smaller than in the non-progressive group.

Balthazar et al. (24–26) established the CT Severity Index (CTSI). It grades and scores the degree of pancreatitis and necrosis to predict the incidence rate and mortality of AP. Later, Mortele et al. (27) improved its limitations by incorporating numerical scores for extrapancreatic complications and establishing MCTSI. Its correlation with clinical outcomes and local complications is superior to CTSI. Banday et al. (28) found that compared to CTSI, MCTSI is simpler and more accurate and has a stronger correlation with clinical outcomes, including hospital stay, infection development trends, organ failure, mortality rate, and the need for intervention treatment. Therefore, we fused MCTSI to the radiomics model to improve the predictive performance. Radiomics can reveal

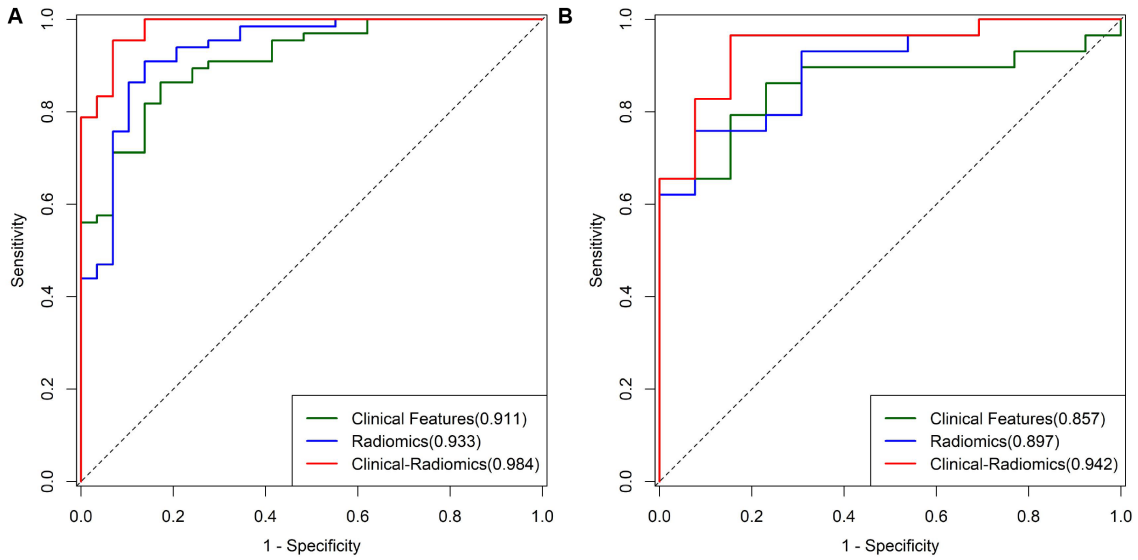


FIGURE 4
(A) ROC curves of different models in the training group. **(B)** ROC curves of different models in the validation group.

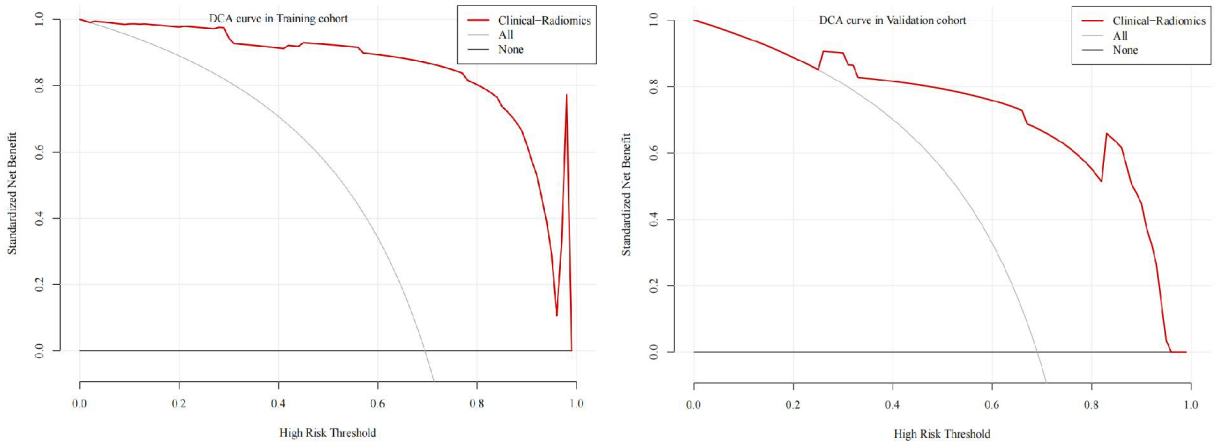


FIGURE 5
DCA of the models in the training set and validation set.

TABLE 5 **Results of univariate and multivariate logistic regression analysis of clinical–radiomics characteristics.**

Clinical–radiomics features	Univariate analysis		Multivariate analysis	
	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value
BISAP	5.167 (2.630–12.094)	<0.001	6.187 (2.259–27.047)	0.003
CRP	1.006 (1.002–1.011)	0.014	1.001 (0.992–1.010)	0.872
PCT	2.034 (1.381–3.511)	0.003	1.923 (1.040–5.568)	0.124
MCTSI	2.658 (1.791–4.477)	<0.001	1.281 (0.690–2.617)	0.455
age	0.978 (0.949–1.007)	0.139		
gender	2.057 (0.819–5.558)	0.136		
Rad-score	5.037 (2.799–10.728)	<0.001	3.841 (1.578–13.481)	0.013

information hidden in conventional images to reflect potential biological and pathological changes.

In the early stage of a conventional CT plain scan, radiomics may reflect damage to the pancreatic parenchyma. In this study, a series of radiomics features of the first order, shape, glcm, glrlm, glszm, gldm, and ngtdm parameters were extracted with high throughput. They provided a comprehensive description of the morphology, radiographic attenuation, and texture information of the lesion. After the dimensionality reduction and radiomics analysis, the 13 most valuable texture features were selected. Then, the rad-score was calculated to obtain. The radiomics feature with the maximum absolute values in LASSO is recursivegaussian_glcm_ClusterShade. Cluster Shade is a measure of the skewness and uniformity of the GLCM. It is used to describe the joint distribution of two pixels with some spatial relationship. ClusterShade is associated with the heterogeneity of voxels in the area of interest. The maximum absolute values in LASSO indicate that the internal structure of the pancreatic parenchyma varies greatly due to inflammation or necrosis caused by leakage of pancreatic fluid. In addition, less first-order feature extraction may be related to the exudate parenchymal necrosis of inflammation resulting in poorer display of anatomical details. Moreover, we obtain the texture features containing wavelet filter decomposition and local binomial transform account for more. The image signal can be decomposed into subbands. Using different algorithms for different subbands aims to highlight approximation and details at different scales. Wavelet transforms have important image analysis capabilities. The wavelet transform coefficients can provide us with edge information of the lesion for extracting relevant radiomics features. The local binomial transform can extract the original feature information and spatial positional relationships of the image pixels. This analysis method has the advantages of rotation and gray-scale invariance (29). It can effectively reflect the invariance of local information in a CT plain scan (30).

Therefore, we establish a radiomics model with rad-score and MCTSI. We use clinical baseline data to establish a single predictive model. By comparing the AUC, we found that the performances of the radiomics model and the clinical model were good and stable. Finally, the radiomics model was fused with clinical baseline data to obtain a joint clinical–radiomics model. The performance has been further improved, with statistically significant differences compared to single clinical models or radiomics models. Moreover, the diagnostic efficiency of the clinical–radiomics model is the highest in both the training and validation sets. Moreover, there is no statistical difference between the two sets, which shows that the model obtained through further research has the best and stable predictive performance. Early prediction of patient progression can be achieved through quantitative analysis of several simple clinical data and patient CT plain scan images.

Limitations of this study

(1) The analysis of this study was based on radiomics features extracted from two-dimensional images of the pancreatic parenchyma at the largest level of the diseased pancreas, and it may be more effective to reflect the lesions if three-dimensional images of the entire pancreatic parenchyma are extracted; (2) a retrospective single-center study was conducted with hospitalized patients. There may be significant bias in sample selection; (3) AP patients in our hospital undergo routine CT plain scans upon admission. If conducting radiomics analysis based on CECT scanning images, a more efficient model can be obtained; (4) patients with mild symptoms do not need image examinations to diagnose and manage AP. Currently, there are few imaging studies on patients with mild symptoms. Moreover, our results tend to analyze more severe and complicated pancreatitis.

In summary, the clinical–radiomics model based on pancreatic parenchymal CT has good predictive performance in both the training and validation sets, helping to identify and judge the progression trend of acute pancreatitis patients as soon as possible, and taking timely and effective measures to improve the patient's prognosis.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by the Ethics Committee of First Affiliated Hospital of Harbin Medical University. The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because (1) this study only retrospectively analyzed the patient's imaging and clinical data, without any intervention on the patient's condition, and there is no risk to the life and health of the subjects. (2) For the collection or research of previously archived data, files, records, and imaging data, researchers are unable to contact the subjects.

Author contributions

MX: Conceptualization, Methodology, Writing – original draft. SL: Data curation, Methodology, Writing – original draft. DX: Data curation, Writing – original draft. HW: Formal analysis, Resources, Writing – original draft. QG: Investigation, Writing – original draft. LZ: Investigation, Writing – original draft. XX: Conceptualization, Project administration, Writing – review & editing. YJ: Methodology, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Misgar RA, Bhat MH, Rather TA, Masoodi SR, Wani AI, Bashir MI, et al. Primary hyperparathyroidism and pancreatitis. *J Endocrinol Investig*. (2020) 43:1493–8. doi: 10.1007/s40618-020-01233-5

2. Banks PA, Bollen TL, Dervenis C, Gooszen HG, Johnson CD, Sarr MG, et al. Acute pancreatitis classification working group. Classification of acute pancreatitis--2012: revision of the Atlanta classification and definitions by international consensus. *Gut*. (2013) 62:102–11. doi: 10.1136/gutjnl-2012-302779

3. Petrov MS, Yadav D. Global epidemiology and holistic prevention of pancreatitis. *Nat Rev Gastroenterol Hepatol*. (2019) 16:175–84. doi: 10.1038/s41575-018-0087-5

4. Zuidema MJ, van Santvoort HC, Besselink MG, van Ramshorst B, Boerma D, Timmer R, et al. The predictive value of proteinuria in acute pancreatitis. *Pancreatology*. (2014) 14:484–9. doi: 10.1016/j.pan.2014.09.004

5. Buxbaum JL, Quezada M, Da B, Jani N, Lane C, Mwengela D, et al. Early aggressive hydration hastens clinical improvement in mild acute pancreatitis. *Am J Gastroenterol*. (2017) 112:797–803. doi: 10.1038/ajg.2017.40

6. Gupta P, Rana P, Bellam BL, Samanta J, Mandavdhare H, Sharma V, et al. Site and size of extrapancreatic necrosis are associated with clinical outcomes in patients with acute necrotizing pancreatitis. *Pancreatology*. (2020) 20:9–15. doi: 10.1016/j.pan.2019.11.010

7. Bradley EL 3rd. A clinically based classification system for acute pancreatitis. *Arch Surg*. (1993) 128:586–90. doi: 10.1001/archsurg.1993.01420170122019

8. Koziel D, Suliga E, Grabowska U, Gluszek S. Morphological and functional consequences and quality of life following severe acute pancreatitis. *Ann Ital Chir*. (2017) 6:403–11.

9. Kim DB, Paik CN, Song DS, Kim HA, Kim YJ, Lee JM, et al. The role of endoscopic ultrasonography and magnetic resonance cholangiopancreatography in patients with acute pancreatitis after negative computed tomography findings of the etiology. *Pancreas*. (2018) 47:1165–71. doi: 10.1097/MPA.0000000000001149

10. Khanna AK, Meher S, Prakash S, Tiwary SK, Singh U, Srivastava A, et al. Comparison of Ranson, Glasgow, MOSS, SIRS, BISAP, APACHE-II, CTSI scores, IL-6, CRP, and Procalcitonin in predicting severity, organ failure, pancreatic necrosis, and mortality in acute pancreatitis. *HPB Surg*. (2013) 2013:367581. doi: 10.1155/2013/367581

11. Harshit Kumar A, Singh GM. A comparison of APACHE II, BISAP, Ranson's score and modified CTSI in predicting the severity of acute pancreatitis based on the 2012 revised Atlanta classification. *Gastroenterol Rep*. (2018) 6:127–31. doi: 10.1093/gastro/gox029

12. Papachristou GI, Muddana V, Yadav D, O'Connell M, Sanders MK, Slivka A, et al. Comparison of BISAP, Ranson's, APACHE-II, and CTSI scores in predicting organ failure, complications, and mortality in acute pancreatitis. *Am J Gastroenterol*. (2010) 105:435–41. doi: 10.1038/ajg.2009.622

13. Alberti P, Pando E, Mata R, Vidal L, Roson N, Mast R, et al. Evaluation of the modified computed tomography severity index (MCTSI) and computed tomography severity index (CTSI) in predicting severity and clinical outcomes in acute pancreatitis. *J Dig Dis*. (2021) 22:41–8. doi: 10.1111/1751-2980.12961

14. Cho JH, Kim TN, Chung HH, Kim KH. Comparison of scoring systems in predicting the severity of acute pancreatitis. *World J Gastroenterol*. (2015) 21:2387–94. doi: 10.3748/wjg.v21.i8.2387

15. Chauhan S, Forsmark CE. The difficulty in predicting outcome in acute pancreatitis. *Am J Gastroenterol*. (2010) 105:443–5. doi: 10.1038/ajg.2009.623

16. Fung C, Svystun O, Fouladi DF, Kawamoto S. CT imaging, classification, and complications of acute pancreatitis. *Abdom Radiol (NY)*. (2020) 45:1243–52. doi: 10.1007/s00261-019-02236-4

17. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. (2016) 278:563–77. doi: 10.1148/radiol.2015151169

18. Chen Y, Chen TW, Wu CQ, Lin Q, Hu R, Xie CL, et al. Radiomics model of contrast-enhanced computed tomography for predicting the recurrence of acute pancreatitis. *Eur Radiol*. (2019) 29:4408–17. doi: 10.1007/s00330-018-5824-1

19. Lin Q, Ji YF, Chen Y, Sun H, Yang DD, Chen AL, et al. Radiomics model of contrast-enhanced MRI for early prediction of acute pancreatitis severity. *J Magn Reson Imaging*. (2020) 51:397–406. doi: 10.1002/jmri.26798

20. Meng X, Xia W, Xie P. Preoperative radiomic signature based on multiparametric magnetic resonance imaging for noninvasive evaluation of biological characteristics in rectal cancer. *Eur Radiol*. (2019) 29:3200–9. doi: 10.1007/s00330-018-5763-x

21. Peery AF, Crockett SD, Murphy CC, Lund JL, Dellon ES, Williams JL, et al. Burden and cost of gastrointestinal, liver, and pancreatic diseases in the United States: update 2018. *Gastroenterology*. (2019) 156:254–272.e11. doi: 10.1053/j.gastro.2018.08.063

22. Singh VK, Wu BU, Bollen TL, Repas K, Maurer R, Mortele KJ, et al. Early systemic inflammatory response syndrome is associated with severe acute pancreatitis. *Clin Gastroenterol Hepatol*. (2009) 7:1247–51. doi: 10.1016/j.cgh.2009.08.012

23. Majidi S, Golembioski A, Wilson SL, Thompson EC. Acute pancreatitis: etiology, pathology, diagnosis, and treatment. *South Med J*. (2017) 110:727–32. doi: 10.14423/SMJ.0000000000000727

24. Balthazar EJ, Robinson DL, Megibow AJ, Ranson JH. Acute pancreatitis: value of CT in establishing prognosis. *Radiology*. (1990) 174:331–6. doi: 10.1148/radiology.174.2.2296641

25. Balthazar EJ. Acute pancreatitis: assessment of severity with clinical and CT evaluation. *Radiology*. (2002) 223:603–13. doi: 10.1148/radiol.2233010680

26. Balthazar EJ, Freeny PC, van Sonnenberg E. Imaging and intervention in acute pancreatitis. *Radiology*. (1994) 193:297–306. doi: 10.1148/radiology.193.2.7972730

27. Mortele KJ, Wiesner W, Intriere L, Shankar S, Zou KH, Kalantari BN, et al. A modified CT severity index for evaluating acute pancreatitis: improved correlation with patient outcome. *Am. J. Radiol.* (2004) 183:1261–5. doi: 10.2214/ajr.183.5.1831261

28. Banday IA, Gattoo I, Khan AM, Javeed J, Gupta G, Latief M. Modified computed tomography severity index for evaluation of acute pancreatitis and its correlation with clinical outcome: a tertiary care hospital based observational study. *J Clin Diagn Res*. (2015) 9:TC01–TC5. doi: 10.7860/JCDR/2015/14824.6368

29. Atto AM, Berthoumieu Y, Bolon P. 2-D wavelet packet spectrum for texture analysis. *IEEE Trans Image Process*. (2013) 22:2495–500. doi: 10.1109/TIP.2013.2246524

30. Hegenbart S, Uhl A. A scale-and orientation-adaptive extension of local binary patterns for texture classification. *Pattern Recogn*. (2015) 48:2633–44. doi: 10.1016/j.patcog.2015.02.024

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 01 December 2023
DOI 10.3389/fmed.2023.1276672

Check for updates

OPEN ACCESS

EDITED BY
Jie-Zhi Cheng,
Shanghai United Imaging Intelligence, Co., Ltd., China

REVIEWED BY
Wei Wu,
Shanghai Jiao Tong University, China
Rui Li,
Shanghai United Imaging Intelligence, Co., Ltd., China
Minjeong Kim,
University of North Carolina at Greensboro, United States

*CORRESPONDENCE
Hao Wang
✉ wanghaohh@vip.126.com
Fu Shen
✉ ssff_53@163.com

†These authors have contributed equally to this work

‡These authors share first authorship

RECEIVED 12 August 2023
ACCEPTED 13 November 2023
PUBLISHED 01 December 2023

CITATION
Ma S, Lu H, Jing G, Li Z, Zhang Q, Ma X, Chen F, Shao C, Lu Y, Wang H and Shen F (2023) Deep learning-based clinical-radiomics nomogram for preoperative prediction of lymph node metastasis in patients with rectal cancer: a two-center study.
Front. Med. 10:1276672.
doi: 10.3389/fmed.2023.1276672

COPYRIGHT
© 2023 Ma, Lu, Jing, Li, Zhang, Ma, Chen, Shao, Lu, Wang and Shen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Deep learning-based clinical-radiomics nomogram for preoperative prediction of lymph node metastasis in patients with rectal cancer: a two-center study

Shiyu Ma[1‡], Haidi Lu[1‡], Guodong Jing[1‡], Zhihui Li[2], Qianwen Zhang[1], Xiaolu Ma[1], Fangying Chen[1], Chengwei Shao[1], Yong Lu[2], Hao Wang[3*†] and Fu Shen[1*†]

[1]Department of Radiology, Changhai Hospital, The Navy Medical University, Shanghai, China, [2]Department of Radiology, Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China, [3]Department of Colorectal Surgery, Changhai Hospital, The Navy Medical University, Shanghai, China

Background: Precise preoperative evaluation of lymph node metastasis (LNM) is crucial for ensuring effective treatment for rectal cancer (RC). This research aims to develop a clinical-radiomics nomogram based on deep learning techniques, preoperative magnetic resonance imaging (MRI) and clinical characteristics, enabling the accurate prediction of LNM in RC.

Materials and methods: Between January 2017 and May 2023, a total of 519 rectal cancer cases confirmed by pathological examination were retrospectively recruited from two tertiary hospitals. A total of 253 consecutive individuals were selected from Center I to create an automated MRI segmentation technique utilizing deep learning algorithms. The performance of the model was evaluated using the dice similarity coefficient (DSC), the 95th percentile Hausdorff distance (HD95), and the average surface distance (ASD). Subsequently, two external validation cohorts were established: one comprising 178 patients from center I (EVC1) and another consisting of 88 patients from center II (EVC2). The automatic segmentation provided radiomics features, which were then used to create a Radscore. A predictive nomogram integrating the Radscore and clinical parameters was constructed using multivariate logistic regression. Receiver operating characteristic (ROC) curve analysis and decision curve analysis (DCA) were employed to evaluate the discrimination capabilities of the Radscore, nomogram, and subjective evaluation model, respectively.

Results: The mean DSC, HD95 and ASD were 0.857±0.041, 2.186±0.956, and 0.562±0.194mm, respectively. The nomogram, which incorporates MR T-stage, CEA, CA19-9, and Radscore, exhibited a higher area under the ROC curve (AUC) compared to the Radscore and subjective evaluation in the training set (0.921 vs. 0.903 vs. 0.662). Similarly, in both external validation sets, the nomogram demonstrated a higher AUC than the Radscore and subjective evaluation (0.908 vs. 0.735 vs. 0.640, and 0.884 vs. 0.802 vs. 0.734).

Conclusion: The application of the deep learning method enables efficient automatic segmentation. The clinical-radiomics nomogram, utilizing preoperative MRI and automatic segmentation, proves to be an accurate method for assessing LNM in RC. This approach has the potential to enhance clinical decision-making and improve patient care.

Research registration unique identifying number (UIN): Research registry, identifier 9158, https://www.researchregistry.com/browse-the-registry#home/registrationdetails/648e813efffa4e0028022796/.

KEYWORDS

rectal cancer, radiomics, magnetic resonance imaging, lymph node metastasis, deep learning

Introduction

Rectal cancer (RC) is a prevalent tumor affecting the gastrointestinal system and poses a significant global burden (1). The presence of lymph node metastasis (LNM) in RC individuals, particularly in cases defined as locally advanced rectal cancer (LARC), is associated with a poor prognosis. In order to manage LARC, the customary clinical strategy comprises administering neoadjuvant chemoradiotherapy (nCRT) before conducting total mesorectal excision (TME) surgery (2). This approach proves effective in diminishing the likelihood of local recurrence or the spread of cancer to distant sites (3). Achieving precision treatment in RC relies on accurate preoperative assessment of LNM (4). Consequently, it becomes crucial to accurately detect of lymph node (LN) involvement before surgery (4–6).

High-resolution magnetic resonance imaging (MRI) holds significant importance in the initial assessment of RC conditions. Nonetheless, achieving a precise preoperative diagnosis of LN involvement remains challenging in clinical practice (5). Relying solely on size as the exclusive criterion provides only acceptable precision. For instance, just 94% of the impacted LN possess a dimension less than 5 mm (6). A large node could be a successful tool to examine dimensions, perimeter, and signal intensity in LN. However, morphological criteria did not enhance the precision of lymph node staging in cases of RC (7). This challenge is further complicated by the absence of agreement regarding the relevant standards for evaluating LN contribution (7–9). Therefore, it is imperative to establish advanced and highly sensitive diagnostic tools to enhance the accuracy of LNM diagnosis in patients with RC.

Recently, several studies have demonstrated that radiomics can assist researchers in tackling diverse clinical tasks. By extracting numerous quantitative features from medical images through high-throughput analysis, radiomics approaches have the potential to empower radiologists to enhance diagnostic accuracy, ultimately benefiting patients (10–16). Radiomics-based models have exhibited promising value in detecting LNM in digestive tumors (10–12, 17–20). However, most existing methodologies rely on manual volume measurements of the entire primary tumor, which can be highly laborious, time-consuming, and subject to operator variability (12, 19, 20).

To the best of our knowledge, there is a lack of clear exploration regarding a deep learning-based image segmentation and clinical-radiomics nomogram for detecting LNM in individuals with RC. Therefore, the objective of this research was to create and validate an MR-based clinical-radiomics nomogram model that utilizes deep learning-based image segmentation. The purpose was to enable preoperative assessment of LNM and assess its clinical applicability in the context of RC.

Abbreviations: RC, Rectal cancer; nCRT, Neoadjuvant chemoradiotherapy; LNM, Lymph node metastasis; MRI, Magnetic resonance imaging; T2WI, T2-weighted imaging; ICC, Intraclass correlation coefficient; VOI, Volume of interest; LASSO, Least absolute shrinkage and selection operator; ROC, Receiver operating characteristic; AUC, Area under the roc curve; DCA, Decision curve analysis.

Materials and methods

Participants

The trial followed the Declaration of Helsinki and had permission from the Ethics Committees of Changhai Hospital and Ruijin Hospital Luwan Branch. Written informed consent was waived as the retrospective design.

From January 2017 to January 2020, a total of 392 consecutive patients with RC diagnosed pathologically at Changhai Hospital (center I) were included in this retrospective trial. The inclusion criteria comprised the following: (1) histological diagnosis of rectal adenocarcinoma based on postoperative pathological examination; (2) presence of a single tumor focus; (3) baseline rectal magnetic resonance imaging (MRI) performed within 14 days prior to surgical resection. Exclusion criteria were as follows: (1) receipt of any local or systemic treatment prior to surgical resection ($n = 86$); (2) previous or concurrent diagnosis of cancers other than RC ($n = 8$); (3) poor image quality ($n = 11$); (4) synchronous distant metastasis ($n = 22$); (5) positive CRM ($n = 7$); (6) history of previous pelvic surgery ($n = 5$). Ultimately, a total of 253 cases were enrolled from center I. Additionally, another 178 patients from Changhai Hospital (temporal external validation center I, EVC1) and 88 patients from Ruijin Hospital Luwan Branch (spatial external validation center II, EVC2), who met the same exclusion criteria as external validation sets 1 and 2, were also included between February 2020 and May 2023 for external validation.

Clinicopathologic data

Patient information and clinicopathologic findings were retrospectively obtained from the clinicopathological databases. This included data such as sex, age, BMI, histological differentiation, carbohydrate antigen 19-9 (CA19-9), carcinoembryonic antigen (CEA), circumferential resection margin (CRM), and pathological T-stage and N-stage. The CEA level was considered negative if it was less than 5 ng/mL, while the CA19-9 level was considered negative if it was less than 37 U/mL. These measurements were recorded at the same time as the baseline MRI. During the surgical procedure, all LN within the mesorectum were obtained from the surgical samples, ensuring a minimum of 12 lymph nodes were extracted per subject. The patients were categorized into different groups based on the National Comprehensive Cancer Network (NCCN) and American Joint Committee on Cancer (AJCC) staging system (21). The N0 group consisted of patients without lymph node metastasis (LNM), while the N1–2 group included patients with LNM.

Image acquisition and analysis

Rectal MRI scans were conducted using either a 1.5 or 3.0 T MR systems (Siemens 1.5, 3.0, and GE 3.0 T) along with a phased array coil. Prior to the scan, a 20 mL glycerin enema was administered to perform intestinal cleansing. The standard imaging protocol included axial diffusion-weighted imaging (DWI) with a b-value of 0 and 1,000 s/mm^2, sagittal T2-weighted imaging (T2WI), axial T1-weighted imaging (T1WI), and gadolinium contrast-enhanced T1WI of the pelvis in sagittal, coronal, and axial planes. Additionally, oblique axial high-resolution T2WI (HR-T2WI) images, which were perpendicular to the long axis of the rectum and included the lesion, were obtained. Supplementary Table 1 provides detailed information on the parameters used for HR-T2WI, which were utilized for the radiomics models.

Subjective evaluation of RC using MR imaging was conducted by three trained radiologists, namely R1, R2, and R3, with 12, 9, and 6 years of expertise, correspondingly. These radiologists were unaware of the pathological data. The assessment encompassed the evaluation of the subsequent tumor attributes: (1) tumor height, described as the measurement from the lower border of the tumor to the anal verge on MRI; (2) MR-reported T stage; (3) MR-reported N stage, and LN metastasis was identified if any of the following criteria was met: LN short-axis diameter superior to 10 mm, internal necrosis, nonuniform signal, LN fusion, nonuniform enhancement, or ill-defined borders (22, 23); (4) involvement of the mesorectal fascia (MRF); (5) presence of extramural venous invasion (EMVI). Any discrepancies among the radiologists' evaluations were resolved through discussion until a consensus was reached by at least two of the experts. The interobserver correlation of subjective evaluation for LN metastasis between any two radiologists was assessed using the Kappa statistic. The intraclass correlation coefficient (ICC) was calculated to evaluate the consistency of subjective evaluation for LN metastasis among all three radiologists.

Deep learning-based image segmentation

Since MR scans were performed using different MR scanners, the acquired DICOM data (oblique axial HR-T2WI) underwent preprocessing in these two centers. We adopted the data preprocessing strategy through data fingerprint information, including resampling strategy, cropping area size, gray value distribution, etc. information, thus forming a so-called "configuration plan." The size of each raw image was first adjusted by cropping to a size of 384 × 384 × 64. Subsequently, all images were resampled to a target spacing of [0.36, 0.36, 0.36] mm to ensure a consistent target spacing. The preprocessed images were subsequently brought into ITK-SNAP software version 4.0.0[1] for manual layer-by-layer segmentation of the entire RC lesion. This segmentation process aimed to obtain the volume of interest (VOI) representing the most accurate boundary fitting the primary tumor's area for each case. These segmented images served as mask images (ground truth, GT) for the training of the segmentation neural network.

The initial cohort of 253 cases from center I was randomly split into a network training set (60%, $n = 152$) and a network test set (40%, $n = 101$) for the development and validation of an automated segmentation method using nnU-Net during Stage I of our research. nnU-Net is a self-configuring approach specifically designed for deep learning-based segmentation of biomedical images (24). The details of the segmentation neural network can be found in Supplementary Figure 1. To mitigate overfitting, we implemented data augmentation along with 5-fold cross-validation. Additionally, the dice similarity coefficient (DSC), the 95th percentile Hausdorff distance (HD95), and average surface distance (ASD) between the automatically segmented images and the GT images were also reported in Supplementary Figure 2.

Then, the tested cases for automatic segmentation ($n = 101$) were also employed as a subsequent training set for the model to facilitate LNM classification in Stage II, thus avoiding excessively time-consuming processes. As for the segmentation task in Stage II, we also learned from the "configuration plan" and selected a parameter setting with a centered distribution. The automatic segmentation process was repeated with a one-week time interval to assess feature consistency. Finally, the Artificial Intelligence Kit software (GE Healthcare) was utilized to extract features from all automatically delineated VOIs derived from the model training set (n = 101), EVC1 ($n = 178$), and EVC2 ($n = 88$).

Radiomics feature extraction and reduction

Based on the automatically delineated VOIs, four categories of features were identified. These included: (1) first-order features, which describe the voxel intensity distribution on MR images, (2) shape features, which capture the 3D properties of the VOIs, (3) texture features, which quantify the dissimilarities in heterogeneity within the region using techniques such as size zone, run length, gray-level co-occurrence, and neighborhood gray-tone difference matrices, and (4) higher-order features, which are derived from transformed first-order data and texture features. This category includes square, square root, logarithm, exponential, gradient, local binary pattern (LBP), and wavelet transformations.

The intraclass correlation coefficient (ICC) was calculated to evaluate the robustness of the features during model training. Only indexes with an ICC value above 0.8 were considered for further analysis. To identify the most relevant features associated with LNM, the Select K Best method and the least absolute shrinkage and selection operator (LASSO) algorithm were employed to develop a Radscore. The detailed process of feature selection can be found in Supplementary Figure 3.

Nomogram model building and validation

The predictive value of clinical features and the Radscore in detecting LNM was assessed through univariable logistic regression evaluation in the model training set. Factors with p lower than 0.05 were then used to develop a nomogram model through multiple factor logistic regression. Receiver operating characteristic (ROC) curve analysis was conducted to evaluate the performance of the Radscore, nomogram, and subjective evaluation model. External validation sets

1 www.itksnap.org

1 and 2 were used to validate the accuracy of the detection (25). The models were compared using the DeLong test, and the goodness-of-fit of the nomogram was determined employing the Hosmer-Lemeshow test and calibration curves. To assess the comprehensive benefits, decision curve analysis (DCA) was employed. The study's workflow is depicted in Figure 1.

Statistical analysis

Statistical analysis was conducted using SPSS software (v. 26.0, IBM) and R package (v. 3.5.1, http://www.Rproject.org). Categorical data were analyzed using the Pearson chi-square test or Fisher's exact test, whereas continuous data (mean ± standard deviation) were assessed using the Student's t-test or Mann–Whitney *U*-test. A significance level of <0.05 (two-sided) was used to determine statistical significance.

Results

Patient features

The three cohorts exhibited no significant variations in demographic characteristics (all $p > 0.05$), as indicated in Table 1. Based on the pathological reports, LNM was identified in 50 out of 253 cases (19.8%) in center I, compared to 36 out of 178 cases (20.2%) in EVC1 and 24 out of 88 cases (27.3%) in EVC2. The interobserver agreement for the subjective evaluation of MR N-stage across all cohorts is presented in Supplementary Table 2.

Automatic segmentation results

The developed deep learning-based automatic segmentation method demonstrates the capability to execute automated configuration for our datasets, effectively encompassing the entire lesion in HR-T2WI (Figure 2). The mean DSC, HD95, and ASD between the automatic segmentation and GT were 0.857 ± 0.041, 2.186 ± 0.956 mm, and 0.562 ± 0.194 mm, respectively (Supplementary Figure 2).

Model building and evaluation

In the model training set, five features were identified and utilized to develop a Radscore, as shown in Table 2 and Supplementary Figure 3. Univariable analysis demonstrated a significant association between LNM and the following factors: MR T-stage, MR N-stage, CEA, CA19-9, and Radscore (Table 3). Subsequently, a nomogram model was constructed using multivariable logistic regression analysis, considering the selected risk factors (MR T-stage, CEA, CA19-9, and Radscore, as indicated in Table 4). The probabilities were calculated using the formula: −4.97107 + 3.72165 * Radscore + 1.85358 * CEA + 2.16416 * CA199 + 2.18032 * MR T-stage, resulting in an AUC of 0.921 (Supplementary Table 3). The generated nomogram, presented in Figure 3, exhibited a higher AUC compared to the Radscore and subjective evaluation in both external validation sets (0.908 vs. 0.735 vs. 0.640, and 0.884 vs. 0.802 vs. 0.734). These statistically significant differences were confirmed by the DeLong test. Detailed ROC analyses can be found in Table 5 and Figure 4. Calibration curves for the nomogram in both validation datasets indicated no significant deviation (Hosmer-Lemeshow test, $p = 0.065$ and 0.610) from an ideal fit (Supplementary Figure 4). DCA demonstrated that utilizing the nomogram model to assess the probability of LNM offered a positive net benefit compared to the Radscore, subjective evaluation, and the all-or-none approach at a significant threshold probability (Figure 5).

Discussion

Here, we focused on the development and validation of a deep learning-based image segmentation method for accurate delineation

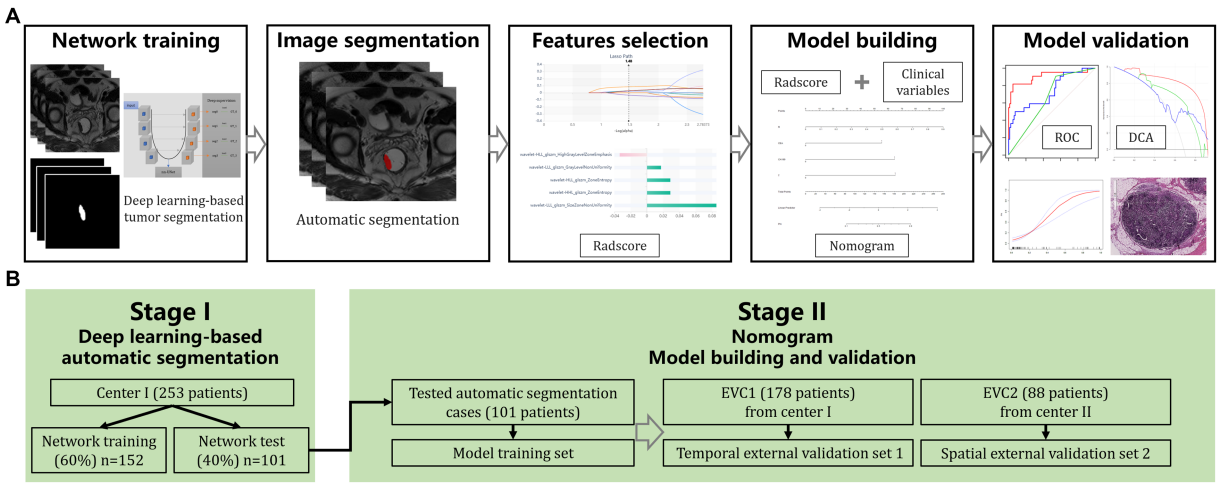


FIGURE 1
Study flowchart **(A)** and modeling methods **(B)**. Center 1, Changhai Hospital; EVC1, temporal external validation center I, Changhai Hospital; EVC2, spatial external validation center II, Ruijin Hospital Luwan Branch.

TABLE 1 Patient demographics.

Variables		Center I	EVC1	EVC2	*p*-value
		n = 253	*n* = 178	*n* = 88	
Gender					0.849
	Male	173 (68.4%)	124 (69.7%)	63 (71.6%)	
	Female	80 (31.6%)	54 (30.3%)	25 (28.4%)	
Age (years)		58.420 ± 12.112	56.750 ± 11.357	57.830 ± 10.254	0.442
BMI (kg/m^2)		23.434 ± 2.944	24.148 ± 2.968	23.334 ± 2.611	1.000
Tumor height (cm)*		4.751 ± 2.043	3.813 ± 1.864	4.773 ± 1.987	0.926
Pathological T-stage					0.187
	T1–2	117 (46.2%)	74 (41.6%)	47 (53.4%)	
	T3–4	136 (53.8%)	104 (58.4%)	41 (46.6%)	
Pathological N-stage					0.308
	N0	203 (80.2%)	142 (79.8%)	64 (72.7%)	
	N1–2	50 (19.8%)	36 (20.2%)	24 (27.3%)	
Differentiation					0.145
	High-moderate	200 (79.1%)	153 (86.0%)	69 (78.4%)	
	Poor	53 (20.9%)	25 (14.0%)	19 (21.6%)	
MR T-stage					0.366
	T1–2	148 (58.5%)	116 (65.2%)	55 (62.5%)	
	T3–4	105 (41.5%)	62 (34.8%)	33 (37.5%)	
MR N-stage					0.069
	N0	120 (47.4%)	67 (37.6%)	44 (50.0%)	
	N1–2	133 (52.6%)	111 (62.4%)	44 (50.0%)	
MRF					0.334
	Negative	202 (79.8%)	152 (85.4%)	72 (81.8%)	
	Positive	51 (20.2%)	26 (14.6%)	16 (18.2%)	
EMVI					0.696
	Negative	162 (64.0%)	121 (68.0%)	58 (65.9%)	
	Positive	91 (36.0%)	57 (32.0%)	30 (34.1%)	
CEA**					0.844
	Negative	165 (65.2%)	112 (62.9%)	55 (62.5%)	
	Positive	88 (34.8%)	66 (37.1%)	33 (37.5%)	
CA19-9**					0.139
	Negative	204 (80.6%)	130 (73.0%)	71 (80.7%)	
	Positive	49 (19.4%)	48 (27.0%)	17 (19.3%)	

BMI, body mass index; CEA, carcinoembryonic antigen; CA19-9, carbohydrate antigen 19-9; MRF, mesorectal fascia; EMVI, extramural venous invasion. *Tumor height was defined as the distance between the lower edge of the tumor and the anal verge by MRI. **Preoperative blood samples at the same time as baseline MRI. Center I, Changhai Hospital; EVC1, external validation from center I, Changhai Hospital; EVC2, external validation from center II, Ruijin Hospital Luwan Branch.

of rectal adenocarcinoma. Subsequently, a clinical-radiomics nomogram was constructed, demonstrating significantly improved performance compared to the Radscore and subjective evaluation when assessing lymph node metastasis (LNM) in patients with rectal cancer (RC). Radiologists and clinicians can utilize this intelligent, noninvasive, intuitive, and convenient approach to obtain personalized predictive information through straightforward calculations prior to surgery.

In patients with RC, preoperative detection of LNM plays a crucial role in tumor staging and treatment decision-making. It provides fundamental information for individualized treatment approaches, which primarily include surgical resection and nCRT, with variations based on the pathological stage of the lesion (2). Precise LN staging in RC is crucial to appropriately select individuals for preoperative procedure, ensuring avoidance of undertreatment and minimization of overtreatment. However, conventional magnetic resonance imaging (MRI) falls short in accurately detecting LN metastasis, exhibiting suboptimal sensitivity, accuracy, and specificity (7–9, 26). This suggests that subjective MRI standards for LNM detection are unreliable,

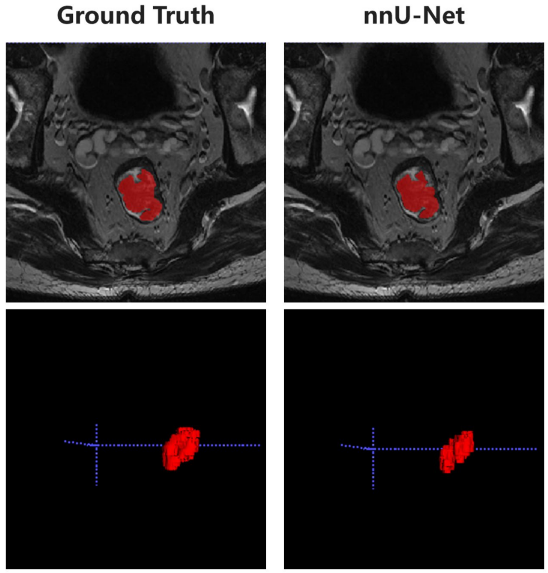


FIGURE 2
Representative diagram of automatic segmentation.

TABLE 2 Description of the selected radiomics features.

Radiomics feature	Radiomics class	Filter
Size zone non-uniformity	GLSZM	wavelet-LLL
Gray level non-uniformity	GLSZM	wavelet-LLL
Zone entropy	GLSZM	wavelet-HLL
High gray level zone emphasis	GLSZM	wavelet-HLL
Zone entropy	GLSZM	wavelet-HHL

GLSZM, Gray-level size zone matrix.

primarily due to the absence of a consensus on appropriate morphological criteria for accurate assessment of LN involvement. The data from our validation sets confirmed that the subjective evaluation demonstrates acceptable sensitivity for detecting LNM, ranging from 86.4% to 83.3%. However, the specificity is relatively low, ranging from 45.6 to 62.5%, which aligns with our clinical experience. Meanwhile, the accuracies of subjective MR N-stage were 54.5% to 71.6%. This negative influence becomes more pronounced when constructing a clinical-radiomics nomogram, leading to the exclusion of subjective MR N-stage from the final nomogram model in current research.

Radiomics represents a novel approach that utilizes routine imaging findings to conduct high-throughput quantitative evaluations. This quantitative method offers a noninvasive tool for the detailed analysis of the biological properties and variability of RC, surpassing the limitations of morphological visual representation. Currently, several studies (10–12) have showcased the viability of radiomics in predicting LNM in CRC. Our previous study (12) developed a radiomics model for primary lesions in RC using a random forest (RF) classifier to LNM. The RF demonstrated an AUC of 0.746, serving as a performance evaluation of diagnostics. However, the sensitivity and specificity of the model still fell below 80%. One potential explanation for this is the absence of clinicopathological risk factors in the model.

It is worth noting that we developed a clinical-radiomics nomogram model that combines MR T-stage, CEA, CA19-9, and Radscore. This model serves as an intuitive visualization tool with enhanced discriminatory ability for preoperative detection of LNM. It demonstrated favorable performance and superior diagnostic efficiency compared to subjective evaluation ($p < 0.05$). Furthermore, our findings suggest that the combination of Radscore and clinical factors outperformed the radiomics signature alone in predicting LNM in rectal adenocarcinoma. The addition of clinical factors resulted in an elevated AUC (0.802 to 0.884), along with significantly higher specificity (96.9%) and PLR of 21.333 in the external validation cohort. Consequently, a preoperative nomogram which can be trained effectively and explained easily was developed to assist radiologists and clinicians in assessment of LNM intuitively and rapidly.

Moreover, this study utilized radiomics features extracted from automatic segmentation based on deep learning. Specifically, we employed 60% of the center 1 dataset for training a neural network called nnU-Net, which enables automated image segmentation in HR-T2WI. Although nnU-net is a unified framework, the original architecture displays strong generalization characteristics requiring neither expert knowledge nor compute resources beyond standard network training in various medical image segmentation challenges (24). Compared to the conventional manual approach, the automated image segmentation offers convenience, eliminates the risk of perceptual errors, and is well-suited for processing substantial amounts of records. As a standardized and dataset-agnostic framework, nnU-Net was proposed as a robust and powerful tool for medical image segmentation (24). This streamlined and efficient procedure has the potential to alleviate the burden of the often laborious and inconsistent manual segmentation process. By leveraging artificial intelligence, this approach enhances the reliability of research and holds promise as a replacement for the time-consuming and non-reproducible manual segmentation method currently in use (27).

The inclusion of two distinct validation cohorts from external sources was another noteworthy aspect of this research. Consistent with the findings in the training set, the temporal and spatial external validation cohorts exhibited favorable discrimination, calibration, and improved clinical utility when utilizing the nomogram. This suggests that incorporating an external dataset can help mitigate the limitations of overfitting associated with a novel model. Consequently, the nomogram model holds the potential to enhance diagnostic confidence for radiologists and offer clinicians a more valuable and objective understanding of overall prognostic factors prior to clinical decision-making.

This investigation had several limitations that should be acknowledged. Firstly, the sample size was small, and the study design was retrospective, which may introduce selection bias and limit the general applicability of the findings. Therefore, larger-scale multicenter studies are required to overcome these limitations and validate the results more robustly. Additionally, the imaging segmentation was conducted automatically based on the primary tumor in RC. While most methodologies emphasize the use of the entire tumor volume, this study only extracted and analyzed radiomics features from the primary tumor itself, without exploring

TABLE 3 Univariate analysis in training set.

Variables	Pathological N stage			Univariate logistic regression	
	Total ($n = 101$)	N0 ($n = 79$)	N1–2 ($n = 22$)	OR (95% CI)	p-value
Gender					0.897
Male	77 (76.2%)	60 (76.0%)	17 (77.3%)	1.0 (reference)	
Female	24 (23.8%)	19 (24.0%)	5 (22.7%)	0.929 (0.302, 2.854)	
Age (years)	56.139 ± 11.335	56.671 ± 10.789	54.227 ± 13.212	0.982 (0.942, 1.022)	0.371
BMI (kg/m^2)	23.402 ± 2.947	23.545 ± 2.904	22.888 ± 3.110	0.925 (0.785, 1.090)	0.355
Tumor height (cm)	4.743 ± 2.057	4.785 ± 2.170	4.591 ± 1.623	0.953 (0.751, 1.210)	0.695
MR T-stage					**<0.001**
T1–2	66 (65.3%)	60 (76.0%)	6 (27.3%)	1.0 (reference)	
T3–4	35 (34.7%)	19 (24.0%)	16 (72.7%)	8.421 (2.886, 24.570)	
MR N-stage					**0.030**
N0	39 (38.6%)	36 (45.6%)	3 (13.6%)	1.0 (reference)	
N1–2	62 (61.4%)	43 (54.4%)	19 (86.4%)	1.887 (1.063, 3.351)	
MRF					0.116
Negative	81 (80.2%)	66 (83.5%)	15 (68.2%)	1.0 (reference)	
Positive	20 (19.8%)	13 (16.5%)	7 (31.8%)	2.369 (0.807, 6.952)	
EMVI					0.208
Negative	65 (64.4%)	54 (68.4%)	11 (50.0%)	1.0 (reference)	0.116
Positive	36 (35.6%)	25 (31.6%)	11 (50.0%)	2.160 (0.826, 5.646)	
CEA					**<0.001**
Negative	66 (65.3%)	61 (77.2%)	5 (22.7%)	1.0 (reference)	
Positive	35 (34.7%)	18 (22.8%)	17 (77.3%)	11.522 (3.732, 35.571)	
CA19-9					**<0.001**
Negative	82 (81.2%)	72 (91.1%)	10 (45.4%)	1.0 (reference)	
Positive	19 (18.8%)	7 (8.9%)	12 (54.6%)	12.343 (3.936, 38.709)	
Radscore	0.263 ± 0.203	0.219 ± 0.153	0.421 ± 0.275	84.761 (8.301, 865.471)	**<0.001**

LNM, lymph node metastasis; OR, odds ratio. The meaning of bold values provided in table was p-value < 0.05.

TABLE 4 Multivariate analysis in training set.

Variables		Training set ($n = 101$)	
		OR (95% CI)	p-value
MR T-stage			**0.005**
	T1–2	1.0 (reference)	
	T3–4	9.344 (1.973, 44.259)	
MR N-stage			0.050
	N0	1.0 (reference)	
	N1–2	2.513 (1.000, 6.317)	
CEA			**0.010**
	Negative	1.0 (reference)	
	Positive	7.810 (1.6252, 37.5291)	
CA19-9			**0.011**
	Negative	1.0 (reference)	
	Positive	9.418 (1.654, 53.641)	
Radscore		39.242 (1.540, 999.808)	**0.026**

LNM, lymph node metastasis; OR, odds ratio. The meaning of bold values provided in table was p-value < 0.05.

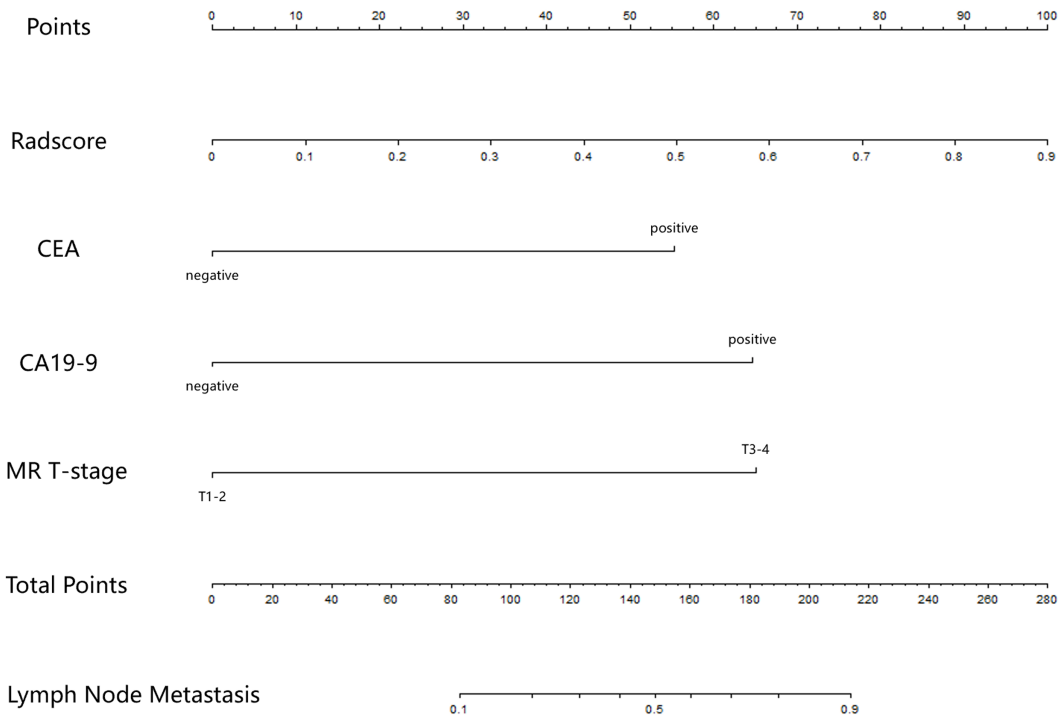


FIGURE 3
The nomogram. In the visual nomogram, first, a vertical line was drawn according to the values of the most influential factors to determine the corresponding numbers of points. The total points were the sum of the above points. Then, a vertical line was drawn according to the value of total points to determine the probability of LNM.

TABLE 5 ROC analysis in validation sets.

	External validation set 1			External validation set 2		
	Subjective evaluation	Radscore	Nomogram	Subjective evaluation	Radscore	Nomogram
AUC	0.640	0.735	0.908	0.734	0.802	0.884
95% CI	0.539 to 0.733	0.638 to 0.818	0.834 to 0.956	0.629 to 0.823	0.704 to 0.879	0.798 to 0.943
Sensitivity	86.4%	50.0%	81.8%	83.3%	66.7%	66.7%
Specificity	45.6%	92.4%	94.9%	62.5%	92.2%	96.9%
Accuracy	54.5%	83.2%	92.1%	71.6%	85.2%	88.6%
PLR	1.670	6.583	16.159	2.222	8.533	21.333
NLR	0.199	0.541	0.191	0.267	0.362	0.344
PPV	0.317	0.647	0.818	0.455	0.762	0.889
NPV	0.947	0.869	0.949	0.909	0.881	0.886
p-value*	<0.001	<0.001		0.018	0.035	

*Compared with nomogram by DeLong test. AUC, area under the curve; PLR, positive likelihood ratio; NLR, negative likelihood ratio; NPV, negative predictive value; PPV, positive predictive value.

the features of the LN. This limitation may lead to incomplete observation data and potentially impact the overall analysis. This point has garnered significant attention in both theoretical and application domains. However, deep learning approaches for the direct identification of LNM have not been developed and validated in this research. Multiple prior studies provide evidence that DL models can effectively predict tumor heterogeneity in rectal cancer, covering aspects like lymph node metastasis, distant metastasis, and patient survival (28–31). Nevertheless, deep learning investigations vary widely, and these models often lack interpretability. Although it is not easy for deep learning models to become explanatory and reasonable, which still puzzles many researchers. The application of artificial intelligence methods has the potential to guide personalized treatment plans, offering an emerging prognostic approach that warrants further investigation in the future (32–34).

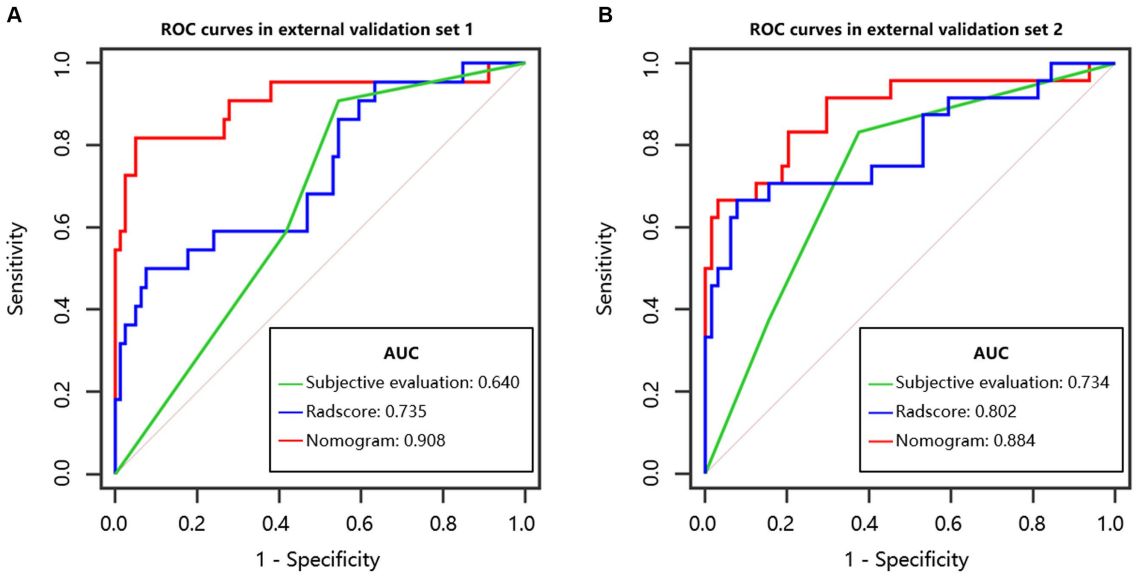


FIGURE 4
ROC curves. **(A)** External validation set 1. **(B)** External validation set 2.

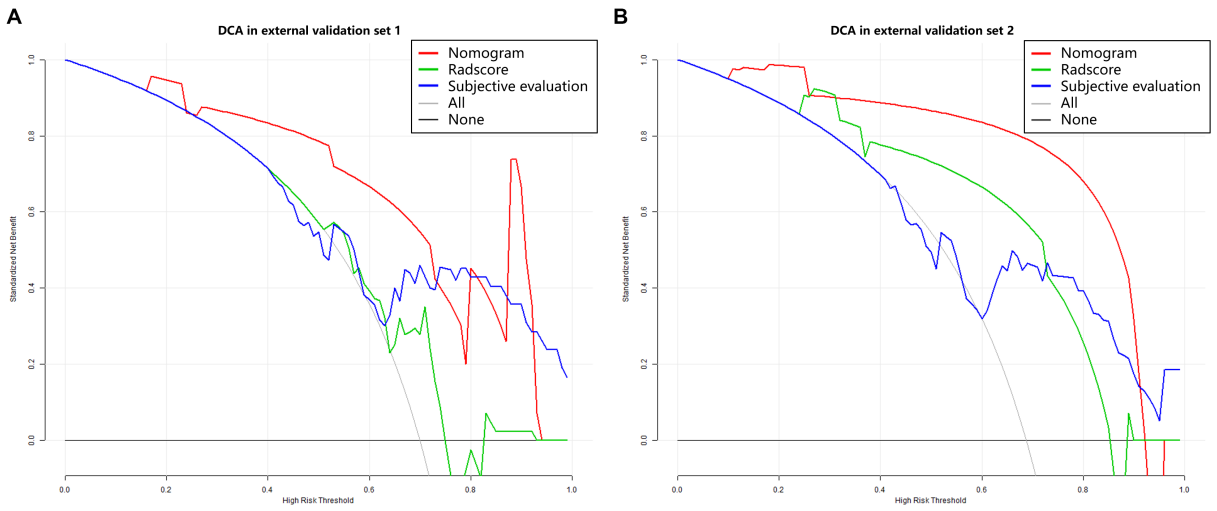


FIGURE 5
Decision curve analysis. **(A)** External validation set 1. **(B)** External validation set 2. The Y-axis represents the net benefit, calculated by adding true positives and subtracting false positives. The X-axis corresponds to the probability threshold (depicted as a solid line with a scale at the bottom). The light and dark gray lines indicate assumptions that all cases or no cases have lymph node metastasis (referred to as the "all" or "none" scheme), respectively. Red, green and blue curves showed that with a large probability range, utilizing the developed nomogram to predict the odds of LNM conferred a positive net benefit vs. the Radscore, subjective evaluation and the all-or-none scheme.

Conclusion

In summary, this study effectively developed and confirmed a clinical-radiomics nomogram by utilizing preoperative rectal MRI and automated segmentation. The nomogram incorporated both the Radscore and clinical risk factors, demonstrating its usefulness in predicting LNM. This innovative nomogram model demonstrated enhanced clinical utility compared to subjective evaluation and the Radscore alone. This noninvasive approach has the potential to intelligently enhance risk stratification in rectal cancer and can be readily applied in a clinical setting.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by the Ethics Committee of the Changhai Hospital, Naval Medical University. The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because written informed consent was waived as the retrospective design. Written informed consent was not obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article because written informed consent was waived as the retrospective design.

Author contributions

SM: Writing – original draft, Data curation, Formal Analysis, Investigation. HL: Writing – original draft, Methodology, Data curation. GJ: Writing – original draft, Data curation, Investigation. ZL: Data curation, Writing – original draft, Validation. QZ: Data curation, Writing – original draft. XM: Methodology, Data curation, Writing – original draft. FC: Data curation, Writing – original draft. CS: Funding acquisition, Supervision, Writing – review & editing. YL: Conceptualization, Funding acquisition, Methodology, Validation, Writing – review & editing. HW: Methodology, Conceptualization, Funding acquisition, Project administration, Resources, Writing – review & editing. FS: Conceptualization, Methodology, Writing – review & editing, Data curation, Funding acquisition, Project administration, Resources, Supervision.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by the Guhai Project of Changhai hospital (GH145-09), the Initial Foundation for Young Scientists of Changhai hospital (#2020QNB01), the "Yi Yuan Xin Xing" young medical talents funding project of Shanghai (project's number: N/A), and the 234-discipline Climbing Program of Changhai Hospital (2019YXK056).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2023.1276672/full#supplementary-material

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2022. *CA Cancer J Clin*. (2022) 72:7–33. doi: 10.3322/caac.21708

2. Glynne-Jones R, Wyrwicz L, Tiret E, Brown G, Rödel C, Cervantes A, et al. Rectal cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. (2018) 29:iv263. doi: 10.1093/annonc/mdy161

3. Benson AB, Venook AP, al-Hawary MM, Cederquist L, Chen YJ, Ciombor KK, et al. Rectal Cancer, version 2.2018, NCCN clinical practice guidelines in oncology. *J Natl Compr Cancer Netw*. (2018) 16:874–901. doi: 10.6004/jnccn.2018.0061

4. Dossa F, Chesney TR, Acuna SA, Baxter NN. A watch-and-wait approach for locally advanced rectal cancer after a clinical complete response following neoadjuvant chemoradiation: a systematic review and meta-analysis. *Lancet Gastroenterol Hepatol*. (2017) 2:501–13. doi: 10.1016/S2468-1253(17)30074-2

5. Nougaret S, Reinhold C, Mikhael HW, Rouanet P, Bibeau F, Brown G. The use of MR imaging in treatment planning for patients with rectal carcinoma: have you checked the "DISTANCE"? *Radiology*. (2013) 268:330–44. doi: 10.1148/radiol.13121361

6. Wang C, Zhou Z, Wang Z, Zheng Y, Zhao G, Yu Y, et al. Patterns of neoplastic foci and lymph node micrometastasis within the mesorectum. *Langenbeck's Arch Surg*. (2005) 390:312–8. doi: 10.1007/s00423-005-0562-7

7. Gröne J, Loch FN, Taupitz M, Schmidt C, Kreis ME. Accuracy of various lymph node staging criteria in rectal cancer with magnetic resonance imaging. *J Gastrointest Surg*. (2018) 22:146–53. doi: 10.1007/s11605-017-3568-x

8. Kwok H, Bissett IP, Hill GL. Preoperative staging of rectal cancer. *Int J Color Dis*. (2000) 15:9–20. doi: 10.1007/s003840050002

9. al-Sukhni E, Milot L, Fruitman M, Beyene J, Victor JC, Schmocker S, et al. Diagnostic accuracy of MRI for assessment of T category, lymph node metastases, and circumferential resection margin involvement in patients with rectal cancer: a systematic review and meta-analysis. *Ann Surg Oncol*. (2012) 19:2212–23. doi: 10.1245/s10434-011-2210-5

10. Li M, Zhang J, Dan Y, Yao Y, Dai W, Cai G, et al. A clinical-radiomics nomogram for the preoperative prediction of lymph node metastasis in colorectal cancer. *J Transl Med*. (2020) 18:46. doi: 10.1186/s12967-020-02215-0

11. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J Clin Oncol*. (2016) 34:2157–64. doi: 10.1200/JCO.2015.65.9128

12. Ma X, Shen F, Jia Y, Xia Y, Li Q, Lu J. MRI-based radiomics of rectal cancer: preoperative assessment of the pathological features. *BMC Med Imaging*. (2019) 19:86. doi: 10.1186/s12880-019-0392-7

13. Liu Z, Wang Y, Shen F, Zhang Z, Gong J, Fu C, et al. Radiomics based on readout-segmented echo-planar imaging (RS-EPI) diffusion-weighted imaging (DWI) for prognostic risk stratification of patients with rectal cancer: a two-center, machine learning study using the framework of predictive, preventive, and personalized medicine. *EPMA J*. (2022) 13:633–47. doi: 10.1007/s13167-022-00303-3

14. Jing G, Chen Y, Ma X, Li Z, Lu H, Xia Y, et al. Predicting mismatch-repair status in rectal cancer using multiparametric mri-based radiomics models: a preliminary study. *Biomed Res Int*. (2022) 2022:1–11. doi: 10.1155/2022/6623574

15. Li Z, Li S, Zang S, Ma X, Chen F, Xia Y, et al. Predicting treatment response to neoadjuvant Chemoradiotherapy in rectal mucinous adenocarcinoma using an MRI-based Radiomics nomogram. *Front Oncol*. (2021) 11:671636. doi: 10.3389/fonc.2021.671636

16. Liu M, Ma X, Shen F, Xia Y, Jia Y, Lu J. MRI-based radiomics nomogram to predict synchronous liver metastasis in primary rectal cancer patients. *Cancer Med*. (2020) 9:5155–63. doi: 10.1002/cam4.3185

17. Dong D, Fang MJ, Tang L, Shan XH, Gao JB, Giganti F, et al. Deep learning radiomic nomogram can predict the number of lymph node metastasis in locally advanced gastric cancer: an international multicenter study. *Ann Oncol*. (2020) 31:912–20. doi: 10.1016/j.annonc.2020.04.003

18. Wang Y, Liu W, Yu Y, Liu JJ, Xue HD, Qi YF, et al. CT radiomics nomogram for the preoperative prediction of lymph node metastasis in gastric cancer. *Eur Radiol*. (2020) 30:976–86. doi: 10.1007/s00330-019-06398-z

19. Chen Q, Zhang L, Liu S, You J, Chen L, Jin Z, et al. Radiomics in precision medicine for gastric cancer: opportunities and challenges. *Eur Radiol*. (2022) 32:5852–68. doi: 10.1007/s00330-022-08704-8

20. Qin Y, Deng Y, Jiang H, Hu N, Song B. Artificial intelligence in the imaging of gastric cancer: current applications and future direction. *Front Oncol*. (2021) 11:631686. doi: 10.3389/fonc.2021.631686

21. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, et al. The eighth edition AJCC Cancer staging manual: continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. *CA Cancer J Clin*. (2017) 67:93–9. doi: 10.3322/caac.21388

22. Amin MB, Edge SB, Greene FL, Byrd DR, Brookland RK, Washington MK, et al. *AJCC cancer staging manual. 8th* Edn. New York: Springer; (2017). p. 337–406.

23. MG H. *Atlas of lymph node anatomy*. New York: Springer (2013).

24. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. (2021) 18:203–11. doi: 10.1038/s41592-020-01008-z

25. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J*. (2014) 35:1925–31. doi: 10.1093/eurheartj/ehu207

26. Li XT, Sun YS, Tang L, Cao K, Zhang XY. Evaluating local lymph node metastasis with magnetic resonance imaging, endoluminal ultrasound and computed tomography in rectal cancer: a meta-analysis. *Color Dis*. (2015) 17:O129–35. doi: 10.1111/codi.12909

27. Jian J, Xiong F, Xia W, Zhang R, Gu J, Wu X, et al. Fully convolutional networks (FCNs)-based segmentation method for colorectal tumors on T2-weighted magnetic resonance images. *Australas Phys Eng Sci Med*. (2018) 41:393–401. doi: 10.1007/s13246-018-0636-9

28. Jiang X, Zhao H, Saldanha OL, Nebelung S, Kuhl C, Amygdalos I, et al. An MRI deep learning model predicts outcome in rectal Cancer. *Radiology*. (2023) 307:e222223. doi: 10.1148/radiol.222223

29. Wan L, Hu J, Chen S, Zhao R, Peng W, Liu Y, et al. Prediction of lymph node metastasis in stage T1-2 rectal cancers with MRI-based deep learning. *Eur Radiol*. (2023) 33:3638–46. doi: 10.1007/s00330-023-09450-1

30. Bedrikovetski S, Dudi-Venkata NN, Kroon HM, Seow W, Vather R, Carneiro G, et al. Artificial intelligence for pre-operative lymph node staging in colorectal cancer: a systematic review and meta-analysis. *BMC Cancer*. (2021) 21:1058. doi: 10.1186/s12885-021-08773-w

31. Liu X, Zhang D, Liu Z, Li Z, Xie P, Sun K, et al. Deep learning radiomics-based prediction of distant metastasis in patients with locally advanced rectal cancer after neoadjuvant chemoradiotherapy: a multicentre study. *EBioMedicine*. (2021) 69:103442. doi: 10.1016/j.ebiom.2021.103442

32. Jiang Y, Liang X, Wang W, Chen C, Yuan Q, Zhang X, et al. Noninvasive prediction of occult peritoneal metastasis in gastric cancer using deep learning. *JAMA Netw Open*. (2021) 4:e2032269. doi: 10.1001/jamanetworkopen.2020.32269

33. Hu C, Chen W, Li F, Zhang Y, Yu P, Yang L, et al. Deep learning radio-clinical signature for predicting neoadjuvant chemotherapy response and prognosis from pretreatment CT images of locally advanced gastric cancer patients. *Int J Surg*. (2023) 2023:1980–92. doi: 10.1097/JS9.0000000000000432

34. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. *Nat Rev Cancer*. (2018) 18:500–10. doi: 10.1038/s41568-018-0016-5

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 21 December 2023
DOI 10.3389/fmed.2023.1279622

Check for updates

OPEN ACCESS

EDITED BY
Jie-Zhi Cheng,
Shanghai United Imaging Intelligence, Co., Ltd., China

REVIEWED BY
Yongqin Zhang,
Northwest University, China
Xiantong Zhen,
Shanghai University of Technology, China
Yaping Wang,
Zhengzhou University, China
Zhiming Cui,
ShanghaiTech University, China

*CORRESPONDENCE
Jin Li
✉ lijin@tjh.tjmu.edu.cn
Jiaxuan Zhang
✉ jiaxuanzhang@126.com

†These authors have contributed equally to this work and share first authorship

RECEIVED 18 August 2023
ACCEPTED 04 December 2023
PUBLISHED 21 December 2023

CITATION
Feng Y, Feng Z, Wang L, Lv W, Liu Z, Min X, Li J and Zhang J (2023) Comparison and analysis of multiple machine learning models for discriminating benign and malignant testicular lesions based on magnetic resonance imaging radiomics.
Front. Med. 10:1279622.
doi: 10.3389/fmed.2023.1279622

COPYRIGHT
© 2023 Feng, Feng, Wang, Lv, Liu, Min, Li and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Comparison and analysis of multiple machine learning models for discriminating benign and malignant testicular lesions based on magnetic resonance imaging radiomics

Yanhui Feng[1†], Zhaoyan Feng[2†], Liang Wang[3], Wenzhi Lv[4], Zhiyong Liu[1], Xiangde Min[2], Jin Li[3*] and Jiaxuan Zhang[2*]

[1]School of Medicine and Health Management, Huazhong University of Science and Technology, Wuhan, China, [2]Department of Radiology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [3]Computer Center, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [4]Britton Chance Center and MoE Key Laboratory for Biomedical Photonics, Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan, China

Objective: Accurate identification of testicular tumors through better lesion characterization can optimize the radical surgical procedures. Here, we compared the performance of different machine learning approaches for discriminating benign testicular lesions from malignant ones, using a radiomics score derived from magnetic resonance imaging (MRI).

Methods: One hundred fifteen lesions from 108 patients who underwent MRI between February 2014 and July 2022 were enrolled in this study. Based on regions-of-interest, radiomics features extraction can be realized through PyRadiomics. For measuring feature reproducibility, we considered both intraclass and interclass correlation coefficients. We calculated the correlation between each feature and the predicted target, removing redundant features. In our radiomics-based analysis, we trained classifiers on 70% of the lesions and compared different models, including linear discrimination, gradient boosting, and decision trees. We applied each classification algorithm to the training set using different random seeds, repeating this process 10 times and recording performance. The highest-performing model was then tested on the remaining 30% of the lesions. We used widely accepted metrics, such as the area under the curve (AUC), to evaluate model performance.

Results: We acquired 1,781 radiomic features from the T2-weighted maps of each lesion. Subsequently, we constructed classification models using the top 10 most significant features. The 10 machine-learning algorithms we utilized were capable of diagnosing testicular lesions. Of these, the XGBoost classification emerged as the most superior, achieving the highest AUC value of 0.905 (95% confidence interval: 0.886–0.925) on the testing set and outstripping the other models that typically scored AUC values between 0.697–0.898.

Conclusion: Preoperative MRI radiomics offers potential for distinguishing between benign and malignant testicular lesions. An ensemble model like the boosting algorithm embodied by XGBoost may outperform other models.

KEYWORDS

boosting algorithm, machine learning, magnetic resonance imaging, radiomics, testicular neoplasms

1 Introduction

Testicular cancer, which is the most common solid tumor in males aged 15–34 years, is anticipated to result in approximately 470 fatalities and usher in an estimated 9,190 new cases in the United States by 2023 (1). Based on a 2020 statistical report, testicular cancer ranks among the top five causes of cancer-related fatalities in males aged 20–39 years in the United States (2). The standard treatment for malignant testicular lesions is inguinal orchiectomy (3, 4). For patients with benign testicular lesions, a more sensible treatment approach often involves conservative care, complemented by regular follow-ups and testicular preservation surgery. This is primarily because orchiectomy can adversely impact the patient's reproductive abilities and mental health, an effect particularly profound among young adult males (4, 5). Accurate identification of testicular tumors through better lesion characterization can help to reduce unnecessary radical surgical procedures (6).

Ultrasonography (US) is often used to confirm the presence of tumors in patients with testicular lesions (7). However, US has limited ability to distinguish benign and malignant testicular lesions effectively or to predict tumor size accurately (8). The advanced multi-directional and multi-sequence scanning capabilities of magnetic resonance imaging (MRI) can effectively depict testicular lesions and their relationship to surrounding tissues. Furthermore, it can infer possible tissue compositions, thereby providing valuable aid in both the diagnosis and differential diagnoses of these lesions (9). Therefore, MRI can afford us more adequate information and help to clarify some uncertainties or ambiguities in the results of the US, thereby reducing unnecessary surgical treatment (10).

Machine learning (ML), a multidisciplinary facet of artificial intelligence, endows computers with the capability to learn, enabling them to perform complex tasks similarly to humans. It is applied to both scientific research and industrial production to make accurate predictions using diverse data sources (11). Since it has achieved excellent prediction results in a wide range of applications, machine learning technology has attracted significant interest from medical researchers and clinicians (12).

In the past decade, the rapid development of medical image analysis has promoted the development of radiomics, which acquires massive quantitative information from image (13–15). It has a great application prospect in diagnosis, grading, staging, and prognosis of many tumors (16–18). Our previous studies established machine learning using radiomic signatures based on histogram analysis of apparent diffusion coefficient (ADC) (19). A previous study combined features and clinical indicators extracted from MRI to create predictive models to diagnose benign and malignant testicular lesions (20). However, to the best of our knowledge, no study to date has compared different modeling methods for the diagnosis of testicular diseases.

Therefore, we intend to utilize MRI imaging data for a comparative analysis of various machine learning algorithms deployed in differentiating between benign and malignant testicular diseases.

2 Materials and methods

2.1 Patients

A total of 394 patients, who underwent routine testicular MRI examinations, were recruited from February 2014 to July 2022. Of these, 286 patients were excluded based on the following criteria: (1) patients with no significant testicular lesions on MRI ($n = 185$); (2) patients who underwent biopsies, surgery, or treatment prior to MRI examination ($n = 77$); (3) patients with no testicular lesions confirmed by pathology ($n = 16$); and (4) patients who lacked MRI data or had MRI data of poor image quality ($n = 8$). Finally, 115 lesions were identified from 108 patients screened, including 44 benign and 71 malignant tumors. In this study, all lesions were diagnosed from testicular tissue sections after surgery or biopsy specimens. A flowchart of the case identification process is shown in Figure 1.

2.2 MRI protocol

We use the advanced type superconducting magnetic resonance system MAGNETOM Skyra, to scan patients with follow specification 3.0 T technology parameters, and set up an 18-element matrix and a 32-channel coil. The MRI protocol was listed in Supplementary Table S1. Due to the limited sample size, diffusion-weighted imaging and dynamic contrast-enhanced MRI were not included in this study.

2.3 Image segmentation

All transverse T2-weighted images (T2WI) were input into ITK-SNAP software (version 3.4.0) to realize the 3D segmentation of the target region manually. The lesions of all patients were manually segmented by radiologists with extensive experience in abdominal imaging. The two readers had 4 years and 5 years of experience, respectively. Segmentation was independently conducted to assess the reproducibility of inter-observer segmentation. Both two readers were blinded to the histopathological results. A radiologist with 4 years of experience (Reader 1) visualized the testicular lesions 1 month later to assess intra-observer segmentation reproducibility.

2.4 Radiomics feature extraction

The PyRadiomics package (version 2.1.2) was adopted to extract features from MRI. All MRI data were resampled with the same resolution ($1.0 \times 1.0 \times 1.0$ mm), and the built-in standardization function of PyRadiomics with a scale of 1 was used to normalize the intensity of MRI data. Nineteen filters were applied to each MRI scan of a lesion, as listed in Supplementary Table S2. All classes of features

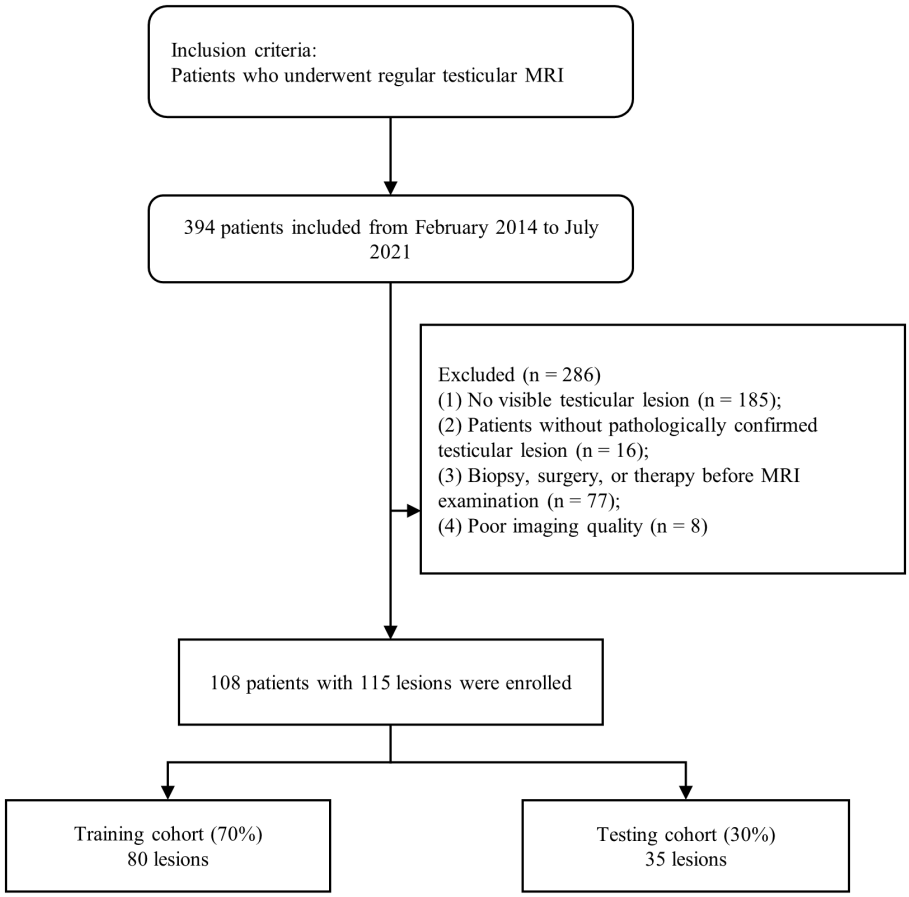


FIGURE 1
Inclusion and exclusion criteria.

(Supplementary Table S3), with the exception of shape, were computed for both the original and derived images.

2.5 Inter- and intra-correlation analysis of features

The robustness of the features was evaluated using ICCs. Randomly selected 34 lesions and the segmentation was operated by Reader 1 (4 years' experience in abdominal imaging). Secondary segmentation of these cases was performed by Reader 1 month later to evaluate the reproducibility within the observer. These images were also assessed by Reader 2 (5 years' experience in abdominal imaging) to assess consistency between observers. Features with ICC ≥0.8 were considered to be robust and were included in the follow-up study. Feature selection was performed with the maximum relevance and minimum redundancy (mRMR) approach (21), and the classification model based on radiomics was established. Figure 2 shows the workflow of radiomics signature development.

2.6 Model construction and evaluation

The included cases were divided into the training and testing set according to the ratio of 7:3. The following machine-learning models were considered: logistic regression (LR), quadratic discriminant analysis (QDA), *k*-nearest neighbor classifier (KNN), decision tree (DT), support vector machine (SVM), Gaussian naive Bayes (GaussianNB), random forest (RF), adaptive boosting (AdaBoost), gradient boosting (GB), and extreme gradient boosting (XGBoost). In the training set used to evaluate prediction performance and stability, different random seeds were set to train each classifier for 10 times. The average performance on the training set was recorded (Supplementary Table S4). The optimal model in the training cohort was subsequently tested in testing set.

When using the XGBoost algorithm, the following parameters are considered for adjustment: The learning_rate refers to the learning rate or step size, which controls the adjustment of model weights in each iteration. A small learning rate may require more training rounds, but it can potentially result in better prediction performance. The n_estimators refers to the number of trees, i.e., the number of sub-models or subtrees in the generated model. Insufficient trees may cause underfitting, while an excess of trees may cause overfitting. The max_depth indicates the maximum distance between the root node and the furthest leaf node in each tree. It affects the complexity of the model, as deeper trees result in a more complex model. Excessively large depths can lead to overfitting. The min_child_weight is used to determine the minimum weight sum of child sub-trees. If the weight of instances in a newly partitioned sub-tree is below this value, further

partitioning will not occur. This parameter helps avoid overfitting. The gamma parameter adjusts the degree of instance importance. A node will only split if the reduction in the loss function value after the split exceeds the specified gamma threshold. The colsample_bytree refers to the subsample ratio of columns, which is feature sampling used to construct each tree. The colsample_bytree is the subsample ratio considered during tree building. The subsample represents the subsample ratio of observed samples, which helps prevent overfitting. Typically, the value is between 0.5 and 1. In the experimental phase of this study, grid search was used to find appropriate parameters that ensured the model maintained optimal performance.

2.7 Statistical analysis

The Python (version 3.7) package was used for statistical analysis. For continuous variables, data are presented as means ± standard deviation. ICCs were computed to evaluate the agreement between features. Indicators covered the area under the receiver operating characteristic curve (AUC), average precision of the curve and five confusion matrix related indicators. These were computed by the bootstrap method (1,000 subsamples, 100 times). To evaluate the efficient of models and clinical practicability, calibration curve and decision curve analysis (DCA) analyses were employed. *p*-values less than 0.05 were considered to be statistically significant.

3 Results

3.1 Patients

After inclusion and exclusion and characteristic analysis, the study included108 patients with 115 testicular lesions (44 benign and 71 malignant). Patients had a wide age range (from 5 to 74 years), and the mean age was 36.25 years. Besides, the mean ages of the patients with benign and malignant lesion were 33.93 years and 46.40 years, respectively. Pathological analysis was performed in each case, and the statistical distribution is presented in Table 1. No significant difference in age was observed between the benign and malignant groups ($p = 0.217$).

3.2 Radiomics feature extraction and selection

T2WI contrast-enhanced sequence was used for radiomics features extraction. For each image space, 356 non-texture and 1,425 texture features were obtained from both the original and filtered images. ICCs were calculated for the inter-observer agreement, and 1,277 and 1,242 features were thought to be highly reproducible in terms of ICC values (ICC ≥0.8). A total of 1,182 features were considered to be robust and were included in the subsequent analysis. Finally, the mRMR method was used to eliminate redundant features and to select a subset of 10 features that were most relevant to the target to build the classification models. The radiomics features ranked by the mRMR method were mostly filter-based (7/10), which played an important role in the establishment of models.

3.3 Performance of models

On the training set, the prediction performance of 10 machine learning models was evaluated. All models performed well on the training set (AUC scores were greater than 0.8), and their performances are listed in Table 2. Among all the models, XGBoost exhibited the best diagnostic performance, which has a highest AUC (0.905, 95% CI, 0.886–0.925), sensitivity (0.895, 95% CI, 0.867–0.928), accuracy (0.886, 95% CI, 0.864–0.908), and NPV (0.875, 95% CI, 0.844–0.901) on the testing set. Other indicators of performance are showed in Table 3.

The prediction probabilities of each model for all lesions are shown in Figure 3A. The positive cases are mainly concentrated at the top, whereas the negative samples are mainly at the bottom, and the

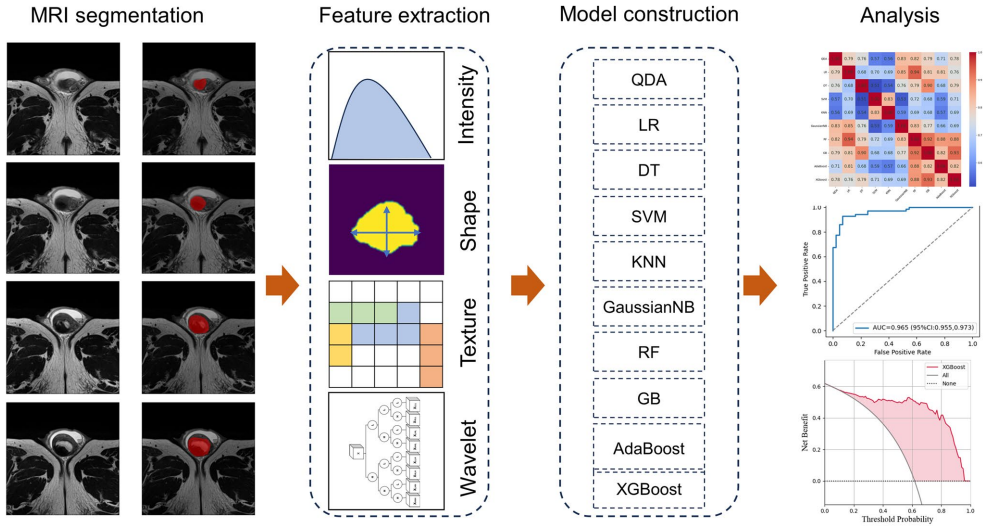


FIGURE 2
Illustration of the study design.

predicted results are more consistent with the reality. However, cases with a prediction probability of about 0.5 are relatively difficult to estimate, and the predicted values of each model are scattered. The correlation coefficients of the probabilities for each model are showed in Figure 3B. The coefficients of the RF, GB, AdaBoost and XGBoost models were 0.82 or higher (range: 0.82–0.93), indicating strong correlations. In addition, the LR, DT, GaussianNB, and RF models had high correlations, with coefficients >0.82, particularly RF and LR (coefficient = 0.94), while the correlation coefficient of SVM and KNN was 0.83.

In all cases, the AUC of XGBoost was 0.965 (95% CI, 0.955–0.973), as shown in Figure 4A. The Brier score of calibration curve is 0.091 (Figure 4B), which means the predicted probability and the actual malignant testicular lesions are approximated. In the decision curve, compared to assuming that all testicular tumors are malignant, the net profit of the prediction using XGBoost will be higher between the prediction probability of 10 to 95 percent (Figure 4C).

4 Discussion

In the present study, we used MRI as the object of feature extraction for predicting benign and malignant testicular lesions. Among all methods, the XGBoost classifier achieved best predictive performance, and the results revealed that machine learning models established based on radiomics features were able to differentiate benign from malignant testicular lesions.

Currently, MRIs serve as powerful tools that offer valuable insights into the characterization of various pathologies. The differentiation of testicular lesions, particularly between benign and malignant lesions, presents significant challenges for clinical experts. For radiologists, the visual differentiation of testicular lesions in MRI often requires a high level of expertise and experience. In terms of visual differential diagnosis, experts typically rely on certain key characteristics observed in MRI. The integration of machine learning models, particularly those

TABLE 1 Distribution of pathological findings in the included cases.

Benign		Malignant	
Mean age (years)	33.90 ± 17.00	**Mean age (years)**	46.40 ± 11.73

Pathology	Patient number	Lesion number	Pathology	Patient number	Lesion number
Epidermoid cyst	12	12	Seminoma	25	25
Sex cord-mesenchymal tumor	8	8	Embryonal carcinoma	9	9
Infection	13	15	Mixed germinoma	10	10
Cyst	1	1	Lymphoma	11	14
Infarction	1	1	Teratoma	6	6
Contusion	3	3	Testicular metastases	3	4
Prepubertal teratoma	1	1	Neuroendocrine carcinoma	1	1
Prepubertal teratoma and epidermoid cyst	1	1	Yolk sac tumor	1	1
Testicular adrenal rest tumor	1	2	Borderline serous tumor	1	1
Summation	41	44		67	71

TABLE 2 Performance of the models in the training cohort.

Model	AUC	Sensitivity	Specificity	Accuracy	Average precision	NPV	PPV
QDA	0.905	0.654	**1.000**	0.775	0.952	0.609	**1.000**
LR	0.826	0.654	0.893	0.738	0.908	0.581	0.919
DT	0.796	0.923	0.643	0.825	0.822	0.818	0.828
SVM	0.897	0.904	0.786	0.863	0.937	0.815	0.887
KNN	0.852	0.538	**1.000**	0.700	0.894	0.538	**1.000**
Gaussian NB	0.828	0.731	0.821	0.763	0.903	0.622	0.884
RF	0.911	0.865	0.821	0.850	0.954	0.767	0.900
GB	0.936	0.788	0.964	0.850	0.964	0.711	0.976
AdaBoost	0.923	0.712	**1.000**	0.813	0.954	0.651	**1.000**
XGBoost	**0.987**	**0.942**	0.964	**0.950**	**0.994**	**0.900**	0.980

The bold values indicate the highest values for specific indicators across different models. AUC, area under the curve; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value.

TABLE 3 Performance of XGBoost in the testing cohort.

Group	AUC (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	Accuracy (95% CI)	Average precision (95% CI)	NPV (95% CI)	PPV (95% CI)
Training set	0.987 (0.982, 0.991)	0.942 (0.928, 0.958)	0.964 (0.941, 0.981)	0.950 (0.938, 0.962)	0.994 (0.992, 0.996)	0.900 (0.877, 0.926)	0.980 (0.968, 0.990)
Testing set	0.905 (0.886, 0.925)	0.895 (0.867, 0.928)	0.875 (0.848, 0.904)	0.886 (0.864, 0.908)	0.934 (0.922, 0.946)	0.875 (0.844, 0.911)	0.895 (0.874, 0.917)

AUC, area under the curve; CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value.

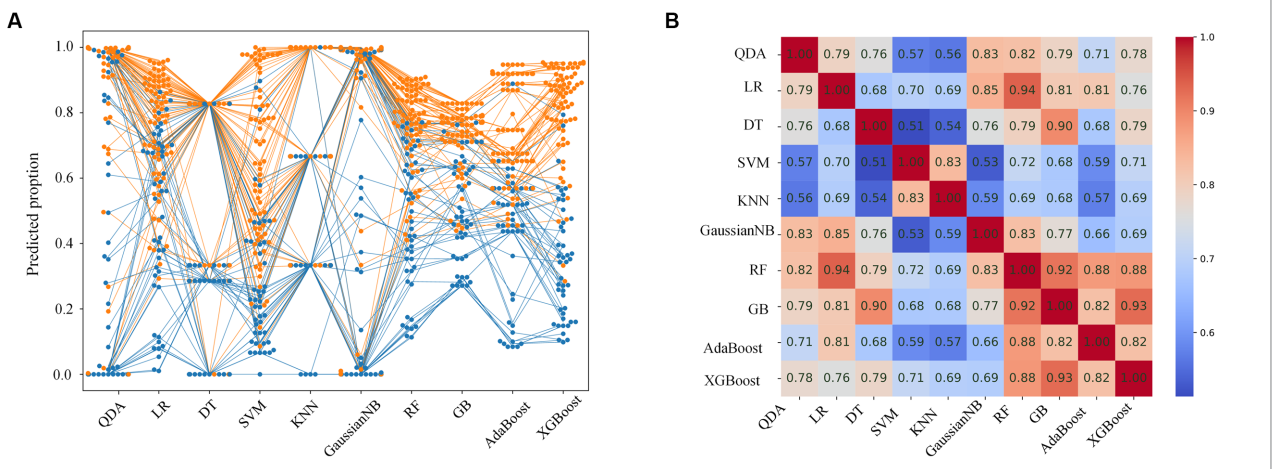


FIGURE 3
Prediction probabilities and correlation coefficients for each model. **(A)** Swarm plot of predicted probability of each model for all cases. Each dot represents a single sample. The orange and blue dots indicate malignant and benign lesions, respectively. **(B)** Correlation coefficients of the predicted probability for each model.

employing radiomic analysis, aims to overcome these challenges by quantitatively analyzing a wider range of features than what the human eye can discern. These advanced techniques offer promising avenues for improving diagnostic accuracy; they are intended to complement the expert judgment of clinical professionals. MRI-based radiomics models are emerging as an innovative approach to aid clinical decision-making. Several previous studies illustrate the efficacy and superior performance of these models. For instance, Zhang et al. (22) carried out a comparative analysis between traditional models and MRI-based radiomics models for diagnosing divergent carotid plaques. The outcomes indisputably denoted enhanced diagnostic performance by the radiomics model. Furthermore, the contribution of the AdaBoost classifier was substantial in differentiating low-grade gliomas from glioblastoma peritumoral regions relying on MRI radiomics (23). In this study, we observed a strong association and impressive correlation among the predicted probabilities of the boosting algorithms, such as gradient boosting (GB), AdaBoost, and extreme gradient boosting (XGBoost), across all examined cases. This signifies their potential for effectively distinguishing between benign and malignant lesions based on multidimensional radiomic data. Furthermore, we unveiled a noteworthy finding: the random forest (RF) model and these boosting algorithms yielded correlation coefficients equal to or higher than 0.88. This talent of the integrated algorithm to capture complex relationships between various features is well reflected in its superior performance. Interestingly, the logistic regression (LR) model was found to have a high correlation coefficient with the RF model. This emphasizes that classical models can powerfully differentiate between benign and malignant tumors. Hence, we should not undermine their potential while exploiting the power of advanced algorithms. Overall, the robust performance of the MRI-based radiomics models in our study, alongside findings from prior research, proposes a promising paradigm for future clinical applications. Particularly for the classification and diagnosis of diverse pathologies, these models could influence a shift from conventional diagnostic methods towards a more integrated and personalized approach.

In this study, the superior performance of XGBoost may be attributed to its gradient boosting framework, which inherently minimizes exponentially the discrepancy between predicted and true outcomes at each iteration. It's this boosting feature that makes it a robust and reliable algorithm for modeling complex patterns and predicting outcomes in healthcare data. Moreover, the significant findings of our study showcase the potential of employing machine learning models built on the basis of radiomic features in clinical radiology. Unlike conventional assessment methods, which rely heavily on subjective impressions or labor-intensive quantitative volumetric analysis, machine learning offers an objective and systematic approach to medical imaging evaluation. By leveraging robust algorithms, it allows for high-throughput detection and quantification of pertinent images' features, offering reproducible and unbiased results. At its core, our findings highlight a paradigm shift in

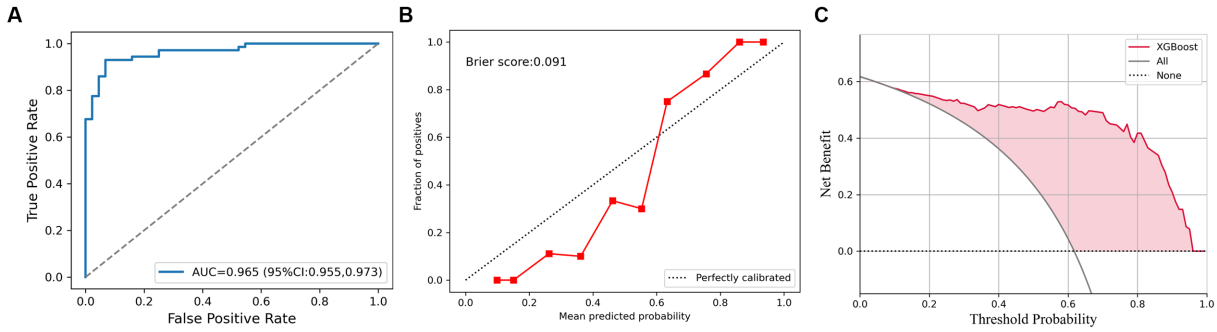


FIGURE 4
The XGBoost classifier for all cases. **(A)** Receiver operating characteristic (ROC) curve for XGBoost for discrimination of testicular lesions. **(B)** Calibration curve shows that the possibility of malignant testicular tumors is consistent with the true incidence. **(C)** Decision curve analysis (DCA) plot of the testicular lesions.

the evaluation of testicular lesions. The coupling of radiomics, machine learning, and, specifically, the use of the XGBoost algorithm underscores the emergence of a new era in clinical diagnostics. Interestingly, our study not only documents an improved method for predicting benign and malignant lesions but also sets a benchmark for future research to further optimize these prediction models, thereby enhancing our understanding and management of testicular diseases.

Correct noninvasive preoperative diagnosis is critically important for proper clinical decision-making and devising appropriate surgical plans, as it seeks to prevent unnecessary orchiectomy and enhance the quality of patient care. MRI has emerged as a promising imaging modality, exhibiting valuable radiomic features particularly relevant to testicular germ cell tumors (24, 25). As the body of literature in this area advances, a greater understanding of these radiomic characteristics can refine diagnostic accuracy and impact clinical practice. Zhang et al. (26) demonstrate the potential of T2-weighted imaging (T2WI)-based radiomics for differentiating seminomas from non-seminomas, yielding an impressive area under the curve (AUC) score of 0.979. In comparison to Zhang's study, our investigation benefits from a larger sample size and demonstrates substantial diagnostic performance by leveraging sophisticated machine learning algorithms. This improved methodology adds validity to our results and bolsters the case for the incorporation of MRI-based machine learning models in disease diagnosis. Similarly, He et al. (27) explore the application of MRI-based radiomic models for distinguishing benign and malignant prostate lesions. The study reports AUCs of 0.775 (T2WI) and 0.863 (apparent diffusion coefficient, ADC) for models based on single sequences. More notably, the integration of clinical characteristics enhances lesion discrimination capabilities, indicating the potential for combining radiomic data with patient profiling to further optimize diagnostic performance. The convergence of MRI and machine learning in these studies represents a paradigm shift in diagnostic approaches, signifying the growing importance of noninvasive and accurate methods in clinical practice. By transcending traditional, subjective assessments, machine-learning-assisted MRI has the potential to provide robust, reproducible, and data-driven insights with the added advantage of efficient, high-throughput analysis.

Notably, other imaging domains, such as ultrasound, may also contribute to distinguishing benign and malignant testicular tumors. Ultrasound imaging is a first-line, non-invasive diagnostic tool used in the evaluation of testicular tumors. It allows us to observe variations in size, shape, and location and to detect any discrete lesions, which can help guide clinical management. Typically, benign testicular tumors are well-defined, have homogeneous consistency, and may exhibit a halo of hypervascularity if there is inflammation or cystic changes. Various benign tumors, such as Leydig cell tumors, Sertoli cell tumors, and granulosa cell tumors, can be identified based on these characteristics. Conversely, malignant testicular tumors often present with a heterogeneous echo texture due to areas of necrosis, hemorrhage, or calcification. Growth patterns, vascularization, and the presence of metastatic tumors in the abdomen or pelvis seen on ultrasound can help identify malignant conditions such as seminomas and non-seminomatous germ cell tumors. Isidori et al. (28) investigated the accuracy of non-enhanced ultrasound combined with enhanced ultrasound in distinguishing benign and malignant lesions of $\leq$1.5 cm in the testes. Their results demonstrated that the combination of unenhanced and contrast-enhanced US achieved high accuracy in the diagnosis of small testicular malignancies (area under the ROC curve performance: 0.927; 95% confidence interval: 0.872, 0.981). This study suggests that the combination of enhanced and non-enhanced ultrasound effectively distinguishes benign and malignant testicular lesions of $\leq$1.5 cm, compensating for the inferior differentiating ability of non-enhanced ultrasound. However, it should be noted that ultrasound findings alone may not definitively distinguish benign from malignant tumors. Correlation with patient history, physical examination, and tumoral markers can further substantiate the diagnosis.

The current study emphasizes the paramount need for an accurate prognosis of testicular lesions in the pursuit of limiting false-negative results, as wrongful identification can pose a significant risk for patients. Orchiectomy stands as the conventional method of treatment for presumptive malignant testicular masses; however, the potential for error underscores the importance of discerning between benign and malignant testicular lesions. Misdiagnosis can result in unnecessary surgical intervention or postpone necessary treatment, thereby influencing patient outcomes and quality of life. Each patient presents a unique probability of predicting malignant testicular lesions, thereby underscoring individual-based therapeutic planning. In our quest to strike a risk-benefit balance, decision curve analysis (DCA) holds immense promise as a means to offer quantitative reference values that can inform the treatment strategy. This study

incorporated DCA as a key component in our evaluation methodology for the listed model's prediction results. By presenting a graphical representation of the model's applicability at varying threshold probabilities, DCA aids in the comprehension of potential benefits against potential harms in decision-making processes. Moreover, it augments the traditional measures of test performance by integrating patient preferences into the analysis. Our model's performance demonstrated significant consistency with the actual rate of testicular cancer across all cases, as revealed by the calibration plot. In essence, the calibration plot offers a visual demonstration of the model's predictive qualities in comparison to the ideal prediction. A curve that aligns closely with the 45-degree line infers perfect calibration, whereas deviation from the line implicates over- or underestimation. Thus, the proximity of the presented calibration curve to the real cancer rate supports the robustness of our model in predicting malignant testicular lesions. Moreover, the results of DCA computations signal that our model is generally applicable for a broad scope of threshold probabilities. It accentuates the rigor of the model predictions and manifests its potential adaptability across a spectrum of clinical settings.

This study focused on T2WI for the diagnosis of testicular diseases, as it is a routine and pivotal sequence in testicular MRI protocols. T2WI offers exceptional tissue contrast resolution, which is crucial for accurately delineating testicular lesions and differentiating between various disease types. This technique highlights differences in tissue composition and internal lesion structure, aiding in the identification of features like cystic components and solid areas. While diffusion-weighted imaging (DWI) and dynamic contrast-enhanced (DCE) sequences have diagnostic value, their limited use in clinical practice restricted their inclusion in our analysis.

Our study had several limitations. First and foremost, this study's reliance on data from a single center limits the scope of its findings. Given the wide spectrum of global health contexts and population dynamics, it should be noted that results derived from a single-center study may not be universally applicable. As a result, our findings should be interpreted with a certain level of caution when extended to other settings with differing population and health system characteristics. Future research could benefit from a multi-center trial, which would allow for a more diverse sampling of patient populations and healthcare settings. This would enhance the generalizability of our findings and further validate the insights we have gleaned from this investigation. Secondly, we must acknowledge the relatively small sample size of our study due to the low incidence of testicular cancer. While this small sample size enabled us to investigate this essential topic, it could nonetheless have affected the statistical power and practicability of our study. Considering this, we propose that future research on this topic strive for larger sample sizes to ensure a more robust analysis of data, gain a nuanced understanding of this cancer variety, and facilitate a more reliable estimate of the examination process's practicability. Lastly, we have recognized that the usage of the mRMR (minimum redundancy maximum relevance) algorithm could potentially underestimate the importance of features that individually bear limited impact on the targeted outcome but collectively can be highly effective. While the mRMR algorithm serves as a valuable tool for selecting relevant features in a dataset, it may not recognize the cumulative effect of feeble features. Future investigations should consider evaluating alternative methods alongside, or instead of, the mRMR algorithm. Employing different feature selection techniques could potentially give a more holistic view of factors affecting clinical outcomes, thereby enhancing the robustness and reliability of the results.

In summary, despite these limitations, our study provides essential insights into the fight against testicular cancer. Patient prognosis and treatment could be improved through further multi-center studies with larger sample sizes and different statistical methods. Nevertheless, it is vital that future research builds on this foundation and continues to explore these avenues to further advance our understanding and capabilities in combating this disease.

5 Conclusion

In conclusion, machine learning models based on MRI could accurately predict benign and malignant testicular lesions in the present study. Compared with a simple machine learning model, the ensemble model may achieve better performance, particularly when using the boosting algorithm represented by XGBoost. Information from a single sequence is limited, prompting the potential combination of different types of images or multiple sequences of a particular kind for machine learning training and prediction in the future. Additionally, integrating different machine learning could enhance predictive effectiveness.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Tongji Hospital, Tongji Medical College, Huazhong University of Science & Technology. The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because written informed consent was waived in view of the retrospective nature of the study and all the procedures being performed were part of the routine examination.

Author contributions

YF: Formal analysis, Methodology, Writing – original draft, Investigation, Software. ZF: Data curation, Funding acquisition, Investigation, Writing – original draft. LW: Writing – review & editing. WL: Formal analysis, Methodology, Writing – review & editing, Conceptualization. ZL: Writing – review & editing. XM: Writing – review & editing, Data curation, Writing – original draft. JL: Funding acquisition, Resources, Supervision, Writing – review & editing, Project administration. JZ: Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the National Natural Science Foundation of China (Nos. 81801668, 62072349, and 82001787).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2023.1279622/full#supplementary-material

References

1. Baird DC, Meyers GJ, Hu JS. Testicular cancer: diagnosis and treatment. *Am Fam Physician*. (2018) 97:261–8.

2. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. *CA Cancer J Clin*. (2023) 73:17–48. doi: 10.3322/caac.21763

3. Gilligan T, Lin DW, Aggarwal R, Chism D, Cost N, Derweesh IH, et al. Testicular Cancer, version 2.2020, NCCN clinical practice guidelines in oncology. *J Natl Compr Cancer Netw*. (2019) 17:1529–54. doi: 10.6004/jnccn.2019.0058

4. Stephenson A, Eggener SE, Bass EB, Chelnick DM, Daneshmand S, Feldman D, et al. Diagnosis and treatment of early stage testicular cancer: AUA guideline. *J Urol*. (2019) 202:272–81. doi: 10.1097/JU.0000000000000318

5. Gentile G, Rizzo M, Bianchi L, Falcone M, Dente D, Ciletti M, et al. Testis sparing surgery of small testicular masses: retrospective analysis of a multicenter cohort. *J Urol*. (2020) 203:760–6. doi: 10.1097/JU.0000000000000579

6. Tsili AC, Bertolotto M, Turgut AT, Dogra V, Freeman S, Rocher L, et al. MRI of the scrotum: recommendations of the ESUR scrotal and penile imaging working group. *Eur Radiol*. (2018) 28:31–43. doi: 10.1007/s00330-017-4944-3

7. Albers P, Albrecht W, Algaba F, Bokemeyer C, Cohn-Cedermark G, Fizazi K, et al. Guidelines on testicular Cancer: 2015 update. *Eur Urol*. (2015) 68:1054–68. doi: 10.1016/j.eururo.2015.07.044

8. Shtricker A, Silver D, Sorin E, Schreiber L, Katlowitz N, Tsivian A, et al. The value of testicular ultrasound in the prediction of the type and size of testicular tumors. *Int Braz J Urol*. (2015) 41:655–60. doi: 10.1590/S1677-5538.IBJU.2013.0077

9. Tsili AC, Argyropoulou MI, Giannakis D, Tsampalas S, Sofikitis N, Tsampoulas K. Diffusion-weighted MR imaging of normal and abnormal scrotum: preliminary results. *Asian J Androl*. (2012) 14:649–54. doi: 10.1038/aja.2011.172

10. Parenti GC, Feletti F, Carnevale A, Uccelli L, Giganti M. Imaging of the scrotum: beyond sonography. *Insights Imaging*. (2018) 9:137–48. doi: 10.1007/s13244-017-0592-z

11. Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. *Science*. (2015) 349:255–60. doi: 10.1126/science.aaa8415

12. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. *BMC Med Res Methodol*. (2019) 19:64. doi: 10.1186/s12874-019-0681-4

13. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036

14. Rizzo S, Botta F, Raimondi S, Origgi D, Fanciullo C, Morganti AG, et al. Radiomics: the facts and the challenges of image analysis. *Eur Radiol Exp*. (2018) 2:36. doi: 10.1186/s41747-018-0068-z

15. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures. *Radiology*. (2016) 278:563–77. doi: 10.1148/radiol.2015151169

16. Zhao J, Zhang W, Zhu YY, Zheng HY, Xu L, Zhang J, et al. Development and validation of noninvasive MRI-based signature for preoperative prediction of early recurrence in Perihilar cholangiocarcinoma. *J Magn Reson Imaging*. (2022) 55:787–802. doi: 10.1002/jmri.27846

17. Wang J, Wu CJ, Bao ML, Zhang J, Wang XN, Zhang YD. Machine learning-based analysis of MR radiomics can help to improve the diagnostic performance of PI-RADS v2 in clinically relevant prostate cancer. *Eur Radiol*. (2017) 27:4082–90. doi: 10.1007/s00330-017-4800-5

18. Gao M, Huang S, Pan X, Liao X, Yang R, Liu J. Machine learning-based radiomics predicting tumor grades and expression of multiple pathologic biomarkers in gliomas. *Front Oncol*. (2020) 10:1676. doi: 10.3389/fonc.2020.01676

19. Fan C, Sun K, Min X, Cai W, Lv W, Ma X, et al. Discriminating malignant from benign testicular masses using machine-learning based radiomics signature of appearance diffusion coefficient maps: comparing with conventional mean and minimum ADC values. *Eur J Radiol*. (2022) 148:110158. doi: 10.1016/j.ejrad.2022.110158

20. Xu M, Fang M, Zou J, Yang S, Yu D, Zhong L, et al. Using biparametric MRI radiomics signature to differentiate between benign and malignant prostate lesions. *Eur J Radiol*. (2019) 114:38–44. doi: 10.1016/j.ejrad.2019.02.032

21. Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans Pattern Anal Mach Intell*. (2005) 27:1226–38. doi: 10.1109/TPAMI.2005.159

22. Zhang R, Zhang Q, Ji A, Lv P, Zhang J, Fu C, et al. Identification of high-risk carotid plaque with MRI-based radiomics and machine learning. *Eur Radiol*. (2021) 31:3116–26. doi: 10.1007/s00330-020-07361-z

23. Malik N, Geraghty B, Dasgupta A, Maralani PJ, Sandhu M, Detsky J, et al. MRI radiomics to differentiate between low grade glioma and glioblastoma peritumoral region. *J Neuro-Oncol*. (2021) 155:181–91. doi: 10.1007/s11060-021-03866-9

24. Wakileh GA, Ruf C, Heidenreich A, Dieckmann KP, Lisson C, Prasad V, et al. Contemporary options and future perspectives: three examples highlighting the challenges in testicular cancer imaging. *World J Urol*. (2022) 40:307–15. doi: 10.1007/s00345-021-03856-6

25. Liu R, Lei Z, Li A, Jiang Y, Ji J. Differentiation of testicular seminoma and nonseminomatous germ cell tumor on magnetic resonance imaging. *Medicine*. (2019) 98:e17937. doi: 10.1097/MD.0000000000017937

26. Zhang P, Feng Z, Cai W, You H, Fan C, Lv W, et al. T2-weighted image-based Radiomics signature for discriminating between seminomas and nonseminoma. *Front Oncol*. (2019) 9:1330. doi: 10.3389/fonc.2019.01330

27. He D, Wang X, Fu C, Wei X, Bao J, Ji X, et al. MRI-based radiomics models to assess prostate cancer, extracapsular extension and positive surgical margins. *Cancer Imaging*. (2021) 21:46. doi: 10.1186/s40644-021-00414-6

28. Isidori AM, Pozza C, Gianfrilli D, Giannetta E, Lemma A, Pofi R, et al. Differential diagnosis of nonpalpable testicular lesions: qualitative and quantitative contrast-enhanced US of benign and malignant testicular tumors. *Radiology*. (2014) 273:606–18. doi: 10.1148/radiol.14132718

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 05 January 2024
DOI 10.3389/fmed.2023.1303501

Check for updates

OPEN ACCESS

EDITED BY
Minjeong Kim,
University of North Carolina at Greensboro, United States

REVIEWED BY
Cheng-Yu Wei,
Chang Bing Show Chwan Memorial Hospital, Taiwan
Weisheng Zhuang,
Henan Provincial People's Hospital, China
Andrea Quattrone,
University Magna Graecia, Italy

*CORRESPONDENCE
Chuanming Li
licm@cqu.edu.cn
Junbang Feng
junbangfeng@163.com
Chao Peng
523945695@qq.com

†These authors have contributed equally to this work

RECEIVED 28 September 2023
ACCEPTED 08 December 2023
PUBLISHED 05 January 2024

CITATION
Li C, Hui D, Wu F, Xia Y, Shi F, Yang M, Zhang J, Peng C, Feng J and Li C (2024) Automatic diagnosis of Parkinson's disease using artificial intelligence base on routine T1-weighted MRI.
Front. Med. 10:1303501.
doi: 10.3389/fmed.2023.1303501

COPYRIGHT
© 2024 Li, Hui, Wu, Xia, Shi, Yang, Zhang, Peng, Feng and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Automatic diagnosis of Parkinson's disease using artificial intelligence base on routine T1-weighted MRI

Chang Li[1†], Dongming Hui[2†], Faqi Wu[3], Yuwei Xia[4], Feng Shi[4], Mingguang Yang[1], Jinrui Zhang[1], Chao Peng[1*], Junbang Feng[1*] and Chuanming Li[1*]

[1]Bioengineering College of Chongqing University, Chongqing University Central Hospital (Chongqing Emergency Medical Center), Chongqing, China, [2]Department of Radiology, Chongqing Western Hospital, Chongqing, China, [3]Department of Medical Service, Yanzhuang Central Hospital of Gangcheng District, Chongqing, China, [4]Department of Research and Development, Shanghai United Imaging Intelligence, Co., Ltd. Shanghai, China

Background: Parkinson's disease (PD) is the second most common neurodegenerative disease. An objective diagnosis method is urgently needed in clinical practice. In this study, deep learning and radiomics techniques were studied to automatically diagnose PD from healthy controls (HCs).

Methods: 155 PD patients and 154 HCs were randomly divided into a training set (246 patients) and a testing set (63 patients). The brain subregions identification and segmentation were automatically performed with a VB-net, and radiomics features of billateral thalamus, caudatum, putamen and pallidum were extracted. Five independent machine learning classifiers [Support Vector Machine (SVM), Stochastic gradient descent (SGD), random forest (RF), quadratic discriminant analysis (QDA) and decision tree (DT)] were trained on the training set, and validated on the testing. Delong test was used to compare the performance of different models.

Results: Our VB-net could automatically identify and segment the brain into 109 regions. 2,264 radiomics features were automatically extracted from the billateral thalamus, caudatum, putamen or pallidum of each patient. After four step of features dimensionality reduction, Delong tests showed that the SVM model based on combined features had the best performance, with AUCs of 0.988 (95% CI: 0.979 ~ 0.998, specificity = 91.1%, sensitivity =100%, accuracy = 89.4% and precision = 88.2%) and 0.976 (95% CI: 0.942 ~ 1.000, specificity = 100%, sensitivity = 87.1%, accuracy = 93.5% and precision = 88.6%) in the training set and testing set, respectively. Decision curve analysis showed that the clinical benefit of the line graph model was high.

Conclusion: The SVM model based on combined features could be used to diagnose PD with high accuracy. Our fully automatic model could rapidly process the MRI data and distinguish PD and HCs in one minute. It greatly improved the diagnostic efficiency and has a great potential value in clinical practice to help the early diagnosis of PD.

KEYWORDS

PD, radiomics, artificial intelligence, T1-weighted, MRI

Introduction

Parkinson's disease (PD) is a complex and progressive neurodegenerative disorder characterized by an insidious onset, high incidence, and significant disability rate (1, 2). It poses a serious threat to the physical and mental health, as well as the overall quality of life, of middle-aged and elderly individuals (3). Its primary clinical manifestations include motor symptoms such as tremors, muscle rigidity, bradykinesia, postural instability; non-motor symptoms such as sleep disorders and olfactory dysfunction; autonomic nervous system dysfunction; cognitive impairment; and psychiatric disturbances (4). Currently, more than 6 million people worldwide suffer from PD, and the number is expected to further increase, bringing a huge burden to families and the society (5).

The rapid and accurate diagnosis of PD is of great significance for targeted treatment, prevention of disease progression, improvement of quality of life and overall prognosis (6). At present, the diagnosis of PD still relies on subjective clinical symptoms (7). Objective diagnosis methods are urgently needed in clinical practice. As a medical imaging technique, MRI has the advantages of non-invasive, non-radiation exposure, and high-resolution capabilities, making it widely used in the diagnosis and staging of neurological diseases (8, 9). Previous studies have identified alterations in both the structure and function of the brain in individuals diagnosed with PD (10, 11). Vogt et al. (12) discovered that the cingulate cortex played a crucial role in identifying new biomarkers for patients with early PD. Nyberg et al. (13) found a significant increasing in the volume of bilateral hippocampal and right nucleus accumbens of PD patients. Kassubek et al. discovered a significant increasing in the gray matter volume of the ventral medial thalamic nucleus on the contralateral side of the tremor limb in patients with PD. Furthermore, they observed a positive correlation between changes in thalamic gray matter volume and tremor amplitude (14). However, these studies mainly focused on macroscopic changes of the brain of PD patients, but overlook the small structural changes.

Radiomics, proposed by Lambin et al. (15), is a new field of computer-aided imaging diagnosis that assists in diagnosing and differentiating diseases by quantifying subtle information in medical images that is difficult to evaluate with the naked eye. Tupe-Waghmare et al. (16) had extracted radiomic features from the subcortical structure, cerebellum, brainstem, and used a random forest machine learning model to identify PD and HCs with an accuracy of 70%. Tafuri et al. (17) used Freesurfer software to extract radiomics features from the subcortical nucleus and used SVM model to distinguish PD and HCs patients, achieving an AUC (area under the receiver operating characteristic curve) of 0.77. However, these methods mostly showed low accuracy and were time-consuming, taking approximately 4 h to process a patient. These shortcomings limited its clinical application. In our previous study, we had developed a CNN-based artificial intelligence model for the automatic segmentation and measurement of the whole brain (109 brain regions), and the entire segmentation and reconstruction process took less than half a minute (18). In this study, we used this network to quickly segment and analyze the radiomics characteristics of the cerebral cortex and neuclei, and try to establish artificial intelligence models to help distinguish PD and HCs.

Materials and methods

Participants

All PD patients included in this study were sourced from the PD Progression Initiative (PPMI) database. The PD subjects in the PPMI were newly diagnosed patients who were not receiving any medication. The clinical diagnostic criteria for PD were based on the Movement Disorder Society guidelines. For the most up-to-date information, please visit https://www.ppmi-info.org/. The T1-weighted MR images were obtained using a 3 T Tesla scanner manufactured by Siemens. The imaging parameters were as follows: Acquisition Type = 3D; Flip Angle = 9.0 degree; field of view (FOV) = 256 × 256; matrix = 256 × 256; TR = 2300.0 ms; TE = 3.0 ms; Slice Thickness = 1.0 mm; interslice gap = 0 mm. A total of 155 PD patients and 154 HCs were enrolled in the study. The HC subjects and PD patients were matched for age and sex. All patients were randomly assigned to a training set and a testing set in a ratio of 8:2.

Brain subregions segmentation

The brain subregions segmentation module was implemented using a deep learning algorithm based on a 3D VB-NET network. The data preprocessing module performed a series of operations, including rotation, resampling, resizing, skull stripping, image non-uniform correction, histogram matching, and gray-scale normalization, on the MRI images used for training and testing. All images are standardized to the size of 256*256*256*1 mm^3 in the standard Cartesian LPI coordinate system, and the gray-scale range was within the interval (−1, 1). The network training module used an end-to-end deep convolutional neural network, taking each sample image and its corresponding brain substructure partition atlas as the training sample. The sample image was the network input, and the output label was the brain atlas correspondent to the sample image. The network parameters were adjusted according to the difference between the output brain division and the actual brain division, and the training continues until the network basically converges and the output label image was substantially consistent with the partition image corresponding to the sample. In the overall network training process, a coarse-to-fine cascading segmentation strategy was used, simplifying the complexity and difficulty of the brain segmentation problem by step decomposition, providing extra information to the lower level network by the upper level network to enhance the network's segmentation performance, and achieving fine segmentation of the large brain region, medium brain region, and brain substructures on a stage-by-stage basis. The model was constructed based on 1,800 subjects and evaluation showed an averaged 0.92 Dice overlap with ground truth on 295 subjects. Detailed segmentation process information was descripted in our previously published literature (18).

The entire brain was automatically divided into 109 subregions, which included 22 subregions in the temporal lobe, 20 subregions in the frontal lobe, 12 subregions in the parietal lobe, 8 subregions in the occipital lobe, 8 subregions in the cingulate gyrus, 2 subregions in the insula, 12 subcortical neuclei, white matter structures, ventricles, cerebellum structures and other structures (Supplementary material 1). The segmentation process took less than half a minute for each patient.

The flow chart of brain subregions segmentation was shown in Figure 1.

Features extraction and dimensionality reduction

The volume of 109 brain subregions of each patient was automatically extracted through the deep learning model. 2,264 radiomics features were automatically extracted from the billateral thalamus, caudatum, putamen or pallidum of each patient. There were 18 first-order statistics and 14 shape features, which reflect the shape and size of the region accurately. Texture features included 21 Gray Level Co-occurrence Matrix (GLCM) features, 16 Gray Level Run Length Matrix (GLRLM) features, 16 Gray Level Size Zone Matrix (GLSZM) features, 5 Neighbouring Gray Tone Difference Matrix (NGTDM) features, and 14 Gray Level Dependence Matrix (GLDM) features. The high level features were obtained through 24 filters (including Box Mean, additive Gaussian Noise, binomial blur, curvature flow, Box-Sigma, normalization, Laplace Sharpening, discrete Gaussian, mean, speck noise, recursive Gaussian, Shot Noise and LoG with sigma values of 0.5, 1, 1.5, and 2), as well as wavelet transformations (LLL, LLH, LHL, LHH, HLL, HLH, HHL, and HHH). All radiomic features were then normalized using z-score normalization, and the reproducibility of these features was assessed using a pipeline that adheres to the recommendations of the Image Biomarker Standardization Initiative.

The relief and least absolute shrinkage and selection operator (LASSO) method were used to select the most robust features. Hyperparameter for LASSO was evaluated using stratified 5-fold cross-validation-based grid search method on the training set. The parameters that provided the highest cross-validation AUC was selected.

Models building and evaluation

Based on the retained features, five independent machine learning classifiers, including Support Vector Machine (SVM), Stochastic gradient descent (SGD), random forest (RF), quadratic discriminant analysis (QDA) and decision tree (DT) algorithm were trained on the training set, and validated on the testing set. The performance of classifier models on the test subset was evaluated by the mean and 95% confidence intervals (CI) of the accuracy, sensitivity/recall, specificity, and precision based on a case probability cut-off value of 0.5, as well as the F-score metric and area under the receiver operating characteristic curve (AUC). The calibration curve was used to evaluate the calibration of the model, and DCA was used to evaluate the clinical applicability of the predictive model. The flow chart of our study was shown in Figure 2.

Statistical analysis

Statistical analyses were performed with R (version 4.0.4) and Python software. Mann Whitney U test or Student's t test was used for the continuous variables according to the test of normal distribution. The chi-square test was used to compare categorical variables. Statistical significance was indicated by a two-tailed p value <0.05. Delong test was used for the comparison of different models.

Results

A total of 155 PD patients were enrolled in the study, including 99 males and 56 females. Meanwhile, 154 HCs were included, including 99 males and 55 females. There was no significant difference in age

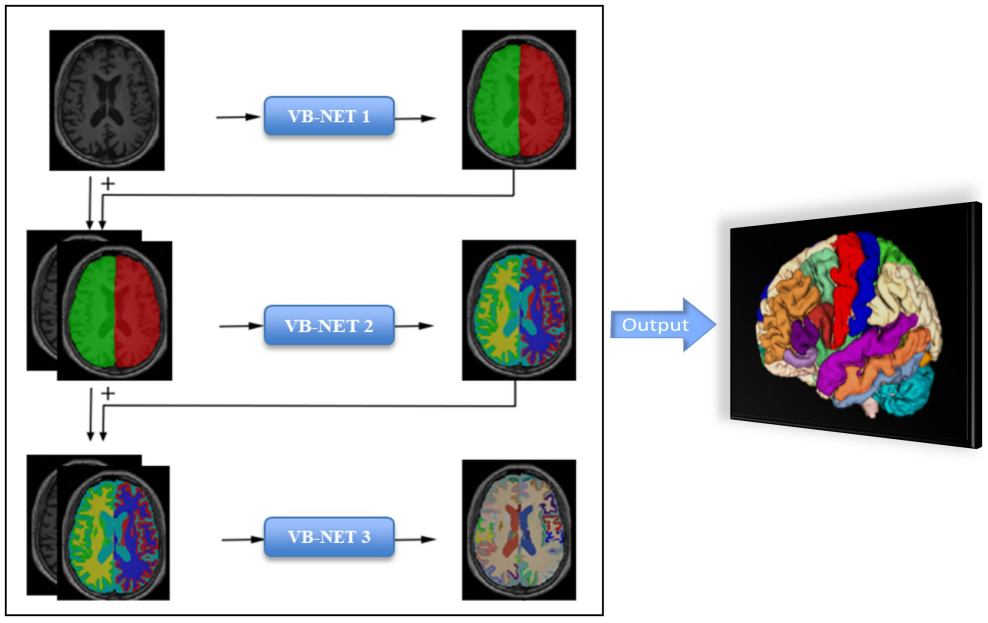


FIGURE 1
The flow chart of Cascading neural network framework.

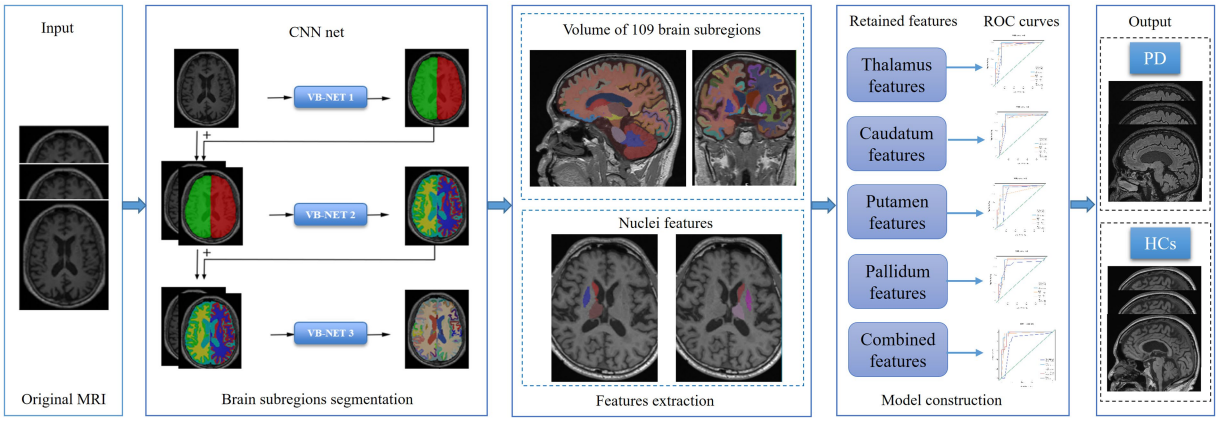


FIGURE 2
The flow chart of brain subregions segmentation, features extraction, features dimension reduction and models construction.

TABLE 1 Clinical and neuropsychological characteristics.

Variable	PD (n = 155)	HC (n = 154)	p
Age (years)	61.2 ± 9.4	70.5 ± 6.5	0.143
Gender (M/F)	99/56	99/55	0.110
Education (years)	15.3 ± 2.9	16.8 ± 2.2	0.109
UPDRS III_score	20.9 ± 9.1		/
MoCA_score	27.6 ± 2.0		/
Modified Hoehn and Yahr Scale	1.6 ± 0.5		/

Two-sample t test was used for continuous variable p-value and Chi-square test was used for discrete variable p-value. *p values less than 0.05 were considered statistically significant.

and gender between the two groups. The demographic and clinical features of PD and HC were shown in Table 1.

In the training set, the volume of 109 brain subregions of each patient was automatically obtained. 100 radiomics features were selected from 4,528 radiomics features of the bilateral thalamus using the relief method, and then 16 optimal features were obtained using the LASSO method. According to the same method, 12, 15, and 10 optimal features were obtained from the caudatum, putamen, and pallidum, respectively. Finally, feature selection of relief method and LASSO were performed again for all the above brain regions together, to obtain the 13 optimal combined features (Figure 3).

Twenty-five models were established based on algorithms of SVM, SGD, RF, QDA, DT and thalamus features, caudatum features, putamen features, pallidum feature and combined features. ROC curves of SVM, SGD, RF, QDA and DT models based on thalamus, caudatum, putamen, pallidum and combined features were shown in Table 2. It was found that the SVM model based on the combined features (5 features from pallidum, 4 features from putamen, 4 features from caudatum) showed the best performance, with AUCs of 0.988 (95% CI: 0.979 ~ 0.998, specificity = 91.1%, sensitivity =100%, accuracy = 89.4% and precision = 88.2%), 0.976 (95% CI: 0.942 ~ 1.000, specificity = 100%, sensitivity = 87.1%, accuracy = 93.5% and precision = 88.6%) in the training set and testing set, respectively (Figure 4). The calibration curve showed a good agreement between the actual and predicted probabilities of the samples (Figure 5A). Decision curve analysis showed that the clinical benefit of the line graph model was high (Figure 5B).

Discussion

It is well known that PD is a complex progressive neurodegenerative disease with high incidence rate (19). Developing an objective, accurate, and effective method to distinguish PD and HC has become an urgent issue in clinical practice. The changes in the cerebral cortex, deep nucleus, cerebellum and ventricles were found closely related to the occurrence of PD (20). Traditional imaging diagnostic methods were time-consuming, subjective and unable to detect these subtle changes in brain structure (21). Deep learning and radiomics were new fields of computer-aided imaging diagnosis (22), which overcomed the limitations of visual diagnosis and could quantify the subtle information in medical images, providing new hope for the rapid and accurate clinical diagnosis of PD.

In this study, a VB-net network evolved from U-Net network was developed to automatically identify and segment the cerebral cortex, cerebellum, ventricle, and 14 subcortical nuclei in 3D-T1-weighted imaging sequences. 2,264 radiomics features were automatically extracted from each region of the bilateral thalamus, caudatum, putamen, or pallidum. Five machine learning algorithms were applied to classify PD and HCs automatically and an independent verification group was set up to verify the performance of the model. We found among SVM, SGD, RF, QDA and DT classifiers, SVM classifier had the highest classification performance, with an AUC of 0.988 in the training set and 0.974 in the testing set. Our results demonstrated that the application of artificial intelligence technology to analyze raw T1-weighted MRI images could accurately differentiation PD patients from HCs. As a completely objective method, it did not rely on patients' personal history or doctors' clinical

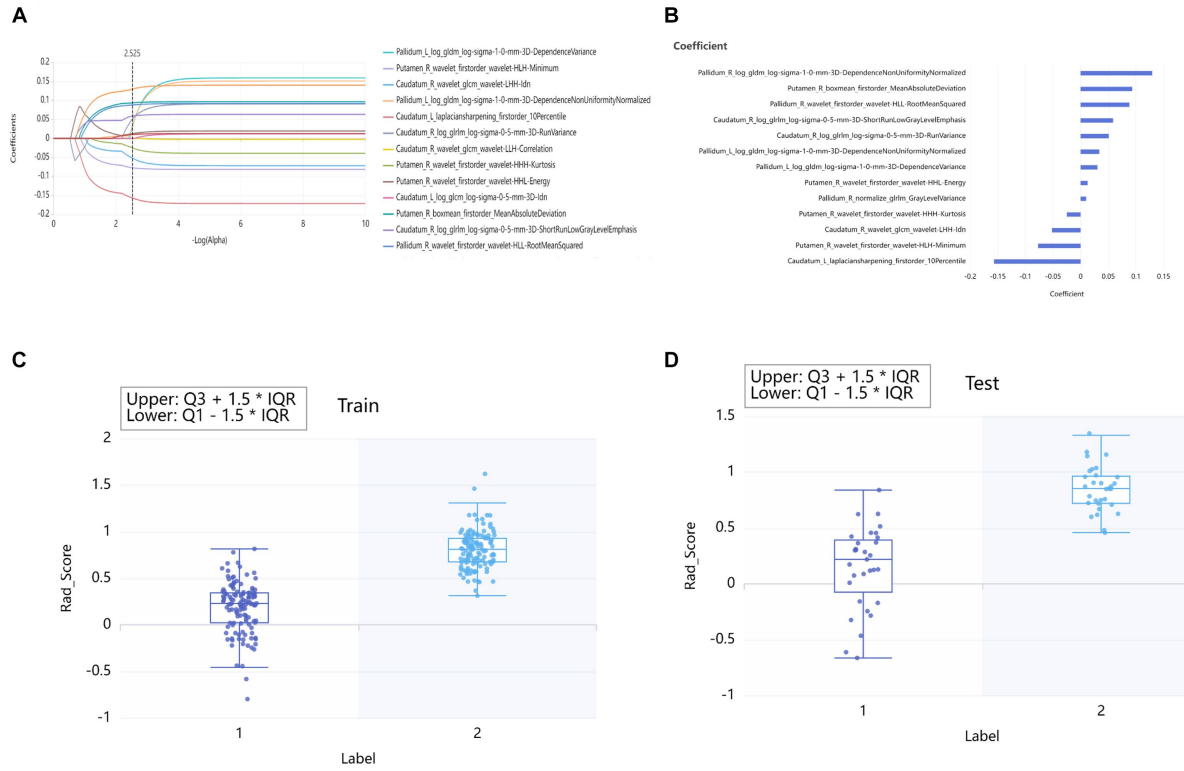


FIGURE 3
Lasso algorithm for features selection of PD patients and HCs **(A)**. The Lasso path displayed coefficient profiles for radiomic features across the entire range of possible values **(B)**. Rad_Score distribution in the training sets **(C)** and testing set **(D)**. "0" group represented HCs. "1" group represented PD.

TABLE 2 **Results of DT, SVM, SGD, RF, and QDA models on training set and testing set.**

	Models	AUC		Sensitivity		Specificity		Accuracy	
		Training set	Testing set	Training set	Testing set	Training set	Testing set	Training set	Testing set
Thalamus features	DT	0.975 (0.957–0.993)	0.892 (0.799–0.985)	0.992	0.968	0.902	0.871	0.947	0.919
	SVM	0.967 (0.945–0.989)	0.970 (0.932–1.000)	0.967	0.968	0.894	0.903	0.931	0.935
	SGD	0.890 (0.851–0.929)	0.903 (0.828–0.978)	0.846	0.903	0.927	0.903	0.886	0.903
	RF	0.958 (0.934–0.981)	0.934 (0.870–0.998)	0.927	0.968	0.862	0.871	0.894	0.919
	QDA	0.866 (0.817–0.916)	0.881 (0.787–0.976)	0.837	0.935	0.813	0.871	0.825	0.903
Caudatum features	DT	0.976 (0.961–0.992)	0.873 (0.772–0.973)	0.911	0.903	0.959	0.839	0.935	0.871
	SVM	0.947 (0.917–0.977)	0.956 (0.907–1.000)	0.919	0.935	0.886	0.871	0.902	0.903
	SGD	0.886 (0.846–0.926)	0.887 (0.807–0.967)	0.87	0.903	0.894	0.871	0.882	0.887
	RF	0.957 (0.935–0.978)	0.937 (0.874–0.999)	0.886	0.903	0.878	0.871	0.882	0.887
	QDA	0.876 (0.828–0.923)	0.921 (0.848–0.994)	0.878	0.968	0.813	0.871	0.846	0.919
Putamen features	DT	0.977 (0.959–0.995)	0.822 (0.699–0.994)	0.951	0.903	0.943	0.774	0.947	0.839
	SVM	0.947 (0.917–0.977)	0.945 (0.885–1.000)	0.935	0.968	0.878	0.871	0.907	0.919
	SGD	0.918 (0.883–0.952)	0.851 (0.762–0.940)	0.935	0.774	0.894	0.935	0.915	0.855
	RF	0.957 (0.935–0.980)	0.939 (0.879–0.998)	0.886	0.935	0.878	0.871	0.882	0.903
	QDA	0.867 (0.817–0.916)	0.912 (0.833–0.990)	0.878	0.935	0.821	0.871	0.85	0.903
Pallidum features	DT	0.958 (0.934–0.981)	0.841 (0.733–0.950)	0.951	0.839	0.886	0.839	0.919	0.839
	SVM	0.941 (0.909–0.972)	0.924 (0.854–0.994)	0.886	0.839	0.87	0.871	0.878	0.855
	SGD	0.849 (0.806–0.893)	0.900 (0.827–0.973)	0.943	1	0.756	0.806	0.85	0.903
	RF	0.947 (0.92–0.973)	0.914 (0.836–0.991)	0.902	0.903	0.854	0.871	0.878	0.887
	QDA	0.877 (0.831–0.923)	0.891 (0.801–0.980)	0.846	0.903	0.813	0.871	0.829	0.887
Combined features	DT	0.976 (0.957–0.996)	0.818 (0.702–0.935)	0.967	0.839	0.951	0.806	0.959	0.823
	SVM	0.988 (0.979–0.998)	0.976 (0.940–1.000)	0.967	0.935	0.943	0.903	0.955	0.919
	SGD	0.947 (0.919–0.975)	0.952 (0.897–1.000)	0.967	0.968	0.927	0.935	0.947	0.952
	RF	0.980 (0.966–0.994)	0.960 (0.917–1.000)	0.959	0.935	0.894	0.871	0.927	0.903
	QDA	0.963 (0.943–0.984)	0.969 (0.931–1.000)	0.927	0.968	0.878	0.774	0.902	0.871

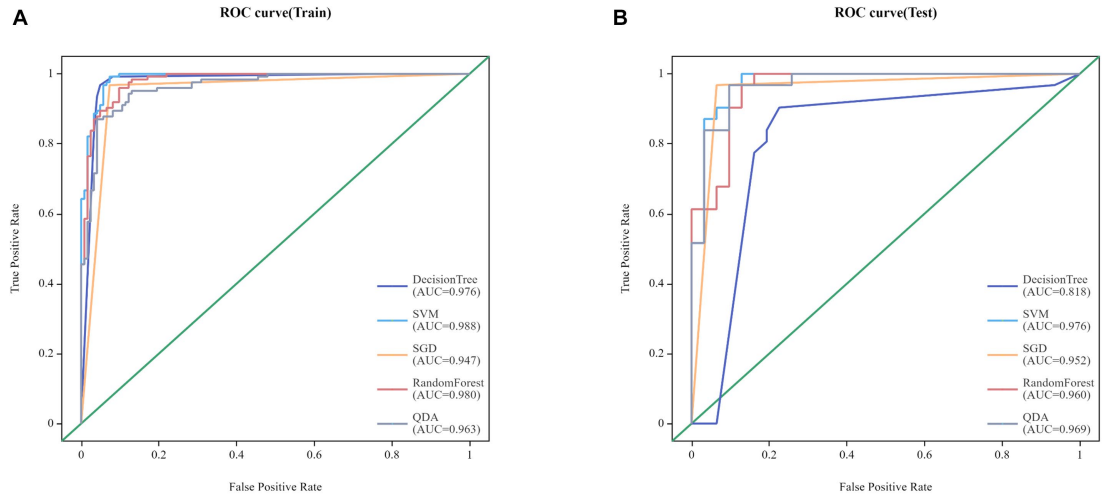


FIGURE 4
Receiver Operating Characteristic (ROC) curves for SVM, SGD, RF, QDA and DT models in training set **(A)** and testing set **(B)**.

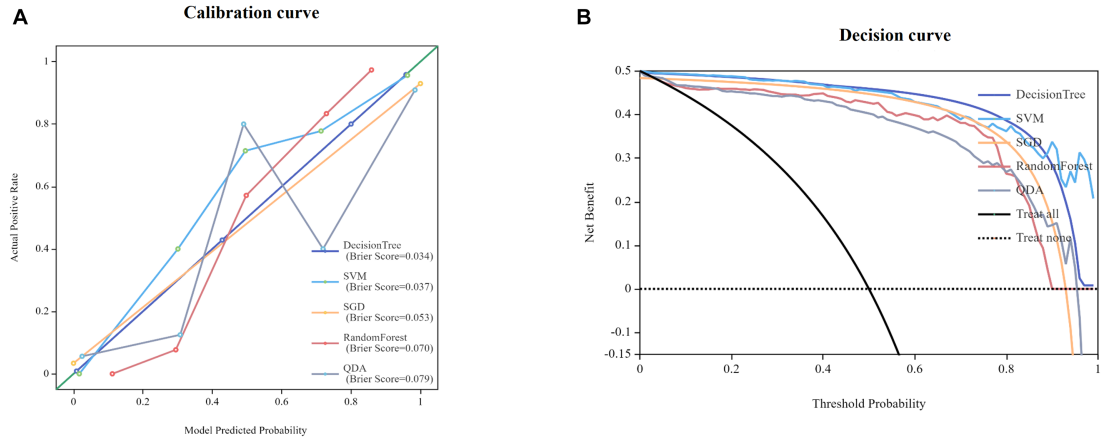


FIGURE 5
Calibration curves for nomogram goodness of fit **(A)**. Decision curves of different models **(B)**.

experience, making it more subjective. To the best of our knowledge, this study achieved the first fully automatic differentation of PD and HCs with high accuracy. Previously, Liu et al. (23) had used a radiomic model based on T2-weighted images of caudate nucleus and putamen to distinguish between PD and HCs, and the AUCs in the training set and testing set were 0.8767 (95% CI: 0.8066 ~ 0.9469) and 0.7143 (95% CI:0.5540 ~ 0.8746), respectively. Their results showed low accuracy and the manual brain segmentation took a long time, greatly reducing its work efficiency and clinical value. Adeli et al. (24) had utilized MRI and Single-Photon Emission Computed Tomography (SPECT) data from 538 subjects in the PPMI database to establish a classification framework, achieving a diagnostic accuracy of 97.5% for distinguishing between PD and HCs. However, the accuracy of their study mainly depended on the SPECT data, while the MRI data only made a small contribution to diagnostic ability. Additionally, the SPECT examination was expensive and involved radiopharmaceuticals that were not easily acceptable by patients (25). In our study, conventional MRI T1 weighted images were used, and all processes were fully automated, resulting in significant clinical significance. Our study was the first to use a deep learning model to automatically detect and segment brain subregions and extract valuable information. Our model could identify PD patients within 1 min and provided high accuracy. With this artificial intelligence system, doctors could quickly and accurately diagnose PD and perform early treatment timely, which could significantly improve prognosis. On the other hand, it could improve medical efficiency, shorten patient waiting time and reduce medical labor costs.

In the optimal model of this study, 13 radiomics features were retained, including 5 features from pallidum, 4 features from putamen and 4 features from caudatum. The 5 features from the pallidum included one Gray-level dependence matrix (GLDM) feature, one Gray-level run length matrix (GLRLM) feature, one first-order feature and two GLDM features. The pallidum receives afferent fibers from the subthalamic nucleus and is the primary region of the basal ganglia that emits efferent fibers (26). It could continuously release inhibitory neurotransmitter γ-aminobutyric acid (GABA), which may cause tremors (27). Hutchison et al. (28) had discovered the presence of neurons that exhibit tremor frequency activity ranging from 4 to 6 Hz in the pallidum, thus

supporting the significant role of the pallidum in generating resting tremors in patients with PD. The first-order feature of it reflected changes in the shape and volume of the pallidum (29). Previous studies had demonstrated significant alterations in pallidum volume between PD patients and HCs (30, 31). The loss and degeneration of pallidus neurons, as well as the proliferation of glial cells, could lead to texture changes in MRI images, which were reflected in GLDM and GLRLM. The four features of the putamen were all first-order characteristics, including Mean Absolute Deviation, Kurtosis, Energy, and Minimum. Previous studies had indicated that the putamen appeared to be the first region affected in PD (32). The four features of the caudatum included one first-order feature, one GLCM feature and two GLRLM features. The caudatum is a key substructure of the unique basal ganglia circuit associated with emotional and psychomotor fatigue. It plays a crucial role in the pathological and physiological regulation of PD (33). Previous studies had proved that dopaminergic dysfunction could lead to damage in the caudatum of PD patients, resulting in morphological and pathological changes (34, 35). The first-order features reflected the asymmetry and flatness of the morphology of the caudate nucleus, while the features of GLCM and GLRLM reflected the roughness and heterogeneity caused by pathological changes in the internal structure of the caudate nucleus.

Our study had several limitations. Firstly, this was a retrospective cross-sectional study that did not reflect the dynamic changes of the brain in PD. A prospective longitudinal follow-up study was needed in the future. Secondly, the sample size of PD patients was relatively small and all data were obtained through 3 T Siemens scanners; therefore, the generalization of the model needed to be further verified. In the future, images from multiple research centers were required for independent external validation, so that it could be adapted to a wider range of clinical scenarios. Finally, although we had confirmed that radiomics features have high accuracy in distinguishing PD from HCs, the pathological basis behind them still needed to be further revealed.

Conclusion

In this study, we established an AI model which could distinguish PD and HCs accurately. It was fully automated and could quickly process the routine MRI data within one minute to obtain accurate results. Our research results have greatly improved the diagnostic efficiency and had a great potential value in clinical practice to help the early diagnosis of PD.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://www.ppmi-info.org/.

Ethics statement

The studies involving humans were approved by The PPMI study was approved by an ethics standards committee on human experimentation at each institution. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

ChaL: Conceptualization, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing, Resources. DH: Data curation, Methodology, Software, Validation, Writing – original draft, Writing – review & editing. FW: Methodology, Writing – original draft, Resources. YX: Formal Analysis, Methodology, Software, Supervision, Validation, Writing-original draft. FS: Formal Analysis, Methodology, Software, Supervision, Validation, Writing-original draft. MY: Writing – original draft, Methodology, Software, Validation. JZ: Data curation, Writing – review & editing. CP: Model validation and analysis of Rad_ Score values. JF: Conceptualization, Funding acquisition, Writing – original draft, Writing – review & editing, Data curation, Investigation, Methodology, Resources, Software, Validation. ChuL: Conceptualization, Funding acquisition, Writing – original draft, Writing – review & editing, Investigation, Methodology, Resources.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Fundamental Research Funds for the Central Universities (project no. 2022CDJYGRH-004); Natural Science Foundation Project of Chongqing (cstc2020jcyj-msxm0030, cstb2023nscq-bhx0074); and Chongqing Medical Scientific Research Project (Joint project of Chongqing Health Commission and Science and Technology Bureau) (2022QNXM013).

Conflict of interest

Authors YX and FS were employed by company Shanghai United Imaging Intelligence, Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2023.1303501/full#supplementary-material

References

1. Tolosa E, Garrido A, Scholz SW, Poewe W. Challenges in the diagnosis of Parkinson's disease. *Lancet Neurol*. (2021) 20:385–97. doi: 10.1016/S1474-4422(21)00030-2

2. Albin RL. Parkinson's disease: background, diagnosis, and initial management. *Clin Geriatr Med*. (2006) 22:735–51. v. doi: 10.1016/j.cger.2006.06.003

3. Qi S, Yin P, Wang L, Qu M, Kan GL, Zhang H, et al. Prevalence of Parkinson's disease: a community-based study in China. *Mov Disord*. (2021) 36:2940–4. doi: 10.1002/mds.28762

4. Bloem BR, Okun MS, Klein C. Parkinson's disease. *Lancet (London, England)*. (2021) 397:2284–303. doi: 10.1016/S0140-6736(21)00218-X

5. Simon DK, Tanner CM, Brundin P. Parkinson disease epidemiology, pathology, genetics, and pathophysiology. *Clin Geriatr Med*. (2020) 36:1–12. doi: 10.1016/j.cger.2019.08.002

6. Mele B, Van S, Holroyd-Leduc J, Ismail Z, Pringsheim T, Goodarzi Z. Diagnosis, treatment and management of apathy in Parkinson's disease: a scoping review. *BMJ Open*. (2020) 10:e037632. doi: 10.1136/bmjopen-2020-037632

7. Waller S, Williams L, Morales-Briceño H, Fung VS. The initial diagnosis and management of Parkinson's disease. *Aust J Gen Pract*. (2021) 50:793–800. doi: 10.31128/AJGP-07-21-6087

8. Donzuso G, Agosta F, Canu E, Filippi M. MRI of motor and nonmotor therapy-induced complications in Parkinson's disease. *Mov Dis*. (2020) 35:724–40. doi: 10.1002/mds.28025

9. Zhou Y, Song Z, Han X, Li H, Tang X. Prediction of Alzheimer's disease progression based on magnetic resonance imaging. *ACS Chem Neurosci*. (2021) 12:4209–23. doi: 10.1021/acschemneuro.1c00472

10. Baggio HC, Junqué C. Functional MRI in Parkinson's disease cognitive impairment. *Int Rev Neurobiol*. (2019) 144:29–58. doi: 10.1016/bs.irn.2018.09.010

11. Xu X, Han Q, Lin J, Wang L, Wu F, Shang H. Grey matter abnormalities in Parkinson's disease: a voxel-wise meta-analysis. *Eur J Neurol*. (2020) 27:653–9. doi: 10.1111/ene.14132

12. Vogt BA. Cingulate cortex in Parkinson's disease. *Handb Clin Neurol*. (2019) 166:253–66. doi: 10.1016/B978-0-444-64196-0.00013-3

13. Nyberg EM, Tanabe J, Honce JM, Krmpotich T, Shelton E, Hedeman J, et al. Morphologic changes in the mesolimbic pathway in Parkinson's disease motor subtypes. *Parkinsonism Relat Disord*. (2015) 21:536–40. doi: 10.1016/j.parkreldis.2015.03.008

14. Kassubek J, Juengling FD, Hellwig B, Spreer J, Lücking CH. Thalamic gray matter changes in unilateral parkinsonian resting tremor: a voxel-based morphometric analysis of 3-dimensional magnetic resonance imaging. *Neurosci Lett*. (2002) 323:29–32. doi: 10.1016/s0304-3940(02)00111-8

15. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036

16. Tupe-Waghmare P, Rajan A, Prasad S, Saini J, Pal PK, Ingalhalikar M. Radiomics on routine T1-weighted MRI can delineate Parkinson's disease from multiple system atrophy and progressive supranuclear palsy. *Eur Radiol*. (2021) 31:8218–27. doi: 10.1007/s00330-021-07979-7

17. Tafuri B, Lombardi A, Nigro S, Urso D, Monaco A, Pantaleo E, et al. The impact of harmonization on radiomic features in Parkinson's disease and healthy controls: a multicenter study. *Front Neurosci*. (2022) 16:1012287. doi: 10.3389/fnins.2022.1012287

18. Liu S, Jie C, Zheng W, Cui J, Wang Z. Investigation of underlying association between whole brain regions and Alzheimer's disease: a research based on an artificial intelligence model. *Front Aging Neurosci*. (2022) 14:872530. doi: 10.3389/fnagi.2022.872530

19. Tysnes OB, Storstein A. Epidemiology of Parkinson's disease. *J Neural Transm*. (1996) 124:901–5. doi: 10.1007/s00702-017-1686-y

20. Li R, Zou T, Wang X, Wang H, Hu X, Xie F, et al. Basal ganglia atrophy-associated causal structural network degeneration in Parkinson's disease. *Hum Brain Mapp*. (2022) 43:1145–56. doi: 10.1002/hbm.25715

21. Cabreira V, Massano J. Parkinson's disease: clinical review and update. *Acta medica portuguesa*. (2019) 32:661–70. doi: 10.20344/amp.11978

22. Loh HW, Hong W, Ooi CP, Chakraborty S, Barua PD, Deo RC, et al. Application of deep learning models for automated identification of Parkinson's disease: a review (2011-2021). *Sensors (Basel)*. (2021) 21:7034. doi: 10.3390/s21217034

23. Liu P, Wang H, Zheng S, Zhang F, Zhang X. Parkinson's disease diagnosis using Neostriatum Radiomic features based on T2-weighted magnetic resonance imaging. *Front Neurol*. (2020) 11:248. doi: 10.3389/fneur.2020.00248

24. Adeli E, Wu G, Saghafi B, An L, Shi F, Shen D. Kernel-based joint feature selection and max-margin classification for early diagnosis of Parkinson's disease. *Sci Rep*. (2017) 7:41069. doi: 10.1038/srep41069

25. Bidesi NSR, Vang Andersen I, Windhorst AD, Shalgunov V, Herth MM. The role of neuroimaging in Parkinson's disease. *J Neurochem*. (2021) 159:660–89. doi: 10.1111/jnc.15516

26. Parent A, Hazrati LN. Functional anatomy of the basal ganglia. II. The place of subthalamic nucleus and external pallidum in basal ganglia circuitry. Brain research. *Brain Res Rev*. (1995) 20:128–54. doi: 10.1016/0165-0173(94)00008-d

27. Hua S, Reich SG, Zirh AT, Perry V, Dougherty PM, Lenz FA. The role of the thalamus and basal ganglia in parkinsonian tremor. *Mov Dis*. (1998) 13:40–2. doi: 10.1002/mds.870131307

28. Hutchison WD, Lozano AM, Tasker RR, Lang AE, Dostrovsky JO. Identification and characterization of neurons with tremor-frequency activity in human globus pallidus. *Exp Brain Res*. (1997) 113:557–63. doi: 10.1007/pl00005606

29. Li G, Zhai G, Zhao X, An H, Spincemaille P, Gillen KM, et al. 3D texture analyses within the substantia nigra of Parkinson's disease patients on quantitative susceptibility maps and R2(*) maps. *NeuroImage*. (2019) 188:465–72. doi: 10.1016/j.neuroimage.2018.12.041

30. Zhang S, Zi S, Xiong S, Peng H, Hu K, He H. Apathy following bilateral deep brain stimulation of subthalamic nucleus and Globus pallidus internus in Parkinson's disease: a Meta-analysis. *Parkinsons Dis*. (2022) 2022:4204564. doi: 10.1155/2022/4204564

31. Santos AN, Kherif F, Melie-Garcia L, Lutti A, Chiappini A, Rauschenbach L, et al. Parkinson's disease may disrupt overlapping subthalamic nucleus and pallidal motor networks. *NeuroImage Clin*. (2023) 38:103432. doi: 10.1016/j.nicl.2023.103432

32. Sigirli D, Ozdemir ST, Erer S, Sahin I, Ercan I, Ozpar R, et al. Statistical shape analysis of putamen in early-onset Parkinson's disease. *Clin Neurol Neurosurg*. (2021) 209:106936. doi: 10.1016/j.clineuro.2021.106936

33. Schwab NA, Tanner JJ, Nguyen PT, Schmalfuss IM, Bowers D, Okun M, et al. Proof of principle: transformation approach alters caudate nucleus volume and structure-function associations. *Brain Imaging Behav*. (2015) 9:744–53. doi: 10.1007/s11682-014-9332-x

34. Shen W, Zhai S, Surmeier DJ. Striatal synaptic adaptations in Parkinson's disease. *Neurobiol Dis*. (2022) 167:105686. doi: 10.1016/j.nbd.2022.105686

35. Sacheli MA, Neva JL, Lakhani B, Murray DK, Vafai N, Shahinfard E, et al. Exercise increases caudate dopamine release and ventral striatal activation in Parkinson's disease. *Mov Dis*. (2019) 34:1891–900. doi: 10.1002/mds.27865

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 11 January 2024
DOI 10.3389/fmed.2023.1337723

Check for updates

OPEN ACCESS

EDITED BY
Jie-Zhi Cheng,
Shanghai United Imaging Intelligence, Co., Ltd., China

REVIEWED BY
Xiangde Min,
Huazhong University of Science and Technology, China
Jingjing Gao,
University of Electronic Science and Technology of China, China

*CORRESPONDENCE
Jianhua Liu
✉ jian_hua@jlu.edu.cn

RECEIVED 13 November 2023
ACCEPTED 19 December 2023
PUBLISHED 11 January 2024

CITATION
Cui F, Khodrog OA, Liu W, Liu J and Yuan Q (2024) Clinical application of CT-based radiomics model in differentiation between laryngeal squamous cell carcinoma and squamous cell hyperplasia.
Front. Med. 10:1337723.
doi: 10.3389/fmed.2023.1337723

COPYRIGHT
© 2024 Cui, Khodrog, Liu, Liu and Yuan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Clinical application of CT-based radiomics model in differentiation between laryngeal squamous cell carcinoma and squamous cell hyperplasia

Fengzhi Cui, Osama A. Khodrog, Wei Liu, Jianhua Liu* and Qinghai Yuan

Department of Radiology, The Second Hospital of Jilin University, Changchun, China

Objective: To evaluate the clinical application of the CT-based radiomics prediction model for discriminating SCC and SCH.

Methods: A total of 254 clinical samples were selected from 291 patients with larynx-occupying lesions who underwent primary surgery. All lesions were validated via histopathological examination at The Second Hospital of Jilin University between June 2004 and December 2019. All patients were randomly allocated to the training (n = 177) and validation (n = 77) cohorts. After the acquisition of CT images, manual 3D tumor segmentation was performed using the CT images of the arterial, venous, and non-contrast phases via ITK-SNAP software. Subsequently, radiomics features were extracted using A.K. software. Based on the above features, three different diagnostic models (CTN, CTA+CTV, and CTN+CTA+CTV) were constructed to classify squamous cell carcinoma (SCC) and squamous cell hyperplasia (SCH). Additionally, receiver operating characteristic (ROC) and decision curve analysis (DCA) curves were measured to evaluate the diagnostic characteristics and clinical safety of the proposed three prognostic models.

Results: In the radiomic prediction Model 1 (CTN), the area under the curve (AUC), accuracy, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of the training cohorts in differentiating SCC and SCH were 0.883, 0.785, 0.645, 1.000, 1.000, and 0.648, while in the testing cohorts, these values were 0.852, 0.792, 0.66, 1.000, 1.000, and 0.652, respectively. In the radiomic prediction Model 2 (CTA+CTV), the AUC, accuracy, sensitivity, specificity, PPV, and NPV values of the training cohorts were 0.965, 0.91, 0.916, 0.9, 0.933, and 0.875, respectively, while in the testing cohorts, the corresponding values were 0.902, 0.805, 0.851, 0.733, 0.833, and 0.759, respectively. In the radiomic prediction Model 3(CTN+CTA+CTV), the AUC, accuracy, sensitivity, specificity, PPV, and NPV values of the training cohorts were 0.985, 0.944, 0.953, 0.929, 0.953, and 0.929, while in the testing cohorts, the corresponding values were 0.965, 0.857, 0.894, 0.8, 0.875, and 0.828, respectively.

Conclusion: The radiomic prediction Model 3, based on the arterial-venous-plain combined scan phase of CT, achieved promising diagnostic performance,

expected to be regarded as a preoperative imaging tool in classifying SCC and SCH to guide clinicians to develop individualized treatment programs.

KEYWORDS

laryngeal squamous cell carcinoma, laryngeal squamous cell hyperplasia, radiomics, CT imaging, differential diagnosis

1 Introduction

As the seventh most common cancer worldwide, squamous cell cancer of the head and neck (HNSCC) is a potential health concern worldwide with increased incidence and death rates. It mainly occurs in the lips, oral cavity, nasal cavity, paranasal sinuses, oropharynx, hypopharynx, larynx, and parotid gland (1). More than 500,000 individuals are diagnosed with HNSCC worldwide each year. Furthermore, it is estimated that the number of new cases is expected to exceed 1 million in 2020, with the number of deaths expected to exceed 500,000. Notably, Central Europe and Eastern Europe had the largest increased incidence among men over 55 years of age (2, 3). In one study, ∼75% of head and neck cancers overall are caused by tobacco smoking and alcohol abuse, with the remaining other ∼25% attributable to HPV infection (4, 5). Therefore, efforts to promote the HPV vaccination and reduce pharmacy tobacco sales could help reduce risk factors in patients with head and neck cancer (6, 7).

There are ∼2,814,000 cancer deaths in China, and the incidence and mortality rates are steadily increasing. In almost all populations examined, the incidence of cancer is higher in men than women, and the mortality rate is almost equal to the morbidity rate (8, 9). Among all cancers, HNSCC occupies an important position which has seriously threatened the health and lives of the Chinese people (10). Another point is that the incidence of HNSCC is ∼3 times more common in men than women. Therefore, this area of research to improve the early detection, early diagnosis, and early treatment of HNSCC is of high importance, which could possibly enable the patients to receive the best treatment within the shortest possible time to reduce injury and prevent death rates by improving their clinical treatment (11).

In recent years, radiomics has been applied in the medical field for medical diagnosis, treatment, and prognosis prediction (12, 13). Many of radiomics features can be extracted from regions of interest on medical images which can be associated with clinical diagnosis and biological characteristics to build a diagnostic model, thereby improving the diagnostic efficacy (14). Integration of big data and medicine has raised new hopes for personalized medicine, and radiomics has become more feasible, extracting large amounts of data from medical images. CT has the characteristics of quick scanning speed and high repeatability, making it the most preferred inspection method to detect early HNSCC. Conventional examination methods are easily susceptible to the influence of a physician's subjective experiences. Furthermore, morphological characteristics supplied by CT are insufficient to evaluate the biological characteristics of the primary tumor. At present, radiomics overcomes the insufficiency of the above traditional imaging techniques, widely used in clinical diagnosis, treatment, and prognosis (15, 16). The qualitative and quantitative assessment of lesion characteristics and intratumoral spatial heterogeneity can contribute to improving the non-invasive preoperative diagnosis accuracy of HNSCC. This approach also provides personalized adjuvant treatment programs (17). What has been inquired into in this research is the evaluation of the clinical application of a CT-based radiomics prediction model for discriminating laryngeal squamous cell carcinoma (SCC) and squamous cell hyperplasia (SCH).

2 Materials and methods

2.1 Patient characteristics

In this retrospective study, a total of 254 clinical samples were selected from 291 patients with larynx-occupying lesions who underwent primary surgery, and all lesions were validated via histopathological examination at The Second Hospital of Jilin University between June 2004 and December 2019. A flowchart is presented in Figure 1 to represent the process of selecting patients. All patients underwent preoperative unenhanced and dual-phase contrast-enhanced CT examination of the neck. Clinical-pathological data including age, gender, smoking and drinking history, pathological grade, tumor size, and clinical stage were collected. They were divided into two groups: the first group of patients were diagnosed with SCC while the second group consisted of those diagnosed with SCH. The study involved 227 men and 27 women with an average age of 44–85 years. Among them, 209 patients were smokers while 45 patients were non-smokers, and 202 patients reported alcohol consumption, while 52 patients did not drink.

The study included the following inclusion criteria: (a) Patients with primary larynx-occupying lesion; (b) No medical history of preoperative chemoradiotherapy; (c) Complete clinical and pathological diagnosis data were acquired; (d) Plain scan plus conventional dual-phase enhanced scan (arterial phase and venous phase) were implemented before the operation with complete image information. The exclusion criteria were as follows: (a) The lesions measured 10mm or less in diameter, partially becoming superficial; (b) No neck CT examination before the operation; (c) Low-quality CT images due to movement or artifacts.

2.2 CT radiomics analysis

The radiomics analysis process mainly includes five phases: CT image acquisition, ROI segmentation, feature extraction, feature selection, forecast model establishment, and diagnostic performance evaluation, as shown in Figure 2.

2.2.1 CT image acquisition

All patients underwent a neck CT (iCT 256, Philips, Netherlands) to collect plain scan plus conventional dual-phase enhanced scans (arterial phase and venous phase). In general, the scanning neck range was from the inferior margin of the foramen magnum to the upper edge of the aortic arch, in which the patients were laid in the supine position with the neck completely exposed during the scanning. The most important parameters for the neck tissue CT scan requested here are as follows: the tube voltage 120kVp, electric current of 200mAs, scan time raging from 1s to 3s, slice thickness of 1 mm, matrix size of 256 × 256, and pitch ratio of 0.342. For an enhanced scan, iodinated contrast material was intravenously injected at a dose of 1.5 mL/kg, with an injection rate of 3.5 mL/s. The time for the arterial phase scan after injection was 35 s, whereas the time for the portal venous scan was 65 s.

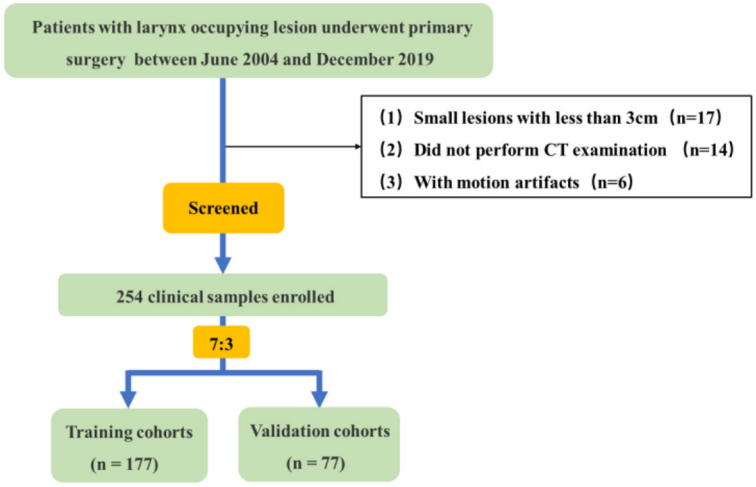


FIGURE 1
Flowchart of the patient selection process.

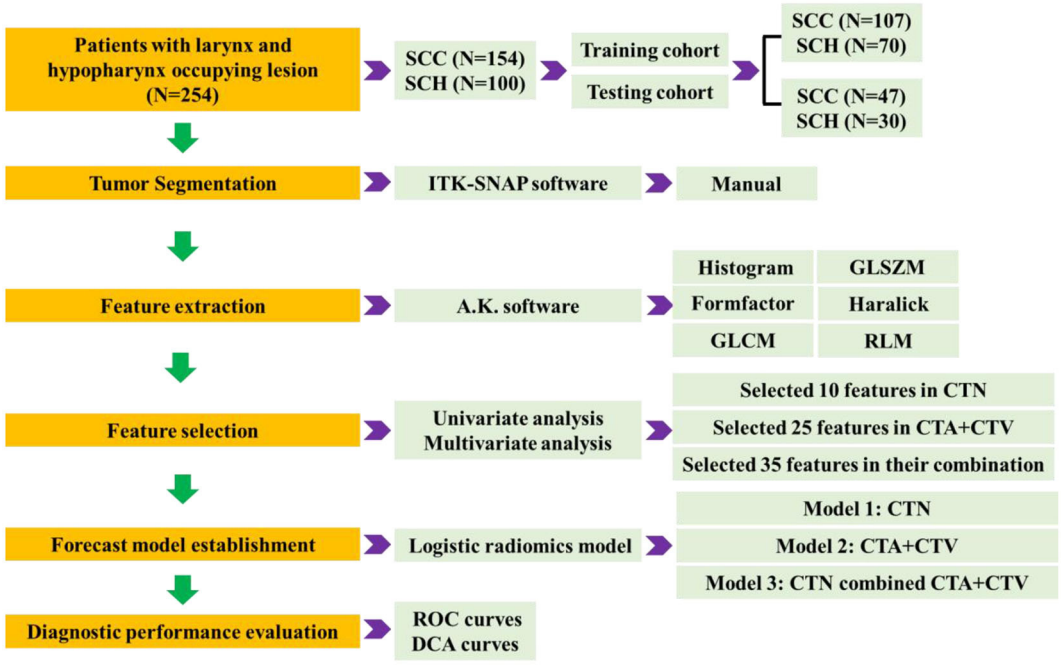


FIGURE 2
Graphical summarization of radiomics models analysis process.

2.2.2 ROI segmentation

Three-dimensional manual segmentation of tumor in axial CT image of plain, arterial, and venous phase utilized via ITK-SNAP software (v.3.8.0; www.itksnap.org). Under the random and double-blind method, the boundaries of each layer were manually delineated on tumors by two radiologists with head and neck CT imaging diagnostic experience, and the region of interest (ROI) was composited. The tumor ROI segmentation in the larynx and hypopharynx was performed by a junior radiologist (with 5 years of experience) and reviewed by another senior radiologist (with 10 years of experience). In case of occurrence of a dispute, the final decision was made after the discussion between two doctors.

2.2.3 Feature extraction

High-throughput quantitative features are extracted from 3D ROI of tumor lesions in three-phasic CT scan, from which mainly include histogram, gray-level size zone matrix (GLSZM), formfactor, haralick, gray-level cooccurrence matrix (GLCM), run length matrix (RLM) by A.K. software (Analysis Kit, GE Healthcare). In total, 1,188 quantitative radiomics features were extracted for each patient, with 396 features from each of the plain, arterial, and venous phases, respectively.

2.2.4 Feature selection and forecast model establishment

We implement a comprehensive feature selection to establish the final forecast model via IPM statistics (V1.1.463 GE Healthcare, Shanghai, China). Before feature selection, the collected 1188 quantitative radiomics features were preprocessed, and feature normalization were employed. We first proposed that variables with zero variance were excluded from analyses. Then, the missing values and outlier values were replaced by the median. Finally, the data were standardized by standardization.

The 254 tumors will be randomly allocated based on a 7:3 ratio between the training cohort and the testing cohort, where 177 patients were used as the training cohort for feature selection and model building. The remaining 23 patients were regarded as the testing cohort for verifying the selected features and forecast model. Feature selection was implemented employing univariate (using Variance, Correlation_xx, and General_Univariate_analysis) and multivariate analyses (using L1) with stepwise selection-based dimensionality reduction algorithm for all the features. To further avoid model overfitting, 10-foldcross-validation was performed to recurve the selection of redundancy features and the least absolute shrinkage and selection operator (LASSO) regression to effectively eliminate a sequence of regression coefficients to exactly zero. Furthermore, a set of optimal features, which were compared by the Wilcoxon test, were obtained.

Subsequently, Model 1 (CTN) was constructed based on 10 optimal features; Model 2 (CTA+CTV) was constructed based on 25 optimal features; and Model 3 (CTN+CTA+CTV) considered all 35 parameters, extracted in the combination of Model 1 and Model 2 using logistic regression for discriminating laryngeal SCC and SCH. The differential diagnostic effectiveness and performance of the proposed three models were measured by the receiver operating characteristic (ROC) curves, area under the curve (AUC), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). The calibration curves were measured to analyze the goodness of fit of the prediction models. In addition, the decision curve analysis (DCA) was adapted to evaluate the clinical efficacy and safety of the three models.

2.3 Statistical analysis

All statistical analyses were conducted based on IPM statistics. The continuous variables were assessed by the Mann–Whitney *U*-test or independent samples *T*-test, and the categorical variables were investigated using the chi-squared or Fisher exact tests. A two-tailed test with a $p < 0.05$ (typically ≤ 0.05) was indicative of a statistically significant difference.

3 Results

3.1 Clinical characteristics of the patients

Table 1 shows the characteristics of the 254 patients, with 154 patients diagnosed with SCC and 100 patients diagnosed with SCH. Statistical analysis for clinical data, such as age, gender, smoking status, alcohol consumption, and tumor location described above, was performed. The comparison of the two groups yielded a $p > 0.05$, revealing that there were similar between SCC and SCH in the training and testing cohorts.

3.2 Radiomic feature selection and model building

There are three steps in building Model 1: For the CT plain scan, a total of 396 features were first subjected to the variance method (threshold: 1.0) to screen out 98 features. Next, we adopted the procedure of Correlation_xx method with a cutoff set to 0.7 to remove redundant features, resulting in remaining 31 parameters. Then, a total of 11 features were retained via General_Univariate_analysis (*p*-value threshold 0.05). Finally, the remaining 10 features were obtained using the L1 method, revealing an obvious difference between SCC and SCH in the training cohorts. A correlation heat map (Figures 3A–J) revealed that strong positive correlation radiomics features were sufficient to receive an obviously differential diagnosis. Figure 3K shows the results of the tenfold cross-validation method, and Figure 3L shows the results of the LASSO regression analysis. Subsequently, 10 features from the CT plain scan were finally selected, and the optimal parameters are as follows: ["GLCMEnergy_angle0_offset7"], ["GLCMEnergy_angle90_offset4"], ["GLCMEnergy_angle90_offset7"], ["GLCMEntropy_angle90_offset1"], ["InverseDifferenceMoment_AllDirection_offset1_SD"], ["InverseDifferenceMoment_AllDirection_offset7_SD"], ["InverseDifferenceMoment_angle135_offset7"], ["HighGreyLevelRunEmphasis_angle90_offset1"], ["LongRunLowGreyLevelEmphasis_AllDirection_offset4_SD"], ["RunLengthNonuniformity_angle90_offset7"].

TABLE 1 The risk factor analysis of larynx-occupying lesions in the training and testing cohorts examining clinical characteristics.

	Training cohort			Testing cohort			*P*-value
	SCC ($N = 107$)	SCH ($N = 70$)	*P*-value	SCC ($N = 47$)	SCH ($N = 30$)	*P*-value	
Age	61.9 (9.00)	60.0 (8.05)	0.150	60.2 (8. 27)	56.9 (6.90)	0.066	0.051
Sex			0.774			1.000	0.762
Female	11 (10.3%)	9 (12.9%)		4 (8.51%)	3 (10.0%)		
Male	96 (89.7%)	61 (87.1%)		43 (91.5%)	27 (90.0%)		
Smoking			0.656			0.643	0.067
No	17 (15.9%)	11 (15.7%)		4 (8.51%)	1 (3.33%)		
Yes	90 (84.1%)	59 (84.3%)		43 (91.5%)	29 (96.7%)		
Alcohol			0.759			0.186	0.651
No	20 (18.7%)	11 (15.7%)		9 (19.1%)	2 (6.67%)		
Yes	87 (81.3%)	59 (84.3%)		38 (80.9%)	28 (93.3%)		
Tumor location			0.476			0.951	0.603
Supraglottis	27 (25.27%)	19 (27.1%)		11 (23.43%)	6 (20.0%)		
Glottis	68 (63.6%)	47 (67.1%)		33 (70.2%)	23 (76.7%)		
Subglottis	12 (11.2%)	4 (5.71%)		3 (6.38%)	1 (3.33%)		

Based on above single factor analysis and multifactor analysis, a total of 10 most predictive features and coefficients for constructing the optimal radiomics signatures Model 1(Figure 5A). Consequently, rad-score was calculated by selected 10 features weighted as below: Rad-score (Model 1)=0.9334 + 0.4575[“GLCMEnergy_angle0_offset7”]-1.2571[“GLCMEnergy_angle90_offset4”]-0.0960[“GLCMEnergy_angle90_offset7”]-1.1963[“GLCMEntropy_angle90_offset1”]-0.6162[“InverseDifferenceMoment_AllDirection_offset1_SD”]-1.3213[“InverseDifferenceMoment_AllDirection_offset7_SD”]-3.6916[“InverseDifferenceMoment_angle135_offset7”]-0.2776[“HighGreyLevelRunEmphasis_angle90_offset1”] +0.8819[“LongRunLowGreyLevelEmphasis_AllDirection_offset4_SD”]+4.2529[“RunLengthNonuniformity_angle90_offset7”].

Using similar feature reduction methods, 25 features from the CT conventional dual-phase enhanced scan were finally selected by specific methods listed in four steps: In the first step, after applying the variance method (threshold: 1.0), the 792 features extracted by the conventional dual-phase enhanced scan were reduced to 331 features. Next, we obtained a feature count of 86 parameters by the Correlation_xx method (cutoff: 0.7). Then, General_Univariate_analysis was utilized to remove 54 features, and 32 radiomics features remained. Finally, the remaining 25 features were obtained using the L1 method. A correlation heat map (Figures 4A–H) revealed that strong positive correlation radiomics features were sufficient to receive an obviously differential diagnosis. Figure 3I shows the results of the tenfold cross-validation method, and Figure 3J shows the results of the LASSO regression analysis. The optimal parameters are as follows: [“V_ClusterProminence_angle0_offset7”], [“V_Correlation_angle0_offset4”], [“V_GLCMEnergy_angle90_offset7”], [“V_HaralickCorrelation_angle45_offset7”], [“V_Inertia_AllDirection_offset4_SD”], [“V_Inertia_AllDirection_offset7”], [“V_InverseDifferenceMoment_angle0_offset7”], [“V_GreyLevelNonuniformity_AllDirection_offset4_SD”], [“V_Compactness2”], [“V_Maximum3DDiameter”], [“A_ClusterProminence_angle90_offset7”], [“A_ClusterShade_angle135_offset1”], [“A_ClusterShade_angle45_offset4”], [“A_ClusterShade_angle90_offset1”], [“A_ClusterShade_angle90_offset7”], [“A_GLCMEnergy_angle90_offset4”], [“A_GLCMEntropy_angle45_offset4”], [“A_HaralickCorrelation_angle135_offset4”], [“A_HaralickCorrelation_angle90_offset7”], [“A_Inertia_AllDirection_offset1_SD”], [“A_HighGreyLevelRunEmphasis_AllDirection_offset7_SD”], [“A_LongRunLowGreyLevelEmphasis_angle45_offset7”], [“A_RunLengthNonuniformity_AllDirection_offset1”], [“A_RunLengthNonuniformity_AllDirection_offset7_SD”], [“A_ShortRunHighGreyLevelEmphasis_AllDirection_offset1_SD”].

Based on the above single-factor analysis and multifactor analysis, a total of 25 most predictive features and coefficients were identified for constructing the optimal radiomics signatures in Model 2 (Figure 5B). The Rad-score was calculated using the selected 25 features weighted as below: Rad-score (Model 2) = 3.1869 + 1.4625[“V_ClusterProminence_angle0_offset7”]-4.1270[“V_Correlation_angle0_offset4”] + 0.2540[“V_GLCMEnergy_angle90_offset7”] + 2.3007[“V_HaralickCorrelation_angle45_offset7”] + 0.1652[“V_Inertia_AllDirection_offset4_SD”]-0.6157[“V_Inertia_AllDirection_offset7”]-1.3833[“V_InverseDifferenceMoment_angle0_offset7”] + 0.8270[“V_GreyLevelNonuniformity_AllDirection_offset4_SD”] + 0.7914[“V_Compactness2”]-0.0179[“V_Max

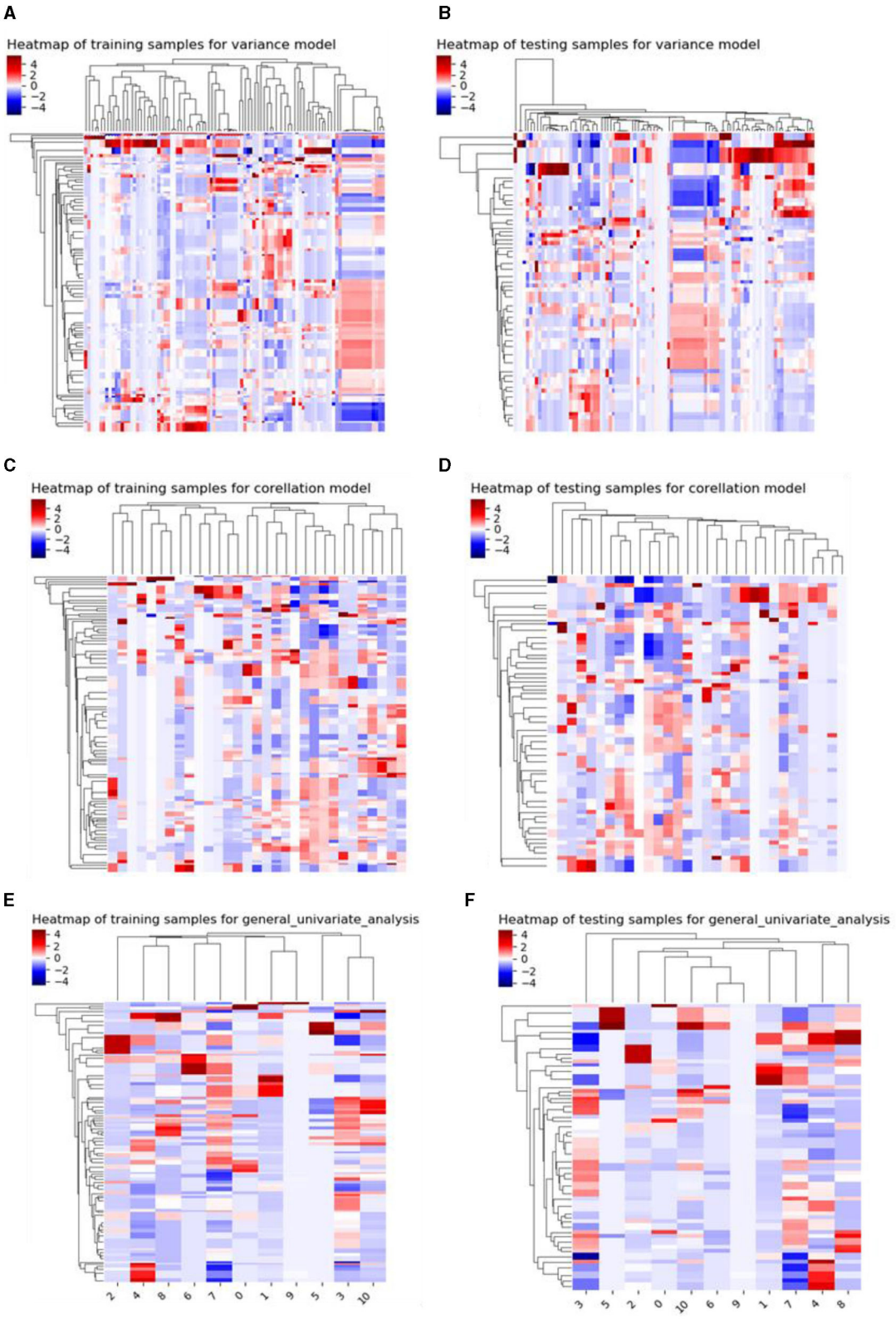


FIGURE 3
Continued

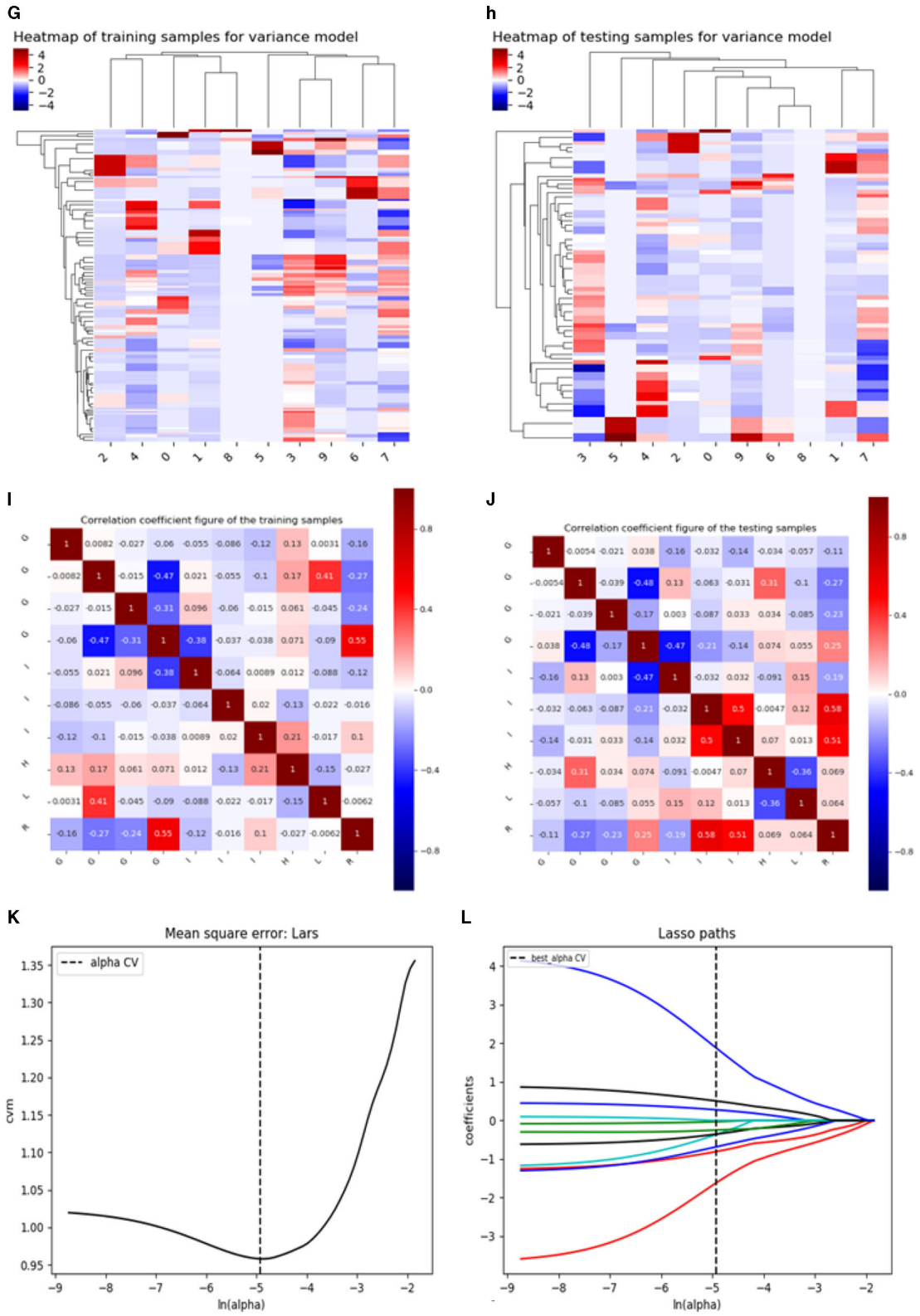


FIGURE 3
(A–J) shows the correlation heat maps of Model 1 demonstrating correlations between features in the training and testing cohort. **(K, L)** shows the results of the 10-fold cross-validation method and LASSO regression analysis, respectively, removing highly redundant features to obtain the optimal features.

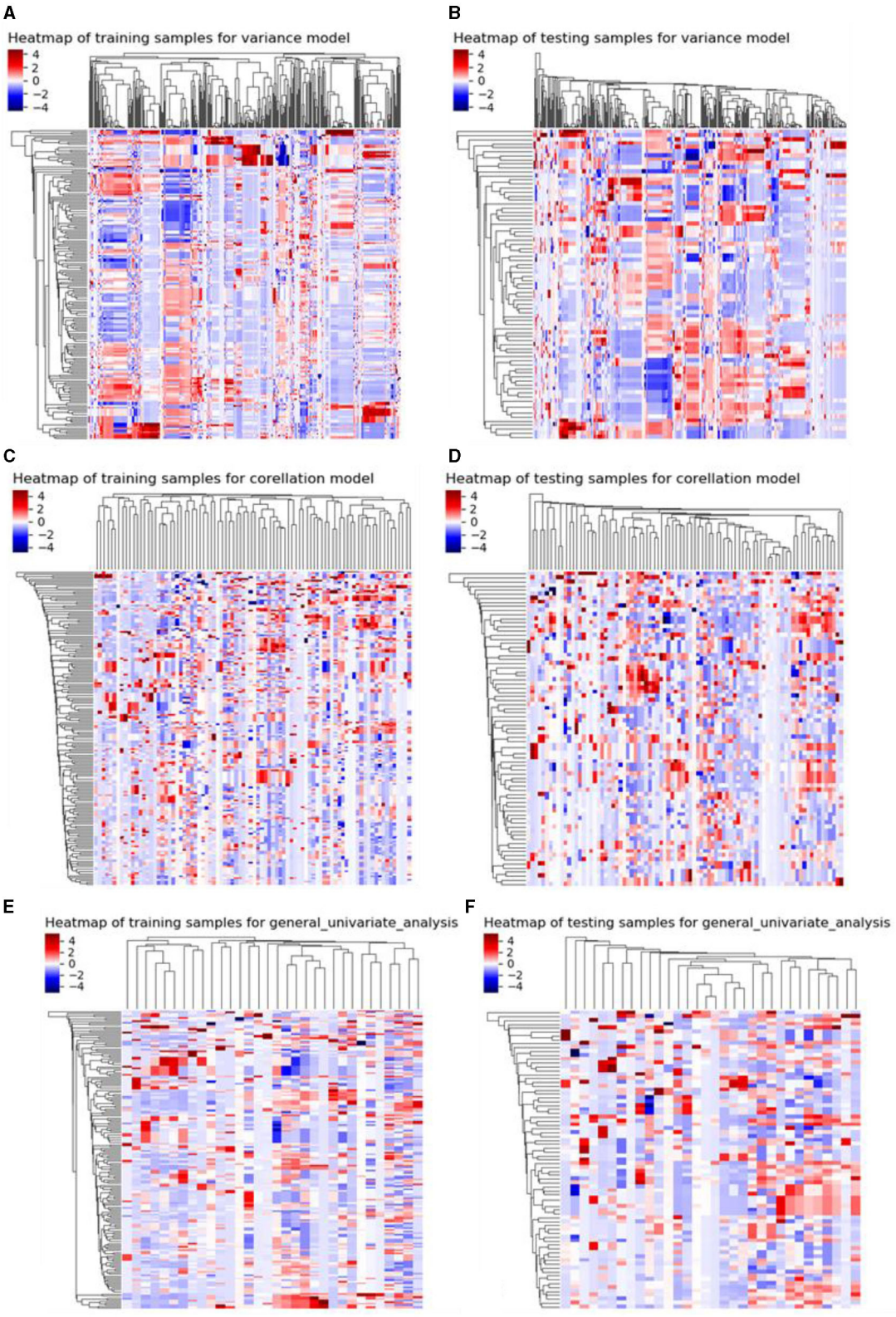


FIGURE 4
Continued

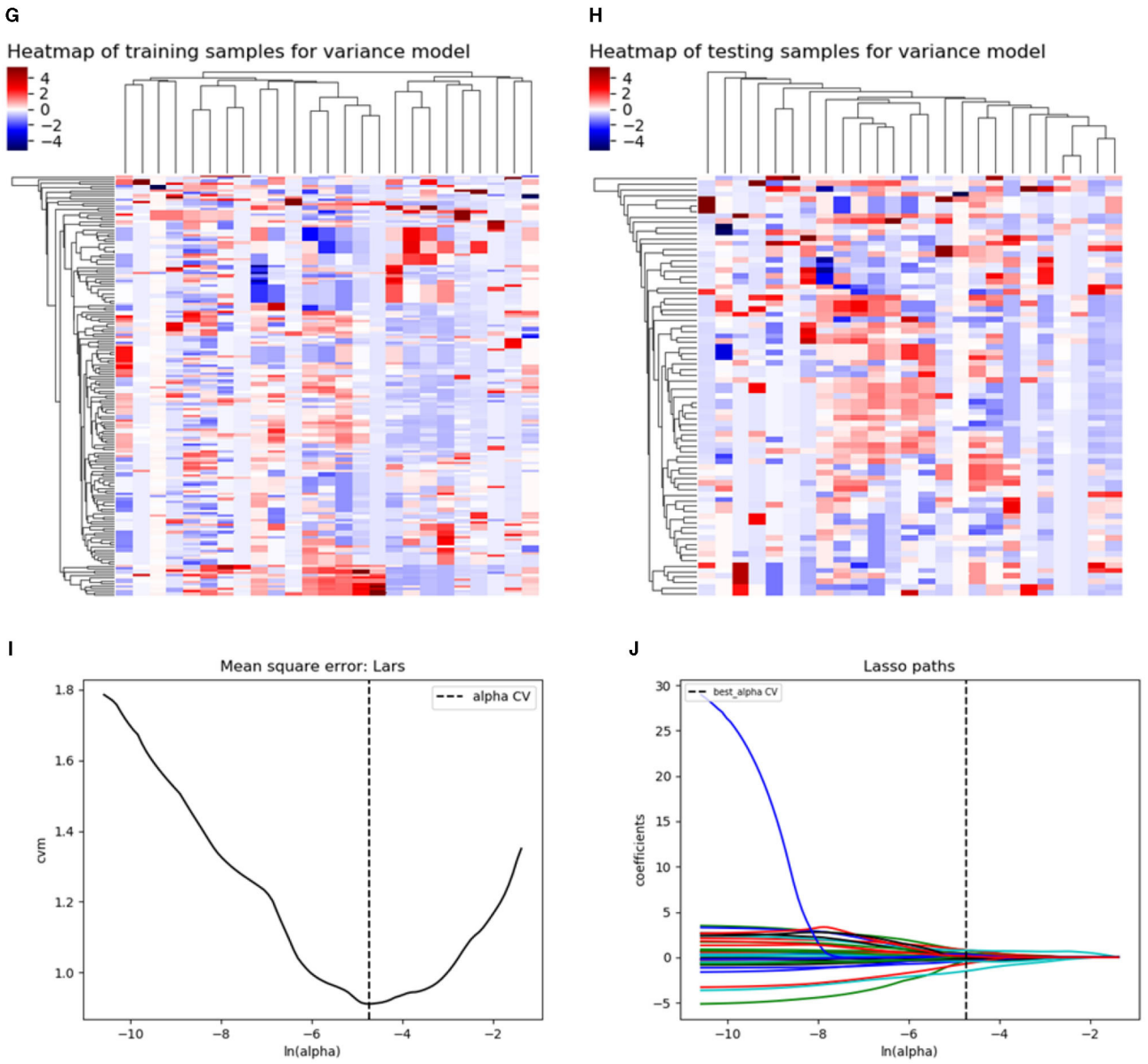


FIGURE 4
(A–H) shows the correlation heat maps of Model 2 demonstrating the correlations between features in the training and testing cohort. **(I, J)** show the results of the 10-fold cross-validation method and LASSO regression analysis, respectively, removing highly redundant features to obtain the optimal features.

imum3DDiameter"] + 3.5277["A_ClusterProminence_angle90_offset7"] + 1.5561[A_ClusterShade_angle135_offset1""] + 0.3104["A_ClusterShade_angle45_offset4"] + 1.5868["A_ClusterShade_angle90_offset1"] + 2.7000["A_ClusterShade_angle90_offset7"]-0.0034["A_GLCMEnergy_angle90_offset4"]-2.0308["A_GLCMEntropy_angle45_offset4"] + 1.0889["A_HaralickCorrelation_angle135_offset4"]-0.4997["A_HaralickCorrelation_angle90_offset7"]-0.2386["A_Inertia_AllDirection_offset1_SD"] + 1.5807["A_HighGreyLevelRunEmphasis_AllDirection_offset7_SD"]-2.7901["A_LongRunLowGreyLevelEmphasis_angle45_offset7"] + 0.6788["A_RunLengthNonuniformity_AllDirection_offset1"] + 3.7087["A_RunLengthNonuniformity_AllDirection_offset7_SD"] + 0.5963["A_ShortRunHighGreyLevelEmphasis_AllDirection_offset1_SD"]

To comprehensively and intuitively demonstrate the characteristics and differences of the dataset, we put together the boxplots of radiomics scores for benign and malignant cases. Based on all collected clinical dataset, Figure 6 shows that the Rad-score was significantly upregulated in malignant patients as compared to benign controls in both Model 1(CTN) and Model 2(CTAV).

As shown in Table 2, the Wilcoxon rank sum test was used to compare the Rad-scores of Model 1 and Model 2 for discriminating SCC and SCH in both the training and testing cohorts (both $P < 0.001$).

The process of constructing the three-period combined Model 3 is presented as follows: The 10 optimal radiomics

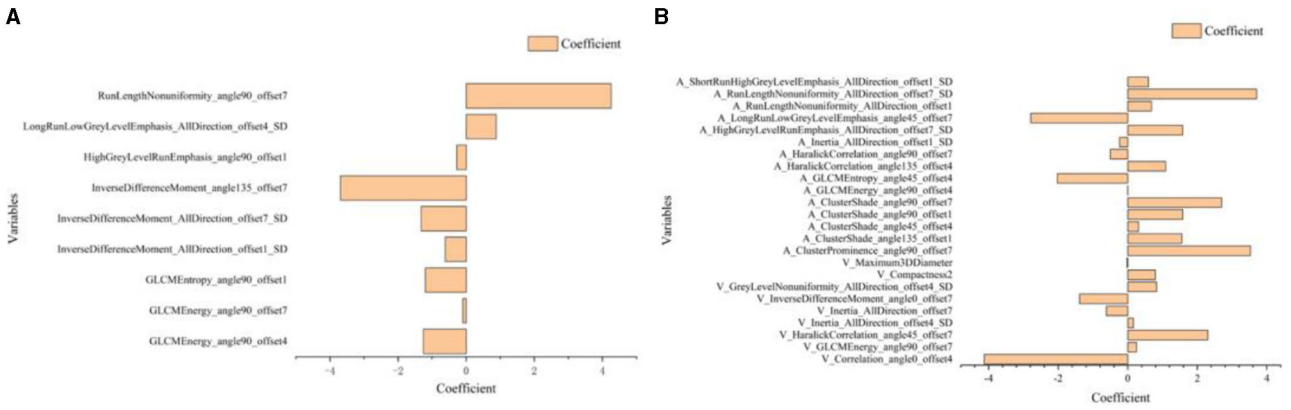


FIGURE 5
Histogram shows the 10 most predictive radiomics features obtained in Model 1 **(A)** and 25 most predictive radiomics features obtained in Model 2 **(B)**.

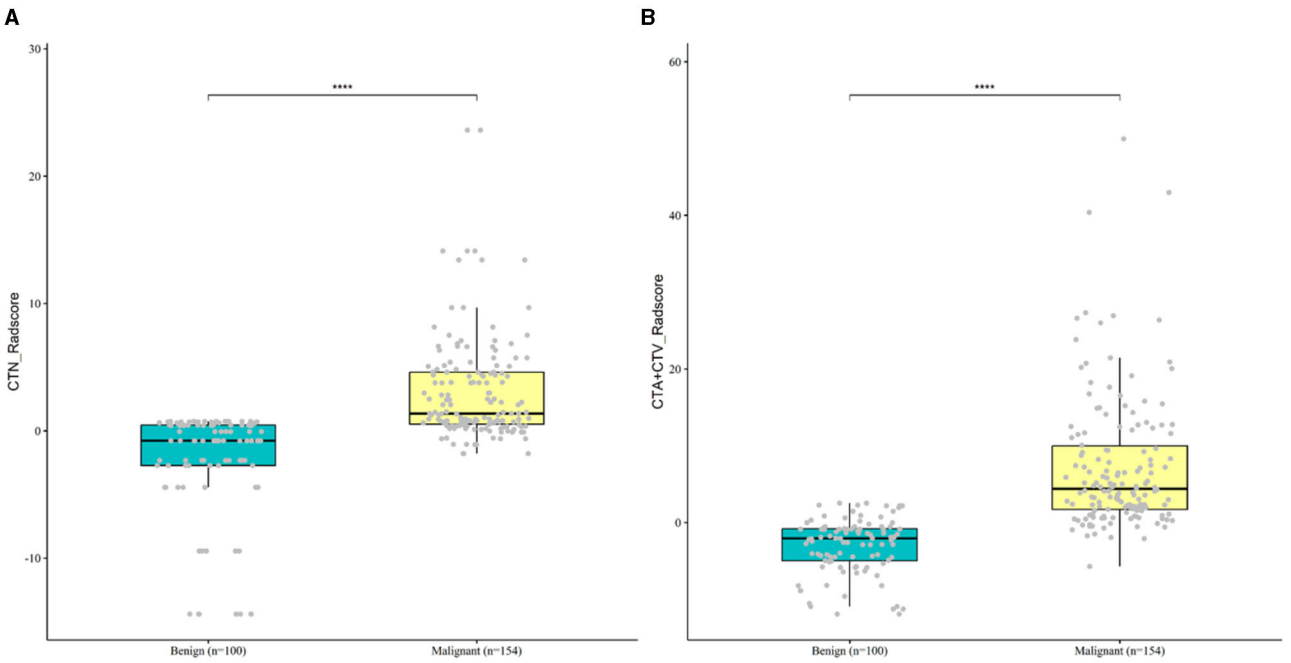


FIGURE 6
Boxplot of differentially Rad-score for malignant and benign case in Model 1 **(A)** and Model 2 **(B)**.

TABLE 2 **The Wilcoxon rank sum test of Rad-scores in Model 1 and Model 2 for training and testing cohorts.**

	Training cohort (*P*-value)	Validation cohort (*P*-value)
Model 1	< 0.001*	< 0.001*
Model 2	< 0.001*	< 0.001*

*Indicates that the difference is significant.

signatures selected from Model 1 and the 25 optimal radiomics signatures from Model 2 were considered, incorporating all 35 parameters to establish Model 3 (CTN+CTA+CTV) using logistic regression for discriminating laryngeal SCC and SCH.

3.3 Comparing the performance of the three different models

The corresponding performance evaluation criteria for differentiating SCC and SCH contained AUC, accuracy, sensitivity, specificity, PPV, and NPV for each model. In radiomic prediction Model 1(CTN), the measured values of the training cohort were 0.883, 0.785, 0.645, 1.000, 1.000, and 0.648, while in the testing

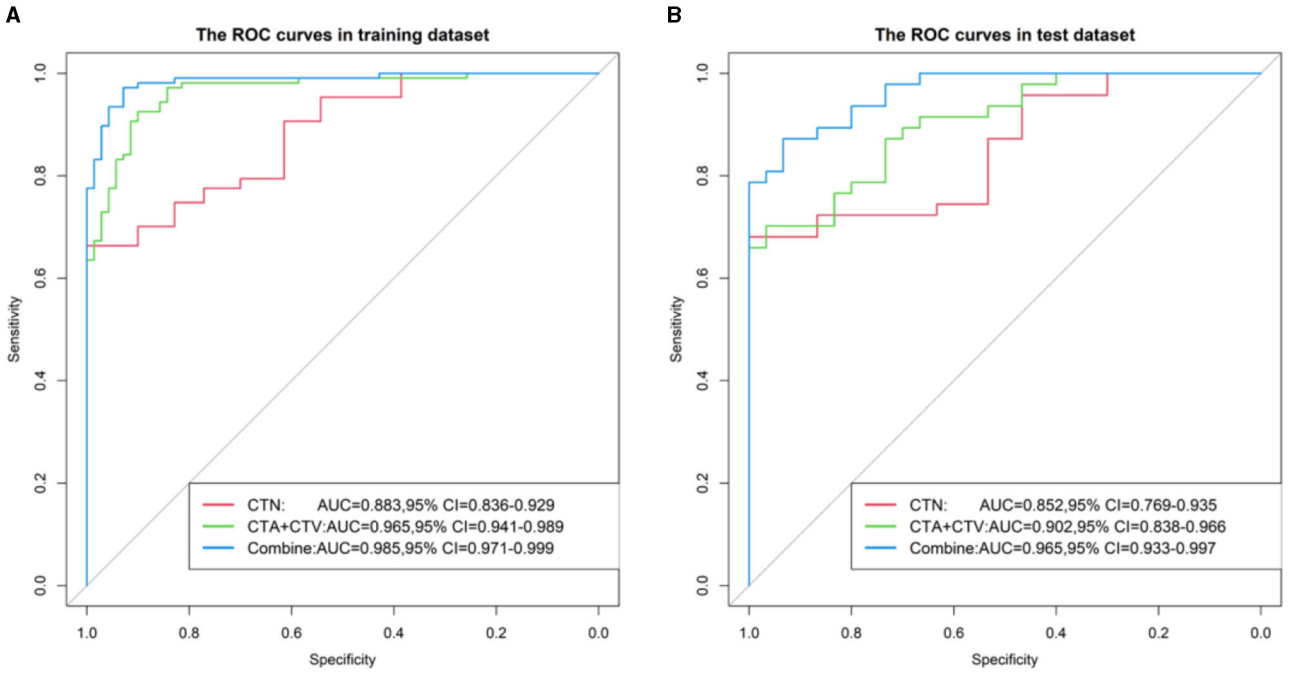


FIGURE 7
ROC curve of three Models (CTN, CTA+CTV, and CTN+CTA+CTV) in the training **(A)** and testing cohorts **(B)**.

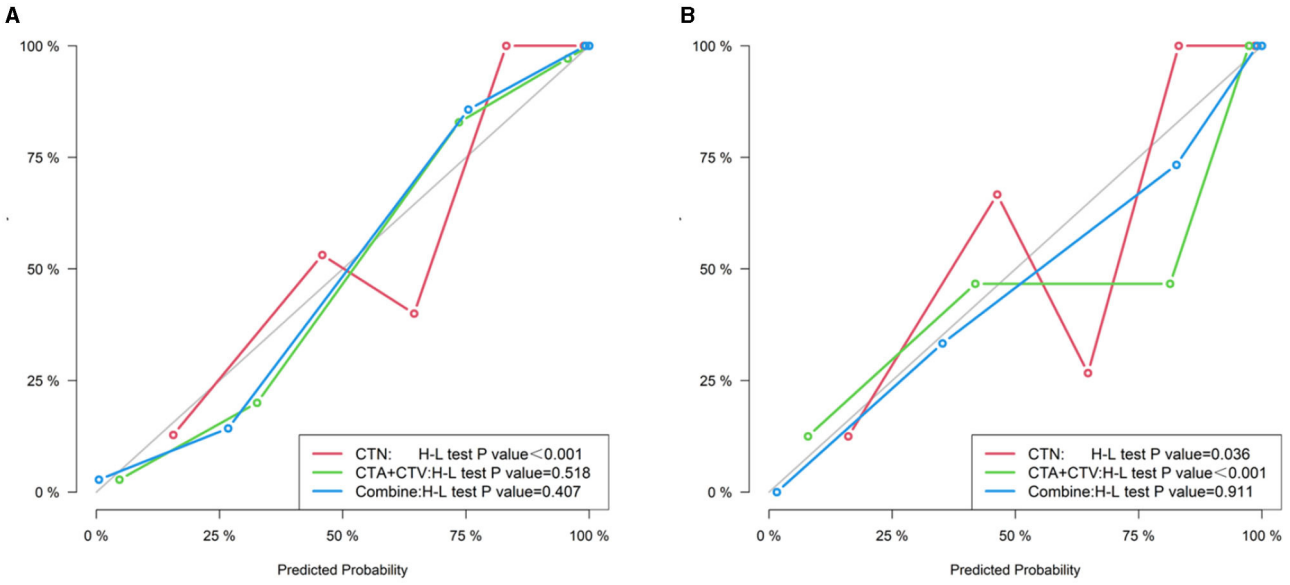


FIGURE 8
Calibration curve of three models (CTN, CTA+CTV, and CTN+CTA+CTV) in the training **(A)** and testing cohorts **(B)**.

cohorts were 0.852, 0.792, 0.66, 1.000, 1.000, and 0.652. In radiomic prediction Model 2 (CTA+CTV), the measured values of training cohorts were 0.965, 0.91, 0.916, 0.9, 0.933, and 0.875, while in the testing cohorts were 0.902, 0.805, 0.851, 0.733, 0.833, and 0.759. In radiomic prediction Model 3 (CTN+CTA+CTV), the measured values of training cohorts were 0.985, 0.944, 0.953, 0.929, 0.953, and 0.929, while in the testing cohorts were 0.965, 0.857, 0.894, 0.8, 0.875, and 0.828, respectively. Among them, Model 3 has the highest performance and can be used for predicting differential clinical diagnosis.

We constructed the calibration curve to describe the degree of calibration of the three models in the training (Figure 7A) and testing (Figure 7B) cohorts, which illustrates that the closer the calibration curves (red, green and blue) are to the standard curve

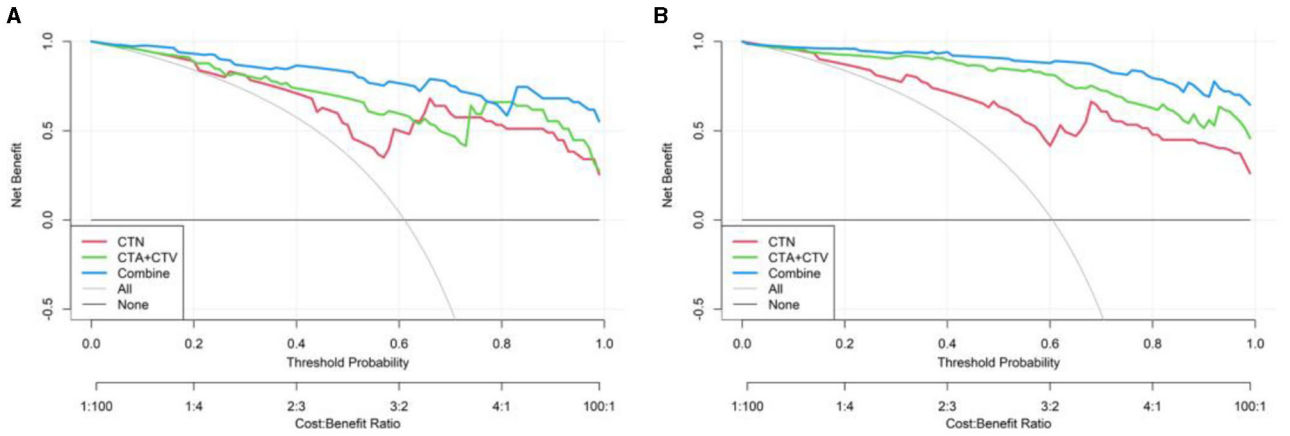


FIGURE 9
Decision curve of three models (CTN, CTA+CTV, and CTN+CTA+CTV) in the training **(A)** and testing cohorts **(B)**.

(black), the better the calibration capability of the model. As shown in Figure 8, the actual prediction performance of the prediction Model 3 has a good consistency.

We also established a calibration curve with a DCA curve for evaluating three models in the training (Figure 9A) and testing (Figure 9B) cohorts, which indicated that a larger area under the decision curve indicated a better clinical practicability. As shown in Figure 9, within the safe range, the DCA indicated that the net benefit of the prediction was higher for all three models, with Model 3, exhibiting the highest net benefit, and the benefit rate of the population would reach its maximum.

4 Discussion and conclusion

Radiomics is an emerging field of image analysis with potential applications for diagnosis, treatment response, and prognosis evaluation in cancer patients via computer-aided diagnosis (CAD) technology, which solved many clinical problems, reduced misdiagnosis rate, and decreased the radiologist workload (18). In our study, radiomics used computational algorithms to convert medical imaging information into high-resolution quantitative mineable "big data" and aimed to further build a more reliable predictive and differential diagnosis model (19).

A multicenter study by Zhang et al. (20) showed that different machines and different CT scanning parameters might influence the radiomics result. Therefore, CT image data extraction is the basis of radiomics. It is important to note that CT scanning parameters could be consistent for radiomics analysis, in which the CT images were collected uniformly. We were doubtful that different CT scanning parameters might affect the radiomics optimal features and prediction model establishment.

The radiomics features of the entire primary larynx-occupying lesion were extracted from the CT images of the arterial, venous, and non-contrast phases, which have a positive effect on establishing a differential diagnosis model for discriminating SCC and SCH. We concluded that the AUC, accuracy, sensitivity, specificity, PPV, and NPV of the optimal Model 3 in training cohorts for differentiating SCC and SCH were 0.985, 0.944, 0.953, 0.929, 0.953, and 0.929, respectively, while in the testing cohorts were 0.965, 0.857, 0.894, 0.8, 0.875, and 0.828, respectively. Based on the research results, a comprehensive comparative analysis of the consequences of these three models revealed that the amalgamation of characteristic parameters from the plain-arterial-venous combined model provided more optimal radiomics parameters than from the plain model or arterial-venous combined model alone. Thus, Model 3 based on arterial-venous-plain combined scan phase of CT has good discriminative performance with high sensitivity and specificity in SCC and SCH.

Additionally, we collected related important clinical characteristics of all patients, including age, gender, smoking status, and alcohol consumption. Among them, smoking status and alcohol consumption were the major risk factors leading to the high incidence of laryngeal cancer. Among the enrolled patients, 209 patients had a smoking history and 45 patients had no smoking history. We analyzed that long-term smokers are more susceptible to developing laryngeal cancer, taking into account consistent smoking initiation and current smoking status. In terms of alcohol consumption, 202 patients had a drinking history and 52 patients have no drinking history. Similarly, we found that long-term drinkers are more susceptible to developing laryngeal cancer, taking into account the drinking start times and current drinking quantity. Therefore, clinicians will be able to advise their patients to quit smoking and drinking, thereby extending overall patient survival time (21, 22).

Chen et al. (23) evaluated the use of venous-phase CT images to develop a radiomics model, a deep learning model, and a combined model to predict preoperative staging in stratifying patients with laryngeal carcinoma. The authors demonstrated that the combined model performed significantly capability than a radiologist in stratifying patients into stage I–II and stage III–IV. The AUCs, which indicated model diagnostic performance, assessed the accuracy of a model. The radiomics model, DL model, and combined model for distinguishing staging ability in the test set were 0.704 (95% CI: 0.588–0.820), 0.724 (95% CI: 0.613–0.835), and 0.849 (95% CI: 0.755–0.943), respectively. This

study confirmed the application value of radiomics in accurate preoperative staging of laryngeal cancer. Kang et al. (24) developed a radiomics nomogram to analyze 114 patients with advanced laryngeal cancer after induction chemotherapy. The experiments demonstrated that the Rad-score was an independent predictor. In addition, clinical factors were incorporated to build radiomics nomogram which predicted the pathological response and overall survival. This study proved that CT radiomics nomogram possesses the best predictive property in the pathological response after induced chemotherapy and overall survival. Therefore, we fused plain scan and conventional dual-phase enhanced scanning to the radiomics model to improve the predictive performance.

While the radiomics models are promising in clinical practice, there are several key limitations to this study that require attention. First, the model was established based on the single-center nature of the research that lacked prospective multicenter external validation of our findings. Second, we only collected the CT examination images; however, not collect MRI data, which may cause potential data bias. Finally, the number of clinical samples size is relatively limited, which expanded sample size to further research.

To the best of our knowledge, with the supplement of related important clinical characteristics, Model 3 based on arterial-venous-plain combined scan phase of CT has important clinical significance in distinguishing SCC from SCH. In conclusion, our results demonstrated that radiomics could provide valuable information and play an important role in preoperative diagnosis and clinical treatment to guide clinicians to develop individualized treatment programs (25, 26).

Data availability statement

The datasets presented in this article are not readily available because protect the privacy of enrolled patient information. Requests to access the datasets should be directed to FC, 574267524@qq.com.

Ethics statement

The studies involving humans were approved by the Ethics Committee of Second Hospital of Jilin University. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin because Since this study was retrospective, it was not necessary to obtain written informed consent from patients. Written informed consent was not obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article because Since this study was retrospective, it was not necessary to obtain written informed consent from patients.

Author contributions

FC: Writing—original draft. OK: Writing—review & editing. WL: Writing—review & editing. JL: Writing—review & editing. QY: Writing—review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Program of Youth Science and Technology Innovation and Entrepreneurship Outstanding Talents (Team) of Jilin Province, China (no. 20230508063RC) and the Excellent Youth Training Foundation of Jilin University, China (no. 419080520665).

Acknowledgments

We sincerely appreciate Min Zhao from the Department of Core Imaging, Life Science, GE healthcare, Shanghai, China, for their assistance in the data processing and statistical analysis.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Mody MD, Rocco JW, Yom SS, Haddad RI, Saba NF. Head and neck cancer. *Lancet.* (2021) 398:2289–99. doi: 10.1016/S0140-6736(21)01550-6

2. Johnson DE, Burtness B, Leemans CR, Lui VWY, Bauman JE, Grandis JR. Head and neck squamous cell carcinoma. *Nat Rev Dis Primers.* (2020) 6:1–22. doi: 10.1038/s41572-020-00224-3

3. Lin C, Cheng W, Liu X, Li H, Song Y. The global, regional, national burden of laryngeal cancer and its attributable risk factors (1990–2019) and predictions to 2035. *Eur J Cancer Care.* (2022) 31:e13689. doi: 10.1111/ecc.13689

4. Giraldi L, Leoncini E, Pastorino R, Wünsch-Filho V, de Carvalho M, Lopez R, et al. Alcohol and cigarette consumption predict mortality in patients with head and neck cancer: a pooled analysis within the international head and neck cancer epidemiology (INHANCE) consortium. *Annal Oncol.* (2017) 28:2843–51. doi: 10.1093/annonc/mdx486

5. Gallaway MS, Henley SJ, Steele CB, Momin B, Thomas CC, Jamal A, et al. Surveillance for cancers associated with tobacco use – United States, 2010–2014. *MMWR Surv Summ.* (2018) 67:1–42. doi: 10.15585/mmwr.ss6712a1

6. Law AB, Schmitt NC. Laryngeal anatomy, molecular biology, cause, and risk factors for laryngeal cancer. *Otolaryngol Clin North Am*. (2023) 56:197–203. doi: 10.1016/j.otc.2022.12.001

7. Mima M, Okabe A, Hoshii T, Nakagawa T, Kurokawa T, Kondo S, et al. Tumorigenic activation around HPV integrated sites in head and neck squamous cell carcinoma. *Int J Cancer*. (2023) 152:1847–62. doi: 10.1002/ijc.34439

8. Hsueh C-Y, Shao M, Cao W, Li S, Zhou L. Pretreatment serum uric acid as an efficient predictor of prognosis in men with laryngeal squamous cell cancer: a retrospective cohort study. *Oxid Med Cell Longev*. (2019) 2019:1821969. doi: 10.1155/2019/1821969

9. Zhang Q, Xiong Y, Lin L, Yuan K. Analysis of related factors of surgical treatment effect on 215 patients with laryngeal cancer. *Exp Ther Med*. (2018) 15:2786–91. doi: 10.3892/etm.2018.5741

10. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. *CA A Cancer J Clin*. (2016) 66:115–32. doi: 10.3322/caac.21338

11. Liu Y, Han Y, Chen B, Zhang J, Yin S, Li D, et al. A new online dynamic nomogram: construction and validation of an assistant decision-making model for laryngeal squamous cell carcinoma. *Front Oncol*. (2022) 12:829761. doi: 10.3389/fonc.2022.829761

12. Chiesa-Estomba CM, Echaniz O, Larruscain E, Gonzalez-Garcia JA, Sistiaga-Suarez JA, Graña M. Radiomics and texture analysis in laryngeal cancer. Looking for new frontiers in precision medicine through imaging analysis. *Cancers*. (2019) 11:1409. doi: 10.3390/cancers11101409

13. Huang EP, O'Connor JPB, McShane LM, Giger ML, Lambin P, Kinahan PE, et al. Criteria for the translation of radiomics into clinically useful tests. *Nat Rev Clin Oncol*. (2023) 20:69–82. doi: 10.1038/s41571-022-00707-0

14. Wang F, Zhang B, Wu X, Liu L, Fang J, Chen Q, et al. Radiomic nomogram improves preoperative T category accuracy in locally advanced laryngeal carcinoma. *Front Oncol*. (2019) 9:1064. doi: 10.3389/fonc.2019.01064

15. Maskeliūnas R, Kulikajevas A, Damaševičius R, Pribuišis K, Ulozaitė-Stanienė N, Uloza V. Lightweight deep learning model for assessment of substitution voicing and speech after laryngeal carcinoma surgery. *Cancers*. (2022) 14:2366. doi: 10.3390/cancers14102366

16. Chen L, Wang H, Zeng H, Zhang Y, Ma X. Evaluation of CT-based radiomics signature and nomogram as prognostic markers in patients with laryngeal squamous cell carcinoma. *Cancer Imaging*. (2020) 20:28. doi: 10.1186/s40644-020-00310-5

17. Rajgor AD, Patel S, McCulloch D, Obara B, Bacardit J, McQueen A, et al. The application of radiomics in laryngeal cancer. *Br J Radiol*. (2021) 94:20210499. doi: 10.1259/bjr.20210499

18. Feng L, Zhou D, Luo C, Shen J, Wang W, Lu Y, et al. Clinically applicable artificial intelligence algorithm the diagnosis, evaluation, and monitoring of acute retinal necrosis. *J Zhejiang Univ Sci B*. (2021) 22:504–11. doi: 10.1631/jzus.B2000343

19. Zhu C, Huang H, Liu X, Chen H, Jiang H, Liao C, et al. A clinical-radiomics nomogram based on computed tomography for predicting risk of local recurrence after radiotherapy in nasopharyngeal carcinoma. *Front Oncol*. (2021) 11:637687. doi: 10.3389/fonc.2021.637687

20. Zhang G, Yang H, Zhu X, Luo J, Zheng J, Xu Y, et al. A CT-based radiomics nomogram to predict complete ablation of pulmonary malignancy: a multicenter study. *Front Oncol*. (2022) 12:841678. doi: 10.3389/fonc.2022.841678

21. Chen X, Gao L, Sturgis EM, Liang Z, Zhu Y, Xia X, et al. HPV16 DNA and integration in normal and malignant epithelium: implications for the etiology of laryngeal squamous cell carcinoma. *Annal Oncol*. (2017) 28:1105–10. doi: 10.1093/annonc/mdx027

22. Chang C-P, Siwakoti B, Sapkota A, Gautam DK, Lee Y-CA, Monroe M, et al. Tobacco smoking, chewing habits, alcohol drinking and the risk of head and neck cancer in Nepal. *Int J Cancer*. (2020) 147:866–75. doi: 10.1002/ijc.32823

23. Chen X, Yu Q, Peng J, He Z, Li Q, Ning Y, et al. A combined model integrating radiomics and deep learning based on contrast-Enhanced CT for preoperative staging of laryngeal carcinoma. *Acad Radiol*. (2023) 30:3022–31. doi: 10.1016/j.acra.2023.06.029

24. Kang C, Sun P, Yang R, Zhang C, Ning W, Liu H. CT radiomics nomogram predicts pathological response after induced chemotherapy and overall survival in patients with advanced laryngeal cancer: a single-center retrospective study. *Front Oncol*. (2023) 13:1094768. doi: 10.3389/fonc.2023.1094768

25. Zhang Y, Yu S, Zhang L, Kang L. Radiomics based on CECT in differentiating kimura disease from lymph node metastases in head and neck: a non-invasive and reliable method. *Front Oncol*. (2020) 10:1121. doi: 10.3389/fonc.2020.01121

26. Usuda K, Ishikawa M, Iwai S, Yamagata A, Iijima Y, Motono N, et al. Pulmonary nodule and mass: superiority of MRI of diffusion-weighted imaging and T2-weighted imaging to FDG-PET/CT. *Cancers*. (2021) 13:5166. doi: 10.3390/cancers13205166

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 12 January 2024
DOI 10.3389/fmed.2024.1305565

Check for updates

OPEN ACCESS

EDITED BY
Xin Yang,
Huazhong University of Science and Technology, China

REVIEWED BY
Xuntao Yin,
Guangzhou Medical University, China
Jun Liu,
Central South University, China

*CORRESPONDENCE
Chuanming Li
licm@cqu.edu.cn
Jinrui Zhang
457162204@qq.com
Junbang Feng
junbangfeng@163.com

†These authors have contributed equally to this work

RECEIVED 02 October 2023
ACCEPTED 02 January 2024
PUBLISHED 12 January 2024

CITATION
Yang M, Meng S, Wu F, Shi F, Xia Y, Feng J, Zhang J and Li C (2024) Automatic detection of mild cognitive impairment based on deep learning and radiomics of MR imaging. *Front. Med.* 11:1305565. doi: 10.3389/fmed.2024.1305565

COPYRIGHT
© 2024 Yang, Meng, Wu, Shi, Xia, Feng, Zhang and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Automatic detection of mild cognitive impairment based on deep learning and radiomics of MR imaging

Mingguang Yang[1†], Shan Meng[2†], Faqi Wu[3], Feng Shi[4], Yuwei Xia[4], Junbang Feng[1*], Jinrui Zhang[1*] and Chuanming Li[1*]

[1]Medical Imaging Department, Chongqing Emergency Medical Center, Chongqing University Central Hospital, Chongqing, China, [2]Department of Radiology, Chongqing Western Hospital, Chongqing, China, [3]Department of Medical Service, Yanzhuang Central Hospital of Gangcheng District, Jinan, China, [4]Department of Research and Development, Shanghai United Imaging Intelligence, Co., Ltd., Shanghai, China

Purpose: Early and rapid diagnosis of mild cognitive impairment (MCI) has important clinical value in improving the prognosis of Alzheimer's disease (AD). The hippocampus and parahippocampal gyrus play crucial roles in the occurrence of cognitive function decline. In this study, deep learning and radiomics techniques were used to automatically detect MCI from healthy controls (HCs).

Method: This study included 115 MCI patients and 133 normal individuals with 3D-T1 weighted MR structural images from the ADNI database. The identification and segmentation of the hippocampus and parahippocampal gyrus were automatically performed with a VB-net, and radiomics features were extracted. Relief, Minimum Redundancy Maximum Correlation, Recursive Feature Elimination and the minimum absolute shrinkage and selection operator (LASSO) were used to reduce the dimensionality and select the optimal features. Five independent machine learning classifiers including Support Vector Machine (SVM), Random forest (RF), Logistic Regression (LR), Bagging Decision Tree (BDT), and Gaussian Process (GP) were trained on the training set, and validated on the testing set to detect the MCI. The Delong test was used to assess the performance of different models.

Result: Our VB-net could automatically identify and segment the bilateral hippocampus and parahippocampal gyrus. After four steps of feature dimensionality reduction, the GP models based on combined features (11 features from the hippocampus, and 4 features from the parahippocampal gyrus) showed the best performance for the MCI and normal control subject discrimination. The AUC of the training set and test set were 0.954 (95% CI: 0.929–0.979) and 0.866 (95% CI: 0.757–0.976), respectively. Decision curve analysis showed that the clinical benefit of the line graph model was high.

Conclusion: The GP classifier based on 15 radiomics features of bilateral hippocampal and parahippocampal gyrus could detect MCI from normal controls with high accuracy based on conventional MR images. Our fully automatic model could rapidly process the MRI data and give results in 1 minute, which provided important clinical value in assisted diagnosis.

KEYWORDS

MCI, AD, deep learning, radiomics, MRI

Introduction

Alzheimer's disease (AD) is an irreversible chronic neurodegenerative brain disease that poses a serious threat to human health. The main clinical manifestations include memory impairment, aphasia, loss of use and recognition, impairment of visual and spatial skills, executive dysfunction, and personality and behavioral changes (1). The occurrence and development of AD is a continuous process, and mild cognitive impairment (MCI) is considered as the preclinical stage of AD (2, 3). Early diagnosis and timely treatment of MCI can delay the disease progression and have important clinical value to improve the prognosis (2–4).

At present, the diagnosis of MCI still relies on subjective clinical symptoms. Objective examination methods are urgently needed in clinical practice. FDG-PET and Amyloid-PET are expensive and need to be exposed to radiation, which limits their usefulness (5). As a medical imaging technique, MRI has the advantages of non-invasive, non-radiation exposure, and high-resolution capabilities, making it widely used in the diagnosis and staging of neurological diseases. As an important part of emotion regulation, the hippocampus and parahippocampal gyrus play key roles in cognitive function, especially emotional memory (6, 7). Recent studies have reported that the morphology and network connectivity changes of the hippocampus and parahippocampal gyrus were important indicators of MCI and AD (8–11). However, these studies mainly focused on macroscopic markers, but overlooked the small structural indicators. Lambin et al. proposed radiomics in 2012, which could help diagnose and differentiate diseases by quantifying the subtle information in medical images that were difficult to assess with the naked eye (12). Radiomics has shown important application value in many neurology diseases, such as PD, AD, epilepsy, and brain tumors (13–17). Previously, a radiomics study by Zhang et al. suggested that 3D textures of the hippocampus and entorhinal cortex might be a diagnostic biomarker for AD (18). Luk et al. used the hippocampus texture features of MRI to predict the conversion of mild cognitive impairment to AD with an accuracy of 76.2% (19). However, most of these literatures used manual methods to segment the brain region and extract relevant parameters, which were time-consuming, taking approximately 4 h to process a patient. These shortcomings limited their clinical use greatly. In this study, we developed a CNN-based artificial intelligence model for the automatic segmentation and radiomics features extraction of bilateral hippocampus and parahippocampal gyrus, and established diagnostic models to help distinguish between MCI and HC in a short time.

Materials and methods

Patient information

All data in this study were collected from the Alzheimer's disease Neuroimaging Initiative (ADNI) database.[1] This study was approved by the ethics standards committee of our institution.

TABLE 1 The demographic data of MCI and HC groups.

	MCI	HC	t/Z/F	p-value
Sample size	115	133	-	-
Gender, female (%)	50 (43.48)	80 (60.15)	6.873	0.009
Age (years)	72.43±7.88	69.13±7.25	−3.441	0.001
Education (years)	16.55±2.36	16.68±2.12	0.480	0.632
MMSE [*M* (Q1, Q3)]	27.75 (27, 29)	29.03 (29, 30)	−5.314	<0.001

MMSE, Mini-Mental State Examination.

Totally 248 subjects with 3D-T1 weighted MR structural images, including 115 MCI patients and 133 normal individuals were included. According to the ADNI protocol, the diagnostic criteria for MCI should meet: (a) Cognitive problems reported by participants or those around them; (b) The patient showed impairment in the subtest logical memory-II on the Wechsler Memory Scale R; (c) Mini-Mental State Examination (MMSE) score ≥ 24. Clinical information was obtained for all participants, including age, sex, education, and MMSE score (Table 1).

Image acquisition and preprocessing

For all participants, 3D-T1-MPRAGE or equivalent protocol with slimly different resolutions was used. ADNI website offered all of the detailed imaging parameters.[2] For scanner 1 (Siemens Medical Solutions, 3.0 T), the scanning parameters were listed below: repetition time (TR) = 2300.0, echo time (TE) = 3.0, matrix = 240 × 256 × 176. For scanner 2 (General Electric Healthcare, 3.0 T), the scanning parameters were: TR = 7.7–7.0, TE = 3.1–2.8, matrix = 256 × 256 × 196. For scanner 3 (Philips Medical Systems, 3.0 T), the MR imaging data were acquired with the following parameters: TR = 6.8, TE = 3.1, matrix = 256 × 256 × 170. The layer thickness of the three different scanners was 1.0 or 1.2 mm, and the layer spacing was 0.

The hippocampus and parahippocampal gyrus segmentation

The hippocampus and parahippocampal gyrus segmentation module was implemented using a deep learning algorithm based on a 3D VB-NET network (20). The data preprocessing module performed a series of operations, including rotation, resampling, resizing, skull stripping, image non-uniform correction, histogram matching, and gray-scale normalization on the MRI images used for training and testing. All images were standardized to the size of 256*256*256*1 mm^3 in the standard Cartesian LPI coordinate system, and the gray-scale range was within the interval (−1, 1). The model was constructed based on 1,800 subjects and evaluation showed an averaged 0.92 Dice overlap with ground truth. The segmentation process took less than 1 minute for each patient.

1 http://adni.loni.usc.edu

2 http://adni.loni.usc.edu/methods/documents/

Radiomics features extraction

Totally 2,264 radiomics features were automatically extracted from the bilateral hippocampus or parahippocampal gyrus of each patient. The radiomics features included four categories of first-order features, shape features, texture features, and wavelet-based features (21). The first-order statistics and shape features could reflect the shape and size of the brain region. Texture features included Gray Level Co-occurrence Matrix (GLCM) features, Gray Level Run Length Matrix (GLRLM) features, Gray Level Size Zone Matrix (GLSZM) features, Neighboring Gray Tone Difference Matrix (NGTDM) features, and Gray Level Dependence Matrix (GLDM) features. The high-level features were obtained through 24 filters (including Box Mean, additive Gaussian Noise, binomial blur, curvature flow, Box-Sigma, normalization, Laplace Sharpening, discrete Gaussian, mean, speck noise, recursive Gaussian, Shot Noise and LoG with sigma values of 0.5, 1, 1.5 and 2), as well as wavelet transformations (LLL, LLH, LHL, LHH, HLL, HLH, HHL, and HHH).

Radiomics features selection, models establishment and validation

All patients were randomly divided into a training group and a testing group in an 8:2 ratio. Four feature selection methods, namely Relief, Minimum Redundancy Maximum Correlation, Recursive Feature Elimination, and LASSO were used to gradually select the optimal radiomics features. Then, five independent machine learning classifiers, including Support Vector Machine (SVM), Random forest (RF), Logistic Regression (LR), Bagging Decision Tree (BDT) and Gaussian Process (GP) algorithm were trained on the training set, and validated on the testing set in the form of 10 fold cross-validation. The flow chart of this study was shown in Figure 1.

Statistics analysis

Statistical analysis was conducted using SPSS software (version 22.0, IBM). Quantitative data was tested for normality using the Kolmogorov–Smirnov method. Continuous variables with normal distribution were expressed as mean standard deviation and compared using independent sample *t*-tests. Continuous variables without normal distribution were expressed as median and compared using the Mann–Whitney U test. Classified variables were expressed in frequency (percentage) and compared using the chi-square test or Fisher's exact test. The statistical significance was considered to be $p < 0.05$. The model performance was evaluated using the receiver operating characteristics (ROC) curve. The area under the curve (AUC), sensitivity, specificity, accuracy, as well as F1 score were calculated. The calibration curve was used to evaluate the calibration of the model, and DCA was used to evaluate the clinical applicability of the model.

Results

Totally 115 MCI patients and 133 healthy controls were included in this study. There was no significant difference in educational level between the MCI and healthy control groups.

In the training set, 200 features were selected from 4,528 radiomic features of bilateral hippocampus by features dimensionality reduction of the Relief, Minimum Redundancy Maximum Correlation and Recursive Feature Elimination methods. Then 13 optimal features were obtained using the LASSO method. According to the same method, 12 optimal features were obtained from the bilateral parahippocampal gyrus. 300 features were selected from radiomic features of both the bilateral hippocampus and parahippocampal gyrus by Relief, Minimum Redundancy Maximum Correlation and Recursive

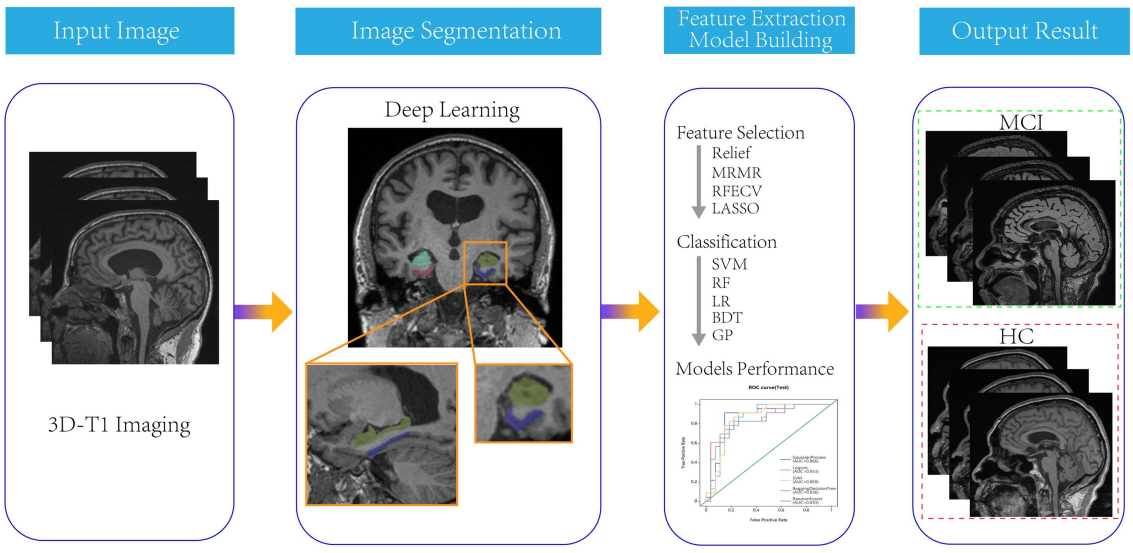


FIGURE 1
The flow chart of segmentation and models construction.

Feature Elimination methods, and then 15 optimal features were obtained as combined features using the LASSO method (Figure 2).

Fifteen models were established based on the optimal features of the hippocampus, parahippocampal gyrus and combined features. ROC curves of GP, LR, SVM, BDT and RF models are shown in Figure 3. The DeLong test showed that GP models based on combined features (11 features from the hippocampus, and 4 features from parahippocampal gyrus) showed the best performance. The AUC of the training set and test set were 0.954 (95% CI: 0.929–0.979) and 0.866 (95% CI: 0.757–0.976), respectively. The sensitivity, specificity, and accuracy of the training set and test set were 0.848, 0.896, 0.874, and 0.870, 0.852, and 0.860, respectively (Table 2). The calibration curve showed a good agreement between the actual and predicted probabilities of the sample (Figure 4). Decision curve analysis showed that the GP model had the highest clinical net benefit (Figure 5).

Discussion

With the aging of the population, the incidence of AD is increasing year by year. It had been proven that AD could be prevented, and the key lied in early detection of mild cognitive impairment (22–24). Therefore, developing a fast and accurate method to distinguish MCI and HC had become an important focus in clinical practice. Previous studies had reported that the morphological changes of hippocampal regions were closely related to the occurrence of MCI (25, 26). In this study, an automatic segmentation framework was established on 3D-T1 (MPRAGE) sequence images based on a 3D VB-NET deep learning model. The bilateral hippocampus and parahippocampal gyrus were automatically segmented and a large number of radiomic features were automatically extracted. We found that among the classifiers of GP, LR, SVM, BDT, and RF algorithms, the GP classifier had the highest classification performance, with an AUC of 0.954 in the training set and 0.866 in the test set. Our results showed that the

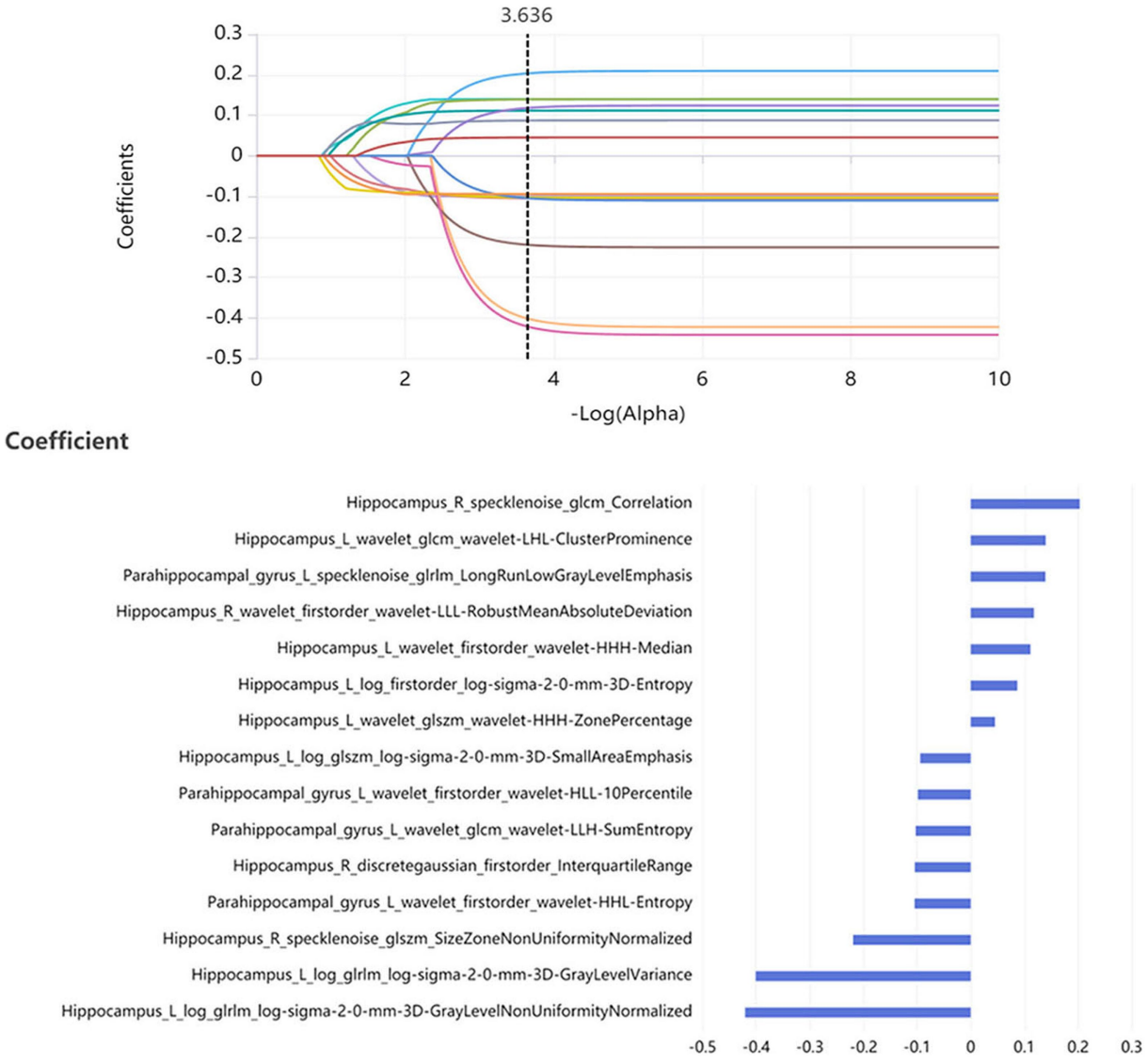


FIGURE 2
15 optimal radiomics features were obtained as the optimal combined features using the LASSO method.

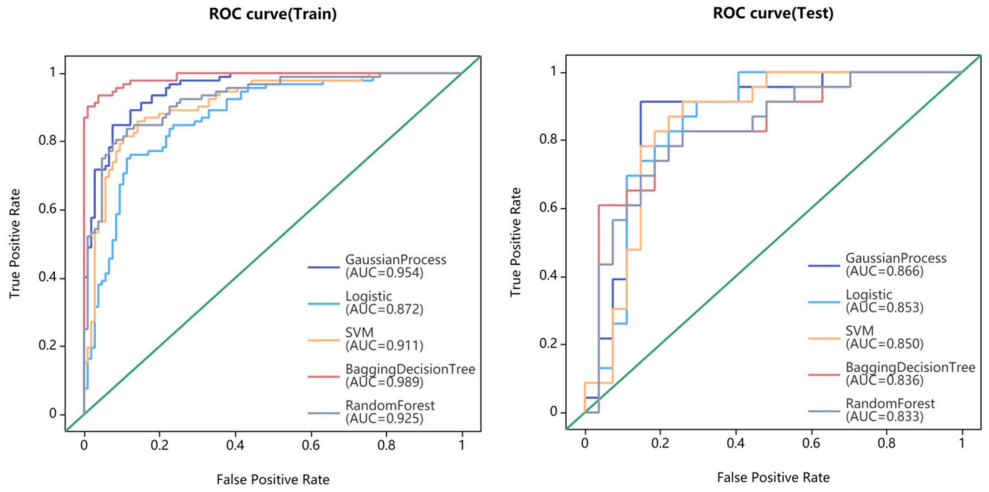


FIGURE 3
The models performance of GP, LR, SVM, BDT, and RF classifiers in the discrimination between MCI and normal controls.

TABLE 2 Performance of GP, BDT, SVM, RF, and LR models on training set and testing set.

		AUC		Sensitivity		Specificity		Accuracy	
	Models	Training set	Testing set	Training set	Testing set	Training set	Testing set	Training set	Testing set
Hippocampus	GP	0.912 (0.873–0.951)	0.839 (0.718–0.960)	0.761	0.783	0.849	0.889	0.808	0.840
	BDT	0.987 (0.974–0.999)	0.855 (0.749–0.961)	0.946	0.739	0.962	0.815	0.955	0.780
	SVM	0.888 (0.843–0.933)	0.810 (0.674–0.946)	0.793	0.739	0.821	0.815	0.808	0.780
	RF	0.873 (0.825–0.921)	0.815 (0.692–0.938)	0.522	0.565	0.925	0.926	0.737	0.760
	LR	0.863 (0.813–0.913)	0.794 (0.655–0.933)	0.750	0.826	0.821	0.778	0.788	0.800
Parahippocampal gyrus	GP	0.938 (0.907–0.968)	0.826 (0.709–0.943)	0.848	0.739	0.877	0.778	0.864	0.760
	BDT	0.968 (0.946–0.990)	0.736 (0.592–0.880)	0.913	0.739	0.877	0.630	0.894	0.680
	SVM	0.845 (0.790–0.900)	0.784 (0.650–0.918)	0.772	0.696	0.811	0.852	0.793	0.780
	RF	0.902 (0.861–0.944)	0.731 (0.586–0.876)	0.696	0.522	0.906	0.815	0.808	0.680
	LR	0.802 (0.741–0.862)	0.684 (0.528–0.840)	0.685	0.652	0.736	0.667	0.712	0.660
Combined features	GP	0.954 (0.929–0.979)	0.866 (0.757–0.976)	0.848	0.870	0.896	0.852	0.874	0.860
	BDT	0.989 (0.979–0.999)	0.836 (0.719–0.952)	0.913	0.739	0.962	0.815	0.939	0.780
	SVM	0.911 (0.870–0.952)	0.850 (0.737–0.964)	0.826	0.783	0.868	0.852	0.848	0.820
	RF	0.925 (0.889–0.961)	0.833 (0.717–0.948)	0.750	0.652	0.953	0.852	0.859	0.760
	LR	0.872 (0.823–0.922)	0.853 (0.740–0.967)	0.772	0.826	0.830	0.778	0.803	0.800

GP, Gaussian Process; BDT, Bagging Decision Tree; SVM, Support Vector Machine; RF, Random forest; LR, Logistic Regression.

application of artificial intelligence technology to analyze raw T1-weighted MRI images could accurately detect MCI from normal controls. This was a very objective method that did not rely on the patient's personal medical history or the doctor's clinical experience. Previously, Ferrarini et al. had used markers based on the shape of the hippocampus to distinguish between AD and MCI with an accuracy of 80% (27). A meta-analysis of the value of hippocampal volume in the diagnosis of MCI showed a sensitivity and specificity of 60 and 75%, respectively (28). Beheshti et al. used a voxel-based morphometric method to construct a structural connection network and a support vector machine (SVM) model to achieve a 70.38% accuracy for MCI and normal control discrimination (29). Feng et al. (30) developed a logistic regression machine learning model to identify MCI from normal control with an accuracy of 0.79 and 0.76 using radiomics features of the hippocampus. In these previous studies, the accuracies were low and many software such as VBM, SPM, Freesufer, 3DSlicer or Python were used to achieve manual brain region segmentation and features extraction, which greatly reduced the work efficiency. Compared to them, our method could achieve fully automated brain segmentation, feature extraction, and diagnostic modeling establishment. Our results were more accurate and the results could be obtained in several minutes. It had the characteristics of objectivity, high speed, low cost, and high accuracy, making it more suitable for clinical application and promotion.

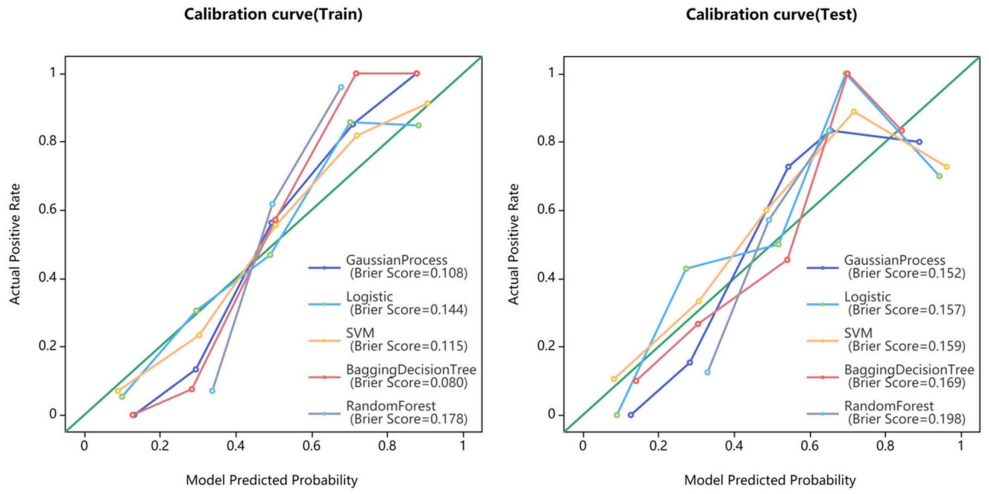


FIGURE 4
Calibration curves for GP, LR, SVM, BDT, and RF models.

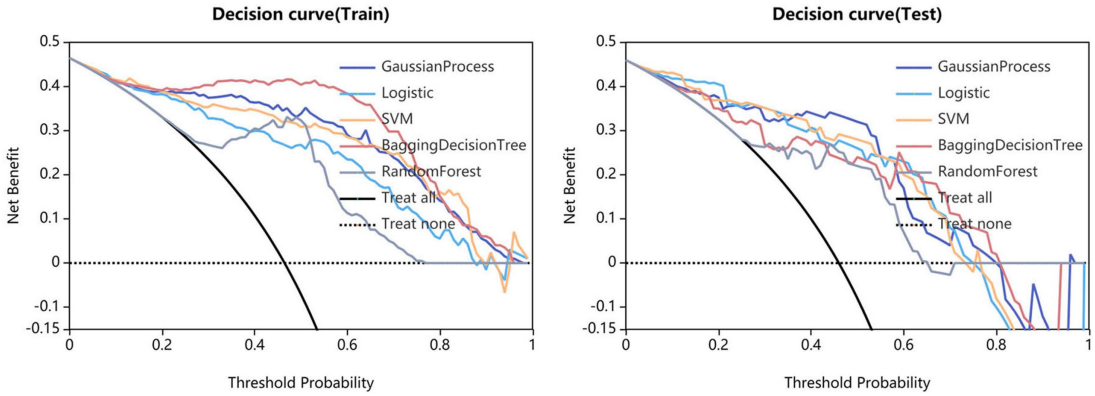


FIGURE 5
Clinical decision curves of GP, LR, SVM, BDT, and RF models.

Radiomics features contain much microstructure information that reflects the underlying early biomarkers of pathophysiology. In this study, the optimal model contained 15 radiomics features, including 11 features from the hippocampus and 4 features from the parahippocampal gyrus. The 11 features from the hippocampus included 4 first-order features, 3 GLSZM features, 2 GLRLM features, and 2 GLCM features. The four radiomic features of the parahippocampal gyrus included 2 first-order features, 1 GLCM features and 1 GLRLM features. The hippocampus is located between the thalamus and the medial temporal lobe of the brain and is part of the limbic system. It is mainly responsible for the storage, conversion, and orientation functions of short-term memory (31). The hippocampus is one of the earliest brain regions affected by Alzheimer's disease. As the disease progresses, hippocampal damage gradually worsens, which can help determine the severity of the disease, monitor the progress of the disease, or evaluate the effectiveness of interventions such as medication, cognitive therapy, and healthy lifestyles (32). The parahippocampal gyrus is an important structure that assists the hippocampus in its function (33). The damage of them can cause abnormalities in emotion, cognition and behavior. The first-order features include mean absolute deviation, kurtosis, energy and minimum, which mainly reflect the basic statistical information of the image from various angles. It could measure the asymmetry and flatness of the morphological layout of the brain regions. Previously, Feng et al. had found hippocampal neuroanatomical abnormalities of size, shape, gray value distribution and spatial heterogeneity in MCI subjects (30). GLSZM, GLRLM, and GLCM belong to texture features. They are based on different grayscale matrices to evaluate the spatial distribution of pixel intensity. These features have been proven to be useful in studying neuropathological heterogeneity. When pathological changes occur in the internal structure of the brain, its smoothness, roughness, and heterogeneity can be reflected through GLSZM, GLRLM, and GLCM features. The texture features of hippocampal microstructure have been proven to reflect cognitive function in direct and indirect ways (9).

Our study had several limitations. Firstly, this was a retrospective cross-sectional study, which did not track the dynamic process of the radiomic features. A prospective longitudinal follow-up study in the

future is needed. Secondly, the sample size of MCI patients was relatively small, and internal cross-validation was adopted; therefore, the generalization of the model needed to be further verified by a larger sample and external validation. Finally, in order to achieve rapid and fully automated diagnosis, this study only considered imaging information. Adding more clinical information and biological indicators could further increase accuracy.

Conclusion

The GP classifier based on 15 radiomics features of bilateral hippocampal and parahippocampal gyrus could detect MCI based on conventional MR images with high accuracy. Our fully automatic model could rapidly process the MRI data and distinguish MCI and HCs in 1 minute. Our method was fast, simple, and accurate, which provided important clinical value in assisted diagnosis.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

Ethics statement

The studies involving humans were approved by Alzheimer's disease Neuroimaging Initiative. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

MY: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – original draft. SM: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Writing – original draft. FW: Conceptualization, Data curation, Formal analysis, Writing – original draft. FS: Data curation, Formal analysis, Methodology, Software, Supervision, Validation, Writing – original draft. YX: Methodology, Project administration, Software, Supervision, Validation, Writing – original draft. JF: Data curation, Formal analysis, Methodology, Supervision, Writing – original draft. JZ: Data curation, Investigation, Methodology, Writing – original draft. CL: Conceptualization, Funding acquisition, Investigation, Resources, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Fundamental Research Funds for the Central Universities (Project No. 2022CDJYGRH-004); Natural Science Foundation Project of Chongqing, (cstc2020jcyj-msxm0030); Chongqing medical scientific research project (Joint project of Chongqing Health Commission and Science and Technology Bureau) (2022QNXM013).

Conflict of interest

FS and YX were employed by Shanghai United Imaging Intelligence, Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Soria Lopez JA, González HM, Léger GC. Alzheimer's disease. *Handb Clin Neurol*. (2019) 167:231–55. doi: 10.1016/B978-0-12-804766-8.00013-3

2. Jongsiriyanyong S, Limpawattana P. Mild cognitive impairment in clinical practice: a review article. *Am J Alzheimers Dis Other Dement*. (2018) 33:500–7. doi: 10.1177/1533317518791401

3. Anderson N. D. State of the science on mild cognitive impairment (MCI). *CNS Spectr*. (2019) 24:78–87. doi: 10.1017/S1092852918001347

4. Suh CH, Shim WH, Kim SJ, Roh JH, Lee JH, Kim MJ. Development and validation of a deep learning-based automatic brain segmentation and classification algorithm for Alzheimer disease using 3D T1-weighted volumetric images. *AJNR Am J Neuroradiol*. (2020) 41:2227–34. doi: 10.3174/ajnr.A6848

5. Albert MS, Dekosky ST, Dickson D, Dubois B, Feldman HH, Fox NC. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. (2011) 7:270–9. doi: 10.1016/j.jalz.2011.03.008

6. Peng X, Burwell RD. Beyond the Hippocampus: the role of Parahippocampal-prefrontal communication in context-modulated behavior. *Neurobiol Learn Mem*. (2021) 185:107520. doi: 10.1016/j.nlm.2021.107520

7. Rolls ET. The cingulate cortex and limbic systems for action, emotion, and memory. *Handb Clin Neurol*. (2019) 166:23–37. doi: 10.1016/B978-0-444-64196-0.00002-9

8. Wei Y, Huang N, Liu Y, Zhang X, Wang S, Tang X. Hippocampal and amygdalar morphological abnormalities in Alzheimer's disease based on three Chinese MRI datasets. *Curr Alzheimer Res*. (2020) 17:1221–31. doi: 10.2174/1567205018666210218150223

9. Du Y, Zhang S, Qiu Q, Zhang J, Fang Y, Zhao L. The effect of hippocampal Radiomic features and functional connectivity on the relationship between hippocampal volume and cognitive function in Alzheimer's disease. *J Psychiatr Res*. (2023) 158:382–91. doi: 10.1016/j.jpsychires.2023.01.024

10. Liu J, Zhang X, Yu C, Duan Y, Zhuo J, Cui Y. Impaired Parahippocampus connectivity in mild cognitive impairment and Alzheimer's disease. *J Alzheimers Dis*. (2016) 49:1051–64. doi: 10.3233/JAD-150727

11. Gong L, Xu R, Lan L, Liu D, Shen J, Zhang B. The CD33 genotype associated cognitive performance was bidirectionally modulated by intrinsic functional connectivity in the Alzheimer's disease spectrum. *Biomed Pharmacother*. (2019) 115:108903. doi: 10.1016/j.biopha.2019.108903

12. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036

13. Sotoudeh H, Sarrami AH, Roberson GH, Shafaat O, Sadaatpour Z, Rezaei A. Emerging applications of Radiomics in neurological disorders: a review. *Cureus*. (2021) 13:E20080. doi: 10.7759/cureus.20080

14. Feng Q, Niu J, Wang L, Pang P, Wang M, Liao Z. Comprehensive classification models based on amygdala Radiomic features for Alzheimer's disease and mild cognitive impairment. *Brain Imaging Behav*. (2021) 15:2377–86. doi: 10.1007/s11682-020-00434-z

15. Bevilacqua R, Barbarossa F, Fantechi L, Fornarelli D, Paci E, Bolognini S, et al. Radiomics and artificial intelligence for the diagnosis and monitoring of Alzheimer's disease: a systematic review of studies in the field. *J Clin Med*. (2023) 12:5432. doi: 10.3390/jcm12165432

16. Kang JJ, Chen Y, Xu GD, Bao SL, Wang J, Ge M, et al. Combining quantitative susceptibility mapping to radiomics in diagnosing Parkinson's disease and assessing cognitive impairment. *Eur Radiol*. (2022) 32:6992–7003. doi: 10.1007/s00330-022-08790-8

17. Jie B, Hongxi Y, Ankang G, Yida W, Guohua Z, Xiaoyue M, et al. Radiomics nomogram improves the prediction of epilepsy in patients with gliomas [J]. *Front Oncol*. (2022) 12:856359. doi: 10.3389/fonc.2022.856359

18. Zhang J, Yu C, Jiang G, Liu W, Tong L. 3D texture analysis on MRI images of Alzheimer's disease [J]. *Brain Imaging Behav*. (2012) 6:61–9. doi: 10.1007/s11682-011-9142-3

19. Luk CC, Ishaque A, Khan M, Ta D, Chenji S, Yang YH, et al. Alzheimer's disease: 3-dimensional MRI texture for prediction of conversion from mild cognitive impairment. *Alzheimers Dement*. (2018) 10:755–63. doi: 10.1016/j.dadm.2018.09.002

20. Wei J, Shi F, Cui Z, Yongsheng P, Yong X, Dinggang S. Consistent segmentation of longitudinal brain MR images with Spatio-temporal constrained networks; proceedings of the medical image computing and computer assisted intervention. *MICCAI*. (2021) 2021:89–98. doi: 10.1007/978-3-030-87193-2_9

21. Zhang X, Zhang Y, Zhang G, Qiu X, Tan W, Yin X, et al. Deep learning with Radiomics for disease diagnosis and treatment: challenges and potential. *Front Oncol*. (2022) 12:773840. doi: 10.3389/fonc.2022.773840

22. Crous-Bou M, Minguillón C, Gramunt N, Molinuevo JL. Alzheimer's disease prevention: from risk factors to early intervention. *Alzheimers Res Ther*. (2017) 9:71. doi: 10.1186/s13195-017-0297-z

23. Stefaniak O, Dobrzyńska M, Drzymała-Czyż S, Przysławski J. Diet in the prevention of Alzheimer's disease: current knowledge and future research requirements. *Nutrients*. (2022) 14:4564. doi: 10.3390/nu14214564

24. Valenzuela PL, Castillo-García A, Morales JS, de la Villa P, Hampel H, Emanuele E, et al. Exercise benefits on Alzheimer's disease: state-of-the-science. *Ageing Res Rev*. (2020) 62:101108. doi: 10.1016/j.arr.2020.101108

25. Miao D, Zhou X, Wu X, Chen C, Tian L. Hippocampal morphological atrophy and distinct patterns of structural covariance network in Alzheimer's disease and mild cognitive impairment. *Front Psychol*. (2022) 13:980954. doi: 10.3389/fpsyg.2022.980954

26. Lee P, Ryoo H, Park J, Jeong YAlzheimer's Disease Neuroimaging Initiative. Morphological and microstructural changes of the Hippocampus in early MCI: a study utilizing the Alzheimer's disease neuroimaging initiative database. *J Clin Neurol*. (2017) 13:144–54. doi: 10.3988/jcn.2017.13.2.144

27. Ferrarini L, Frisoni GB, Pievani M, Reiber JHC, Ganzola R, Milles J. Morphological hippocampal markers for automated detection of Alzheimer's disease and mild cognitive impairment converters in magnetic resonance images. *J Alzheimers Dis*. (2009) 17:643–59. doi: 10.3233/JAD-2009-1082

28. Park HY, Suh CH, Heo H, Shim WH, Kim SJ. Diagnostic performance of hippocampal volumetry in Alzheimer's disease or mild cognitive impairment: a meta-analysis. *Eur Radiol*. (2022) 32:6979–91. doi: 10.1007/s00330-022-08838-9

29. Beheshti I, Maikusa N, Daneshmand M, Matsuda H, Demirel H, Anbarjafari G, et al. Classification of Alzheimer's disease and prediction of mild cognitive impairment conversion using histogram-based analysis of patient-specific anatomical brain connectivity networks. *J Alzheimers Dis*. (2017) 60:295–304. doi: 10.3233/JAD-161080

30. Feng Q, Song Q, Wang M, Pang PP, Liao Z, Jiang H, et al. Hippocampus Radiomic biomarkers for the diagnosis of amnestic mild cognitive impairment: a machine learning method. *Front Aging Neurosci*. (2019) 11:323. doi: 10.3389/fnagi.2019.00323

31. Nemati SS, Sadeghi L, Dehghan G, Sheibani N. Lateralization of the hippocampus: a review of molecular, functional, and physiological properties in health and disease. *Behav Brain Res*. (2023) 454:114657. doi: 10.1016/j.bbr.2023.114657

32. Broadhouse KM, Singh MF, Suo C, Gates N, Wen W, Brodaty H, et al. Hippocampal plasticity underpins long-term cognitive gains from resistance exercise in MCI. *Neuroimage Clin*. (2020) 25:102182. doi: 10.1016/j.nicl.2020.102182

33. LaFlamme EM, Waguespack HF, Forcelli PA, Malkova L. The parahippocampal cortex and its functional connection with the hippocampus are critical for nonnavigational spatial memory in macaques. *Cereb Cortex*. (2021) 31:2251–67. doi: 10.1093/cercor/bhaa358

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 02 February 2024
DOI 10.3389/fmed.2024.1336640

Check for updates

OPEN ACCESS

EDITED BY
Jie-Zhi Cheng,
Shanghai United Imaging Intelligence, Co., Ltd., China

REVIEWED BY
Xiaoming Liu,
Jilin University, China
Baojun Xie,
Renmin Hospital of Wuhan University, China
Xuechun Wang,
Huazhong University of Science and Technology, China

*CORRESPONDENCE
Feng Li
✉ xfkite@163.com
Xiaomin Qin
✉ 13487120205@163.com

†These authors share first authorship

RECEIVED 11 November 2023
ACCEPTED 15 January 2024
PUBLISHED 02 February 2024

CITATION
Wu F, Zhang R, Li F, Qin X, Xing H, Lv H, Li L and Ai T (2024) Radiomics analysis based on multiparametric magnetic resonance imaging for differentiating early stage of cervical cancer.
Front. Med. 11:1336640.
doi: 10.3389/fmed.2024.1336640

COPYRIGHT
© 2024 Wu, Zhang, Li, Qin, Xing, Lv, Li and Ai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Radiomics analysis based on multiparametric magnetic resonance imaging for differentiating early stage of cervical cancer

Feng Wu[1†], Rui Zhang[1†], Feng Li[1*], Xiaomin Qin[2*], Hui Xing[2], Huabing Lv[2], Lin Li[2] and Tao Ai[3]

[1]Department of Radiology, Xiangyang Central Hospital, Affiliated Hospital of Hubei University of Arts and Science, Xiangyang, China, [2]Department of Obstetrics and Gynaecology, Xiangyang Central Hospital, Affiliated Hospital of Hubei University of Arts and Science Xiangyang, China, [3]Department of Radiology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

Objective: To investigate the performance of multiparametric magnetic resonance imaging (MRI)—based radiomics models in differentiating early stage of cervical cancer (Stage I-IIa vs. IIb-IV).

Methods: One hundred patients with cervical cancer who underwent preoperative MRI between June 2020 and March 2022 were retrospectively enrolled. Training ($n = 70$) and testing cohorts ($n = 30$) were assigned by stratified random sampling. The clinical and pathological features, including age, histological subtypes, tumor grades, and node status, were compared between the two cohorts by *t*-test or chi-square test. Radiomics features were extracted from each volume of interest (VOI) on T2-weighted images (T2WI) and apparent diffusion coefficient (ADC) maps. The data balance of the training cohort was resampled by synthesizing minority oversampling techniques. Subsequently, the adiomics signatures were constructed by the least absolute shrinkage and selection operator algorithm and minimum-redundancy maximum-relevance with 10-fold cross-validation. Logistic regression was applied to predict the cervical cancer stages (low [I-IIa]) and (high [IIb–IV] FIGO stages). The receiver operating characteristic curve (area under the curve [AUC]) and decision curve analysis were used to assess the performance of the radiomics model.

Results: The characteristics of age, histological subtypes, tumor grades, and node status were not significantly different between the low [I-IIa] and high [IIb–IV] FIGO stages ($p > 0.05$ for both the training and test cohorts). Three models based on T2WI, ADC maps, and the combined were developed based on six radiomics features from T2WI and three radiomics features from ADC maps, with AUCs of 0.855 (95% confidence interval [CI], 0.777–0.934) and 0.823 (95% CI, 0.727–0.919), 0.861 (95% CI, 0.785–0.936) and 0.81 (95% CI, 0.701–0.918), 0.934 (95% CI, 0.884–0.984) and 0.902 (95% CI, 0.832–0.972) in the training and test cohorts.

Conclusion: The radiomics models combined T2W and ADC maps had good predictive performance in differentiating the early stage from locally advanced cervical cancer.

KEYWORDS

Radiomics, magnetic resonance imaging, cervical cancer, treatment, multiparametric

1 Introduction

Cervical cancer (CC) is one of the most common malignant among women worldwide and also the second leading cause of cancer deaths among women in China (1–3). Although evidence shows that the incidence of CC in developed countries is declining (3, 4), the age-standardized morbidity and mortality of CC in China have shown a significant upward trend (3). In China, the incidence rate of CC has increased from 10 to 40% over the past 30 years (3).

International Federation of Obstetrics and Gynecology (FIGO) staging of CC has always been the staging system commonly used in clinical diagnosis and treatment. CC is primarily managed by surgical treatment or radiotherapy, with chemotherapy as a valuable adjunct. Surgery is the first choice for treating stage IA1, IA2, IB1, IB2, and IIA1 lesions (5). Concurrent chemoradiation is the standard treatment for stages IB3, IIA2, III, and IV diseases. A radical trachelectomy can be performed for young women interested in preserving their fertility, and it is suitable for stage IA2–IB1 tumors with a maximum diameter of no more than 2 cm (5). Given the excellent prognosis of early-stage CC and that as many as 40% of the women affected by these tumors are of reproductive age, fertility-sparing surgery has become a priority (6).

Magnetic resonance imaging (MRI) can noninvasively assess tumor size and the extent of invasion owing to its merits of multiparametric and multidirectional imaging with high soft tissue resolution (7, 8). Therefore, imaging is complementary to clinical assessment with MRI, which is accepted as the optimal modality for staging CC. Important information about the morphology and extent of interstitial invasion of CC can be obtained from T2-weighted (T2W) images (9). The apparent diffusion coefficient (ADC) maps provide information about water fluidity and tissue cell structure to characterize cancer quantitatively (10).

The heterogeneity of CC is inconsistent among different FIGO stages, histological subtypes, and tumor grades (9). It is an essential factor that can predict tumor aggressiveness and could also be reflected in MRI. However, these heterogeneities may be considered similar just by visual assessment of MRI with the radiologist's naked eye. Radiomics is an evolving field that involves extracting many quantitative features from images, such as MRI, computed tomography, and ultrasound, and using a high-throughput process that effectively transforms images into quantitative data to provide more valuable information (11). A series of quantitative features that have been generated can be further used to measure intra-tumor heterogeneity. Currently, several recent studies have described the use of radiomics in CC, mostly on clinicopathological characteristics (9, 12), parametrial invasion (13), pelvic lymph node metastases (14, 15), and predictive performance (16, 17). However, few studies have assessed the performance of radiomics in predicting the stage of cervical cancer, which essentially influences treatment decision-making in the clinical setting. Thus, this study aimed to investigate the predictive performance of multiparametric MRI-based radiomics models in differentiating the low (I-IIa) and high (IIb–IV) FIGO stages of cervical cancer.

2 Materials and methods

2.1 Patients enrollment

This study was approved by the institutional ethics review board of our hospital (approval no. 2022–027), and the informed consent requirement was waived due to the retrospective study. The patients were enrolled through the following inclusion criteria: (a) patients with histologically confirmed CC; (b) patients who have not undergone therapy (neoadjuvant chemotherapy, radiotherapy, or conization) before MRI examination; (c) patients undergoing T2-weighted imaging (T2WI) with fat suppression, and DWI with ADC maps; and (d) classify the cases based on the 2018 FIGO system. The exclusion criteria were as follows: (a) patients with tumors that were too small for the region of interest (ROI) to be accurately drawn, (b) patients with poor MRI image quality resulting from artifacts, (c) patients with incomplete clinicopathological data, and (d) patients with rare histological subtypes. Between June 2020 and March 2022, 100 patients with CC participated in our study. According to the proportion of 7:3, the 100 patients were randomly divided into two independent cohorts, a training and a test cohort (Figure 1).

All patients' clinical and pathologic features, including age, treatment, FIGO stage, pathological information, and serum squamous cell carcinoma antigen (SCC-Ag) levels before treatment, were derived from the medical records. The treatments were divided into surgical and non-surgical treatments. FIGO stages were dichotomized into low (I-IIa) and high (IIb–IV) FIGO stages. The assessed pathological information comprised histological subtypes, tumor grades, invasion depth, and lymphovascular space invasion (LVSI) according to the World Health Organization Classification of Tumors of Female Reproductive Organs. There are two histological subtypes of squamous cell carcinoma (SCC) and adenocarcinoma (ACA). Tumor grades were divided into three groups: well (G1), moderately (G2), and poorly differentiated (G3). Invasion depth was classified into inner, middle, and outer layers. After reviewing the MRI of all patients, the node status was recorded by two radiologists with five and more than ten years of experience in gynecological cancer diagnosis, respectively. Any disagreements were resolved by discussion and consensus. Nodal status was based on T2WI. The positive lymph node was defined as the short axis of the lymph node >10 mm, spiculated or lobulated margin, or internal necrosis (9).

2.2 MRI acquisition

All preoperative MR examinations were performed with a 3.0 T platform with respiratory gating technology and an eight-channel phased array body coil (Siemens Medical Solutions, Verio 3.0, Germany). All recruited patients underwent T2W fat-suppressed and diffusion-weighted imaging (DWI) sequences acquired before surgery or chemoradiation. The ADC was calculated according to the traditional single exponential model, and patients were advised to fast for 5–6 h before examination (9). Conventional MRI comprised oblique axial T2W images (echo time [TE], 82 ms; repetition time [TR], 3,800 ms; gap, 2 mm; slice thickness, 5 mm; field of view [FOV], 320 × 320 mm) with fat suppression and transverse DWI (TE, 52 ms; TR, 3900 ms; gap, 2 mm; slice thickness, 5 mm; FOV, 320 × 256 mm; and b values, 50 and 800 s/mm^2).

2.3 Image segmentation and radiomics feature extraction

The solid lesions' three-dimensional volumes of interest (VOIs) were manually segmented using ITK-SNAP (version 3.8.0), a free and

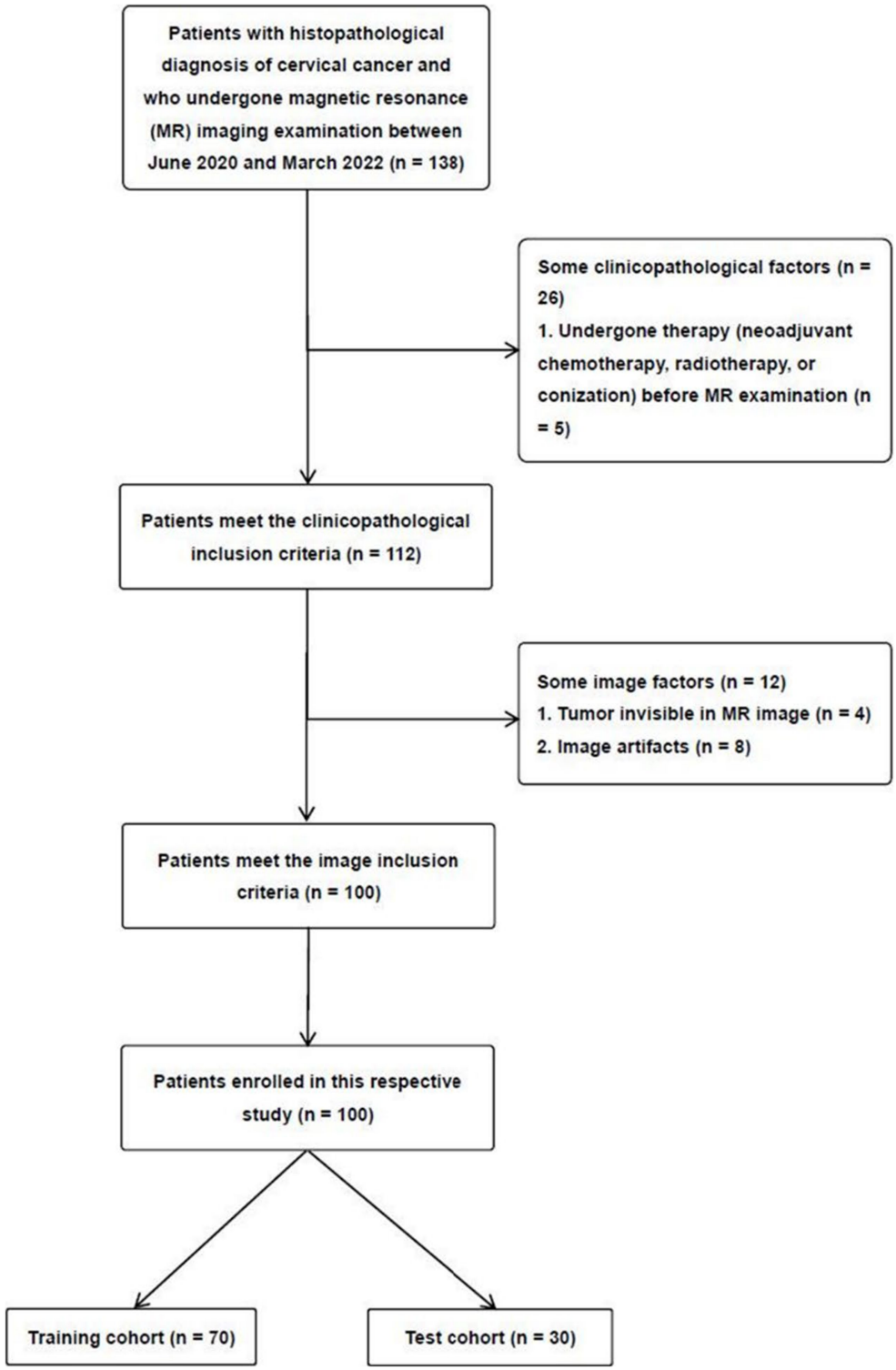


FIGURE 1
Schematic diagram for patient selection.

open-source software. A junior radiologist (with five years of experience in diagnosing gynecologic cancer) manually delineated the low-signal rim of the tumor from adjacent normal tissue, excluding high-signal areas within the lesion, on high-spatial-resolution axial T2W images. The VOI segmentation was performed on DWI with a b value of 800 s/mm^2 and then mapped into the ADC image. All segmented VOIs were confirmed and corrected by a senior radiologist (with >10 years' experience in gynecological tumor diagnosis). The radiologists were blinded to the clinicopathological results. Another junior radiologist (three years of experience in diagnosing gynecological diseases) independently performed manual segmentation of these lesions to analyze interobserver reproducibility.

The radiologists performed manual segmentation blinded to diagnostic information such as clinical and histopathology.

Python (version 3.7.5) with the PyRadiomics package (https://github.com/AIM-Harvard/pyradiomics.git, version 3.0.1) extracted radiomics features from T2W and ADC images. All radiomics features were extracted from the original image and wavelet-filtered image, which could be divided into three groups: 18 first-order statistics, 14 shape features, and 75 textural features. The feature extraction method is provided on an official website.[1] Before radiomics feature extraction, all T2W and ADC maps were normalized using the z-score method and voxel size resampling by $1 \times 1 \times 1$ mm. Finally, 851 features are extracted in each VOI of the T2W and ADC images, respectively.

2.4 Feature selection and radiomics model building

The interobserver reproducibility of each radiomics feature was assessed using the interclass correlation coefficient (ICC). The ICC > 0.80 was considered excellent and included in subsequent analyses. The synthetic minority oversampling technique (SMOTE) method dealt with the balance of balanced categories in the training cohort to prevent bias in the construction of the predictive model because the sample size of the high FIGO stage was 33.3% less than the sample size of the low FIGO stage. The training dataset used for model building analysis was the training dataset after SMOTE processing (training-SMOTE cohort). The low/high FIGO stage of CC patients were 1:1 (42 low FIGO stage patients and 42 high FIGO stage patients) in the SMOTE-training cohort.

To reduce redundancy, spearman correlation analysis was used to eliminate features with a Spearman correlation coefficient > 0.9. The top 10 features with low redundancy and high correlation with CC were selected for the following analysis using the minimum redundancy maximum correlation (mRMR) algorithm. Then, the least absolute shrinkage and selection operator (LASSO) method was used to screen the radiomics features that helped predict the CC therapy method in the training cohort with SMOTE. A total of 10 cross-validation methods were used to identify the model's generalization performance in the LASSO method. The single radiomics model using T2W images (T2 model) and ADC maps (ADC model) was weighted using coefficients of selected features with optimal adjustment weights in LASSO logistic regression. The combined model was developed based on multivariate analysis's T2 and ADC models.

2.5 Statistical analyses

We conducted differences analysis of the characteristics of patients between the training and test cohorts, using the chi-square test for categorical variables and Student's *t*-test for continuous variables. The interobserver reproducibility of the radiomics features evaluating the interobserver agreement among radiologists was accessed using the ICC, with a coefficient greater than 0.8, indicating good reproducibility. Receiver operating characteristic (ROC) curve analysis was used to assess the diagnostic performance of the radiomics models for predicting CC. The optimal cutoff value for predictive diagnosis for radiomics models was determined by maximizing the Youden index in the training cohort with SMOTE. The areas under the ROC (AUC), accuracies, specificities, sensitivities, negative predictive values, and positive predictive values were used to quantify the diagnostic performance of the radiomics models. Decision curve analysis (DCA) was used to assess the clinical usefulness of the models. This study's statistical analysis was performed using R (version 3.6.1, https://www.r-project.org). $p < 0.05$ (two-tailed) was considered to be statistically significant.

3 Results

3.1 Clinical characteristics

The clinical characteristics of the patients are summarized in Table 1. In the invasion depth, LVSI, and tumor grade groups, the pathological information of some patients not undergoing surgery (such as those receiving chemotherapy) was missing. In 100 patients with CC (mean age, 53.48 ± 10.58 years), 90 and 10 had SCC and ACA, respectively. All patients in the training and test cohorts were further divided into the low ($n = 61$) and high ($n = 39$) FIGO stage cohorts. The rates of low FIGO stages in the training and test cohorts remained balanced (0.600 and 0.633, respectively, $p = 0.374$). The clinical and pathologic characteristics, including age and histological subtype, were not significantly different between the two cohorts ($p > 0.05$). Furthermore, the MR-reported nodal status was not significantly different between the low and high FIGO stage cohorts ($p = 0.385$).

3.2 Radiomics model construction

The low/high FIGO stage cohort in the training cohort was converted from 42/28 to 42/42 using the SMOTE method. In total, 221 features were screened from each VOI in a T2W image, and the features were reduced to six CC-related features after the application of the mRMR and LASSO algorithms in the training-SMOTE cohort (Figures 2A–C). Similarly, the 230 ADC radiomics features were reduced to three imaging biomarkers after applying the mRMR and LASSO algorithms in the training-SMOTE cohort (Figures 2D–F). The ICC ranges for T2W and ADC image radiomics features were 0.34–0.99 and 0.21–0.99, respectively. The T2 and ADC model calculation formulae were as follows:

T2 radiomics signature =
−1.624 − 5.923 × wavelet.HHH_firstorder_Uniformity
−3.922 × wavelet.LLH_glcm_ClusterTendency
−2.531 × wavelet.HHL_glcm_Correlation
−0.163 × wavelet.HHL_glszm_LowGrayLevelZoneEmphasis
+0.271 × original_shape_Flatness
+0.660 × wavelet.
HHH_glszm_SmallAreaHighGrayLevelEmphasis

ADC radiomics signature =
0.431 + 0.431 × wavelet.LHL_glcm_Imc1
+0.523 × wavelet.HLL_glcm_Idn
+0.782 × wavelet.HLL_glszm_SmallAreaEmphasis

1 https://pyradiomics.readthedocs.io/en/latest

TABLE 1 Clinical characteristics of the patients with CC in the training and test cohorts.

Characteristics	Overall	Training cohort	Test cohort	*p*	SMD	Missing
	100	70	30			
Age (mean [SD])	53.480 (10.580)	55.260 (10.610)	50.980 (11.250)	0.380	0.070	0
Nodal status				0.385	0.255	0
Positive (%)	24 (24.000)	19 (27.143)	5 (16.667)			
Negative (%)	76 (76.000)	51 (72.857)	25 (83.333)			
Before SCC (median [IQR])	1.960 (0.880–5.850)	2.040 (0.850–5.910)	3.080 (1.270–4.890)	0.958	0.041	0
Neoadjuvant chemotherapy				0.269	0.316	0
Positive (%)	22 (22.000)	18 (25.714)	4 (13.333)			
Negative (%)	78 (78.000)	52 (74.286)	26 (86.667)			
Surgery				0.374	0.243	0
Positive (%)	68 (68.000)	50 (71.429)	18 (60.000)			
Negative (%)	32 (32.000)	20 (28.571)	12 (40.000)			
Histological subtype				0.716	0.153	0
Adenocarcinoma (%)	10 (10.000)	6 (8.571)	4 (13.333)			
Squamous cell carcinoma (%)	90(90.000)	64(91.429)	26 (86.667)			
Tumor grades (%)				0.393	0.322	18
G1	6 (6.000)	3 (4.286)	3 (10.000)			
G2	15 (15.000)	12 (17.143)	3 (10.00)			
G3	61 (61.00)	43 (61.428)	18 (60.000)			
Depth (%)				0.801	0.164	22
Inner	23 (23.000)	15 (21.428)	8 (26.667)			
Middle	15 (15.000)	11 (15.714)	4 (13.333)			
Outer	40 (40.000)	25 (35.714)	15 (50.000)			
LVSI				0.954	0.076	22
Positive (%)	42 (42.000)	29 (41.428)	13 (43.333)			
Negative (%)	36 (36.000)	26 (37.143)	10 (33.333)			
FIGO stage (%)				0.337	0.502	0
I	42 (42.000)	30 (42.857)	12 (40.000)			
IIA	19 (19.000)	12 (17.142)	7 (23.333)			
IIB	8 (8.000)	4 (5.714)	4 (13.333)			
III	27 (27.000)	20 (28.571)	7 (23.333)			
IV	4 (4.000)	4 (5.714)	0 (0.000)			
FIGO group				0.374	0.056	
Low FIGO (I-IIa)	61 (61.000)	42 (60.000)	19 (63.333)			
High FIGO (IIb-IV)	39 (39.000)	28 (40.000)	11 (36.667)			

SD, standard deviation; SMD, standardized mean difference; IQR, interquartile range; LVSI, lymphovascular space invasion; FIGO, international federation of obstetrics and gynecology.

3.3 Radiomics model performance

In the training cohort, the AUCs of the T2W, ADC, and combined models of predicting CC were 0.823 (95% CI, 0.727–0.919), 0.810 (95%CI, 0.701–0.918), and 0.902 (95% CI, 0.832–0.972), respectively. In the test cohort, the AUCs were 0.829 (95% CI, 0.658–0.999), 0.773 (95% CI, 0.578–0.969), and 0.856 (95% CI, 0.707–1.000). The sensitivities of the three models for predicting CC were 0.929 and 0.889, 0.762 and 0.722, 0.857 and 0.833 in the training and test cohorts. In the training and test cohorts, the specificities of the three models were 0.536 and 0.583, 0.714 and 0.667, and 0.857 and 0.833. The accuracies were 0.771 and 0.767, 0.743 and 0.700, and 0.857 and 0.833 in the training and test cohorts. The AUCs, accuracies, sensitivities, and specificities of the three models are shown in Table 2. Figure 3 shows the ROC curves of the three models. The results showed that the combined model had better diagnostic

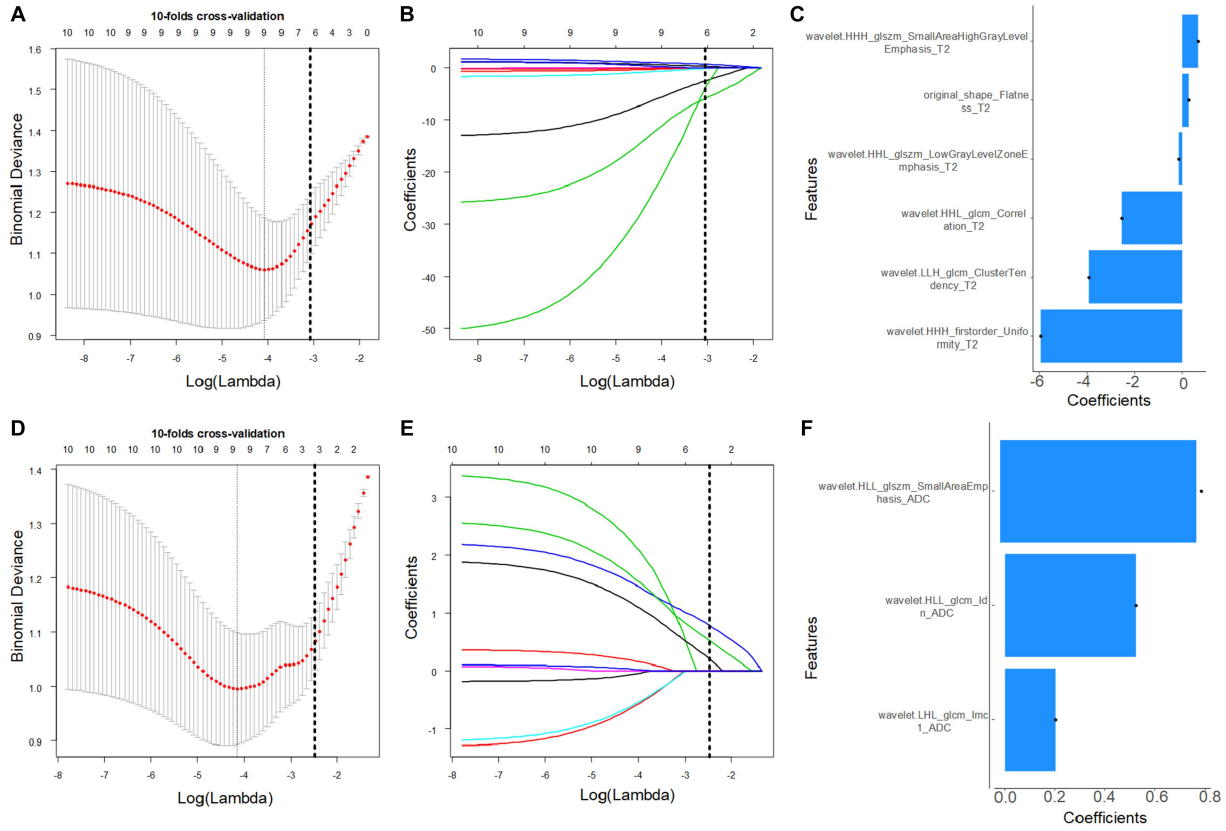


FIGURE 2
Magnetic resonance-based radiomics feature selection using the least absolute shrinkage and selection operator (LASSO) method in the training cohort. **(A,D)** The optimal penalty coefficient lambda (λ) for the feature of the T2W **(A)** and apparent diffusion coefficient (ADC) **(D)** images was obtained based on 10-fold cross-validation. **(B,E)** Changes in the corresponding coefficients of T2W and ADC image features during Lasso analysis. The vertical dashed line represents the optimal λ, corresponding to six (T2) and three (ADC) nonzero feature coefficients.

TABLE 2 **Performance of the sequences models.**

		AUC	SEN	SPE	ACC	NPV	PPV
T2 model	Training cohort	0.823 (0.727–0.919)	0.929 (0.833–1.000)	0.536 (0.357–0.714)	0.771 (0.656–0.863)	0.833 (0.661–1.000)	0.750 (0.632–0.868)
	Test cohort	0.787 (0.604–0.969)	0.889 (0.722–1.000)	0.583 (0.333–0.833)	0.767 (0.577–0.901)	0.778 (0.506–1.000)	0.761 (0.579–0.944)
ADC model	Training cohort	0.810 (0.701–0.918)	0.762 (0.619–0.881)	0.714 (0.536–0.857)	0.743 (0.624–0.840)	0.667 (0.624–0.840)	0.800 (0.676–0.924)
	Test cohort	0.806 (0.634–0.977)	0.722 (0.500–0.889)	0.667 (0.417–0.917)	0.700 (0.506–0.853)	0.615 (0.807–0.993)	0.764 (0.563–0.996)
Combined model	Training cohort	0.902 (0.832–0.972)	0.857 (0.738–0.952)	0.857 (0.714–0.964)	0.857 (0.753–0.929)	0.800 (0.657–0.943)	0.900 (0.807–0.993)
	Test cohort	0.852 (0.683–1.000)	0.833 (0.667–1.000)	0.833 (0.583–1.000)	0.833 (0.653–0.944)	0.769 (0.540–0.998)	0.882 (0.729–1.000)

AUC, area under the curve; SEN, sensitivity; SPE, specificity; ACC, accuracy; PPV, positive predictive value; NPV, negative predictive value.

and predictive performance than the T2W and ADC models alone. The DCA was applied to evaluate the clinical usefulness, showing that the combined model could provide benefits in the training cohort (Figure 4A), with the threshold probability greater than 0.200, and in the test cohort (Figure 4B), with the threshold probability between 0.150 and 0.850.

4 Discussion

In our study, we successfully constructed a radiomics model for the preoperative prediction of CC therapy (surgical and non-surgical treatment). The radiomics model incorporated the ADC map and T2W radiomics signature, and the results were validated in the test

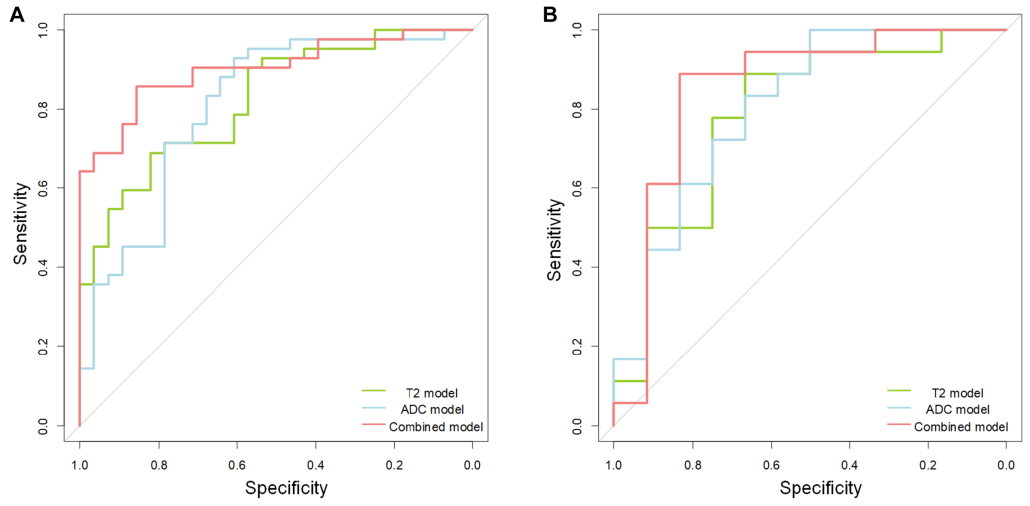


FIGURE 3
The receiver operating characteristic (ROC) curves of the T2 (green line), apparent diffusion coefficient (ADC) (blue line), and combined (red line) model. The receiver operating characteristic curve shows the combined model is better than the separate T2 and ADC model in the training **(A)** and test **(B)** cohorts.

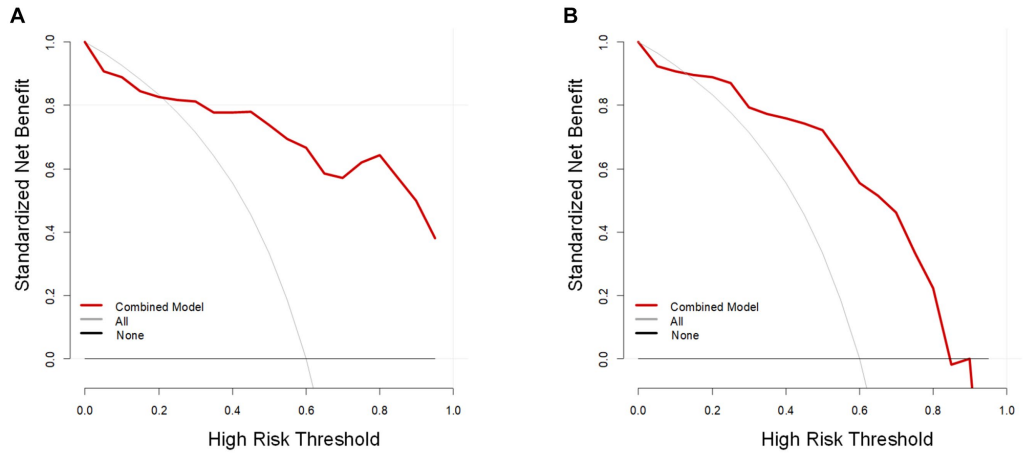


FIGURE 4
The decision curve analysis (DCA) of the combined model (red line). The vertical and horizontal axes represent net benefit and threshold probability, respectively. The DCA revealed that the combined model could provide benefits in the training cohort **(A)**, with the threshold probability between 0.200 and 1, and in the test cohort **(B)**, with the threshold probability between 0.150 and 0.850.

cohort. The established radiomics model demonstrated good discrimination and predictive power in both cohorts.

Treatment options for early-, intermediate-, and advanced-stage CC vary widely based on the latest FIGO staging system from 2018. Despite technological advancements in imaging-based diagnostics, many studies have shown that the morphological evaluation of MR images involves many subjective viewpoints. Many patients are misclassified, such as evaluating lymph node metastasis and parametrial invasion (18, 19).

Multidimensional characterization of radiomics features and quantification of detailed information in tumor images can reflect heterogeneity among different tumors. Therefore, preoperative radiomics signatures can provide a more objective and accurate assessment. Preoperative MR imaging can improve clinical staging accuracy by assessing the tumor's location and size, parametrial invasion, and lymph node metastasis to select more appropriate treatment plans. Previous studies involving the radiomics method have noted their predictive value for diagnosing tumors and therapeutic effects accurately. Some of these studies have shown that the performance of radiomics models can be improved by using the high-throughput features of multiparametric images of tumor lesions (20, 21). Therefore, in our study, we extracted and screened some radiomics features from T2WI and ADC maps and finally obtained radiomics signatures based on T2WI and ADC. The results demonstrated that radiomics features from T2WI and ADC have roughly similar discrimination performance for therapy method prediction.

The T2WI-ADC-combined radiomics features contained more wavelet filtered features, most likely because the filter could map the

image to several transform domains and better conveyed the tumor's biological information (22). The wavelet transform can also gradually convert image information into low- and high-frequency information, which improves local features, increases information content in tumor images, and provides more information about the biological behaviors and heterogeneity of different tumors at multiple scales (22, 23). There was an original feature named original_shape_Flatness in the T2WI radiomics features. The lesions in the high FIGO stage cohort were more irregular in shape because of the larger size of the tumor, deep stromal invasion, or parametrial invasion than those in the low FIGO stage cohort (24). Lee, in a review, stated that the 2018 revision of FIGO requires a more accurate description of the size of primary tumors and should be measured using MRI, especially for cervical resection plans (25).

DWI describes water mobility within the lesion tissue and enables quantitative evaluation of the diffusion properties of diseased tissues according to the calculated ADC (10). This quantitative parameter has been used in many studies to characterize tumors or assess their response to treatment (24, 26). Several studies have found that in CC, the minimum ADC values of tumors have been related to SCC, tumor grades, parametrial invasion, and poor survival rate. Furthermore, the changes in ADC values of lesions during radiotherapy and chemotherapy are also associated with the treatment response of tumors (10, 24, 27, 28). Haldorsen et al. considered that the ADC value of the tumor provides additional information about the microstructure of the tumor that may be relevant for staging and prediction of CC (24).

T2WI can provide detailed morphological features of CC in patients, and the features extracted from T2WI in this study have high sensitivity and low specificity. The low specificity of T2WI may be due to high-signal edema or inflammation within the paracervical fascia, which is indistinguishable from high-signal tumors (29). However, a combination of T2 and ADC prediction models can solve this problem. Many previous studies have found that radiomics features extracted from T2WI can help predict cervical lymph node metastasis and parametrial infiltration (9, 16, 21).

In the 2018 edition of the FIGO staging system, preoperative MR lymph node status is directly involved in IIIC staging (5). We used this as an important independent factor in the study; the results did not perform relatively well. Previous studies have used different tumor diameters to forecast the risk of parametrial invasion in patients with early-stage CC, and the measurement and selecting standards for tumor diameter have also varied from the different studies (19, 30). Some studies have found no direct correlation between tumor diameter and lymph node metastasis (21). Prediction of a therapeutic method relying on tumor diameter might not apply in clinical settings. Previous studies have shown that patient age is also an important independent factor for para-uterine invasion and lymph node metastasis prediction of CC (31, 32); however, the results did not perform relatively well in predicting the CC therapy method in our study. Gravdal et al. reported that the incidence of CC in women aged <30 years has increased in European countries over the past 20 years, but overall, the cancer does not tend to be more advanced when detected (33). The same study from the UK concluded that CC in younger women (aged 20–24 years) tended to be more advanced than in older women and is often a rarer histological type (34). In our study, there were only four patients aged <30 years with pathology of SCC, and statistical differences may not have been noted. SCC-Ag is currently the most widely used biomarker for diagnosing and estimating the effect of chemotherapy in patients with CC (21, 32). Shou et al. found that the serum SCC-Ag level was statistically associated with advanced FIGO stage (35). However, some studies have shown no relationship between SCC-Ag level and clinical stage (36). We also included preoperative SCC-Ag levels as a clinical factor, and there was no correlation between SCC-Ag levels and prediction of the CC therapy method. However, further studies with larger sample sizes are warranted.

Our study has some limitations. First, MRI acquisition and segmentation were independently obtained by two radiologists using a consensus, and further studies are needed to validate inter- and intra-observer repeatability. Second, in our study, radiomics features extracted from T2W images had low specificity, and further research with larger sample sizes or wider range of clinical and imaging features are required. Furthermore, all subjects in our study had ACAs and squamous carcinomas. Different histological subtypes of CC should be thoroughly studied in the future.

In conclusion, radiomics models were constructed from the ADC maps and T2WI, which were robust in differentiating the low (I-IIa) and high (IIb–IV) FIGO stages of cervical cancer, which may be valuable for the therapy decision-making in cervical cancer. The results also suggest that the combination model based on T2WI and ADC maps had the best performance in predicting the CC stage.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

This study was approved by the institutional ethics review board of our hospital (approval no. 2022-027), and the informed consent requirement was waived due to the retrospective study.

Author contributions

FW: Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft. RZ: Data curation, Writing – original draft. FL: Validation, Writing – original draft. XQ: Methodology, Supervision, Validation, Writing – original draft. HX: Methodology, Supervision, Writing – review & editing. HL: Data curation, Writing – review & editing. LL: Formal analysis, Investigation, Writing – review & editing. TA: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research received funding from the Keypoint Research and Development Program of Hubei Province (Grant Number: 2022BCE019).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Lin S, Gao K, Jin M, Simeng G, Liuqing Y, Sangni Q, et al. Worldwide trends in cervical cancer incidence and mortality, with predictions for the next 15 years. *Cancer*. (2021) 127:4030–9. doi: 10.1002/cncr.33795

2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*. (2019) 69:7–34. doi: 10.3322/caac.21551

3. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. *CA Cancer J Clin*. (2016) 66:115–32. doi: 10.3322/caac.21338

4. Denny L, Sanjose SD, Mutebi M, Anderson BO, Kim J, Jeronimo J, et al. Interventions to close the divide for women with breast and cervical cancer between low-income and middle-income countries and high-income countries. *Lancet*. (2017) 389:861–70. doi: 10.1016/S0140-6736(16)31795-0

5. Olawaiye AB, Baker TP, Washington MK, Mutch DG. The new (version 9) american joint committee on cancer tumor, node, metastasis staging for cervical cancer. *CA A Cancer J Clin*. (2021) 71:287–98. doi: 10.3322/caac.21663

6. Kuznicki ML, Chambers LM, Morton M, Ji S, Rose PG. Fertility-sparing surgery for early-stage cervical cancer: a systematic review of the literature. *J Minim Invasive Gynecol*. (2020) 28:513–26. doi: 10.1016/j.jmig.2020.10.013

7. Spindler M, Thiel CM. Quantitative magnetic resonance imaging for segmentation and white matter extraction of the hypothalamus. *J Neurosci Res*. (2022) 100:564–77. doi: 10.1002/jnr.24988

8. Jiang X, Li J, Kan Y, Tao Y, Chang S, Sha X, et al. MRI based radiomics approach with deep learning for prediction of vessel invasion in early-stage cervical cancer. *IEEE/ACM Trans Comput Biol Bioinform*. (2021) 18:995–1002. doi: 10.1109/TCBB.2019.2963867

9. Wang M, Perucho JAU, Tse KY, Chu M, Lee E, Philip I. MRI texture features differentiate clinicopathological characteristics of cervical carcinoma. *Eur Radiol*. (2020) 30:5384–91. doi: 10.1007/s00330-020-06913-7

10. Karunya RJ, Tharani P, John S, Kumar RM, Das S. Role of functional magnetic resonance imaging derived parameters as imaging biomarkers and correlation with clinicopathological features in carcinoma of uterine cervix. *J Clin Diagn Res*. (2017) 11:Xc06-11. doi: 10.7860/JCDR/2017/29165.10426

11. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. (2016) 278:563–77. doi: 10.1148/radiol.2015151169

12. Liu Y, Zhang Y, Cheng R, Liu S, Qu F, Yin X, et al. Radiomics analysis of apparent diffusion coefficient in cervical cancer: a preliminary study on histological grade evaluation. *J Magn Reson Imaging*. (2019) 49:280–90. doi: 10.1002/jmri.26192

13. Charoenkwan P, Shoombuatong W, Nantasupha C, Muangmool T, Suprasert P, Charoenkwan K. Ipmi: machine learning-aided identification of parametrial invasion in women with early-stage cervical cancer. *Diagnostics (Basel)*. (2021) 11:1454–67. doi: 10.3390/diagnostics11081454

14. Hou L, Zhou W, Ren J, Du X, Xin L, Zhang R, et al. Radiomics analysis of multiparametric MRI for the preoperative prediction of lymph node metastasis in cervical cancer. *Front Oncol*. (2020) 10:1393–404. doi: 10.3389/fonc.2020.01393

15. Yu YY, Zhang R, Dong RT, Hu QY, Yu T, Liu F, et al. Feasibility of an ADC-based radiomics model for predicting pelvic lymph node metastases in patients with stage IB–IIA cervical squamous cell carcinoma. *Br J Radiol*. (2019) 92:20180986. doi: 10.1259/bjr.20180986

16. Benjamin WW, Simon JD, Thomas EI, James DA, James P, Nandita MD. Radiomic features of cervical cancer on T2-and diffusion-weighted MRI: prognostic value in low-volume tumors suitable for trachelectomy. *Gynecol Oncol*. (2020) 156:107–14. doi: 10.1016/j.ygyno.2019.10.010

17. Zheng X, Guo W, Dong J, Qian L. Prediction of early response to concurrent chemoradiotherapy in cervical cancer: value of multi-parameter MRI combined with clinical prognostic factors. *Magn Reson Imaging*. (2020) 72:159–66. doi: 10.1016/j.mri.2020.06.014

18. Luo Q, Luo L, Tang L. A network Meta-analysis on the diagnostic value of different imaging methods for lymph node metastases in patients with cervical Cancer. *Technol Cancer Res Treat*. (2018) 17:1–10. doi: 10.1177/1533034617742311

19. Woo S, Moon MH, Cho JY, Kim SH, Kim SY. Diagnostic performance of MRI for assessing Parametrial invasion in cervical Cancer: a head-to-head comparison between oblique and true axial T2-weighted images. *Korean J Radiol*. (2019) 20:378–84. doi: 10.3348/kjr.2018.0248

20. Liu Z, Wang S, Dong D, Wei J, Fang C, Tian J, et al. The applications of Radiomics in precision diagnosis and treatment of oncology: opportunities and challenges. *Theranostics*. (2019) 9:1303–22. doi: 10.7150/thno.30309

21. Wang T, Gao T, Guo H, Wang Y, Zhang M. Preoperative prediction of parametrial invasion in early-stage cervical cancer with MRI-based radiomics nomogram. *Eur Radiol*. (2020) 30:3585–93. doi: 10.1007/s00330-019-06655-1

22. Wu Y, Wang S, Chen Y, Liao Y, Yin X, Li T, et al. A multicenter study on preoperative assessment of Lymphovascular space invasion in early-stage cervical Cancer based on multimodal MR Radiomics. *J Magn Reson Imaging*. (2023) 58:1638–48. doi: 10.1002/jmri.28676

23. Zheng Z, Gu Z, Xu F, Maskey N, He Y, Yan Y, et al. Magnetic resonance imaging-based radiomics signature for preoperative prediction of Ki67 expression in bladder cancer. *Cancer Imaging*. (2021) 21:65–79. doi: 10.1186/s40644-021-00433-3

24. Haldorsen IS, Lura N, Blaakær J, Fischerova D, Werner HMJ. What is the role of imaging at primary diagnostic work-up in uterine cervical Cancer? *Curr Oncol Rep*. (2019) 21:77–92. doi: 10.1007/s11912-019-0824-0

25. Lee SI, Atri M. 2018 FIGO staging system for uterine cervical Cancer: enter cross-sectional imaging. *Radiology*. (2019) 292:15–24. doi: 10.1148/radiol.2019190088

26. Fournier LS, Bats AS, Durdux C. Diffusion MRI: technical principles and application to uterine cervical cancer. *Cancer Radiother*. (2020) 24:368–73. doi: 10.1016/j.canrad.2020.02.008

27. Kuang F, Ren J, Zhong Q, Liyuan F, Huan Y, Chen Z. The value of apparent diffusion coefficient in the assessment of cervical cancer. *Eur Radiol*. (2013) 23:1050–8. doi: 10.1007/s00330-012-2681-1

28. Onal C, Erbay G, Guler OC. Treatment response evaluation using the mean apparent diffusion coefficient in cervical cancer patients treated with definitive chemoradiotherapy. *J Magn Reson Imaging*. (2016) 44:1010–9. doi: 10.1002/jmri.25215

29. Moloney F, Ryan D, Twomey M, Hewitt M, Barry J. Comparison of MRI and high-resolution transvaginal sonography for the local staging of cervical cancer. *J Clin Ultrasound*. (2016) 44:78–84. doi: 10.1002/jcu.22288

30. Lee JY, Youm J, Kim JW, Cho JY, Kim MA, Kim TH, et al. Identifying a low-risk group for parametrial involvement in microscopic stage IB1 cervical cancer using criteria from ongoing studies and a new MRI criterion. *BMC Cancer*. (2015) 15:167–74. doi: 10.1186/s12885-015-1184-2

31. Hsu HC, Tai YJ, Chen YL, Chiang YC, Chen CA, Cheng WF, et al. Factors predicting parametrial invasion in patients with early-stage cervical carcinomas. *PLoS One*. (2018) 13:e0204950. doi: 10.1371/journal.pone.0204950

32. Olthof EP, Aa MAVD, Adam JA, Stalpers L, Wenzel HHB, Mom CH, et al. The role of lymph nodes in cervical cancer: incidence and identification of lymph node metastases—a literature review. *Int J Clin Oncol*. (2021) 26:1600–10. doi: 10.1007/s10147-021-01980-2

33. Gravdal BH, Lönnberg S, Skare GB, Sulo G, Bjørge T. Cervical cancer in women under 30 years of age in Norway: a population-based cohort study. *BMC Womens Health*. (2021) 21:110–8. doi: 10.1186/s12905-021-01242-3

34. Castanon A, Leung VM, Landy R, Lim AW, Sasieni P. Characteristics and screening history of women diagnosed with cervical cancer aged 20-29 years. *Br J Cancer*. (2013) 109:35–41. doi: 10.1038/bjc.2013.322

35. Shou H, Yasuo Y, Yuan S, Lou H, Ni J. Association of pretreatment suvmax of cervix and scc-antigen with figo2018 stage in stage IIB–IVB squamous cervical cancer and relationship to prognosis. *Int J Gynaecol Obstet*. (2020) 152:112–7. doi: 10.1002/ijgo.13465

36. Jin Y, Kim SC, Kim HJ, Ju W, Kim YH, Kim HJ. Use of protein-based biomarkers of exfoliated cervical cells for primary screening of cervical cancer. *Arch Pharm Res*. (2018) 41:438–49. doi: 10.1007/s12272-018-1015-5

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 06 February 2024
DOI 10.3389/fmed.2024.1334062

Check for updates

OPEN ACCESS

EDITED BY
Minjeong Kim,
University of North Carolina at Greensboro, United States

REVIEWED BY
Yongai Li,
Changzhi People's Hospital, China
Wei Xia,
Chinese Academy of Sciences (CAS), China

*CORRESPONDENCE
Yu Wang
renjiwangyu@126.com
Jiejun Cheng
drchjj@163.com
Jianli Yu
yjl0617@sohu.com

†These authors have contributed equally to this work

RECEIVED 06 November 2023
ACCEPTED 11 January 2024
PUBLISHED 06 February 2024

CITATION
Fu L, Wang W, Lin L, Gao F, Yang J, Lv Y, Ge R, Wu M, Chen L, Liu A, Xin E, Liu F, Yu J, Cheng J and Wang Y (2024) Multitask prediction models for serous ovarian cancer by preoperative CT image assessments based on radiomics.
Front. Med. 11:1334062.
doi: 10.3389/fmed.2024.1334062

COPYRIGHT
© 2024 Fu, Wang, Lin, Gao, Yang, Lv, Ge, Wu, Chen, Liu, Xin, Liu, Yu, Cheng and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Multitask prediction models for serous ovarian cancer by preoperative CT image assessments based on radiomics

Le Fu[1†], Wenjing Wang[1†], Lingling Lin[1†], Feng Gao[1], Jiani Yang[2], Yunyun Lv[1], Ruiqiu Ge[1], Meixuan Wu[2], Lei Chen[3], Aie Liu[3], Enhui Xin[3], Jianli Yu[1*], Jiejun Cheng[1*] and Yu Wang[2*]

[1]Department of Radiology, Shanghai First Maternity and Infant Hospital, School of Medicine, Tongji University, Shanghai, China, [2]Department of Obstetrics and Gynecology, Shanghai First Maternity and Infant Hospital, School of Medicine, Tongji University, Shanghai, China, [3]Shanghai United Imaging Intelligence Co., Ltd., Shanghai, China

Objective: High-grade serous ovarian cancer (HGSOC) has the highest mortality rate among female reproductive system tumors. Accurate preoperative assessment is crucial for treatment planning. This study aims to develop multitask prediction models for HGSOC using radiomics analysis based on preoperative CT images.

Methods: This study enrolled 112 patients diagnosed with HGSOC. Laboratory findings, including serum levels of CA125, HE-4, and NLR, were collected. Radiomic features were extracted from manually delineated ROI on CT images by two radiologists. Classification models were developed using selected optimal feature sets to predict R0 resection, lymph node invasion, and distant metastasis status. Model evaluation was conducted by quantifying receiver operating curves (ROC), calculating the area under the curve (AUC), De Long's test.

Results: The radiomics models applied to CT images demonstrated superior performance in the testing set compared to the clinical models. The area under the curve (AUC) values for the combined model in predicting R0 resection were 0.913 and 0.881 in the training and testing datasets, respectively. De Long's test indicated significant differences between the combined and clinical models in the testing set (p =0.003). For predicting lymph node invasion, the AUCs of the combined model were 0.868 and 0.800 in the training and testing datasets, respectively. The results also revealed significant differences between the combined and clinical models in the testing set (p =0.002). The combined model for predicting distant metastasis achieved AUCs of 0.872 and 0.796 in the training and test datasets, respectively. The combined model displayed excellent agreement between observed and predicted results in predicting R0 resection, while the radiomics model demonstrated better calibration than both the clinical model and combined model in predicting lymph node invasion and distant metastasis. The decision curve analysis (DCA) for predicting R0 resection favored the combined model over both the clinical and radiomics models, whereas for predicting lymph node invasion and distant metastasis, DCA favored the radiomics model over both the clinical model and combined model.

Conclusion: The identified radiomics signature holds potential value in preoperatively evaluating the R0, lymph node invasion and distant metastasis in

patients with HGSC. The radiomics nomogram demonstrated the incremental value of clinical predictors for surgical outcome and metastasis estimation.

KEYWORDS

radiomics, preoperative evaluation, serous ovarian cancer, computer tomography, nomogram

1 Introduction

High-grade serous ovarian cancer (HGSOC) is the malignant tumor with the highest mortality rate in female reproductive system at present. The early clinical symptoms of HGSOC are not obvious, and most patients are already in the middle to late stage when detected. The current standard treatment methods for HGSOC are platinum-based chemotherapy after primary tumor reduction surgery (PDS), or intermittent tumor reduction surgery after neoadjuvant chemotherapy (1). Studies have shown that these two treatment methods can achieve similar prognosis in patients with stage IIIC-IV ovarian cancer (2). However, for patients who are suitable for early surgical intervention, the biggest risk of undergoing surgical treatment after chemotherapy is the possibility of losing the opportunity for early surgery and developing tolerance to chemical drugs (3). Additionally, the residual lesion size after tumor reduction surgery is one of the most important independent risk factors for the prognosis and survival of ovarian cancer patients (4). Therefore, precise preoperative evaluation of tumors is crucial for selecting treatment plans. Factors such as lymph node invasion, distant metastasis, and whether complete resection of all visible diseases (R0 resection) can be achieved are important considerations for preoperative evaluation.

Radiomics is a powerful and promising image mining method that utilizes high-throughput feature selection based on imaging data. It has been proven to improve diagnostic accuracy, evaluate treatment response, and predict prognosis (5, 6). Several published radiomics prediction models have been established based on computer tomography (CT) (7, 8). However, these radiomics models mainly focused on the location, size, and the metastasis of the abdomen, and they all focused on a single prediction point. In this study, we aimed to establish multitask prediction models for HGSOC by utilizing preoperative CT image assessments based on radiomics.

2 Materials and methods

2.1 Study population

This retrospective study was approved by our hospital ethics committee, and the requirement for patient informed consent was waived. A total of 112 consecutive patients (age range: 36–84 years) with confirmed serous ovarian cancer based on pathology were enrolled in our study. The enrollment period spanned from November 2012 to January 2022.

Patients' laboratory findings, including serum cancer antigen-125 (CA125), serum human epididymis protein 4 (HE-4) level, and neutrophil-to-lymphocyte ratio (NLR), were collected from the electronic medical record. Additionally, preoperative unenhanced CT scans of the abdomen and pelvis were obtained. A total of 112 patients were included in the study to predict R0 status, lymph node invasion, and distant metastasis.

2.2 Radiomics analysis

The workflow of radiomics analysis consists of five steps: obtaining ROI, computing features, selecting features, constructing the model, and evaluation. Radiomics analysis was performed using the uAI Research Portal (United Imaging Intelligence, China), which is a clinical research platform implemented in Python programming language (version 3.7.3). the widely used package PyRadiomics package[1] (9, 10) was utilized for this analysis.

The volume of the entire ovarian lesion was manually delineated on CT images by two radiologists (L. Fu with 12 years of imaging experience and WJ. Wang with 14 years of imaging experience) using the uAI Research Portal, denoted as ROI (region of interest). A total of 2,264 radiomic features were extracted from the ROI on each CT image, including 104 original features grouped as: 18 the first-order statistics, 72 texture, and 14 shape features. Among 14 shape features, selection was done only on original images, while the others were based on both original images and images processed through 25 filters such as boxmean, wavelet, laplacian, etc. To account for any difference in index dimension, the extracted radiomic features for each sequence were standardized into normal distributed z-scores. For the three classification tasks, the top 10, 9, and 8 highest-ranking radiomic features were selected, respectively, on CT images using feature selection methods such as K best and least absolute shrinkage and selection operator regression (LASSO), as shown in Figure 1.

To evaluate the performance of the classifier and protect against overfitting due to the limited amount of data, we used the cross-validation method. Specifically, we employed 5-fold cross-validation (9). The feature set was randomly split into five partitions, ensuring that each partition maintained the same ratio of positive and negative images. During each fold, the classifier was trained on four-fifths of the dataset and validated on the remaining partition. This process was repeated five times with different subgroups, resulting in five distinct training/testing sets. The average performance across these folds was then calculated to obtain an overall result. To maximize the discrimination ability of the radiomics algorithm, we implemented machine learning classifiers including logistic regression (LR),

1 https://pyradiomics.Readthedocs.io/en/latest/index.html

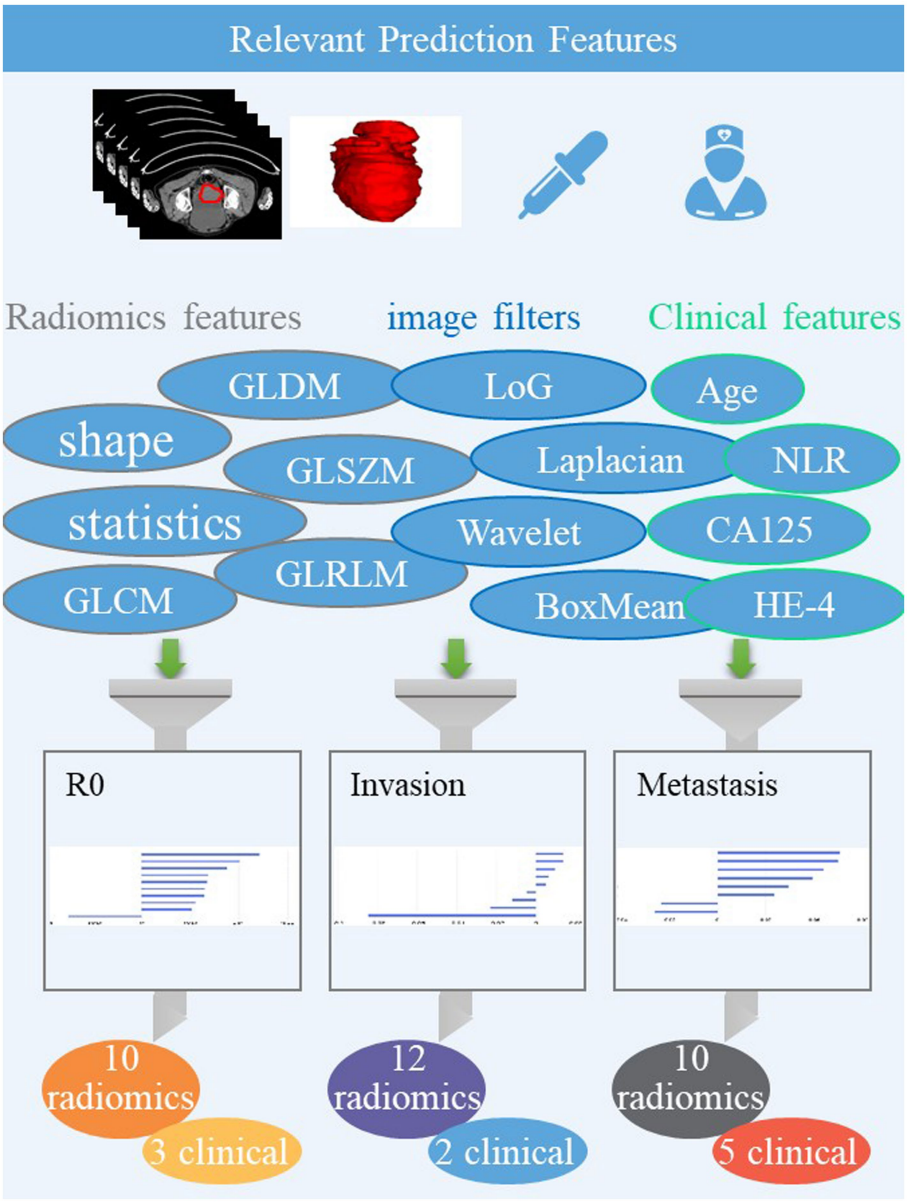


FIGURE 1
Flowchart of radiomics analysis.

random forest, decision tree, and support vector machine (SVM) for model construction.

Using the selected optimal feature sets, we constructed classification models to predict R0 status, lymph node invasion, and distant metastasis. Finally, we plotted receiver operating characteristic (ROC) curves and calculated the area under the curve (AUC), sensitivity, specificity, and accuracy to evaluate the performance of the models.

2.3 Statistical analysis

All statistical analyses for the present study were performed using SPSS software (version 26.0), R (version 4.1.0), and Python (version 3.5.6). A significance level of $p < 0.1$ was considered statistically significant. DeLong's test was used to compare the AUC values of different models.

3 Results

3.1 Clinical characteristics

The clinical baseline data of patients are presented in Table 1 for three postoperative predictors. Firstly, a total of 112 patients were included to predict R0 status, with 80 patients classified as non-R0 status and 32 patients as R0 resection. The cancer antigen-125 (CA125), human epididymis protein 4 (HE-4), and lymphocyte levels showed significant differences between the R0 and non-R0 groups ($p < 0.05$), while no significant differences were observed in patient

TABLE 1 Patient characteristics for three predicting postoperative results.

	Predicting R0			Predicting lymph node invasion			Predicting distant metastasis		
	Non-R0 (*n* = 80)	R0 (*n* = 32)	*p*	Non-invasion (*n* = 58)	Invasion (*n* = 54)	*p*	Non-metastasis (*n* = 79)	Metastasis (*n* = 33)	*p*
Age, mean (SD)	61.7 ± 8.9	60.1 ± 9.9	0.426	61.8 ± 8.5	60.5 ± 9.9	0.475	61.1 ± 9.4	61.3 ± 8.9	0.904
CA125, median (IQR)	880 (415, 2,155)	509 (187, 1774)	0.039	657 (291, 1860)	1,000 (421, 2024)	0.094	635 (288, 1,355)	1,432 (834, 2,283)	0.000
HE-4, median (IQR)	452 (381, 856)	394 (197, 642)	0.013	452 (198, 706)	452 (376, 944)	0.091	452 (220, 643)	642 (404, 1,011)	0.017
NLR, median (IQR)	3.58 (3.0, 5.59)	3.55 (2.71, 4.96)	0.074	3.58 (2.69, 4.67)	4.39 (2.91, 6.48)	0.026	3.58 (2.69, 4.97)	4.62 (3.13, 6.55)	0.002
Neutrophils, median (IQR)	4.70 (3.36, 6.44)	4.43 (4.02, 5.15)	0.282	4.64 (3.08, 5.19)	4.71 (4.08, 6.48)	0.045	4.32 (3.33, 5.58)	4.91 (4.42, 5.80)	0.023
Lymphocyte, median (IQR)	1.23 (0.95, 1.39)	1.42 (1.12, 1.74)	0.007	1.27 (1.07, 1.48)	1.23 (0.86, 1.42)	0.183	1.30 (1.07, 1.50)	1.10 (0.85, 1.30)	0.032

*Negative/positive; *p*-values in bold indicated that the corresponding variables were closely related to the two types in the univariable logistic regression ($p < 0.1$). OR, odd ratio; CI, confidence interval; CA125, cancer antigen-125; HE-4, human epididymis protein 4; NLR, neutrophil-to-lymphocyte ratio.

TABLE 2 The 5-fold mean performance of the clinical model, radiomics model and radiomics + clinical model for predicting postoperative results.

	Model	AUC	Sensitivity	Specificity	Accuracy
Predicting R0	Clinical	0.675 (0.567–0.784)	0.652	0.648	0.649
	Radiomics	0.872 (0.806–0.938)	0.819	0.835	0.827
	Radiomics+ Clinical	**0.881** (0.806–0.956)	0.833	0.784	0.802
Predicting Lymph node Invasion	Clinical	0.591 (0.406–0.885)	0.647	0.489	0.563
	Radiomics	0.770 (0.563–0.972)	0.704	0.702	0.705
	Radiomics+ Clinical	0.800 (0.614–0.970)	0.776	0.689	0.732
Predicting Distant Metastasis	Clinical	0.729 (0.425–0.925)	0.686	0.697	0.696
	Radiomics	0.795 (0.572–0.998)	0.729	0.718	0.722
	Radiomics+ Clinical	0.796 (0.623–0.981)	0.781	0.711	0.731

The highest AUC was bold in Table 2.

age, neutrophil-to-lymphocyte ratio (NLR), and Neutrophils (all $p > 0.05$). Secondly, among the 112 patients collected for predicting lymph node invasion, 58 patients had non- Invasion status and 54 patients had Invasion. The NLR and Neutrophils showed significant differences ($p < 0.05$), while no significant difference was detected in other characteristics. Thirdly, the 112 patients were collected to predict distant metastasis, with 79 patients having non- Metastasis status and 33 patients having Metastasis. The age did not show a significant difference ($p > 0.01$), while significant differences were noted in other characteristics (all $p < 0.05$).

3.2 Performances of clinical and CT radiomics models

The predictive performance of each model is presented in Table 2. The clinical models exhibited relatively poor predictive performance in the testing set, utilizing six clinical characteristics (AUC_R0 = 0.675, AUC_Invasion = 0.591, AUC_Metastasis = 0.729). In contrast, the radiomics models on CT images demonstrated better performance in the testing set (AUC_R0 = 0.872, AUC_Invasion = 0.770, AUC_Metastasis = 0.795) than the clinical models. De Long's test indicated significant differences between the radiomics and clinical models for predicting R0 and lymph node invasion in the testing set ($p = 0.003$ and 0.011, respectively). However, there was no statistical difference between the radiomics and clinical models for predicting distant metastasis ($p = 0.367$).

3.3 Performances of radiomics-clinical comprehensive models

The model for predicting R0 was developed using 10 radiomics features and three clinical features, including CA125, HE-4, lymphocyte. The AUCs of the combined model were 0.913 and 0.881 in the training and testing datasets, respectively (Figure 1 and Table 2). De Long's test revealed significant differences between the combined and clinical models in the testing set ($p = 0.003$). However, there was no statistical difference between the combined and radiomics models ($p = 0.756$).

The predicting lymph node invasion model was established based on 12 radiomics features and two clinical features, including NLR and

Neutrophils. In the training and test datasets, the combined model achieved AUCs of 0.868 and 0.800, respectively. Significantly different results were observed between the combined and clinical models in the testing set according to De Long's test ($p = 0.002$). However, no statistical difference was found between the combined and radiomics models ($p = 0.185$).

In establishing the predicting distant metastasis model, we utilized 10 radiomics features and five clinical features, including CA125, HE-4, NLR, Neutrophils, and lymphocyte. The combined model achieved AUCs of 0.872 and 0.796 in the training and test datasets, respectively. However, there was no statistical difference observed among the three models, as all *p*-values were greater than 0.05.

3.4 The calibration and clinical utility of all models

The calibration curves of all models are shown in Figures 2A–C. The calibration of the radiomics model for predicting lymph node invasion was superior to that of the clinical model and combined model. Similarly, the calibration of the radiomics model for predicting distant metastasis outperformed that of the clinical model and combined model. On the other hand, the combined model for predicting R0 demonstrated excellent agreement between the observed and predicted results, surpassing both the clinical model and radiomics model.

The DCA of all models is displayed in Figures 2D–F, illustrating the clinical utility. The DCA of the radiomics model for predicting lymph node invasion outperformed the clinical model and combined model. Furthermore, the DCA of the radiomics model for predicting distant metastasis exhibited better results compared to both the clinical model and combined model. Similarly, the DCA of the combined model for predicting R0 demonstrated superior performance compared to both the clinical model and radiomics model.

4 Discussion

Accurate and non-invasive prediction of R0, lymph node invasion, and distant metastasis is crucial for implementing individualized management and improving the prognosis of patients with HGSOC. In this study, we developed and validated three multitask prediction models for HGSOC that integrate CT radiomics features and clinical information. The combined model demonstrated excellent performance in predicting R0, with AUCs of 0.913 and 0.881 in the training and testing datasets, respectively. For predicting lymph node invasion, the combined model achieved AUCs of 0.807 and 0.800 in the training and testing datasets, respectively. Additionally, for predicting distant metastasis, the combined model achieved AUCs of 0.807 and 0.800 in the training and testing datasets, respectively. Notably, the AUC values of the combined models were consistently higher than those of the clinical models in both the R0 cohort and lymph node invasion cohort.

Most studies have only focused on the features of metastases for R0 prediction (8, 11). The observation of extensive metastases in the abdomen, affected by bowl and ascites, might make it hard to accurately predict R0 in practice. Studies had also shown that the likelihood of metastases can be predicted through the assessment of

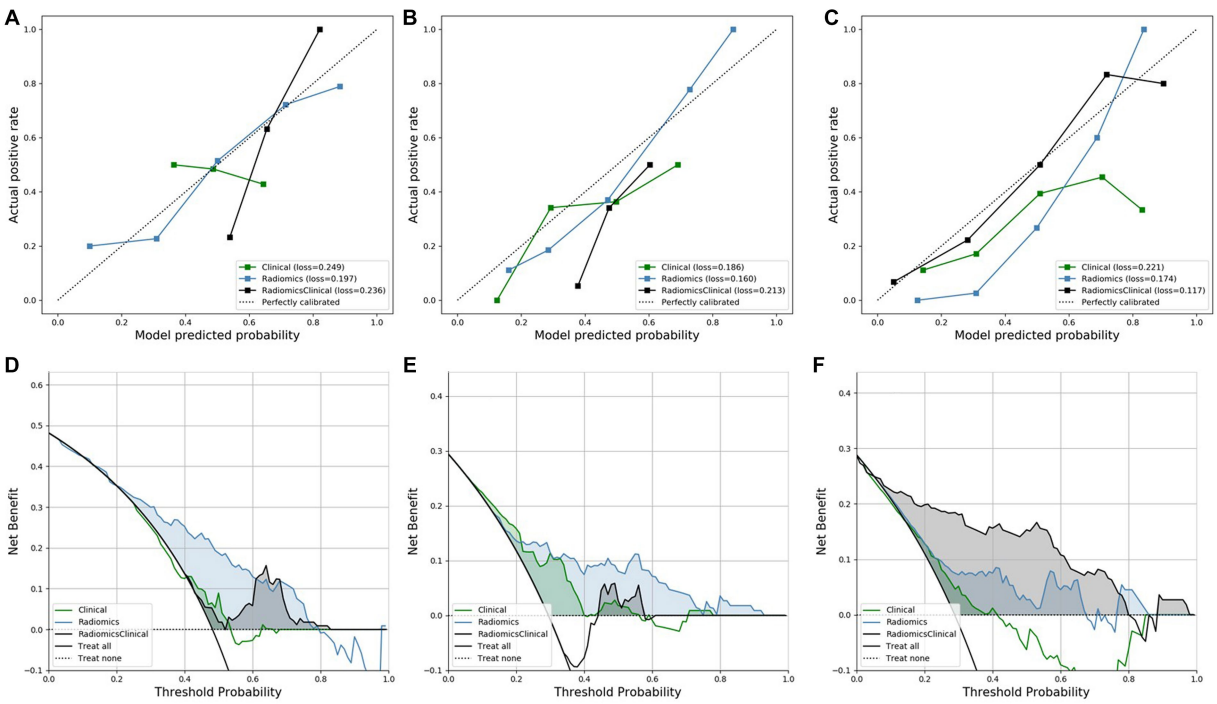


FIGURE 2
The calibration curves and DCA curves of all models. **(A–C)** The calibration curves for predicting lymph node invasion, distant metastasis, and R0, respectively. **(D–F)** The DCA curves for predicting lymph node invasion, distant metastasis, and R0, respectively.

primary tumors (12, 13). Radiomics has been shown to be a high-performance method for accurately predicting treatment response by assessing tumor heterogeneity (14). Rizzo et al. found that patients with values below the median for F2-Shape/Compactness1, F1-GrayLevelCooccurenceMatrix25/0-1InformationMeasureCorr2, and above the median for F1-GrayLevelCooccurenceMatrix25/−333-1InverseVariance showed a higher risk of residual tumor (36, 36, and 35%, respectively, as opposed to 18, 18, and 18%). However, models were not developed for prediction, or considering the value of clinical information (15). A radiomics signature–based nomogram was developed for the preoperative prediction of R0 in patients with advanced HGSOC. It demonstrated favorable performance in both training and validation sets with an AUC of 0.815 and 0.803, respectively (16). The metastatic situation of HOSG determines the FIOG stages and operation range, actual assessment carries significant weight. The radiomics nomogram demonstrated favorable calibration and discrimination in both the training cohort (AUC = 0.821) and test cohort (AUC = 0.843) (17). Previous studies on radiomic for predicting R0, lymph node invasion and distant metastasis have been based on CT enhanced images and MR enhanced images. However, some individuals may be allergic to contrast agents or have renal dysfunction, making them unsuitable for enhanced CT/MRI examination. Additionally, some individuals may have claustrophobia or metal implants that are not suitable for MRI examination. Therefore, we explored the performance of radiomic based on CT examination without injecting contrast agent in predicting R0, lymph node invasion, and distant metastasis. The three clinical-radiomic combined models showed good performance, indicating that CT examination without contrast agent can provide favorable evidence for the preoperative evaluation of ovarian cancer.

Wavelet features belong to higher-order statistical features, which can more comprehensively reflect the heterogeneity of the original image and may also result in more valuable features than the original image (18). The three prediction models in this study include many Wavelet feature (Figure 1). This suggests that higher-order texture features have a good correlation with surgical outcomes and metastasis in advanced serous cancer patients. A model for predicting platinum resistance was constructed based on the radiomic features proposed by T2WI, DWI, and CE-T1WI sequences of 114 EOCs. In the validation set, the AUC was 0.89 (accuracy = 85.0%, sensitivity = 87.0%, and specificity = 80.0%) (19). The platinum resistance prediction model also includes 9 wavelet features, which is similar to our research results and once again confirms the good performance of wavelet features for preoperative evaluation of advanced serous ovarian cancer.

This study selected clinical features and laboratory examination indicators such as age, CA125, HE4, lymphocytes, neutrophils, and NLR to explore the differences between different groups. The clinical model composed of CA125, HE-4, and lymphocytes showed poor performance. However, with the addition of radiomic features, the performance of the model for predicting R0 significantly increased. The clinical model composed of NLR and neutrophils showed average performance but with the addition of radiomic features, the efficiency of the model for predicting lymph node infiltration significantly increased. This suggests a good correlation between radiomic features and surgical outcomes, as well as lymph node metastasis. The clinical models for predicting distant metastasis composed of CA125, HE-4, NLR, neutrophils, and lymphocyte generally performed good. However, the addition of radiomic features did not significantly improve the performance of the model possibly due to a small sample size. The predictive value of some blood inflammatory composite markers in OC has been extensively reported (20). They can be used for early detection and differential diagnosis of OC and can also predict survival, treatment response, and recurrence in the affected patients. Our results confirmed the NLR is related to R0 status, lymph node invasion, and distant metastasis in HGSC patient. This suggests a close correlation between NLR and surgical outcomes as well as metastasis. Further exploration of these correlations will be conducted at the molecular level in future studies. A radiomic-clinical nomogram based on MRI for predicting R0 also included CA125, LDH, and NLR (16).

The DCA of the combined model for predicting R0 confirmed the incremental clinical utility of the proposed model for individualized prediction. This finding is consistent with several previous radiomics studies on HGSOC and cervical carcinoma (21–23). However, the DCA of the combined models for predicting lymph node invasion and distant metastasis did not show the superior performance compared to a single model. The combined models were less sensitive in evaluating tumor heterogeneity, and other artificial intelligence methods will be explored to optimize the model fusion.

This study had several limitations. First, selection bias was inevitable due to the retrospective nature of the study and strict inclusion and exclusion criteria. Second, the retrospective datasets were relatively small, with an unbalanced distribution of patients. Third, radiomics has inevitable limitations in terms of reproducible application as it heavily relies on artificial segmentation and handcrafted features (24). Moreover, although volumetric tumor segmentation can provide a robust way to characterize tumor heterogeneity, it may be time-consuming, especially for larger ovarian tumors.

5 Conclusion

The identified radiomics signature holds potential value in preoperatively evaluating the R0, lymph node invasion, and distant metastasis in patients with HGSC. The radiomics nomogram demonstrated the incremental value of clinical predictors for surgical outcomes and metastasis estimation. However, further external validation is required before its wide clinical application.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

This study was approved by the Ethics Committee of the Shanghai First Maternity and Infant Hospital (registration number:KS22281). The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because This is a retrospective study.

Author contributions

LF: Conceptualization, Methodology, Writing – original draft. WW: Methodology, Writing – original draft. LL: Formal analysis, Writing – original draft. FG: Writing – review & editing. JYa: Writing – review & editing, Data curation. YL: Writing – review & editing. RG: Writing – review & editing. MM: Writing – review & editing. LC: Writing – review & editing. AL: Writing – review & editing. EX: Writing – review & editing. JYu: Writing – review & editing. JC: Funding acquisition, Writing – review & editing. YW: Writing – original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by the grants from the Shenkang-United-Imaging Development Plan (2022SKLY-12), Shanghai Science and Technology Innovation Action Plan (21S31904200, 22S31903700), Pudong Obstetrics and Gynecology Joint Venture Cooperation Project (PDYLT2022-20), and Shang First Maternity and Infant Hospital Talent Project (2022RC04).

Conflict of interest

LC, AL, and EX was employed by the Shanghai United Imaging Intelligence Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Webb PM, Jordan SJ. Epidemiology of epithelial ovarian cancer. *Best Pract Res Clin Obstet Gynaecol*. (2017) 41:3–14. doi: 10.1016/j.bpobgyn.2016.08.006

2. Onda T, Satoh T, Ogawa G, Saito T, Kasamatsu T, Nakanishi T, et al. Comparison of survival between primary debulking surgery and neoadjuvant chemotherapy for stage III/IV ovarian, tubal and peritoneal cancers in phase III randomised trial. *Eur J Cancer*. (2020) 130:114–25. doi: 10.1016/j.ejca.2020.02.020

3. Feng Z, Wen H, Li R, Liu S, Fu Y, Chen X, et al. Comparison of survival between primary debulking surgery versus neoadjuvant chemotherapy for ovarian cancers in a personalized treatment cohort. *Front Oncol*. (2020) 10:632195. doi: 10.3389/fonc.2020.632195

4. Medina-Franco H, Cortes-Gonzalez R, Lambreton-Hinojosa F, Fimbres-Morales A, Vargas-Siordia JC. Neoadjuvant chemotherapy increases R0 cytoreduction rate but does not improve final outcome in advanced epithelial ovarian cancer. *Ann Surg Oncol*. (2017) 24:1330–5. doi: 10.1245/s10434-016-5704-3

5. Peng H, Dong D, Fang MJ, Li L, Tang LL, Chen L, et al. Prognostic value of deep learning PET/CT-based radiomics: potential role for future individual induction chemotherapy in advanced nasopharyngeal carcinoma. *Clin Cancer Res*. (2019) 25:4271–9. doi: 10.1158/1078-0432.CCR-18-3065

6. Qian L, Ren J, Liu A, Gao Y, Hao F, Zhao L, et al. MR imaging of epithelial ovarian cancer: a combined model to predict histologic subtypes. *Eur Radiol*. (2020) 30:5815–25. doi: 10.1007/s00330-020-06993-5

7. Qayyum A, Coakley FV, Westphalen AC, Hricak H, Okuno WT, Powell B. Role of CT and MR imaging in predicting optimal cytoreduction of newly diagnosed primary epithelial ovarian cancer. *Gynecol Oncol*. (2005) 96:301–6. doi: 10.1016/j.ygyno.2004.06.054

8. Suidan RS, Ramirez PT, Sarasohn DM, Teitcher JB, Mironov S, Iyer RB, et al. A multicenter prospective trial evaluating the ability of preoperative computed tomography scan and serum CA-125 to predict suboptimal cytoreduction at primary debulking surgery for advanced ovarian, fallopian tube, and peritoneal cancer. *Gynecol Oncol*. (2014) 134:455–61. doi: 10.1016/j.ygyno.2014.07.002

9. Chong HH, Yang L, Sheng RF, Yu YL, Wu DJ, Rao SX, et al. Multi-scale and multi-parametric radiomics of gadoxetate disodium-enhanced MRI predicts microvascular invasion and outcome in patients with solitary hepatocellular carcinoma </= 5 cm. *Eur Radiol*. (2021) 31:4824–38. doi: 10.1007/s00330-020-07601-2

10. Wu J, Xia Y, Wang X, Wei Y, Liu A, Innanje A, et al. uRP: an integrated research platform for one-stop analysis of medical images. *Front Radiol*. (2023) 3:1153784. doi: 10.3389/fradi.2023.1153784

11. Gu Y, Qin M, Jin Y, Zuo J, Li N, Bian C, et al. A prediction model for optimal primary Debulking surgery based on preoperative computed tomography scans and clinical factors in patients with advanced ovarian cancer: a multicenter retrospective cohort study. *Front Oncol*. (2020) 10:611617. doi: 10.3389/fonc.2020.611617

12. Bertucci F, Birnbaum D. Distant metastasis: not out of reach any more. *J Biol*. (2009) 8:28. doi: 10.1186/jbiol128

13. Messal HA, Alt S, Ferreira RMM, Gribben C, Wang VM, Cotoi CG, et al. Tissue curvature and apicobasal mechanical tension imbalance instruct cancer morphogenesis. *Nature*. (2019) 566:126–30. doi: 10.1038/s41586-019-0891-2

14. Sala E, Mema E, Himoto Y, Veeraraghavan H, Brenton JD, Snyder A, et al. Unravelling tumour heterogeneity using next-generation imaging: radiomics, radiogenomics, and habitat imaging. *Clin Radiol*. (2017) 72:3–10. doi: 10.1016/j.crad.2016.09.013

15. Rizzo S, Botta F, Raimondi S, Origgi D, Buscarino V, Colarieti A, et al. Radiomics of high-grade serous ovarian cancer: association between quantitative CT features, residual tumour and disease progression within 12 months. *Eur Radiol*. (2018) 28:4849–59. doi: 10.1007/s00330-018-5389-z

16. Li H, Zhang R, Li R, Xia W, Chen X, Zhang J, et al. Noninvasive prediction of residual disease for advanced high-grade serous ovarian carcinoma by MRI-based radiomic-clinical nomogram. *Eur Radiol*. (2021) 31:7855–64. doi: 10.1007/s00330-021-07902-0

17. Chen HZ, Wang XR, Zhao FM, Chen XJ, Li XS, Ning G, et al. The development and validation of a CT-based Radiomics nomogram to preoperatively predict lymph node metastasis in high-grade serous ovarian Cancer. *Front Oncol*. (2021) 11:711648. doi: 10.3389/fonc.2021.711648

18. Soufi M, Arimura H, Nagami N. Identification of optimal mother wavelets in survival prediction of lung cancer patients using wavelet decomposition-based radiomic features. *Med Phys*. (2018) 45:5116–28. doi: 10.1002/mp.13202

19. Lu J, Li HM, Cai SQ, Zhao SH, Ma FH, Li YA, et al. Prediction of platinum-based chemotherapy response in advanced high-grade serous ovarian cancer: ADC histogram analysis of primary tumors. *Acad Radiol*. (2021) 28:e77–85. doi: 10.1016/j.acra.2020.01.024

20. Zhang CL, Jiang XC, Li Y, Pan X, Gao MQ, Chen Y, et al. Independent predictive value of blood inflammatory composite markers in ovarian cancer: recent clinical evidence and perspective focusing on NLR and PLR. *J Ovarian Res*. (2023) 16:36. doi: 10.1186/s13048-023-01116-2

21. Xiao ML, Fu L, Wei Y, Liu AE, Cheng JJ, Ma FH, et al. Intratumoral and peritumoral MRI radiomics nomogram for predicting parametrial invasion in patients with early-stage cervical adenocarcinoma and adenosquamous carcinoma. *Eur Radiol*. (2023). doi: 10.1007/s00330-023-10042-2

22. Ai Y, Zhang J, Jin J, Zhang J, Zhu H, Jin X. Preoperative prediction of metastasis for ovarian cancer based on computed tomography radiomics features and clinical factors. *Front Oncol*. (2021) 11:610742. doi: 10.3389/fonc.2021.610742

23. Yi X, Liu Y, Zhou B, Xiang W, Deng A, Fu Y, et al. Incorporating SULF1 polymorphisms in a pretreatment CT-based radiomic model for predicting platinum resistance in ovarian cancer treatment. *Biomed Pharmacother*. (2021) 133:111013. doi: 10.1016/j.biopha.2020.111013

24. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. *CA Cancer J Clin*. (2019) 69:127–57. doi: 10.3322/caac.21552

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 01 March 2024
DOI 10.3389/fmed.2024.1328073

Check for updates

OPEN ACCESS

EDITED BY
Gongning Luo,
Harbin Institute of Technology, China

REVIEWED BY
Alexandru Florea,
University Hospital RWTH Aachen, Germany
Nandavar Shobha,
Manipal Hospitals, India

*CORRESPONDENCE
Jing Liu
✉ liujing0198@163.com
Peijun Wang
✉ tongjipjwang1960@163.com

RECEIVED 26 October 2023
ACCEPTED 20 February 2024
PUBLISHED 01 March 2024

CITATION
Wei L, Pan X, Deng W, Chen L, Xi Q, Liu M, Xu H, Liu J and Wang P (2024) Predicting long-term outcomes for acute ischemic stroke using multi-model MRI radiomics and clinical variables.
Front. Med. 11:1328073.
doi: 10.3389/fmed.2024.1328073

COPYRIGHT
© 2024 Wei, Pan, Deng, Chen, Xi, Liu, Xu, Liu and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Predicting long-term outcomes for acute ischemic stroke using multi-model MRI radiomics and clinical variables

Lai Wei[1,2], Xianpan Pan[3], Wei Deng[3], Lei Chen[3], Qian Xi[4], Ming Liu[5], Huali Xu[6], Jing Liu[7]* and Peijun Wang[1,2]*

[1]Department of Medical Imaging, Tongji Hospital, School of Medicine, Tongji University, Shanghai, China, [2]Institute of Medical Imaging Artificial Intelligence, Tongji University School of Medicine, Shanghai, China, [3]Department of Research United Imaging Intelligence Co., Ltd., Shanghai, China, [4]Department of Radiology, Shanghai East Hospital, Tongji University School of Medicine, Shanghai, China, [5]Department of Radiology, Xinhua Hospital Affiliated to Shanghai Jiaotong University School of Medicine, Shanghai, China, [6]Department of Radiology, Putuo Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China, [7]Department of Radiology, Zhabei Central Hospital, Shanghai, China

Purpose: The objective of this study was to create and validate a novel prediction model that incorporated both multi-modal radiomics features and multi-clinical features, with the aim of accurately identifying acute ischemic stroke (AIS) patients who faced a higher risk of poor outcomes.

Methods: A cohort of 461 patients diagnosed with AIS from four centers was divided into a training cohort and a validation cohort. Radiomics features were extracted and selected from diffusion-weighted imaging (DWI) and apparent diffusion coefficient (ADC) images to create a radiomic signature. Prediction models were developed using multi-clinical and selected radiomics features from DWI and ADC.

Results: A total of 49 radiomics features were selected from DWI and ADC images by the least absolute shrinkage and selection operator (LASSO). Additionally, 20 variables were collected as multi-clinical features. In terms of predicting poor outcomes in validation set, the area under the curve (AUC) was 0.727 for the DWI radiomics model, 0.821 for the ADC radiomics model, 0.825 for the DWI + ADC radiomics model, and 0.808 for the multi-clinical model. Furthermore, a prediction model was built using all selected features, the AUC for predicting poor outcomes increased to 0.86.

Conclusion: Radiomics features extracted from DWI and ADC images can serve as valuable biomarkers for predicting poor clinical outcomes in patients with AIS. Furthermore, when these radiomics features were combined with multi-clinical features, the predictive performance was enhanced. The prediction model has the potential to provide guidance for tailoring rehabilitation therapies based on individual patient risks for poor outcomes.

KEYWORDS

acute ischemic stroke (AIS), prognosis, radiomics, diffusion-weighted imaging (DWI), apparent diffusion coefficient (ADC), machine learning (ML)

1 Introduction

Acute ischemic stroke (AIS) is a globally prevalent condition that ranks among the leading causes of disability and mortality, accounting for a staggering 60–80% of all stroke incidents (1, 2). The middle cerebral artery (MCA) territory is the most common site for AIS (3, 4). The outcomes of AIS are influenced by various factors related to patient differences, such as demographics, general health conditions, and the extent of cerebral infarction (5, 6). Predicting the prognosis of AIS quickly and accurately is essential for determining appropriate clinical management strategies (7, 8).

Radiomics (RA) is a discipline that extracts quantitative and high-dimensional features from medical images (9, 10). These features are indistinguishable to the naked eyes, but they may contain information related to the pathophysiology of diseases (11, 12). Currently, the role of RA was explored in the prediction of early outcome and long-term prognosis of AIS (13–17). However, most studies primarily predicted the outcomes of AIS based on a limited sample size of patients from either a single hospital or two hospitals (13–17). Additionally, the majority of these studies have focused on Computed Tomography (CT) images or a single-modality Magnetic Resonance Imaging (MRI) images (13, 14). Very few studies have investigated long-term outcomes for AIS using multi-modalities MRI images (15, 17), and only a few studies have developed combination models that integrated clinical and radiomic features with standard validation process (15, 16).

In this retrospective multicenter study, we developed a prediction model for long-term outcome of AIS in middle cerebral artery territory (MCA-AIS) using a combination of multi-model MRI images and multiple clinical variables from four medical centers. We extracted and selected 49 radiomics features from Diffusion-weighted imaging (DWI) and Apparent diffusion coefficient (ADC) images, and incorporated various clinical variables, such as general information, medical history, neurological scores, neuroimaging score, and laboratory examinations. Using machine learning (ML) techniques, we established models to rapidly and accurately predict the long-term outcomes of AIS. To ensure the robustness of our model, we furtherly validated it using comprehensive evaluations.

2 Methods

2.1 Study population

This study was a retrospective multi-center investigation. The study received approval from the Ethics Review Committee of Tongji Hospital in Shanghai (Approval No. K-2020 021), written informed consent for participation was not required for this study in accordance with national legislation and the institutional requirements. We pooled individual patient-level data from patients with AIS admitted to Tongji Hospital affiliated to Tongji University, Xinhua Hospital affiliated to the School of Medicine of Shanghai Jiaotong University, East Hospital affiliated to Tongji University and Putuo Hospital affiliated to Shanghai University of Traditional Chinese Medicine from January 2018 to December 2021. The admission criteria are as follows: (1) patients who had brain MRI (including DWI and ADC images) examination within 3 days after symptom onset; (2) initially diagnosed MCA-AIS patients who were admitted to the hospital for treatment; (3) patients who underwent DWI imaging for depicting lesions with a maximum diameter more than 1.5 cm; (4) initially diagnosed MCA-AIS patients who were admitted to the hospital for standard stroke treatment. The exclusion criteria were as follows: (1) patients with AIS involving posterior circulation area; (2) patients with AIS involving the anterior cerebral artery region; (3) patients with lacunar infarcts; (4) patients with poor quality images. A total of 1,675 AIS patients were included, and 1,316 patients were excluded due to posterior cerebral AIS ($n = 293$), anterior and posterior AIS ($n = 112$), anterior lacunar AIS ($n = 516$), anterior cerebral artery cerebral AIS ($n = 262$), and image artifacts ($n = 31$). Finally, 461 cases met the inclusion criteria. All included patients were randomly divided into training cohort (411) and validation cohort (50). A flowchart of the patient selection and study process was provided in Figure 1.

2.2 Data collection

2.2.1 Multi-model MRI images

The MRI-DWI images were obtained using four different MRI scanners. The acquisition parameters were as follows: (1) Philips Ingenia 3.0 T: TR = 2,584 ms, TE = 96.7 ms, slice thickness 6 mm, slice spacing 7 mm, field of view 23 cm × 23 cm, matrix 256 × 256, excitation times 2, echo gap 0.75 ms, b value 1,000 s/mm^2; (2) Siemens Verio 3.0 T: TR = 4,600 ms, TE = 89 ms, slice thickness 5 mm, scanning without spacing, field of view 24 cm × 24 cm, matrix 256 × 256, echo gap 0.75 ms, b value 1,000 s/mm^2; (3) uMR 1.5 T: TR = 5,400 ms, TE = 94 ms, slice thickness 5 mm, layer spacing 6 mm, field of view 23 cm × 23 cm, echo gap 0.75 ms, b value 1,000 s/mm^2; (4) GE SIGNA EXCITE 1.5 T: TR = 6,000 ms, TE = 81.1 ms, slice thickness 7 mm, slice spacing 8 mm, field of view 23 cm × 23 cm, matrix 256 × 256, excitation times 2, echo gap 0.75 ms, b value 1,200 s/mm^2. The ADC images were automatically created from DWI scans using built-in software.

2.2.2 Multi-clinical variables

The following 20 clinical data were collected: (1) General information: gender and age; (2) Medical history: history of smoking, history of alcohol, history of diabetes, history of myocardial infarction, history of coronary atherosclerosis, history of atrial fibrillation, history of hypertension and history of stroke; (3) Neurological score scale: National Institutes of Health Stroke Scale (NIHSS) on admission; (4) Neuroimaging score: diffusion weighted imaging-Alberta Stroke Program Early CT Score (DWI-ASPECTS); (5) Laboratory tests on admission: prothrombin time (PT), fibrinogen, D-dimer, serum Troponin I, blood glucose, blood lipids, and plasma brain natriuretic peptide (BNP).

2.3 Image preprocessing and delineation

Three attending neuro-radiologists manually delineated the ischemic lesions on MRI-DWI images using ITK-SNAP software (Version 3.8.0, available at http://www.itksnap.org). The ischemic lesion volume of interest (VOI) was also replicated from the DWI images onto another parametric map (ADC) and further refined by the radiologists. Finally, all the delineations were reviewed by two

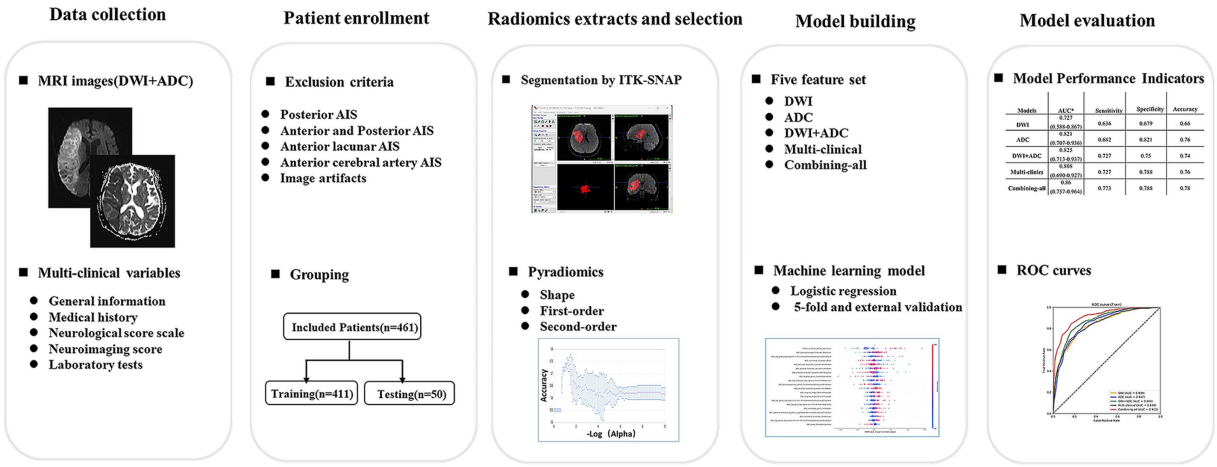


FIGURE 1
Flowchart of study patients and process. DWI = diffusion-weighted imaging, AIS = acute ischemic stroke.

chief radiologists with 8 years of experience in brain imaging. All parametric maps underwent a normalization using maximum and minimum truncation processing.

2.4 Radiomics extraction and selection

The flowchart of radiomics analysis was shown in Figure 1. 14 image filters (such as BoxMean, AdditiveGaussianNoise, BinomialBlurImage, CurvatureFlow, BoxsigmaImage, LoG with sigma values of 0.5, 1, 1.5, and 2), Wavelet filters (LLL, LLH, LHL, LHH, HLL, HLH, HHL, HHH), Normalize, LaplacianSharpening, DiscreteGaussian, Mean, SpeckleNoise RecursiveGaussian and ShotNoise were used to generate derived images. From these derived images, first-order statistics and texture features were extracted. A total of 2,264 radiomics features were automatically extracted from each ischemic lesion. These features can be categorized into three groups: 14 shape features, 450 first-order features that quantify the distribution of voxel intensities in the images, and 1800 texture features. The texture features consist of 525 gray level co-occurrence matrix (GLCM) features, 350 gray level run length matrix (GLRLM) features, 400 gray level size zone matrix (GLSZM) features, 400 neighboring gray tone difference matrix (NGTDM) features, and 125 gray level dependent matrix (GLDM) features. These texture features capture regional heterogeneity differences. All radiomics features were normalized using Z-score.

We employed LASSO selection to identify the most reliable predictive radiomic features. Initially, we performed feature selection separately for each sequence of DWI and ADC modalities. Then, an additional round of LASSO selection was conducted to combine the selected features from both modalities, resulting in a set of multi-modality RA features. These multi-modality RA features, along with clinical features, were subsequently merged and subjected to another round of LASSO selection to obtain a comprehensive combined feature set.

Based on Harrell's guideline, the number of selected features should be less than 10% of the sample size. Consequently, in our experiment involving the DWI sequence, ADC sequence, multi-modality sequence, and the final combination of radiomics with clinical features, the final number of selected features was approximately 30.

2.5 Prediction model

2.5.1 Predictive task

The objective of our predictive task was to accurately predict the long-term prognosis of initially diagnosed MCA-AIS patients. The long-term prognosis was defined based on a 90-day modified Rankin Scale (90d-mRS) score, where scores of 0–2 indicated a good outcome and scores of 3–6 indicated a poor outcome. The majority of the 90-d mRS data were collected through telephone interviews, outpatient care, and clinical medical records. During phone interviews, patients were asked about their functional recovery 90 days after therapy.

2.5.2 Development and validation of the predictive model

Based on multi-model MRI RA features and/or multi-clinical features, three machine learning models were constructed for binary classification (good outcome or poor outcome) by using three classifiers, namely random forest (RF), support vector machine (SVM), and logistic regression (LR). The prediction model utilized input data from one of five feature sets: (1) DWI RA features with 25 variables, (2) ADC RA features with 24 variables, (3) DWI + ADC RA features with 35 variables, (4) Multi-clinical features with 12 variables, and (5) Combining-all features with 30 variables. To optimize performance, a grid search was conducted on different features and classification algorithms for parameter tuning.

2.6 Statistical analysis

Mann–Whitney U test and chi-square test were used for evaluating significant differences in the variables (such as age, NIHSS score) between the training set and the validation set. The receiver operating characteristic curve (ROC) was drawn, and various performance metrics including sensitivity (SEN), specificity (SPE), accuracy (ACC), F1-Score, and area under the curve (AUC) were

calculated to assess the model's performance. The Shapley additive explanation (SHAP) diagram was utilized for model explanation. A two-tailed statistical test was used and p-value lower than 0.05 was considered to be statistically significant. The R software package (version 4.0.3) was used to process the demographic data for evaluating significant differences in the variables between the training set and the validation set. Python (version 3.6) was used for programming the training, validation of the prediction model, as well as conducting statistical analysis.

3 Results

3.1 Basic characteristics

As shown in Table 1, the basic variables of most of the patients showed no statistical differences ($p > 0.05$) between the training set and the validation set, such as general conditions (gender and age), medical history (hypertension, diabetes), neurological score scales (NIHSS), and laboratory tests (BNP, etc.).

3.2 Assessment of radiomic features

A total of 4,528 radiomics features were extracted from DWI and ADC images. The optimal feature subset for the machine learning models consisted of 49 radiomic features, with 25 features selected from DWI and 24 features selected from ADC. These features were comprised of 4 shape features, 16 first-order features, and 29 texture features. The detailed information about the features based on DWI + ADC model was presented in Figure 2A. Rad-score was calculated according to the coefficient of the selected features, and the distribution of rad-score between good and poor outcome was shown both in train (Figure 2B) and test (Figure 2C) sets.

3.3 Comparison of prediction models

3.3.1 Comparison between different models

The LR model achieved the best classification results in all feature sets in our study. Table 2 and Figure 3 illustrated the AUC values along with other diagnostic performance metrics such as specificity,

TABLE 1 Basic patient information.

	Training set (n = 411)	Validation set (n = 50)	p-values
Basic characteristics			
Age (Median, IQR)	71 (63, 82)	65 (55.25, 83.75)	0.194
Male (Percentile: %)	259 (56.2%)	25 (5.4%)	0.074
Neurological score scale (Median, IQR)			
NIHSS on admission	5 (5, 10)	5 (3.25, 8.75)	0.253
Location (Left: Percentile: %)	228 (49.5%)	26 (5.6%)	0.641
Neuroimaging score scale (Median, IQR)			
DWI-ASPECTS	8 (6, 9)	8 (6, 8.75)	0.752
History (Percentile: %)			
Alcohol	106 (23%)	11 (2.4%)	0.561
Smoking	167 (36.2%)	17 (3.7%)	0.366
Myocardial infarction	405 (87.9%)	49 (10.6%)	0.555
Coronary atherosclerosis	86 (18.7%)	7 (1.5%)	0.249
Atrial fibrillation	74 (16.1%)	10 (2.2%)	0.730
Hypertension	295 (64%)	32 (6.9%)	0.253
Stroke	103 (22.3%)	14 (3%)	0.652
Diabetes	135 (29.3%)	14 (3%)	0.489
Laboratory test (Median, [IQR])			
Prothrombin time	11.5 (10.9, 12.3)	11.25 (10.8, 11.675)	0.023
Fibrinogen	2.97 (2.55, 3.78)	2.8 (2.405, 3.475)	0.072
D-dimer	0.56 (0.27, 1.355)	0.54 (0.25, 1.145)	0.526
Serum troponin I	0.01 (0.01, 0.0305)	0.01 (0.01, 0.019)	0.048
Blood sugar	6.46 (5.605, 8.595)	6.185 (5.235, 7.145)	0.035
Blood lipids	1.21 (0.96, 1.61)	1.21 (1.04, 1.5775)	0.839
Brain natriuretic peptide	103.1 (64.15, 269.3)	103.1 (58.275, 154.95)	0.395
Long-term outcome			
Poor outcome (90d-mRS>2) (Percentile: %)	184 (39.9%)	22 (4.8%)	0.918

IQR: interquartile range; NIHSS: National Institute of Health stroke scale; 90d-mRS:90 days-modified Rankin scale.

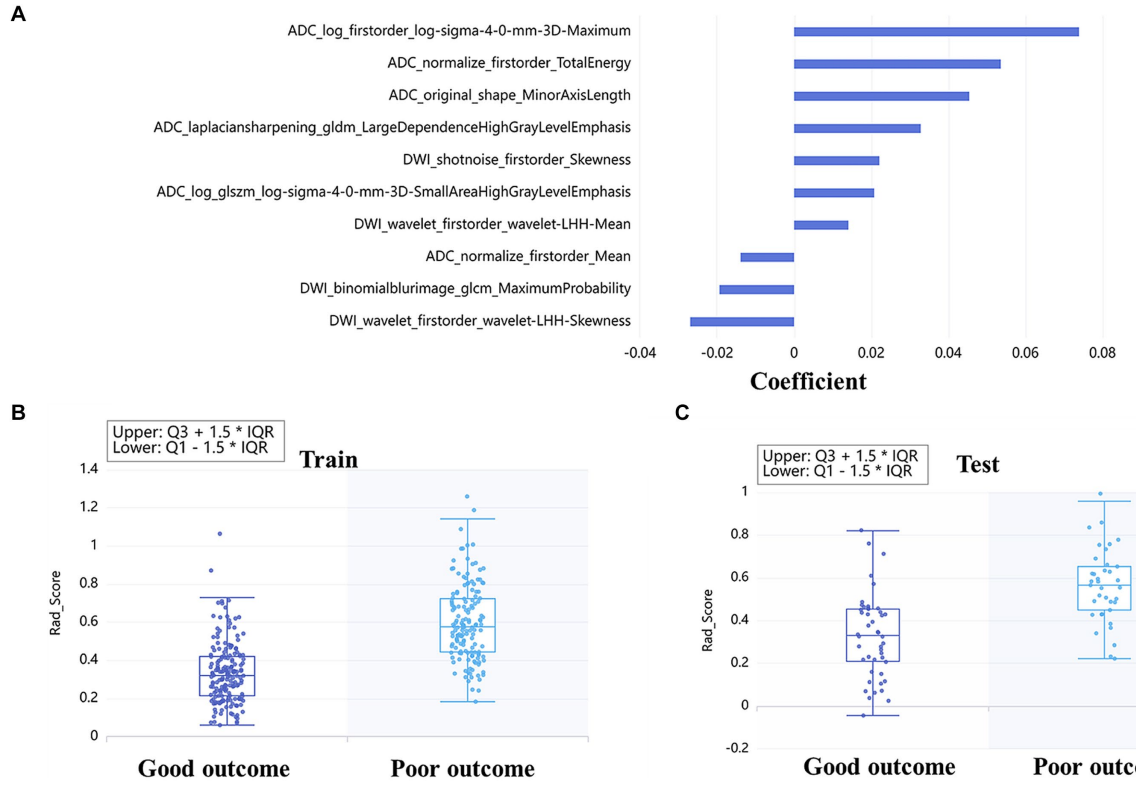


FIGURE 2
Radiomics based on DWI + ADC feature set for long-term prognosis prediction: the coefficient of each feature was taken to make a bar chart, which showed the importance of the top 10 feature in the prediction **(A)**. The box diagrams of rad-score for discriminating good and poor outcomes in the train **(B)** and test sets **(C)**.

sensitivity, accuracy, and F1 Score, which demonstrated the indicator results for predicting poor outcomes in the training set, test set, and validation set. In the test set for predicting poor outcome using LR model, the AUCs were as follows: DWI RA model 0.805, ADC RA model 0.823, DWI + ADC RA model 0.838, multi-clinical model 0.808. When combining multi-clinical features and RA features, the AUC was significantly increased, reaching to 0.873. In the validation set for predicting poor outcome, the AUCs were as follows: DWI RA model 0.727, ADC RA model 0.821, ADC + DWI RA model 0.825, multi-clinical model 0.808. When combining all the features, the AUC value was increased to 0.86, which means the model with combining-all features achieved superior diagnostic performance compared to other models.

3.3.2 Model interpretability

We generated a nomogram to predict the probability of long-term outcomes using the multi-clinical feature set (Figure 4). It showed that patients with higher NIHSS on admission, a history of myocardial infarction, and lower DWI-ASPECTS were at greater risks for poor outcome.

The Shap values corresponding to each feature in combining-all model were also calculated. In each prediction, a positive Shap value denoted an elevated risk of poor outcome, while a negative value suggested the opposite. The accompanying Figure 5A presented the average Shap values for each feature within the test set. Notably, the NIHSS admission emerged as the most influential predictor in forecasting long-term outcomes. Alongside, several RA features, such as ADC_log_boxsigmaimage_firstorder_maximum, ADC_normalize_firstorder_mean, and DWI_Wavelet_firstorder_wavelet-lhh-mean also played an important role in this predictive model. The detailed SHAP values of the most important variables for one typical patient from the validation group (poor outcome) was illustrated in Figure 5B.

4 Discussion

In this retrospective multicenter study, we developed a logistic regression model based on DWI RA features, ADC RA features, and multi-clinical factors to predict long-term outcomes in patients with AIS. Our model was applicable to MCA-AIS patients receiving different therapies and provided preferable accuracy. It was worth mentioning that our study just conformed to the "big data" trend of medicine which took radiomics as a block of "big data".

RA can provide quantitative morphological and texture features based on voxel level while our naked eye can only distinguish 16 gray scales (18). In this context, regarding the heterogeneity of AIS lesions, radiomics seems to be superior to conventional imaging visual analysis (19, 20). As we know, ADC images can more accurately reflect diffusion restriction than DWI images without the influence of T2 shining-through effect. In this study, we found that the prediction model with ADC RA features performed better than the model with DWI RA features that was consistent with the principle of diffusion sequence imaging mentioned above, and a previous study has also yielded similar

TABLE 2 The performance of the prediction models.

	Models	AUC*		Sensitivity		Specificity		Accuracy		F1 score	
Training (n = 411)		Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
	DWI	0.839 (0.797–0.882)	0.805 (0.711–0.903)	0.776	0.73	0.763	0.748	0.769	0.74	0.751	0.71
	ADC	0.847 (0.806–0.889)	0.823 (0.732–0.915)	0.791	0.762	0.746	0.734	0.766	0.747	0.752	0.728
	DWI + ADC	0.868 (0.83–0.907)	0.838 (0.753–0.925)	0.85	0.838	0.738	0.721	0.788	0.774	0.783	0.768
	Multi-clinics	0.838 (0.796–0.882)	0.808 (0.714–0.903)	0.759	0.724	0.764	0.752	0.762	0.74	0.742	0.712
	Combining-all	0.912 (0.883–0.944)	0.873 (0.802–0.949)	0.847	0.805	0.815	0.801	0.83	0.803	0.817	0.783
Validation (n = 50)											
	DWI	0.727 (0.588–0.867)		0.636		0.679		0.66		0.622	
	ADC	0.821 (0.707–0.936)		0.682		0.821		0.76		0.714	
	DWI + ADC	0.825 (0.713–0.937)		0.727		0.75		0.74		0.711	
	Multi-clinics	0.808 (0.690–0.927)		0.727		0.788		0.76		0.727	
	Combining-all	0.86 (0.757–0.964)		0.773		0.788		0.78		0.756	

*AUC = area under the receiver operating characteristic curve; DWI = diffusion-weighted imaging; ADC = apparent diffusion coefficient.

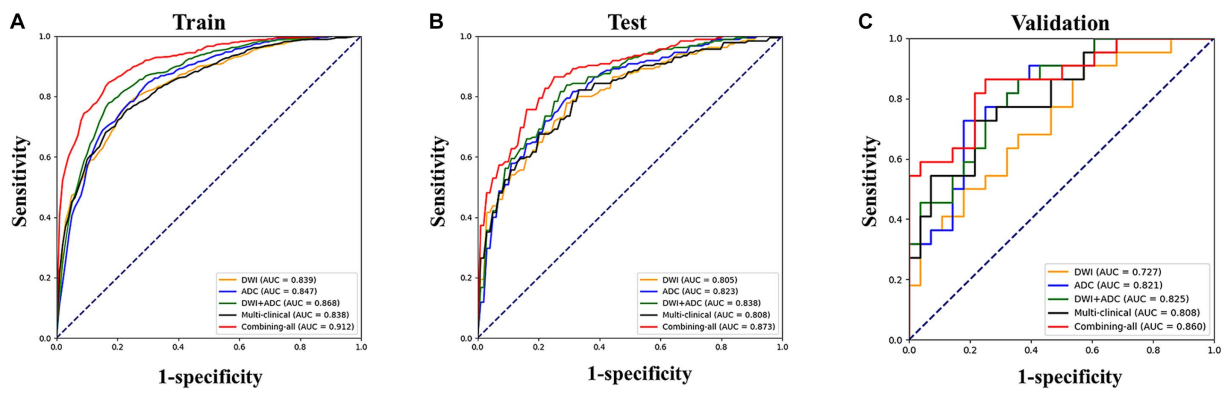


FIGURE 3
Performances of machine learning models for prediction of outcome: Receiver operating characteristic (ROC) curves of five feature sets when using the logistic regression classifier in the train **(A)**, test **(B)** and validation **(C)** set, respectively.

results (21). First-order features (ADC_boxsigmaimage_firstorder_Maximum) was positively correlated to infarction core volume, which was considered to be critical factors for stroke severity and treatment plan in the guidelines. And Vogt et al. have reported that the initial lesion volume of cerebral infarction acted as an independent predictor of prognosis (90d-Rankin score) (22). Our study included MCA-AIS cases in 72 h from onset, predominantly capturing patients in the acute-subacute phase. There was vascular edema and/or cytotoxic edema in acute-subacute cerebral infarction, that was the pathophysiology mechanism of the signal elevation on DWI images and signal reduction on ADC maps. First-order feature (ADC_normalize_firstorder_Mean, DWI_wavelet_firstorder_wavelet-LHH-Mean) showed the average gray level intensity within the infarction core, and both higher gray level intensity on DWI and lower gray level intensity on ADC reflected more severe overall diffusion restriction within the lesion, suggesting a higher grade of overall edema. Consequently, we hypothesized that the voxel-based diffusion restriction heterogeneity represented the progress rate of blood–brain barrier destruction. First-order feature (ADC_normalize_firstorder_totalenergy) measured the magnitude of voxel values in images, with larger values indicating a higher sum of the squares of these values. This metric suggested that the infarction core on ADC images of patients with poor outcomes had more heterogeneity. A two-center study showed that infarction lesion homogeneity of DWI images indicated favorable outcomes, which was similar with our results (16).

This study collected multiple-dimension clinical variables, including general information, medical history, neuroimaging scores, and laboratory test, which was different from previous studies (13–17). We observed that NIHSS score on admission remained associated with the risks for poor outcomes whether in multi-clinical model or combing-all model. The National Institute of Health Stroke Scale (NIHSS) is the most commonly used clinical score (23), which quantitatively and comprehensively evaluates the functional impairment in stroke patients. A history of myocardial infarction, indicative of underlying atherosclerotic disease, and a shortened prothrombin time, suggestive of hypercoagulability, are both significant risk factors for the onset and progression of AIS (24). We also found that DWI-ASPECTS played an important role in the multi-clinical model. It is a 10-point semi-quantitative scoring system

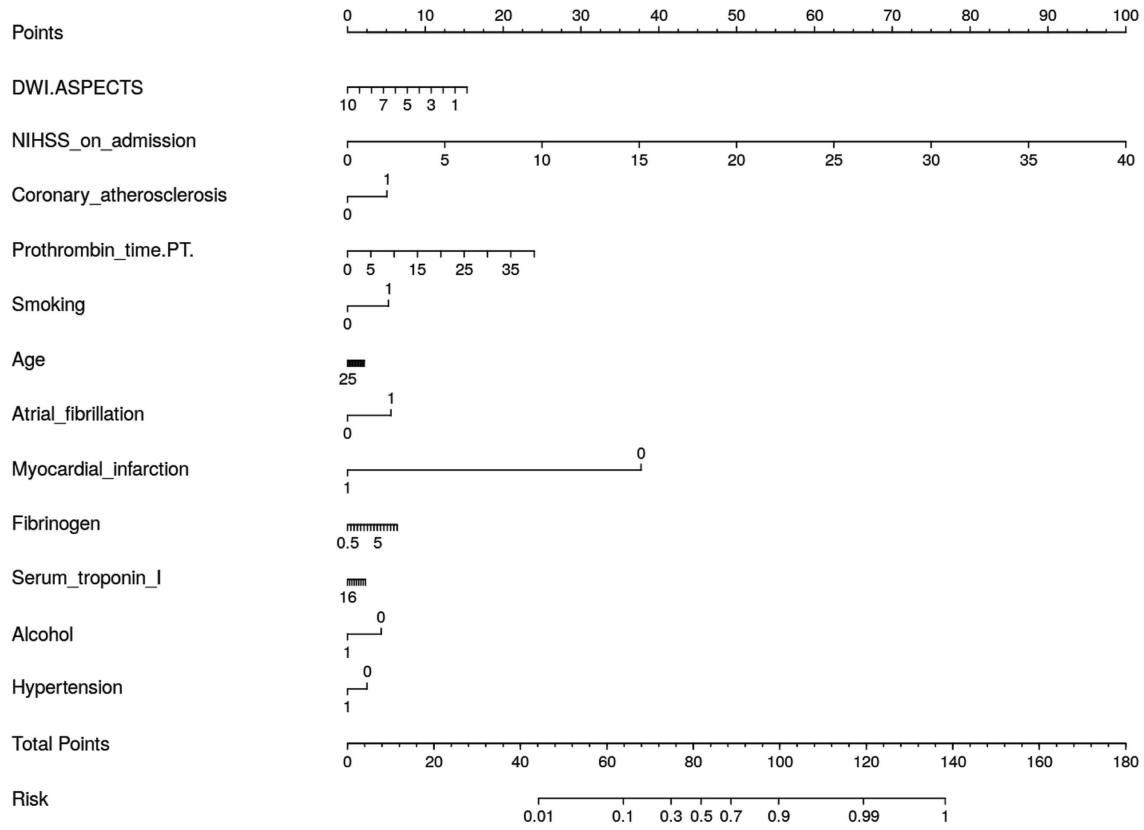


FIGURE 4
Nomogram for the Multi-Clinical model for long-term prognosis prediction.

for assessing the degree of ischemic changes (25, 26). ASPECTS has been widely utilized to identify patients that presumed to have a large ischemic core and high risks for intracerebral hemorrhage and poor clinical outcome (27, 28). These findings are consistent with the current guidelines and consensus for the diagnosis and therapy of AIS (29, 30). However, medical history and laboratory test contributed little to the prediction models in our study.

Despite the favorable prognostic efficacy of the combining model, our research still has some limitations. First, a more extensive and prospective study cohort is needed to generalize the performance of the prediction model in the future. However, compared with most previous studies (13–17), our sample size had certain advantages, especially the four-center characteristic. Second, reperfusion factors, such as collateral circulation and vascular recanalization, have not been investigated as a variable. Third, when collecting the data, the lacunar cerebral infarction patients with good prognosis were excluded, which meant that this study did not include all clinically common cases of AIS.

5 Conclusion

Our findings highlighted the utility of radiomics based on DWI and ADC images in predicting long-term outcomes in patients with MCA-AIS. The prediction model, which incorporated multi-clinical variables along with ADC+DWI RA features, demonstrated the highest efficiency in the prediction of long-term outcomes for AIS. This model has the potential to assist clinicians in offering personalized management strategies for optimal patient care.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Ethics Committee of Shanghai Tongji Hospital. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

LW: Conceptualization, Data curation, Formal analysis, Methodology, Writing – original draft, Investigation, Visualization. XP: Formal analysis, Methodology, Validation, Visualization, Writing – original draft. WD: Methodology, Validation, Visualization, Writing – original draft. LC: Conceptualization,

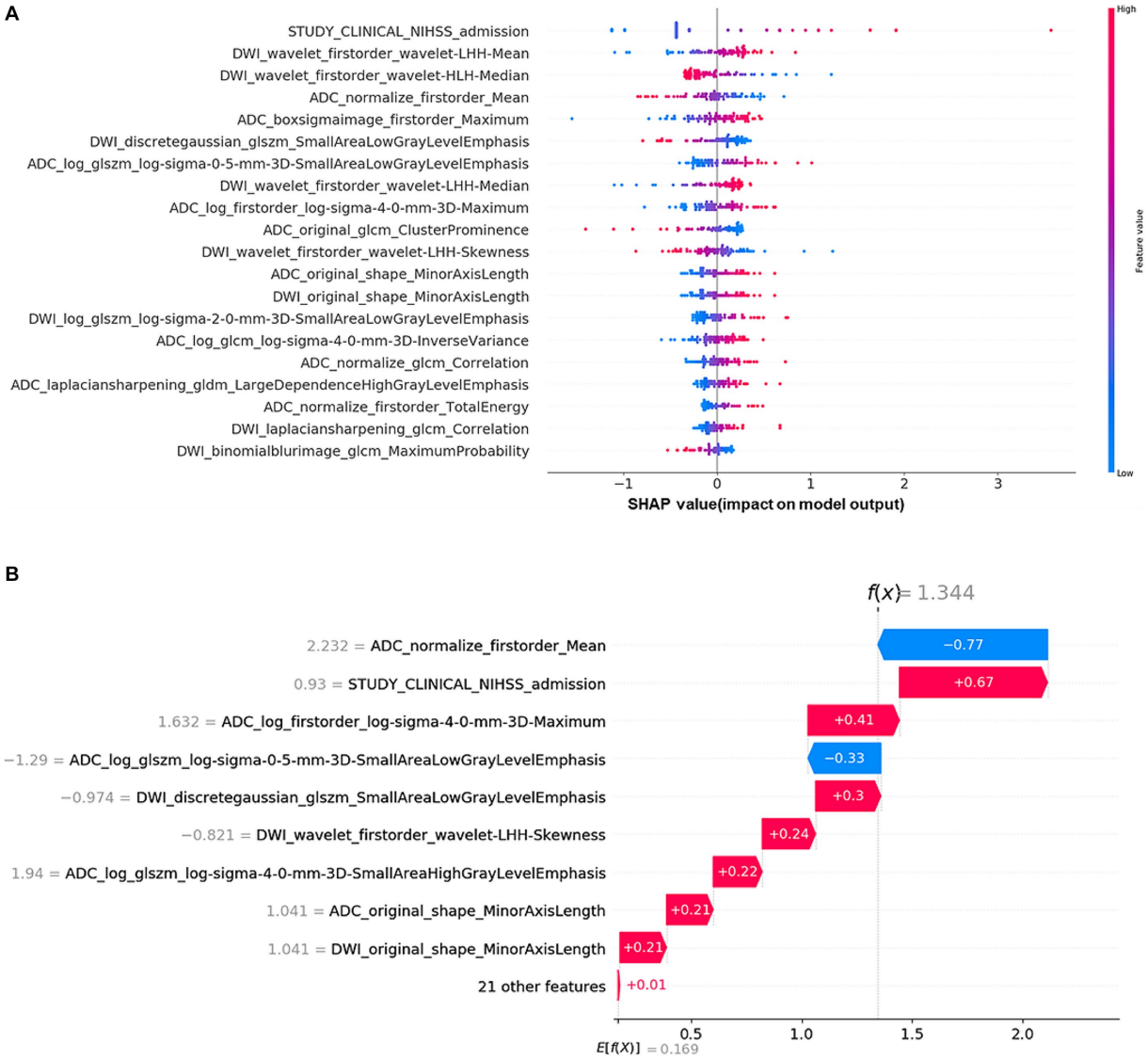


FIGURE 5
Shapley additive explanation (SHAP) diagram of variable contributions for the optimal (combining-all) predictive model: **(A)** The relative contributions of RA features and multi-clinical variables for long-term prognosis prediction. The color intensity of the graph revealed a discernible pattern: an increment in the NIHSS admission score corresponded to an escalating probability of a poor outcome. For the DWI_Wavelet_firstorder_wavelet-lhh-mean feature, the lower the value of the feature, the higher likelihood of a poor outcome. **(B)** SHAP values of a typical patient from the positive group (poor outcome), illustrated with the important variables.

Methodology, Supervision, Writing – original draft. QX: Data curation, Investigation, Writing – review & editing. ML: Data curation, Investigation, Writing – review & editing. HX: Data curation, Investigation, Writing – review & editing. JL: Conceptualization, Project administration, Supervision, Writing – review & editing. PW: Conceptualization, Funding acquisition, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by the Major project of the National Natural Scientific Foundation of China (grant no. 82227807), the Research Project of Shanghai Municipal Health Commission (grant no. 2022JC017) and the National Key R&D Program of the Ministry of Science and Technology (grant no. 2022YFC2009904).

Conflict of interest

XP, WD, and LC are employed by Shanghai United Imaging Intelligence Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Darwish EAF, Abdelhameed-El-Nouby M, Geneidy E. mapping the ischemic penumbra and predicting stroke progression in acute ischemic stroke: the overlooked role of susceptibility weighted imaging. *Insights Imaging*. (2020), 11:6. doi: 10.1186/s13244-019-0810-y

2. Sohn B, Won SY. Quality assessment of stroke radiomics studies: promoting clinical application. *Eur J Radiol*. (2023) 161:110752. doi: 10.1016/j.ejrad.2023.110752

3. Nam KW, Kwon HM, Lee YS. Infarct growth velocity predicts early neurological outcomes in single subcortical infarction. *Sci Rep*. (2023) 13:4511. doi: 10.1038/s41598-023-31727-0

4. Dragoș HM, Stan A, Pintican R, Feier D, Lebovici A, Panaitescu PS, et al. MRI Radiomics and predictive models in assessing ischemic stroke outcome—a systematic review. *Diagnostics (Basel)*. (2023) 13:857. doi: 10.3390/diagnostics13050857

5. Handelsmann H, Herzog L, Kulcsar Z, Luft AR. Wegener S (2021) Predictors for affected stroke territory and outcome of acute stroke treatments are different for posterior versus anterior circulation stroke. *Sci Rep*. (2021) 11:10544. doi: 10.1038/s41598-021-89871-4

6. Bu N, Khlif MS, Lemmens R, Wouters A, Fiebach JB, Chamorro A, et al. Imaging markers of brain frailty and outcome in patients with acute ischemic stroke. *Stroke*. (2021) 52:1004–11. doi: 10.1161/STROKEAHA.120.029841

7. Bretzner M, Bonkhoff AK, Schirmer MD, Hong S, Dalca A, Donahue K, et al. Radiomics-derived brain age predicts functional outcome after acute ischemic stroke. *Neurology*. (2023) 100:e822–33. doi: 10.1212/WNL.0000000000201596

8. Yu H, Wang Z, Sun Y, Bo W, Duan K, Song C, et al. Prognosis of ischemic stroke predicted by machine learning based on multi-modal MRI radiomics. *Front Psychiatry*. (2022) 13:1105496. doi: 10.3389/fpsyt.2022.1105496

9. Zhang K, Wei L, Zhou X, Yang B, Meng J, Wang P. Risk factors for poor outcomes of spontaneous supratentorial cerebral hemorrhage after surgery. *J Neurol*. (2022) 269:3015–25. doi: 10.1007/s00415-021-10888-w

10. Zhang K, Zhou X, Xi Q, Wang X, Yang B, Meng J, et al. Outcome prediction of spontaneous Supratentorial intracerebral hemorrhage after surgical treatment based on non-contrast computed tomography: a multicenter study. *J Clin Med*. (2023) 12:1580. doi: 10.3390/jcm12041580

11. Wei L, Cao Y, Zhang K, Xu Y, Zhou X, Meng J, et al. Prediction of progression to severe stroke in initially diagnosed anterior circulation ischemic cerebral infarction. *Front Neurol*. (2021) 12:652757. doi: 10.3389/fneur.2021.652757

12. Lin R, Lin J, Xu Y, Yu Y, Foster E, Lin M, et al. Development and validation of a novel radiomics-clinical model for predicting post-stroke epilepsy after first-ever intracerebral haemorrhage. *Eur Radiol*. (2023) 33:4526–36. doi: 10.1007/s00330-023-09429-y

13. Cui H, Wang X, Bian Y, Song S, Feng DD. Ischemic stroke clinical outcome prediction based on image signature selection from multimodality data. *Annu Int Conf IEEE Eng Med Biol Soc*. (2018) 2018:722–5. doi: 10.1109/EMBC.2018.8512291

14. Tang TY, Jiao Y, Cui Y, Zhao DL, Zhang Y, Wang Z, et al. Penumbra-based radiomics signature as prognostic biomarkers for thrombolysis of acute ischemic stroke patients: a multicenter cohort study. *J Neurol*. (2020) 267:1454–63. doi: 10.1007/s00415-020-09713-7

15. Quan G, Ban R, Ren JL, Liu Y, Wang W, Dai S, et al. FLAIR and ADC image-based Radiomics features as predictive biomarkers of unfavorable outcome in patients with acute ischemic stroke. *Front Neurosci*. (2021) 15:730879. doi: 10.3389/fnins.2021.730879

16. Jiang L, Miao Z, Chen H, Geng W, Yong W, Chen YC, et al. Radiomics analysis of diffusion-weighted imaging and long-term unfavorable outcomes risk for acute stroke. *Stroke*. (2023) 54:488–98. doi: 10.1161/STROKEAHA.122.040418

17. Wang H, Lin J, Zheng L, Zhao J, Song B, Dai YM. Texture analysis based on ADC maps and T2-FLAIR images for the assessment of the severity and prognosis of ischaemic stroke. *Clin Imaging*. (2020) 67:152–9. doi: 10.1016/j.clinimag.2020.06.013

18. Pujadas ER, Raisi-Estabragh Z, Szabo L, McCracken C, Morcillo CI, Campello VM, et al. Prediction of incident cardiovascular events using machine learning and CMR radiomics. *Eur Radiol*. (2023) 33:3488–500. doi: 10.1007/s00330-022-09323-z

19. Liu J, Tao W, Wang Z, Chen X, Wu B, Liu M. Radiomics-based prediction of hemorrhage expansion among patients with thrombolysis/thrombectomy related-hemorrhagic transformation using machine learning. *Ther Adv Neurol Disord*. (2021) 14:175628642110600. doi: 10.1177/17562864211060029

20. Bretzner M, Bonkhoff AK, Schirmer MD, Hong S, Dalca AV, Donahue KL, et al. MRI Radiomic signature of white matter Hyperintensities is associated with clinical phenotypes. *Front Neurosci*. (2021) 15:691244. doi: 10.3389/fnins.2021.691244

21. Zhang YQ, Liu AF, Man FY, Zhang YY, Li C, Liu YE, et al. MRI radiomic features-based machine learning approach to classify ischemic stroke onset time. *J Neurol*. (2022) 269:350–60. doi: 10.1007/s00415-021-10638-y

22. Vogt G, Laage R, Shuaib A, Schneider ACollaboration V. Initial lesion volume is an independent predictor of clinical stroke outcome at day 90: an analysis of the virtual international stroke trials archive (VISTA) database. *Stroke*. (2012) 43:1266–72. doi: 10.1161/STROKEAHA.111.646570

23. Choi V, Kate M, Kosior JC, Buck B, Steve T, McCourt R, et al. National Institutes of Health stroke scale score is an unreliable predictor of perfusion deficits in acute stroke. *Int J Stroke*. (2015) 10:582–8. doi: 10.1111/ijs.12438

24. Sommer P, Posekany A, Serles W, Marko M, Scharer S, Fertl E, et al. Is functional outcome different in posterior and anterior circulation stroke? *Stroke*. (2018) 49:2728–32. doi: 10.1161/STROKEAHA.118.021785

25. Meng Y, Wang H, Wu C, Liu X, Qu L, Shi Y. Prediction model of hemorrhage transformation in patient with acute ischemic stroke based on multiparametric MRI Radiomics and machine learning. *Brain Sci*. (2022) 12:858. doi: 10.3390/brainsci12070858

26. Heo J, Yoon JG, Park H, Kim YD, Nam HS, Heo JH. Machine learning-based model for prediction of outcomes in acute stroke. *Stroke*. (2019) 50:1263–5. doi: 10.1161/STROKEAHA.118.024293

27. Goyal M, Demchuk AM, Menon BK, Eesa M, Rempel JL, Thornton J, et al. Randomized assessment of rapid endovascular treatment of ischemic stroke. *N Engl J Med*. (2015) 372:1019–30. doi: 10.1056/NEJMoa1414905

28. Farzin B, Fahed R, Guilbert F, Poppe AY, Daneault N, Durocher AP, et al. Early CT changes in patients admitted for thrombectomy: Intrarater and interrater agreement. *Neurology*. (2016) 87:249–56. doi: 10.1212/WNL.0000000000002860

29. Nogueira RG, Jadhav AP, Haussen DC, Bonafe A, Budzik RF, Bhuva P, et al. Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct. *N Engl J Med*. (2018) 378:11–21. doi: 10.1056/NEJMoa1706442

30. Albers GW, Marks MP, Kemp S, Christensen S, Tsai JP, Ortega-Gutierrez S, et al. Thrombectomy for stroke at 6 to 16 hours with selection by perfusion imaging. *N Engl J Med*. (2018) 378:708–18. doi: 10.1056/NEJMoa1713973

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 06 March 2024
DOI 10.3389/fmed.2024.1356752

Check for updates

OPEN ACCESS

EDITED BY
Minjeong Kim,
University of North Carolina at Greensboro, United States

REVIEWED BY
Qingxia Wu,
Northeastern University, China
Jie-Zhi Cheng,
Shanghai United Imaging Intelligence Co., Ltd., China
Yuwei Xia,
Northwestern Polytechnical University, China
in collaboration with reviewer J-ZC

*CORRESPONDENCE
Shengxiang Rao
raoxray@163.com
Jianjun Zhou
zhoujianjunzs@126.com

†These authors have contributed equally to this work and share first authorship

‡These authors have contributed equally to this work

RECEIVED 16 December 2023
ACCEPTED 29 January 2024
PUBLISHED 06 March 2024

CITATION
Lin C, Guo Y, Huang X, Rao S and Zhou J (2024) Esophageal cancer detection via non-contrast CT and deep learning. *Front. Med.* 11:1356752. doi: 10.3389/fmed.2024.1356752

COPYRIGHT
© 2024 Lin, Guo, Huang, Rao and Zhou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Esophageal cancer detection via non-contrast CT and deep learning

Chong Lin[1,2†], Yi Guo[1,2†], Xu Huang[3], Shengxiang Rao[4*‡] and Jianjun Zhou[1,2,4*‡]

[1]Department of Radiology, Zhongshan Hospital (Xiamen), Fudan University, Shanghai, Fujian, China,
[2]Xiamen Municipal Clinical Research Center for Medical Imaging, Xiamen, Fujian, China,
[3]Departments of Thoracic Surgery, Zhongshan Hospital, Fudan University, Shanghai, China,
[4]Department of Radiology, Zhongshan Hospital, Fudan University, Shanghai, China

Background: Esophageal cancer is the seventh most frequently diagnosed cancer with a high mortality rate and the sixth leading cause of cancer deaths in the world. Early detection of esophageal cancer is very vital for the patients. Traditionally, contrast computed tomography (CT) was used to detect esophageal carcinomas, but with the development of deep learning (DL) technology, it may now be possible for non-contrast CT to detect esophageal carcinomas. In this study, we aimed to establish a DL-based diagnostic system to stage esophageal cancer from non-contrast chest CT images.

Methods: In this retrospective dual-center study, we included 397 primary esophageal cancer patients with pathologically confirmed non-contrast chest CT images, as well as 250 healthy individuals without esophageal tumors, confirmed through endoscopic examination. The images of these participants were treated as the training data. Additionally, images from 100 esophageal cancer patients and 100 healthy individuals were enrolled for model validation. The esophagus segmentation was performed using the no-new-Net (nnU-Net) model; based on the segmentation result and feature extraction, a decision tree was employed to classify whether cancer is present or not. We compared the diagnostic efficacy of the DL-based method with the performance of radiologists with various levels of experience. Meanwhile, a diagnostic performance comparison of radiologists with and without the aid of the DL-based method was also conducted.

Results: In this study, the DL-based method demonstrated a high level of diagnostic efficacy in the detection of esophageal cancer, with a performance of AUC of 0.890, sensitivity of 0.900, specificity of 0.880, accuracy of 0.882, and F-score of 0.891. Furthermore, the incorporation of the DL-based method resulted in a significant improvement of the AUC values w.r.t. of three radiologists from 0.855/0.820/0.930 to 0.910/0.955/0.965 ($p = 0.0004/<0.0001/0.0068$, with DeLong's test).

Conclusion: The DL-based method shows a satisfactory performance of sensitivity and specificity for detecting esophageal cancers from non-contrast chest CT images. With the aid of the DL-based method, radiologists can attain better diagnostic workup for esophageal cancer and minimize the chance of missing esophageal cancers in reading the CT scans acquired for health check-up purposes.

KEYWORDS

deep learning, no new net, non-contrast chest computed tomography, esophageal cancer, diagnosis

1 Introduction

Esophageal cancer is the seventh most frequently diagnosed cancer with a high mortality rate and the sixth leading cause of cancer deaths in the world (1–3). The prevalence of esophageal cancer is increasing due to the rising world population, longer longevity, and the popularity of risk factors such as tobacco and alcohol consumption (2, 4, 5). This cancer originates from the inner layer of the esophagus wall and progresses outward, which makes early detection difficult as symptoms are often absent, resulting in late-stage diagnosis and poor prognosis (2, 6). Given its high malignancy and unfavorable outcomes, timely identification is of utmost importance. While endoscopy serves as the gold standard for diagnosing esophageal cancer, its invasiveness and high cost necessitate the exploration of alternative methods to expand the reach of testing (7).

Esophageal carcinomas can manifest in several forms (8). They may appear as a focal area of mural thickening, either with or without ulceration. Another form is a flat or polypoid lesion. Finally, they can also present as generalized mural thickening. According to these characteristics, computed tomography (9) offers opportunities to detect esophageal carcinomas. With the development of medical technology, CT examination is a central modality in modern radiology contributing to diagnostic medicine in almost every medical subspecialty and has become increasingly convenient and common (10). Traditionally, contrast CT was used to detect esophageal carcinomas (8), but with the development of deep learning (DL) technology, it may now be possible for CT to detect early-stage esophageal carcinomas.

DL (6) is a type of representation learning method with complex multi-layer neural network architecture and has emerged as the state-of-the-art machine learning method in many applications (11, 12). In radiology, DL techniques have the most significant impact: lesion or disease detection (13–15), classification (16, 17), quantification, and segmentation (12, 17, 18). Examples of these applications include the identification of pulmonary nodules (19, 20) and breast cancer (21), classification of benign or malignant lung nodules (22) and breast tumors (23), utilization of texture-based radiomic features for predicting therapy response in gastrointestinal cancer (24), and segmentation of brain anatomy (25, 26).

The applications of DL methods are gradually common. However, the early detection of esophageal cancer with DL methods is relatively limited. On the other hand, since the esophagus is a hollow organ with contractile and diastolic functions, there are still several challenges in the clinical early diagnosis of esophageal cancer. The benefits and disadvantages of CT with DL to detect esophageal carcinomas are worth exploring.

In this study, we aimed to establish a DL-based diagnostic system to detect esophageal cancer from non-contrast chest CT images. There were 397 esophageal cancer patients and 250 healthy individuals enrolled to train the model. Then, 100 esophageal cancer patients and 100 healthy individuals were included for validation. We compared the diagnostic efficacy of the DL model with that of radiologists at different expertise levels, both with and without the reference to the DL model.

2 Materials and methods

2.1 Data sets

This retrospective dual-center study included non-contrast chest CT images of 397 primary esophageal cancer patients and 250 healthy individuals, collected from July 2017 to December 2022 at Zhongshan Hospital (Xiamen), for the purpose of training the model, then 100 esophageal cancer patients and 100 healthy individuals were enrolled from October 2015 to August 2019 at Zhongshan Hospital for validation (Table 1). The inclusion criteria of esophageal cancer patients were as follows: patients with pathologically proven esophageal cancer through endoscopic biopsy or surgical pathology with non-contrast chest CT images from the thoracic inlet to the esophagogastric junction and patients who had no other disease that could cause thickening of the esophageal wall, such as varicocele caused by liver cirrhosis. Non-esophageal cancer subjects were enrolled randomly from the health checkup centers and were imaged with chest CT scans. These subjects were confirmed to be negative for esophageal cancer in the following 2 years. Patients were excluded from the dataset if any clinical data was incomplete, or the quality of chest CT scans was poor.

2.2 Computed tomography image acquisition

All images were scanned by Revolution CT, GE Discovery CT750 HD, 512-slice LightSpeed VCT (GE Medical Systems), Aquilian one (Canon Medical Systems Corporation), and uCT 760, 128-slice (United imaging) with parameter setting: tube voltage as 120 kVp, tube current as 100 ~ 750 mA, image slice matrix as 512 × 512, and slice thickness as 5 mm.

2.3 CT-image convolutional neural network

The nnU-Net is a powerful neural network specifically designed for medical image segmentation. The nnU-Net is based on 2D and 3D U-Net models geared with several technical improvements (27). For instance, in terms of preprocessing and post-processing, the nnU-Net applies various methods such as denoising, enhancement, cropping, thresholding, and fusion to improve image quality and segmentation results, while also enhancing the visualization and interpretability of segmentation outcomes. For model optimization, the nnU-Net employs an optimizer with adaptive learning rate and momentum to

TABLE 1 Patient background information.

	Esophageal cancer negative		Esophageal cancer positive	
	Training (n = 250)	Validation (n = 100)	Training (n = 397)	Validation (n = 100)
Ages (years)	54.79 ± 12.42	58.475 ± 9.83	64.86 ± 9.58	61.91 ± 7.91
Female/Male	100/150	45/55	93/304	10/90
Main location	N/A	N/A		
Upper thoracic	N/A	N/A	34	12
Middle thoracic	N/A	N/A	209	63
Lower thoracic	N/A	N/A	154	25
Length of esophageal cancers (mm)	N/A	N/A	44.92 ± 22.26	25.41 ± 12.08
T stage	N/A	N/A		
T_1	N/A	N/A	18	5
T_2	N/A	N/A	63	15
T_3	N/A	N/A	185	53
T_4	N/A	N/A	26	26
T_x	N/A	N/A	105	1
Pathology	N/A	N/A		
SCC	N/A	N/A	349	95
Adenocarcinoma	N/A	N/A	37	2
Other	N/A	N/A	11	3

expedite the training process and enhance the performance of the model. In model training, the cross-validation scheme is implemented for the selection of the best-performing model. These technical improvements promise that the nnU-Net can yield more robust models.

In previous research, the nnU-Net has been widely used for the segmentation of the aorta (28), carotid artery (29), liver (30), and fetal brain (31), with promising performance in terms of accuracy, reliability, and efficiency. Accordingly, the nnU-Net is employed for the segmentation of the esophagus in the CT images with the evaluation metrics of Dice coefficient and Hausdorff Distance.

In the experiment, we trained a 3d U-Net model to segment the esophagus (see Figure 1). After preprocessing the training data, the networks automatically cropped the image patch with the sizes 80, 192, and 160 for training. The initial learning rate was 0.01, which continuously decreased with the increase in the number of iterations, and it no longer decreased when it reached 0.001. The networks were optimized with SGD and the training loss was dice loss.

Specifically, the nnU-Net demarcates the esophagus, and a post-processing of the appropriate thresholding for the removal of the air portions within the esophagus is applied to delineate the esophageal wall. Afterward, the average diameter and wall thickness of the esophagus can be calculated through distance transform, see Figure 1. In clinical definition, the esophagus is typically divided into upper, middle, and lower segments. In such cases, each segment may need different analytical methods and treatments. To mimic the clinical analytical paradigm, the center line is computed from the esophagus and further straightened to facilitate the automatic division of the upper, middle, and lower segments with intervals of 5 cm, 10 cm, and the remaining length from the starting point, respectively. For each segment, the measurement variances of the diameter and wall thickness of sampled transversal cut-planes are further computed. With these measurement variances, a decision tree is applied to determine if esophageal cancer is presented in the corresponding segment, see Figure 2.

2.4 The clinical application of the DL model

To assess the efficacy of the model in clinical application for the detection of esophageal cancer, three radiologists participated in this study. The participants reviewed the CT images in the validation dataset independently, which were presented in a randomized sequence, and made diagnoses either on their own or with the assistance of the model. The detailed reading protocol is elaborated as follows.

Two junior radiologists, Radiologist 1 and Radiologist 2, with 5 years of image diagnosis experience, and one senior radiologist, Radiologist 3, with 13 years of experience were invited to this study. All three radiologists were involved in the reading and diagnosis of the validation set tests. None of them had any knowledge of the study's purpose or any clinical information. Each radiologist independently reviewed the CT images of the validation dataset and made routine diagnostic practices. The diagnostic efficiency of each radiologist, including sensitivity, specificity, accuracy, F1 score, and AUC, was then calculated.

After a 3-month memory washout period, the three radiologists reevaluated the CT images of the validation dataset with the assistance of the DL model and made another round of diagnoses. The diagnostic workups of each radiologist, with the aid of the model, were further assessed with the same evaluation metrics. Finally, a quantitative comparison was performed to illustrate the diagnostic efficacy among the image diagnostic workups of radiologists, with and without the assistance of the DL model, as well as the pure prediction results from the DL model. The total flow diagram of the study is shown in Figure 3.

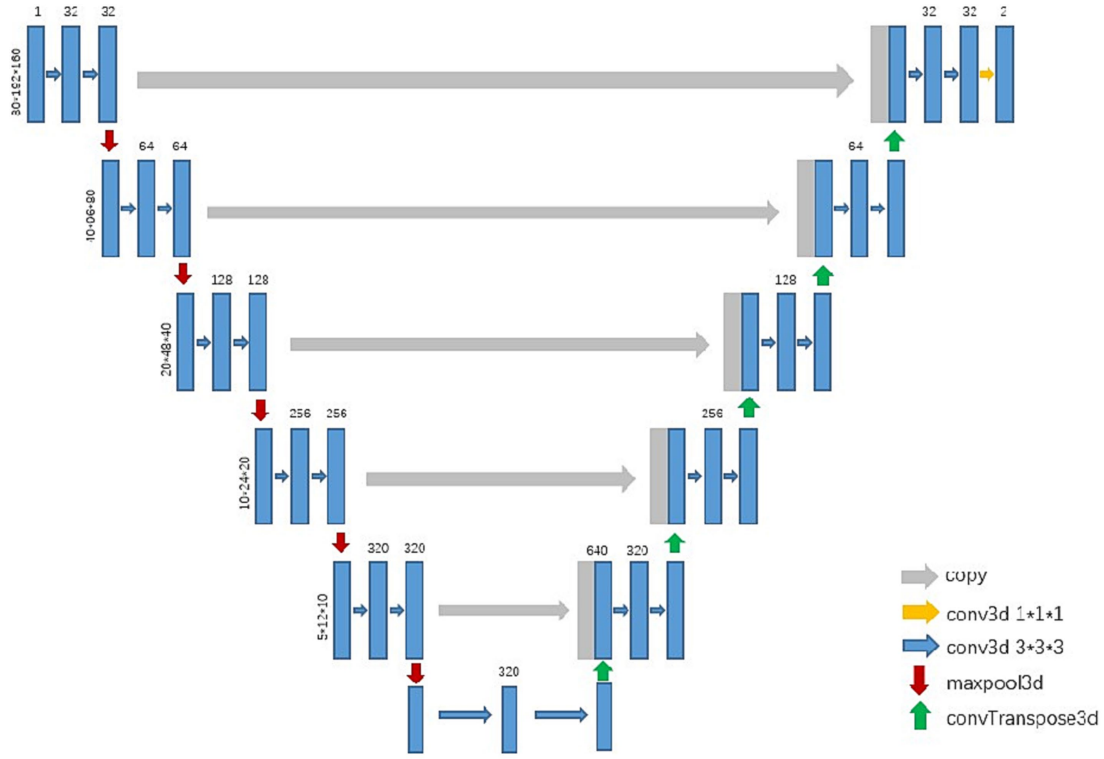


FIGURE 1
Flow diagram of the nnU-net.

2.5 Statistical analysis

In the classic evaluation paradigm for a classification model, four basic metrics of true positive (TP), true negative (TN), false negative (FN), and false positive (FP) may commonly need to be calculated for the computation of sensitivity and specificity. In this study, a TP suggests true cancer identification, whereas TN is the true non-cancerous classification. The FN indicates a missing cancer finding by either the model or the radiologist, while the FP represents a false cancer finding from the radiologist or DL model. In addition to sensitivity and specificity, the metrics of precision, false negative rate (FNR), false positive rate (FPR), and F1 score are computed to support extensive and quantitative performance comparison. The mentioned evaluation metrics are defined as follows.

$$\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN}), \quad (1)$$

$$\text{Specificity} = \text{TN} / (\text{TN} + \text{FP}), \quad (2)$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}), \quad (3)$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}), \quad (4)$$

$$\text{FNR} = 1 - \text{sensitivity}, \quad (5)$$

$$\text{FPR} = 1 - \text{specificity}, \quad (6)$$

$$\text{F1 score} = 2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision}). \quad (7)$$

Meanwhile, the area under the receiver operating characteristic (ROC) curve (AUC) was also employed as another quantitative metric (32). We used the intraclass correlation efficient (ICC) to compare the diagnosis consistency between the two junior radiologists. The ICC (95%CI) was 0.942 (0.924 and 0.955), which showed good diagnosis consistency. To further compare the performance of the DL model as well as the readers' performance with and without the referencing of the DL model, DeLong's test for AUC was adopted (33). The overall statistical analyses were carried out with software packages of SPSS 26.0 and MedCalc 22.016. Continuous variables were presented as mean ± standard deviation. Statistical significance was defined at a value of p of less than 0.05.

3 Results

In this study, CT scans of 397 primary esophageal cancer patients and 250 healthy individuals were involved in training the DL model, whereas independent images of 100 esophageal cancer patients and 100 healthy individuals were used for validation. Table 1 summarizes the background of all 497 primary esophageal cancer patients and 350 healthy individuals.

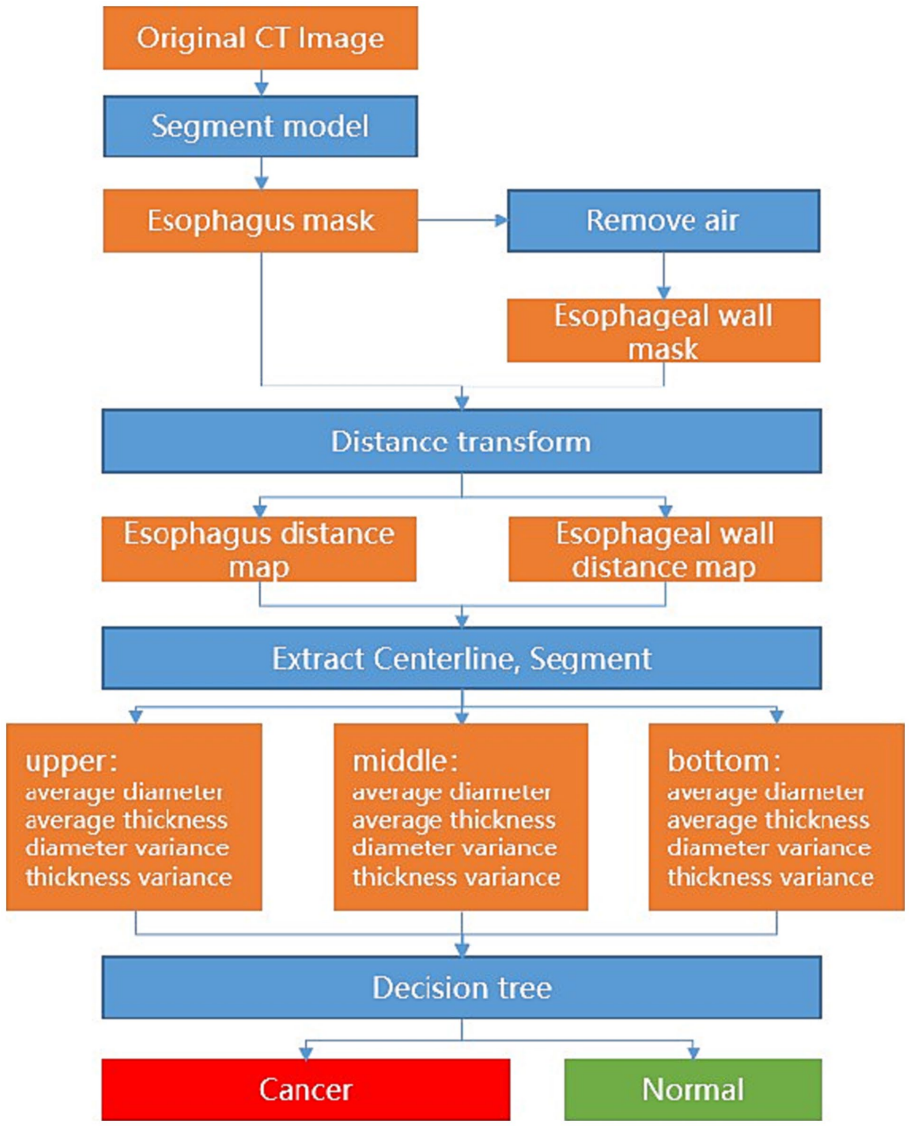


FIGURE 2
Flow diagram of the deep learning model.

3.1 The diagnostic efficiency of the DL model in the validation data set

The nnU-Net-based DL model was evaluated in a five-fold cross-validation scheme. The DICE for the esophagus segmentation in the validation data set was 0.875 ± 0.0728 and Hausdorff Distance was 1.765 ± 0.154. The performance of the DL model in the validation data set was summarized in Table 2. In the validation data set, the AUC of the model was 0.890, whereas the metrics of sensitivity, specificity, accuracy, and F1 score were 0.900, 0.880, 0.882, and 0.891, respectively.

Among the 10 CT examinations segmented by all three radiologists, the segmentations created by the different radiologists were shown to be similar. As shown in Table 3, Median interreader DSC ranged from 0.80 to 0.89 for all CT examinations. Median model-reader DSC ranged from 0.76 to 0.88 for all scans. The interreader DSC was not different than the model-reader DSC, indicating that the segmentation performance of the machine-learning algorithm did not differ significantly from that of the radiologists.

3.2 The diagnostic efficiency of radiologists with and without referring to the results of the DL model

The diagnostic efficiency of the radiologists in the validation data set is shown in Table 2. The AUC of Radiologist 1 independently in the validation set was 0.855, whereas the metrics of sensitivity Equation (1), specificity Equation (2), accuracy, and F1 score Equations (3, 4, 7) were 0.860, 0.850, 0.855, and 0.856, respectively. The AUC of Radiologist 2 independently in the validation set was 0.820, with the sensitivity, specificity, and F1 score of 0.780, 0.870, and 0.817, respectively. The AUC of Radiologist 3 independently in the validation set was 0.930, with the sensitivity, specificity, and F1 score of 0.950, 0.910, and 0.931, respectively. The diagnostic performance of the DL model is better than Radiologist 1 and Radiologist 2 independently with statistical significance in the AUC; however, it was lower than Radiologist 3 significantly. The other metrics of sensitivity, specificity, and F1 score were also attained higher by the DL model

than Radiologist 1 and Radiologist 2, but lower than Radiologist 3. With the help of the model, Radiologist 1 and Radiologist 2 showed significant improvement in the AUC, as well as the other metrics. Meanwhile, the performance of Radiologist 3 also improved with the DL model when compared to the performance in the independent reading session. Figure 4 visually compares the ROC curves of the DL model and the radiologists.

3.3 Comparison of the rates of misdiagnosis and missed diagnosis between DL model and radiologists

In the validation set, the DL model missed 10% of esophageal cancer cases [FNR = 0.100, Equation (5)], which was lower than the average FNR of 13.7% for all radiologists in the independent reading session (without the DL model). With the incorporation of DL modeling in the reading session, the average FNR by all radiologists was lowered to 5%. In such cases, the DL model can improve radiologists' workups in finding esophageal cancers. On the other hand, the DL model yielded 12% false positives in the validation set, which was similar to the average FPR Equation (6) of 12.3% by all radiologists in independent reading sessions. With the aid of the DL model, the average FPR by all radiologists was reduced to 6%, see Table 2. Accordingly, the DL model can on average improve radiologists' performance and reduce the FP and FN rates in half.

Further analysis was conducted for the FPs yielded by the DL model. The majority of FPs were acute and chronic esophagitis (75%, nine cases), and a small proportion were esophageal papillomas, esophageal hyperplastic polyps, and gastric mucosal ectopies (25%, one case for each abnormality). For the FN cases by the DL model, most of them were early-stage cancers, involving seven cases (70%) of esophageal cancer at T1-2 and three cases (30%) of T3-4 esophageal cancer. The DL model missing the T3-4 cancers may be because the nearby soft tissues around the cancers are complicated which further confused the model to an incorrect differentiation. Additionally, a challenging case involving a 77-year-old man diagnosed with T1 stage esophageal cancer was missed by the radiologists but successfully detected by the DL model (Figure 5), which revealed the excellent performance of the DL model. There were still some cases that were too early and did not have detectable changes in the images to be detected, see Figure 6.

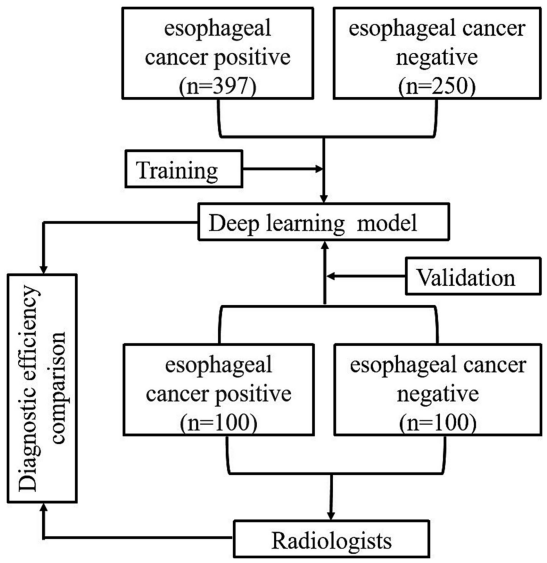


FIGURE 3
Experimental flow chart of the study.

4 Discussion

In this retrospective dual-center study, a DL-based method was developed to detect esophageal cancer to assist the clinical reading. The model was trained with non-contrast chest CT scans acquired from 397 esophageal cancer-positive patients and 250 individuals with no esophageal cancer. In the validation, the DL-based method showed a satisfactory diagnostic efficacy in detecting esophageal cancer with an AUC of 0.890 and an accuracy of 0.882, which were higher than the two junior radiologists, i.e., Radiologist 1 and Radiologist 2, but lower than the senior radiologist (Radiologist 3). Referring to the previous study, the underlying reasons the DL model outperformed the junior radiologists may be two-fold (34). First, the DL model was trained by the esophageal cancer cases which were validated by pathology. The junior radiologists did not get systematic and sufficient training in the reading and diagnosis of esophageal cancer in *non-contrast chest* CT

TABLE 2 Diagnostic efficiency comparison between deep learning model and radiologists.

	Sensitivity	Specificity	Accuracy	F1 score	AUC	FNR	FPR	P^a	P^b
Deep learning model	0.900	0.880	0.882	0.891	0.890	0.100	0.120		
Radiologists 1 independently	0.860	0.850	0.855	0.856	0.855	0.140	0.150	0.0040	
Radiologists 2 independently	0.780	0.870	0.857	0.817	0.820	0.220	0.130	0.0001	
Radiologists 3 independently	0.950	0.910	0.913	0.931	0.930	0.050	0.090	0.0040	
Radiologists average 1	0.860	0.877	0.873	0.866	0.867	0.137	0.123		
Radiologists 1 with model	0.920	0.900	0.902	0.911	0.910	0.080	0.100	0.0444	0.0004
Radiologists 2 with model	0.960	0.950	0.950	0.955	0.955	0.040	0.050	0.0002	<0.0001
Radiologists 3 with model	0.960	0.970	0.970	0.965	0.965	0.040	0.030	0.0001	0.0068
Radiologists average 2	0.947	0.940	0.941	0.944	0.943	0.053	0.060		

P^a: Comparisons of AUC value between the deep learning model and each radiologist with or without the deep learning model.
P^b: Comparisons of AUC value between the radiologist with and without the deep learning model.

TABLE 3 Median interreader and radiologists-model DSCs for 10 cases in the test set.

Reader no.	Radiologists 1	Radiologists 2	Radiologists 3	Model
Radiologists 1	1	0.80 (0.65–0.96)	0.89 (0.79–0.98)	0.87 (0.73–0.99)
Radiologists 2		1	0.81 (0.68–0.88)	0.76 (0.64–0.83)
Radiologists 3			1	0.87 (0.76–0.98)

Data are Dice similarity coefficients (DSCs), with minimum and maximum values in parentheses.

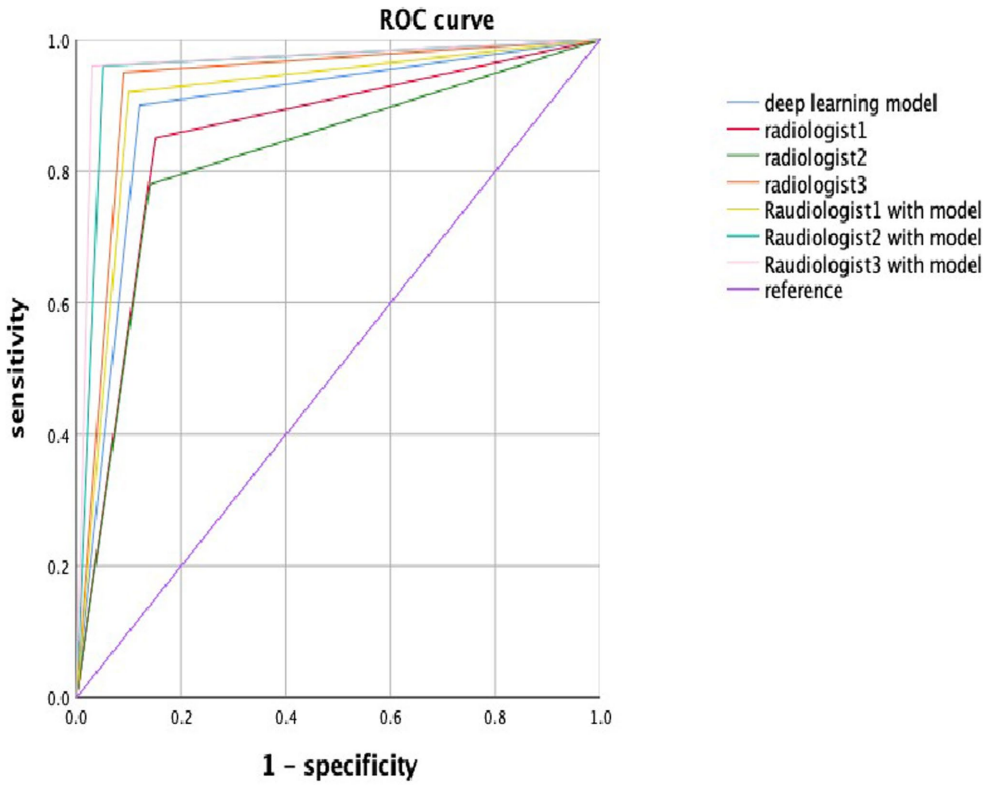


FIGURE 4
The ROC curve in the deep learning model and radiologists with or without the deep learning model. Blue, red, green, orange, lemon-yellow, blue-green, and pink lines indicate the ROC curve of the deep learning model, radiologist1, radiologist2, radiologist3, radiologist1 with the model, radiologist2 with the model, and radiologist3 with the model.

images. Second, DL algorithms have a higher sensitivity to subtle image changes than human eyes, and hence yield better detection results for the easy-missing lesions like esophageal cancers (32). With the help of the DL model, the junior and senior radiologists achieved better diagnostic workups in detecting esophageal cancers. Accordingly, the computerized DL system may be potentially valuable in the context of health checkup non-contrast CT examination for the early detection of esophageal cancers.

In this study, our model reached the performance of sensitivity, specificity, accuracy, F1 score, and AUC values of 0.900, 0.880, 0.882, 0.891, and 0.890, respectively, which was better than the previous study with V-net, where the sensitivity, specificity, and AUC were 0.690, 0.610, and 0.650, respectively (35). It may be because our method is equipped with a more robust segmentation model and better cancer identification post-processing scheme for better results. Compared to the method with the pure image classification model of VGG16 on the *contrast-enhanced CT*, the reported performance of sensitivity, specificity, accuracy, and F1 score were 0.717, 0.90, 0.842, and 0.742, respectively (36). Accordingly, a segmentation model may be helpful to improve the detection performance with slightly lower specificity. On the other hand, another image classification CNN for the *contrast-enhanced chest images* suggested a performance with metrics of sensitivity, specificity, accuracy, and AUC as 0.87, 0.92, 0.92, and 0.95 (33), respectively, since the contrast-enhanced CT may better depict the esophageal cancers and may ease the algorithmic difficulty for DL models. However, our experimental results suggested that the DL can also assist radiologists in improving the workups of esophageal cancers by reducing FPs and FNs in non-contrast chest CT scans. In particular, the DL model may improve the performance of junior radiologists to the senior level, which resonates with the conclusion of the studies (33, 35). Accordingly, this may shed light on the early detection of esophageal cancers, especially in the context of health check-up examinations.

There are several limitations in our study. First, the distribution of sex and age were uneven in the training and validation data, but the

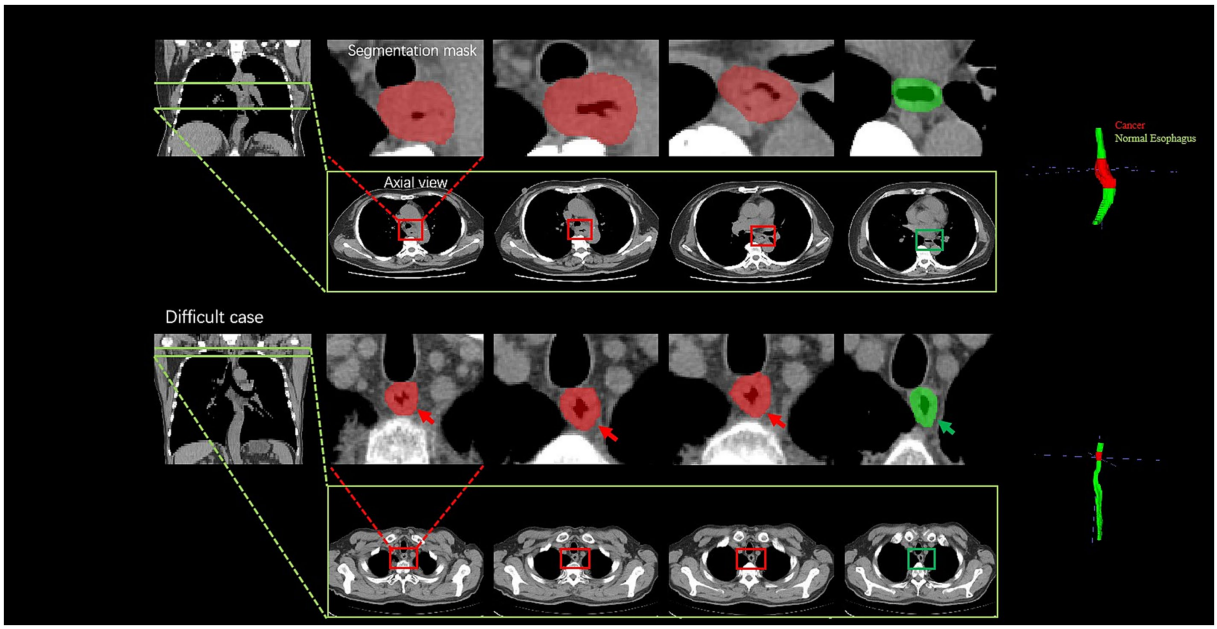


FIGURE 5
Visualization of two cancer cases. For the easy case, there is a significant thickening of the diameter and thickening of the esophageal wall; the difficult case is a 77-year-old man diagnosed with T1 stage esophageal cancer; the radiologists failed to accurately diagnose the cancer, whereas the deep learning model successfully detected it by the subtle variation of the esophageal wall thickness. The cancer part is indicated by the red color, and the green color part presents a normal esophagus.

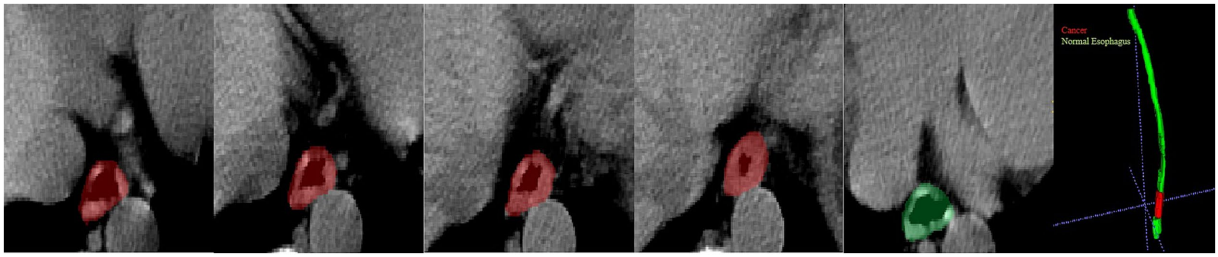


FIGURE 6
Visualization of a missed diagnosed case by DL. A 62-year-old man diagnosed with T1 stage esophageal cancer under an endoscope; the pathology showed the cancer was confined to the lamina propria of the mucosa and very close to the cardia. The cancer part is indicated by the red color, and the green color part presents a normal esophagus.

esophageal cancers mainly occurred in men and ages >50 years (9); therefore, the enrolled individuals were suitable for model training. Second, we enrolled some early-stage esophageal cancer in this study. However, the DL model and the radiologists failed to identify all these cases. The detection of early-stage esophageal cancer can be very challenging for both the radiologist and the DL model (such as Figure 6), but it is important for clinical practice. Referencing the other studies (33, 36), the contrast-enhanced CT images may provide more information about esophageal cancer from early to late stage than the non-contrast images. Accordingly, we will consider incorporating contrast-enhanced CT to augment the capability of the DL model. Third, for some patients with neoadjuvant chemotherapy before the surgical operation, we obtained the pathology from endoscopic biopsy and did not get the true cancer stage. Fourth, this study involved a medium number of patients. A further expansion of the cohort is needed.

5 Summary statement

The DL model can detect esophageal cancer from non-contrast chest images with good sensitivity and specificity. With the help of the DL model, the radiologist can improve the diagnostic efficacy in detecting esophageal csancer, shorten the training time for junior radiologists, and reduce the missed diagnosis of esophageal cancer in routine physical examinations of individuals with only non-contrast chest CT images.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding authors.

Ethics statement

The studies involving humans were approved by the Ethics Committee of Fudan University Zhongshan Hospital (Xiamen). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

CL: Data curation, Formal analysis, Investigation, Software, Writing – original draft. YG: Data curation, Formal analysis, Investigation, Methodology, Software, Writing – original draft, Validation. XH: Data curation, Writing – original draft. SR: Supervision, Writing – original draft, Writing – review & editing. JZ: Supervision, Writing – original draft, Writing – review & editings.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the Science and Technology Guided Project of Fujian Province (grant number: 2019D025) and the Scientific Research Cultivation and Medical Innovation Project of Fujian Province (grant number: 2019CXB33).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2024.1356752/full#supplementary-material

References

1. Uhlenhopp DJ, Then EO, Sunkara T, Gaduputi V. Epidemiology of esophageal. cancer: update in global trends, etiology and risk factors. *Clin J Gastroenterol*. (2020) 13:1010–21. doi: 10.1007/s12328-020-01237-x

2. Liu CQ, Ma YL, Qin Q, Wang PH, Luo Y, Xu PF, et al. Epidemiology of. Esophageal cancer in 2020 and projections to 2030 and 2040. Thorac. *Cancer*. (2023) 14:3–11. doi: 10.1111/1759-7714.14745

3. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. (2021) 71:209–49. doi: 10.3322/caac.21660

4. Huang FL, Yu SJ. Esophageal cancer: risk factors, genetic association, and treatment. *Asian J Surg*. (2018) 41:210–5. doi: 10.1016/j.asjsur.2016.10.005

5. Kamangar F, Nasrollahzadeh D, Safiri S, Sepanlou SG, Fitzmaurice C, Ikuta KS, et al. The global, regional, and national burden of oesophageal cancer and its attributable risk factors in 195 countries and territories, 1990-2017: a systematic analysis for the global burden of disease study 2017. *Lancet Gastroenterol Hepatol*. (2020) 5:582–97. doi: 10.1016/S2468-1253(20)30007-8

6. Wei MT, Friedland S. Early esophageal Cancer what the gastroenterologist needs to know. *Gastroenterol Clin N Am*. (2021) 50:791–08. doi: 10.1016/j.gtc.2021.07.004

7. Smyth EC, Lagergren J, Fitzgerald RC, Lordick F, Shah MA, Lagergren P, et al. Oesophageal cancer. *Nat Rev Dis Primers*. (2017) 3:3. doi: 10.1038/nrdp.2017.48

8. Ba-Ssalamah A, Zacherl J, Noebauer-Huhmann IM, Uffmann M, Matzek WK, Pinker K, et al. Dedicated multi-detector CT of the esophagus: spectrum of diseases. *Abdom Imaging*. (2009) 34:3–18. doi: 10.1007/s00261-007-9290-5

9. Obermannova R, Alsina M, Cervantes A, Leong T, Lordick F, Nilsson M, et al. Oesophageal cancer: ESMO clinical practice guideline for diagnosis, treatment and follow-up. *Ann Oncol*. (2022) 33:992–04. doi: 10.1016/j.annonc.2022.07.003

10. Arndt C, Guttler F, Heinrich A, Burckenmeyer F, Diamantis I, Teichgraber U. Deep Learning CT image reconstruction in clinical practice. *Rofo*. (2021) 193:252–61. doi: 10.1055/a-1248-2556

11. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. (2015) 521:436–44. doi: 10.1038/nature14539

12. McBee MP, Awan OA, Colucci AT, Ghobadi CW, Kadom N, Kansagra AP, et al. Deep Learning in radiology. *Acad Radiol*. (2018) 25:1472–80. doi: 10.1016/j.acra.2018.02.018

13. Wang QF, Zhou XH, Wang C, Liu ZQ, Huang J, Zhou Y, et al. WGAN-based synthetic minority over-sampling technique: improving semantic fine-grained classification for lung nodules in CT images. *IEEE Access*. (2019) 7:18450–63. doi: 10.1109/ACCESS.2019.2896409

14. Ouyang X, Karanam S, Wu ZY, Chen TC, Huo JY, Zhou XS, et al. Learning hierarchical attention for weakly-supervised chest X-ray abnormality localization and diagnosis. *IEEE Trans Med Imaging*. (2021) 40:2698–10. doi: 10.1109/TMI.2020.3042773

15. Li ZR, Cui ZM, Wang S, Qi YJ, Ouyang X, Chen QT, et al. Domain generalization for mammography detection via multi-style and multi-view contrastive Learning. Medical image computing and computer assisted intervention - Miccai 2021. *Electr Eng Sys Sci*. (2021) 12907:98–108. doi: 10.1007/978-3-030-87234-2_10

16. Cheng JZ, Ni D, Chou YH, Qin J, Tiu CM, Chang YC, et al. Computer-aided. Diagnosis with Deep Learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans. *Sci Rep*. (2016) 6:24454. doi: 10.1038/srep24454

17. Sihong C, Jing Q, Xing J, Baiying L, Tianfu W, Dong N, et al. Automatic scoring of multiple semantic attributes with multi-task feature leverage: a study on pulmonary nodules in CT images. *IEEE Trans Med Imaging*. (2017) 36:802–14. doi: 10.1109/TMI.2016.2629462

18. Lei BY, Huang S, Li H, Li R, Bian C, Chou YH, et al. Self-co-attention neural. Network for anatomy segmentation in whole breast ultrasound. *Med Image Anal*. (2020) 64:101753. doi: 10.1016/j.media.2020.101753

19. Wang C, Elazab A, Wu J, Hu Q. Lung nodule classification using deep feature. Fusion in chest radiography. *Comput Med Imaging Graph*. (2017) 57:10–8. doi: 10.1016/j.compmedimag.2016.11.004

20. Saihood A, Karshenas H, Nilchi ARN. Deep fusion of gray level co-occurrence. Matrices for lung nodule classification. *PLoS One*. (2022) 17:e0274516. doi: 10.1371/journal.pone.0274516

21. Chae EY, Kim HH, Jeong JW, Chae SH, Lee S, Choi YW. Decrease in. Interpretation time for both novice and experienced readers using a concurrent computer-aided detection system for digital breast tomosynthesis. *Eur Radiol*. (2019) 29:2518–25. doi: 10.1007/s00330-018-5886-0

22. Ciompi F, Chung K, van Riel SJ, Setio AAA, Gerke PK, Jacobs C, et al. Towards. Automatic pulmonary nodule management in lung cancer screening with deep learning. *Sci Rep*. (2017) 7:46479. doi: 10.1038/srep46479

23. Thirumalaisamy S, Thangavilou K, Rajadurai H, Saidani O, Alturki N, Mathivanan SK, et al. Breast Cancer classification using synthesized Deep Learning model with metaheuristic optimization algorithm. *Diagnostics (Basel)*. (2023) 13:2925. doi: 10.3390/diagnostics13182925

24. Wong PK, Chan IN, Yan HM, Gao S, Wong CH, Yan T, et al. Deep learning. Based radiomics for gastrointestinal cancer diagnosis and treatment: a minireview. *World J Gastroenterol*. (2022) 28:6363–79. doi: 10.3748/wjg.v28.i45.6363

25. Akkus Z, Galimzianova A, Hoogi A, Rubin DL, Erickson BJ. Deep Learning for. Brain MRI segmentation: state of the art and future directions. *J Digit Imaging*. (2017) 30:449–59. doi: 10.1007/s10278-017-9983-4

26. Yamanakkanavar N, Choi JY, Lee B. MRI segmentation and classification of. Human brain using Deep Learning for diagnosis of Alzheimer's disease: a survey. *Sensors (Basel)*. (2020) 20:3243. doi: 10.3390/s20113243

27. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. (2021) 18:203–11. doi: 10.1038/s41592-020-01008-z

28. Li F, Sun L, Lam KY, Zhang S, Sun Z, Peng B, et al. Segmentation of human. Aorta using 3D nnU-net-oriented deep learning. *Rev Sci Instrum*. (2022) 93:114103. doi: 10.1063/5.0084433

29. Zhu Y, Chen L, Lu W, Gong Y, Wang X. The application of the nnU-Net-based. Automatic segmentation model in assisting carotid artery stenosis and carotid atherosclerotic plaque evaluation. *Front Physiol*. (2022) 13:1057800. doi: 10.3389/fphys.2022.1057800

30. Pettit RW, Marlatt BB, Corr SJ, Havelka J, Rana A, nnU-Net Deep Learning . Method for segmenting parenchyma and determining liver volume from computed tomography images. *Ann Surg Open*. (2022) 3:e155. doi: 10.1097/AS9.0000000000000155

31. Peng Y, Xu Y, Wang M, Zhang H, Xie J. The nnU-Net based method for. Automatic segmenting fetal brain tissues. *Health Inf Sci Syst*. (2023) 11:17. doi: 10.1007/s13755-023-00220-3

32. Balaur E, O'Toole S, Spurling AJ, Mann GB, Yeo B, Harvey K, et al. Colorimetric. Histology using plasmonically active microscope slides. *Nature*. (2021) 598:65–71. doi: 10.1038/s41586-021-03835-2

33. Yasaka K, Hatano S, Mizuki M, Okimoto N, Kubo T, Shibata E, et al. Effects of. Deep learning on radiologists' and radiology residents' performance in identifying esophageal cancer on CT. *Br J Radiol*. (2023) 96:685. doi: 10.1259/bjr.20220685

34. Cao K, Xia Y, Yao J, Han X, Lambert L, Zhang T, et al. Large-scale pancreatic. cancer detection via non-contrast CT and deep learning. *Nat Med*. (2023) 29:3033–43. doi: 10.1038/s41591-023-02640-w

35. Sui H, Ma R, Liu L, Gao Y, Zhang W, Mo Z. Detection of incidental esophageal. Cancers on chest CT by Deep Learning. *Front Oncol*. (2021) 11:700210. doi: 10.3389/fonc.2021.700210

36. Takeuchi M, Seto T, Hashimoto M, Ichihara N, Morimoto Y, Kawakubo H, et al. Performance of a deep learning-based identification system for esophageal cancer from CT images. *Esophagus*. (2021) 18:612–20. doi: 10.1007/s10388-021-00826-0

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 11 March 2024
DOI 10.3389/fmed.2024.1357981

Check for updates

OPEN ACCESS

EDITED BY
Minjeong Kim,
University of North Carolina at Greensboro, United States

REVIEWED BY
Lisheng Xu,
Northeastern University, China
Alexandru Florea,
University Hospital RWTH Aachen, Germany

*CORRESPONDENCE
Xigang Xiao
✉ xxgct_417@126.com

RECEIVED 19 December 2023
ACCEPTED 23 February 2024
PUBLISHED 11 March 2024

CITATION
Jia Y, Zou L, Xue M, Zhang X and Xiao X (2024) Evaluation of peri-plaque pericoronary adipose tissue attenuation in coronary atherosclerosis using a dual-layer spectral detector CT.
Front. Med. 11:1357981.
doi: 10.3389/fmed.2024.1357981

COPYRIGHT
© 2024 Jia, Zou, Xue, Zhang and Xiao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Evaluation of peri-plaque pericoronary adipose tissue attenuation in coronary atherosclerosis using a dual-layer spectral detector CT

Yulin Jia[1], Lei Zou[1,2], Ming Xue[1], Xiaoyu Zhang[1] and Xigang Xiao[1]*

[1]Department of Radiology, The First Affiliated Hospital of Harbin Medical University, Harbin, China, [2]Department of Radiology, Zigong Fourth People's Hospital, Zigong, China

Purpose: This study aimed to evaluate the differences between pericoronary adipose tissue (PCAT) attenuation at different measured locations in evaluating coronary atherosclerosis using spectral computed tomography (CT) and to explore valuable imaging indicators.

Methods: A total of 330 patients with suspicious coronary atherosclerosis were enrolled and underwent coronary CT angiography with dual-layer spectral detector CT (SDCT). Proximal and peri-plaque fat attenuation index (FAI) of stenosis coronary arteries were measured using both conventional images (CIs) and virtual monoenergetic images (VMIs) ranging from 40 keV to 100 keV. The slopes of the spectral attenuation curve (λ) of proximal and peri-plaque PCAT at three different monoenergetic intervals were calculated. Additionally, peri-plaque FAI on CI and virtual non-contrast images, and effective atomic number were measured manually.

Results: A total of 231 coronary arteries with plaques and lumen stenosis were finally enrolled. Peri-plaque FAI_{CI} and FAI_{VMI} were significantly higher in severe stenosis than in mild and moderate stenosis ($p < 0.05$), while peri-plaque λ, proximal FAI, and proximal λ were not statistically different. Proximal FAI, peri-plaque FAI, and peri-plaque λ were significantly higher in low-density non-calcified plaque (LD-NCP) and non-calcified plaque (NCP) than in calcified plaque ($p < 0.01$). Peri-plaque FAI was the highest in the LD-NCP group, while proximal FAI was the highest in the NCP group. In severe stenosis and in LD-NCP, peri-plaque FAI was significantly higher than proximal FAI ($p < 0.05$). The manually measured parameters related to peri-plaque PCAT attenuation had a positive correlation with the results of peri-plaque FAI measured automatically.

Conclusion: Peri-plaque PCAT has more value in assessing coronary atherosclerosis than proximal PCAT. Peri-plaque PCAT attenuation is expected to be used as a standard biomarker for evaluating plaque vulnerability and hemodynamic characteristics.

KEYWORDS

pericoronary adipose tissue, fat attenuation index, coronary inflammation, longitudinal location, dual-layer spectral detector CT

Introduction

According to the 2021 World Health Organization report, cardiovascular disease (CVD) is the world's leading cause of death, with 85% of deaths resulting from heart attacks and strokes (1). As the most common CVD, coronary atherosclerotic heart disease (CAD) is receiving increasing attention, especially in the aspect of accurate diagnosis and treatment. Traditionally, the assessment of CAD using coronary computed tomography angiography (CCTA) focuses mainly on the morphological evaluation of plaque types and lumen stenosis. Recent studies have shown that CAD is a chronic inflammatory disease that can release mediators into pericoronary adipose tissue (PCAT), causing morphological changes and increased attenuation of PCAT (2, 3). PCAT is visceral adipose tissue, which contacts the coronary artery directly. PCAT was defined as an epicardial adipose tissue (EAT) with a radial distance of 1 mm from the outer wall of the coronary artery, and all voxel values were within the range of −190 to −30 HU (4). PCAT has a complex bidirectional paracrine pathway with the coronary artery wall. The dysfunctional PCAT, which reflects the inflammatory status of the PCAT and the coronary artery, can also be detected by CCTA according to the increased PCAT attenuation (4, 5). Thus, some studies measured the perivascular fat attenuation index (FAI) of coronary arteries to assess the plaque risk or severity of CAD (6–9). Most of them adopted FAI around the proximal right coronary artery (RCA) due to the abundance of adipose tissue, which can be obtained and measured easily compared to the left anterior descending (LAD) artery and left circumflex (LCX) artery. Other studies used FAI of proximal segments of three major coronary arteries or around the plaque that caused the maximal degree of vascular stenosis as the evaluation indicator (10–12). At present, there is no unified standard in terms of measurement locations of PCAT attenuation for the evaluation of CAD-related ischemic severity or risk. Therefore, the aim of this study is to evaluate the diagnostic value of PCAT attenuation at different positions based on spectral CT for CAD.

Materials and methods

Patients

A total of 330 consecutive patients with suspected CAD underwent CCTA using dual-layer spectral detector CT (SDCT) at the First Affiliated Hospital of Harbin Medical University from April to November 2021. A total of 170 patients were excluded according to the following exclusion criteria: no coronary atherosclerotic plaque or luminal stenosis, or only the left main coronary artery or coronary branches of three main coronary arteries with atherosclerotic plaques; a history of percutaneous coronary intervention (PCI) or coronary artery bypass graft (CABG); malignant tumor or severe inflammatory disease; poor CCTA image quality; and incomplete clinical data. The remaining 160 patients had 480 main coronary arteries, including the LAD artery, LCX artery, and RCA. Among them, only 231 branches had atherosclerotic plaques and luminal stenosis. Therefore, the other 249 branches without plaques or stenosis were excluded. All CCTA images were independently evaluated for plaque types and luminal stenosis according to previous studies (13, 14) by two radiologists who had more than 10 years of experience in cardiac imaging diagnosis. The consensus was reached when there were different opinions. Finally, there were 123 coronary arteries with mild stenosis, 69 arteries with moderate stenosis, and 39 arteries with severe stenosis. In terms of plaque types, there were 66 coronary arteries with calcified plaques (CPs), 92 arteries with non-calcified plaque (NCP), and 73 arteries with low-density non-calcified plaque (LD-NCP). The study flow chart is shown in Figure 1. This retrospective study was approved by the ethics committee of the First Affiliated Hospital of Harbin Medical University (No. 202214).

CCTA scan protocol and image reconstruction

The retrospective ECG-gated CCTA scans were performed on SDCT (IQon Spectral CT, Philips Healthcare, Best, The Netherlands). A bolus of 60–80 mL non-ionic contrast media (iobitridol; 350 mg iodine/ml, Guerbet, Roissy, France) was injected into the median cubital vein at a flow rate of 4–5 mL/s via a high-pressure injector with binocular tube (Ulrich, Germany), followed by 30 mL of saline flush. When the descending aorta reached the 110HU threshold, the CCTA scan was triggered automatically. The scan protocol was as follows: 120kVp tube voltage, automatic tube current modulation, 0.9 mm slice thickness and slice interval, 64 × 0.625 mm detector collimation, 270 ms tube rotation time, and 512 × 512 matrix. The original data were reconstructed, including both conventional images (CIs) and virtual monoenergetic images (VMIs) from 40 keV to 100 keV with a 10 keV interval. All data were transferred to the post-processing workstation (IntelliSpace Portal Vision 10, Philips Healthcare, Best, The Netherlands).

PCAT attenuation parameter measurement

Proximal and peri-plaque FAI of stenosis coronary arteries were measured both on CI and VMI by semi-automatic software (Shukun Technology, Beijing, China), which can automatically segment coronary arteries and identify PCAT. The measurement starting point, length, and width of proximal PCAT were set manually, completely consistent with the method used in our previous research (10). The starting point of the RCA measurement was set at 10 mm from the right coronary sinus ostium. LAD and LCX were measured from the bifurcation of the left main coronary artery. The measurement length was proximal 40 mm, and the measurement width was the mean diameter of the proximal coronary. Peri-plaque FAI was measured surrounding the plaque at the obvious stenosis. The measurement length was set according to the length of the plaque, and the measurement width was set according to the mean diameter of normal lumens at the proximal and distal ends of the plaque. Adipose tissue voxels on CI, 40 keV, 50 keV, 60 keV, 70 keV, 80 keV, 90 keV, and 100 keV images were identified with different thresholds, respectively: -190 ~ -30HU, −280 ~ -40HU, −260 ~ -40HU, −240 ~ -40HU, −220 ~ -30HU, −210 ~ -30HU, −200 ~ -30HU, and −190 ~ -30HU. The slope of the spectral attenuation curve (λ) was computed: $\lambda_{40\text{-}70\ KeV} = (CT_{40keV}\text{-}CT_{70keV})/30$, $\lambda_{40\text{-}100\ keV} = (CT_{40keV}\text{-}CT_{100keV})/60$, and $\lambda_{70\text{-}100\ KeV} = (CT_{70keV}\text{-}CT_{100keV})/30$. The automatic PCAT attenuation parameter measurement is shown in Figure 2. In

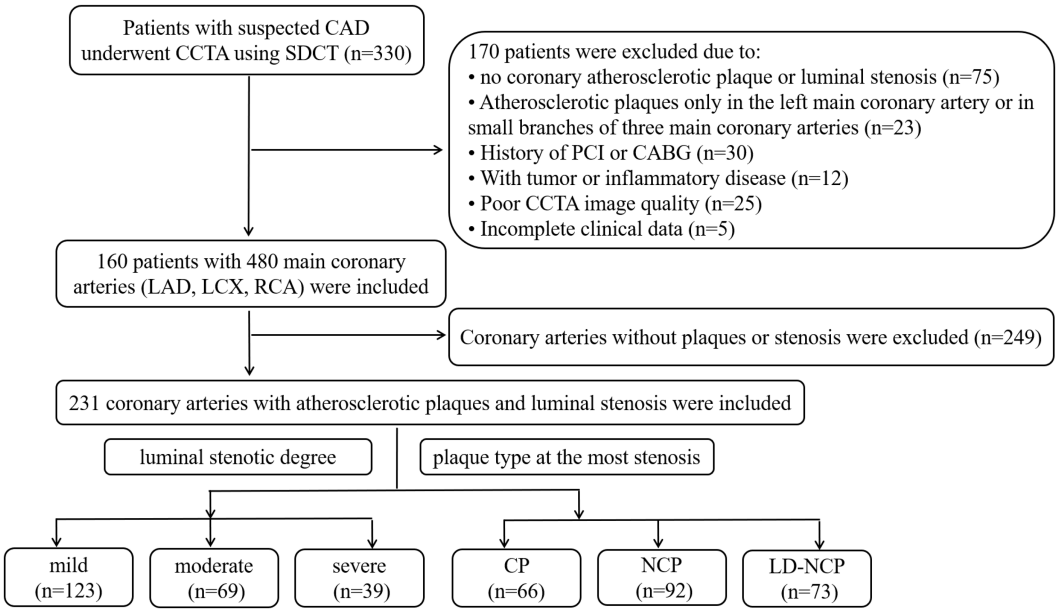


FIGURE 1
Study flowchart. CAD, coronary atherosclerotic heart disease; CCTA, coronary CT angiography; SDCT, dual-layer spectral detector computed tomography; PCI, percutaneous coronary intervention; CABG, coronary artery bypass graft; LAD, left anterior descending artery; LCX, left circumflex artery; RCA, right coronary artery; CP, calcified plaques; NCP, non-calcified plaque; and LD-NCP, low-density non-calcified plaque.

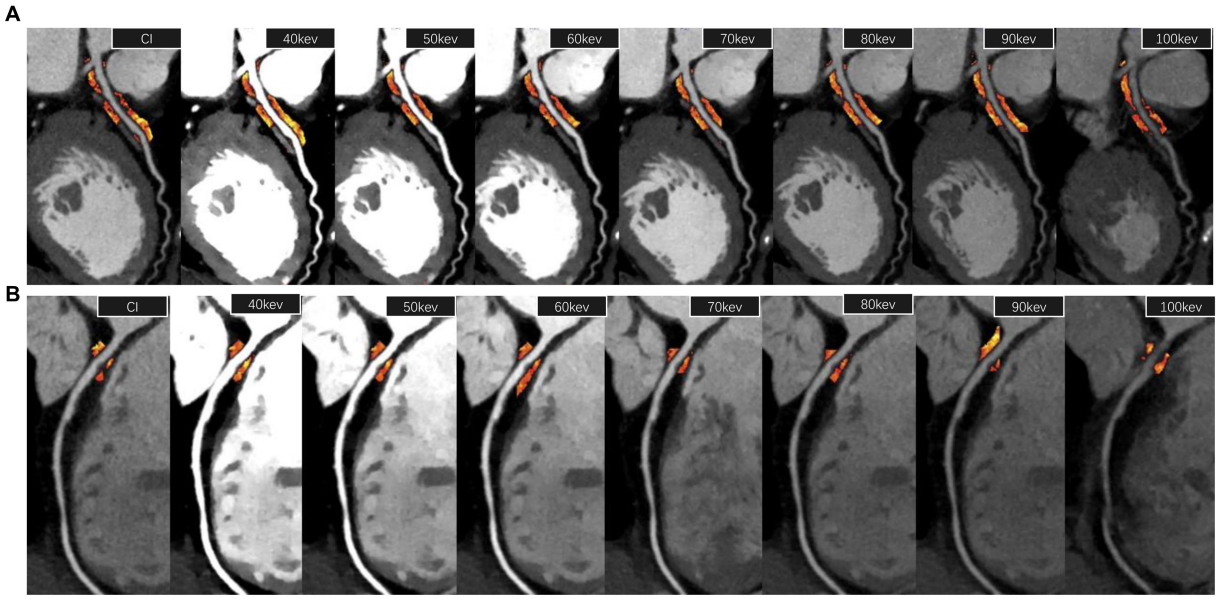


FIGURE 2
PCAT attenuation parameter measurement. **(A)** The measurement of proximal FAI of the left anterior descending artery (40 mm length from the bifurcation of the left main coronary artery) on a 120 kVp conventional image and 40–100 keV virtual monoenergetic images, respectively. **(B)** The measurement of peri-plaque FAI around a low-density non-calcified plaque at the most stenosis of the lumen of the right coronary artery. FAI, fat attenuation index.

addition, peri-plaque FAI on CI, virtual non-contrast (VNC) image, and effective atomic number (Zeff) were also measured manually. The mean value of each parameter was obtained from three consecutive regions of interests (ROIs) with consistent size in the PCAT at the highest luminal stenosis.

Statistical analysis

Continuous variables were expressed as the average ± standard deviation (SD). Normal distribution and homogeneity of variance tests on the data were performed. If the data were a normal distribution

with homogeneity of variance, analysis of variance (ANOVA) was used for comparison among groups. Otherwise, a multi-variant ANOVA was performed. SPSS 26.0 (SPSS, Inc., Chicago, IL, USA) was used for statistical analysis. A p-value of <0.05 was considered statistically significant.

Results

FAI gradually increased with the increase in monoenergetic level on VMI, whether it was proximal or peri-plaque FAI, whether it was mild, moderate, or severe lumen stenosis, whether it was CP, NCP, or LD-NCP (Supplementary Figure S1).

Among the different degrees of stenosis, both proximal and peri-plaque FAI increased with the stenosis degree of the coronary artery. Peri-plaque FAI_{CI} and FAI_{VMI} were significantly higher in severe stenosis than in mild and moderate stenosis ($p < 0.05$), while λ of proximal and peri-plaque PCAT attenuation was neither statistically different ($p > 0.05$) nor was the proximal FAI of vessels with different degrees of stenosis ($p > 0.05$). In severe stenosis, peri-plaque FAI was significantly higher than proximal FAI ($p < 0.05$), while they were not significantly different in mild and moderate stenosis ($p > 0.05$) (Figure 3) (Supplementary Tables S1, S2).

Among different plaque types, peri-plaque FAI_{CI}, FAI_{VMI}, and λ were significantly higher in LD-NCP and NCP than in CP ($p < 0.01$). The values were the highest in the LD-NCP group. Proximal FAI_{CI} and FAI_{VMI} were also significantly different in different plaque types ($p < 0.05$), but the highest FAI values were in the NCP group. λ of proximal PCAT attenuation was not statistically different among different plaque types ($p > 0.05$). In LD-NCP, peri-plaque FAI was significantly higher than the proximal FAI ($p < 0.05$), while they were not significantly different in CP and NCP ($p > 0.05$) (Figure 4) (Supplementary Tables S3, S4).

The manually measured peri-plaque FAI on CI (FAI_{MA}) was significantly positively correlated with the peri-plaque FAI_{CI} obtained automatically ($r = 0.7265$, $p < 0.0001$). Similar to the differences in automatically measured peri-plaque PCAT attenuation parameters between groups, the three manually measured peri-plaque PCAT spectral parameters also showed differences. The values of FAI_{MA}, FAI_{VNC}, and Zeff were as follows: severe stenosis > moderate stenosis > mild stenosis ($p < 0.01$) and LD-NCP > NCP > CP ($p < 0.001$) (Figure 5).

Discussion

The present study explored the different roles between measured locations of PCAT attenuation in assessing coronary atherosclerosis. Peri-plaque PCAT attenuation was more valuable than proximal PCAT attenuation, and manual measurement of peri-plaque PCAT spectral quantitative parameters derived from SDCT also had important value. Therefore, peri-plaque measurement is recommended to obtain PCAT attenuation and spectral quantitative parameters, which are used to monitor and identify inflammation in

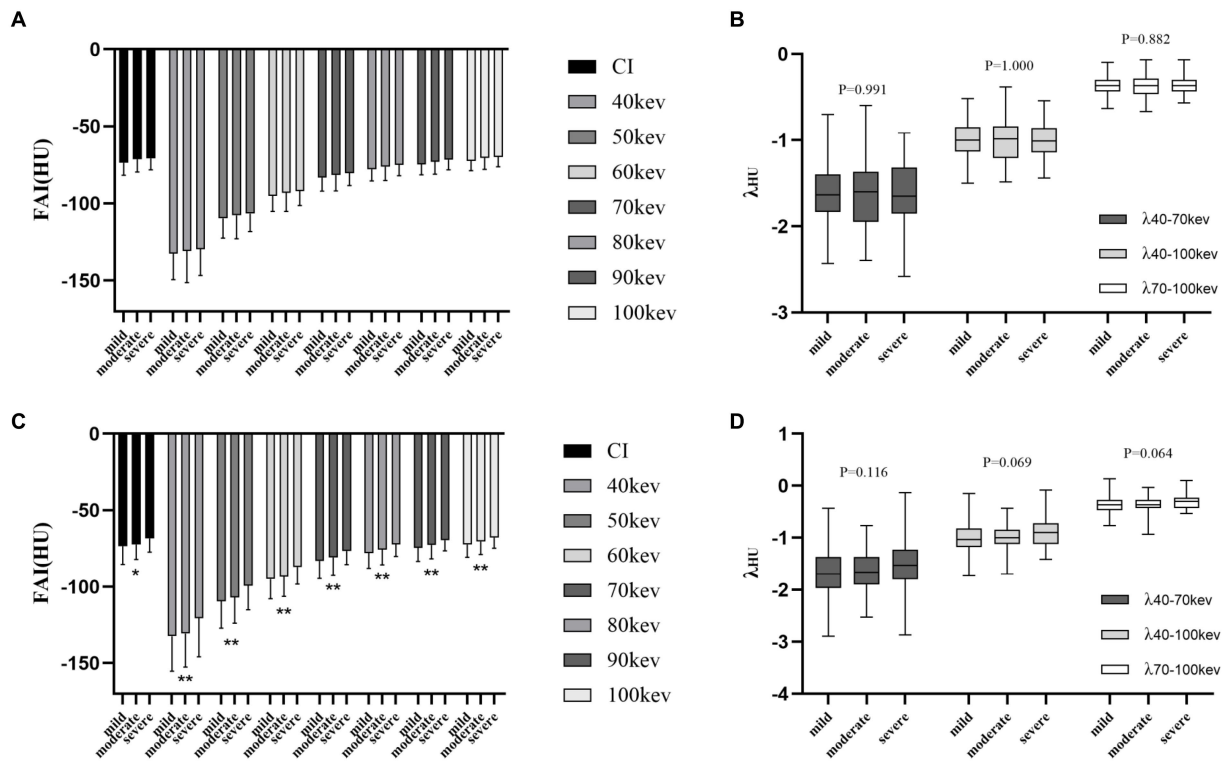


FIGURE 3
Differences in PCAT parameters at the luminal stenosis level. **(A)** Proximal FAI on CI and VMI. **(B)** Spectral attenuation slope of proximal PCAT at three monoenergetic intervals. **(C)** Peri-plaque FAI on CI and VMI. **(D)** Spectral attenuation slope of peri-plaque PCAT at three monoenergetic intervals. PCAT, pericoronary adipose tissue; FAI, fat attenuation index; CI, conventional image; and VMI, virtual monoenergetic image. *, $p < 0.05$; **, $p < 0.01$.

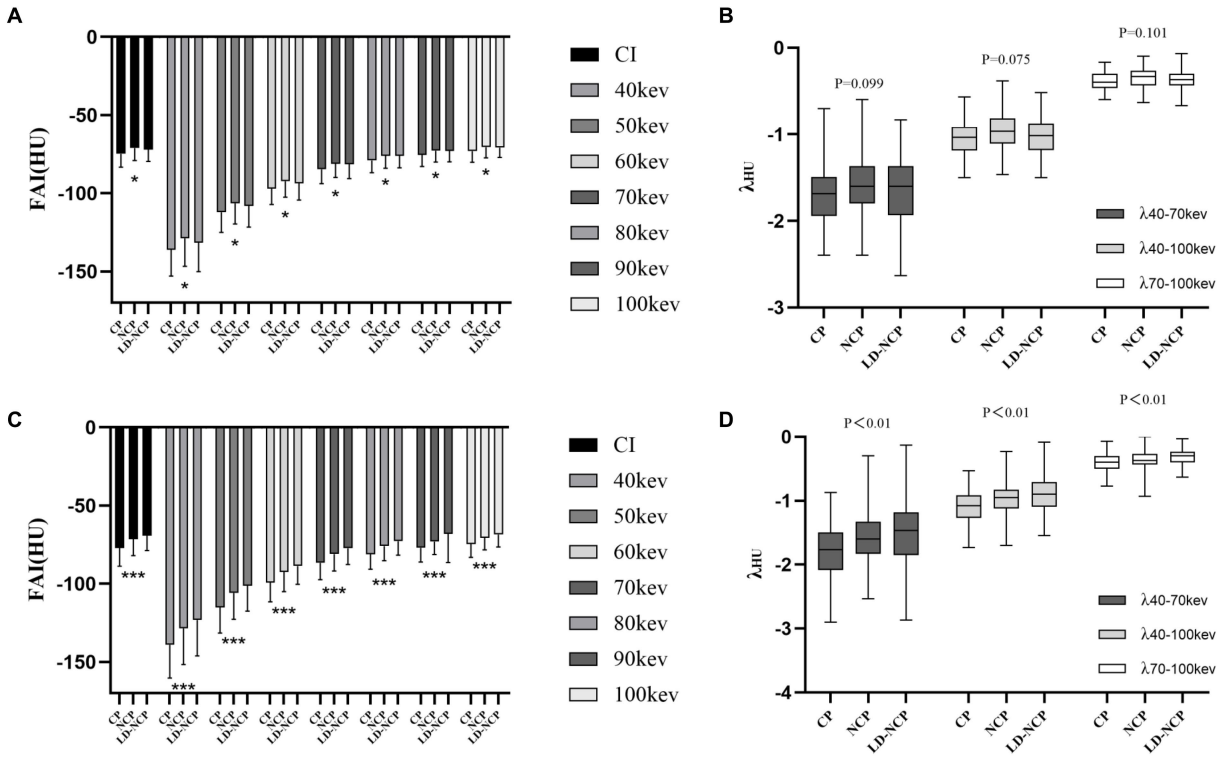


FIGURE 4
Differences in PCAT parameters at the plaque-type level. **(A)** Proximal FAI on CI and VMI. **(B)** Spectral attenuation slope of proximal PCAT at three monoenergetic intervals. **(C)** Peri-plaque FAI on CI and VMI. **(D)** Spectral attenuation slope of peri-plaque PCAT at three monoenergetic intervals. PCAT, pericoronary adipose tissue; FAI, fat attenuation index; CI, conventional image; VMI, virtual monoenergetic image; CP, calcified plaques; NCP, non-calcified plaque; LD-NCP, low-density non-calcified plaque. *, $p < 0.05$; ***, $p < 0.001$.

patients with CAD. To the best of our knowledge, this is the first study to explore the measured location and measured method of PCAT quantitative parameters on SDCT.

CAD is caused by coronary atherosclerosis and luminal stenosis or occlusion. Due to the substantial risk of recurrent cardiovascular (CV) events, CAD remains the leading single cause of death worldwide (1, 15). The key driver of atherogenesis and atherosclerotic plaque rupture resulting in acute coronary syndrome (ACS) has been confirmed to be vascular inflammation (16). Therefore, the detection of coronary inflammation has important implications for CV risk stratification and targeted medical therapy, in order to improve patient prognosis. The routine non-invasive method for systemic inflammation monitoring is based on laboratory tests. It is convenient and sensitive, whereas it lacks specificity for coronary inflammation (17). Thus, it cannot be used for assessing the process of coronary atherosclerosis or the presence of vulnerable plaques. Several studies have proved that the CT attenuation index can represent the functional status of PCAT, which is closely related to coronary inflammation (3–5, 7). Hence, the study of evaluating coronary inflammation and related clinical issues through PCAT attenuation has become one of the research hotspots in recent years.

As part of perivascular adipose tissue (PVAT) and EAT, PCAT is also a functional sensor of coronary inflammation. PCAT is an active metabolic fat pool that is closely located around coronary arteries. It has a complex bidirectional paracrine pathway with a coronary artery wall. Dysfunctional PCAT secretes pro-inflammatory adipocytokines, causing vascular inflammation and leading to the formation of coronary atherosclerosis plaques. Vascular inflammation prevents lipid accumulation in PCAT by inhibiting preadipocyte differentiation (4, 18, 19). They are affected by each other's inflammation and functional status. Therefore, it is possible to detect vascular inflammation non-invasively through PCAT attenuation using CCTA. In this study, we first confirmed the differences in PCAT attenuation among different degrees of luminal stenosis and among different coronary atherosclerotic plaque types, respectively. The result was particularly significant among plaque types. Therefore, the stages of plaque development may be an ideal target for better and earlier monitoring of atherosclerosis (20). In this study, we divided atherosclerotic plaques that caused the maximal degree of luminal stenosis into three types according to the attenuation characteristics on CCTA images. Among them, NCP and LD-NCP are the foundations of CV events and help to determine an individual's CV risk (21). This study on the analysis of the relationship between plaque types and PCAT attenuation found that the degree of inflammation around NCP and LD-NCP was higher than that of the CP group. This finding indicates that the inflammatory state of PCAT and the coronary artery is closely related to the instability of atherosclerotic plaques, while CP is relatively stable and the degree of inflammation around it is relatively mild.

In addition, different measured locations of PCAT have different results, which is the main finding of this study. In the CRISP-CT study, PCAT around the proximal RCA was selected as a representative imaging marker of coronary inflammation for the analysis of PCAT in each participant (6). The main reason is that the adipose tissue

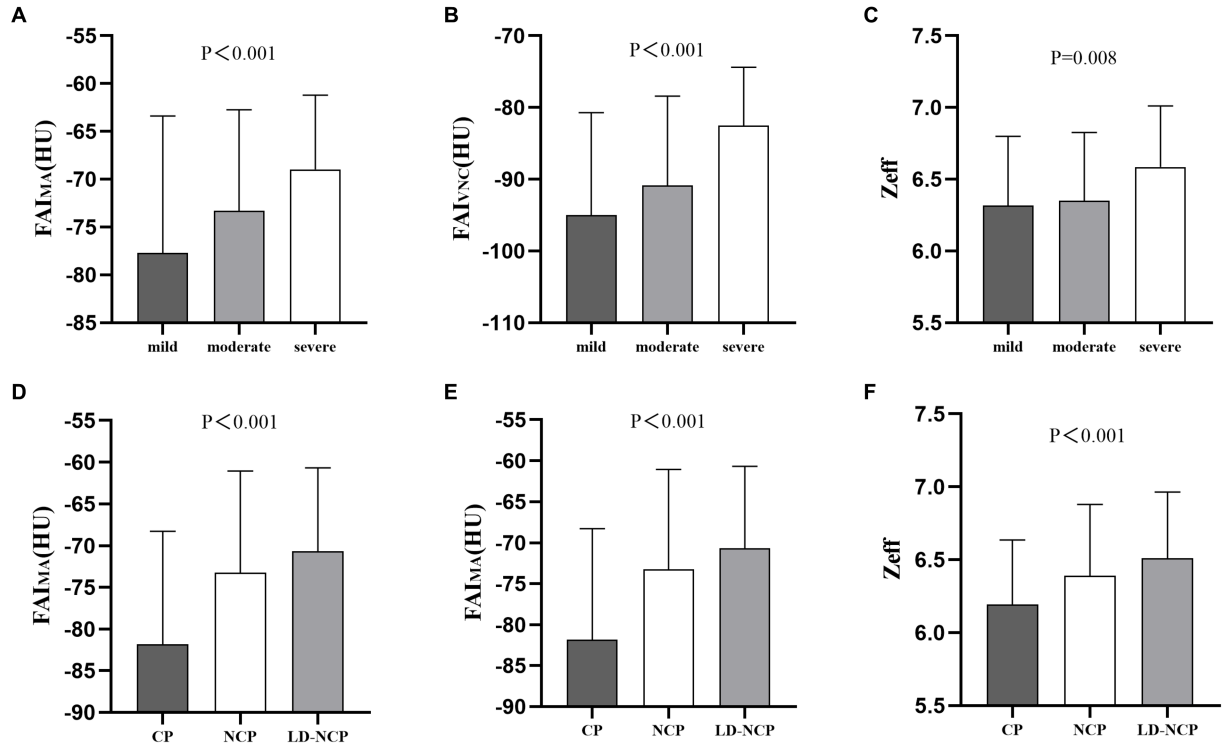


FIGURE 5
Differences in manually measured peri-plaque PCAT spectral parameters. **(A–C)** At luminal stenosis level. **(D–F)** At plaque-type level. FAI_{MA}, fat attenuation index measured manually; FAI_{VNC}, fat attenuation index on a virtual non-contrast image; Zeff, effective atomic number; CP, calcified plaques; NCP, non-calcified plaque; and LD-NCP, low-density non-calcified plaque.

around the proximal segment of RCA is relatively abundant, and it is easy to sketch, extract, and measure. However, this method may have limitations for the evaluation of some diffuse or focal lesions. It may not be sufficient for evaluating diffuse lesions, while for evaluating focal lesions, it may not be accurate enough because it covers normal adipose tissue. Therefore, some scholars have selected to adopt PCAT around the proximal segments of three major coronary arteries for attenuation measurement. However, they found that PCAT of LAD and LCX were not associated with event-free survival, although both PCAT of LAD and LCX and the occurrence of CV were correlated with RCT PCAT (6). In Alexios SA et al.'s study, they examined the relationships between FAI and radial measurement distance from the vascular wall in patients with coronary atherosclerosis compared to healthy individuals. The results showed that there was no difference in FAI 2 cm away from the coronary wall between the two groups. However, they found that FAI around the culprit lesion was higher than FAI proximal to the lesion. Therefore, they proposed that there is a gradient in adipocyte size and the expression of adipose genes moving from PVAT adjacent to coronary arteries to adipose tissue further away from the coronary wall, which makes a parallel shift of FAI to more negative values moving away from the coronary wall (4). Another study also found different FAI values with the radial distance for PCAT proximal to the coronary artery. As shown, FAI_{ref} (within the radial distance from the outer vessel wall equal to the coronary vessel diameter) proximal to LAD could better represent the inflammation of culprit lesions in patients with ACS than FAI with any other radial distances proximal to the coronary arteries or FAI around culprit lesions (22). In our study, we confirmed that there are indeed differences in the PCAT attenuation at different locations. What is different is that we have confirmed that the change in PCAT attenuation around the plaque is more significant than that of proximal coronary arteries, especially in severe stenostic vessels and LD-NCP. A possible explanation for this result may be that the lesion is mainly on a single coronary artery in most participants in this study. The paracrine signal sent to PCAT from the vascular inflammation is easier to be limited around the plaque and then transmitted to PCAT around the inflamed coronary arteries and other locations. Therefore, peri-plaque PCAT attenuation is a more robust and easily accessible measurement of coronary inflammation. It has a higher diagnostic value in CAD patients as compared to proximal PCAT attenuation.

Furthermore, SDCT was used, and several spectral parameters of PCAT were obtained to assess coronary arteriosclerosis in this study, which is another innovation. In our previous study (10), we confirmed that PCAT attenuation parameters obtained using dual-layer SDCT can aid in distinguishing patients with and without CAD, which might predict the formation of atherosclerotic plaques before they appear. Most previous studies on PCAT attenuation in evaluating coronary atherosclerosis and CV events were performed by conventional polyenergetic CT. In this study, PCAT attenuation parameters from CI and VMI were both adopted. In addition, manually measured peri-plaque PCAT attenuation on CI and VNC images and Zeff of peri-plaque PCAT were also used for assessing coronary atherosclerosis. The results confirmed that peri-plaque PCAT spectral parameters measured manually have significant value in estimating coronary atherosclerosis, which is similar to the results of automatically measured peri-plaque PCAT attenuation. As the most commonly used

imaging detection method, conventional CCTA can offer information on the morphological manifestation of coronary plaques and the structural change in vascular walls. However, SDCT can provide more functional information and quantitative parameters. It achieves accurate "homologous, simultaneous, codirectional and synchronous" spectral scanning based on collecting low- and high-energy data simultaneously using a dual-layer detector (23, 24). Therefore, it can derive more real and accurate data for measurement and diagnosis based on non-invasive multimodal images. Besides, it can also supply more spectral quantitative parameters. In this study, the combination of multiple PCAT spectral parameters can better diagnose and assess the severity of luminal stenosis and atherosclerotic plaques.

This study has some limitations. First, this is a single-center retrospective study, which is based on vessel level and the specific device. Therefore, culprit plaques, follow-up prognosis, and the influence of different types and brands of CT imaging devices were not analyzed, which requires further and prospective research in the future. Second, the laboratory test was not included in this study. The calcified score and invasive coronary angiography were not adopted, either. Thus, the level of coronary inflammation and the severity of coronary atherosclerosis could not be assessed comprehensively and accurately. Third, only the relationship between the longitudinal location of PCAT parameters and coronary atherosclerosis was analyzed, while PCAT attenuation with different radial distances was not measured. Although peri-plaque PCAT attenuation parameters are expected to be used as a standard biomarker for evaluating the severity of coronary atherosclerosis, further randomized controlled trials with a large sample and detailed analysis are required to verify this hypothesis.

In conclusion, peri-plaque PCAT is more valuable than proximal PCAT in assessing coronary atherosclerosis. Spectral parameters of peri-plaque PCAT can further improve the diagnostic accuracy of coronary atherosclerosis. The quantitative parameters related to peri-plaque PCAT attenuation are expected to become standard biomarkers and predictors for assessing plaque vulnerability and the hemodynamic characteristics of coronary atherosclerosis.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by the First Affiliated Hospital of Harbin Medical University (No. 202214). Written informed consent was not required as per local legislation and institutional requirements.

Author contributions

YJ: Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Data curation, Formal analysis. LZ: Data curation, Methodology, Writing – original draft, Formal analysis. MX: Data curation, Formal analysis. XZ: Data curation, Formal analysis, Writing – original draft. XX: Conceptualization, Project administration, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Acknowledgments

The authors thank RT Feng Yin and Jing Zhu for patient scanning and Dr. Xiaomei Lu for technical assistance with the imaging.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2024.1357981/full#supplementary-material

References

1. World Health Organization. Cardiovascular diseases (CVDs). (2021). Available at:https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases.

2. Mazurek T, Zhang L, Zalewski A, Mannion JD, Diehl JT, Arafat H, et al. Human epicardial adipose tissue is a source of inflammatory mediators. *Circulation*. (2003) 108:2460–6. doi: 10.1161/01.CIR.0000099542.57313.C5

3. Mancio J, Oikonomou EK, Antoniades C. Perivascular adipose tissue and coronary atherosclerosis. *Heart*. (2018) 104:1654–62. doi: 10.1136/heartjnl-2017-312324

4. Antonopoulos AS, Sanna F, Sabharwal N, Thomas S, Oikonomou EK, Herdman L, et al. Detecting human coronary inflammation by imaging perivascular fat. *Sci Transl Med*. (2017) 9:eaal2658. doi: 10.1126/scitranslmed.aal2658

5. Goeller M, Achenbach S, Duncker H, Dey D, Marwan M. Imaging of the Pericoronary adipose tissue (PCAT) using cardiac computed tomography: modern clinical implications. *J Thorac Imaging*. (2021) 36:149–61. doi: 10.1097/RTI.0000000000000583

6. Oikonomou EK, Marwan M, Desai MY, Mancio J, Alashi A, Hutt Centeno E, et al. Non-invasive detection of coronary inflammation using computed tomography and prediction of residual cardiovascular risk (the CRISP CT study): a post-hoc analysis of prospective outcome data. *Lancet*. (2018) 392:929–39. doi: 10.1016/S0140-6736(18)31114-0

7. Goeller M, Tamarappoo BK, Kwan AC, Cadet S, Commandeur F, Razipour A, et al. Relationship between changes in pericoronary adipose tissue attenuation and coronary plaque burden quantified from coronary computed tomography angiography. *Eur Heart J Cardiovasc Imaging*. (2019) 20:636–43. doi: 10.1093/ehjci/jez013

8. Goeller M, Rahman Ihdayhid A, Cadet S, Lin A, Adams D, Thakur U, et al. Pericoronary adipose tissue and quantitative global non-calcified plaque characteristics from CT angiography do not differ in matched south Asian, east Asian and European-origin Caucasian patients with stable chest pain. *Eur J Radiol*. (2020) 125:108874. doi: 10.1016/j.ejrad.2020.108874

9. van Diemen PA, Bom MJ, Driessen RS, Schumacher SP, Everaars H, de Winter RW, et al. Prognostic value of RCA Pericoronary adipose tissue CT-attenuation beyond high-risk plaques, plaque volume, and ischemia. *JACC Cardiovasc Imaging*. (2021) 14:1598–610. doi: 10.1016/j.jcmg.2021.02.026

10. Zou L, Xiao X, Jia Y, Yin F, Zhu J, Gao Q, et al. Predicting coronary atherosclerosis heart disease with pericoronary adipose tissue attenuation parameters based on dual-layer spectral detector computed tomography: a preliminary exploration. *Quant Imaging Med Surg*. (2023) 13:2975–88. doi: 10.21037/qims-22-1019

11. Ma R, van Assen M, Ties D, Pelgrim GJ, van Dijk R, Sidorenkov G, et al. Focal pericoronary adipose tissue attenuation is related to plaque presence, plaque type, and stenosis severity in coronary CTA. *Eur Radiol*. (2021) 31:7251–61. doi: 10.1007/s00330-021-07882-1

12. Dang Y, Chen X, Ma S, Ma Y, Ma Q, Zhou K, et al. Association of Pericoronary Adipose Tissue Quality Determined by dual-layer spectral detector CT with severity of coronary artery disease: a preliminary study. *Front Cardiovasc Med*. (2021) 8:720127. doi: 10.3389/fcvm.2021.720127

13. Cheng VY, Wolak A, Gutstein A, Gransar H, Wong ND, Dey D, et al. Low-density lipoprotein and noncalcified coronary plaque composition in patients with newly diagnosed coronary artery disease on computed tomographic angiography. *Am J Cardiol*. (2010) 105:761–6. doi: 10.1016/j.amjcard.2009.11.007

14. Cury RC, Abbara S, Achenbach S, Agatston A, Berman DS, Budoff MJ, et al. CAD-RADS(TM) coronary artery disease -reporting and data system. An expert consensus document of the Society of Cardiovascular Computed Tomography (SCCT), the American College of Radiology (ACR) and the north American Society for Cardiovascular Imaging (NASCI). Endorsed by the American College of Cardiology. *J Cardiovasc Comput Tomogr*. (2016) 10:269–81. doi: 10.1016/j.jcct.2016.04.005

15. Cholesterol Treatment Trialists' (CTT) CollaborationBaigent C, Blackwell L, Emberson J, Holland LE, Reith C, et al. Efficacy and safety of more intensive lowering of LDL cholesterol: a meta-analysis of data from 170,000 participants in 26 randomised trials. *Lancet*. (2010) 376:1670–81. doi: 10.1016/S0140-6736(10)61350-5

16. Libby P, Ridker PM, Maseri A. Inflammation and atherosclerosis. *Circulation*. (2002) 105:1135–43. doi: 10.1161/hc0902.104353

17. Nakajima A, Sugiyama T, Araki M, Seegers LM, Dey D, McNulty I, et al. Compared with plaque Erosion, is associated with a higher level of Pancoronary inflammation. *JACC Cardiovasc Imaging*. (2022) 15:828–39. doi: 10.1016/j.jcmg.2021.10.014

18. Lin A, Dey D, Wong DTL, Nerlekar N. Perivascular adipose tissue and coronary atherosclerosis: from biology to imaging phenotyping. *Curr Atheroscler Rep*. (2019) 21:47. doi: 10.1007/s11883-019-0817-3

19. Ridker PM, Everett BM, Thuren T, MacFadyen JG, Chang WH, Ballantyne C, et al. Antiinflammatory therapy with Canakinumab for atherosclerotic disease. *N Engl J Med*. (2017) 377:1119–31. doi: 10.1056/NEJMoa1707914

20. Darçot E, Colotti R, Pellegrin M, Wilson A, Siegert S, Bouzourene K, et al. Towards quantification of inflammation in atherosclerotic plaque in the clinic -characterization and optimization of Fluorine-19 MRI in mice at 3 T. *Sci Rep*. (2019) 9:17488. doi: 10.1038/s41598-019-53905-9

21. Channon KM, Newby DE, Nicol ED, Deanfield J. Cardiovascular computed tomography imaging for coronary artery disease risk: plaque, flow and fat. *Heart*. (2022) 108:1510–5. doi: 10.1136/heartjnl-2021-320265

22. Dong X, Zhu C, Li N, Shi K, Si N, Wang Y, et al. Identification of patients with acute coronary syndrome and representation of their degree of inflammation: application of pericoronary adipose tissue within different radial distances of the proximal coronary arteries. *Quant Imaging Med Surg*. (2023) 13:3644–59. doi: 10.21037/qims-22-864

23. Rassouli N, Etesami M, Dhanantwari A, Rajiah P. Detector-based spectral CT with a novel dual-layer technology: principles and applications. *Insights Imaging*. (2017) 8:589–98. doi: 10.1007/s13244-017-0571-4

24. Ehn S, Sellerer T, Muenzel D, Fingerle AA, Kopp F, Duda M, et al. Assessment of quantification accuracy and image quality of a full-body dual-layer spectral CT system. *J Appl Clin Med Phys*. (2018) 19:204–17. doi: 10.1002/acm2.12243

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 19 April 2024
DOI 10.3389/fmed.2024.1328687

Check for updates

OPEN ACCESS

EDITED BY
Nicolas A. Karakatsanis,
Cornell University, United States

REVIEWED BY
Naoaki Sakata,
Fukuoka University, Japan
Lixia Wang,
Cedars Sinai Medical Center, United States

*CORRESPONDENCE
Yaping Lv
13977287067@126.com
Lei Chen
lei.chen@uii-ai.com

†These authors have contributed equally to this work and share first authorship

RECEIVED 27 October 2023
ACCEPTED 03 April 2024
PUBLISHED 19 April 2024

CITATION
Jiang W, Pan X, Luo Q, Huang S, Liang Y, Zhong X, Zhang X, Deng W, Lv Y and Chen L (2024) Radiomics analysis of pancreas based on dual-energy computed tomography for the detection of type 2 diabetes mellitus. *Front. Med.* 11:1328687. doi: 10.3389/fmed.2024.1328687

COPYRIGHT
© 2024 Jiang, Pan, Luo, Huang, Liang, Zhong, Zhang, Deng, Lv and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Radiomics analysis of pancreas based on dual-energy computed tomography for the detection of type 2 diabetes mellitus

Wei Jiang[1†], Xianpan Pan[2†], Qunzhi Luo[1], Shiqi Huang[1], Yuhong Liang[1], Xixi Zhong[1], Xianjie Zhang[2], Wei Deng[2], Yaping Lv[1*] and Lei Chen[2*]

[1]Department of Radiology, Liuzhou Municipal Liutie Central Hospital, Liuzhou, China, [2]Shanghai United Imaging Intelligence Co., Ltd., Shanghai, China

Objective: To utilize radiomics analysis on dual-energy CT images of the pancreas to establish a quantitative imaging biomarker for type 2 diabetes mellitus.

Materials and methods: In this retrospective study, 78 participants (45 with type 2 diabetes mellitus, 33 without) underwent a dual energy CT exam. Pancreas regions were segmented automatically using a deep learning algorithm. From these regions, radiomics features were extracted. Additionally, 24 clinical features were collected for each patient. Both radiomics and clinical features were then selected using the least absolute shrinkage and selection operator (LASSO) technique and then build classifies with random forest (RF), support vector machines (SVM) and Logistic. Three models were built: one using radiomics features, one using clinical features, and a combined model.

Results: Seven radiomic features were selected from the segmented pancreas regions, while eight clinical features were chosen from a pool of 24 using the LASSO method. These features were used to build a combined model, and its performance was evaluated using five-fold cross-validation. The best classifier type is Logistic and the reported area under the curve (AUC) values on the test dataset were 0.887 (0.73–1), 0.881 (0.715–1), and 0.922 (0.804–1) for the respective models.

Conclusion: Radiomics analysis of the pancreas on dual-energy CT images offers potential as a quantitative imaging biomarker in the detection of type 2 diabetes mellitus.

KEYWORDS

pancreas, dual-energy CT, type 2 diabetes mellitus, radiomics analysis, deep learning

1 Introduction

The anatomy of the human pancreas is closely related to its endocrine and exocrine functions. In individuals diagnosed with type 1 or type 2 diabetes, alterations in the pancreas have been observed (1). Chronic inflammation of the pancreas can cause damage to the insulin-producing cells, potentially leading to the development of diabetes. Pancreatitis and

type 2 diabetes share similar risk factors (1). Several imaging techniques, such as computed tomography (CT), magnetic resonance imaging (MRI), and ultrasound (US), have been used in various studies to investigate pancreatic changes in individuals with diabetes. These studies aim to assess the size (diameter, area, or volume) as well as the fat content of the pancreas using these imaging modalities. The comparisons are made between individuals with type 1 diabetes (T1DM) and/or type 2 diabetes (T2DM) and healthy controls (2).

These imaging studies hold significant potential for providing reliable insights into diabetes mellitus (DM). Between 2008 and 2013, a total of 1,478 lean individuals without diabetes underwent CT scans (3). The presence of fatty pancreas was evaluated using a validated histological method, which measured the attenuation of the pancreas on CT scans conducted at the beginning of the study. Lower pancreas attenuation values indicate higher fat content in the pancreas (3). Furthermore, a study aimed to investigate abdominal CT biomarkers for type 2 diabetes mellitus using advanced automated deep learning techniques within a substantial clinical dataset (4). The analysis demonstrated a correlation between the diagnosis of type 2 diabetes mellitus and specific abdominal CT biomarkers, particularly measurements related to pancreatic CT attenuation and visceral fat (4). Dual-energy CT (DECT) has shown promise in providing more precise quantitative information compared to conventional methods by utilizing two different X-ray beams with distinct absorption characteristics (5).

Radiomics analysis has made significant advancements in the field of medical image analysis in recent years (6). This approach enables the extraction of numerous features from medical images, facilitating the quantification of phenotypic characteristics of tumors (7). These quantifications play a crucial role in various areas such as diagnosis, clinical prognosis, treatment selection, and decision support. By optimizing the selection of features and utilizing machine learning algorithms, it becomes possible to effectively differentiate between patients with similar outcome conditions and establish personalized prediction models based on scientific and data-driven analyses for treatment outcomes. The emerging field of radiomics also holds promise in identifying previously undetectable characteristics and assisting in the diagnosis and prediction of diabetes through pancreas imaging. For instance, a study focused on evaluating the diagnostic accuracy of a Dual-Energy Computed Tomography (DECT)-based technique that utilizes iodine quantification and fat fraction analysis for early detection of acute pancreatitis (8). The results indicated that DECT with iodine quantification exhibits higher sensitivity in diagnosing early acute pancreatitis compared to standard image evaluation methods (8). Furthermore, another study aimed to assess whether an AI model based on pancreas radiomics can identify the CT imaging pattern associated with type 2 diabetes (9). The findings showed that the model successfully detects the imaging pattern linked to type 2 diabetes. However, further enhancements and validation are necessary to evaluate its potential for identifying type 2 diabetes in the millions of CT scans conducted annually.

Our previous study investigated the clinical value of pancreatic fat fraction measured on DECT images for the detection of type 2 diabetes mellitus (10). Given the burgeoning evidence linking pancreatic imaging characteristics with DM and the advancements in imaging and analytical technologies, we hypothesize that radiomics analysis of dual-energy CT images of the pancreas can identify specific imaging biomarkers that quantitatively differentiate individuals with type 2 diabetes mellitus from non-diabetic controls. This approach aims to establish a novel, quantitative imaging biomarker for T2DM, leveraging the precision of DECT and the analytical power of radiomics to enhance early detection and intervention strategies for diabetes mellitus.

2 Materials and methods

2.1 Study population

This retrospective study received approval from the Ethics Committee of Liuzhou Municipal Liutie Central Hospital (Approval No. 2021033), and the requirement for informed consent was waived. We conducted a search in our medical information database for cases that occurred between September 2021 and July 2022. Eligible patients meeting the following criteria were included in the study: (1) those who underwent dual-energy abdominal CT with a non-contrast phase; (2) individuals who had blood tests done within 3 days before and after DECT; and (3) patients whose electronic medical records clearly indicated either "type 2 diabetes" or "non-diabetes." Patients with malignant tumors ($n = 12$), pancreatitis ($n = 5$), or significant pancreatic atrophy without visible pancreatic parenchyma ($n = 4$) were excluded from the study.

2.2 Datasets

2.2.1 DECT images

Abdominal DECT was conducted using a Dual-source CT scanner (SOMATOM Drive, Siemens Healthineers, Forchheim, Germany). The tube voltages used were 100 kV and Sn140 kV, with average effective tube currents of 250 mAs and 193 mAs. Automatic tube current modulation technology (CARE Dose 4D) was employed for dose control. The CT parameters were as follows: detector collimation of 32×0.6 mm, pitch of 0.6, gantry rotation time of 0.5 s, and matrix size of 512×512. Reconstruction of the CT image was performed using an iterative reconstruction technique called Adaptive Model-based Iterative Reconstruction (ADMIRE) with the Q30f algorithm, resulting in a reconstructed image thickness of 1.5 mm (see Figure 1).

2.2.2 Clinical variables

The following 24 clinical data were collected: (1) categorical variables: gender, lipid turbidity index, hemolysis index, jaundice index, and clinical manifestations (now the criteria for the diagnosis of diabetes include the patient exhibiting symptoms of polydipsia, polyuria, polyphagia, and weight loss (referred to as "three excesses and one deficiency"), along with elevated blood glucose levels and the presence of glucose in the urine (which should not be found in normal urine). So if the subject has these manifestations the category of the subject is "positive"); (2) normally distributed variables: age, albumin, globulin, systolic blood pressure, total cholesterol, and high-density lipoprotein; (3) non-normally distributed variables: hematocrit, low-density lipoprotein. Others are listed in Table 1.

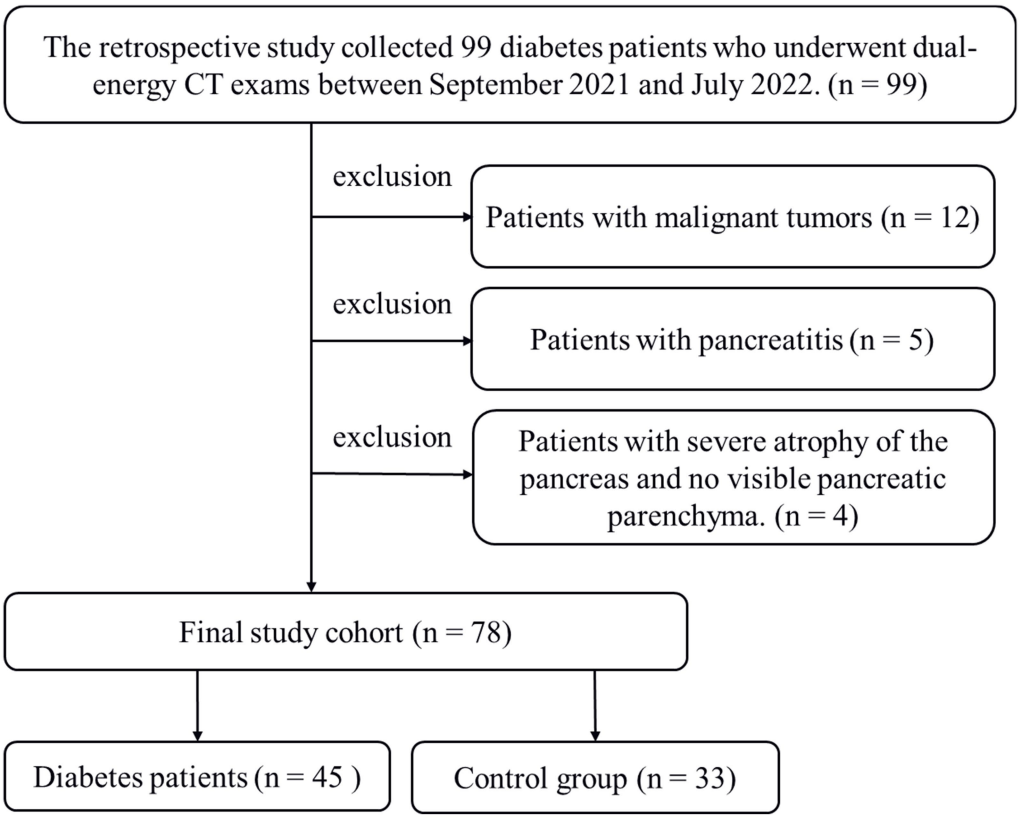


FIGURE 1
Flowchart of the study including diabetes patients enrolled with dual-energy CT images.

2.3 Pancreas segmentation on DECT images

In our previous work (11), we implemented a deep learning framework that utilized a cascade coarse-to-fine segmentation approach with an attention mechanism to segment organs. In this study, we applied this deep learning framework specifically to segment the pancreas from fusion DECT images. The segmentation process was carried out using the uAI Research Portal (United Imaging Intelligence, China) (12), which is a clinical research platform based on the Python programming language (version 3.7.3). Figure 2 displays the original pancreas DECT image and the corresponding segmentation results. In our previous work (11), we conducted a quantitative analysis comparing the DECT results between the diabetes and control groups. Figure 2 demonstrates that the head of the pancreas exhibited lower density and higher fat fraction compared to the body and tail of the pancreas in the diabetes group. Finally, all delineations were reviewed by a highly experienced chief radiologist with 8 years of expertise in abdominal imaging.

2.4 Radiomics extraction and selection

A total of 2,264 radiomic features were extracted from each pancreas region. These features consisted of 104 original features, which were further categorized into 18 first-order statistics, 14 shape features, and 21 texture features. The texture features included Gray-Level Co-occurrence Matrix (GLCM), Gray-Level Run-Length Matrix (GLRLM), Gray-Level Size-Zone Matrix (GLSZM), Gray-Level Dependence Matrix (GLDM), and Neighboring Gray-Tone Difference Matrix (NGTDM). Additionally, 14 image filters including Box Mean, Additive Gaussian Noise, Binomial Blur Image, Curvature Flow, Box Sigma Image, LoG with sigma values of 0.5, 1, 1.5, and 2 were applied to generate derived images. Derived images were further processed using Wavelet filters (LLL, LLH, LHL, LHH, HLL, HLH, HHL, HHH), Normalize, Laplacian Sharpening, Discrete Gaussian Mean Speckle Noise Recursive Gaussian and Shot Noise to extract first-order statistics and texture features within the pancreas regions. This resulted in a total of 2,160 derived features.

Figure 3 illustrates the flowchart of the radiomics analysis. First, radiomic features were calculated from the region of interest (ROI) on each DECT image. All the radiomic features were then normalized using *Z*-score. Next, the least absolute shrinkage and selection operator (LASSO) technique was applied to sift through these normalized radiomic features, with the aim of identifying those with the highest predictive reliability for type 2 diabetes mellitus (T2DM). This critical step focused on isolating features that are predictive of the binary outcome of diabetes presence (yes or no). Then a similar LASSO selection process was conducted for the clinical features gathered from the study participants. The radiomic and clinical features deemed significant through these selection processes were then amalgamated. A final round of LASSO selection was performed on this combined set to refine and identify a comprehensive feature set.

TABLE 1 Clinical variables of the study population.

Indicators	Control group (n = 33)	Diabetic group (n = 45)	p-value	Note
Categorical variables				
Gender, Female:Male	12:21	19:26	0.601	
Clinical manifestations None:Present	33:0	36:9	0.006	
Lipid turbidity index	6:26:1:0	4:38:2:1	0.535	0:1:2:3
Hemolysis index	20:12:0:1	36:7:2:0	0.065	0:1:2:3
Jaundice index	2:28:2:1	2:37:6:0	0.483	0:1:2:3:4
Normal distribution variables				
Age (years)	59 ± 14	64 ± 10	0.083	
Albumin	28.2 ± 5.0	28.6 ± 7.4	0.771	
*Globulin	40.6 ± 3.6	35.8 ± 5.0	0.000	
Systolic blood pressure	135.8 ± 22.6	137.3 ± 21.5	0.755	mmHg
Total cholesterol	4.8 ± 0.9	5.0 ± 1.6	0.416	mmol/L
High-density lipoprotein	1.3 ± 0.3	1.3 ± 0.3	0.721	mmol/L
Non-normally distributed variables				
Endogenous creatinine clearance rate (%)	72.2 (38.9)	90.2 (38.2)	0.004	
Hematocrit (%)	42.6 (5.2)	40.6 (9.6)	0.055	
Albumin/Globulin ratio (%)	1.4 (0.5)	1.3 (0.4)	0.005	
Low-density lipoprotein	2.7 (1.1)	2.7 (1.7)	0.903	mmol/L
Indirect bilirubin	6.7 (4.6)	7.1 (4.5)	0.980	
AST	18.0 (8.0)	17.0 (13.5)	0.598	IU/L
ALT	18.0 (9.5)	19.0 (17.5)	0.567	IU/L
*Total protein	69.5 (9.9)	63.8 (7.4)	0.002	
Triglycerides	1.3 (0.9)	1.5 (1.2)	0.482	mmol/L
Total bilirubin	12.2 (7.7)	11.9 (5.2)	0.561	μmol/L
Direct bilirubin	4.4 (2.1)	4.3 (2.7)	0.675	μmol/L
Creatinine	73.0 (17.5)	71.0 (36.0)	0.992	μmol/L
Diastolic blood pressure	78.0 (16.5)	80.0 (15.0)	0.689	mmHg

ALT, alanine transaminase; AST, aspartate transaminase; clinical manifestations, polyphagia, polydipsia, polyuria, and weight loss; categorical variables, described using sample proportions; normal distribution variables, described using mean ± standard deviation; non-normally distributed variables, described using median (interquartile range).

LASSO selection was employed to identify the most reliable predictive radiomic features. Finally, another round of LASSO selection was performed to evaluate clinical features. The selected radiomic and clinical features were merged and subjected to another round of LASSO selection to obtain a comprehensive combined feature set.

2.5 Model construction

Three machine learning models were developed for binary classification (diabetes or not) using three classifiers: random forest (RF), support vector machine (SVM), and logistic regression (LR). The models were built based on the selected features and/or clinical features. The input data for the detection model came from one of the three feature sets: (1) radiomic features with 7 variables, (2) clinical features with 8 variables, or (3) a combination of all features with 8 variables. To enhance performance, a grid search was performed to fine-tune the parameters for different features and classification algorithms.

2.6 Statistical analysis

The receiver operating characteristic curve (ROC) was generated to evaluate the performance of the detection model, and several performance metrics such as sensitivity (SEN), specificity (SPE), accuracy (ACC), F1-Score, and area under the curve (AUC) were computed. The confidence intervals for cross-validated AUC were computed to estimate the performance of each model. The demographic data was processed using the uAI Research Portal to examine significant differences in variables between the training set and the validation set. Python (version 3.6) was utilized for programming the training, validation of the prediction model, and conducting statistical analysis.

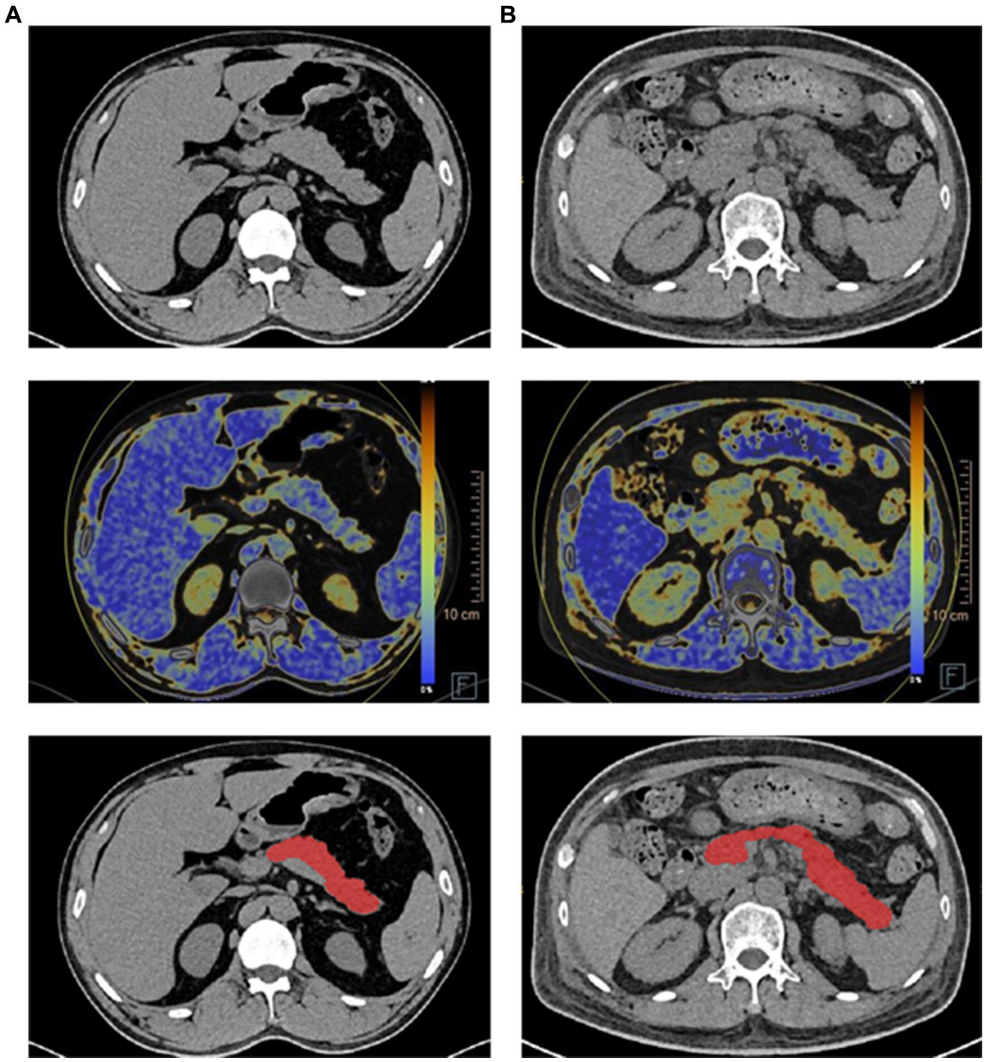


FIGURE 2
Illustration of pancreas DECT images, dual-energy fat map, and the segmentation results (each row from top to bottom) of the healthy subject and diabetes patient. **(A)** Healthy subject. **(B)** Diabetes patient.

3 Results

Based on Harrell's guideline, the number of selected features should be less than 10% of the sample size. Consequently, in our experiment involving the radiomic and clinical features, as well as the final combination of radiomic and clinical features, the number of selected features was less than 10% of the sample size (13, 14).

3.1 Assessment of radiomic and clinical features

A total of 2,264 radiomic features were computed using the uAI Research Portal for each pancreas region. These features were then normalized using the *Z*-score approach. The feature selection method, Lasso, was employed to reduce dimensionality. Ultimately, 7-dimensional features were selected for the subsequent classification modeling. The names and corresponding importance coefficients of these 7 features calculated by Lasso are illustrated in Figure 4A. The selection of coefficient values was computed using the least square method. Each coefficient signifies the average impact of the corresponding feature on the selection results. In simpler terms, a higher value indicates a greater importance of the feature for the detection model.

Additionally, for each patient, a total of 24 clinical features were digitized and normalized using the *Z*-score method. Employing a similar feature selection procedure with Lasso, 8 clinical features were chosen. The names and coefficients of each clinical feature are featured in Figure 4B.

Moreover, a combination of all extracted radiomic features (2,264) and clinical features (24) resulted in a total of 2,288 features. Through Lasso computation, an optimal subset of 8 features was obtained, consisting of 6 radiomic features and 2 clinical features. Detailed information regarding all selected features is presented in Figure 4C.

3.2 Evaluation of models

The feature selection step automatically chose the most significant features for this classification task. In this step, all

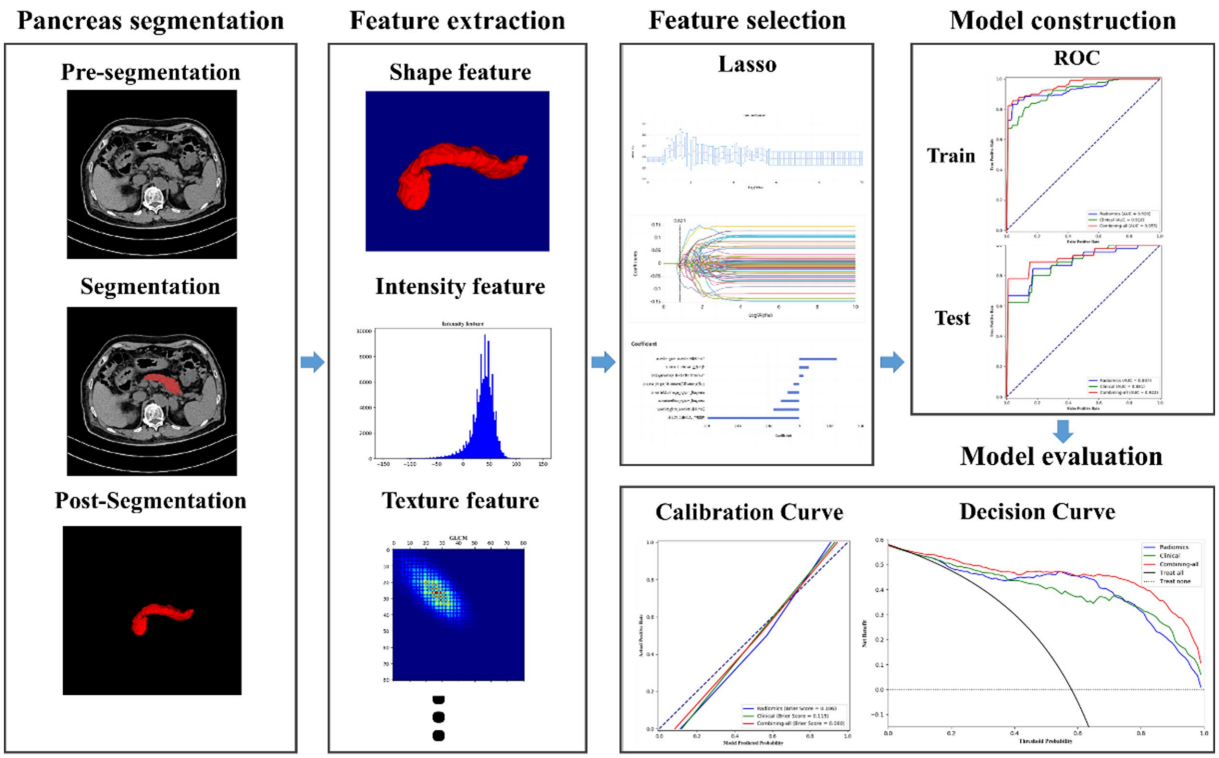


FIGURE 3
Flowchart of radiomics analysis on pancreas DECT images.

features were given equal priority and underwent *Z*-score normalization as a preprocessing step before using LASSO. To avoid overfitting, the LASSO parameter α was set to 0.05. The logistic regression classifier was then used to construct the detection model with the following parameter settings: a penalty factor C of 1.0, no class weight, L2 penalty type, a threshold of 0.5, and a tolerance of 1×10^{-4}.

To effectively assess the performance of the detection models, the five-fold cross-validation technique was implemented due to the limited amount of data. Cross-validation is a widely accepted statistical technique used to evaluate predictive models by partitioning the original dataset into a training set to train the model, and a testing set to evaluate it. In the context of our study, we divided the complete dataset into five equal (or nearly equal) parts randomly. During the validation process, four of these subsets are combined to form the training set, and the remaining one subset is used as the testing set. This process is repeated five times (folds), with each of the five subsets serving as the testing set exactly once.

LR, RF, and SVM classifiers were constructed for each fold using the selected features, which consisted of radiomic features, clinical features, and combined features. The overall performance was summarized by calculating the mean AUC, sensitivity, specificity, accuracy, and F1 score for both the training set and testing set. The results are presented in Table 2, we can found that LR get the best performance among these three models (RF, SVM, and Logistic) in AUC 0.922, Specificity 0.886, Accuracy 0.862, and F1 Score 0.871 indexes on test dataset with combined model, which demonstrate LR model having a robust classification ability. Then we do some other tests in the LR model, according to the Delong test, there was no statistically significant difference in the area under the ROC curve between the training set and testing set (as shown in Figure 5).

3.3 Regenerate response

The performance of the models built with radiomic features, clinical features, and combined features shown in Table 2.

4 Discussion

In our previous study (10), we calculated the fat fraction and CT value of the head, body, and tail of the pancreas from dual-energy CT images of 45 patients with type 2 diabetes mellitus (T2DM) and 33 control subjects. The experimental results demonstrated a significant association between the fat content of the pancreas and diabetes (10). While our focus was on fat fraction measurements, there are numerous other quantitative parameters that can be derived from dual-energy CT data. Hence, further studies could evaluate additional parameters such as radiomic features.

This retrospective study aims to establish a quantitative imaging biomarker for type 2 diabetes mellitus using dual-energy CT images of the pancreas. We have constructed three models using radiomic features, clinical features, and combined features, obtained through the LASSO regression approach, as shown in Table 3. The coefficients listed in Table 3 were derived from LASSO regression and represent the average impact of each corresponding feature on the classification results. A higher coefficient value indicates greater importance of the feature for the detection model.

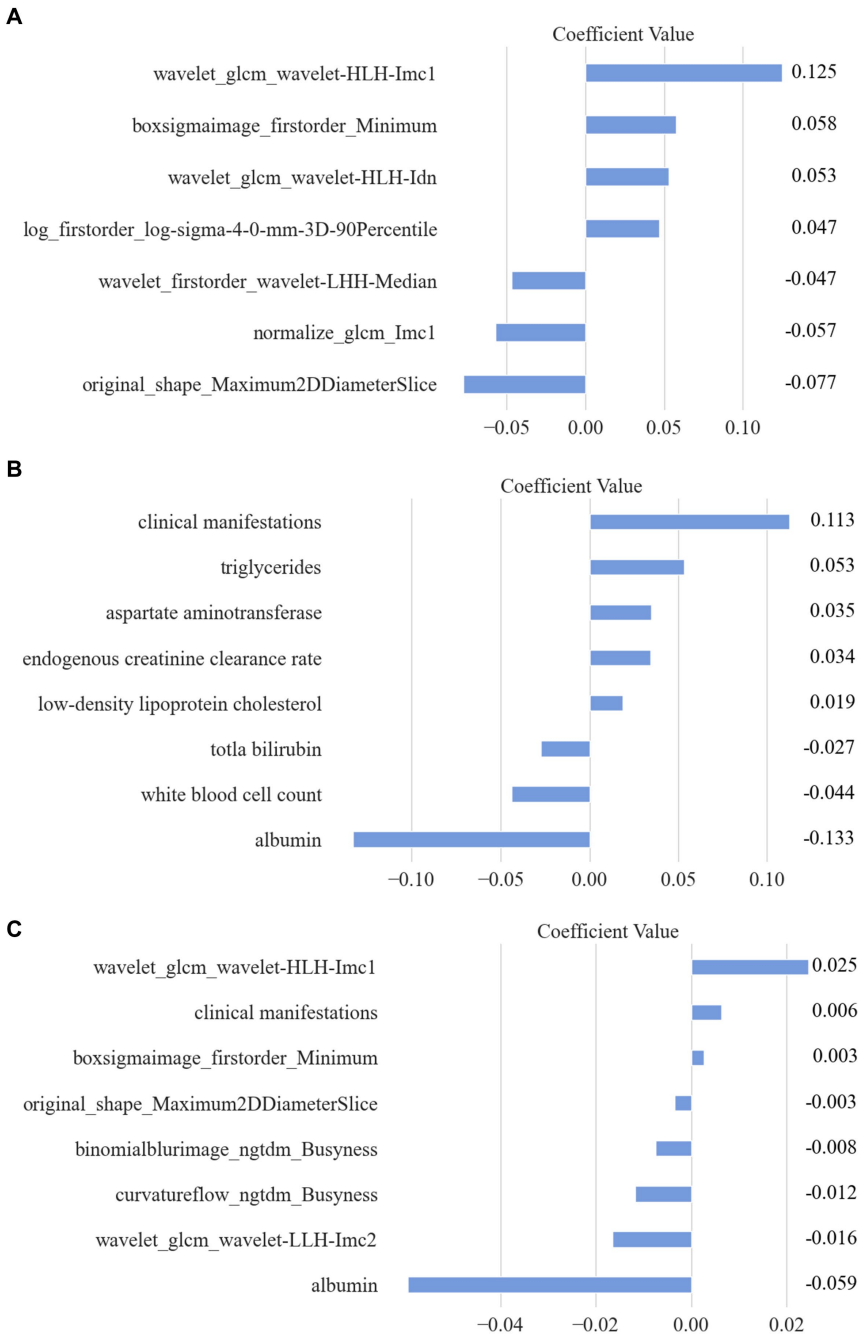


FIGURE 4
The selected features and their corresponding importance coefficients calculated by Lasso. **(A)** Selected Radiomic features. **(B)** Selected clinical features. **(C)** Combined features.

The description and explanation of each feature for the three models are as follows: these radiomics features are linked to the focus of the doctor's attention, such as image texture, gray level intensity, local homogeneity, and etc.

1. Informational Measure of Correlation (IMC) 1: IMC1 evaluates the correlation between the probability distributions of variables i and j, using mutual information $I(i,j)$. This feature quantifies the complexity of the texture within the region of interest (ROI).
2. Minimum: This is a first-order feature that represents the minimum gray level intensity within the ROI.
3. Inverse Difference Normalized (IDN): IDN is a measure of the local homogeneity within the ROI.
4. 90 Percentile: This is a first-order feature that represents the number of voxels exceeding 90% of the voxel values in the set of all voxel values within the ROI.
5. Median: This first-order feature represents the median gray level intensity within the ROI.
6. Maximum 2D Diameter Slice: This shape feature is defined as the largest pairwise Euclidean distance between organ surface mesh vertices in the row-column plane (typically axial plane).

TABLE 2 The performance of the models built with radiomic features, clinical features, and combined features.

Methods	Methods	AUC (95% CI)	Sensitivity	Specificity	Accuracy	F1 score
A. Performance of the models on train dataset						
Clinical model	LR	0.918 (0.857–0.985)	0.85	0.795	0.827	0.85
	RF	0.871 (0.788–0.958)	0.778	0.788	0.782	0.805
	SVM	**0.96 (0.925–1)**	**0.933**	**0.812**	**0.882**	**0.901**
Radiomics model	LR	0.928 (0.87–0.993)	0.883	**0.894**	**0.888**	**0.901**
	RF	**0.945 (0.897–0.997)**	**0.894**	0.818	0.862	0.883
	SVM	0.861 (0.764–0.958)	0.833	0.749	0.798	0.827
Combined model	LR	0.955 (0.918–0.999)	0.861	**0.94**	**0.894**	0.904
	RF	**0.97 (0.94–1)**	**0.922**	0.879	0.904	**0.917**
	SVM	0.891 (0.801–0.981)	0.878	0.849	0.866	0.883
B. Performance of the models on test dataset						
Clinical model	LR	**0.881 (0.715–1)**	**0.822**	**0.757**	**0.795**	**0.819**
	RF	0.719 (0.441–0.967)	0.711	0.61	0.67	0.709
	SVM	0.876 (0.704–1)	0.8	0.752	0.782	0.807
Radiomics model	LR	**0.887 (0.73–1)**	**0.844**	**0.876**	**0.86**	**0.873**
	RF	0.794 (0.596–0.975)	0.822	0.638	0.745	0.779
	SVM	0.856 (0.679–1)	0.844	0.762	0.809	0.835
Combined model	LR	**0.922 (0.804–1)**	0.844	**0.886**	**0.862**	**0.871**
	RF	0.889 (0.748–1)	0.822	0.733	0.784	0.81
	SVM	0.902 (0.773–1)	**0.867**	0.767	0.823	0.85

The bold meaning is the best performance.

For the clinical model, diabetes is primarily characterized by hyperglycemia and metabolic disturbances. Its key clinical manifestations include increased appetite, excessive thirst, frequent urination, and unintended weight loss (i.e., more than three kilograms but less than one kilogram). In terms of serum biochemical markers, triglycerides, aspartate aminotransferase, endogenous creatinine clearance, low-density lipoprotein cholesterol, total bilirubin, and albumin are all intricately linked to human metabolism.

For the combined model, two additional features were listed as follows:

1. Informational Measure of Correlation (IMC) 2: IMC2 also evaluates the correlation between the probability distributions of variables i and j, similarly quantifying the complexity of the texture within the region of interest (ROI).
2. Busyness: This feature measures the change from a voxel to its neighbor. A high value for busyness indicates a "busy" image, with rapid changes in intensity between voxels and their neighborhood.

These additional features contribute to capturing more detailed information about texture complexity and local intensity variations in relation to diabetes detection.

From the selected features above, it is evident that some of them quantify the complexity of the texture of the pancreas. This suggests that there are rapid changes in intensity between voxels and their neighborhood in the medical images of patients with type 2 diabetes mellitus. These texture complexities may be indicative of underlying structural or compositional changes in the pancreas associated with diabetes. The identification and quantification of such changes can potentially provide valuable insights into disease progression and aid in the development of imaging biomarkers for diabetes diagnosis and monitoring.

Radiomics has indeed gained widespread utilization in clinical diagnosis. It has emerged as a valuable tool for auxiliary diagnosis by converting clinical images into mining data with high fidelity, repeatability, and minimal redundancy. This is achieved through the extraction of mathematical structural features from quantitative images. With the focus on personalized precision diagnosis and treatment, radiomics has played a pivotal role in advancing medical imaging beyond being just a diagnostic tool. It has become a fundamental instrument that provides specific and effective guidance for clinical diagnosis and treatment. By leveraging radiomic features, medical professionals can gain deeper insights into diseases, enabling them to make more accurate and personalized decisions regarding patient care (15). The integration of radiomics into clinical practice holds great promise for improving patient outcomes, optimizing treatment strategies, and facilitating precision medicine approaches.

Radiomics has been widely utilized in various diseases. van Griethuysen et al. (16) extracted radiomic features from 429 different lung lesions, including 48 texture features, 310 logarithmic features, and 158 wavelet features, to differentiate between benign and malignant nodules in the lungs. Their analysis revealed a correlation between image-based subtypes and the benign or malignant status of lung lesions. In another study, Grimm et al. (17) analyzed 275 preoperative breast MRI images of breast cancer patients and extracted a total of 56 imaging features encompassing morphology, texture, and dynamic characteristics for each patient's image. Utilizing multivariate analysis, they determined the correlation between these imaging

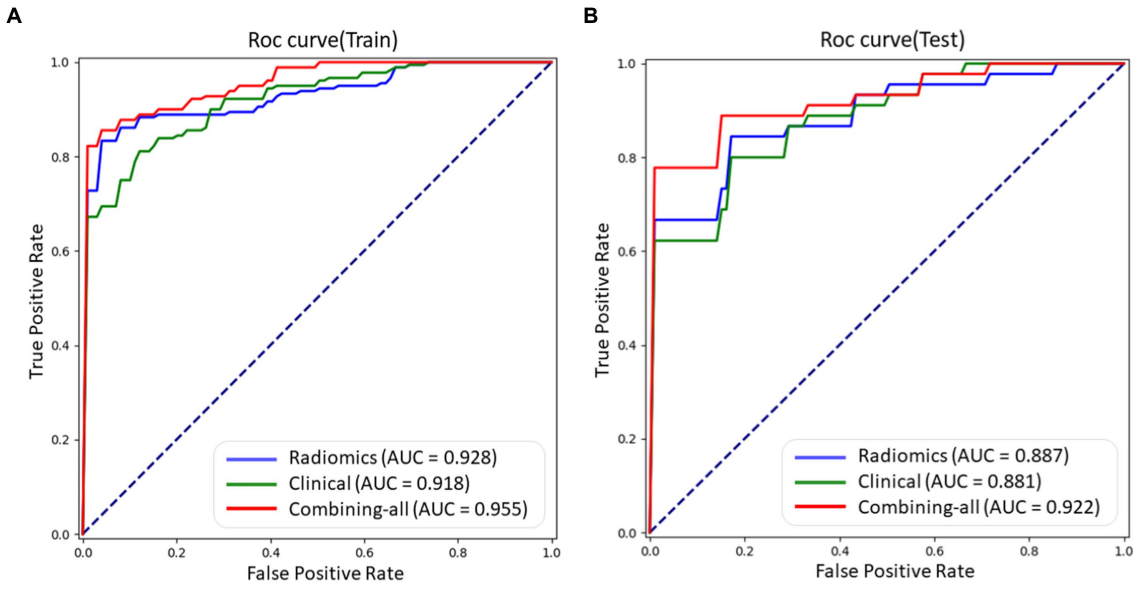


FIGURE 5
The ROC curves of different models using LR method, including radiomic features, clinical features, and combined features on train and test datasets, respectively. **(A)** ROC curve on train dataset. **(B)** ROC curve on test dataset.

features and molecular subtypes of breast cancer. The findings revealed a significant association between radiomic features derived from dynamic contrast-enhanced MRI and the molecular subtypes of luminal A and luminal B hormone receptor-positive breast cancers.

Kaissis et al. (18) retrospectively analyzed preoperative CT images of 207 patients diagnosed with pancreatic ductal adenocarcinoma. They developed a random forest machine learning algorithm to accurately predict the molecular subtype of pancreatic cancer based on radiomic features. The classification algorithm demonstrated high sensitivity (0.84 ± 0.05) and specificity (0.92 ± 0.01). Furthermore, the area under the receiver operating characteristic curve (AUC-ROC) was determined to be 0.93 ± 0.01, indicating that preoperative CT image radiomics analysis utilizing machine learning holds promise in predicting molecular subtypes closely associated with the survival outcomes of pancreatic cancer patients. Xue et al. (19) retrospectively analyzed CTA images and clinical data of 170 cases involving the head and radiomic features in conjunction with elevated levels of homocysteine and hypertension. Both the Rad-score model and joint model were established to investigate associations with symptomatic carotid plaque. Their findings demonstrated that hyper-homocysteinemia and hypertension exhibited independent associations with symptomatic carotid plaque. These studies highlight the potential of radiomics in various disease contexts, including lung lesions, breast cancer, pancreatic cancer, and carotid plaque. Radiomics analysis can provide valuable insights into disease characterization, subtype classification, and prediction of clinical outcomes.

In this study, a total of 2,264 radiomic features were extracted from pancreatic dual-energy CT images. A feature selection procedure similar to Lasso was employed to identify 7 optimal features. Simultaneously, 24 clinical features were digitized and normalized. Out of these, 8 representative clinical features were selected. The radiomic features were integrated with the clinical features, resulting in an optimal subset comprising of 8 calculated features determined by the Lasso algorithm. Subsequently, three models were constructed: one solely based on radiomic features, another solely based on clinical features, and the third model incorporating both types of features. The performance evaluation of these models was conducted using cross-validation. In the test set, the area under the curve (AUC) values were 0.887 (0.73–1), 0.881 (0.715–1), and 0.922 (0.804–1), respectively. When compared with previous studies (10), the sensitivity (0.844, 0.822, 0.844), specificity (0.876, 0.757, 0.886), accuracy (0.86, 0.795, 0.862), and F1 score (0.873, 0.819, 0.871) of this study showed significant improvement in performance.

Furthermore, we evaluated the performance of the combined model using three clinical features: average blood glucose, duration of diabetes, and diabetic complications. The feature "average blood glucose" was used for the diagnosis of diabetes patients, which is consistent with the ground truth. Regarding diabetes duration, all patients who were classified incorrectly by the combined model had a disease duration of less than 10 years. This interesting finding may be due to the selected feature not being able to capture the difference between healthy controls and diabetes patients with a duration of less than 10 years as shown in Figure 6. Additionally, when classifying patients with (25 patients) or without (20 patients) diabetic complications, 2 patients were classified into the complication group.

There are several limitations to our study. Firstly, the sample size was relatively small, and further expansion of the sample size is necessary to enhance the credibility of the results. Secondly, we only analyzed the imaging features of all patients using dual-energy pancreatic plain scans without considering differences in pancreatic imaging features during different enhanced phases. Additionally, we solely selected axial pancreatic images for analysis, potentially resulting in the omission of certain characteristic radiomics data.

5 Conclusion

After a comprehensive analysis, the three models constructed based on CT radiomic features demonstrate the potential of pancreatic dual-energy CT images as quantitative imaging biomarkers for detecting type 2 diabetes.

TABLE 3 The significance of radiomic features, clinical features, and combined features to build three models.

Models	Features	Coefficient
Radiomics model	**wavelet_glcm_wavelet-HLH-Imc1**	**0.125**
	boxsigmaimage_firstorder_Minimum	0.057
	wavelet_glcm_wavelet-HLH-Idn	0.052
	log_firstorder_log-sigma-4-0-mm-3D-90Percentile	0.046
	wavelet_firstorder_wavelet-LHH-Median	−0.046
	normalize_glcm_Imc1	**−0.056**
	original_shape_Maximum2DDiameterSlice	−0.077
Clinical model	Clinical manifestations	0.112
	Triglycerides	0.053
	Aspartate aminotransferase	0.034
	Endogenous creatinine clearance	0.034
	LDL cholesterol	0.018
	Total bilirubin	−0.027
	White/ball	−0.043
	*albumin	−0.132
Combined model	**wavelet_glcm_wavelet-HLH-Imc1**	0.024
	Clinical manifestations	0.006
	boxsigmaimage_firstorder_Minimum	0.002
	original_shape_Maximum2DDiameterSlice	−0.003
	binomialblurimage_ngtdm_Busyness	−0.007
	curvatureflow_ngtdm_Busyness	−0.011
	wavelet_glcm_wavelet-LLH-Imc2	−0.016
	*albumin (clinical feature)	−0.059

The specific calculation method of each feature can be found from Pyradiomics website (https://pyradiomics.readthedocs.io/en/latest/features.html).
The bold meaning is the best performance.

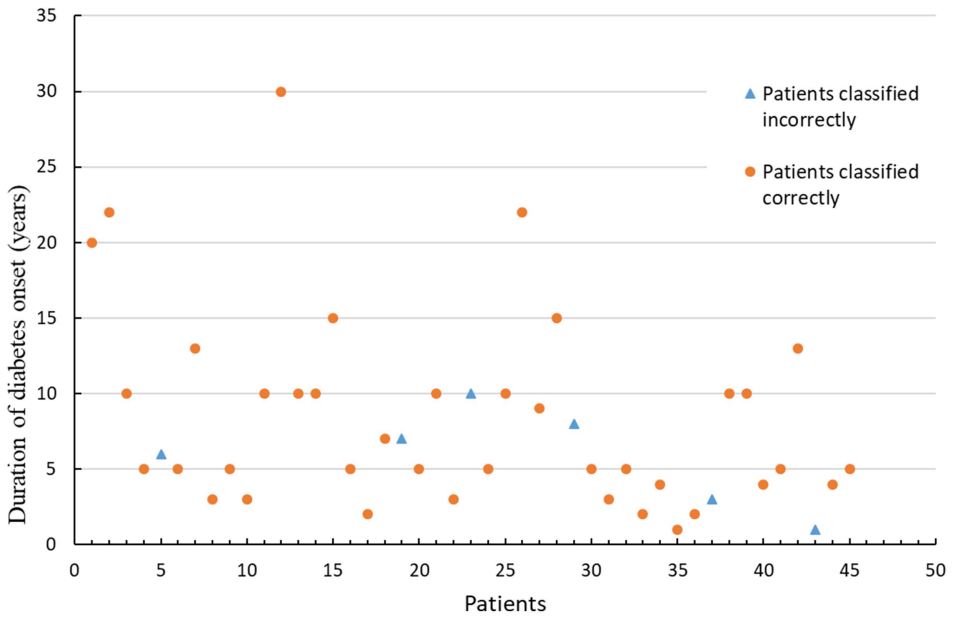


FIGURE 6
Statistical graph of the disease duration for 45 diabetes patients.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Author contributions

WJ: Writing – original draft, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Writing – review & editing. XP: Writing – original draft, Data curation, Formal analysis, Methodology, Software, Writing – review & editing. QL: Writing – original draft, Writing – review & editing. SH: Writing – original draft, Writing – review & editing. YuL: Writing – original draft, Writing – review & editing. XiZ: Writing – original draft, Writing – review & editing. XeZ: Writing – original draft, Writing – review & editing. WD: Writing – original draft, Writing – review & editing. YaL: Writing – original draft, Writing – review & editing. LC: Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

XP, XeZ, WD, and LC were employed by Shanghai United Imaging Intelligence Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2024.1328687/full#supplementary-material

References

1. Marshall SM. The pancreas in health and in diabetes. *Diabetologia*. (2020) 63:1962–5. doi: 10.1007/s00125-020-05235-z

2. Garcia TS, Rech TH, Leitão CB. Pancreatic size and fat content in diabetes: a systematic review and meta-analysis of imaging studies. *PLoS One*. (2017) 12:e0180911. doi: 10.1371/journal.pone.0180911

3. Yamazaki H, Tauchi S, Wang J, Dohke M, Hanawa N, Kodama Y, et al. Longitudinal association of fatty pancreas with the incidence of type-2 diabetes in lean individuals: a 6-year computed tomography-based cohort study. *J Gastroenterol*. (2020) 55:712–21. doi: 10.1007/s00535-020-01683-x

4. Tallam H, Elton DC, Lee S, Wakim P, Pickhardt PJ, Summers RM. Fully automated abdominal CT biomarkers for type 2 diabetes using deep learning. *Radiology*. (2022) 304:85–95. doi: 10.1148/radiol.211914

5. Sofue K, Itoh T, Takahashi S, Schmidt B, Shimada R, Negi N, et al. Quantification of cisplatin using a modified 3-material decomposition algorithm at third-generation dual-source dual-energy computed tomography: an experimental study. *Investig Radiol*. (2018) 53:673–80. doi: 10.1097/RLI.0000000000000491

6. Zhou Y, Zhang J, Chen J, Yang C, Gong C, Li C, et al. Prediction using T2-weighted magnetic resonance imaging-based radiomics of residual uterine myoma regrowth after high-intensity focused ultrasound ablation. *Ultrasound Obstet Gynecol*. (2022) 60:681–92. doi: 10.1002/uog.26053

7. An C, Kim DW, Park YN, Chung YE, Rhee H, Kim MJ. Single hepatocellular carcinoma: preoperative MR imaging to predict early recurrence after curative resection. *Radiology*. (2015) 276:433–43. doi: 10.1148/radiol.15142394

8. Martin SS, Trapp F, Wichmann JL, Albrecht MH, Lenga L, Durden J, et al. Dual-energy CT in early acute pancreatitis: improved detection using iodine quantification. *Eur Radiol*. (2019) 29:2226–32. doi: 10.1007/s00330-018-5844-x

9. Wright DE, Mukherjee S, Patra A, Khasawneh H, Korfiatis P, Suman G, et al. Radiomics-based machine learning (ML) classifier for detection of type 2 diabetes on standard-of-care abdomen CTs: a proof-of-concept study. *Abdom Radiol*. (2022) 47:3806–16. doi: 10.1007/s00261-022-03668-1

10. Huang S, Liang Y, Zhong X, Luo Q, Yao X, Nong Z, et al. Pancreatic fat fraction in dual-energy computed tomography as a potential quantitative parameter in the detection of type 2 diabetes mellitus. *Eur J Radiol*. (2023) 159:110668. doi: 10.1016/j.ejrad.2022.110668

11. Shi F, Hu W, Wu J, Han M, Wang J, Zhang W, et al. Deep learning empowered volume delineation of whole-body organs-at-risk for accelerated radiotherapy. *Nat Commun*. (2022) 13:6566. doi: 10.1038/s41467-022-34257-x

12. Wu J, Xia Y, Wang X, Wei Y, Liu A, Innanje A, et al. uRP: an integrated research platform for one-stop analysis of medical images. *Front Radiol*. (2023) 3:1153784. doi: 10.3389/fradi.2023.1153784

13. Zheng BH, Liu LZ, Zhang ZZ, Shi JY, Dong LQ, Tian LY, et al. Radiomics score: a potential prognostic imaging feature for postoperative survival of solitary HCC patients. *BMC Cancer*. (2018) 18:1–12. doi: 10.1186/s12885-018-5024-z

14. Harrell FE Jr, Lee KL, Califf RM, Pryor DB, Rosati RA. Regression modelling strategies for improved prognostic prediction. *Stat Med*. (1984) 3:143–52. doi: 10.1002/sim.4780030207

15. Feng XY. Precision medicine, medical imaging first. *Natl Med J China*. (2016) 1:1–2. doi: 10.3760/cma.j.issn.1005-1201.2016.01.001

16. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computationalradiomics system to decode the radiographic pheno type. *Cancer Res*. (2017) 77:e104–7. doi: 10.1158/0008-5472.CAN-17-0339

17. Grimm LJ, Zhang J, Mazurowski MA. Computational approach to radiogenomics of breast cancer: luminal A and luminal B molecular subtypes are associated with imaging features on routine breast MRI extracted using computer vision algorithms. *J Magn Reson Imaging*. (2015) 42:902–7. doi: 10.1002/jmri.24879

18. Kaissis GA, Ziegelmayer S, Lohofer FK, Harder FN, Jungmann F, Sasse D, et al. Image-based molecular phenotyping of pancreatic ductal adenocarcinoma. *J Clin Med*. (2020) 9:724. doi: 10.3390/jcm9030724

19. Xue LL, Wang LJ, Shi CY, Zhang Q, Qiao Y, Zhang H. Radiomics analysis of CTA for identification of symptomatic carotid artery plaques. *Chin J Integr Med Cardio/Cerebrovasc Dis*. (2023) 21:2083–8. doi: 10.12102/j.issn.1672-1349.2023.11.032

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 20 May 2024
DOI 10.3389/fmed.2024.1409477

Check for updates

OPEN ACCESS

EDITED BY
Minjeong Kim,
University of North Carolina at Greensboro, United States

REVIEWED BY
Xiaoming Liu,
Jilin University, China
Peng Zhang,
Shanxi University, China

*CORRESPONDENCE
Xiao-Ping Yin
✉ yinxiaoping78@sina.com

†These authors have contributed equally to this work

RECEIVED 30 March 2024
ACCEPTED 30 April 2024
PUBLISHED 20 May 2024

CITATION
Meng H, Wang T-D, Zhuo L-Y, Hao J-W, Sui L-y, Yang W, Zang L-L, Cui J-J, Wang J-N and Yin X-P (2024) Quantitative radiomics analysis of imaging features in adults and children Mycoplasma pneumonia. *Front. Med.* 11:1409477. doi: 10.3389/fmed.2024.1409477

COPYRIGHT
© 2024 Meng, Wang, Zhuo, Hao, Sui, Yang, Zang, Cui, Wang and Yin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Quantitative radiomics analysis of imaging features in adults and children Mycoplasma pneumonia

Huan Meng[1,2,3†], Tian-Da Wang[1,2,3†], Li-Yong Zhuo[1,2,3], Jia-Wei Hao[1,2,3], Lian-yu Sui[1,2,3], Wei Yang[4], Li-Li Zang[5], Jing-Jing Cui[6], Jia-Ning Wang[1,2,3] and Xiao-Ping Yin[1,2,3*]

[1]Clinical Medicine School of Hebei University, Baoding, China, [2]Department of Radiology, Affiliated Hospital of Hebei University, Baoding, China, [3]Hebei Key Laboratory of Precise Imaging of Inflammation Related Tumors, Baoding, China, [4]Department of Radiology, Baoding First Central Hospital, Baoding, China, [5]Department of Radiology, Baoding Children's Hospital, Baoding, China, [6]Department of Research and Development, United Imaging Intelligence (Beijing) Co., Beijing, China

Purpose: This study aims to explore the value of clinical features, CT imaging signs, and radiomics features in differentiating between adults and children with Mycoplasma pneumonia and seeking quantitative radiomic representations of CT imaging signs.

Materials and methods: In a retrospective analysis of 981 cases of mycoplasmal pneumonia patients from November 2021 to December 2023, 590 internal data (adults:450, children: 140) randomly divided into a training set and a validation set with an 8:2 ratio and 391 external test data (adults:121; children:270) were included. Using univariate analysis, CT imaging signs and clinical features with significant differences ($p < 0.05$) were selected. After segmenting the lesion area on the CT image as the region of interest, 1,904 radiomic features were extracted. Then, Pearson correlation analysis (PCC) and the least absolute shrinkage and selection operator (LASSO) were used to select the radiomic features. Based on the selected features, multivariable logistic regression analysis was used to establish the clinical model, CT image model, radiomic model, and combined model. The predictive performance of each model was evaluated using ROC curves, AUC, sensitivity, specificity, accuracy, and precision. The AUC between each model was compared using the Delong test. Importantly, the radiomics features and quantitative and qualitative CT image features were analyzed using Pearson correlation analysis and analysis of variance, respectively.

Results: For the individual model, the radiomics model, which was built using 45 selected features, achieved the highest AUCs in the training set, validation set, and external test set, which were 0.995 (0.992, 0.998), 0.952 (0.921, 0.978), and 0.969 (0.953, 0.982), respectively. In all models, the combined model achieved the highest AUCs, which were 0.996 (0.993, 0.998), 0.972 (0.942, 0.995), and 0.986 (0.976, 0.993) in the training set, validation set, and test set, respectively. In addition, we selected 11 radiomics features and CT image features with a correlation coefficient r greater than 0.35.

Conclusion: The combined model has good diagnostic performance for differentiating between adults and children with mycoplasmal pneumonia, and different CT imaging signs are quantitatively represented by radiomics.

KEYWORDS

mycoplasma pneumonia, radiomics, adults, children, computed tomography (CT)

1 Introduction

Mycoplasma pneumonia (MP) accounts for 10%–30% of community-acquired pneumonia (CAP) and often occurs in autumn, especially in children and adolescents (1). In recent years, the adult incidence rate has also increased. This disease spreads approximately every 3–7 years (2). During the epidemic period, this microbe can cause up to 20%–40% of CAP cases in the general population and even up to 70% in closed populations (3). The diagnosis of mycoplasmal pneumonia depends on the detection of specific antibodies. Due to its often negative early diagnosis, computed tomography (CT) imaging plays an important guiding role in the early diagnosis and treatment of mycoplasmal pneumonia. Previous studies have shown that children tend to present with large patchy consolidation on CT imaging compared to adults (4), but the discovery of this difference often depends on the reading habits and clinical experience of the reader. Moreover, as the disease progresses, the imaging manifestations at different stages of the same disease are not always the same, and there are often overlapping manifestations. In recent years, the term radiomics has received increasing attention (5). Radiomics has been successfully applied in the identification, staging, and evaluation of lung cancer (6, 7). However, radiomics methods are relatively less applied in the prediction and diagnosis of non-tumor diseases of the lung. Yanling et al. (8) applied radiomics nomograms to identify pneumonia and acute paraquat lung injury. Xie et al. (9) applied CT radiomics to conduct a comparative analysis of ground-glass density shadows in COVID-19 and non-COVID-19 and proposed that the CT radiomics model can help to differentiate between COVID-19 and non-COVID-19 ground-glass density shadows. At the same time, Honglin Li et al. (10) confirmed that radiomics-clinical nomograms have good discriminative effects on mycoplasmal pneumonia and bacterial pneumonia, which is helpful for clinical decision-making. In addition, radiomics also plays an important role in grading severity (11) and prognostic evaluation (12) of pneumonia.

The above results provide confidence and reference for our research. Considering that there is no research on differentiating the radiomic features of adult and child mycoplasmal pneumonia in domestic and foreign studies, this article will analyze and compare the clinical features, CT imaging signs, and radiomic features of adult and child patients and conduct external validation. It will provide a quantitative representation of different CT imaging signs using radiomics, thus providing evidence for early clinical diagnosis and precise treatment.

2 Materials and methods

2.1 Study population

In a retrospective analysis of clinical and imaging data of patients diagnosed with MP in two hospitals from November 2021 to December 2023, 590 patients (450 adults and 140 children) with internal data, which were divided into a training set and a validation set according to an 8:2 ratio, and 391 patients (121 adults and 270 children) with external data, which were used as an external test set, were included. Based on age, patients were divided into adult group (>14 years old) and child group (≤14 years old). The inclusion criteria were as follows: (1) patients with mycoplasmal pneumonia confirmed by throat swab or fiberoptic bronchoscopy with alveolar lavage nucleic acid testing; and (2) patients with clear lesions detected by chest CT. Exclusion criteria were as follows: (1) poor image quality; and (2) previous bronchial asthma, chronic obstructive pulmonary disease, recurrent respiratory tract infections, severe pneumonia without a history of cure, congenital or secondary immune suppression or immune deficiency, and connective tissue disease (Figure 1). This study was approved by the ethics committee of the Affiliated Hospital of Hebei University, and because this is a retrospective study, written informed consent is waived. This study was conducted in accordance with the principles of the Helsinki Declaration.

2.2 CT image acquisition

Philips Brilliance 256-row, GE Discovery HD750 CT, and United Imaging uCT550 spiral CT scanner were used. The patient was placed in a supine position with both hands raised above his head. The scanning range was from the thoracic inlet to the level of the diaphragm, and deep breath-holding scanning was performed after deep inspiration. Scanning parameters: tube voltage 120 kV, tube current automatic milliamp technology, pitch 0.900; 0.984; 1.175, Rotation time 0.5; 0.6; 0.6 s, matrix 512 × 512, layer thickness 5 mm, interlayer spacing 5 mm, and field of view 40 cm × 40 cm. Axial reconstruction of lung window (window width 1500HU, window level −600 HU) and mediastinal window (window width 350HU, window level 40HU).

2.3 CT image analysis

The CT images were independently reviewed by two physicians mainly engaged in chest imaging diagnosis. In case of disagreement, the two physicians reached a consensus through consultation. The CT characteristics of each patient were recorded, including consolidation pattern, consolidation with ground-glass opacity (GGO), bronchial wall thickening, air bronchogram, atelectasis, interlobular septal thickening, number of involved lung lobes, mediastinal enlargement of lymph nodes, pleural effusion, and other imaging features, as well as quantitative characteristics such as mean lesion density, lesion volume, and CTLP.

2.4 Radiomics feature extraction, feature selection, and machine learning models building

Before radiomics feature extraction, the images were normalized by subtracting the window level (WL: 40) and dividing by the window width (WW: 300). The auto-segmentation, radiomics feature extraction, feature selection, and machine learning models building were established on the uAI Research Portal V1.1 (Shanghai United Imaging Intelligence, Co., Ltd.) (13–16). The radiomics features were automatically extracted from ROIs using an open-source Python package, Pyradiomics V3.0 (17). The PCC, LASSO, LR, and other methods used the package of Scikit-learn (18). All analyses were implemented in Python (Python Software Foundation, http://python.org). Two physicians modified the ROI of the automatically segmented lesions layer-by-layer to avoid non-lesion areas such as blood vessels and ribs, confirmed and submitted it, and obtained the volume of interest (VOI) of the lesion (Figure 2). The features were divided into seven groups, and the shape features were extracted from the original

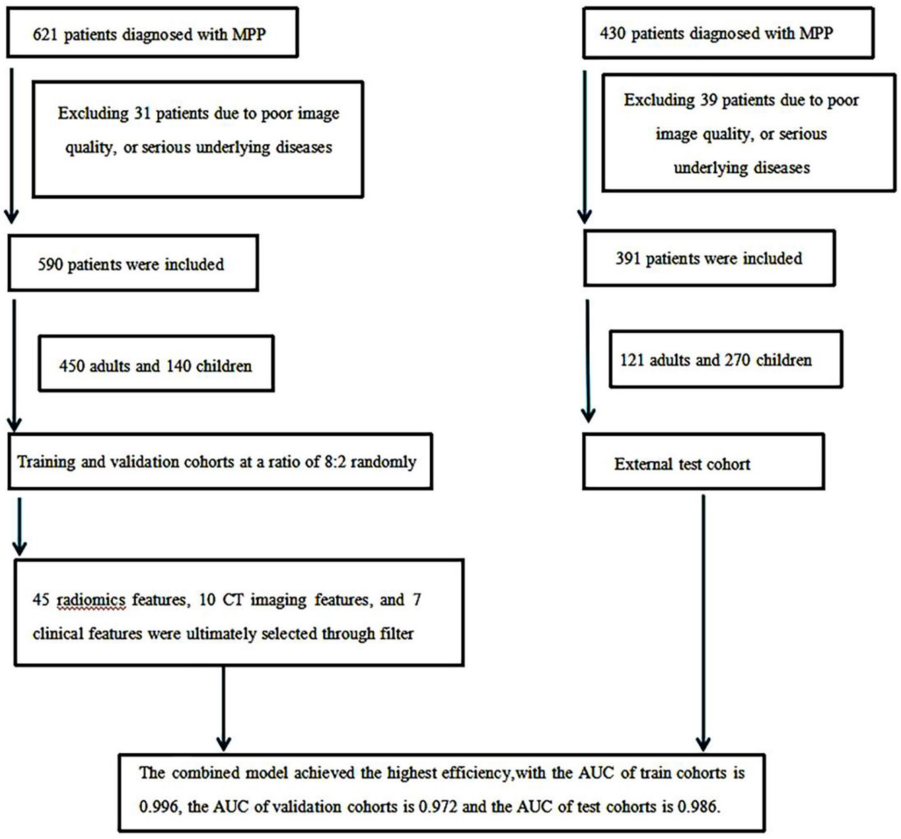


FIGURE 1
Research flowchart.

image based on the ROI. The texture features and grayscale statistics features were extracted from the original image and 15 filtered images, with a total of 1,904 features extracted, which were:

1. Shape feature: 14;
2. Grayscale statistics feature: 450;
3. Gray Level Cooccurence Matrix, GLCM: 525;
4. Gray Level Run Length Matrix, GLRLM: 350;
5. Gray Level Size Zone Matrix, GLSZM: 400;
6. Neighboring Gray Tone Difference Matrix, NGTDM: 400;
7. Gray Level Dependence Matrix, GLDM: 125.

Using univariate analysis to select CT imaging signs and clinical features with significant differences ($p < 0.05$), we constructed the clinical model. Z-score was used to normalize radiomics features before feature selection and model construction. Pearson correlation coefficient (PCC) and least absolute shrinkage and selection operator (LASSO) were used to screen and reduce the dimensionality of radiomic features, and RadScore was calculated by weighting the features based on the coefficients obtained by LASSO. In addition, multivariable logistic regression analysis is used to construct radiomic models based on the features selected. Finally, the combined model was constructed using Radscore, CT imaging signs, and clinical features selected.

Importantly, the radiomics features and quantitative and qualitative CT imaging signs were analyzed using Pearson correlation analysis and analysis of variance, respectively.

2.5 Statistical analysis

All data were analyzed using SPSS 26.0. For quantitative data, independent sample t-tests (when normal distribution) or Mann–Whitney U-tests (when non-normal distribution) were performed. For count data, $\chi 2$ tests were performed. Logistic regression analysis was performed on the clinical features, CT imaging signs, and radiomics features that showed statistical differences between the groups. Single-phase models and combined models were established, and the predictive performance of each model was evaluated using AUC, sensitivity, specificity, and accuracy. The Delong test was used to compare the AUCs between the models. Pearson correlation analysis and variance analysis were used to analyze quantitative and qualitative CT imaging signs and radiomics features, respectively, to find the quantitative radiomics of CT imaging signs.

3 Results

3.1 Clinical features

Statistical analysis was conducted on the clinical data of the training set, validation set, and test set. There were significant differences in the type of fever, LC, CK-MB, LDH, D-dimer, and CRP between adult and child groups with mycoplasmal pneumonia ($p < 0.05$), but there was no significant difference in PLT. The

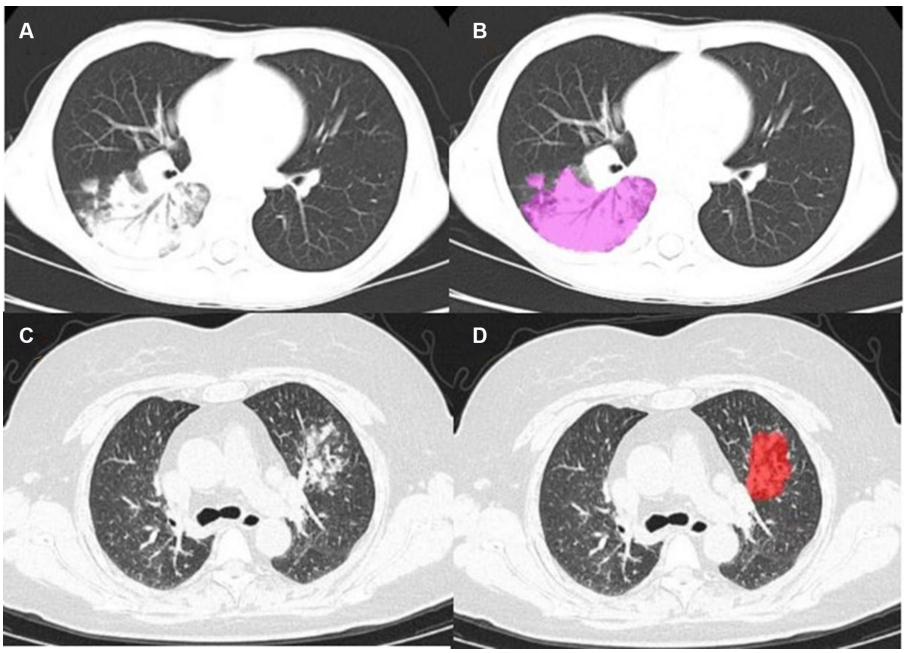


FIGURE 2
(A) Mycoplasmal pneumonia in a child (male, 9 years old), mainly manifested as large patchy consolidation, with air bronchogram sign visible; **(C)** Mycoplasmal pneumonia in an adult (female, 57 years old), characterized by focal and small patchy consolidation; **(B–D)** Lesion annotation.

proportion of severe cases in the training set was 30.8% in the adult group and 36.6% in the child group; in the validation set, it was 32.2% in the adult group and 35.7% in the child group; in the test set, it was 21.5% in the adult group and 36.7% in the child group, with statistically significant differences ($p < 0.05$). The details are shown in Table 1.

3.2 CT imaging signs

Statistical analysis was conducted on the CT image features of the training set, validation set, and test set. Segmental and Wedge-shaped consolidation showed significant differences between the adult group and the child group, with Segmental and Wedge-shaped consolidation in the child group, with statistical significance ($p < 0.05$; Figure 2), consolidation mixed GGO and air bronchogram signs were significantly different in children, with statistical significance ($p < 0.05$). In addition, there were statistically significant differences between adults and children in interlobular septal thickening, number of lobes involved, mean lesion density, and CTLP ($p < 0.05$), while there was no statistical difference in bronchial wall thickening ($p > 0.05$). For details, see Table 1 and Figure 2.

3.3 Models construction

The seven most clinically relevant features extracted from the patient's clinical characteristics are type of fever, LC, CRP, PLT, CK-MB, LDH, and D-dimer ($p < 0.05$); and 10 CT imaging signs, are consolidation pattern, consolidation mixed GGO, bronchial wall thickening, air bronchogram sign, interlobular septal thickening, number of lobes involved, pleural effusion, mediastinal enlargement of lymph nodes, mean lesion density, and CTLP, with significant differences ($p < 0.05$). Based on these features, we constructed the clinical model and the CT imaging model. For the radiomics analysis, 45 features with the highest correlation were obtained after PCC and LASSO, and Figure 3 shows the top 20 features with a correlation coefficient greater than 0.02 in the LASSO. Based on this, the radiomic model was constructed. In addition, we build the combined model using the clinical features, CT imaging signs, and the radiomics selected.

For the three models, the AUC for the testing set were 0.893(0.863,0.921), 0.744(0.698,0.783), and 0.969(0.953,0.982), the AUC for training set and validation set is shown in Table 2, and the ROC curve and prediction performance results were plotted (Table 2; Figure 4). The results showed that the combined model showed higher predictive performance in distinguishing adult and child Mycoplasma pneumonia than any single model. According to the Delong test, there was a statistical difference ($p < 0.05$) in the AUC between the CT imaging model, radiomics model, and combination model in the external test set (Table 3).

3.4 Correlation analysis between CT imaging signs and radiomics features

Pearson correlation analysis evaluated the correlation between CT features and radiomics features; the correlation map is shown in Figure 5, and the case presentation is shown in Figure 6. Those with a correlation coefficient r greater than 0.35 were included in the charts (Table 4). For the quantitative and qualitative CT images, we visualized the data distribution using box plots and correlation plots, respectively. Mean_lesion_density, Consolidation_pattern, Air_bronchogram_sign, and Interlobular_septal_thickening demonstrated a high correlation with texture features.

TABLE 1 General information of adult and child patients with mycoplasmal pneumonia.

	Train *N* = 472			Validation *N* = 118			Test *N* = 391			Overall *N* = 981
Characteristic	0	1	*p*-value[1]	0	1	*p*-value[1]	0	1	*p*-value[1]	*p*-value[2]
	N = 360(76%)	*N* = 112(24%)		*N* = 90(76%)	*N* = 28(24%)		*N* = 121(31%)	*N* = 270(69%)		
Age	58 (17)	7 (3)	<0.001	57 (16)	7 (4)	<0.001	46 (19)	7 (3)	<0.001	<0.001
Gender			0.48			0.88			0.043	0.033
Female	200 (55.6%)	58 (51.8%)		40 (44.4%)	12 (42.9%)		48 (39.7%)	137 (50.7%)		
Male	160 (44.4%)	54 (48.2%)		50 (55.6%)	16 (57.1%)		73 (60.3%)	133 (49.3%)		
Type_of_fever			<0.001			<0.001			<0.001	<0.001
None	276 (76.7%)	19 (17.0%)		69 (76.7%)	8 (28.6%)		26 (21.5%)	4 (1.5%)		
Grade 1(37.1–38°C)	24 (6.7%)	6 (5.4%)		6 (6.7%)	0 (0.0%)		24 (19.8%)	8 (3.0%)		
Grade 2(38.1–39°C)	35 (9.7%)	38 (33.9%)		11 (12.2%)	8 (28.6%)		50 (41.3%)	103 (38.1%)		
Grade 3(39.1–41°C)	25 (6.9%)	48 (42.9%)		4 (4.4%)	12 (42.9%)		21 (17.4%)	153 (56.7%)		
Grade 3(>41°C)	0 (0.0%)	1 (0.9%)					0 (0.0%)	2 (0.7%)		
LC	1.56 (1.14,2.03)	2.16 (1.61,2.88)	<0.001	1.51 (1.14,2.06)	2.43 (1.63,4.05)	<0.001	1.50 (1.10,2.01)	2.12 (1.61,2.70)	<0.001	<0.001
PLT	247 (192,312)	298 (236,383)	<0.001	245 (193,329)	300 (255,368)	0.001	256 (195,309)	284 (228,354)	<0.001	0.031
CK-MB	0.70 (0.40,1.10)	1.10 (0.60,2.20)	<0.001	0.60 (0.39,0.91)	1.51 (0.60,2.22)	0.001	12.8 (10.7,17.3)	2.3 (1.9,2.9)	<0.001	<0.001
LDH	191 (161,231)	261 (222,323)	<0.001	188 (155,227)	277 (206,337)	<0.001	173 (127,211)	284 (242,336)	<0.001	<0.001
D-dimer	178 (137,343)	137 (1,235)	<0.001	162 (137,298)	137 (104,234)	0.064	1.31 (0.42,9.77)	0.23 (0.15,0.38)	<0.001	<0.001
CRP	6 (2,30)	6 (1,19)	0.13	6 (2,41)	7 (2,18)	0.43	19 (6,48)	7 (2,16)	<0.001	0.024
Severe			0.25			0.73			0.003	0.98
No	249(69.2%)	71(63.4%)		61(67.8%)	18(64.3%)		95(78.5%)	171(63.3%)		
Yes	111(30.8%)	41(36.6%)		29(32.2%)	10(35.7%)		26(21.5%)	99(36.7%)		
Consolidation_pattern			0.001			0.86			<0.001	<0.001
None	134 (37.2%)	26 (23.2%)		33 (36.7%)	9 (32.1%)		28 (23.1%)	41 (15.2%)		
Patchy	77 (21.4%)	16 (14.3%)		16 (17.8%)	7 (25.0%)		42 (34.7%)	83 (30.7%)		
Segmental	77 (21.4%)	38 (33.9%)		27 (30.0%)	8 (28.6%)		40 (33.1%)	63 (23.3%)		
Wedge-shaped	72 (20.0%)	32 (28.6%)		14 (15.6%)	4 (14.3%)		11 (9.1%)	83 (30.7%)		
Consolidation_mixed_GGO			<0.001			0.026			0.003	<0.001
No	253 (70.3%)	58 (51.8%)		68 (75.6%)	15 (53.6%)		70 (57.9%)	113 (41.9%)		
Yes	107 (29.7%)	54 (48.2%)		22 (24.4%)	13 (46.4%)		51 (42.1%)	157 (58.1%)		
Pleural_effusion			<0.001			0.24			0.005	

(Continued)

10.3389/fmed.2024.1409477

TABLE 1 (Continued)

Characteristic	Train *N* = 472			Validation *N* = 118			Test *N* = 391			Overall *N* = 981
	0	1	*p*-value[1]	0	1	*p*-value[1]	0	1	*p*-value[1]	*p*-value[2]
	N = 360(76%)	*N* = 112(24%)		*N* = 90(76%)	*N* = 28(24%)		*N* = 121(31%)	*N* = 270(69%)		
None	276 (76.7%)	107 (95.5%)		71 (78.9%)	27 (96.4%)		101 (83.5%)	253 (93.7%)		
Minor	58 (16.1%)	4 (3.6%)		8 (8.9%)	1 (3.6%)		15 (12.4%)	11 (4.1%)		
Moderate	14 (3.9%)	1 (0.9%)		8 (8.9%)	0 (0.0%)		4 (3.3%)	3 (1.1%)		
Massive	12 (3.3%)	0 (0.0%)		3 (3.3%)	0 (0.0%)		1 (0.8%)	3 (1.1%)		
Mediastinal_enlargement_of_lymph_nodes			0.001			0.020			0.002	0.004
No	295 (81.9%)	106 (94.6%)		75 (83.3%)	28 (100.0%)		104 (86.0%)	257 (95.2%)		
Yes	65 (18.1%)	6 (5.4%)		15 (16.7%)	0 (0.0%)		17 (14.0%)	13 (4.8%)		
Air_bronchogram_sign			0.003			0.25			<0.001	<0.001
No	228 (63.3%)	53 (47.3%)		56 (62.2%)	14 (50.0%)		63 (52.1%)	89 (33.0%)		
Yes	132 (36.7%)	59 (52.7%)		34 (37.8%)	14 (50.0%)		58 (47.9%)	181 (67.0%)		
bronchial_wall_thickening			0.010			0.29			0.93	0.28
No	247 (68.6%)	91 (81.3%)		70 (77.8%)	19 (67.9%)		83 (68.6%)	184 (68.1%)		
Yes	113 (31.4%)	21 (18.8%)		20 (22.2%)	9 (32.1%)		38 (31.4%)	86 (31.9%)		
Interlobular_septal_thickening			<0.001			0.029			0.019	<0.001
No	244 (67.8%)	102 (91.1%)		66 (73.3%)	26 (92.9%)		94 (77.7%)	235 (87.0%)		
Yes	116 (32.2%)	10 (8.9%)		24 (26.7%)	2 (7.1%)		27 (22.3%)	35 (13.0%)		
Number_of_lobes_involved						0.40			<0.001	
1	133 (36.9%)	60 (53.6%)		28 (31.1%)	14 (50.0%)		29 (24.0%)	97 (35.9%)		
2	37 (10.3%)	25 (22.3%)		12 (13.3%)	4 (14.3%)		22 (18.2%)	65 (24.1%)		
3	50 (13.9%)	6 (5.4%)		10 (11.1%)	2 (7.1%)		16 (13.2%)	52 (19.3%)		
4	42 (11.7%)	7 (6.3%)		9 (10.0%)	2 (7.1%)		17 (14.0%)	20 (7.4%)		
5	93 (25.8%)	8 (7.1%)		27 (30.0%)	4 (14.3%)		37 (30.6%)	36 (13.3%)		
0	5 (1.4%)	6 (5.4%)		4 (4.4%)	2 (7.1%)					
Mean_lesion_density	−487 (155)	−432 (180)	0.005	−506 (140)	−475 (137)	0.33	−488 (141)	−400 (207)	<0.001	<0.001
CTLP	0.05 (0.08)	0.07 (0.12)	0.050	0.09 (0.15)	0.04 (0.05)	0.28	0.06 (0.11)	0.09 (0.10)	<0.001	<0.001

[1]Pearson's Chi-squared test; Wilcoxon rank sum test; Fisher's exact test; [2]Pearson's Chi-squared test; Kruskal–Wallis rank sum test. 0, adult; 1, child.

10.3389/fmed.2024.1409477

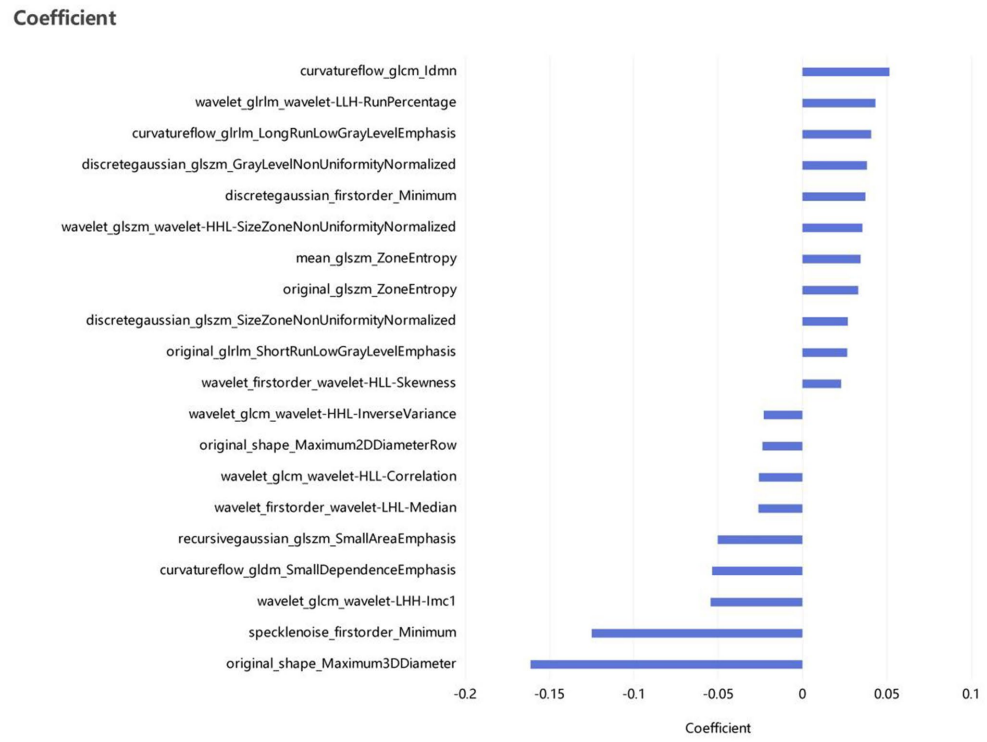


FIGURE 3
The correlation ranking of the top 20 radiomic features in the training set.

TABLE 2 Predictive ability of four models for distinguishing adult and childhood mycoplasmal pneumonia.

Model	Cohort	AUC(95% CI)	Sensitivity	Specificity	Accuracy	Precision	threshold_ train
Clinical model	Training	0.915(0.886,0.941)	0.865	0.839	0.879	0.827	0.315
	Validation	0.889(0.802,0.956)	0.825	0.750	0.864	0.810	0.315
	Testing	0.893(0.863,0.921)	0.744	0.537	0.665	0.720	0.315
CT image model	Training	0.831(0.794,0.863)	0.748	0.768	0.737	0.689	0.291
	Validation	0.736(0.660,0.825)	0.714	0.750	0.695	0.659	0.291
	Testing	0.744(0.698,0.783)	0.689	0.511	0.621	0.670	0.291
Radiomics model	Training	0.995(0.992,0.998)	0.965	0.955	0.970	0.954	0.354
	Validation	0.952(0.921,0.978)	0.896	0.893	0.898	0.850	0.354
	Testing	0.969(0.953,0.982)	0.764	0.544	0.680	0.739	0.354
Combined model	Training	0.996(0.993,0.998)	0.970	0.982	0.964	0.937	0.270
	Validation	0.972(0.942,0.995)	0.890	0.857	0.907	0.864	0.270
	Testing	0.986(0.976,0.993)	0.824	0.656	0.760	0.779	0.270

4 Discussion

In this study, we established clinical models, CT imaging models, radiomics models, and combined models and confirmed their effectiveness in differentiating adult and children mycoplasmal pneumonia. For the individual model, the radiomics model achieved the highest AUC. In addition, the radiomics features were well correlated with CT imaging signs, which could quantitatively represent different CT imaging signs to a certain extent.

Through the analysis of the CT imaging signs of the two groups of patients, it was found that there was no or patchy consolidation in the adult group and segmental or wedge-shaped consolidation in the child group, indicating that the condition of adult Mycoplasma pneumonia was mild and slow, and children had the characteristics

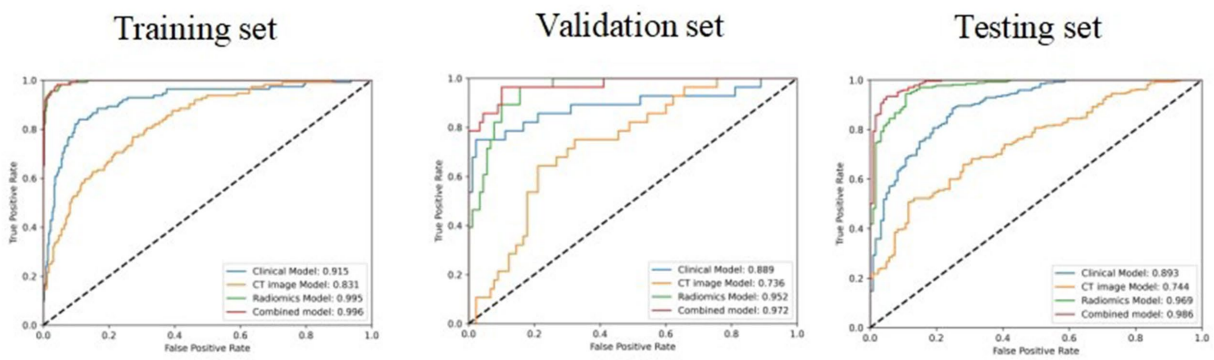


FIGURE 4
The ROC curves of the four models in the training set, validation set, and test set.

TABLE 3 Comparison of AUC between the three individual models and the combined model on the test set.

Group	model_ name_1	model_ name_2	auc_1	auc_cov_1	auc_2	auc_cov_2	*P*-value
Test	Clinical model	CT image model	0.893	0.000307134	0.744	0.000658474	$P > 0.05$
Test	Clinical model	Radiomics model	0.893	0.000307134	0.969	7.25E-05	$P > 0.05$
Test	Clinical model	Combined model	0.893	0.000307134	0.986	2.79E-05	$P > 0.05$
Test	CT image model	Radiomics Model	0.744	0.000658474	0.969	7.25E-05	$P < 0.05$
Test	CT image model	Combined model	0.744	0.000658474	0.986	2.79E-05	$P < 0.05$
Test	Radiomics model	Combined model	0.969	7.25E-05	0.986	2.79E-05	$P < 0.05$

of rapid progress, serious disease, and high incidence of complications, which was consistent with previous studies (4). The reason for this analysis is that mycoplasma, as the smallest microorganism between bacteria and viruses, can induce cellular and humoral immune responses after infection. Due to the immature and incomplete development of the lungs in children, the number of pulmonary alveoli is relatively small compared to adults, and the immune system is relatively incomplete. The elastic fibers of the bronchial tube are not strong. After mycoplasma infection, the disease progresses faster, the function of defending inflammation is weaker, and the inflammatory manifestations are more obvious than those in adults. If it invades the bronchioles and interstitial lung tissue near the lung field, it will cause congestion, edema, infiltration, and exudation of inflammatory cells, and the exudate will stimulate the pleura, causing pleural reactive effusion, leading to pleural effusion (4). Based on the different imaging manifestations and progression of adult and children mycoplasmal pneumonia, once mycoplasmal pneumonia is diagnosed, especially in children, active treatment should be taken to prevent complications or the possibility of progression to severe disease. In addition, after feature selection, a CT imaging model was established, and a ROC curve was drawn. The internal training set AUC value was 0.831 (0.794, 0.863), the validation set AUC value was 0.736 (0.660, 0.825), and the external test set AUC value was 0.744 (0.698, 0.783). It has good discriminative power, indicating that typical CT imaging signs are important in distinguishing between adult and pediatric mycoplasmal pneumonia. At the same time, Dongdong Wang et al. (19) used radiomics to analyze the diagnostic value of distinguishing between mycoplasmal pneumonia (MPP) and *streptococcus pneumoniae* pneumonia (SPP) in children under 5 years old and divided them into a testing set and a validation set at a ratio of 7:3. In the validation cohort, the consolidation + surrounding halo sign was used to distinguish between MP and SPP, resulting in an AUC value of 0.822 and sensitivity and specificity of 0.81 and 0.81, respectively. Through the decision curve, RF was found to be superior to other classifiers.

Radiomics is an artificial intelligence technology that extracts features such as shape, intensity, texture, and wavelet from images based on images and converts them into high-dimensional quantifiable quantitative feature data to further reflect the biological information of lesions. It can provide relevant information for disease diagnosis, prognosis evaluation, and efficacy prediction (20–22). To date, few studies have used radiomics to solve the problem of pneumonia identification. Mei et al. (23) used artificial intelligence algorithms to combine chest CT findings with clinical symptoms, exposure history, and laboratory tests to diagnose COVID-19. Wang et al. (24) combined deep learning-radiomics models to distinguish COVID-19 from non-COVID-19 viral pneumonia. Honglin Li (10) confirmed that radiomics-clinical nomograms have good discriminative power for mycoplasmal pneumonia and bacterial pneumonia. These studies demonstrate the feasibility of using radiomics to identify lung inflammation. On this basis, we distinguish between adult and children mycoplasmal pneumonia. Logistic regression is a multiple regression analysis method that studies the relationship between a binary or multi-class response variable and multiple influencing factors (25). This study used the LASSO logistic regression model to screen and model 1,904 imaging features and calculated the Radscore for each patient,

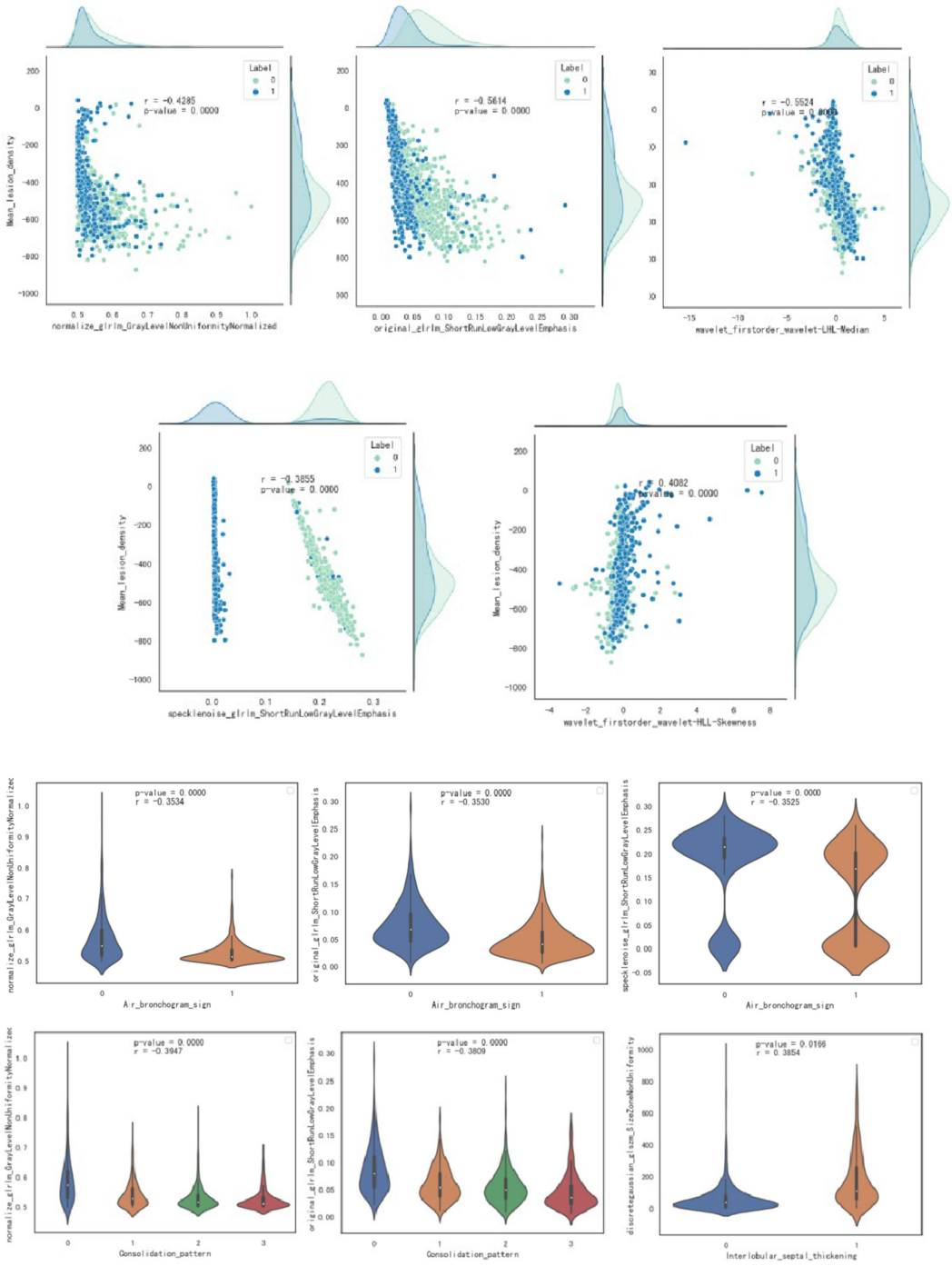
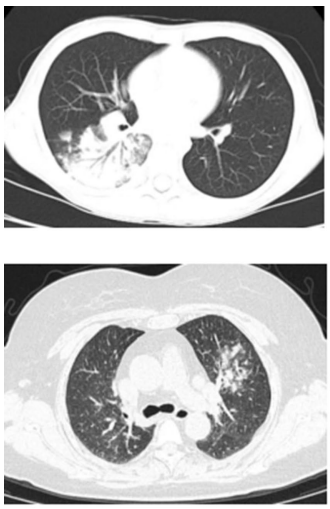


FIGURE 5
Visualization through scatter plot and box plot analysis of the correlation between quantitative and qualitative imaging features and radiomics features.

which can more intuitively reflect the imaging differences between adults and children with Mycoplasma pneumonia. The internal training set AUC value of the radiomics feature model in this group is 0.995 (0.992, 0.998), the validation set AUC value is 0.952 (0.921, 0.978), and the external test set AUC value is 0.969 (0.953, 0.982), indicating good differential diagnostic performance. To explore the relationship between radiomics features, CT imaging signs, and clinical features, a combined model nomogram was established based on radiomics, combining clinical and CT imaging signs. The internal training set had an AUC value of 0.996 (0.993, 0.998), the validation set had an AUC value of 0.972 (0.942, 0.995), and the external test set had an AUC value of 0.986 (0.976, 0.993), which is higher than that of the single model. Consistent with the study by Honglin Li et al. (10), a combined nomogram combining radiological and clinical features was established and validated for distinguishing Mycoplasma pneumonia and bacterial pneumonia with similar CT manifestations. In the radiomics model, the AUC of the training set was 0.877 and the AUC of the test set was 0.810.



A

Child (male, 9 years old) with mycoplasmal pneumonia:
Consolidation pattern:3;
original_glrlm_ShortRunLowGrayLevelEmphasis:0.031;
normalize_glrlm_GrayLevelNonUniformityNormalized:0.513

B

Adult (female, 57 years old) with mycoplasmal pneumonia:
Consolidation pattern:1;
original_glrlm_ShortRunLowGrayLevelEmphasis:0.046;
normalize_glrlm_GrayLevelNonUniformityNormalized:0.534

FIGURE 6
Correlation analysis between radiomics and CT imaging signs: case presentation. **(A)** Child (male, 9 years old) with mycoplasmal pneumonia: Consolidation pattern:3; original_glrlm_ShortRunLowGrayLevelEmphasis:0.031; normalize_glrlm_GrayLevelNonUniformityNormalized:0.513. **(B)** Adult (female, 57 years old) with mycoplasmal pneumonia: Consolidation pattern:1; original_glrlm_ShortRunLowGrayLevelEmphasis:0.046; normalize_glrlm_GrayLevelNonUniformityNormalized:0.534.

TABLE 4 Correlation analysis results between CT imaging signs and radiomics features.

CT feature	Radiomics feature	r	*P*
Mean_lesion_density	original_glrlm_ShortRunLowGrayLevelEmphasis	−0.561384238	1.51E-82
Mean_lesion_density	wavelet-LHL firstorder Median	−0.552387518	1.88E-79
Mean_lesion_density	normalize_glrlm_GrayLevelNonUniformityNormalized	−0.428460907	4.51E-45
Mean_lesion_density	specklenoise_glrlm_ShortRunLowGrayLevelEmphasis	−0.385539074	4.06E-36
Mean_lesion_density	wavelet-HLL firstorder Skewness	0.408192167	1.11E-40
Consolidation_pattern	original_glrlm_ShortRunLowGrayLevelEmphasis	−0.380854025	3.22E-35
Consolidation_pattern	normalize_glrlm_GrayLevelNonUniformityNormalized	−0.394684625	6.44E-38
Air_bronchogram_sign	original_glrlm_ShortRunLowGrayLevelEmphasis	−0.353032225	3.61E-30
Air_bronchogram_sign	normalize_glrlm_GrayLevelNonUniformityNormalized	−0.353370265	3.16E-30
Air_bronchogram_sign	specklenoise_glrlm_ShortRunLowGrayLevelEmphasis	−0.352479244	4.50E-30
Interlobular_septal_thickening	discretegaussian_glszm_SizeZoneNonUniformity	0.385360317	4.39E-36

In the radiomics-clinical model, the AUC of the training set is 0.905 and the AUC of the test set is 0.847. Decision curve analysis shows that both models can improve the clinical benefits of patients, and the radiomics-clinical combination model achieves higher clinical benefits than the radiomics model.

The features of radiomics, including shape, grayscale, and texture, help to build radiomics models (26). This study establishes the correlation between radiomics features and CT imaging signs, and the study reveals that "mean lesion density" is negatively correlated with "original glrlm ShortRunLowGrayLevelEmphasis," "wavelet-LHL firstorder Median," "normalize glrlm GrayLevelNonUniformityNormalized," and "specklenoise glrlm ShortRunLowGrayLevelEmphasis"; and is positively correlated with "wavelet-HLL firstorder Skewness"; "consolidation pattern" is negatively correlated with "original glrlm ShortRunLowGrayLevelEmphasis" and "normalize glrlm GrayLevelNonUniformityNormalized"; "air bronchogram sign" is negatively correlated with "original glrlm ShortRunLowGrayLevelEmphasis," "normalize glrlm GrayLevelNonUniformityNormalized," and "specklenoise glrlm ShortRunLowGrayLevelEmphasis"; "Interlobular_septal_thickening" is negatively correlated with "discretegaussian glszm SizeZoneNonUniformity"; and the correlation coefficients were all greater than 0.35. Most of these radiomics features are texture features and grayscale statistics features, indicating that texture features and grayscale statistics features are largely quantitative representations of CT image features. Moreover, based on the close correlation between radiomics features and traditional CT image features, the advantage of radiomics lies in its ability to transform images into a large amount of high-throughput imaging information that can be mined. Through selection and comparison of the information, optimal features are selected, resulting in more

objective and accurate results (27). Radiomics is non-invasive, quantitative, easily accessible, and reproducible. When combined with CT imaging signs and clinical features, it can provide more comprehensive information about the biological characteristics and microenvironment changes of diseases and has broad prospects in disease diagnosis and prognosis evaluation. This study achieved good results in external validation, indicating that multiple centers and different scanners are beneficial for universality.

There are certain limitations in this study: (1) There are common shortcomings in retrospective studies, such as selection bias; (2) Due to the vague outline of pneumonia lesions, it is difficult to accurately delineate the ROI, and even some smaller lesions are easily missed; (3) Without classifying patients into mild and severe groups before extracting features, further research is needed to investigate the impact of different disease severities.

In summary, this study proposes that radiomics features, CT imaging signs, and clinical features facilitate the identification of differences between adults and children with mycoplasmal pneumonia. For the individual model, the radiomics model achieved the highest AUC. The radiomics features are well-correlated with CT imaging signs, which can provide a quantitative representation of different CT imaging signs using radiomics to a certain extent.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by the Ethics Committee of Affiliated Hospital of Hebei University Affiliated Hospital of Hebei University. The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because since this study is a retrospective study, written informed consent is waivered.

Author contributions

HM: Conceptualization, Data curation, Writing – original draft, Writing – review & editing. T-DW: Data curation, Writing – original draft. L-YZ: Data curation, Writing – original draft. J-WH: Writing – original draft. L-yS: Writing – review & editing. WY: Writing – original draft. L-LZ: Writing – original draft. J-JC: Data curation, Writing – review & editing. J-NW: Writing – review & editing. X-PY: Conceptualization, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. The Youth Scientific research fund of Affiliated Hospital of Hebei University (2023QA06).

Acknowledgments

We express our sincere gratitude to Dan Shen for her invaluable contributions in conceptualizing this work and for her insightful suggestions on the manuscript. The authors thank all the patients for participating in this research.

Conflict of interest

J-JC was employed by United Imaging Intelligence (Beijing) Co.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Waites KB, Talkington DF. Mycoplasma pneumoniae and its role as a human pathogen. *Clin Microbiol Rev*. (2004) 17:697–728. doi: 10.1128/CMR.17.4.697-728.2004

2. Ding Y, Chu C, Li Y, Li G, Lei X, Zhou W, et al. High expression of HMGB1 in children with refractory *Mycoplasma pneumoniae* pneumonia. *BMC Infect Dis*. (2018) 18:439. doi: 10.1186/s12879-018-3346-8

3. Waites KB, Xiao L, Liu Y, Balish MF, Atkinson TP. *Mycoplasma pneumoniae* from the respiratory tract and beyond. *Clin Microbiol Rev*. (2017) 30:747–809. doi: 10.1128/CMR.00114-16

4. Miyashita N, Ouchi K, Kawasaki K, Oda K, Kawai Y, Shimizu H, et al. *Mycoplasma pneumoniae* pneumonia in the elderly. *Med Sci Monit*. (2008) 14:Cr387–91.

5. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, They Are Data. *Radiology*. (2016) 278:563–77. doi: 10.1148/radiol.2015151169

6. Wang T, Deng J, She Y, Zhang L, Wang B, Ren Y, et al. Radiomics signature predicts the recurrence-free survival in stage I non-small cell lung Cancer. *Ann Thorac Surg*. (2020) 109:1741–9. doi: 10.1016/j.athoracsur.2020.01.010

7. Wang X, Duan H, Li X, Ye X, Huang G, Nie S. A prognostic analysis method for non-small cell lung cancer based on the computed tomography radiomics. *Phys Med Biol*. (2020) 65:045006. doi: 10.1088/1361-6560/ab6e51

8. Yanling W, Duo G, Zuojun G, Zhongqiang S, Yankai W, Shan L, et al. Radiomics nomogram analyses for differentiating pneumonia and acute Paraquat lung injury. *Sci Rep*. (2019) 9:15029. doi: 10.1038/s41598-019-50886-7

9. Xie C, Ng MY, Ding J, Leung ST, Lo CSY, Wong HYF, et al. Discrimination of pulmonary ground-glass opacity changes in COVID-19 and non-COVID-19 patients using CT radiomics analysis. *Eur J Radiol Open*. (2020) 7:100271. doi: 10.1016/j.ejro.2020.100271

10. Li H, Li T, Cai Q, Wang X, Liao Y, Cheng Y, et al. Development and validation of a Radiomics nomogram for differentiating Mycoplasma pneumonia and bacterial pneumonia. *Diagnostics*. (2021) 11:1330. doi: 10.3390/diagnostics11081330

11. Tang Z, Zhao W, Xie X, Zhong Z, Shi F, Ma T, et al. Severity assessment of COVID-19 using CT image features and laboratory indices. *Phys Med Biol*. (2021) 66:035015. doi: 10.1088/1361-6560/abbf9e

12. Wu Q, Wang S, Li L, Wu Q, Qian W, Hu Y, et al. Radiomics analysis of computed tomography helps predict poor prognostic outcome in COVID-19. *Theranostics*. (2020) 10:7231–44. doi: 10.7150/thno.46428

13. Wu J, Xia Y, Wang X, Wei Y, Liu A, Innanje A, et al. uRP: an integrated research platform for one-stop analysis of medical images. *Front Radiol*. (2023) 3:1153784. doi: 10.3389/fradi.2023.1153784

14. Cheng Z, Qin L, Cao Q, Dai J, Pan A, Yang W, et al. Quantitative computed tomography of the coronavirus disease 2019 (COVID-19) pneumonia. *Radiol Infect Dis*. (2020) 7:55–61. doi: 10.1016/j.jrid.2020.04.004

15. Shan F, Gao Y, Wang J, Shi W, Shi N, Han M, et al. Abnormal lung quantification in chest CT images of COVID-19 patients with deep learning and its application to severity prediction. *Med Phys*. (2021) 48:1633–45. doi: 10.1002/mp.14609

16. Gu D, Chen L, Shan F, Xia L, Liu J, Mo Z, et al. Computing infection distributions and longitudinal evolution patterns in lung CT images. *BMC Med Imaging*. (2021) 21:57. doi: 10.1186/s12880-021-00588-2

17. Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics system to decode the radiographic phenotype. *Cancer Res*. (2017) 77:e104–7. doi: 10.1158/0008-5472.CAN-17-0339

18. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res*. (2011) 12:2825–30.

19. Wang D, Zhao J, Zhang R, Yan Q, Zhou L, Han X, et al. The value of CT radiomic in differentiating *mycoplasma pneumoniae* pneumonia from *streptococcus pneumoniae* pneumonia with similar consolidation in children under 5 years. *Front Pediatr*. (2022) 10:953399. doi: 10.3389/fped.2022.953399

20. Jia X, Wan L, Chen X, Ji W, Huang S, Qi Y, et al. Risk stratification for 1- to 2-cm gastric gastrointestinal stromal tumors: visual assessment of CT and EUS high-risk features versus CT radiomics analysis. *Eur Radiol*. (2023) 33:2768–78. doi: 10.1007/s00330-022-09228-x

21. Tobaly D, Santinha J, Sartoris R, Dioguardi Burgio M, Matos C, Cros J, et al. CT-based Radiomics analysis to predict malignancy in patients with Intraductal papillary mucinous neoplasm (IPMN) of the pancreas. *Cancers*. (2020) 12:3089. doi: 10.3390/cancers12113089

22. Hsu JB, Lee GA, Chang TH, Huang SW, Le NQK, Chen YC, et al. Radiomic Immunophenotyping of GSEA-assessed Immunophenotypes of glioblastoma and its implications for prognosis: a feasibility study. *Cancers*. (2020) 12:3039. doi: 10.3390/cancers12103039

23. Mei X, Lee HC, Diao K, Huang M, Lin B, Liu C, et al. Artificial intelligence–enabled rapid diagnosis of patients with COVID-19. *Nat Med*. (2020) 26:1224–8. doi: 10.1038/s41591-020-0931-3

24. Wang H, Wang L, Lee EH, Zheng J, Zhang W, Halabi S, et al. Decoding COVID-19 pneumonia: comparison of deep learning and radiomics CT image signatures. *Eur J Nucl Med Mol Imaging*. (2021) 48:1478–86. doi: 10.1007/s00259-020-05075-4

25. Xu WH, Li ML, Gao S, Ni J, Yao M, Zhou LX, et al. Middle cerebral artery intraplaque hemorrhage: prevalence and clinical relevance. *Ann Neurol*. (2012) 71:195–8. doi: 10.1002/ana.22626

26. Wang G, Kang B, Cui J, Deng Y, Zhao Y, Ji C, et al. Two nomograms based on radiomics models using triphasic CT for differentiation of adrenal lipid-poor benign lesions and metastases in a cancer population: an exploratory study. *Eur Radiol*. (2023) 33:1873–83. doi: 10.1007/s00330-022-09182-8

27. Park VY. Expanding applications of MRI-based radiomics in HER2-positive breast cancer. *EBioMedicine*. (2020) 61:103085. doi: 10.1016/j.ebiom.2020.103085

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 05 July 2024
DOI 10.3389/fmed.2024.1402967

Check for updates

OPEN ACCESS

EDITED BY
Minjeong Kim,
University of North Carolina at Greensboro, United States

REVIEWED BY
Raffaella Massafra,
National Cancer Institute Foundation (IRCCS), Italy
Meng Wang,
Liaoning Technical University, China

*CORRESPONDENCE
Guoxu Zhang
✉ zhangguoxu_502@163.com

RECEIVED 18 March 2024
ACCEPTED 14 June 2024
PUBLISHED 05 July 2024

CITATION
Lu G, Tian R, Yang W, Liu R, Liu D, Xiang Z and Zhang G (2024) Deep learning radiomics based on multimodal imaging for distinguishing benign and malignant breast tumours.
Front. Med. 11:1402967.
doi: 10.3389/fmed.2024.1402967

COPYRIGHT
© 2024 Lu, Tian, Yang, Liu, Liu, Xiang and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Deep learning radiomics based on multimodal imaging for distinguishing benign and malignant breast tumours

Guoxiu Lu[1,2], Ronghui Tian[1], Wei Yang[3], Ruibo Liu[1], Dongmei Liu[4], Zijie Xiang[5] and Guoxu Zhang[2]*

[1]College of Medicine and Biological Information Engineering, Northeastern University, Shenyang, Liaoning, China, [2]Department of Nuclear Medicine, General Hospital of Northern Theater Command, Shenyang, Liaoning, China, [3]Department of Radiology, Cancer Hospital of China Medical University, Liaoning Cancer Hospital and Institute, Shenyang, Liaoning, China, [4]Department of Ultrasound, Beijing Shijitan Hospital, Capital Medical University, Beijing, China, [5]Biomedical Engineering, Shenyang University of Technology, Shenyang, Liaoning, China

Objectives: This study aimed to develop a deep learning radiomic model using multimodal imaging to differentiate benign and malignant breast tumours.

Methods: Multimodality imaging data, including ultrasonography (US), mammography (MG), and magnetic resonance imaging (MRI), from 322 patients (112 with benign breast tumours and 210 with malignant breast tumours) with histopathologically confirmed breast tumours were retrospectively collected between December 2018 and May 2023. Based on multimodal imaging, the experiment was divided into three parts: traditional radiomics, deep learning radiomics, and feature fusion. We tested the performance of seven classifiers, namely, SVM, KNN, random forest, extra trees, XGBoost, LightGBM, and LR, on different feature models. Through feature fusion using ensemble and stacking strategies, we obtained the optimal classification model for benign and malignant breast tumours.

Results: In terms of traditional radiomics, the ensemble fusion strategy achieved the highest accuracy, AUC, and specificity, with values of 0.892, 0.942 [0.886–0.996], and 0.956 [0.873–1.000], respectively. The early fusion strategy with US, MG, and MRI achieved the highest sensitivity of 0.952 [0.887–1.000]. In terms of deep learning radiomics, the stacking fusion strategy achieved the highest accuracy, AUC, and sensitivity, with values of 0.937, 0.947 [0.887–1.000], and 1.000 [0.999–1.000], respectively. The early fusion strategies of US+MRI and US+MG achieved the highest specificity of 0.954 [0.867–1.000]. In terms of feature fusion, the ensemble and stacking approaches of the late fusion strategy achieved the highest accuracy of 0.968. In addition, stacking achieved the highest AUC and specificity, which were 0.997 [0.990–1.000] and 1.000 [0.999–1.000], respectively. The traditional radiomic and depth features of US+MG + MR achieved the highest sensitivity of 1.000 [0.999–1.000] under the early fusion strategy.

Conclusion: This study demonstrated the potential of integrating deep learning and radiomic features with multimodal images. As a single modality, MRI based on radiomic features achieved greater accuracy than US or MG. The US and MG models achieved higher accuracy with transfer learning than the single-mode or radiomic models. The traditional radiomic and depth features of US+MG + MR achieved the highest sensitivity under the early fusion strategy, showed

higher diagnostic performance, and provided more valuable information for differentiation between benign and malignant breast tumours.

KEYWORDS

deep learning, radiomics, multimodality imaging, breast tumours, deep learning radiomics, MRI, Mammography, Ultrosonography

1 Introduction

Breast cancer is the most prevalent cancer and the second leading cause of cancer-related deaths among women in the United States (1). In 2023, an estimated 55,720 women were diagnosed with carcinoma *in situ*, whilst 297,790 were diagnosed with invasive carcinoma, and 43,170 women died from breast cancer (2). Early diagnosis and classification are critical for effective treatment. Currently, many imaging modalities, such as ultrasonography (US), mammography (MG), and magnetic resonance imaging (MRI), are commonly used for the classification and diagnosis of breast cancer (3). MG is the predominant tool used for breast cancer screening (4–6), showing high sensitivity for calcification, but its low specificity is one of its limitations. Consequently, a large number of unnecessary biopsies are carried out, leading to healthcare resource waste and stress for patients (7, 8). These disadvantages have led to increased use of other adjunct imaging modalities in clinical practise, including US and MRI (9). US can effectively distinguish between cysts and solid masses and is more sensitive in dense breasts than MG (10). As an adjunct to MG, US provides highly accurate breast mass information and facilitates annotations (11, 12), but it often misses certain types of breast masses, such as invasive micropapillary carcinoma, ductal carcinoma *in situ*, invasive lobular carcinoma, fat-surrounded isoechoic lesions, heterogeneous echoic lesions with heterogeneous backgrounds, subareolar lesions, and deep lesions in large breasts. Additionally, lesions may be missed due to poor operator skills (12–15). MRI, which has high sensitivity, supports multiplanar scanning and 3D reconstruction, allowing for better visualisation of breast lesion size, shape, and location (16). MRI is valuable for screening high-risk individuals, diagnosing occult cases, staging, and assessing the response to chemotherapy (17, 18). However, MRI scans are expensive, and the examination requires more time than other tests (19).

Early and precise detection of malignant breast lesions is crucial for timely intervention and improvement of patient prognosis. Conventional diagnostic methods such as US, MG, and MRI are available but have inherent limitations, including indistinct boundaries, false-positive results, and potential sampling errors. In recent years, deep learning radiomics (DLR) in breast cancer has gained attention as a promising field (20, 21). Although deep learning (DL) models have achieved considerable progress in the automatic segmentation and classification of breast cancer (22, 23), data on how they are improving the overall management of breast cancer, starting from screening to diagnosis and ultimately to survival, are lacking (24). MG, US, and MRI are routinely used during breast cancer screening and are commonly used to identify and characterise breast lesions and guide biopsy. Several studies have focussed on MG and US. Cruz et al. (25) proposed a method consisting of different steps, including segmentation and extraction of deep learning features performed by a CNN—specifically, DenseNet 201. They analysed deep learning and handcrafted features during the fusion stage and then applied several classifiers (XGBoost, AdaBoost, and multilayer perceptron) based on stochastic measures. Ultimately, they achieved strong performance in multimodal imaging studies (US and MG). Lamb et al. (26) reported a higher cancer detection rate for patients who underwent breast screening by MRI than for patients identified as high risk with the traditional risk model using a retrospective mammogram-based model of 2,168 women. Natalia et al. (27) tested three different clinical imaging modalities (dynamic contrast-enhanced MRI, full-field digital mammography, and ultrasound) by pretraining a CNN and fusing it with deep learning methods for radiomic computer-aided diagnosis. They found that compared to previous breast cancer methods, computer-aided diagnosis methods achieved better performance in distinguishing between malignant and benign lesions. However, open questions remain on how to use the DL risk assessment model in clinical practise, and few studies have focussed on multimodality imaging based on deep learning and radiomics with MG, US, and MRI.

The aim of this study was to develop a comprehensive deep learning radiomic framework utilising multimodal imaging data, including MG, US, and MRI data. By integrating deep learning radiomic technology with multimodal imaging, complementary information from different imaging modes can be leveraged to fully characterise the imaging features of breast tumours, thereby achieving a greater differential diagnosis capability for benign and malignant tumours than single-mode radiomics, which will ultimately lead to a reduction in unnecessary biopsies.

2 Materials and methods

2.1 Patient population

This retrospective study obtained approval from the institutional review board of our hospital (Approval No. Y(2404)-030), and the requirement for informed consent was waived. This study enrolled 1,564 female patients who preoperatively underwent multimodality (US, MG, and MRI) examinations at our centre between January 2018 and May 2023. The inclusion criteria were (a) complete imaging and clinical data availability, (b) multimodality breast examination performed within 4 weeks before breast surgery, and (c) no treatment performed before the aforementioned examination. The exclusion criteria were as follows: (a) a history of preoperative therapy, including radiotherapy or neoadjuvant chemotherapy; (b) poor image tumour segmentation due to blurred boundaries; (c) missing US, MG, and MRI data; and (d) no available pathological results. Ultimately, 322 patients (112 with benign breast tumours and 210 with malignant

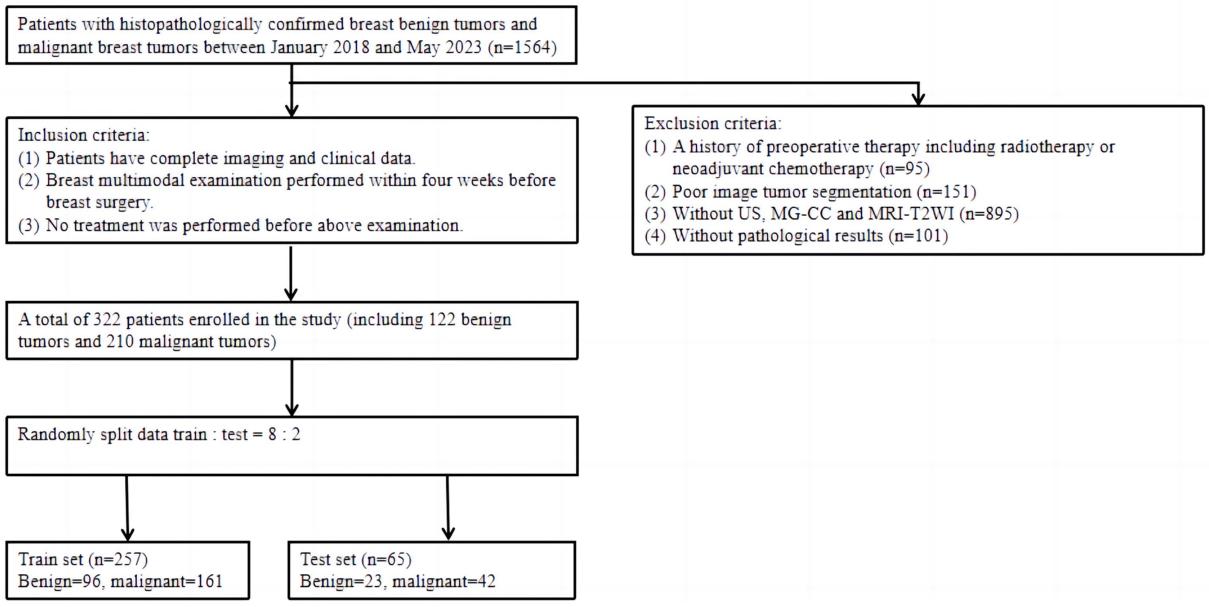


FIGURE 1
Flowchart of patient recruitment.

breast tumours) were included, with 257 patients (96 with benign breast tumours and 161 with malignant breast tumours) enrolled in the training cohort and 65 patients (23 with benign breast tumours and 42 with malignant breast tumours) enrolled in the internal testing cohort. The enrolment process is shown in Figure 1.

2.2 Image acquisition

Each patient's multimodality imaging examinations were as follows: MG-CC, MRI-T2WI, and US. In all patients, routine digital mammography was performed with the Hologic Selenia Dimensions system using standard, craniocaudal (CC), and mediolateral (MLO) views, and we analysed the former images. Routine ultrasound, including Doppler US, was performed using Philips IU22 and EPIQ7 instruments with 12–5-MHz transducers. All contrast-enhanced MRI examinations were performed on a 3.0 T MR system (Skyra, Siemens Healthcare, 3.0 T GE Discovery MR750) in the prone position with no breast compression using a dedicated four-channel breast coil and the following sequences: T2-weighted imaging (T2WI), dynamic contrast-enhanced (DCE) imaging, and diffusion-weighted imaging (DWI). Within 2 min after intravenous injection of gadolinium contrast agent (0.2 mL/kg), the first postcontrast images were acquired, followed by five subsequent postcontrast images were acquired. Axial DWI scans were acquired with two b-values (0 and 1,000 s/mm^2). All patients had undergone core biopsy or surgery of the abnormal area. The final histopathological results were all recorded.

2.3 Region of interest segmentation

Primary breast tumours were selected for region of interest (ROI) segmentation on the largest layer of the tumour. Two radiologists and one diagnostic ultrasound physician with extensive experience (reader 1 with 12 years, reader 2 with 10 years, and reader 3 with 12 years) in breast imaging diagnosis manually delineated each ROI along the tumour margin from the first to the last layer of the whole tumour using ITK-SNAP software (version 3.80). They completed ROI segmentation under the supervision of a senior radiologist with 30 years of experience in breast imaging diagnosis. The radiologists were blinded to the histopathological information of the malignant breast tumours and benign tumours from the US, MG-CC, and MRI-T2W images. We traced abnormal areas in these images and attempted to delineate the burr at the edge of each tumour as completely as possible. All lesion images were included, as shown in Figure 2.

2.4 Feature extraction and selection

A total of 108 radiomic features were extracted using the PyRadiomics (3.0.1) open-source Python package. In this study, the following features were extracted: first-order statistics (FOSs), shape-based 2D (S-2D/3D) features, grey-level co-occurrence matrices (GLCMs), grey-level run length matrices (GLRLMs), grey-level size matrices (GLSZMs), neighbourhood grey tone difference matrices (NGTDMs), and grey-level dependence matrices (GLDMs) (28). Feature selection and fusion techniques were applied to reduce dimensionality and integrate complementary information. The Mann–Whitney U-test and Spearman's rank correlation coefficient were used to determine the statistical significance and repeatability of the features, respectively. Finally, the least absolute shrinkage and selection operator (LASSO) regression model was used to construct the feature signature for the entire dataset.

Classification models for single-mode and multimodal fusion were established from multimodal imaging (MG-CC, US, and MRI-T2WI). The classification model was then constructed using different strategies, including support vector machine (SVM),

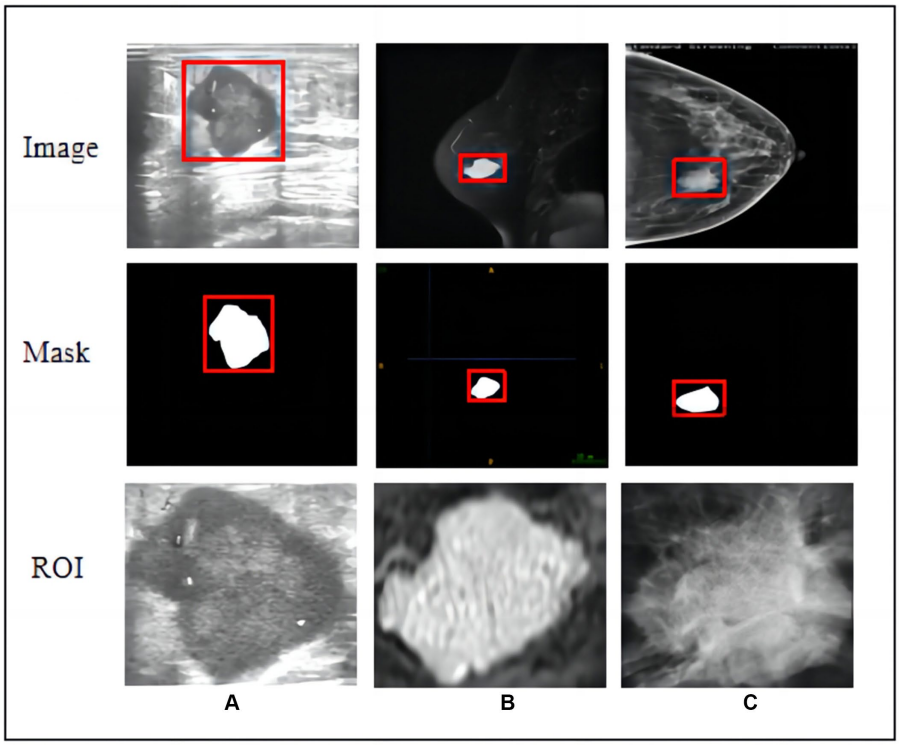


FIGURE 2
Raw images, hand-crafted masks, and cropped ROIs of three modal images. **(A)** Ultrasound, **(B)** T2-weighted magnetic resonance imaging, and **(C)** mammography (MG) craniocaudal view.

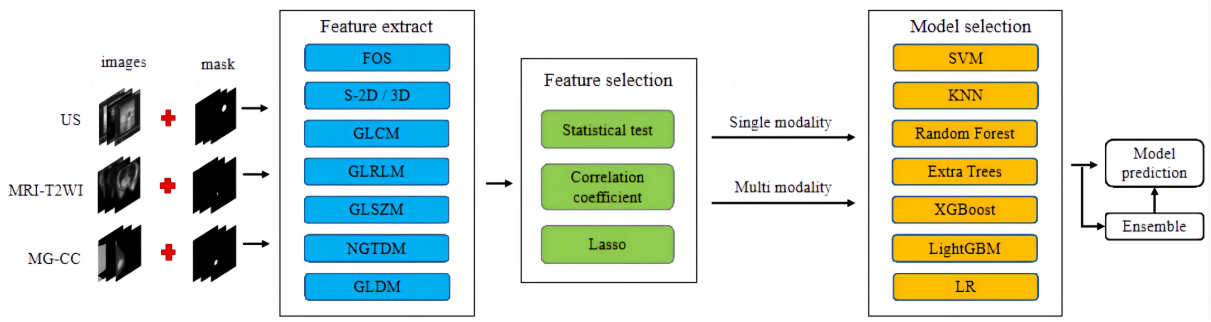


FIGURE 3
Workflow of conventional radiomics from multimodal data. We extracted conventional radiomic features from US, MR, and MG-CC images. Feature selection and fusion techniques were applied to reduce dimensionality and integrate complementary information. The classification model was constructed using seven machine learning algorithms.

K-nearest neighbour (KNN), random forest (RF), extremely randomised trees (ExtraTree), extreme gradient boosting (XG Boost), light gradient boosting machine (LightGBM), and logistic regression (LR), and the optimal fusion method was selected. The workflow for classification model construction is shown in Figure 3.

For US, MG-CC, and MRI-T2WI multimodality imaging, in terms of deep learning, we used a pretrained ResNet-50 model to perform transfer learning tasks on rectangular ROI images acquired from the three imaging modes, as shown in Figure 4 (Step 1). Specifically, the convolution layer parameters of the ResNet-50 model were fixed, and the output of the fully connected layer was 2. During the model training stage, the optimal parameter settings (batch size = 32, learning rate = 0.001, epochs = 200, and optimiser = sgd) were obtained through hyperparameter fine-tuning. Next, we input the images from the three imaging modes into their respective optimal models and derived the deep feature values of the average pooling layer. Since the size of the feature map of the pooling layer was fixed, the number of dimensions of the deep feature values for all the modes was 2048. For feature selection, the PCA algorithm was used to reduce the dimension of the depth feature value. The classification model was then established using different strategies, as shown in Figure 4 (Step 2). For model interpretability, Grad-CAM was utilised to

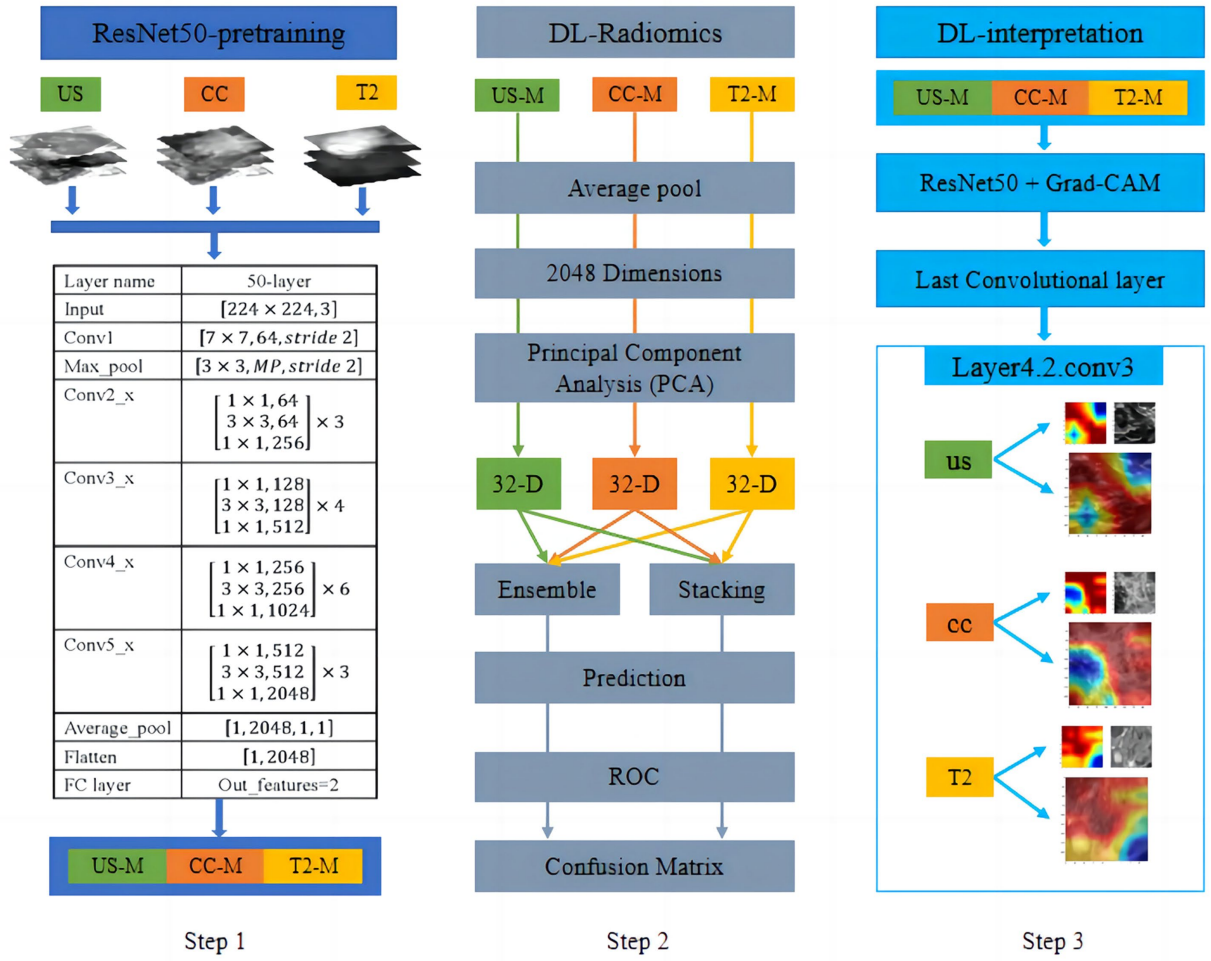


FIGURE 4
Workflow of transfer learning from multimodal data (Step 1 and Step 2) and visualisation of the CNN decision process (Step 3). For US, MG-CC, and MRI-T2WI images, we employed transfer learning using a pretrained ResNet-50 model. After model training, deep features were extracted from the average pooling layer and reduced in dimensionality with PCA. A classification model was then constructed using two different late-fusion strategies. For model interpretability, Grad-CAM was utilised to visualise and explain the validity of the CNN models.

visualise and explain the validity of the multimodal models, as shown in Figure 4 (Step 3).

2.5 Feature fusion

The fusion workflow of deep features and conventional radiomic features from multimodal data were below: The feature fusion methods were divided into early fusion and late fusion (ensemble and stacking) approaches. For early fusion, features from different modalities were concatenated before modelling to create an integrated feature representation as input to the classifier. For the ensemble approach, accuracy-weighted average integration based on softmax normalisation weighting was used. For the stacking approach, separate models were first built on each modality, and then their outputs were combined via the ensemble method. Stacking involved using a machine learning model to fuse the results from the training and testing sets and using another machine learning algorithm for classification, as shown in Figure 5.

2.6 Evaluation indicators

The model performance evaluation adopted four evaluation metrics, namely accuracy, sensitivity, specificity, and AUC value. Accuracy refers to the proportion of correctly classified samples to the total number of samples. Sensitivity represents the proportion of correctly classified positive samples to the actual number of positive samples. Specificity represents the proportion of correctly classified negative samples to the actual number of negative samples. The AUC is the area under the ROC curve, and the ROC curve is the curve obtained by plotting the True Positive Rate on the Y-axis and the False Positive Rate on the X-axis. The value of AUC ranges from 0.5 to 1, and the higher the AUC, the better the performance of the classifier. TP is the number of true-positive results, FP is the number of false-positive results, TN is the number of true-negative results, and FN is the number of false-negative results.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

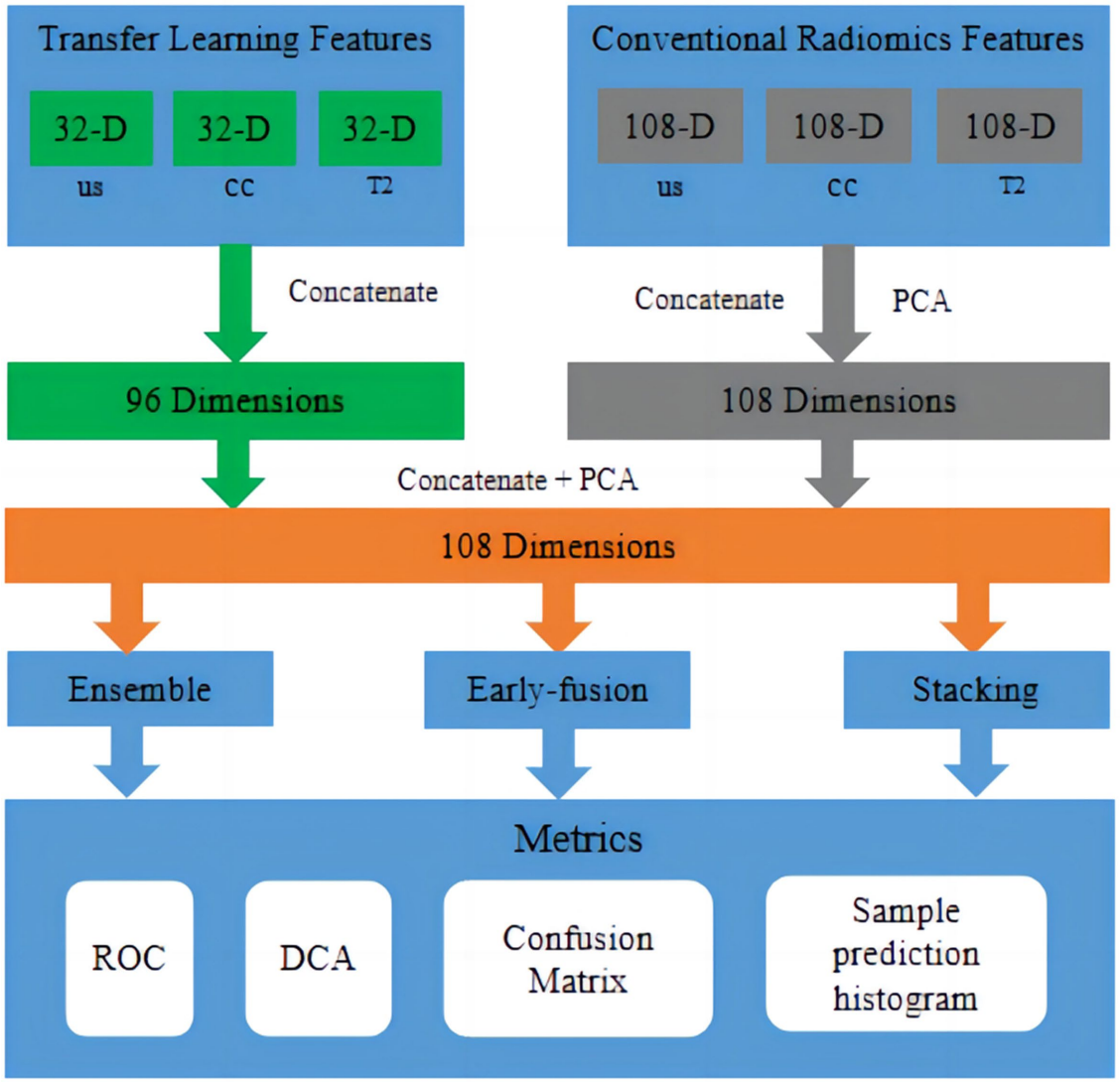


FIGURE 5
Fusion workflow of deep features and conventional radiomic features from multimodal data. The feature fusion methods were divided into early fusion and late fusion (ensemble and stacking) approaches. In early fusion, features from different modalities were concatenated before modelling to create an integrated feature representation as input to the classifier. In late fusion, separate models were built on each modality first, and then, their outputs were combined via ensemble or stacking techniques to produce final predictions.

$$Sensitivity = \frac{TP}{TP + FN} \quad (2)$$

$$Specificity = \frac{FP}{FP + TN} \quad (3)$$

3 Results

3.1 Clinical characteristics

In this study, 322 female patients with a mean age of 50.48 ± 11.57 years were enrolled. The patients were divided into a training set (257 patients) and a test set (65 patients). A total of 112 benign breast tumours and 210 malignant breast tumours were included. The clinicopathological data and corresponding multimodal imaging data resulted in 6440 data points. No significant difference in clinical features was noted amongst the cohorts ($p > 0.05$), as shown in Table 1.

3.2 Radiomic model for multimodal imaging

A total of 108 groups of feature values from US, MRI-T2WI, and MG-CC rectangular ROI images were extracted. After feature selection, we retained 42 sets of feature values for the training of the machine learning model. As shown in Figure 3, the experiments were divided into a single-mode radiomic model, a prefusion model (two-mode image fusion model and three-mode image fusion model), and an ensemble fusion modal method. For both single-mode features and multimode

TABLE 1 Characteristics of breast tumours in this study.

Characteristics	Training (n = 257)	Testing (n = 65)	Values	p
Menstrual status	89 (34.6%)	23 (35.4%)	$\chi 2 = 3.078$	0.079
Age (years)	50.31 ± 11.57	51.08 ± 10.72	$t = 0.486$	0.627
Diameter (mm)	19.94 ± 11.27	22.78 ± 10.01	$t = 1.476$	0.141
CA-153	19.76 ± 8.97	20.52 ± 10.27	$t = 0.593$	0.554
BI-RADS category			$\chi 2 = 6.080$	0.108
1–3	57 (22.2%)	24 (36.9%)	–	–
4(4a,4b,4c)	138 (53.7%)	28 (43.1%)	–	–
5	44 (17.1%)	8 (12.3%)	–	–
6	18 (7.0%)	5 (7.7%)		
Pathology			$\chi 2 = 0.087$	0.768
Benign	96 (37.4%)	23 (35.4%)	–	–
Malignant	161 (62.6%)	42 (64.6%)	–	–

TABLE 2 Results of radiomic classification utilising conventional features.

Methods	Accuracy	AUC	Sensitivity	Specificity	Classifier
US	0.784	0.707 [0.555–0.858]	0.904 [0.815–0.993]	0.565 [0.362–0.767]	SVM
MR	0.800	0.795 [0.674–0.915]	0.857 [0.751–0.962]	0.695 [0.507–0.883]	SVM
MG	0.753	0.748 [0.612–0.883]	0.714 [0.577–0.850]	0.826 [0.671–0.980]	XGBoost
US+MR	0.815	0.858 [0.763–0.952]	0.833 [0.720–0.946]	0.782 [0.614–0.951]	SVM
US+MG	0.692	0.718 [0.578–0.857]	0.642 [0.497–0.787]	0.782 [0.614–0.951]	LightGBM
MR + MG	0.815	0.746 [0.603–0.889]	0.881 [0.783–0.978]	0.727 [0.507–0.883]	XGBoost
US+MR + MG	0.843	0.812 [0.693–0.929]	0.952 [0.887–1.000]	0.636 [0.435–0.837]	XGBoost
Ensemble	0.892	0.942 [0.886–0.996]	0.857 [0.751–0.962]	0.956 [0.873–1.000]	SVM + LightGBM#

# denotes the classifiers (SVM + LightGBM), and [] represents the 95% confidence intervals (CI).

fusion features, seven classifiers were tested in the experiment, and finally, the optimal classification model was obtained. For the generation of classification models, 20% of the images were randomly selected for testing, and the other 80% were selected for training. Notably, when training the first classification model, we set up random seeds to fix the instances of the training set and the test set. The established training set and test set ensured the consistency of training and testing of all classification models and thus the fairness of model evaluation.

The combined modalities integrating multimodal imaging (MG-CC, US, and MRI) showed good validity and stability. We described the diagnostic indices of the different modalities for all patients in the primary cohort and validation cohorts. Table 2 and Figure 6A show the evaluation performance of the optimal classification model under different traditional image radiomic feature sets. With respect to the conventional radiomic features, for the single-modal images, MRI-T2WI achieved the best accuracy (80.0%) and an AUC of 0.785 [0.674–0.915]. US had the best sensitivity of 90.4% [81.5–99.3%]. MG-CC had the best specificity of 82.6% [67.1–98.0%] (lines 1 to 3). Amongst the two multimodal methods, US+MRI had the highest AUC of 0.858 [0.763–0.952] and a specificity of 78.2% [61.4–95.1%] (lines 4 to 6). For the three-mode imaging method, the highest accuracy was 84.3%, the AUC was 0.812, the sensitivity was 95.2% [88.7–100.0%], and the specificity was 63.6% (line 7). The ensemble fusion modal method performed the best, with an accuracy of 89.2%, an AUC of 0.942 [0.886–0.996], a sensitivity of 85.7%, and a specificity of 95.6% [87.3–100.0%] (line 8).

3.3 Deep learning models for multimodal imaging

In summary, we used the pretrained ResNet-50 model to extract 2048 sets of feature values from US, MRI-T2WI, and MG-CC rectangular ROI images, as shown in Figure 4 (Step 1). The difference was that feature selection with PCA was used for dimension reduction. The experiment reduced the eigenvalue of each mode to 32 dimensions. In terms of the model, we generated single-mode, multimode, prefusion, and postfusion classification models, as shown in Figure 4 (Step 2). Similarly, seven classification models were tested to determine the optimal classifier. In addition, the experimental setup was also consistent with that described above.

The performance of the deep features from the transfer learning model, when combined with multimodal imaging, outperformed that of the single-mode models, as shown in Table 3 and Figure 6B. For the single-mode images, US achieved the best accuracy and sensitivity of 78.1 and 88.0%, respectively [78.3–97.8%]. MRI had the best AUC and specificity (0.830 [0.723–0.935] and 81.8 [65.7–97.9%], respectively) (lines 1–3). For the two multimodal imaging methods, the accuracies were 85.9,

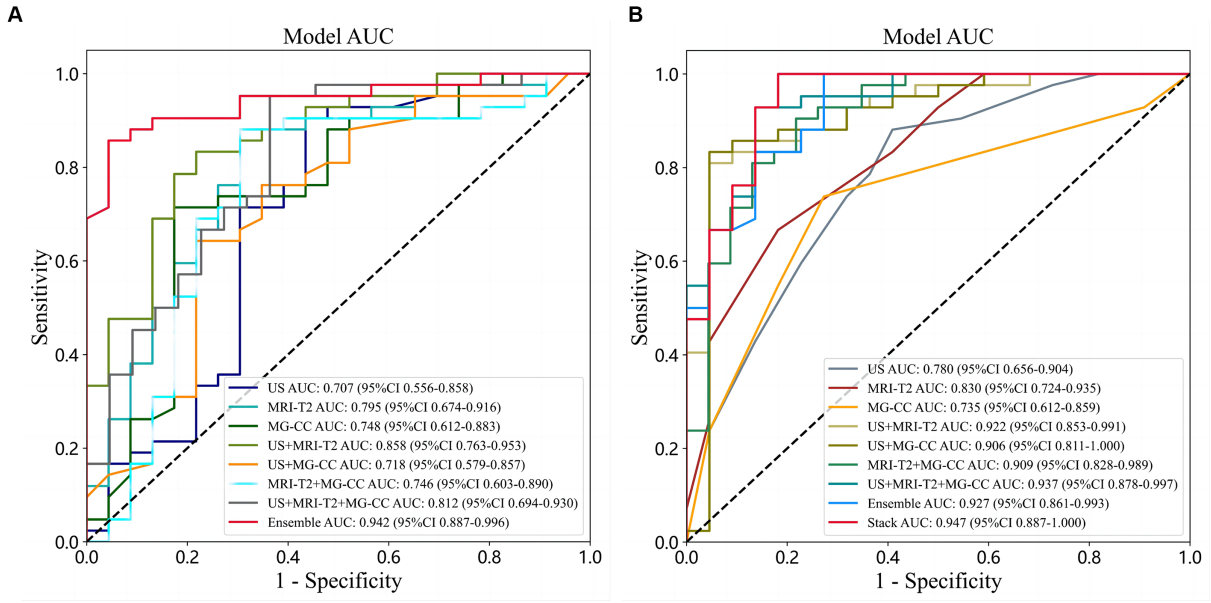


FIGURE 6
Comparison of ROC curves under different classifiers with different sources of features. **(A)** Conventional radiomic features; **(B)** deep features from transfer learning.

TABLE 3 **Results of transfer learning classification utilising deep features.**

Methods	Accuracy	AUC	Sensitivity	Specificity	Classifier
US	0.781	0.780 [0.655–0.903]	0.880 [0.783–0.978]	0.619 [0.385–0.796]	RF
MRI	0.719	0.830 [0.723–0.935]	0.667 [0.524–0.809]	0.818 [0.657–0.979]	ExtraTrees
MG	0.734	0.735 [0.611–0.859]	0.738 [0.605–0.871]	0.761 [0.541–0.913]	KNN
US+MR	0.859	0.922 [0.852–0.991]	0.809 [0.690–0.928]	0.954 [0.867–1.000]	XGBoost
US+MG	0.875	0.906 [0.811–1.000]	0.833 [0.720–0.946]	0.954 [0.867–1.000]	SVM
MR + MG	0.861	0.909 [0.828–0.989]	0.904 [0.815–0.993]	0.782 [0.614–0.951]	SVM
US+MR + MG	0.906	0.937 [0.877–0.996]	0.928 [0.850–1.000]	0.863 [0.720–1.000]	XGBoost
Ensemble	0.906	0.927 [0.861–0.992]	1.000 [0.999–1.000]	0.727 [0.541–0.913]	SVM+KNN+LightGBM#
Stacking	0.937	0.947 [0.887–1.000]	1.000 [0.999–1.000]	0.818 [0.657–0.979]	XGBoost

# denotes the classifiers (SVM + KNN + LightGBM).

87.5, and 86.1%; the AUCs were 0.922, 0.906, and 0.909; the sensitivities were 80.9, 83.3, and 90.4%; and the specificities were 95.4, 95.4, and 78.2%, respectively. US+MR and US+MG had the same specificity of 95.4% [86.7–100.0%] (lines 4–6). For the three multimodal imaging methods, the highest accuracy was 90.6%, the AUC was 0.937 [0.877–0.996], the sensitivity was 92.8% [85.0–100.0%], and the specificity was 86.3% (line 7). Overall, the postfusion model's performance was better than that of the multimodal models. The ensemble model had an accuracy of 90.6%, an AUC of 0.927, a sensitivity of 100.0%, and a specificity of 72.7% (line 8). The stacking model performed best, with an accuracy of 93.7%, an AUC of 0.947 [0.877–1.000], a sensitivity of 100.0% [99.9–100.0%], and a specificity of 81.8% (line 9).

3.4 Deep learning radiomic models for multimodality imaging

The classification model with both conventional image radiomic features and deep learning features showed robust performance. We tried to integrate the conventional image radiomic features and deep learning features from multimodal imaging of breast tumours and further improve the performance of the classification model.

The deep learning feature values of US, MRI-T2WI, and MG-CC were spliced in the same dimension. Figure 5 shows the specific process of feature fusion. In traditional image radiomics, after the three sets of eigenvalues are spliced in the same dimension, the number of dimensions is reduced to 108 according to PCA. After the two types of features were generated, we repeated the above operation, first splicing and then PCA dimension reduction. Finally, we obtained 108-dimensional features containing 51 sets of deep feature values and 57 sets of traditional image radiomic feature values. The 108 sets of features represented a valid feature set for each patient and formed the basis for our classification model. The experiment implemented three fusion methods—early fusion, ensemble, and stacking—for the classification model. Research has shown that the radiomic and deep features of these multimodal

images play a decisive role in the final performance of the model (Figure 6).

Table 4 shows the performance evaluation indices of the three fusion models. We noted that the SVM classifier used in the stacking model achieved the best overall performance, yielding the highest accuracy, AUC, and specificity of 0.968, 0.997 [0.990–1.000], and 1.000 [0.999–1.000], respectively (Figure 7).

3.5 Comparison with different classification models

By setting random seeds, we fixed the training sets and test sets of cases for the deep learning radiomic model (stacking). This study compared benign and malignant breast tumour classification models based on VGG19 (29), GoogLeNet (30), ResNet-101 (31), and Inception-v3 (32). Specifically, the same training set was used for model migration training and fixed convolution layer and modified fully connected layer parameters (the fully connected layer parameter was set to 2). After the model was generated, the same test set was used for the performance evaluation. Table 4 shows the classification results of the deep learning radiomic models and existing deep learning models. The experiments showed that the deep learning image model, which combined traditional imaging radiomic and deep learning features, was superior to the model based on deep learning in the classification of benign and malignant breast tumours, as shown Figure 8.

TABLE 4 Feature fusion results of conventional radiomic features and deep features from transfer learning.

Methods	Accuracy	AUC	Sensitivity	Specificity	classifier
Rad + DF	0.953	0.986 [0.966–1.000]	1.000 [0.999–1.000]	0.863 [0.720–1.000]	SVM
Ensemble	0.968	0.994 [0.982–1.000]	0.976 [0.930–1.000]	0.954 [0.867–1.000]	SVM+XGBoost+LightGBM#
Stacking	0.968	0.997 [0.990–1.000]	0.952 [0.887–1.000]	1.000 [0.999–1.000]	SVM
VGG-19 (29)	0.846	0.867 [0.775–0.959]	0.938 [0.850–1.000]	0.695 [0.507–0.883]	Softmax
GoogLeNet (30)	0.828	0.807 [0.678–0.935]	0.928 [0.863–0.952]	0.636 [0.435–0.837]	Softmax
ResNet-101 (31)	0.796	0.770 [0.640–0.899]	0.952 [0.827–0.987]	0.500 [0.291–0.708]	Softmax
Inception-v3 (32)	0.875	0.892 [0.803–0.980]	0.952 [0.887–1.000]	0.727 [0.541–0.913]	Softmax

In addition, we compared against existing deep learning classification models. Rad+DF represents fused radiomic and deep features. # represents the classifiers (SVM + XGBoost + LightGBM).

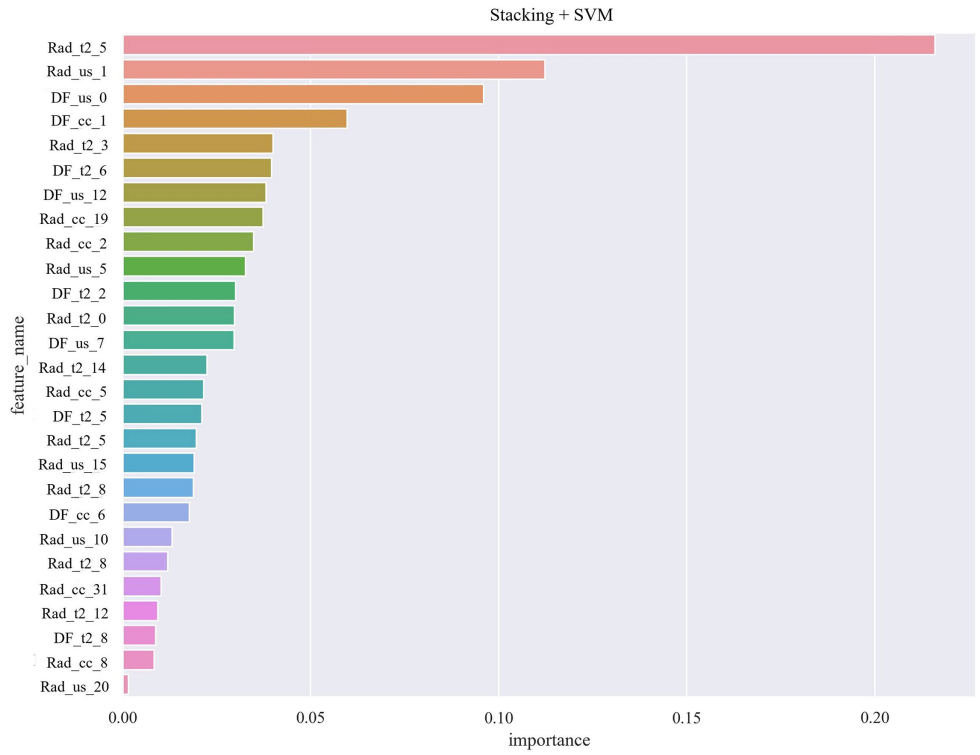


FIGURE 7
Radiomic and deep features of multimodal images role in the final performance of the model.

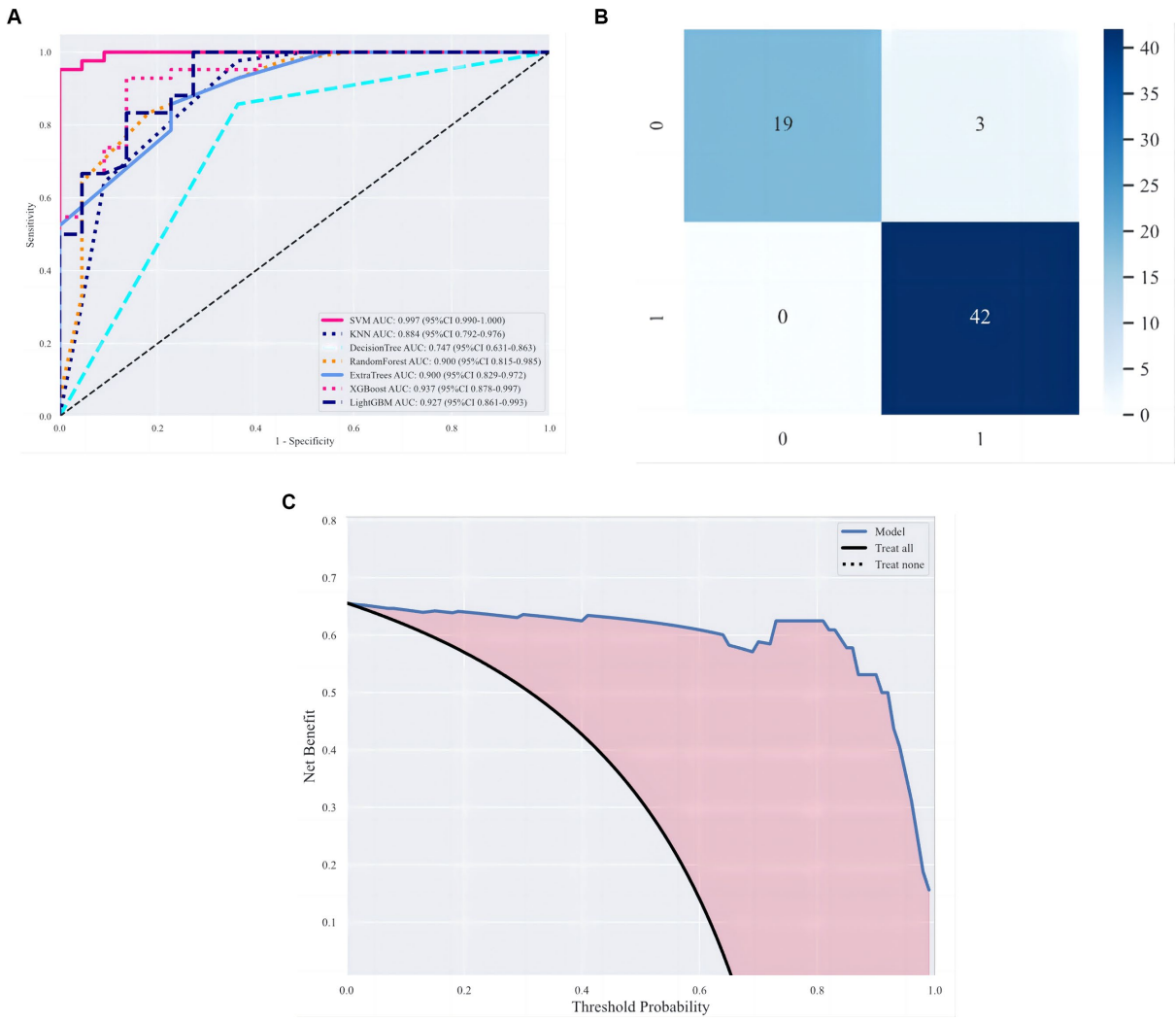


FIGURE 8
Evaluation of classification models. **(A)** ROC curves under different classifiers. **(B, C)** Optimal classification model evaluation, namely, stacking, including confusion matrix and decision curve analysis (DCA).

4 Discussion

Breast cancer shows profound disease heterogeneity, metastasis, and therapeutic resistance and is a leading cause of cancer-related mortality in women. The accuracy and sensitivity of diagnostic tools for differentiating breast tumours need to be further improved, although several diagnostic methods have been developed. Compared to a traditional radiomic model and deep learning feature model, the deep learning radiomic model showed better performance in the classification of benign and malignant breast tumours (33). In our study, we compared a traditional radiomic model, a deep learning model, and a deep learning radiomic model for multimodal imaging. The experimental results showed that the deep learning radiomic fusion model of multimodal imaging exhibited an outstanding performance in distinguishing between benign and malignant breast tumours and achieved the best classification performance. We obtained an AUC of 0.937 with the multimodal model with deep features and transfer learning. With the support of multimodality imaging, the model integrating traditional imaging radiomic and deep learning eigenvalues could more accurately capture key information from the tumour images. Therefore, this model improved the accuracy and robustness of classification.

Hu et al. (34) developed a computer-aided diagnosis method based on dynamic contrast-enhanced (DCE) and T2-weighted MR sequences. The study classified lesions as benign or malignant using support vector machine (SVM) classifiers, and the area under the curve (AUC) of the multiparametric schemes was 0.86 for classifier fusion. The best result was obtained with the feature fusion method. Compared with the prefusion, postfusion added more features into analysed information, so it got the best performance of all the modals.

Huang et al. (35) constructed a deep learning radiopathomic model based on preoperative US images and haematoxylin and eosin (H&E)-stained biopsy slide feature fusion. The deep learning radiopathomic model yielded high performance, with an AUC of 0.929, outperforming the deep learning radiomic model based only on US images and the deep learning pathomic model based only on WSIs. Their study achieved good diagnostic efficacy, which was superior to that of the MG and MRI modalities alone, whilst our study

focussed on US, MG, and MRI multimodal and obtained better performance than single-modal or two-fused modals. We also obtained a high AUC of 0.937, similar to that reported by Huang using H&E staining. Therefore, in the future with the mulitmodal image DLR, we may achieve a non-invasive means of examination, aimed at reducing the need for breast mass biopsy.

Based on the performance of the three models, both the deep radiomic model and the feature fusion model outperform the traditional radiomic models in classifying benign and malignant breast tumours. We attribute this to the incorporation of deep feature values into the traditional radiomic features. To assess the efficacy of the deep feature values, the final convolutional layer of the ResNet-50 model was visualised using the Grad-CAM method, as depicted in Figure 4 (Step 3). The visualisation reveals that the deep feature values contributing to the decision-making are distributed within and around the tumour. We observed that the highlighted areas on the heatmap align with those observed by clinicians, underscoring the significance of integrating deep feature values into traditional radiomic features.

The excellent performance of the deep learning radiomic model provides important technical support and guidance for the early diagnosis and treatment of breast tumours. The advantage of deep learning radiomic modalities in breast tumour classification could not only be reflected in the classification performance but also in the full use of multimodal imaging. The fusion of multimodal imaging could provide more comprehensive and multidimensional information for the model such that the model had more diagnostic value and clinical application prospects. Therefore, we believe that the deep learning radiomic imaging model has the best performance in distinguishing benign and malignant breast tumours and plays an important role in the field of medical imaging.

Previous studies have also explored the use of radiomic models and DLR nomograms with promising results. For example, Gao et al. achieved an AUC of 0.82 with a radiomic model using combined craniocaudal + lateral oblique MG features (36). Zhang et al. (37) developed an ultrasound-based DLR nomogram that showed excellent performance in predicting axillary lymph node load. The AUCs of the training and test sets were 0.900 and 0.821, respectively. In our study, in the field of traditional radiomic models, the integrated features in the ensemble model showed better overall performance than the single-mode models. The advantage of this multimodal fusion was not only the integration of information from different imaging modes but also the exquisite design of the ensemble model. The ensemble model seamlessly integrated information from various imaging modes, making full use of the advantages of each mode, thus improving the overall classification performance. Classifiers such as SVM and LightGBM were selected not only because of their applicability in processing multimodal data but also based on their performance and stability in different situations. Through this clever combination, ensemble models were able to maintain high accuracy whilst maintaining modal robustness and generalisability. In the forecasting process, the ensemble model adopted the weighted voting strategy, which synthesised the opinions of various classifiers, effectively reducing the error rate and improving the reliability of the classification results. In summary, the application of the ensemble model to the traditional radiomic model showed its unique advantages in integrating multimodal information and improving classification performance.

Chen et al. (38) used deep learning features from DWI-ADC imaging and DCE-MRI to predict axillary lymph node metastasis with high accuracy (AUC = 0.80 and 0.71) in training and testing cohorts, respectively. In our study, amongst the deep learning models, the model with amalgamated deep learning features demonstrated superior performance compared to the single-mode models. The fusion strategy of the stacking method significantly enhanced the performance and robustness of the model compared to the single-mode deep learning feature model, rendering it more competitive in practical applications. Integrating predicted probabilities into feature sets through the stacking model enhanced new stacking relationships and data labelling, providing a novel idea for further optimisation of deep feature models (39, 40). Unlike traditional radiomics, we also used a stacking-based deep learning feature model, which enhanced the classification performance, particularly the stacking model and XGBoost classifier, amongst the various classifiers. The unique advantage of the stacking model is its ability to effectively fuse deep learning features from each mode and achieve more precise classification using efficient classifiers such as XGBoost. Kwon et al. (41) compared the performance of every meta-learner model with a stacking ensemble approach as a supporting tool for breast cancer classification. The study showed that using specific models as a meta-learner resulted in better performance than that of single classifiers. Mohammed et al. (42) took the output of the submodels (base-learners) as input and then merged the input predictions to determine the final prediction, which was better than that of each of the base-classifiers. In this study, we achieved high accuracy and perfect specificity (100%) with the stacking deep learning model, which may benefit from our multimodal images.

Currently, differentiating malignant breast tumours from benign breast tumours is very important for guiding future clinical treatment and avoiding unnecessary biopsies. Although several diagnostic methods have been developed (34, 43, 44), the accuracy and sensitivity of those tools for differentiating breast tumours need to be further improved. Patterns of breast calcifications visible on mammograms may be useful for differentiating between benign and malignant lesions. A radiomic feature analysis revealed several statistically significant correlations of the tumour and near and far regions in mammograms with intensity-based histogram features, edge frequency features, and Fourier-based power-law beta features (45). Yamamoto et al. studied 353 patients and identified 21 MRI features, finding that they correlated with 71% of the gene expression profiles of breast cancer (46). Cai et al. created a deep learning (DL)-based CNN capable of discriminating amongst benign and malignant microcalcifications of radiological features of the breast (47). Two model datasets are commonly used by several authors in the state of the art (48–50). Our study analyzed multimodal imaging (MG, US and MRI) of breast tumor with deep learning model. And we got good diagnostic efficacy, which was superior to single model image (MG, US or MRI) and two models of fused image (MG+MRI, MG+US or US+MRI). Therefore, the deep learning radiomic method has a certain value in the differential diagnosis of breast tumours, and multimodal image data can complement each other. We also found that multimodality methods had a strong advantage when the

maximum diameter was less than 1 cm compared to using one or two model images alone.

In the present study, we also detected a noticeable difference in multimodality imaging between DLR and BI-RADS categories 3 and 4, whilst this difference was not detected for BI-RADS categories 5 and 6. The BI-RADS category sometimes varied amongst the model images. MRI images were much more common than US and MG images in our study, especially for BI-RADS 4 and 5. Multimodality imaging provides the best evaluation of the exact BI-RADS category, so multimodality imaging is recommended for diagnosis or surgical consultation for patients in the BI-RADS category 4 or 5. Witowski et al. studied 13,463 patients with breast carcinoma and developed a CNN model based on T1-weighted MR images to generate a three-dimensional (3D) mask of the breast area, achieving the highest sensitivity for BI-RADS 5 (92.5%) and a low value for BI-RADS 3 (33.3%), indicating that BI-RADS 3 represents an uncertain category not only for radiologists but also for DL approaches. This approach also prevents biopsies from yielding benign results in up to 20% of all patients with BI-RADS category 4 lesions (51). Both of the above results showed that the DLR model could serve as a helpful tool in the reporting system to increase the specificity of cancer screening. We still need to further developped broadly accessible, reliable, and accurate multimodality imagings with DLR tools. In this way breast tumor could be detected earlily and get more measures for prevention.

Our study has several limitations: (1) the classification proposed in this study focussed only on the differentiation of benign or malignant breast tumours; thus, it did not accurately distinguish pathological subtypes, which is a topic for future research. (2) In the present study, we utilised only the CC view in the MG and T2 MR images, whilst other piesces of information, such as lateral oblique images from MG and T1-weighted imaging (T1WI), diffusion-weighted imaging (DWI), apparent diffusion coefficient (ADC) images, and DCE-MRI sequences, were not fully analysed. Exploring these additional data may provide more insights into DLR applications. (3) This was a retrospective analysis with a relatively small sample size. For our future study, we plan to use a multicentre external validation dataset and prospective validation to further confirm these findings. (4) Manual segmentation of ROIs on each image slice increases the workload. Further studies should focus on developing deep learning-based segmentation methods for automatic lesion segmentation via multimodal imaging.

5 Conclusion

In this study, we demonstrated the potential of integrating deep learning and radiomic features with multimodal images. As a single modality, MRI based on radiomic features achieved greater accuracy than US or MG. The US and MG models achieved higher accuracy with transfer learning than the single-mode or radiomic models. Our findings may contribute to the growing body of research on the use of DLR in breast cancer diagnosis and classification with MG, US, and MRI. The traditional radiomic and depth features of US+MG + MR achieved the highest sensitivity under the early fusion strategy, exhibited higher diagnostic performance, and provided more valuable information for differentiation between benign and malignant breast tumours. By incorporating multimodal images and DLR analysis, we demonstrated the potential for improved accuracy and clinical relevance in distinguishing breast mass characteristics. In future investigations and validation, we plan to employ the designed fusion approach to other medical images, for example, PET/CT or PET/MRI.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by General Hospital of Northern Theater Command, No. Y(2404)-030. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

GL: Conceptualization, Formal analysis, Writing – original draft, Writing – review & editing. RT: Methodology, Writing – original draft. WY: Methodology, Resources, Writing – review & editing. RL: Data curation, Writing – review & editing. DL: Writing – review & editing, Data curation. ZX: Data curation, Writing – original draft. GZ: Conceptualization, Funding acquisition, Project administration, Supervision, Validation, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was partially supported by grants from Applied Basic Research Program of Liaoning Province (2022JH2/101500011); Livelihood Science and Technology Plan Joint Plan Project of Liaoning Province (2021JH2/10300098).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Zaluzec EK, Sempere LF. Systemic and local strategies for primary prevention of breast Cancer. *Cancers*. (2024) 16:248. doi: 10.3390/cancers16020248

2. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. *CA Cancer J Clin*. (2023) 73:17–48. doi: 10.3322/caac.21763

3. Radak M, Lafta HY, Fallahi H. Machine learning and deep learning techniques for breast cancer diagnosis and classification: a comprehensive review of medical imaging studies. *J Cancer Res Clin Oncol*. (2023) 149:10473–91. doi: 10.1007/s00432-023-04956-z

4. Gegios AR, Peterson MS, Fowler AM. Breast Cancer screening and diagnosis: recent advances in imaging and current limitations. *PET Clin*. (2023) 18:459–71. doi: 10.1016/j.cpet.2023.04.003

5. Tangaro S, Bellotti R, de Carlo F, Gargano G, Lattanzio E, Monno P, et al. MAGIC-5: an Italian mammographic database of digitised images for research. *Radiol Med*. (2008) 113:477–85. doi: 10.1007/s11547-008-0282-5

6. Basile TMA, Fanizzi A, Losurdo L, Bellotti R, Bottigli U, Dentamaro R, et al. Microcalcification detection in full-field digital mammograms: A fully automated computer-aided system. *Phys Med*. (2019) 64:1–9. doi: 10.1016/j.ejmp.2019.05.022

7. Li H, Robinson K, Lan L, Baughan N, Chan CW, Embury M, et al. Temporal machine learning analysis of prior mammograms for breast Cancer risk prediction. *Cancers (Basel)*. (2023) 15:2141. doi: 10.3390/cancers15072141

8. Wang ZH, Li M, Wang HX, Jiang H, Yao Y, Zhang H, et al. Breast Cancer detection using extreme learning machine based on feature fusion with CNN deep features. *IEEE Access*. (2019) 7:105146–58. doi: 10.1109/ACCESS.2019.2892795

9. Schaffter T, Buist DSM, Lee CI, Nikulin Y, Ribli D, Guan Y, et al. Evaluation of combined artificial intelligence and radiologist assessment to interpret screening mammograms. *JAMA Netw Open*. (2020) 3:e200265. doi: 10.1001/jamanetworkopen.2020.0265

10. Berg WA, Zhang Z, Lehrer D, Jong RA, Pisano ED, Barr RG, et al. Detection of breast cancer with addition of annual screening ultrasound or a single screening MRI to mammography in women with elevated breast cancer risk. *JAMA*. (2012) 307:1394–404. doi: 10.1001/jama.2012.388

11. Afrin H, Larson NB, Fatemi M, Alizad A. Deep learning in different ultrasound methods for breast Cancer, from diagnosis to prognosis: current trends, challenges, and an analysis. *Cancers (Basel)*. (2023) 15:3139. doi: 10.3390/cancers15123139

12. Verras GI, Tchabashvili L, Mulita F, Grypari IM, Sourouni S, Panagodimou E, et al. Micropapillary breast carcinoma: from molecular pathogenesis to prognosis. *Breast Cancer*. (2022) 14:41–61. doi: 10.2147/BCTT.S346301

13. Kamitani K, Kamitani T, Ono M, Toyoshima S, Mitsuyama S. Ultrasonographic findings of invasive micropapillary carcinoma of the breast: correlation between internal echogenicity and histological findings. *Breast Cancer*. (2012) 19:349–52. doi: 10.1007/s12282-011-0293-2

14. Yun SU, Choi BB, Shu KS, Kim SM, Seo YD, Lee JS, et al. Imaging findings of invasive micropapillary carcinoma of the breast. *J Breast Cancer*. (2012) 15:57–64. doi: 10.4048/jbc.2012.15.1.57

15. Uematsu T. Ultrasonographic findings of missed breast cancer: pitfalls and pearls. *Breast Cancer*. (2014) 21:10–9. doi: 10.1007/s12282-013-0498-7

16. Patel BK, Lobbes M, Lewin J. Contrast enhanced spectral mammography: a review. *Semin Ultrasound CT MRI*. (2018) 39:70–9. doi: 10.1053/j.sult.2017.08.005

17. Song D, Yang F, Zhang Y, Guo Y, Qu Y, Zhang X, et al. Dynamic contrast-enhanced MRI radiomics nomogram for differentiating axillary lymph node metastasis in breast cancer. *Cancer Imaging*. (2022) 22:17. doi: 10.1186/s40644-022-00450-w

18. Gu YL, Pan SM, Ren J, Yang ZX, Jiang GQ. Role of magnetic resonance imaging in detection of pathologic complete remission in breast cancer patients treated with neoadjuvant chemotherapy: a meta-analysis. *Clin Breast Cancer*. (2017) 17:245–55. doi: 10.1016/j.clbc.2016.12.010

19. Acciavatti RJ, Lee SH, Reig B, Moy L, Conant EF, Kontos D, et al. Beyond breast density: risk measures for breast Cancer in multiple imaging modalities. *Radiology*. (2023) 306:e222575. doi: 10.1148/radiol.222575

20. Dileep G, Gianchandani Gyani SG. Artificial intelligence in breast Cancer screening and diagnosis. *Cureus*. (2022) 14:e30318. doi: 10.7759/cureus.30318

21. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. (2018) 68:394–424. doi: 10.3322/caac.21492

22. Ma Z, Qi Y, Xu C, Zhao W, Lou M, Wang Y, et al. ATFE-net: axial transformer and feature enhancement-based CNN for ultrasound breast mass segmentation. *Comput Biol Med*. (2023) 153:106533. doi: 10.1016/j.compbiomed.2022.106533

23. Jabeen K, Khan MA, Alhaisoni M, Tariq U, Zhang YD, Hamza A, et al. Breast Cancer classification from ultrasound images using probability-based optimal deep learning feature fusion. *Sensors*. (2022) 22:807. doi: 10.3390/s22030807

24. Fruchtman Brot H, Mango VL. Artificial intelligence in breast ultrasound: application in clinical practice. *Ultrasonography*. (2024) 43:3–14. doi: 10.14366/usg.23116

25. Cruz-Ramos C, García-Avila O, Almaraz-Damian JA, Ponomaryov V, Reyes-Reyes R, Sadovnychiy S. Benign and malignant breast tumor classification in ultrasound and mammography images via fusion of deep learning and handcraft features. *Entropy*. (2023) 25:991. doi: 10.3390/e25070991

26. Lamb LR, Mercaldo SF, Ghaderi KF, Carney A, Lehman CD. Comparison of the diagnostic accuracy of mammogram-based deep learning and traditional breast cancer risk modals in patients who underwent supplemental screening with MRI. *Radiology*. (2023) 308:e223077. doi: 10.1148/radiol.223077

27. Antropova N, Huynh BQ, Giger ML. A deep feature fusion methodology for breast cancer diagnosis demonstrated on three imaging modality datasets. *Med Phys*. (2017) 44:5162–71. doi: 10.1002/mp.12453

28. Kontos D, Kalpathy-Cramer J. Breast Cancer risk assessment in the AI era: the importance of modal validation in ethnically diverse cohorts. *Radiol Artif Intell*. (2023) 5:e230462. doi: 10.1148/ryai.230462

29. Suzuki Y, Hanaoka S, Tanabe M, Yoshikawa T, Seto Y. Predicting breast Cancer risk using Radiomics features of mammography images. *J Pers Med*. (2023) 13:1528. doi: 10.3390/jpm13111528

30. Saleh GA, Batouty NM, Gamal A, Elnakib A, Hamdy O, Sharafeldeen A, et al. Impact of imaging biomarkers and AI on breast Cancer management: A brief review. *Cancers*. (2023) 5, 15:5216. doi: 10.3390/cancers15215216

31. Alqudah A, Alqudah AM. Sliding window based deep ensemble system for breast cancer classification. *J Med Eng Technol*. (2021) 45:313–23. doi: 10.1080/03091902.2021.1896814

32. Rahaman MM, Millar EKA, Meijering E. Breast cancer histopathology image-based gene expression prediction using spatial transcriptomics data and deep learning. *Sci Rep*. (2023) 13:13604. doi: 10.1038/s41598-023-40219-0

33. Anwar SM, Majid M, Qayyum A, Awais M, Alnowami M, Khan MK. Medical image analysis using convolutional neural net-works: a review. *J Med Syst*. (2018) 42:226. doi: 10.1007/s10916-018-1088-1

34. Hu Q, Whitney HM, Giger ML. A deep learning methodology for improved breast cancer diagnosis using multiparametric MR. *Sci Rep*. (2020) 10:10536. doi: 10.1038/s41598-020-67441-4

35. Huang Y, Yao Z, Li L, Mao R, Huang W, Hu Z, et al. Deep learning radiopathomics based on preoperative US images and biopsy whole slide images can distinguish between luminal and non-luminal tumors in early-stage breast cancers. *EBioMedicine*. (2023) 94:104706. doi: 10.1016/j.ebiom.2023.104706

36. Gao W, Chen J, Zhang B, Wei X, Zhong J, Li X, et al. Automatic deep learning method for detection and classification of breast lesions in dynamic contrast-enhanced magnetic resonance imaging. *Quant Imaging Med Surg*. (2023) 13:2620–33. doi: 10.21037/qims-22-323

37. Zhang H, Zhao T, Zhang S, Sun J, Zhang F, Li X, et al. Prediction of axillary lymph node metastatic load of breast Cancer based on ultrasound deep learning Radiomics nomogram. *Technol Cancer Res Treat*. (2023) 22:153303382311662. doi: 10.1177/15330338231166218

38. Chen Y, Wang L, Dong X, Luo R, Ge Y, Liu H, et al. Deep learning Radiomics of preoperative breast MRI for prediction of axillary lymph node metastasis in breast Cancer. *J Digit Imaging*. (2023) 36:1323–31. doi: 10.1007/s10278-023-00818-9

39. Ali MD, Saleem A, Elahi H, Khan MA, Khan MI, Yaqoob MM, et al. Breast Cancer classification through Meta-learning ensemble technique using convolution neural networks. *Diagnostics*. (2023) 13:2242. doi: 10.3390/diagnostics13132242

40. Hosni M, Abnane I, Idri A, Carrillo de Gea JM, Fernández Alemán JL. Reviewing ensemble classification methods in breast cancer. *Comput Methods Prog Biomed*. (2019) 177:89–112. doi: 10.1016/j.cmpb.2019.05.019

41. Kwon H, Park J, Lee Y. Stacking ensemble technique for classifying breast Cancer. *Healthc Inform Res*. (2019) 25:283–8. doi: 10.4258/hir.2019.25.4.283

42. Mohammed M, Mwambi H, Mboya IB, Elbashir MK, Omolo B. A stacking ensemble deep learning approach to cancer type classification based on TCGA data. *Sci Rep*. (2021) 11:15626. doi: 10.1038/s41598-021-95128-x

43. Al-Jabbar M, Alshahrani M, Senan EM, Ahmed IA. Analyzing histological images using hybrid techniques for early detection of multi-class breast Cancer based on fusion features of CNN and handcrafted. *Diagnostics*. (2023) 13:1753. doi: 10.3390/diagnostics13101753

44. Du X, Zhao Y. Multimodal adversarial representation learning for breast cancer prognosis prediction. *Comput Biol Med*. (2023) 157:106765. doi: 10.1016/j.compbiomed.2023.106765

45. Li H, Mendel KR, Lan L, Sheth D, Giger ML. Digital mammography in breast Cancer: additive value of Radiomics of breast parenchyma. *Radiology*. (2019) 291:15–20. doi: 10.1148/radiol.2019181113

46. Yamamoto S, Maki DD, Korn RL, Kuo MD. Radiogenomic analysis of breast cancer using MRI: a preliminary study to define the landscape. *AJR Am J Roentgenol*. (2012) 199:654–63. doi: 10.2214/AJR.11.7824

47. Cai H, Huang Q, Rong W, Song Y, Li J, Wang J, et al. Breast microcalcification diagnosis using deep convolutional neural network from digital mammograms. *Comput Math Methods Med*. (2019) 2019:1–10. doi: 10.1155/2019/2717454

48. Daoud MI, Abdel-Rahman S, Bdair TM, al-Najar MS, al-Hawari FH, Alazrai R. Breast tumor classification in ultrasound images using combined deep and handcrafted features. *Sensors*. (2020) 20:6838. doi: 10.3390/s20236838

49. Muduli D, Dash R, Majhi B. Automated diagnosis of breast Cancer using multi-modalDatasets: A deep convolution neural network based approach. *Biomed Signal Process Control*. (2021) 71:102825. doi: 10.1016/j.bspc.2021.102825

50. Zhang YD, Satapathy SC, Guttery DS, Górriz JM, Wang SH. Improved breast Cancer classification through combining graph convolutional network and convolutional neural network. *Inf Process Manag*. (2021) 58:102439. doi: 10.1016/j.ipm.2020.102439

51. Mahoro E, Akhloufi MA. Applying deep learning for breast Cancer detection in radiology. *Curr Oncol*. (2022) 29:8767–93. doi: 10.3390/curroncol29110690

Frontiers in Medicine

Translating medical research and innovation into improved patient care

A multidisciplinary journal which advances our medical knowledge. It supports the translation of scientific advances into new therapies and diagnostic tools that will improve patient care.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact