

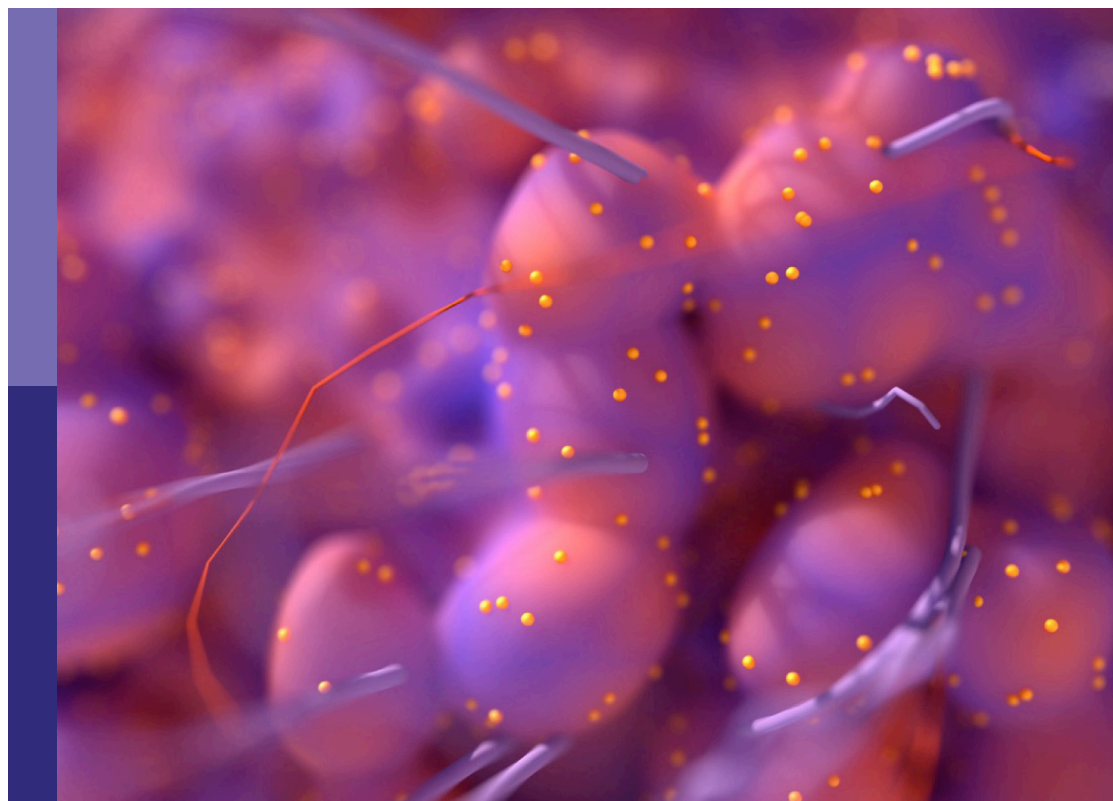
# Artificial intelligence and imaging for oncology

**Edited by**

Poonam Yadav, Zhimin Li and Yuxiang Zhou

**Published in**

Frontiers in Oncology



## 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 978-2-8325-6019-8  
DOI 10.3389/978-2-8325-6019-8

## 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](https://frontiersin.org/about/contact)



# Artificial intelligence and imaging for oncology

## Topic editors

Poonam Yadav — Northwestern University, United States

Zhimin Li — Northwestern University, United States

Yuxiang Zhou — Mayo Clinic Arizona, United States

## Citation

Yadav, P., Li, Z., Zhou, Y., eds. (2025). *Artificial intelligence and imaging for oncology*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-6019-8

## Table of contents

- 05 **Editorial: Artificial intelligence and imaging for oncology**  
Yuxiang Zhou, Zhimin Li and Poonam Yadav
- 09 **A contrast-enhanced CT-based radiomic nomogram for the differential diagnosis of intravenous leiomyomatosis and uterine leiomyoma**  
Jiang Shao, Chaonan Wang, Keqiang Shu, Yan Zhou, Ninghai Cheng, Zhichao Lai, Kang Li, Leyin Xu, Junye Chen, Fenghe Du, Xiaoxi Yu, Zhan Zhu, Jiaxian Wang, Yuyao Feng, Yixuan Yang, Xiaolong Liu, Jinghui Yuan and Bao Liu
- 20 **Using deep DenseNet with cyclical learning rate to classify leukocytes for leukemia identification**  
Essam H. Houssein, Osama Mohamed, Nagwan Abdel Samee, Noha F. Mahmoud, Rawan Talaat, Aymen M. Al-Hejri and Riyadh M. Al-Tam
- 34 **Synthetic MRI plus FSE-PROPELLER DWI for differentiating malignant from benign head and neck tumors: a preliminary study**  
Baohong Wen, Zanzia Zhang, Jing Zhu, Liang Liu, Zijun Liu, Xiaoyue Ma, Kaiyu Wang, Lizhi Xie, Yong Zhang and Jingliang Cheng
- 44 **Radiomic nomogram for discriminating parotid pleomorphic adenoma from parotid adenolymphoma based on grayscale ultrasonography**  
Yi Mao, LiPing Jiang, Jing-Ling Wang, Fang-Qun Chen, Wie-Ping Zhang, Zhi-Xing Liu and Chen Li
- 54 **Machine learning-based CT texture analysis in the differentiation of testicular masses**  
Can Hu, Xiaomeng Qiao, Zhenyu Xu, Zhiyu Zhang and Xuefeng Zhang
- 64 **Brain tumor classification from MRI scans: a framework of hybrid deep learning model with Bayesian optimization and quantum theory-based marine predator algorithm**  
Muhammad Sami Ullah, Muhammad Attique Khan, Anum Masood, Olfa Mzoughi, Oumaima Saidani and Nazik Alturki
- 85 **A study on the radiomic correlation between CBCT and pCT scans based on modified 3D-RUNet image segmentation**  
Yanjuan Yu, Guanglu Gao, Xiang Gao, Zongkai Zhang, Yipeng He, Liwan Shi and Zheng Kang
- 93 **An efficient decision support system for leukemia identification utilizing nature-inspired deep feature optimization**  
Muhammad Awais, Md. Nazmul Abdal, Tallha Akram, Areej Alasiry, Mehrez Marzougui and Anum Masood
- 109 **A deep learning model for the localization and extraction of brain tumors from MR images using YOLOv7 and grab cut algorithm**  
Srigiri Krishnapriya and Yepuganti Karuna

- 123 **Initial investigation on ultrasound-guided percutaneous biopsy of lesions in the first hepatic hilum with fusion of ultrasound and multimodal imaging cognitive guidance**  
Xian-Tao Zeng, Xia Liang, Zhi-Liang Hong, Sheng Chen, Jian-Chuan Yang, Yu-cheng Lin and Song-Song Wu
- 134 **MAMILNet: advancing precision oncology with multi-scale attentional multi-instance learning for whole slide image analysis**  
Qinqing Wang, Qiu Bi, Linhao Qu, Yuchen Deng, Xianhong Wang, Yijun Zheng, Chenrong Li, Qingyin Meng and Kun Miao
- 144 **Leveraging radiomics and AI for precision diagnosis and prognostication of liver malignancies**  
Maryam Haghshomar, Darren Rodrigues, Aparna Kalyan, Yury Velichko and Amir Borhani
- 155 **Deep learning driven diagnosis of malignant soft tissue tumors based on dual-modal ultrasound images and clinical indexes**  
Haiqin Xie, Yudi Zhang, Licong Dong, Heng Lv, Xuechen Li, Chenyang Zhao, Yun Tian, Lu Xie, Wangjie Wu, Qi Yang, Li Liu, Desheng Sun, Li Qiu, Linlin Shen and Yusen Zhang
- 166 **Radiogenomics nomogram based on MRI and microRNAs to predict microvascular invasion of hepatocellular carcinoma**  
Guangchao Hu, Jianyi Qu, Jie Gao, Yuqian Chen, Fang Wang, Haicheng Zhang, Han Zhang, Xuefeng Wang, Heng Ma, Haizhu Xie, Cong Xu, Naixuan Li and Qianqian Zhang
- 178 **Prediction of clear cell renal cell carcinoma  $\leq 4\text{cm}$ : visual assessment of ultrasound characteristics versus ultrasonographic radiomics analysis**  
Fan Yang, Dai Zhang, Li-Hui Zhao, Yi-Ran Mao, Jie Mu, Hai-Ling Wang, Liang Pang, Shi-Qiang Yang, Xi Wei and Chun-Wei Liu
- 190 **PI-YOLO: dynamic sparse attention and lightweight convolutional based YOLO for vessel detection in pathological images**  
Cong Li, Shuanlong Che, Haotian Gong, Youde Ding, Yizhou Luo, Jianing Xi, Ling Qi and Guiying Zhang
- 205 **2.5D peritumoural radiomics predicts postoperative recurrence in stage I lung adenocarcinoma**  
Haimei Lan, Chaosheng Wei, Fengming Xu, Eqing Yang, Dayu Lu, Qing Feng and Tao Li
- 218 **Differentiation of pheochromocytoma and adrenal lipoid adenoma by radiomics: are enhanced CT scanning images necessary?**  
Shi he Liu, Pei Nie, Shun li Liu, Dapeng Hao, Juntao Zhang, Rui Sun, Zhi tao Yang, Chuan yu Zhang and Qing Fu
- 227 **Radiomics-based machine learning models for differentiating pathological subtypes in cervical cancer: a multicenter study**  
Huiling Liu, Mi Lao, Yalin Zhang, Cheng Chang, Yong Yin and Ruozheng Wang



## OPEN ACCESS

EDITED AND REVIEWED BY  
Zaver Bhujwalla,  
Johns Hopkins Medicine, United States

## \*CORRESPONDENCE

Yuxiang Zhou  
✉ zhou.yuxiang@mayo.edu  
Zhimin Li  
✉ zhimin.li@nm.org  
Poonam Yadav  
✉ poonam.yadav@nm.org

RECEIVED 19 December 2024  
ACCEPTED 10 January 2025  
PUBLISHED 06 February 2025

## CITATION

Zhou Y, Li Z and Yadav P (2025)  
Editorial: Artificial intelligence and  
imaging for oncology.  
*Front. Oncol.* 15:1547968.  
doi: 10.3389/fonc.2025.1547968

## COPYRIGHT

© 2025 Zhou, Li and Yadav. 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: Artificial intelligence and imaging for oncology

Yuxiang Zhou<sup>1\*</sup>, Zhimin Li<sup>2\*</sup> and Poonam Yadav<sup>3\*</sup>

<sup>1</sup>Department of Radiology, Mayo Clinic at Arizona, Phoenix, AZ, United States, <sup>2</sup>Department of Radiology, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States, <sup>3</sup>Department of Radiation Oncology, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

## KEYWORDS

artificial intelligence, CT, MRI, ultrasound, PET, pathological image, deep learning, radiomics

## Editorial on the Research Topic

### Artificial intelligence and imaging for oncology

## Introduction

Recent advancements in artificial intelligence (AI) and imaging technologies have significantly transformed the diagnostic and therapeutic landscapes of oncology (1–3). Cutting-edge imaging modalities, such as CT, PET, US, and MRI, are being increasingly utilized for tumor imaging (4–7), with emerging interdisciplinary fields like MR-LINAC gaining considerable traction (8, 9). This accelerating convergence of imaging and therapy in oncology highlights the urgent need to further explore the role of AI and imaging across various oncology specialties, including radiation therapy, to enhance cancer care. In response to this need, the topic titled “Artificial Intelligence and Imaging in Oncology” has been proposed, bringing together 19 contributions from 149 authors/experts in the field. These contributions delve into the potential of AI and imaging in tumor diagnosis and treatment, explore emerging AI-driven models for oncology diagnosis and prediction, and highlight the extraction of quantitative features from medical images to predict tumor behavior, therapy response, and patient prognosis.

## AI and imaging in tumor diagnosis and treatment

AI is revolutionizing cancer diagnosis and treatment by enhancing the accuracy and efficiency of medical image analysis. By analyzing medical images like CT scans, MRIs, and X-rays, AI algorithms can detect tumors earlier, differentiate between benign and malignant growths, and assist in treatment planning and monitoring.

Shao et al. demonstrated the potential of radiomics-based nomograms in enhancing the diagnostic capabilities of CT imaging. By extracting quantitative features from CT images, these nomograms can more accurately differentiate between conditions like intravenous leiomyomatosis and uterine leiomyoma, offering a significant clinical advantage over traditional CT image interpretation.

Zeng et al. explored the potential of fusing multimodal imaging with ultrasound to enhance the accuracy of interventional diagnostic procedures. By integrating machine

learning techniques, they demonstrated the clinical utility of this approach in guiding percutaneous biopsies of liver and adjacent organs, leading to improved diagnostic success rates.

Yu et al. showed that UNet based deep learning models when applied to positional CT and CBCT images and extracted radiomics features show clinical significance of CBCT images. The work showed that dice coefficient results of CBCT are within 85% of the results of pCT for rectal cancer imaging. CBCT images are frequently utilized on radiation treatment modalities.

Yang et al. explored the potential of combining ultrasound imaging with radiomics analysis to differentiate small clear cell renal cell carcinoma (ccRCC) from renal angiomyolipoma (RAML). By developing and validating models that incorporate both clinical and radiomic features, the study seeks to enhance diagnostic accuracy and support more precise treatment decisions for patients with small renal tumors. The findings suggest that this innovative approach could significantly improve the clinical utility of ultrasound in managing renal neoplasms.

Wen et al. explored an innovative approach to differentiate benign from malignant head and neck tumors using synthetic MRI in conjunction with FSE-PROPELLER DWI. In their study, the authors employed both synthetic MRI and FSE-PROPELLER diffusion-weighted imaging (DWI) to investigate the characteristics of malignant and benign head and neck tumors. The study involved 48 subjects, who were retrospectively classified into malignant and benign groups. The results were promising, demonstrating that both synthetic MRI and FSE-PROPELLER DWI can quantitatively distinguish malignant from benign tumors based on T2 and ADC values. Notably, combining T2 and ADC values provided improved accuracy in tumor differentiation.

Liu et al. focused on the differential diagnosis of two common adrenal tumors that are often misdiagnosed in clinical practice. Their research utilized radiomics techniques, enhancing diagnostic accuracy without the need for enhanced CT scans.

Haghshomar et al. reviewed recent advancements in the application of artificial intelligence (AI) in liver oncology imaging. They specifically highlighted the evolution of manual radiomic techniques and the increasing use of deep learning-based representations for more accurate assessments. They demonstrated radiomics, a framework that complements conventional radiological interpretation, has emerged as a powerful tool for extracting and quantifying texture characteristics derived from tumor heterogeneity.

## Emerging AI-driven models for oncology diagnosis and prediction

Emerging AI-driven models are revolutionizing oncology by enabling earlier and more accurate cancer diagnosis. By analyzing vast amounts of medical data, these models can identify subtle patterns and predict disease progression, leading to more personalized and effective treatment plans.

Xie et al., conducted a study to establish this deep learning (DL) driven Artificial intelligence (AI) system for predicting

malignant STTs based on US images and clinical indexes of the patients. The AI system could extract more morphological features of the system and heatmaps of images for classifying malignant soft tissue tumors. The system utilized a ResNet based architect on both grey scale and color ultrasound images for tumor feature extraction. The model can assist clinicians in diagnosing soft tissue tumors.

Ullah et al. studied brain tumor on MRI images diligently incorporating linear stretching in contrast enhancement and data augmented images fed to variants of efficient Net and Inception ResNet. The study utilized bayesian optimization on their deep learning process and showed an accuracy improvement over limited clinical dataset for brain tumor classification. The study showed that cubic SVM can increase accuracy by 0.5% over a bilayered neural network.

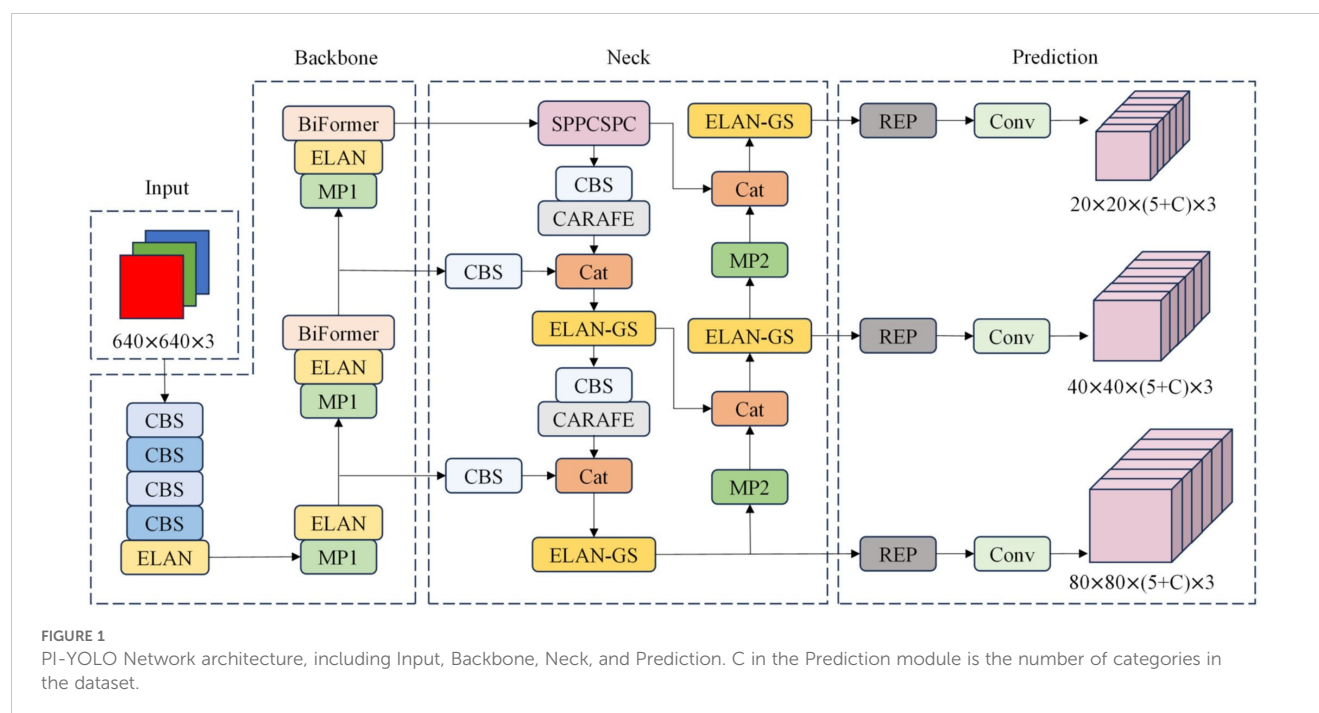
Wang et al. studied performance of MAMIL Net by histologic features in predicting breast cancer in sentinel lymph node, differentiating lung adenocarcinoma from squamous cell carcinoma, and predicting therapeutic response of high-grade ovarian serous carcinoma by retrospective case series. They found that MAMILNet performed excellent for lung cancer, good for breast cancer and fair for ovarian cancer based on AUC and accuracy values, suggesting that this learning framework has the potential in disease diagnosis and prognosis.

Li et al. developed PI-YOLO, a novel deep learning model designed for automated blood vessel detection in pathology images. This model effectively addresses the challenges posed by complex backgrounds, small targets, and dense distributions in these images. By incorporating the BiFormer attention mechanism, PI-YOLO efficiently captures long-range dependencies and reduces computational costs. Additionally, the use of GSConv convolution further enhances the model's performance by reducing parameters and improving inference speed. The results demonstrate that PI-YOLO achieves a significant mAP of 87.48%, outperforming existing methods. This advancement in automated blood vessel detection holds significant medical value, particularly in the field of anti-tumor vascular therapy research. Figure 1 showed a typical network framework including four main components: Input, Backbone, Neck, and Prediction.

Krishnapriya and Karuna performed a study to show that deep learning-based YOLO architecture can predict bounding boxes for prediction and have added enhancements compared to its peers for analogous inference tasks. The grab cut algorithm assisted segmentation is likely to improve dice coefficients by 0.1 in the presented dataset and is worth exploring in brain tumor detection pipelines.

Awais et al. presented a novel decision support system for identifying acute lymphoblastic leukemia (ALL). By combining techniques like neighborhood pixel transformation, transfer learning from deep neural networks, and a customized binary Grey Wolf Algorithm for feature optimization, the system achieves outstanding accuracy in both binary and subtype classification of ALL. This approach holds great promise in aiding medical professionals in the early and precise diagnosis of this aggressive leukemia, leading to better patient outcomes.





Houssein et al. investigated a new and efficient deep learning technique for classifying white blood cells (WBCs) in blood smear images, crucial for diagnosing leukemia. Leveraging DenseNet-161 and optimizing the training process with cyclical learning rates, the method achieves exceptional accuracy in classifying various WBC types, surpassing current state-of-the-art techniques. This innovative approach has great potential to aid medical professionals in the early and accurate diagnosis of leukemia, ultimately improving patient outcomes.

## Radiomics and radiogenomics for predicting tumor behavior, treatment response, and patient outcomes

Radiomics and radiogenomics are emerging fields that extract quantitative features from medical images to predict tumor behavior, treatment response, and patient outcomes. By analyzing these features, clinicians can make more informed decisions about treatment strategies and monitor disease progression.

Lan et al., Radiomics has shown promising applicability in cancer prediction, especially in recurrence. Lan et al. utilized ROIs delineated on CT images for extracting over 1100 radiomic features. To incorporate post-surgical data they used ten features based on relevance. This work shows employing clinical data over imaging parameters can be effectively used for predicting stage 1 lung adenocarcinoma prediction.

Mao et al. explored a novel radiomic nomogram that effectively differentiates parotid pleomorphic adenoma (PA) from adenolymphoma (AL) using grayscale ultrasonography. By combining advanced image analysis techniques with machine learning algorithms, this non-invasive nomogram provides a

highly accurate method for distinguishing between these two common parotid gland tumors. This innovative approach has the potential to greatly enhance diagnostic precision and guide more effective treatment planning for patients with parotid gland lesions.

Liu et al., demonstrated the potential of radiomics-based machine learning models using 18F-FDG PET/CT imaging data to distinguish between adenocarcinoma and squamous cell carcinoma in cervical cancer. By extracting and analyzing numerous quantitative features from medical images, these models offer valuable insights into tumor biology and assist in personalized treatment planning. The study highlights the promising role of radiomics in improving the diagnosis and management of cervical cancer.

Hu et al. introduced an innovative approach for predicting microvascular invasion (MVI) in hepatocellular carcinoma (HCC), a critical factor influencing the disease's aggressiveness. By integrating MRI imaging data with microRNA analysis, the researchers developed a radiogenomics nomogram that significantly outperforms existing models. This tool offers a promising path for more accurate risk assessment and personalized treatment strategies for HCC patients. With its high sensitivity and specificity, the nomogram shows great potential in improving clinical decision-making and enhancing patient outcomes.

Hu et al., explored a novel approach to testicular tumor diagnosis using computed tomography (CT) texture analysis (CTTA). This technique involves analyzing the texture patterns within CT images to identify subtle differences between benign and malignant tumors. By extracting specific texture features, researchers were able to develop machine learning models that can accurately classify tumors with high precision. One of the most promising findings of this study is the ability of CTTA to

differentiate between primary testicular lymphoma and other malignant tumors. This distinction is particularly important as it can influence treatment strategies. Additionally, CTTA can help identify seminoma, the most common type of testicular germ cell tumor, from other types of germ cell tumors.

## Summary

This Research Topic explores the transformative role of artificial intelligence (AI) and imaging advancements in oncology, focusing on how these technologies are reshaping the field. The articles highlight the growing integration of AI and imaging across various oncology specialties, demonstrating their potential to revolutionize cancer diagnosis, treatment planning, and prognostication. By leveraging cutting-edge imaging modalities, such as CT, PET, US, and MRI, along with AI-driven models, these innovations are improving the accuracy of tumor detection, enabling personalized treatment strategies, and predicting patient outcomes with greater precision. The Research Topic emphasizes the need for continued research and development in these areas, with the promise of enhancing patient care and outcomes across diverse cancer types.

## References

1. Lotter W, Hassett MJ, Schultz N, Kehl KL, Van Allen EM, Cerami E. Artificial intelligence in oncology: current landscape, challenges, and future directions. *Cancer Discovery*. (2024) 14:711–26. doi: 10.1158/2159-8290.CD-23-1199
2. Amisha Malik P, Pathania M, Rathaur VK. Overview of artificial intelligence in medicine. *J Family Med Prim Care*. (2019) 8:2328–31. doi: 10.4103/jfmpc.jfmpc\_440\_19
3. Lococo F, Ghaly G, Flamini S, Campanella A, Chiappetta M, Bria E, et al. Artificial intelligence applications in personalizing lung cancer management: state of the art and future perspectives. *J Thorac Dis*. (2024) 16:7096–110. doi: 10.21037/jtd-24-244
4. Loy LM, Aftab S, Ang YLS, Ding SLC, Ho CL. Intracranial “white” epidermoid tumours - An imaging approach and systematic review. *Clin Imaging*. (2022) 91:1–8. doi: 10.1016/j.clinimag.2022.08.002
5. Paudyal R, Shah AD, Akin O, Do RKG, Konar AS, Hatzoglou V, et al. Artificial intelligence in CT and MR imaging for oncological applications. *Cancers*. (2023) 15:2573. doi: 10.3390/cancers15092573
6. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med*. (2019) 17:195. doi: 10.1186/s12916-019-1426-2
7. Koh D-M, Papanikolaou N, Bick U, Illing R, Kahn CE, Kalpathi-Cramer J, et al. Artificial intelligence and machine learning in cancer imaging. *Commun Med*. (2022) 2:133. doi: 10.1038/s43856-022-00199-0
8. Liu X, Li Z, Yin Y. Clinical application of MR-Linac in tumor radiotherapy: a systematic review. *Radiat Oncol*. (2023) 18:52. doi: 10.1186/s13014-023-02221-8
9. Nachbar M, Lo Russo M, Gani C, Boeke S, Wegener D, Paulsen F, et al. Automatic AI-based contouring of prostate MRI for online adaptive radiotherapy. *Z Med Phys*. (2024) 34:197–207. doi: 10.1016/j.zemedi.2023.05.001

## Author contributions

YZ: Writing – original draft, Writing – review & editing. ZL: Writing – original draft, Writing – review & editing. PY: Writing – original draft, Writing – review & editing.

## 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.



## OPEN ACCESS

## EDITED BY

Poonam Yadav,  
Northwestern University, United States

## REVIEWED BY

Subathra Adithan,  
Jawaharlal Institute of Postgraduate  
Medical Education and Research (JIPMER),  
India  
Sikandar Shaikh,  
Shadan Hospital and Institute of Medical  
Sciences, India

## \*CORRESPONDENCE

Bao Liu  
✉ liubao72@aliyun.com

†These authors have contributed equally to  
this work

RECEIVED 12 June 2023

ACCEPTED 31 July 2023

PUBLISHED 23 August 2023

## CITATION

Shao J, Wang C, Shu K, Zhou Y, Cheng N,  
Lai Z, Li K, Xu L, Chen J, Du F, Yu X, Zhu Z,  
Wang J, Feng Y, Yang Y, Liu X, Yuan J and  
Liu B (2023) A contrast-enhanced CT-  
based radiomic nomogram for the  
differential diagnosis of intravenous  
leiomyomatosis and uterine leiomyoma.  
*Front. Oncol.* 13:1239124.  
doi: 10.3389/fonc.2023.1239124

## COPYRIGHT

© 2023 Shao, Wang, Shu, Zhou, Cheng, Lai,  
Li, Xu, Chen, Du, Yu, Zhu, Wang, Feng, Yang,  
Liu, Yuan and Liu. This is an open-access  
article distributed under the terms of the  
[Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/)  
(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.

# A contrast-enhanced CT-based radiomic nomogram for the differential diagnosis of intravenous leiomyomatosis and uterine leiomyoma

Jiang Shao<sup>1†</sup>, Chaonan Wang<sup>1,2†</sup>, Keqiang Shu<sup>1†</sup>, Yan Zhou<sup>1,3</sup>,  
Ninghai Cheng<sup>4</sup>, Zhichao Lai<sup>1</sup>, Kang Li<sup>1</sup>, Leyin Xu<sup>1</sup>,  
Junye Chen<sup>1,5</sup>, Fenghe Du<sup>1,6</sup>, Xiaoxi Yu<sup>1,3</sup>, Zhan Zhu<sup>1,3</sup>,  
Jiaxian Wang<sup>1,3</sup>, Yuyao Feng<sup>1</sup>, Yixuan Yang<sup>1,3</sup>, Xiaolong Liu<sup>1</sup>,  
Jinghui Yuan<sup>1</sup> and Bao Liu<sup>1\*</sup>

<sup>1</sup>Department of Vascular Surgery, Peking Union Medical College Hospital, Peking Union Medical College, Chinese Academy of Medical Science, Beijing, China, <sup>2</sup>Plastic Surgery Hospital, Chinese Academy of Medical Science, Peking Union Medical College, Beijing, China, <sup>3</sup>Eight-year Program of Clinical Medicine, Peking Union Medical College Hospital, Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing, China, <sup>4</sup>National Clinical Research Center for Obstetric & Gynecologic Diseases, Department of Obstetrics and Gynecology, Peking Union Medical College Hospital, Peking Union Medical College, Chinese Academy of Medical Sciences, Beijing, China, <sup>5</sup>State Key Laboratory of Medical Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, Department of Pathophysiology, Peking Union Medical College, Beijing, China, <sup>6</sup>Peking Union Medical College, MD Program, Beijing, China

**Objective:** Uterine intravenous leiomyomatosis (IVL) is a rare and unique leiomyoma that is difficult to surgery due to its ability to extend into intra- and extra-uterine vasculature. And it is difficult to differentiate from uterine leiomyoma (LM) by conventional CT scanning, which results in a large number of missed diagnoses. This study aimed to evaluate the utility of a contrast-enhanced CT-based radiomic nomogram for preoperative differentiation of IVL and LM.

**Methods:** 124 patients (37 IVL and 87 LM) were retrospectively enrolled in the study. Radiomic features were extracted from contrast-enhanced CT before surgery. Clinical, radiomic, and combined models were developed using LightGBM (Light Gradient Boosting Machine) algorithm to differentiate IVL and LM. The clinical and radiomic signatures were integrated into a nomogram. The diagnostic performance of the models was evaluated using the area under the curve (AUC) and decision curve analysis (DCA).

**Results:** Clinical factors, such as symptoms, menopausal status, age, and selected imaging features, were found to have significant correlations with the differential diagnosis of IVL and LM. A total of 108 radiomic features were extracted from contrast-enhanced CT images and selected for analysis. 29 radiomics features were selected to establish the Rad-score. A clinical model was developed to discriminate IVL and LM (AUC=0.826). Radiomic models were used to effectively differentiate IVL and LM (AUC=0.980). This radiological

nomogram combined the Rad-score with independent clinical factors showed better differentiation efficiency than the clinical model (AUC=0.985,  $p=0.046$ ).

**Conclusion:** This study provides evidence for the utility of a radiomic nomogram integrating clinical and radiomic signatures for differentiating IVL and LM with improved diagnostic accuracy. The nomogram may be useful in clinical decision-making and provide recommendations for clinical treatment.

#### KEYWORDS

intravenous leiomyomatosis, contrast-enhanced CT, radiomics, preoperative differential, nomogram

## 1 Introduction

Intravenous leiomyomatosis (IVL) is a rare benign type of uterine leiomyoma. Although histologically benign, it can spread to the extrauterine venous system or even the heart and pulmonary arterial system (1, 2). The current information on IVL mainly comes from case reports and case series, and its clinical presentation is nonspecific and may lead to right heart obstruction, pulmonary embolism and even sudden death (3). The development of IVL is insidious, and the clinical symptoms and pathological imaging features lack specificity and can cause serious consequences, especially in patients presenting with cardiac symptoms.

In addition, the pathological presentation of IVL is the same as that of common uterine leiomyoma (LM), and it may be difficult for pathologists to distinguish it from LM in patients with primary LM combined with IVL, especially if the lesions are confined to the uterus without invasion of the extrauterine veins. Some patients were only diagnosed with IVL after a previous hysterectomy to remove a primary uterine tumor. Pathological tissue findings of invasion of the parauterine veins may be a marker for IVL diagnosis. As a result, IVL is often underestimated due to the ease of misdiagnosis and the lack of specific identifying biomarkers.

The imaging presentation of IVL depends on the location and extent of its involvement. Typical imaging methods for the diagnosis of IVL include ultrasonography, computed tomography (CT) and magnetic resonance imaging (MRI). When a mass is confined to the pelvis, it is difficult to completely distinguish between IVL and LM on the basis of traditional radiology alone unless it has invaded the extrauterine vessels and is growing invasively (4). Radiomics refers to quantitative methods of extracting image features from conventional

radiographic images and analyzing the data to create models with features to aid in diagnosis, prediction and prognosis (5). Previous studies have demonstrated the value of radiomic features as imaging predictors that can be used to treat and diagnose various types of tumors (6). A study applied a radiomic model generated from features extracted from the region of interest covering the uterus with good diagnostic performance for uterine sarcomas and leiomyomas (7). However, no research has been performed to determine whether contrast-enhanced CT-based radiomics can be used to differentiate IVL and LM.

Therefore, this study aimed to use radiomics features extracted from clinically acquired abdominal pelvic CT scans to predict whether LM patients have IVL features prior to treatment.

## 2 Materials and methods

### 2.1 Patients

The Peking Union Medical College Hospital (PUMCH) ethics committee approved the study and waived informed consent from the patients (No. JS-2964). We reviewed the PUMCH surgical database. Patients who underwent gynecologic surgery between January 2011 and December 2020 were pathologically confirmed to have IVL. The inclusion criteria were as follows: 1) surgically and pathologically confirmed IVL or LM; 2) abdominal pelvic contrast-enhanced CT within the 20 days prior to gynecologic surgery; and 3) no relevant treatment prior to CT examination. The exclusion criteria were as follows: 1) no pathological findings, 2) poor image quality or significant image artifacts affecting the visualization, 3) incomplete clinical data, 4) intravascular leiomyosarcoma, 5) and a lack of CT images. Patients with uterine LM were matched to those who underwent surgery for uterine neoplasms by BMI, risk factors, and CT tube voltage. Ultimately, CT results from 124 patients (37 IVL and 87 LM) were included in the study. Figure 1 shows the flow chart of patient enrollment.

### 2.2 CT scan protocol

Patients who underwent contrast-enhanced CT examinations of the abdomen and pelvis were examined using GE Discovery CT (GE

**Abbreviations:** AUC, Area under the curve; BMI, Body mass index; CT, Computed tomography; DCA, Decision curve analysis; ER, Estrogen receptor; ICC, Intraclass correlation coefficient; IVC, Inferior vena cava; IVL, Intravenous leiomyomatosis; LASSO, Least absolute shrinkage and selection operator; LM, Uterine leiomyoma; MRI, Magnetic resonance imaging; NPV, Negative predictive value; PR, Progesterone receptor; PPV, Positive predictive value; PUMCH, Peking Union Medical College Hospital; Rad-score, Radiomics score; RA, Right atrium; ROC, Receiver operating characteristic; ROI, Region of interest.

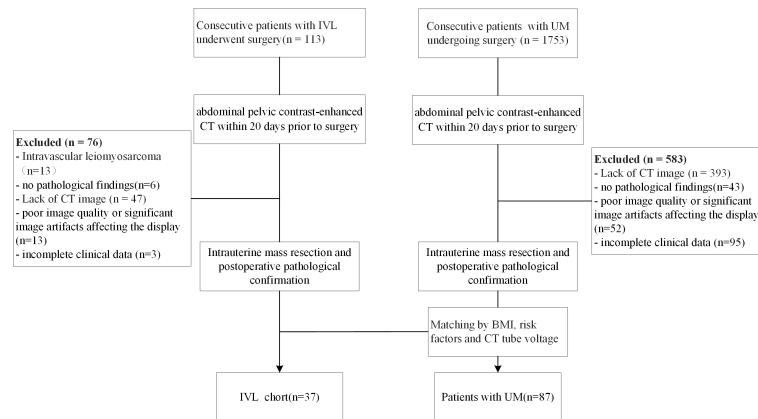


FIGURE 1

Flow chart demonstrating the inclusion and exclusion criteria for the study participants with IVL and UM. IVL, intravenous leiomyomatosis; UM, uterine leiomyoma; BMI, body mass index; CTA, computed tomography.

Medical, Piscataway, NJ, USA) or Somatom Definition Flash CT (Siemens Medical Solutions, Germany). All patients underwent thin-slice image scanning using a soft tissue algorithm, and CT images were obtained for the arterial (30 seconds postinjection), venous (60 seconds postinjection), and delayed (120 seconds postinjection) phases. The scanning parameters were as follows: tube voltage, 120 kV with automatic tube current modulation initiated; collimation, Somatom Definition Flash CT 128 × 0.6 mm, GE Discovery CT 64 × 0.6 mm; slice thickness, 0.625–1 mm; slice interval, 0.625–1 mm.

## 2.3 Image segmentation

The target of image segmentation is the intrauterine mass. When there were multiple masses in the uterus, the largest mass was chosen as the region of interest (ROI). Image segmentation was performed independently by two radiologists with extensive experience in gynecologic tumor imaging diagnosis. They were blinded to the patients' histopathology. One of the radiologists (radiologist A, with 7 years of experience in diagnostic imaging of gynecologic tumors) manually drew the ROI slice by slice using the open-source software 3D Slicer 4.11.0 (<https://www.slicer.org/>) (8). Another radiologist (radiologist B, with 10 years of experience in diagnostic imaging of gynecologic tumors) reviewed all ROIs manually segmented by radiologist A.

## 2.4 Data preprocessing

The dataset was randomly assigned in a 3:1 ratio to either the training dataset or test dataset. All cases in the training dataset were used to train the predictive model, while cases in the test dataset were used to independently evaluate the model's performance. Medical volumes are common with heterogeneous voxel spacing because of different scanners or different acquisition protocols. Such spacing refers to the physical distance between two pixels in an

image. Spatial normalization is often employed to reduce the effect of voxel spacing variation. The fixed resolution resampling method was used in our experiment to handle the aforementioned problems. All images were resampled to a voxel size of 3×3×3 mm to standardize the voxel spacing. Finally, the data were standardized using z score standardization (zero-mean normalization).

## 2.5 Radiomics feature extraction

The handcrafted features can be divided into three groups: (I) geometry, (II) intensity and (III) texture. The geometric features describe the three-dimensional shape characteristics of the tumor. The intensity features describe the first-order statistical distribution of the voxel intensities within the tumor. The texture features describe the patterns or the second- and high-order spatial distributions of the intensities. Here, the texture features were extracted using several different methods, including the gray-level cooccurrence matrix (GLCM), gray-level run length matrix (GLRLM), gray level size zone matrix (GLSZM) and neighborhood gray-tone difference matrix (NGTDM) methods. A total of 107 categories of handcrafted features were extracted, including 18 geometry features, 14 intensity features, and 75 texture features. All handcrafted features were extracted with an in-house feature analysis program implemented using Pyradiomics (<http://pyradiomics.readthedocs.io>).

## 2.6 Radiomics feature selection

### 2.6.1 Intraclass correlation coefficient

First, the robustness of the image features was evaluated. As the feature calculation depends on the ROI subregion contours, image features that are robust against ROI segmentation uncertainties were selected. Here, both test-retest analysis and interrater analysis were used to determine the feature robustness. Based on 35 patients randomly chosen from the discovery dataset, the test-retest analysis



was performed, where for each patient, the tumor subregions were segmented twice by one rater. The dataset used for interrater analysis included another 35 randomly chosen patients, where for each patient, the ROI subregions were segmented by two raters independently. The features extracted from these multiple-segmented subregions were assessed using the intraclass correlation coefficient (ICC). Features with an ICC of  $\geq 0.85$  were considered robust against intra- and interrater uncertainties. Pipeline of radiomics in Figure 2.

## 2.6.2 Spearman correlation

For features with high repeatability, Spearman's rank correlation coefficient was also used to calculate the correlation between features (Supplementary Figure 1 Spearman correlation of each feature), and one of the features with a correlation coefficient greater than 0.9 between any two features was retained. To retain the ability to depict features to the greatest extent, we use a stringent recursive deletion strategy for feature filtering; that is, the feature with the greatest redundancy in the current set is deleted each time.

## 2.6.3 LASSO and radiomics signature

The least absolute shrinkage and selection operator (LASSO) Cox regression model was used on the discovery dataset for signature construction. Depending on the regulation weight  $\lambda$ , LASSO shrinks all regression coefficients toward zero and sets the coefficients of many irrelevant features exactly to zero. To find an

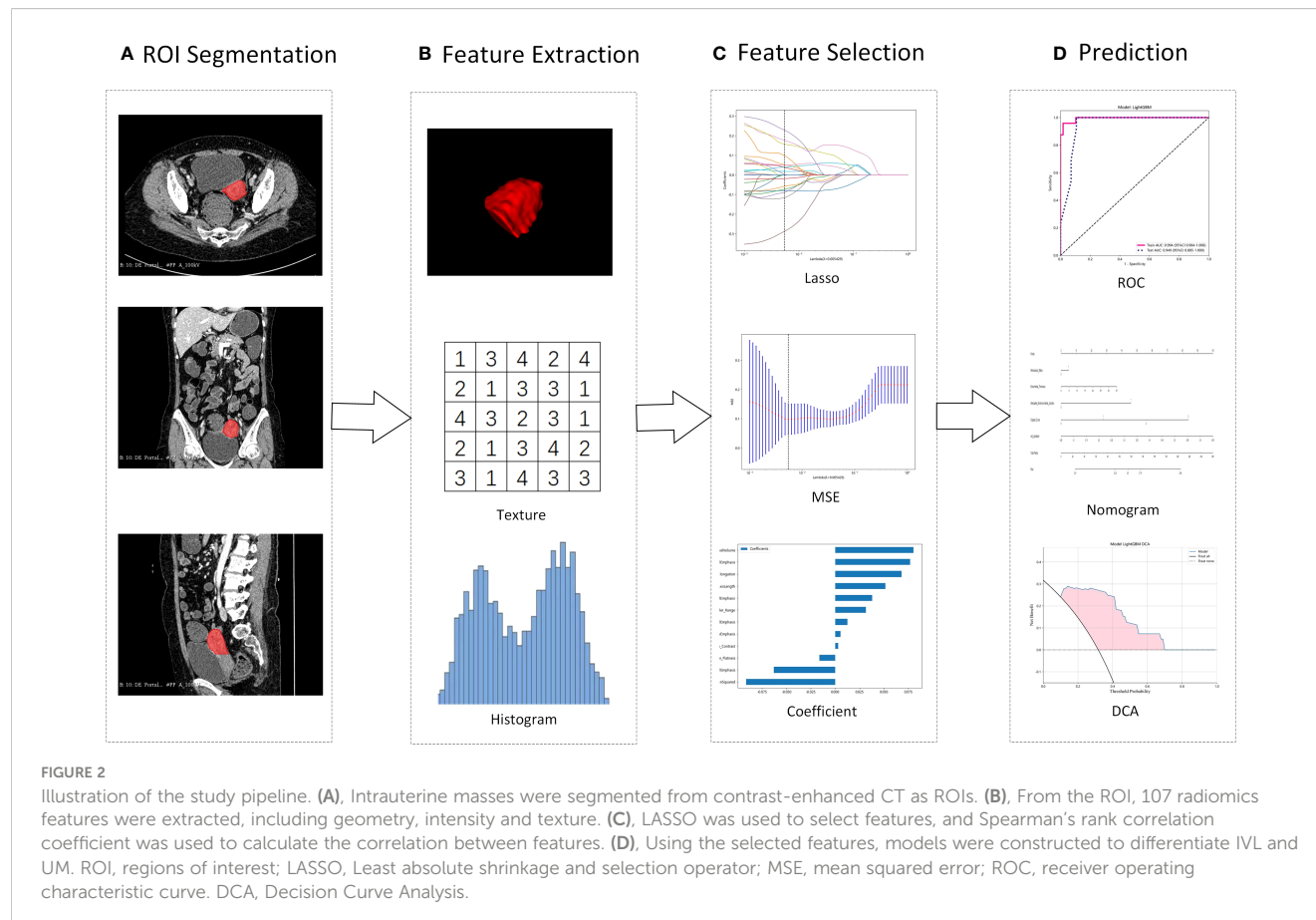
optimal  $\lambda$ , 10-fold cross validation with minimum criteria was employed, where the final value of  $\lambda$  yielded the minimum cross validation error (Figure 3). The retained features with nonzero coefficients were used for regression model fitting and combined into a radiomics signature. Subsequently, we obtained a radiomics score (Rad-score) for each patient by a linear combination of retained features weighed by their model coefficients. The Python scikit-learn package was used for LASSO regression modeling. The histogram of the Rad-score is shown in Figure 3.

## 2.7 Clinical factor model construction

Age, body mass index (BMI), weight, height, symptoms, reproductive history, menopausal history, estrogen receptor (ER) status, progesterone receptor (PR) status, and diabetes were selected as clinical factors for the IVL and LM groups and analyzed for differences between groups. The selected clinical factors were fed into the LightGBM model for clinical signature building.

## 2.8 Radiomics model construction

After Lasso feature screening, we input the final features into the LightGBM model for risk model construction. Here, we adopt 3-fold cross verification to obtain the final radiomics signature.



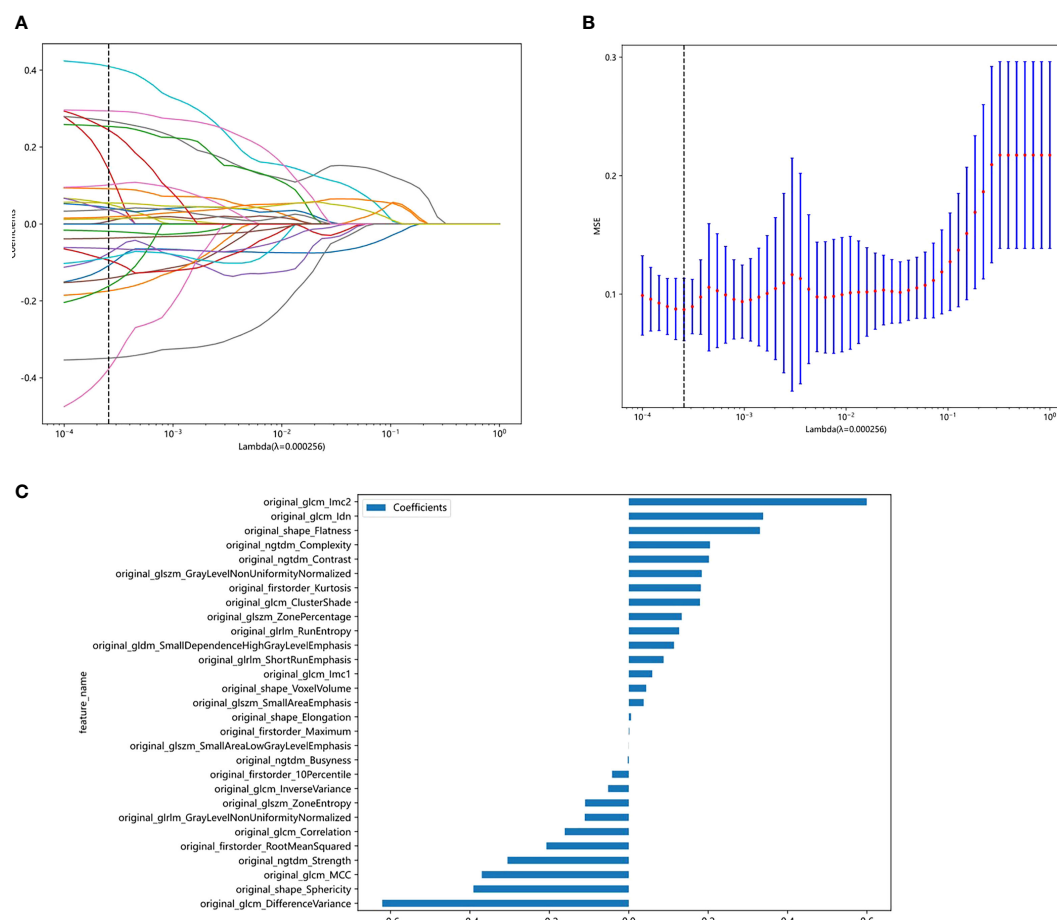


FIGURE 3

Figures of logistic LASSO regression. (A), Lasso path plot of the model in the training dataset. (B), Cross-validation plot for the penalty term. (C), Spearman correlation coefficients between features were calculated, and 27 features with correlations were retained.

Receiver operating characteristic (ROC) curves were plotted to assess the diagnostic performance of the predictive models, and the corresponding area under the curve (AUC), diagnostic accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were analyzed.

## 2.9 Construction of the nomogram

Furthermore, to assess the incremental prognostic value of the radiomics signature to the clinical risk factors intuitively and efficiently, a radiomics nomogram was presented on the validation dataset. The nomogram combined the radiomics signature and the clinical risk factors based on logistic regression analysis. To compare the agreement between the IVL prediction of the nomogram and the actual observation, the calibration curve (Hosmer–Lemeshow H test) was calculated. The AUC was calculated simultaneously for the training and test groups to quantify the discriminability of the nomogram. The discriminability of the model was tested using the Delong test. Finally, decision curve analysis (DCA) was used to assess the clinical utility of this nomogram by quantifying the net benefit of the

training and test sets of the combined model at different threshold probabilities.

## 2.10 Statistical analysis

The Python statmodels (version 0.13.2) package was used to perform statistical analysis, and a  $p$  value  $< 0.05$  was considered statistically significant. We analyzed the differences between the IVL and uterine LM groups using Student's  $t$  test or Mann–Whitney  $U$  tests for continuous variables; the chi-square test or Fisher's exact test was applied for categorical variables.

## 3 Results

### 3.1 Patient characteristics

A total of 124 patients, including 37 IVL and 87 LM patients, were included in our study. Patients were divided into a training set (82 patients) and an independent test set (42 patients) based on treatment duration. A pathologist reviewed the pathological data.



All patients underwent surgical treatment; there were 18 (48.6%) patients with IVL and 44 (50.6%) patients with LM in the training group and 19 (51.4%) patients with IVL and 43 (49.4%) patients with LM in the test group. The characteristics of the patients in the cohort are shown in [Table 1](#). [Table 1](#) summarizes the patients' baseline characteristics and postoperative pathological findings in the training and validation sets. The comparison of BMI, weight, height, ER, PR, diabetes, and fertility history showed no significant difference between the two groups and within each group ( $p > 0.05$ ), ensuring a reasonable classification. Significant differences between the cohorts were found in symptoms, menopause history, age, mass size, hypertension, and history of surgery ( $p < 0.05$ ).

### 3.2 Feature selection and radiomics signature development

Features with an ICC of  $\geq 0.85$  were considered robust against intra- and interrater uncertainties. After robustness evaluation, 62 categories out of the initial 108 image features remained. Spearman correlation coefficients between features were calculated, and features with correlations were retained ([Supplementary Table 1](#)). As shown in [Figure 3](#), 29 features of nonzero coefficients were selected to establish the Rad-score with a LASSO logistic regression model ( $\lambda = 0.005429$ ). The formula used to calculate the Rad-score is described in the [Supplementary Materials](#) ([Supplementary Table 1](#)).

### 3.3 Clinical factor model

Analysis of differences between groups showed that symptoms, menopausal history, and age were independent clinical risk factors for IVL ([Table 1](#)). A clinical signature was composed of three factors selected, namely, symptoms, menopausal history, and age. In the training group, the AUC value of the radiomics model was 0.865

(95% CI 0.786–0.944); in the test group, the AUC value of the model was 0.826 (95% CI of 0.669–0.983) ([Table 2](#), [Figure 4A](#)).

### 3.4 Diagnostic performance of radiomics features

Our results show that the radiomics features have good predictive performance for both the training and test sets. The AUCs of the radiomics model were significantly larger than those of the clinical model in both the training dataset (AUC=0.998 95% CI: 0.995–1.000) and the validation dataset (AUC=0.98; 95% CI: 0.936–1.000) ([Table 2](#), [Figure 4A](#)).

### 3.5 Combined models and radiomics nomogram

A combined model was developed by integrating the Rad-score and clinical predictors. A good performance was shown for the combined nomogram model in both the training dataset (AUC = 0.999 95% CI: 0.998–1.000) and the validation dataset (AUC = 0.985; 95% CI: 0.951–1.000) ([Table 2](#), [Figure 4A](#)). The diagnostic accuracy, sensitivity, specificity, PPV, NPV, precision and recall of the three models are also demonstrated in [Table 2](#).

The calibration curve showed that the IVL predicted by the combined model was very close to the actual results in both datasets ([Figure 4B](#)). The DCA also revealed the improvement in the combined model in both datasets ([Figure 4C](#)). This showed that when the threshold probability was between 1% and 99%, the combined model was more beneficial than the Rad-score and clinical models.

We also developed a nomogram to visualize the model for the combination ([Figure 5](#)). In the nomogram, points for each variable can be added to the corresponding axis to determine the risk of IVL. A higher total score is associated with a greater risk of IVL.

TABLE 1 Demographic and clinical characteristics of study populations.

| Characteristic       | Total (n=124)     | IVL (n=37)        | UM (n=87)         | p-value |
|----------------------|-------------------|-------------------|-------------------|---------|
| Age                  | 46.51 $\pm$ 8.17  | 43.32 $\pm$ 8.09  | 47.87 $\pm$ 7.86  | 0.004   |
| BMI                  | 24.01 $\pm$ 3.95  | 25.01 $\pm$ 3.03  | 23.59 $\pm$ 4.22  | 0.069   |
| Weight (kg)          | 61.45 $\pm$ 10.93 | 63.82 $\pm$ 8.87  | 60.43 $\pm$ 11.59 | 0.114   |
| height (cm)          | 159.89 $\pm$ 5.27 | 159.68 $\pm$ 4.94 | 159.98 $\pm$ 5.43 | 0.768   |
| Symptoms             | 68 (0.5484)       | 35 (0.9459)       | 33 (0.3793)       | <0.001  |
| Reproductive history | 114 (0.9194)      | 34 (0.9189)       | 80 (0.9195)       | 0.991   |
| menopause            | 40 (0.3226)       | 5 (0.1351)        | 35 (0.4023)       | 0.003   |
| ER- Positive         | 100 (0.8065)      | 33 (0.8919)       | 67 (0.7701)       | 0.118   |
| PR- Positive         | 123 (0.9919)      | 36 (0.9730)       | 87 (1.0000)       | 0.126   |
| Hypertension         | 28 (0.2258)       | 4 (0.1081)        | 24 (0.2759)       | 0.041   |
| Diabetes             | 17 (0.1371)       | 5 (0.1351)        | 12 (0.1379)       | 0.967   |

BMI, body mass index; ER, estrogen receptor; PR, progesterone receptor.

TABLE 2 Main consequence of 3 models based on Light Gradient Boosting Machine(LightGBM)algorithm.

| Model     | Dataset | Accuracy | AUC   | 95% CI        | Sensitivity | Specificity | PPV      | NPV      | Precision | Recall   | F1       | Threshold |
|-----------|---------|----------|-------|---------------|-------------|-------------|----------|----------|-----------|----------|----------|-----------|
| Clinical  | Train   | 0.793478 | 0.865 | 0.786 – 0.944 | 0.964286    | 0.718750    | 0.600000 | 0.978723 | 0.600000  | 0.964286 | 0.739726 | 0.240545  |
|           | Test    | 0.806452 | 0.826 | 0.669 – 0.983 | 0.888889    | 0.772727    | 0.615385 | 0.944444 | 0.615385  | 0.888889 | 0.727273 | 0.268398  |
| Radiomics | Train   | 0.989247 | 0.998 | 0.995 – 1.000 | 0.964286    | 1.000000    | 1.000000 | 0.984848 | 1.000000  | 0.964286 | 0.981818 | 0.730348  |
|           | Test    | 0.967742 | 0.980 | 0.936 – 1.000 | 0.888889    | 0.772727    | 0.615385 | 0.944444 | 0.615385  | 0.888889 | 0.727273 | 0.268398  |
| Combined  | Train   | 0.999442 | 0.999 | 0.998 – 1.000 | 1.000000    | 0.984375    | 0.965517 | 1.000000 | 0.965517  | 1.000000 | 0.982456 | 0.232776  |
|           | Test    | 0.967742 | 0.985 | 0.951 – 1.000 | 0.888889    | 1.000000    | 1.000000 | 0.956522 | 1.000000  | 0.888889 | 0.941176 | 0.586624  |

AUC, Area Under Curve; 95% CI, 95% Confidence Interval; NPV, Negative predictive value; PR, Progesterone receptor; PPV, Positive predictive value; F1, F1 Score.

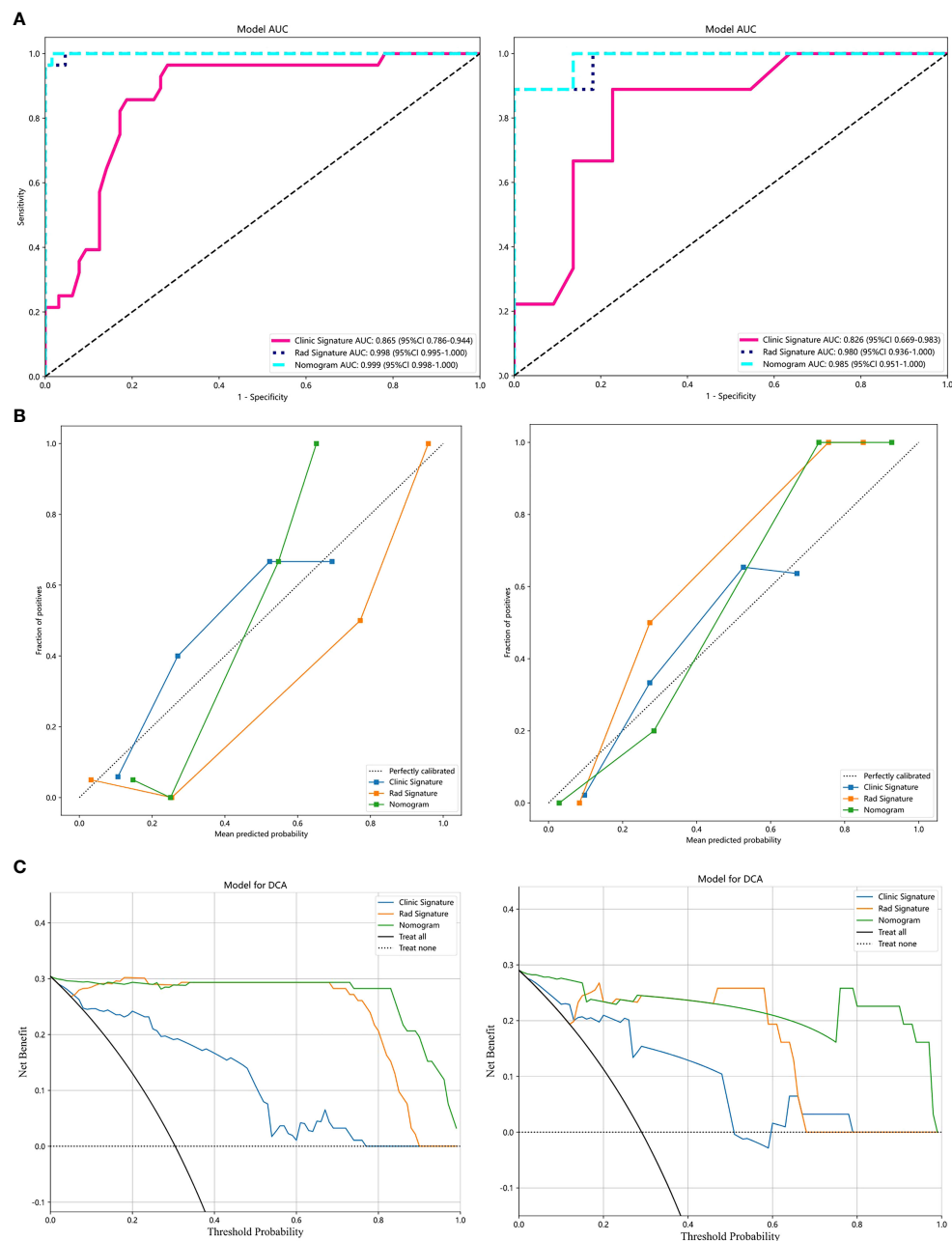
According to the DeLong test, the AUCs of the nomogram-based models in the training and test sets were significantly different from those of the clinical model ( $P=0.046$ ) (Supplementary Table 2). Therefore, we found that the nomogram method performed well on both sets of data. Furthermore, the Hosmer–Lemeshow test showed no statistically significant difference between the training and testing subsets ( $p>0.05$ ) (Table 3).

4 Discussion

In this retrospective study, we constructed for the first time a comprehensive model incorporating the Rad-score, symptoms, menopausal history and age and established a preoperative distinction between IVL and LM based on contrast-enhanced CT images.

The combined model consisting of radiomic features and clinical factors exhibited the best discriminatory ability and fit, indicating a good diagnostic performance. The AUC values of the model were 0.999 and 0.985 in the training and test groups, respectively.

LM is the most common uterine neoplasm in gynecology, with a prevalence of up to 20-30% in women of childbearing age. It has typical imaging features and clinical manifestations, and the radiological diagnosis of classic LM is definitive (9, 10). Occasionally, however, LM with rare growth patterns occurs, mostly in women of reproductive age, and IVL is one type of LM with an unusual growth pattern that presents as serpentine growth within the inferior vena cava (IVC) and genital veins and may spread to the right atrium (RA), making its identification clinically and radiologically more challenging (4). Worldwide, fewer than 300 cases of IVL and fewer than 100 cases of cardiac involvement have been reported. The imaging features of IVL are unclear and are often misdiagnosed preoperatively. It is mostly evaluated clinically using multimodal imaging techniques such as echocardiography, contrast-enhanced CT and MRI, which can provide important information revealing the extent and location of the mass and are used to determine surgical options (11). Echocardiography can assess the extension of the tumor into the RA, and CT and MRI can show the continuity of intraluminal tumor growth from the pelvic veins. It has been suggested that MRI is a particularly valuable imaging technique for the preoperative evaluation of IVL, which in the inferior vena cava looks similar to a sieve on axial images and to a sponge on T2-weighted images with several fissures parallel to the IVL, which may lead to turbulent blood flow (12–15). However, MRI has poor spatial resolution and is time-consuming and unsuitable for patients with metal in their bodies. Enhanced CT can produce multilevel enhanced CT data in a short period, directly displaying the full extent of the tumor, with a sponge and sieve appearance similar to MRI (16). In addition, the combined scan of the chest, abdomen and pelvis can clearly show the changes in the uterus and the extent of tumor invasion (17, 18). According to previous reports in the literature, radiologists are prone to misdiagnose IVL located in the venous system or RA as an occupying lesion, mainly leiomyosarcoma, RA myxoma, endometrial stromal sarcoma, and intravenous thrombosis (19,



**FIGURE 4**  
Results of the LightGBM models: **(A)**, Receiver operator characteristic curves of the 3 LightGBM models for identifying patients with IVL and uterine LM in the training and test datasets. **(B)**, The calibration curve of the 3 models. **(C)**, The decision curve analysis (DCA) of the three models of the training and test datasets.

20). However, these patients usually do not have a history of LM. It is almost impossible for radiologists to distinguish IVL confined to parauterine veins without distant venous system invasion, and in the early stages, LM is difficult to distinguish completely from IVL clinically and radiologically.

Previous studies did not find significant differences between IVL and LM in terms of histomorphology and immunophenotype, such as both expressing ER and PR and smooth muscle cell markers, and no elevated proliferation index or nuclear division number was found, suggesting that both have more of the same intrinsic molecular basis. Our data and recent reports suggest that IVL

accounts for approximately 1% of LM surgical specimens and its incidence is increasing. Some providers have an inadequate understanding of IVL, therefore, there are more missed diagnoses and its incidence is seriously underestimated (21–23). Some scholars compared the transcriptomic data of IVL and LM and found that antiapoptosis and angiogenesis-related genes may be novel biomarkers of IVL, indicating that IVL is very different from LM on a molecular and genetic basis. Further analysis of their gene expression profiles revealed that IVL and LM share some molecular genetic features and that IVL has a similar expression profile to leiomyosarcoma, further supporting that IVL has a quasi-malignant

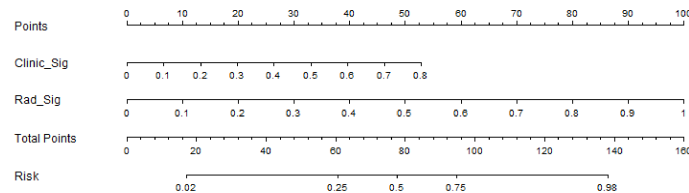


FIGURE 5  
Nomogram based on the combined model.

behavior and is not a distinct variant of LM (24, 25). However, these molecular genetic features are not independent predictors, and although they are associated with the occurrence of IVL, they do not distinguish IVL from LM.

By extracting high-dimensional imaging features from different modality images and mining the data, radiomics can be used for molecular typing of tumors, differential diagnosis, treatment option selection, efficacy detection and prognosis assessment (6).

These high-dimensional features are indistinguishable by the human eye and contain biological information determined by genes, proteins and tissue microcomponents, which radiomics can measure (26, 27). A radiomic model with features extracted from a ROI containing the whole uterus was shown to have good diagnostic performance for uterine fibroids and uterine sarcomas with an AUC of 0.83 (7). Some studies have used radiomic features to distinguish uterine sarcomas from atypical fibroids, showing better diagnostic efficacy than MRI features alone. Radiologists achieved an AUC of 0.752 for MRI-based diagnostic efficacy, and the radiomic model achieved an AUC of 0.830 (28). One study established an MRI-based radiomic nomogram for detecting deep myometrial invasion in early-stage endometrioid adenocarcinoma, showing superior diagnostic accuracy to radiologists, with an AUC of 0.883 (29). This suggests that radiomic methods can better predict and differentiate the type of uterine tumors compared to traditional clinical features. However, there are no relevant radiomic studies to better differentiate and distinguish uterine smooth muscle tumors with unusual growth patterns, which are often rare and require multiple imaging techniques to aid in the differential diagnosis.

In our study, the nomogram was constructed using the Radscore and contrast-enhanced CT with radiological methods. The Radscore is described as the probability of principal component analysis calculated from the radiomic signature, which is constructed based on sixteen selective radiomic features. The AUCs for predicting the radiomic features of IVL were 0.998 (training group) and 0.980 (test group). Nomograms constructed from radiological and clinical features show good discrimination

between IVL and LM. The AUC values of the training and test groups were 0.999 (95% CI: 0.998–1.000) and 0.985 (95% CI: 0.951–1.000), respectively. The results showed that the nomogram effectively predicted IVL in both the training and validation groups, exceeding the predictive accuracy of the radiomics and clinical models. The decision curve suggests that patients could benefit more from using the radiological nomogram in this study if they have a threshold probability of 1% to 99%. The combined model has better predictive performance than clinical factors or radiological features alone. The model is clear, simple, and easy to understand, which makes it more suitable for clinical application.

In the analysis of clinical factors between the IVL and LM groups, there were significant differences in age, symptoms, and menopausal history, so we introduced these factors into the clinical model and they demonstrated some predictive capacity. IVL often has no specific symptoms before causing cardiac insufficiency, and its clinical manifestations are usually related to the scope and size of the tumor (25). In clinical practice, we have found that IVL extending to the extrauterine venous system often accompanies large pelvic LM and causes related symptoms. However, only a tiny percentage of LMs develop at unusual locations beyond the uterus. All IVL cases occur in women, and the literature reports that the mean age of onset is 47 years; 90% are premenopausal, and 64% have uterine fibroids or a history of hysterectomy (30). The mean age of the cases in this group was 43.3 years; patients with a history of menopause in IVL were significantly younger than those in the LM group, which is similar to the literature.

This study still has some limitations. First, the sample size was relatively small, and it was a single-center study because the study population was a rare disease. Second, this study was retrospective, which may lead to patient selection bias. Third, manual ROI segmentation has inherent inter- and intra-observer differences. Fourth, we only built a radiomic model based on enhanced CT without using other imaging, so it is impossible to gauge the quality of each image. In the future, we will include more patients and make further technical improvements, such as fully automated image segmentation, deep learning and multiparametric modeling, to explore more accurate radiological diagnoses.

TABLE 3 Hosmer-Lemeshow test.

|     | Clinic Signature | Rad Signature | Nomogram |
|-----|------------------|---------------|----------|
| UM  | 0.099372         | 0.598583      | 0.046526 |
| IVL | 0.459798         | 0.913913      | 0.560273 |

Hosmer-Lemeshow test showed no statistically significant difference between the training and testing subsets ( $p > 0.05$ )

## 5 Conclusion

In conclusion, our study confirmed that a radiomics nomogram model and radiomics signature based on contrast-enhanced CT can help differentiate between IVL and LM patients and predict whether

IVL will invade the extrauterine vessels when it is still confined to the uterus to guide clinical treatment.

## Data availability statement

The datasets presented in this article are not readily available because this dataset is not publicly available due to patient privacy. Requests to access the datasets should be directed to Jiang Shao, esan2002@vip.sina.com.

## Ethics statement

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

All authors contributed significantly to this work; all authors revised the manuscript and gave final approval of the manuscript to be submitted. JS and BL: Conceptualization, Methodology. CW, KS and NC: Data curation, Writing-Original draft preparation. JC, FD, XY, and ZZ: Visualization, Investigation. ZL: Supervision. JW, YF, XL, JY, YY: Software, Validation. YZ, KS and BL: Writing-Reviewing and Editing.

## Funding

This study received funding from the National High Level Hospital Clinical Research Funding (2022-PUMCH-A-078), National Natural Science Foundation of China (NO. 82070498),

Non-profit Central Research Institute Fund of Chinese Academy of Medical Science (no. 2022-I2M-C&T-A-002), Beijing Science and Technology Innovation Foundation for University or College students (no:2023zglc06082) and Fundamental Research Funds for the Central Universities (no. 3332020009).

## Acknowledgments

We sincerely thank Platform Onekey AI and their developers' help with the code consultation of the study.

## 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/fonc.2023.1239124/full#supplementary-material>

## References

1. Dvorská D, Škovierová H, Braný D, Halašová E, Danková Z. Liquid biopsy as a tool for differentiation of leiomyomas and sarcomas of corpus uteri. *Int J Mol Sci* (2019) 20(15):3825. doi: 10.3390/ijms20153825
2. Li H, Xu J, Lin Q, Zhang Y, Zhao Y, Tong H, et al. Surgical treatment strategies for extra-pelvic intravenous leiomyomatosis. *Orphanet J Rare Dis* (2020) 15(1):153. doi: 10.1186/s13023-020-01394-9
3. Burke M, Opekin K. Death due to intravenous leiomyomatosis extending to the right pulmonary artery. *Pathology* (2004) 36(2):202–3. doi: 10.1080/00313020410001672075
4. Fasih N, Prasad Shanbhogue AK, Macdonald DB, Fraser-Hill MA, Papadatos D, Kiear AZ, et al. Leiomyomas beyond the uterus: unusual locations, rare manifestations. *Radiographics* (2008) 28(7):1931–48. doi: 10.1148/rg.287085095
5. Gardin I, Grégoire V, Gibon D, Kirišli H, Pasquier D, Thariat J, et al. Radiomics: Principles and radiotherapy applications. *Crit Rev Oncol Hematol* (2019) 138:44–50. doi: 10.1016/j.critrevonc.2019.03.015
6. Liu Z, Wang S, Dong D, Wei J, Fang C, Zhou X, et al. The applications of radiomics in precision diagnosis and treatment of oncology: opportunities and challenges. *Theranostics* (2019) 9(5):1303–22. doi: 10.7150/thno.30309
7. Xie H, Zhang X, Ma S, Liu Y, Wang X. Preoperative differentiation of uterine sarcoma from leiomyoma: comparison of three models based on different segmentation volumes using radiomics. *Mol Imaging Biol* (2019) 21(6):1157–64. doi: 10.1007/s11307-019-01332-7
8. van 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(21):e104–7. doi: 10.1158/0008-5472.CAN-17-0339
9. Szklaruk J, Tamm EP, Choi H, Varavithya V. MR imaging of common and uncommon large pelvic masses. *Radiographics* (2003) 23(2):403–24. doi: 10.1148/rg.232025089
10. Buttram VC, Reiter RC. Uterine leiomyomata: etiology, symptomatology, and management. *Fertil Steril* (1981) 36(4):433–45. doi: 10.1016/s0015-0282(16)45789-4
11. Xu ZF, Yong F, Chen YY, Pan AZ. Uterine intravenous leiomyomatosis with cardiac extension: Imaging characteristics and literature review. *World J Clin Oncol* (2013) 4(1):25–8. doi: 10.5306/wjco.v4.i1.25
12. Wray RC, Dawkins H. Primary smooth muscle tumors of the inferior vena cava. *Ann Surg* (1971) 174(6):1009–18. doi: 10.1097/0000658-197112000-00021
13. Jalaguier-Coudray A, Allain-Nicolai A, Thomassin-Piana J, Villard-Mahjoub R, Delarbre B, Rua S, et al. Radio-surgical and pathologic correlations of pelvic intravenous leiomyomatosis. *Abdom Radiol (NY)* (2017) 42(12):2927–32. doi: 10.1007/s00261-017-1225-1
14. Robert-Ebadi H, Terraz S, Mach N, Dubuisson JB, Kalangos A, Bounameaux H. Intravenous leiomyomatosis of the uterus: link with new fertilisation methods? *Swiss Med Wkly* (2009) 139(29–30):436. doi: 10.4414/smw.2009.12750

15. Kang LQ, Zhang B, Liu BG, Liu FH. Diagnosis of intravenous leiomyomatosis extending to heart with emphasis on magnetic resonance imaging. *Chin Med J (Engl)* (2012) 125(1):33–7. doi: 10.3760/cma.j.issn.0366-6999.2012.01.007
16. Wang H, Nie P, Chen B, Hou F, Dong C, He F, et al. Contrast-enhanced CT findings of intravenous leiomyomatosis. *Clin Radiol* (2018) 73(5):503.e1–6. doi: 10.1016/j.crad.2017.12.016
17. Foster BR, Anderson SW, Uyeda JW, Brooks JG, Soto JA. Integration of 64-detector lower extremity CT angiography into whole-body trauma imaging: feasibility and early experience. *Radiology* (2011) 261(3):787–95. doi: 10.1148/radiol.11100604
18. Peng HJ, Zhao B, Yao QW, Qi HT, Xu ZD, Liu C. Intravenous leiomyomatosis: CT findings. *Abdom Imaging* (2012) 37(4):628–31. doi: 10.1007/s00261-011-9798-6
19. Gui T, Qian Q, Cao D, Yang J, Peng P, Shen K. Computerized tomography angiography in preoperative assessment of intravenous leiomyomatosis extending to inferior vena cava and heart. *BMC Cancer* (2016) 16:73. doi: 10.1186/s12885-016-2112-9
20. Bilyeu SP, Bilyeu JD, Parthasarathy R. Intravenous lipoleiomyomatosis. *Clin Imaging* (2006) 30(5):361–4. doi: 10.1016/j.clinimag.2006.03.034
21. Wang J, Yang J, Huang H, Li Y, Miao Q, Lu X, et al. Management of intravenous leiomyomatosis with intracaval and intracardiac extension. *Obstet Gynecol* (2012) 120(6):1400–6. doi: 10.1097/AOG.0b013e31826ebb90
22. Ma G, Miao Q, Liu X, Zhang C, Liu J, Zheng Y, et al. Different surgical strategies of patients with intravenous leiomyomatosis. *Med (Baltimore)* (2016) 95(37):e4902. doi: 10.1097/MD.0000000000004902
23. Du J, Zhao X, Guo D, Li H, Sun B. Intravenous leiomyomatosis of the uterus: a clinicopathologic study of 18 cases, with emphasis on early diagnosis and appropriate treatment strategies. *Hum Pathol* (2011) 42(9):1240–6. doi: 10.1016/j.humpath.2010.10.015
24. Wang W, Wang Y, Chen F, Zhang M, Jia R, Liu X, et al. Intravenous leiomyomatosis is inclined to a solid entity different from uterine leiomyoma based on RNA-seq analysis with RT-qPCR validation. *Cancer Med* (2020) 9(13):4581–92. doi: 10.1002/cam4.3098
25. Ordulu Z, Nucci MR, Dal Cin P, Hollowell ML, Otis CN, Hornick JL, et al. Intravenous leiomyomatosis: an unusual intermediate between benign and Malignant uterine smooth muscle tumors. *Mod Pathol* (2016) 29(5):500–10. doi: 10.1038/modpathol.2016.36
26. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology* (2016) 278(2):563–77. doi: 10.1148/radiol.2015151169
27. Rogers W, Thulasi Seetha S, Refaee TAG, Lieverse RIY, Granzier RWY, Ibrahim A, et al. Radiomics: from qualitative to quantitative imaging. *Br J Radiol* (2020) 93(1108):20190948. doi: 10.1259/bjr.20190948
28. Xie H, Hu J, Zhang X, Ma S, Liu Y, Wang X. Preliminary utilization of radiomics in differentiating uterine sarcoma from atypical leiomyoma: Comparison on diagnostic efficacy of MRI features and radiomic features. *Eur J Radiol* (2019) 115:39–45. doi: 10.1016/j.ejrad.2019.04.004
29. Wang Y, Bi Q, Deng Y, Yang Z, Song Y, Wu Y, et al. Development and validation of an MRI-based radiomics nomogram for assessing deep myometrial invasion in early stage endometrial adenocarcinoma. *Acad Radiol* (2022) S1076-6332(22):00320–8. doi: 10.1016/j.acra.2022.05.017
30. Stegmann T, Garcia-Gallont R, Döring W. Intravascular leiomyomatosis: report of a case and review of the literature. *Thorac Cardiovasc Surg* (1987) 35(3):157–60. doi: 10.1055/s-2007-1020220





## OPEN ACCESS

## EDITED BY

Yuxiang Zhou,  
Mayo Clinic Arizona, United States

## REVIEWED BY

Palash Ghosal,  
Sikkim Manipal University, India  
Pradeep Kumar Das,  
VIT University, India

## \*CORRESPONDENCE

Nagwan Abdel Samee  
✉ nmabdelsamee@pnu.edu.sa

RECEIVED 28 May 2023

ACCEPTED 15 August 2023

PUBLISHED 12 September 2023

## CITATION

Houssein EH, Mohamed O, Abdel Samee N, Mahmoud NF, Talaat R, Al-Hejri AM and Al-Tam RM (2023) Using deep DenseNet with cyclical learning rate to classify leukocytes for leukemia identification. *Front. Oncol.* 13:1230434. doi: 10.3389/fonc.2023.1230434

## COPYRIGHT

© 2023 Houssein, Mohamed, Abdel Samee, Mahmoud, Talaat, Al-Hejri and Al-Tam. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

# Using deep DenseNet with cyclical learning rate to classify leukocytes for leukemia identification

Essam H. Houssein<sup>1</sup>, Osama Mohamed<sup>2</sup>,  
Nagwan Abdel Samee<sup>3\*</sup>, Noha F. Mahmoud<sup>4</sup>, Rawan Talaat<sup>5</sup>,  
Aymen M. Al-Hejri<sup>6</sup> and Riyadh M. Al-Tam<sup>6</sup>

<sup>1</sup>Faculty of Computers and Information, Minia University, Minia, Egypt, <sup>2</sup>Faculty of Computers and Artificial Intelligence, Beni-Suef University, Beni-Suef, Egypt, <sup>3</sup>Department of Information Technology, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia, <sup>4</sup>Rehabilitation Sciences Department, Health and Rehabilitation Sciences College, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia, <sup>5</sup>Biotechnology & Genetics Department, Agriculture Engineering, Ain Shams University, Cairo, Egypt, <sup>6</sup>School of Computational Sciences, Swami Ramanand Teerth Marathwada University, Nanded, Ma-harashtra, India

**Background:** The examination, counting, and classification of white blood cells (WBCs), also known as leukocytes, are essential processes in the diagnosis of many disorders, including leukemia, a kind of blood cancer characterized by the uncontrolled proliferation of carcinogenic leukocytes in the marrow of the bone. Blood smears can be chemically or microscopically studied to better understand hematological diseases and blood disorders. Detecting, identifying, and categorizing the many blood cell types are essential for disease diagnosis and therapy planning. A theoretical and practical issue. However, methods based on deep learning (DL) have greatly helped blood cell classification.

**Materials and Methods:** Images of blood cells in a microscopic smear were collected from GitHub, a public source that uses the MIT license. An end-to-end computer-aided diagnosis (CAD) system for leukocytes has been created and implemented as part of this study. The introduced system comprises image preprocessing and enhancement, image segmentation, feature extraction and selection, and WBC classification. By combining the DenseNet-161 and the cyclical learning rate (CLR), we contribute an approach that speeds up hyperparameter optimization. We also offer the one-cycle technique to rapidly optimize all hyperparameters of DL models to boost training performance.

**Results:** The dataset has been split into two sets: approximately 80% of the data (9,966 images) for the training set and 20% (2,487 images) for the validation set. The validation set has 623, 620, 620, and 624 eosinophil, lymphocyte, monocyte, and neutrophil images, whereas the training set has 2,497, 2,483, 2,487, and 2,499, respectively. The suggested method has 100% accuracy on the training set of images and 99.8% accuracy on the testing set.

**Conclusion:** Using a combination of the recently developed pretrained convolutional neural network (CNN), DenseNet, and the one fit cycle policy, this study describes a technique of training for the classification of WBCs for leukemia detection. The proposed method is more accurate compared to the state of the art.

#### KEYWORDS

leukemia, leukocytes, DenseNet, transfer learning, cyclical learning rate

## 1 Introduction

Medical images are a massive data source for the healthcare sector. With developments in imaging technology and processing capabilities, the demand for increasingly complex tools to interpret images has developed. More accurate image analysis will save healthcare costs and improve the quality of diagnosis, ultimately leading to better patient outcomes. Anemia, leukemia, and malaria are just a few of the blood disorders that can be detected with improved pathologists' ability to recognize, count, and classify blood cells (1–3). Improved understanding will facilitate treatment, reduce potentially dangerous drug interactions, and facilitate health monitoring. The three types of cells that make up human blood are the erythrocytes (red blood cells), leukocytes (white blood cells (WBCs)), and thrombocytes (platelets). All three are derived from lymphoid and bone marrow stem cells. Erythrocytes, which are non-nucleated biconcave diskettes, transport both carbon dioxide (CO<sub>2</sub>) and oxygen (O<sub>2</sub>) around the body. Blood is composed of roughly 40%–45% red blood cells and 1% WBCs (4–6). Organs in the body rely on each of the three types of blood cells for specific tasks. Nevertheless, WBCs are made in the bone marrow and are a crucial part of the blood's immune system. The immune system is the body's primary line of defense against invaders, most notably pathogens, and is mostly the work of WBCs (7).

Thrombocytes, often known as platelets, are smaller than erythrocytes and lack a nucleus. Giemsa staining produces a vivid purple tint in platelets (8). Platelets are crucial to the body's clotting process, which guards against bacterial invasion and keeps the body from bleeding out continuously following injuries (9). Leukocytes may be divided into five major types based on a variety of characteristics, including cell size, nucleus shape, type of nucleus lobes, granule cytoplasm-to-nucleus ratio (CNR) staining qualities, and function.

Lymphocytes, monocytes, neutrophils, eosinophils, and basophils are the five most common types of WBCs. Another thing is the band identification for a certain nucleus shape. [Figure 1](#) illustrates several common types of leukocytes. A decrease in leukocytes below the threshold is medically referred to as leukopenia. It is evidence of the frailty of the immune system and a potential reason for disease.

Leukocyte counts can be low for one of two major reasons: either the bone marrow has ceased producing leukocytes or an infection is present and causing cells to be destroyed more quickly than they can be replaced. Leukocytosis, a proliferative condition, is characterized by a rise of leukocytes over the upper limit, which is typically an indication of an inflammatory reaction. It occasionally takes place because of normal immunological responses. Nonetheless, if the neoplasm has an abnormally high or low cell count, or if autoimmunity causes immunological reactions, it will be classified as abnormal. Leukocyte disorders can also be classified in this fashion (10) based on the nature and function of affected cells. Hematologists can discover a great deal about blood diseases such as anemia, bleeding disorders, leukemia, and HIV positivity from a complete blood count (CBC) and differential blood count (DBC). The CBC can be performed automatically by a cytometer as blood flows past the detector, with parameters including hematocrit and hemoglobin measured (11). DBC, which may count the different types of leukocytes in peripheral blood, was previously performed by a blood pathologist physically inspecting blood smears under a light microscope. Nonetheless, this process is sensitive, and it is essential that there be no (or just very few) inspection errors made by the human professional. However, after several hours of examination, specialists might often feel exhausted and make false identifications of the various WBCs. This can happen rather frequently (3, 12). As a result of the development of both theoretical and practical applications for the technology that is available today, several different methods of blood analysis that are either fully or partially automated and are based on the image analysis of blood smears or the principles of flow cytochemistry have been developed. Image processing and artificial intelligence (AI) (13) have lately been used to develop several new methods that researchers have designed to automate the leukocyte classification process. Within the scope of this investigation, a fully automated computer-aided diagnosis (CAD) system of leukocytes has been developed and implemented. The proposed CAD system includes four primary stages, which are the image preprocessing and enhancement stage, the image segmentation stage, the feature extraction and selection stage, and the WBC classification stage.

The medical imaging industry makes extensive use of the recently developed and powerful pretrained convolutional neural network (CNN) DenseNet-161. However, compared to other

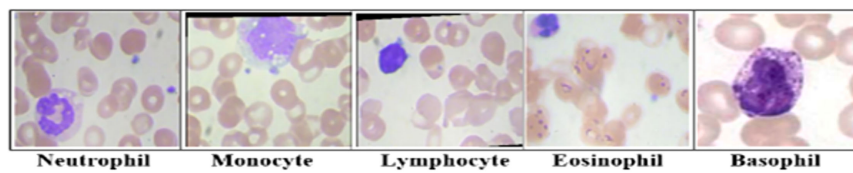


FIGURE 1

The main types of leukocytes cell images. Lymphocytes, Monocytes, Neutrophils, Eosinophils, and basophils are the five most common types of white blood cells. Each type has a certain nucleus shape.

pretrained CNNs, it has a high processing time and cannot generalize. Thus, we are exploring the one cycle policy (14, 15), a technique used to shorten training time while simultaneously enhancing performance and tuning all hyperparameters of deep learning (DL) models (15, 16). As can be shown in Figure 2, a cyclical learning rate (CLR) can produce better training results than the default learning rate (LR).

In contrast to blood cell segmentation algorithms that rely on watershed segmentation, this article presents a segmentation algorithm that uses Bounded Opening followed by Fast Radial Symmetry (BO-FRS)-based seed-point detection and hybrid Ellipse Fitting (EF)-based contour estimation. These methods accurately extract seed points and precisely segment overlapping cells, even from low-contrast inhomogeneous visual features. This makes the method suitable for complex blood cell segmentation problems. The proposed Least Squares (LS)-based geometric ellipse fitting approach leads to better accuracy (ACC) and more localization compared to algebraic Ellipse Fitting Methods (EFMs), which are prone to biased fitting parameters and inaccurate boundaries. The proposed method combines the benefits of geometric and algebraic EFMs and is computationally efficient. It also solves the noise problem with an Laplacian of Gaussian (LoG)-based modified high-boosting operation and avoids oversegmentation. This approach can also be applied to other medical applications such as MRI, CT, ultrasound, and X-ray images, as well as cybernetic applications and the segmentation of overlapping objects. Notably, the proposed algorithm does not require training data, making it more suitable than DL-based techniques when little or no data are available for training (17).

The following is a list of the contributions that were made to bring attention to the significance of the work that we will be presenting:

- We present an improved, lightweight, and effective CAD system that can automatically classify four types of leukocytes (neutrophils, eosinophils, lymphocytes, and monocytes), which is a significant contribution to the field of medical image analysis.
- We investigate the potential of DenseNet-161 pretrained CNN for the suggested CAD system, which is a modern approach to developing the system.
- The authors train the DenseNet efficiently with a single cycle policy, cutting down on epochs and iterations, and thereby making use of big datasets. This is a significant contribution to the field of DL, as it demonstrates a more efficient approach to training CNNs.
- The proposed model is tested experimentally on a variety of real-world datasets, which is a significant contribution to the field of medical image analysis, as it demonstrates the effectiveness of the model on a range of different datasets.
- The results of the study show that the proposed model outperforms the gold standard classification model, which is a significant contribution to the field of medical image analysis.
- The achieved ACC in categorization is approximately 99.8%, which is a significant contribution to the field of medical image analysis, as it demonstrates the high ACC of the proposed model.

The sections of this paper are as follows: Section 2 (Literature Review) details the related work. In Section 3 (Materials and Methods), we provide some the datasets and methods utilized for the proposed model. The analysis and results of the experiments are presented in Section 5 (Results and Discussion). In the end, the paper was concluded in Section 6 (Conclusion).

## 2 Literature review

Many attempts at automatically segmenting, categorizing, and analyzing leukocytes have been published. The automatic analysis of medical images such as microscopic blood smears has attracted the attention of many researchers. Numerous scientists have argued for employing machine learning (ML) and AI to automatically detect and diagnose abnormalities in microscopic images of leukocytes. CAD of leukocytes can be broken down into two categories: those that use ML (18) and those that use DL (19). Both ML and DL are described and summarized here. Table 1

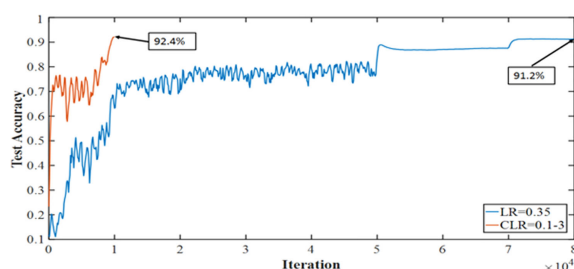


FIGURE 2

Accuracy achieved with one cycle of training against the conventional method of training Convolutional Neural Networks.

TABLE 1 Overview of research using DL techniques for leukocyte classification or segmentation.

| Author                      | Method                                 | Accuracy | Dataset          | Data volume |
|-----------------------------|----------------------------------------|----------|------------------|-------------|
| Maryam et al. (20)          | Optimized CNN                          | 99%      | BCCD             | 12,444      |
| Bani-Hani et al. (21)       | GA-optimized CNN                       | 91%      | BCCD             | 12,435      |
| Liang et al. (22)           | Hybrid CNN-RNN                         | 90.80%   | BCCD             | 12,444      |
| Rao (23)                    | CNN and ResNeXt                        | 99.24%   | BCCD             | 12,444      |
| Rao (24)                    | ANN and CNN                            | 97.70%   | BCCD             | 1,600       |
| Baydilli, and Atila (25)    | Capsul Networks                        | 96.90%   | LISC             | 263         |
| Ghosh and Bhattacharya (26) | CNN and FCN on noise-free cell images  | 98.40%   | BCCD             | 12,500      |
| Wang et al. (27)            | Single Shot Multibox Detector and YOLO | 90.09%   | Private database | 11,600      |
| Ma et al. (28)              | DC-GAN, and ResNet                     | 91.70%   | BCCD             | 12,447      |
| banik et al. (29)           | CNN                                    | 96%      | BCCD             | 12,811      |
| Sahlool et al. (30)         | VGGNet, CNN, and SESSA                 | 83.20%   | C-NMC            | 10,661      |
| Proposed model              | DenseNet with Cyclical Learning Rate   | 99.80%   | BCCD             | 12,447      |

CNN, Convolutional Neural Network; GA, Genetic Algorithms; ANN, Artificial Neural Network; FCN, Fully Connected Network.

provides a summary of DL-based methods and serves to contrast our proposed work with the state-of-the-art DL-based methods. Table 1 summarizes the current state of the field and the limitations of each technique based on recent studies that employed DL algorithms to identify abnormalities in leukocytes. The most noteworthy aspects of the new system are highlighted in the table together with the results of the performance evaluation in terms of ACC.

The following studies represent leukocyte diagnosis research that has been conducted using classical ML. Sanei et al. (14) have utilized the Bayesian classifier for the classification of leukocytes. They have split the blood microscopic image into three sections. Instead of relying on the image's geometric or physical properties, they used a Bayesian classifier to isolate the Eigen cells. Decisions were based on the relative density of various colors. First, the input photographs were rescanned, segmented, and rotated, and the three vectors representing intensity and color were identified. Leukocyte images from 10 patients were employed by Sarrafzadeh et al. (31), who trained a support vector machine (SVM) using a set of parameters that includes six geometrical qualities, six color attributes, six statistical features, and seven-moment invariances (invariants). The classifier reported an ACC rate of over 93%. Leukocyte borders in images are defined manually to reduce the impact of segmentation errors. The cytoplasm and nucleus of leukocytes were separately identified by the Fuzzy C-means clustering method. Thereafter, the cytoplasm, nucleus, and other components of the cell that are of interest are removed Ko et al. (32) used SVMs to classify the 480 blood smear images into training and testing sets. They claimed that random forest performed better than multilayer SVM when it came to classification. In a previous work, the snake algorithm has been utilized to divide leukocytes. They used the shape, color, and texture of the image as criteria for classification. Gaussian normalization was then utilized to transform the feature vectors from 0 to 1 after feature extraction (32). Ramoser et al. employed SVM to automatically grade

leukocytes. The study of 1,166 images split into 13 categories found that segmentation was performed with 95% ACC (94/100) and classification was performed with 75%–90% ACC. In their study, Theera-Umpon and Dhompangsa (33) analyzed if it was possible to classify leukocytes using only data from their nuclei. To prevent segmentation errors from affecting the results of the investigations, the cell nuclei were removed manually. Bayes classifiers and CNNs were used for classification. They determined that the information obtained from cell nucleus 100 was adequate because their classification was correct 77% of the time. WBC subtype detection by flow cytometry was proposed by Adjouadi et al. Parametric datasets were analyzed in a multidimensional space using SVMs (34). To classify WBCs, Rodrigues et al. created a two-stage artificial neural network. To reduce the 106 problems, they first employed the Back Propagation Neural Network (BPNN) for preclassification and then presented a hybrid model based on the SVM and the pulse-coupled neural network (PCNN). As a result, they looked for ways to lessen the negative effects (35).

Both Otsu's automated thresholding methodology and the image enhancement and arithmetic strategy were proposed by Joshi et al. for separating leukocytes from red blood cells. The K Nearest Neighbor (K-NN) classifier was used to separate blast cells from typical lymphocyte cells. Their ACC was determined to be 93% based on the results of the tests (36). Image processing methods were used by Tantikitti et al. (37) for classifying WBCs, extracting features from edges, changing colors, and fragmenting images. Patients with dengue virus infections were sorted using a decision tree analysis. The results showed that a total of 167 cell shots were able to accurately classify leukocytes (92.2% ACC) and that 264 blood cell photos correctly classified dengue (72.3%). One hundred fifteen images were used by Hiremath et al. (38) as input parameters for AI-based algorithms that categorized WBCs based on their color, texture, and geometric properties. Histogram equalization, edge extraction, and threshold-based automatic

segmentation for lymphocytes, monocytes, and neutrophils are the focus of that study. Several images of blood smears were used in the trials, with geometric features of the images being utilized in the classification process. Habibzadeh et al. (39) employed the shape, density, and texture of microscopic images of blood to classify and count leukocytes. The parameters of the SVM classifier were the wavelet characteristics that were generated for the classification process using the dual-tree complex wavelet transform (DT-CWT) approach.

Ramesh et al. (40) proposed a simple classification method that incorporates morphological characteristics and color data. As the first step in a two-stage classification process, leukocyte cell nuclei and leukocyte boundaries have been meticulously established. The second stage involved applying the linear discriminant analysis method to implement the features found in the cytoplasm and nucleus of leukocytes. In another study, Su et al. (41) classified leukocytes into five distinct groups, each with its own set of distinguishing features. In this location, they aimed to use morphological mechanisms to segment the elliptical nuclei and cytoplasm of leukocytes. These photo chunks were mined for geometric elements, color characteristics, and texture qualities based on LDP (local directional pattern) and then used to train three distinct neural networks. For the testing, they used 450 images of leukocytes, and the highest identification ACC was 99.11%.

The microscopic analysis of blood cells is crucial for the early diagnosis of life-threatening hematological disorders such as leukemia. This paper presents an effective and computationally efficient approach for automatically detecting and classifying acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML). Das et al. (42) proposed an approach that uses transfer learning, which has been successful in medical image analysis due to its excellent performance in small databases. The proposed system employs a lightweight transfer learning-based feature extraction followed by SVM-based classification technique for efficient ALL and AML detection. The system is faster and more efficient due to the depthwise separable convolution, tunable multiplier, and inverted residual bottleneck structure. Moreover, the SVM-based classification technique improves the overall performance by optimizing the hyperplane location. The experimental results demonstrate that the proposed system outperforms others in all three publicly available standard databases, including ALLIDB1, ALLIDB2, and ASH.

Breast cancer is a leading cause of cancer-related deaths among women worldwide, and early detection is crucial for successful treatment. In this work, the authors have developed five new deep hybrid CNN-based frameworks for breast cancer detection. Sahu et al. (43) proposed that hybrid schemes exhibit better performance than the respective base classifiers by combining the benefits of both networks. A probability-based weight factor and threshold value are essential for efficient hybridization. An experimentally selected optimum threshold value makes the system faster and more accurate. Notably, unlike traditional DL methods, the proposed framework yields excellent performance even with small datasets. The proposed scheme is validated with datasets of breast cancer: mini-DDSM (mammogram), BUSI, and BUS2 (ultrasound). The experimental results demonstrate the superiority of the proposed

ShuffleNet-ResNet scheme over the current state-of-the-art methods in all of the mentioned datasets. Moreover, the proposed scheme achieves high ACC rates of 99.17% and 98.00% for abnormality and malignancy detection in mini-DDSM, respectively, and 96.52% and 93.18% for abnormality and malignancy detection in BUSI, respectively. In BUS2, the proposed scheme delivers 98.13% ACC for malignancy detection.

Sahu et al. (44) introduce a breast cancer detection framework based on DL that utilizes EfficientNet to achieve high performance even in cases of small databases. The framework incorporates uniform and adaptive scaling of depth, width, and resolution to ensure an optimal balance between classification performance and computational cost. Furthermore, a Laplacian of Gaussian-based modified high boosting (LoGMHB) is employed as a preprocessing step, along with data augmentation, to enhance the system's performance. The study evaluated the proposed method on mammogram and ultrasound modalities and demonstrated its superiority over other methods in all performance measures. The experimental results were obtained using 5-fold cross-validation and showed promising results for automatic and accurate detection of breast cancer at an early stage, which could lead to proper treatment and greatly reduce mortality rates.

The early detection of leukemia is crucial for proper treatment planning and improving patient outcomes. Microscopic analysis of WBCs is a cost-effective and less painful approach for detecting leukemia. However, automatic detection of leukemia using DL and ML techniques is a challenging task. Das et al. (45) present a systematic review of recent advancements in DL- and ML-based ALL detection. The review categorizes various AI-based ALL detection approaches into signal and image processing-based techniques, conventional ML-based techniques, and DL-based techniques, including supervised and unsupervised ML and CNN, recurrent neural network (RNN), and autoencoder-based classification methods. Furthermore, the review categorizes CNN-based classification schemes into conventional CNN, transfer learning, and other advancements. The article provides a critical analysis of recent research, discussing the merits and demerits of the different approaches and highlighting the challenges and future research directions in this field. Overall, this systematic review provides a comprehensive understanding of DL- and ML-based ALL detection, which may assist researchers in formulating new research problems in this domain.

Das et al. (46) propose an efficient deep convolutional neural network (DCNN) framework for accurate diagnosis of ALL, a challenging task. The framework features depthwise separable convolutions, linear bottleneck architecture, inverted residual, and skip connections. It uses a probability-based weight factor to efficiently hybridize MobilenetV2 and ResNet18, preserving their benefits. The approach achieves the best ACC in ALLIDB1 and ALLIDB2 datasets, with superior performance compared to transfer learning-based techniques.

In the field of biomedical image processing, DCNNs have received a lot of attention for various detection and classification tasks. The outcomes of many of them are comparable to or even superior to those of radiologists and neurologists. However, the need for a large dataset makes using such DCNNs difficult to



achieve decent results. Paul et al. (47) present a novel single model-based strategy for classifying brain tumors on a short dataset. To avoid overfitting, a modified DCNN known as the RegNetY-3.2G is coupled with regularization DropOut and DropBlock. Additionally, to mitigate the issue of tiny datasets, the RandAugment is an improved augmentation technique. Last but not least, the MWNL (Multi-Weighted New Loss).

Many studies have been introduced using DL techniques for the classification of leukocytes because of the outstanding performance of DL methods for the classification of medical images. The grid search (GS) and random search (RS) hyperparameter optimization methods were used by Hosseini et al. (20) to categorize images of four different categories of leukocytes. ACC of 99% on the training set and of 97% on the validation set was effectively obtained by the given hybrid technique. Through this study (21), the authors highlight the potential of DL, specifically CNNs, in automating the classification of different types of WBCs based on microscopic images. The use of CNNs allows for the detection of significant features that help distinguish different classes of leukocytes, which can assist hematologists in diagnosing diseases such as AIDS and leukemia. The study applied genetic algorithms to optimize the CNN's hyperparameters and trained the model on a dataset containing 9,957 images and tested it on another dataset of 2,478 images. The optimized CNN achieved high classification ACC, sensitivity, and specificity, indicating its potential as a substitute for manual WBC counting by pathologists. Overall, this study demonstrates the potential of DL techniques in the field of hematology and medical diagnosis. By automating the classification of WBCs, it could lead to more efficient and accurate diagnoses, ultimately improving patient outcomes. CNNs have been presented by Liang et al. (22). This approach can help to strengthen the explanation of input images and discover the structured features of images, and it can also begin end-to-end training of leukocyte images. In particular, they implemented the transfer learning method in order to transfer the feature weights to the CNN segment. Additionally, they implemented a configurable loss function in order to enable the network to train and converge at a faster rate and with more precise parameterization. The findings of their experiments demonstrated that their proposed model for the network has achieved an ACC of 90.8%. The optimized CNN achieved a classification ACC of 99% on the training set, which was 91% for the validation set. In (23), Bairaboina et al. present a DL model developed to classify mature and immature WBCs from peripheral blood smear images. Traditional methods of manual classification by hematologists can be laborious, expensive, and time-consuming. The proposed model uses a combination of W-Net, GhostNet, ResNeXt, and DCGAN-based data augmentation techniques to achieve high ACC levels of 99.16%, 99.24%, and 98.61% for three datasets. The model has potential for clinical application in blood cell microscopic analysis. Another a hybrid approach of recurrent neural networks (RNNs). Leukocyte segmentation was implemented using a network based on W-Net, a CNN-based technique for WBC classification implemented by Rao and Rao (24). Afterward, a DL system based on GhostNet was used to retrieve important feature maps. Then, a ResNeXt approach was used to classify them. The proposed method has attained an

ACC of 99.24% on the Blood Cell Count and Detection (BCCD). Rao and Rao (24) presented another DL-based framework for the classification of leukocytes based on the MobilenetV3-ShufflenetV2 DL paradigm. At first, an effective Pyramid Scene Parsing Network (PSPNet) is used to segment the images. When the images have been segmented, the global and local features are extracted and selected using MobilenetV3 and an Artificial Gravitational Cuckoo Search (AGCS)-based technique. Images are then classified into five groups using a ShufflenetV2 model. The proposed method achieves 99.19% and 99% ACC when tested on the BCCD and Raabin-Wbc datasets. Baydilli and Atila (25) have presented a capsule deep neural network (DNN)-based DL system for classifying leukocytes. They have attained an ACC of 96.9% on the benchmarking dataset, LISC. Ghosh and Bhattacharya (26) came up with two distinct models of CNNs that improve and categorize input images of blood cells. On the BCCD benchmarking dataset, they have achieved an ACC of 98.4%. Wang et al. (27) have applied two unique object detection strategies to the problem of leukocyte recognition. These strategies are known as Single Shot Multibox Detector and You Only Look Once (YOLO). In order to enhance the performance of recognition, several essential elements affecting these object detection strategies have been investigated, and detection models have been constructed utilizing a private dataset. The level of ACC that was achieved was 90.09%. Ma et al. (28) have come up with a new framework for the classification of blood cell images. This framework is built on a deep convolutional generative adversarial network (DCGAN) as well as a residual neural network (ResNet). They have accomplished a level of precision on the BCCD dataset that is 91.7% accurate. By bringing together the ideas of merging the features of the first and last convolutional layers and propagating the input image to the convolutional layer, Banik et al. (29) created a novel CNN model. They additionally employed a dropout layer to mitigate the model's overfitting issue. On the BCCD test database, they have obtained an average ACC of 96%. Sahlol et al. (30) have used VGGNet, a robust CNN architecture, already trained on ImageNet, to extract features from images of leukocytes. The statistically improved Salp Swarm Algorithm was then used to filter the extracted features. This optimization method takes biological principles as its inspiration, picking the most important features while discarding those that are excessively linked or noisy. ACC of 83.2% was attained when the proposed method was used on the C-NMC public Leukemia reference dataset.

## 3 Materials and methods

### 3.1 Dataset

The BCCD public dataset (25) contains 12,453 augmented images of leukocytes in JPEG format and cell type labels in CSV format. There are 3,120, 3,103, 3,107, and 3,123 augmented images for each class of the four cell types of eosinophil, lymphocyte, monocyte, and neutrophil, respectively, as compared with the 88, 33, 21, and 207 original images (Mooney, 2018). The basophil images are removed from the dataset as that type typically makes up less than 1% of the leukocytes. A drop of blood is placed on a glass



slide and smeared with a spreader slide. The blood is stained with a Romanowsky stain such as May-Grünwald Giemsa, Wright, or Wright–Giemsa. Image quality, illumination, and different staining techniques affect the outcome. The taken picture of cells is magnified 100× and converted to standard RGB channels. The dataset has been preprocessed, as each image was augmented and repositioned before it is made available to the public for the input of the CNN to avoid overfitting. The BCCD database is split into two sets: approximately 80% of the data (9,966 images) for the training set and 20% (2,487 images) for the validation set. The training set is composed of 2,497, 2,483, 2,487, and 2,499 images of eosinophil, lymphocyte, monocyte, and neutrophil, while the validation set contains 623, 620, 620, and 624 images of eosinophil, lymphocyte, monocyte, and neutrophil.

Neutrophils are the most numerous types of leukocytes constituting 50%–70% of the circulating leukocytes (44). The nucleus is relatively small and often multilobed. The stained nucleus is dark blue, and its CNR is 2:1. They are capable of phagocytizing viruses, toxins, fungi, and bacteria. They are the first line of defense once microbial infection strikes.

Eosinophils compose 1%–5% of the leukocytes; however, their counts fluctuate under different conditions (44). The cytoplasm is pink-stained while the nucleus is purple-stained and frequently is bilobed connected by a band of nuclear material. They protect against parasitic infections and cancer cells. They produce histamine as an inflammatory response to allergy-inducing agents, damaged tissue, or pathogen invasion.

Lymphocytes constitute 20%–45% of leukocytes and are much more common in the lymphatic system than in blood (22). They are agranular cells with a large dark purple-stained nucleus and a relatively small pale-colored amount of cytoplasm (38). They create antibodies to regulate immune system responses against bacteria, viruses, and other potentially harmful agents. The main types of lymphocytes are T cells, B cells, and natural killer cells.

Monocytes make up approximately 2%–10% of leukocytes and are the biggest leukocyte (22). Monocytes are granular and have a kidney-shaped nucleus with plenty of light blue cytoplasm. They share the phagocytic ability of neutrophils, break down bacteria, and remove waste from the blood. They have a longer life span compared with other leukocytes (20).

The BCCD database is augmented before becoming publicly available on the Kaggle website because, practically, the amount of training data is usually limited or not sufficient. Augmentation expands the training set with artificial data so it can be used by researchers. For the classification tasks, that means receiving a high-dimensional input such as images and producing a related output. A good classifier is immune to a wide-ranging variation. CNN as a framework well-established for image data can discriminate relevant minor features in the image while it is invariant to unrelated large variations in the image (26). For image datasets, augmentation can be done by modifying the images a few pixels to improve the generalization ability and avoid overfitting. Among available transformations are flipping, scaling, zooming, and rotating the image in several directions. Augmentation helps increase the correct classification rate regardless of size, position, or degree of distortion of an image. Using random transformation

exposes the network to more features in the data so it can generalize better. One thing to consider when using an augmentation approach is that one should take care of not altering the correct class by using the wrong transformations (42, 44).

## 3.2 Convolutional neural network and transfer learning

The CNN model is made up of multiple layers, including an input layer, convolutional layers, batch normalization layers, pooling layers, ReLU layers, Softmax layers, and one output layer. The dimensions  $a$ ,  $b$ , and  $c$  of the input image make up what is known as the input layer. The total number of channels is specified by  $c$ . The main and first convolutional layer of the network takes in data via three separate inputs labeled  $a$ ,  $b$ , and  $c$ . The convolutional layer is the one that is responsible for mapping out the features. The activation layer makes use of these features, which are also put to use for visualization purposes. Transfer learning makes use of an already trained and reused model as the foundation for a new task and model. The model used for one task can be repurposed for other tasks as an optimization to improve performance. By applying transfer learning, the model can be trained with a small volume of data. It is helpful to save time and achieve good results. In the transfer learning approach, we transfer knowledge from the source mammogram input images to the target domain mammogram mass images IT. The target classifier  $T_c(M_t)$  is to be trained from the input mammogram image  $I_s$  to the target image IT to get the classifier prediction about BMNT<sub>i</sub>, which stands for benign, malignant, and normal. To extract the features, a transfer layer is used. The top layer from the classifier retrained the new target classes, while the other layers were kept frozen as defined in Equation 1.

$$BMNT_i = T_c(M_t) \quad (1)$$

### 3.2.1 DenseNet

DenseNets are the subsequent stage to increase the depth of deep convolutional networks. When CNNs go deeper, the problems arise. This happens because of the big path for information from input to output layer. DenseNet-161 is a simple connectivity pattern because it connects all layers directly with each other to be sure that information flow is maximum between layers in the network. Feed forward nature is maintained by obtaining each layer additional inputs from the preceding layers. Figure 3 presents the architectures of DenseNet for ImageNet. Features are combined by concatenation. DenseNet is not as the same as traditional architecture because it introduces  $\frac{L(L+1)}{2}$  connections in an  $L$ -layer network in lieu of  $L$ . Handling problems of vanish gradient, reusing feature, lacking parameter's number, and propagating features is the most important feature of DenseNets.

### 3.2.2 Mathematical model of DenseNet deep networks

DNNs have reached state-of-the-art performance in a variety of computer vision applications. Moreover, the interpretation of

| Layers               | Output Size      | DenseNet-121 ( $k = 32$ )                                                                    | DenseNet-169 ( $k = 32$ )                                                                    | DenseNet-201 ( $k = 32$ )                                                                    | DenseNet-161 ( $k = 48$ )                                                                    |
|----------------------|------------------|----------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------|
| Convolution          | $112 \times 112$ |                                                                                              |                                                                                              |                                                                                              |                                                                                              |
| Pooling              | $56 \times 56$   |                                                                                              |                                                                                              |                                                                                              |                                                                                              |
| Dense Block (1)      | $56 \times 56$   | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$  | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$  | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$  | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 6$  |
| Transition Layer     | $56 \times 56$   | $1 \times 1 \text{ conv}$                                                                    |                                                                                              |                                                                                              |                                                                                              |
|                      | $28 \times 28$   | $2 \times 2 \text{ average pool, stride } 2$                                                 |                                                                                              |                                                                                              |                                                                                              |
| Dense Block (2)      | $28 \times 28$   | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$ | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$ | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$ | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 12$ |
| Transition Layer     | $28 \times 28$   | $1 \times 1 \text{ conv}$                                                                    |                                                                                              |                                                                                              |                                                                                              |
|                      | $14 \times 14$   | $2 \times 2 \text{ average pool, stride } 2$                                                 |                                                                                              |                                                                                              |                                                                                              |
| Dense Block (3)      | $14 \times 14$   | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$ | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$ | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 48$ | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 36$ |
| Transition Layer     | $14 \times 14$   | $1 \times 1 \text{ conv}$                                                                    |                                                                                              |                                                                                              |                                                                                              |
|                      | $7 \times 7$     | $2 \times 2 \text{ average pool, stride } 2$                                                 |                                                                                              |                                                                                              |                                                                                              |
| Dense Block (4)      | $7 \times 7$     | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 16$ | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$ | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 32$ | $\begin{bmatrix} 1 \times 1 \text{ conv} \\ 3 \times 3 \text{ conv} \end{bmatrix} \times 24$ |
| Classification Layer | $1 \times 1$     | $7 \times 7 \text{ global average pool}$                                                     |                                                                                              |                                                                                              |                                                                                              |
|                      |                  | 1000D fully-connected, softmax                                                               |                                                                                              |                                                                                              |                                                                                              |

FIGURE 3

DenseNet architectures for ImageNet. DenseNets are broken up into DenseBlocks, and while the dimensions of the feature maps stay the same inside each block, the number of filters that are used varies from one block to the next. These layers in between them are referred to as Transition Layers.

DNNs has been examined from the perspective of visualization as well as resilience. The groundbreaking studies that highlight the potential of DNNs include AlexNet and VGGNet. The community's research focus has changed from feature engineering to network design engineering as a direct result of the success of these key efforts. As a result, various new network architectures have been developed to improve the performance of DNNs. ResNets have achieved state-of-the-art performance on a variety of benchmark datasets, including ImageNet and the COCO detection dataset. This was accomplished by reusing previous features in conjunction with the identification shortcut. One of the factors that contribute to ResNet's phenomenally high level of popularity is its straightforward design strategy, which includes just one identity shortcut. The shortcomings of the identity shortcut have been investigated in subsequent publications, despite the tremendous success that it has enjoyed. Because the identity shortcut bypasses the residual blocks to maintain characteristics, it is possible that the network's capacity for representation is diminished as a result.

The ResNet has brought about a fundamental shift in how it was thought to parametrize the functions of DNNs. The DenseNet can be thought of as a kind of logical extension of this. Both the connection pattern in which each layer connects to all the preceding layers and the concatenation operation (as opposed to the addition operations in ResNet) to retain and reuse features of previous layer are defining characteristics of the DenseNet architecture. Let us make a brief detour into mathematics to comprehend how one might possibly arrive at such a conclusion. Looking back to functions' Taylor expansion. To clarify, for a point  $y=0$ , it might be expressed as shown in Equation 2. One of the most important features of ResNet is that it can break down a function into a series of terms with progressively higher orders. In a manner analogous to this, ResNet disassembles functions, as demonstrated in Equation 3. In other words, the ResNet breaks down a function  $f(y)$  into a straightforward linear component and a complex nonlinear one. However, if we were to write down more information than just the two components, but not necessarily add anything new. DenseNet is one example of such a solution. Figure 4 illustrates the primary difference between ResNet (shown on the left) and DenseNet (shown on the right) in terms of cross-layer connections: the utilization of addition versus the utilization of concatenation. As can be seen in Figure 4, the primary distinction between ResNet and DenseNet is that, in the latter case, outputs are

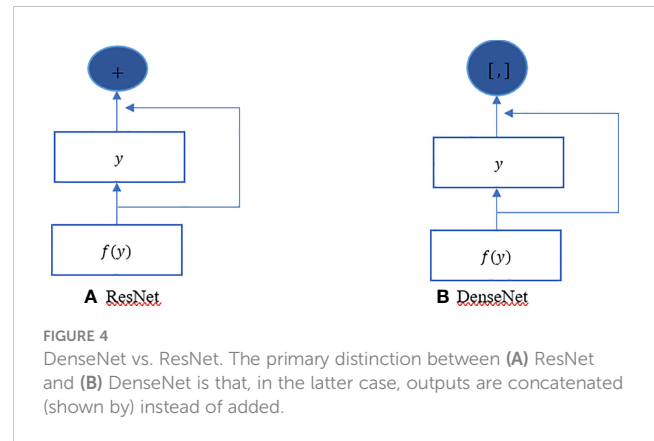


FIGURE 4

DenseNet vs. ResNet. The primary distinction between (A) ResNet and (B) DenseNet is that, in the latter case, outputs are concatenated (shown by) instead of added.

concatenated (shown by) instead of added. This is denoted by the notation. As a consequence of this, we apply an increasingly complex chain of functions before performing a mapping from the values it contains, as depicted in Equation 4. The number of features is further reduced by combining all of these functions in Multi-layer Perceptron (MLP). The mechanics of this are straightforward; instead of adding, we just string together the terms. DenseNet gets its name from how packed the dependency tree between the variables gets. The last layer in this structure has numerous connections to its predecessors. Figure 5 depicts these complex interconnections.

$$f(y) = f(0) + y \cdot [f'(0) + y \cdot [\frac{f''(0)}{2!} + y \cdot [\frac{f'''(0)}{3!} + \dots]]] \quad (2)$$

$$f(y) = y + g(y) \quad (3)$$

$$y \rightarrow [y, f_1(y), f_2([y, f_1(y)]), f_3([y, f_1(y), f_2([y, f_1(y))])], \dots] \quad (4)$$

### 3.3 The proposed CAD system for leukocyte images

The image preprocessing and enhancement stage, the image segmentation stage, the feature extraction and selection stage, and the WBC classification stage are the four primary stages that are included in the proposed CAD system. These stages are illustrated in Figure 6, which also contains the information that is mentioned

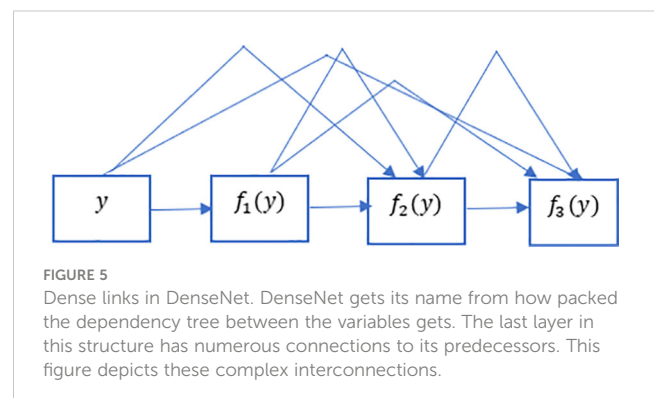


FIGURE 5

Dense links in DenseNet. DenseNet gets its name from how packed the dependency tree between the variables gets. The last layer in this structure has numerous connections to its predecessors. This figure depicts these complex interconnections.

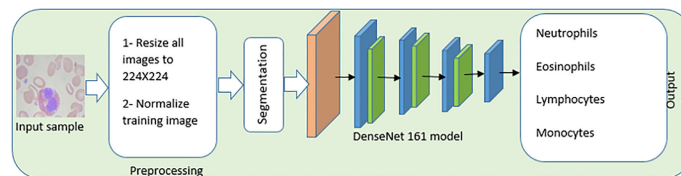


FIGURE 6

An automated End-to-End CAD, system of white blood cells. The image preprocessing and enhancement stage, the image segmentation stage, the feature extraction and selection stage, and the white blood cell classification stage are the four primary stages that are included in the proposed CAD system.

in the *Introduction* section. In addition, the findings of this research contribute to existing state-of-the-art models by suggesting the implementation of a one fit cycle strategy, which makes the process of training simpler. As a result, there is no requirement to adjust any of the hyperparameters of the network that is being used.

### 3.3.1 Image preprocessing phase

In order to process the input histopathological image sample, images are resized to  $244 \times 244$ , and training images are the only ones that are normalized. Changing the range of intensity values for individual pixels is the core idea behind image normalization. The purpose of image normalization is to transform the pixel range values into ranges that are more intuitive to the senses.

### 3.3.2 Image segmentation phase

Figure 7 illustrates an example for an input image with its corresponding output image, segmented one. Color Image Segmentation was used on the images to separate each individual pixel using the HSV color space. The images will be segmented using information derived from the HSV color space. HSV is an abbreviation that stands for hue, saturation, and value as illustrated in Figure 8.

The following Algorithm 1 is an outline of the primary steps that are involved in the image segmentation phase:

- i) First, convert the RGB image in HSV form as depicted in Figure 6.
- ii) using the color bar at the right to choose the thresholds.

iii) set up the thresholds for the masks.

- Lower Mask (refer to the hue channel)
- Upper Mask (refer to the hue channel)
- Saturation Mask (refer to the transparency channel)

Ex: to segment the NEUTROPHIL cell, the lower and upper mask values that are appropriate would be 0.0 and 1.0 After that, the saturation threshold is decided. This is a bit tricky because you need to consider the colors that are seen in the object. In this case, the values are 0.45

iv) Create the mask by multiplying all masks of the thresholds.

mask = upper\_mask\*lower\_mask\*saturation\_mask

v) Then, multiply this mask by each value in the rgb image.

red = img[:, :, 0]\*mask

green = img[:, :, 1]\*mask

blue = img[:, :, 2]\*mask

i) Lastly, apply the morphology operation to remove the noise or halls.

Algorithm 1. Image segmentation phases.

### 3.3.3 Feature selection and classification using the DenseNet model

The DenseNet-161 DL model is used in the implementation of both the feature extraction and classification stages. Adjouadi et al. (34) developed DenseNet that had the best classification results on the available datasets such as ImageNet. DenseNet does not use direct

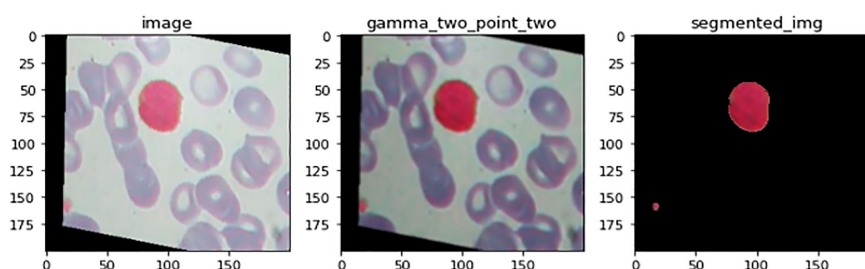


FIGURE 7

An illustration of image before and after the segmentation phase. This is an example for an input image with its corresponding output image, segmented one.

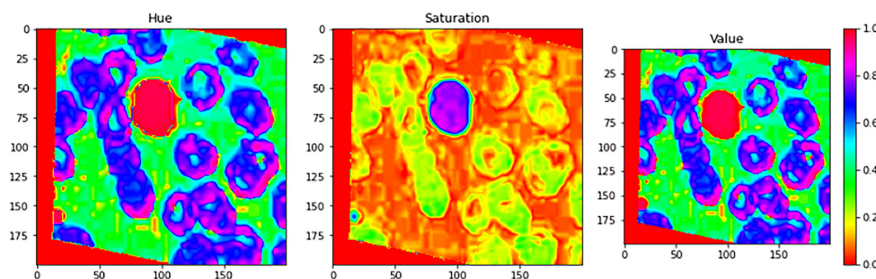


FIGURE 8

Color Image Segmentation was used on the images to separate each individual pixel using the HSV color space. The images will be segmented using information derived from the HSV color space. HSV is an abbreviation that stands for Hue, Saturation, and Value.

connections among hidden layers, but it uses dense connection to build a model. Its construction was based on linking each to a subsequent layer. In any layer, any important features learned are involved within the network. Due to the extracted features, deep network training became more efficient and the performance of the model increased. The number of parameters has become less than CNN because feature maps are sent directly to all subsequent layers. The DenseNet has a very important feature, which is that it reduces overfitting in the model because of the use of dense connections. Training DL models with very large numbers of parameters takes much time. More and more data and powerful GPU are required to train these models from scratch. Transfer learning is used to overcome the pervious problem. By using transfer learning, you are saving time. Transfer learning is a method of machine learning in which a model that was developed for one task can be utilized as a foundation for a model that will be used on a different task. Learned features are often transferable to different data. For example, a model trained in Dataset for animal images that includes learned features such as edges and lines can be used on other dataset using transfer learning technique. In transfer learning, feature extractor is done by fully connected layer after removing it from the model used.

In this study, DenseNet-161 with ImageNet is utilized; however, the final layer, which is designated as the “completely connected” layer, has had its number of classes reduced from 1,000 to 4. The strategy known as one fit cycle policy is utilized to implement DenseNet-161.

### 3.3.4 One fit cycle policy

It is known that training of DNN is a difficult optimization problem. Tuning of hyperparameters such as LR is very important. The performance of the network will be enhanced by carefully selecting the hyperparameters for LR, momentum, and weight decay. The traditional approach involves running a grid or random search, which can be time-consuming and computationally intensive. The impacts of these hyperparameters are also closely related to the architecture, the data, and each other. This section provides more effective guidelines for selecting certain hyperparameters (27). A small LR leads to very slow training, while a large LR hinders the convergence. A low LR is good, but it takes a long time to train perfectly. When training speed is increased, LR is increased until LR gets too large and diverge. To obtain the exact LR, you need to do many

experiments and be patient. A new method was discovered by Leslie N. Smith for setting up LR named CLR. CLR made LR values between minimum and maximum range instead of having fixed values during the training. CLR cycle has two steps, one of them being an increase in LR and the other one being a decrease in LR. CLR eliminates the need to find the optimal LR but the optimal rate between minimum and maximum range. Figure 9 shows classification ACC while training CIFAR-10. The red curve is CLR. As depicted in Figure 9, the CLR achieves the same ACC as the original LR but in iteration less than the original LR method (15). In Leslie N. Smith’s research (48), super-convergence is the method that uses CLR, but with one cycle that contains two LR steps. The total number of iterations must be larger than the size of the cycle. After completing the cycle, LR is decreased much further for the remaining iterations. Leslie N. Smith named this method one fit cycle policy. In super-convergence, LR starts from a small value and is increased to a very large value then returns to a value lower than its initial one. The impact of LR many values is a training ACC curve. In super-convergence, training ACC is moved fast as LR is increased (15, 44), becomes oscillated as LR is very large, and then jumps again to an extreme point of ACC.

To utilize CLR, one must provide a step size and minimum and maximum LR bounds. A cycle consists of two such steps, one in which the LR linearly grows from the lowest to the maximum and the other in which it progressively falls. The step size is the number of iterations (or epochs) utilized for each step. Smith (2015) explored a variety of methods for varying the LR between the two boundary values, discovered that they were all equivalent, and thus advised the

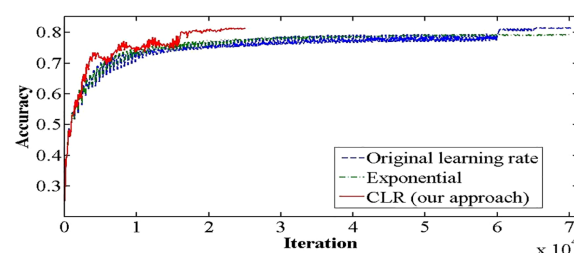


FIGURE 9

CLR method and original learning rate.



most straightforward method—letting the LR change linearly—even though suggested discrete jumps and found similar outcomes (28).

Training for the LR range test begins with a modest LR and gradually rises linearly over the course of a pretraining run. This single run offers useful insight into the maximal LR and how well the network can be taught over a variety of LRs. The network starts to converge at a low LR, and as the LR rises, it finally reaches an unmanageable size, which lowers ACC and increases test/validation loss. By using a constant LR, a smaller number is required since otherwise the network will not start to converge. The LR at these extrema is the highest value that can be utilized as the LR for the maximum bound with CLRs. The minimal LR constraint can be chosen in a variety of ways: 1) by a factor of 3 or 4 less than the maximum bound, 2) by a factor of 10 or 20 less than the maximum bound if only one cycle is used, or 3) by a quick test of hundreds of iterations with a few initial LRs and choosing the largest one that permits convergence to start without overfitting. If the initial LR is too large, the training will not start to converge. Be aware that the LR can only rise to a certain point before the training becomes unstable. This affects your decision about the lowest and maximum LRs (i.e., raise the step size to widen the gap between the minimum and maximum).

## 4 Results and discussion

The experiments are applied on a BCCD public dataset. Our studies were carried out on it with the help of Google Colab. The evaluation criteria are used to evaluate the performance of classification model, including image test ACC, Macro-F1, Micro-F1, and Kappa criteria, and average time. Macro-F1 takes the average of the precision and recall of each class. ACC is defined by the ratio of Ncor (the number of correctly classified images in testing set) to Nall (Total number of images in testing set). Equation 5 defines the image test ACC. Precision is calculated as the sum of true positives across all classes divided by the sum of true positives and false positives across all classes. Recall is calculated as the sum of true positives across all classes divided by the sum of true positives and false negatives across all classes. Equation 6 defines Micro-F1. Kappa measure, based on confusion matrix calculation, can handle problems such as imbalanced datasets and multiclass problems. Precision is defined by Equation 7, and it means the percentage of your results that are relevant. On the other hand, recall as described by Equation 8 refers to the percentage of total relevant results correctly classified by your algorithm. Equation 9 defines Kappa coefficient, where  $p_0$  is the image test ACC as defined in Equation 5, and  $p_e$  is the summation of the product of the number of images in each type of cancer and the predicted number of images in each type of cancer to the square of the total number of images in the testing set.

$$Accuracy(ACC) = \frac{N_{cor}}{N_{all}} \quad (5)$$

$$F1 - score = \frac{2X(precision \times recall)}{(precision + recall)} \quad (6)$$

$$Precision = \frac{\text{Sum } c \text{ in } C \text{ TruePositives\_c}}{\text{Sum } c \text{ in } C (\text{TruePositives\_c} + \text{FalsePositives\_c})} \quad (7)$$

$$Recall = \frac{\text{Sum } c \text{ in } C \text{ TruePositives\_c}}{\text{Sum } c \text{ in } C (\text{TruePositives\_c} + \text{FalseNegatives\_c})} \quad (8)$$

$$Kappa = \frac{p_0 - p_e}{1 - p_e}, p_0 = \frac{N_{cor}}{N_{all}}, p_e = \frac{\sum N_{true} X N_{pre}}{N_{all} X N_{all}} \quad (9)$$

### 4.1 Classification results

The next subsection discusses the classification result on the BCCD dataset based on the default one fit cycle policy approach. The experimental result is applied on a raw dataset. Moreover, the results of our research experiments are compared with the results of other researchers. The experiments are performed over a desktop computer system having an Intel Core i7-7700 CPU, 16 GB RAM, and one 8-GB GPU. This research used DenseNet-161 to perform the classification of microscopic images into neutrophils, eosinophils, lymphocytes, and monocytes by using a pretrained model in terms of ACC, F1, AUC, and Kappa. Our experimental result of multiclassification problem on raw data is shown in Table 1 according to ACC, Macro-F1, Micro-F1 and Kappa. We ran the raw data on 30 epochs. All classification results are given in Tables 2, 3. The loss curves are shown in Figure 10, and the confusion matrices are shown in Figure 11.

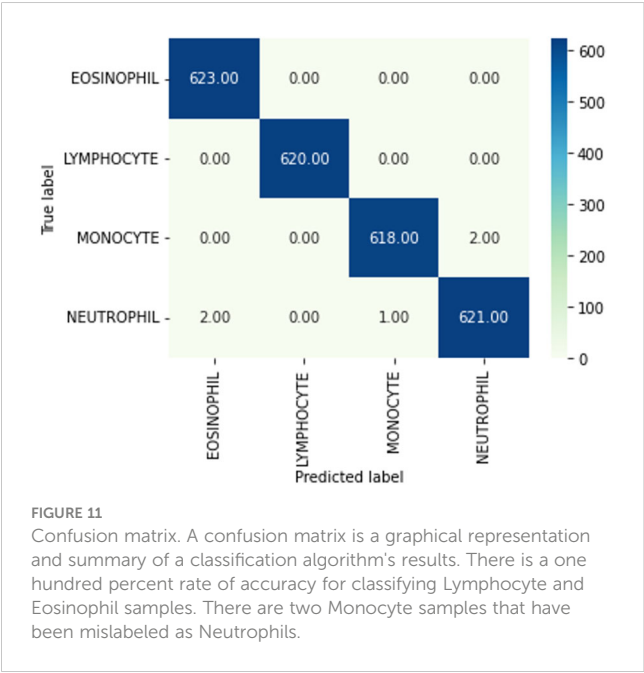
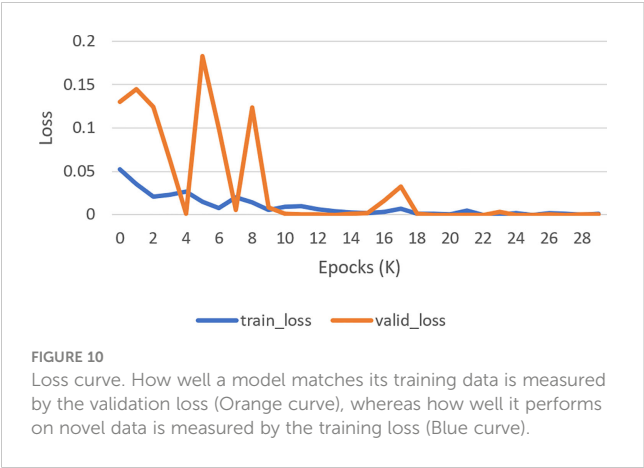
The experimental results in Table 3 show that all evaluation metrics on 40× magnification factor (which is indicated by the black

TABLE 2 The result of each evaluation is on raw data.

| Network      | Criteria       | Result | Average Time |
|--------------|----------------|--------|--------------|
| DenseNet-161 | Accuracy (ACC) | 0.985  | 4.30 h       |

TABLE 3 Precision, Recall, and F1-score for raw data.

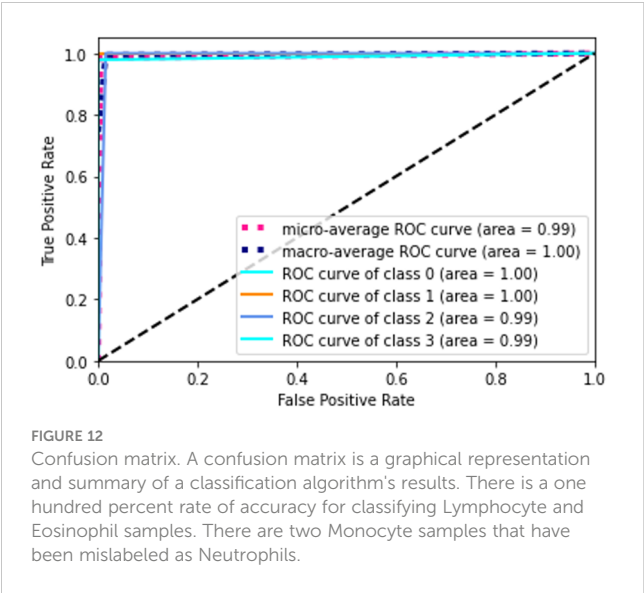
| Criteria     | Types      | precision | recall | f1-score |
|--------------|------------|-----------|--------|----------|
|              | Eosinophil | 1.00      | 1.00   | 1.00     |
|              | Lymphocyte | 1.00      | 1.00   | 1.00     |
|              | Monocyte   | 0.80      | 1.00   | 0.89     |
|              | Neutrophil | 1.00      | 0.98   | 0.99     |
| accuracy     | –          | –         | –      | 0.99     |
| macro avg    | –          | 0.95      | 0.99   | 0.99     |
| weighted avg | –          | 0.99      | 0.99   | 0.99     |



underline) are better than the other magnification factors. The reason for 40× achieving the best ACC is because it contains more significant features of breast cancer. From Table 3, precision, recall, and F1-score values show that our model classification result is perfect.

TABLE 4 A comparison between research results and the state of the art.

| Criteria       | Author                      | Methods                               | Result |
|----------------|-----------------------------|---------------------------------------|--------|
| Accuracy (ACC) | Maryam et al. (20)          | Optimized CNN                         | 99%    |
|                | Bani-Hani et al. (21)       | GA-optimized CNN                      | 91%    |
|                | Liang et al. (22)           | Hybrid CNN-RNN                        | 90.8%  |
|                | Rao (23)                    | CNN and ResNeXt                       | 99.24% |
|                | Rao (24)                    | ANN and CNN                           | 97.7%  |
|                | Ghosh and Bhattacharya (25) | CNN and FCN on noise-free cell images | 98.4%  |
|                | Ma et al. (26)              | DC-GAN, and ResNet                    | 91.7%  |
|                | Banik et al. (27)           | CNN                                   | 96%    |
|                | Proposed model              | DenseNet-161 with CLR Approach        | 99.8%  |



The receiver operating characteristic (ROC) metric is used to evaluate the output quality. ROC is a probability curve, while AUC, area under the curve, is a metric for assessing how well two groups may be distinguished. It reveals the extent to which the model can differentiate between categories. If the AUC is high, then the model is very good at predicting the correct classes. The AUC value (see in Figure 12) for class 0 and class 1 is 1.00 and for class 2 and class 3 is 0.99. Ideally, the ROC for the false positive rate should be zero and one for the true positive rate.

4.2 Comparisons with other models

This section compares our experimental results with the other experiments carried out by other research papers on raw data and augmented data. We evaluated the ACC of the newly introduced method to that of the most recent and cutting-edge classification frameworks for leukocyte histology by using the BCCD database. As a consequence of this, we were in a position to evaluate the significance of the hybrid DenseNet and CLR approach. The differences between the suggested method and the state-of-the-art methods that are



currently in use are outlined in Table 4. To facilitate this comparison, ACC was utilized as a performance metric.

DenseNet with the CLR approach, the suggested classification framework, outperforms the DL systems established by Bani-Hani et al. (40), Liang et al. (22), Paul et al. (47), Bairaboina and Battula (23), Rao and Rao (24), and Banik et al. (29) when applied to the BCCD dataset. In addition to this, it has accomplished a level of ACC that is on par with that which Habibzadeh et al. (39), Rao and Rao (24), and Ghosh and Bhattacharya (26) have accomplished.

In general, it can be deduced from the comparison in Table 4 that the suggested system is capable of recording a performance that is better than that of all other systems.

From the experimental results applied on raw and augmented data, Tables 2–4 show that the evaluation criteria-specified ACC achieved the best results by applying a new method in training called one fit cycle policy and with small number of batches and the fewest number of epochs. When we have trained the CNN using 32 batch size and 60 epoch, we did not attain high performance. On the contrary, we use 32 batch sizes and 30 epochs on raw data, and this helped us to reduce the time of training and achieve better ACC than the other research.

## 5 Conclusions

Using a combination of the recently developed pretrained CNN, DenseNet, and the one fit cycle policy, this study describes a technique of training for the classification of WBCs. The proposed method is more accurate and requires less cycles to train CNN—thanks to the one fit cycle policy. It fixes how difficult it is to adjust the DL model's hyperparameters. DenseNet-161 was used in the experiment, and the results are analyzed in terms of various performance indicators. ACC, precision, and recall are presented as indicators of the suggested model's efficacy. We solved the multiclass classification problem with a raw data ACC of 99.8%. As a result, the outcomes of our experiments are more reliable than those obtained in the existing state of the art for the classification of WBCs. In the future work, the proposed model can be applied to diagnosis-specific diseases such as cancer and liver disease.

## 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.

## References

1. Ghosh S, Majumder M, Kudesia A. LeukoX: leukocyte classification using least entropy combiner (LEC) for ensemble learning. *IEEE Trans Circuits Syst II: Express Briefs* (2021) 68:2977–81. doi: 10.1109/TCSII.2021.3064389
2. Khan MA, Qasim M, Lodhi HMJ, Nazir M, Javed K, Rubab S, et al. Automated design for recognition of blood cells diseases from hematopathology using classical

## Author contributions

Supervision and methodology, EH; software, EH and OM; conceptualization, EH, NS, RT, AA-H, RA-T, and NM; validation, EH, NS, RT, AA-H, RA-T, and NM; formal analysis, EH and OM; investigation, EH and OM; resources, NS and NM; data curation, NS and NM; visualization, EH, OM, NS, RT, AA-H, RA-T, and NM; writing—original draft preparation, OM; writing—review and editing, EH, OM, NS, RT, AA-H, RA-T, and NM; funding acquisition, NS and NM. All authors discussed the results and approved the final paper. All authors have read and agreed to the published version of the article.

## Funding

This work was supported by the Princess Nourah bint Abdulrahman University Researchers Supporting Project Number (PNURSP2023R206), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

## Acknowledgments

The authors express their gratitude to Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2023R206), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

## 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.

features selection and ELM. *Microsc Res Tech* (2021) 84:202–16. doi: 10.1002/JEMT.23578

3. Almezghwi K, Serte S. Improved classification of white blood cells with the generative adversarial network and deep convolutional neural network. *Comput Intell Neurosci* (2020) 2020. doi: 10.1155/2020/6490479

4. Herron C. Know your WBCs. *Nurs Made Incredibly Easy* (2012) 10:11–5. doi: 10.1097/01.NME.0000408238.43869.E2
5. Badior KE, Casey JR. Molecular mechanism for the red blood cell senescence clock. *IUBMB Life* (2018) 70:32–40. doi: 10.1002/IUB.1703
6. Liu S, Deng Z, Li J, Wang J, Huang N, Cui R, et al. Measurement of the refractive index of whole blood and its components for a continuous spectral region. *J Biomed Opt* (2019) 24:035003. doi: 10.1117/1.JBO.24.3.035003
7. Maton A. *Human biology and health*. 1st ed. Englewood Cliffs, United States: Prentice Hall (1993).
8. Tai WL, Hu RM, Hsiao HCW, Chen RM, Tsai JJP. Blood cell image classification based on hierarchical SVM. *Proc - 2011 IEEE International Symposium Multimedia ISM* (2011) 2011:129–36. doi: 10.1109/ISM.2011.29
9. Pore YN, Kalshetty Y. Review on blood cell image segmentation and counting. *Int J Innovat Eng Manag* (2014).
10. Lichtman M, Kaushansky K, Prchal J, Levi M, Burns L, Linch D. *Williams manual of hematology*. 10e ed. United States: McGraw-Hill (2011).
11. Ushizima DM, Lorena AC, De Carvalho ACPLF. (2005). Support vector machines applied to white blood cell recognition, in: *Proceedings - HIS 2005: Fifth International Conference on Hybrid Intelligent Systems*, Vol. 2005. pp. 379–84. doi: 10.1109/ICHIS.2005.100
12. Chen H, Liu J, Hua C, Feng J, Pang B, Cao D, et al. Accurate classification of white blood cells by coupling pre-trained ResNet and DenseNet with SCAM mechanism. *BMC Bioinf* (2022) 23. doi: 10.1186/s12859-022-04824-6
13. Pan SJ, Yang Q. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering* (2010). doi: 10.1109/TKDE.2009.191
14. Sanei S, Lee TKM. *CELL RECOGNITION BASED ON PCA AND BAYESIAN CLASSIFICATION*. Japan: ICA (2003).
15. Smith L. A disciplined approach to neural network hyper-parameters: Part 1 – learning rate, batch size, momentum, and weight decay. (2018).
16. Chola C, Muad AY, Bin Heyat MB, Benifa JVB, Naji WR, Hemachandran K, et al. BCNet: A deep learning computer-aided diagnosis framework for human peripheral blood cell identification. *Diagnostics* (2022) 12:2815. doi: 10.3390/diagnostics12112815
17. Das PK, Meher S, Panda R, Abraham. A. (2021). An efficient blood-cell segmentation for the detection of hematological disorders, in: *IEEE Transactions on Cybernetics*, Vol. 52. pp. 10615–26.
18. Abdel Samee NM. Classical and deep learning paradigms for detection and validation of key genes of risky outcomes of HCV. *Algorithms* (2020) 13:73. doi: 10.3390/a13030073
19. Samee NA, Mahmoud NF, Aldahri EA, Rafiq A, Muthanna MSA, Ahmad I. RNN and biLSTM fusion for accurate automatic epileptic seizure diagnosis using EEG signals. *Life* (2022) 12:1946. doi: 10.3390/life12121946
20. Hosseini M, Bani-Hani D, Lam SS. Leukocytes image classification using optimized convolutional neural networks. *Expert Syst Appl* (2022) 205:117672. doi: 10.1016/J.ESWA.2022.117672
21. Bani-Hani D, Khan N, Alsultan F, Karanjkar S, Nagarur NN. *Classification of leucocytes using convolutional neural network optimized through genetic algorithm*. Conference: Society for Industrial and Systems Engineering, Binghamton, NY, USA.
22. Liang G, Hong H, Xie W, Zheng L. Combining convolutional neural network with recursive neural network for blood cell image classification. *IEEE Access* (2018) 6:36188–97. doi: 10.1109/ACCESS.2018.2846685
23. Bairaboina SSR, Battula SR. Ghost-resNeXt: an effective deep learning based on mature and immature WBC classification. *Appl Sci* (2023) 13:4054. doi: 10.3390/AP13064054
24. Rao BSS, Rao BS. An effective WBC segmentation and classification using mobilenetV3-shufflenetV2 based deep learning framework. *IEEE Access* (2023) 11:27739–48. doi: 10.1109/ACCESS.2023.3259100
25. Baydilli YY, Atila Ü. Classification of white blood cells using capsule networks. *Computerized Med Imaging Graphics* (2020) 80:101699. doi: 10.1016/J.COMPMEDIMAG.2020.101699
26. Ghosh S, Bhattacharya S. Classification of RBC and WBC in noisy microscopic images of blood smear. *Lecture Notes Networks Syst* (2020) 79:195–200. doi: 10.1007/978-981-32-9453-0\_20/FIGURES/3
27. Wang Q, Bi S, Sun M, Wang Y, Wang D, Yang S. Deep learning approach to peripheral leukocyte recognition. *PloS One* (2019) 14. doi: 10.1371/JOURNAL.PONE.0218808
28. Ma L, Shuai R, Ran X, Liu W, Ye C. Combining DC-GAN with ResNet for blood cell image classification. *Med Biol Eng Comput* (2020) 58:1251–64. doi: 10.1007/S11517-020-02163-3
29. Banik PP, Saha R, Kim KD. An automatic nucleus segmentation and CNN model based classification method of white blood cell. *Expert Syst Appl* (2020) 149:113211. doi: 10.1016/J.ESWA.2020.113211
30. Sahlol AT, Kollmannsberger P, Ewees AA. Efficient classification of white blood cell leukemia with improved swarm optimization of deep features. *Sci Rep* (2020) 10:1–11. doi: 10.1038/s41598-020-59215-9
31. Sarrafzadeh O, Rabbani H, Talebi A, Banaem HU. Selection of the best features for leukocytes classification in blood smear microscopic images. *Prog Biomed Opt Imaging - Proc SPIE* (2014) 9041:159–66. doi: 10.1117/12.2043605
32. Ko BC, Gim JW, Nam JY. Cell image classification based on ensemble features and random forest. *Electron Lett* (2011) 47:638–9. doi: 10.1049/EL.2011.0831/CITE/REFWORKS
33. Theera-Umporn N, Dhompongs S. Morphological granulometric features of nucleus in automatic bone marrow white blood cell classification. *IEEE Trans Inf Technol BioMed* (2007) 11:353–9. doi: 10.1109/TITB.2007.892694
34. Adjouadi M, Zong N, Ayala M. Multidimensional pattern recognition and classification of white blood cells using support vector machines. *Particle Particle Syst Characterization* (2005) 22:107–18. doi: 10.1002/PPSC.200400888
35. Rodrigues P, Ferreira M, Monteiro J. Segmentation and classification of leukocytes using neural networks: A generalization direction. *Stud Comput Intell* (2008) 83:373–96. doi: 10.1007/978-3-540-75398-8\_17/COVER
36. Joshi MD, Karode AH, Suralkar SR. White blood cells segmentation and classification to detect acute leukemia ms. (2013). Available at: <https://api.semanticscholar.org/CorpusID:18894374>.
37. Tantikitti S, Tumswadi S, Premchaiswadi W. Image processing for detection of dengue virus based on WBC classification and decision tree. *Int Conf ICT Knowledge Eng* (2015), 84–9. doi: 10.1109/ICTKE.2015.7368476
38. Hiremath PS, Bannigidad P, Geeta S. Automated identification and classification of white blood cells (Leukocytes) in digital microscopic images. *Int J Comput Appl* (2010).
39. Habibzadeh M, Krzyzak A, Fevens T. Comparative study of shape, intensity and texture features and support vector machine for white blood cell classification. *J Theor Appl Comput Sci* (2013) 7.
40. Ramesh N, Dangott B, Salama ME, Tasdizen T. Isolation and two-step classification of normal white blood cells in peripheral blood smears. *J Pathol Inform* (2012) 3:13. doi: 10.4103/2153-3539.93895
41. Su MC, Cheng CY, Wang PC. A neural-network-based approach to white blood cell classification. *Sci World J* (2014) 2014. doi: 10.1155/2014/796371
42. Das PK, Nayak B, Meher. S. A lightweight deep learning system for automatic detection of blood cancer. *Measurement* (2022) 191:110762.
43. Sahu A, Das PK, Meher. S. High accuracy hybrid CNN classifiers for breast cancer detection using mammogram and ultrasound datasets. *Biomed Signal Process Control* (2023) 80:104292.
44. Sahu A, Das PK, Meher S, Panda R, Abraham. A. (2022). An efficient deep learning-based breast cancer detection scheme with small datasets, in: *International Conference on Intelligent Systems Design and Applications*, pp. 39–48.
45. Das PK, Diya VA, Meher S, Panda R, Abraham. A. A systematic review on recent advancements in deep and machine learning based detection and classification of acute lymphoblastic leukemia. *IEEE Access* (2022).
46. Das PK, Meher. S. An efficient deep convolutional neural network based detection and classification of acute lymphoblastic leukemia. *Expert Syst Appl* (2021) 183:115311.
47. Indraneel P, Sahu A, Das PK, Meher. S. (2023). Deep convolutional neural network-based automatic detection of brain tumour, in: *2023 2nd International Conference for Innovation in Technology (INOCON)*, pp. 1–6.
48. Smith LN, Topin N. Super-convergence: very fast training of neural networks using large learning rates. *Conference: Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications* (2019) 11006:369–86. doi: 10.1117/12.2520589



## OPEN ACCESS

## EDITED BY

Zhimin Li,  
Northwestern University, United States

## REVIEWED BY

Han Lv,  
Capital Medical University, China  
Guolin Ma,  
China-Japan Friendship Hospital, China  
Ruiyang Zhao,  
GE Healthcare, United States

## \*CORRESPONDENCE

Yong Zhang

✉ zzuzhangyong2013@163.com

Jingliang Cheng

✉ fccchengjl@zzu.edu.cn

<sup>†</sup>These authors have contributed equally to this work

RECEIVED 24 May 2023

ACCEPTED 04 September 2023

PUBLISHED 27 September 2023

## CITATION

Wen B, Zhang Z, Zhu J, Liu L, Liu Z, Ma X, Wang K, Xie L, Zhang Y and Cheng J (2023) Synthetic MRI plus FSE-PROPELLER DWI for differentiating malignant from benign head and neck tumors: a preliminary study. *Front. Oncol.* 13:1225420. doi: 10.3389/fonc.2023.1225420

## COPYRIGHT

© 2023 Wen, Zhang, Zhu, Liu, Liu, Ma, Wang, Xie, Zhang and Cheng. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

# Synthetic MRI plus FSE-PROPELLER DWI for differentiating malignant from benign head and neck tumors: a preliminary study

Baohong Wen<sup>1†</sup>, Zanzia Zhang<sup>1†</sup>, Jing Zhu<sup>1</sup>, Liang Liu<sup>1</sup>, Zijun Liu<sup>1</sup>, Xiaoyue Ma<sup>1</sup>, Kaiyu Wang<sup>2</sup>, Lizhi Xie<sup>2</sup>, Yong Zhang<sup>1\*</sup> and Jingliang Cheng<sup>1\*</sup>

<sup>1</sup>Department of MRI, the First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, <sup>2</sup>MR Research China, GE Healthcare, Beijing, China

**Background:** Preoperative classification of head and neck (HN) tumors remains challenging, especially distinguishing early cancerogenic masses from benign lesions. Synthetic MRI offers a new way for quantitative analysis of tumors. The present study investigated the application of synthetic MRI and stimulus and fast spin echo diffusion-weighted imaging with periodically rotated overlapping parallel lines with enhanced reconstruction (FSE-PROPELLER DWI) to differentiate malignant from benign HN tumors.

**Materials and methods:** Forty-eight patients with pathologically confirmed HN tumors were retrospectively recruited between August 2022 and October 2022. The patients were divided into malignant (n = 28) and benign (n = 20) groups. All patients were scanned using synthetic MRI and FSE-PROPELLER DWI. T1, T2, and proton density (PD) values were acquired on the synthetic MRI and ADC values on the FSE-PROPELLER DWI.

**Results:** Benign tumors (ADC:  $2.03 \pm 0.31 \times 10^{-3} \text{ mm}^2/\text{s}$ , T1:  $1741.13 \pm 662.64 \text{ ms}$ , T2:  $157.43 \pm 72.23 \text{ ms}$ ) showed higher ADC, T1, and T2 values compared to malignant tumors (ADC:  $1.46 \pm 0.37 \times 10^{-3} \text{ mm}^2/\text{s}$ , T1:  $1390.06 \pm 241.09 \text{ ms}$ , T2:  $97.64 \pm 14.91 \text{ ms}$ ) (all  $P < 0.05$ ), while no differences were seen for PD values. ROC analysis showed that T2+ADC (cut-off value,  $> 0.55$ ; AUC, 0.950) had optimal diagnostic performance vs. T1 (cut-off value,  $\leq 1675.84 \text{ ms}$ ; AUC, 0.698), T2 (cut-off value,  $\leq 113.24 \text{ ms}$ ; AUC, 0.855) and PD (cut off value,  $> 80.67 \text{ pu}$ ; AUC, 0.568) alone in differentiating malignant from benign lesions (all  $P < 0.05$ ); yet, the difference in AUC between ADC and T2+ADC or T2 did not reach statistical significance.

**Conclusion:** Synthetic MRI and FSE-PROPELLER DWI can quantitatively differentiate malignant from benign HN tumors. T2 value is comparable to ADC value, and T2+ADC values could improve diagnostic efficacy., apparent diffusion coefficient, head and neck tumors

#### KEYWORDS

synthetic, magnetic resonance imaging, diffusion-weighted image, apparent diffusion coefficient, head and neck tumors

## Highlights

- Synthetic MRI, FSE-PROPELLER DWI, and the combination of the two methods can all be used to quantitatively diagnose differential head and neck (HN) tumors.
- Synthetic MRI could constitute a new adjunct in diagnosing HN tumors.
- Synthetic MRI is comparable to FSE-PROPELLER DWI.
- The combined effect of the two methods was better than synthetic MRI used alone.

## 1 Introduction

Head and neck (HN) cancer is the 6<sup>th</sup> most common cancer and the 9<sup>th</sup> most common cause of cancer-related death (1, 2). Surgery is the most effective treatment for managing primary HN cancer. Yet, many patients present with advanced-stage tumors at the time of diagnosis and thus require more invasive treatment, including radiochemotherapy, immunotherapy, and targeted therapy (3). Also, diagnosis remains challenging considering its specific location (masses originating from the larynx, the nasopharynx, oropharynx, oral cavity hypopharynx, salivary glands, etc.). In addition, HN cancer might cause various symptoms that commonly accompany benign conditions (1). Thus, the differentiation of benign from cancerous masses is very important.

Tissue biopsy and pathologic examination remain the gold standard for evaluating the nature of HN lesions; nevertheless,

only a part of the tissue can be obtained using this method. In addition, this approach is invasive and not always well accepted by the patient (4).

Endoscopy, head MRI, computed tomography (CT of the sinuses and head, dental cone beam CT), panoramic dental x-ray, and positron emission tomography (PET)/CT or chest imaging are the most common imaging methods used to assess the HN region. MRI is frequently used to detect, differentiate, grade, or draw the extent of HN tumors (1, 5). Among different MRI models, diffusion-weighted imaging (DWI) can quantitatively evaluate the Brownian motion diffusion of water molecules in tissues at a cellular level expressed as an apparent diffusion coefficient (ADC). DWI with a single-shot echo-planar sequence (SS-EP-DWI), which is commonly applied to investigate HN regions (6), is sensitive to chemical shifts, signal loss and geometric distortion, metallic dental implant-related magnetic susceptibility artifacts, and motion artifacts (7). Moreover, stimulus and fast spin echo DWI with periodically rotated overlapping parallel lines with enhanced reconstruction (FSE-PROPELLER DWI) is useful to distinguish parotid pleomorphic adenoma from Warthin tumor with less distortion of tumors than SS-EP-DWI (7). However, the value of FSE-PROPELLER DWI in distinguishing malignant from benign HN tumors has not been fully explored.

The major limitations of DWI include low signal-to-noise ratios and prolonged acquisition time. Over the years, a new synthetic MRI sequence based on a quantitative approach has been developed. This tool can estimate absolute physical properties, proton density (PD), and longitudinal and transverse relaxation times (T1, T2), which are independent of the MRI scanners or scanning parameters at a given field strength (8). Also, quantitative values (PD, T1, and T2) can be simultaneously acquired on the synthetic MRI, which enables a significant reduction in examination time with good accuracy and reproducibility for use in clinical practice (individual patient follow-up and comparison analysis (9–11). This approach has been used in the study of multiple systemic diseases of the brain (12), knee (13), spine (14), prostate (15), breast (16), bladder (17), and nasopharynx (18).

In this study, we further assessed the value of synthetic MRI in differentiating malignant from benign HN tumors compared with FSE-PROPELLER DWI and a combination of these two methods. To the best of our knowledge, this is the first study that focused on

**Abbreviations:** ADC, apparent diffusion coefficient; AUC, area under the curve; DWI, diffusion-weighted imaging; EPI, echo-planar imaging; FOV, field of view; FSE, fast spin echo; ICC, intraclass correlation coefficient; MAGIC, magnetic resonance image compilation; NA, not applicable; NPV, negative predictive value; PD, proton density; PPV, positive predictive value; PROPELLER, periodically rotated overlapping parallel lines with enhanced reconstruction; ROC, receiver operating characteristic; ROI, region of interest; SS, single-shot; T1, longitudinal relaxation time; T2, transverse relaxation time; T1WI, T1-weighted imaging; T2WI, T2-weighted imaging; TE, echo time; TR, repetition time; TSE, turbo spin-echo.

synthetic MRI and FSE-PROPELLER DWI to characterize HN tumors.

## 2 Materials and methods

### 2.1 Patients

MRI data from 48 consecutive patients (mean age  $\pm$  standard deviation [SD], 48.08 years  $\pm$  15.01 [range, 18–76 years]) with HN tumors who were treated at our hospital between August 2022 and October 2022 were collected. The inclusion criteria were: (1) no tumor treatments before MR examinations; (2) all pathological examinations of samples were obtained by surgical resection or biopsy of the tumor; (3) synthetic MRI and FSE-PROPELLER DWI were acquired before surgical resection and biopsy; (4) the maximum tumor diameter was  $\geq 6$  millimeters. The exclusion criteria were: (1) MR images with obvious artifacts and poor quality; (2) patients previously treated. Subjects were divided into benign and malignant groups.

This study was approved by our institutional review board. Informed consent was waived.

### 2.2 Data collection

Demographic data included gender and age. All MRI acquisitions were performed on a 3T MR scanner (Premier, GE Healthcare, Milwaukee, WI, USA) in a supine position with a 21-channel head-neck coil. The following sequences were acquired in this study: axial T1-weighted image (T1WI), T2-weighted image (T2WI), synthetic MRI, and FSE-PROPELLER DWI with two b-values (0 and 800s/mm<sup>2</sup>). Detailed acquisition parameters are listed in Table 1.

### 2.3 Image analysis

Acquired data from synthetic MRI sequences were analyzed using magnetic resonance image compilation (MAGIC) software. Then, quantitative T1, T2, and PD maps were created and used for measurements to yield synthetic images and match the conventional images (19). The two radiologists (with 10 and 8 years of experience in head and neck MR imaging independently analyzed MR images) who were blind to the grouping manually drew the regions of interest (ROIs) on synthetic T2WI to obtain the PD, T1, and T2 values. Postprocessing of FSE-PROPELLER DWI was performed using the ADW 4.7 workstation (GE Healthcare). The axial routine MR images and DWI were used as references. ROIs were drawn on synthetic T2WI and ADC maps with care by avoiding necrosis, cystic degeneration, and bleeding areas at the slice with the largest tumor diameter and directly colocalized on the T1, T2, and PD maps. The size of ROIs was  $>25$  mm<sup>2</sup>. Two radiologists measured three times. The average value was obtained by both radiologists in the analysis. Additionally, the largest lesion was selected for analysis if more than one HN lesion were present.

### 2.4 Statistical analysis

Shapiro-Wilk test was used to assess normality, while Levene's test was used for variance homogeneity. The normally distributed variables were expressed as the means  $\pm$  SD. Non-normally distributed variables were expressed as medians (interquartile ranges, IQRs). Differences in sex between the two groups were compared using a chi-square test. An independent samples t-test was used to compare the discrepancy in age between the two groups. The intraclass correlation coefficient (ICC) was used to assess the intraobserver agreement for quantitative parameters (19):

TABLE 1 MRI Sequence Parameters.

| Parameters                       | T1WI    | T2WI      | MAGIC         | DWI           |
|----------------------------------|---------|-----------|---------------|---------------|
| Imaging technique                | FSE     | Flex      | Synthetic MRI | FSE-PROPELLER |
| Orientation                      | Axial   | Axial     | Axial         | Axial         |
| TR (ms)                          | 693     | 3339      | 4000          | 3620          |
| TE (ms)                          | 6.7     | 80        | 13.3          | 50            |
| Field of view (mm <sup>2</sup> ) | 220×220 | 220×220   | 220×220       | 220×220       |
| Slice thickness (mm)             | 4       | 4         | 4             | 4             |
| No. of slices                    | 24      | 24        | 24            | 24            |
| Nex                              | 1       | 2         | 1             | 4             |
| Fat suppression                  | NA      | Dixon     | NA            | Fat sat       |
| Acquisition matrix               | 320×224 | 280×224   | 224×224       | 120×50        |
| Flip angle (degree)              | 111     | 110       | NA            | 110           |
| Acquisition time                 | 40s     | 2 min 27s | 3 min 38 s    | 3 min 20 s    |

T1WI, T1-weighted imaging; T2WI, T2 weighted imaging; DWI, diffusion-weighted imaging; TSE, turbo spin-echo; MAGIC, magnetic resonance image compilation; PROPELLER, periodically rotated overlapping parallel lines; TR, repetition time; TE, echo time; NA, not applicable.



value <0.40, 0.41-0.59, 0.60-0.74, and  $\geq 0.75$  indicated poor, fair, good, and excellent consistency, respectively. Pearson's correlation coefficient was used to evaluate the correlation among parameters. The receiver operating characteristic (ROC) curve was conducted, and the area under the curve (AUC), sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV) were further calculated to ascertain the diagnostic performance of quantitative parameters for differentiating malignant from the benign HN tumors. The diagnostic value of the combined ADC and T2 values (T2+ADC) was based on the logistic regression analysis. The method developed by DeLong et al. (20) was applied to compare AUCs. Statistical analysis was performed using MedCalc statistical software (version 19.6, MedCalc) and SPSS software (version 17.0, Chicago, IL, USA).  $P < 0.05$  was considered statistically significant.

## 3 Results

### 3.1 General data

A total of 48 patients with histologically diagnosed HN tumors were assessed. Demographics are listed in Tables 2 and 3. In addition, representative images of benign and malignant tumors are depicted in Figures 1 and 2. There was no difference in age ( $t = -1.392$ ,  $P > 0.05$ ) and gender between the two groups ( $\chi^2 = 0.689$ ,  $P = 0.406$ ).

### 3.2 Interobserver reliability

ICC analyses showed excellent consistency in the ADC, T1, T2, and PD values assessed by the two physicians: the ICC values were 0.976

TABLE 2 Histologic types and locations of head and neck tumors.

| Benign/malignant | Histologic types        | Locations            | Gender (M/F) | Age (y) | T1 (ms) | T2 (ms) | PD (pu) | ADC ( $\times 10^{-3}$ mm <sup>2</sup> /s) |
|------------------|-------------------------|----------------------|--------------|---------|---------|---------|---------|--------------------------------------------|
| Benign           | Pleomorphic adenoma     | Parotid gland        | M            | 19      | 1854    | 147.17  | 89.4    | 2.15                                       |
| Benign           | Pleomorphic adenoma     | Parotid gland        | M            | 50      | 1230.67 | 117.83  | 68.74   | 1.67                                       |
| Benign           | Pleomorphic adenoma     | Parotid gland        | F            | 48      | 2991    | 317.33  | 90.99   | 2.59                                       |
| Benign           | Pleomorphic adenoma     | Submandibular gland  | F            | 54      | 1837    | 152.33  | 84.97   | 2.15                                       |
| Benign           | Pleomorphic adenoma     | Parotid gland        | F            | 49      | 1334.5  | 110.83  | 80.67   | 1.86                                       |
| Benign           | Pleomorphic adenoma     | Parotid gland        | F            | 32      | 916.5   | 114.17  | 86.25   | 1.76                                       |
| Benign           | Pleomorphic adenoma     | Parotid gland        | M            | 35      | 1292.5  | 99.17   | 74.17   | 1.84                                       |
| Benign           | Pleomorphic adenoma     | Parotid gland        | F            | 24      | 1712.84 | 143     | 86.15   | 2.03                                       |
| Benign           | Pleomorphic adenoma     | Parapharyngeal space | M            | 59      | 1287.33 | 153.5   | 80.47   | 1.95                                       |
| Benign           | Pleomorphic adenoma     | Parotid gland        | F            | 63      | 1854.67 | 174.17  | 87.07   | 2.18                                       |
| Benign           | Pleomorphic adenoma     | Parotid gland        | M            | 52      | 1573.5  | 140.5   | 88.60   | 1.94                                       |
| Benign           | Pleomorphic adenoma     | Parotid gland        | F            | 32      | 3332.5  | 304.5   | 97.52   | 2.42                                       |
| Benign           | Pleomorphic adenoma     | Parotid gland        | M            | 21      | 2810.34 | 274.34  | 98.15   | 2.58                                       |
| Benign           | Pleomorphic adenoma     | Parotid gland        | F            | 42      | 1548.84 | 118.17  | 86.09   | 1.86                                       |
| Benign           | Pleomorphic adenoma     | Parotid gland        | F            | 54      | 2227.5  | 265.17  | 89.15   | 2.40                                       |
| Benign           | Basal cell adenoma      | Parotid gland        | M            | 53      | 1382.33 | 93.17   | 80.3    | 1.54                                       |
| Benign           | Basal cell adenoma      | Parotid gland        | F            | 57      | 1431    | 96.5    | 77.84   | 1.79                                       |
| Benign           | Basal cell adenoma      | Parotid gland        | M            | 44      | 1401.83 | 97.5    | 79.19   | 1.71                                       |
| Benign           | Basal cell adenoma      | Parotid gland        | M            | 62      | 817     | 120.67  | 89.25   | 2.34                                       |
| Benign           | Basal cell adenoma      | Parotid gland        | F            | 41      | 1986.84 | 108.5   | 87.99   | 1.90                                       |
| Malignant        | Squamous cell carcinoma | Tongue               | M            | 54      | 1329.5  | 98.84   | 83.22   | 1.55                                       |
| Malignant        | Squamous cell carcinoma | Hypopharynx          | M            | 64      | 1166.17 | 78.33   | 89.83   | 1.72                                       |
| Malignant        | Squamous cell carcinoma | Tongue               | M            | 32      | 1850    | 93.5    | 90.17   | 1.84                                       |

(Continued)

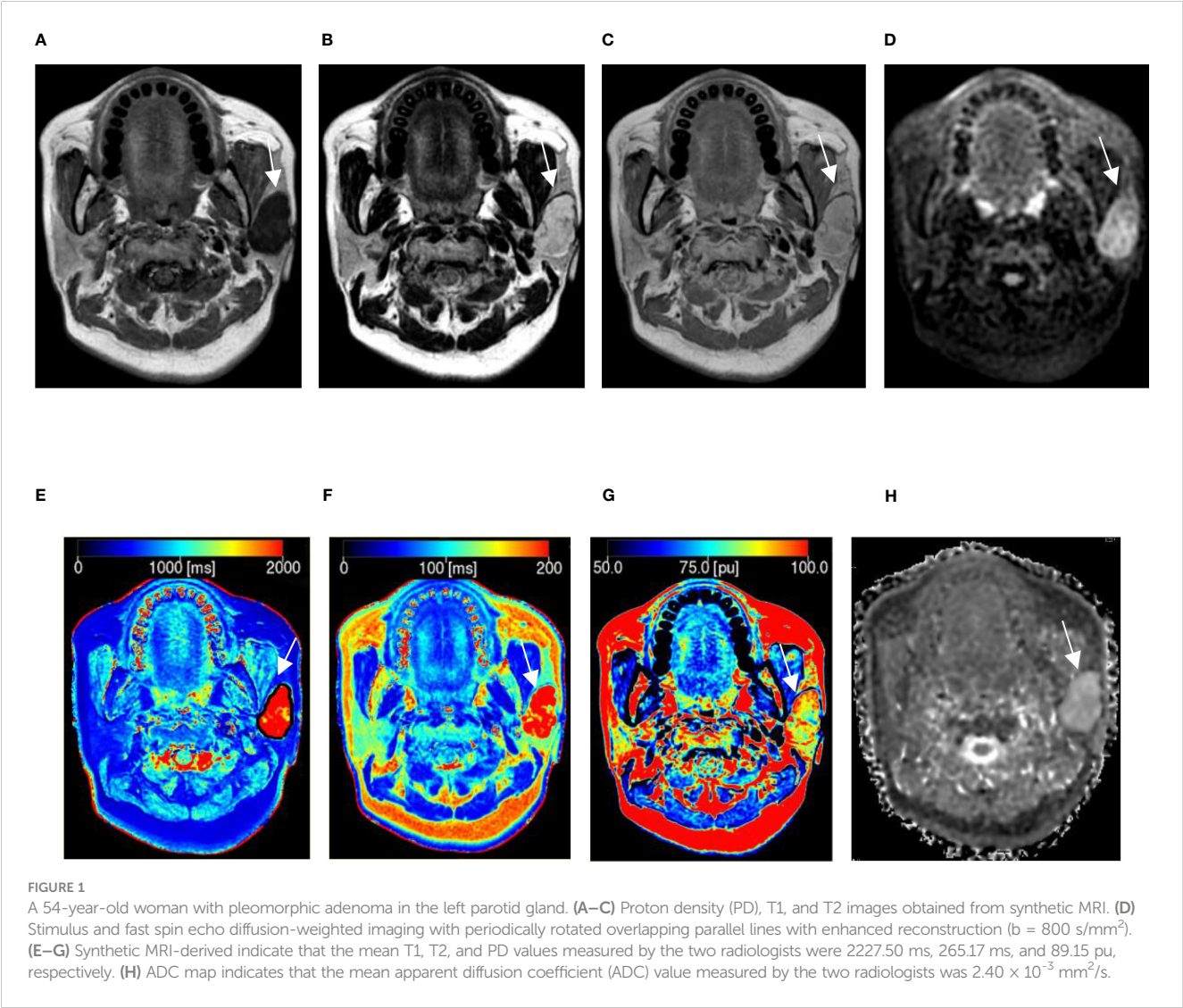
TABLE 2 Continued

| Benign/malignant | Histologic types        | Locations              | Gender (M/F) | Age (y) | T1 (ms) | T2 (ms) | PD (pu) | ADC ( $\times 10^{-3}$ mm <sup>2</sup> /s) |
|------------------|-------------------------|------------------------|--------------|---------|---------|---------|---------|--------------------------------------------|
| Malignant        | Squamous cell carcinoma | Nasopharynx            | M            | 18      | 1422.67 | 100     | 85.38   | 1.04                                       |
| Malignant        | Squamous cell carcinoma | Tongue                 | F            | 62      | 1143.67 | 86.67   | 80.5    | 2.49                                       |
| Malignant        | Squamous cell carcinoma | Nasopharynx            | M            | 56      | 1326.17 | 95.67   | 85.85   | 1.15                                       |
| Malignant        | Squamous cell carcinoma | Tongue                 | M            | 53      | 1244.34 | 91.84   | 82.6    | 1.54                                       |
| Malignant        | Squamous cell carcinoma | Nasopharynx            | F            | 36      | 1175.34 | 80.17   | 87.25   | 1.34                                       |
| Malignant        | Squamous cell carcinoma | Vocal cords            | M            | 52      | 2067.17 | 88.67   | 90.14   | 2.41                                       |
| Malignant        | Squamous cell carcinoma | Nasopharynx            | M            | 37      | 1197.17 | 69.33   | 81.54   | 1.56                                       |
| Malignant        | Squamous cell carcinoma | Tongue                 | F            | 48      | 1334.17 | 101.5   | 85.83   | 1.44                                       |
| Malignant        | Squamous cell carcinoma | Nasopharynx            | M            | 27      | 1054.5  | 82.17   | 80.92   | 1.07                                       |
| Malignant        | Squamous cell carcinoma | Nasopharynx            | M            | 59      | 1214.17 | 93.5    | 89.35   | 1.14                                       |
| Malignant        | Squamous cell carcinoma | Parotid gland          | M            | 76      | 1471.17 | 113.34  | 82.87   | 1.50                                       |
| Malignant        | Squamous cell carcinoma | Buccal mucosa          | F            | 47      | 1348.84 | 102.4   | 87.78   | 1.43                                       |
| Malignant        | Squamous cell carcinoma | Tongue                 | M            | 41      | 1214.67 | 105.17  | 80.39   | 1.52                                       |
| Malignant        | Squamous cell carcinoma | Buccal mucosa          | F            | 65      | 1276.84 | 92.17   | 79.27   | 1.60                                       |
| Malignant        | Squamous cell carcinoma | Tongue                 | M            | 69      | 1283.67 | 78.83   | 84.54   | 1.43                                       |
| Malignant        | Lymphoma                | Parapharyngeal space   | F            | 49      | 1308.5  | 83.83   | 83      | 0.98                                       |
| Malignant        | Lymphoma                | Nasopharynx            | M            | 66      | 1398    | 89.67   | 84.97   | 1.03                                       |
| Malignant        | Lymphoma                | Parotid gland          | F            | 52      | 1483    | 125.33  | 82.25   | 1.54                                       |
| Malignant        | Lymphoma                | Submandibular gland    | F            | 73      | 1571.84 | 109.34  | 87.9    | 1.24                                       |
| Malignant        | Lymphoma                | Tonsil                 | F            | 50      | 1649    | 110.84  | 89.48   | 1.16                                       |
| Malignant        | Acinar cell carcinoma   | Parotid gland          | F            | 58      | 1675.84 | 121.17  | 86.32   | 1.77                                       |
| Malignant        | Acinar cell carcinoma   | Parotid gland          | M            | 32      | 1181    | 103.33  | 89.43   | 1.58                                       |
| Malignant        | Rhabdomyosarcoma        | Parotid gland and neck | F            | 20      | 1833.67 | 133.84  | 83.92   | 1.31                                       |
| Malignant        | Melanoma                | Paranasal sinus        | M            | 74      | 1239.5  | 99.84   | 89.54   | 1.17                                       |
| Malignant        | Plasmacytoma            | Parapharyngeal space   | F            | 47      | 1461.17 | 104.5   | 85.95   | 1.20                                       |

TABLE 3 Demographics and the parameters of patients with head and neck tumors.

| Group           | Gender (M/F)       | Age(y)              | T1 (ms)              | T2 (ms)            | PD (pu)             | ADC ( $\times 10^{-3}$ mm <sup>2</sup> /s) |
|-----------------|--------------------|---------------------|----------------------|--------------------|---------------------|--------------------------------------------|
| Benign tumor    | 9/11               | 44.55 $\pm$ 13.44   | 1741.13 $\pm$ 662.64 | 157.43 $\pm$ 72.23 | 85.15 $\pm$ 7.17    | 2.03 $\pm$ 0.31                            |
| Malignant tumor | 16/12              | 50.61 $\pm$ 15.80   | 1390.06 $\pm$ 241.09 | 97.64 $\pm$ 14.91  | 85.36 $\pm$ 3.37    | 1.46 $\pm$ 0.37                            |
| $\chi^2/t/t'$   | 0.689 <sup>a</sup> | -1.392 <sup>b</sup> | 2.265 <sup>c</sup>   | 3.647 <sup>c</sup> | -0.125 <sup>c</sup> | 5.762 <sup>c</sup>                         |
| P value         | 0.406              | 0.171               | 0.033                | 0.002              | 0.901               | <0.001                                     |

Unless otherwise indicated, data are mean  $\pm$  standard deviation (SD). ADC, apparent diffusion coefficient; PD, proton density, <sup>a</sup> Data is  $\chi^2$ , <sup>b</sup> Data is t, <sup>c</sup> Data is t'.



(95% CI 0.958 - 0.987,  $P = 0.000$ ), 0.936 (95% CI 0.988 - 0.997,  $P = 0.000$ ), 0.996 (95% CI 0.993 - 0.998,  $P = 0.001$ ), and 0.823 (95% CI 0.706 - 0.897,  $P = 0.000$ ), respectively.

0.001), T2 and PD values ( $r = 0.495$ ,  $P < 0.001$ ), and T2 and ADC values ( $r = 0.646$ ,  $P < 0.001$ ), respectively. There was no significant positive correlation between PD and ADC values ( $r = 0.281$ ,  $P = 0.053$ ).

3.3 Correlation analysis

Pearson’s correlation analysis showed a significant positive correlation between the T1 and T2 values ( $r = 0.854$ ,  $P < 0.001$ ), T1 and PD values ( $r = 0.574$ ,  $P < 0.001$ ), T1 and ADC values ( $r = 0.565$ ,  $P <$

3.4 MRI values between the two groups

The T1 value ( $1741.13 \pm 662.64$ ms), T2 value ( $157.43 \pm 72.23$ ms), and ADC value ( $2.03 \pm 0.31 \times 10^{-3}$  mm<sup>2</sup>/s) of the

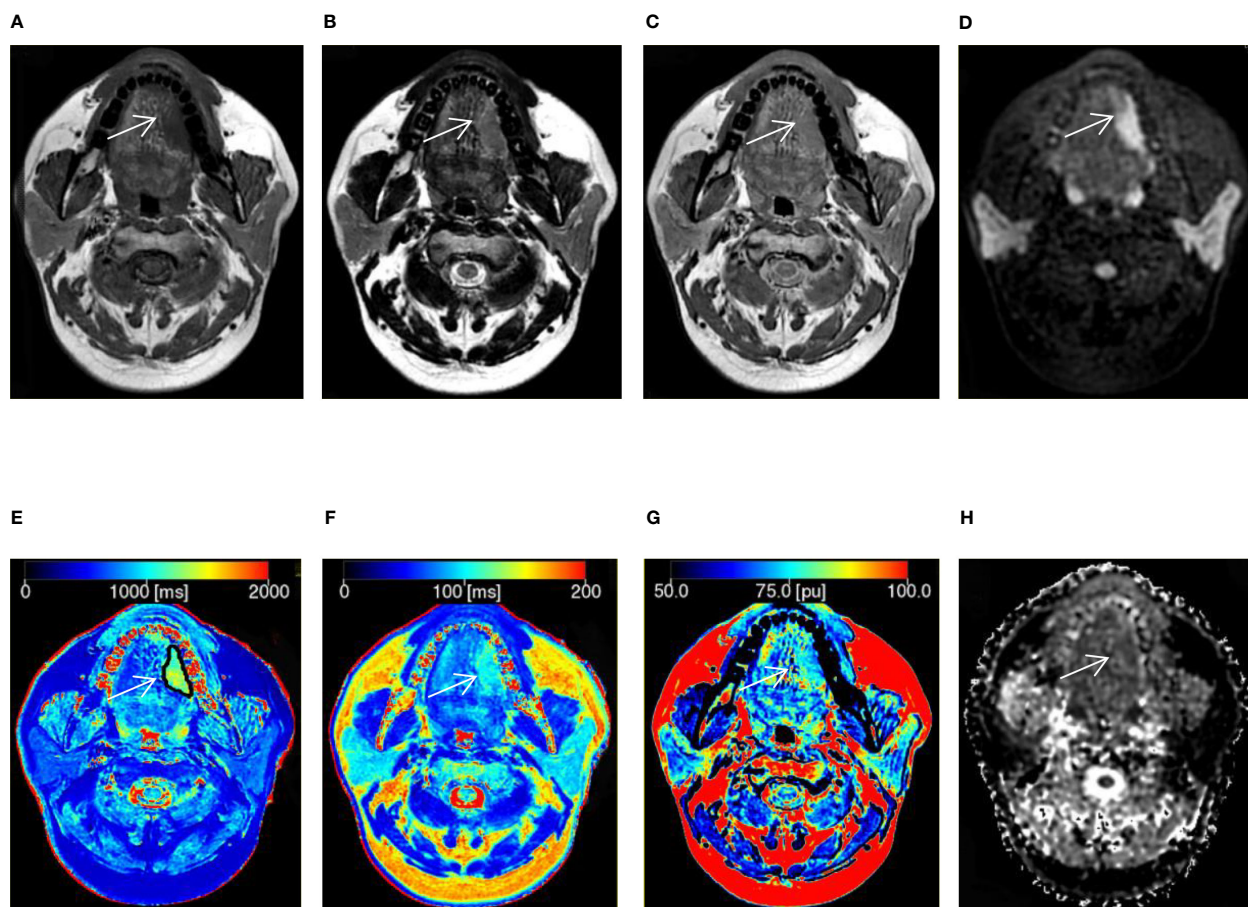


FIGURE 2

A 41-year-old man with squamous cell carcinoma in the tongue. (A–C) T1, T2, and proton density (PD) weighted images obtained from synthetic MRI. (D) Stimulus and fast spin echo diffusion-weighted imaging with periodically rotated overlapping parallel lines with enhanced reconstruction ( $b = 800 \text{ s/mm}^2$ ). (E–G) Synthetic MRI-derived maps indicate that the mean T1, T2, and PD values measured by the two radiologists were 1214.67 ms, 105.17 ms, and 80.39 pu, respectively. (H) ADC map indicates that the mean apparent diffusion coefficient (ADC) value measured by the two radiologists was  $1.52 \times 10^{-3} \text{ mm}^2/\text{s}$ .

benign group was higher compared to the malignant group (T1:  $1390.06 \pm 241.09 \text{ ms}$ ,  $t' = 2.265$ ,  $P = 0.033$ ; T2:  $97.64 \pm 14.91 \text{ ms}$ ,  $t' = 3.647$ ,  $P = 0.002$ ; ADC:  $1.46 \pm 0.37 \times 10^{-3} \text{ mm}^2/\text{s}$ ,  $t' = 5.762$ ,  $P < 0.001$ ). Yet, no significant differences were found in PD values between the two groups ( $t' = -0.125$ ,  $P = 0.901$ ).

### 3.5 Comparison of ROC curves

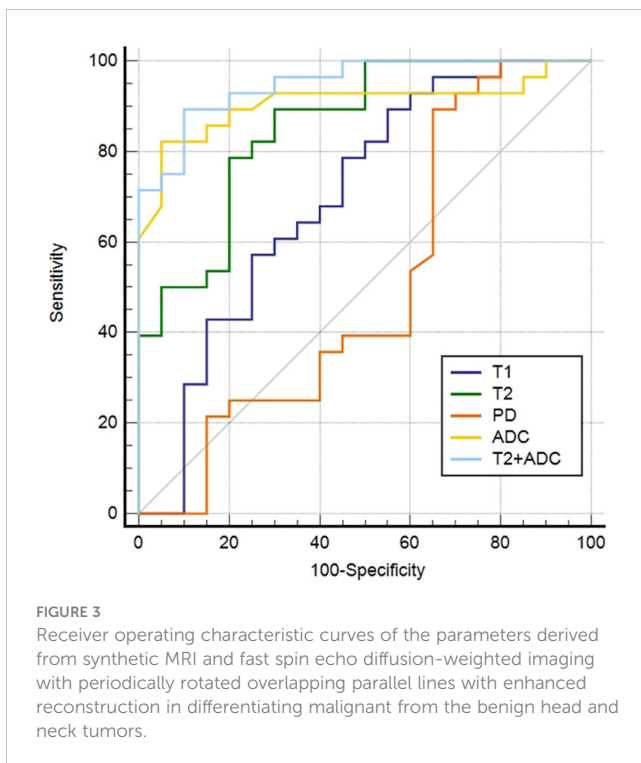
The AUC, cut-off, sensitivity, specificity, PPV, and NPV of each parameter discriminating malignant from benign lesions are summarized in Table 4.

TABLE 4 Diagnostic performance of MRI values and combined values for differentiating malignant from benign lesions.

| Variable                                       | Cut off | AUC (95%CI)        | Sensitivity (%) | Specificity (%) | PPV (%) | NPV (%) | *P value |
|------------------------------------------------|---------|--------------------|-----------------|-----------------|---------|---------|----------|
| T1 (ms)                                        | 1675.84 | 0.698(0.549-0.822) | 89              | 45              | 69      | 75      | 0.0161   |
| T2 (ms)                                        | 113.34  | 0.855(0.546-0.801) | 89              | 70              | 81      | 82      | <0.0001  |
| PD (pu)                                        | 80.67   | 0.568(0.429-0.700) | 89              | 35              | 66      | 70      | 0.408    |
| ADC ( $\times 10^{-3} \text{ mm}^2/\text{s}$ ) | 1.60    | 0.906(0.787-0.971) | 82              | 95              | 96      | 79      | <0.0001  |
| T2+ADC                                         | 0.55    | 0.950(0.845-0.992) | 89              | 90              | 93      | 86      | <0.0001  |

ADC, apparent diffusion coefficient; PD, proton density; AUC, area under the curve; NPV, negative predictive value; PPV, positive predictive value, \*p values are for the differences between benign and malignant head and neck tumors.





ROC curves for differentiating malignant from benign lesions are depicted in Figure 3. T2+ADC (cut-off value,  $> 0.55$ ; AUC, 0.950) showed optimal diagnostic performance, which was better than that of T1 (cut-off value,  $\leq 1675.84$  ms; AUC, 0.698), T2 (cut-off value,  $\leq 113.24$  ms; AUC, 0.855) and PD (cut off value,  $> 80.67$  pu; AUC, 0.568) ( $P = 0.0030$ ,  $0.0464$ , and  $P < 0.0001$ , respectively). The diagnostic performance of ADC was better than T1 and PD ( $P = 0.0138$  and  $0.0005$ , respectively), but the difference in AUC between ADC (cut-off value,  $\leq 1.60 \times 10^{-3}$  mm<sup>2</sup>/s; AUC, 0.906) and T2+ADC or T2 did not reach significance ( $P = 0.2648$  and  $0.4604$ , respectively). The diagnostic performance of T2 was better than PD ( $P = 0.0075$ ); however, the difference in AUC between T2 and T1 did not reach statistical significance ( $P = 0.0549$ ).

## 4 Discussion

We examined the tissue magnetic property parameters acquired on the synthetic MRI with MAGIC and ADC acquired on the FSE-PROPELLER DWI in discriminating malignant from benign HN lesions. The purpose of evaluating parameter correlation is to identify preferred parameters and facilitate the translation of scientific research into clinical practice. Except for PD values, T1, T2, and ADC values were lower in malignant than in benign HN tumors. ADC, T1, and T2 values are widely used parameters for differentiating malignant from benign HN tumors. The diagnostic performance of the T2 value is comparable to the ADC value. However, the diagnostic performance of the T1 and PD values was not as good as that of the ADC value. We also found that T2+ADC showed optimal diagnostic performance.

In this study, the malignant tumor had a lower T1 value related to hyper-cellularity, smaller extracellular space, and lower free water

content (18). Meng et al. found that the T1 value for nasopharyngeal carcinoma was significantly lower than that for benign hyperplasia in the nasopharynx, regardless of the ROI used (18). Contrary Gao and his team found no difference in the T1 values between malignant and benign breast lesions (21). The different types of tumors in these studies may contribute to the discrepancy.

T2 value can be affected by various factors, including the main magnetic field strength and the intrinsic properties of the tissue and the environment. Tissue water content is the most important influencing factor (22). A previous study reported a linear relationship between the T2 value and water content (23); thus, increased T2 values indicate increased tissue water content (24). In our study, 75% of benign HN tumors were pleomorphic adenomas followed by basal cell adenomas, and 64% of malignant HN tumors were squamous cell carcinomas followed by lymphomas. We speculate that the higher T2 values of benign HN tumors could be ascribed to their tissue composition, lower cell density, and higher free water content (5). Nevertheless, the lower T2 values in malignant tumors are due to the increased solid components, smaller extracellular spaces, and lower free water content (5, 18, 25). Several studies have reported higher T2 values in benign breast lesions vs. in malignant lesions (21). This difference could be attributed to increased tissue water content or interaction between water and alkaline metal cations in the pathological tissue (22).

PD value, which primarily reflects the water content in tissue, is frequently used in brain imaging (26). This study found that the difference in PD value in the malignant compared with benign HN tumors was not statistically significant. Yet, Gao et al. demonstrated that the PD value was significantly lower in malignant than that benign breast lesions (21). The different types of tumors enrolled in these studies may contribute to the contradiction.

Here we found that the ADC values of malignant tumors were significantly lower than those of benign tumors. Higher ADC values correlate with lower cellularity (27). Malignant tumors demonstrate lower ADC values than benign tumors due to their relatively higher cellularity (4). Srinivasan et al. also found that malignant lesions showed lower ADC values than benign lesions (28).

In this study, the overall diagnostic performance of synthetic MRI-derived parameters in discriminating malignant from the benign HN lesions was inferior to the ADC value. However, the diagnostic performance showed no significant differences between ADC and T2 values. Also, T2+ADC showed optimal diagnostic efficacy in distinguishing malignant from benign tumors; T2+ADC showed a significantly higher differential performance vs. T1, T2, or PD value alone, but it did not improve the diagnostic performance of the ADC value. Despite this, the PD, T1, and T2 are intrinsic magnetic properties and independent from the MRI scanners or scanning parameters at a given field strength (15), predicting the potential advantage of using synthetic MRI-derived parameters compared to the ADC value alone (17). In addition, synthetic MRI can generate multiple contrast-weighted images and quantification maps in a single scan, greatly improving work efficiency (10). Thus, synthetic MRI plus FSE-PROPELLER DWI might be a promising tool for differentiating benign from malignant HN lesions.



The present study has a few limitations. First, this is a single-center retrospective study with small sample size, next, we will expand the sample size for further study. Second, ROIs were manually drawn at the slice with the largest tumor diameter, leading to potential operator errors. In the future, the whole tumor should be selected to determine whether the tumor volume is more meaningful and accurate for tumor characterizing. Finally, test-retest repeatability was not assessed.

## 5 Conclusion

The quantitative T1, T2, and PD values obtained by synthetic MRI and ADC value obtained by FSE-PROPELLER DWI helped discriminate malignant from benign HN tumors. The overall diagnostic performance of synthetic MRI was inferior to FSE-PROPELLER DWI. However, the T2 value was comparable to the ADC value, and the combination of synthetic MRI and FSE-PROPELLER DWI could provide improved diagnostic efficacy.

## Data availability statement

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

## Author contributions

BW: Conceptualization, Methodology, Software, Data curation, Writing - original draft, Visualization, Investigation, Writing - review & editing. ZZ: Data curation, Writing - original draft, Visualization, Investigation, Writing - review & editing. JZ: data analysis, statistical analysis. LL: Data curation. ZL: Data curation, Writing - original draft. XM: Visualization, Investigation. KW:

Conceptualization, Methodology, Software, Writing - review & editing. LX: Conceptualization, Methodology, Software, Validation, Writing - review & editing. YZ: Supervision. JC: Supervision. All authors contributed to the article and approved the submitted version.

## Funding

Henan Medical Science and Technology Research Plan (LHGJ20190157) supported this study.

## Acknowledgments

We are particularly grateful to everyone who helped us with our article. This work was supported by the First Affiliated Hospital of Zhengzhou University, especially the Department of MRI.

## 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. Law BKH, King AD, Ai QY, Poon DMC, Chen W, Bhatia KS, et al. Head and neck tumors: Amide proton transfer MRI. *Radiology* (2018) 288:782–90. doi: 10.1148/radiol.2018171528
2. Noij DP, Martens RM, Marcus JT, de Bree R, Leemans CR, Castelijns JA, et al. Intravoxel incoherent motion magnetic resonance imaging in head and neck cancer: A systematic review of the diagnostic and prognostic value. *Oral Oncol* (2017) 68:81–91. doi: 10.1016/j.oraloncology.2017.03.016
3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. *CA Cancer J Clin* (2017) 67:7–30. doi: 10.3322/caac.21387
4. Payabvash SMD. Quantitative diffusion magnetic resonance imaging in head and neck tumors. *Quant Imaging Med Surg* (2018) 8:1052–65. doi: 10.21037/qims.2018.10.14
5. Baohong W, Jing Z, Zanzia Z, Kun F, Liang L, Eryuan G, et al. T2 mapping and readout segmentation of long variable echo-train diffusion-weighted imaging for the differentiation of parotid gland tumors. *Eur J Radiol* (2022) 151:110265. doi: 10.1016/j.ejrad.2022.110265
6. Cheng J, Shao S, Chen W, Zheng N. Application of diffusion kurtosis imaging and dynamic contrast-enhanced magnetic resonance imaging in differentiating benign and Malignant head and neck lesions. *J Magn Reson Imaging* (2022) 414–23. doi: 10.1002/jmri.27885
7. Liu YJ, Lee YH, Chang HC, Chung HW, Wang CW, Juan CH, et al. Imaging quality of PROPELLER diffusion-weighted MR imaging and its diagnostic performance in distinguishing pleomorphic adenomas from Warthin tumors of the parotid gland. *NMR Biomed* (2020) 33:e4282. doi: 10.1002/nbm.4282
8. Granberg T, Uppman M, Hashim F, Cananau C, Nordin LE, Shams S, et al. Clinical feasibility of synthetic MRI in multiple sclerosis: A diagnostic and volumetric validation study. *AJNR Am J Neuroradiol* (2016) 37:1023–9. doi: 10.3174/ajnr.A4665
9. Park S, Kwack KS, Lee YJ, Gho SM, Lee HY. Initial experience with synthetic MRI of the knee at 3T: Comparison with conventional T1 weighted imaging and T2 mapping. *Br J Radiol* (2017) 90:20170350. doi: 10.1259/bjr.20170350
10. Krauss W, Gunnarsson M, Andersson T, Thunberg P. Accuracy and reproducibility of a quantitative magnetic resonance imaging method for concurrent measurements of tissue relaxation times and proton density. *Magn Reson Imaging* (2015) 33:584–91. doi: 10.1016/j.mri.2015.02.013
11. Boudabbous S, Neroladaki A, Bagetakos I, Hamard M, Delattre BM, Vargas MI. Feasibility of synthetic MRI in knee imaging in routine practice. *Acta Radiol Open* (2018) 7:205846011876968. doi: 10.1177/2058460118769686
12. Betts AM, Leach JL, Jones BV, Zhang B, Serai S. Brain imaging with synthetic MR in children: clinical quality assessment. *Neuroradiology* (2016) 58:1017–26. doi: 10.1007/s00234-016-1723-9

13. Chougar L, Hagiwara A, Andica C, Hori M, Aoki S. Synthetic MRI of the knee: New perspectives in musculoskeletal imaging and possible applications for the assessment of bone marrow disorders. *Br J Radiol* (2018) 91:20170886. doi: 10.1259/bjr.20170886
14. Zhang W, Zhu J, Xu X, Fan G. Synthetic MRI of the lumbar spine at 3.0 T: feasibility and image quality comparison with conventional MRI. *Acta Radiol* (2020) 61:461–70. doi: 10.1177/0284185119871670
15. Cui Y, Han S, Liu M, Wu PY, Zhang W, Zhang J, et al. Diagnosis and grading of prostate cancer by relaxation maps from synthetic MRI. *J Magn Reson Imaging* (2020) 52:552–64. doi: 10.1002/jmri.27075
16. Jung Y, Gho SM, Back SN, Ha T, Kang DK, Kim TH. The feasibility of synthetic MRI in breast cancer patients: Comparison of T2 relaxation time with multiecho spin echo T2 mapping method. *Br J Radiol* (2019) 92:20180479. doi: 10.1259/bjr.20180479
17. Cai Q, Wen Z, Huang Y, Li M, Ouyang L, Ling J, et al. Investigation of synthetic magnetic resonance imaging applied in the evaluation of the tumor grade of bladder cancer. *J Magn Reson Imaging* (2021) 54:1989–97. doi: 10.1002/jmri.27770
18. Meng T, He H, Liu H, Lv X, Huang C, Zhong L, et al. Investigation of the feasibility of synthetic MRI in the differential diagnosis of non-keratinising nasopharyngeal carcinoma and benign hyperplasia using different contoured methods for delineation of the region of interest. *Clin Radiol* (2021) 76:238.e9–238.e15. doi: 10.1016/j.crad.2020.10.010
19. Matsuda M, Fukuyama N, Matsuda T, Kikuchi S, Shiraishi Y, Takimoto Y, et al. Utility of synthetic MRI in predicting pathological complete response of various breast cancer subtypes prior to neoadjuvant chemotherapy. *Clin Radiol* (2022) 77:855–63. doi: 10.1016/j.crad.2022.06.019
20. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* (1988) 44:837–45.
21. Gao W, Zhang S, Guo J, Wei X, Li X, Diao Y, et al. Investigation of synthetic relaxometry and diffusion measures in the differentiation of benign and Malignant breast lesions as compared to BI-RADS. *J Magn Reson Imaging* (2021) 53:1118–27. doi: 10.1002/jmri.27435
22. Liu L, Yin B, Shek K, Geng D, Lu Y, Wen J, et al. Role of quantitative analysis of T2 relaxation time in differentiating benign from Malignant breast lesions. *J Int Med Res* (2018) 46:1928–35. doi: 10.1177/0300060517721071
23. Chaland B, Mariette F, Marchal P, De Certaines J. <sup>1</sup>H nuclear magnetic resonance relaxometric characterization of fat and water states in soft and hard cheese. *J Dairy Res* (2000) 67:609–18. doi: 10.1017/S0022029900004398
24. Lüssea S, Claassen H, Gehrke T, Hassenpflug J, Schünke M, Heller M, et al. Evaluation of water content by spatially resolved transverse relaxation times of human articular cartilage. *Magn Reson Imaging* (2000) 18:423–30. doi: 10.1016/S0730-725X(99)00144-7
25. Wu Q, Zhu LN, Jiang JS, Bu SS, Xu XQ, Wu FY. Characterization of parotid gland tumors using T2 mapping imaging: initial findings. *Acta Radiol* (2020) 61:629–35. doi: 10.1177/0284185119875646
26. Gracien RM, Reitz SC, Hof SM, Fleischer V, Zimmermann H, Droby A, et al. Changes and variability of proton density and T1 relaxation times in early multiple sclerosis: MRI markers of neuronal damage in the cerebral cortex. *Eur Radiol* (2016) 26:2578–86. doi: 10.1007/s00330-015-4072-x
27. Chen L, Liu M, Bao J, Xia Y, Zhang J, Zhang L, et al. The correlation between apparent diffusion coefficient and tumor cellularity in patients: A meta-analysis. *PloS One* (2013) 8:e79008. doi: 10.1371/journal.pone.0079008
28. Srinivasan A, Dvorak R, Perni K, Rohrer S, Mukherji SK. Differentiation of benign and Malignant pathology in the head and neck using 3T apparent diffusion coefficient values: Early experience. *AJNR Am J Neuroradiol* (2008) 29:40–4. doi: 10.3174/ajnr.A0743



## OPEN ACCESS

## EDITED BY

Sharon R. Pine,  
University of Colorado Anschutz Medical  
Campus, United States

## REVIEWED BY

Ailin Zhao,  
Sichuan University, China  
Qin Genggeng,  
Southern Medical University, China

## \*CORRESPONDENCE

Zhi-Xing Liu  
✉ 1214582369@qq.com  
Chen Li  
✉ 1727237899@qq.com

<sup>†</sup>These authors have contributed equally to  
this work and share first authorship

<sup>‡</sup>These authors have contributed equally to  
this work

RECEIVED 28 July 2023

ACCEPTED 18 December 2023

PUBLISHED 11 January 2024

## CITATION

Mao Y, Jiang L, Wang J-L, Chen F-Q,  
Zhang W-P, Liu Z-X and Li C (2024) Radiomic  
nomogram for discriminating parotid  
pleomorphic adenoma from parotid  
adenolymphoma based on  
grayscale ultrasonography.  
*Front. Oncol.* 13:1268789.  
doi: 10.3389/fonc.2023.1268789

## COPYRIGHT

© 2024 Mao, Jiang, Wang, Chen, Zhang, Liu  
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.

# Radiomic nomogram for discriminating parotid pleomorphic adenoma from parotid adenolymphoma based on grayscale ultrasonography

Yi Mao<sup>1†</sup>, LiPing Jiang<sup>1†</sup>, Jing-Ling Wang<sup>1</sup>, Fang-Qun Chen<sup>1</sup>,  
Wie-Ping Zhang<sup>1</sup>, Zhi-Xing Liu<sup>1,2\*‡</sup> and Chen Li<sup>1\*‡</sup>

<sup>1</sup>Department of Ultrasound, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University, Nanchang, Jiangxi, China, <sup>2</sup>Department of Ultrasound, GanJiang New District Peoples Hospital, Nanchang, Jiangxi, China

**Objectives:** To differentiate parotid pleomorphic adenoma (PA) from adenolymphoma (AL) using radiomics of grayscale ultrasonography in combination with clinical features.

**Methods:** This retrospective study aimed to analyze the clinical and radiographic characteristics of 162 cases from December 2019 to March 2023. The study population consisted of a training cohort of 113 patients and a validation cohort of 49 patients. Grayscale ultrasonography was processed using ITP-Snap software and Python to delineate regions of interest (ROIs) and extract radiomic features. Univariate analysis, Spearman's correlation, greedy recursive elimination strategy, and least absolute shrinkage and selection operator (LASSO) correlation were employed to select relevant radiographic features. Subsequently, eight machine learning methods (LR, SVM, KNN, RandomForest, ExtraTrees, XGBoost, LightGBM, and MLP) were employed to build a quantitative radiomic model using the selected features. A radiomic nomogram was developed through the utilization of multivariate logistic regression analysis, integrating both clinical and radiomic data. The accuracy of the nomogram was assessed using receiver operating characteristic (ROC) curve analysis, calibration, decision curve analysis (DCA), and the Hosmer–Lemeshow test.

**Results:** To differentiate PA from AL, the radiomic model using SVM showed optimal discriminatory ability (accuracy = 0.929 and 0.857, sensitivity = 0.946 and 0.800, specificity = 0.921 and 0.897, positive predictive value = 0.854 and 0.842, and negative predictive value = 0.972 and 0.867 in the training and validation cohorts, respectively). A nomogram incorporating rad-Signature and clinical features achieved an area under the ROC curve (AUC) of 0.983 (95% confidence interval [CI]: 0.965–1) and 0.910 (95% CI: 0.830–0.990) in the training and validation cohorts, respectively. Decision curve analysis showed that the

nomogram and radiomic model outperformed the clinical-factor model in terms of clinical usefulness.

**Conclusion:** A nomogram based on grayscale ultrasonic radiomics and clinical features served as a non-invasive tool capable of differentiating PA and AL.

#### KEYWORDS

ultrasonography, radiomics, parotid tumor, nomogram, wavelet transformation

## Introduction

Parotid gland tumors are the most common type of salivary gland tumors, with approximately 80%–85% of them being benign. The primary types of these tumors are pleomorphic adenoma (PA) and adenolymphoma (AL) (1), and both of them share similar characteristics, such as slow growth, painlessness, and well-defined borders. However, differentiations between PA and AL are crucial for clinical diagnosis and treatment. On radiological examinations, AL shows more heterogeneous density and signal compared to PA, often accompanied by multiple small cystic changes and increased blood flow (2). PA is more likely to occur in the deep regions of the parotid gland, typically presenting as lobulated, with a higher risk of malignant transformation and recurrence (3). Therefore, PA usually requires a tumor and superficial parotidectomy, along with facial nerve dissection. AL, however, typically only requires partial parotidectomy. However, in PA cases, tumor cells can be detected at the resection margins in 41.9% of cases (4). This could be one of the reasons why PA is more prone to relapse. To distinguish between the two types of parotid gland tumors at an early stage, a fine-needle aspiration biopsy (FNAB) is commonly used as an auxiliary diagnostic tool (5). It has high accuracy in the diagnosis of both benign and malignant tumors. However, FNAB is an invasive procedure and carries the risk of needle-track seeding and facial nerve palsy (6).

Ultrasonic examinations can reflect differences in signal scattering and speckling patterns, which correlate with variations in parotid gland morphology and increasing tissue stiffness (2). Compared to FNAB, grayscale ultrasonic examination is a non-invasive, cost-effective, and user-friendly imaging technique. However, differentiating between PA and AL using grayscale ultrasonic examination can be challenging for sonographers. Some morphological features, long-to-short diameter ratio (L/S) ratio, and ultrasonographic shear wave elastography have limited utility in distinguishing between the two types (7, 8). Therefore, visible differences discernible by the naked eye do not significantly improve the diagnostic accuracy of medical imaging.

Radiomics is a rapidly growing discipline that utilizes machine learning to extract quantitative information from medical images

like CT, MR, US, and predict outcomes in cancer research (9–11). For head and neck tumors, radiomic features from T2-weighted MR imaging (T2WI) and contrast-enhanced T1-weighted MR imaging (CE-T1WI) can predict cancer staging pre-operatively (12). Additionally, radiomic features from CT and PET scans can accurately determine if oropharyngeal squamous cell carcinoma is infected with the HPV (P16) virus (13). Radiomics has also shown success in assessing early treatment effects (14) and radiotherapy complications in nasopharyngeal cancer (15). In summary, radiomic analysis of various medical imaging modalities holds potential for improving diagnosis, prognosis, and personalized treatment of head and neck cancers.

Wavelet transformation is created *via* dilatation and translation of the mother wavelet (16). These modifications provide a spatial/frequency representation of the signal, indicating that the wavelet coefficients act as a projection of the original signal onto a multi-resolution subspace. The high-pass filter also draws attention to the grayscale changes in the image, improving the presentation of image details and texture information. The low-pass filter, however, blurs the differences in grayscale, obscuring the finer details of the image and emphasizing its main characteristics (17). The radiomic model's texture features can be separated further. Studies have shown that, compared to the original radiomics, wavelet-transformed radiomics perform better in assessing COVID-19 lung lesions (18).

The purpose of our study is to investigate whether radiology based on grayscale ultrasonography can distinguish PA and AL and whether the nomogram combined with clinical and radiological features can facilitate and accurately help to distinguish these two benign tumors.

## Materials and methods

### Ethics statement

This study adhered to the principles outlined in the Declaration of Helsinki and received approval from the local ethics review board. Written, informed consent was obtained from all participants.

## Selection of participants

We retrospectively analyzed patients with parotid tumors undergoing grayscale ultrasonic examination at the local hospital from December 2019 to March 2023. The inclusion criteria were as follows: 1) preoperative two-dimensional ultrasonography confirmed the presence of a parotid tumor. 2) A postoperative histopathological examination confirmed the diagnosis of PA and AL. 3) There was no history of fine-needle aspiration (FNA), radiotherapy, or other treatments. 4) Complete clinical and data records were available. 5) A preoperative ultrasound examination was performed within 1 week. The exclusion criteria were as follows: 1) the maximum diameter of the tumor was less than 1 cm; 2) the images were not clear, with incomplete visualization of the tumor and significant artifacts; 3) the concurrent presence of other organ tumors.

## Ultrasonography procedures

The bilateral parotid glands were scanned using high-end ultrasound diagnostic equipment such as Siemens ACUson Sequoia, GE LOGIQ E11, and Philips EPIQ 7. A high-frequency linear array probe was used for the examination. The maximum diameter of the parotid gland masses was saved in the machine's memory in a cross-sectional view and exported in DICOM format for subsequent analysis.

## Image segmentation

All ultrasound images were imported into the ITK-SNAP (<http://www.itksnap.org>) software. Two ultrasound physicians with 6 years of experience in the field delineated the tumor margins by carefully outlining them and selecting the maximum section of the tumor to delineate a region of interest (ROI). The delineation was subsequently reviewed and approved by a senior physician. In case of any disagreements, a group discussion was held to reach a consensus.

## Feature extraction

The images and ROIs extracted from the ITK-SNAP software were imported into Python (version 3.11) for further analysis. Handcrafted features were extracted using an in-house feature analysis program implemented in Pyradiomics (<https://pyradiomics.readthedocs.io>). These features can be categorized into three groups: I) geometry, II) intensity, and III) texture. There were 14 geometry features, 306 intensity features, and 1,241 texture features comprised of the Gray Level Co-Occurrence Matrix (GLCM), Gray Level Dependence Matrix (GLDM), Gray Level Run Length Matrix (GLRLM), Gray Level Size Zone Matrix (GLSZM), and Neighborhood Gray Tone Difference Matrix (NGTDM).

The ROIs were delineated by two sonographers, and the interobserver agreement was evaluated using the interclass

correlation coefficient (ICC) analysis. ICC values higher than 0.75 were considered to have good consistency and were selected for further analysis. Patients were randomly divided into two cohorts with a ratio of 7:3 for training and validation purposes, respectively.

## Feature selection

After applying z-score normalization, the t-test and Mann–Whitney U test were performed on all radiomic features. Only features with a p-value <0.05 were retained. For features exhibiting high repeatability, Spearman's rank correlation coefficient was used to assess the correlation between features. If the correlation coefficient between any two features exceeded 0.9, only one of them was retained.

To identify the optimal feature subset, the least absolute shrinkage and selection operator (LASSO) (19) algorithm was employed. LASSO shrinks all regression coefficients toward zero and sets the coefficients of irrelevant features to exactly zero. A 10-fold cross-validation with minimum criteria was used to determine the optimal lambda ( $\lambda$ ) value, which yielded the lowest cross-validation error.

## Model construction and validation

### Radiomic and clinical models

After performing LASSO feature screening, the final selected features were input into machine learning models such as LR, SVM, RandomForest, and XGBoost. The coefficients of the features were used to calculate a radiomic quality signature (rad-Signature). Clinical features used for building the same machine learning models were selected based on a baseline statistic with a p-value <0.05.

### Radiomic nomogram

A radiomic nomogram was developed by combining the radiomic signature and clinical features. The diagnostic efficacy of the radiomic nomogram was tested in the validation cohort, and receiver operating characteristic (ROC) curves were plotted to evaluate its diagnostic performance. Calibration curves were used to evaluate the calibration efficiency of the nomogram, and the Hosmer–Lemeshow analytical fit was employed to assess its calibration ability. Additionally, decision curve analysis (DCA) was used to evaluate the clinical utility of the predictive models.

## Statistical analyses

Statistical analysis of the data was performed using SPSS 26.0 and Python 3.11. Continuous variables are presented as mean  $\pm$  standard deviation, while categorical variables are reported as counts (n). The independent samples t-test was used to analyze clinical data, and the chi-square test was applied for categorical variables. A significance level of  $p < 0.05$  was considered statistically significant.



## Results

### Clinical characteristics

The flowchart depicting the process of patient selection is presented in Figure 1. Table 1 displays the clinical and imaging data of the 162 patients included in this study. Out of the total, 105 were confirmed to have PA, and 57 were diagnosed with adenoid cystic carcinoma (AL). The clinical characteristics of all 162 subjects are summarized in Table 1. In the PA group, the average age was  $43.49 \pm 15.67$  years, with a male-to-female gender ratio of 0.91:1. Among the AL patients, the average age was  $61.50 \pm 10.08$  years, and the male-to-female gender ratio was 10.5:1.

The 162 subjects were randomly divided into training and validation cohorts in a 7:3 ratio. Therefore, the training cohort comprised 113 cases (76 PA and 37 AL), while the remaining 49 patients (29 PA and 20 AL) were assigned to the validation cohort. There were no significant differences in clinical features between the training and validation cohorts ( $p$ -value  $< 0.05$ ).

### Feature selection, model construction, and validation

The course of processing radiomics is shown in Figure 2. From the grayscale ultrasonography for each participant, 1,561 radiomics were extracted; 294 features were selected after univariate analysis

and ICC; and 70 features were retained after being filtered using Spearman's correlation (Figure 3; Spearman's correlation of each feature). The radiomic feature selection was performed using LASSO logistic regression, resulting in 18 selected radiomic features. The coefficients and mean standard error (MSE) from the 10-fold validation are presented in Figure 4. These features were utilized to construct the radiomic signature. The final formula for calculating rad-Signature and the corresponding coefficients is depicted in Figure 5.

### Nomogram performance and validation

The model constructed using clinical features such as age, maximum diameter, and smoking status showed good predictive performance (accuracy = 0.850 and 0.776, sensitivity = 0.811 and 0.950, specificity = 0.868 and 0.679, positive predictive value = 0.750 and 0.655, and negative predictive value = 0.904 and 0.950 in the training and validation cohorts, respectively). Similarly, the imaging-based radiomic features (Table 2), especially the SVM model, demonstrated excellent predictive performance (accuracy = 0.929 and 0.857, sensitivity = 0.946 and 0.800, specificity = 0.921 and 0.897, positive predictive value = 0.854 and 0.842, and negative predictive value = 0.972 and 0.867 in the training and validation cohorts, respectively). Furthermore, incorporating the results of the radiomic model into the clinical model improved the predictive performance of the combined model (accuracy = 0.947 and 0.857,

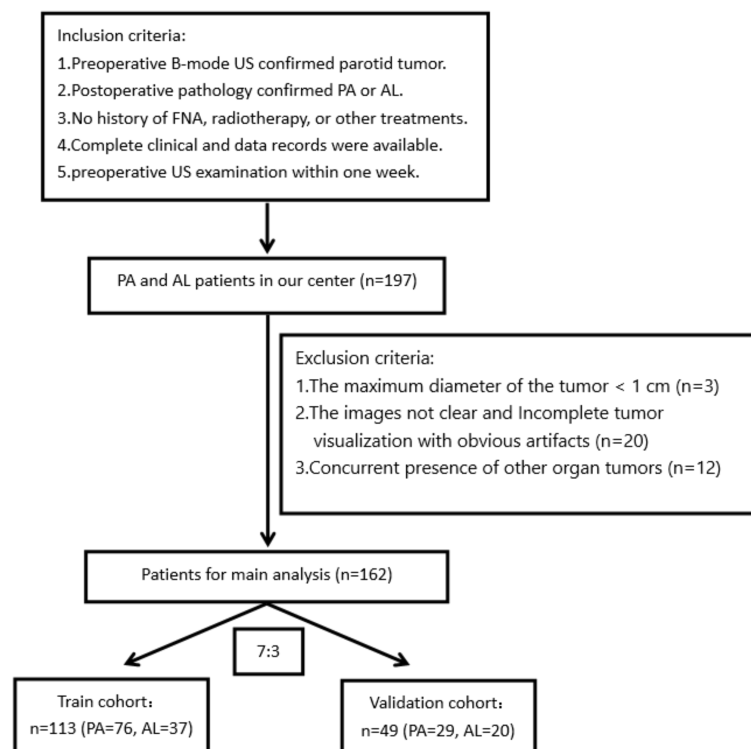


FIGURE 1  
The flowchart of the patient selection process.

TABLE 1 Patient's characteristics at baseline.

| Feature name |                 | Train         | PA            | AL            | p-Value     | Test          | PA            | AL           | p-Value     |
|--------------|-----------------|---------------|---------------|---------------|-------------|---------------|---------------|--------------|-------------|
| Age          |                 | 49.17 ± 16.74 | 42.82 ± 15.40 | 62.22 ± 10.87 | <0.001      | 50.86 ± 16.16 | 44.28 ± 17.12 | 60.40 ± 8.10 | <0.001      |
| Max D        |                 | 25.04 ± 9.33  | 23.30 ± 8.94  | 28.62 ± 9.21  | 0.004009086 | 25.88 ± 8.35  | 22.38 ± 7.51  | 30.95 ± 6.88 | <0.001      |
| Sex          |                 |               |               |               | <0.001      |               |               |              | 0.001820478 |
|              | Male subjects   | 67 (59.29)    | 33 (43.42)    | 34 (91.89)    |             | 33 (67.35)    | 14 (48.28)    | 19 (95.00)   |             |
|              | Female subjects | 46 (40.71)    | 43 (56.58)    | 3 (8.11)      |             | 16 (32.65)    | 15 (51.72)    | 1 (5.00)     |             |
| Smoking      |                 |               |               |               | <0.001      |               |               |              | <0.001      |
|              | No              | 77 (68.14)    | 63 (82.89)    | 14 (37.84)    |             | 29 (59.18)    | 24 (82.76)    | 5 (25.00)    |             |
|              | Yes             | 36 (31.86)    | 13 (17.11)    | 23 (62.16)    |             | 20 (40.82)    | 5 (17.24)     | 15 (75.00)   |             |
| Number       |                 |               |               |               | <0.001      |               |               |              | 0.013690557 |
|              | Single          | 93 (82.30)    | 70 (92.11)    | 23 (62.16)    |             | 39 (79.59)    | 27 (93.10)    | 12 (60.00)   |             |
|              | Multiple        | 20 (17.70)    | 6 (7.89)      | 14 (37.84)    |             | 10 (20.41)    | 2 (6.90)      | 8 (40.00)    |             |
| Position     |                 |               |               |               | 0.629061941 |               |               |              | 0.842154851 |
|              | Right           | 62 (54.87)    | 40 (52.63)    | 22 (59.46)    |             | 29 (59.18)    | 18 (62.07)    | 11 (55.00)   |             |
|              | Left            | 51 (45.13)    | 36 (47.37)    | 15 (40.54)    |             | 20 (40.82)    | 11 (37.93)    | 9 (45.00)    |             |

PA, parotid pleomorphic adenoma; AL, adenolymphoma.

sensitivity = 1 and 0.950, specificity = 0.921 and 0.931, positive predictive value = 0.860 and 0.882, and negative predictive value = 1 and 0.844 in the training and validation cohorts, respectively). The performance comparison of the three models is presented in Table 3 and Figure 6 (DeLong test,  $p < 0.005$  for the training and validation cohorts). A nomogram combining clinical features and rad-Signature was developed (Figure 6), and its calibration curve demonstrated consistent predictive and observed effects in both the training and validation cohorts.

To assess the calibration ability of the developed nomogram, the Hosmer–Lemeshow test (20) was employed. The results indicated a good model fit ( $p$ -value  $> 0.05$ ), suggesting that the nomogram accurately captured the observed data and that there was no

significant difference between the predicted and observed outcomes. The DCA of the nomogram is depicted in Figure 6F. Furthermore, the DCA of the nomogram demonstrated a larger area under the curve compared to the clinical model. This indicates that both the nomogram and radiomic model have a greater net benefit in distinguishing between PA and AL.

## Discussion

In this study, we utilized radiomic features extracted from grayscale ultrasonography to assist in the early preoperative differentiation of two benign tumors, PA and AL, with the goal of aiding clinicians in selecting

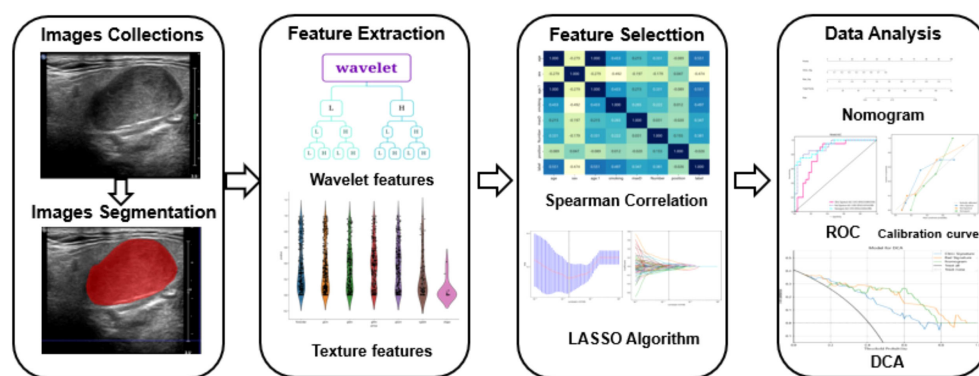


FIGURE 2

The flowchart detailing the radiomic processing steps employed in this study. The collected images were exported to ITK software for region of interest (ROI) delineation and image segmentation. Ultrasound radiomics were then extracted using Python software. Models were developed based on the clinical features of patients with pleomorphic adenoma (PA) or adenoid cystic carcinoma (AL). The models underwent calibration and validation processes to evaluate their performance.

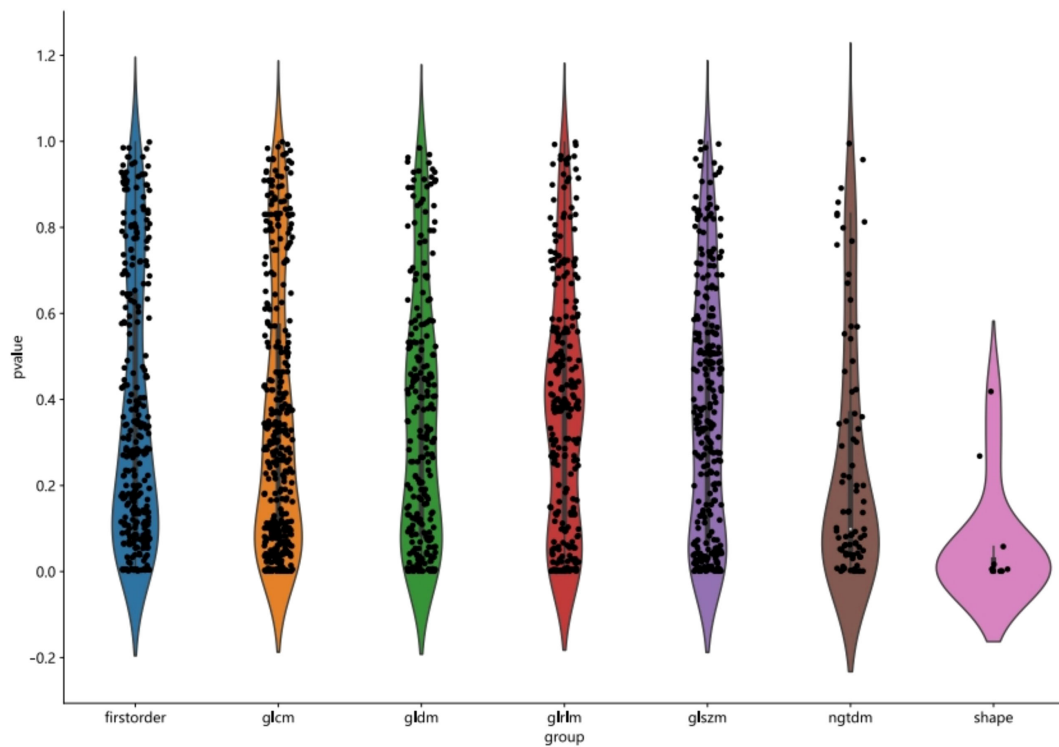


FIGURE 3  
Statistics of radiomic features.

appropriate diagnostic and treatment approaches. Previous studies have demonstrated that there are important differences between PA and AL in terms of clinical features and traditional parameters, including smoking history, age, and the presence of multiple lesions (15). However, the effectiveness of these factors in a comprehensive analysis is inconsistent, with varying areas under the ROC curve (AUC) values ranging from 0.68 to 0.95, leading to significant uncertainty in clinical diagnosis and treatment. Additionally, common ultrasound features such as the L/S and ultrasound grayscale ratio

(UGSR) have also shown poor performance ( $AUC = 0.74$ ) (7). In contrast, the radiomic SVM-based model that we constructed has demonstrated excellent performance in distinguishing between PA from AL, with AUC values of 0.956 in the training cohort and 0.903 in the validation cohort. Additionally, other models in our study have also demonstrated good performance in distinguishing between the two types of tumors, but for the RandomForest, ExtraTrees, and XGBoost models, there are significant differences in AUC between the training and validation cohorts, indicating overfitting of the models (11, 21).

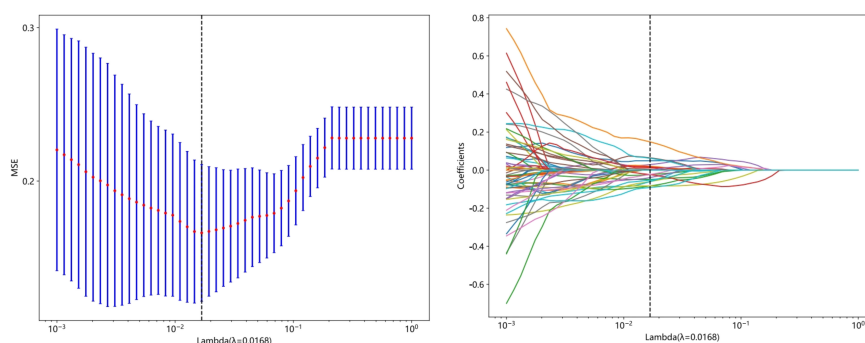
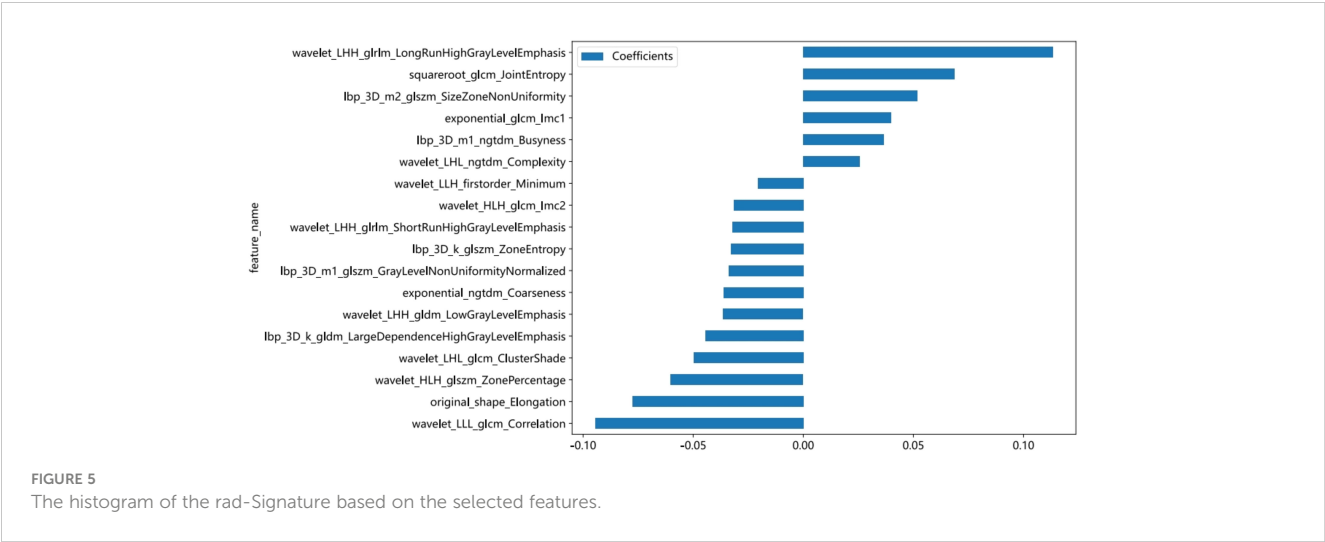


FIGURE 4  
The least absolute shrinkage and selection operator (LASSO) algorithm was employed for feature selection. In the LASSO model, a 10-fold cross-validation approach was utilized to determine the optimal tuning parameter ( $\lambda$ ). The minimum criterion was used to select the best values, and vertical lines were drawn to indicate the true selection points. Additionally, a 10-fold cross-validation was performed to identify the selected value in the  $\lambda$  sequence, resulting in 18 features with non-zero coefficients.



However, it is important to note that models with a large number of input parameters or high degrees of freedom may have a tendency to overfit the data by memorizing it. Consequently, when analyzing the features, the model may react to random fluctuations in the data, which is undesirable in accurate feature analysis.

Ultrasound images displaying both PA and AL appear on ultrasound as localized enlargements of the salivary gland with regular morphology and well-defined borders, presenting as well-circumscribed hypoechoic masses. Matsuda (22) indicated that 63.2% of PA cases belonged to the category of no anechoic area homogeneous tumors, while 53.3% of AL cases were classified as

multiple and sponge-like anechoic area heterogeneous tumors. However, Jiang (23) and Rong (2) believed that there were no statistically significant differences observed in the sonographic features of boundaries, echo pattern, homogeneity, calcification, and distal acoustic enhancement between PA and AL. We believe that this discrepancy is only related to the number of samples. However, it is undeniable that AL is more susceptible to infection and cystic degeneration, characterized by a loose tissue texture with numerous small cysts that create echo-free areas. In this study, most of the features used for modeling were obtained through wavelet transformation, revealing more layered variation and information

TABLE 2 Performance contributions of various radiological classifier models in classification.

| Model name   | Cohort     | ACC   | AUC   | 95% CI        | SEN   | SPE   | PPV   | NPV   | Precision | Recall | F1    | Threshold |
|--------------|------------|-------|-------|---------------|-------|-------|-------|-------|-----------|--------|-------|-----------|
| LR           | Train      | 0.858 | 0.898 | 0.838, −0.958 | 0.730 | 0.921 | 0.818 | 0.875 | 0.818     | 0.730  | 0.771 | 0.417     |
|              | Validation | 0.878 | 0.936 | 0.870, −1     | 1     | 0.793 | 0.769 | 1     | 0.769     | 1      | 0.870 | 0.181     |
| SVM          | Train      | 0.929 | 0.956 | 0.909, −1     | 0.946 | 0.921 | 0.854 | 0.972 | 0.854     | 0.946  | 0.897 | 0.230     |
|              | Validation | 0.857 | 0.903 | 0.818, −0.989 | 0.800 | 0.897 | 0.842 | 0.867 | 0.842     | 0.800  | 0.821 | 0.420     |
| KNN          | Train      | 0.814 | 0.895 | 0.840, −0.951 | 0.784 | 0.829 | 0.690 | 0.887 | 0.690     | 0.784  | 0.734 | 0.400     |
|              | Validation | 0.776 | 0.863 | 0.764, −0.962 | 0.800 | 0.786 | 0.696 | 0.846 | 0.696     | 0.800  | 0.744 | 0.400     |
| RandomForest | Train      | 1     | 1     | 1, −1         | 1     | 1     | 1     | 1     | 1         | 1      | 1     | 0.500     |
|              | Validation | 0.714 | 0.791 | 0.661, −0.920 | 0.850 | 0.621 | 0.607 | 0.857 | 0.607     | 0.850  | 0.708 | 0.300     |
| ExtraTrees   | Train      | 1     | 1     | 1, −1         | 1     | 1     | 1     | 1     | 1         | 1      | 1     | 1         |
|              | Validation | 0.816 | 0.849 | 0.741, −0.957 | 0.850 | 0.793 | 0.739 | 0.885 | 0.739     | 0.850  | 0.791 | 0.400     |
| XGBoost      | Train      | 1     | 1     | 1, −1         | 1     | 1     | 1     | 1     | 1         | 1      | 1     | 0.539     |
|              | Validation | 0.857 | 0.893 | 0.795, −0.992 | 0.750 | 0.931 | 0.882 | 0.844 | 0.882     | 0.750  | 0.811 | 0.384     |
| LightGBM     | Train      | 0.920 | 0.977 | 0.957, −0.998 | 0.973 | 0.895 | 0.818 | 0.986 | 0.818     | 0.973  | 0.889 | 0.314     |
|              | Validation | 0.857 | 0.902 | 0.819, −0.985 | 0.700 | 0.966 | 0.933 | 0.824 | 0.933     | 0.700  | 0.800 | 0.443     |
| MLP          | Train      | 0.862 | 0.923 | 0.877, −0.969 | 0.804 | 0.893 | 0.804 | 0.893 | 0.804     | 0.804  | 0.804 | 0.392     |
|              | Validation | 0.875 | 0.861 | 0.716, −1     | 0.818 | 0.950 | 0.818 | 0.905 | 0.818     | 0.818  | 0.818 | 0.491     |

ACC, accuracy; SEN, sensitivity; SPE, specificity; PPV, positive predictive value; NPV, negative predictive value; AUC, area under the receiver operating characteristic curve.

TABLE 3 Performance contributions of three different models in classification.

| Cohort     | Signature | ACC   | AUC   | 95% CI        | SEN   | SPE   | PPV   | NPV   | Precision | Recall | F1    | Threshold |
|------------|-----------|-------|-------|---------------|-------|-------|-------|-------|-----------|--------|-------|-----------|
| Train      | Clinic    | 0.850 | 0.853 | 0.774, −0.931 | 0.811 | 0.868 | 0.750 | 0.904 | 0.750     | 0.811  | 0.779 | 0.370     |
|            | Rad       | 0.929 | 0.956 | 0.909, −1     | 0.946 | 0.921 | 0.854 | 0.972 | 0.854     | 0.946  | 0.897 | 0.230     |
|            | Nomogram  | 0.947 | 0.983 | 0.965, −1     | 1     | 0.921 | 0.860 | 1     | 0.860     | 1      | 0.925 | 0.277     |
| Validation | Clinic    | 0.776 | 0.812 | 0.690, −0.934 | 0.950 | 0.679 | 0.655 | 0.950 | 0.655     | 0.950  | 0.776 | 0.194     |
|            | Rad       | 0.857 | 0.903 | 0.818, −0.989 | 0.800 | 0.897 | 0.842 | 0.867 | 0.842     | 0.800  | 0.821 | 0.420     |
|            | Nomogram  | 0.857 | 0.910 | 0.830, −0.990 | 0.750 | 0.931 | 0.882 | 0.844 | 0.882     | 0.750  | 0.811 | 0.440     |

ACC, accuracy; SEN, sensitivity; SPE, specificity; PPV, positive predictive value; NPV, negative predictive value; AUC, area under the receiver operating characteristic curve.

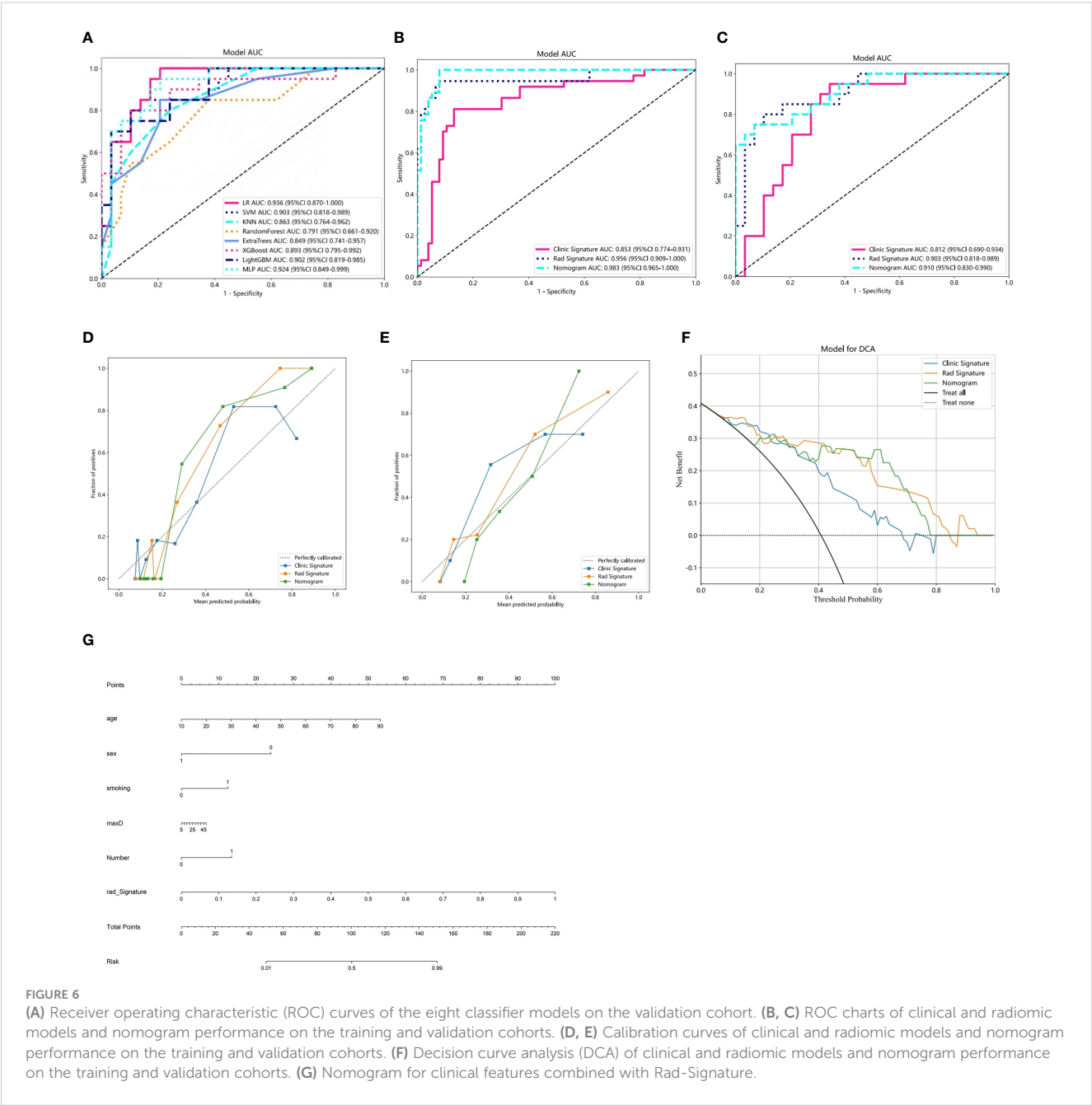


FIGURE 6 (A) Receiver operating characteristic (ROC) curves of the eight classifier models on the validation cohort. (B, C) ROC charts of clinical and radiomic models and nomogram performance on the training and validation cohorts. (D, E) Calibration curves of clinical and radiomic models and nomogram performance on the training and validation cohorts. (F) Decision curve analysis (DCA) of clinical and radiomic models and nomogram performance on the training and validation cohorts. (G) Nomogram for clinical features combined with Rad-Signature.



content in these feature maps. Among the 18 features used to construct the radiomic model, the most influential feature is wavelet\_LHH\_glrIm\_LongRunHighGrayLevelEmphasis, which describes the texture feature of long and high gray-level runs in the image. A higher value indicates the presence of longer and higher gray-level continuous texture features in the image (24). In our study, the feature value PA > AL can be seen in both the training and validation cohorts (Appendix 1, the average feature values of 18 modeling features in the training and validation sets). We believe that the characteristic of cystic lesions in AL results in a lower value of this feature compared to PA.

Previous studies have shown that radiomic research using CT and MR images performs well in differentiating PA and AL. Zheng (25) gathered 76 instances of PA and 34 cases of AL and built a model based on CT images with an AUC of 0.89 and an accuracy of 83.3%. Song (26) built a T1-2WI model based on MR images with an AUC of 0.90 and an accuracy of 86% after collecting 140 instances of PA and 112 cases of AL. The mutual information (MI) feature model that Fruehwald-Pallamar et al. (27) developed using CE-T1WI pictures had an accuracy of 81.8%. She gathered 13 cases of PA and 11 cases of AL. Similarly, Piludu et al. (28) enrolled 35 parotid PA and 20 AL to construct an SVM model using T2WI and ADC pictures, which was successful with an accuracy of 91.7%. Additionally, according to their studies, AL and PA could possibly be distinguished from one another on T1WI, T2WI, and ADC images by the characteristics of AL's cystic components.

Our study still established and validated a novel prognostic model using a nomogram-based approach to differentiate between PA and AL. The nomogram, as a predictive statistical model, not only provides a visual display of the relevant indicators influencing the outcomes in multiple regression analysis but also enables a simple graphical representation to predict survival probability, making the prediction simpler and more convenient (29, 30). We combined clinical features and rad-Signature and utilized a nomogram for prediction. The results showed that in both the training and validation cohorts, the AUC was higher than that of the single model. However, in the validation cohort, the specificity was 0.931 while the sensitivity was only 0.750, indicating high accuracy in identifying AL patients. Therefore, this prognostic model has certain clinical applicability. Zheng (25) developed and validated a novel prognostic model using a nomogram-based approach to differentiate between PA and AL. This model incorporated the CT Rad-score and independent clinical factors. The nomogram exhibited excellent discriminative performance, with an AUC of 0.98 in the training cohort and 0.95 in the validation cohort. However, when compared to the CT radiomic model (with an AUC of 0.89 in both the training and validation cohorts), the grayscale ultrasonography-based radiomic model in this study demonstrates higher accuracy and stability.

Nevertheless, our study has several limitations that should be noted. First, due to difficulties in disease epidemiology and obtaining qualified patient images, the sample size was limited, and we did not conduct an independent external validation. Future research should involve a larger dataset for further investigation. Second, our radiomic study only utilized conventional grayscale ultrasonography, which is the most commonly used scanning method. In the future, we plan to

incorporate more scanning technologies, such as Sound-Touch Elastography (STE) and contrast-enhanced ultrasound, to construct a multimodal radiomic model to further assist clinical diagnosis and treatment. Third, all images in our study were obtained from a single center. Therefore, we intend to include more types of devices and data centers in future studies to establish a multicenter radiomic model.

## Conclusion

Evaluating the imaging features of grayscale ultrasonography can significantly improve the diagnostic ability of clinical indicators for distinguishing between PA and AL. Based on this, the construction of a nomogram combining radiological features with clinical characteristics is also a highly accurate and non-invasive tool for distinguishing these two benign tumors.

## Data availability statement

The datasets presented in this article are not readily available because the institution requires full protection of patient privacy. Requests to access the datasets should be directed to CL or Z-XL at 1727237899@qq.com, 1214582369@qq.com.

## Ethics statement

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

YM: Data curation, Methodology, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. LJ: Data curation, Funding acquisition, Investigation, Supervision, Validation, Writing – original draft, Writing – review & editing. J-LW: Writing – original draft. F-QC: Data curation, Investigation, Software, Writing – original draft. W-PZ: Data curation, Investigation, Software, Writing – original draft. Z-XL: Data curation, Funding acquisition, Software, Supervision, Validation, Writing – review & editing. CL: Investigation, 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 is supported by the Key Research and Development Program of Jiangxi Province (20203BBGL73196PSS) and Youth Science Foundation of Jiangxi Province (GJJ210218) and Key Science and Technology Research Project in Jiangxi Province Department of Education (GJJ2200114).

## 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/fonc.2023.1268789/full#supplementary-material>

## References

- Eveson JW, Cawson RA. Salivary gland tumours. A review of 2410 cases with particular reference to histological types, site, age and sex distribution. *J Pathol* (1985) 146(1):51–8. doi: 10.1002/path.1711460106
- Rong X, Zhu Q, Ji H, Li J, Huang H. Differentiation of pleomorphic adenoma and Warthin's tumor of the parotid gland: ultrasonographic features. *Acta Radiol* (2014) 55(10):1203–9. doi: 10.1177/0284185113515865
- Comoglu S, Ozturk E, Celik M, Avci H, Sonmez S, Basaran B, et al. Comprehensive analysis of parotid mass: A retrospective study of 369 cases. *Auris Nasus Larynx* (2018) 45(2):320–7. doi: 10.1016/j.anl.2017.04.003
- Lombardi D, Tomasoni M, Nicolai A, Paderno A, Grammatica A, Arcuri M, et al. Parotid pleomorphic and non-pleomorphic adenomas: a mono-institutional series of 512 patients. *Eur Arch Otorhinolaryngol* (2022) 279(5):2543–51. doi: 10.1007/s00405-021-07018-7
- Maahs GS, Oppermann Pde O, Maahs LG, MaChado Filho G, Ronchi AD. Parotid gland tumors: a retrospective study of 154 patients. *Braz J Otorhinolaryngol* (2015) 81(3):301–6. doi: 10.1016/j.bjorl.2015.03.007
- Seyhun N, Doğan U, Çalış ZAB, Kaya MF, Hasççek SÖ, Turgut S. The role of fine needle aspiration biopsy in deep lobe parotid tumors: Comparison of preoperative cytology and postoperative histopathologic results. *Am J Otolaryngol* (2021) 42(1):102590. doi: 10.1016/j.amjoto.2020.102590
- Xia F, Qin W, Feng J, Zhou X, Sun E, Xu J, et al. Differential diagnostic value of tumor morphology, long/short diameter ratio, and ultrasound gray-scale ratio for 3 parotid neoplasms. *Oral Surg Oral Med Oral Pathol Oral Radiol* (2022) 134(4):484–91. doi: 10.1016/j.oooo.2022.05.012
- Miao LY, Xue H, Ge HY, Wang JR, Jia JW, Cui LG. Differentiation of pleomorphic adenoma and Warthin's tumour of the salivary gland: is long-to-short diameter ratio a useful parameter? *Clin Radiol* (2015) 70(11):1212–9. doi: 10.1016/j.crad.2015.06.085
- Zhou H, Vallières M, Bai HX, Su C, Tang H, Oldridge D, et al. MRI features predict survival and molecular markers in diffuse lower-grade gliomas. *Neuro Oncol* (2017) 19(6):862–70. doi: 10.1093/neuonc/now256
- Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* (2017) 14(12):749–62. doi: 10.1038/nrclinonc.2017.141
- Mayerhoefer ME, Materka A, Langs G, Häggström I, Szczypiński P, Gibbs P, et al. Introduction to radiomics. *J Nucl Med* (2020) 61(4):488–95. doi: 10.2967/jnumed.118.222893
- Ren J, Tian J, Yuan Y, Dong D, Li X, Shi Y, et al. Magnetic resonance imaging based radiomics signature for the preoperative discrimination of stage I-II and III-IV head and neck squamous cell carcinoma. *Eur J Radiol* (2018) 106:1–6. doi: 10.1016/j.ejrad.2018.07.002
- Liao KY, Chiu CC, Chiang WC, Chiou YR, Zhang G, Yang SN, et al. Radiomics features analysis of PET images in oropharyngeal and hypopharyngeal cancer. *Med (Baltimore)* (2019) 98(18):e15446. doi: 10.1097/MD.00000000000015446
- Wang G, He L, Yuan C, Huang Y, Liu Z, Liang C. Pretreatment MR imaging radiomics signatures for response prediction to induction chemotherapy in patients with nasopharyngeal carcinoma. *Eur J Radiol* (2018) 98:100–6. doi: 10.1016/j.ejrad.2017.11.007
- Gabrys HS, Buettner F, Sterzing F, Hauswald H, Bangert M. Design and selection of machine learning methods using radiomics and dosiomics for normal tissue complication probability modeling of xerostomia. *Front Oncol* (2018) 8:35. doi: 10.3389/fonc.2018.00035
- Addison PS. *The illustrated wavelet transform handbook: introductory theory and applications in science, engineering, medicine and finance*. 2nd ed. CRC Press (2017).
- Hajek M, Dezortova M, Materka A, Lerski R eds. *Texture analysis for magnetic resonance imaging*. Prague, Czechoslovakia: Med4Publishing s.r.o (2006).
- Jiang Z, Yin J, Han P, Chen N, Kang Q, Qiu Y, et al. Wavelet transformation can enhance computed tomography texture features: a multicenter radiomics study for grade assessment of COVID-19 pulmonary lesions. *Quant Imaging Med Surg* (2022) 12(10):4758–70. doi: 10.21037/qims-22-252
- Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Stat Med* (2007) 26(30):5512–28. doi: 10.1002/sim.3148
- Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* (2010) 21(1):128–38. doi: 10.1097/EDE.0b013e3181c30fb2
- Wagner MW, Namdar K, Biswas A, Monah S, Khalvati F, Ertl-Wagner BB. Radiomics, machine learning, and artificial intelligence-what the neuroradiologist needs to know. *Neuroradiology* (2021) 63(12):1957–67. doi: 10.1007/s00234-021-02813-9
- Matsuda E, Fukuhara T, Donishi R, Kawamoto K, Hirooka Y, Takeuchi H. Usefulness of a novel ultrasonographic classification based on anechoic area patterns for differentiating warthin tumors from pleomorphic adenomas of the parotid gland. *Yonago Acta Med* (2018) 60(4):220–6. doi: 10.24563/yam.2017.12.002
- Jiang LP. Characteristics of pleomorphic adenomas, adenolymphomas, and Malignant tumors of the salivary glands on color doppler ultrasonography and contrast enhanced ultrasonography. *Eur Rev Med Pharmacol Sci* (2020) 24(22):11509–17. doi: 10.26355/eurrev\_202011\_23792
- Haralick RM, Shanmugam K. Textural features for image classification. *IEEE Trans Syst Man Cybern* (1973) SMC-3:610–21. doi: 10.1109/TSMC.1973.4309314
- Zheng M, Chen Q, Ge Y, Yang L, Tian Y, Liu C, et al. Development and validation of CT-based radiomics nomogram for the classification of benign parotid gland tumors. *Med Phys* (2023) 50(2):947–57. doi: 10.1002/mp.16042
- Song LL, Chen SJ, Chen W, Shi Z, Wang XD, Song LN, et al. Radiomic model for differentiating parotid pleomorphic adenoma from parotid adenolymphoma based on MRI images. *BMC Med Imaging* (2021) 21(1):54. doi: 10.1186/s12880-021-00581-9
- Fruehwald-Pallamar J, Czerny C, Holzer-Fruehwald L, Nemec SF, Mueller-Mang C, Weber M, et al. Texture-based and diffusion-weighted discrimination of parotid gland lesions on MR images at 3.0 Tesla. *NMR BioMed* (2013) 26(11):1372–9. doi: 10.1002/nbm.2962
- Piludu F, Marzi S, Ravanelli M, Pellini R, Covello R, Terrenato I, et al. MRI-based radiomics to differentiate between benign and Malignant parotid tumors with external validation. *Front Oncol* (2021) 11:656918. doi: 10.3389/fonc.2021.656918
- Raghav K, Hwang H, Jácome AA, Bhang E, Willett A, Huey RW, et al. Development and validation of a novel nomogram for individualized prediction of survival in cancer of unknown primary. *Clin Cancer Res* (2021) 27(12):3414–21. doi: 10.1158/1078-0432.CCR-20-4117
- Liu S, Yu X, Yang S, Hu P, Hu Y, Chen X, et al. Machine learning-based radiomics nomogram for detecting extramural venous invasion in rectal cancer. *Front Oncol* (2021) 11:610338. doi: 10.3389/fonc.2021.610338



## OPEN ACCESS

## EDITED BY

Sharon R. Pine,  
University of Colorado Anschutz Medical  
Campus, United States

## REVIEWED BY

Sung Bin Park,  
Chung-Ang University Hospital, Republic of  
Korea  
Jaya Lakshmi Thangaraj,  
University of California, San Diego,  
United States

## \*CORRESPONDENCE

Xuefeng Zhang  
✉ zhangxuefeng626@sina.com  
Zhiyu Zhang  
✉ abner\_666@126.com

<sup>†</sup>These authors have contributed  
equally to this work and share  
first authorship

RECEIVED 27 August 2023

ACCEPTED 26 December 2023

PUBLISHED 16 January 2024

## CITATION

Hu C, Qiao X, Xu Z, Zhang Z and Zhang X  
(2024) Machine learning-based CT  
texture analysis in the differentiation  
of testicular masses.  
*Front. Oncol.* 13:1284040.  
doi: 10.3389/fonc.2023.1284040

## COPYRIGHT

© 2024 Hu, Qiao, Xu, Zhang and Zhang. This is  
an open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

# Machine learning-based CT texture analysis in the differentiation of testicular masses

Can Hu<sup>1,2†</sup>, Xiaomeng Qiao<sup>3†</sup>, Zhenyu Xu<sup>4†</sup>,  
Zhiyu Zhang<sup>1\*</sup> and Xuefeng Zhang<sup>1\*</sup>

<sup>1</sup>Department of Urology, The First Affiliated Hospital of Soochow University, Suzhou, Jiangsu, China, <sup>2</sup>Department of Urology, Suzhou Xiangcheng People's Hospital, Suzhou, China, <sup>3</sup>Department of Radiology, The First Affiliated Hospital of Soochow University, Suzhou, Jiangsu, China, <sup>4</sup>Department of Urology, The Affiliated Hospital of Nanjing University of Traditional Chinese Medicine: Traditional Chinese Medicine Hospital of Kunshan, Kunshan, China

**Purpose:** To evaluate the ability of texture features for distinguishing between benign and malignant testicular masses, and furthermore, for identifying primary testicular lymphoma in malignant tumors and identifying seminoma in testicular germ cell tumors, respectively.

**Methods:** We retrospectively collected 77 patients with an abdominal and pelvic enhanced computed tomography (CT) examination and a histopathologically confirmed testicular mass from a single center. The ROI of each mass was split into two parts by the largest cross-sectional slice and deemed to be two samples. After all processing steps, three-dimensional texture features were extracted from unenhanced and contrast-enhanced CT images. Excellent reproducibility of texture features was defined as intra-class correlation coefficient  $\geq 0.8$  (ICC  $\geq 0.8$ ). All the groups were balanced via the synthetic minority over-sampling technique (SMOTE) method. Dimension reduction was based on pearson correlation coefficient (PCC). Before model building, minimum-redundancy maximum-relevance (mRMR) selection and recursive feature elimination (RFE) were used for further feature selection. At last, three ML classifiers with the highest cross validation with 5-fold were selected: autoencoder (AE), support vector machine(SVM), linear discriminant analysis (LAD). Logistics regression (LR) and LR-LASSO were also constructed to compare with the ML classifiers.

**Results:** 985 texture features with ICC  $\geq 0.8$  were extracted for further feature selection process. With the highest AUC of 0.946 ( $P < 0.01$ ), logistics regression was proved to be the best model for the identification of benign or malignant testicular masses. Besides, LR also had the best performance in identifying primary testicular lymphoma in malignant testicular tumors and in identifying seminoma in testicular germ cell tumors, with the AUC of 0.982 ( $P < 0.01$ ) and 0.928 ( $P < 0.01$ ), respectively.

**Conclusion:** Until now, this is the first study that applied CT texture analysis (CTTA) to assess the heterogeneity of testicular tumors. LR model based on CTTA might be a promising non-invasive tool for the diagnosis and differentiation of testicular masses. The accurate diagnosis of testicular masses would assist urologists in correct preoperative and perioperative decision making.

#### KEYWORDS

contrast enhanced computerized tomography, CT texture analysis, testicular masses, machine learning, urology and radiology

## Introduction

Testicular tumor is one of the most common malignancy in men aged 14–44 years worldwide, accounting for approximately 1% of all male tumors and 5% of genitourinary neoplasms. In recent years, the morbidity and mortality of testicular cancer has risen continuously, especially in Western countries (1–3). Testicular tumor is a heterogeneous group of diseases with various pathological subtypes and clinical behavior. Among them, 90%–95% are testicular germ cell tumors (TGCTs), including seminoma, embryoma, teratoma and choriocarcinoma, of which about 55% are seminoma of the testis. The other part of testicular tumor subtypes includes hematological neoplasm, sex cord stromal tumors, and other exceedingly rare types of tumors. As the different pathophysiology and molecular mechanisms, diverse biological behaviors were observed in these testicular masses, which leads to different management and clinical decision (4, 5). Of course, different treatment strategies are applied in benign or malignant testicular tumors and primary testicular lymphoma (6). Furthermore, as to these local or systemic progressed TGCTs, the main treatment is radiotherapy or chemotherapy instead of surgery (radical orchiectomy) (7). Under this circumstance, we cannot reach exact pathological results from the surgical specimens. Thus, a pre-operative diagnostic tool that allows histological subtype classification of testicular masses will be of great importance to precise treatment and clinical prognosis judgement. Although ultrasound examination is the preferred examination for testicular masses, the widespread use of ultrasound has led to more and more impalpable or ambiguous results (8). As mentioned by the EAU Guidelines 2022 (<http://uroweb.org/guidelines/compilations-of-all-guidelines/>) (7), although magnetic resonance imaging (MRI) provides higher sensitivity and specificity than ultrasound in the diagnosis of testicular tumor, MRI is not superior to contrast enhanced computerized tomography (CECT) in detecting retroperitoneal lymph node metastasis in general and is more expensive, which does not justify its routine use in the diagnosis of testicular tumor (8, 9). Besides, it should only be considered when ultrasound is inconclusive, as local staging for testis-sparing surgery. However,

CECT is recommended in all patients for staging before orchidectomy (7, 10). Therefore, CT has become an indispensable imaging method for patients with testicular masses. In addition, testicular biopsy is used in few centers and has not gained widespread acceptance because of narrow indication and possible increased local recurrence rate, with which it is difficult to assess intratumoral heterogeneity for its limitation (7, 11). In recent years, CT texture analysis (CTTA) has become a promising technique for evaluating tumor heterogeneity in a quantitative manner. CTTA could provide a measure of heterogeneity of testicular masses with various mathematical methods that can be used to evaluate the gray-level intensity and position of the pixels within contrast-enhanced CT images (12).

Up to now, no study has paid attention on the utility of CTTA in histological subtyping of testicular masses. This is the first study that explores the value of texture features in testicular masses.

## Materials and methods

### Patients

This study was approved by the Institutional Review Board in the First Affiliated Hospital of Soochow University with a waiver of informed consent. We retrospectively collected the imaging data and clinical data of consecutive 94 patients diagnosed with testicular masses from January 2015 to April 2022. Inclusion criteria were as follows: (a) patients with available three-phase CT scan prior to any treatment and operation; (b) pathologically proven testicular masses after surgery treatment; (c) the interval between CT and surgery was less than three months and no treatment received. Exclusion criteria included: (a) lack of pretreatment contrast-enhanced CT; (b) the absence of a certain phase of CT; (c) poor image quality. After conducting the criteria, 77 men were identified to constitute our study cohort and divided into a benign group (n=21) and a malignant group (n=56) according to their histological results. And then, in the malignant group, we divided them into primary testicular lymphomas group (n=10) and non-lymphomas group (n=46). Finally, we screened out all the testicular

germ cell tumors from malignancy ( $n=43$ ) and divided them into seminoma group ( $n=30$ ) and non-seminoma group ( $n=13$ ) for the differentiation.

## Study design

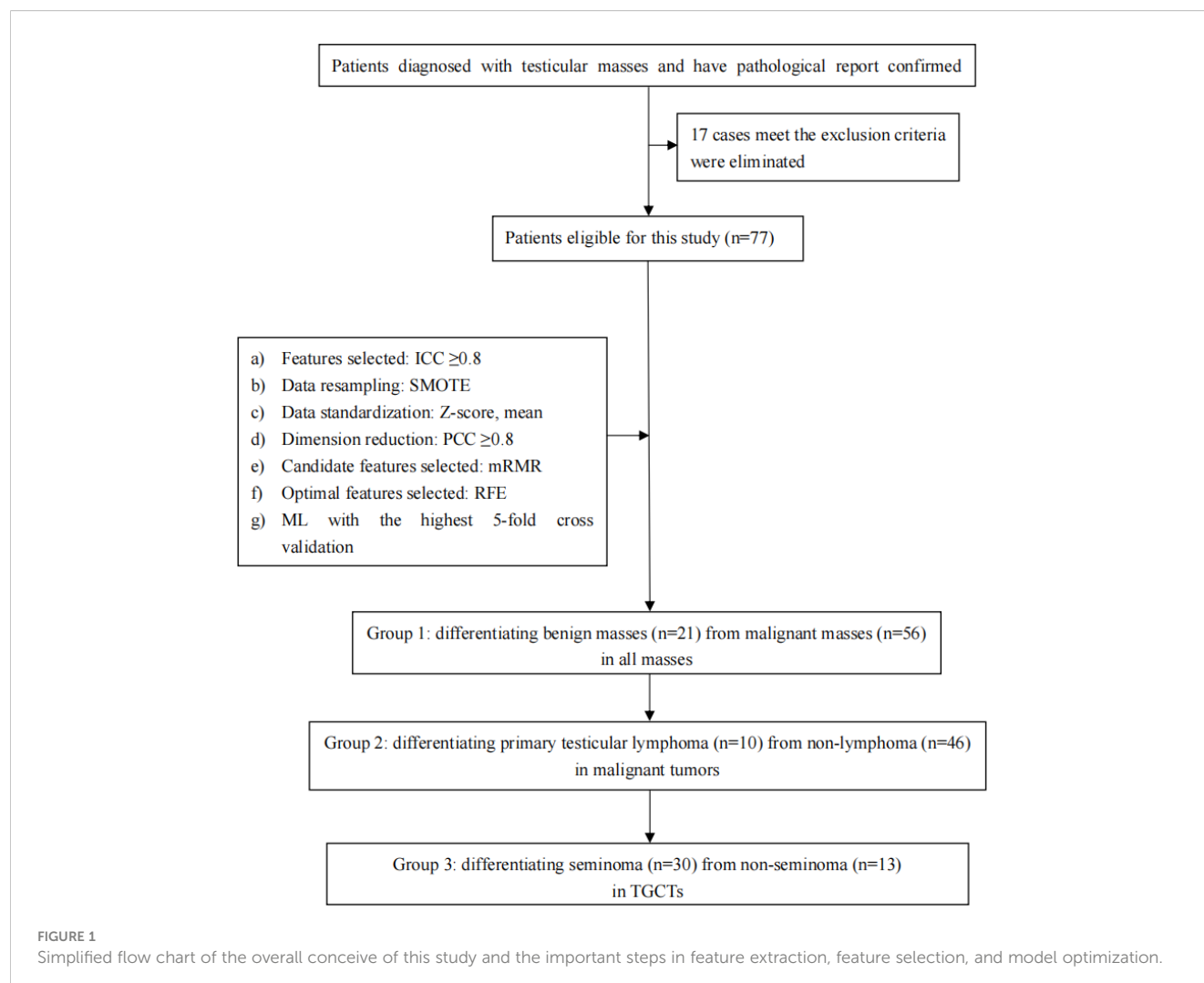
To make this article clear, a flow chart including specific technical steps was provided to the readers (Figure 1).

## Image acquisition

All patients underwent contrast-enhanced CT (GE Healthcare and Siemens Healthcare), including three phases: unenhanced phase (UP), arterial phase (AP, 9s delay after contrast injection) and portal venous phase (PP, 30s delay after contrast injection). Similar protocols were applied when scanning: tube voltage of 120 kVp, tube current of 180–450 mA, matrix of 512, field of view of 380–500 mm, and 5 mm reconstructed section thickness. Contrast medium (iopromide) was injected intravenously at a rate of 3.0 mL/s.

## ROIs delineation and data augmentation

For the mass without a distinct border or with invasion of the whole testicle, the region of interest (ROI) was defined as the whole testicular tissue on the diseased side. Meanwhile, for the mass with a distinct border, ROI was presumed to be the whole mass (Figure 2). One radiologist (with 5 years of experience) and one urologist (with 3 years of experience) blinded to the histopathology results first identified the border of each mass in consensus and then manually delineated the ROIs around the margin of the testicular masses with the ITK-SNAP (v 3.6.0) software (Can Hu and Xiaomeng Qiao). The ROIs were carefully drawn with an approximate distance of 1–3 mm from the margin of tumors to prevent the effect of fat and air (13). Due to the low morbidity of testicular tumors, sample size was inevitably limited in our study. Hence, as a scheme of data augmentation, the ROI of each patient was split into the upper and lower part by the largest slice and counted as two samples (for bilateral tumors, we counted one patient as four samples) (14). The histopathology results of augmented samples were in line with the original patients. After 2 weeks, the same task was repeated by the radiologist for the evaluation of intra-observer variation.





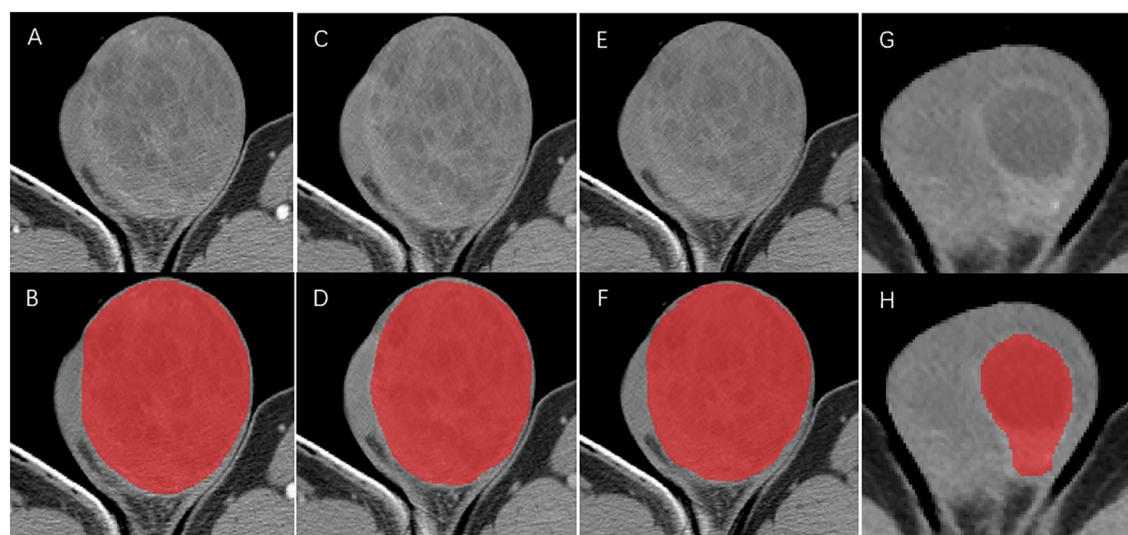


FIGURE 2

ROIs delineation in arterial phase (A, B), portal venous phase (C, D) and unenhanced phase (E, F) CT for the mass without a distinct border. ROIs delineation in arterial phase (G, H) for mass with a distinct border.

## CT texture feature evaluation and selection

Texture features were extracted from three-phases CT images. All CT images were anonymous before they were uploaded to the commercial texture analysis software (TexRAD, version 3.9, Feedback Medical Ltd) stored in DICOM format. A total of 572 features were extracted from each of the CT phases, including 18 first order features, 14 shape-based features, 24 features of grey level co-occurrence matrix (GLCM), 14 of grey level dependence matrix (GLDM), 16 of grey level run length matrix (GLRLM), 16 of grey level size zone matrix (GLSZM) and 5 of neighborhood grey tone difference matrix (NGTDM). The first order features and second order features were extracted from the original images and derived images via filtering based on the Laplacian of Gaussian. The spatial scale factor (SSF) at 6 levels (0 mm: no filtration; 2 mm: fine texture scale; 3 mm, 4 mm and 5 mm: medium texture scales; and 6 mm: coarse texture scale) were used. These features have been used in previous quantitative analysis studies and mathematical formula been described in the website in detail (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7581467/>) (15–18).

Inter- and intra-observer intra-class correlation coefficient (ICC) was firstly utilized to assess reproducibility and repeatability for each texture feature. We retained features with ICCs greater than 0.8. A total of 985 texture features with ICCs  $\geq 0.8$  were included in the further feature selection process. In order to avoid the classifiers overtrained owing to highly-correlated features, feature selection dimension reduction was conducted to identify candidate and optimal features for model building (19). A synthetic minority oversampling technique (SMOTE) was adopted to deal with the adverse impact of the imbalanced data in this study. In addition, we also standardized the data by the method of Z-score and mean to compare the AUC of the model established by these two standardization methods for better model selecting. Dimension reduction was based on pearson

correlation coefficient (PCC). Features demonstrating a strong correlation (PCC  $\geq 0.8$ ) were removed one by one to achieve better performance. Moreover, after the application of minimum-redundancy maximum-relevance (mRMR), each of the three groups for intra-group comparisons were reduced to 20 features. Before build the model, we also used recursive feature elimination (RFE) to further select optimal features with excellent discrimination ability from the above 20 texture features (20). Finally, with the highest 5-fold cross validation, five models were built by machine learning (ML) algorithms including auto encoder (AE), support vector machine (SVM), linear discriminant analysis (LAD), logistics regression (LR) and logistics regression-least absolute shrinkage and selection operator (LR-LASSO).

## Statistical analysis

Statistical analysis was performed using IBM SPSS v.23.0, Python software v2.7.13 (<https://www.python.org>) and R software v.4.1.1. Non-normal distribution continuous variables were expressed as medians (interquartile range). The group differences were assessed using a Mann–Whitney U test. Receiver operating characteristic (ROC) curve analysis, accuracy, sensitivity, specificity, PPV and NPV were calculated to comprehensively assess the models. Significance between the AUC of models were compared using the Delong test. A two-sided p value  $< 0.05$  indicated statistical significance.

## Results

### Demographics

Specific pathological subtypes of all these testicular masses were provided in Table 1. Patient characteristics between the three

TABLE 1 Specific pathological subtypes of testicular masses.

|                 | Pathological subtypes       | Specific subtypes             | n  |
|-----------------|-----------------------------|-------------------------------|----|
| Malignant, n=56 | TGCTs                       |                               | 43 |
|                 |                             | seminoma                      | 30 |
|                 |                             | mixed TGCTs                   | 9  |
|                 |                             | embryonal carcinomas          | 2  |
|                 |                             | yolk sac tumors               | 2  |
|                 | Sex cord-stromal tumor      |                               | 2  |
|                 |                             | granulosa cell tumor          | 1  |
|                 |                             | Sertoli-Leydig cell tumor     | 1  |
|                 | Primary testicular lymphoma |                               | 10 |
|                 |                             | diffuse large B-cell lymphoma | 8  |
|                 |                             | NK/T-cell lymphoma            | 2  |
|                 |                             | embryonal rhabdomyosarcoma    | 1  |
| Benign, n=21    |                             | inflammation or abscess       | 5  |
|                 |                             | adenomatoid tumor             | 3  |
|                 |                             | angiomas or leiomyoma         | 8  |
|                 |                             | others                        | 5  |

TGCTs, testicular germ cell tumors.

groups were summarized in Table 2. Among them, 36 patients with lesions on the left side while 41 patients on the right side. Only one patient with granulosa cell tumor was bilateral. Thus, a total of 77 patients with 156 masses (76\*2 + 1\*4) were enrolled in the study according to our special method of data augmentation. For group 1, 21 benign cases and 56 malignant cases were counted. Statistical

significance could be observed in age and all the serum tumor markers. For group 2, there were 10 primary testicular lymphomas (8 diffuse large B-cell lymphomas and 2 NK/T-cell lymphomas) and 46 non-lymphomas. The mean age of the lymphomas subgroup was statistically significantly higher than the non-lymphoma subgroup (33 (29, 39) vs 68 (58, 76),  $P < 0.001$ ). In the three serum tumor markers, LDH between the two subgroups had no significant difference. For group 3, there were 30 seminomas and 13 non-seminomas (9 mixed TGCTs, 2 embryonal carcinomas and 2 yolk sac tumors). Statistical significance could be observed in age, HCG and AFP. The average time interval between CT and serum tumor markers was 5 days.

## Reproducibility and Feature selection

572 features were extracted from each of the CT phases. A total of 985 texture features with an ICC  $\geq 0.8$  were included in the further feature selection process. After mRMR, each of the three groups for intra-group comparisons were reduced to 20 features. Before model building, RFE was applied in all models to further select optimal features with excellent discrimination ability from the above 20 texture features (range from 9 to 15) (Supplementary 1).

## ML-based classifications

The predictive performance and ROC curves of all ML and the two LR-based models using two data standardization methods for the three groups were summarized in Tables 3A–C, respectively. As a whole, z-score had a better performance than mean in the three groups. For group 1 (Table 3A), the LR and LR-LASSO were the two best-performing classifiers that achieved similar AUC values (AUC = 0.946,  $P = 1.000$ ). However, considering the AUC of LR was slightly higher than LR-LASSO by the method of z-score, LR was selected for the best model. The overall accuracy, sensitivity, specificity, PPV, NPV and AUC of the best model were 87.3%, 86.1%, 90.5%, 95.6%, 73.1% and 0.946 (95% CI 0.896–0.995), respectively. For group 2 (Table 3B), although SVM and LR-LASSO had high AUC of 0.986 and 0.985, respectively, LR was chosen as the most appropriate model, achieved an accuracy of 90.4% (sensitivity 100%, specificity 88.3%, PPV 64.5% and NPV 100%) with an AUC of 0.982 (95% CI 0.963–1.000). For group 3

TABLE 2 Patients' demographics between the three groups.

| Variables, (M, IQR) | Benign vs Malignant              | P      | non-lymphomas vs lymphomas       | P      | non-seminoma vs seminoma         | P      |
|---------------------|----------------------------------|--------|----------------------------------|--------|----------------------------------|--------|
| Age                 | 47(35, 68) vs 35(29, 48)         | 0.041  | 33(29, 39) vs 68(58, 76)         | <0.001 | 29(37, 34) vs 36(31, 43)         | 0.007  |
| HCG                 | 0(0, 0.2) vs 1.8(0.3, 38.8)      | <0.001 | 3.8(1.0, 92.3) vs 0.18(0, 0.5)   | <0.001 | 49.8(2.8, 246) vs 3.6(0.9, 38.8) | 0.023  |
| AFP                 | 2.5(1.8, 3.1) vs 3.3 (2.1, 18.3) | 0.011  | 3.4(2.5, 112.2) vs 2.1(1.1, 5.0) | 0.02   | 273(44, 381) vs 3.1(2.2, 4.2)    | <0.001 |
| LDH                 | 173(151, 184) vs 219 (172, 278)  | <0.001 | 224(176, 290) vs 193(160, 270)   | 0.404  | 288(196, 327) vs 218(180, 251)   | 0.054  |

HCG, human chorionic gonadotropin; AFP, alpha fetoprotein; LDH, lactic dehydrogenase; IQR, interquartile range.

TABLE 3A Performance of ML classifiers, LR and LR-LASSO in differentiating benign masses from malignant masses with the method of Z-score and mean.

| Model    | Standardization | ACC   | SEN   | SPE   | PPV   | NPV   | Youden | AUC (95 % CI)       |
|----------|-----------------|-------|-------|-------|-------|-------|--------|---------------------|
| AE       | mean            | 77.5% | 74.0% | 85.7% | 92.5% | 58.1% | 0.59   | 0.833 (0.711-0.954) |
|          | z-score         | 87.3% | 92.3% | 76.2% | 90.2% | 80.0% | 0.68   | 0.866 (0.758-0.972) |
| SVM      | mean            | 86.0% | 84.1% | 90.1% | 95.4% | 70.4% | 0.74   | 0.922 (0.852-0.990) |
|          | z-score         | 91.6% | 94.2% | 85.7% | 94.0% | 85.7% | 0.80   | 0.900 (0.804-0.995) |
| LDA      | mean            | 88.7% | 87.9% | 90.5% | 95.7% | 76.0% | 0.78   | 0.910 (0.825-0.996) |
|          | z-score         | 88.7% | 88.3% | 90.5% | 95.7% | 76.0% | 0.79   | 0.910 (0.825-0.996) |
| LR       | mean            | 87.3% | 85.8% | 90.5% | 95.6% | 73.1% | 0.76   | 0.944 (0.892-0.995) |
|          | z-score         | 87.3% | 86.1% | 90.5% | 95.6% | 73.1% | 0.77   | 0.946 (0.896-0.995) |
| LR-LASSO | mean            | 88.7% | 89.9% | 85.7% | 93.8% | 78.2% | 0.76   | 0.912 (0.836-0.988) |
|          | z-score         | 88.7% | 88.4% | 90.5% | 95.7% | 75.0% | 0.78   | 0.946 (0.894-0.996) |

ML, machine learning; LR, logistics regression; LR-LASSO, logistics regression-least absolute shrinkage and selection operator; AE, autoencoder; SVM, support vector machine; LDA, linear discriminant analysis; ACC, accuracy; SEN, sensitivity; SPE, specificity; PPV, positive predictive value; NPV, negative predictive value; AUC, area under the curve; CI, confidence interval.

(Table 3C), LR also outperformed other models, achieving an accuracy of 90.7% (sensitivity 90.0%, specificity 92.3%, PPV 96.4% and NPV 80.0%) with a high AUC of 0.928 (95% CI 0.858-0.996). Overall, LR was the best choice for the histological classification of testicular masses. The ROC curves of LR among the three groups were demonstrated in Figure 3.

Discussion

This is the first study that applied CT texture analysis (CTTA) to assess the pathological subtypes of testicular tumors. All the patients were divided into three groups to evaluate the ability of texture features for identifying benign and malignant testicular masses, identifying primary testicular lymphoma in malignant tumors and identifying seminoma in testicular germ cell tumors, respectively. For all three groups, the most appropriate model was LR rather than ML-based classifiers by the data standardization of z-score.

Testicular tumor is a heterogeneous group of diseases with various pathological subtypes and clinical behavior, which leads to different response to treatment (21). Firstly, the treatment of benign and malignant mass is different. Radical orchiectomy was the standard operation of TGCTs while symptomatic treatment is often used in benign masses. As to clinical stage I non-seminomas without vascular and lymphatic infiltration, retroperitoneal lymph node dissection (RPLND) is the standard treatment for patients without follow-up conditions. CTTA facilitates clinical evaluation and psychological development of patients, and to some extent RPLND could even be performed immediately after orchiectomy, avoiding the need for a second operation. As to clinical stage II TGCTs, seminomas tend to have sensitive response to radiotherapy while non-seminomas tend to benefit more from RPLND or neo-adjuvant chemotherapy. As to metastatic testicular tumors, urologists could only apply different chemotherapy regimens according to the prognosis (7, 22). Under the circumstance, exact pathological results cannot be reached from

the surgical specimens. Conventionally, ultrasound examination is the preferred choice for testicular masses. Despite its high sensitivity in the mass detection, it shows low specificity in distinguishing between benign and malignant masses, let alone other pathological subtypes (23, 24). Furthermore, testicular biopsy is used in some centers but has not gained widespread acceptance because of narrow indications and concerns for tumor seeding along the biopsy tract. Germ cell neoplasia *in situ* (GCNIS) could be diagnosed by testicular biopsy using immunohistochemistry with high sensitivity and specificity. However, a certain amount of false-negative biopsy was brought inevitably (25). Thus, non-invasive test for the evaluation of testicular masses may open the possibility of allowing histological subtype classification.

CT is recommended for the pre-surgical assessment of testicular masses, and at the same time, could evaluate retroperitoneal lymph node metastases. However, the heterogeneity of tumors is not particularly obvious on imaging and the diagnostic accuracy depends on the experience of radiologists. In the present study, we found that quantitative CTTA potentially allowed for detection of subtle differences and was able to differentiate various histological subtype classifications beyond visual assessment. To date, as far as we know, there have been no CTTA related studies on testicular tumors. Previous research has focused on tumors such as epithelial ovarian carcinoma, renal cell carcinoma or lung carcinoma (26–29). In the study of An et al. (26), they demonstrated that CTTA was instrumental in the identification of high-grade serous carcinoma (HGSC) or non-HGSC in 205 patients. Erdim et al. (28) investigated that renal masses with unclear pathological diagnosis could be distinguished through ML-based CTTA in 79 patients. Furthermore, Ceyda et al. (27) has confirmed the ability of different ML-based classifiers in the prediction of Fuhrman nuclear grade of clear cell renal cell carcinomas in 53 patients. Yang et al. (29) evaluated the value of 2D and 3D CTTA in predicting lymphatic vascular invasion in lung adenocarcinoma.

Our study is not only focused on the differentiation of benign and malignant lesions but also on identifying primary testicular

**TABLE 3B** Performance of ML classifiers, LR and LR-LASSO in differentiating primary testicular lymphoma from non-lymphoma in malignant tumors with the method of Z-score and mean.

| Model    | Standardization | ACC   | SEN   | SPE   | PPV   | NPV   | Youden | AUC (95 % CI)       |
|----------|-----------------|-------|-------|-------|-------|-------|--------|---------------------|
| AE       | mean            | 93.7% | 90.0% | 94.7% | 78.3% | 97.8% | 0.85   | 0.965 (0.930-1.000) |
|          | z-score         | 87.6% | 89.8% | 86.2% | 58.1% | 97.6% | 0.76   | 0.921 (0.847-0.995) |
| SVM      | mean            | 93.7% | 90.0% | 94.7% | 78.3% | 97.8% | 0.85   | 0.973 (0.947-0.998) |
|          | z-score         | 93.0% | 100%  | 91.5% | 71.4% | 100%  | 0.91   | 0.986 (0.970-1.000) |
| LDA      | mean            | 87.7% | 100%  | 85.1% | 58.8% | 100%  | 0.85   | 0.979 (0.957-1.000) |
|          | z-score         | 87.7% | 100%  | 85.1% | 58.8% | 100%  | 0.85   | 0.979 (0.957-1.000) |
| LR       | mean            | 90.4% | 100%  | 88.3% | 64.5% | 100%  | 0.88   | 0.982 (0.963-1.000) |
|          | z-score         | 90.4% | 100%  | 88.3% | 64.5% | 100%  | 0.88   | 0.982 (0.963-1.000) |
| LR-LASSO | mean            | 96.5% | 85.0% | 98.9% | 94.4% | 96.7% | 0.84   | 0.978 (0.959-1.000) |
|          | z-score         | 91.2% | 100%  | 89.4% | 66.7% | 100%  | 0.89   | 0.985 (0.967-1.000) |

ML, machine learning; LR, logistics regression; LR-LASSO, logistics regression-least absolute shrinkage and selection operator; AE, autoencoder; SVM, support vector machine; LDA, linear discriminant analysis; ACC, accuracy; SEN, sensitivity; SPE, specificity; PPV, positive predictive value; NPV, negative predictive value; AUC, area under the curve; CI, confidence interval.

lymphoma in malignant tumors and identifying seminoma in TGCTs. The differential diagnosis of TGCTs or non-TGCTs was not included in our study for the reason that most testicular tumors were germ cell neoplasms (accounting for 95%), and the remaining few were of no great discriminative value and had a low incidence. To avoid confounding bias, we also did not identify lymphoma and seminoma across all tumor types. We think the above process may be more appropriate and in line with the clinical practice. The performance of most classifiers in all three groups are satisfactory. Despite the ACC of SVM is slightly higher than LR in group 1, we chose LR as the best classifier for the better stability of the model (the AUC of LR was higher than SVM) (30). For group 2, SVM and LR-LASSO seem to outperform LR ( $P > 0.05$ ). Nevertheless, compared to LR, the AUC of the two classifiers had a relatively large reduction when using the data standardization of mean. For group 3, LR was obviously superior than other models ( $P < 0.05$ ). Therefore, in view of the fact that the diagnostic performance of

each model was not significantly different, we still tend to choose LR as the last model for uniformity. In general, CTTA could be potentially valuable in guiding treatment and provide a reliable reference for clinicians.

The result of optimal features indicated that the entropy of the gray-level cooccurrence matrix (GLCM) for AP, energy of the first-order texture feature for PP and 90<sup>th</sup> percentile of the first-order texture feature for UP were features with the largest coefficient for the three groups, respectively. For group 1, malignant testicular tumors were characterized by a greater entropy for AP ( $P = 0.028$ ). Entropy represents the randomness or complexity of the texture in the image and a greater entropy tends to reflect heterogeneity, which exactly demonstrated the invasive growth pattern with poorly defined boundaries in malignant tumors (31–33). In addition, malignant testicular tumors appear to be more irregular on cells for the different degree of the disturbed formation of the germ cells (22, 34). Energy is the sum of the squares of voxel values

**TABLE 3C** Performance of ML classifiers, LR and LR-LASSO in differentiating seminoma from non-seminoma in TGCTs with the method of Z-score and mean.

| Model    | Standardization | ACC   | SEN   | SPE   | PPV   | NPV   | Youden | AUC (95 % CI)       |
|----------|-----------------|-------|-------|-------|-------|-------|--------|---------------------|
| AE       | mean            | 79.1% | 85.0% | 65.4% | 85.0% | 65.4% | 0.50   | 0.765 (0.639-0.891) |
|          | z-score         | 72.1% | 75.0% | 65.4% | 83.3% | 53.1% | 0.40   | 0.673 (0.528-0.819) |
| SVM      | mean            | 84.9% | 83.3% | 88.5% | 94.3% | 69.7% | 0.72   | 0.890 (0.805-0.974) |
|          | z-score         | 87.2% | 83.3% | 96.1% | 98.0% | 71.4% | 0.79   | 0.919 (0.847-0.989) |
| LDA      | mean            | 88.4% | 91.7% | 80.8% | 91.7% | 80.8% | 0.72   | 0.912 (0.839-0.984) |
|          | z-score         | 88.4% | 91.7% | 80.8% | 91.7% | 80.8% | 0.72   | 0.912 (0.839-0.984) |
| LR       | mean            | 90.7% | 90.0% | 92.3% | 96.4% | 80.0% | 0.82   | 0.925 (0.858-0.996) |
|          | z-score         | 90.7% | 90.0% | 92.3% | 96.4% | 80.0% | 0.82   | 0.928 (0.855-0.994) |
| LR-LASSO | mean            | 86.1% | 83.3% | 92.3% | 96.1% | 70.6% | 0.76   | 0.894 (0.808-0.978) |
|          | z-score         | 86.2% | 81.7% | 96.2% | 98.0% | 69.4% | 0.78   | 0.919 (0.848-0.988) |

ML, machine learning; LR, logistics regression; LR-LASSO, logistics regression-least absolute shrinkage and selection operator; AE, autoencoder; SVM, support vector machine; LDA, linear discriminant analysis; ACC, accuracy; SEN, sensitivity; SPE, specificity; PPV, positive predictive value; NPV, negative predictive value; AUC, area under the curve; CI, confidence interval.

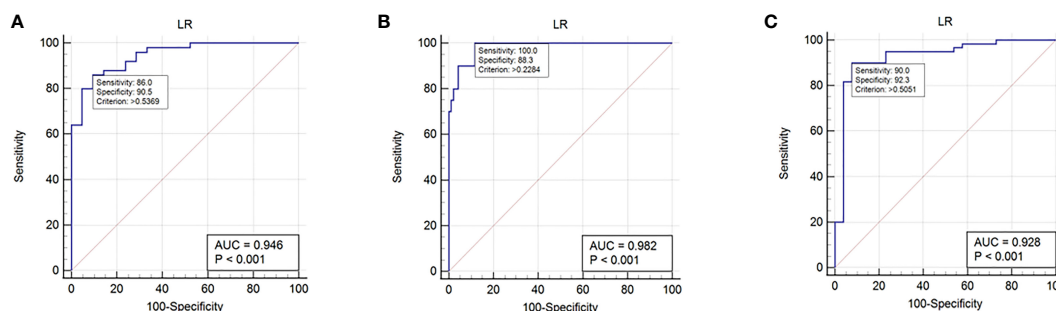


FIGURE 3

ROC curves in histological subtype classification of testicular masses using the selected logistics regression model. (A): ROC curve for group 1. (B): ROC curve for group 2. (C): ROC curve for group 3.

and reflects the uniformity of image gray distribution and texture thickness (35, 36). Primary testicular lymphoma displayed a lower energy ( $P < 0.001$ ) and it may be associated with a worse overall survival and more aggressive tumors (36, 37). We also found that higher 90<sup>th</sup> percentile was correlated with seminomas ( $P = 0.020$ ), demonstrating a phenomenon of hyper-attenuation in UP (38). Possible explanation for this is that seminomas typically have homogenous internal attenuation while non-seminomas show inhomogeneous soft-tissue density (39). Moreover, as the representation of low attenuation, hemorrhage and necrosis of seminomas may present but are usually limited (40, 41).

There are several limitations in our study. First, owing to the low morbidity, the sample size of the study is small inevitably. We had to apply the method of data augmentation to expand the sample size, which may aggravate selection bias. Secondly, no comparison was made with MRI and ultrasound in terms of diagnostic efficacy because not all patients had complete imageological examinations. Besides, as a comparative analysis with CTTA with other experimental methods like flow cytometry, H&E, IHC that would help to accurately diagnose the tumors based on CTTA. We look forward to further research on MRI and detecting techniques in the identification of testicular tumors. Thirdly, the potential impact of this methodical difference on clinical findings is largely unexplored. the reproducibility of texture analysis has yet to be established widely. Some issues like image acquisition and image quality, and their effect on texture analysis need to be regulated and resolved. Fourthly, our study was retrospective and lack of external validation. Although 5-fold cross validation was used, the risk of overfitting could not be avoided. Fifthly, a three- dimensional CTTA may be time-consuming, but this exactly the advantage of our study. Lastly, we chose only a few representative ML classifiers. Lastly, different devices and software may have different consequences. Thus, large-scale and well-designed studies are warranted to validate the performance of the models.

## Conclusion

In conclusion, LR model based on CTTA might be a promising non-invasive tool for the diagnosis and differentiation of testicular masses. The accurate diagnosis of testicular masses would

assist urologists in correct preoperative and perioperative decision making.

## 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/s.

## Ethics statement

This study was approved by the Institutional Review Board in the First Affiliated Hospital of Soochow University. 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

CH: Data curation, Formal analysis, Investigation, Resources, Writing – original draft, Writing – review & editing. XQ: Data curation, Formal analysis, Methodology, Software, Writing – original draft. ZX: Data curation, Formal analysis, Investigation, Writing – original draft. ZZ: Project administration, Supervision, Validation, Visualization, Writing – review & editing, Resources. XZ: Conceptualization, Funding acquisition, Investigation, Project administration, Supervision, Validation, Visualization, 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.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2023.1284040/full#supplementary-material>

## References

- 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(3):209–49. doi: 10.3322/caac.21660
- Huang J, Chan SC, Tin MS, Liu X, Lok VT, Ngai CH, et al. Worldwide distribution, risk factors, and temporal trends of testicular cancer incidence and mortality: A global analysis. *Eur Urol Oncol* (2022) 5(5):566–76. doi: 10.1016/j.euo.2022.06.009
- Chovanec M, Cheng L. Advances in diagnosis and treatment of testicular cancer. *BMJ* (2022) 379:e070499. doi: 10.1136/bmj-2022-070499
- Cheng L, Lyu B, Roth LM. Perspectives on testicular germ cell neoplasms. *Hum Pathol* (2017) 59:10–25. doi: 10.1016/j.humpath.2016.08.002
- Al-Obaidy KI, Idrees MT. Testicular Tumors: A contemporary update on morphologic, immunohistochemical and molecular features. *Adv Anat Pathol* (2021) 28(4):258–75. doi: 10.1097/PAP.0000000000000302
- Cheah CY, Wirth A, Seymour JF. Primary testicular lymphoma. *Blood* (2014) 123(4):486–93. doi: 10.1182/blood-2013-10-530659
- EAU Guidelines Office A. *The Netherlands. EAU guidelines* (2022). Available at: <http://uroweb.org/guidelines/compilations-of-all-guidelines/>.
- Ager M, Donegan S, Boeri L, de Castro JM, Donaldson JF, Omar MI, et al. Radiological features characterising indeterminate testes masses: a systematic review and meta-analysis. *BJU Int* (2023) 131(3):288–300. doi: 10.1111/bju.15869
- Mittal PK, Abdalla AS, Chatterjee A, Baumgarten DA, Harri PA, Patel J, et al. Spectrum of extratesticular and testicular pathologic conditions at scrotal MR imaging. *Radiographics* (2018) 38(3):806–30. doi: 10.1148/rg.2018170150
- Baird DC, Meyers GJ, Hu JS. Testicular cancer: diagnosis and treatment. *Am Fam Phys* (2018) 97(4):261–8.
- Mondal K, Mandal R, Saha A, Shahabuddin MD, Sarkar R. Fine needle aspiration cytology of epididymal nodules and its corroboration with ultrasonographic-histological findings. *Diagn Cytopathol* (2020) 48(2):118–27. doi: 10.1002/dc.24336
- Collaborators GBDMD. Global, regional, and national burden of 12 mental disorders in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet Psychiatry* (2022) 9(2):137–50. doi: 10.1016/S2215-0366(21)00395-3
- Kocak B, Ates E, Durmaz ES, Ulsan MB, Kilickesmez O. Influence of segmentation margin on machine learning-based high-dimensional quantitative CT texture analysis: a reproducibility study on renal clear cell carcinomas. *Eur Radiol* (2019) 29(9):4765–75. doi: 10.1007/s00330-019-6003-8
- Tang Z, Yu D, Ni T, Zhao T, Jin Y, Dong E. Quantitative analysis of multiphase contrast-enhanced CT images: A pilot study of preoperative prediction of fat-poor angiolipoma and renal cell carcinoma. *AJR Am J Roentgenol* (2020) 214(2):370–82. doi: 10.2214/AJR.19.21625
- Liu Z, Li Z, Qu J, Zhang R, Zhou X, Li L, et al. Radiomics of multiparametric MRI for pretreatment prediction of pathologic complete response to neoadjuvant chemotherapy in breast cancer: A multicenter study. *Clin Cancer Res* (2019) 25(12):3538–47. doi: 10.1158/1078-0432.CCR-18-3190
- Yuan M, Zhang YD, Pu XH, Zhong Y, Li H, Wu JF, et al. Comparison of a radiomic biomarker with volumetric analysis for decoding tumour phenotypes of lung adenocarcinoma with different disease-specific survival. *Eur Radiol* (2017) 27(11):4857–65. doi: 10.1007/s00330-017-4855-3
- Haider MA, Vosough A, Khalvati F, Kiss A, Ganeshan B, Bjarnason GA. CT texture analysis: a potential tool for prediction of survival in patients with metastatic clear cell carcinoma treated with sunitinib. *Cancer Imaging* (2017) 17(1):4. doi: 10.1186/s40644-017-0106-8
- Bian Y, Guo S, Jiang H, Gao S, Shao C, Cao K, et al. Radiomics nomogram for the preoperative prediction of lymph node metastasis in pancreatic ductal adenocarcinoma. *Cancer Imaging* (2022) 22(1):4. doi: 10.1186/s40644-021-00443-1
- Xu JW, Suzuki K. Max-AUC feature selection in computer-aided detection of polyps in CT colonography. *IEEE J BioMed Health Inform* (2014) 18(2):585–93. doi: 10.1109/JBHI.2013.2278023
- Dasgupta S, Goldberg Y, Kosorok MR. Feature elimination in kernel machines in moderately high dimensions. *Ann Stat* (2019) 47(1):497–526. doi: 10.1214/18-AOS1696
- Oliver TRW, Chappell L, Sanghvi R, Deighton L, Ansari-Pour N, Dentre SC, et al. Clonal diversification and histogenesis of Malignant germ cell tumours. *Nat Commun* (2022) 13(1):4272. doi: 10.1038/s41467-022-31375-4
- Cheng L, Albers P, Berney DM, Feldman DR, Daugaard G, Gilligan T, et al. Testicular cancer. *Nat Rev Dis Primers* (2018) 4(1):29. doi: 10.1038/s41572-018-0029-0
- Lotti F, Bertolotto M, Maggi M. Historical trends for the standards in scrotal ultrasonography: What was, what is and what will be normal. *Andrology* (2021) 9(5):1331–55. doi: 10.1111/andr.13062
- Tsili AC, Bougia CK, Pappa O, Argyropoulou MI. Ultrasonography of the scrotum: Revisiting a classic technique. *Eur J Radiol* (2021) 145:110000. doi: 10.1016/j.ejrad.2021.110000
- Rajpert-De Meyts E, Jorgensen N, Petersen JH, Almstrup K, Aaklaede L, Lauritsen J, et al. Optimized detection of germ cell neoplasia in situ in contralateral biopsy reduces the risk of second testis cancer. *BJU Int* (2022) 130(5):646–54. doi: 10.1111/bju.15774
- An H, Wang Y, Wong EMF, Lyu S, Han L, Peruchio JAU, et al. CT texture analysis in histological classification of epithelial ovarian carcinoma. *Eur Radiol* (2021) 31(7):5050–8. doi: 10.1007/s00330-020-07565-3
- Bektas CT, Kocak B, Yardimci AH, Turkcanoglu MH, Yucetas U, Koca SB, et al. Clear cell renal cell carcinoma: machine learning-based quantitative computed tomography texture analysis for prediction of fuhrman nuclear grade. *Eur Radiol* (2019) 29(3):1153–63. doi: 10.1007/s00330-018-5698-2
- Erdim C, Yardimci AH, Bektas CT, Kocak B, Koca SB, Demir H, et al. Prediction of benign and Malignant solid renal masses: machine learning-based CT texture analysis. *Acad Radiol* (2020) 27(10):1422–9. doi: 10.1016/j.acra.2019.12.015
- Yang G, Nie P, Zhao L, Guo J, Xue W, Yan L, et al. 2D and 3D texture analysis to predict lymphovascular invasion in lung adenocarcinoma. *Eur J Radiol* (2020) 129:109111. doi: 10.1016/j.ejrad.2020.109111
- Ling CX, Huang J, Zhang H. AUC: a Better Measure than Accuracy in Comparing Learning Algorithms. Available at: <https://site.uottawa.ca/~stan/csi7162/presentations/William-presentation.pdf>.
- Deng Y, Soule E, Samuel A, Shah S, Cui E, Asare-Sawiri M, et al. CT texture analysis in the differentiation of major renal cell carcinoma subtypes and correlation with Fuhrman grade. *Eur Radiol* (2019) 29(12):6922–9. doi: 10.1007/s00330-019-06260-2
- Sidhu HS, Benigno S, Ganeshan B, Dikaos N, Johnston EW, Allen C, et al. "Textural analysis of multiparametric MRI detects transition zone prostate cancer". *Eur Radiol* (2017) 27(6):2348–58. doi: 10.1007/s00330-016-4579-9
- Marko J, Wolfman DJ, Aubin AL, Sesterhenn IA. Testicular seminoma and its mimics: from the radiologic pathology archives. *Radiographics* (2017) 37(4):1085–98. doi: 10.1148/rg.2017160164
- Cheng H, Shang D, Zhou R. Germline stem cells in human. *Signal Transduct Target Ther* (2022) 7(1):345. doi: 10.1038/s41392-022-01197-3
- van 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(21):e104–e7. doi: 10.1158/0008-5472.CAN-17-0339

36. Zhang X, Zhang Q, Guo J, Zhao J, Xie L, Zhang J, et al. Added-value of texture analysis of ADC in predicting the survival of patients with 2018 FIGO stage IIICr cervical cancer treated by concurrent chemoradiotherapy. *Eur J Radiol* (2022) 150:110272. doi: 10.1016/j.ejrad.2022.110272
37. Chiu FY, Yen Y. Imaging biomarkers for clinical applications in neuro-oncology: current status and future perspectives. *biomark Res* (2023) 11(1):35. doi: 10.1186/s40364-023-00476-7
38. Feng Z, Rong P, Cao P, Zhou Q, Zhu W, Yan Z, et al. Machine learning-based quantitative texture analysis of CT images of small renal masses: Differentiation of angiomyolipoma without visible fat from renal cell carcinoma. *Eur Radiol* (2018) 28(4):1625–33. doi: 10.1007/s00330-017-5118-z
39. Takahashi K, Al-Janabi NJ. Computed tomography and magnetic resonance imaging of mediastinal tumors. *J Magn Reson Imaging* (2010) 32(6):1325–39. doi: 10.1002/jmri.22377
40. Marx A, Chan JKC, Chalabreysse L, Dacic S, Detterbeck F, French CA, et al. The 2021 WHO classification of tumors of the thymus and mediastinum: what is new in thymic epithelial, germ cell, and mesenchymal tumors? *J Thorac Oncol* (2022) 17(2):200–13. doi: 10.1016/j.jtho.2021.10.010
41. Ozgun G, Nappi L. Primary mediastinal germ cell tumors: A thorough literature review. *Biomedicines* (2023) 11(2):6. doi: 10.3390/biomedicines11020487



## OPEN ACCESS

## EDITED BY

Poonam Yadav,  
Northwestern University, United States

## REVIEWED BY

Summrina Kanwal,  
Independent Researcher, Helsingborg, Sweden  
L. J. Muhammad,  
Bayero University Kano, Nigeria

## \*CORRESPONDENCE

Anum Masood  
✉ anum.masood@ntnu.no

RECEIVED 09 November 2023

ACCEPTED 12 January 2024

PUBLISHED 08 February 2024

## CITATION

Ullah MS, Khan MA, Masood A, Mzoughi O, Saidani O and Alturki N (2024) Brain tumor classification from MRI scans: a framework of hybrid deep learning model with Bayesian optimization and quantum theory-based marine predator algorithm.  
*Front. Oncol.* 14:1335740.  
doi: 10.3389/fonc.2024.1335740

## COPYRIGHT

© 2024 Ullah, Khan, Masood, Mzoughi, Saidani and Alturki. 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.

# Brain tumor classification from MRI scans: a framework of hybrid deep learning model with Bayesian optimization and quantum theory-based marine predator algorithm

Muhammad Sami Ullah<sup>1</sup>, Muhammad Attique Khan<sup>1</sup>,  
Anum Masood<sup>2\*</sup>, Olfa Mzoughi<sup>3</sup>, Oumaima Saidani<sup>4</sup>  
and Nazik Alturki<sup>4</sup>

<sup>1</sup>Department of Computer Science, HITEC University, Taxila, Pakistan, <sup>2</sup>Department of Physics, Norwegian University of Science and Technology, Trondheim, Norway, <sup>3</sup>Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Al-Kharj, Saudi Arabia, <sup>4</sup>Department of Information Systems, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Riyadh, Saudi Arabia

Brain tumor classification is one of the most difficult tasks for clinical diagnosis and treatment in medical image analysis. Any errors that occur throughout the brain tumor diagnosis process may result in a shorter human life span. Nevertheless, most currently used techniques ignore certain features that have particular significance and relevance to the classification problem in favor of extracting and choosing deep significance features. One important area of research is the deep learning-based categorization of brain tumors using brain magnetic resonance imaging (MRI). This paper proposes an automated deep learning model and an optimal information fusion framework for classifying brain tumor from MRI images. The dataset used in this work was imbalanced, a key challenge for training selected networks. This imbalance in the training dataset impacts the performance of deep learning models because it causes the classifier performance to become biased in favor of the majority class. We designed a sparse autoencoder network to generate new images that resolve the problem of imbalance. After that, two pretrained neural networks were modified and the hyperparameters were initialized using Bayesian optimization, which was later utilized for the training process. After that, deep features were extracted from the global average pooling layer. The extracted features contain few irrelevant information; therefore, we proposed an improved Quantum Theory-based Marine Predator Optimization algorithm (QTbMPA). The proposed QTbMPA selects both networks' best features and finally fuses using a serial-based approach. The fused feature set is passed to neural network classifiers for the

final classification. The proposed framework tested on an augmented Figshare dataset and an improved accuracy of 99.80%, a sensitivity rate of 99.83%, a false negative rate of 17%, and a precision rate of 99.83% is obtained. Comparison and ablation study show the improvement in the accuracy of this work.

#### KEYWORDS

brain tumor, MRI, contrast enhancement, deep learning, hyperparameters optimization, feature selection

## 1 Introduction

One of the deadliest brain disorders is a brain tumor, which develops from an abnormal development of tissue inside the skull. Primary and secondary forms can be distinguished among them. 70% of cases of primary brain tumors only spread within the brain (1). In contrast, secondary tumors start in an organ like the breast, kidney, or lung before metastasizing to the brain (2). The World Health Organization (WHO) divides malignant gliomas into two categories: grade IV/IV tumors, which include glioblastoma multiforme (GBM), and grade III/IV tumors, which include anaplastic astrocytoma, anaplastic oligodendroglioma, anaplastic oligoastrocytoma, and anaplastic ependymomas. With an incidence rate of 3.19 cases per 100,000 person a year and a median age of 64, GBM is the most prevalent malignant brain tumor. It makes up 80% of all primary malignant CNS tumors and 45.2% of all malignant CNS tumors. GBM is 1.5 times more common in men than in women, and it is twice as common in white people as it is in black people (3).

Meningioma is the most common primary tumor of the central nervous system, with 5/100,000 annual occurrence. Radiation therapy and hormone use are risk factors. According to the WHO's 2016 histological criteria, the majority of meningiomas are grade I benign tumors; however, up to 15% can be atypical and 2% can be anaplastic (4). Pituitary adenomas usually are benign tumors that develop from unusual pituitary gland cell development. They appear either by producing too much hormone or by putting pressure on the surrounding structures, which causes less hormone to be secreted. Prolactinomas, non-functioning adenomas, adenomas that secrete growth hormone, and adenomas that secrete adrenocorticotrophic hormones are the four primary forms. Less frequent kinds include gonadotroph adenomas with clinically significant luteinizing hormone, follicle-stimulating hormone secretion, and thyroid-stimulating hormone-secreting adenomas. Pituitary incidentalomas are a subtype that was unintentionally found while undergoing brain MRI. They can be divided into macroadenomas (bigger, accounting for roughly 40% of occurrences) and microadenomas (less than 1 cm in diameter). Macroadenomas can strain essential structures and regions like the optic chiasm (5).

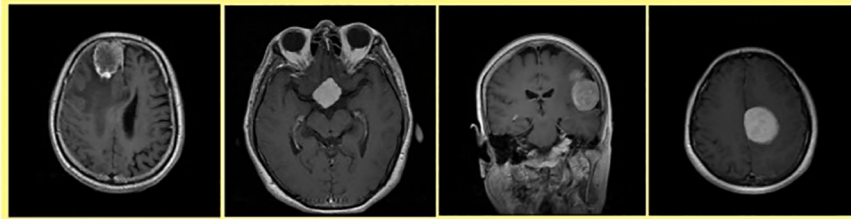
Gliomas and meningioma emerge from neuroglial and brain membranes, respectively; both are the most frequent primary brain

cancers. Pituitary gland and nerve sheath tumors are also included in this group. High-grade gliomas are a common form of malignant tumor. Meningiomas are typically benign; however, they can occasionally turn cancerous (6). Gliomas are more common in men, whereas meningiomas are more common in women; other brain cancers affect both sexes equally (7). Pituitary tumors, whether benign or malignant, can have severe consequences due to their location. Malignant tumors spread quickly, whereas benign tumors develop slowly and are generally entirely eradicated through surgery (8).

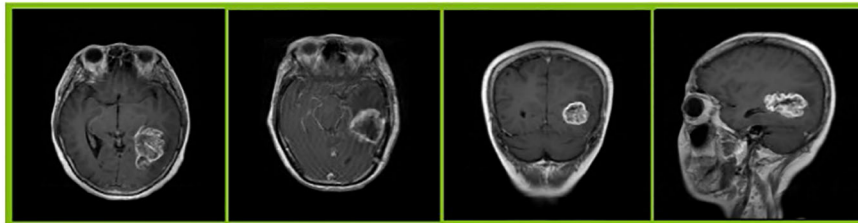
Radiologists and clinicians have substantial difficulties in detecting brain tumors. Brain tumor images produced in medical settings might be challenging to analyze. As a result, there is a need for computer-aided procedures with increased early detection accuracy. Currently, there is a lot of interest in using multimodal images to classify brain tumors (9). Magnetic resonance imaging (MRI) is frequently used to diagnose brain malignancies. A tumor can be found via MRI, commonly used to identify brain tissues based on their size, shape, or location (10). Figshare is a publicly available MRI image-based brain tumor dataset containing 3,064 T1-weighted contrast-enhanced images. These are obtained from 233 patients. A total of 1,426 slices of glioma, 708 slices of meningioma, and 930 slices of pituitary tumors are included in said dataset (11, 12). A few sample images are shown in Figure 1.

In recent years, interest in computer vision has grown across various fields of studies, from medical to industrial robotics. Computer science and advances in image processing techniques have greatly aided computer vision (13). Deep learning is a diverse set of techniques that includes neural networks, hierarchical probabilistic models, and a wide range of unsupervised and supervised feature learning algorithms. Deep learning approaches have recently gained popularity because of their ability to beat prior state-of-the-art techniques in various tasks and the amount of complex data from various sources (e.g., visual, auditory, medical, social, and sensor) (14). Deep learning has made significant advances in a wide range of computer vision tasks, including object recognition (15), motion tracking (16), and medical image classification and detection (17, 18). Classification of brain tumors for medical specialists is an important field where computer vision and deep learning techniques work together and bring prosperity to patients with non-invasive diagnosis of brain tumors using MRI.

## Meningioma



## Glioma



## Pituitary

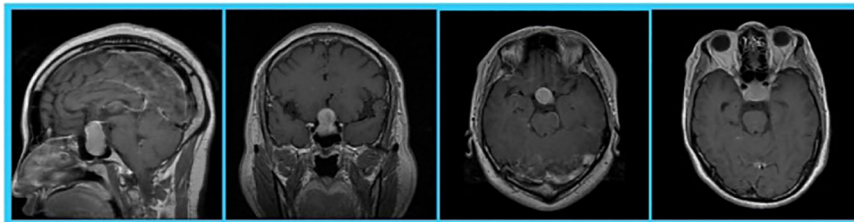


FIGURE 1  
Sample MRI images of brain Meningioma, glioma, and pituitary tumors.

### 1.1 Aims and objectives

Image acquisition from MRI has loss of information that leads to improper feature visibility. A technique is required to employ that can enhance the contrast of MRI images so that loss of information during the acquisition process can be minimized. Hence, feature visibility can be improved and classification problems can be addressed, which has a close relationship with features visibility. In order to address classification problems for MRI images of brain tumors, there is an immense need to introduce a technique using state-of-the-art deep learning methods. In a quest to fulfill this need, a deep learning technique should acquire brain tumor MRI images from publically available benchmark datasets. The selected dataset explained in a related section of this manuscript has significant imbalance classes, so it is important to incorporate a data augmentation technique that can gracefully fill the gap of imbalanced dataset classes. After enhancement of contrast and data augmentation steps, lightweight pretrained deep learning models need to be deployed and modified based on the low complexity for training of the balanced dataset. Optimization of hyperparameters to train deep learning models is required to select the optimal combination of values for model training on the selected dataset. Extracted features can be optimized using some

optimization algorithms and then be fused together. Feature fusion greatly impacts the overall classification accuracy. The subsequent section presents the major challenges in order to develop an aimed technique and contribution to address these challenges in proposed work.

### 1.2 Major challenges and contributions

This imbalance in the training dataset impacts the performance of deep learning models because it causes the classifier performance to become biased in favor of the majority class. The authors tried to resolve this issue by using few traditional techniques such as flip image and rotate image, and few of the authors performed contrast enhancement. However, these techniques are not enough, and the images are highly duplicated. Therefore, it is essential to address this challenge by employing some of the latest techniques, such as GAN and encoders. Still, most currently used feature selection techniques ignore certain features that have particular significance and relevance to the classification problem in favor of extracting and choosing deep significance features. We proposed a hybrid deep learning framework with BO and QTbMPA feature selection algorithms to address these challenges. Our major contributions are listed below.



- Sparse Autoencoder architecture was proposed for the generation of new images based on the training data for the augmentation process.
- Two lightweight pretrained deep learning models were fine-tuned based on the additional layers and removal of pooling layers. The models were trained from scratch on an augmented dataset.
- A Bayesian optimization technique was implemented to initialize the hyperparameters of the fine-tuned deep models for improved learning.
- An efficient Quantum Theory-based Marine Predator Optimization algorithm was proposed for the selection of best features for the final classification.
- A detailed ablation study was performed on the proposed framework for the validation of the proposed framework.

## 2 Literature review

A wide range of classification approaches have been introduced for the Figshare dataset. Several techniques have been introduced in the literature for the classification of brain tumor from MRI images. Researchers used deep learning models for the feature extraction and later performed classification using Softmax and machine learning classifiers. A novel deep transfer learning-based model was identified by Alanazi et al. (19). It entails creating several convolutional neural network models and then utilizing transfer learning to repurpose a 22-layer model for subclass classification. The proposed model achieved 95.75% accuracy on three classes of the Figshare dataset. Moreover, the technique was also validated for an unseen dataset and achieved an accuracy of 96.89%. Another DeepTumorNet hybrid deep learning model was suggested by Raza et al. (20). The last five layers of GoogleNet were eliminated while creating the hybrid DeepTumorNet technique, and 15 new layers were added. They used the feature map's leaky ReLU activation function to make the model more expressive. The suggested model was evaluated on the Figshare dataset and achieved 99.67% accuracy. Tummala et al. (21) used ensemble-oriented vision transformer-based pretrained models to classify the modalities of the Figshare dataset. An ensemble of B/16, B/32, L/16, and L/32 was used. The selected approach achieved an overall accuracy of 98.70%. Attention mechanism, patch-oriented input, and token embedding are techniques used in vision transformers, which make them more computationally expensive, and processing requires a tensor processing unit (TPU) environment.

Another work by Polat et al. (22) introduced a novel divergence-based feature extractor which is used for classification by decreasing weights for deep neural networks. The achieved accuracy was 99.18%. They have reduced the input image dimensions considerably (i.e.,  $512 \times 512$  to  $128 \times 128$ ), which can result in loss of spatial information. Loss of information at the input level can result in compromised accuracy. A technique that uses a multilevel attention network (MANet) (23) was suggested by Shaik et al. in which the model has an attention mechanism with several tiers of attention blocks and can concentrate on crucial spatial and category-specific properties.

Prioritizing tumor details in the image is done by the first attention block, and the second attention layer is highlighted by the tumor-specific descriptors using ConvLSTM. MRI images are represented as input to the model using pretrained features from the Xception network. The resultant accuracy of 96.51 for the Figshare dataset was obtained. In the presented technique, only those glioma images with tumor in it will be classified. A CNN-based approach was created by Haq et al. (24); they performed classification as well as segmentation. A classification accuracy of 98% was achieved. The proposed algorithm has a long running time and needs an improvement to reduce the running time. In another technique, Rahman et al. (25) implemented a Parallel Deep Convolutional Neural Network (PDCNN) technique. It operates in two concurrent stages to capture both global and local features. The model includes dropout regularization and batch normalization to alleviate the overfitting issue. The classification accuracy is 97.60%. The proportion of 80:20 training and testing data was respectively used. A major proportion of training data may lead to overfitting as it becomes specialized for known data but not for unseen or unknown data.

The authors Talukder et al. (22) presented a technique to classify brain tumors. They used different pretrained models and obtained an accuracy of 99.68% on ResNet50V2. The lack of sharp images is the main shortcoming of this study. In their work, Aloraini et al. (26) presented another technique in which the authors utilized a hybrid method combining a transformer with an attention mechanism to capture global features. Local features were extracted using a convolutional neural network (CNN). The approach attained an accuracy of 99.10% for the Figshare dataset. Few misclassifications were reported due to visual similarity between classes. In their work, authors Athisayamani et al. (27) introduced a new adaptive Canny Mayfly algorithm for edge identification. An algorithm that reduces the dimension of retrieved features, the enhanced chimpanzee optimization algorithm (EChOA), is utilized to choose features. The feature classification process is then done using the Softmax classifier and ResNet-152. The proposed technique achieved an accuracy of 98.85%. In their presented work, the authors Mishra et al. (28) provided a method for classifying brain tumors using a K-NN classifier, where the parameter  $k$  is adjusted and the best feature set is selected using the binary version of the comprehensive learning elephant herding optimization (CLEHO) algorithm. The presented method obtained an accuracy of 98.97%, better than the recent techniques. A pretrained model-based approach was suggested by the authors Malla et al. (25), in which a transfer learning DCNN framework known as VGGNet was used. They employed transfer learning aspects such as fine-tuning the convolutional network and freezing layers for better performance. Features were extracted from the Global Average Pooling (GAP) layer. The technique resulted in an accuracy of 98.93% on the Figshare dataset. In the given approach, the feature dimensionality issue was not addressed, and that intended to address it in future research.

In another work, authors Cinar et al. (29) presented a Convolutional Neural Network (CNN) architecture for brain tumor classification. The model was compared with ResNet50, VGG19, DenseNet121, and InceptionV3 pretrained models. The presented model achieved an average classification accuracy of 98.32% on the prescribed dataset. The authors determined to

enhance their technique using area and size-oriented metrics. In another technique, the authors Deepak et al. (30) coined an approach in which they trained CNN using three different methods: cross-entropy loss, class-weighted loss, and weighted local loss. They fused the features, and classification was performed with an accuracy of 95.40%. Another approach by authors by authors Zulfiqar et al. (31) suggested an approach in which five variations of the EfficientNets family's pretrained models, EfficientNetB0 through EfficientNetB4, were fine-tuned. They also investigated how data augmentation affects the model's accuracy. The best model's attention maps are finally visualized using Grad-CAM, successfully highlighting the tumorous region of the brain cell. The achieved accuracy was 98.86%.

### 3 Methodology

The proposed methodology of brain tumor classification is illustrated in Figure 2. This section starts with the preprocessing phase in which the Figshare brain tumor dataset (32) is obtained. The contrast enhancement step is crucial to improving the quality of

low-contrast images, and it was performed using a statistical technique presented in (33). Data augmentation is performed on contrast-enhanced images. This step is taken into account due to the high imbalance of classes in the original dataset. Augmentation of the data is performed using sparse autoencoders (34). The said technique augments the data by learning the most important features of the original data and leaving behind the least important features. Two pretrained models named InceptionResNetV2 (35) and EfficientNetB0 (36) are used and fine-tuned for the input of preprocessed data. Dynamic and optimized selection of hyperparameters of both models is carried out using Bayesian-based optimization (37). Features are extracted from each optimized resultant model. To further optimize the features, a nature-inspired algorithm named the Marine Predators Algorithm (MPA) (38) is used on the obtained features of each model. Feature fusion is carried out, final classification is performed.

#### 3.1 Dataset of this work

The Figshare dataset includes 3,064 T1 weighted contrast-enhanced MRI scans collected from 233 patients. There are three

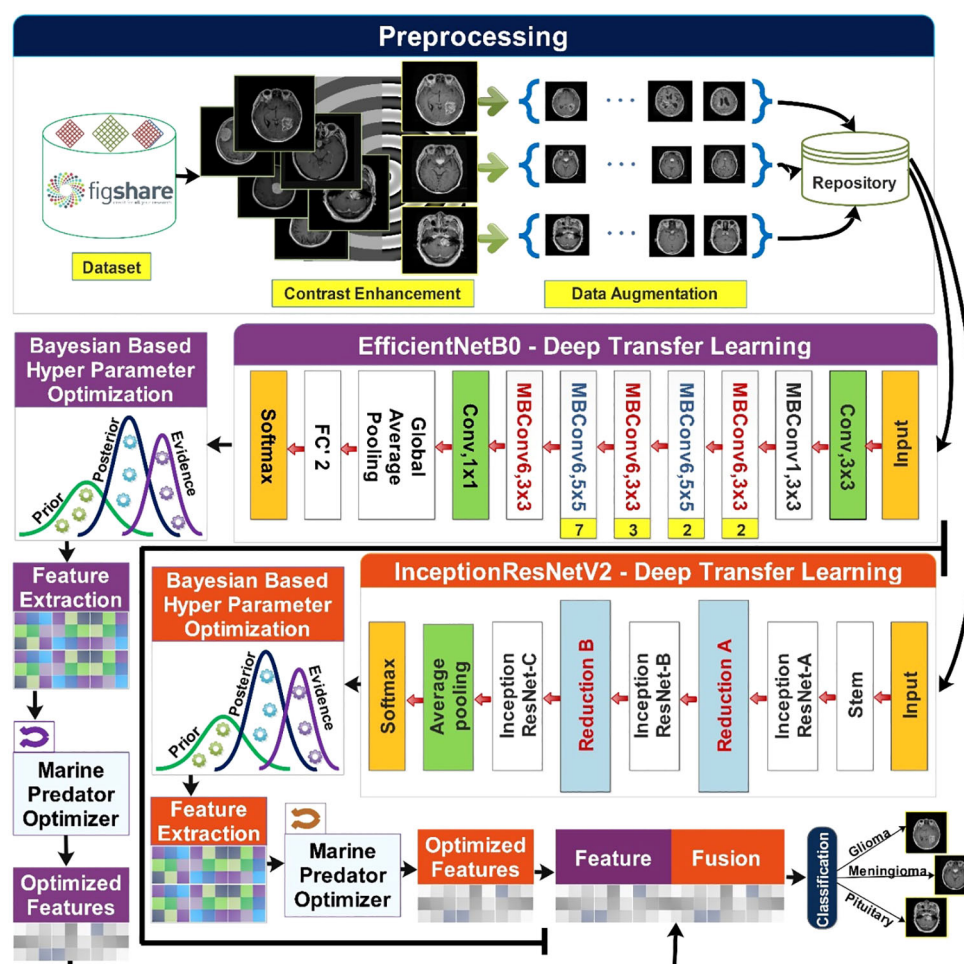


FIGURE 2  
Proposed methodology of brain tumor classification using deep learning and optimization algorithm.

classes of these scans named meningioma, glioma, and pituitary, with 708, 1,426, and 930 MRI scans, respectively, in each class (32). Meningiomas are the most prevalent intracranial tumor, accounting for more than one-third of all primary central nervous system (CNS) tumors. They are typically benign tumors that can be observed or preferentially treated with extensive complete resection, which results in satisfactory outcomes. Meningioma with complex histology or in vulnerable areas has proven difficult to treat and predict prognostic outcomes (39).

Gliomas are divided into different categories based on the cells of their origin. They make up around 80% of all malignant primary brain tumors and are most frequent malignancies of the central nervous system (CNS). The most dangerous and common variety of glioma is called glioblastoma multiforme (GBM). More than 60% of adult brain tumors are caused by it. Despite the wide range of contemporary treatments available, GBM remains a fatal condition with a very bad prognosis. The median survival time for patients is typically 14 to 15 months after diagnosis of the deadly disease (40).

The anterior pituitary gland is the site of tumors called pituitary adenomas. They rank as the third most frequent adult cause of central nervous system malignancies (CNS). Most benign adenomas cause either a large-scale effect or an increase in hormone release. Depending on their size and hormone produced, pituitary adenomas appear differently in clinical evaluations (41). Samples of meningioma, glioma, and pituitary brain tumors from the Figshare dataset are presented in Figure 1.

## 3.2 Contrast enhancement

Analyzing medical images is challenging because of the inherent qualities present in medical images, such as poor contrast, speckle noise, signal dropouts, and complicated anatomical formation. Contrast enhancement is a vital component of subjective evaluation of image quality that aims to improve the overall excellence of medical imagery for feature visualization and clinical measurement (42). In fact, despite technological advancements in healthcare systems, they still produce images that demonstrate a deficiency in contrast due to improper locales and equipment limitations. To enhance the contrast of MRI images of the dataset discussed above, an existing technique for contrast enhancement (33) is employed. It uses basic statistics and some basic image processing methods. The approach adjusts the global and local contrast of a given image separately, then combines both results using logarithmic image processing (LIP), producing an output that is further analyzed by an adaptive linear stretching method to produce the improved version of the image. The overall process of contrast enhancement is defined as follows:

Letting a low-contrast image  $g_{(x,y)}$ , at first, the local contrast is altered using contrast stretching transformation (CST). The CST process is defined in Equation 1.

$$k_{(x,y)} = \frac{1}{1 + (m/g_{(x,y)})^E} \quad (1)$$

In the above equation,  $k_{(x,y)}$  is the output of the CST procedure where  $x, y$  represents the dimensions of an image. The slope of the function is set by constant  $E$ , and its value is set to 0.5 for this experiment. The mean value of the input image is represented by  $m$ . A standard logic function is applied to the original image to change its global contrast. Mathematically, it is defined in Equation 2.

$$j_{(x,y)} = \frac{\exp(g_{(x,y)})}{1 + \exp(g_{(x,y)})} \quad (2)$$

The resultant images with altered local and global contrast will be combined. The Logarithmic Image Processing (LIP) method devised in (43) is for this purpose and is mathematically defined as follows:

$$f_{(x,y)} = \frac{k_{(x,y)} + j_{(x,y)}}{1 + (k_{(x,y)} * j_{(x,y)})} \quad (3)$$

An exponent  $W$  is used to control the enhancement, and the entire equation is raised to its power of it. The scalar parameter ( $W > 0$ ) and its higher value lead to achieving a good level of contrast enhancement. Mathematically, it is defined in Equation 4.

$$f_{(x,y)} = \left[ \frac{k_{(x,y)} + j_{(x,y)}}{1 + (k_{(x,y)} * j_{(x,y)})} \right]^W \quad (4)$$

Contrast enhancement of the image has been achieved after employing Equation 4, but the image  $f_{(x,y)}$  does not correspond to the natural range of pixel values. A linear stretching method with adaptive form (40) brings a natural range of pixel values to the image. Mathematically, it is defined by Equation 5.

$$t_{(x,y)} = \alpha * f_{(x,y)} - \beta \quad (5)$$

where  $t_{(x,y)}$  is a resultant image, and  $\alpha$  and  $\beta$  are the control variables for the stretching process. The value of these control variables is adjusted manually, but here, Equation 6 and Equation 7 are used to select the values of these variables automatically.

$$\alpha = \frac{1}{\max(f_{(x,y)}) - \min(f_{(x,y)})} \quad (6)$$

$$\beta = \frac{\min(f_{(x,y)})}{\max(f_{(x,y)}) - \min(f_{(x,y)})} \quad (7)$$

In the above equations, the variables  $\max$  and  $\min$  represent the upper and lower bounds of values for pixels of an image  $f_{(x,y)}$ , respectively. The pseudocode of the above mathematical description is given under Pseudo-code 1. A few visual images are also illustrated in Figure 3.

**Input:** Original image  $g_{(x,y)}$  and a parameter  $W$ .  
 Computation of CST method by using Equation 1.  
 Estimation of SL function by using Equation 2.  
 Calculation of modified LIP method by using Equation 4.  
 Computation of parameters  $\alpha$  and  $\beta$  by using Equations 6 and 7.  
 Processing of contrast by using Equation 5.  
**Output:** Contrast Enhanced image  $t_{(x,y)}$

Pseudo Code 1. Proposed Contrast Enhancement Technique.

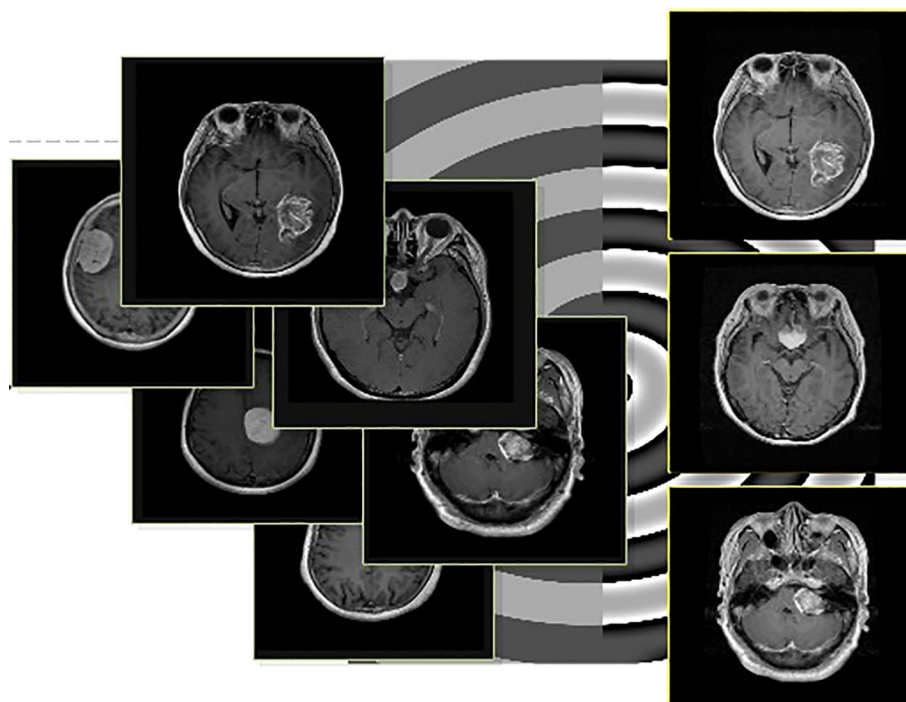


FIGURE 3

Visual illustration of the contrast enhancement process. The left images are original, and the right images are generated using contrast enhancement.

### 3.3 Data augmentation

Classification performance is negatively impacted by class imbalance. The impact of imbalance on classification performance gets more robust with increasing task size. The effect of imbalance depends on the distribution of observations (i.e., images) throughout the classes and cannot be solely attributed to a lower overall number of training cases (44). In Section 3.1, it is noted that our dataset has a high-class imbalance. Hence, creating a dataset bias may lead to an overfitting problem for some classes. To fill that gap, we employed a sparse autoencoder (45) to augment the dataset instead of traditional methods.

Sparse Autoencoders learn a compressed representation of the input data. The following hyperparameters are used to train a sparse autoencoder network:

| Hyperparameters        | Value |
|------------------------|-------|
| Hidden size            | 300   |
| Maximum epochs         | 2000  |
| L2WeightRegularization | 0.001 |
| SparsityRegularization | 4     |
| SparsityProportion     | 0.15  |

Hidden size parameter represents the number of neurons in layers. Few dozen neurons are enough for simpler tasks, but in order

to use it with complex tasks, a few hundred neurons are used. A hidden size of 300 might be able to prevent overfitting while still having sufficient capacity to learn from the data, particularly in situations where bigger hidden sizes could cause overfitting.

One whole cycle through the whole training dataset is referred to as an epoch. The hyperparameter for maximum epochs indicates the maximum number of times the training dataset will be processed by the learning algorithm. In the proposed technique, the training dataset for augmentation took 2,000 epochs to converge at a suitable result for MRI images.

The intensity or weight of L2 regularization given to a neural network's weights during training is commonly denoted by the hyperparameter L2WeightRegularization, which has a value of 0.001. The selection of 0.001 maintains a balance between letting the model learn from the data and regularizing it to avoid overfitting. It is also referred to as weight decay.

The sparsity regularization weight that is given to a neural network during training is indicated by the hyperparameter SparsityRegularization, and the chosen value for it is 4. By encouraging the model to have fewer active (non-zero) weights, the objective is to cause the weight matrices to become sparse, which means that during the training phase, a large number of the weights are driven to be zero or almost zero. Sparsity regularization helps to create a more effective and sparse representation for better feature selection.

The hyperparameter of "SparsityProportion" with a value of 0.15 commonly refers to a threshold sparsity level, which is used



with sparsity regularization. The target of around 15% of the neural network's weights becoming zero or almost zero is indicated by the value of 0.15. The sparsity regularization hyperparameter sets a threshold of 4, and weights that are below threshold are settled to zero. The value of 0.15 represents the proportion of weights that should actually fall below the specified threshold value during the training process.

The specified values for each hyperparameter are adjusted for augmentation of MRI images. The resultant images obtained from this step are used to augment the data.

The overall representation of sparse autoencoders is provided in Figure 4.

The total number of observations for each class increased to 2,000 after employing the proposed sparse encoder network. The newly generated images have been utilized to train selected deep learning models.

### 3.4 Hyperparameter selection for modified EfficientNetB0 and InceptionResNetV2

The augmented dataset is used to train fine-tuned deep-learning models. Three hyperparameters for both models are optimized using Bayesian optimization to train the models. These hyperparameters are named InitialLearnRate, Momentum, and L2Regularization. The dynamic tuning of hyperparameters is a crucial task for deep learning models. In this case, dynamically selected values for specific hyperparameters are used until a specific best-value threshold is achieved. The particular model is then trained, and features are extracted for classification tasks.

Bayesian optimization (BO) is an effective technique for hyperparameter tuning. Implementation (46) can be achieved by setting an optimization goal. The Equations 8 and 9 below describes the BO process.

$$x' = \arg \max_{x \in A} f(x) \quad (8)$$

In the equation, search space is  $A$  for input  $x$ . BO is based on the Bayes theorem that is mathematically defined as follows:

$$P(D | F) \propto P(F | D)P(D) \quad (9)$$

Given that an event or hypothesis  $F$  has occurred, it is the likelihood that the event or hypothesis  $D$  will also occur, where  $F$  denotes the evidence data,  $D$  denotes the model, and  $P(D | F)$  is the posterior probability that is proportional to the likelihood  $P(F | D)$  and is multiplied with a probability of  $D$ . The foundation of BO is the combination of sample data (evidence) and the prior distribution of the function  $f(x)$  to produce the posterior of the function. Then, based on the criterion, the posterior information is used to determine the location where the function  $f(x)$  is maximized. The criterion is also called an acquisition function ( $v$ ) and is used to estimate the next sample point. Sampling points are searched using exploration and exploitation sampling methods while searching the sampling space. Exploration tends to search for sampling areas with high uncertainty. Exploitation searches for those samples that are of high value. These methods improve the performance, even with multiple local maxima solutions.

The prior distribution of the function  $f(x)$ , a crucial component in the statistical inference of the posterior distribution, is a requirement for Bayesian optimization in addition to sample information. The posterior distribution is updated using the Gaussian process to better align with the data, improving our forecasts' accuracy and knowledge. Algorithm 1 describes the working of BO.

```

1: For  $i = 1, 2, \dots$ 
2: Find  $x_i$  by optimizing the acquisition function  $v$  over
   function  $f: x_i = \arg \max_x v(x | D_{1:i-1})$ . (Equation 10)
3: Sample the objective function:  $y_i = f(x_i)$ .
4: Augment the data  $D_{1:i} = \{D_{1:i-1}, (x_i, y_i)\}$  and update the
   posterior of function  $f$ .
5: End For.

```

Algorithm 1. Bayesian optimization.

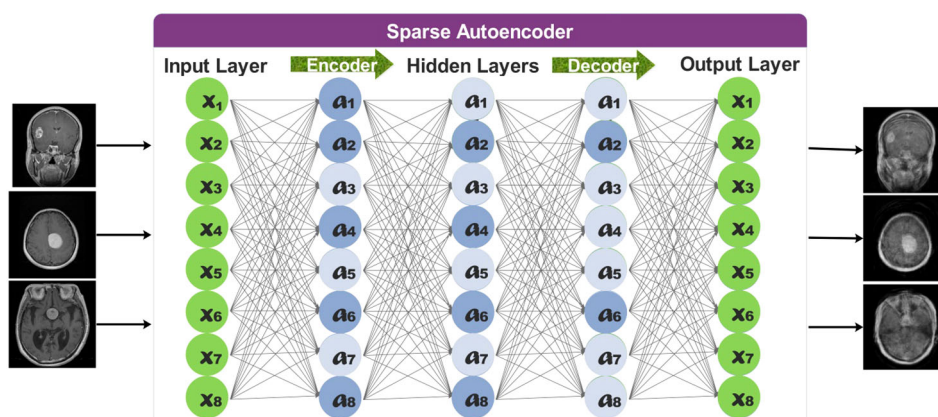


FIGURE 4  
Representation of sparse autoencoder for data augmentation.



The algorithm consists of two parts: acquisition function maximization using step 2 and posterior distribution update using steps 3 and 4. Furthermore, the training dataset is denoted by  $D_{1:i-1} = \{x_n, y_n\}_{n=1}^{i-1}$  with  $i-1$  observations of function  $f$ . Each processed observation updates the posterior distribution. The updated distribution helps to find the highest value of the acquisition function at some point, which is then added to the training dataset. This process continues until the maximum number of iterations is reached or the difference between the current and best values so far is less than a predetermined threshold. The following starting and stopping criteria are selected for experiments. Number of seed points = 4; Maximum Objective Evaluation = 30, and Maximum time = Infinite by default.

A Gaussian process prior with additional Gaussian noise in the observations serves as the fundamental probabilistic model for the objective function  $f$ . Therefore, the Gaussian process with mean  $\mu(x; \theta)$  and covariance kernel function  $k(x, x', \theta)$  represent the prior distribution on  $f(x)$ . Here,  $x$  represents the initial value,  $x'$  denotes the updated value, and  $\theta$  is a parameter containing a kernel vector. Therefore, looking into more detail, we show a set of points  $x = x'$  with associated objective function  $F = f_i$  and the prior joint probability distribution of the function value  $k(x, x')$  where  $k_{ij} = k(x_i, x_j)$  and initially  $\mu = 0$ . Moreover, Gaussian noise is added, which is denoted by  $\sigma^2$  so the prior distribution has covariance  $k(x, x', \theta) + \sigma^2 x$ , and therefore, the final Gaussian process regression is depicted by the following Equation 11.

$$k(x_i, x_j, \theta) = \sigma_f^2 \exp \left[ -\frac{1}{2} \sum_{m=1}^d \frac{(x_i, x_j)^2}{\sigma_m^2} \right] \quad (11)$$

where  $\sigma_m$  is length scale prediction  $m$  and  $m = 1, 2, 3, \dots, d$ ,  $\sigma_f$  is the signal standard deviation,  $\theta_m = \log(\sigma_m)$ ,  $\theta_{d+1} = \log(\sigma_f)$ , and  $k(x, x', \theta)$  is a Kernel function that significantly affects the quality of Gaussian process regression. In Bayesian optimization, the ARD Matern 5/2 kernel is optimized by default and is given in the following Equation 12.

$$k(x_i, x_j | \theta) = \sigma_f^2 \left( 1 + \sqrt{5} r + \frac{5}{3} r^2 \right) \exp(-\sqrt{5} r) \quad (12)$$

where  $r = \sum_{m=0}^d \sqrt{\frac{(x_{im} - x_{jm})^2}{\sigma_m^2}}$ . BO employs the acquisition function to derive the highest value of the function  $f$  after collecting the posterior distribution of the objective function. Typically, we believe the large value of the objective function  $f$  matches the high value of the acquisition function. Therefore, the increasing the acquisition function is the same as increasing the function  $f$ , as presented in Equation 13:

$$x' = \arg \max_{x \in A} u(x | D) \quad (13)$$

The acquisition function named expected improvement per second plus is employed for hyperparameter optimization. The family of acquisition functions known as “expected improvement” assesses the expected rate of improvement in the objective function while ignoring values that increase the objective. The equation for expected improvement is defined in Equation 14:

$$EI(x, Q) = EQ[\max(0, \mu Q(x_{best}) - f(x))] \quad (14)$$

where  $x_{best}$  is the location of the lowest posterior mean and  $\mu Q(x_{best})$  is the lowest value of the posterior mean. The anticipated improvement per second used by the acquisition function during the objective function assessment is formulated in Equation 15:

$$EIPs(x) = \frac{EIQ(x)}{\mu S(x)} \quad (15)$$

where  $\mu S(x)$  is the posterior mean of the timing Gaussian process model. Finally, the maximization process was performed and returned the best hyperparameter value. The initial learn rate (InitialLearnRate) range is 0.01–0.9, the momentum value is selected between 0.8 and 0.98, and the L2Regularization range is  $1e-10$  (0.0000000001) to  $1e-2$  (0.01). To find the values of the hyperparameters, the search space needs to be transformed logarithmically. A logarithmic transformation is used to improve the search process order-of-magnitude balance. Results for optimizing the hyperparameters for EfficientNetB0 are provided in Figure 5A. While optimizing the hyperparameters, the best objective function value is achieved during iteration number 5.

Optimizing results for hyperparameters of the InceptionResNetV2 model are provided in Figure 5B. The best object value (i.e., optimized hyperparameters) is achieved at iteration number 5, the best and last iteration per already defined termination criteria.

### 3.5 Training and feature extraction

Both fine-tuned models have been trained on the augmented dataset, and deep features are extracted from the global average pooling layer. The sigmoid activation function has been employed in the feature extraction process and obtained a feature vector of  $N \times 1280$  and  $N \times 1536$  from fine-tuned EfficientNetB0 and fine-tuned InceptionResNetV2, respectively. The complex patterns are captured from the deeper layers of the above models, and higher spatial dimensions are achieved. The Global Average Pooling layer reduces the higher dimensions to a fixed-size vector; however, optimizing the features' size for accurate classification is necessary.

### 3.6 Improved MPA optimization

In this work, we proposed an improved Quantum Theory-based Marine Predator Algorithm to select the best features. The MPA is a metaheuristic algorithm. Random walk describes the behavior of particles or objects in various physical and biological domains. These are effective methods for studying the movement of organisms such as bacteria or animals looking for food. The random character of each step in these circumstances allows for a realistic picture of how these organisms explore and navigate their surroundings. Lévy and Brown's movements are random walks. Different velocity ratios are extracted and used in the three phases of MPA. These are strategies behind MPA (38). MPA is based on population as many other metaheuristic algorithms. The initial

**A**

| Optimized Hyperparameters for EfficientNetB0 |           |                    |          |                  |
|----------------------------------------------|-----------|--------------------|----------|------------------|
| Iteration                                    | Objective | InitialLearnerRate | Momentum | L2Regularization |
| 1                                            | 0.58933   | 0.50404            | 0.95861  | 0.000000063133   |
| 2                                            | 0.66667   | 0.83335            | 0.82047  | 0.0000000015273  |
| 3                                            | 0.098     | 0.018088           | 0.92972  | 0.000000021129   |
| 4                                            | 0.66467   | 0.17542            | 0.80687  | 0.0087329        |
| 5                                            | 0.03      | 0.010001           | 0.88505  | 0.0000000051622  |
| 6                                            | 0.046     | 0.010024           | 0.90823  | 0.000000012423   |
| 7                                            | 0.034     | 0.010047           | 0.8024   | 0.0000000020056  |

**B**

| Optimized Hyperparameters for InceptionResNetV2 |           |                    |          |                  |
|-------------------------------------------------|-----------|--------------------|----------|------------------|
| Iteration                                       | Objective | InitialLearnerRate | Momentum | L2Regularization |
| 1                                               | 0.66467   | 0.50404            | 0.95861  | 0.000000063133   |
| 2                                               | 0.66933   | 0.83335            | 0.82047  | 0.0000000015273  |
| 3                                               | 0.11467   | 0.018088           | 0.92972  | 0.000000021129   |
| 4                                               | 0.67067   | 0.17542            | 0.80687  | 0.0087329        |
| 5                                               | 0.048     | 0.010006           | 0.88776  | 0.0000047067     |

FIGURE 5

Summary of best selected hyperparameter values using BO. (A) Bayesian optimized (BO) hyperparameters for training of EfficientNetB0. (B) Bayesian optimized (BO) hyperparameters for training of InceptionResNetV2.

solution is homogeneously disseminated over the entire search space through the first sample see, Equation 16.

$$Y_0 = Y_{min} + rand(Y_{max} - Y_{min}) \quad (16)$$

Upper and lower bounds of variables are represented with  $Y_{max}$  and  $Y_{min}$ , respectively, whereas uniform random vector is denoted by  $rand$  whose range is between 0 and 1.

According to the notion of survival of the fittest, top natural predators are better foragers. As a result, the top predator, also known as elite, in the  $E$  matrix is chosen as the fittest solution. The chosen matrix is constructed, and arrays of the matrix provide a detail of searching and finding the prey based on the position of the prey. The matrix is given in Equation 17:

$$E = \begin{bmatrix} Y_{1,1}^I & Y_{1,2}^I & \dots & Y_{1,d}^I \\ Y_{2,1}^I & Y_{2,2}^I & \dots & Y_{2,d}^I \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n,1}^I & Y_{n,2}^I & \dots & Y_{n,d}^I \end{bmatrix}_{n \times d} \quad (17)$$

In the above matrix,  $\overrightarrow{Y^I}$  is a vector representing the top predator, and it is repeated  $n$  times to create the  $E$  matrix.

Dimensions are represented by  $d$ , whereas search agents are denoted by  $n$ . Predators and prey are considered search agents because predators look for its prey and the prey is looking for its food. The  $E$  matrix is updated once a better predator replaces the existing top predator.

Another matrix with the same dimensions is constructed depending on the position of the prey. Predator updates the position based on the prey's position matrix. The matrix is named  $P$  and is given in Equation 18:

$$P = \begin{bmatrix} Y_{1,1} & Y_{1,2} & \dots & Y_{1,d} \\ Y_{2,1} & Y_{2,2} & \dots & Y_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{n,1} & Y_{n,2} & \dots & Y_{n,d} \end{bmatrix}_{n \times d} \quad (18)$$

In  $P$  matrix  $Y_{i,j}$  the  $j$  represents the  $j^{\text{th}}$  dimension, and  $i$  represents the  $i^{\text{th}}$  prey. These two matrices are the backbone for optimization.

There are three phases of MPA. These are based on the predator and prey's life cycle and velocity criteria. These three phases are discussed separately as follows: In the first phase, the predator is considered moving faster than the prey, which is also called as the

high-velocity ratio ( $velocity \geq 10$ ) phase. The ideal predator strategy is to remain still. The mathematical model for this phase is defined in Equation 19:

$$\begin{aligned} \text{While } Iteration < \frac{1}{3} \text{ Max\_Iteration} \\ \overrightarrow{stepsize}_i &= \overrightarrow{R}_B \otimes (\overrightarrow{E}_i - \overrightarrow{R}_B \otimes \overrightarrow{P}_i) \quad i = 1, \dots, n \\ \overrightarrow{P}_i &= \overrightarrow{P}_i + T \cdot \overrightarrow{R} \otimes \overrightarrow{stepsize}_i \end{aligned} \quad (19)$$

This scenario occurs in the first third of iterations. *Iteration* represents the current iteration, whereas *Max\_Iteration* represents maximum iterations.  $R_B$  is a vector containing random values from the normal distribution exhibiting Brownian movement. Entry-wise multiplications are denoted by  $\otimes$ . Movement of prey is simulated by the multiplication of  $\overrightarrow{R}_B \otimes \overrightarrow{P}_i$ . Here,  $T$  is a constant, and its value is 0.5.  $R$  denotes a vector of uniform random numbers between 0 and 1.

The second phase occurs in unit velocity ratio or when the prey and predator move at the same speed. It means that the predator is actively looking for prey, and the prey is actively looking for its food. This optimization stage is where the transition from exploration to exploitation occurs. The prey does exploitation, whereas exploration is the predator's primary goal. Half of the population is designated for exploitation and the other half for exploration. If the velocity ratio ( $velocity \approx 1$ ), then the prey moves in Lévy and the predator follows the Brownian motion. A mathematical model for this is given below:

$$\text{While } \frac{1}{3} \text{ Max\_Iteration} < Iteration < \frac{2}{3} \text{ Max\_Iteration}$$

The first half of the population can be modeled by Equation 20:

$$\begin{aligned} \overrightarrow{stepsize}_i &= \overrightarrow{R}_L \otimes (\overrightarrow{E}_i - \overrightarrow{R}_L \otimes \overrightarrow{P}_i) \quad i = 1, \dots, \frac{n}{2} \\ \overrightarrow{P}_i &= \overrightarrow{P}_i + T \cdot \overrightarrow{R} \otimes \overrightarrow{stepsize}_i \end{aligned} \quad (20)$$

In the above equation,  $\overrightarrow{R}_L$  represents the Lévy movement of the first half of the population. The multiplication of  $\overrightarrow{R}_L \otimes \overrightarrow{P}_i$  describes the Lévy movement of the prey, and adding the step size of the prey position determines its movement. The second half of the population can be modeled in the given below Equation 21:

$$\begin{aligned} \overrightarrow{stepsize}_i &= \overrightarrow{R}_B \otimes (\overrightarrow{R}_B \otimes \overrightarrow{E}_i - \overrightarrow{P}_i) \quad i = \frac{n}{2}, \dots, n \\ \overrightarrow{P}_i &= \overrightarrow{E}_i + T \cdot \overrightarrow{CF} \otimes \overrightarrow{stepsize}_i \end{aligned} \quad (21)$$

where  $CF = (1 - \frac{Iteration}{Max\_Iteration})^{(2 \frac{Iteration}{Max\_Iteration})}$  is regarded as an adaptive parameter to regulate the predator's movement's step size. The multiplication of  $\overrightarrow{R}_B \otimes \overrightarrow{E}_i$  determines the step size in the Brownian movement of the predator, whereas the prey modifies its position in relation with the predator's movement. The third phase starts with a low velocity ratio or when a predator has a faster pace than the prey. It is the last phase of optimization. High exploitation capability is demonstrated in this phase. In such a

low-velocity ratio of  $velocity = 0.1$ , the predator adopts the Lévy strategy. The mathematical model is provided in Equation 22:

$$\begin{aligned} \text{While } Iteration > \frac{2}{3} \text{ Max\_Iteration} \\ \overrightarrow{stepsize}_i &= \overrightarrow{R}_L \otimes (\overrightarrow{R}_L \otimes \overrightarrow{E}_i - \overrightarrow{P}_i) \quad i = 1, \dots, n \\ \overrightarrow{P}_i &= \overrightarrow{E}_i + T \cdot \overrightarrow{CF} \otimes \overrightarrow{stepsize}_i \end{aligned} \quad (22)$$

In the Lévy method, multiplying  $\overrightarrow{R}_L \otimes \overrightarrow{E}_i$  simulates the predator's movement, whereas adding the step size to the Elite position assists in updating the position of the prey. Fish aggregating devices (FADs), considered local optima in their search space, are where sharks spend most of their time (i.e., more than 80% of the time). They make longer jumps in diverse directions during the remaining 20% of their time, probably to locate different prey distributions. To ensure a more dynamic search during the simulation, these lengthier hops help prevent them from being stuck in local optima. The FAD effect's mathematical elaboration can be represented as the following Equation 23:

$$\overrightarrow{P}_i = \begin{cases} \overrightarrow{P}_i + CF[\overrightarrow{Y}_{min} + \overrightarrow{R} \otimes (\overrightarrow{Y}_{max} - \overrightarrow{Y}_{min})] \otimes \overrightarrow{U} & \text{if } r \leq FADs \\ \overrightarrow{P}_i + [FADs (1-r) + r] (\overrightarrow{P}_{r1} - \overrightarrow{P}_{r2}) & \text{if } r \geq FADs \end{cases} \quad (23)$$

The likelihood that FADs may affect the optimization process is represented by the probability, given as 0.2. A binary vector  $\overrightarrow{U}$  is made up of zeros and ones. It is created by a random vector with values between [0,1], with a zero set for values below 0.2 and one for values above 0.2. Additionally,  $r$  stands for a random number uniformly distributed between [0,1].  $\overrightarrow{Y}_{min}$  and  $\overrightarrow{Y}_{max}$  denote lower and upper bounds of dimensions.  $P$  matrix's random indexes are denoted by subscripts  $r_1$  and  $r_2$ .

#### Novelty in this method

The problem of the MPA algorithm is finding an optimal global position; therefore, we added a concept of Quantum Theory that improves populations' motion behavior. The initial population in the modified version is defined as follows:

$$Z_i(k+1) = Z_{min} + r \times [Z_{max} - Z_{min}]$$

where  $Z_i$  denotes the  $i$ th iteration value,  $r$  is a random value between (0,1),  $k$  is a current iteration, and  $Z_{max}$  and  $Z_{min}$  denote the upper and lower limits, respectively. The fitness value is computed to find the best solution in the next step. The following Equations 24–27 is utilized for this purpose:

$$\overrightarrow{L}_Q = Z_{ij}(k+1) = \begin{cases} C_i - b \star (J_{best} - Z_{ij}(k)) \times \ln(\frac{1}{u}), & \text{if } T \geq Entropy \\ C_i + b \star (J_{best} - Z_{ij}(k)) \times \ln(\frac{1}{u}), & \text{if } T < Entropy \end{cases} \quad (24)$$

$$Q(Z_{i,k+1}^j) = \frac{1}{Length_{i,k}^j} \exp\left(-\frac{2|Z_{i,k+1}^j - s_{i,k}^j|}{Length_{i,k}^j}\right) \quad (25)$$

$$C_i = \theta \times cBest_i + (1 - \theta) \times gBest \quad (26)$$

$$J_{best} = \frac{1}{N} \sum_{i=1}^N cBest_i \quad (27)$$

The notation  $cBest_i$  denotes the best position in the  $i$ th iteration for the predator, and  $gBest$  is the best position for all predators at each iteration. The average best predator is denoted by  $J_{best}$ , and  $\theta$  is the distribution of a chaotic number on (0,1). The  $b$  denotes the contraction expression phase, and it is used to control the convergence rate. Mathematically,  $b$  is defined by Equation 28:

$$b = b_{max} - \left[ \left\{ \frac{b_{max} - b_{min}}{Iterations_{max}} \right\} \times Iterations \right] \quad (28)$$

Hence, the final equation is formulated as the following Equation 29:

$$\vec{Step} = \vec{C}_Q \otimes (\vec{Elite}_i - \vec{C}_Q \otimes \vec{Prey}_i), \quad i = 1, 2, \dots, \frac{n}{2} \quad (29)$$

Every solution in the current iteration is compared with its equivalent in the prior iteration for fitness. If the current solution is found to be a superior match, the previous one is superseded. This iterative procedure improves solution quality with time and imitates the behavior of predators that return to locations with abundant prey after successful foraging attempts. After completion of optimization, a feature vector of dimension  $N \times 380$  and a feature vector of  $N \times 313$ , respectively, are obtained.

### 3.7 Feature fusion and classification

The selected features are finally fused and later classified using machine learning classifiers. The fusion process improves an object's information that directly relates to better accuracy. In this

work, a simple serial-based fusion has been chosen to combine the selected feature vectors in a single vector.

Using the following equation, we can determine the dimension of the serial-based fusion vector if we have two feature vectors,  $f1$  and  $f2$ , with dimensions of  $N \times 380$  and  $N \times 313$ , respectively, where  $N$  denotes the total number of observations as defined by Equation 30.

$$R = \begin{pmatrix} f1 \\ f2 \end{pmatrix}_{N \times 380 + N \times 313} \quad (30)$$

The resultant feature vector is obtained of dimension  $N \times 693$ . The fused feature vector is finally classified using traditional machine learning classifiers named as Cubic SVM and Weighted KNN and neural network-based classifiers such as narrow, wide, tri-layered, bi-layered, and medium. The hyperparameters used to train these classifiers are provided in Table 1 as follows:

### 3.8 Dataset and performance evaluation

The augmented Figshare dataset is used for our experiments and is contributed by (11). The dataset is publicly available for research purposes. A model or algorithm's ability to predict outcomes based on the available data is measured using performance metrics in machine learning. The calculated measures contain each classifier's sensitivity rate, false negative rate (FNR), precision rate, and area under the curve (AUC). Time and accuracy measures are also used to interpret the performance of each classifier. Table 2 provides more details on these performance metrics.

TABLE 1 Classifiers and training hyperparameters of each classifier.

| Classifier                      | Training hyperparameters                                                                                                                                                                          | Classifier                       | Training hyperparameters                                                                                                                                                                                                  |
|---------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Cubic SVM</b>                | Kernel function: Cubic<br>Kernel scale: Automatic<br>Box constraint level: 1<br>Multiclass method: one-vs-one<br>Standardize data: true                                                           | <b>Weighted KNN</b>              | Number of neighbors: 10<br>Distance metric: Euclidean<br>Distance weight: squared inverse<br>Standardize data: true                                                                                                       |
| <b>Wide neural network</b>      | Number of fully connected layers: 1<br>First layer size: 100<br>Activation: ReLU<br>Iteration limit: 1000<br>Regularization strength (Lambda): 0<br>Standardize data: yes                         | <b>Medium neural network</b>     | Number of fully connected layers: 1<br>First layer size: 25<br>Activation: ReLU<br>Iteration limit: 1000<br>Regularization strength (Lambda): 0<br>Standardize data: yes                                                  |
| <b>Bilayered neural network</b> | Number of fully connected layers: 2<br>First layer size: 10<br>Second layer size: 10<br>Activation: ReLU<br>Iteration limit: 1000<br>Regularization strength (Lambda): 0<br>Standardize data: yes | <b>Trilayered neural network</b> | Number of fully connected layers: 3<br>First layer size: 10<br>Second layer size: 10<br>Third layer size: 10<br>Activation: ReLU<br>Iteration limit: 1000<br>Regularization strength (Lambda): 0<br>Standardize data: yes |
| <b>Narrow neural network</b>    | Number of fully connected layers: 1<br>First layer size: 10<br>Activation: ReLU<br>Iteration limit: 1000<br>Regularization strength (Lambda): 0<br>Standardize data: yes                          |                                  |                                                                                                                                                                                                                           |

TABLE 2 Performance measures used to validate the proposed methodology.

| Name                | Accuracy (%)                                                      | Time                          | Sensitivity rate (%)                                   | False negative rate (%)                           | Precision rate (%)                      | Area under the curve                   |
|---------------------|-------------------------------------------------------------------|-------------------------------|--------------------------------------------------------|---------------------------------------------------|-----------------------------------------|----------------------------------------|
| Performance measure | $\frac{TP + TN}{TP + TN + FP + FN}$                               | Seconds                       | $\frac{TP}{TP + FP}$                                   | $\frac{FN}{TP + FP}$                              | $\frac{TP}{TP + FN}$                    | $\int_a^b f(x)dx$                      |
| Description         | It is a characteristic or condition of being precise or accurate. | Precise time to finish a task | Measures how successfully a test finds true positives. | Probability of failure to detect a true positive. | Degree of false positives in the result | x-axis integral over a particular time |

TP is for true positive, TN is for true negative, FP is for false positive, and FN is for false negative.

The reason to choose each measure provided in Table 1 is given below:

- Accuracy is the ratio of accurately predicted occurrences to total instances. This gives a general idea of how well a model is predicting in every class. Accuracy by itself, though, could not be enough if the classes are unbalanced. In our proposed technique, classes are balanced. Balance among classes is achieved by data augmentation process.
- Time required to finish a specific task is given in seconds.
- Sensitivity quantifies the percentage of actual positive instances that the model accurately predicted. In order to reduce false negatives, it is very crucial. For example, in the medical domain, high sensitivity indicates that the model is effective in identifying positive cases.
- False negative rate refers to the percentage of true positive cases that were mistakenly forecast as negative. It stands for the probability of overlooking favorable examples. When the cost of missing positive occurrences is large, it is essential to reduce false negative rate.
- Precision gauges how well the model predicts positive occurrences. If you wish to reduce false positives, accuracy is crucial. For instance, high precision in medical diagnosis indicates that the model is likely to be accurate when it predicts a positive case.
- The area under the receiver operating characteristic (ROC) curve is known as the area under the curve (AUC). The trade-off between true positive rate (sensitivity) and false positive rate is represented graphically by the ROC curve. AUC offers a single scalar value that sums up the model's overall performance. Perfect categorization is indicated by an AUC of 1.0; random chance is suggested by an AUC of 0.5.

## 4 Results and discussion

### 4.1 Experimental setup

In this section, detailed experimental setup is discussed. The data augmentation is performed using a sparse auto-encoder. A single

hidden layer with 300 neurons is selected while the training parameters like a maximum epochs are 2,000, L2Weight Regularization is set equal to 0.001, SparsityRegularization is equal to 4, and finally, SparsityProportion is set to 0.15. Hyperparameter optimization is performed to optimize the parameters of fine-tuned deep models such as EfficientNetB0 and InceptionResNetV2. The original dataset is split into a ratio of 50:50 in training and test proportions. After that, the training and testing images are separately augmented and trained models. The gradient vectors are accelerated via stochastic gradient descent (SGDM) for quicker convergence at the convolutional layers. During the algorithm learning phase for both models, a mini-batch size 128 is chosen. Additionally, the experiments are carried out using MATLAB R2023a on a machine equipped with 128 GB of RAM and CPU Intel(R) Core(TM) i7-6700 @ 3.40 GHz and 12 GB RTX3000.

### 4.2 Proposed framework results (fine-tuned models)

In this section, results of the first step of the proposed framework are presented. The hyperparameter optimization using the Bayesian method is performed separately for EfficientNetB0 and InceptionResNetV2 models and numerical results are computed.

#### 4.2.1 Fine-tuned Bayesian optimization-based EfficientNetB0

Table 3 describes the classification performance of fine-tuned EfficientNetB0 deep architecture with an accuracy value of 99.10%, achieved by the Cubic SVM classifier. The wide neural network obtained the second best accuracy of 98.90%. The rest of the classifiers obtained accuracies of 98.80%, 98.70%, 98.60%, and 98.50%. The sensitivity and precision rates of each classifier are also noted, and the maximum value of Cubic SVM is 99.10%. In addition, the performance of Cubic SVM can be confirmed by a confusion matrix, given in Figure 6A. The diagonal numbers in the figure represent the number of true observations and the true positive rate for glioma, meningioma, and pituitary classes. The computational time of each classifier is also noted during the classification process, and it is observed that the minimum noted time is 12.512 (seconds) for Medium Neural Network.

#### 4.2.2 Fine-tuned Bayesian optimization-based InceptionResNetV2

In the second step, the classification results are computed using fine-tuned InceptionResNetV2 with the initialization of BO-based



TABLE 3 Classification results using the BO-based EfficientNetB0 model.

| Classifier                | Accuracy (%) | Time (s)      | Sensitivity rate (%) | False negative rate (%) | Precision rate (%) | Area under curve (%) |
|---------------------------|--------------|---------------|----------------------|-------------------------|--------------------|----------------------|
| Cubic SVM                 | <b>99.10</b> | 20.227        | <b>99.10</b>         | <b>0.90</b>             | <b>99.10</b>       | <b>1.00</b>          |
| Wide neural network       | 98.90        | 14.505        | 98.93                | 1.07                    | 98.93              | 1.00                 |
| Medium neural network     | 98.80        | <b>12.512</b> | 98.76                | 1.24                    | 98.76              | 1.00                 |
| Bilayered neural network  | 98.70        | 15.886        | 98.66                | 1.34                    | 98.66              | 0.99                 |
| Weighted KNN              | 98.60        | 26.907        | 98.63                | 1.37                    | 98.63              | 1.00                 |
| Narrow neural network     | 98.60        | 15.377        | 98.60                | 1.40                    | 98.60              | 0.99                 |
| Trilayered neural network | 98.50        | 16.059        | 98.53                | 1.47                    | 98.53              | 1.00                 |

Bold denotes the best values.

hyperparameters. The results of this step are given in Table 4, which shows the maximum accuracy of 98.10 for the Cubic SVM classifier. The minimum computational time of this step is 20.543 (second) for the Narrow Neural Network classifier. In addition, the performance of the Cubic SVM classifier can be confirmed by a confusion matrix, illustrated in Figure 6B. Compared with the performance of this step with step 1, it is observed that the accuracy of this step is degraded by approximately 1%. Moreover, the increase in time shows the drawbacks of this step. In order to reduce the drawbacks of this step, a feature selection method is employed, which selects only important features for classification.

4.3 Feature selection using proposed QTbMPA feature selection

The third and fourth steps correspond to the best feature selection.

4.3.1 QTbMPA feature selection on fine-tuned EfficientNetB0

In the third step, the proposed feature selection method is applied to deep extracted features; in return, the best optimal

features are obtained. The results of the feature selection method on fine-tuned EfficientNetB0 are presented in Table 5. In this table, the maximum accuracy of 99.00% by the Cubic SVM classifier is shown. The sensitivity and precision rate of this classifier are also 99% that the confusion matrix in Figure 7A can confirm. Wide neural network obtained the second best accuracy of 98.80%. Each classifier’s computational time is noted, and its minimum reported time is 3.8078 (sec). In step 1, the minimum time was 12.52 (s), which is now reduced by almost 300%. Moreover, the accuracy of this step is consistent, which can be a strength of the proposed feature selection method.

4.3.2 QTbMPA feature selection on fine-tuned InceptionResNetV2

In the fourth step, features of the fine-tuned InceptionResNetV2 model are selected using the proposed QTbMPA method and classification is performed. Table 6 describes the results of this step, showing an maximum accuracy of 97.70% by Cubic SVM. The sensitivity and precision rate of this classifier is also 97.70%. The confusion matrix in Figure 7B can further confirm these values. The computational time of each classifier is also given in this table, and the minimum reported

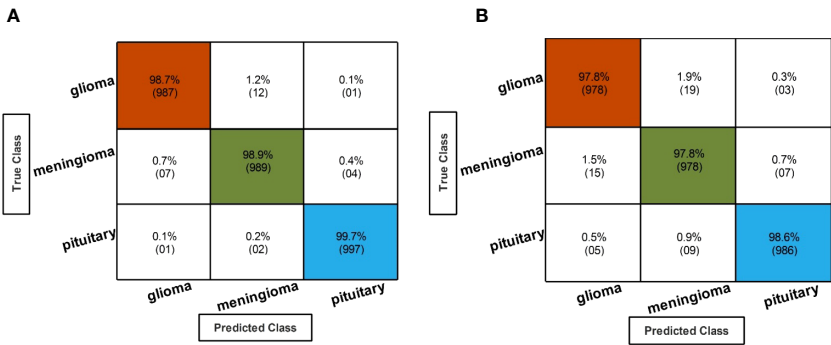


FIGURE 6 Confusion matrix of EfficientNetB0 and InceptionResNetV2 hyperparameter optimization using BO. (A) Confusion matrix of fine-tuned EfficientNetB0 hyperparameter optimization using BO. (B) Confusion matrix of fine-tuned InceptionResNetV2 hyperparameter optimization using BO.

TABLE 4 Classification results of using BO-based InceptionResNetV2.

| Classifier                | Accuracy (%) | Time (s)      | Sensitivity rate (%) | False negative rate (%) | Precision rate (%) | Area under curve (%) |
|---------------------------|--------------|---------------|----------------------|-------------------------|--------------------|----------------------|
| Cubic SVM                 | <b>98.10</b> | 21.966        | <b>98.06</b>         | <b>1.94</b>             | <b>98.06</b>       | <b>1.00</b>          |
| Narrow neural network     | 97.90        | <b>20.543</b> | 97.93                | 2.07                    | 97.93              | 0.99                 |
| Wide neural network       | 97.90        | 49.824        | 97.93                | 2.07                    | 97.93              | 1.00                 |
| Bilayered neural network  | 97.90        | 36.736        | 97.93                | 2.07                    | 97.93              | 0.99                 |
| Trilayered neural network | 97.90        | 36.711        | 97.86                | 2.14                    | 97.86              | 0.99                 |
| Medium neural network     | 97.80        | 29.453        | 97.80                | 2.20                    | 97.80              | 1.00                 |
| Weighted KNN              | 97.10        | 34.365        | 97.13                | 2.87                    | 97.13              | 1.00                 |

Bold denotes the best values.

time is 6.1486 (s) for Cubic SVM. Compared with the computational time of this step with the second step, the time is reduced almost 300%.

4.3.3 Fusion of best selected features

Finally, the best-selected features of both models, in the third and fourth steps, are fused using a serial approach. The cubic SVM classifier obtained the maximum accuracy of 99.80% and the sensitivity and precision rates of 99.83% (*can be seen in Table 7*). The confusion matrix in *Figure 8* can further confirm these values. A minor increase in computational time is observed after the fusion process; however, the accuracy is significantly improved for all classifiers. In comparison, with all previous steps, noted accuracy has significantly improved and is the highest among all early noted accuracies. Moreover, *Table 8* shows a detailed comparison of the proposed method with state-of-the-art techniques and shows significant improvement.

4.4 Discussion

A t-test is a statistical technique used to compare the mean values of two groups. It is frequently used in hypothesis testing to see whether a particular process or treatment has a noticeable effect on the target population or whether there is a significant difference between the two groups. In order to test the significant difference between the classifiers, t-test is applied.

In the proposed technique, t-test is conducted to check any considerable gap between accuracies at different stages of our proposed model. The gap is resulted when we have unbalanced classes of dataset (54). The augmented step balanced the classes of dataset; however, to validate our augmentation step, t-test is applied on all phases of the proposed technique. The test starts by setting a null hypothesis as below:

$$H_0 = \text{The accuracy of the chosen classifiers differs significantly over the phases of proposed technique.}$$

TABLE 5 Proposed classification results after employing the QTbMPA selection method on features returned from the Bayesian-based EfficientNetB0 model.

| Classifier                | Accuracy (%) | Time (s)      | Sensitivity rate (%) | False negative rate (%) | Precision rate (%) | Area under curve (%) |
|---------------------------|--------------|---------------|----------------------|-------------------------|--------------------|----------------------|
| Cubic SVM                 | <b>99.00</b> | 5.4668        | <b>99.00</b>         | <b>1.00</b>             | <b>99.00</b>       | <b>1.00</b>          |
| Wide neural network       | 98.80        | 4.5414        | 98.83                | 1.17                    | 98.83              | 1.00                 |
| Medium neural network     | 98.60        | 3.8831        | 98.63                | 1.37                    | 98.63              | 1.00                 |
| Weighted KNN              | 98.60        | 7.4494        | 98.56                | 1.44                    | 98.56              | 1.00                 |
| Trilayered neural network | 98.50        | 4.9574        | 98.53                | 1.47                    | 98.53              | 0.99                 |
| Narrow neural network     | 98.50        | 3.9573        | 98.50                | 1.5                     | 98.50              | 1.00                 |
| Bilayered neural network  | 98.50        | <b>3.8078</b> | 98.50                | 1.5                     | 98.50              | 0.99                 |

Bold denotes the best values.

Additionally, two best-performing classifiers at all phases of the proposed technique are chosen. The accuracy achieved by respective classifier at each phase is selected to conduct experiments.

A detailed overview of test is given below:

where  $I$  is the number of experiments and the noted mean value after this step is 0.48. After calculating the *Mean* ( $\mu$ ), the *standard deviation* ( $\sigma$ ) is calculated by using the following equation:

| Phases →<br>Classifiers ↓ | BO based<br>EfficientNetB0 | BO based<br>InceptionResNetV2 | MPA Optimization<br>for EfficientNetB0 | MPA Optimization<br>for InceptionResNetV2 | Feature<br>fusion |
|---------------------------|----------------------------|-------------------------------|----------------------------------------|-------------------------------------------|-------------------|
| Cubic SVM                 | 99.10                      | 98.10                         | 99.00                                  | 99.70                                     | 99.80             |
| Weighted KNN              | 98.60                      | 97.10                         | 98.60                                  | 97.20                                     | 99.80             |

The mean of the differences for all experiments are calculated using the following Equations 31–34:

$Difference\ (J) = |Accuracy(k) - Accuracy(l)|$  (31)

$standard\ deviation\ (\sigma) = \sqrt{\frac{(\sum_{k=1}^I (J_k) - \mu)^2}{I - 1}}$  (33)

$Mean\ (\mu) = \frac{1}{N} \sum_{k=1}^I |J_k|$  (32)

The resultant standard deviation value is 0.357, later used in the  $T$  Selection formula.

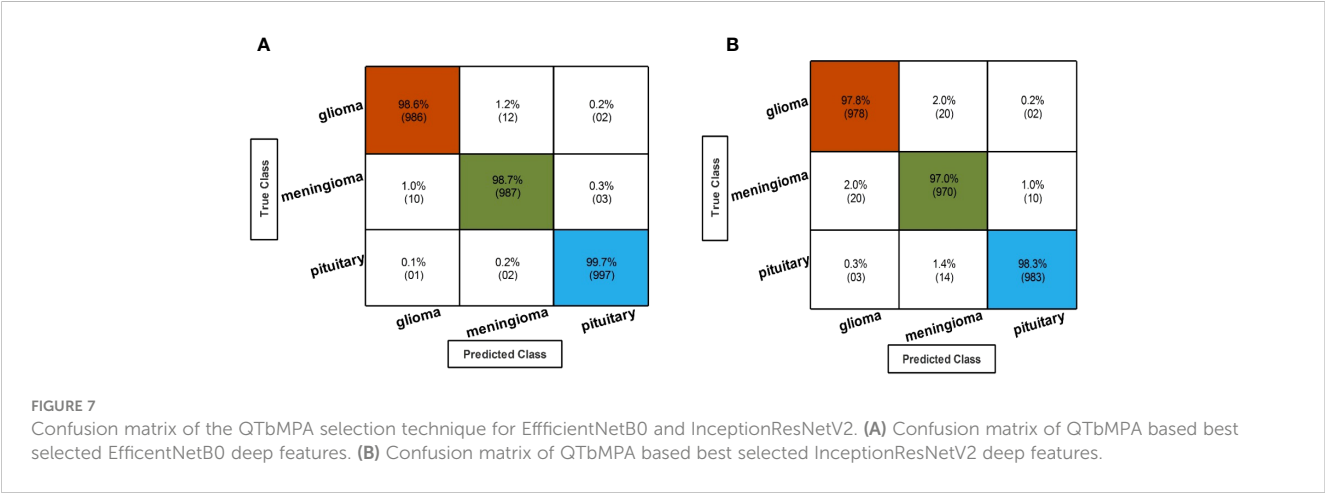


TABLE 6 Proposed classification results after employing the QTbMPA selection method on features returned from the Bayesian-based InceptionResNetV2 model.

| Classifier                | Accuracy (%) | Time (s) | Sensitivity rate (%) | False negative rate (%) | Precision rate (%) | Area under curve (%) |
|---------------------------|--------------|----------|----------------------|-------------------------|--------------------|----------------------|
| Cubic SVM                 | 97.70        | 6.1486   | 97.70                | 2.30                    | 97.70              | 1.00                 |
| Narrow neural network     | 97.60        | 7.8251   | 97.60                | 2.40                    | 97.60              | 0.99                 |
| Medium neural network     | 97.60        | 7.3787   | 97.60                | 2.40                    | 97.60              | 0.99                 |
| Bilayered neural network  | 97.50        | 7.7592   | 97.50                | 2.50                    | 97.5               | 0.98                 |
| Wide neural network       | 97.30        | 11.376   | 97.20                | 2.80                    | 97.20              | 0.99                 |
| Trilayered neural network | 97.30        | 8.9962   | 97.26                | 2.74                    | 97.26              | 0.99                 |
| Weighted KNN              | 97.20        | 6.8459   | 97.16                | 2.84                    | 97.16              | 0.99                 |

Bold denotes the best values.

TABLE 7 Classification results after fusing of the best selected features of both models.

| Classifier                | Accuracy (%) | Time (s)      | Sensitivity rate (%) | False negative rate (%) | Precision rate (%) | Area under curve (%) |
|---------------------------|--------------|---------------|----------------------|-------------------------|--------------------|----------------------|
| Cubic SVM                 | <b>99.80</b> | 11.198        | <b>99.83</b>         | <b>0.17</b>             | <b>99.83</b>       | <b>1.00</b>          |
| Weighted KNN              | 99.80        | 13.699        | 99.80                | 0.20                    | 99.80              | 1.00                 |
| Wide neural network       | 99.80        | 6.776         | 99.76                | 0.24                    | 99.76              | 1.00                 |
| Medium neural network     | 99.70        | 5.6234        | 99.70                | 0.30                    | 99.70              | 1.00                 |
| Bilayered neural network  | 99.70        | <b>5.0002</b> | 99.66                | 0.34                    | 99.66              | 1.00                 |
| Trilayered neural network | 99.70        | 5.6399        | 99.66                | 0.34                    | 99.66              | 1.00                 |
| Narrow neural network     | 99.60        | 6.2125        | 99.60                | 0.40                    | 99.60              | 1.00                 |

Bold denotes the best values.

$$T_{\text{ Selection}} = \frac{\sqrt{I} \times \mu}{\sigma} \tag{34}$$

The value 3.012 was obtained after calculation using the above formula. The obtained value will be considered as a decisive point to conduct the *Student's T – Test*. Moreover, the *degree of freedom (df)* is calculated using the formula:  $df = n - 1$ ; the resultant value is four and selected *p value* = 0.05 (55). After looking at the corresponding output value in the t-test chart, the value is (– 2.776, +2.776). The decisive *T Selection* value is 3.012; based on the given below formulation in [equation 35](#), it is established that  $H_0$  is rejected, and there is no noteworthy difference between the accuracy of the selected classifiers.

$$\text{If } (T_{\text{ Selection}} \geq -2.776 \text{ and } \leq + 2.776) \tag{35}$$

Hypothesis test establishes that throughout the phases of the proposed technique, there is a consistency in accuracy of each phase; it means that the class imbalance problem is accurately addressed. Inconsistent accuracies are the result of imbalance

classes of dataset, which lead to loss of accuracies. The proposed data augmentation step helps to properly address class imbalance problem.

Heat map-based analysis: Heat map-based techniques are employed to express the decisive features of classification for each class. Grad-CAM, LIME, and Occlusion Sensitivity are three methods commonly used to represent decision features for classification of an image. Grad-CAM uses gradients to determine the classification score about the final convolutional feature map. It draws attention to that part of input image that has the biggest influence on this score. The method uses a global average pooling layer to extract features. [Equation 36](#) serves as the basis for this procedure, which is illustrated below:

$$b_g^c = \frac{1}{N} \sum_i \sum_j \frac{\partial y}{\partial B_{ij}^g} \tag{36}$$

where  $b_g^c$  represents class scores of  $g$  features from the Global Average Pooling layer,  $N$  represents total pixels in a feature map,  $c$

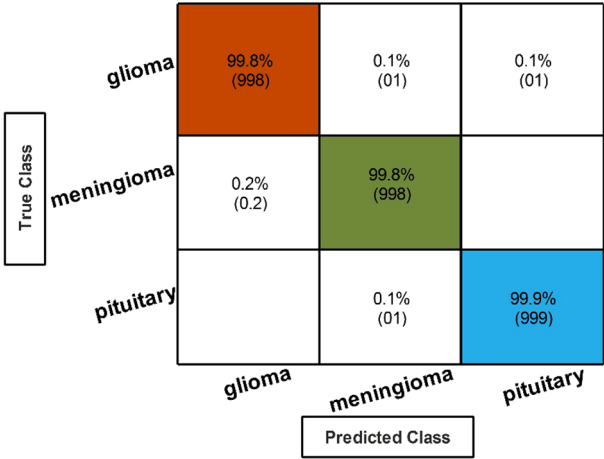


FIGURE 8  
Classification results after the fusion of selected feature features.

TABLE 8 Summary of recent state-of-the-art (SOTA) techniques for brain tumor classification using Figshare dataset.

| Serial no. | Reference                | Year | Dataset  | Accuracy (%) |
|------------|--------------------------|------|----------|--------------|
| 1          | Alanazi et al. (19)      | 2022 | Figshare | 95.75        |
| 2          | Raza et al. (20)         | 2022 | Figshare | 99.67        |
| 3          | Tummala et al. (21)      | 2022 | Figshare | 98.70        |
| 4          | Polat et al. (22)        | 2022 | Figshare | 99.18        |
| 5          | Shaik et al. (23)        | 2022 | Figshare | 96.51        |
| 6          | Haq et al. (24)          | 2022 | Figshare | 98.00        |
| 7          | Rahman et al. (25)       | 2023 | Figshare | 97.60        |
| 8          | Talukder et al. (47)     | 2023 | Figshare | 99.68        |
| 9          | Aloraini et al. (26)     | 2023 | Figshare | 99.10        |
| 10         | Athisayamani et al. (27) | 2023 | Figshare | 98.85        |
| 11         | Mishra et al. (28)       | 2023 | Figshare | 98.97        |
| 12         | Agrawal et al. (48)      | 2023 | Figshare | 96.40        |
| 13         | Malla et al. (49)        | 2023 | Figshare | 98.93        |
| 14         | Asif et al. (50)         | 2023 | Figshare | 98.69        |
| 15         | Cinar et al. (29)        | 2023 | Figshare | 98.32        |
| 16         | Deepak et al. (30)       | 2023 | Figshare | 95.40        |
| 17         | Zulfiqar et al. (31)     | 2023 | Figshare | 98.86        |
| 18         | Shyamala et al. (51)     | 2023 | Figshare | 94.70        |
| 19         | Yapici et al. (52)       | 2023 | Figshare | 99.47        |
| 20         | Sahoo et al. (53).       | 2023 | Figshare | 97.00        |
|            | <b>Proposed</b>          |      | Figshare | <b>99.80</b> |

Bold denotes the best values.

depicts the class score, and  $y$  is the considered output. The whole expression  $\partial B_{ij}^g$  represents the convolution map. In the expression,  $i$  and  $j$  represent two dimensions and  $B$  represents gradients. Features with negative weight can be possible using the above equation; therefore, the Relu activation function is used to convert the negative weights to positive and is represented using the given below Equation 37:

$$M = \text{Relu}(\sum b_g^c \cdot B^g) \tag{37}$$

Mathematical details of LIME and Occlusion Sensitivity can be seen from (56) and (57), respectively. Figure 9 represents the visualization of important features of each class using the explained methods of the heat map.

## 5 Conclusion

This article presents a novel deep learning framework with an efficient QTbMPA feature selection technique for the classification of brain tumor types such as meningioma, glioma, and pituitary from MRI images. Instead of manual data augmentation, a sparse autoencoder architecture was

proposed and generated new images based on the training set. Two lightweight deep learning architectures were modified and trained with the help of BO hyperparameter initialization. The deeper layer (global average pool) was employed for feature extraction and performed classification. The classification process shows that there exist few irrelevant features, which are impacted on the classification computational time. Therefore, we proposed an efficient QTbMPA feature selection algorithm that almost 300% reduced the computational time and maintained the classification accuracy. The selected features were finally fused and classified using ML and neural network classifiers. On the augmented dataset, the proposed framework obtained an improved accuracy of 99.80% than the SOTA technique.

The goal of the proposed research is to create a deep learning (DL) model for brain tumor classification, utilizing DL's capabilities to classify various forms of brain tumors more accurately. This finding could have a significant clinical impact in neuro-oncology and have a wide variety of potential applications. The proposed research can assist doctors and radiologists in making accurate diagnoses when using medical imaging data, such as MRI scans, to identify brain tumors. It offers dependable and consistent tumor categorization results,



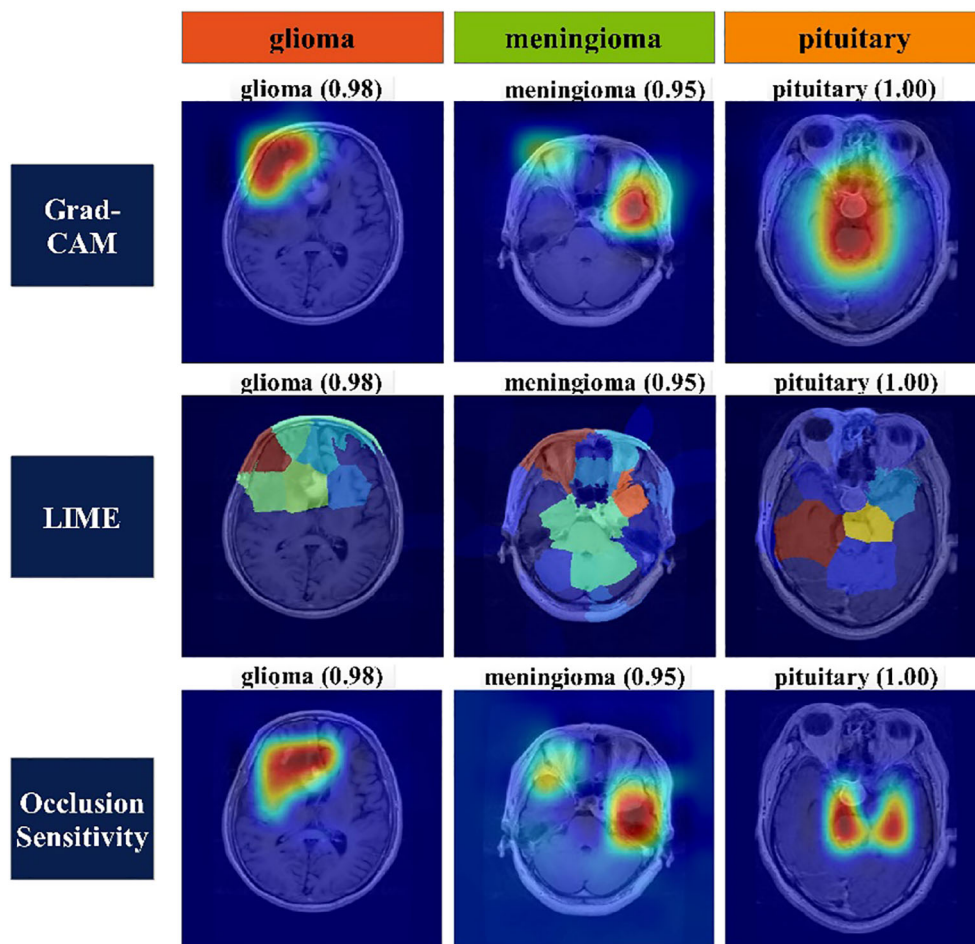


FIGURE 9  
Heat map of classes using Grad-CAM, LIME, and occlusion sensitivity.

lowering the misdiagnosis risk and enabling early brain tumor discovery. Furthermore, the accurate classification of brain tumors might help in developing customized treatment plans for patients. The model assists physicians in developing customized treatment regimens that lead to more accurate and successful treatment outcomes by aiding in the identification of the exact type of tumor.

## 5.1 Limitations and future work

Although we obtained the maximum accuracy, there are few limitations that make the proposed architecture more consistent. The limitations of this work are selection of pretrained models and best feature selection. The pretrained models have been selected based on the Top-5 accuracy on ImageNet dataset and total number of parameters. In addition, the selection process reduces the overfitting, but still there are few irrelevant features selected for

the classification. The proposed architecture has been evaluated on brain tumor MRIs of the Figshare dataset; however, in future, it will be tested on BRATS datasets. Moreover, a new self-attention and vision transformer model will be proposed for the improved accuracy and efficiency.

## Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset>.

## Author contributions

MU: Data curation, Methodology, Conceptualization, Investigation, Project administration, Software, Writing –

original draft. MK: Conceptualization, Investigation, Methodology, Software, Supervision, Writing – original draft. AM: Conceptualization, Formal analysis, Methodology, Software, Writing – original draft. OM: Data curation, Formal analysis, Project administration, Visualization, Writing – review & editing. OS: Conceptualization, Data curation, Formal analysis, Funding acquisition, Project administration, Visualization, Writing – review & editing. NA: Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R333), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. This study is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2023/R/1445).

## References

- Ohgaki H. Epidemiology of brain tumors. In: Verma M, editor. *Cancer Epidemiology: Modifiable Factors*. Totowa, NJ: Humana Press (2009). p. 323–42.
- Kibriya H, Amin R, Alshehri AH, Masood M, Alshamrani SS, Alshehri A. A novel and effective brain tumor classification model using deep feature fusion and famous machine learning classifiers. *Comput Intell Neurosci* (2022) 2022. doi: 10.1155/2022/7897669
- Bush NAO, Chang SM, Berger MS. Current and future strategies for treatment of glioma. *Neurosurgical Rev* (2017) 40(1):1–14. doi: 10.1007/s10143-016-0709-8
- Apra C, Peyre M, Kalamarides M. Current treatment options for meningioma. *Expert Rev Neurother* (2018) 18(3):241–9. doi: 10.1080/14737175.2018.1429920
- Daly AF, Beckers A. The epidemiology of pituitary adenomas. *Endocrinol Metab Clinics* (2020) 49(3):347–55. doi: 10.1016/j.ecl.2020.04.002
- Baba AI, Cătoi C. *Comparative oncology*. Netherland: Publishing House of the Romanian Academy Bucharest (2007).
- Carrano A, Juarez JJ, Incontri D, Ibarra A, Cazares HG. Sex-specific differences in glioblastoma. *Cells* (2021) 10(7):1783. doi: 10.3390/cells10071783
- Yildirim M, Cengil E, Eroglu Y, Cinar A. Detection and classification of glioma, meningioma, pituitary tumor, and normal in brain magnetic resonance imaging using deep learning-based hybrid model. *Iran J Comput Sci* (2023) 6(4):455–64. doi: 10.1007/s42044-023-00139-8
- Naseer A, Yasir T, Azhar A, Shakeel T, Zafar K. Computer-aided brain tumor diagnosis: performance evaluation of deep learner CNN using augmented brain MRI. *Int J BioMed Imaging* (2021) 2021:5513500. doi: 10.1155/2021/5513500
- Gull S, Akbar S. Artificial intelligence in brain tumor detection through MRI scans: advancements and challenges. *Artif Intell Internet Things* (2021) 10:241–76.
- Cheng J, Huang W, Cao S, Yang R, Yang W, Yun Z, et al. Enhanced performance of brain tumor classification via tumor region augmentation and partition. *PloS One* (2015) 10(10):e0140381. doi: 10.1371/journal.pone.0140381
- Cheng J, Yang W, Huang M, Huang W, Jiang J, Zhou Y, et al. Retrieval of brain tumors by adaptive spatial pooling and fisher vector representation. *PloS One* (2016) 11(6):e0157112. doi: 10.1371/journal.pone.0157112
- Xu S, Wang J, Shou W, Ngo T, Sadick A-M, Wang X. Computer vision techniques in construction: A critical review. *Arch Comput Methods Eng* (2021) 28(5):3383–97. doi: 10.1007/s11831-020-09504-3
- Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E. Deep learning for computer vision: A brief review. *Comput Intell Neurosci* (2018) 2018. doi: 10.1155/2018/7068349
- Balamurugan D, Aravindh S, Reddy PCS, Rupani A, Manikandan A. Multiview objects recognition using deep learning-based wrap-CNN with voting scheme. *Neural Process Lett* (2022) 54(3):1495–521. doi: 10.1007/s11063-021-10679-4

## Acknowledgments

This research was funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R333), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia. This study is supported via funding from Prince Sattam bin Abdulaziz University project number (PSAU/2023/R/1445).

## 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.

- Bjerge K, Mann HM, Høye TT. Real-time insect tracking and monitoring with computer vision and deep learning. *Remote Sens Ecol Conserv* (2022) 8(3):315–27. doi: 10.1002/rse2.245
- Kim HE, Cosa-Linan A, Santhanam N, Jannesari M, Maros ME, Ganslandt T. Transfer learning for medical image classification: a literature review. *BMC Med Imaging* (2022) 22(1):69. doi: 10.1186/s12880-022-00793-7
- Zhang J, Rao VM, Tian Y, Yang Y, Acosta N, Wan Z, et al. Detecting schizophrenia with 3d structural brain mri using deep learning. *Sci Rep* (2023) 13(1):14433. doi: 10.1038/s41598-023-41359-z
- Alanazi MF, Ali MU, Hussain SJ, Zafar A, Mohatram M, Irfan M, et al. Brain tumor/mass classification framework using magnetic-resonance-imaging-based isolated and developed transfer deep-learning model. *Sensors* (2022) 22(1):372. doi: 10.3390/s22010372
- Raza A, Ayub H, Khan JA, Ahmad I, Salama AS, Daradkeh YI, et al. A hybrid deep learning-based approach for brain tumor classification. *Electronics* (2022) 11(7):1146. doi: 10.3390/electronics11071146
- Tummala S, Kadry S, Bukhari SAC, Rauf HT. Classification of brain tumor from magnetic resonance imaging using vision transformers ensembling. *Curr Oncol* (2022) 29(10):7498–511. doi: 10.3390/currenol29100590
- Polat Ö, Dokur Z, Ölmez T. Brain tumor classification by using a novel convolutional neural network structure. *Int J Imaging Syst Technol* (2022) 32(5):1646–60. doi: 10.1002/ima.22763
- Shaik NS, Cherukuri TK. Multi-level attention network: application to brain tumor classification. *Signal Image Video Process* (2022) 16(3):817–24. doi: 10.1007/s11760-021-02022-0
- Haq EU, Jianjun H, Huarong X, Li K, Weng L. A hybrid approach based on deep cnn and machine learning classifiers for the tumor segmentation and classification in brain MRI. *Comput Math Methods Med* (2022) 2022. doi: 10.1155/2022/6446680
- Rahman T, Islam MS. MRI brain tumor detection and classification using parallel deep convolutional neural networks. *Measurement: Sensors* (2023) 26:100694.
- Aloraini M, Khan A, Aladhadh S, Habib A, Alsharekh MF, Islam M. Combining the transformer and convolution for effective brain tumor classification using MRI images. *Appl Sci* (2023) 13(6):3680. doi: 10.3390/app13063680
- Athisayamani S, Antonyswamy RS, Sarveshwaran V, Almeshari M, Alzamil Y, Ravi V. Feature extraction using a residual deep convolutional neural network (ResNet-152) and optimized feature dimension reduction for MRI brain tumor classification. *Diagnostics* (2023) 13(4):668. doi: 10.3390/diagnostics13040668
- Mishra HK, Kaur M. Classification of brain tumour based on texture and deep features of magnetic resonance images. *Expert Syst* (2023) 21:e13294. doi: 10.1111/exsy.13294

29. Cinar N, Kaya M, Kaya B. A novel convolutional neural network-based approach for brain tumor classification using magnetic resonance images. *Int J Imaging Syst Technol* (2023) 33(3):895–908. doi: 10.1002/ima.22839
30. Deepak S, Ameer PM. Brain tumor categorization from imbalanced MRI dataset using weighted loss and deep feature fusion. *Neurocomputing* (2023) 520:94–102. doi: 10.1016/j.neucom.2022.11.039
31. Zulfiqar F, Ijaz Bajwa U, Mehmood Y. Multi-class classification of brain tumor types from MR images using EfficientNets. *Biomed Signal Process Control* (2023) 84:104777. doi: 10.1016/j.bspc.2023.104777
32. Cheng J. Brain magnetic resonance imaging tumor dataset. *Figshare MRI dataset version* (2017) 5.
33. Al-Ameen Z. (2020). Contrast enhancement of medical images using statistical methods with image processing concepts, in: *2020 6th International Engineering Conference "Sustainable Technology and Development" (IEC)*, IEEE.
34. Vincent P, Larochelle H, Bengio Y, Manzagol P-A. *Extracting and Composing Robust Features with Denoising Autoencoders*. Springer (2021).
35. Szegedy C, Ioffe S, Vanhoucke V, Alemi A. (2017). Inception-v4, inception-resnet and the impact of residual connections on learning, in: *Proceedings of the AAAI conference on artificial intelligence*.
36. Tan M, Le Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks, in: *International conference on machine learning*, PMLR.
37. Hutter F, Hoos HH, Leyton-Brown K. (2011). Sequential model-based optimization for general algorithm configuration, in: *Learning and Intelligent Optimization: 5th International Conference, LION 5, Rome, Italy, January 17–21, 2011, Selected Papers* 5Springer.
38. Faramarzi A, et al. Marine Predators Algorithm: A nature-inspired metaheuristic. *Expert Syst Appl* (2020) 152:113377. doi: 10.1016/j.eswa.2020.113377
39. Ogasawara C, Philbrick BD, Adamson DC. Meningioma: a review of epidemiology, pathology, diagnosis, treatment, and future directions. *Biomedicine* (2021) 9(3):319. doi: 10.3390/biomedicine9030319
40. Hanif F, et al. Glioblastoma multiforme: a review of its epidemiology and pathogenesis through clinical presentation and treatment. *Asian Pacific J Cancer prevention: APJCP* (2017) 18(1):3.
41. Banskota S, Adamson DC. Pituitary adenomas: from diagnosis to therapeutics. *Biomedicine* (2021) 9(5):494. doi: 10.3390/biomedicine9050494
42. Dawood FA, Abood ZM. The importance of contrast enhancement in medical images analysis and diagnosis. *Int J Eng Res Technol (IJERT)* (2018) 7(12). doi: 10.17577/IJERTV7IS120006
43. Florea C, Florea L. Parametric logarithmic type image processing for contrast based auto-focus in extreme lighting conditions. *Int J Appl Mathematics Comput Sci* (2013) 23(3). doi: 10.2478/amcs-2013-0048
44. Buda M, Maki A, Mazurowski MA. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks* (2018) 106:249–59. doi: 10.1016/j.neunet.2018.07.011
45. Ng A. Sparse autoencoder. *CS294A Lecture Notes* (2011) 72(2011):1–19.
46. Wu J, et al. Hyperparameter optimization for machine learning models based on Bayesian optimization. *J Electronic Sci Technol* (2019) 17(1):26–40.
47. Talukder MA, Islam MM, Uddin MA, Akhter A, Pramanik MAJ, Aryal S, et al. An efficient deep learning model to categorize brain tumor using reconstruction and fine-tuning. *Expert Syst Appl* (2023) 120534. doi: 10.1016/j.eswa.2023.120534
48. Agrawal T, Choudhary P, Shankar A, Singh P, Diwakar M. MultiFeNet: Multi-scale feature scaling in deep neural network for the brain tumour classification in MRI images. *Int J Imaging Syst Technol* (2019).
49. Malla PP, Sahu S, Alutaibi AI. Classification of tumor in brain MR images using deep convolutional neural network and global average pooling. *Processes* (2023) 11(3):679. doi: 10.3390/pr11030679
50. Asif S, Zhao M, Chen X, Zhu Y. BMRI-NET: A deep stacked ensemble model for multi-class brain tumor classification from MRI images. *Interdiscip Sciences: Comput Life Sci* (2023) 17:1–16. doi: 10.1007/s12539-023-00571-1
51. Shyamala B, Brahmananda SH Dr. Brain tumor Classification using Optimized and Relief-based Feature Reduction and Regression Neural Network. *Biomed Signal Process Control* (2023) 86:105279. doi: 10.1016/j.bspc.2023.105279
52. Yapici M, Karakiş R, Gürkahraman K. Improving brain tumor classification with deep learning using synthetic data. *Computers Materials Continua* (2023) 74(3). doi: 10.32604/cmc.2023.035584
53. Kumar Sahoo A, Parida P, Muralibabu K, Dash S. Efficient simultaneous segmentation and classification of brain tumors from MRI scans using deep learning. *Biocybernetics Biomed Eng* (2023) 43(3):616–33. doi: 10.1016/j.bbe.2023.08.003
54. Wang L, et al. Review of classification methods on unbalanced data sets. *IEEE Access* (2021) 9:64606–28. doi: 10.1109/ACCESS.2021.3074243
55. Di Leo G, Sardanelli F. Statistical significance: p value, 0.05 threshold, and applications to radiomics—reasons for a conservative approach. *Eur Radiol Exp* (2020) 4(1):1–8.
56. Ribeiro MT, Singh S, Guestrin C. (2016). "Why should I trust you?": explaining the predictions of any classifier, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery: San Francisco, California, USA. pp. 1135–44.
57. Zeiler MD, Fergus R. *Visualizing and Understanding Convolutional Networks*. Cham: Springer International Publishing (2014).



## OPEN ACCESS

## EDITED BY

Sharon R. Pine,  
University of Colorado Anschutz Medical  
Campus, United States

## REVIEWED BY

Juefeng Wan,  
Fudan University, China  
Zhenyu Shu,  
Zhejiang Provincial People's Hospital, China

## \*CORRESPONDENCE

Xiang Gao

✉ gx@xmu.edu.cn

Zheng Kang

✉ txzzyx07@126.com

<sup>†</sup>These authors have contributed equally to this work

RECEIVED 25 September 2023

ACCEPTED 06 February 2024

PUBLISHED 22 February 2024

## CITATION

Yu Y, Gao G, Gao X, Zhang Z, He Y, Shi L and Kang Z (2024) A study on the radiomic correlation between CBCT and pCT scans based on modified 3D-RUNet image segmentation. *Front. Oncol.* 14:1301710. doi: 10.3389/fonc.2024.1301710

## COPYRIGHT

© 2024 Yu, Gao, Gao, Zhang, He, Shi and Kang. 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.

# A study on the radiomic correlation between CBCT and pCT scans based on modified 3D-RUNet image segmentation

Yanjuan Yu<sup>1†</sup>, Guanglu Gao<sup>2†</sup>, Xiang Gao<sup>2\*</sup>, Zongkai Zhang<sup>2</sup>, Yipeng He<sup>2</sup>, Liwan Shi<sup>2</sup> and Zheng Kang<sup>2\*</sup>

<sup>1</sup>College of Electronic Engineering, Zhangzhou Institute of Technology, Zhangzhou, Fujian, China,

<sup>2</sup>Department of Radiation Oncology, the First Affiliated Hospital of Xiamen University, Xiamen, Fujian, China

**Purpose:** The present study is based on evidence indicating a potential correlation between cone-beam CT (CBCT) measurements of tumor size, shape, and the stage of locally advanced rectal cancer. To further investigate this relationship, the study quantitatively assesses the correlation between positioning CT (pCT) and CBCT in the radiomics features of these cancers, and examines their potential for substitution.

**Methods:** In this study, 103 patients diagnosed with locally advanced rectal cancer and undergoing neoadjuvant chemoradiotherapy were selected as participants. Their CBCT and pCT images were used to divide the participants into two groups: a training set and a validation set, with a 7:3 ratio. An improved conventional 3D-RUNet (CLA-UNet) deep learning model was trained on the training set data and then applied to the validation set. The DSC, HD95 and ASSD were calculated for quantitative evaluation purposes. Then, radiomics features were extracted from 30 patients of the test set.

**Results:** The experiments demonstrate that, the modified model achieves an average DSC score 0.792 for pCT and 0.672 for CBCT scans. 1037 features were extracted from each patient's CBCT and pCT images, 73 image features were found to have R values greater than 0.9, including three features related to the staging and prognosis of rectal cancer.

**Conclusion:** In this study, we proposed an automatic, fast, and consistent method for rectal cancer GTV segmentation for pCT and CBCT scans. The findings of radiomic results indicate that CBCT images have significant research value in the field of radiomics.

## KEYWORDS

radiomics, cone-beam CT, rectal cancer, 3D-RUNet, CLAHE

**Abbreviations:** pCT, Planning Computerized Tomography; CBCT, Cone Beam CT; TME, Total Mesorectal Excision; nCRT, Neoadjuvant Chemoradiotherapy; pCR, Pathologic Complete Response; CLAHE, Contrast Limited Adaptive Histogram Equalization; DSC, Dice similarity coefficient; 95%HD, Hausdorff-95 distance; ASSD, average symmetric surface distance; PCC, Pearson correlation coefficient.



## Introduction

The standard-of-care treatment for locally advanced rectal cancer (LARC, T3+ or N+) is currently total mesorectal excision (TME) followed by neoadjuvant chemoradiotherapy (nCRT) (1–3). After nCRT, approximately 15%–27% of patients can show a pathologic complete response (pCR) (4, 5). And several prior studies have shown that these patients typically have outstanding long-term outcomes without surgery (6–9). Habr-Gama and colleagues suggested a “wait and see” policy, while Maas and colleagues approach a reasonable solution that could avoid surgery and preserve organs (6, 10). The pCR, however, could only be performed using histopathological analysis of surgically resected specimens. So, it remains a major challenge to develop a non-invasive, validated way to reliably classify pCR patients after chemoradiotherapy.

Tumor segmentation and the subsequent quantitative of rectal cancer in medical images provide valuable information for the analysis of pathologies and prediction of patient outcomes. Numerous studies have shown that image radiomic features extracted from multi-modality imaging techniques, such as CT (11), MRI (12), and PET-CT (13), can be used to predict the treatment response and prognosis of locally advanced rectal cancer. Machine learning models based on CT and MRI image radiomics have also demonstrated good reproducibility and robustness (14, 15). However, these imaging techniques are typically used for disease diagnosis before or after radiotherapy, and are unable to monitor the changes in tumor heterogeneity during the treatment process (16). In contrast, cone-beam CT (CBCT) scans, which are routinely obtained during radiotherapy to examine patient position changes, do not require patients to undergo additional radiation exposure. The features extracted from CBCT may provide valuable information on the changes in tumors during the treatment process without exposing patients to additional radiation hazards. The goal of this study is to examine whether CBCT features can be used for clinical staging or prognosis assessment of tumors by comparing the linear relationship between CBCT and pCT-extracted imaging features.

Precise segmentation of rectal cancer as the mask is particularly important for radiomics extraction and affects the robustness of radiomic features. The current image segmentation methods include manual, semiautomatic, and fully segmentation. The U-Net (17) based models have proven effectiveness over traditional medical segmentation algorithms. However, the 2D U-Net model for segmenting tumors only obtain a single tumor slice in CT scan, while tumors are usually distributed in continuous CT slices (18). To solve the issues, we extend the 2D U-Net to a 3D version with Resnet architecture to capture the inter-slice continuity of the tumor.

## Methods

### Patients

The article under consideration presents a retrospective analysis of 103 patients who underwent neoadjuvant chemoradiotherapy in

the Department of Oncology Radiotherapy at the Affiliated Hospital of Xiamen University between January 2019 and October 2020. The study followed the ethical principles outlined in the Helsinki Declaration and its subsequent relevant revisions for all procedures involving human participants. The inclusion criteria for the retrospective analysis were as follows:

1. Biopsy-confirmed primary rectal adenocarcinoma
2. Locally advanced disease (T stage  $\geq 3$ ) prior to treatment
3. No prior receipt of chemoradiotherapy, radiotherapy, or chemotherapy

Only patients who met these criteria were included in the analysis. In compliance with the Helsinki Declaration of 1964 and its later corresponding revisions, all the procedures carried out in this study involving human participants were compliant. CBCT was scanned during the whole treatment period.

### Image acquisition

The CT scans utilized in this study for lesion localization were conducted using a GE Lightspeed device (manufactured by GE Medical System, USA). The scans were performed using parameters of 120 kV tube voltage, 200 mA tube current, a 512×512 reconstruction matrix, and 5mm slice thickness. In addition, CBCT scans were performed using a Truebeam linear accelerator (manufactured by Varian Medical System), with a 512×512 reconstruction matrix and 3mm slice thickness.

### CLA-UNet structure analysis

In this article, we extend the traditional 2D U-Net to 3D U-Net equipped with ResNet architecture to capture the inter-slice continuity of the tumor, and we propose a CLAHE (19) processed U-Net (CLA-UNet) model to further improve the clarity of the anatomy structure, texture, and boundary in the CT image before segmentation. This CLA-UNet designed to accurately segment the lesion area in positioning CT (pCT) and cone-beam CT (CBCT) images of rectal cancer tumor. The CLA-UNet network combines the popular 3D-UNet structure with a residual module (Res-net) to improve the accuracy of tumor location and boundary description, ensuring a precise radiation target area.

The structure of the CLA-UNet network is illustrated in Figure 1 of the article. The network is designed to automatically segment the lesion area both in pCT and CBCT images, which also providing valuable information for radiation treatment planning and evaluation.

### CLAHE algorithm processing

Prior to importing the 3D CT data into the 3D-RUNet network for training, a preprocessing step is carried out to crop the CT



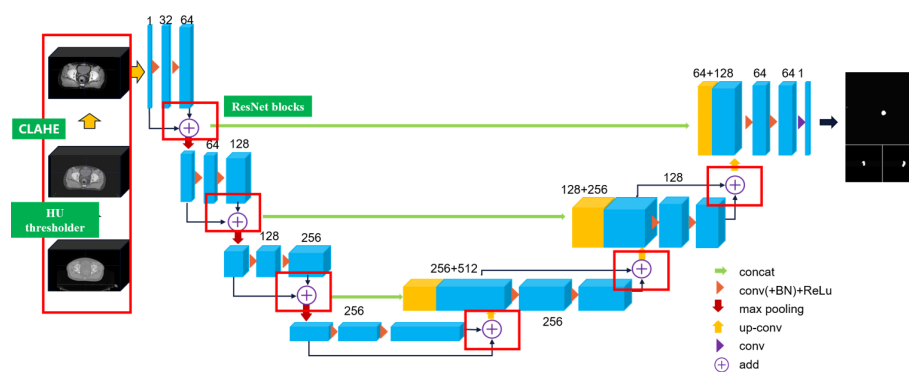


FIGURE 1  
Layers of the proposed CLA-UNet model.

volume and enhance the image contour. This involves removing any blank areas or areas without drawn target regions, resulting in the cropped CT volume being resized to  $256 \times 256 \times 128$  voxels using linear interpolation. As rectal tumors are considered soft tissue, the CT value range is restricted to  $(-200, 300)$ . To focus the network training on information that is relevant to rectal tumors, the 3D CT image is thresholded such that any image values outside the specified range are replaced with corresponding boundary values.

To further improve the clarity of the anatomy structure, texture, and boundary in the CT image, the thresholded CT image is processed using the Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm.

This results in the rectal structure and boundary becoming clearer and an overall improvement in image quality, as demonstrated in a comparison of the image before and after thresholding and CLAHE processing shown in Figure 2.

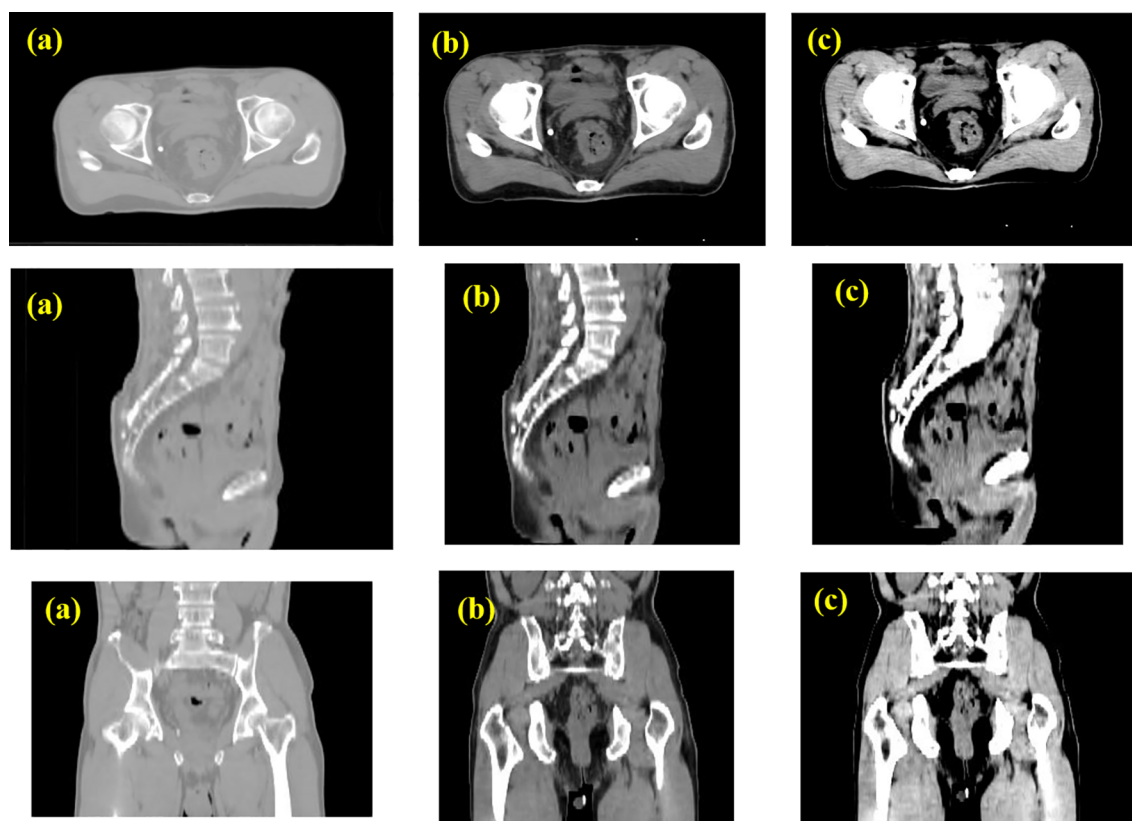


FIGURE 2  
Data pre-processing with CLAHE. (A) Original CT image; (B) CT image intercepted by threshold with  $HU=(-200 \sim 300)$ ; (C) CT image transformed by CLAHE.

## Model training and evaluation

The CLA-UNet network is based on the PyTorch kernel platform and the training machine is configured with a Windows 10 operating system and a Quadro P5000 GPU device. The Adam optimizer is used for training with an initial learning rate of 0.0001, and the batch size for network training is set to 2 with a maximum training cycle of 150. This study is trained on a 103 rectal cancer CT dataset provided by the First Affiliated Hospital of Xiamen University, with 70% of the data randomly selected as the training set and 30% as the testing set. The network is trained using the Tversky Loss function, shown in Equation 1.

$$T(\alpha, \beta) = \frac{\sum_{i=1}^N p_{0i} g_{0i}}{\sum_{i=1}^N p_{0i} g_{0i} + \alpha \sum_{i=1}^N p_{0i} g_{1i} + \beta \sum_{i=1}^N p_{1i} g_{0i}} \quad (1)$$

$p_{0i}$  is the probability that the  $i$ -th voxel is a tumor,  $p_{1i}$  is the probability that the  $i$ -th voxel is not a tumor,  $g_{0i}$  is 1 if the voxel is abnormal, 0 if it is not abnormal,  $g_{1i}$  is the opposite of  $g_{0i}$ . Tversky Loss effectively solves the problem of data imbalance by finding a better balance between accuracy and recall.

The model is evaluated using the Dice similarity coefficient (DSC), Hausdorff-95 distance (95% HD), and average symmetric surface distance (ASSD) evaluation metric to compare the segmentation results with those of CBCT scans.

The DSC is defined as follows, shown in Equation 2:

$$\text{DSC} = \frac{2|P \cap G|}{|P| + |G|} \quad (2)$$

Where the  $P$  represents the ground truth,  $G$  denotes the prediction results and the  $P \cap G$  is the intersection of  $P$  and  $G$ . The range of DSC evaluation is  $[0, 1]$ , and the higher the score is close to 1.0, the more accurate the prediction is.  $P$  and  $G$  represent the target structure drawn by the physician and the model, respectively.

The HD(A,B) is defined as follows, shown in Equation 3:

$$\text{HD}(A, B) = \max(\max_{a \in A}(\min_{b \in B} d(a, b)), \max_{b \in B}(\min_{a \in A} d(b, a))) \quad (3)$$

Where  $d(a, b)$  is the distance between the point  $a$  and  $b$ .

The ASSD is shown in Equation 4:

$$\text{ASSD} = \frac{1}{S(A) + S(B)} (\sum_{S_A \in S(A)} d(S_A, S(B)) + \sum_{S_B \in S(B)} d(S_B, S(A))) \quad (4)$$

Where  $S(A)$  represents the surface voxels in set  $A$ , and  $d(S_A, S(B))$  represents the shortest distance from  $S_A$  to  $S(B)$ .

## Radiomics correlation analysis

The open source radiomics extraction software Pyradiomics 3.0 (<https://pyradiomics.readthedocs.io/en/latest/>) was used to extract high-throughput features from patient images. In the test group, a total of 30 cases were automatically segmented from pCT and CBCT images using the 3D CLA-UNet model. All the images were filtered by Laplacian of Gaussian (LoG) filter and performed wavelet

transformation, so there are four types of images, namely, "Original Images", "texture Images", "LoG Images", and "Wavelet Images".

After that, the Pearson correlation test was used to analyze the correlation between the image radiomic feature values of pCT and CBCT, if the Pearson correlation coefficient (PCC)  $R$  is greater than 0.9, it is considered that the feature value has strong consistency and substitutability in machine learning (14). Pearson correlation coefficient is a method for measuring the similarity of vectors, the range of correlation is  $[-1, 1]$ , it is defined as the ratio of the covariance and standard deviation of two feature variables, calculated as follows shown in Equation 5:

$$\rho = \frac{\text{Cov}(X, Y)}{\sigma_x \sigma_y} = \frac{E[(X - \mu_x)(Y - \mu_y)]}{\sigma_x \sigma_y} \quad (5)$$

Among them,  $X$  and  $Y$  are two different groups of eigenvalue variables, and  $\mu_x$  and  $\sigma_x$  are mean and standard deviation respectively. This process was implemented using the Pearson algorithm in the R language (R language 3.6).

## Results

### Study population

The radiomics analysis were conducted on a test set of 30 patients, patients' radiomic characteristics were grouped by LN metastasis and compared in Supplementary Table 1. The clinical information includes gender, age, pathology, and clinical-stage information. All the patients received pCRT followed by TME, and group differences were examined.

### Model performance

The trend of the average Loss and average Dice values during the CLA-UNet training process is depicted in Figure 3. It is evident that as the number of training rounds, also known as epochs, increases, the Loss values (a) on both the training and validation sets rapidly decrease, while the Dice values (b) steadily improve. When the number of epochs reaches 50, the trend stabilizes, with the Loss value reaching close to 0 and the Dice value reaching a stable value around 0.8, details were shown in Table 1.

The results of the CLA-UNet network training on 103 samples showed that the network could segment the rectal tumor with good accuracy, details shown in Figure 4. As seen in the transverse sections, the performance of the automatic segmentation was satisfactory for the majority of the levels. However, there were some regions, particularly near the cecum and anus, where larger discrepancies were observed between the manual annotations and the machine segmentations. This was likely due to the close proximity of densities in these areas, making it more challenging to distinguish between the different tissues. In such cases, manual annotations by doctors may require additional imaging modalities, such as MRI or PET-CT, or the use of their experience to assist in the outlining process. Despite these limitations, the average Dice

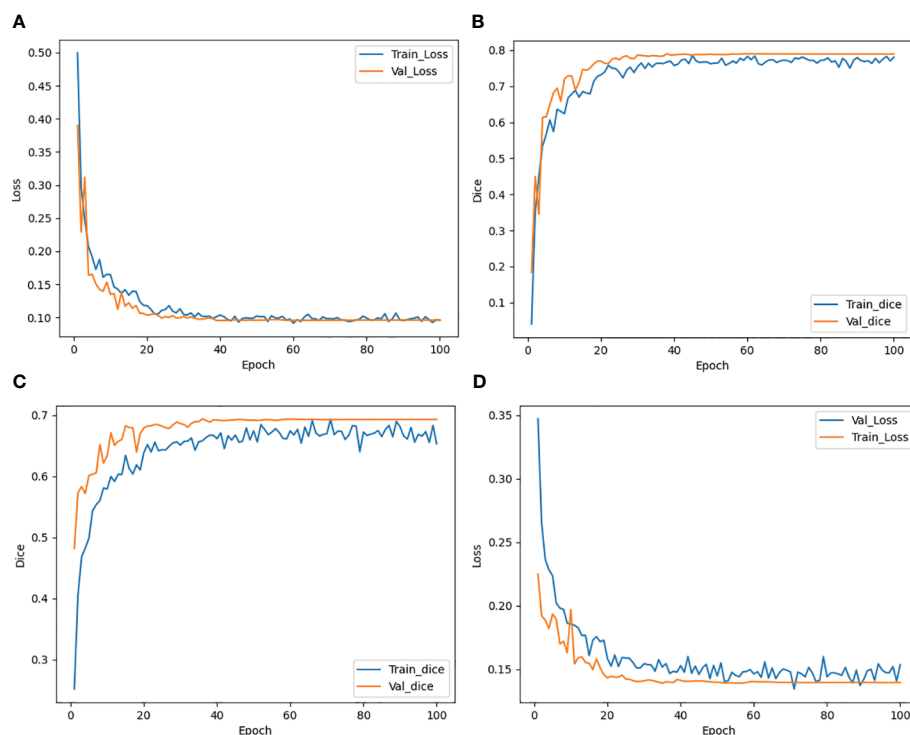


FIGURE 3

Trend of Loss and Dice value with epoch during training both for CTs (A, B) and CBCTs (C, D).

score for the CLA-UNet outlining ranged from 0.72 to 0.86, which is generally in line with the clinical requirements.

## Inter-group correlations calculation

Automatically contouring 30 patient images based on deep learning algorithms to ensure consistency in contouring results. 1037 features were extracted from pCT and CBCT modalities using machine learning algorithms, including shape features (n=14), first-order features (n=19), texture features (n=172), wavelet features (n=728), and loG features (n=104). Pearson's correlation analysis was used to analyze the correlation of two sets of features, and strong correlated features were extracted. 73 features had Pearson correlation coefficients  $R$  greater than 0.9, meaning that these 73 features can be interchangeable. The three features confirmed in previous literature to be related to rectal cancer staging and new adjuvant therapy effectiveness (20), including original first-order Energy, wavelet-HLH\_glrIm Gray Level Non Uniformity, and original\_glrIm Gray Level Non Uniformity, are included in the

strong correlated features. The correlation coefficients  $R$  of these three features are 0.9521, 0.9406, and 0.9191, respectively, the data of the subsequent two radiomics were shown in detail in Figure 5.

## Discussion

The usage of imaging radiomics in CT scans for rectal cancer diagnosis and prognosis analysis has been well documented in previous studies (21). However, the application of CBCT in this regard has been less explored. In this study, a modified deep learning algorithm, CLA-UNet, was developed to automatically segment the rectal cancer tumor location. With our previous work, we had trained the deep-learning model and used it in our clinical practice. The results indicate that the CLA-UNet model is feasible and time-saving to perform fully automatic segmentation for the rectal tumor both on CBCT and pCT images. To evaluate the accuracy of the 3D mask of the CLA-UNet mode, we compared the coverage of the predicated area with the manual segmentation with an average Dice value, which was 0.792 for pCT and 0.672 for CBCT scans.

Subsequently, imaging radiomic features were extracted and analyzed from both pCT and CBCT scans using machine learning algorithms. The results showed that 73 features had a correlation coefficient ( $R$ ) greater than 0.9. Our findings also demonstrate that some of the prognostically significant features of radiomics have a strong linear relationship between the pCT and CBCT images based on automatic image segmentation, which indicate a measure of interchangeability between the two scans. These high-correlation

TABLE 1 Comparison results on the test set for the pCT and CBCT scans.

|          | pCT               | CBCT              |
|----------|-------------------|-------------------|
| DSC      | 0.792 $\pm$ 0.056 | 0.672 $\pm$ 0.084 |
| HD95(mm) | 15.4 $\pm$ 9.5    | 20.2 $\pm$ 12.4   |
| ASSD     | 4.3 $\pm$ 2.1     | 5.4 $\pm$ 2.6     |

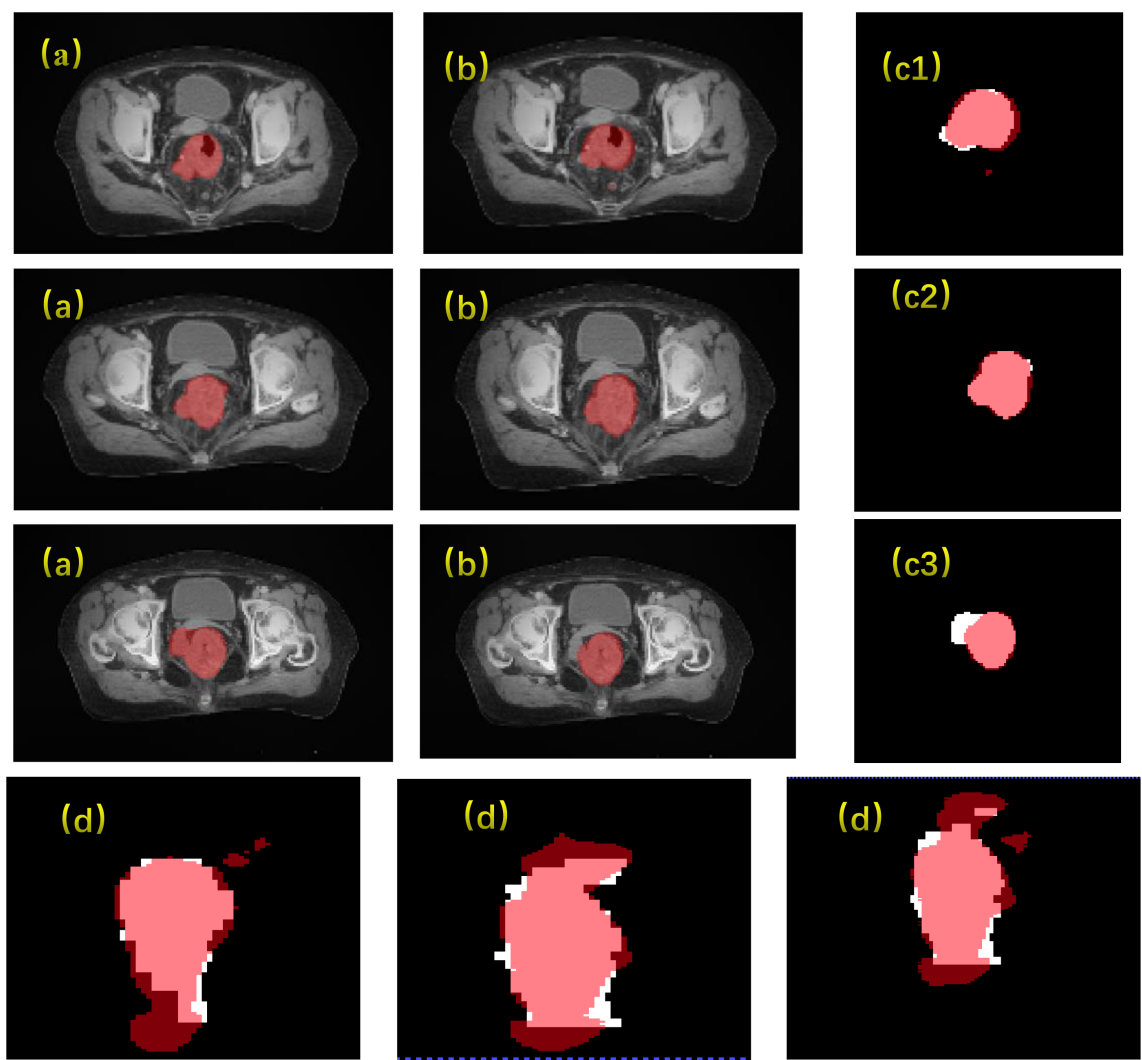


FIGURE 4  
Gross tumor volume contouring with Dice = 0.78. (A) manual; (B) contouring of CLA-UNet; comparing of segmentations between manual and deep-learning both for transverse (C) and coronal (D) planes. The red was contouring by CLA-UNet and the white was contouring by manual.

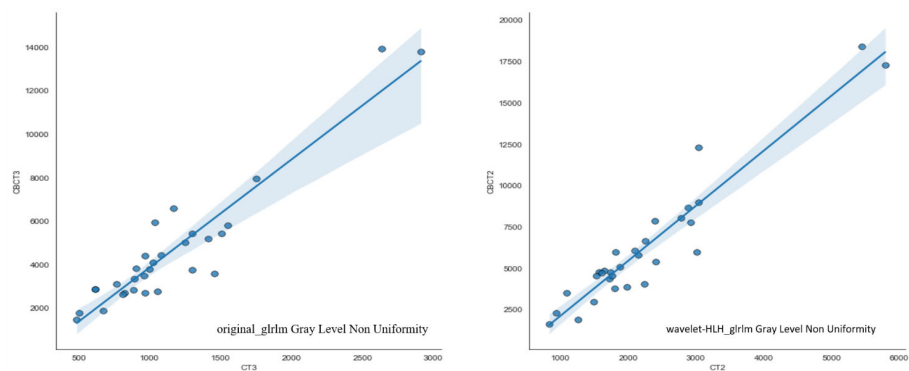


FIGURE 5  
Scatter diagram of features extracted from CBCT and CT scans.

features include those previously reported in literature as important indicators for rectal cancer (22, 23). This highlights the potential value of CBCT as an early biomarker for treatment response evaluation (24).

Moreover, high-dimensional features were confirmed in previous literature to be related to rectal cancer staging and new adjuvant therapy effectiveness, in the present study, most of the key features were wavelet features, which are challenging to decipher with the naked eyes. However, high-dimensional features hold more detailed information about the tumor and more sensitive when assessing pCR, as was also demonstrated in recent study (20).

However, there are still some challenges in using CBCT images for radiomics, cause the extracted textural features typically depend on the reconstruction and scanning parameters (25). To be consistent throughout in this study all the CBCTs were resampled into an equal size of 5mm as pCTs, and the influence of slice thickness on the radiomic parameters needs further investigation. Nevertheless, other unknown factors may also influence the consistency evaluation between pCT and CBCT radiomics. Potentially, a radiomics approved reconstruction or corrections could in general improve the consistency and utility of radiomics in medical imaging. Besides this, the detector size of the CBCT has a limited field of view (FOV) that may not be large enough for off-axis patient positions and extensive tumors.

In conclusion, this study provides a preliminary exploration of the correlation between pCT and CBCT imaging radiomics in locally advanced rectal cancer. The CLA-UNet algorithm was successfully applied to segment the rectal tumors, then the correlation between the extracted imaging radiomic features was analyzed. The results showed that radiomic features have a high correlation between pCT and CBCT images, indicating the potential use of CBCT images as an early biomarker for the evaluation of treatment response. However, there are still some limitations in the use of CBCT images. First, the patient sample size was small, a larger sample size test is needed to achieve robust results. Second, the differences in reconstruction algorithms and scan parameters, FOV limitations, and sensitivity to motion artifacts which will influence the consistency evaluation between pCT and CBCT radiomics. Further research is needed to explore the potential applications of CBCT in the diagnosis and prognosis of rectal cancer. Future studies could also focus on multi-center data collection and validation, and on reducing the number of features for clinical predictions.

## Conclusion

In this study, we have presented a modified 3D-UNet segmentation method, CLA-UNet, based on deep learning to automatic segmentation the rectal cancer tumor both for pCTs and CBCTs. Subsequently, radiomic features were extracted and analyzed to find out the inter-group correlation, and the results indicate that some of the prognostically significant features of radiomics have a strong linear relationship between the pCT and CBCT images, which indicate a measure of interchangeability between the two scans.

## 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 First Affiliated Hospital of Xiamen University medical ethics committee. 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

YY: Writing – original draft. GG: Data curation, Software, Writing – review & editing. XG: Conceptualization, Funding acquisition, Writing – review & editing. YH: Formal analysis, Validation, Writing – review & editing. LS: Resources, Visualization, Writing – review & editing. ZK: Funding acquisition, Methodology, Supervision, Writing – original draft. ZZ: Formal analysis, Validation, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Contract grant sponsor (contract grant number): The Nature Scientific Foundation of China (Grant No. 81772893); Fujian Province Young and Middle-aged Teacher Education Research Project (No. JAT210852); Zhangzhou Institute of Technology school-level projects (No. ZZY2021B031).

## 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/fonc.2024.1301710/full#supplementary-material>



## References

- Kapiteijn E, Marijnen CA, Nagtegaal ID, Putter H, Steup WH, Wiggers T, et al. Preoperative radiotherapy combined with total mesorectal excision for resectable rectal cancer. *N Engl J Med* (2001) 345:638–46. doi: 10.1056/NEJMoa010580
- van de Velde CJ, Boelens PG, Borras JM, Coebergh J-W, Cervantes A, Blomqvist L, et al. EURECCA colorectal: multidisciplinary management: European consensus conference colon & rectum. *EJC* (2014) 50:1.e1–1.e34. doi: 10.1016/j.ejca.2013.06.048
- van Gijn W, Marijnen CA, Nagtegaal ID, Kranenburg EM-K, Putter H, Wiggers T, et al. Preoperative radiotherapy combined with total mesorectal excision for resectable rectal cancer: 12-year follow-up of the multicentre, randomised controlled TME trial. (2011) 12:575–82. doi: 10.1016/S1470-2045(11)70097-3
- Sanghera P, Wong D, McConkey C, Geh J, Hartley A. Chemoradiotherapy for rectal cancer: an updated analysis of factors affecting pathological response. *Clin Oncol* (2008) 20:176–83. doi: 10.1016/j.clon.2007.11.013
- Maas M, Nelemans PJ, Valentini V, Das P, Rödel C, Kuo L-J, et al. Long-term outcome in patients with a pathological complete response after chemoradiation for rectal cancer: a pooled analysis of individual patient data. *Lancet Oncol* (2010) 11:835–44. doi: 10.1016/S1470-2045(10)70172-8
- Habr-Gama A, Perez RO, Proscurshim I, Campos FG, Nadalin W, Kiss D, et al. Patterns of failure and survival for nonoperative treatment of stage c0 distal rectal cancer following neoadjuvant chemoradiation therapy. *J Gastrointest Surg* (2006) 10:1319–29. doi: 10.1016/j.gassur.2006.09.005
- Borschitz T, Wachtlin D, Möhler M, Schmidberger H, Junginger T. Neoadjuvant chemoradiation and local excision for T2-3 rectal cancer. *Ann Surg Oncol* (2008) 15:712–20. doi: 10.1245/s10434-007-9732-x
- Sebag-Montefiore D, Stephens RJ, Steele R, Monson J, Grieve R, Khanna S, et al. Preoperative radiotherapy versus selective postoperative chemoradiotherapy in patients with rectal cancer (MRC CR07 and NCIC-CTG C016): a multicentre, randomised trial. *Lancet* (2009) 373(9666):811–20. doi: 10.1016/S0140-6736(09)60484-0
- Renahan AG, Malcomson L, Emsley R, Gollins S, Maw A, Myint AS, et al. Watch-and-wait approach versus surgical resection after chemoradiotherapy for patients with rectal cancer (the OnCoRe project): a propensity-score matched cohort analysis. *Lancet Oncol* (2016) 17(2):174–83. doi: 10.1016/S1470-2045(15)00467-2
- Marijnen CAM. Organ preservation in rectal cancer: have all questions been answered? *Lancet Oncol* (2015) 16(1):e13–22. doi: 10.1016/S1470-2045(14)70398-5
- Ganeshan B, Panayiotou E, Burnand K, Dizdarevic S, Miles K. Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: a potential marker of survival. *European radiology* (2012) 22:796–802. doi: 10.1007/s00330-011-2319-8
- Yu X, Song W, Guo D, Liu H, Zhang H, He X, et al. Preoperative prediction of extramural venous invasion in rectal cancer: comparison of the diagnostic efficacy of radiomics models and quantitative dynamic contrast-enhanced magnetic resonance imaging. *Front Oncol*. (2020) 10:459. doi: 10.3389/fonc.2020.00459
- Cook GJ, Yip C, Siddique M, Goh V, Chicklore S, Roy A, et al. Are pretreatment 18F-FDG PET tumor textural features in non-small cell lung cancer associated with response and survival after chemoradiotherapy? *J Nucl Med* (2013) 54(1):19–26. doi: 10.2967/jnumed.112.107375
- Gillies RJ, Kinahan PE, Hricak HJR. Radiomics: images are more than pictures, they are data. *Radiology* (2016) 278:563–77. doi: 10.1148/radiol.2015151169
- Xie K, Sun H, Lin T, Gao L, Sui J, Ni X. Research progresses in feature extraction of radiomics. *Chin J Med Imaging Technol*. (2017) 033:1792–6. doi: 10.13929/j.1003-3289.201707022
- García-Figueiras R, Baleato-González S, Padhani AR, Luna-Alcalá A, Vallejo-Casas JA, et al. How clinical imaging can assess cancer biology. *Insights Imaging*. (2019) 10(1):28. doi: 10.1186/s13244-019-0703-0
- Ronneberger O, Fischer P, Brox T. U-NET: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells W, Frangi A (eds). *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science*. vol. 9351. (Cham: Springer) (2015). doi: 10.48550/arXiv.1505.04597
- Huang Q, Sun J, Ding H, Wang X, Wang G. Robust liver vessel extraction using 3D U-Net with variant dice loss function. *Comput Biol Med*. (2018) 101:153–62. doi: 10.1016/j.compbiomed.2018.08.018
- Mishra A. Contrast limited adaptive histogram equalization (CLAHE) approach for enhancement of the microstructures of friction stir welded joints. *Advances in Mechanics* [preprint] (2021) 9(3):216–40. doi: 10.21203/rs.3.rs-607179/v1 In Review.
- Liu Z, Zhang XY, Shi YJ, Wang L, Zhu HT, Tang Z, et al. Radiomics analysis for evaluation of pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer. *Clin Cancer Res*. (2017) 23:7253–62. doi: 10.1158/1078-0432.CCR-17-1038
- Wang J, Shen L, Zhong H, Zhou Z, Hu P, Gan J, et al. Radiomics features on radiotherapy treatment planning CT can predict patient survival in locally advanced rectal cancer patients. *Sci Rep*. (2019) 9(1):15346. doi: 10.1038/s41598-019-51629-4
- Wang F, Tan BF, Poh SS, Siow TR, Lim FLWT, Yip C, et al. Predicting outcomes for locally advanced rectal cancer treated with neoadjuvant chemoradiation with CT-based radiomics. *Sci Rep*. (2022) 12(1):6167. doi: 10.1038/s41598-022-10175-2
- Bibault JE, Giraud P, Housset M, Durdur C, Taieb J, Berger A, et al. Deep learning and radiomics predict complete response after neo-adjuvant chemoradiation for locally advanced rectal cancer. *Sci Rep*. (2018) 8(1):12611. doi: 10.1038/s41598-018-30657-6
- Wang J, Shen L, Zhong H, Hu P, Zhang Z. Radiomics prediction model for locally advanced rectal cancer. *Int J Radiat Oncol Biol Phys*. (2017) 99(2):S184. doi: 10.1016/j.ijrobp.2017.06.460
- van Timmeren JE, Leijenaar RTH, van Elmpt W, Reymen B, Lambin P. Feature selection methodology for longitudinal cone-beam CT radiomics. *Acta Oncol*. (2017) 56(11):1537–43. doi: 10.1080/0284186X.2017.1350285



## OPEN ACCESS

## EDITED BY

Yuxiang Zhou,  
Mayo Clinic Arizona, United States

## REVIEWED BY

Bikash Kumar Paul,  
Mawlana Bhashani Science and Technology  
University, Bangladesh  
Jaya Lakshmi Thangaraj,  
University of California, San Diego,  
United States

## \*CORRESPONDENCE

Anum Masood

✉ anum.masood@ntnu.no

Muhammad Awais

✉ muhammadawais@ciitwah.edu.pk

RECEIVED 26 October 2023

ACCEPTED 25 January 2024

PUBLISHED 05 March 2024

## CITATION

Awais M, Abdal MN, Akram T, Alasiry A,  
Marzougui M and Masood A (2024) An  
efficient decision support system for leukemia  
identification utilizing nature-inspired deep  
feature optimization.

*Front. Oncol.* 14:1328200.

doi: 10.3389/fonc.2024.1328200

## COPYRIGHT

© 2024 Awais, Abdal, Akram, Alasiry, Marzougui  
and Masood. 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.

# An efficient decision support system for leukemia identification utilizing nature-inspired deep feature optimization

Muhammad Awais<sup>1,2\*</sup>, Md. Nazmul Abdal<sup>3</sup>, Tallha Akram<sup>1</sup>,  
Areej Alasiry<sup>4</sup>, Mehrez Marzougui<sup>4</sup> and Anum Masood<sup>5\*</sup>

<sup>1</sup>Department of Electrical and Computer Engineering, COMSATS University Islamabad, Wah, Pakistan,

<sup>2</sup>Department of Computer Engineering, TED University, Ankara, Türkiye, <sup>3</sup>Department of Computer Science and Engineering, University of Liberal Arts Bangladesh, Dhaka, Bangladesh, <sup>4</sup>College of Computer Science, King Khalid University, Abha, Saudi Arabia, <sup>5</sup>Department of Physics, Norwegian University of Science and Technology, Trondheim, Norway

In the field of medicine, decision support systems play a crucial role by harnessing cutting-edge technology and data analysis to assist doctors in disease diagnosis and treatment. Leukemia is a malignancy that emerges from the uncontrolled growth of immature white blood cells within the human body. An accurate and prompt diagnosis of leukemia is desired due to its swift progression to distant parts of the body. Acute lymphoblastic leukemia (ALL) is an aggressive type of leukemia that affects both children and adults. Computer vision-based identification of leukemia is challenging due to structural irregularities and morphological similarities of blood entities. Deep neural networks have shown promise in extracting valuable information from image datasets, but they have high computational costs due to their extensive feature sets. This work presents an efficient pipeline for binary and subtype classification of acute lymphoblastic leukemia. The proposed method first unveils a novel neighborhood pixel transformation method using differential evolution to improve the clarity and discriminability of blood cell images for better analysis. Next, a hybrid feature extraction approach is presented leveraging transfer learning from selected deep neural network models, InceptionV3 and DenseNet201, to extract comprehensive feature sets. To optimize feature selection, a customized binary Grey Wolf Algorithm is utilized, achieving an impressive 80% reduction in feature size while preserving key discriminative information. These optimized features subsequently empower multiple classifiers, potentially capturing diverse perspectives and amplifying classification accuracy. The proposed pipeline is validated on publicly available standard datasets of ALL images. For binary classification, the best average accuracy of 98.1% is achieved with 98.1% sensitivity and 98% precision. For ALL subtype classifications, the best accuracy of 98.14% was attained with 78.5% sensitivity and 98% precision. The proposed feature selection method shows a better convergence behavior as compared to classical population-based meta-heuristics. The suggested

solution also demonstrates comparable or better performance in comparison to several existing techniques.

#### KEYWORDS

bio-inspired, CNN, transfer learning, leukemia classification, deep learning, metaheuristics optimization

## 1 Introduction

Blood is a vital fluid for the human body. It performs a number of crucial physiological functions, including the distribution of oxygen and nutrients from organs to cells, delivery of waste products from cells to organs for elimination, the maintenance of the human immune system, clotting and wound healing process, and the regulation of body temperature and fluid balance. The body's main source of blood production is the bone marrow, a spongy tissue-like structure located within the bone cavities. A complex process known as hematopoiesis involves the maturation of stem cells into other blood cell types.

Figure 1 demonstrates the categorization of stem cells during hematopoiesis of a normal human being. The hematopoietic stem cells develop into either of two types of cells, i.e., a) lymphoid stem cells and b) myeloid stem cells. The lymphoid stem cells are then converted into the lymphoid blast, which then matures into B and T

subtypes of lymphocytes. In contrast, the myeloid type of stem cells matures to synthesize erythrocytes, platelets, and various types of granulocytes (i.e., basophils, eosinophils, neutrophils, and monocytes).

The body produces the blood cells in a controlled manner as per its requirements. Each cell type has a specific function in preserving a person's general state of health. An abnormality in the production and structure of blood cells leads to certain medical conditions. For example, white blood cells (WBCs), also referred to as leukocytes, constitute an integral part of the general immune and inflammatory response system (1, 2). Leukemia is a blood malignancy that is caused by the unregulated production of immature leukocytes in the bone marrow. Figure 2 shows a broad classification of leukemia, which is primarily of two types, i.e., acute and chronic, depending upon its progress rate. Chronic leukemia is slow-growing and takes months or years to manifest its symptoms, whereas acute leukemia develops rather swiftly. Each type of leukemia is further categorized

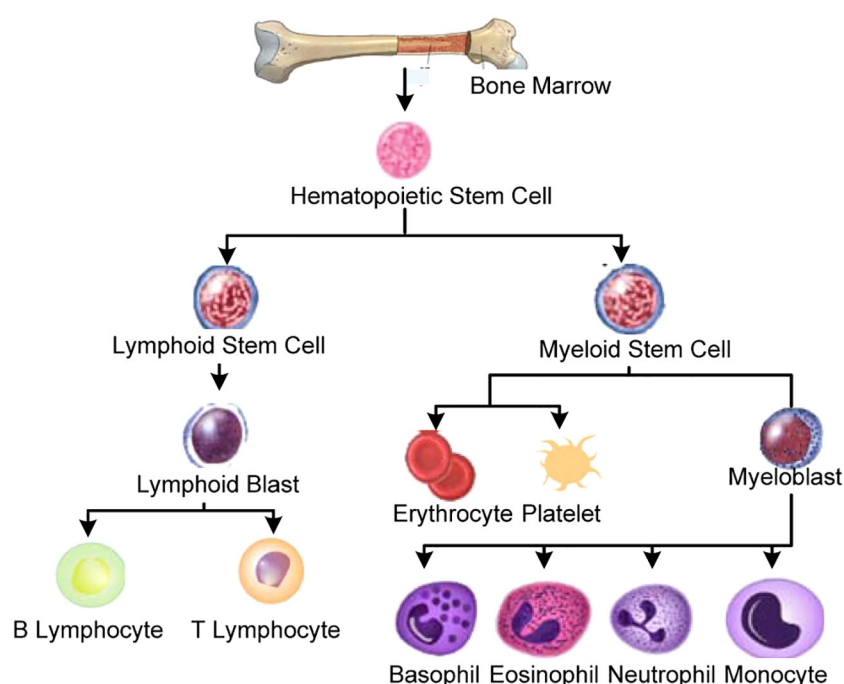
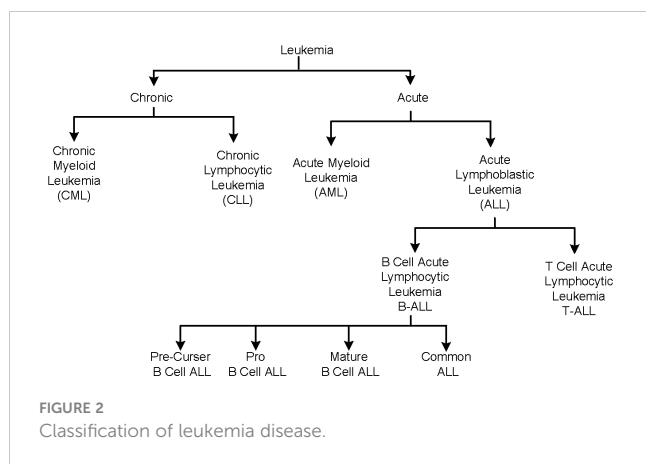


FIGURE 1  
Human hematopoiesis.



on the basis of affected leukocytes. In the chronic leukemia category, chronic myeloid leukemia (CML) affects the myeloid type of cells, whereas chronic lymphocytic leukemia affects the lymphoid cells. Similarly, the acute leukemia category is further classified as acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML) categories. ALL is further categorized into T-cell or B-cell subtypes. B-cell ALL is the most prevalent type of leukemia, impacting the B-cell lymphocytes; it is further divided into pre-cursor, pro, mature, and common B-cell ALL subtypes.

The existing medical approach for leukemia diagnosis involves a series of tests, ranging from simple blood count tests to more invasive tests such as needle biopsy or bone marrow aspiration. A blood test that shows a high value of white blood cell count suggests leukemia diagnosis. An important diagnostic tool in the evaluation of leukemia is the peripheral blood smear test. It involves the smearing of blood on the slide and its visual inspection under the microscope. A blood smear of a leukemia patient shows a significant number of purple-stained lymphoblasts in the bloodstream, with poorly defined boundaries. Traditionally, hematologists perform this ocular inspection of blood smears. This manual method not only consumes much time and effort of medical experts but can also be error-prone due to several external factors. Blood analysis is usually the first step of leukemia diagnosis and is carried out in conjunction with more detailed analysis methods such as RNA sequencing and molecular genetics. Computer-aided automation of blood analysis can be a significant step in reducing the time and cost of leukemia diagnosis.

Thanks to the landmark advancement in the domain of digital electronics and imaging technologies, automated blood analysis has been made possible. In particular, computer vision-based blood disease diagnosis has seen an increased research focus in recent years. However, due to morphological similarities across various blood entities as well as their structural anomalies, accurate machine learning-based blood analysis, particularly leukemia subtype detection, is still a challenge. A breakthrough in modern computer vision approaches, i.e., deep convolutional neural networks (CNNs), has shown a promising solution for a variety of classification scenarios (3, 4). They are capable of extracting a diverse range of features from the images. However, a large and well-labeled dataset is typically required to achieve a certain acceptable accuracy level. In the biomedical domain, a detailed dataset for CNN training from scratch is not readily available. Transfer learning is an often adopted

approach in which deep CNN, pre-trained on another dataset, is retrained for a specific task (5, 6). Some well-known pre-trained CNNs include ResNet (7), MobileNet (8), DarkNet (9), Inception (10), and Xception (11). Modern methods also propose ensembles of feature vectors extracted from multiple CNNs (12). Apart from the wide use of deep CNNs in the computer vision domain, one limiting factor is the very high size of their extracted feature vectors. The present focus of research is to investigate approaches to improve the accuracy of classification systems while reducing their computational complexity and memory requirements.

## 2 Literature review

Existing research on leukemia detection can be broadly categorized into two types of approaches. In the first category, the studies are included, which perform white blood cell classification as an important preliminary step. The second category of work is focused on considering the stained images containing leukemia blasts and proposing an efficient method for leukemia subtype classification. Some considerable studies from the first category are summarized as follows. Sanei and Lee proposed a method that computes eigenvectors of blood cell images using the minimization of similarity approach (13). Using the density and color information of eigencells, a Bayesian classifier was used to perform cell classification. Kumar et al. (14) used various image pre-processing techniques with a random forest classifier for blood cancer detection. Su et al. (15) suggested a segmentation method based on detecting a discriminating zone of WBCs on the hue, saturation, and intensity (HSI) space. WBC classification was performed using geometrical, color, and local directional pattern (LDP) features. Sharma et al. (16) used DenseNet121 CNN for WBC classification. Almezghwi and Serte (17) proposed an image augmentation approach using generative adversarial networks, and classification was performed using DenseNet. Yildirim and Çinar (18) proposed Gaussian and median filtering approaches for image pre-processing. Then, multiple CNN architectures were trained for WBC classification.

Table 1 summarizes some recently published leukemia classification methods using blood smear images containing cell blasts. Bhattacharjee and Saini (19) applied different morphological operations to perform image segmentation. This was followed by classification using multiple baseline classifiers to diagnose the presence of ALL. The proposed solution achieved the best accuracy of 95.23% with the artificial neural network (ANN) classifier. Goutam and Sailaja (20) applied K-means clustering for segmentation, followed by the local directional path technique in order to extract features, and, finally, classification using support vector machines (SVMs). The F-measure achieved by this approach was 93.44. Patel and Mishra (21) applied histogram equalization along with the zack algorithm group wbcS in the smear images. Next, various morphological features including area, color, mean, and standard deviation were extracted and classified using SVM. The overall accuracy achieved by the system was 93.57%. Rawat et al. (22) proposed a method in which leukocytes and lymphocytes were sequentially separated from other blood cells. The shape and grey-level occurrence matrices were classified using a binary SVM

TABLE 1 Summary of some published studies on leukemia identification.

| Author                       | Method                                           | Leukemia type | Results             |
|------------------------------|--------------------------------------------------|---------------|---------------------|
| Bhattacharjee and Saini (19) | Morphological segmentation                       | ALL           | Accuracy: 96.67%    |
|                              | Classification: SVM                              |               | Accuracy: 90.47%    |
|                              | ANN                                              |               | Accuracy: 95.23%    |
|                              | K-means                                          |               | Accuracy: 85.71%    |
| Goutam and Sailaja (20)      | K-means clustering classification: SVM           | AML           | F-measure 93.44     |
| Patel and Mishra (21)        | Zack algorithm segmentation classification: SVM  | ALL           | Accuracy: 93.57     |
| Rawat et al. (22)            | K-means clustering classification: SVM           | ALL           | Accuracy: 89.8%     |
| Mishra et al. (23)           | DCT feature extraction                           | ALL           | Accuracy: 81.66     |
|                              | Classification: SVM                              |               |                     |
| Di Ruberto et al. (24)       | Watershed segmentation classification: CNN, SVM  | ALL           | Accuracy: 94.1.8%   |
| Bodzas et al. (25)           | Classification: ANN, SVM                         | ALL           | Specificity: 95.31% |
| Anwar and Alam (26)          | Automated feature extraction classification: CNN | ALL           | Accuracy: 99.5%     |

ALL, acute lymphoblastic leukemia; AML, acute myeloid leukemia; SVM, support vector machine; ANN, artificial neural network; DCT, discrete cosine transform; CNN, convolutional neural network.

classifier. Mishra et al. (23) performed nucleus feature extraction from RGB images using discrete cosine transform (DCT), followed by SVM classification. Di Ruberto et al. (24) utilized a multiscale blob detection scheme followed by the watershed algorithm for segmentation and, finally, classification using CNN and SVM classifiers. The proposed solution achieves a binary classification accuracy of 94.18%. Anwar and Alam (26) proposed a three-phase filtering algorithm to perform image segmentation. Next, 16 robust features were extracted, and classification was performed using ANN and SVM classifiers, yielding a specificity of 95.41%. Bodzas et al. (25) utilized different data augmentation techniques and performed training on their custom-proposed CNN architecture to obtain an overall accuracy of 99.5% for the binary classification of leukemic cells. Batool and Byun (27) proposed a lightweight deep learning-based EfficientNet-B3 model, which employs depth-wise separable convolutions for ALL classification. The proposed method achieves an accuracy of 96.81% for leukemia subtype classification using public datasets. Elhassan et al. (28) proposed an approach of AML detection from WBC images. First, a CMYK moment-based region of interest (ROI) localization method was used, followed by deep learning-based feature extraction and classification using several baseline classifiers. The proposed system achieves the best accuracy of 97.57%. In our previous

work (29), we utilized a quantum-inspired deep feature selection method for WBC classification for leukemia detection.

Modern transfer learning-based deep CNN techniques are characterized by their ability to extract a high number of characteristics from the input images. Due to the unreasonably huge feature sets that must be stored and processed, this has enormous computational costs and memory needs (30, 31). Most frequently, a large portion of these extracted deep characteristics are redundant and provide nothing to help with categorization. By selecting just potent, discriminating characteristics, feature selection is essential to reduce the complexity of feature vectors. This shortens the processing time while simultaneously improving the accuracy of the classification system. Several studies have investigated efficient feature selection methods, which include two kinds of approaches, namely, the filter approach and the wrapper approach. The filter approach quickly converges to the critical features, but it ignores the relationship between the classification algorithm and the feature subset. The wrapper approach, in contrast, considers a tight relationship between a subset of selected features and accuracy. While nature-inspired metaheuristics have been extensively applied in a wide range of combinatorial optimization problems (32–34), they have been recently investigated for feature selection optimization (35–38).

### 3 Contributions

This work proposes an improved pipeline for ALL subtype identification. The following are the main contributions of this study.

1. First, an efficient neighborhood pixel-based contrast enhancement technique was proposed based on a differential evolution algorithm, whose parameters were optimized using a greedy metaheuristic.
2. Next, two CNNs, namely, InceptionV3 and DenseNet201, were used for feature extraction using deep transfer learning.
3. A combined feature vector was created by performing a fusion of extracted feature vectors.
4. As a main contribution, the deep feature selection problem was modeled as an optimization problem and solved using a nature-inspired Grey Wolf Optimization (GWO) algorithm. The suggested approach selects only the most pertinent features, efficiently excluding correlated and noisy information.
5. The classification performance of various baseline classifiers was validated on the selected feature set to obtain the best-performing classifiers.
6. The proposed system achieves better performance metrics as compared to several existing feature selection methods, with a significant reduction in feature vector size.

### 4 Materials and method

The key components of the suggested methodology are elaborated upon in the subsequent sections.



## 4.1 Description of datasets

This study utilized different datasets of blood smear images. The initial dataset utilized in this study was the ALL-IDB2 dataset introduced by Scotti et al. (39). This dataset consisted of a total of 260 pictures, encompassing both healthy individuals and subjects diagnosed with ALL. The dataset was generated by employing an optical microscope that was attached to a Canon Power Shot G5 camera. The IDB2 dataset comprises images in which the region of interest has been cropped to include the area of interest for both normal and blast cells. All images were stored in the Tagged Image File Format (TIFF) and had a resolution of 2,592 pixels in width and 1,944 pixels in height. Figure 3 displays a selection of sample images from the ALL-IDB2 dataset.

This study used multiple datasets of blood smear images. First, the ALL-IDB2 dataset was used, which was composed of 260 images of healthy and ALL subjects. The dataset was prepared using an optical microscope connected to a Canon Power Shot G5 camera. The IDB2 dataset contained images in which the area of interest of normal and blast cells was cropped as the region of interest. All images were in TIFF format with a resolution of  $2,592 \times 1,944$  pixels. The figure demonstrates some sample images of the ALL-IDB2 dataset.

Another dataset prepared by the bone marrow laboratory of Taleqani Hospital, Iran (40) was also used. The dataset consisted of 3,242 peripheral blood smear images belonging to two classes, i.e., benign and malignant. The latter class was further divided into three sub-classes of ALL, i.e., early, precursor B-cell ALL, and pro-B-cell ALL. The images were captured using a Zeiss camera integrated with a microscope setting with  $\times 100$  magnification. The resolution of images was  $224 \times 224$ . Figure 4 shows some sample images of the dataset of (40), whereas Table 2 shows the class distribution.

## 4.2 Proposed system pipeline

In Figure 5, a pipeline is presented for the proposed system. The main steps of computation are discussed in the following.

### 4.2.1 Contrast enhancement

In most of the existing works, image enhancement is mainly accomplished using transforms, points, and spatial operations (41). Among the transforms, various kinds of filtering operations are included such as homomorphic or linear operations. Point operations include contrast enhancement, thresholding, and histogram adjustment. The main limitation of most of these methods is that they perform a global operation on the input image without considering region-specific contrast stretching. Spatial transformation includes neighborhood-based methods such as filtering or masking. These techniques sometimes produce unnecessary noise enhancement of images or increase the smoothness of image regions where sharpness is required (42).

In this work, we performed image contrast stretching using a greedy differential evolution approach, which consisted of the following steps.

1. Convert the input image from RGB image space to HSI image.
2. Perform contrast stretching of the I-channel of the image using the proposed greedy differential approach
3. Convert the HSI image to RGB image space.

The proposed contrast stretching method was based on neighborhood pixel transformations instead of image-wise global operators. Considering an input intensity image  $I$  with dimensions  $M \times N$ , we used the following function for pixel-wise update based on local neighborhood (41). Mathematically, it is formulated by Equation 1.

$$f(m, n) = \left( \frac{\mu \cdot \delta}{\beta + \sigma_p(m, n)} \right) [I(m, n) - \mu_p(m, n)] + \mu_p(m, n)^\alpha \quad \forall m \in M, n \in N, \quad (1)$$

where  $I(m, n)$  is the current pixel value of the intensity image with coordinates  $m$  and  $n$ ,  $\mu$  is the global mean of the image,  $\mu_p(m, n)$  is the local mean, and  $\sigma_p(m, n)$  is the local variance of a window comprising of  $p \times p$  neighborhood pixels around the central pixel  $I$

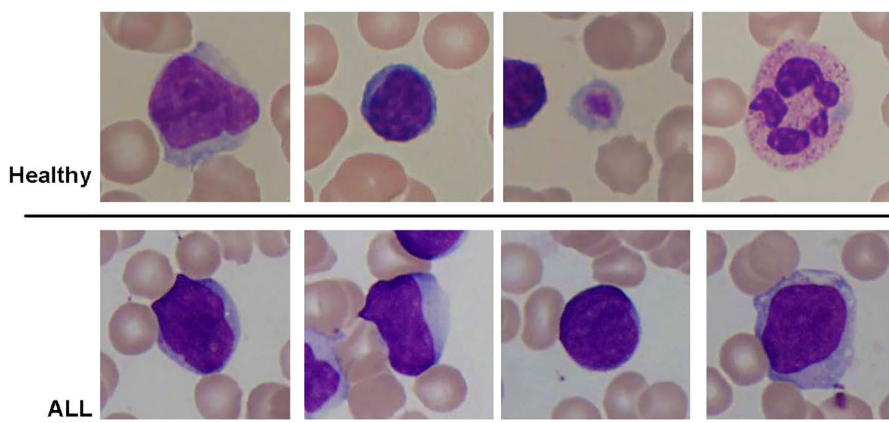


FIGURE 3  
Some samples of images from ALL-IDB2 dataset of Scotti et al. (39) used in this study.

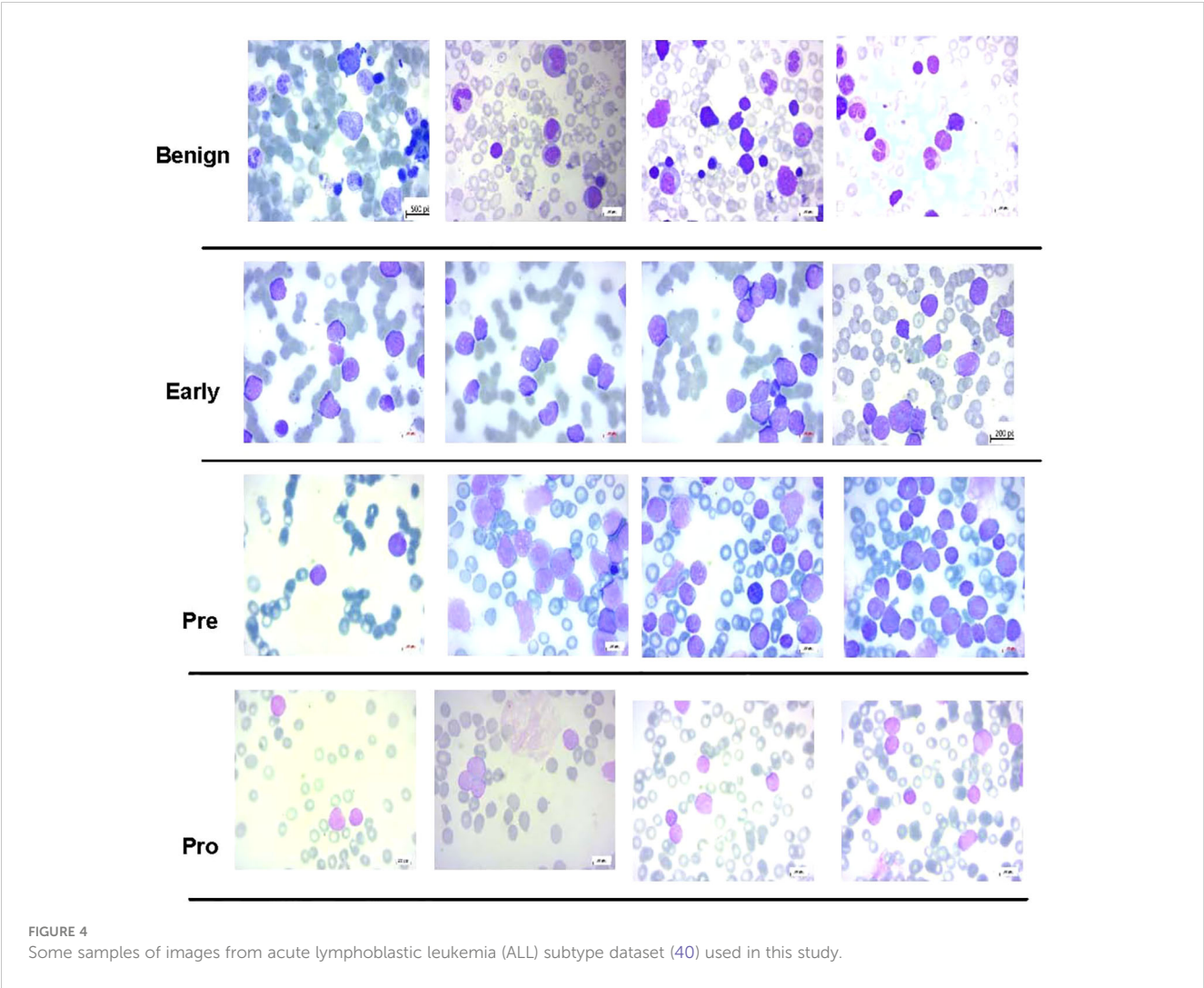


FIGURE 4  
Some samples of images from acute lymphoblastic leukemia (ALL) subtype dataset (40) used in this study.

( $i, j$ ). Munteanu and Rosa (41) used the non-zero constants  $\alpha$ ,  $\beta$ , and  $\gamma$ . The non-zero value of  $\beta$  allows to have zero standard deviation  $\sigma_p(i, j)$  of the pixel window. The final term of Equation 2 was added to achieve smoothness while preserving the brightness.

In this work, we performed automatic estimation of decision variables  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\delta$  using a meta-heuristic algorithm. The following objective function was used in Equation 2:

$$F(I^*) = \frac{\log(\log(G(I^*)) \times n_e(I^*) \times H(I^*))}{M \times N}, \tag{2}$$

TABLE 2 Class distribution of image dataset of Ghaderzadeh et al. (40).

| Class                | No. of images |
|----------------------|---------------|
| Benign               | 512           |
| Precursor B-cell ALL | 955           |
| Pro-B-cell           | 796           |
| Early pre-B          | 979           |

ALL, acute lymphoblastic leukemia.

where  $I^*$  denotes an enhanced intensity image obtained using (2) on the input intensity image, and  $n_e(I^*)$  and  $H(I^*)$ , respectively, denote the number of pixels on the boundary and value of entropy of  $I^*$ . The value of function  $G(I^*)$  was obtained by applying a Sobel kernel on the enhanced intensity image  $I^*$ .

4.2.2 Optimization of decision variables using differential evolution

The estimation of variables  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\delta$  was performed using bounded exploration of search space. First, a population matrix  $P$  consisting of  $N_p$  row vectors was generated, where each vector was composed of four variables, i.e.,  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\delta$ . Each entity of the population matrix was generated randomly, as Equations 3–6 (43).

$$P(i, j) = l_b + r_1 \times (u_b - l_b),$$
$$\forall i \in \{1, \dots, N_p\}, j \in \{1, \dots, 4\}, \tag{3}$$

where  $l_b$  and  $u_b$  are respectively the minimum and maximum values of the decision variable as given in Table 3, and  $r_1$  denotes a random number in  $[0, 1]$ . All vectors of population  $P$  then participated in computing the value of intensity transformation function  $f(i, j)$  of Equation 2 and objective function  $F(I^*)$  of

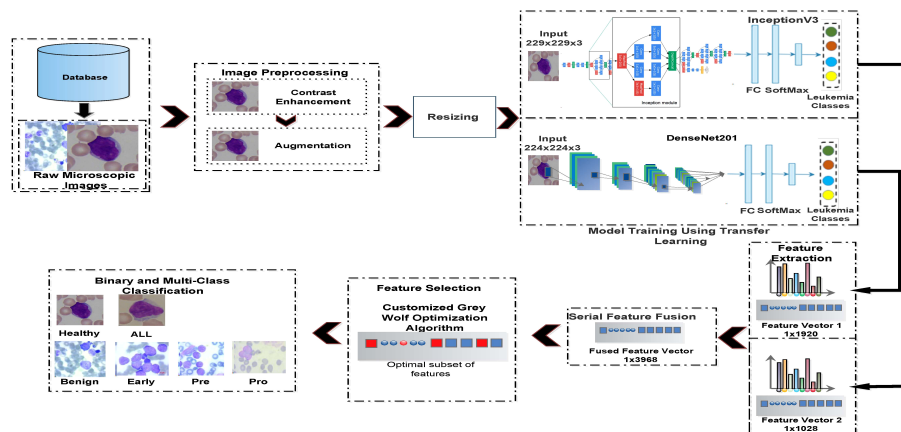


FIGURE 5

Proposed pipeline for binary and multiclass identification of acute lymphoblastic leukemia.

**Equation 2.** The vector yielding the maximum value of the objective function was selected as the population best, i.e.,  $P_b$ .

In the next step, each population vector  $P_i$   $i \in \{1, \dots, N_p\}$  underwent mutation operation to generate its corresponding mutation vector  $M_i$  such that (43)

$$M_i(t) = P_b(t) + A(P_{r1} - P_{r2}) + A(P_{r3} - P_{r4}), \quad (4)$$

where  $t$  denotes the value of the current iteration,  $P_b(t)$  denotes the iteration's best individual vector, and scaling factor  $A$  is a random number in  $[0, 1]$ . The indices of population vectors (i.e.,  $r1$ ,  $r2$ ,  $r3$ , and  $r4$ ) were randomly chosen such that they are all distinct from each other and the index  $i$  of the current population vector.

The population vector  $P_i$  and its corresponding mutation vector  $M_i$  then underwent binomial crossover operation to generate vector  $C_i$  such that (43)

$$C_i(t) = \begin{cases} P_i(t) & i \neq x_1 \quad r > x_2 \\ C_i(t) & i = x_1 \quad \text{or} \quad r \leq x_2 \end{cases}, \quad (5)$$

where  $x_1$  is a randomly generated index within the interval  $[1, N_p]$  and  $x_2$  is a random number in  $[0, 1]$ . Next, all decision variables each vector  $C_i(t)$  are bounded within limits  $l_b$  and  $u_b$ . In differential evolution, a greedy selection of survival of the fittest was carried out using the following criteria to update the population matrix (43).

**TABLE 3** Minimum and maximum values of decision variables used in differential evolution.

| Decision variable | Minimum value ( $l_b$ ) | Maximum value ( $u_b$ ) |
|-------------------|-------------------------|-------------------------|
| $\alpha$          | 0                       | 1.6                     |
| $\beta$           | 0                       | 0.5                     |
| $\gamma$          | 0                       | 0.8                     |
| $\delta$          | 0.5                     | 1.5                     |

$$P_i(t+1) = \begin{cases} P_i(t) & \text{if } F(C_i) < F(P_i) \\ C_i(t) & \text{otherwise} \end{cases}, \quad (6)$$

where  $F(P_i)$  and  $F(C_i)$  denote the cost of the objective function (2) using the vectors  $P_i$  and  $C_i$ , respectively. This completes one iteration of the algorithm, which was repeated for  $T$  iterations.

#### 4.2.3 Feature extraction

The contrast-enhanced images of datasets were used in the feature extraction step. For this purpose, we employed transfer learning as a feature extraction using two well-known deep CNNs, namely, InceptionV3 and ResNetV2.

InceptionV3 is a deep CNN that belongs to the Inception family of CNNs. It is pre-trained on the ImageNet database (44) consisting of 1,000 object classes. The network has 316 layers and 350 connections (45). The size of the first layer, i.e., the input layer, is  $299 \times 299 \times 3$ . A convolution layer consists of different filters and stride sizes. Each convolution layer also incorporates batch normalization and ReLU layers for adding non-linearity. A pooling layer is also added between the convolution layers to obtain active neurons. The addition of Inception modules is a distinguishing characteristic of this network. They are designed for multiscale feature extraction. Each inception module is built using multiple parallel convolution layers with different filter sizes and a pooling layer. The outputs of these layers are concatenated along the depth dimension.

To use InceptionV3 for feature extraction, its last learnable layer, "predictions", was replaced with a fully connected layer, which had outputs matching the number of classes of our datasets [i.e., two classes for ALL-IDB2 and four classes for the dataset of (40)]. The softmax layer named "predictions softmax" was replaced by the new softmax layer. A label-free classification layer was added to the network, which replaced the "ClassificationLayer predictions" layer. Prior to the network training, the dataset image resizing was performed per the requirement of the network's input layer. Then, specific augmentation steps were performed. The activations were applied

on the “avg pool” layer, and a feature vector of length 2,048 was obtained per image of the training and testing datasets.

DenseNet201 (46) has a depth of 201 layers and was initially trained on the ImageNet (44) dataset. The DenseNet201 is based on the idea of layer concatenation; i.e., each layer obtains data from all of its previous layers and transfers its computed feature maps to all its next layers. As a result, a thinner and more compact network is realized, which is computationally efficient and achieves considerable memory savings.

To use DenseNet201 as a feature extractor, its “fc1000” layer was substituted with a fully connected one that contains an equal number of classes from our datasets. Similarly, a new softmax and classification layer without labels was inserted in the network in place of their respective layers. After performing image resizing and augmentation, feature activation was applied on the global average pool layer, and a feature vector of size 1,920 was extracted per image of the dataset.

#### 4.2.4 Feature fusion

In this work, we performed a simple horizontal concatenation of the individual feature vectors extracted from the abovementioned deep CNNs and formed a fused feature vector of size  $1 \times (a + b)$ , where  $a = 2,048$  and  $b = 1,920$  are the number of features extracted from InceptionV3 and DenseNet201 networks, respectively.

#### 4.2.5 Meta-heuristic for feature selection

The combined feature vector obtained from the transfer learning steps above has a considerably large size. Directly using the extracted fused feature vector to train the baseline classifiers requires a huge amount of processing power and memory. As a main contribution, this work modeled the optimization problem of feature selection with the objective of maximizing classification accuracy with the minimum feature set. The optimization problem was then solved with the help of a population-based meta-heuristic named Grey Wolf Algorithm. This technique (47) mimics the hunting behavior of grey wolves. A pack of grey wolf apex predators consists of an average of five to 12 individuals. The grey wolf population is composed of four distinct individuals categorized as alpha, beta, delta, and omega, based on their dominant order. The alpha wolf is the individual within a population that holds the highest rank and assumes the role of decision-maker and dominant figure. The subsequent position in the dominance hierarchy is occupied by the beta wolf. It is subordinate to the alpha and helps in the decision-making. The delta wolf ranks third in the hierarchy and only dominates the least significant omega group.

In the mathematical framework of GWO, the most optimal solution is referred to as the alpha wolf ( $\alpha$ ). Subsequently, the second and third most optimal solutions are denoted as the beta ( $\beta$ ) and delta ( $\delta$ ) wolves, respectively. The main steps of grey wolf hunting are as follows:

- Search the prey (exploration).
- Encircle the prey.
- Attack the prey (exploitation).

The prey encircle behavior of a pack of wolves is mathematically modeled as defined in Equations 7–18.

$$D = |C \cdot X_p - V(t)|, \quad (7)$$

$$V(t+1) = V_p(t) - A \cdot D, \quad (8)$$

where  $V_p$  denotes the prey vector position at iteration  $t$ ,  $V(t)$  is the current position of the grey wolf,  $A$  and  $C$  are the vectors of coefficients:

$$A = 2 \cdot a \cdot x_1 - a, \quad (9)$$

$$C = 2 \cdot x_2 \quad (10)$$

where  $x_1$  and  $x_2$  are vectors containing random vectors in  $[0, 1]$ , and  $a$  is the encircling coefficient that mimics the encircling behavior by decreasing linearly from 2 to 0, which is linearly decreased from 2 to 0 with iterations as (47)

$$a = 2 - 2 \left( \frac{t}{t_{max}} \right) \quad (11)$$

where  $t_{max}$  is the maximum number of algorithm iterations. During an iteration  $t$ , each wolf updates its position using the  $\alpha$ ,  $\beta$ , and  $\delta$  wolves such that (47)

$$V(t+1) = \frac{V_1 + V_2 + V_3}{3} \quad (12)$$

$$V_1 = |V_\alpha - A_1 \cdot D_\alpha| \quad (13)$$

$$V_2 = |V_\beta - A_1 \cdot D_\beta| \quad (14)$$

$$V_3 = |V_\delta - A_1 \cdot D_\delta| \quad (15)$$

where  $V_\alpha$ ,  $V_\beta$ , and  $V_\delta$  denote the position vectors of  $\alpha$ ,  $\beta$ , and  $\delta$  wolves, respectively, at iteration  $t$ ;  $A_1$ ,  $A_2$ , and  $A_3$  are computed using Equation 9. The vectors  $D$ ,  $D_\beta$ , and  $D_\delta$  are computed as

$$D_\alpha = |C_1 V_\alpha - V| \quad (16)$$

$$D_\beta = |C_1 V_\beta - V| \quad (17)$$

$$D_\delta = |C_1 V_\delta - V| \quad (18)$$

The coefficients  $C_1$ ,  $C_2$ , and  $C_3$  are computed using Equation 10. The original GWO algorithm of Mirjalili et al. (47) is generally applicable to continuous optimization problems with variables  $X \in \mathcal{R}$ .

##### 4.2.5.1 Binary Grey Wolf Algorithm

This work used a binary GWO algorithm of Emery et al. (48), in which the position update of the wolf is determined using the crossover operation of individual genes, and mathematically formulated by Equations 19–23.

$$V(t+1) = (V_1 \otimes V_2 \otimes V_3) \quad (19)$$



where  $V_1$ ,  $V_2$ , and  $V_3$  are binary vectors for dimension  $d_{max}$  and computed as

$$V_i^d = \begin{cases} 1, & \text{if } (\Delta_j^d + V_j^d) \geq 1 \\ 0, & \text{otherwise} \end{cases}, \forall d \in D \quad (20)$$

For  $i = 1, 2, 3$ ,  $V_j^d$  is equal to  $V_\alpha$ ,  $V_\beta$ , and  $V_\delta$ , whereas  $\Delta_j^d$  is equal to  $\Delta_\alpha^d$ ,  $\Delta_\beta^d$ , and  $\Delta_\delta^d$  respectively.  $\Delta_j^d$  is computed as (48):

$$\Delta_j^d = \begin{cases} 1, & \text{if } S_j^d \geq r_4 \\ 0, & \text{otherwise} \end{cases} \quad (21)$$

where  $r_4$  is the vector of random numbers in  $[0, 1]$ . The continuous step size  $S_j^d$  is computed as (48)

$$S_j^d = \frac{1}{\exp(-10(A_1^d \cdot D_j^d - 0.5)) + 1} \quad (22)$$

$D_j^d$  is equal to  $D_\alpha$ ,  $D_\beta$ , and  $D_\delta$ , respectively, for  $i = 1, 2$ , and  $3$ .  $A_1^d$  is computed using Equation 9, whereas  $D_j^d$  is computed using Equations 16–18. The new position of the wolf is updated using the following crossover operation (48).

$$V^d(t+1) = \begin{cases} V_1^d, & \text{if } r_6 < \frac{1}{3} \\ V_2^d, & \text{if } \frac{1}{3} \leq r_6 < \frac{2}{3} \\ V_3^d, & \text{otherwise} \end{cases} \quad (23)$$

where  $r_6$  is a random variable that follows a uniform distribution in the interval  $[0, 1]$ .

#### 4.2.5.2 Wrapper feature selection using binary GWO

This study presents the application of the binary GWO method for the purpose of deep feature selection within the leukemia classification pipeline. The computational steps of the suggested feature selection strategy are presented in Algorithm 1.

The main inputs to the binary GWO algorithm include the fused feature matrix  $\mathbb{F}$ , the vector  $L$ , which contains the labels of the training image set; the maximum count of iterations  $t_{max}$  the size of grey wolf population  $n_p$ ; and dimension size  $d_{max}$ , which represents the total number of variables (features) of each wolf (solution) of population. The size of matrix  $\mathbb{F}$  is  $n_t \times d_{max}$ , where  $n_t$  and  $d_{max}$  respectively denote the number of training images and the dimension of fused feature vector per image.

Phase 1 initializes the main parameters including iteration counter  $t$ , and alpha, beta, and delta grey wolves  $X_\alpha$ ,  $X_\beta$ , and  $X_\delta$  along with their fitness values  $f_\alpha$ ,  $f_\beta$ , and  $f_\delta$  respectively. In Phase 2, an initial population is generated and stored in matrix  $\mathbb{X}$  of size  $n_p \times d_{max}$ . The  $randn(1, n_p, 1: d_{max})$  function generates a matrix of dimensions  $n_p \times d_{max}$  of binary values of uniform distribution in  $[0, 1]$ . The execution phase of the GWO algorithm proceeds in Steps 6–75. The while loop is executed for  $t_{max}$  iterations. In an iteration, first, a prey is extracted from the population matrix (Step 8), and its fitness is evaluated (Step 9). The *Fitness* function receives three inputs, namely, the fused feature set  $\mathbb{F}$ , the vector  $L$  of labels, and one member of the population, i.e., a binary vector  $X$ . In the *Fitness* function routine, Steps 85–86 obtain the features from  $\mathbb{F}$ , which are indexed by non-zero values of  $X$ . The updated feature matrix  $\mathbb{F}_2$  is

then divided into testing and training parts. In Steps 87–93 of the *Fitness* function, the classification error of the K-nearest neighbor (KNN) classifier is used as a fitness value (cost). This value is then used to update the alpha, beta, and delta  $X_\alpha$ ,  $X_\beta$ , and  $X_\delta$  grey wolf vectors, respectively, in Steps 10–23 of the main function. Steps 26–74 of the main routine perform the position update of each grey wolf of the population according to Equations 19, 20 of the binary GWO algorithm. After the execution of the while loop for  $t_{max}$  iterations, the global best solution, i.e., alpha wolf  $X_\alpha$ , contains the indices of features to be selected from the fused feature vector.

1: **External Inputs:**  $\mathbb{F}$ ,  $L$ ,  $d_{max}$ ,  $t_{max}$ ,  $n_p$

2: **Phase 1: Initialization of Main Parameters**  $t \leftarrow 1$ ,

$V_\alpha(1: d_{max}) \leftarrow 0$ ,  $f_\alpha \leftarrow \infty$

$V_\beta(1: d_{max}) \leftarrow 0$ ,  $f_\beta \leftarrow \infty$

$V_\delta(1: d_{max}) \leftarrow 0$ ,  $f_\delta \leftarrow \infty$

3: **Phase 2: Generate Initial Population of Grey Wolves**

4:  $(1: n_p, 1: d_{max}) \leftarrow randn(1: n_p, 1: d_{max})$

5: **Execution**

6: **while**  $i < t_{max}$  **do**

7:   **for**  $j = 1: n_p$  **do**

8:      $V \leftarrow (j, 1: d_{max})$

9:      $f \leftarrow Fitness(\mathbb{F}, L, V)$

10:    **if**  $f < f_\alpha$  **then**

11:      $V_\beta \leftarrow V_\alpha$

12:      $f_\beta \leftarrow f_\alpha$

13:      $V_\alpha \leftarrow V$

14:      $f_\alpha \leftarrow f$

15:    **else if**  $f < f_\beta$  **then**

16:      $V_\delta \leftarrow V_\beta$

17:      $f_\delta \leftarrow f_\beta$

18:      $V_\beta \leftarrow V$

19:      $f_\beta \leftarrow f$



```

20:   else
21:        $V_{\delta} \leftarrow V$ 
22:        $f_{\delta} \leftarrow f$ 
23:   end if
24: end for
25: Population Update
26: for  $j = 1 : n_p$  do
27:   for  $d = 1 : d_{max}$  do
28:        $a \leftarrow 2 - 2 \frac{j}{t_{max}}$ 
29:        $a_1 \leftarrow 2 \cdot a \cdot rand(1, 1) - a$ 
30:        $a_2 \leftarrow 2 \cdot a \cdot rand(1, 1) - a$ 
31:        $a_3 \leftarrow 2 \cdot a \cdot rand(1, 1) - a$ 
32:        $c_1 \leftarrow 2 \cdot rand(1, 1) - a$ 
33:        $c_2 \leftarrow 2 \cdot rand(1, 1) - a$ 
34:        $c_3 \leftarrow 2 \cdot rand(1, 1) - a$ 
35:        $D_{\alpha}(1, d) \leftarrow |c_1 \cdot V_{\alpha}(1, d) - \mathbb{V}(j, d)|$ 
36:        $D_{\beta}(1, d) \leftarrow |c_1 \cdot V_{\alpha}(1, d) - \mathbb{V}(j, d)|$ 
37:        $D_{\delta}(1, d) \leftarrow |c_1 \cdot V_{\alpha}(1, d) - \mathbb{V}(j, d)|$ 
38:        $S_1(1, d) \leftarrow \frac{1}{1 + \exp(-10(a_1 \cdot D_{\alpha}(1, d) - 0.5))}$ 
39:        $S_2(1, d) \leftarrow \frac{1}{1 + \exp(-10(a_2 \cdot D_{\beta}(1, d) - 0.5))}$ 
40:        $S_3(1, d) \leftarrow \frac{1}{1 + \exp(-10(a_3 \cdot D_{\delta}(1, d) - 0.5))}$ 
41:       if  $S_1(1, d) \geq randn(1, 1)$  then
42:            $\Delta_1(1, d) \leftarrow 1$ 
43:       else
44:            $\Delta_1(1, d) \leftarrow 0$ 
45:       end if
46:       if  $S_2(1, d) \geq randn(1, 1)$  then
47:           else
48:                $\Delta_2(1, d) \leftarrow 0$ 
49:           end if
50:       if  $S_3(1, d) \geq randn(1, 1)$  then
51:            $\Delta_3(1, d) \leftarrow 1$ 
52:       else
53:            $\Delta_3(1, d) \leftarrow 0$ 
54:       end if
55:       if  $(V_{\alpha}(1, d) + \Delta_1(1, d)) \geq 1$  then
56:            $Z_1(1, d) \leftarrow 1$ 
57:       else
58:            $Z_1(1, d) \leftarrow 0$ 
59:       end if
60:       if  $(V_{\beta}(1, d) + \Delta_2(1, d)) \geq 1$  then
61:            $Z_2(1, d) \leftarrow 1$ 
62:       else
63:            $Z_2(1, d) \leftarrow 0$ 
64:       end if
65:       if  $(V_{\delta}(1, d) + \Delta_3(1, d)) \geq 1$  then
66:            $Z_3(1, d) \leftarrow 1$ 
67:       else
68:            $Z_3(1, d) \leftarrow 0$ 
69:       end if
70:        $r \leftarrow rand(1, 1)$ 
71:       if  $r < \frac{1}{3}$  then
72:            $\mathbb{V}(j, d) \leftarrow Z_1(1, d)$ 
73:       else if  $r \geq \frac{1}{3}$  and  $r < \frac{2}{3}$  then
74:            $\mathbb{V}(j, d) \leftarrow Z_2(1, d)$ 
75:       else
76:            $\mathbb{V}(j, d) \leftarrow Z_3(1, d)$ 
77:       end if
78:   end do
79: end do

```

```

70:     else
71:          $V(j, d) \leftarrow Z_3(1, d)$ 
72:     end if
73: end for
74: end for
75: end while
76: Select Features
77:  $I \leftarrow 1 : d_{max}$ 
78:  $S_F \leftarrow I((V_\alpha == 1))$ 
    OUTPUT:  $S_F$ 
79: Function: Fitness
80: Inputs:  $V, L, \mathbb{F}$ 
81: Parameters:  $k = 5, h_o = 0.2, \alpha_1 = 0.99, \alpha_2 = 0.01$ 
82: if ( $\text{sum}(V == 1) == 0$ ) then
83:      $\Gamma = 1$ 
84: else
85:      $\mathbb{F}_2 \leftarrow \mathbb{F}(:, (V == 1))$ 
86:      $A_{train}, L_{train}, A_{test}, L_{test} \leftarrow \text{partition}(\mathbb{F}_2, L, h_o)$ 
87:      $Model \leftarrow \text{trainKNN}(A_{train}, L_{train}, k)$ 
88:      $L_{pred} \leftarrow \text{predict}(Model, A_{test})$ 
89:      $a \leftarrow \text{sum}(L_{pred} == L_{test}) / \text{length}(L_{test})$ 
90:      $e \leftarrow 1 - a$ 
91:      $q_s \leftarrow \text{sum}(a == 1)$ 
92:      $q_t \leftarrow \text{length}(V)$ 
93:      $\Gamma \leftarrow \alpha_1 \times e + \alpha_2 \times (\frac{q_s}{q_t})$ 
94: end if
95: Return:  $\Gamma$ 

```

ALGORITHM 1. Feature selection based on binary GWO algorithm.

#### 4.2.6 Classification

The set of selected features from the binary GWO algorithm was then used along with the label  $L$  for training and classification of outer classifiers. Multiple classifiers were used in this work, and the best-performing classifiers were selected.

## 5 Results and discussion

The proposed decision support system for leukemia identification was implemented on an Intel Core i5 CPU with and 64-bit Windows 10 operating system and 16GB RAM.

### 5.1 Experiment 1: binary classification

First, the proposed pipeline was implemented for binary detection of leukemia using the ALL-IDB2 dataset. The classification performance of CNN was influenced by the quality and size of the training dataset. A small dataset leads to overfitting and poor generalization of the model. Hence, augmentation of contrast stretched ALL-IDB2 dataset was performed using the operations of random rotation, flipping, intensity modification, and brightness correction. Table 4 shows the class distribution of ALL-IDB2 as a result of augmentation. In the next step, the augmented dataset was divided into training and test parts with a 70:30 split ratio, as shown in Table 5. Then, the training dataset was used for transfer learning of InceptionV3 and DenseNet201 models with parameters listed in Table 6.

InceptionV3 and DenseNet201 return deep feature vectors of sizes 2,048 and 1,920, respectively, which are horizontally concatenated to obtain a fused feature vector of size 3,968. This vector is then subjected to the proposed feature selection step using the GWO algorithm. After a fixed number of iterations, the GWO algorithm returns its best solution, i.e., a reduced vector of the most

TABLE 4 Class distribution of ALL-IDB2 dataset before and after augmentation.

| Class   | Frequency |       |
|---------|-----------|-------|
|         | Before    | After |
| Healthy | 130       | 593   |
| ALL     | 130       | 601   |

ALL, acute lymphoblastic leukemia.

TABLE 5 Training and testing ALL-IDB2 dataset for binary classification of leukemia.

| Class   | Training Images | Testing Images | Total |
|---------|-----------------|----------------|-------|
| Healthy | 415             | 178            | 593   |
| ALL     | 420             | 181            | 601   |
| Total   | 835             | 359            | 1,194 |

ALL, acute lymphoblastic leukemia.

TABLE 6 Parameter settings for training of InceptionV3 and DenseNet201 models.

| Parameter             | Value              | Parameter    | Value |
|-----------------------|--------------------|--------------|-------|
| Kernel type           | sdgm               | Max epochs   | 10    |
| Initial learning rate | $1 \times 10^{-4}$ | Environment  | Auto  |
| Validation frequency  | 30                 | Stride size  | 1     |
| Mini-batch size       | 20                 | Dropout rate | 0.1   |

important selected features, which are then used to train several baseline classifiers with multiple settings of their kernel. Table 7 shows the performance results of the proposed binary classification pipeline. The KNN classifier with cosine kernel achieves the best performance metrics with a reduced feature vector of 797 features, which is approximately 80% smaller than the original fused feature vector of size 3,986. The confusion matrix of the KNN cosine classifier is demonstrated in Table 8.

In Figure 6, the error rate of the GWO algorithm is plotted along with the standard genetic algorithm (GA), as a function of iterations with a constant value of population size  $n_p = 20$ . A better convergence behavior is demonstrated by the GWO algorithm as compared to GA, which reveals that GWO performs better exploration of feature search space.

5.2 Experiment 2: leukemia subtype classification

In the second step, the proposed pipeline was implemented for multiclass problems, i.e., leukemia subtype identification using the

dataset of Ghaderzadeh et al. (40). As discussed earlier, the dataset consists of four classes, i.e., benign, precursor, pro-B-cell, and early pre-B. Following the contrast stretching phase (Section 4.2.1) on the dataset, the augmentation was carried out using the same methodology as binary classification. The class distribution of the augmented dataset is shown in Table 9. Next, with a splitting ratio of 70:30, the training and testing parts of the dataset were extracted as shown in Table 10.

In Table 11, the leukemia subtype classification results are presented. Several classifiers with various kernels were tested. The performance metrics, i.e., accuracy, precision, recall, sensitivity, specificity, and F1 score, were computed through macro averaging of the individual class metrics. In this case, the SVM classifier with Gaussian kernel achieved the best average accuracy of 98.05%, whereas the maximum average accuracy values achieved by KNN, decision tree, and neural network (NN) classifiers were 97.9%, 82.4%, and 95.8%, respectively. The testing confusion matrix with the SVM Gaussian classifier is demonstrated in Table 12, whereas the class-wise statistics are mentioned in Table 13. The maximum accuracy of 98.66% was achieved by the pro-B-cell class, whereas all other classes achieved an accuracy of above 90%.

TABLE 7 Results of binary classification of leukemia on ALL-IDB2 dataset.

| Classifier             | Selected feature vector size | Accuracy % | Sensitivity | F1 score | Precision | Recall |
|------------------------|------------------------------|------------|-------------|----------|-----------|--------|
| KNN cosine             | 797                          | 98.1       | 0.981       | 0.987    | 0.98      | 0.987  |
| KNN coarse             |                              | 97.8       | 0.971       | 0.981    | 0.972     | 0.972  |
| KNN cubic              |                              | 97.9       | 0.981       | 0.976    | 0.972     | 0.972  |
| KNN fine               |                              | 97.5       | 0.964       | 0.965    | 0.989     | 0.977  |
| SVM (regression)       |                              | 85.2       | 0.887       | 0.842    | 0.890     | 0.80   |
| SVM (Gaussian)         |                              | 86.4       | 0.894       | 0.891    | 0.901     | 0.86   |
| SVM (quadratic)        |                              | 72.2       | 0.734       | 0.741    | 0.882     | 0.72   |
| Decision tree (medium) |                              | 72.4       | 0.742       | 0.725    | 0.73      | 0.726  |
| NN wide                |                              | 94.8       | 0.925       | 0.911    | 0.932     | 0.951  |

KNN, K-nearest neighbor; SVM, support vector machine; NN, neural network.

TABLE 8 Confusion matrix of binary classification experiment of ALL-IDB2 with KNN-cosine classifier.

|            |         | Predicted class |         | TPR   | FNR  |
|------------|---------|-----------------|---------|-------|------|
|            |         | ALL             | Healthy |       |      |
| True Class | ALL     | 177             | 4       | 97.7% | 2.2% |
|            | Healthy | 2               | 176     | 98.8% | 1.2% |

TPR, true-positive rate; FNR, false-negative rate; KNN, K-nearest neighbor; ALL, acute lymphoblastic leukemia.

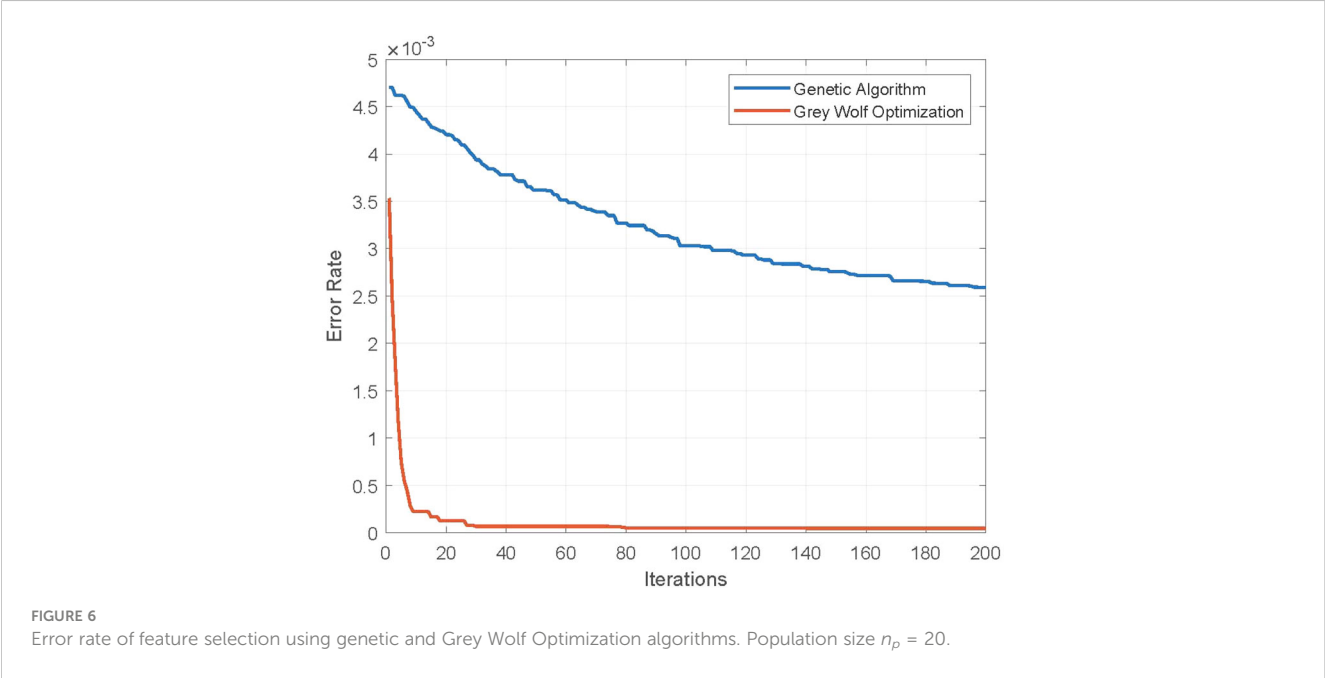


TABLE 9 Class distribution of dataset of Ghaderzadeh et al. (40) before and after augmentation.

| Class       | Frequency |       |
|-------------|-----------|-------|
|             | Before    | After |
| Benign      | 512       | 1,024 |
| Precursor   | 955       | 1,000 |
| Pro-B-cell  | 796       | 1,050 |
| Early pre-B | 979       | 1,020 |

TABLE 10 Class distribution of training and testing parts of dataset of Ghaderzadeh et al. (40) for leukemia subtype classification.

| Class       | Training images | Testing images | Total |
|-------------|-----------------|----------------|-------|
| Benign      | 716             | 308            | 1,024 |
| Precursor   | 700             | 300            | 1,000 |
| Pro-B-cell  | 735             | 315            | 1,050 |
| Early pre-B | 714             | 306            | 1,020 |
| Total       | 2,149           | 1,229          | 4,094 |

TABLE 11 Results of leukemia subtype classification using the dataset of Ghaderzadeh et al. (40).

| Classifier             | Selected feature vector size | Accuracy % | Sensitivity | F1 score | Precision | Recall |
|------------------------|------------------------------|------------|-------------|----------|-----------|--------|
| KNN cosine             | 797                          | 97.9       | 0.783       | 0.78     | 0.978     | 0.783  |
| KNN coarse             |                              | 97.6       | 0.781       | 0.78     | 0.976     | 0.781  |
| KNN cubic              |                              | 97.4       | 0.709       | 0.71     | 0.974     | 0.974  |
| KNN fine               |                              | 96.4       | 0.771       | 0.771    | 0.96      | 0.771  |
| SVM (regression)       |                              | 85.2       | 0.721       | 0.642    | 0.890     | 0.659  |
| SVM (Gaussian)         |                              | 98.14      | 0.785       | 0.78     | 0.981     | 0.785  |
| SVM (quadratic)        |                              | 96.2       | 0.638       | 0.61     | 0.96      | 0.61   |
| Decision tree (medium) |                              | 82.4       | 0.68        | 0.689    | 0.82      | 0.826  |
| NN wide                |                              | 95.8       | 0.71        | 0.86     | 0.952     | 0.086  |

KNN, K-nearest neighbor; SVM, support vector machine; NN, neural network.

Table 14 presents an accuracy comparison of three feature extraction methods, i.e., a full set of deep features extracted from InceptionV3 and DenseNet201 CNNs, a reduced set of features selected by the proposed GWO algorithm, and a standard genetic algorithm. The table demonstrates that the GWO algorithm achieves a better or comparable accuracy as compared to the other two feature selection methods with a significantly small feature set.

In Table 15, a comparison is presented of the performance of our proposed method with some existing studies on leukemia identification. For a fair comparison, we selected the published studies that have used identical or almost similar datasets. Our proposed pipeline for leukemia binary detection and subtype identification achieves better or comparable performance metrics as compared to several other relevant studies with smaller feature sizes. This shows the validity and applicability of the proposed approach.

## 6 Conclusion

Leukemia, a kind of hematologic malignancy, is frequently diagnosed in both pediatric and geriatric populations. An automated, computer-aided system of leukemia diagnosis is essential to aid medical

TABLE 12 Confusion matrix of leukemia subtype identification using SVM Gaussian classifier on dataset of Ghaderzadeh et al. (40).

|            |            | Predicted class |           |            |           | TPR    | FNR   |
|------------|------------|-----------------|-----------|------------|-----------|--------|-------|
|            |            | Benign          | Precursor | Pro-B-cell | Early pre |        |       |
| True Class | Benign     | 1,004           | 11        | 5          | 4         | 98.04% | 1.9%  |
|            | Precursor  | 5               | 982       | 11         | 2         | 98.2%  | 1.8%  |
|            | Pro-B-cell | 2               | 2         | 1,036      | 10        | 98.6%  | 1.4%  |
|            | Early pre  | 6               | 13        | 5          | 996       | 97.67% | 2.35% |

TPR, true-positive rate; FNR, false-negative rate; SVM, support vector machine.

TABLE 13 Statistics of individual classes using SVM Gaussian kernel.

| Class      | Accuracy % | Sensitivity % | Precision % | Recall |
|------------|------------|---------------|-------------|--------|
| Benign     | 98.04      | 98.72         | 98.72       | 0.9804 |
| Precursor  | 98.2       | 97.42         | 97.42       | 0.982  |
| Pro-B-cell | 98.66      | 98.01         | 98.01       | 0.986  |
| Early pre  | 97.64      | 98.41         | 98.41       | 0.976  |

SVM, support vector machine.

TABLE 14 Performance comparison of leukemia classification using three feature selection approaches, i.e., proposed GWO feature selection algorithm, feature selection using genetic algorithm, and full feature set.

| Classifier           | Full feature set |            | Genetic algorithm |            | Grey Wolf Algorithm |            |
|----------------------|------------------|------------|-------------------|------------|---------------------|------------|
|                      | No. of features  | Accuracy % | No. of features   | Accuracy % | No. of features     | Accuracy % |
| KNN cosine           | 3,986            | 97.2       | 1,520             | 96.2       | 797                 | 97.9       |
| KNN coarse           |                  | 96.9       |                   | 94.3       |                     | 97.6       |
| KNN cubic            |                  | 98.1       |                   | 97.2       |                     | 97.4       |
| KNN fine             |                  | 95.1       |                   | 96.1       |                     | 96.4       |
| SVM Gaussian         |                  | 98.5       |                   | 97.58      |                     | 98.14      |
| SVM regression       |                  | 90.2       |                   | 89.2       |                     | 85.2       |
| NN wide              |                  | 96.2       |                   | 94.5       |                     | 95.8       |
| Decision tree medium |                  | 84         |                   | 81.2       |                     | 82.4       |

KNN, K-nearest neighbor; SVM, support vector machine; NN, neural network.

TABLE 15 Comparison of classification accuracy of proposed leukemia identification pipeline with some existing relevant works.

| Work      | Proposed method                                                                                                          | Dataset                     | Disease type                   | Performance result                      |
|-----------|--------------------------------------------------------------------------------------------------------------------------|-----------------------------|--------------------------------|-----------------------------------------|
|           | Classification: SVM, ANN                                                                                                 |                             |                                |                                         |
| (24)      | Multiscale blob detection<br>deep feature extraction: AlexNet classification: SVM                                        | ALL-IDB                     | ALL                            | Accuracy = 94.1%                        |
| (25)      | Preprocessing<br>segmentation: three-phase filtering morphological feature extraction                                    | Self-collected              | ALL                            | Specificity = 93.5%                     |
| (49)      | Active contours for nucleus detection<br>Shape and texture feature extraction classification: NN, SVM                    | Self-collected              | Leukemia                       | Accuracy = 98.8%                        |
| (50)      | Preprocessing<br>Feature extraction: hybrid CNN<br>Classification: bagging ensemble                                      | ALL-IDB<br>MiMMSBI<br>SN-AM | ALL<br>AML<br>Multiple myeloma | ALL classification accuracy<br>= 97.04% |
| This work | Contrast stretching using DE<br>Deep feature extraction: InceptionV3 and DenseNet201 feature selection:<br>GWO algorithm | ALL-IDB2<br>(40)            | Leukemia<br>ALL subtypes       | Accuracy = 97.9% Accuracy<br>= 98.14%   |

KNN, K-nearest neighbor; ANN, artificial neural network; ALL, acute lymphoblastic leukemia; NN, neural network; AML, acute myeloid leukemia; GWO, Grey Wolf Optimization.



professionals in making informed decisions about the disease and making an effective prognosis and treatment plan. In this work, we have demonstrated the effectiveness of deep feature optimization taking as a relevant design case, the detection, and classification of leukemia disease from blood smear images. We have proposed a hybrid deep learning methodology utilizing transfer learning as feature extraction. The problem of feature selection has been modeled as a combinatorial optimization problem and solved using a customized Grey Wolf Optimization algorithm. Our proposed leukemia identification system can be used as a supporting evidence tool in conjunction with other more detailed analysis methods such as RNA sequencing and molecular testing. We believe that the proposed expert system can also be integrated with more complex and rather practical image analysis systems such as image flow cytometry.

## 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.

## Author contributions

MA: Conceptualization, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing. MNA: Methodology, Software, Writing – review & editing. TA: Investigation, Methodology, Validation, Writing – review & editing. AA: Resources, Supervision, Writing – review & editing. MM: Resources, Supervision, Writing – review & editing. AM: Funding acquisition, Resources, Validation, Writing – review & editing.

## References

- Farag MR, Alagawany M. Erythrocytes as a biological model for screening of xenobiotics toxicity. *Chemico-biol. Interact.* (2018) 279:73–83. doi: 10.1016/j.cbi.2017.11.007.
- Rezatofighi SH, Soltanian-Zadeh H. Automatic recognition of five types of white blood cells in peripheral blood. *Computer. Med Imaging Graphics.* (2011) 35:333–43. doi: 10.1016/j.compmedimag.2011.01.003.
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* (2015) 521:436–44. doi: 10.1038/nature14539.
- Bengio Y, Courville A, Vincent P. Representation learning: A review and new perspectives. *IEEE Trans Pattern Anal Mach Intell.* (2013) 35:1798–828. doi: 10.1109/TPAMI.2013.50.
- Nguyen LD, Lin D, Lin Z, Cao J. (2018). Deep cnns for microscopic image classification by exploiting transfer learning and feature concatenation, in: *2018 IEEE international symposium on circuits and systems (ISCAS)*, Vol. 1–5. IEEE.
- Ahamed KU, Islam M, Uddin A, Akhter A, Paul BK, Yousuf MA, et al. A deep learning approach using effective preprocessing techniques to detect covid-19 from chest ct-scan and x-ray images. *Comput Biol Med.* (2021) 139:105014. doi: 10.1016/j.combiomed.2021.105014.
- He K, Zhang X, Ren S, Sun J. (2016). Deep residual learning for image recognition, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, (IEEE). pp. 770–8.
- Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861.* (2017).
- Redmon J, Farhadi A. *arXiv preprint arXiv:1804.02767.* (2018).
- Chollet F. (2017). Xception: Deep learning with depthwise separable convolutions, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, (IEEE). pp. 1251–8.
- Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging.* (2016) 35:1285–98. doi: 10.1109/TMI.2016.2511012.
- Biswas S, Mostafiz R, Paul BK, Uddin KMM, Rahman MM, Shariful F. Dfu multinet: A deep neural network approach for detecting diabetic foot ulcers through multi-scale feature fusion using the dfu dataset. *Intelligence-Based Med.* (2023), 100128. doi: 10.1016/j.ibmed.2023.100128.
- Sanei S, Lee TK. (2003). T: Cell recognition based on pca and bayesian classification, in: *4th International Symposium, ICA 2003 (Citeseer)*, .
- Kumar P. Matlab based potent algorithm for wbc cancer detection and classification. *Biomed Pharmacol J.* (2021) 14:2277–84. doi: 10.13005/bpj.
- Su MC, Cheng CY, Wang PC. A neural-network-based approach to white blood cell classification. *Sci World J.* (2014). doi: 10.1155/2014/796371.
- Sharma S, Gupta S, Gupta D, Juneja S, Gupta P, Dhiman G, et al. Deep learning model for the automatic classification of white blood cells. *Comput Intell Neurosci.* (2022). doi: 10.1155/2022/7384131.
- Almezghwi K, Serte S. Improved classification of white blood cells with the generative adversarial network and deep convolutional neural network. *Comput Intell Neurosci.* (2020). doi: 10.1155/2020/6490479.
- Yildirim M, Çinar A. Classification of white blood cells by deep learning methods for diagnosing disease. *Rev d'Intel. Artif.* (2019) 33:335–40. doi: 10.18280/ria.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through small group Research Project under grant number RGP1/417/44.

## 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/fonc.2024.1328200/full#supplementary-material>

### SUPPLEMENTARY TABLE 1

Description of main symbols used in Algorithm 1.

19. Bhattacharjee R, Saini LM. (2015). Robust technique for the detection of acute lymphoblastic leukemia, in: *2015 IEEE Power, Communication and Information Technology Conference (PCITC)*, pp. 657–62. IEEE.
20. Goutam D, Sailaja S. (2015). Classification of acute myelogenous leukemia in blood microscopic images using supervised classifier, in: *2015 IEEE International Conference on Engineering and Technology (ICETECH)*, , Vol. 1–5. IEEE.
21. Patel N, Mishra A. Automated leukaemia detection using microscopic images. *Proc Comput Sci.* (2015) 58:635–42. doi: 10.1016/j.procs.2015.08.082.
22. Rawat J, Singh A, Bhadauria H, Virmani J. Computer aided diagnostic system for detection of leukemia using microscopic images. *Proc Comput Sci.* (2015) 70:748–56. doi: 10.1016/j.procs.2015.10.113.
23. Mishra S, Sharma L, Majhi B, Sa PK. (2017). Microscopic image classification using dct for the detection of acute lymphoblastic leukemia (all), in: *Proceedings of International Conference on Computer Vision and Image Processing: CVIP 2016, Volume 1*. pp. 171–80. Springer.
24. Di Ruberto C, Loddio A, Puglisi G. Blob detection and deep learning for leukemic blood image analysis. *Appl Sci.* (2020) 10:1176. doi: 10.3390/app10031176.
25. Bodzas A, Kodytek P, Zidek J. Automated detection of acute lymphoblastic leukemia from microscopic images based on human visual perception. *Front Bioeng. Biotechnol.* (2020) 8:1005. doi: 10.3389/fbioe.2020.01005.
26. Anwar S, Alam A. A convolutional neural network-based learning approach to acute lymphoblastic leukaemia detection with automated feature extraction. *Med Biol Eng comput.* (2020) 58:3113–21. doi: 10.1007/s11517-020-02282-x.
27. Batool A, Byun YC. Lightweight efficientnetb3 model based on depthwise separable convolutions for enhancing classification of leukemia white blood cell images. *IEEE Access.* (2023) 11:37203–15. doi: 10.1109/ACCESS.2023.3266511
28. Elhassan TAM, Rahim MSM, Swee TT, Hashim SZM, Aljurf M. Feature extraction of white blood cells using cmyk-moment localization and deep learning in acute myeloid leukemia blood smear microscopic images. *IEEE Access.* (2022) 10:16577–91. doi: 10.1109/ACCESS.2022.3149637
29. Ahmad R, Awais M, Kausar N, Tariq U, Cha JH, Balili J. Leukocytes classification for leukemia detection using quantum inspired deep feature selection. *Cancers.* (2023) 15:2507. doi: 10.3390/cancers15092507.
30. Ahmed N, Yigit A, Isik Z, Alpkocak A. Identification of leukemia subtypes from microscopic images using convolutional neural network. *Diagnostics.* (2019) 9:104. doi: 10.3390/diagnostics9030104.
31. Al Amin M, Paul BK, Bithi NI. (2022). Real time detection and localization of colorectal polyps from colonoscopy images: A deep learning approach, in: *2022 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE)*, . pp. 58–61. IEEE.
32. Ahmed A, Khan Q, Naeem M, Iqbal M, Anpalagan A, Awais M. An insight to the performance of estimation of distribution algorithm for multiple line outage identification. *Swarm evolution. Comput.* (2018) 39:114–22. doi: 10.1016/j.swevo.2017.09.006.
33. Ahmed A, Awais M, Naeem M, Iqbal M, Ejaz W, Anpalagan A, et al. Multiple power line outage detection in smart grids: probabilistic bayesian approach. *IEEE Access.* (2017) 6:10650–61. doi: 10.1109/ACCESS.2017.2710285.
34. Awais M, Ahmed A, Naeem M, Iqbal M, Qadri N, Anpalagan A. Multiple line outages identification: A customized quantum inspired approach. *Electric Power Syst Res.* (2016) 134:47–55. doi: 10.1016/j.epsr.2015.12.001.
35. Gupta D, Agrawal U, Arora J, Khanna A. Bat-inspired algorithm for feature selection and white blood cell classification. *Nature-Inspired Comput Swarm Intell (Elsevier).* (2020), 179–97. doi: 10.1016/B978-0-12-819714-1.00022-1.
36. Shahzad A, Raza M, Shah JH, Sharif M, Nayak RS. Categorizing white blood cells by utilizing deep features of proposed 4b-additionnet-based cnn network with ant colony optimization. *Complex Intel. Syst.* (2022), 1–17. doi: 10.1007/s40747-021-00564-x.
37. Liu J, Lin Y, Li Y, Weng W, Wu S. Online multi-label streaming feature selection based on neighborhood rough set. *Pattern Recogn.* (2018) 84:273–87. doi: 10.1016/j.patcog.2018.07.021.
38. Ahmad R, Awais M, Kausar N, Akram T. White blood cells classification using entropy-controlled deep features optimization. *Diagnostics.* (2023) 13:352. doi: 10.3390/diagnostics13030352.
39. Scotti F, Labati R, Piuri V. Acute lymphoblastic leukemia image database for image processing. *Department Comput Sci. Univ Milan Italy.* (2005).
40. Ghaderzadeh M, Aria M, Hosseini A, Asadi F, Bashash D, Abolghasemi H. A fast and efficient cnn model for b-all diagnosis and its subtypes classification using peripheral blood smear images. *Int J Intel. Syst.* (2022) 37:5113–33. doi: 10.1002/int.22753.
41. Munteanu C, Rosa A. (2000). Towards automatic image enhancement using genetic algorithms, in: *Proceedings of the 2000 Congress on Evolutionary Computation. CEC00 (Cat. No. 00TH8512)*, , Vol. 2. pp. 1535–42. IEEE.
42. Gonzalez RC. *Digital image processing* (Pearson education India). (2009).
43. Malik S, Akram T, Awais M, Khan MA, Hadjouni M, Alasiry A, et al. Improved skin lesion boundary estimation for enhanced-intensity images using hybrid metaheuristics. *Diagnostics.* (2023) 13(7):1285.
44. Imagenet. Available online at: <http://www.image-net.org>.
45. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. (2016). Rethinking the inception architecture for computer vision, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, (IEEE). pp. 2818–26.
46. Huang G, Liu Z, van der Maaten L, Weinberger KQ. (2017). Densely connected convolutional networks, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, (IEEE). pp. 4700–8.
47. Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer. *Adv Eng software.* (2014) 69:46–61. doi: 10.1016/j.advengsoft.2013.12.007.
48. Emary E, Zawbaa HM, Hassanien AE. Binary grey wolf optimization approaches for feature selection. *Neurocomputing.* (2016) 172:371–81. doi: 10.1016/j.neucom.2015.06.083.
49. Hegde RB, Prasad K, Hebbar H, Singh BMK, Sandhya I. Automated decision support system for detection of leukemia from peripheral blood smear images. *J digital Imaging.* (2020) 33:361–74. doi: 10.1007/s10278-019-00288-y.
50. Baig R, Rehman A, Almuhaimeed A, Alzahrani A, Rauf HT. Detecting Malignant leukemia cells using microscopic blood smear images: a deep learning approach. *Appl Sci.* (2022) 12:6317. doi: 10.3390/app12136317.



## OPEN ACCESS

## EDITED BY

Poonam Yadav,  
Northwestern University, United States

## REVIEWED BY

Rahul Joshi,  
Symbiosis International University, India  
Xirui Hou,  
Johns Hopkins University, United States

## \*CORRESPONDENCE

Yepuganti Karuna

✉ karun@vit.ac.in

RECEIVED 30 November 2023

ACCEPTED 25 March 2024

PUBLISHED 12 April 2024

## CITATION

Krishnapriya S and Karuna Y (2024) A deep learning model for the localization and extraction of brain tumors from MR images using YOLOv7 and grab cut algorithm. *Front. Oncol.* 14:1347363. doi: 10.3389/fonc.2024.1347363

## COPYRIGHT

© 2024 Krishnapriya and Karuna. 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.

# A deep learning model for the localization and extraction of brain tumors from MR images using YOLOv7 and grab cut algorithm

Srigiri Krishnapriya and Yepuganti Karuna\*

School of Electronics Engineering, Vellore Institute of Technology, Vellore, India

**Introduction:** Brain tumors are a common disease that affects millions of people worldwide. Considering the severity of brain tumors (BT), it is important to diagnose the disease in its early stages. With advancements in the diagnostic process, Magnetic Resonance Imaging (MRI) has been extensively used in disease detection. However, the accurate identification of BT is a complex task, and conventional techniques are not sufficiently robust to localize and extract tumors in MRI images. Therefore, in this study, we used a deep learning model combined with a segmentation algorithm to localize and extract tumors from MR images.

**Method:** This paper presents a Deep Learning (DL)-based You Look Only Once (YOLOv7) model in combination with the Grab Cut algorithm to extract the foreground of the tumor image to enhance the detection process. YOLOv7 is used to localize the tumor region, and the Grab Cut algorithm is used to extract the tumor from the localized region.

**Results:** The performance of the YOLOv7 model with and without the Grab Cut algorithm is evaluated. The results show that the proposed approach outperforms other techniques, such as hybrid CNN-SVM, YOLOv5, and YOLOv6, in terms of accuracy, precision, recall, specificity, and F1 score.

**Discussion:** Our results show that the proposed technique achieves a high dice score between tumor-extracted images and ground truth images. The findings show that the performance of the YOLOv7 model is improved by the inclusion of the Grab Cut algorithm compared to the performance of the model without the algorithm.

## KEYWORDS

brain tumor, deep learning, YOLOv7, grab cut algorithm, magnetic resonance imaging (MRI), gamma correction

## 1 Introduction

Brain tumors (BT) result in an unusual growth of brain cells, which is caused by the uncontrolled division of cells in the brain. In general, BTs are categorized as malignant (cancerous) or benign (noncancerous). (1). Benign or normal tumors do not cause any damage to the brain cells and can be easily treated. On the other hand, malignant tumors are dangerous and can spread to other organs if not treated in the early stages. The tumors are also classified as primary and secondary tumors wherein primary BT is developed from the existing cells and secondary tumors are developed from the cancerous cells (2). Benign tumors develop slowly and can be identified easily. These tumors can be removed by determining the brain region where they are located. Conversely, brain tumors can have serious consequences on human health and do not have any specific boundaries. Hence, they can affect other healthy cells in the brain and thereby completely disrupt the functioning of the brain (3).

There are several imaging modalities such as Perfusion magnetic resonance imaging (4), computed tomography (CT) (5), and positron emission tomography (PET) (6). Among the different techniques, MRI is a potential technique for identifying irregularities in brain patterns and works effectively on soft tissue (7). MRI is an invasive technique that generates high-quality brain images with better resolution. Usually, brain tumors are treated using advanced treatment processes such as chemotherapy, radiotherapy, and surgery which can destroy cancerous cells completely if the location of the tumor is identified correctly (8–10).

Manual identification and diagnosis of brain tumors can be a tedious and labor-intensive task. Since these techniques depend on manual intervention, the accuracy and precision of the tumor detection process are questionable. Hence, there is a need for a qualitative approach that can detect tumors and their location in the early stage with high accuracy and precision (11–13). The use of machine learning (ML) and deep learning (DL) techniques for precisely detecting BT has been emphasized in several studies (14, 15).

ML algorithms such as support vector machines (SVM) (16), Random Forest (RF) (17), Decision Trees (DT), K-nearest neighbor (KNN) (18), etc. have been used in previous works. However, these algorithms depend on manual feature extraction wherein the detection models are trained using these features. Hence, the accuracy of the detection and classification of brain tumors depends on the quality of the extracted features. In addition, constructing ML classifiers requires more resources, and their computational time is very high while processing large-scale datasets. As a result, these models exhibit a low classification accuracy (19).

A substantial amount of research has been dedicated to brain tumor detection and segmentation processes and various researchers have attempted to address the complexities associated with the detection process (20–22). One of the main challenges

related to brain tumor identification is the classification of neoplastic tissues which are heterogeneous in nature. These tissues overlap with the healthier tissues most of the time and conventional techniques used for tumor detection fail to distinguish them. Texture analysis is one such effective technique that can be used to determine the textural features of the tumors such as regularity, and orientation of the tumor, and thereby identify multiple indistinct areas in an image (23, 24).

The extraction of textural features helps the classifier to determine both visible and non-visible tumor regions with the aid of advanced techniques such as MRI. Conventional ML classifiers use gray-level and pixel-level-based features for classifying malignant and benign tumors. Various algorithms are used to automatically segment BT using MRI images and these techniques fail to achieve desired solutions for the issues related to BT detection techniques (25). The hybrid Convolutional Neural Network (CNN) and Deep Neural Network (DNN) were suggested (26) for addressing the drawbacks of ML algorithms such as high computational time and reduced classification accuracy. In this process, the CNN model was used to extract features that were classified using a fully connected network. The DNN employed in this work enhanced the performance of CNN by accurately classifying the tumor regions with an accuracy and F1 score of 96.08% and 97.3% respectively.

An ensemble model is implemented for distinguishing BT from MRI images (27). A pre-trained Inception ResNetV2 model is adopted for tumor detection and a ML-based RF model is employed for determining the stage and type of brain cancer (28). A cycle generative adversarial networks (C-GAN) model is used to augment the size of the dataset. The results exemplify that the proposed ensemble approach achieved detection and classification accuracies of 99% and 98%, respectively. The authors Dipu et al. (29) implemented a YOLOv5 model for detecting BT along with a DL library known as FastAi. The model was trained using data collected from the BRATS 2018 dataset, which consisted of 1,992 brain MRI images. It attained an overall accuracy of 85.95% and the FastAi model exhibited an improved accuracy of 95.78%. These two techniques validated the effectiveness of DL in the early detection of brain cancer.

The work mentioned in (30) implemented a YOLOv3 for identifying cancerous BTs. The YOLOv3 model was combined with a CNN model to boost the performance. This hybrid model attained an accuracy of 97%. However, YOLOv3 significantly requires more memory and this can be a challenging factor while working with limited resources. A YOLOv4 model is employed in (31) for BT detection. It is trained using a transfer learning (TL) approach and a pre-trained COCO dataset was used to maximize the tumor detection performance. Compared to the traditional YOLO model, the YOLOv4 model achieves better performance but with a high localization error.

A YOLOv5 was used by Paul et al. and Shelatkar et al. (32, 33) for segmenting brain cancer images and diagnosing brain tumors.



The YOLOv5 was trained on the BRATS 2021 dataset and the model achieved an average precision of 88%. It was observed from the results that the YOLOv5 model provided a slightly lesser accuracy compared to other classification models. It was also inferred that the complexity of the model increases the training time. The authors Arunachalam & Sethumathavan and Hossain et al. (34, 35) implemented YOLOv5 to detect abnormalities from brain images. The YOLOv5 model performed better compared to previous versions of YOLO and exhibited excellent tumor detection performance. However, the model was not tested for detecting malignant tumors from brain images.

As inferred from the existing approaches, most of the techniques used in the brain tumor detection process employ pre-processing and segmentation to identify and distinguish BTs and these techniques are not effective in recognizing normal or malicious tumor areas. In addition, conventional YOLO models namely YOLOv3, YOLOv4, and YOLOv5 suffer from certain drawbacks such as high computational complexity, compromised accuracy to maintain fast execution speed restricts their adaptability in disease detection tasks. Besides, these models rely on larger datasets, and collecting such datasets for rare tumor classes can be a tedious and time-consuming task. These drawbacks motivate this research to employ an advanced version of the YOLO model to automatically segment BTs with enhanced accuracy.

To address these drawbacks, in this work, we implemented a deep learning (DL) model for the accurate detection of brain tumors with better performance. The detection and segmentation of brain tumors from MRI images using the hybrid DL-based YOLOv7 and Grab Cut algorithms are presented here. The model was trained using a Br35H Brain tumor dataset, and its effectiveness is validated through a comparative analysis.

The primary contributions of the proposed work are as follows:

- The data processing technique used in this work consists of different processes such as RGB to Gray conversion, Otsu's thresholding, Brain Skull Removal, Image Resizing, and Median filtering.
- We proposed an efficient object detection-based YOLOv7 algorithm for diagnosing brain tumors in the early stages to mitigate the effect and speed up the diagnosis process.
- A gamma correction technique and a Grab Cut algorithm are used to extract the Gamma-corrected image.
- The performance of the YOLOv7 model is evaluated with and without the GrabCut algorithm and the proposed model performed better than the other existing algorithms in both cases.

The remaining portion of the paper is organized as: Section 2 includes the suggested methodology to train the model with the sourced dataset for detecting tumors from brain MRI images. This section also discusses the implementation of YOLOv7 and the GrabCut algorithm for the detection and extraction of tumors. Section 3 evaluates the results of the experiments conducted based on the proposed methodology. Lastly, Section 4 outlines the conclusion based on the produced results with future scope.

## 2 Materials and methods

This paper aims to achieve a highly accurate recognition of BTs from MRI images. DL-based YOLOv7 model (36) is used for achieving faster and more accurate results for tumor detection and classification. The automatic recognition of BT is a challenging task because of the similarities and irregularities in tumor images obtained from MRI scans. These issues make it difficult for the classifier to recognize and classify the tumors with better precision. Hence, it employs a Gamma correction mechanism to improve the quality of the images.

This work implements a structured approach to classify BTs. In the initial stage, the data from the brain tumor dataset is collected for analysis, and in the second stage, the images are preprocessed and subjected to Gamma correction in the third stage. In the fourth step, the YOLOv7 model is implemented to detect and locate the tumor. In the fifth stage, the Grab cut algorithm (37) is used for extracting the foreground of the tumor image. The process flow of the proposed approach is shown in Figure 1.

### 2.1 Dataset collection and preparation

The dataset used for training the YOLOv7 model is collected from BR35H: Brain Tumor Detection 2020 (BR35H) (38). The dataset incorporated 1437 brain MRI images of which 734 were malignant and 703 were normal tumors. The dataset consists of both T1 and T2 weighted images and all images are two-dimensional (2D images) and have a dimension of  $256 \times 256$  pixels. All the images are skull-stripped and labeled as 'YES' if the tumor is present; otherwise, labeled as 'NO'. The description of the dataset is given in Table 1.

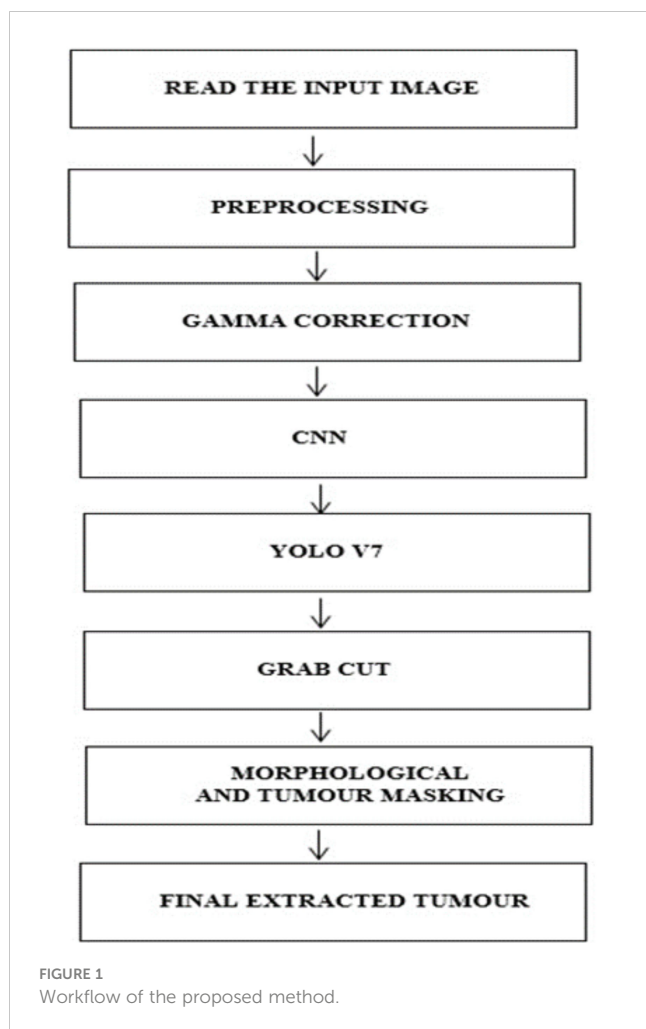
The dataset is split into a ratio of 70:15:15 where 70% of the data is used for training the model, 15% for testing, and the remaining 15% for validation.

### 2.2 Data preprocessing

The data is preprocessed to enhance the quality of the images and make them suitable for the classification process. Preprocessing significantly improves the classification performance of the DL models by filtering out the uncertainties. In this work, preprocessing is performed using different stages such as RGB to Gray conversion, Otsu's thresholding, Brain Skull Removal, Image Resizing, and Median filtering.

- **RGB to Gray conversion:** The RGB images consist of red, green, and blue-scale images arranged on top of each other. A grayscale image is a single-layered image denoted as an  $M \times N$  array, whose values are used to represent the intensity of an image. To convert the RGB images into gray images, the components of the red, green, and blue images were extracted and represented in three different two-dimensional matrices. A new matrix is created with similar dimensions, where the





number of rows and columns is equal to that of the RGB images. Each pixel of the RGB image was converted at location  $(i, j)$  to grayscale values by determining the weighted sum of the RGB color components and assigning it to the respective location  $(i, j)$  in the new matrix. This process simplifies the classification process and reduces computational complexity.

- **Otsu's thresholding process:** This process is used to determine a threshold value to reduce the overlap between foreground and background images. In other words, Otsu's algorithm returns a single intensity threshold value that separates the pixels into two different sets: foreground and background.
- **Brain Skull Removal:** Skull stripping was performed to eliminate non-brain tissue from the MRI images. This improved the speed and accuracy of the segmentation process. At this stage, flood-filling and masking

operations were considered for the skull removal process. The flood-fill algorithm is used to identify or modify adjacent values in the image based on their similarity to the original values. Furthermore, masking was performed to identify a specific Region of Interest (ROI) for analyzing the tumor. This process discards image regions that are not characterized by tumors.

- **Image Resizing:** It is performed to minimize the size of the image without altering the actual image information. In this work, all the input images were resized to  $(250 \times 250)$  pixels to avoid overfitting.
- **Noise Removal:** The noise in the input images was removed using a median filter with a kernel size of  $(3 \times 3)$ . Median filters are highly effective in filtering noise while preserving the edges. The filter computes the intensity of the pixel surrounded by the central pixel. The obtained median value was replaced with the intensity of the center pixel.

## 2.3 Gamma correction

The preprocessed images are subjected to Gamma correction to control the overall brightness of the tumor image. In this process, the images that are too dark or bright are corrected. The CNN performs automatic classification of the image attributes considering the statistical features. The contrast of the image is enhanced by dynamically modifying the parameters. In this stage, the Gamma correction is applied for each ROI of the image and this contributes to the overall image enhancement process. Overall, by incorporating gamma correction into the pipeline, the model can benefit from improved image quality, enhanced feature representation, and better generalization capabilities, ultimately leading to improved classification performance for brain tumor images. However, this did not change the underlying content or category of the image.

## 2.4 YOLOv7 model for tumor detection

Considering the benefits of the supervised learning of DL based YOLO model, this research employs an advanced version of the traditional YOLO model known as the YOLOv7 model. The YOLOv7 model is designed to develop an appropriate technique for identifying BT from brain MRI images. The working operation of YOLOv7 is unique and indistinct from fundamental methods used for detecting BT. In this process, the model simultaneously predicts the class and puts a bounding box around the tumor area. Each bounding box consists of five components  $(x, y, w, h, \text{ and } \text{the})$

TABLE 1 Description of the Brain Tumor MRI Dataset.

|            | No of Images | No of patients | Training Samples | Testing Samples | Validation Samples | Label   |
|------------|--------------|----------------|------------------|-----------------|--------------------|---------|
| With Tumor | 734          | 68             | 514              | 110             | 110                | Yes (1) |
| No Tumor   | 703          | 70             | 493              | 105             | 105                | No (0)  |

confidence score) with the first four components corresponding to the center coordinates (x, y, width, and height) of the respective bounding box, and the fundamental motive of YOLO is object detection and localization via bounding boxes. Therefore, two sets of bounding box vectors are required, i.e., vector 'y' is the representative of ground truth, and vector 'Y' is the predicted vector which is shown in Equation 1.

$$Y = [p_c, b_x, b_y, b_h, b_w, c] \quad (1)$$

- $p_c$  corresponds to the objectness score (the probability score of the grid containing an object).
- $b_x, b_y$  are the x and y coordinates of the center of the bounding box for the enveloping grid cell.
- $b_h, b_w$  correspond to the height and the width of the bounding box for the enveloping grid cell.
- 'c' corresponds to the class.

The MRI images are arranged in a grid of dimensions  $D \times D$  for each grid cell. In the case where the center of the object of interest falls into one of the grid cells, that particular grid cell would be responsible for the detection of that object. This permitted the other cells to neglect the object in the case of multiple appearances. Each grid cell predicts B bounding boxes along with the dimensions and confidence scores. The confidence score was indicative of the absence or presence of an object within the bounding box. Therefore, the confidence score can be expressed as Equation 2:

$$C = P_r(\text{Object}) \times IOU_{pred}^{truth} \quad (2)$$

where  $P_r(\text{Object})$  dignified the probability of the object being present, within a range of 0–1, with 0 indicating that the object does not exist and  $IOU_{pred}^{truth}$  notes the intersection-over-union with the predicted bounding box for the ground truth bounding box. To address multiple bounding boxes containing no object or the same object, YOLO opts for non-maximum suppression (NMS). By defining a threshold value for the NMS, all overlapping predicted bounding boxes with an IoU lower than the defined NMS value are eliminated.

The losses associated with YOLOv7 are bounding box loss and objectness loss. Bounding box Loss (Localization loss) is represented in Equation 3:

$$L_{i\text{ box}} = (x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 + (\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2 \quad (3)$$

here  $(\hat{x}_i, \hat{y}_i, \hat{w}_i, \hat{h}_i)$  represent ground truth values and  $(x_i, y_i, w_i, h_i)$  represent predicted values. Objectness Loss(confidence loss) is expressed as in Equation 4:

$$L_{i\text{ object}} = (c_i - \hat{c}_i)^2 \quad (4)$$

In this process, the features are learned from labeled data, and the YOLOv7 is initialized using the learned features. In this work, the model is trained using both low-level and high-level features of the brain tumor, and the model is updated after every iteration. This

allows fine-tuning of the learned parameters and enables the layers of YOLOv7 to capture features that are highly discriminative in nature. The architecture of the YOLOv7 model is illustrated in Figure 2.

The proposed YOLOv7 architecture incorporates three layers namely; (i) Backbone: E-ELAN, (ii) Neck: PANet, and (iii) Head: YOLO Layer. The backbone layer is the first layer responsible for extricating important tumor features from MRI images. A cross-stage partial network is utilized for extracting representational features.

The backbone of YOLOv7 consists of an Extended Efficient Layer Aggregation Network (E-ELAN) architecture (39) that uses expand, shuffle, and merge cardinality to improve the learning ability of the model without affecting gradient flow paths. E-ELAN modifies the YOLOv7 architecture in the computational block and the architecture remains the same in the transition layer. E-ELAN incorporates a group convolution method to maximize the channel capacity and cardinality of the computation block. The channel multiplier is applied to all blocks in the computation layer, and a feature map is created for each block. The feature maps from all blocks are concatenated, and the obtained feature map is used to merge the cardinality, as shown in Figure 3.

The second layer is the PANet layer, also known as the neck of the model. The main reason behind selecting PANet is its capacity to restore the spatial data and thereby contribute significantly to the improvement of the localization process which in turn helps in creating the mask around the image. This layer employs anchor boxes for constructing feature vectors with bounding boxes for tumor detection. The neck aggregates the feature maps obtained from the Backbone and creates feature pyramids. The neck is made up of multiple paths and the features extracted from the backbone model are used to create the FPN as shown in Figure 4.

The last layer in the YOLOv7 architecture is the head of the model which computes the final predictions as classification and localization. The head predicts classes and bounding boxes, classification scores, and objectness scores of objects based on the features collected from the neck. In YOLOv7, the head generates the final output, which is called the Lead Head, and assists in training the middle layers, called the Auxiliary Head. With the help of assistant loss, the weights of the auxiliary heads are updated, which enables deep supervision and thereby allows the model to learn better. The head of the YOLOv7 model is presented in Figure 5.

## 2.5 Grab cut algorithm

The detected tumor from the YOLOv7 model is analyzed using the Grab Cut algorithm which extracts the feature from the gamma-corrected image. This algorithm is used to extract the foreground of an image by drawing a rectangular box around it. This box helps in coordinating the image regions. However, the image contains both foreground and background regions and hence it is essential to eliminate the redundant background regions. This is achieved through a segmentation process wherein the pixels located in the foreground and background images are segmented and thereby

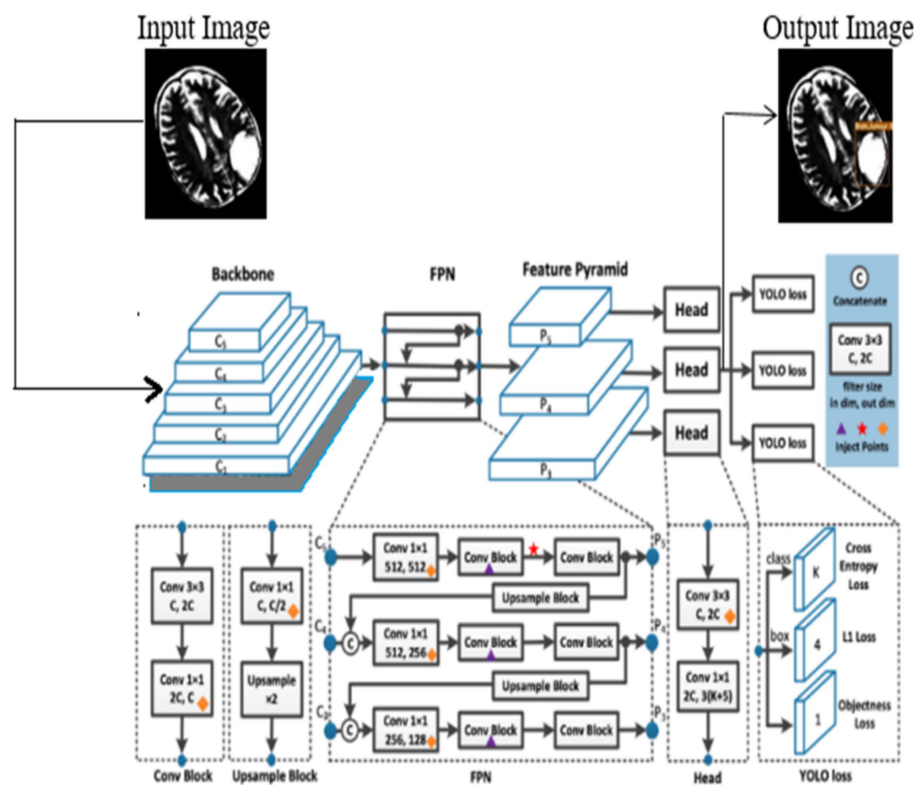


FIGURE 2  
The proposed architecture of the YOLOv7 model.

helping in extracting only foreground images to achieve a better tumor detection performance. An input image is accepted whose value is 1 and for that, a bounding box is assigned. This determines the object in an image that needs to be segmented. The steps involved in the process are defined as follows:

Step 1: A Gaussian Mixture Model (GMM) is used for estimating the color distribution across the foreground and background images.

Step 2: A Markov random field is constructed over the pixel's labels.

Step 3: The final segmented output images are obtained via the Grab Cut algorithm.

In Grab Cut, the model for monochrome images is replaced by GMM for color images. Soft segmentation is performed wherein a new vector  $k = \{k_1, \dots, k_n, \dots, k_N\}$  is imputed to each pixel of GMM's  $k_n$ th component, where  $k_n = 1, 2, \dots, K$  (Normally  $K = 5$ ), and  $\alpha_n = 0, 1$  is assigned to each pixel to show that it belongs to either the foreground or background GMM. The energy function of the GrabCut algorithm is defined as shown in Equation 5:

$$E(\alpha, k, \theta, z) = U(\alpha, k, \theta, z) + V(\alpha, z) \quad (5)$$

and GMM is defined using Equation 6:

$$G(z) = \sum_{i=1}^K \omega_k g_k(z; \mu_k, \Sigma_k), \quad \sum_{i=1}^K \omega_k = 1, \text{ and } 0 < \omega_k < 1 \quad (6)$$

Where  $g_k(z; \mu_k, \Sigma_k)$  is the Gaussian distribution function for each component  $t_k, k = 1, 2, \dots, K$  is given by Equation 7.

$$g(z; \mu; \Sigma) = \frac{1}{\sqrt{2\pi^D} |\Sigma|} \exp\left[-\frac{1}{2} (z - \mu)^T \Sigma^{-1} (z - \mu)\right] \quad (7)$$

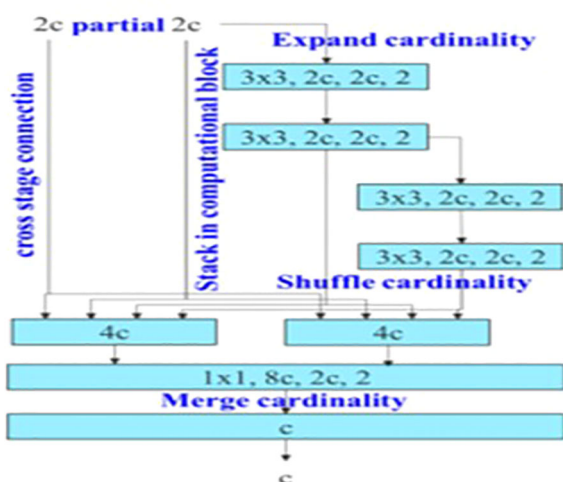


FIGURE 3  
E-ELAN as Backbone Layer for YOLOv7 architecture.

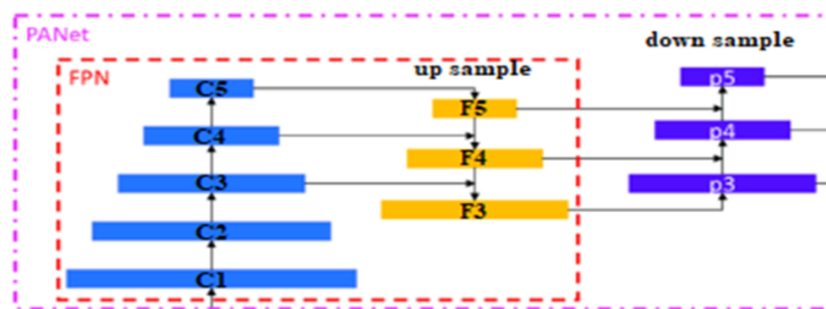


FIGURE 4  
PANet layer in the YOLOv7 architecture.

and  $\omega_k$  is the weighting coefficient,  $\mu_k$  is the means,  $\Sigma_k$  is the covariance matrix for  $k^{\text{th}}$  component and  $D$  is the number of dimensions of variable  $z$ . Combining equations 2, 3, and 4, the term  $U$  is computed as in Equation 8.

$$U(\alpha, k, \theta, z) = \sum_n G(\alpha, k, \theta, z) \quad (8)$$

Where  $G(a, k, 0, z)$  is expressed as in Equation 9.

$$G(\alpha, k, \theta, z) = -\log \omega(\alpha_n, k_n) + \frac{1}{2} \log |\Sigma(\alpha_n, k_n)| + \frac{1}{2} [z_n - \mu(\alpha_n, k_n)^T \Sigma(\alpha_n, k_n)^{-1} [z_n - \mu(\alpha_n, k_n)]] \quad (9)$$

And the term  $\theta$  is defined as in Equation 10.

$$\theta = \{\pi(\alpha, k), \mu(\alpha, k), \Sigma(\alpha, k), \alpha = 0, 1, k = 1, \dots, K\} \quad (10)$$

Grab Cut minimizes the energy function by modifying the iterative minimization cut algorithm. In the initial stage, the algorithm considers two-pixel sets wherein one set represents the background ( $\alpha_n = 0$ ) and another one for object classes ( $\alpha_n = 1$ ). Two GMMs were initialized along with the two sets to start the iteration. GrabCut is an interactive version of

graph cut where the user quickly marks some pixels as background, some as foreground, and then graph cut sorts out the rest (constraining the marked pixels to belong to the background and foreground or source-side and sink-side respectively). In this algorithm, the minimum cut is obtained by determining the maximum flow of data in the graph. In a graph, the connectivity is formed by removing the set of edges which also forms two individual subsets namely a maximum and a minimum cut. The max-flow min-cut theorem states that the maximum flow through any network from a given source to a given sink is equal to the minimum sum of a cut. The results of the simulation analysis are discussed in the below sections.

### 3 Results

This section provides localization and segmentation results on a dataset made publicly available on Kaggle (38). We conducted our experiments on the PYTHON 3.10.2 platform and executed on a system with an Intel(R) Core (TM) i5-1035G1 CPU, 8 GB RAM, and 3.3 GHz. We trained the model using the following hyperparameters: a learning rate (lr0) of 0.01, weight decay of 0.0005, and batch size of 16. We used the ADAM optimizer for 100 epochs.

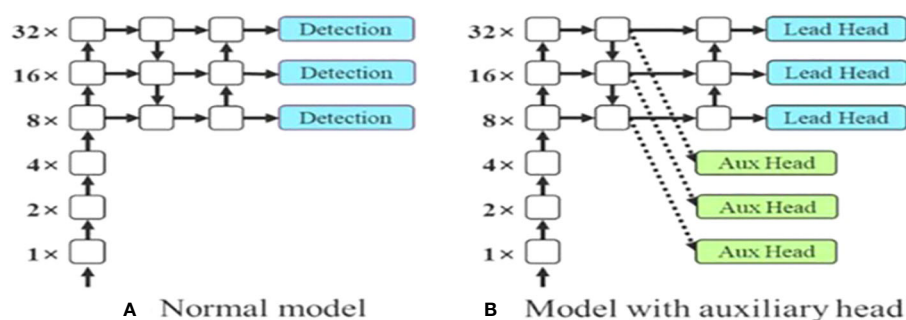


FIGURE 5  
Head layer in YOLOv7 architecture. (A) Normal model (B) Model with auxiliary head.

### 3.1 Performance evaluation metrics

The efficacy of the YOLOv7 model was determined using the following metrics.

Accuracy is defined as the percentage of accurately detected brain tumors and is calculated as shown in Equation 11.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

Recall is defined as the ratio of brain tumor images that were accurately classified as shown in Equation 12.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (12)$$

The F1 score is determined as the weighted harmonic mean of its precision and recall are given by Equation 13.

$$\text{F1 score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (13)$$

Similarly, precision is defined as the accuracy of the positive predictions which is shown in Equation 14.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (14)$$

Based on the YOLOv7 model we trained, we achieved good results in terms of the overall mAP and individual class performance. The model achieved an overall mAP50 of 0.9391 and mAP 50-95 of 0.4981 on the validation set. This means that the model was able to accurately localize the tumor region with a high degree of confidence.

Figure 6 shows the loss values for the box loss, and object loss at each epoch during the training process. The box loss represents the difference between the predicted and ground-truth bounding box coordinates, and the object loss represents the confidence score for each object detected in an image. The goal of training an object

detection model is to minimize the total loss, which is a combination of box loss, and object loss. The loss values should exhibit a decreasing trend as the training progresses, indicating an improvement in the model's ability to localize the tumor region.

Moreover, from Figure 6, it appears that the precision, recall, and mean average precision (mAP) are all increasing with training epochs. This could indicate that the model improves over time and becomes more accurate at identifying the correct location of the tumor region. Our proposed model achieved a mean average precision (mAP50) of 0.9304 and 0.9391, respectively, indicating a high level of accuracy in identifying and localizing tumor regions in the images. The model accurately localized tumor regions with a precision (P) of 99% and recall (R) of 100%, demonstrating its ability to localize tumor regions even in challenging image conditions.

Overall, the results of our YOLOv7s model suggest that it performed well in accurately localizing the tumor region in the brain MR images we used for training and validation. Hence, we can infer that these results demonstrate the potential of the YOLOv7 and Grab cut model for localizing and extracting brain tumor in MR medical images.

### 3.2 Simulation results

The input image for YOLOv7 and the tumor detected image is shown in Figure 7.

The values of different performance metrics obtained from simulation for the proposed method are tabulated in Table 2.

It can be inferred from the table that the proposed detection model achieved an optimal accuracy of 99.5% for training and testing datasets, and 99% for validation datasets. In addition to the performance evaluation metrics listed in Table 2, the performance of the proposed approach was validated in terms of training and

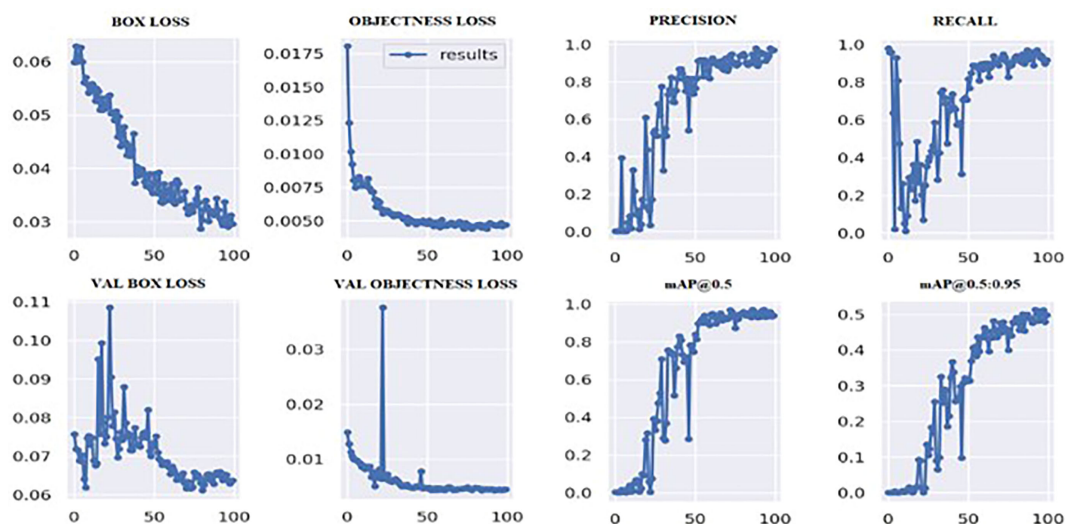


FIGURE 6  
Outcomes of the training process.



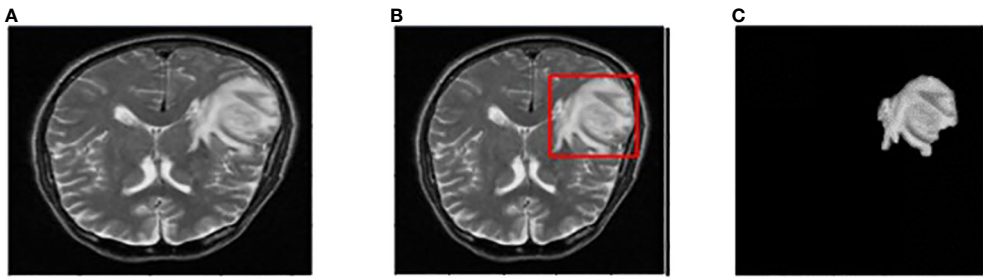


FIGURE 7  
(A) Input image (B) Tumor detected using YOLOv7 (C) Extracted tumor using Grab Cut algorithm.

validation loss, objectness loss, precision, and recall which are illustrated in the figures below.

The loss function of YOLOv7 is computed as a combination of two individual loss functions, that is, Bounding Box Regression (which measures how well predicted bounding boxes capture ground truth bounding boxes) and cross-entropy loss (which measures how well a job the detector did in predicting the correct class). The box loss represented in Figure 8A shows the effectiveness of the algorithm in terms of locating the center of the object (tumor image) and how well the predicted bounding box covers an object. The validation objectness loss is shown in Figure 8B. Objectness loss measures the probability that a tumor exists in the proposed ROI. If objectness is high, the image window is likely to contain an object.

TABLE 2 Performance metrics for the proposed method.

|             | Training | Testing | Validation |
|-------------|----------|---------|------------|
| Accuracy    | 99.5%    | 99.5%   | 99%        |
| Precision   | 99.0%    | 99.0%   | 98.03%     |
| Recall      | 100%     | 100%    | 100%       |
| Specificity | 100%     | 100%    | 100%       |
| F1 score    | 99.5%    | 99.5%   | 99%        |

As observed in Figure 8B, the proposed approach exhibits a high objectness score and hence helps in locating the tumor from the given image. The precision and recall graphs for the proposed model are shown in Figure 9.

The training measurement values used to train the YOLOv7 model are tabulated in Table 3. A graphical representation of the training process outcomes is shown in Figure 6.

Figure 6 shows that the mAP values obtained during the validation for 100 epochs was IOU = 0.5 and mAP for IOU from 0.5 to 0.95. For mAP at 0.5 and 0.95 are measured as the step values for different values such as 0.05 (0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95).

In addition, the comparison is done between ground truth test images and predicted test images. In this work, four segmentation techniques like Fuzzy C means segmentation, K-means clustering, Otsu thresholding, and Grab cut algorithm are applied for finding the DICE similarity measurement between ground truth images and predicted test images. The resultant images for the segmentation process are shown in Figure 10.

The dice similarity measurements for all four segmentation techniques are illustrated in Table 4. It is understood from the results that the proposed Grab cut algorithm yields enhanced results compared to other techniques by achieving high dice similarity measurement. The performance of the YOLOv7 model in combination with the Grab cut algorithm is also compared with

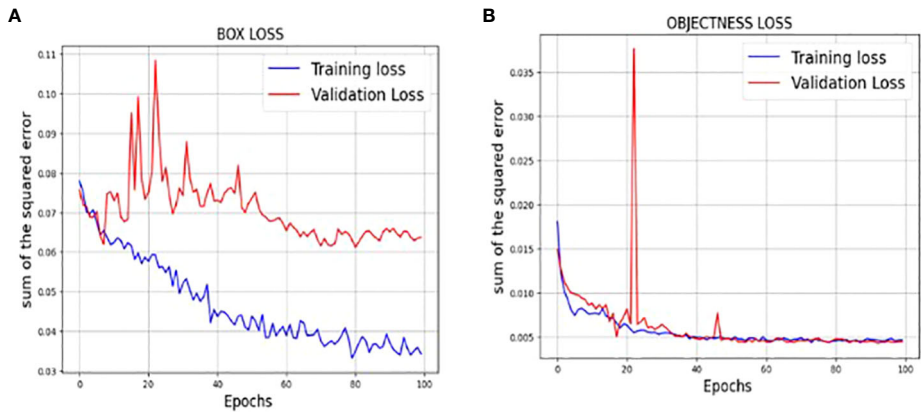
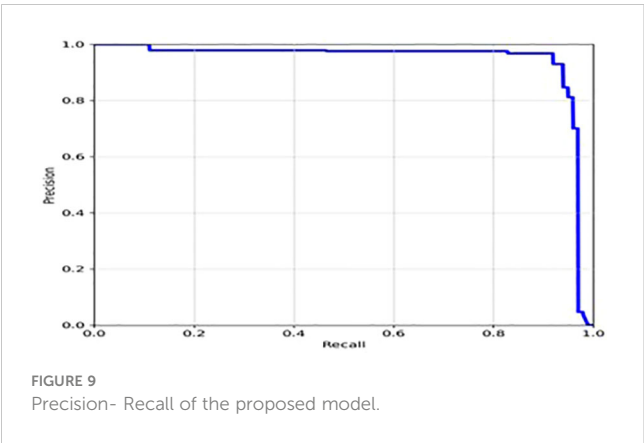


FIGURE 8  
(A) Training and validation Box loss of the proposed model (B) Training and validation Objectness loss of the proposed model.



that of other tumor extraction mechanisms, and the results are illustrated in Figure 11.

4 Discussion

Although several studies have been conducted on the application of deep learning for tumor localization and extraction, the combination of Grab cut and YOLOv7 has not been widely employed in this field. Actually, as far as we are aware, no studies have used YOLOv7 plus Grab cut for this purpose. Therefore, by using YOLOv7 in combination with Grab cut for tumor localization and extraction, our work represents a novel contribution to the field.

TABLE 3 Findings from training procedure.

| Epochs | Box loss | Val Box loss | Objectness loss | Val Objectness loss | Precision | Recall | mAP @ 0.5 | mAP@ 0.5: 0.95 |
|--------|----------|--------------|-----------------|---------------------|-----------|--------|-----------|----------------|
| 1      | 0.07792  | 0.07574      | 0.01807         | 0.01491             | 0.00332   | 0.9798 | 0.0032    | 0.00048        |
| 10     | 0.06179  | 0.07527      | 0.00763         | 0.00861             | 0.08104   | 0.2626 | 0.03995   | 0.00600        |
| 20     | 0.05863  | 0.07332      | 0.00650         | 0.00729             | 0.6102    | 0.2688 | 0.2806    | 0.09291        |
| 30     | 0.04947  | 0.07618      | 0.00532         | 0.00612             | 0.7755    | 0.5859 | 0.7094    | 0.2555         |
| 40     | 0.0455   | 0.07286      | 0.05112         | 0.00469             | 0.8157    | 0.7152 | 0.773     | 0.323          |
| 50     | 0.03831  | 0.06884      | 0.00461         | 0.00460             | 0.9157    | 0.8889 | 0.9211    | 0.4099         |
| 60     | 0.04114  | 0.06723      | 0.00506         | 0.00475             | 0.8183    | 0.8687 | 0.8892    | 0.4378         |
| 70     | 0.04045  | 0.06302      | 0.00479         | 0.00465             | 0.9261    | 0.9091 | 0.9291    | 0.4515         |
| 80     | 0.03322  | 0.06304      | 0.00456         | 0.00465             | 0.9384    | 0.9293 | 0.9464    | 0.4771         |
| 90     | 0.03923  | 0.06588      | 0.00482         | 0.00449             | 0.8846    | 0.9291 | 0.9304    | 0.4881         |
| 100    | 0.03426  | 0.06369      | 0.00469         | 0.00449             | 0.9681    | 0.9191 | 0.9391    | 0.4981         |

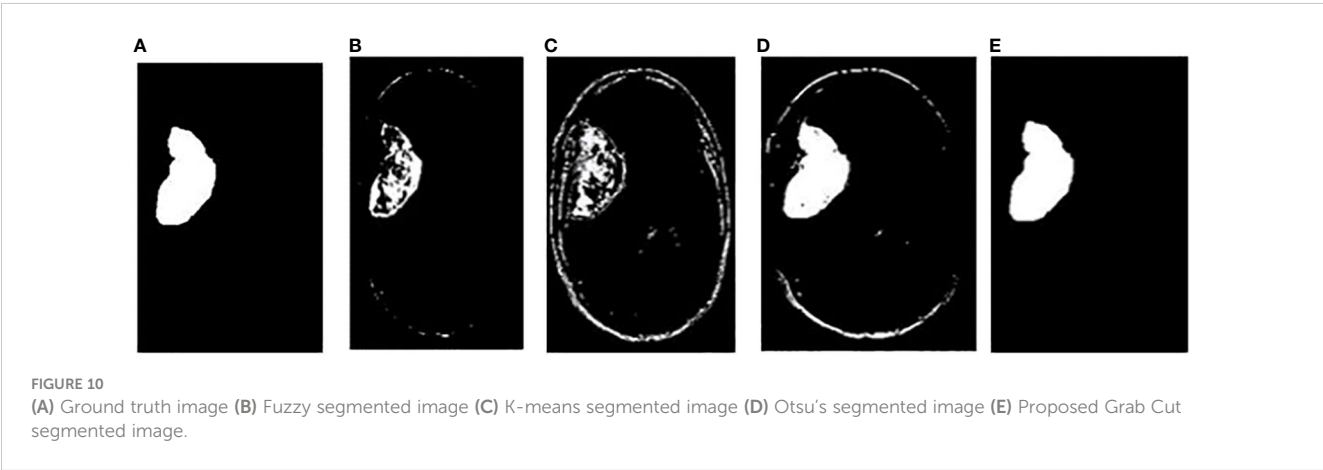


TABLE 4 Dice similarity measurement of various segmentation techniques.

| Segmentation techniques                      | Dice similarity measurements |
|----------------------------------------------|------------------------------|
| Fuzzy Segmentation(Pitchai, R et al.,2021)   | 0.9240                       |
| K-means Clustering(Sinaga, K. P et al.,2020) | 0.9354                       |
| Otsu's Thresholding(Huang, C et al., 2021)   | 0.8765                       |
| <b>Proposed Grab cut Algorithm</b>           | <b>0.9831</b>                |

Bold text and values represent the proposed work.

Moreover, it is important to remember that accuracy on its own might not be a good enough statistic for object detection tasks because it ignores false positives and false negatives. Rather, mean average precision, or mAP, is frequently employed to assess how well object detection models perform. The mAP offers a more thorough assessment of the model's performance by accounting for precision and recall at various intersection over union (IoU) thresholds. Our study's mAP50 of 0.9391 shows that our model does a good job of identifying the tumor location.

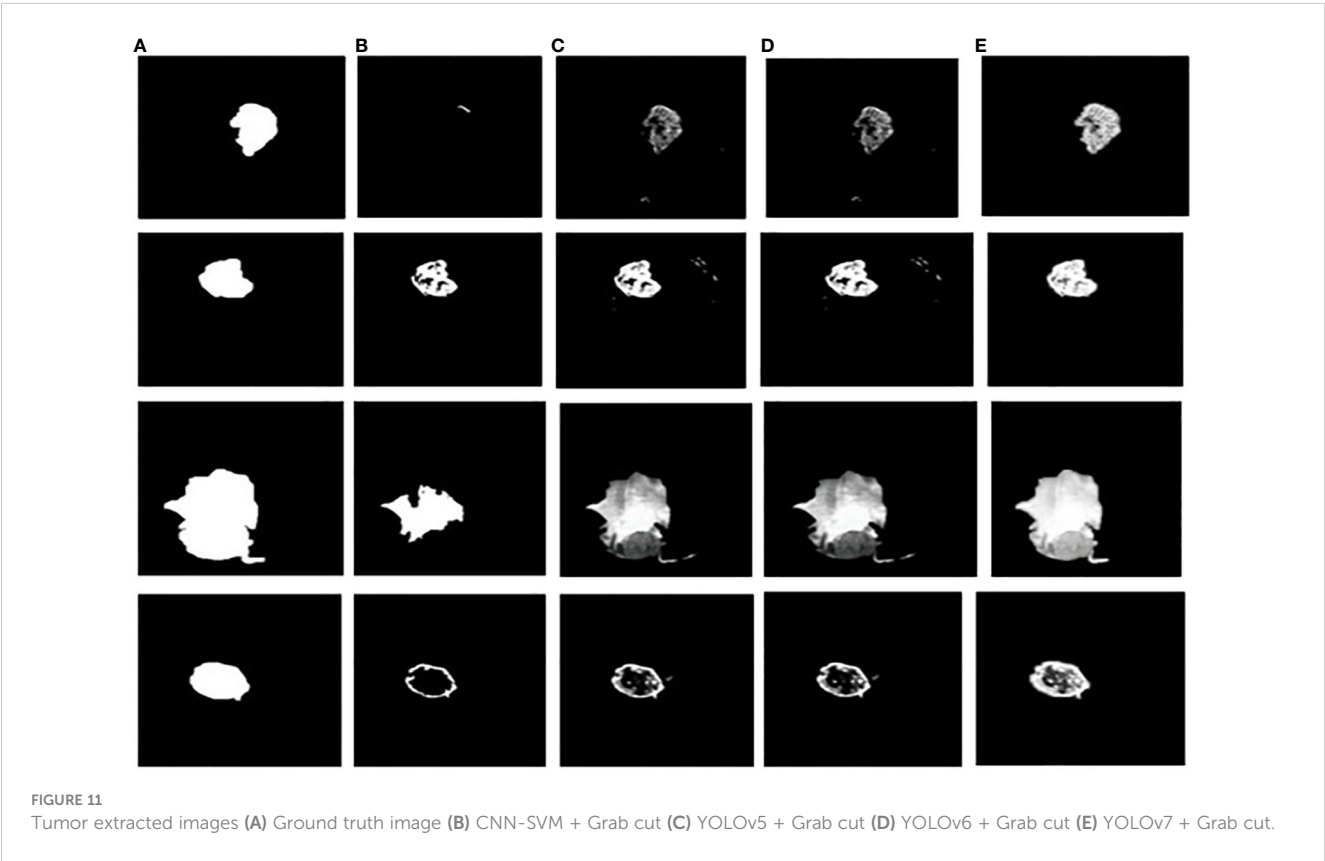
The performance of the proposed model is compared with other techniques, such as the hybrid CNN-SVM, YOLOv5, and YOLOv6 models. The outcomes are shown in Figure 12, and the

obtained values are listed in Table 5. The table clearly shows that our model outperforms the benchmark models in terms of metrics of dice similarity, accuracy, precision, recall, specificity, and F1 score.

In Figure 10, the first, third, and fourth rows represent the tumor input image and the second row represents a non-tumor input image. As inferred from the comparative results (Table 4) the proposed YOLOv7 achieves excellent results compared to the existing methodologies. The accuracy of 99% is obtained by using the proposed approach and the accuracy of hybrid CNN-SVM is 69%, YOLOv5 and YOLOv6 are 97.5% respectively. A highest precision of 98.03% is achieved by the YOLOv7 model and the precision values are 69.79%, 97.02% and 97.79% for hybrid CNN-SVM, YOLOv5 and YOLOv6 techniques respectively.

In addition, the YOLOv7 model is also tested with and without the application of the Grab cut algorithm, as shown in Figures 13A, B, respectively. The average mean dice similarity score value between the predicted test images and corresponding ground truth images using Grab cut algorithm for tumor extraction is shown in Table 6. The outcomes of YOLOv7 with and without the combination of Grab Cut are shown in Figure 14.

Results show that the proposed technique attains a high dice score between tumor extracted images and Ground truth images. The findings show that the performance of the YOLOv7 model is improved by the inclusion of the Grab Cut algorithm as compared to the performance of the model without the algorithm.



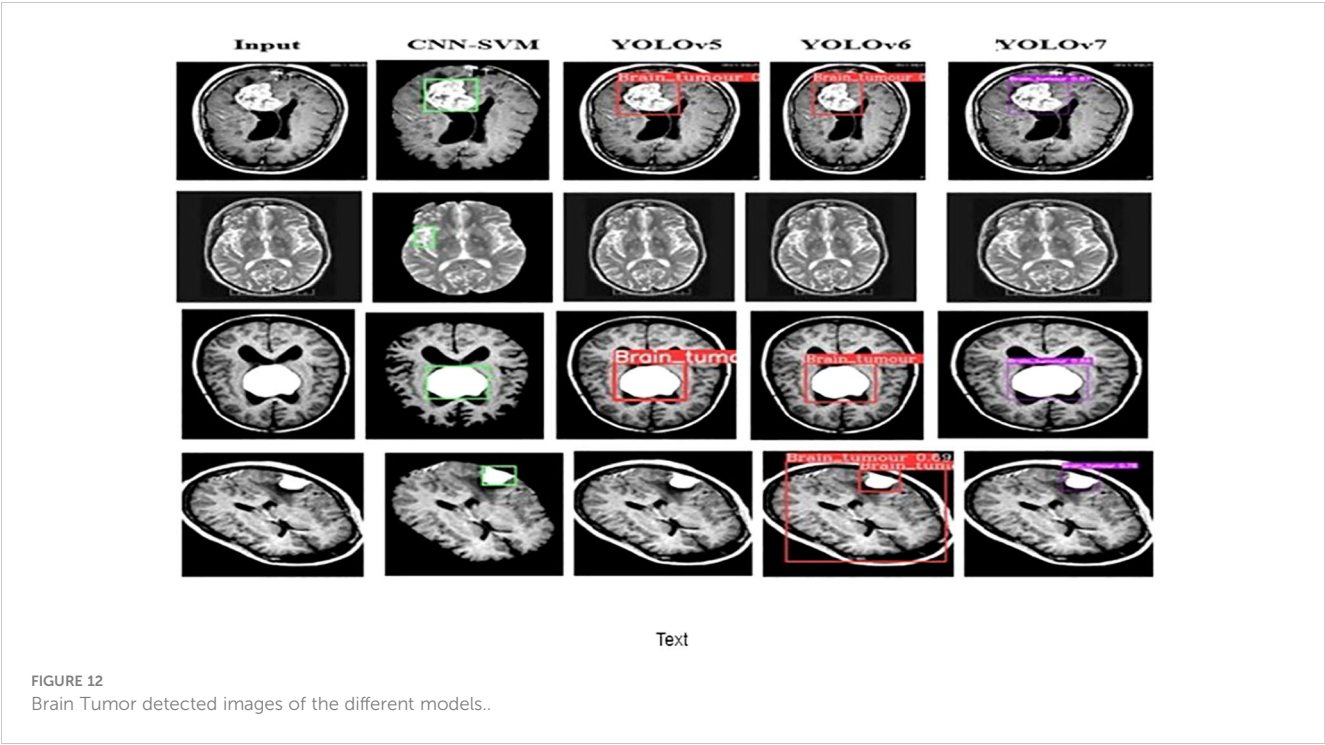
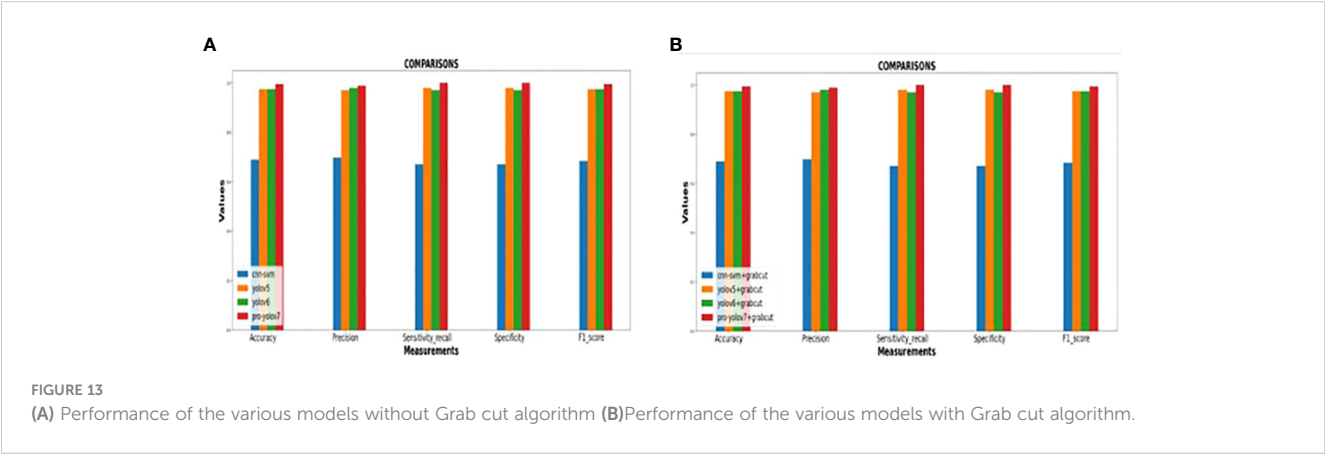


TABLE 5 Comparison of the proposed method with other methods.

|             | Hybrid CNN-SVM | YOLOv5 | YOLOv6 | Proposed YOLOv7 |
|-------------|----------------|--------|--------|-----------------|
| Accuracy    | 69%            | 97.5%  | 97.5%  | <b>99%</b>      |
| Precision   | 69.79%         | 97.02% | 97.79% | <b>98.03%</b>   |
| Recall      | 67%            | 98%    | 97%    | <b>100%</b>     |
| Specificity | 67%            | 98%    | 97%    | <b>100%</b>     |
| F1 score    | 68.36%         | 97.51% | 97.48% | <b>99%</b>      |

Bold text and values represent the proposed work.



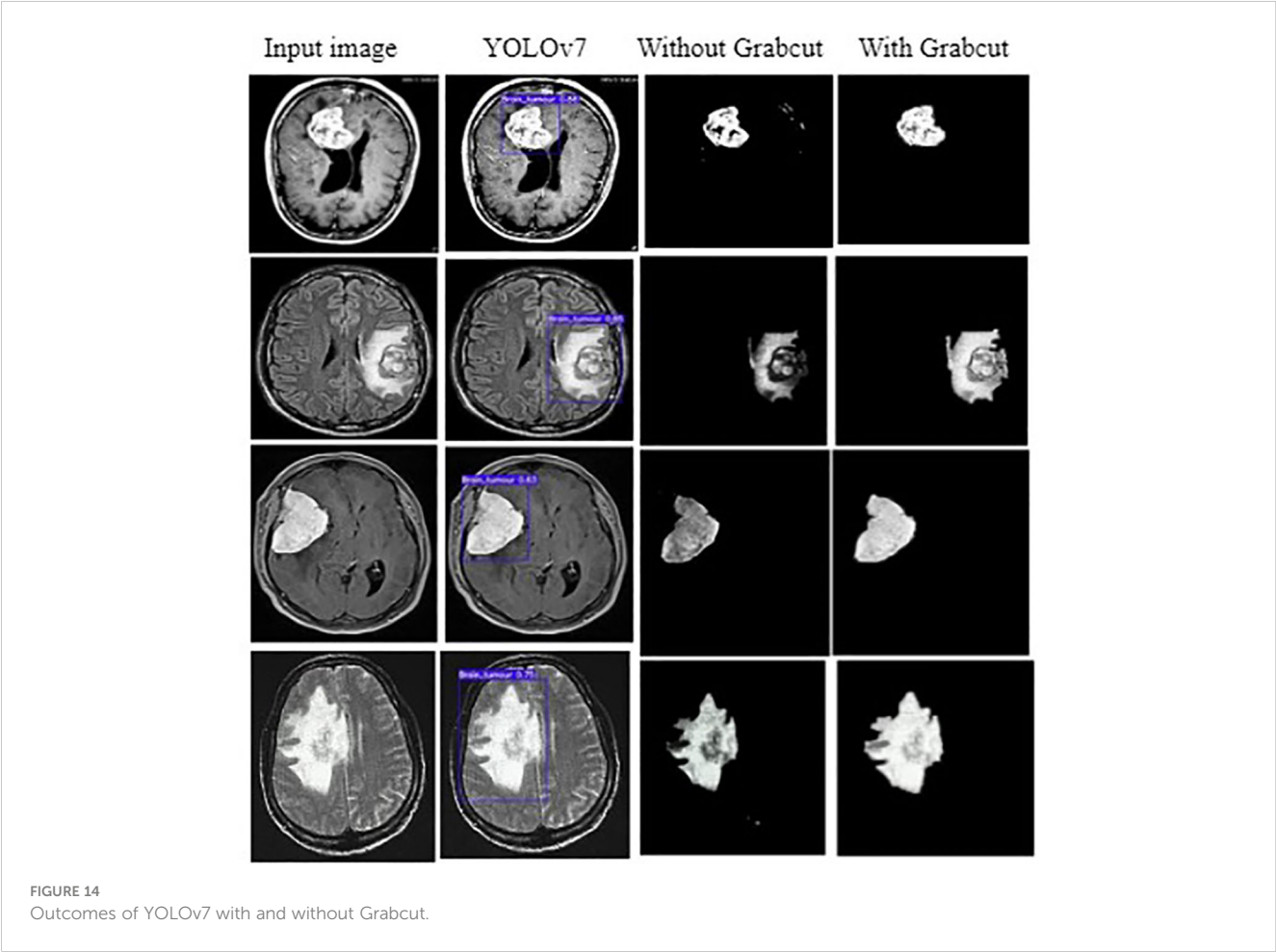
## 5 Conclusion

This paper deploys a new BT detection and extraction method using deep convolution neural network YOLOv7 in combination with Grab cut algorithm. This approach detects the salient images for accurate results. The proposed model involved different stages of

preprocessing like noise removal, image resizing, thresholding and RGB to gray conversion. The tumor image is converted to grayscale before being segmented and corrected using the Gamma correction process based on the threshold level. Our methodology provides better resolution and dimension-independent segmentation outcomes than the prior deep learning-based detection

TABLE 6 Dice similarity measurement of various models with Grab Cut algorithm.

|                                 | Hybrid CNN-SVM +<br>Grab Cut | YOLOv5 +<br>Grab Cut | Yolov6 +<br>Grab Cut | Proposed Yolov7 +<br>Grab Cut |
|---------------------------------|------------------------------|----------------------|----------------------|-------------------------------|
| Dice<br>Similarity Measurements | 0.3328                       | 0.8105               | 0.8190               | 0.9147                        |



techniques. We evaluated our method using BR35H: Brain Tumor Detection 2020 (BR35H) dataset. Results show that the YOLOv7 model in combination with grab cut achieves an outstanding accuracy of 99% in comparison to existing hybrid CNN-SVM, YOLOv5 and YOLOv6 models. The outcome of the analysis reveals that the YOLOv7 model is fast compared to the other models. In addition, the YOLOv7 accurately detects and extracts BT in the presence of the Grab Cut algorithm. This approach is best identified for BT detection when implemented for larger datasets. This model can be extended in the future to explore various types of tumors from the extracted tumor for accurate diagnosis.

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.

Author contributions

SK: Conceptualization, Formal analysis, Methodology, Writing – original draft. YK: Supervision, Validation, 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

- Jalali V, Kaur D. A study of classification and feature extraction techniques for brain tumor detection. *Int J Multimedia Inf Retrieval*. (2020) 9:271–90. doi: 10.1007/s13735-020-00199-7
- Amran GA, Alsharam MS, Blajam AOA, Hasan AA, Alfaifi MY, Amran MH, et al. Brain tumor classification and detection using hybrid deep tumor network. *Electronics*. (2022) 11:3457. doi: 10.3390/electronics11213457
- Sarah AVA, Yepuganti K, Kripa K, Srigiri Krishna P, Saritha S. "Brain Tumor Detection Using Deep Learning with Synthetic Data Augmentation." In *2023 IEEE 12th International Conference on Communication Systems and Network Technologies (CSNT)*, pp. 164–170. IEEE, 2023.
- Metwali H, Raemaekers M, Kniese K, Kardavani B, Fahlbusch R, Samii A. Reliability of functional magnetic resonance imaging in patients with brain tumors: a critical review and meta-analysis. *World Neurosurg*. (2019) 125:183–90. doi: 10.1016/j.wneu.2019.01.194
- Ramakrishnan T, Sankaragomathi B. A professional estimate on the computed tomography brain tumor images using SVM-SMO for classification and MRG-GWO for segmentation. *Pattern Recognition Lett*. (2017) 94:163–71. doi: 10.1016/j.patrec.2017.03.026
- Lohmann P, Werner JM, Shah NJ, Fink GR, Langen KJ, Galldiks N. Combined amino acid positron emission tomography and advanced magnetic resonance imaging in glioma patients. *Cancers*. (2019) 11:153. doi: 10.3390/cancers11020153
- Windisch P, Weber P, Fürweger C, Ehret F, Kufeld M, Zwahlen D, et al. Implementation of model explainability for a basic brain tumor detection using convolutional neural networks on MRI slices. *Neuroradiology*. (2020) 62:1515–8. doi: 10.1007/s00234-020-02465-1
- Amin J, Sharif M, Gul N, Raza M, Anjum MA, Nisar MW, et al. Brain tumor detection by using stacked autoencoders in deep learning. *J Med Syst*. (2020) 44:1–12. doi: 10.1007/s10916-019-1483-2
- Roberts TA, Hyare H, Agliardi G, Hipwell B, d'Esposito A, Ianus A, et al. Noninvasive diffusion magnetic resonance imaging of brain tumour cell size for the early detection of therapeutic response. *Sci Rep*. (2020) 10:1–13. doi: 10.1038/s41598-020-65956-4
- Sarhan AM. Brain tumor classification in magnetic resonance images using deep learning and wavelet transform. *J Biomed Sci Eng*. (2020) 13:102. doi: 10.4236/jbise.2020.136010
- Wang Y, Zu C, Hu G, Luo Y, Ma Z, He K, et al. Automatic tumor segmentation with deep convolutional neural networks for radiotherapy applications. *Neural Process Lett*. (2018) 48:1323–34. doi: 10.1007/s11063-017-9759-3
- Mubashar M, Ali H, Grönlund C, Azmat S. R2U++: a multiscale recurrent residual U-Net with dense skip connections for medical image segmentation. *Neural Computing Appl*. (2022) 34:17723–39. doi: 10.1007/s00521-022-07419-7
- Raza A, Ayub H, Khan JA, Ahmad I, S. Salama A, Daradkeh YI, et al. A hybrid deep learning-based approach for brain tumor classification. *Electronics*. (2022) 11:1146. doi: 10.3390/electronics11071146
- Khan MA, Ashraf I, Alhaisoni M, Damaševičius R, Scherer R, Rehman A, et al. Multimodal brain tumor classification using deep learning and robust feature selection: A machine learning application for radiologists. *Diagnostics*. (2020) 10:565. doi: 10.3390/diagnostics10080565
- Sadat T, Rehman A, Munir A, Saba T, Tariq U, Ayesha N, et al. Brain tumor detection and multi-classification using advanced deep learning techniques. *Microscopy Res Technique*. (2021) 84:1296–308. doi: 10.1002/jemt.23688
- Vani N, Sowmya A, Jayamma N. Brain tumor classification using support vector machine. *Int Res J Eng Technol (IRJET)*. (2017) 4:792–6.
- Anitha R, Siva Sundhara Raja D. Development of computer-aided approach for brain tumor detection using random forest classifier. *Int J Imaging Syst Technol*. (2018) 28:48–53. doi: 10.1002/ima.22255
- Garg G, Garg R. Brain tumor detection and classification based on hybrid ensemble classifier. *arXiv*. (2021). doi: 10.48550/arXiv.2101.00216
- Ahmad I, Liu Y, Javed D, Shamshad N, Sarwr D, Ahmad S. "A review of artificial intelligence techniques for selection & evaluation," In: *IOP Conference Series: Materials Science and Engineering*, IOP Publishing. (2020), Vol. 853. p. 012055.
- Bhatele KR, Bhadauria SS. Brain structural disorders detection and classification approaches: a review. *Artif Intell Rev*. (2020) 53:3349–401. doi: 10.1007/s10462-019-09766-9
- Sri Sabarimani K, Arthi R. A brief review on Brain Tumour detection and classifications. *Bio-inspired Neurocomputing*. (2021), 61–72. doi: 10.1007/978-981-15-5495-7\_4
- Gull S, Akbar S. Artificial intelligence in brain tumor detection through MRI scans: Advancements and challenges. *Artif Intell Internet Things*. (2021), 241–76. doi: 10.1201/9781003097204-10
- Bhagat PK, Choudhary P, Singh KM. "A comparative study for brain tumor detection in MRI images using texture features." In: *Sensors for health monitoring*. Academic Press (2019). p. 259–87. doi: 10.1016/B978-0-12-819361-7.00013-0
- Durgamahanthi V, Anita Christaline J, Shirley Edward A. "GLCM and GLRLM based texture analysis: application to brain cancer diagnosis using histopathology images," In: *Intelligent Computing and Applications: Proceedings of ICICA 2019*. Springer Singapore (2021). p. 691–706.
- Krishnapriya S, Yepuganti K. "A survey of deep learning for MRI brain tumor segmentation methods: Trends, challenges, and future directions." *Health and Technology* 13, no. 2 (2023): 181–201.
- Choudhury CL, Mahanty C, Kumar R, Mishra BK. (2020). "Brain tumor detection and classification using convolutional neural network and deep neural network," In: *2020 international conference on computer science, engineering and applications (ICCSEA)*, IEEE. pp. 1–4. doi: 10.1109/ICCSEA49143.2020.9132874
- Gupta RK, Bharti S, Kunhare N, Sahu Y, Pathik N. Brain tumor detection and classification using cycle generative adversarial networks. *Interdiscip Sciences: Comput Life Sci*. (2022) 14:485–502. doi: 10.1007/s12539-022-00502-6
- Srigiri K, Yepuganti K. "Pre-trained deep learning models for brain MRI image classification". *Front. Hum. Neurosci.*, (2023) Brain-Computer Interfaces Volume 17 – 2023.
- Dipu NM, Shohan SA, Salam KMA. "Deep learning-based brain tumor detection and classification," In: *2021 International Conference on Intelligent Technologies (CONIT)*, IEEE. (2021). pp. 1–6.
- Chanu MM, Singh NH, Muppala C, Prabu RT, Singh NP, Thongam K. Computer-aided detection and classification of brain tumor using YOLOv3 and deep learning. *Soft Computing*. (2023) 27:9927–40. doi: 10.1007/s00500-023-08343-1
- Dixit A, Singh P. Brain tumor detection using fine-tuned YOLO model with transfer learning. In: *Artificial Intelligence on Medical Data: Proceedings of International Symposium, ISCM 2021*. Springer Nature Singapore, Singapore (2022). p. 363–71.
- Paul S, Ahad DMT, Hasan MM. Brain cancer segmentation using YOLOv5 deep neural network. *arXiv*. (2022).
- Shelatkar T, Urvashi D, Shorfuazzaman M, Alsufyani A, Lakshmana K. Diagnosis of brain tumor using light weight deep learning model with fine-tuning approach. *Comput Math Methods Med*. (2022). 2022 doi: 10.1155/2022/2858845
- Arunachalam S, Sethumathavan G. An effective tumor detection in MR brain images based on deep CNN approach: i-YOLOV5. *Appl Artif Intell*. (2022) 36:2151180. doi: 10.1080/08839514.2022.2151180
- Hossain A, Islam MT, Almutairi AF. A deep learning model to classify and detect brain abnormalities in portable microwave based imaging system. *Sci Rep*. (2022) 12:6319. doi: 10.1038/s41598-022-10309-6
- Labcharoenwongs P, Vongansansup S, Chunhapran O, Noolek D, Yampaka T. An automatic breast tumor detection and classification including automatic tumor volume estimation using deep learning technique. *Asian Pacific J Cancer Prev*. (2023) 24:1081–8. doi: 10.31557/APJCP.2023.24.3.1081
- Saeed S, Abdullah A, Jhanjhi N, Naqvi M, Masud M, AlZain MA. Hybrid GrabCut hidden Markov model for segmentation. *Comput Mater Contin*. (2022) 72:851–69. doi: 10.32604/cmc.2022.024085
- Hamada. Br35h: Brain Tumor Detection 2020, version 5 (2020). Available online at: <https://www.kaggle.com/datasets/ahmedhamada0/brain-tumor-etecntion?select=yes>.
- Wang CY, Bochkovskiy A, Liao HYM. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, in: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 7464–75.



## OPEN ACCESS

## EDITED BY

Poonam Yadav,  
Northwestern University, United States

## REVIEWED BY

Rahul Gupta,  
Synergy Institute of Medical Sciences, India  
Bilgin Kadri Aribas,  
Bülent Ecevit University, Türkiye

## \*CORRESPONDENCE

Song-Song Wu  
✉ wu\_songsong@126.com

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

RECEIVED 19 September 2023

ACCEPTED 21 March 2024

PUBLISHED 24 April 2024

## CITATION

Zeng X-T, Liang X, Hong Z-L, Chen S,  
Yang J-C, Lin Y-c and Wu S-S (2024) Initial  
investigation on ultrasound-guided  
percutaneous biopsy of lesions in the first  
hepatic hilum with fusion of ultrasound and  
multimodal imaging cognitive guidance.  
*Front. Oncol.* 14:1297153.  
doi: 10.3389/fonc.2024.1297153

## COPYRIGHT

© 2024 Zeng, Liang, Hong, Chen, Yang, Lin  
and Wu. 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.

# Initial investigation on ultrasound-guided percutaneous biopsy of lesions in the first hepatic hilum with fusion of ultrasound and multimodal imaging cognitive guidance

Xian-Tao Zeng<sup>1,2,3†</sup>, Xia Liang<sup>1,2,3†</sup>, Zhi-Liang Hong<sup>1,2,3†</sup>,  
Sheng Chen<sup>1,2,3</sup>, Jian-Chuan Yang<sup>1,2,3</sup>, Yu-cheng Lin<sup>1,4</sup>  
and Song-Song Wu<sup>1,2,3\*</sup>

<sup>1</sup>Shengli Clinical Medical College of Fujian Medical University, Fuzhou, China, <sup>2</sup>Fujian Provincial Key Laboratory of Critical Care Medicine, Fuzhou, China, <sup>3</sup>Department of Ultrasound, Fujian Provincial Hospital, Fuzhou, China, <sup>4</sup>Department of Ultrasonography, Affiliated Fuzhou First Hospital of Fujian Medical University, Fuzhou, China

**Purpose:** This study aims to evaluate the efficacy and safety of ultrasound-guided percutaneous biopsy of the first hepatic hilum lesion, and examine its clinical value of diagnosis and treatment.

**Methods:** We conducted a retrospective study on patients diagnosed with the first hepatic hilum lesions at Fujian Provincial Hospital between February 2015 and October 2022. We selected patients who had lesions in the first hepatic hilum (including a 2cm surrounding area of the left/right hepatic ducts and upper-middle segment of the common bile duct) and the liver periphery (in the peripheral area of the liver, outside of the above-mentioned first hepatic porta region). These patients underwent percutaneous ultrasound-guided core needle biopsy (PUS-CNB) with cognitive fusion guidance using CT, MRI, or PET-CT. We compared the safety and efficacy of PUS-CNB in the first hepatic hilum and the liver periphery to explore the value of PUS-CNB in optimizing the clinical treatment of the first hepatic hilum lesions.

**Results:** The study includes 38 cases of the first hepatic hilum cases (18 females; 20 males), 23 presented with mass-forming tumors while the remaining 15 exhibited diffuse infiltrative tumors, with an average diameter of  $4.65 \pm 2.51$  cm. The percutaneous biopsy procedure, conducted under ultrasound guidance, had an average operation time of  $14.55 \pm 2.73$  minutes, and resulted in a postoperative bleeding volume of approximately  $10.79 \pm 2.79$  ml. The diagnostic success rate was noted to be as high as 92.11% among the participants who underwent percutaneous biopsy of the first hepatic hilum. Procedural complications, such as bleeding, bile leakage, intestinal perforation, infection or needle tract seeding, did not occur during or after the biopsy procedure. Affected by biopsy results, 5 altered their clinical treatment plans accordingly, 24 patients received non-surgical treatment, 9 underwent surgical treatment, 5 underwent radiofrequency ablation for the lesions. The study

comprised a total of 112 cases for percutaneous biopsy of the liver periphery. The safety and effectiveness of the two biopsy techniques were comparable, with diagnostic success rates of 92.11% VS. 94.34%, respectively ( $p = 0.61$ ).

**Conclusion:** Cognitive fusion of ultrasound and multi-modal imaging for the first hepatic hilum lesion puncture biopsy is a safe and effective diagnostic procedure, with better diagnostic rate, may improve clinical value of diagnosis and treatment of various diseases.

#### KEYWORDS

multi-modal imaging, cognitive fusion, first hepatic hilum, ultrasound-guided, core needle biopsy

## 1 Introduction

The first hepatic hilum, located on the visceral surface of the liver, is a complex anatomical region where multiple structures converge, including the portal vein, hepatic artery, bile duct, lymphatics, nerves, and connective tissue. Lesions in this area pose a diagnostic challenge as imaging features of benign and malignant tumors can overlap, resulting in approximately 13-15% of preoperatively suspected malignancies being diagnosed as benign (1, 2). Clinicians should be aware of these challenges when evaluating patients with suspected lesions in the first hepatic hilum. Surgical resection of lesions located in the first hepatic hilum represents a formidable challenge, with reported rates of severe complications ranging from 37-64%, and surgical mortality rates of 8-10% (3). As such, obtaining an accurate histological diagnosis prior to surgical intervention is essential for ensuring proper diagnosis and appropriate surgical management. This crucial step can help mitigate the risk of adverse outcomes and improve patient outcomes (4). The most commonly used biopsy techniques for diagnosing suspected malignancies in the bile duct include cytology using brush cytology during endoscopic retrograde cholangiopancreatography (ERCP) (5), forceps biopsy, and endoscopic ultrasound-guided fine-needle aspiration (EUS-FNA). However, ERCP is primarily used to sample the bile ducts and has a detection rate of 44% to 80% for suspected cholangiocarcinoma in that area (6, 7). The malignant tumor detection rate for forceps biopsy is 43% to 81% (8, 9). EUS-FNA demonstrates slightly higher diagnostic sensitivity (45%-86%) (10, 11). But lower negative predictive value (9%-38%) compared to previous techniques for obtaining pathological tissue of the first hepatic hilum lesions (12, 13). False negatives may occur, highlighting the need for alternative approaches to safely and effectively diagnose advanced malignant tumors and obtain tissue for genetic testing to facilitate targeted and immunotherapy.

Hence, there is an urgent need for novel strategies that can guarantee the safe and effective acquisition of pathological tissue from lesions in the first hepatic hilum, thereby enabling precise

preoperative diagnosis. We conducted a study to explore a novel and safe percutaneous biopsy technique for detecting the first hepatic hilum lesions. The primary objective is to bolster diagnostic precision, furnish valuable insights for clinical deliberation, and promote better treatment selection, particularly in the context of malignant tumors.

## 2 Materials and methods

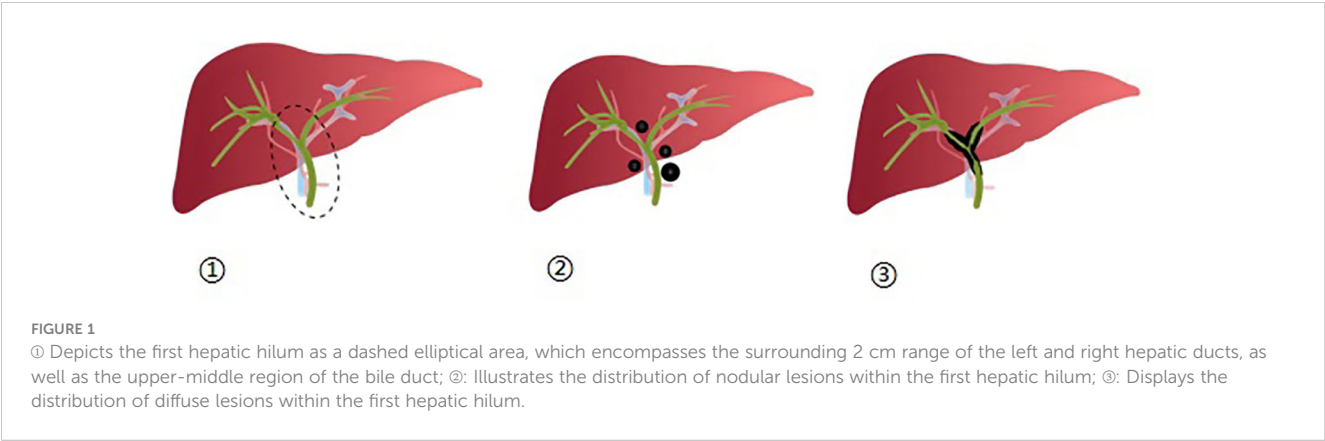
### 2.1 The subjects of study

This retrospective study encompassed cases between February 2015 and October 2022 at Fujian Provincial Hospital. The patients who were included in the study exhibited lesions within the first hepatic hilum (including a 2cm surrounding area of the left/right hepatic ducts and upper-middle segment of the common bile duct) (Figure 1) and the liver periphery (in the peripheral area of the liver, outside of the above-mentioned first hepatic porta region). Inclusion criteria: (a) Imaging findings indicating the first hepatic hilum or perihepatic masses requiring a definitive pathological diagnosis; (b) Availability of a safe and suitable transabdominal puncture route; (c) Platelet count  $>50 \times 10^9/L$ ; (d) Prothrombin time ratio  $\geq 70\%$ . Exclusion criteria: ① Severe coagulation dysfunction; ② Severe underlying diseases or physical weakness (Figure 2).

This retrospective analysis has been approved by the Institutional Review Board of Fujian Provincial Hospital. Written informed consent was obtained from each patient prior to treatment, with all patients being informed of the treatment indications, potential therapeutic effects, as well as possible complications that may arise.

### 2.2 Instruments and methods

In our clinical setting, We utilize the Philips iU22 and GE Vivid 7 Dimension color Doppler ultrasound diagnostic machines for our



ultrasound equipment. The linear array probe operates at a frequency of 5-12MHz, while the convex array probe operates at a frequency of 2-5MHz. For our biopsy procedures, we employ either the 18G fully automatic biopsy gun or the semi-automatic option (BARD Magnum, MN18-20, CR).

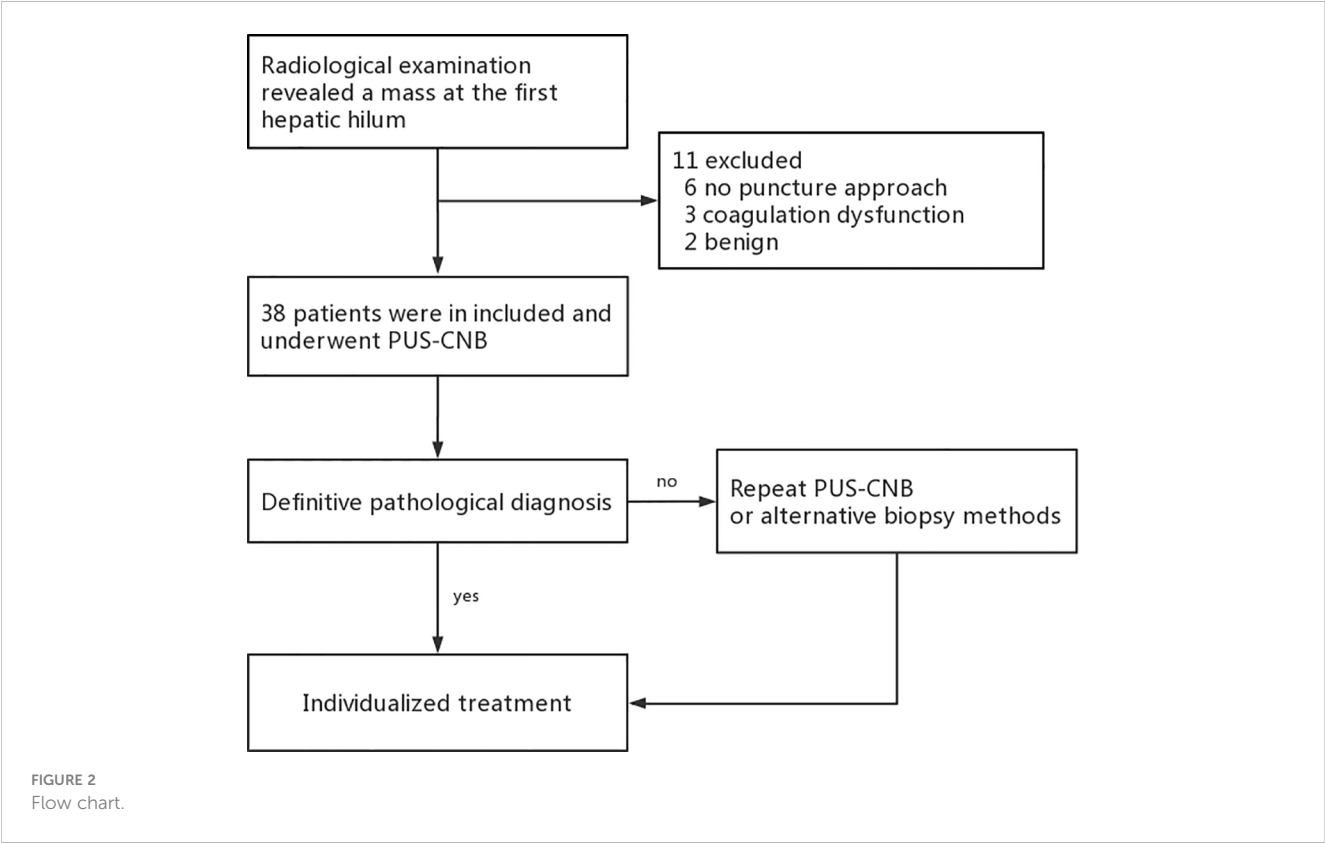
2.3 Pre-puncture preparation

All patients underwent routine laboratory examinations, including complete blood counts, biochemical tests, and coagulation function assessments. Prior to the surgery, all patients underwent CT/MRI/PET-CT imaging to evaluate the first hepatic hilum and the surrounding areas using multimodal imaging. The surgical team

performed a multimodal imaging assessment prior to the puncture procedure to reconstruct the specific location of the first hepatic hilum lesion and its surrounding anatomical relationships. The lesion was scanned using an ultrasound probe prior to the puncture, and then the imaging image in the mind was fused with the ultrasonic image to determine the puncture passage (a. intercostal approach above the rib arch, b. subcostal approach through the liver parenchyma).

2.4 Ultrasound-guided percutaneous biopsy procedure

The procedure is performed in the ultrasound intervention room: The patient was positioned supine, The ideal puncture site



is identified through the process of cognitive fusion, using contrast-enhanced ultrasonography to visualize areas of abnormal enhancement (indicating an active lesion) and rapid washout (suggesting a suspicious region for malignant tumor). These findings guide the selection of the suspected active target area. The puncture site was sterilized and draped, and local anesthesia was administered to the area. Under real-time ultrasound guidance, the needle tip was observed to sequentially pass through the skin, subcutaneous tissue, and liver until reaching the lesion (taking care to avoid major blood vessels). The trigger (biopsy device) was activated to obtain tissue, and then the needle tip was withdrawn. All biopsies were performed under local anesthesia with 2% lidocaine. The biopsy procedure was carried out by two experienced ultrasound intervention physicians with over 10 years of experience. An 18G biopsy needle was used for the puncture, and after tissue sampling, the needle was quickly withdrawn. The quality of the tissue samples within the needle notch was observed, and if deemed unsatisfactory, additional samples were taken. Each time, 2–4 tissue samples measuring approximately 15mm–20mm in length were obtained for pathological diagnosis. After applying sterile dressing to the skin puncture site, patients were instructed to apply pressure to the site for 5 minutes and were observed for 30–60 minutes to check for any active bleeding. Additionally, patients were advised to rest in bed for 6 hours after the procedure, and no medication intervention was administered.

## 2.5 Evaluation indicators

**Diagnostic success rate assessment:** The diagnostic success rate was assessed as follows: Cases in which the biopsy pathology diagnosis was confirmed and matched the clinical discharge diagnosis or postoperative pathology diagnosis were considered successful cases. Cases in which the biopsy pathology diagnosis was confirmed but did not match the postoperative pathology diagnosis were considered non-matching cases. Cases in which the biopsy was unsuccessful and could not provide a definitive pathology diagnosis, requiring a repeat biopsy or surgical procedure for a definitive pathology diagnosis, were also considered non-matching cases. The diagnostic success rate was calculated as the number of cases with a confirmed pathology diagnosis on the first biopsy divided by the total number of cases.

**The effectiveness measure includes:** The average time spent during the operation.

**Safety indicators include (Complications of puncture):** Post-puncture, the amount of bleeding from the skin puncture site is evaluated. Following the surgery and after one hour, careful observation is conducted to assess the presence of fluid collections within the abdominal and pelvic cavities caused by active bleeding, as well as to check for any apparent damage to surrounding organs. The patients are monitored for symptoms of infection, bile leakage, intestinal perforation, and other related symptoms. During the follow-up period, which lasts for three months after the surgery, the presence of needle tract implantation metastasis and other long-term complications related to the procedure are observed using methods such as reviewing medical records, real-time

communication with the attending physician, and conducting telephone follow-ups after the patient's discharge.

## 2.6 Statistical processing

This study employed SPSS 26.0 statistical software and considered  $p < 0.05$  (two-tailed test) as the level of statistical significance. Independent sample t-tests and Mann-Whitney U tests were utilized for comparing quantitative data, while count data was represented using frequency with between-group differences assessed via chi-square tests and Fisher's exact probability method.

## 3 Results

### 3.1 General information

In this study, a total of 49 patients with lesions in the first hepatic hilum were selected for percutaneous ultrasound-guided core needle biopsy (PUS-CNB). Among them, 11 cases were excluded due to the lack of a safe puncture pathway or abnormal coagulation function. The remaining 38 cases underwent PUS-CNB (20 males and 18 females) with a mean age of  $61.08 \pm 11.29$  years (ranging from 30 to 77 years). The clinical symptoms upon admission included abdominal pain in 28 cases, jaundice in 15 cases, vomiting in 5 cases, constipation in 4 cases, and a less common symptom of recurrent melena in 1 case. The types of lesions included 23 cases of mass-type and 15 cases of diffuse infiltration, with an average diameter of  $4.65 \pm 2.51$  cm. Prior to the puncture, 30 cases underwent MRI examination, 7 cases underwent CT examination, and 2 cases underwent PET-CT examination. Based on cognitive fusion guidance, the selected puncture approaches were intercostal transhepatic in 32 cases and subcostal transhepatic in 6 cases (See Table 1).

### 3.2 Safety and effectiveness of the puncture

The average duration of PUS-CNB procedure for the first hepatic hilum lesions was  $14.55 \pm 2.73$  minutes. The postoperative bleeding volume was approximately  $10.79 \pm 2.79$  ml. All patients experienced mild discomfort, and no complications such as bleeding, bile leakage, intestinal perforation, or infection were observed during or after the procedure. There were no evident needle tract metastases during the six-month follow-up. The success rate of tissue sampling through the biopsy procedure was 100%. The biopsy results revealed 19 cases of cholangiocarcinoma, 2 cases of gallbladder carcinoma, 5 cases of hepatocellular carcinoma, 1 case of diffuse large B-cell lymphoma, 4 cases of inflammatory lesions, 1 case of parasitic infection, 5 cases of metastatic tumors (non-hepatic or biliary in origin), and 1 case of malignant pheochromocytoma. Among them, three cases initially diagnosed as inflammatory



TABLE 1 General information for both groups.

| index                               | first hepatic hilum | liver periphery |
|-------------------------------------|---------------------|-----------------|
| Total number of cases               | 38                  | 112             |
| Gender                              |                     |                 |
| Male                                | 20                  | 66              |
| Female                              | 18                  | 46              |
| Age (years)                         |                     |                 |
| Average                             | 61.08 ± 11.29       | 58.79 ± 12.42   |
| Focal type                          |                     |                 |
| The mass type                       | 23                  | 94              |
| diffuse type                        | 15                  | 18              |
| Diameter of the lesion              | 4.65 ± 2.51         | 5.59 ± 3.85     |
| puncture channel                    |                     |                 |
| intercostal liver puncture approach | 32                  | 106             |
| Subcostal space puncture approach   | 6                   | 6               |

lesions were confirmed as cholangiocarcinoma through surgery, resulting in a biopsy success rate of 92.11% (See [Tables 2, 3](#)).

3.3 Compared with perihepatic puncture biopsy

A total of 112 patients underwent ultrasound-guided percutaneous needle biopsy of the liver periphery with an average operation time of 14.24 ± 3.01 minutes and an average postoperative bleeding volume of approximately 10.16 ± 2.30 ml. All patients tolerated the procedure well, experiencing only mild pain, and no major complications such as significant bleeding, bile

leakage, intestinal perforation, or infection were observed during or after the procedure. During the six-month follow-up, there were no evident needle tract metastases. The overall success rate of tissue sampling through the biopsy procedure was 94.64%, which did not show a significant difference compared to the success rate of the first hepatic hilum biopsies (P = 0.57). (See [Table 3](#)).

3.4 Post-puncture changes in diagnosis and treatment

After undergoing the first hepatic hilum biopsy, 5 patients had a change in their treatment plans ([Table 4](#)):

One patient with postoperative lung cancer was found to have the first hepatic hilum lesion on follow-up imaging, which raised suspicion of a possible metastatic tumor. However, PUS-CNB confirmed it as malignant pheochromocytoma, leading to a referral for urological treatment ([Figure 3](#)).

One patient with postoperative gastric cancer was found to have a first hepatic hilum lesion on follow-up imaging, along with a significant increase in AFP levels. Clinical considerations pointed to primary liver cancer, but PUS-CNB confirmed it as a metastatic carcinoma, resulting in chemotherapy treatment.

One patient who was suspected of having a primary tumor in the first hepatic hilum with multiple liver metastases underwent PUS-CNB, which diagnosed a parasitic infection. The treatment plan was then changed to internal medicine, and follow-up at 1 year showed the disappearance of the lesion ([Figure 4](#)).

One patient who was being evaluated for fever underwent lymph node excision biopsy in the neck, which revealed reactive proliferation on pathology. However, the treatment outcome was unsatisfactory. Subsequent PET-CT revealed a high metabolic lesion in the first hepatic hilum, and PUS-CNB confirmed it as diffuse large B-cell lymphoma. The patient was then referred to hematology for chemotherapy treatment.

One patient who was suspected of having cholangiocarcinoma based on imaging findings had a PUS-CNB that diagnosed IgG4-

TABLE 2 Safety and efficacy comparison of the two groups.

| index                     | first hepatic hilum | liver periphery | t    | p    |
|---------------------------|---------------------|-----------------|------|------|
| Total number of cases     | 38                  | 112             |      |      |
| validity index            |                     |                 |      |      |
| puncture time(min)        | 14.55 ± 2.73        | 14.24 ± 3.01    | 0.59 | 0.56 |
| Diagnostic success rate   | 92.11%              | 94.64%          | /    | 0.57 |
| Safety indicator          |                     |                 |      |      |
| bleeding Volume(ml)       | 10.79 ± 2.79        | 10.16 ± 2.30    | 1.25 | 0.22 |
| infect                    | 0                   | 0               |      |      |
| Perforation               | 0                   | 0               |      |      |
| Bile leakage              | 0                   | 0               |      |      |
| needle track implantation | 0                   | 0               |      |      |
| postoperative hemorrhage  | 0                   | 0               |      |      |

TABLE 3 Pathological results of the two groups.

| index                         | first hepatic hilum | liver periphery |
|-------------------------------|---------------------|-----------------|
| cholangiocarcinoma            | 19                  | 9               |
| hepatocellular carcinoma      | 5                   | 23              |
| diffuse large B cell lymphoma | 1                   | 0               |
| Inflammatory lesions          | 4                   | 9               |
| parasitic infection           | 1                   | 0               |
| metastatic tumor              | 5                   | 69              |
| Malignant pheochromocytoma    | 1                   | 0               |
| carcinoma of gallbladder      | 2                   | 0               |
| Neuroendocrine tumor          | 0                   | 1               |
| angioma                       | 0                   | 1               |
| Total number of cases         | 38                  | 112             |

related sclerosing cholangitis. The treatment plan was then changed to internal medicine.

In the case of an advanced-stage cholangiocarcinoma patient, gene testing after PUS-CNB indicated high PD-L1 expression (90%). The patient received immunotherapy with Keytruda, receiving a total of 22 doses, with the first two doses at 200mg and the subsequent 20 doses at 100mg. During this period, the patient also underwent one session of radiation therapy with a dose of 500GY\*6. Follow-up PET-CT showed that the lesions had mostly disappeared, with no metabolic enhancement, indicating a significant treatment response. The patient has survived for 3 years and maintains a high quality of life (Figure 5).

Five patients with metastatic cancer underwent radiofrequency ablation treatment, and follow-up MR scans showed complete lesion inactivation in all cases. In total, 6 patients with

cholangiocarcinoma, 2 patients with hepatocellular carcinoma, and 1 patient with chronic inflammation underwent surgical treatment. The remaining 24 patients had advanced-stage malignant tumors and received non-surgical treatment (Figure 6).

## 4 Discussion

Malignant tumors are commonly found in the first hepatic hilum, with cholangiocarcinoma, gallbladder carcinoma, metastatic carcinoma, and lymphoma being the main types. CT and MRI are widely accepted methods for further diagnosing the first hepatic hilum tumors, with MRI being the preferred imaging modality for evaluating histological characteristics and the presence of diffusion. However, even with imaging, it is not possible to completely differentiate between primary cholangiocarcinoma, metastatic carcinoma, hepatocellular carcinoma, and lymphoma (14, 15). Different tumor types require different treatment strategies, and an accurate diagnosis is crucial for clinical staging and treatment planning. Currently, the main methods of biopsy include brush cytology under ERCP, forceps biopsy, and fine-needle aspiration under endoscopic ultrasound guidance. Unfortunately, these biopsy methods are mainly focused on cytological diagnosis and have a lower sensitivity. Therefore, there is a need to explore a new, safe, and effective biopsy method.

Ultrasonography-guided percutaneous core needle biopsy (PUS-CNB) has been validated as a feasible modality for obtaining liver-peripheral tissue samples (16, 17). However, the complex anatomical structure and proximity to neighboring organs in the first hepatic hilum pose challenges in acquiring satisfactory pathological specimens. Our research has revealed that the fusion of ultrasonography with multimodal imaging cognition enables PUS-CNB to emerge as a viable technique, with a diagnostic success rate of 92.11%. In a large series of 1300 patients, accuracy rate was found as 92.8% (499/538) in total, 85% (159/187) in small needle group vs. 96.9% (340/351) in large needle group with the liver metastases ( $P<0.001$ ). These rates were 91% (91/100) in total, 85.5% (47/55) in small needle group vs. 97.9% (44/45 in large needle group with hepatocellular carcinomas ( $P=0.039$ ). Among 100 hepatocellular carcinomas, 18% were well-differentiated, 26% and 56% were moderate-differentiated and poor-differentiated, respectively. Biopsies of hepatocellular carcinomas were performed with large needles in 45% (45/100) and with small needles in 55% (55/100) (18).In comparison, the diagnostic success rate of endoscopic ultrasound-guided fine-needle aspiration (EUS-FNA), as reported by Krister Jones et al., stands at 32% (19). PUS-CNB significantly outperforms cytology-based biopsy methods (8–11). Moreover, our study has demonstrated the safety and efficacy of PUS-CNB for the first hepatic hilum lesions, as evidenced by the absence of major complications such as significant hemorrhage, bile leakage, intestinal perforation, infection, or needle tract seeding. In our experience, two factors have contributed to our success: Firstly, our ultrasonography interventionalists possess over a decade of expertise in procedural interventions, exhibiting a high level of proficiency and skill. Secondly, ultrasonography provides real-time dynamic visualization, facilitating the precise navigation around

TABLE 4 Effect of PUS-CNB on the treatment.

| Type | Effect of ultrasound-guided needle biopsy on therapy                               | Example number | Percentage of cases |
|------|------------------------------------------------------------------------------------|----------------|---------------------|
| ET1  | Leading to inappropriate treatment options.                                        | 0              | 00.00               |
| ET2  | There was no effect on the choice of the treatment method                          | 1              | 2.63                |
| ET3  | No change the choice of treatment but increased clinician confidence in the choice | 27             | 71.05               |
| ET4  | Has an important role in the choice of treatment.                                  | 5              | 13.16               |
| ET5  | Change in the treatment approach chosen                                            | 5              | 13.16               |
|      | total                                                                              | 38             |                     |

ET, Effect on treatment.

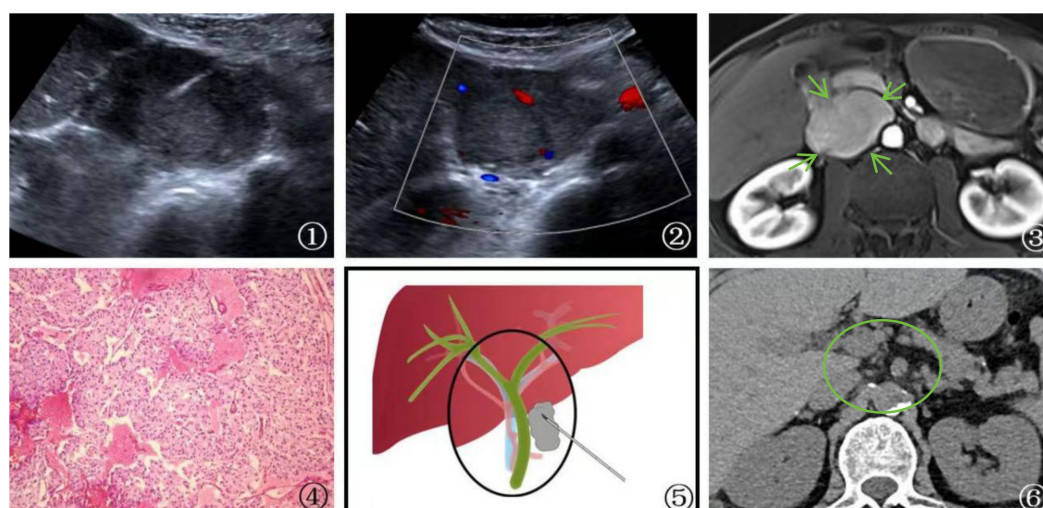


FIGURE 3

A 58-year-old male patient Presented with malignant pheochromocytoma. ①: Percutaneous biopsy of the first hepatic hilum was performed under ultrasound guidance. ②: Ultrasonography revealed a hyperechoic area in the first hepatic hilum with clear borders, regular shape, heterogeneous internal echogenicity, and slight blood flow signal. ③: MRI T1 arterial phase showed mild to moderate enhancement of the first hepatic hilum lesion (arrow), suggesting the possibility of metastatic tumor. ④: Schematic diagram of percutaneous biopsy of the first hepatic hilum under ultrasound guidance. ⑤: The neoplasm displays a fascicular and lobular architecture with active proliferative activity and local tissue invasion. Cystic degeneration is noted, and malignant tumor emboli are identified within the neoplastic thrombosed vessels and adjacent fibrovascular stroma. These features fulfill the criteria for malignant pheochromocytoma per the 2016 World Health Organization Classification of Neuroendocrine Tumors (magnification, x10). ⑥: After 1 year of follow-up CT examination post-surgery, the lesion was completely excised (circle) and no evidence of recurrence was detected.

vital anatomical structures during the needle puncture. Efforts should be made to identify the shortest trajectory for needle insertion, avoiding critical organ structures and favoring a lateral approach along blood vessels and bile ducts. Nonetheless, there

were three cases in which the intended pathological diagnosis was not successfully obtained, resulting in false negatives. In these instances, the final histopathological findings revealed cholangiocarcinoma, characterized by an abundance of fibrous

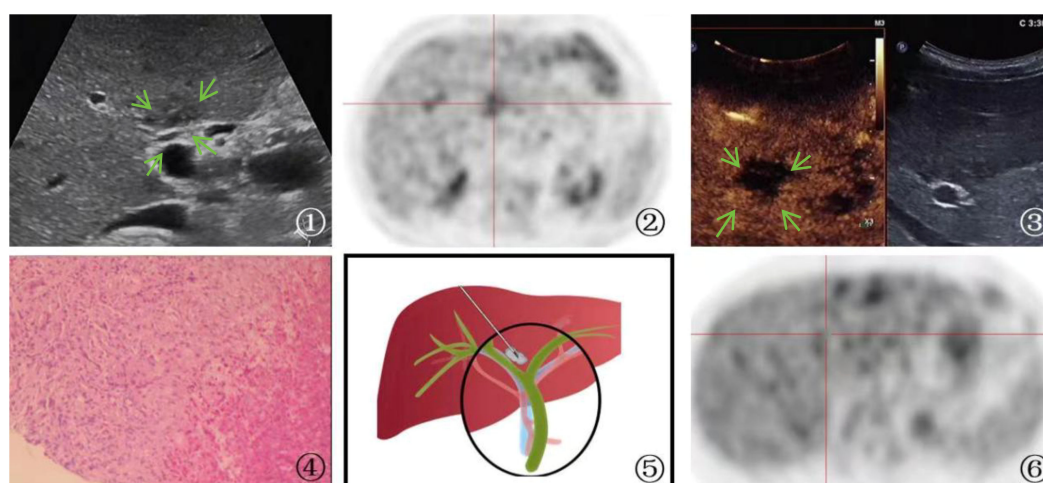


FIGURE 4

A 65-year-old man Presented with parasitic infection. ①: On ultrasound imaging, multiple patchy hypoechoic lesions (arrow) were identified surrounding the bile duct, with one particularly located at the first hepatic hilum. The boundary of the lesion is indistinct, and its irregular shape is accompanied by internal echogenicity that appears uneven. ②: PET-CT scan showed hypermetabolic activity in the first hepatic hilum lesion, indicating the possibility of tumor. ③: Ultrasound contrast demonstrated rapid wash-in and wash-out enhancement (arrow) in the first hepatic hilum lesion. ④: A schematic diagram was provided to illustrate the percutaneous biopsy procedure of the first hepatic hilum under ultrasound guidance. ⑤: High magnification microscope, HE staining (magnification, x40); The biopsy sample of the liver demonstrated the occurrence of focal necrosis accompanied by the formation of granulomatous nodules, with a substantial infiltration of eosinophilic granulocytes and liver cell edema. The portal area displayed a chronic infiltration of inflammatory cells and fibrous tissue hyperplasia, indicating a parasitic infection. ⑥: After the therapeutic intervention, a subsequent PET-CT scan revealed a lack of metabolic elevation in the previously identified lesion located at the first hepatic hilum, indicative of its regression.

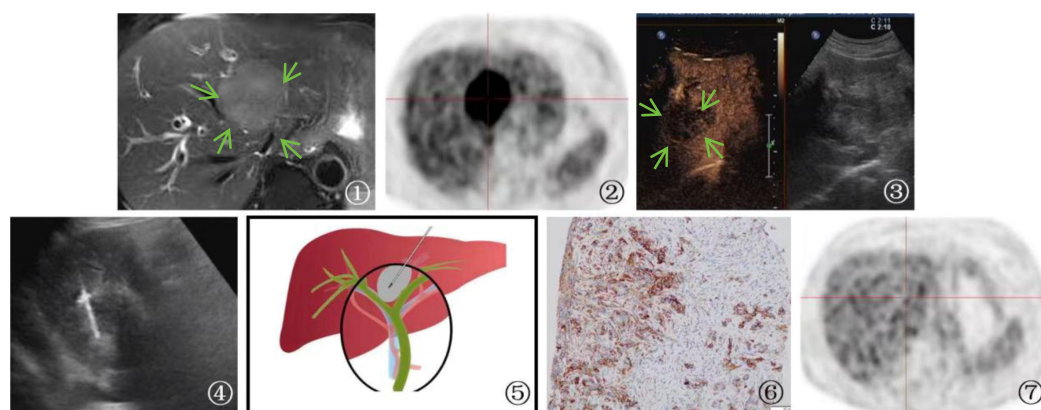


FIGURE 5

A 70-year-old male patient Presented with cholangiocarcinoma (a nodular type). ①: MRI T2-weighted magnetic resonance imaging showed a slightly elevated signal in the first hepatic hilum region(arrow), indicating a high likelihood of cholangiocarcinoma. ②: PET-CT shows high metabolic activity in the lesion, suggesting the possibility of malignant tumor. ③: An irregular hypoechoic mass(arrow) was detected in the first hepatic hilum during ultrasound examination. Further contrast-enhanced ultrasound revealed rapid wash-in and wash-out in the first hepatic hilum area, along with arterial phase enhancement. These findings suggest the presence of a tumor or lesion in the first hepatic hilum region. ④: Percutaneous biopsy of the first hepatic hilum was performed under ultrasound guidance. ⑤: A schematic diagram was provided to illustrate the percutaneous biopsy procedure of the first hepatic hilum under ultrasound guidance. ⑥: The immunohistochemical analysis for programmed death ligand 1 (PD-L1) demonstrated that the tumor cells expressed PD-L1 in a positive manner (TC+), with a positivity rate of 90%. ⑦: The imaging follow-up after immuno-oncology treatment presented a compelling outcome, revealing complete disappearance of the first hepatic hilum lesion on the PET-CT scan.

tissue and chronic inflammatory cells, which may have contributed to the failure of our biopsy procedure. Insufficient tumor cell representation within the biopsy specimens can lead to both overdiagnosis and underdiagnosis. Perhaps the utilization of a larger gauge biopsy needle, such as 16G, could ameliorate this issue. Additionally, increasing the number of biopsy attempts and adopting a multidirectional approach may also serve as potential improvements. Pre-procedural ultrasonography contrast enhancement could aid in identifying suitable biopsy sites, prioritizing areas with enhanced activity while avoiding necrotic regions. These strategies may potentially enhance the overall success rate of the biopsy procedure. It needs further investigation in this setting.

In this study, 5 cases had a change in treatment plan after PUS-CNB, highlighting the important role of PUS-CNB in guiding subsequent therapies. PUS-CNB had a positive impact on diagnosis and enhanced physicians' confidence in 97.37% of cases. Among the tumors in the first hepatic hilum, cholangiocarcinoma accounted for the majority (57.89%). Cholangiocarcinoma is a malignant tumor originating from the bile ducts and is the second most common primary malignant liver tumor after hepatocellular carcinoma, representing 10%-20% of all liver tumors (20). It has a poor prognosis, with an overall 5-year survival rate of 2%-30% (21, 22). Even with curative resection, the 5-year survival rate is less than 40% (23). Importantly, many cases of cholangiocarcinoma are diagnosed at an advanced stage, limiting the opportunity for surgical intervention. However, there are various treatment options available for advanced tumors, including biliary stent placement, percutaneous transhepatic cholangiodrainage (PTCD), chemotherapy, immunotherapy, targeted therapy, and more (24, 25). A definitive histopathological

diagnosis helps guide personalized treatment approaches (26, 27). In one case of our study, a patient was diagnosed with advanced-stage cholangiocarcinoma in the first hepatic hilum based on imaging findings, precluding surgical intervention. After PUS-CNB and subsequent genetic testing, the patient was found to have high PD-L1 expression (90%), and with the combination of immunosuppressive agents and radiotherapy, the patient achieved near-complete regression of the lesion during a 3-year follow-up. Immune checkpoint inhibitors have emerged as a novel first-line treatment option for advanced cholangiocarcinoma, with multiple studies demonstrating their efficacy in this patient population (28, 29). A study by Japanese scholars further supported the reliable and well-tolerated efficacy of combination therapy using immune checkpoint inhibitors and chemotherapy compared to monotherapy with chemotherapy or immunotherapy drugs (30). Therefore, for patients with advanced-stage tumors, obtaining histopathological tissue is a crucial step in exploring non-surgical treatment options.

For lymphoma, patients can often avoid surgery due to their sensitivity to chemotherapy. Effective tumor control and long-term survival can be achieved through standardized chemotherapy, and a specific subtype can only be determined with a definite pathology to establish appropriate clinical chemotherapy protocols. In this study, in the case of diffuse large B-cell lymphoma diagnosed after PUS-CNB, obtaining tissue specimens was crucial. For patients with metastatic cancer, radiofrequency ablation (RFA) serves as a palliative treatment modality. The percutaneous RFA procedure for lesions in the first hepatic hilum under ultrasound guidance is similar to PUS-CNB, making ultrasound-guided RFA feasible for lesions in the first hepatic hilum. In this study, there were 5 patients who were diagnosed with metastatic cancer after PUS-CNB confirmed the first



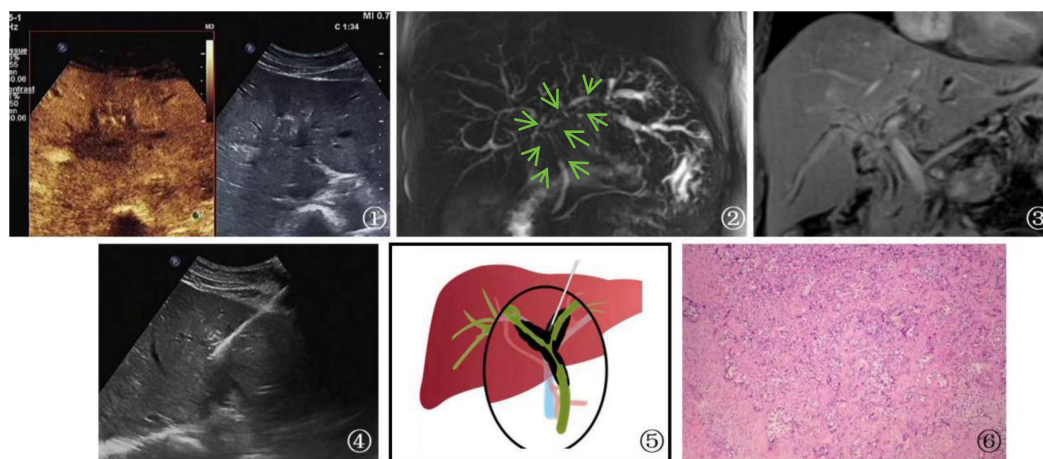


FIGURE 6

A 53-year-old female patient Presented with cholangiocarcinoma (a diffuse type). ①: According to an ultrasound examination, there is a hypoechoic lesion (arrow) with irregular margins located in the left hepatic duct at the first hepatic hilum. Additionally, a contrast-enhanced ultrasound shows that this lesion exhibits rapid wash-in and wash-out during the arterial phase. Ultrasound contrast revealed rapid wash-in and wash-out in the lesion at the first hepatic hilum. ②, ③: Cross-sectional images obtained from MRCP and MRI of the bile duct at the first hepatic hilum reveal irregular thickening of the bile duct and upper common bile duct that is surrounded by infiltrative soft tissue signals (arrow). These findings are consistent with a focal, irregular stricture in the bile duct lumen, as well as intrahepatic bile duct dilatation resembling a soft, vine-like structure with uniform and obvious enhancement. Furthermore, there is evidence of slightly longer T1 and T2 signals with uneven and obvious enhancement. Given the imaging findings, cholangiocarcinoma should be considered as a possible diagnosis. ④: Schematic diagram of percutaneous biopsy of the first hepatic hilum under ultrasound guidance. ⑤: Percutaneous biopsy of the first hepatic hilum was performed under ultrasound guidance. ⑥: High magnification microscope, HE staining, moderately differentiated cholangiocarcinoma (magnification,  $\times 40$ ).

hepatic hilum lesions. All of them underwent ultrasound-guided RFA for the first hepatic hilum lesions, and during a follow-up period of six months, the lesions were in an inactive state. Fusion imaging can also reduce false-positive lesion detection during US-guided RFA and consistently improve the detection of HCCs, especially when these are smaller than 2 cm. The ability of fusion imaging to reduce false positives also applies to the evaluation of local tumor progression after RFA and TACE (31, 32).

In this study, the integration of ultrasound with multimodal image cognition in percutaneous ultrasound-guided the first hepatic hilum puncture biopsy offers the following advantages: 1. High safety: Prior to the procedure, other imaging examinations are performed, and with the integration of ultrasound cognition, large blood vessels and bile ducts can be avoided. The appropriate puncture path can be selected, reducing the risk of major bleeding and bile leakage. This study found that punctures above the common hepatic duct in the first hepatic hilum can be performed through the intercostal approach, while lesions in the upper segment of the common bile duct and the surrounding hepatic hilum can be accessed through the subcostal approach, reducing the puncture distance. Real-time dynamic monitoring of the needle insertion process ensures a safe and controlled procedure. Additionally, multiplane imaging provides confidence for operators. 2. Simplicity, speed, and ease of use: This technique is easy to learn, and experienced ultrasound interventionists can quickly master it after short training. The puncture biopsy procedure is relatively short, with a minimum time of 9 minutes and an average time of  $14.55 \pm 2.73$  minutes. 3. High accuracy: The use of an 18-gauge thick needle allows for an adequate amount of pathological tissue to be

obtained, resulting in higher diagnostic accuracy compared to previous fine needle aspiration biopsy methods. 4. Ultrasound examination features no radiation, is flexible, and has low cost: Compared to fluoroscopy or CT-guided puncture biopsy, ultrasound's radiation-free imaging is an absolute advantage. Moreover, the overall procedure incurs minimal economic costs.

However, there are certain limitations to consider. Firstly, this procedure is not without risks. Despite the integration with other imaging modalities, potential complications such as vascular injury, bile leakage, and needle tract seeding still exist. It also relies on the experience and radiological expertise of the operator. Post-procedure observation for one hour and close collaboration with multiple clinical disciplines are necessary to promptly identify and manage any complications that may arise. Secondly, this study is retrospective and has a relatively small number of cases. Patients without a puncture path were not included in the study, which may have inflated the success rate of the procedure. Additionally, the study lacks a comparison with other imaging-guided techniques. Retrospective data collected from a single institution might introduce bias. Therefore, large-scale, high-quality prospective studies are needed to validate these findings and obtain more accurate conclusions.

## 5 Conclusion

The integration of ultrasound and multimodal imaging cognition in the percutaneous ultrasound-guided biopsy of the



first hepatic hilum lesions can reduce unnecessary surgical exploration or resection. It can confirm the resectability or treatability of the lesions and has the potential to become a new, superior, and faster method. This approach can optimize the triage and treatment strategies for indeterminate hepatic hilum lesions, benefiting the patients. It is particularly valuable for patients with unclear lesions in the first hepatic hilum who require a definitive diagnosis, as well as for those with bile duct carcinoma who have lost the opportunity for surgical intervention and can benefit from immunotherapy.

## Data availability statement

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

## Author contributions

X-TZ: Supervision, Writing – original draft, Writing – review & editing. XL: Data curation, Writing – original draft, Writing – review & editing. Z-LH: Conceptualization, Investigation, Writing – original draft, Writing – review & editing. SC: Visualization, Writing – original draft. J-CY: Investigation, Writing – original draft. Y-cL: Supervision, Writing – original draft. S-SW: Conceptualization, Writing – original draft, Writing – review & editing.

## References

- Gerhards MF, Vos P, Van Gulik TM, Rauws EA, Bosma A, Gouma DJ. Incidence of benign lesions in patients resected for suspicious hilar obstruction. *Br J Surg.* (2001) 88:48–51. doi: 10.1046/j.1365-2168.2001.01607.x
- Vauthey JN, Blumgart LH. Recent advances in the management of cholangiocarcinomas. *Sem Liver Dis.* (1994) 14:109–14. doi: 10.1055/s-2007-1007302
- Fritscher-Ravens A, Broering DC, Knoefel WT, Rogiers X, Swain P, Thonke F, et al. EUS-guided fine-needle aspiration of suspected hilar cholangiocarcinoma in potentially operable patients with negative brush cytology. *Am J Gastroenterol.* (2004) 99:45–51. doi: 10.1046/j.1572-0241.2003.04006.x
- Reddy SB, Patel T. Current approaches to the diagnosis and treatment of cholangiocarcinoma. *Curr Gastroenterol Rep.* (2006) 8:30–7. doi: 10.1007/s11894-006-0061-1
- Leblanc JK, Ciaccia D, Al-Assi MT, McGrath K, Imperiale T, Tao LC, et al. Optimal number of EUS-guided fine needle passes needed to obtain a correct diagnosis. *Gastrointest Endosc.* (2004) 59:475–81. doi: 10.1016/S0016-5107(03)02863-3
- De Bellis M, Sherman S, Fogel EL, Cramer H, Chappo J, McHenry L Jr, et al. Tissue sampling at ERCP in suspected Malignant biliary strictures (part 1). *Gastrointest Endosc.* (2002) 56:552–61. doi: 10.1067/mge.2002.128132
- Glasbrenner B, Ardan M, Boeck W, Preclik G, Moller P, Adler G. Prospective evaluation of brush cytology of biliary strictures during endoscopic retrograde cholangiopancreatography. *Endoscopy.* (1999) 31:712–7. doi: 10.1055/s-1999-73
- Ponchon T, Gagnon P, Berger F, Labadie M, Liaras A, Chavaillon A, et al. Value of endobiliary brush cytology and biopsies for the diagnosis of Malignant bile duct stenosis: results of a prospective study. *Gastrointest Endosc.* (1995) 42:565–72. doi: 10.1016/S0016-5107(95)70012-9
- Pugliese V, Conio M, Nicolo G, Saccomanno S, Gatteschi B. Endoscopic retrograde forceps biopsy and brush cytology of biliary strictures: a prospective study. *Gastrointest Endosc.* (1995) 42:520–6. doi: 10.1016/S0016-5107(95)70004-8
- Eloubeidi MA, Chen VK, Jhala NC, Eltoun IE, Jhala D, Chhieng DC, et al. Endoscopic ultrasound-guided fine needle aspiration biopsy of suspected

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. Fujian Medical innovation project (2022CXA006)

## Acknowledgments

The authors are thankful to Fujian Provincial Hospital and Fujian Medical University for their management of our patient database. The authors are thankful to S-SW for helping critically revise the manuscript for important intellectual content and helping collect data and design the study.

## 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.

- cholangiocarcinoma. *Clin Gastroenterol Hepatol.* (2004) 2:209–13. doi: 10.1016/S1542-3565(04)00005-9
- Byrne MF, Gerke H, Mitchell RM, Stiffler HL, McGrath K, Branch MS, et al. Yield of endoscopic ultrasound-guided fine needle aspiration of bile duct lesions. *Endoscopy.* (2004) 36:715–9. doi: 10.1055/s-2004-825657
- Jailwala J, Fogel EL, Sherman S, Gottlieb K, Flueckiger J, Bucksot LG, et al. Triple-tissue sampling at ERCP in Malignant biliary obstruction. *Gastrointest Endosc.* (2000) 51:383–90. doi: 10.1016/S0016-5107(00)70435-4
- Howell DA, Beveridge RP, Bosco J, Jones M. Endoscopic needle aspiration biopsy at ERCP in the diagnosis of biliary strictures. *Gastrointest Endosc.* (1992) 38:531–5. doi: 10.1016/S0016-5107(92)70510-0
- Nix GA, Van Overbeeke IC, Wilson JHP, Ten Kate FJW. ERCP diagnosis of tumors in the region of the head of the pancreas. *Dig Dis Sci.* (1988) 33:577–86. doi: 10.1007/BF01798360
- Silvis SE, Rohrmann CA, Vennes JA. Diagnostic accuracy of endoscopic retrograde cholangiopancreatography in hepatic, biliary, and pancreatic Malignancy. *Ann Intern Med.* (1976) 84:438–40. doi: 10.7326/0003-4819-84-4-438
- Huang PC, Liu CY, Chuang WY, Shih LY, Wan YL. Ultrasound-guided core needle biopsy of cervical lymphadenopathy in patients with lymphoma: The clinical efficacy and factors associated with unsuccessful diagnosis. *Ultrasound Med Biol.* (2010) 36:1431–6. doi: 10.1016/j.ultrasmedbio.2010.05.018
- Jue TL, Sharaf RN, Appalaneni V, Anderson MA, Ben-Menachem T, Decker GA, et al. Role of EUS for the evaluation of mediastinal adenopathy. *Gastrointest Endosc.* (2011) 74:239–45. doi: 10.1016/j.gie.2011.03.1255
- Aribas BK, Arda K, Ciledag N, Aktas E, Yakut F, Kavak S, et al. Accuracy and safety of percutaneous US-guided needle biopsies in specific focal liver lesions: comparison of large and small needles in 1300 patients. *Panminerva Med.* (2012) 54:233–9.
- Jones K, Biederman L, Draganova-Tacheva R, Solomides C, Bibbo M. Diagnostic yield of endoscopic ultrasound-guided fine-needle aspiration cytology of porta hepatis

lesions: A retrospective study. *Acta Cytologica*. (2016) 60:154–60. doi: 10.1159/000445764

20. Mantripragada S, Chawla A. Cholangiocarcinoma: part 1, pathological and morphological subtypes, spectrum of imaging appearances, prognostic factors and staging. *Curr Probl Diagn Radiol*. (2022) 51:351–61. doi: 10.1067/j.cpradiol.2021.03.008

21. Sharma P, Yadav S. Demographics, tumor characteristics, treatment, and survival of patients with Klatskin tumors. *Ann Gastroenterol*. (2018) 31:231–6. doi: 10.20524/aog.2018.0233

22. Shroff RT, Kennedy EB, Bachini M, Bekaii-Saab T, Crane C, Edeline J, et al. Adjuvant therapy for resected biliary tract cancer ASCO clinical practice guideline. *J Clin Oncol*. (2019) 37:1015–27. doi: 10.1200/JCO.18.02178

23. Rea DJ, Munoz Juares M, Farnell MB, Donohue JH, Que FG, Crownhart B, et al. Major hepatic resection for hilar cholangiocarcinoma: analysis of 46 patients. *Arch Surg*. (2004) 139:514–23. doi: 10.1001/archsurg.139.5.514

24. Chen VK, Eloubeidi MA. Endoscopic ultrasound-guided fine needle aspiration is superior to lymph node echotextures: a prospective evaluation of mediastinal and peritoneal lymphadenopathy [review]. *Am J Gastroenterol*. (2004) 99:628–33. doi: 10.1111/j.1572-0241.2004.04064.x

25. Tanaka N, Yamakado K, Nakatsuka A, Fujii A, Matsumura K, Takeda K. Arterial chemoinfusion therapy through an implanted port system for patients with unresectable intrahepatic cholangiocarcinoma: initial experience. *Eur J Radiol*. (2002) 41:42–8. doi: 10.1016/S0720-048X(01)00414-4

26. Neuhaus P, Jonas S, Bechstein WO, Lohmann R, Radke C, Kling N, et al. Extended resections for hilar cholangiocarcinoma. *Ann Surg*. (1999) 230:808–19. doi: 10.1097/0000658-199912000-00010

27. Kosuge T, Yamamoto J, Shimada K, Yamasaki S, Makuuchi M. Improved surgical results for hilar cholangiocarcinoma with procedures including major hepatic resection. *Ann Surg*. (1999) 230(5):663–71. doi: 10.1097/0000658-199911000-00008

28. Valle JW, Vogel A, Denlinger CS, He AR, Bai LY, Orlova R, et al. Addition of ramucirumab or merestinib to standard first-line chemotherapy for locally advanced or metastatic biliary tract cancer: a randomised, double-blind, multicentre, phase 2 study. *Lancet Oncol*. (2021) 22:1468–82. doi: 10.1016/S1470-2045(21)00409-5

29. Piha-Paul SA, Oh DY, Ueno M, Malka D, Chung HC, Nagrial A, et al. Efficacy and safety of pembrolizumab for the treatment of advanced biliary cancer: Results from the KEYNOTE-158 and KEYNOTE-028 studies. *Int J Cancer*. (2020) 147:2190–8. doi: 10.1002/ijc.33013

30. Ueno M, Ikeda M, Morizane C, Kobayashi S, Ohno I, Kondo S, et al. Nivolumab alone or in combination with cisplatin plus gemcitabine in Japanese patients with unresectable or recurrent biliary tract cancer: a non-randomised multicentre, open-label, phase 1 study. *Lancet Gastroenterol Hepatol*. (2019) 4:611–21. doi: 10.1016/S2468-1253(19)30086-X

31. Lee MW, Rhim H, Cha DI, Kim YJ, Lim HK. Planning US for percutaneous radiofrequency ablation of small hepatocellular carcinomas (1–3 cm): value of fusion imaging with conventional US and CT/MR images. *J Vasc Interv Radiol*. (2013) 24:958–65. doi: 10.1016/j.jvir.2013.04.007

32. Min JH, Lee MW, Rhim H, Cha DI, Lim S, Choi S-Y, et al. Local tumor progression after locoregional therapy of hepatocellular carcinomas: value of fusion imaging-guided radiofrequency ablation. *Clin Radiol*. (2014) 69:286–93. doi: 10.1016/j.crad.2013.10.015



## OPEN ACCESS

## EDITED BY

Zhimin Li,  
Northwestern University, United States

## REVIEWED BY

Xuxin Chen,  
University of Oklahoma, United States  
Zhangyang Ai,  
Hunan University, China  
Xiuxu Chen,  
Loyola University Medical Center,  
United States

## \*CORRESPONDENCE

Kun Miao  
✉ mkqp@163.com

<sup>†</sup>These authors have contributed equally to this work and share first authorship

<sup>†</sup>These authors have contributed equally to this work

RECEIVED 10 August 2023

ACCEPTED 08 April 2024

PUBLISHED 30 April 2024

## CITATION

Wang Q, Bi Q, Qu L, Deng Y, Wang X, Zheng Y, Li C, Meng Q and Miao K (2024) MAMILNet: advancing precision oncology with multi-scale attentional multi-instance learning for whole slide image analysis. *Front. Oncol.* 14:1275769. doi: 10.3389/fonc.2024.1275769

## COPYRIGHT

© 2024 Wang, Bi, Qu, Deng, Wang, Zheng, Li, Meng and Miao. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

# MAMILNet: advancing precision oncology with multi-scale attentional multi-instance learning for whole slide image analysis

Qinqing Wang<sup>1†</sup>, Qiu Bi<sup>2†</sup>, Linhao Qu<sup>3†</sup>, Yuchen Deng<sup>4</sup>,  
Xianhong Wang<sup>4</sup>, Yijun Zheng<sup>4</sup>, Chenrong Li<sup>4</sup>,  
Qingyin Meng<sup>5\*</sup> and Kun Miao<sup>6\*†</sup>

<sup>1</sup>Department of Pathology, The First People's Hospital of Yunnan Province, The Affiliated Hospital of Kunming University of Science and Technology, Kunming, Yunnan, China, <sup>2</sup>Department of MRI, The First People's Hospital of Yunnan Province, The Affiliated Hospital of Kunming University of Science and Technology, Kunming, Yunnan, China, <sup>3</sup>Fudan University, Shanghai, China, <sup>4</sup>School of Clinical Medicine, The Affiliated Hospital of Kunming University of Science and Technology, Kunming, Yunnan, China, <sup>5</sup>Department of Pathology, The Third Affiliated Hospital of Kunming Medical University (Tumor Hospital of Yunnan Province), Kunming, Yunnan, China, <sup>6</sup>Department of Medical Oncology, The First People's Hospital of Yunnan Province, The Affiliated Hospital of Kunming University of Science and Technology, Kunming, Yunnan, China

**Background:** Whole Slide Image (WSI) analysis, driven by deep learning algorithms, has the potential to revolutionize tumor detection, classification, and treatment response prediction. However, challenges persist, such as limited model generalizability across various cancer types, the labor-intensive nature of patch-level annotation, and the necessity of integrating multi-magnification information to attain a comprehensive understanding of pathological patterns.

**Methods:** In response to these challenges, we introduce MAMILNet, an innovative multi-scale attentional multi-instance learning framework for WSI analysis. The incorporation of attention mechanisms into MAMILNet contributes to its exceptional generalizability across diverse cancer types and prediction tasks. This model considers whole slides as “bags” and individual patches as “instances.” By adopting this approach, MAMILNet effectively eliminates the requirement for intricate patch-level labeling, significantly reducing the manual workload for pathologists. To enhance prediction accuracy, the model employs a multi-scale “consultation” strategy, facilitating the aggregation of test outcomes from various magnifications.

**Results:** Our assessment of MAMILNet encompasses 1171 cases encompassing a wide range of cancer types, showcasing its effectiveness in predicting complex tasks. Remarkably, MAMILNet achieved impressive results in distinct domains: for breast cancer tumor detection, the Area Under the Curve (AUC) was 0.8872, with an Accuracy of 0.8760. In the realm of lung cancer typing diagnosis, it achieved an AUC of 0.9551 and an Accuracy of 0.9095. Furthermore, in predicting drug therapy responses for ovarian cancer, MAMILNet achieved an AUC of 0.7358 and an Accuracy of 0.7341.

**Conclusion:** The outcomes of this study underscore the potential of MAMILNet in driving the advancement of precision medicine and individualized treatment planning within the field of oncology. By effectively addressing challenges related to model generalization, annotation workload, and multi-magnification integration, MAMILNet shows promise in enhancing healthcare outcomes for cancer patients. The framework's success in accurately detecting breast tumors, diagnosing lung cancer types, and predicting ovarian cancer therapy responses highlights its significant contribution to the field and paves the way for improved patient care.

#### KEYWORDS

whole slide image analysis, multiple instance learning, cancer diagnosis, multi-scale attention, deep learning

## 1 Introduction

In recent years, computational pathology has emerged as a transformative discipline with immense potential to revolutionize cancer diagnosis and treatment planning. The advent of digital pathology and whole slide imaging has led to vast histopathological data repositories, presenting an unprecedented opportunity for deep learning networks in this field Srinidhi et al. (1) Qu et al. (2). Whole Slide Image (WSI) analysis, enabled by deep learning algorithms, shows promise in tumor detection, typing, and drug treatment response prediction, heralding a new era of precision medicine in oncology Cheplygina et al. (3) Rony et al. (4).

Tumor detection is critical for timely and accurate cancer diagnoses. Conventional methods, relying on manual examination by pathologists, can be time-consuming and subjective, leading to diagnostic errors and variability. Deep learning networks revolutionize tumor detection, using Convolutional Neural Networks (CNNs) to meticulously analyze digitized histopathological images, identifying malignancy with precision and efficiency. Integrating deep learning expedites diagnostic processes, enhances accuracy, and ensures reproducibility in clinical settings. Tumor typing, categorizing cancers into subtypes, is vital for personalized oncology. Deep learning networks address challenges in tumor typing, comprehensively learning from annotated histopathological datasets. They discern subtle differences between tumor subtypes with remarkable accuracy, facilitating efficient and precise tumor typing, leading to tailored therapies and improved patient outcomes. In the realm of cancer treatment, establishing deep-learning models to predict drug therapy response from WSIs has emerged as a transformative frontier. Traditional methods relying on manual evaluation of biopsy samples struggle to capture the true heterogeneity of tumor responses. In contrast, deep learning networks in WSI analysis offer a promising and powerful approach. By analyzing large-scale histopathological datasets, these models can detect subtle alterations induced by therapies, enabling accurate and timely

prediction of treatment response. This groundbreaking development holds the potential to revolutionize cancer treatment and elevate patient outcomes to new heights.

Deep learning networks have a significant impact on computational pathology, particularly in WSI analysis for tumor detection, typing, and treatment response prediction, advancing precision medicine and patient care. However, integrating deep learning algorithms faces significant challenges in this domain. A primary issue is the limited generalizability of models across different cancer types and tasks, compromising their practicality for routine clinical use. Many current approaches achieve remarkable results on specific cancer types or tasks, but their performance tends to deteriorate when confronted with diverse cancers. The burden of patch-level annotation is another major challenge. WSIs are massive and need to be divided into smaller patches for deep learning training. Manual annotation of these patches is time-consuming and labor-intensive, making comprehensive annotation impractical, hindering the adoption of fully-supervised algorithms. Moreover, handling multi-magnification information is vital. Current studies often focus on single magnification analysis, neglecting the diagnostic information embedded in multiple magnifications. This limitation may lead to incomplete understanding of pathological patterns, reducing the efficacy of deep learning algorithms in capturing the full complexity of the images Srinidhi et al. (1) Qu et al. (2) Cheplygina et al. (3) Rony et al. (4) Wang et al. (5) Qu et al. (6).

This paper presents MAMILNet, a novel multi-scale attentional multi-instance learning framework for whole slide pathological image processing. MAMILNet offers several key advantages over existing methods. Firstly, it exhibits high generalization across multiple cancer types and prediction tasks by skillfully integrating the attention mechanism. This adaptability ensures robust performance in diverse scenarios. Secondly, MAMILNet employs a multi-instance learning (MIL) architecture, treating slides as “bags” and their cut patches as “instances,” effectively representing slides as a whole. This eliminates the need for fine-

grained patch-level labeling, reducing the burden of manual labeling for pathologists. Additionally, MAMILNet utilizes a multi-scale “consultation” training and prediction strategy, training with multiple magnifications and aggregating test results from different scales using a probability ensemble method. This approach significantly enhances prediction accuracy by considering multiple magnifications during diagnosis and prediction. Overall, MAMILNet is a promising framework for achieving high-accuracy predictions in WSI analysis with weak labeling at the slide level.

We present a comprehensive evaluation of MAMILNet’s performance. We focus on three critical diagnosis tasks, involving different cancer types, from four distinct medical centers. These tasks include automatic recognition of sentinel lymph node cancer in breast cancer, automatic typing of lung adenocarcinoma and lung squamous cell carcinoma, and drug resistance diagnosis of high-grade serous ovarian cancer. Through a meticulous analysis of 1711 patients and WSIs, MAMILNet demonstrates remarkable accuracy in predicting these complex tasks. The successful outcomes achieved by MAMILNet in this diverse dataset hold significant implications for cancer diagnosis and personalized treatment planning. These findings further reinforce the potential of deep learning networks in advancing WSI processing, paving the way for improved healthcare outcomes in oncology.

## 2 Related work

### 2.1 Deep-learning-based WSI analysis

Numerous noteworthy studies have been dedicated to addressing significant clinical challenges in the WSI analysis field. For instance, Coudray et al. (7) developed deep-learning models capable of accurately predicting cancer subtypes and genetic mutations, sparking the entire field. Naik et al. (8) presented a deep-learning framework for directly predicting estrogen receptor status from H&E slides. Another notable clinical endeavor was undertaken by Tomita et al. (9), who proposed a grid-based network for performing 4-class classification of high-resolution endoscopic esophagus and gastroesophageal junction mucosal biopsy images from 379 patients. Skrede et al. (10) developed a deep model to analyze conventional H&E-stained slides and effectively predict the prognosis of patients after colorectal cancer surgery. Similarly, in a gastrointestinal tract oncology study, Kather et al. (11) employed a deep model to predict microsatellite instability (MSI) directly from H&E-stained slides. Currently, deep-learning models for WSI analysis have been applied across a wide range of cancer types, including breast, colorectal, lung, liver, cervical, thyroid, and bladder cancers Coudray et al. (7) Bejnordi et al. (12) Chaudhary et al. (13) Campanella et al. (14) Saillard et al. (15) Woerl et al. (16) Anand et al. (17) Velmahos et al. (18) Wessels et al. (19) Li et al. (20) Yang et al. (21).

In contrast to the majority of studies that have focused on specific tasks for individual cancers, our proposed MAMILNet takes a broader approach, exploring network architectures for multiple tasks across multiple cancer species. With MAMILNet, we have successfully achieved high accuracy in predicting multiple tasks for various cancer types.

### 2.2 Multi-instance learning techniques

As an effective weakly supervised learning algorithm, multi-instance learning has emerged as the mainstream method for WSI analysis based on deep learning Campanella et al. (14) Ilse et al. (22) Shi et al. (23) Li et al. (24) Qu et al. (25) Qu et al. (26). Due to the substantial size of WSIs, often reaching  $100,000 \times 100,000$  pixels, direct utilization as input for deep-learning models is impractical. To alleviate the computational burden, WSIs are typically divided into numerous small patches for processing. In multi-instance learning, each WSI is treated as a “bag,” while the segmented patches are regarded as “instances” belonging to that bag. If a bag is labeled as negative, all instances within it are considered negative; conversely, if a bag is labeled as positive, at least one instance within it is positive. Multi-instance learning leverages neural networks to extract features from each instance and aggregates them into a feature representation of the bag. Subsequently, the classifier is trained at the bag level, enabling direct slide-level classification without the need for doctors to label patches with fine granularity.

However, current studies primarily focus on MIL-based WSI analysis at a single magnification level, while pathologists often switch between multiple magnifications to perform comprehensive diagnoses. Neglecting the multiplex information may lead to an incomplete understanding and interpretation of pathological patterns, thereby limiting the effectiveness of deep learning algorithms in capturing the full complexity of these images. Embracing the varied information present in different magnifications is essential to enhance the diagnostic accuracy and enable deep-learning models to encompass the richness of information contained within WSIs.

## 3 Materials and methods

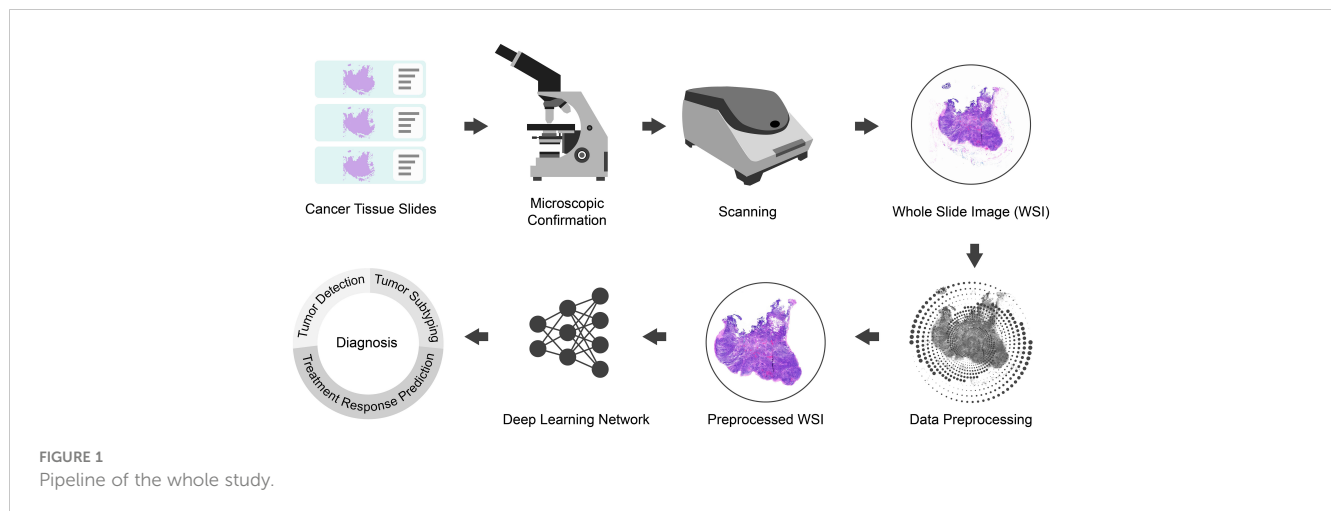
### 3.1 Study design and workflow

The present study focuses on advancing WSI processing through the integration of deep learning techniques. As illustrated in Figure 1, our methodology commences with the expertise of skilled pathologists, who meticulously prepare film and microscope reprints of tumor tissue sections. Subsequently, high-quality sections with clear labels are carefully chosen for digital scanning, yielding comprehensive WSI datasets. The utilization of WSIs is essential as it allows for a holistic view of the tissue, enabling a more comprehensive and accurate analysis.

To optimize the input data for deep learning analysis, the acquired WSIs undergo preprocessing (refer to Section 3.2). This step involves WSI patching, data normalization, and data augmentation, among other techniques, ensuring standardized and consistent data for subsequent analysis.

The main focus of our study involves the development and implementation of a sophisticated deep learning network, referred to as MAMILNet (detailed in section 3.2.3). By incorporating attention mechanisms, multi-instance learning, and multi-scale ensemble strategies, MAMILNet is strategically designed to effectively address the complexities associated with multiple





prediction tasks across various cancer types. To comprehensively evaluate MAMILNet's performance on these diverse prediction tasks, we utilized three data cohorts from four different centers, encompassing 1711 cases and WSIs. The predictive tasks for different cancer types include: automatic recognition of sentinel lymph node cancer in breast cancer, automatic typing of lung adenocarcinoma and lung squamous cell carcinoma, and evaluation of drug resistance in high-grade serous ovarian cancer.

## 3.2 Data collection and preprocessing

In this research, we conducted a comprehensive evaluation of MAMILNet's performance on various prediction tasks for different cancer types using a total of 1711 cases and WSIs from three data cohorts across four centers. The predicted cancer types and tasks included the automatic recognition of sentinel lymph node cancer in breast cancer, automatic typing of lung adenocarcinoma and lung squamous cell carcinoma, and drug-resistance diagnosis of high-grade serous ovarian cancer. For details on the collection of relevant data queues, please refer to section 3.2.1, and for information on the pre-processing of data queues, see section 3.2.2.

### 3.2.1 Data collection

#### 3.2.1.1 Breast cancer cohort

For the breast cancer cohort, we utilized the Camelyon 16 Dataset Bejnordi et al. (12), which is a prominent public benchmark in computational pathology, specifically focusing on sentinel lymph nodes. This dataset consists of a total of 399 whole-slide images (WSIs) collected from the Radboud University Medical Center in Nijmegen and the University Medical Center Utrecht in the Netherlands. Expert pathologists have annotated the tissue slides, labeling slides containing cancer as positive and those without cancer as negative. The raw data queue contains both slide-level weak labels and pixel-level labels for cancer regions. However, to adhere to the requirements of weakly supervised scenarios, we utilized only slide-level weak labels for training and testing purposes. This approach ensures the appropriate use of data

while addressing the challenges posed by weak supervision in the context of this research.

#### 3.2.1.2 Lung cancer cohort

The TCGA lung cancer dataset comprises a total of 1050 H&E stained WSIs from the public Cancer Genome Atlas (TCGA) data portal. This dataset includes two subtypes of lung cancer: Lung Adenocarcinoma and Lung Squamous Cell Carcinoma. Expert pathologists have carefully annotated the WSIs, providing slide-level labels to indicate whether each WSI corresponds to Lung Adenocarcinoma (negative) or Lung Squamous Cell Carcinoma (positive). The meticulous annotations by expert pathologists ensure the accuracy and reliability of the dataset for further analysis and research in the field of lung cancer.

#### 3.2.1.3 Ovarian cancer cohort

The Ovarian Cancer Cohort comprises a total of 262 cases and WSIs from two centers: 228 patients from Yunnan Cancer Hospital, China, and 34 patients from Yunnan First People's Hospital, China. After obtaining institutional review board approval, we retrospectively selected patients who received standardized treatment for ovarian cancer at Yunnan Cancer Hospital and Yunnan First People's Hospital between 2015 and 2022.

Inclusion criteria for patient selection were as follows: (1) confirmation of high-grade serous ovarian cancer through operation and pathology; (2) treatment modalities including primary tumor cell reduction plus first-line platinum drug chemotherapy, or neoadjuvant chemotherapy plus tumor cell reduction plus first-line platinum drug chemotherapy; (3) availability of at least one pathological H&E-stained slide with focal lesions for each patient; (4) at least 6 months of available follow-up records after chemotherapy. Exclusion criteria included: (1) history of other malignant tumors, pelvic surgery, or platinum chemotherapy; (2) poor quality of tissue slides (Cases with poor slide quality, such as broken cap fragments or stains on the surface, insufficient tumor tissue, or tissue folding, were excluded.); (3) maximum diameter of the lesion less than 1 cm; (4) incomplete or substandard chemotherapy regimen; (5) incomplete clinical and

pathological data. To clarify, at Yunnan Cancer Hospital, out of 270 patients considered, 42 were excluded based on the predefined criteria, resulting in 228 patients being included in the study. Similarly, at Yunnan First People's Hospital, from an initial pool of 48 patients, 14 did not meet the inclusion criteria, leaving 34 patients to be enrolled in the study.

We defined platinum resistance as disease progression or recurrence within 6 months after the end of chemotherapy, and platinum sensitivity if there was no disease progression or recurrence within this timeframe. Tumor recurrence was determined based on histopathology or the presence of two of the following manifestations: sustained elevation of CA125, pleural effusion or ascites, physical examination finding a lump, imaging findings of a mass, or unexplained intestinal obstruction. For each patient, two professional pathologists Qingqing Wang and Qingyin Meng evaluated the slides, selecting 1-3 representative primary lesion slides. The images were then digitized through an off-field 20-magnification scan ( $0.48 \mu\text{m}/\text{pixel}$ ) using a portable scanner (Ocus, Grundium, Finland).

### 3.2.2 Data preprocessing and partitioning

In this research, we employed the Python language (Version 3.7) and utilized the Openslide library (Version 3.4.1) to export all data queues at three magnifications: 20x, 10x, and 5x. We saved the resulting image sets separately, dividing them into non-overlapping  $224 \times 224$  small image blocks. To ensure data quality, image blocks with an entropy of less than 5 were excluded, as they are likely to represent the background. For each resolution set, we performed image normalization using the mean and variance of all slices within the corresponding set. Data augmentation techniques, including random flipping, rotation, color transformation, and random cropping, were applied to enhance the dataset's diversity. To achieve this, we utilized Python (Version 3.7) with libraries such

as Pillow (Version 8.4.0), OpenCV (Version 4.1.0), and the PyTorch deep learning framework (Version 1.7.1). By employing these procedures and tools, we ensured that our dataset was prepared with standardized resolution and enhanced with data augmentation, setting a solid foundation for robust and reliable deep-learning model training and evaluation.

In the Breast cancer cohort, we conducted a random division to create a training set consisting of 240 cases and slides, and a test set containing 129 cases and slides. For the Lung Cancer Cohort, we applied a random division resulting in a training set comprising 840 cases and slides, and a test set comprising 210 cases and slides. Similarly, for the Ovarian Cancer Cohort, we randomly divided it into a training set with 183 cases and slides, and a test set with 79 cases and slides. Importantly, each of the training sets also includes validation sets.

### 3.2.3 Multi-scale attentional multi-instance learning network

We present MAMILNet, a multi-instance deep convolutional neural network architecture incorporating a multi-scale attentional mechanism, designed to handle multiple prediction tasks for various cancers. The network's training process is illustrated in Figure 2A. During training, we create separate models for the 20x, 10x, and 5x image sets, as depicted in Figure 2A. Each set of patches from the same WSI constitutes a bag. Before each iteration, we apply random data augmentation techniques to each patch in the bag, including random noise, rotation, clipping, and color transformation. Next, we utilize a pre-trained ResNet He et al. (27) network as the primary feature extractor to obtain the features of each patch within the bag. Subsequently, an attention module is employed, where a learnable attention weight is assigned to the features of each image block. This attention-pooling process aggregates the features within the bag to obtain the bag-level

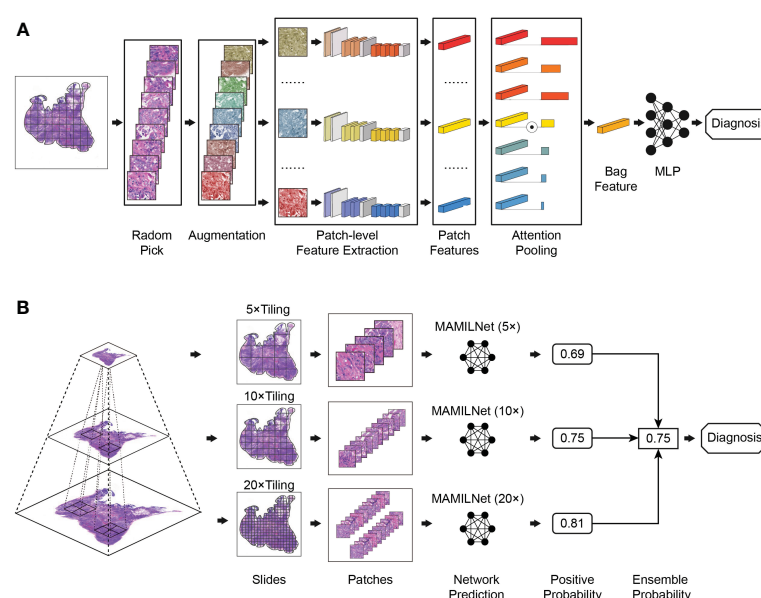


FIGURE 2  
(A) Training process of the MAMILNet; (B) Inference process of the MAMILNet.

features. Finally, a bag-level Multilayer Perceptron (MLP) serves as the bag classifier, directly predicting the negative and positive risks of the WSI. The cross-entropy loss, calculated against the true labels, serves as the loss function during training, and stochastic gradient descent drives the parameter updates in the network.

The attention module consists of two main steps. First, a linear fully connected layer reduces the dimension of each feature vector to 128, followed by the application of the pixel-level hyperbolic tangent function ( $\tanh()$ ). This non-linear output is then scaled to be between -1 and 1. The resulting values are multiplied with another linear layer to calculate the attention weight of each image block using the softmax function. In the second step, we use the feature matrix of the image block and the attention weight matrix to obtain the bag-level features. This step involves element-wise multiplication of the feature matrix with the attention weight matrix, effectively emphasizing the important regions within the bag based on their calculated attention weights. The resulting features represent a compact representation of the bag, capturing the salient information required for accurate bag-level predictions.

The inference process of the network is depicted in Figure 2B. During the testing phase, we propose a multi-scale integration strategy inspired by pathologists who often zoom in and out of slices for diagnosis. This strategy allows us to make the final prediction for the test cohort. Specifically, we employ the 20x, 10x, and 5x models obtained during the training process to calculate the predicted risk probability for each WSI at each magnification level. Next, we use the mean-pooling method to aggregate the prediction probabilities from the three magnification levels. This aggregation process yields the final prediction risk probability for each WSI.

We utilized the PyTorch deep learning framework (Version 1.7.1) in Python (Version 3.7) to perform all training and testing processes of the network. The Adam optimizer was employed to train the model, utilizing the cross-entropy loss as the loss function. The total number of training iterations was set to 500, with a learning rate of  $1e-4$ . We applied a weight decay factor of  $1e-5$ , and the batch size was set to 64. For computational resources, we conducted the training using the 11th Gen Intel(R) Core(TM) i7-11700K CPU in combination with the Nvidia 3090 GPU. These hardware configurations ensure efficient and high-performance processing during the training and testing phases of our deep-learning model.

### 3.2.4 Statistical analysis

In our specific experiments, we employed the cross-validation method to train the model and select the best-performing model for final internal verification and independent testing. During this process, the verification set was utilized to identify the model with the optimal performance, while the independent test set remained unseen during both the training and verification stages, ensuring a fair evaluation of the model's performance. Specifically, for each dataset, we first divide it into a training set (including a validation set) and an independent test set at a ratio close to 4:1, where the independent test set remains unseen during the training and validation process. The details of the division can be found in Section 3.2.2 Data Preprocessing and Partitioning. For the training

set, we employed a 5-fold cross-validation method. This technique divides the training dataset into five parts, using four parts for training and one part for validation in each iteration. This process ensures that each data point is used for both training and validation once, thus obtaining a more reliable estimate of model performance. Then we select the best model and parameters from the cross-validation to test on the independent test set and report the results of the independent test set as the final outcome. This also better simulates the prediction scenario for more new unseen clinical data in the future.

To assess the model's performance, we employed several metrics, including the area under the ROC curve (AUC), Accuracy, False Positive Rate (FPR), and False Negative Rate (FNR). These metrics were reported along with 95% confidence intervals (CI) to provide a comprehensive understanding of the model's effectiveness. All metric calculations and statistical analyses were conducted using the scikit-learn package (Version 1.3.0) within the Python (Version 3.7) environment. The scikit-learn package offers robust and reliable tools for evaluating machine learning models, ensuring the accuracy and consistency of our model assessments.

## 4 Results

### 4.1 Prediction results of sentinel lymph node tumor detection in breast cancer

In the task of tumor diagnosis of sentinel lymph nodes of breast cancer, as shown in Table 1, our innovative MAMILNet demonstrated remarkable success, achieving an impressive AUC of 0.8872 (95%CI 0.86-0.90) on the independent test set. Moreover, our model exhibited high accuracy (0.8760, 95%CI 0.85-0.89) and demonstrated low false positive rate (FPR=0.1406, 95%CI 0.16-0.12) and false negative rate (FNR=0.1077, 95%CI 0.08-0.12) performances. These compelling results underscore the efficacy of MAMILNet in accurately diagnosing tumors based on H&E-stained WSIs, and hold significant promise for enhancing breast cancer diagnostics and patient outcomes.

Moreover, by comparing our multi-scale model with MAMILNet variants and advanced competitors, we observed

TABLE 1 Prediction results on the independent test set of sentinel lymph node tumor detection in breast cancer.

| Deep-learning Model           | AUC           | Accuracy      | FNR           | FPR           |
|-------------------------------|---------------|---------------|---------------|---------------|
| 5x only MAMILNet              | 0.7684        | 0.7520        | 0.2258        | 0.2698        |
| 10x only MAMILNet             | 0.8379        | 0.8217        | 0.1384        | 0.2187        |
| 20x only MAMILNet             | 0.8653        | 0.8450        | 0.1538        | 0.1562        |
| MILRNN Campanella et al. (14) | 0.8178        | 0.8062        | 0.1428        | 0.2542        |
| CLAM Lu et al. (28)           | 0.8762        | 0.8527        | 0.1142        | 0.2000        |
| <b>MAMILNet (ours)</b>        | <b>0.8872</b> | <b>0.8760</b> | <b>0.1077</b> | <b>0.1406</b> |

Bold values refer to the best results.

further improvements through our proposed multi-scale integrated prediction strategy, validating its effectiveness in enhancing model performance. These findings signify significant strides in the field of deep learning-based tumor diagnosis, propelling advancements in early detection and precision medicine for breast cancer patients.

The ROC curve of MAMILNet on the breast cancer sentinel lymph node tumor detection task on the independent test set is shown in [Figure 3A](#).

## 4.2 Prediction results of automatic subtyping of lung adenocarcinoma and lung squamous cell carcinoma

In the domain of WSI processing for lung cancer typing diagnosis, as can be seen from [Table 2](#), our MAMILNet demonstrated outstanding performance. The achieved AUC of 0.9551 (95%CI 0.94-0.96) on the independent test set, coupled with accuracy of 0.9095 (95% CI 0.89-0.91), FPR of 0.0961 (95%CI 0.110.07), and FNR of 0.0857 (95%CI 0.09-0.07), affirms the model's accurate and efficient classification of lung tumor types based on H&E-stained WSIs. Additionally, a comparative analysis with other single-scale variants and competitors underscores the superiority of our innovative multi-scale strategy, further validating its efficacy in enhancing classification accuracy and diagnostic performance. These findings represent a significant advancement in the field of deep learning-based lung cancer typing, offering promising avenues for improving patient care and treatment outcomes.

The ROC curve on lung cancer tumor typing task on the independent test set is shown in [Figure 3B](#).

## 4.3 Prediction results of drug resistance in high-grade serous ovarian cancer

We explored a novel and challenging clinical task—predicting patients' drug therapy response using the Ovarian Cancer Cohort. This task, which surpasses the interpretability of H&E-stained slides even for medical experts, represents a pressing problem in contemporary medical research. As can be seen from [Table 3](#), our

MAMILNet demonstrated promising results in this endeavor, achieving an AUC of 0.7358 (95%CI 0.74-0.72) on the independent test set, along with accuracy of 0.7341 (95%CI 0.72-0.74), FPR of 0.2982 (95%CI 0.30-0.28), and FNR of 0.1818 (95%CI 0.17-0.19). These performance indicators signify the potential of our MAMILNet to make significant advancements in drug response therapy prediction, ushering in a new era of personalized medicine and improved patient outcomes in ovarian cancer management.

The ROC curve on the ovarian cancer treatment resistance prediction task on the independent test set is shown in [Figure 3C](#).

## 5 Discussion and conclusions

This research introduces MAMILNet, a novel multi-scale attentional multi-instance learning framework, which achieves remarkable performance in critical tasks like breast cancer tumor detection, lung cancer subtype diagnosis, and ovarian cancer drug resistance prediction, even with weak slide-level labeling.

MAMILNet's innovative design and capabilities hold significant promise for advancing medical image analysis, improving diagnostic accuracy, and guiding cancer treatment decisions.

MAMILNet demonstrates its robust performance through three key components. Firstly, it effectively employs attention mechanisms to adaptively learn critical visual features associated with various cancer species and tasks. By assigning higher attention weights to clinically relevant visual features and lower weights to unrelated background and noise, MAMILNet acts as a dynamic filter, enhancing its learning ability for different tasks. Consequently, the network exhibits exceptional generalization across multiple cancer types and prediction tasks. Secondly, MAMILNet adopts a feature aggregation-based multi-instance learning architecture, enabling remarkable performance even with slide-level weak annotation. It treats slides as "bags" and their cut patches as "instances," employing an efficient instance-level feature extractor to derive informative representations. An attention mechanism is then used to aggregate these instance features effectively into bag-level features. This approach culminates in a powerful bag classifier, enabling accurate slide-level classification. The combination of instance-level efficient feature representation, attention-based feature aggregation, and a robust bag-level classifier

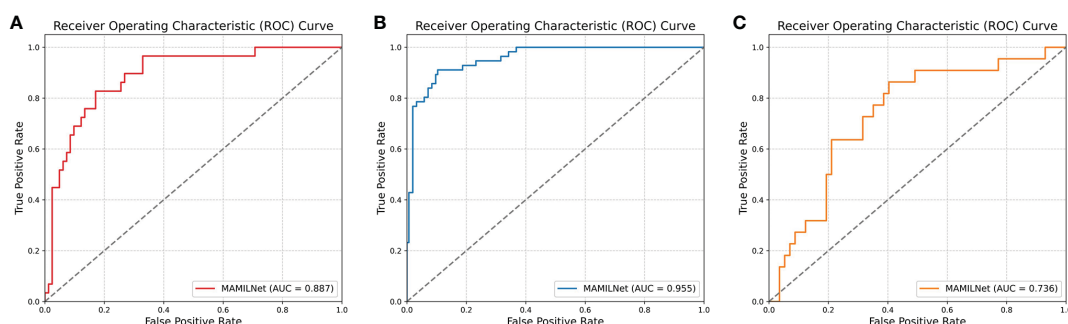


FIGURE 3

(A) The ROC curve of MAMILNet on the breast cancer sentinel lymph node tumor detection task (independent test set). (B) The ROC curve on lung cancer tumor typing task (independent test set). (C) The ROC curve on the ovarian cancer treatment resistance prediction task (independent test set).



TABLE 2 Prediction results on the independent test set of automatic subtyping of lung adenocarcinoma and lung squamous cell carcinoma.

| Deep-learning Model           | AUC           | Accuracy      | FNR           | FPR           |
|-------------------------------|---------------|---------------|---------------|---------------|
| 5× only MAMILNet              | 0.9269        | 0.8619        | 0.1333        | 0.1429        |
| 10× only MAMILNet             | 0.9324        | 0.8904        | 0.0952        | 0.1250        |
| 20× only MAMILNet             | 0.9488        | 0.9000        | 0.0857        | 0.1153        |
| MILRNN Campanella et al. (14) | 0.9236        | 0.8667        | 0.1111        | 0.1733        |
| CLAM Lu et al. (28)           | 0.9411        | 0.8857        | 0.0714        | 0.2000        |
| <b>MAMILNet (ours)</b>        | <b>0.9551</b> | <b>0.9095</b> | <b>0.0857</b> | <b>0.0961</b> |

Bold values refer to the best results.

ensures MAMILNet’s efficiency. Lastly, inspired by pathologists’ “zoom in and out” reading approach, MAMILNet introduces a multi-scale “consultation” training and prediction strategy. During training, it uses multiple magnifications to fully model and learn pathological features. During testing, a probabilistic set approach aggregates results from different scales, harnessing the advantages of diverse magnifications for optimal prediction performance, akin to a medical consultation process. This innovative strategy further enhances MAMILNet’s predictive capabilities.

Tumor detection and pathologic subtyping are vital in WSI analysis through deep learning. The deep learning model offers faster and more detailed diagnostic references directly from H&E-stained slides, significantly reducing examination time. Moreover, its highly generalized nature facilitates diagnosis in regions with limited medical resources. This study demonstrates satisfactory performance in both tasks, paving the way for the widespread application of deep-learning models in this domain. It is also among the pioneering efforts to explore the direct prediction of drug response therapeutic efficacy from WSI using deep learning. Assessing a patient’s response to drug therapy is critical for treatment decisions and prognosis. However, determining drug resistance from H&E-stained slides is challenging, even for experienced physicians. Studies Vamathevan et al. (29) Ballester et al. (30) Farahmand et al. (31) suggest that a patient’s sensitivity to specific drug therapy may be evident in H&E-stained sections, presenting opportunities for deep-learning model applications. The deep-learning model effectively captures

pathological patterns related to drug treatment responses in a data-driven manner, enabling accurate predictions. The research conducted a preliminary trial on High-grade Serous Ovarian Cancer, yielding promising results. These findings support the future prediction of treatment responses using deep learning across various cancer types with different drugs. This breakthrough holds significant potential for advancing personalized medicine and enhancing patient outcomes.

The study has several limitations that require careful consideration. Firstly, it adopts a retrospective analysis approach, which may inherently constrain the diversity and representativeness of the data. Future investigations aim to build a larger and more comprehensive dataset to enhance the model’s robustness. Secondly, for drug resistance prediction tasks, systematic pathologic patterns have not been identified. The deep-learning model relies on data-driven iteration and training, making it challenging to understand the underlying pathological basis of its judgments. Although the designed attention mechanism can highlight patches of high and low risk, further experiments are needed to systematically generalize authoritative pathological patterns. While this study provides a preliminary glimpse into the potential of utilizing deep-learning models for drug therapy response prediction in ovarian cancer, full generalization of this knowledge requires more extensive trials and investigation. Thirdly, despite conducting trials involving three cancer species from four centers, the validation across multiple centers remains insufficient for each task. Limited multi-center training and verification constrain the generalization and robustness of deep learning networks. To address this, future endeavors will focus on collecting more extensive data to facilitate large-scale, multi-center training and validation, ensuring more comprehensive and reliable results.

In our study, we investigated the predictors of therapeutic outcomes in ovarian cancer, acknowledging that these are influenced by a range of factors beyond tumor morphology, such as pathological stage, resection margins, patient performance status, and comorbidities. We developed a deep learning model, using pathological slides as the sole input, to predict drug resistance, exploring the potential of pathological sections as an independent biomarker. While integrating clinical and pathological data could improve predictive accuracy, our initial focus on pathological slides due to scope and time constraints represents a deliberate first step towards a comprehensive research strategy. Plans to include clinical data in future analyses acknowledge the opportunity to enhance drug resistance predictions. We analyzed additional clinical data, including Federation International of Gynecology and Obstetrics (FIGO) stage, age, and BMI, through logistic regression to assess their relationship with platinum resistance. The p-values for these factors (0.2052, 0.9191, and 0.3393, respectively) suggest they are poor predictors of platinum resistance, as evidenced by AUC values of 0.58, 0.51, and 0.54 in independent tests. Conversely, our deep learning analysis of pathological images with MAMILNet demonstrated higher predictive accuracy for treatment response, achieving an AUC of 0.7358, with significant accuracy, FPR, and FNR rates. We aim to extend our research to include broader clinical parameters, enhancing prediction accuracy and understanding of therapeutic outcome determinants in ovarian cancer. This multidimensional approach promises to refine our predictive models and contribute valuable insights into the complex dynamics of cancer treatment response.

TABLE 3 Prediction results on the independent test set of drug resistance in high-grade serous ovarian cancer.

| Deep-learning Model           | AUC           | Accuracy      | FNR           | FPR           |
|-------------------------------|---------------|---------------|---------------|---------------|
| 5× only MAMILNet              | 0.6741        | 0.6582        | 0.3636        | 0.3333        |
| 10× only MAMILNet             | 0.6968        | 0.6835        | 0.2727        | 0.3333        |
| 20× only MAMILNet             | 0.7126        | 0.7088        | 0.2727        | 0.2982        |
| MILRNN Campanella et al. (14) | 0.6537        | 0.6329        | 0.3333        | 0.4118        |
| CLAM Lu et al. (28)           | 0.6884        | 0.6709        | 0.2727        | 0.4000        |
| <b>MAMILNet (ours)</b>        | <b>0.7358</b> | <b>0.7341</b> | <b>0.1818</b> | <b>0.2982</b> |



In summary, deep-learning-based WSI analysis has emerged as a crucial approach for cancer diagnosis. This study introduces a novel multi-scale attentional multi-instance network architecture (MAMILNet), presenting a fresh perspective and method for WSI analysis using deep learning. Notably, MAMILNet demonstrates promising results in breast cancer tumor detection, lung cancer tumor typing, and ovarian cancer drug resistance prediction tasks. These achievements offer valuable insights for the wider application of deep-learning models in these areas and inspire new avenues for utilizing deep learning in diverse cancer types and diagnostic tasks. MAMILNet's performance signifies its potential as a powerful tool for enhancing cancer diagnosis and treatment in clinical settings.

## 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 humans were approved by the First People's Hospital of Yunnan Province. The studies were conducted in accordance with the local legislation and institutional requirements. The human samples used in this study were acquired from primarily isolated as part of your previous study for which ethical approval was obtained. 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. The manuscript presents research on animals that do not require ethical approval for their study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

## References

1. Srinidhi CL, Ciga O, Martel AL. Deep neural network models for computational histopathology: A survey. *Med Image Anal.* (2021) 67:101813. doi: 10.1016/j.media.2020.101813
2. Qu L, Liu S, Liu X, Wang M, Song Z. Towards label-efficient automatic diagnosis and analysis: a comprehensive survey of advanced deep learning-based weakly-supervised, semi-supervised and self-supervised techniques in histopathological image analysis. *Phys Med Biol.* (2022) 67(20):20TR01. doi: 10.1088/1361-6560/ac910a
3. Cheplygina V, de Bruijne M, Pluim JP. Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Med Image Anal.* (2019) 54:280–96. doi: 10.1016/j.media.2019.03.009
4. Rony J, Belharbi S, Dolz J, Ayed IB, McCaffrey L, Granger E. Deep weaklysupervised learning methods for classification and localization in histology images: a survey. *arXiv preprint arXiv:1909.03354.* (2019).
5. Wang Z, Bi Y, Pan T, Wang X, Bain C, Bassed R, et al. Multiplex-detection based multiple instance learning network for whole slide image classification. *arXiv preprint arXiv:2208.03526.* (2022).
6. Qu L, Ma Y, Luo X, Wang M, Song Z. Rethinking multiple instance learning for whole slide image classification: A good instance classifier is all you need. *arXiv preprint arXiv:2307.02249.* (2023).
7. Coudray N, Ocampo PS, Sakellaropoulos T, Narula N, Snuderl M, Fenyö D, et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. *Nat Med.* (2018) 24:1559–67. doi: 10.1038/s41591-018-0177-5
8. Naik N, Madani A, Esteva A, Keskar NS, Press MF, Ruderman D, et al. Deep learningenabled breast cancer hormonal receptor status determination from base-level h&e stains. *Nat Commun.* (2020) 11:5727. doi: 10.1038/s41467-020-19334-3
9. Tomita N, Abdollahi B, Wei J, Ren B, Suriawinata A, Hassanpour S. Attention-based deep neural networks for detection of cancerous and precancerous esophagus tissue on histopathological slides. *JAMA Network Open.* (2019) 2:e1914645–e1914645. doi: 10.1001/jamanetworkopen.2019.14645
10. Skrede O-J, De Raedt S, Kleppe A, Hveem TS, Liestøl K, Maddison J, et al. Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. *Lancet.* (2020) 395:350–60. doi: 10.1016/S0140-6736(19)32998-8
11. Kather JN, Pearson AT, Halama N, Jäger D, Krause J, Loosen SH, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. *Nat Med.* (2019) 25:1054–6. doi: 10.1038/s41591-019-0462-y
12. Bejnordi BE, Veta M, Van Diest PJ, Van Ginneken B, Karssemeijer N, Litjens G, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA.* (2017) 318:2199–210. doi: 10.1001/jama.2017.14585

## Author contributions

KM: Conceptualization, Data curation, Investigation, Project administration, Resources, Supervision, Writing – review & editing. QW: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – original draft. QB: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Validation, Writing – review & editing. LQ: Conceptualization, Methodology, Software, Writing – review & editing. YD: Data curation, Investigation, Resources, Writing – review & editing. XW: Data curation, Investigation, Resources, Writing – review & editing. YZ: Data curation, Investigation, Resources, Writing – review & editing. CL: Data curation, Investigation, Resources, Writing – review & editing. QM: Conceptualization, Resources, Supervision, 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.

13. Chaudhary K, Poirion OB, Lu L, Garmire LX. Deep learning-based multi-omics integration robustly predicts survival in liver cancer. *Clin Cancer Res.* (2018) 24:1248–59. doi: 10.1158/1078-0432.CCR-17-0853
14. Campanella G, Hanna MG, Geneslaw L, Miraflor A, Werneck Krauss Silva V, Busam KJ, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med.* (2019) 25:1301–9. doi: 10.1038/s41591-019-0508-1
15. Saillard C, Schmauch B, Laifa O, Moarii M, Toldo S, Zaslavskiy M, et al. Predicting survival after hepatocellular carcinoma resection using deep learning on histological slides. *Hepatology.* (2020) 72:2000–13. doi: 10.1002/hep.31207
16. Woerl A-C, Eckstein M, Geiger J, Wagner DC, Daher T, Stenzel P, et al. Deep learning predicts molecular subtype of muscle-invasive bladder cancer from conventional histopathological slides. *Eur Urol.* (2020) 78:256–64. doi: 10.1016/j.eururo.2020.04.023
17. Anand D, Yashashwi K, Kumar N, Rane S, Gann PH, Sethi A. Weakly supervised learning on unannotated h&e-stained slides predicts braf mutation in thyroid cancer with high accuracy. *J Pathol.* (2021) 255:232–42. doi: 10.1002/path.5773
18. Velmahos CS, Badgeley M, Lo Y-C. Using deep learning to identify bladder cancers with fgfr-activating mutations from histology images. *Cancer Med.* (2021) 10:4805–13. doi: 10.1002/cam4.4044
19. Wessels F, Schmitt M, Krieghoff-Henning E, Jutzi T, Worst TS, Waldbillig F, et al. Deep learning approach to predict lymph node metastasis directly from primary tumour histology in prostate cancer. *BJU Int.* (2021) 128:352–60. doi: 10.1111/bju.15386
20. Li F, Yang Y, Wei Y, He P, Chen J, Zheng Z, et al. Deep learning-based predictive biomarker of pathological complete response to neoadjuvant chemotherapy from histological images in breast cancer. *J Trans Med.* (2021) 19:1–13. doi: 10.1186/s12967-021-03020-z
21. Yang J, Ju J, Guo L, Ji B, Shi S, Yang Z, et al. Prediction of her2-positive breast cancer recurrence and metastasis risk from histopathological images and clinical information via multimodal deep learning. *Comput Struct Biotechnol J.* (2022) 20:333–42. doi: 10.1016/j.csbj.2021.12.028
22. Ilse M, Tomczak J, Welling M. Attention-based deep multiple instance learning. In *International conference on machine learning.* (2018) 2127–36.
23. Shi X, Xing F, Xie Y, Zhang Z, Cui L, Yang L. Loss-based attention for deep multiple instance learning. In *Proceedings of the AAAI conference on artificial intelligence.* (2020) 34(04):5742–9.
24. Li B, Li Y, Eliceiri KW. Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.* (2021) 14318–28.
25. Qu L, Luo X, Liu S, Wang M, Song Z. Dgmil: Distribution guided multiple instance learning for whole slide image classification. In: *Medical image computing and computer assisted intervention (MICCAI).* Cham: Springer Nature Switzerland (2022). p. 24–34.
26. Qu L, Wang M, Song Z. Bi-directional weakly supervised knowledge distillation for whole slide image classification. *Adv Neural Inf Process Syst.* (2022) 35:15368–81.
27. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* (2016) 770–8.
28. Lu MY, Williamson DF, Chen TY, Chen RJ, Barbieri M, Mahmood F. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat Biomed Eng.* (2021) 5:555–70. doi: 10.1038/s41551-020-00682-w
29. Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. *Nat Rev Drug Discovery.* (2019) 18:463–77. doi: 10.1038/s41573-019-0024-5
30. Ballester PJ, Stevens R, Haibe-Kains B, Huang RS, Aittokallio T. Artificial intelligence for drug response prediction in disease models. *Briefings in bioinformatics.* (2022) 23(1):bbab450. doi: 10.1093/bib/bbab450
31. Farahmand S, Fernandez AI, Ahmed FS, Rimm DL, Chuang JH, Reisenbichler E, et al. Deep learning trained on hematoxylin and eosin tumor region of interest predicts her2 status and trastuzumab treatment response in her2+ breast cancer. *Modern Pathol.* (2022) 35:44–51. doi: 10.1038/s41379-021-00911-w



## OPEN ACCESS

## EDITED BY

Liang Qiao,  
The University of Sydney, Australia

## REVIEWED BY

Jaya Lakshmi Thangaraj,  
University of California, San Diego,  
United States  
Xiao-Wan Bo,  
Tongji University, China

## \*CORRESPONDENCE

Amir Borhani

✉ amir.borhani@nm.org

RECEIVED 28 December 2023

ACCEPTED 12 April 2024

PUBLISHED 08 May 2024

## CITATION

Haghshomar M, Rodrigues D, Kalyan A,  
Velichko Y and Borhani A (2024) Leveraging  
radiomics and AI for precision diagnosis and  
prognostication of liver malignancies.  
*Front. Oncol.* 14:1362737.  
doi: 10.3389/fonc.2024.1362737

## COPYRIGHT

© 2024 Haghshomar, Rodrigues, Kalyan,  
Velichko and Borhani. 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.

# Leveraging radiomics and AI for precision diagnosis and prognostication of liver malignancies

Maryam Haghshomar, Darren Rodrigues, Aparna Kalyan,  
Yury Velichko and Amir Borhani\*

Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

Liver tumors, whether primary or metastatic, have emerged as a growing concern with substantial global health implications. Timely identification and characterization of liver tumors are pivotal factors in order to provide optimum treatment. Imaging is a crucial part of the detection of liver tumors; however, conventional imaging has shortcomings in the proper characterization of these tumors which leads to the need for tissue biopsy. Artificial intelligence (AI) and radiomics have recently emerged as investigational opportunities with the potential to enhance the detection and characterization of liver lesions. These advancements offer opportunities for better diagnostic accuracy, prognostication, and thereby improving patient care. In particular, these techniques have the potential to predict the histopathology, genotype, and immunophenotype of tumors based on imaging data, hence providing guidance for personalized treatment of such tumors. In this review, we outline the progression and potential of AI in the field of liver oncology imaging, specifically emphasizing manual radiomic techniques and deep learning-based representations. We discuss how these tools can aid in clinical decision-making challenges. These challenges encompass a broad range of tasks, from prognosticating patient outcomes, differentiating benign treatment-related factors and actual disease progression, recognizing uncommon response patterns, and even predicting the genetic and molecular characteristics of the tumors. Lastly, we discuss the pitfalls, technical limitations and future direction of these AI-based techniques.

## KEYWORDS

radiomics, AI, liver tumors, HCC, review

## Introduction

Liver tumors, both primary and metastatic, have become a growing global health concern with significant implications. Treating HCC remains challenging given the heterogeneity and complexity of the disease. Most HCC patients have underlying cirrhosis or chronic inflammation. The microscopic changes, in the setting of chronic inflammation, makes HCC an ideal disease state to consider for targeted therapy. While immunotherapy has changed the first line treatment paradigms, there remains a paucity of treatment options in patients who either progress on immunotherapy or are intolerant of these agents. Historically, treatments for HCC have been based on the Barcelona Clinic liver cancer staging system, with the assessment of tumor burden, liver function, and general health status guiding the selection of the best treatment modality (1). However, in the era of precision medicine, tumor biomarkers and treatment selection challenge the one-size-fits-all concept in HCC.

The low sensitivity and specificity of biomarkers has rendered selection of treatment to be difficult. While Alpha-fetoprotein (AFP) has historically been used for detection of early, potentially curable tumors, it is limited by its sensitivity to make treatment decisions (2). Biomarkers that predict response to systemic therapy are urgently needed. Presently, AFP is the only biomarker to predict response, and only in a subset of patients who receive ramucirumab as a second-line agent. Using cell free DNA's genomic and epigenetic changes potentially offers a more sensitive and promising biomarker, especially for detecting minimal residual disease (2). Genetic changes detected by means of circulating tumor DNA allows improved understanding of tumor biology and disease heterogeneity.

Imaging plays a vital role in detecting liver tumors but sometimes conventional methods often lack the precision needed for proper characterization, leading to the need for invasive tissue biopsy. Conventional imaging methods provide limited information on the prognostic factors of liver tumors, such as genetic mutations, molecular markers, and potential treatment response. This information gap delays personalized treatment planning and prognostication.

Rapid advancements in imaging and post-processing techniques have revolutionized high-throughput image analysis, enabling a more precise and comprehensive evaluation of liver diseases. Artificial intelligence (AI) and radiomics have emerged as promising methods with the potential to revolutionize liver lesion characterization. AI and radiomics can analyze medical images at a high level of detail, identifying subtle patterns that correlate with specific tumor types, stages, and biological characteristics. These methods are fast, affordable and readily available. AI and radiomics can do simple tasks and handle a huge amount of data with the same accuracy, meaning that missing manual steps, fatigue, or data overload won't affect the findings.

We provide an outline of radiomics and AI contributions to diagnosis and staging, treatment response assessment, and prognosis prediction in liver malignancies in this review. We describe the progress and potential of AI in the liver oncology

imaging, focusing specifically on radiomic and deep-learning techniques.

## Radiomics and artificial intelligence

Radiomics, a framework that complements conventional radiological interpretation, has emerged as a powerful tool for extracting and quantifying texture characteristics derived from tumor heterogeneity (3–6) (Figure 1). Radiomics employs a wide range of method, each designed to capture specific aspects of tissue architecture and texture. These features provide a detailed representation of tumor heterogeneity, enabling researchers to study and compare tumor characteristics across different patients. The standardization of radiomics features ensures the reproducibility and generalizability of radiomics studies, fostering collaboration and wider adoption of this technique. Development of openly available imaging datasets further creates opportunities to test and benchmark radiomics algorithms and facilitate the translation of radiomics findings into the clinical practice.

Radiomics features extracted from large datasets enable the development of advanced statistical models, including machine learning and artificial intelligence algorithms. These models can enhance various aspects of liver imaging assessment, including tumor origin identification, therapy response prediction, and prognosis assessment. For instance, radiomics provides valuable insights into tumor characteristics, such as aggressiveness and prognosis, which can inform treatment decisions. Another example includes delta-radiomic models, which allow for longitudinal assessment of changes in tumor texture to assess tumor response to treatment. This enables timely adjustments to treatment regimens and improves overall treatment efficacy. Furthermore, radiomics-based predictive models can personalize treatment strategies for individual patients, tailoring treatment to their specific tumor characteristics and maximizing treatment success.

Harnessing the power of neural networks, AI in medical imaging extracts intricate patterns from large datasets and can improve informed predictions. The convolution operation, a cornerstone of many neural networks, employs diverse kernels to transform raw data into meaningful representations, enabling neural networks to learn from and make predictions on complex datasets. Deep learning, a powerful subfield of AI, utilizes many interconnected layers that transform information, enabling more sophisticated information processing. Deep learning's ability to automatically learn features and representations from data stands out as a key strength, eliminating the need for explicit feature engineering by human experts. This capability makes deep learning particularly well-suited for various clinical tasks. For instance, deep learning algorithms can accurately detect and localize objects within images, enabling the identification of anatomical structures or abnormalities in medical scans. Other models can be trained to precisely segment objects in images, allowing for the delineation of organs and lesions. Segmented organs or lesions can be effectively classified into distinct categories, aiding in disease diagnosis and

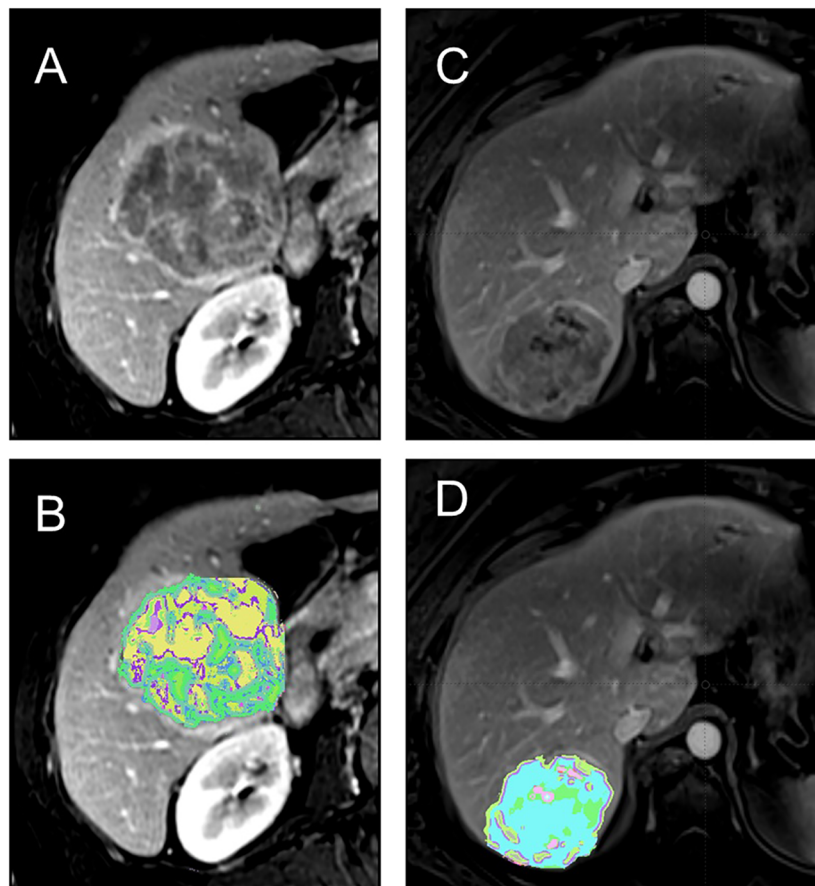


FIGURE 1

(A, B) 64 y/o M with history of cirrhosis and HCC. Contrast-enhanced T1-weighted MRI shows a heterogeneous tumor (A) with associated texture heterogeneity map demonstrating tumor habitats (B). The patient had poor outcome with several recurrent lesions after surgical resection suggestive of poor tumor biology. (C, D) 54 y/o F with history of cirrhosis and HCC. Contrast-enhanced T1-weighted MRI shows a less heterogeneous tumor (C) with associated texture heterogeneity map showing the tumor habitats (D). The patient good outcome after resection with no recurrence.

treatment monitoring. Radiomics can be used to identify the origin of segmented lesions. The integration of radiomics with deep learning has emerged as a promising strategy for enhancing classification performance in medical imaging. Deep learning algorithms possess the ability to complement radiomic features with kernel-based features and then extract patterns from the high-dimensional imaging data. This synergistic combination has yielded noticeable advancements in classification accuracy for a wide range of medical imaging tasks.

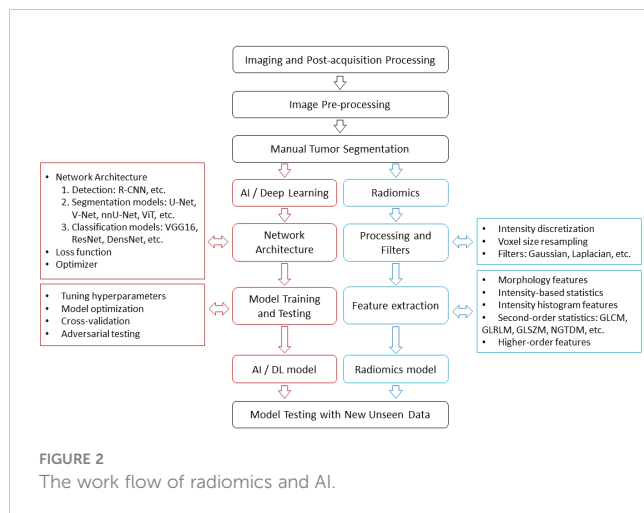
While AI-based approaches offer a diverse toolbox for image analysis, both radiomics and deep learning share a similar workflow including collection and standardization of imaging data, image pre-processing, and segmentation of relevant regions depending on the task. In liver disease analysis, this involves whole liver segmentation, segmentation of various structures like lesions, gallbladder, bile ducts, and vascular components. The whole liver segmentation allows for evaluation of liver morphology, liver surface, and parenchymal changes such as fibrosis and cirrhosis. Segmentation and detailed analysis of individual lesions, including their count, size, heterogeneity, necrosis, and vascular involvement, can provide valuable insights for staging, treatment planning, and prognosis (Figure 2).

## Segmentation

Segmentation of the liver or its vasculature through CT and MRI has importance in diagnosing lesions, planning radiotherapy, conducting liver vascular surgeries, preparing for liver transplantation, and analyzing tumor vascularization, among other applications. The process of manual segmentation is both time-consuming and susceptible to human errors. Several researchers have explored the use of deep learning models to automate this process, aiming to overcome these limitations. Segmentation studies exhibit remarkable specificity in liver imaging, compared to other fields. The mainstream segmentation technology at present is U-Net, a convolutional neural network (CNN), and its derivatives. The segmentation accuracy for the entire liver, as assessed on the SLIVER07 dataset, is exceptionally high, achieving a maximum Dice Similarity Coefficient (DSC; which is a metric of accuracy of the method) of 0.9827 (7).

Said et al. employed CNNs for HCC segmentation in MRI, achieving fair to good performance, notably excelling in single-slice segmentation with mean DSC ranging from 0.442 to 0.778 in 292 patients (8). Another study focused on automating colorectal liver





metastasis and liver ablation zone segmentation on CT, with the Hybrid-WNet model demonstrating high accuracy. Trained on 92 patients, the model achieved a median DSC of 0.73 (9). Additionally, a feasibility study utilized a deep convolutional neural network (DCNN) to automate the application of LI-RADS (Liver Imaging Reporting and data System) algorithm on multiphasic MRI, showcasing efficient liver and HCC segmentation. The average DSC for automatically identified lesions using the DCNN+ random forest classifier employing radiomic features and thresholding was 0.64 in the validation set and 0.68 in the test set (10). One paper presented an automatic algorithm for rapid and accurate liver and lesion segmentation in CT scans. Achieving a 94.2% volume overlap and 3.7 mm accuracy for liver surface segmentation, the method demonstrated a short processing time of 11.4 seconds per slice. Tumor lesion detection showed 82.6% sensitivity and 87.5% specificity (11). A separate study used a deep-learning model for HCC segmentation and classification using gadoxetic acid-enhanced MRI. The 3D U-Net-based model achieved high performance (average DSC of 0.884) for HCC segmentation (12). Lastly, a dual-energy CT radiomics pilot study successfully differentiated between benign and malignant hepatic lesions, outperforming iodine quantification. This involved semiautomatic segmentation of both the volume and rim of individual liver lesions, along with extracting contrast enhancement, iodine concentrations, and radiomic features from each image (13).

Deep learning models can perform HCC segmentation with a high accuracy. This has multiple clinical applications. These models can handle a huge amount of data with the same accuracy, meaning that missing manual steps, fatigue, or data overload won't affect the findings. Computer-based processing is affordable and readily availability.

## Early detection and accurate tumor classification

Identifying liver cancers in their early stages, when they are localized and amenable to curative treatment, is the ultimate goal.

Ideally, the cancer should be diagnosed when patient is asymptomatic as the emergence of symptoms often signifies late-stage, incurable disease in many cases. Moreover, early small tumors should be characterized correctly to allow for correct treatment since different tumor pathologies have distinct course and different treatment approaches. Unfortunately, primary liver cancers may have nonspecific imaging features in their early stages due to their smaller size. Equally significant is the early identification of liver metastases, highlighting the importance timely and precise detection. At the same time addressing and preventing false positives, overdiagnosis, and overtreatment is essential (14). As an example, combined HCC and cholangiocarcinoma (CC) originates from hepatic progenitor cells and can display both hepatocytic and cholangiocytic differentiation (15, 16). HCC the most common hepatic malignancy is histologically derived from the hepatocytes and CC the second most common hepatic malignancy is derived from the biliary epithelial cells. Studies have demonstrated that the clinical features and prognosis of combined HCC-CC markedly differ from those of intrahepatic CC (IHCC) and HCC (17, 18). Hence, it holds significant clinical implications to differentiate these tumors preoperatively accurately.

Radiomics integrated with machine learning algorithms has promising role in distinguishing diverse focal hepatic lesions. The features extracted may allow for noninvasive diagnosis and characterization of liver malignancies and provide vital details such as microvascular invasion within tumors. AI has also been extensively employed for classifying different liver lesions. CNNs specifically designed for image recognition tasks have attracted considerable attention for liver cancer diagnosis.

Numerous large scale studies utilizing CT or MR imaging have employed radiomics to distinguish various liver lesions, yielding areas under ROC curves (AUC) ranging from 0.7 to 0.95 (19–29). These investigations demonstrated robust performance not only on the training set but also on testing and validation sets. The scope of these studies encompassed a wide range of classification tasks and discriminating lesions, including HCC, hemangioma, cysts, adenoma, hepatic focal nodular hyperplasia, CC, combined HCC-CC, inflammatory masses, and metastasis. Clinical variables were integrated into certain models to enhance their performance (19–29). A multitude of AI studies has endeavored to predict liver malignancies, focusing on diverse aspects such as detecting HCC (30–32), classifying major features of LI-RADS (12, 33, 34), and discerning classic HCC form other malignant and nonmalignant liver lesions. AUC values in either the training or validation sets varied from 0.6 to 0.942 (12, 30–34).

Presence of microvascular invasion (MVI) is identified as an independent risk factor for the postoperative recurrence of HCC (35). The definitive assessment for MVI is based on histologic examination of surgical specimen, which is only available after resection of tumor. As a result, assessing the MVI status before surgery will play a crucial role in guiding decisions regarding the optimal extent of surgical resection or ablation treatment for individuals with HCC. Several studies using AI or radiomic features extracted from gadoxetic acid-enhanced MRI, dynamic contrast enhanced MR, or contrast enhanced CT images tried to

predict microvascular invasion in HCC and mass-forming CC (36). The AUCs ranged from 0.75 to 0.98 with most of the studies achieving AUCs higher than 0.85 (36–47). Notably, studies focused on peritumoral areas within the 5 cm to 10 cm range. One study underscored that patients without MVI experienced significantly prolonged recurrence-free survival (RFS). Validation sets were incorporated in all studies (36–47). As mentioned above, accurately predicting MVI before surgery can significantly influence surgical planning, including decisions regarding the extent of resection or the suitability of ablation treatments. Such high AUCs and predictive capabilities mean that presence of MVI can be successfully determined with AI and radiomics prior to surgery allowing for a more personalized surgical approach, potentially improving postoperative outcomes and recurrence-free survival for patients with HCC.

## Grading, association with molecular profile, immunophenotype, etc.

HCC histopathological grading has been identified to be closely associated with the prognosis of the tumor, serving as an indicator of the tumor's biological behavior. Extensive research indicates that both progression-free survival and overall survival are notably lower in poorly differentiated HCC compared to well-differentiated HCC. Certain subtypes of HCCs, such as macrotrabecular-massive subtype, are also correlated with worse prognosis. Pre-operative knowledge of tumor grading affects treatment plan and surgical approach, when surgery is indicated. For example, the recommendation is to opt for an extended resection margin when conducting liver surgery for poorly differentiated HCC to minimized the risk of early recurrence. Some data suggest recommendation against liver transplantation for patients with HCC that is both poorly differentiated and exceeds 3 cm in size. Preoperative knowledge of tumor grading is classically achieved by histologic examination of biopsy specimen. Biopsy however is an invasive procedure and is not feasible in all patients (due to patient's factors and location/size of the tumor). Additionally, given the high success of imaging-based criteria for noninvasive diagnosis of HCC (such as LI-RADS criteria), biopsy is not routinely performed in this population.

Several radiomics models utilizing gadoteric acid-enhanced MRI, some augmented by AI, have aimed for HCC subtyping and grading to overcome these issues. They have achieved AUCs ranging from 0.6 to 0.912 (48–53). Notably, lower grades were correlated with longer progression-free survival in one cohort. Additionally, the radiomics model demonstrated associations with dysregulated humoral immunity, encompassing B-cell infiltration and immunoglobulin synthesis, offering valuable insights into the immune microenvironment of HCC (48–53).

Comprehensive knowledge of the molecular profile and immunophenotype of liver cancers is also relevant for advancing precision oncology. The tumor microenvironment and immune status are integral factors influencing the success of immunotherapies and locoregional treatments in HCC (54). Gene expression analysis has revealed distinct immune classes among HCC patients and immune

profiling of HCC can predict response to immunotherapy (55). Preliminary works have indicated the potential of radiomics quantification in immune profiling for HCC. Notably, these works studied expression of vascular endothelial growth factor (VEGF) (56), angiopoietin-2 (57), Forkhead Box M1 (FOXM1) (58), and Ki-67 (59, 60). Additionally, the presence of  $\beta$ -catenin mutation (61), intra-tumoral tertiary lymphoid structures (62), cytokeratin 19 (63, 64), glypican-3 (GPC3) (65), immunohistochemical cell type markers for T-cells (CD3), macrophages (CD68) and endothelial cells (CD31), PD1 and CTLA4 at mRNA expression level (66), as well as density of CD3+ and CD8+ T cells (67) were studied. All the aforementioned molecules have relevant task in carcinogenesis. VEGF and Angiopoietin-2 regulate tumor growth by influencing angiogenesis. FOXM1 governs cell cycle genes, Ki-67 marks proliferation, and  $\beta$ -Catenin mutation leads to uncontrolled cell growth. Intra-tumoral Tertiary Lymphoid Structures impact the anti-tumor immune response. Cytokeratin 19 maintains cell structure, while GPC3 serves as a tumor diagnostic marker. Immune cell markers like CD3, CD68, and CD31 reveal cell distribution and density, reflecting the local immune response. PD1 and CTLA4 mRNA levels influence responses to checkpoint inhibitors. Each of these immune subtypes plays a critical role in unraveling the complex immune response within HCC, providing insights for prognostication and targeted therapeutic interventions. AUCs of these tasks fell somewhere between 0.76 to 0.95 (56–67). Notably, when clinical factors were integrated with radiomics signatures, models' performance significantly improved. In the MRI studies, the hepatobiliary phase consistently demonstrated the best performance.

While deep learning models haven't been as widely applied as radiomics for this particular task, they undoubtedly hold significant potential. Xie et al. introduced a non-invasive method for predicting PD-1 and PD-L1 expression in HCC. Using a cohort of 87 HCC patients and analyzing 3094 CT images, the Contrastive Learning Network (CLNet) was proposed. Trained with self-supervised contrastive learning, CLNet achieved superior performance, demonstrated an AUC of 86.6 for PD-1 expression and 83.9 for PD-L1 expression (68) (Table 1).

High AUC values in both radiomics and deep learning tasks indicate strong predictive performance, meaning these models are highly effective in identifying molecular profiles, immunophenotypes and grades of HCC.

## Assessment of tumor response

Several locoregional therapeutic strategies have been developed and implemented over past decades, and a considerable number of these are currently considered as the standard of care for liver malignancies (69). These involve a range of percutaneous and trans-arterial methods designed to induce cell death in tumors. This can be achieved through percutaneous approach, as seen in radiofrequency and microwave ablation, or achieved via targeted trans-catheter trans-arterial administration of embolic agents (known as trans-arterial bland embolization, TAE), chemotherapeutic substances (referred to as trans-arterial chemoembolization, TACE), or radioembolizing agents (as in trans-arterial Yttrium-90 radioembolization, TARE) (70). While typically less invasive

compared to surgical removal and transplantation, these therapies can lead to complications. Hence, careful patient selection and thorough evaluation of treatment response are crucial clinical considerations. Evaluating the response to treatment following TARE and external beam radiation, particularly in the initial months post-treatment, poses challenges due to the intrinsic characteristics and timeline of cytotoxic effects induced by radiation. Sustained enhancement in the arterial phase and subsequent washout can be observed in treated lesions for several months following the treatment, even though a complete response is ultimately achieved (71). Considering these limitations, it is essential to explore alternative approaches for evaluating treatment response. Numerous studies have explored the potential of utilizing radiomic features extracted from post-treatment CT and MRI in assessing the treatment response of HCC (72–79). There are limited studies on other liver tumors (80, 81). Radiomics features were

extracted from diverse imaging modalities, including MRI and CT scans, and involved different treatment methods. In some studies key clinical information, such as albumin levels, AFP levels, and Child-Pugh score were integrated into predictive models to enhance accuracy. The studies anticipated diverse outcomes encompassing early response, early recurrence, aggressive intrasegmental recurrence, tumor refractoriness, and local tumor progression across varied locoregional strategies. The AUC values of these studies ranged from 0.8 to 0.95 (72–81). These studies collectively underscore the potential of radiomics in tailoring treatment strategies and prognostic assessments for liver cancer patients, providing a non-invasive means to predict outcomes and guide personalized interventions based on comprehensive imaging analyses and relevant clinical parameters.

The utilization of deep learning to evaluate locoregional therapeutic responses in HCC is relatively limited in the current

TABLE 1 Summary of HCC grading, molecular profiling, and immunophenotyping.

| Author  | Marker                                     | Subjects | Modality | Model         | Accuracy training                                                                                                 | Accuracy testing                | Accuracy validation                                                                        |
|---------|--------------------------------------------|----------|----------|---------------|-------------------------------------------------------------------------------------------------------------------|---------------------------------|--------------------------------------------------------------------------------------------|
| Chen    | Immunoscore                                | 207 HCC  | MRI      | Radiomics/ ML | AUC, accuracy, sensitivity, specificity = 0.904, 0.787, 93.8%, 74.6% combined- AUC = 0.926                        | NA                              | AUC, accuracy, sensitivity, specificity = 0.899, 0.772, 92.3%, 72.7% combined- AUC = 0.934 |
| Chen    | FOXM1 expression                           | 286 HCC  | CT       | Radiomics/ ML | AUC = 0.918                                                                                                       | AUC = 0.837                     | NA                                                                                         |
| Fan     | VEGF expression                            | 202 HCC  | MRI      | Radiomics/ ML | AUC = 0.892 combined- AUC = 0.936                                                                                 | AUC: 0.8/ combined- AUC = 0.836 | NA                                                                                         |
| Hectors | Immunoprofiling and genomics               | 48 HCC   | MRI      | Radiomics/ ML | Tumor size ≥ 5 cm - HCC recurrence (OR = 3.01, p = 0.004, AUC = 0.76).                                            | NA                              | NA                                                                                         |
| Li      | Intra-tumoral tertiary lymphoid structures | 142 HCC  | CT       | Radiomics/ ML | AUC = 0.79                                                                                                        | NA                              | AUC = 0.75                                                                                 |
| Wang    | cytokeratin 19 expression                  | 227 HCC  | MRI      | Radiomics/ ML | AUC = 0.892 combined- AUC, sensitivity, specificity, C-index = 0.951, 0.818, 0.974, 0.959                         | NA                              | AUC = 0.73 combined- AUC, sensitivity, specificity, C-index = 0.822, 0.769, 0.818, 0.846   |
| Wu      | Ki-67 expression                           | 172 HCC  | CT       | Radiomics/ ML | AUC = 0.854 combined- AUC = 0.884                                                                                 | NA                              | AUC: 0.744 combined- AUC = 0.819                                                           |
| Yan     | Ki67 expression                            | 110 HCC  | MRI      | Radiomics/ ML | AUC = 0.833 combined- AUC = 0.901                                                                                 | NA                              | AUC: 0.772 combined- AUC = 0.781                                                           |
| Zeng    | β-catenin mutation                         | 98 HCC   | MRI      | Radiomics/ ML | AUC, accuracy, sensitivity, specificity = 0.86, 0.75, 1.0, 0.65 combined- AUC = 0.86                              | NA                              | AUC, accuracy, sensitivity, specificity = 0.82, 0.73, 0.67, 0.76 combined- AUC = 0.76      |
| Zhang   | cytokeratin 19 expression                  | 311 HCC  | MRI      | Radiomics/ ML | C-index, 0.914                                                                                                    | C-index, 0.855                  | C-index, 0.795                                                                             |
| Zhang   | glypican-3 expression                      | 137 HCC  | MRI      | Radiomics/ ML | AUC, sensitivity, specificity = 0.822, 0.816, 0.706 combined- AUC, sensitivity, specificity = 0.888, 0.777, 0.912 | NA                              | combined- AUC, sensitivity, specificity = 0.800, 0.58.5, 1.0                               |
| Zheng   | angiopoietin-2 expression                  | 52 HCC   | MRI      | Radiomics/ ML | AUC = 0.8 combined- AUC = 0.933                                                                                   | NA                              | NA                                                                                         |
| Xie     | PD-1 and PD-L1 expression                  | 87 HCC   | CT       | AI-DL         | AUC = 0.866 for PD-1 expression AUC = 0.839 for PD-L1 expression                                                  | NA                              | NA                                                                                         |

body of research, yet, it's important to note that the studies presented are novel, and there's considerable unexploited potential in this evolving field. Three studies employed deep learning to explore the response of TACE in HCC (82–84). In a study involving 414 patients, hazard ratios for time to progression (TTP) were 0.32 (training), 0.28 (validation), and 0.55 (test). The research also indicated improved overall survival (OS) with a hazard ratio of 0.58 and a median survival of 38.8 months, compared to 20.9 months (82). Another investigation with 789 patients achieved an 84.3% accuracy, showing AUCs of 0.97, 0.96, 0.95, and 0.96 for complete response (CR), partial response (PR), stable disease (SD), and progressive disease (PD), respectively. The deep learning model displayed accuracies of 85.1% and 82.8% across CR, PR, SD, and PD in two validation sets (83). The deep learning signature showed strong predictive performance, with a C-index of 0.717 in the training set and 0.714 in the validation set (84).

One study developed an automatic and non-invasive deep learning radiomic nomogram (DLRN) to predict hepatic arterial infusion chemotherapy response in HCC. Utilizing contrast-enhanced CT images from 458 patients across three hospitals, the DLRN achieved high AUC values of 0.988 (training), 0.915 (internal validation), and 0.896 (external validation), outperforming other models. The DLRN also successfully stratified survival risk, with the predictive objective response group exhibiting significantly longer overall survival (26.0 vs. 12.3 months) (85).

The ability of AI and radiomics to predict early treatment response and recurrence can improve the management of liver cancer. By identifying patients at risk of aggressive recurrence or poor response to treatment early on, clinicians can adjust therapeutic strategies promptly have shown that these methods have consistently high accuracies.

## Prognostication

The prognosis of HCC continues to be unfavorable, even following curative-intent treatments like liver resection or transplantation. After liver resection, the early recurrence rate stands at 50–70%, while following a median post-transplant period of 13–14 months, the rate is 10–20% (86, 87).

By analyzing detailed features from medical images, radiomics helps predict recurrence, understand tumor growth, and estimate progression-free survival. When combined with essential clinical details, radiomics can become a powerful tool for predicting aggressive disease and customizing treatments. This approach offers a non-invasive, precise way to enhance prognostic assessments, bringing a new level of accuracy to liver cancer care.

Multiple studies - including ten studies on HCC (88–97), four studies on Mass-forming CC (98–101), and three studies on colorectal liver metastases (102–104)- utilized various radiomics approaches to predict outcomes and guide treatment decisions. The studies involved diverse cohorts, including patients undergoing liver transplantation, surgical resection, or chemotherapy. The endpoint outcomes ranged from overall survival (OS), recurrence free survival (RFS), progression-free survival (PFS), event-free survival (EFS), early recurrence (ER), 1-year survival and 5-year

survival, post-hepatectomy liver failure (PHLF), and lymph node metastasis. The AUCs for predictive models varied, ranging between 0.70 to 0.98 (88–104). Moreover, the integration of radiomics with clinical factors consistently improved predictive performance, demonstrating the potential for personalized risk assessment. Notably, radiomics models were applied to predict survival in various contexts, offering valuable insights for prognosis and treatment planning in HCC and other liver cancers.

Radiomics analysis and the integration of CNNs with CT and MRI images and clinical data have been developed to predict the prognosis of HCC patients. Machine learning and CNNs have exhibited a good accuracy in predicting patient survival following surgical treatment. In a bicentric study, a deep learning nomogram based on gadoteric acid MRI features was developed to predict early recurrence in 285 HCC patients post-hepatectomy. Extracting deep learning features using VGGNet-19 from contrast-enhanced MRI images, the deep learning nomogram, incorporating multiphase deep learning signatures, performed well on both the training (AUC: 0.949) and validation sets (AUC: 0.909). Independent predictors for early recurrence included microvascular invasion, tumor number, and the deep learning signature (105). Lv et al. introduced an AI -powered approach for predicting the 3-year recurrence of HCC using contrast-enhanced CT radiomic profiles. In a single-center retrospective cohort of 224 HCC patients, radiomic signatures from arterial and portal venous phases were utilized to establish three models: radiological model (RM), deep learning-based radiomics model (DLRM), and clinical & deep learning-based radiomics model (CDLRM). CDLRM, incorporating clinical factors and DLR features, demonstrated superior accuracy (AUC: 0.98 in training, 0.83 in testing) compared to DLRM and RM (106). In a proof-of-concept study for HCC patients initially eligible for liver transplant, machine learning models were developed using pretreatment clinical and MRI features to predict posttreatment recurrence. The study included 120 patients, and three machine learning models (clinical, imaging, combined) predicted recurrence with AUCs ranging from 0.60 to 0.86 across six timeframes. The imaging model outperformed the clinical model (mean AUC 0.76 vs. 0.68,  $p = 0.03$ ). Kaplan-Meier analysis demonstrated significant differences in recurrence risk prediction between low and high-risk groups for all three models (107). A retrospective study, involving 55 patients with stage 4 colon cancer and hepatic metastasis, explored the role of MRI-based measures of intra-tumor heterogeneity in predicting survival. Extracting a heterogeneity phenotype vector from 94 hepatic lesions, the study identified 22 texture features associated with patient survival. A random forest machine learning model, combining clinical variables with imaging-based features, improved survival prediction performance, yielding an area under the ROC curve of 0.94 compared to 0.83 with clinical variables alone (108).

By analyzing the complex patterns within imaging data, these approaches allow for a deeper understanding of tumor biology and patient-specific disease progression. The predictive capability of radiomics and AI models, as evidenced by their high accuracy in various studies, emphasizes the need for ongoing research to further validate and integrate these technologies into clinical practice.



## Pitfalls and technical limitations

Despite the promising results in radiomics research for liver cancers, a notable gap persists between numerous numerical data generated and their practical clinical application. These studies provide a myriad of quantitative metrics and predictive models, showcasing radiomics' potential in augmenting diagnostic and prognostic evaluations. However, translation of these findings into routine clinical practice remains uncertain. Challenges, including protocol variability and interobserver discrepancies, present significant obstacles in bridging the research-clinical gap. Noteworthy is the absence of clear guidelines on the integration of radiomic data into the real-world clinical decision-making. The intrinsic heterogeneity of liver tumors and the dynamic nature of cancer progression amplify the intricacies of developing robust and generalizable radiomic models. Additionally, challenges related to overfitting, model validation, and potential false correlations in high-dimensional data emphasize the need for rigorous methodology standardization. These technical challenges collectively underscore the substantial work required before radiomics can claim its role as a dependable and clinically impactful tool in liver cancer management.

While machine learning has shown remarkable promise in the radiologic assessment of primary and metastatic liver malignancies, it is not without its pitfalls. One significant challenge lies in the quality and quantity of training data. The performance of machine learning models heavily relies on the availability of diverse and representative datasets, and issues such as imbalances, biases, or insufficient samples can lead to suboptimal generalization and performance. Additionally, the interpretability of machine learning models in radiology remains a concern. The "black-box" nature of some sophisticated algorithms makes it challenging for clinicians to understand the rationale behind specific predictions, limiting their trust and acceptance. Another notable pitfall is the potential for overfitting, where a model may perform exceptionally well on the training data but fails to generalize effectively to new, unseen cases. Moreover, the dynamic nature of medical imaging and evolving standards in radiologic practices pose challenges in keeping machine learning models up-to-date and adaptable to changes in the field. Addressing these pitfalls is crucial to harness the full potential of machine learning in improving the accuracy and efficiency of radiologic assessments for liver malignancies.

## Future direction

It's important to acknowledge the gap between research advancements in radiomics and AI and their clinical implementation. This gap mainly exists because the low external validity of these technologies limits their adoption in routine clinical practice. The primary challenge for clinical translation is ensuring the generalizability of AI and radiomics models. There is a need for further clarification of true role of radiomics and machine learning tools in clinical applications. This involves external validation of machine learning models and the assessment of diagnostic performance for specific diseases using deep learning radiomics.

External validation, particularly through large multi-institutional datasets gathered over a longer period, is essential to confirm the models' generalizability. To enhance the clinical translation and applicability of radiomics studies, it is also crucial to address important issues such as access, cost-effectiveness analysis, and the promotion of open data practices. Generally, achieving sufficient clinical performance in training a CNN necessitates a large amount of training data. In the development of AI imaging models, the cost of annotation is a significant concern, and the future is expected to see a focus on acquiring substantial amounts of high-quality training data while simultaneously minimizing annotation costs. The ultimate goal is to leverage AI and radiomics in clinic for the precise classification and detection of liver tumors and to enable personalized treatment by accurately predicting treatment responses.

## Conclusion

In this review we identified several potentials of AI and radiomics in clinical decision-making in liver oncology imaging, including improving the precision of tumor detection, characterization and classification, enabling the prediction of treatment response, identifying patient-specific prognostic indicators for personalized therapy, and possibly reducing the reliance on invasive procedures like biopsies by non-invasively determining tumor genetics, immune phenotype and behavior.

## Author contributions

MH: Writing – original draft. DR: Writing – original draft. AK: Writing – original draft. YV: Writing – original draft, Writing – review & editing. AB: 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. Publication costs are covered by Northwestern medicine.

## 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

- Galun D, Mijac D, Filipovic A, Bogdanovic A, Zivanovic M, Masulovic D. Precision medicine for hepatocellular carcinoma: Clinical perspective. *J Personalized Med.* (2022) 12:149. doi: 10.3390/jpm12020149
- Johnson P, Zhou Q, Dao DY, Lo YD. Circulating biomarkers in the diagnosis and management of hepatocellular carcinoma. *Nat Rev Gastroenterol Hepatology.* (2022) 19:670–81. doi: 10.1038/s41575-022-00620-y
- Tomaszewski MR, Gillies RJ. The biological meaning of radiomic features. *Radiology.* (2021) 298:505–16. doi: 10.1148/radiol.202102553
- Aerts HJ. The potential of radiomic-based phenotyping in precision medicine: a review. *JAMA Oncol.* (2016) 2:1636–42. doi: 10.1001/jamaoncol.2016.2631
- 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
- 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
- Mohagheghi S, Foruzan AH. Developing an explainable deep learning boundary correction method by incorporating cascaded x-Dim models to improve segmentation defects in liver CT images. *Comput Biol Med.* (2022) 140:105106. doi: 10.1016/j.compbimed.2021.105106
- Said D, Carbonell G, Stocker D, Hectors S, Vietti-Viola N, Bane O, et al. Semiautomated segmentation of hepatocellular carcinoma tumors with MRI using convolutional neural networks. *Eur Radiology.* (2023), 1–13. doi: 10.1007/s00330-023-09613-0
- Anderson BM, Rigaud B, Lin Y-M, Jones AK, Kang HC, Odisio BC, et al. Automated segmentation of colorectal liver metastasis and liver ablation on contrast-enhanced CT images. *Front Oncol.* (2022) 12:886517. doi: 10.3389/fonc.2022.886517
- Bousabarrah K, Letzen B, Tefera J, Savic L, Schobert I, Schlachter T, et al. Automated detection and delineation of hepatocellular carcinoma on multiphase contrast-enhanced MRI using deep learning. *Abdominal Radiology.* (2021) 46:216–25. doi: 10.1007/s00261-020-02604-5
- Massoptier L, Casciaro S. A new fully automatic and robust algorithm for fast segmentation of liver tissue and tumors from CT scans. *Eur radiology.* (2008) 18:1658–65. doi: 10.1007/s00330-008-0924-y
- Park J, Bae JS, Kim J-M, Witanto JN, Park SJ, Lee JM. Development of a deep-learning model for classification of LI-RADS major features by using subtraction images of MRI: a preliminary study. *Abdominal Radiology.* (2023), 1–10. doi: 10.1007/s00261-023-03962-6
- Homayounieh F, Singh R, Nitiwarangkul C, Lades F, Schmidt B, Sedlmair M, et al. Semiautomated segmentation and radiomics for dual-energy CT: a pilot study to differentiate benign and Malignant hepatic lesions. *Am J Roentgenology.* (2020) 215:398–405. doi: 10.2214/AJR.19.22164
- Srivastava S, Koay EJ, Borowsky AD, De Marzo AM, Ghosh S, Wagner PD, et al. Cancer overdiagnosis: a biological challenge and clinical dilemma. *Nat Rev Cancer.* (2019) 19:349–58. doi: 10.1038/s41568-019-0142-8
- Ogasawara S, Akiba J, Nakayama M, Nakashima O, Torimura T, Yano H. Epithelial cell adhesion molecule-positive human hepatic neoplastic cells: development of combined hepatocellular-cholangiocarcinoma in mice. *J Gastroenterol Hepatology.* (2015) 30:413–20. doi: 10.1111/jgh.12692
- Jung DH, Hwang S, Hong SM, Chung YK, Song GW, Lee YJ, et al. Post-resection prognosis of combined hepatocellular carcinoma-cholangiocarcinoma according to the 2010 WHO classification. *World J surgery.* (2017) 41:1347–57. doi: 10.1007/s00268-016-3837-y
- Lee J-H, Chung GE, Yu SJ, Hwang SY, Kim JS, Kim HY, et al. Long-term prognosis of combined hepatocellular and cholangiocarcinoma after curative resection comparison with hepatocellular carcinoma and cholangiocarcinoma. *J Clin gastroenterology.* (2011) 45:69–75. doi: 10.1097/MCG.0b013e3181ce5dfa
- Song P, Midorikawa Y, Nakayama H, Higaki T, Moriguchi M, Aramaki O, et al. Patients' prognosis of intrahepatic cholangiocarcinoma and combined hepatocellular-cholangiocarcinoma after resection. *Cancer Med.* (2019) 8:5862–71. doi: 10.1002/cam4.2495
- Zhao X, Liang P, Yong L, Jia Y, Gao J. Radiomics study for differentiating focal hepatic lesions based on unenhanced CT images. *Front Oncol.* (2022) 12:650797. doi: 10.3389/fonc.2022.650797
- Zhou Y, Zhou G, Zhang J, Xu C, Zhu F, Xu P. DCE-MRI based radiomics nomogram for preoperatively differentiating combined hepatocellular-cholangiocarcinoma from mass-forming intrahepatic cholangiocarcinoma. *Eur Radiology.* (2022) 32:5004–15. doi: 10.1007/s00330-022-08548-2
- Xue B, Wu S, Zhang M, Hong J, Liu B, Xu N, et al. A radiomic-based model of different contrast-enhanced CT phase for differentiate intrahepatic cholangiocarcinoma from inflammatory mass with hepatolithiasis. *Abdominal Radiology.* (2021) 46:3835–44. doi: 10.1007/s00261-021-03027-6
- Bae H, Lee H, Kim S, Han K, Rhee H, D-k K, et al. Radiomics analysis of contrast-enhanced CT for classification of hepatic focal lesions in colorectal cancer patients: its limitations compared to radiologists. *Eur Radiology.* (2021) 31:8786–96. doi: 10.1007/s00330-021-07877-y
- Li Z-f, Kang L-q, Liu F-h, Zhao M, Guo S-y, Lu S, et al. Radiomics based on preoperative rectal cancer MRI to predict the metachronous liver metastasis. *Abdominal Radiology.* (2023) 48:833–43. doi: 10.1007/s00261-022-03773-1
- Mokrane F-Z, Lu L, Vavasseur A, Otal P, Peron J-M, Luk L, et al. Radiomics machine-learning signature for diagnosis of hepatocellular carcinoma in cirrhotic patients with indeterminate liver nodules. *Eur radiology.* (2020) 30:558–70. doi: 10.1007/s00330-019-06347-w
- Yuan Z, Shu Z, Peng J, Wang W, Hou J, Han L, et al. Prediction of postoperative liver metastasis in pancreatic ductal adenocarcinoma based on multiparametric magnetic resonance radiomics combined with serological markers: a cohort study of machine learning. *Abdominal Radiology.* (2023), 1–14. doi: 10.1007/s00261-023-04047-0
- Liu X, Khalvati F, Namdar K, Fischer S, Lewis S, Taouli B, et al. Can machine learning radiomics provide pre-operative differentiation of combined hepatocellular cholangiocarcinoma from hepatocellular carcinoma and cholangiocarcinoma to inform optimal treatment planning? *Eur Radiol.* (2021) 31:244–55. doi: 10.1007/s00330-020-07119-7
- Chen X, Chen Y, Chen H, Zhu J, Huang R, Xie J, et al. Machine learning based on gadoteric acid-enhanced MRI for differentiating atypical intrahepatic mass-forming cholangiocarcinoma from poorly differentiated hepatocellular carcinoma. *Abdominal Radiology.* (2023), 1–12. doi: 10.1007/s00261-023-03870-9
- Taghavi M, Trebeschi S, Simões R, Meek DB, Beckers RC, Lambregts DM, et al. Machine learning-based analysis of CT radiomics model for prediction of colorectal metachronous liver metastases. *Abdominal Radiology.* (2021) 46:249–56. doi: 10.1007/s00261-020-02624-1
- Nakai H, Sakamoto R, Kakigi T, Coeur C, Isoda H, Nakamoto Y. Artificial intelligence-powered software detected more than half of the liver metastases overlooked by radiologists on contrast-enhanced CT. *Eur J Radiology.* (2023) 163:110823. doi: 10.1016/j.ejrad.2023.110823
- Ling Y, Ying S, Xu L, Peng Z, Mao X, Chen Z, et al. Automatic volumetric diagnosis of hepatocellular carcinoma based on four-phase CT scans with minimum extra information. *Front Oncol.* (2022) 12:960178. doi: 10.3389/fonc.2022.960178
- Oestmann PM, Wang CJ, Savic LJ, Hamm CA, Stark S, Schobert I, et al. Deep learning-assisted differentiation of pathologically proven atypical and typical hepatocellular carcinoma (HCC) versus non-HCC on contrast-enhanced MRI of the liver. *Eur radiology.* (2021) 31:4981–90. doi: 10.1007/s00330-020-07559-1
- Kim DW, Lee G, Kim SY, Ahn G, Lee J-G, Lee SS, et al. Deep learning-based algorithm to detect primary hepatic Malignancy in multiphase CT of patients at high risk for HCC. *Eur Radiology.* (2021) 31:7047–57. doi: 10.1007/s00330-021-07803-2
- Xu Y, Zhou C, He X, Song R, Liu Y, Zhang H, et al. Deep learning-assisted LI-RADS grading and distinguishing hepatocellular carcinoma (HCC) from non-HCC based on multiphase CT: a two-center study. *Eur Radiology.* (2023), 1–10. doi: 10.1007/s00330-023-09857-w
- Yamashita R, Mittendorf A, Zhu Z, Fowler KJ, Santillan CS, Sirlin CB, et al. Deep convolutional neural network applied to the liver imaging reporting and data system (LI-RADS) version 2014 category classification: a pilot study. *Abdominal Radiology.* (2020) 45:24–35. doi: 10.1007/s00261-019-02306-7
- Wang H, Wu MC, Cong WM. Microvascular invasion predicts a poor prognosis of solitary hepatocellular carcinoma up to 2 cm based on propensity score matching analysis. *Hepatol Res.* (2019) 49:344–54. doi: 10.1111/hepr.13241
- Zhou Y, Zhou G, Zhang J, Xu C, Wang X, Xu P. Radiomics signature on dynamic contrast-enhanced MR images: a potential imaging biomarker for prediction of microvascular invasion in mass-forming intrahepatic cholangiocarcinoma. *Eur Radiology.* (2021) 31:6846–55. doi: 10.1007/s00330-021-07793-1
- Zheng R, Zhang X, Liu B, Zhang Y, Shen H, Xie X, et al. Comparison of non-radiomics imaging features and radiomics models based on contrast-enhanced ultrasound and Gd-EOB-DTPA-enhanced MRI for predicting microvascular invasion in hepatocellular carcinoma within 5 cm. *Eur Radiology.* (2023), 1–11. doi: 10.1007/s00330-023-09789-5
- Xia T-y, Zhou Z-h, Meng X-p, Zha J-h, Yu Q, Wang W-l, et al. Predicting microvascular invasion in hepatocellular carcinoma using CT-based radiomics model. *Radiology.* (2023) 307:e222729. doi: 10.1148/radiol.222729
- Zhang K, Zhang L, Li W-C, Xie S-S, Cui Y-Z, Lin L-Y, et al. Radiomics nomogram for the prediction of microvascular invasion of HCC and patients' benefit from postoperative adjuvant TACE: a multi-center study. *Eur Radiol.* (2023), 1–12. doi: 10.1007/s00330-023-09824-5
- Chong H-H, Yang L, Sheng R-F, Yu Y-L, Wu D-J, Rao S-X, et al. Multi-scale and multi-parametric radiomics of gadoteric acid-enhanced MRI predicts microvascular invasion and outcome in patients with solitary hepatocellular carcinoma ≤ 5 cm. *Eur Radiology.* (2021) 31:4824–38. doi: 10.1007/s00330-020-07601-2
- Ma X, Wei J, Gu D, Zhu Y, Feng B, Liang M, et al. Preoperative radiomics nomogram for microvascular invasion prediction in hepatocellular carcinoma using contrast-enhanced CT. *Eur radiology.* (2019) 29:3595–605. doi: 10.1007/s00330-018-5985-y

42. Feng S-T, Jia Y, Liao B, Huang B, Zhou Q, Li X, et al. Preoperative prediction of microvascular invasion in hepatocellular cancer: a radiomics model using Gd-EOB-DTPA-enhanced MRI. *Eur radiology*. (2019) 29:4648–59. doi: 10.1007/s00330-018-5935-8
43. Chen Y, Xia Y, Tolat PP, Long L, Jiang Z, Huang Z, et al. Comparison of conventional gadoxetate disodium-enhanced MRI features and radiomics signatures with machine learning for diagnosing microvascular invasion. *Am J Roentgenology*. (2021) 216:1510–20. doi: 10.2214/AJR.20.23255
44. Liu B, Zeng Q, Huang J, Zhang J, Zheng Z, Liao Y, et al. IVIM using convolutional neural networks predicts microvascular invasion in HCC. *Eur Radiology*. (2022) 32:7185–95. doi: 10.1007/s00330-022-08927-9
45. Lewin M, Laurent-Bellue A, Desterke C, Radu A, Feghali JA, Farah J, et al. Evaluation of perfusion CT and dual-energy CT for predicting microvascular invasion of hepatocellular carcinoma. *Abdominal Radiology*. (2022) 47:2115–27. doi: 10.1007/s00261-022-03511-7
46. Li X, Qi Z, Du H, Geng Z, Li Z, Qin S, et al. Deep convolutional neural network for preoperative prediction of microvascular invasion and clinical outcomes in patients with HCCs. *Eur Radiology*. (2021), 1–12. doi: 10.1007/s00330-021-08198-w
47. Wang T, Li Z, Yu H, Duan C, Feng W, Chang L, et al. Prediction of microvascular invasion in hepatocellular carcinoma based on preoperative Gd-EOB-DTPA-enhanced MRI: Comparison of predictive performance among 2D, 2D-expansion and 3D deep learning models. *Front Oncol*. (2023) 13:987781. doi: 10.3389/fonc.2023.987781
48. Han YE, Cho Y, Kim MJ, Park BJ, Sung DJ, Han NY, et al. Hepatocellular carcinoma pathologic grade prediction using radiomics and machine learning models of gadoxetic acid-enhanced MRI: a two-center study. *Abdominal Radiology*. (2023) 48:244–56. doi: 10.1007/s00261-022-03679-y
49. Feng Z, Li H, Liu Q, Duan J, Zhou W, Yu X, et al. CT radiomics to predict macrotrabecular-massive subtype and immune status in hepatocellular carcinoma. *Radiology*. (2022) 307:e21291. doi: 10.1148/radiol.221291
50. Wu M, Tan H, Gao F, Hai J, Ning P, Chen J, et al. Predicting the grade of hepatocellular carcinoma based on non-contrast-enhanced MRI radiomics signature. *Eur radiology*. (2019) 29:2802–11. doi: 10.1007/s00330-018-5787-2
51. Zhu Y, Weng S, Li Y, Yan C, Ye R, Wen L, et al. A radiomics nomogram based on contrast-enhanced MRI for preoperative prediction of macrotrabecular-massive hepatocellular carcinoma. *Abdominal Radiology*. (2021) 46:3139–48. doi: 10.1007/s00261-021-02989-x
52. Mao B, Zhang L, Ning P, Ding F, Wu F, Lu G, et al. Preoperative prediction for pathological grade of hepatocellular carcinoma via machine learning-based radiomics. *Eur Radiology*. (2020) 30:6924–32. doi: 10.1007/s00330-020-07056-5
53. Li M, Fan Y, You H, Li C, Luo M, Zhou J, et al. Dual-energy CT deep learning radiomics to predict macrotrabecular-massive hepatocellular carcinoma. *Radiology*. (2023) 308:e230255. doi: 10.1148/radiol.230255
54. Torkian P, Haghsomari M, Farsad K, Wallace S, Goltzarian J, Young SJ. Cancer immunology: impact of radioembolization of hepatocellular carcinoma on immune response modulation. *Am J Roentgenology*. (2023) 220:863–72. doi: 10.2214/AJR.22.28800
55. Sia D, Jiao Y, Martinez-Quetglas I, Kuchuk O, Villacorta-Martin C, de Moura MC, et al. Identification of an immune-specific class of hepatocellular carcinoma, based on molecular features. *Gastroenterology*. (2017) 153:812–26. doi: 10.1053/j.gastro.2017.06.007
56. Fan T, Li S, Li K, Xu J, Zhao S, Li J, et al. A potential prognostic marker for recognizing VEGF-positive hepatocellular carcinoma based on magnetic resonance radiomics signature. *Front Oncol*. (2022) 12:857715. doi: 10.3389/fonc.2022.857715
57. Zheng J, Du P-Z, Yang C, Tao Y-Y, Li L, Li Z-M, et al. DCE-MRI-based radiomics in predicting angiopoietin-2 expression in hepatocellular carcinoma. *Abdominal Radiology*. (2023) 48:3343–52. doi: 10.1007/s00261-023-04007-8
58. Chen X, Tang Y, Wu D, Li R, Lin Z, Zhou X, et al. From imaging to clinical outcome: dual-region CT radiomics predicting FOXM1 expression and prognosis in hepatocellular carcinoma. *Front Oncol*. (2023) 13:1278467. doi: 10.3389/fonc.2023.1278467
59. Wu C, Chen J, Fan Y, Zhao M, He X, Wei Y, et al. Nomogram based on CT radiomics features combined with clinical factors to predict Ki-67 expression in hepatocellular carcinoma. *Front Oncol*. (2022) 12:943942. doi: 10.3389/fonc.2022.943942
60. Yan C, Han Z, Chen X, Gao L, Ye R, Li Y. Diffusion-weighted imaging as a quantitative imaging biomarker for predicting proliferation rate in hepatocellular carcinoma: developing a radiomics nomogram. *J Comput Assisted Tomography*. (2023) 10:1097. doi: 10.1097/RCT.0000000000001448
61. Zeng F, Dai H, Li X, Guo L, Jia N, Yang J, et al. Preoperative radiomics model using gadobenate dimeglumine-enhanced magnetic resonance imaging for predicting  $\beta$ -catenin mutation in patients with hepatocellular carcinoma: A retrospective study. *Front Oncol*. (2022) 12:916126. doi: 10.3389/fonc.2022.916126
62. Li P, Liang Y, Zeng B, Yang G, Zhu C, Zhao K, et al. Preoperative prediction of intra-tumoral tertiary lymphoid structures based on CT in hepatocellular cancer. *Eur J Radiology*. (2022) 151:110309. doi: 10.1016/j.ejrad.2022.110309
63. Zhang L, Zhou H, Zhang X, Ding Z, Xu J. A radiomics nomogram for predicting cytokeratin 19-positive hepatocellular carcinoma: a two-center study. *Front Oncol*. (2023) 13:1174069. doi: 10.3389/fonc.2023.1174069
64. Wang W, Gu D, Wei J, Ding Y, Yang L, Zhu K, et al. A radiomics-based biomarker for cytokeratin 19 status of hepatocellular carcinoma with gadoxetic acid-enhanced MRI. *Eur radiology*. (2020) 30:3004–14. doi: 10.1007/s00330-019-06585-y
65. Zhang N, Wu M, Zhou Y, Yu C, Shi D, Wang C, et al. Radiomics nomogram for prediction of glypican-3 positive hepatocellular carcinoma based on hepatobiliary phase imaging. *Front Oncol*. (2023) 13:1209814. doi: 10.3389/fonc.2023.1209814
66. Hectors SJ, Lewis S, Besa C, King MJ, Said D, Putra J, et al. MRI radiomics features predict immuno-oncological characteristics of hepatocellular carcinoma. *Eur radiology*. (2020) 30:3759–69. doi: 10.1007/s00330-020-06675-2
67. Chen S, Feng S, Wei J, Liu F, Li B, Li X, et al. Pretreatment prediction of immunoscore in hepatocellular cancer: a radiomics-based clinical model based on Gd-EOB-DTPA-enhanced MRI imaging. *Eur Radiology*. (2019) 29:4177–87. doi: 10.1007/s00330-018-5986-x
68. Xie T, Wei Y, Xu L, Li Q, Che F, Xu Q, et al. Self-supervised contrastive learning using CT images for PD-1/PD-L1 expression prediction in hepatocellular carcinoma. *Front Oncol*. (2023) 13:1103521. doi: 10.3389/fonc.2023.1103521
69. Inchingolo R, Posa A, Mariappan M, Spiliopoulos S. Locoregional treatments for hepatocellular carcinoma: Current evidence and future directions. *World J gastroenterology*. (2019) 25:4614. doi: 10.3748/wjg.v25.i32.4614
70. Bruix J, Reig M, Sherman M. Evidence-based diagnosis, staging, and treatment of patients with hepatocellular carcinoma. *Gastroenterology*. (2016) 150:835–53. doi: 10.1053/j.gastro.2015.12.041
71. Mendiratta-Lala M, Masch W, Owen D, Aslam A, Maurino C, Devasia T, et al. Natural history of hepatocellular carcinoma after stereotactic body radiation therapy. *Abdominal Radiology*. (2020) 45:3698–708. doi: 10.1007/s00261-020-02532-4
72. Zhang X, Wang C, Zheng D, Liao Y, Wang X, Huang Z, et al. Radiomics nomogram based on multi-parametric magnetic resonance imaging for predicting early recurrence in small hepatocellular carcinoma after radiofrequency ablation. *Front Oncol*. (2022) 12:1013770. doi: 10.3389/fonc.2022.1013770
73. Fan XL, Wang YH, Chen YH, Chen BX, Cai JN, Yang JS, et al. Computed tomography texture analysis combined with preoperative clinical factors serve as a predictor of early efficacy of transcatheter arterial chemoembolization in hepatocellular carcinoma. *Abdominal Radiology*. (2023) 48:2008–18. doi: 10.1007/s00261-023-03868-3
74. Lv X, Chen M, Kong C, Shu G, Meng M, Ye W, et al. Construction of a novel radiomics nomogram for the prediction of aggressive intrasegmental recurrence of HCC after radiofrequency ablation. *Eur J Radiology*. (2021) 144:109955. doi: 10.1016/j.ejrad.2021.109955
75. Kong C, Zhao Z, Chen W, Lv X, Shu G, Ye M, et al. Prediction of tumor response via a pretreatment MRI radiomics-based nomogram in HCC treated with TACE. *Eur radiology*. (2021) 31:7500–11. doi: 10.1007/s00330-021-07910-0
76. Sheen H, Kim JS, Lee JK, Choi SY, Baek SY, Kim JY. A radiomics nomogram for predicting transcatheter arterial chemoembolization refractoriness of hepatocellular carcinoma without extrahepatic metastasis or macrovascular invasion. *Abdominal Radiology*. (2021) 46:2839–49. doi: 10.1007/s00261-020-02884-x
77. Kim J, Choi SJ, Lee S-H, Lee HY, Park H. Predicting survival using pretreatment CT for patients with hepatocellular carcinoma treated with transarterial chemoembolization: comparison of models using radiomics. *Am J Roentgenology*. (2018), 1026–34. doi: 10.2214/AJR.18.19507
78. Kuang Y, Li R, Jia P, Ye W, Zhou R, Zhu R, et al. MRI-Based Radiomics: Nomograms predicting the short-term response after transcatheter arterial chemoembolization (TACE) in hepatocellular carcinoma patients with diameter less than 5 cm. *Abdominal Radiology*. (2021) 46:3772–89. doi: 10.1007/s00261-021-02992-2
79. Wang D, Zhang L, Sun Z, Jiang H, Zhang J. A radiomics signature associated with underlying gene expression pattern for the prediction of prognosis and treatment response in hepatocellular carcinoma. *Eur J Radiology*. (2023) 167:111086. doi: 10.1016/j.ejrad.2023.111086
80. Staal F, Taghavi M, van der Reijnd D, Gomez F, Imani F, Klompenhouwer E, et al. Predicting local tumour progression after ablation for colorectal liver metastases: CT-based radiomics of the ablation zone. *Eur J Radiology*. (2021) 141:109773. doi: 10.1016/j.ejrad.2021.109773
81. Mosconi C, Cucchetti A, Bruno A, Cappelli A, Bargellini I, De Benedittis C, et al. Radiomics of cholangiocarcinoma on pretreatment CT can identify patients who would best respond to radioembolisation. *Eur Radiology*. (2020) 30:4534–44. doi: 10.1007/s00330-020-06795-9
82. Wang H, Liu Y, Xu N, Sun Y, Fu S, Wu Y, et al. Development and validation of a deep learning model for survival prognosis of transcatheter arterial chemoembolization in patients with intermediate-stage hepatocellular carcinoma. *Eur J Radiology*. (2022) 156:110527. doi: 10.1016/j.ejrad.2022.110527
83. Peng J, Kang S, Ning Z, Deng H, Shen J, Xu Y, et al. Residual convolutional neural network for predicting response of transcatheter arterial chemoembolization in hepatocellular carcinoma from CT imaging. *Eur radiology*. (2020) 30:413–24. doi: 10.1007/s00330-019-06318-1
84. Zhang L, Xia W, Yan Z-P, Sun J-H, Zhong B-Y, Hou Z-H, et al. Deep learning predicts overall survival of patients with unresectable hepatocellular carcinoma treated by transarterial chemoembolization plus sorafenib. *Front Oncol*. (2020) 10:593292. doi: 10.3389/fonc.2020.593292
85. Xu Z, An C, Shi F, Ren H, Li Y, Chen S, et al. Automatic prediction of hepatic arterial infusion chemotherapy response in advanced hepatocellular carcinoma with

deep learning radiomic nomogram. *Eur Radiology*. (2023), 1–14. doi: 10.1007/s00330-023-09953-x

86. Chan AW, Zhong J, Berhane S, Toyoda H, Cucchetti A, Shi K, et al. Development of pre and post-operative models to predict early recurrence of hepatocellular carcinoma after surgical resection. *J hepatology*. (2018) 69:1284–93. doi: 10.1016/j.jhep.2018.08.027

87. Portolani N, Coniglio A, Ghidoni S, Giovanelli M, Benetti A, Tiberio GAM, et al. Early and late recurrence after liver resection for hepatocellular carcinoma: prognostic and therapeutic implications. *Ann surgery*. (2006) 243:229. doi: 10.1097/01.sla.0000197706.21803.a1

88. Wang L, Ma X, Feng B, Wang S, Liang M, Li D, et al. Multi-sequence MR-based radiomics signature for predicting early recurrence in solitary hepatocellular carcinoma ≤ 5 cm. *Front Oncol*. (2022) 12:899404. doi: 10.3389/fonc.2022.899404

89. Li C, Wang Q, Zou M, Cai P, Li X, Feng K, et al. A radiomics model based on preoperative gadolinic acid-enhanced magnetic resonance imaging for predicting post-hepatectomy liver failure in patients with hepatocellular carcinoma. *Front Oncol*. (2023) 13:1164739. doi: 10.3389/fonc.2023.1164739

90. Wang L, Feng B, Wang S, Hu J, Liang M, Li D, et al. Diagnostic value of whole-tumor apparent diffusion coefficient map radiomics analysis in predicting early recurrence of solitary hepatocellular carcinoma ≤ 5 cm. *Abdominal Radiology*. (2022) 47:3290–300. doi: 10.1007/s00261-022-03582-6

91. Yu Y, Fan Y, Wang X, Zhu M, Hu M, Shi C, et al. Gd-EOB-DTPA-enhanced MRI radiomics to predict vessels encapsulating tumor clusters (VETC) and patient prognosis in hepatocellular carcinoma. *Eur Radiol*. (2022), 1–12. doi: 10.1007/s00330-021-08250-9

92. Ji G-W, Zhu F-P, Xu Q, Wang K, Wu M-Y, Tang W-W, et al. Radiomic features at contrast-enhanced CT predict recurrence in early stage hepatocellular carcinoma: a multi-institutional study. *Radiology*. (2020) 294:568–79. doi: 10.1148/radiol.2020191470

93. Guo D, Gu D, Wang H, Wei J, Wang Z, Hao X, et al. Radiomics analysis enables recurrence prediction for hepatocellular carcinoma after liver transplantation. *Eur J radiology*. (2019) 117:33–40. doi: 10.1016/j.ejrad.2019.05.010

94. Ning P, Gao F, Hai J, Wu M, Chen J, Zhu S, et al. Application of CT radiomics in prediction of early recurrence in hepatocellular carcinoma. *Abdominal Radiology*. (2020) 45:64–72. doi: 10.1007/s00261-019-02198-7

95. Wang X-H, Long L-H, Cui Y, Jia AY, Zhu X-G, Wang H-Z, et al. MRI-based radiomics model for preoperative prediction of 5-year survival in patients with hepatocellular carcinoma. *Br J cancer*. (2020) 122:978–85. doi: 10.1038/s41416-019-0706-0

96. Zhou Y, He L, Huang Y, Chen S, Wu P, Ye W, et al. CT-based radiomics signature: a potential biomarker for preoperative prediction of early recurrence in hepatocellular carcinoma. *Abdominal radiology*. (2017) 42:1695–704. doi: 10.1007/s00261-017-1072-0

97. Zhang L, Hu J, Hou J, Jiang X, Guo L, Tian L. Radiomics-based model using gadolinic acid disodium-enhanced MR images: associations with recurrence-free

survival of patients with hepatocellular carcinoma treated by surgical resection. *Abdominal Radiology*. (2021) 46:3845–54. doi: 10.1007/s00261-021-03034-7

98. Park HJ, Park B, Park SY, Choi SH, Rhee H, Park JH, et al. Preoperative prediction of postsurgical outcomes in mass-forming intrahepatic cholangiocarcinoma based on clinical, radiologic, and radiomics features. *Eur radiology*. (2021) 31:8638–48. doi: 10.1007/s00330-021-07926-6

99. Ji G-W, Zhu F-P, Zhang Y-D, Liu X-S, Wu F-Y, Wang K, et al. A radiomics approach to predict lymph node metastasis and clinical outcome of intrahepatic cholangiocarcinoma. *Eur radiology*. (2019) 29:3725–35. doi: 10.1007/s00330-019-06142-7

100. Silva M, Maddalo M, Leoni E, Giuliotti S, Milanese G, Ghetti C, et al. Integrated prognostication of intrahepatic cholangiocarcinoma by contrast-enhanced computed tomography: the adjunct yield of radiomics. *Abdominal Radiology*. (2021) 46:4689–700. doi: 10.1007/s00261-021-03183-9

101. Chu H, Liu Z, Liang W, Zhou Q, Zhang Y, Lei K, et al. Radiomics using CT images for preoperative prediction of futile resection in intrahepatic cholangiocarcinoma. *Eur Radiology*. (2021) 31:2368–76. doi: 10.1007/s00330-020-07250-5

102. Wei S, Han Y, Zeng H, Ye S, Cheng J, Chai F, et al. Radiomics diagnosed histopathological growth pattern in prediction of response and 1-year progression free survival for colorectal liver metastases patients treated with bevacizumab containing chemotherapy. *Eur J Radiology*. (2021) 142:109863. doi: 10.1016/j.ejrad.2021.109863

103. Muehlberg A, Holch JW, Heinemann V, Huber T, Moltz J, Maurus S, et al. The relevance of CT-based geometric and radiomics analysis of whole liver tumor burden to predict survival of patients with metastatic colorectal cancer. *Eur radiology*. (2021) 31:834–46. doi: 10.1007/s00330-020-07192-y

104. Rahmim A, Bak-Fredslund KP, Ashrafinia S, Lu L, Schmidtlein CR, Subramaniam RM, et al. Prognostic modeling for patients with colorectal liver metastases incorporating FDG PET radiomic features. *Eur J radiology*. (2019) 113:101–9. doi: 10.1016/j.ejrad.2019.02.006

105. Yan M, Zhang X, Zhang B, Geng Z, Xie C, Yang W, et al. Deep learning nomogram based on Gd-EOB-DTPA MRI for predicting early recurrence in hepatocellular carcinoma after hepatectomy. *Eur Radiology*. (2023), 1–13. doi: 10.1007/s00330-023-09419-0

106. Lv C, He N, Yang JJ, Xiao JJ, Zhang Y, Du J, et al. Prediction of 3-year recurrence rate of hepatocellular carcinoma after resection based on contrast-enhanced CT: A single-centre study. *Br J Radiol*. (2023) 96:20220702. doi: 10.1259/bjr.20220702

107. Iseke S, Zeevi T, Kucukkaya AS, Raju R, Gross M, Haider SP, et al. Machine learning models for prediction of posttreatment recurrence in early-stage hepatocellular carcinoma using pretreatment clinical and MRI features: a proof-of-concept study. *AJR Am J roentgenology*. (2023) 220:245. doi: 10.2214/AJR.22.28077

108. Daye D, Tabari A, Kim H, Chang K, Kamran SC, Hong TS, et al. Quantitative tumor heterogeneity MRI profiling improves machine learning-based prognostication in patients with metastatic colon cancer. *Eur Radiology*. (2021) 31:5759–67. doi: 10.1007/s00330-020-07673-0



## OPEN ACCESS

## EDITED BY

Poonam Yadav,  
Northwestern University, United States

## REVIEWED BY

Aurel Ottlakan,  
University of Szeged, Hungary  
Jaya Lakshmi Thangaraj,  
University of California, San Diego,  
United States

## \*CORRESPONDENCE

Linlin Shen

✉ llshen@szu.edu.cn

Yusen Zhang

✉ 395041386@qq.com

RECEIVED 26 December 2023

ACCEPTED 08 May 2024

PUBLISHED 23 May 2024

## CITATION

Xie H, Zhang Y, Dong L, Lv H, Li X, Zhao C,  
Tian Y, Xie L, Wu W, Yang Q, Liu L, Sun D,  
Qiu L, Shen L and Zhang Y (2024) Deep  
learning driven diagnosis of malignant soft  
tissue tumors based on dual-modal  
ultrasound images and clinical indexes.  
*Front. Oncol.* 14:1361694.  
doi: 10.3389/fonc.2024.1361694

## COPYRIGHT

© 2024 Xie, Zhang, Dong, Lv, Li, Zhao, Tian,  
Xie, Wu, Yang, Liu, Sun, Qiu, Shen and Zhang.  
This is an open-access article distributed under  
the terms of the [Creative Commons Attribution  
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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 driven diagnosis of malignant soft tissue tumors based on dual-modal ultrasound images and clinical indexes

Haiqin Xie<sup>1</sup>, Yudi Zhang<sup>2</sup>, Licong Dong<sup>1</sup>, Heng Lv<sup>1</sup>, Xuechen Li<sup>3</sup>,  
Chenyang Zhao<sup>1</sup>, Yun Tian<sup>1</sup>, Lu Xie<sup>1</sup>, Wangjie Wu<sup>1</sup>, Qi Yang<sup>1</sup>,  
Li Liu<sup>1</sup>, Desheng Sun<sup>1</sup>, Li Qiu<sup>4</sup>, Linlin Shen<sup>2\*</sup> and Yusen Zhang<sup>1\*</sup>

<sup>1</sup>Shenzhen Hospital, Peking University, Shenzhen, China, <sup>2</sup>College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, Guangdong, China, <sup>3</sup>National Engineering Laboratory for Big Data System Computing Technology, Shenzhen University, Shenzhen, China, <sup>4</sup>West China Hospital, Sichuan University, Chengdu, Sichuan, China

**Background:** Soft tissue tumors (STTs) are benign or malignant superficial neoplasms arising from soft tissues throughout the body with versatile pathological types. Although Ultrasonography (US) is one of the most common imaging tools to diagnose malignant STTs, it still has several drawbacks in STT diagnosis that need improving.

**Objectives:** The study aims to establish this deep learning (DL) driven Artificial intelligence (AI) system for predicting malignant STTs based on US images and clinical indexes of the patients.

**Methods:** We retrospectively enrolled 271 malignant and 462 benign masses to build the AI system using 5-fold validation. A prospective dataset of 44 malignant masses and 101 benign masses was used to validate the accuracy of system. A multi-data fusion convolutional neural network, named ultrasound clinical soft tissue tumor net (UC-STTNet), was developed to combine gray scale and color Doppler US images and clinic features for malignant STTs diagnosis. Six radiologists (R1-R6) with three experience levels were invited for reader study.

**Results:** The AI system achieved an area under receiver operating curve (AUC) value of 0.89 in the retrospective dataset. The diagnostic performance of the AI system was higher than that of one of the senior radiologists (AUC of AI vs R2: 0.89 vs. 0.84,  $p=0.022$ ) and all of the intermediate and junior radiologists (AUC of AI vs R3, R4, R5, R6: 0.89 vs 0.75, 0.81, 0.80, 0.63;  $p < 0.01$ ). The AI system also



achieved an AUC of 0.85 in the prospective dataset. With the assistance of the system, the diagnostic performances and inter-observer agreement of the radiologists was improved (AUC of R3, R5, R6: 0.75 to 0.83, 0.80 to 0.85, 0.63 to 0.69;  $p < 0.01$ ).

**Conclusion:** The AI system could be a useful tool in diagnosing malignant STTs, and could also help radiologists improve diagnostic performance.

#### KEYWORDS

deep learning, artificial intelligence, ultrasound, soft tissue tumor, malignant tumor

## Highlights

- The deep-learning driven system has a high accuracy in diagnosing malignant soft tissue tumors.
- The deep-learning system showed superior performance than junior radiologists.
- The system is a useful tool for radiologists in discerning malignant soft tissue tumors.

## Introduction

Soft tissue tumors (STTs) are a group of superficial neoplasms with heterogeneous clinical manifestations and diverse pathological types. The ratio of benign to malignant is close to 100:1 (1). Soft tissue sarcomas are the most common malignant STTs, accounting for only 1% of all adult cancers (2). Despite the rarity of malignant STTs compared with other malignant entities, their hazards cannot be ignored due to the substantial mortality and morbidity (3, 4). The overall five-year survival rate of malignant STTs is about 50% (1). They also present high metastasis and postoperative recurrence rates, up to 39% for soft tissue sarcomas (5). Soft tissue sarcoma is one of the leading causes of death for young adults, particularly for certain subtypes (6). As a result, accurate diagnosis and timely treatment for malignant STTs is crucial to improve the prognosis of the patients. However, as STTs substantially vary in clinical manifestations, morphological changes, and biological behaviors, it is difficult to make precise classification of malignant STTs, which might lead to delayed diagnosis. It is reported that the diagnosis of soft tissue sarcoma was usually delayed for up to 94.6 weeks (7), which might cause disastrous consequences on patients' outcome, such as a shorter survival time (8, 9).

Ultrasonography (US) is considered to be the first-line imaging method for STTs, due to its fast speed, high resolution, lower cost,

availability, dynamic observation, and no contraindications (2). Gray-scale US can map the locations and morphological changes of STTs, including size, margin, shape, and internal components. The mobility, compressibility, and its anatomical associations with adjacent structures can also be confirmed by dynamic US. Color Doppler US can further display the distribution of intra-tumoral and peri-tumoral blood vessels (10–14). However, US presented an unstable and relatively low diagnostic performance for classifying benign and malignant STTs. The reported accuracy rate of US varied among studies, ranging from 69% to 93% (10–12, 15, 16). And most of the previous studies only involved a small number of cases for evaluation, compromising their reliability. Meanwhile, conventional US is characterized by high operator-dependence and relatively low inter-observer agreement, which also degrade its performance in classifying malignant STTs. How to improve the diagnosis accuracy of US for malignant STTs, at the same time decrease operator dependence, is a very important research topic.

Advanced techniques like Artificial intelligence (AI), especially deep learning (DL) algorithms, possess an excellent ability in image recognition tasks. DL is emerging as a promising tool to resolve various radiology tasks using US images, including screening breast cancer (17, 18), classifying thyroid nodules (19–21), diagnosing liver diseases (22–24), and assessing musculoskeletal abnormality (25). Apart from showing good diagnostic performances in some diseases, DL can also assist radiologists in enhancing their accuracy and reliability in reading US images (26). Currently, AI-based malignant STTs diagnosis based on US images is still in the initial stage (27–29). These studies applied non-DL methods to develop assistant tools for malignant STTs diagnosis, with limited enrolled STTs cases. And clinical indexes were not fully utilized for the model construction in the previous studies.

To overcome the barrier of US diagnosis of malignant STTs, we established this DL-driven AI system, named ultrasound clinical soft tissue tumor net (UC-STTNet), for predicting STTs based on US images and clinical indexes of the patients. First, one of the highlights of the study is the application of two modalities of US imaging, gray-scale US and color-Doppler US, in model construction, which could provide more morphological information of STTs masses. And basic clinical indexes were also

**Abbreviations:** deep learning (DL), ultrasonography (US), soft tissue tumors (STTs), ultrasound clinical soft tissue tumor net (UC-STTNet).



incorporated in the system for a more comprehensive diagnosis of the tumors. Also, we used 5-fold cross validation method in the model building based on a large database. The AI system could also provide heatmaps of US images illustrating the features relevant to model predictions for radiologists to make diagnosis. The AI system could successfully improve the performances and stability of the radiologists in classifying malignant STTs. To the best of our knowledge, our work is the first one applying DL technology for US diagnosis of malignant STTs.

## Materials and methods

### Ethical approval

The study was designed as a retrospective study and approved by the ethics committee of Peking University Shenzhen Hospital (Approval number: 202200901). The informed content was not waived since the retrospective study was observational and did not involve any interventional procedures. And all the information of the patients is anonymized throughout the study. The ethics committee approved the omission of informed content.

### Study participants enrollment

In this work we employed five-fold cross validation for network evaluation. Among 5 folds, 4 and 1 folds were employed for training and testing, respectively. To build the training and testing dataset, we retrospectively reviewed the clinical and imaging data of the patients with STTs from July 2013 to December 2021. The patients with dual-modal US images and pathological results from surgical resections or biopsies were enrolled. To further evaluate the performance of the AI system, we collected a prospective testing dataset from April 2022 to September 2022 in our hospital. Tumors that occurred in superficial organs, including thyroid gland, breast, salivary gland, and lymph nodes, were excluded in both of the retrospective and prospective workflows.

### US imaging and clinical data collection

All US images were derived from US imaging database at Peking University Shenzhen Hospital. The US examinations were performed by radiologists with over five-year experiences of US using commercial US equipment with 5-15MHz probes. Two representative pictures of each patient, one gray scale image showing the largest section of tumor and one color Doppler flow image with the most abundant blood vessels, were selected for model building. The US images of STTs were reviewed and selected from the patients by two radiologists with five-year experiences in US together for image quality control. When disagreement occurred between the two radiologists, they would refer to a third radiologist with over 10-year experiences for the final decision. Two clinicians collected the clinical data for the enrolled patients, including sex, age, duration, locations, layer, the maximum and minimum diameter of lesions, depth from skin, history of malignancy, and surgical history.

## DL architecture development

We designed a multi-data fusion convolutional neural network, named as UC-STTNet, to combine gray scale and color Doppler US images and clinic features for malignant STTs diagnosis. Detailed descriptions about UC-STTNet are shown in Supplementary materials (Supplementary 1; Supplementary Figure 1; Supplementary Table 1).

The image feature extraction consisted of a tumor area enhancement block and a tumor feature extraction block. The tumor area enhancement block was an encoder-decoder network, which employed ResNet18 as backbone and with five down- and up-sample layers. The encoder was employed to extract the region of interest (ROI) feature of STTs, and the decoder was used to generate a ROI feature map which represented the possibility of tumor area (abbr. ROI-map).

The clinical data was directly digitized as a feature vector, which was then processed by a multi-layer perceptron and directly input into the multi-data fusion block. The multi-data fusion block consisted of feature concatenation and attention mechanism. The segmentation and tumor area features were concatenated together, and then the combined features were input into an attention block.

Global average pooling was used to align the image features to linear space and then concatenated with the features of clinic data to generate a multi-data fusion feature for the final STTs classification. Gradient-weighted Class Activation Mapping (Grad-CAM) was adopted in the classification tasks on deep learning to explain the performance of the proposed UC-STTNet. And we used a weighted combination for the forward activation map and activated the result by Rectified Linear Activation function (ReLU) to get the visualization heatmap.

### Reader study and AI-assisted reader study

Six radiologists with three experience levels were invited to review the dual-modal US images and clinical manifestations independently and make diagnosis. The six radiologists participated the reader study included two senior radiologists with 21 and 24 years of experience (R1 and R2), two intermediate radiologists with 10 and 12 years of experience (R3 and R4), and two junior radiologists with 4 and 7 years of experience (R5 and R6). The radiologists were blind to the pathologic results of the tumors. One month after the original reader study, the same STTs cases were re-presented to the six radiologists for a second diagnosis, along with the AI-predicted results and heatmaps as reference. The radiologists were blind to their first-time results and pathological results of the tumors.

### Statistical analysis

The 5-fold cross validation was used for model training and testing. The split was randomly repeated for five times and the average performances were recorded. The receiver operating curve (ROC), area under ROC curve (AUC), accuracy, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) with 95% confidence interval (CI) were used to evaluate

the diagnostic performance of the model, the radiologists, and the radiologists with AI assistance. AUC values of the same dataset and different datasets were compared to use the methods reported by DeLong et al (30) and Hanley and McNeil (31), respectively. We further calculated the intra-class correlation (ICC) with 95% CI to evaluate the inter-observer variability of the six radiologists before and after the assistance of AI results.  $P < 0.05$  was considered as statistically significant. The statistical analyses were performed by using Medcalc (Version 20.0, MedCalc Software Ltd, Belgium).

## Results

In this study, we developed and presented UC-STTNet, an AI system based on a deep-learning architecture for malignant STTs diagnosis. The study flow of the construction and validation of UC-STTNet is shown in Figure 1.

From July 2013 to December 2021, a total of 15120 adult patients with STTs were received from pathology data, which came from core biopsy or surgery. There were 546 patients with malignancy and 14574 patients with benign. Finally, 733 masses of 693 patients, including 271 malignant masses of 231 patients and 462 benign masses of 462 patients, were enrolled in this study to build the model. For the prospective validation of the model, we recruited 145 masses of 145 patients, including 44 malignant masses of 44 patients and 101 benign masses of 101 patients from April 2022 to September 2022. The workflow of the retrospective and prospective patient recruitment is shown in Figure 2.

## Clinical data and pathological results

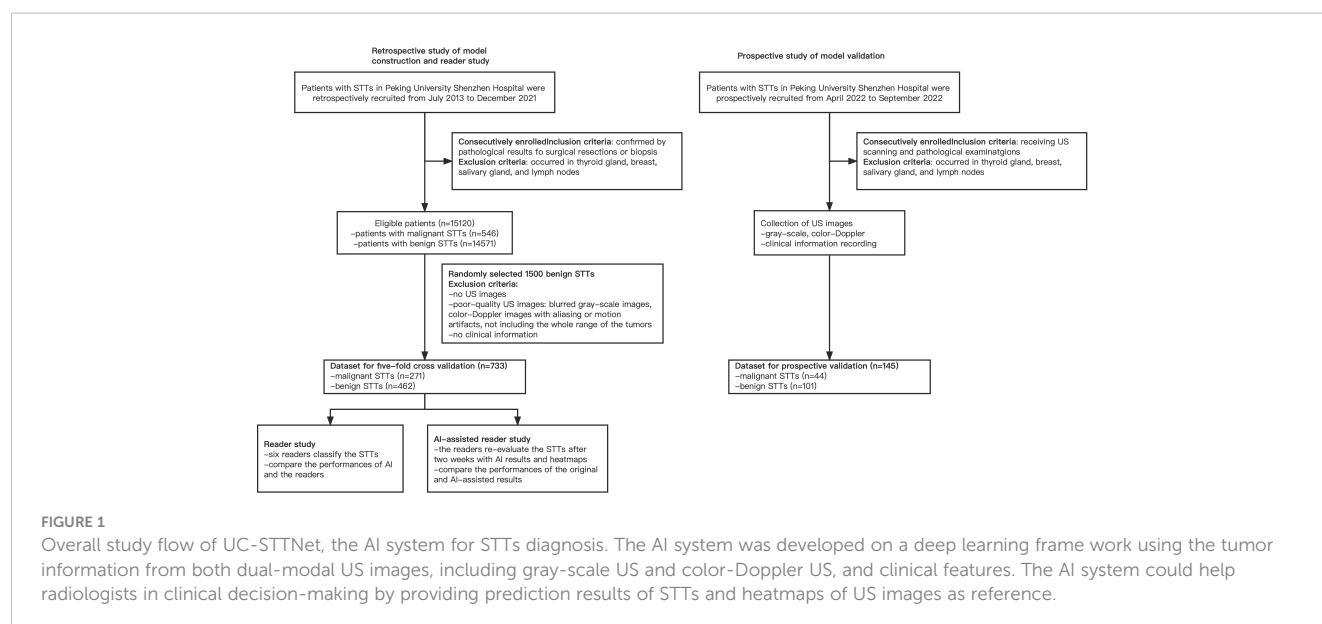
The clinical features and pathological results of the retrospectively enrolled and prospectively enrolled STTs were

listed in Table 1. Except sex, the other clinical characteristics between benign and malignant masses, were significantly different.

## Performance of the AI system on the retrospective dataset for model building

The performance of the AI system was evaluated using 5-fold cross validation. Of the five validation sets, the highest AUC was 0.91 (95% CI: 0.84, 0.95), with accuracy of 0.89 (95% CI: 0.84, 0.94), sensitivity of 0.82 (95% CI: 0.72, 0.82), specificity of 0.93 (95% CI: 0.88, 0.98), PPV of 0.88 (95% CI: 0.79, 0.97), NPV of 0.90 (95% CI: 0.84, 0.96), respectively. The average AUC, accuracy, sensitivity, specificity, PPV and NPV of the model in the five validations were 0.89 (95% CI: 0.87, 0.92), 0.84 (95% CI: 0.82, 0.87), 0.76 (95% CI: 0.71, 0.81), 0.90 (95% CI: 0.87, 0.92), 0.81 (95% CI: 0.76, 0.86) and 0.86 (95% CI: 0.83, 0.89), respectively. The AI system showed higher specificity than sensitivity, indicating that the majority of the benign cases (above 90%) were accurately recognized. While there were around 20% of the malignant cases mistakenly classified as benign. Similarly, the NPV value of the model was slightly higher than the PPV value, indicating that the AI system had more confidence (around 3%) in predicting benign cases.

Figure 3A presents the ROCs of all five folds. According to the figure, UC-STTNet appears to be robust and stable when trained and tested with different folds of data. The AUC values of the AI system in the five validations ranged from 0.84 to 0.91, with the standard deviation 0.028. Figure 3B depicted the performances of radiologists with three different experience levels. The diagnostic performance of UC-STTNet was higher than that of one of the senior radiologists (AUC of UC-STTNet vs AUC of R2: 0.89 vs. 0.84,  $p=0.022$ ) and all the intermediate and junior radiologists (AUC of UC-STTNet vs AUC of R3, R4, R5, and R6: 0.89 vs 0.75, 0.81, 0.80, 0.63;  $p < 0.01$ ), and was comparable to one of the



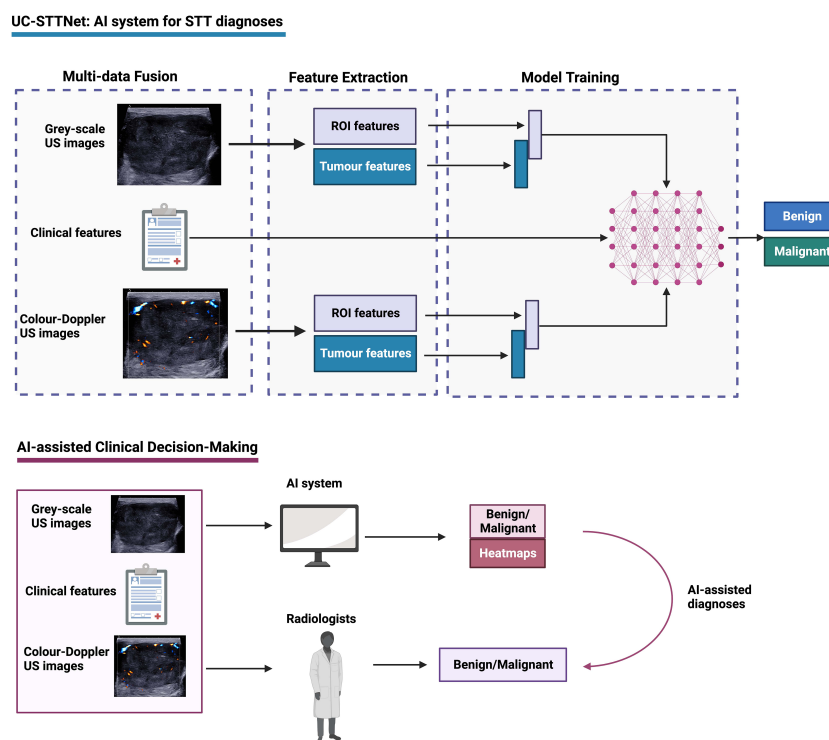


FIGURE 2  
Flow chart of the retrospective and prospective patients' recruitment.

high-level radiologists (AUC of UC-STTNet vs AUC of R1: 0.89 vs 0.87,  $p=0.30$ ). And there were no significant differences in accuracy, specificity and PPV between the AI system and the intermediated-level radiologists ( $p=0.09$ , 0.96, and 0.72, respectively). And the AI system showed better sensitivity and NPV than the intermediated-level radiologists ( $p=0.01$  and 0.04, respectively).

## Assistant role of the AI system for radiologists

The change in diagnostic performance of each radiologist after the assistance from the AI system was displayed in Figure 4 | Table 2. For junior radiologists (R5 and R6) and one of the intermediate radiologists (R3), the AUC values after the AI assistance were significantly improved (R3: 0.75 to 0.83,  $p<0.01$ ; R5: 0.80 to 0.85,  $p<0.01$ ; R6: 0.63 to 0.69,  $p<0.01$ ), indicating that the diagnostic performances of the radiologists could be enhanced via the aid of the AI system.

Subsequently, we calculated the ICC value among the six radiologists in classifying the malignant STTs. The original ICC value of the radiologists before referring to the AI system was 0.87 (0.84–0.89), which increased to 0.92 (0.91–0.93) after AI assistance, indicating the diagnostic agreement of the radiologists could be improved via the aid of the AI system.

## Explainability of the AI system

Explainability of UC-STTNet was demonstrated as heatmaps that highlights the significant areas attended by the model for

malignant STTs diagnosis. The examples of the AI prediction of malignant STTs were illustrated in Figure 5. UC-STTNet gave the prediction result of a malignant STTs mass based on its dual modal US images and clinical indexes. The heatmap of the mass was generated by the AI system and used as reference for radiologists.

## Performance of the AI system on the prospective dataset

The AUC, accuracy, sensitivity, specificity, PPV and NPV of the AI system on the prospective dataset were 0.85 (95% CI: 0.82, 0.89), 0.83 (95% CI: 0.77, 0.90), 0.63 (95% CI: 0.49, 0.78), 0.91 (95% CI: 0.86, 0.97), 0.75 (95% CI: 0.62, 0.90) and 0.85 (95% CI: 0.79, 0.92), respectively. The AUC value of the AI system on the prospective dataset had no statistical difference with the average AUC value on the model-building dataset (0.89 vs 0.85,  $p=0.282$ ). The diagnostic performance of the AI system in the prospective dataset is shown in Figure 6.

## Discussion

In this study, a total of 733 and 145 masses were collected retrospectively and prospectively. We successfully built a DL-driven AI system, named UC-STTNet, for distinguishing the malignant STTs from benign ones based on dual modal US images and clinical manifestations. The AI system achieved the average AUC value of 0.89 in the retrospective dataset, showing a diagnostic performance

TABLE 1 Clinical characteristics of 733 retrospective STTs masses and 145 prospective STTs masses.

|           | AUC                 | Accuracy             | Sensitivity (%)      | Specificity (%)         | PPV                   | NPV                   |
|-----------|---------------------|----------------------|----------------------|-------------------------|-----------------------|-----------------------|
| AI system | 0.89<br>(0.87-0.92) | 0.84<br>(0.82-0.87)  | 0.76<br>(0.71-0.81)  | 0.90<br>(0.87-0.92)     | 0.81<br>(0.76-0.86)   | 0.86<br>(0.83-0.89)   |
| R1        | 0.87<br>(0.85-0.90) | 0.88<br>(0.86-0.91)  | 83.4<br>(78.4-87.6)  | 91.34<br>(88.4 - 93.7)  | 0.85<br>(0.81 ~ 0.89) | 0.90<br>(0.88 ~ 0.93) |
| R1+AI     | 0.88<br>(0.86-0.91) | 0.89<br>(0.87-0.91)  | 86.4<br>(81.7-90.2)  | 90.26<br>(87.2 - 92.80) | 0.84<br>(0.79 ~ 0.88) | 0.92<br>(0.89 ~ 0.94) |
| R2        | 0.84<br>(0.81-0.87) | 0.87<br>(0.85 -0.89) | 72.7<br>(67.0-77.9)  | 95.24<br>(92.9 - 97.0)  | 0.90<br>(0.86 ~ 0.94) | 0.86<br>(0.83 ~ 0.89) |
| R2+AI     | 0.82<br>(0.79-0.85) | 0.86<br>(0.83-0.89)  | 68.3<br>(62.4-73.8)  | 96.32<br>(94.2 - 97.8)  | 0.92<br>(0.88 ~ 0.95) | 0.84<br>(0.81 ~ 0.87) |
| R3        | 0.75<br>(0.72-0.78) | 0.80<br>(0.78-0.83)  | 54.6<br>(48.5-60.6)  | 95.45<br>(93.1 - 97.2)  | 0.88<br>(0.83 ~ 0.93) | 0.78<br>(0.75 ~ 0.82) |
| R3+AI     | 0.83<br>(0.80-0.86) | 0.85<br>(0.82-0.88)  | 75.3<br>(69.7-80.3)  | 90.69<br>(87.7 - 93.2)  | 0.83<br>(0.78 ~ 0.87) | 0.86<br>(0.83 ~ 0.89) |
| R4        | 0.81<br>(0.78-0.83) | 0.82<br>(0.79-0.85)  | 76.1<br>(70.5-81.0)  | 85.06<br>(81.5 - 88.2)  | 0.75<br>(0.70 ~ 0.80) | 0.86<br>(0.83 ~ 0.89) |
| R4+AI     | 0.81<br>(0.78-0.84) | 0.83<br>(0.81-0.86)  | 71.6<br>(65.8-76.9)  | 90.04<br>(86.9 - 92.6)  | 0.81<br>(0.76 ~ 0.86) | 0.84<br>(0.81 ~ 0.88) |
| R5        | 0.80<br>(0.77-0.83) | 0.80<br>(0.77-0.83)  | 83.4<br>(78.4-87.6)  | 77.27<br>(73.2 - 81.0)  | 0.68<br>(0.63 ~ 0.73) | 0.89<br>(0.86 ~ 0.92) |
| R5+AI     | 0.85<br>(0.82-0.88) | 0.86<br>(0.84-0.89)  | 80.07<br>(74.8-84.7) | 90.04<br>(86.9 - 92.6)  | 0.83<br>(0.78 ~ 0.87) | 0.89<br>(0.86 ~ 0.91) |
| R6        | 0.63<br>(0.59-0.66) | 0.71<br>(0.68-0.74)  | 31<br>(25.5-36.9)    | 94.16<br>(91.6 - 96.1)  | 0.76<br>(0.68 ~ 0.84) | 0.70<br>(0.66 ~ 0.74) |
| R6+AI     | 0.69<br>(0.65-0.72) | 0.76<br>(0.73-0.79)  | 40.22<br>(34.3-46.3) | 96.97<br>(95.0 - 98.3)  | 0.89<br>(0.83 ~ 0.94) | 0.73<br>(0.70 ~ 0.77) |

STTs, soft tissue tumors.

comparable to high-level radiologists, superior to intermediate and junior radiologists. With the assistance of the system, the diagnostic performances and inter-observer agreement of the radiologists could be further enhanced. To note, the number of STTs patients

and cases involved in our study is so far the largest, among all available literature works.

Diagnostic models for classifying malignant STTs based on US images have been developed by several studies. Despite of their high

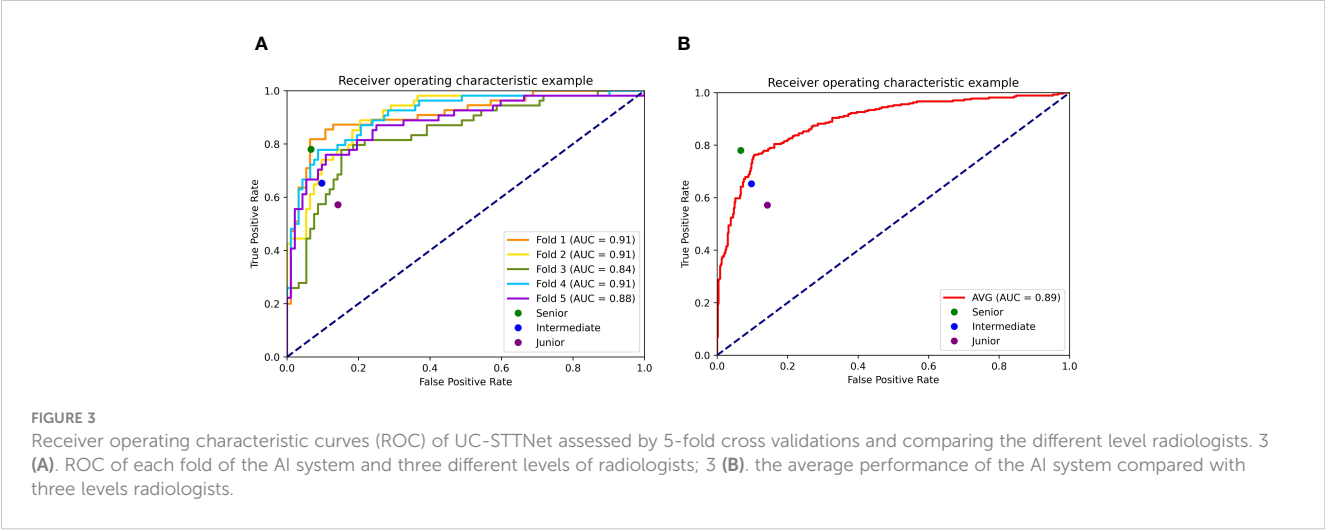
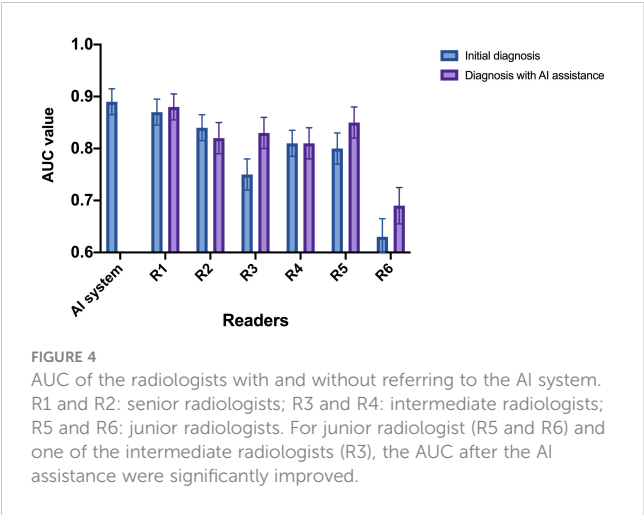


FIGURE 3 Receiver operating characteristic curves (ROC) of UC-STTNet assessed by 5-fold cross validations and comparing the different level radiologists. 3 (A). ROC of each fold of the AI system and three different levels of radiologists; 3 (B). the average performance of the AI system compared with three levels radiologists.



accuracy, the previous models have some disadvantages and are not suitable for clinical promotion. Chen et al. developed a computer-aid-diagnosis (CAD) system using US images to improve the accuracy of 89.5% for malignant STTs (27). However, manual identifications of lesions were required in their CAD system, which is time-consuming and not convenient for clinical application. Wu MJ et al. established a STTs diagnostic nomogram integrating ultrasound and clinical features via multivariable regression analysis, which achieved an AUC value of 0.896 (29). The sample size for model building was also relatively small. Compared with previous studies, our study has the following strengths. Firstly, we used deep learning algorithm to develop the AI system for diagnosis, which was more intelligent and robust than the hand-crafted systems, and could made automatic diagnosis of the masses. Secondly, the AI system utilized the imaging data of two US modalities, the gray scale and color Doppler US, as well as

**TABLE 2** Performance of the AI system, the radiologists with three experience levels, and AI-assisted radiologists.

| Clinical Characteristics | Retrospective STTs masses |               |                  |         | Prospective STTs masses |               |                  |         |
|--------------------------|---------------------------|---------------|------------------|---------|-------------------------|---------------|------------------|---------|
|                          | Overall                   | Benign masses | Malignant masses | P-Value | Overall                 | Benign masses | Malignant masses | P-Value |
| Sex                      |                           |               |                  | 0.8825  |                         |               |                  | 0.100   |
| Male                     | 338                       | 214           | 124              |         | 61                      | 38            | 23               |         |
| Female                   | 395                       | 248           | 147              |         | 84                      | 63            | 21               |         |
| Age                      | 27.6±15.4                 | 40.3±13.9     | 51.7±15.3        | <0.001  | 41.96±15.5              | 38.8±13.4     | 49.2±17.6        | 0.057   |
| Malignant History        |                           |               |                  | <0.001  |                         |               |                  | <0.001  |
| Yes                      | 164                       | 17            | 147              |         | 13                      | 0             | 13               |         |
| None                     | 579                       | 445           | 124              |         | 132                     | 101           | 31               |         |
| Surgical History         |                           |               |                  | <0.001  |                         |               |                  | <0.001  |
| Yes                      | 189                       | 63            | 126              |         | 12                      | 101           | 32               |         |
| None                     | 544                       | 399           | 146              |         | 133                     | 0             | 12               |         |
| Tumor Duration (months)  | 27.6±47.8                 | 33.9±52.1     | 15.5±42.2        | <0.001  | 28.3±42.8               | 31.4±46.2     | 21.9±34.7        | 0.06    |
| Tumor Position           |                           |               |                  | <0.001  |                         |               |                  | <0.001  |
| Head or neck             | 118                       | 91            | 27               |         | 39                      | 30            | 9                |         |
| Trunk                    | 286                       | 109           | 177              |         | 24                      | 10            | 14               |         |
| Upper limb               | 193                       | 172           | 21               |         | 48                      | 43            | 5                |         |
| Lower limb               | 136                       | 90            | 46               |         | 34                      | 18            | 16               |         |
| Tumor Side               |                           |               |                  | 0.012   |                         |               |                  | 0.688   |
| Left                     | 303                       | 186           | 117              |         | 67                      | 49            | 18               |         |
| Right                    | 342                       | 208           | 134              |         | 58                      | 39            | 19               |         |
| Mid                      | 88                        | 68            | 20               |         | 20                      | 13            | 7                |         |
| Tumor Long Diameter (mm) | 33.2±29.4                 | 25.9±19.4     | 45.7±38.1        | <0.001  | 34.7±38.5               | 22.8±15.6     | 62.3±57.3        | 0.026   |

(Continued)



TABLE 2 Continued

| Clinical Characteristics | Retrospective STTs masses |               |                  |         | Prospective STTs masses |               |                  |         |
|--------------------------|---------------------------|---------------|------------------|---------|-------------------------|---------------|------------------|---------|
|                          | Overall                   | Benign masses | Malignant masses | P-Value | Overall                 | Benign masses | Malignant masses | P-Value |
| Tumor Short Diameter(mm) | 16.1±16.2                 | 11.4±10.8     | 24.2±20.2        | <0.001  | 16.3±20.7               | 9.5±8.0       | 32.0±30.4        | <0.001  |
| Tumor Depth(mm)          | 4.5±4.2                   | 3.6±3.3       | 6.2±4.9          | <0.001  | 4.3±5.0                 | 2.8±2.1       | 7.8±7.5          | 0.005   |
| Anatomical Level         |                           |               |                  | <0.001  |                         |               |                  | <0.001  |
| Superficial fascia layer | 613                       | 412           | 201              |         | 113                     | 95            | 18               |         |
| Deep fascia layer        | 120                       | 50            | 70               |         | 32                      | 6             | 26               |         |
| Pathological types       | Retrospective STTs masses |               |                  |         | Prospective STTs masses |               |                  |         |
| Malignant types          | 271                       |               |                  |         | 44                      |               |                  |         |
| Sarcoma                  | 74                        |               |                  |         | 33                      |               |                  |         |
| Metastasis               | 137                       |               |                  |         | 0                       |               |                  |         |
| Lymphoma                 | 13                        |               |                  |         | 3                       |               |                  |         |
| Squamous-cell carcinoma  | 13                        |               |                  |         | 5                       |               |                  |         |
| Melanoma                 | 9                         |               |                  |         | 3                       |               |                  |         |
| Others                   | 25                        |               |                  |         | 0                       |               |                  |         |
| Benign types             | 462                       |               |                  |         | 101                     |               |                  |         |
| lipoma                   | 109                       |               |                  |         | 22                      |               |                  |         |
| hemangioma               | 57                        |               |                  |         | 20                      |               |                  |         |
| epidermoid cyst          | 51                        |               |                  |         | 10                      |               |                  |         |
| schwannoma               | 35                        |               |                  |         | 8                       |               |                  |         |
| giant cell tumor         | 25                        |               |                  |         | 0                       |               |                  |         |
| Others                   | 185                       |               |                  |         | 41                      |               |                  |         |

AI, Artificial intelligence; R1 and R2: senior radiologists R3 and R4: intermediate radiologists R5 and R6: junior radiologists.

clinical information, to make a more comprehensive diagnosis of the tumors. Also, the AI system was built on a relatively large number of cases, and its accuracy and robustness were validated on a prospective dataset. The diagnostic performance of UC-STTNet was comparable with a meta-analysis of elastography in assessment of malignant STTs (16). The average AUC and accuracy of our system were 0.89 and 0.84, demonstrating a better performance than the contrast-enhanced ultrasound (CEUS) for predicting the malignancy of STTs, whose AUC and accuracy were 0.86 and 0.81, respectively (32).

We also verified the assistant role of the AI system for radiologists in making diagnosis of malignant STTs. While our results showed that UC-STTNet was superior to the performance of intermediate and junior radiologists, our AI system could help these less experience radiologists make more accurate diagnosis. Meanwhile, the inter-observer agreement of the radiologists was also improved when they referred to the diagnostic results of the AI system. UC-STTNet not only provided the final predictive results of the masses, but also generated heatmaps representing the active areas for diagnosis for the radiologists. Therefore, the AI

system could be utilized as an assistant tool for the radiologists to enhance their diagnostic performance and stability in STTs, as well as to decrease operator dependence. To note, compared with other models for diagnosing STTs, including the model based on hand-crafted ultrasound features and the model based on radiomics, the process of using our AI system is more clinical applicable. The AI system can directly generate the result for prediction and does not need lesion delineation and feature extraction. For further clinical promotion of the AI system in the future, we will attempt to integrate the DL architecture into commercial US devices as an on-board software to help to improve the diagnosis performance and decrease workforce for radiologists.

The AI system tends to misdiagnose the benign masses with large size, usually more than 30mm in longitude. The benign STTs that possessing abundant blood vessels on color Doppler US imaging, such as glomangioma, could also be misdiagnosed by the system. On the other hand, the malignant tumors with small size and scarce vasculature might be classified as benign ones. In addition, a total of 6 cases of dermatofibrosarcoma protuberans

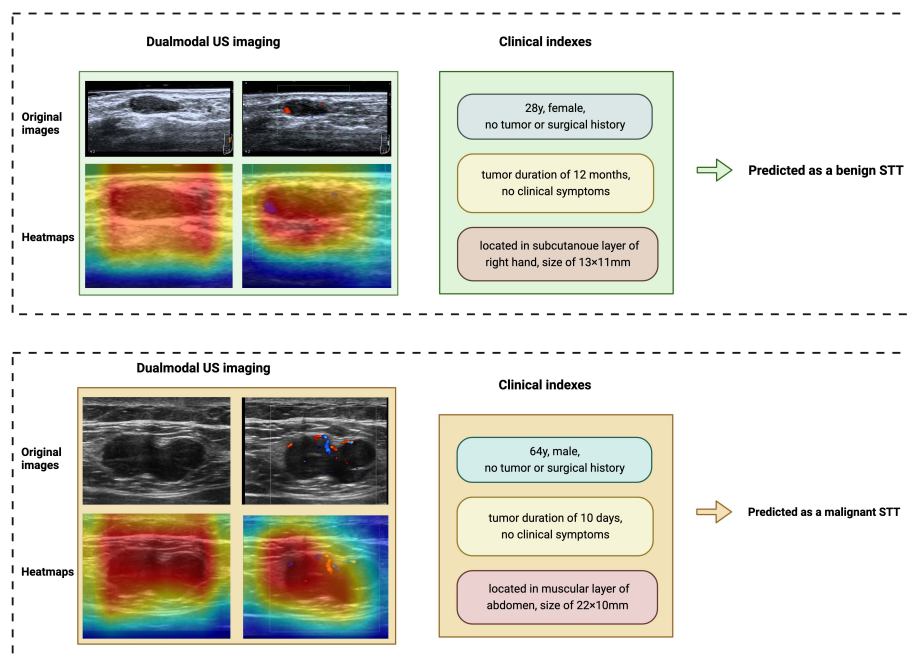


FIGURE 5

Examples of the AI system classifying benign and malignant STTs. The AI system diagnosed STTs based on dual-modal US images and clinical features. Heatmaps of the two modalities of US were also provided by the system. The above case is a 28-year-old female with a STTs mass on the subcutaneous layer of the right hand. She had no tumor or surgical history. The tumor was found 12 months ago and had a size of 13×11mm. The AI system diagnosed it as a benign tumor, which was identified as a benign schwannoma by pathology. The other case is a 64-year-old male with a STTs mass on the muscular layer of abdomen. The patient also reported no tumor or surgical history. The tumor was found 10 days ago and had a size of 22×11mm. The AI system diagnosed it as a malignant STTs tumor, which was identified as a metastatic malignant melanoma by pathology.

(DFSP) were predicted as benign by the AI system. For DFSPs, skin changes should also be taken into account during diagnosis. Additionally, misdiagnosis often occurs in patients with a history of malignancy. To prevent the aforementioned misdiagnosis scenarios, more cases should be supplied for model development in the future study. [Supplementary Figure 2](#) demonstrated the examples of the misdiagnosed STTs cases of the AI system.

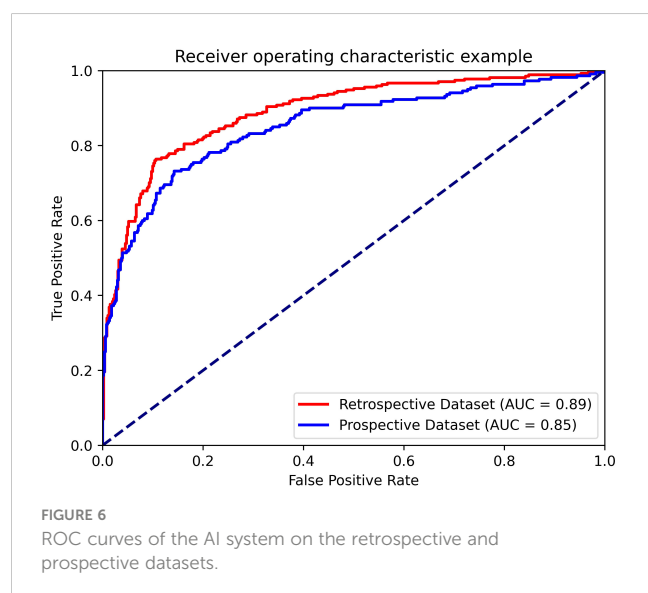


FIGURE 6

ROC curves of the AI system on the retrospective and prospective datasets.

Our study has several limitations. First, it was a single center research. The AI system was not verified by external validation from multi-center datasets. The sensitivity of the AI system was relatively low in the prospective validation dataset, which should be improved by enrolling more malignant cases in the training dataset in further studies. Also, we only used two modalities of US images to build the model, and other available US modalities, including US elastography and CEUS, were not incorporated in our study. The two US modalities will be added to the system in our future study to improve its diagnostic accuracy. Moreover, we compared the performance of the DL model with the radiologists on the retrospective dataset due to its relatively large sample size. In the future study, the accuracy of the model will be further explored on a large prospective data.

## Conclusions

A DL-driven AI system based on dual-modal US images and clinical features for malignant STTs diagnosis was developed on a retrospective dataset of STTs. It achieved a high accuracy in predicting malignant STTs on both retrospective and prospective datasets. The performance of the AI system was comparable to senior radiologists, and better than junior and intermediate radiologists. The developed AI system could also assist radiologists in improving their diagnostic accuracy and stability in classifying malignant STTs.

## 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 Peking University Shenzhen 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

HX: Writing – review & editing, Supervision, Project administration. YZ: Writing – original draft, Software, Formal analysis. LD: Writing – original draft, Software, Methodology, Data curation. HL: Writing – original draft, Software, Formal analysis, Data curation. XL: Writing – original draft, Software, Methodology, Formal analysis. CZ: Writing – original draft, Software, Formal analysis, Data curation. YT: Writing – original draft, Data curation. LX: Writing – original draft, Data curation. WW: Writing – original draft, Data curation. QY: Writing – original draft, Data curation. LL: Writing – review & editing, Supervision. DS: Writing – review & editing, Supervision. LQ: Writing – review & editing, Supervision. LS: Writing – review & editing, Supervision, Project administration. YZ: Writing – original draft, Data curation.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This work

was supported by the General Program for Clinical Research at Peking University Shenzhen Hospital (LCYJ2021003), the Science Technology and Innovation Commission of Shenzhen Municipality in China (JCYJ20210324110015040), Research Foundation of Peking University Shenzhen Hospital (JCYJ2020012), Shenzhen Key Medical Discipline Construction Fund (SZXK051), Sanming Project of Medicine in Shenzhen (SZSM202111011), National Key Research and Development Program of China (2023YFC3402605), the National Natural Science Foundation of China (82261138629), Guangdong Basic and Applied Basic Research Foundation (2023A1515010688), Shenzhen Municipal Science and Technology Innovation Council (JCYJ20220531101412030).

## 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/fonc.2024.1361694/full#supplementary-material>

## References

- Jo VY, Doyle LA. Refinements in sarcoma classification in the current 2013 world health organization classification of tumours of soft tissue and bone. *Surg Oncol Clin N Am.* (2016) 25:621–43. doi: 10.1016/j.soc.2016.05.001
- Colleran G, Madewell J, Foran P, Shelly M, O'Sullivan PJ. Imaging of soft tissue and osseous sarcomas of the extremities. *Semin Ultrasound CT MR.* (2011) 32:442–55. doi: 10.1053/j.sult.2011.05.001
- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin.* (2018) 68:7–30. doi: 10.3322/caac.21442
- Toro JR, Travis LB, Wu HJ, Zhu K, Fletcher CD, Devesa SS. Incidence patterns of soft tissue sarcomas, regardless of primary site, in the surveillance, epidemiology and end results program, 1978–2001: An analysis of 26,758 cases. *Int J Cancer.* (2006) 119:2922–30. doi: 10.1002/ijc.22239
- Tan MC, Brennan MF, Kuk D, Agaram NP, Antonescu CR, Qin LX, et al. Histology-based classification predicts pattern of recurrence and improves risk stratification in primary retroperitoneal sarcoma. *Ann Surg.* (2016) 263:593–600. doi: 10.1097/SLA.0000000000001149
- Pastore G, Peris-Bonet R, Carli M, Martinez-Garcia C, Sanchez de Toledo J, Steliarova-Foucher E. Childhood soft tissue sarcomas incidence and survival in European children (1978–1997): report from the Automated Childhood Cancer Information System project. *Eur J Cancer.* (2006) 42:2136–49. doi: 10.1016/j.ejca.2006.05.016
- Soomers V, Husson O, Young R, Desai I, van der Graaf W. The sarcoma diagnostic interval: a systematic review on length, contributing factors and patient outcomes. *ESMO Open.* (2020) 5(1):e000592. doi: 10.1136/esmoopen-2019-000592
- Ferrari A, Miceli R, Casanova M, Meazza C, Favini F, Luksch R, et al. The symptom interval in children and adolescents with soft tissue sarcomas. *Cancer.* (2010) 116:177–83. doi: 10.1002/cncr.24695
- Nakamura T, Matsumine A, Matsubara T, Asanuma K, Uchida A, Sudo A. The symptom-to-diagnosis delay in soft tissue sarcoma influence the overall survival and the development of distant metastasis. *J Surg Oncol.* (2011) 104:771–5. doi: 10.1002/jso.22006
- Hung EHY, Griffith JF, Yip SWY, Ivory M, Lee JCH, Ng AWH, et al. Accuracy of ultrasound in the characterization of superficial soft tissue tumors: a prospective study. *Skeletal Radiol.* (2020) 49:883–92. doi: 10.1007/s00256-019-03365-z
- Nagano S, Yahiro Y, Yokouchi M, Setoguchi T, Ishidou Y, Sasaki H, et al. Doppler ultrasound for diagnosis of soft tissue sarcoma: efficacy of ultrasound-based screening score. *Radiol Oncol.* (2015) 49:135–40. doi: 10.1515/raon-2015-0011

12. Ozturk M, Selcuk MB, Polat AV, Ozbalci AB, Baris YS. The diagnostic value of ultrasound and shear wave elastography in the differentiation of benign and Malignant soft tissue tumors. *Skeletal Radiol.* (2020) 49:1795–805. doi: 10.1007/s00256-020-03492-y
13. Woo T, Lalam R, Cassar-Pullicino V, Degrieck B, Verstraete K, Donati DM, et al. Imaging of upper limb tumors and tumorlike pathology. *Radiol Clin North Am.* (2019) 57:1035–50. doi: 10.1016/j.rcl.2019.03.008
14. Rimondi E, Benassi MS, Bazzocchi A, Balladelli A, Facchini G, Rossi G, et al. Translational research in diagnosis and management of soft tissue tumours. *Cancer Imaging.* (2016) 16:13. doi: 10.1186/s40644-016-0071-7
15. De Marchi A, Pozza S, Charrier L, Cannone F, Cavallo F, Linari A, et al. Small subcutaneous soft tissue tumors (<5 cm) can be sarcomas and contrast-enhanced ultrasound (CEUS) is useful to identify potentially Malignant masses. *Int J Environ Res Public Health.* (2020) 17(23):8868. doi: 10.3390/ijerph17238868
16. Wu M, Ren A, Xu D, Peng X, Ye X, Li A. Diagnostic performance of elastography in Malignant soft tissue tumors: A systematic review and meta-analysis. *Ultrasound Med Biol.* (2021) 47:855–68. doi: 10.1016/j.ultrasmedbio.2020.12.017
17. Jiang M, Li CL, Luo XM, Chuan ZR, Lv WZ, Li X, et al. Ultrasound-based deep learning radiomics in the assessment of pathological complete response to neoadjuvant chemotherapy in locally advanced breast cancer. *Eur J Cancer.* (2021) 147:95–105. doi: 10.1016/j.ejca.2021.01.028
18. Zheng X, Yao Z, Huang Y, Yu Y, Wang Y, Liu Y, et al. Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. *Nat Commun.* (2020) 11:1236. doi: 10.1038/s41467-020-15027-z
19. Bai Z, Chang L, Yu R, Li X, Wei X, Yu M, et al. Thyroid nodules risk stratification through deep learning based on ultrasound images. *Med Phys.* (2020) 47:6355–65. doi: 10.1002/mp.14543
20. Lee JH, Chai YJ. A deep-learning model to assist thyroid nodule diagnosis and management. *Lancet Digit Health.* (2021) 3:e409. doi: 10.1016/S2589-7500(21)00114-X
21. Peng S, Liu Y, Lv W, Liu L, Zhou Q, Yang H, et al. Deep learning-based artificial intelligence model to assist thyroid nodule diagnosis and management: a multicentre diagnostic study. *Lancet Digit Health.* (2021) 3:e250–9. doi: 10.1016/S2589-7500(21)00041-8
22. Chaudhary K, Poirion OB, Lu L, Garmire LX. Deep learning-based multi-omics integration robustly predicts survival in liver cancer. *Clin Cancer Res.* (2018) 24:1248–59. doi: 10.1158/1078-0432.CCR-17-0853
23. Wang K, Lu X, Zhou H, Gao Y, Zheng J, Tong M, et al. Deep learning Radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis B: a prospective multicentre study. *Gut.* (2019) 68:729–41. doi: 10.1136/gutjnl-2018-316204
24. Xi IL, Wu J, Guan J, Zhang PJ, Horii SC, Soulen MC, et al. Deep learning for differentiation of benign and Malignant solid liver lesions on ultrasonography. *Abdom Radiol (NY).* (2021) 46:534–43. doi: 10.1007/s00261-020-02564-w
25. Christensen ABH, Just SA, Andersen JKH, Savarimuthu TR. Applying cascaded convolutional neural network design further enhances automatic scoring of arthritis disease activity on ultrasound images from rheumatoid arthritis patients. *Ann Rheum Dis.* (2020) 79:1189–93. doi: 10.1136/annrheumdis-2019-216636
26. Zhao C, Xiao M, Jiang Y, Liu H, Wang M, Wang H, et al. Feasibility of computer-assisted diagnosis for breast ultrasound: the results of the diagnostic performance of S-detect from a single center in China. *Cancer Manag Res.* (2019) 11:921–30. doi: 10.2147/CMAR
27. Chen CY, Chiou HJ, Chou SY, Chiou SY, Wang HK, Chou YH, et al. Computer-aided diagnosis of soft-tissue tumors using sonographic morphologic and texture features. *Acad Radiol.* (2009) 16:1531–8. doi: 10.1016/j.acra.2009.07.024
28. Chen CY, Chiou HJ, Chou YH, Chiou SY, Wang HK, Chou SY, et al. Computer-aided diagnosis of soft tissue tumors on high-resolution ultrasonography with geometrical and morphological features. *Acad Radiol.* (2009) 16:618–26. doi: 10.1016/j.acra.2008.12.016
29. Wu M, Hu Y, Ren A, Peng X, Ma Q, Mao C, et al. Nomogram based on ultrasonography and clinical features for predicting Malignancy in soft tissue tumors. *Cancer Manag Res.* (2021) 13:2143–52. doi: 10.2147/CMAR.S296972
30. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* (1988) 44:837–45. doi: 10.2307/2531595
31. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology.* (1983) 148:839–43. doi: 10.1148/radiology.148.3.6878708
32. Wu M, Hu Y, Hang J, Peng X, Mao C, Ye X, et al. Qualitative and quantitative contrast-enhanced ultrasound combined with conventional ultrasound for predicting the Malignancy of soft tissue tumors. *Ultrasound Med Biol.* (2022) 48:237–47. doi: 10.1016/j.ultrasmedbio.2021.10.007



## OPEN ACCESS

## EDITED BY

Yuxiang Zhou,  
Mayo Clinic Arizona, United States

## REVIEWED BY

Jaya Lakshmi Thangaraj,  
University of California, San Diego,  
United States  
Tingfan Wu,  
Shanghai United Imaging Medical Technology  
Co., Ltd., China

## \*CORRESPONDENCE

Qianqian Zhang  
✉ wfcyzhangqian@126.com  
Naixuan Li  
✉ xuannaili@163.com  
Cong Xu  
✉ 616574369@qq.com

<sup>†</sup>These authors have contributed  
equally to this work and share  
first authorship

RECEIVED 16 January 2024

ACCEPTED 27 June 2024

PUBLISHED 11 July 2024

## CITATION

Hu G, Qu J, Gao J, Chen Y, Wang F, Zhang H,  
Zhang H, Wang X, Ma H, Xie H, Xu C, Li N and  
Zhang Q (2024) Radiogenomics nomogram  
based on MRI and microRNAs to predict  
microvascular invasion of hepatocellular  
carcinoma.  
*Front. Oncol.* 14:1371432.  
doi: 10.3389/fonc.2024.1371432

## COPYRIGHT

© 2024 Hu, Qu, Gao, Chen, Wang, Zhang,  
Zhang, Wang, Ma, Xie, Xu, 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.

# Radiogenomics nomogram based on MRI and microRNAs to predict microvascular invasion of hepatocellular carcinoma

Guangchao Hu<sup>1†</sup>, Jianyi Qu<sup>2†</sup>, Jie Gao<sup>3†</sup>, Yuqian Chen<sup>4</sup>,  
Fang Wang<sup>5</sup>, Haicheng Zhang<sup>6</sup>, Han Zhang<sup>6</sup>, Xuefeng Wang<sup>3</sup>,  
Heng Ma<sup>6</sup>, Haizhu Xie<sup>6</sup>, Cong Xu<sup>7\*</sup>, Naixuan Li<sup>8\*</sup>  
and Qianqian Zhang<sup>6\*</sup>

<sup>1</sup>Department of Radiology, Qingdao Municipal Hospital, Qingdao, Shandong, China, <sup>2</sup>Department of Radiology, Zhongshan Hospital, Fudan University, Shanghai, China, <sup>3</sup>Department of Hepatobiliary Surgery, Yantai Yuhuangding Hospital, Qingdao University, Yantai, Shandong, China, <sup>4</sup>School of Information and Electronic Engineering, Shandong Technology and Business University, Yantai, Shandong, China, <sup>5</sup>Department of Pathology, Yantai Yuhuangding Hospital, Qingdao University, Yantai, Shandong, China, <sup>6</sup>Department of Radiology, Yantai Yuhuangding Hospital, Qingdao University, Yantai, Shandong, China, <sup>7</sup>Department of Physical Examination Center, Yantai Yuhuangding Hospital, Qingdao University, Yantai, Shandong, China, <sup>8</sup>Department of Interventional Vascular Surgery, Yantai Affiliated Hospital of Binzhou Medical University, Yantai, Shandong, China

**Purpose:** This study aimed to develop and validate a radiogenomics nomogram for predicting microvascular invasion (MVI) in hepatocellular carcinoma (HCC) on the basis of MRI and microRNAs (miRNAs).

**Materials and methods:** This cohort study included 168 patients (training cohort: n = 116; validation cohort: n = 52) with pathologically confirmed HCC, who underwent preoperative MRI and plasma miRNA examination. Univariate and multivariate logistic regressions were used to identify independent risk factors associated with MVI. These risk factors were used to produce a nomogram. The performance of the nomogram was evaluated by receiver operating characteristic curve (ROC) analysis, sensitivity, specificity, accuracy, and F1-score. Decision curve analysis was performed to determine whether the nomogram was clinically useful.

**Results:** The independent risk factors for MVI were maximum tumor length, rad-score, and miRNA-21 (all P < 0.001). The sensitivity, specificity, accuracy, and F1-score of the nomogram in the validation cohort were 0.970, 0.722, 0.884, and 0.916, respectively. The AUC of the nomogram was 0.900 (95% CI: 0.808–0.992) in the validation cohort, higher than that of any other single factor model (maximum tumor length, rad-score, and miRNA-21).

**Conclusion:** The radiogenomics nomogram shows satisfactory predictive performance in predicting MVI in HCC and provides a feasible and practical reference for tumor treatment decisions.

## KEYWORDS

hepatocellular carcinoma (HCC), microvascular invasion (MVI), radiogenomics, nomogram, MicroRNAs, dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI)



## Introduction

Hepatocellular carcinoma (HCC) is a malignancy with third highest world mortality rate (8.3%), after lung cancer (18%) and colorectal cancer (9.4%) (1). The main treatment of HCC is surgical resection, but recurrence is common, with a five-year recurrence rate of up to 40%–70%. The 5-year survival rate is only 18% (2). Microvascular invasion (MVI) is one of the most important prognostic factors for HCC after surgical treatment, and it has been established as a risk factor for early recurrence and poor outcome. The term MVI refers to the cancer cell nests that are established within the lining of blood vessels by microscopic endothelial cells. It is considered as a sign of strong tumor invasion ability, and it could only be diagnosed through pathology at present. MVI-positive patients often require improved prognosis by expanding surgical margins, and patients with microvascular infiltration are also considered unsuitable for liver transplantation (3–5). Therefore, developing a method for non-invasive prediction of microvascular invasion is necessary to guide the treatment of HCC.

The imaging characteristics of HCC, such as a non-smooth tumor margin, arterial peritumoral enhancement, and peritumoral hypo-intensity on hepatobiliary phase imaging (HBP), have been confirmed to be noninvasive imaging biomarkers for MVI prediction (4–7). However, such qualitative studies are vulnerable to subjective factors, image quality, and interobserver variation. Therefore, more objective quantitative methods are needed to predict MVI. In 2012, Lambin et al. (8) proposed the concept of radiomics, where medical images are converted into useful data by using high-throughput quantitative features to predict the disease treatment efficacy and prognosis. Using radiomics to predict MVI in HCC is a major research area in recent years. Many studies with satisfactory results have been conducted (9–11). Xu's (10) study has achieved predicted satisfactory results by developing a radiomics nomogram model on the basis of computed tomography (CT). MRI examination technology has the advantage of multimodal/multisequence imaging and high soft-tissue resolution. Theoretically MRI multicolumn multimodal imaging provides more characteristic elements. Therefore, in the present study, predictive models based on radiomic features in MRI were developed for predicting MVI.

MicroRNAs (miRNAs) are a kind of endogenous, non-coding RNAs. Thousands of miRNAs play a role in regulating various molecular biological processes by inhibiting the translation of different messenger RNAs (mRNAs) in the cell (12). A dysregulation of miRNAs is often associated with malignancy, and it regulates the proliferation, migration, invasion, and development of tumors in HCC by promoting or suppressing them (13, 14). Previous studies have shown that combining radiomics and genomics could remarkably improve the performance of predictive models (15, 16). Zhou (17) et al. screened 7 plasma miRNAs (miRNAs) out of 723 HCC-associated miRNAs (miR-122, miR-192, miR-21, miR-223, miR-26a, miR-27a, miR-801), which had high diagnostic performance in the early diagnosis of hepatocellular carcinoma. Therefore, we extracted these 7 mi-RNAs from the patients' plasma, but among

them, miR-192 and miR-801 had large differences in expression, and the data were not stable enough to be screened out, so only 5 mi-RNAs (miR-122, miR-21, miR-223, miR-26a, miR-27a) were included in the analysis. Therefore, we measured these miRNAs and explored their relationship with HCC microvascular invasion, and combined them with radiomics to explore the performance of the joint model. Moreover, routine laboratory tests for HCC and radiological characteristics based on MRI were added; the independent risk factors for MVI were determined through multivariate logistic regression, combined radiomics, genomics, and clinico-radiological factors; predictive models were established; and the performance of these models was verified.

This study aimed to develop and validate a radiogenomics nomogram model for preoperative prediction of MVI in HCC. The nomogram is helpful for clinicians to assist in determining individual therapeutic strategies for patients with HCC.

## Materials and methods

### Patients

This retrospective study was approved by the institutional review board, with a waiver for patient informed consent. We included all patients who underwent preoperative MRI and plasma mi-RNAs between December 2018 and November 2021. The inclusion criteria were as follows: (a) all patients who underwent radical hepatectomy with postoperative pathologic confirmation of hepatocellular carcinoma and complete clinical data; (b) MRI examination and plasma miRNA testing within two weeks prior to surgery; (c) Histopathology report containing a complete description of hepatocellular carcinoma (tumor size, number, MVI status and category, etc.); (d) The images were free of artifacts, sequence loss, and high image quality, meeting the basic requirements for image segmentation. The exclusion criteria were as follows: (a) the patient underwent any form of anticancer treatment (surgery, drugs, etc.) before surgery; (b) vascular or vascular invasion or the presence of distant metastasis was detected by the naked eye in preoperative imaging; (c) combined with other primary tumors.

Then, 168 patients (142 males and 26 females) comprised the final cohort. The included patients were divided into training ( $n = 116$ ; 100 males and 16 females) and validation cohorts ( $n = 52$ ; 43 males and 9 females), with a ratio of 7:3. The flowchart of patient enrollment and grouping in Figure 1. All patients received routine laboratory tests and plasma miRNA examinations prior to curative resection. Further information on the patients is available in Table 1.

### Histopathological examination

Seven-point baseline sampling method was used to take 1:1 samples at the junction between the cancer and the paracancerous liver tissues at the clock positions of twelve, three, six, and nine points of the tumor. At least one piece of tissue was taken inside the

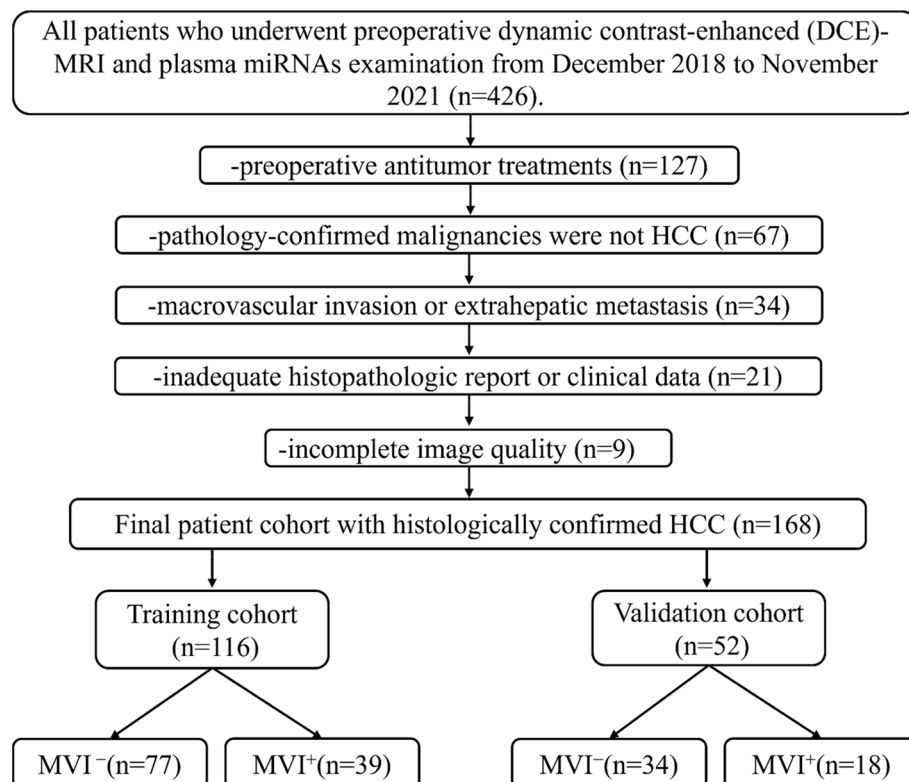


FIGURE 1  
Flowchart of patient enrollment and grouping.

tumor, and one piece of liver tissue was taken  $\leq 1$  cm (near paracancerous) and  $> 1$  cm (distal paracancerous) from the tumor margin (18). Histopathological features (tumor size, number, MVI status, and category) were consistently assessed by two experienced abdominal pathologists.

## Collection of plasma samples and miRNA extraction

Venous blood samples were collected from all patients with HCC prior to any means of processing. Before the sample collection was conducted, a written consent was obtained for each patient to donate a sample for the purpose of the study. For specific steps on collection of plasma samples and miRNA extraction, please refer to the [Supplementary Materials 1](#).

## MRI examination

MRI examinations were conducted using a GE DISCOVERY 750W 3.0 T MRI scanner, with axial in-phase and opposed-phase T1 weighted imaging (T1WI), axial T2-weighted imaging with fat suppression (T2WI-FS), diffusion-weighted imaging (DWI), and DCE-MRI (dynamic contrast-enhanced magnetic resonance imaging) sequences for all patients. Please refer to [Supplementary Material 2](#) for specific MRI parameters.

## Analysis of radiological characteristics

The radiological characteristics were independently evaluated by two abdominal radiologists A (RA) and B (RB), with 7 and 15 years of experience, respectively. If any differences occurred, senior radiologist C (RC) with 20 years of experience would join the discussion to reach a consensus. All radiologists were aware that the lesions were HCCs but blinded to all other laboratory and histopathological information. The largest tumor was used to analyze patients with multiple tumors. The evaluation was based on the Liver Imaging Reporting and Data System (LI-RADS version 2018) (19), and the important morphological features reported in the relevant literature (7). The qualitative features of the images were assessed refer to [Supplementary Material 3](#).

## Analysis of radiomics

### Image segmentation

HCC image segmentation was performed by RA and RB with the use of three-dimensional (3D) slicer software (version 5.0.2). The volumes of interest (VOIs) were delineated in the axial T2WI-FS, DWI (with b value of  $800 \text{ s/mm}^2$ ), AP, PP, and DP images. For assessment of the reproducibility and reliability of image segmentation, images of 30 randomly selected patients were first segmented by RA and RB separately. Then, 30 patients were re-segmented by RB after 2 weeks, and the images of the remaining patients were segmented by RA. The segmentation results were validated by RC.

TABLE 1 Comparisons of patient characteristics in training and validation cohorts.

| Characteristics      | Training cohort         |                         | P       | Validation cohort       |                         | P       |
|----------------------|-------------------------|-------------------------|---------|-------------------------|-------------------------|---------|
|                      | MVI <sup>+</sup> (n=39) | MVI <sup>-</sup> (n=77) |         | MVI <sup>+</sup> (n=18) | MVI <sup>-</sup> (n=34) |         |
| Age (years)          | 58.97 ± 8.49            | 59.48 ± 8.90            | 0.769   | 58.89 ± 10.22           | 56.03 ± 11.33           | 0.375   |
| Sex                  |                         |                         |         |                         |                         |         |
| Male                 | 29 (74.36%)             | 33 (84.62%)             | 0.726   | 13 (72.22%)             | 29 (85.29%)             | 0.304   |
| Female               | 10 (25.64%)             | 6 (15.38%)              |         | 5 (27.78%)              | 5 (14.71%)              |         |
| HBV                  |                         |                         |         |                         |                         |         |
| Present              | 38 (97.44%)             | 73 (94.81%)             | 0.514   | 18 (100.0%)             | 32 (94.12%)             | 0.160   |
| Absent               | 1 (2.56%)               | 4 (5.19%)               |         | 0(0.00%)                | 2 (5.88%)               |         |
| AFP (ng/mL)          |                         |                         |         |                         |                         |         |
| ≤ 20                 | 19 (48.7%)              | 39 (50.6%)              | 0.079   | 11 (61.1%)              | 15 (44.1%)              | 0.383   |
| 20–400               | 8 (20.5%)               | 22 (28.6%)              |         | 2 (11.1%)               | 12 (35.3%)              |         |
| >400                 | 12 (30.8%)              | 16 (20.8%)              |         | 5 (27.8%)               | 7 (20.6%)               |         |
| ALT (μl)             |                         |                         |         |                         |                         |         |
| ≤ 40                 | 25 (64.1%)              | 55 (71.4%)              | 0.318   | 10 (55.6%)              | 23 (67.6%)              | 0.372   |
| >40                  | 14 (36.9%)              | 22 (28.6%)              |         | 8 (44.4%)               | 11 (32.4%)              |         |
| AST (μl)             |                         |                         |         |                         |                         |         |
| ≤ 35                 | 22 (56.4%)              | 42 (45.5%)              | 0.364   | 10 (55.6%)              | 22 (64.7%)              | 0.415   |
| >35                  | 17 (43.6%)              | 35 (54.5%)              |         | 8 (44.4%)               | 12 (35.3%)              |         |
| TBIL (μmol/l)        |                         |                         |         |                         |                         |         |
| ≤ 20                 | 26 (66.7%)              | 53 (68.8%)              | 0.379   | 12 (66.7%)              | 22 (64.7%)              | 0.418   |
| >20                  | 13 (33.3%)              | 24 (31.2%)              |         | 6 (33.3%)               | 12 (35.3%)              |         |
| ALB (g/l)            |                         |                         |         |                         |                         |         |
| ≤ 40                 | 17 (43.6%)              | 42 (54.5%)              | 0.635   | 9 (50.0%)               | 19 (55.9%)              | 0.869   |
| >40                  | 22 (56.4%)              | 35 (45.4%)              |         | 9 (50.0%)               | 15 (44.1%)              |         |
| PIVKA-II (mAu/mL)    |                         |                         |         |                         |                         |         |
| ≤ 40                 | 4 (10.3%)               | 11 (14.3%)              | 0.910   | 4 (22.2%)               | 5 (14.7%)               | 0.144   |
| >40                  | 35 (89.7%)              | 66 (85.7%)              |         | 14 (77.8%)              | 29 (85.3%)              |         |
| PT (s)               |                         |                         |         |                         |                         |         |
| ≤ 14                 | 35 (89.7%)              | 68 (88.3%)              | 0.941   | 17 (94.4%)              | 29 (85.3%)              | 0.679   |
| >14                  | 4 (10.3%)               | 9 (11.7%)               |         | 1 (5.6%)                | 5 (14.7%)               |         |
| INR                  |                         |                         |         |                         |                         |         |
| ≤ 1.0                | 18 (46.2%)              | 18 (23.4%)              | 0.382   | 8 (44.4) %              | 8 (23.5%)               | 0.376   |
| >1.0                 | 21 (53.8%)              | 59 (76.6%)              |         | 10 (55.6%)              | 26 (76.5%)              |         |
| MiRNA-21             | 29.60 ± 1.40            | 31.32 ± 0.85            | < 0.001 | 29.18 ± 1.19            | 31.13 ± 0.76            | < 0.001 |
| MiRNA-26a            | 32.08 ± 1.67            | 30.27 ± 1.69            | < 0.001 | 32.32 ± 1.53            | 30.40 ± 1.60            | < 0.001 |
| MiRNA-27a            | 27.35 ± 2.04            | 28.51 ± 0.95            | 0.002   | 26.50 ± 1.97            | 28.46 ± 1.03            | < 0.001 |
| MiRNA-122            | 28.96 ± 1.62            | 29.89 ± 0.95            | 0.001   | 28.35 ± 1.48            | 29.81 ± 0.97            | < 0.001 |
| MiRNA-223            | 31.88 ± 1.50            | 31.13 ± 1.06            | 0.001   | 32.46 ± 1.30            | 31.25 ± 1.07            | < 0.001 |
| Maximum tumor length | 6.43 ± 3.36             | 5.16 ± 2.58             | 0.044   | 6.11 ± 2.88             | 5.24 ± 3.05             | 0.320   |
| Tumor margin         |                         |                         |         |                         |                         |         |
| Smooth               | 3 (7.7%)                | 23 (29.9%)              | 0.007   | 1 (5.6%)                | 12 (35.3%)              | 0.043   |
| Non-smooth           | 36 (92.3%)              | 54 (70.1%)              |         | 17 (94.4%)              | 22 (64.7%)              |         |
| Number               |                         |                         |         |                         |                         |         |
| =1                   | 36 (92.3%)              | 68 (88.3%)              | 0.730   | 16 (88.9%)              | 31 (91.2%)              | 1.000   |
| >1                   | 3 (7.7%)                | 9 (11.7%)               |         | 2 (11.1%)               | 3 (8.8%)                |         |
| Enhancement pattern  |                         |                         |         |                         |                         |         |
| Typical              | 35 (89.7%)              | 73 (94.8%)              | 0.530   | 2 (11.1%)               | 32 (94.1%)              | 0.900   |
| Atypical             | 4 (10.3%)               | 4 (5.2%)                |         | 16 (88.9%)              | 2 (5.9%)                |         |
| Radiologic capsule   |                         |                         |         |                         |                         |         |
| Present              | 7 (17.9%)               | 40 (51.9%)              | < 0.001 | 3 (16.7%)               | 20 (58.8%)              | 0.004   |
| Absent               | 32 (82.1%)              | 37 (48.1%)              |         | 15 (83.3%)              | 14 (41.2%)              |         |

(Continued)

TABLE 1 Continued

| Characteristics                  | Training cohort         |                         | <i>P</i> | Validation cohort       |                         | <i>P</i> |
|----------------------------------|-------------------------|-------------------------|----------|-------------------------|-------------------------|----------|
|                                  | MVI <sup>+</sup> (n=39) | MVI <sup>−</sup> (n=77) |          | MVI <sup>+</sup> (n=18) | MVI <sup>−</sup> (n=34) |          |
| Arterial peritumoral enhancement |                         |                         |          |                         |                         |          |
| Present                          | 36 (92.3%)              | 44 (57.1%)              | < 0.001  | 17 (94.4%)              | 15 (44.1%)              | 0.001    |
| Absent                           | 3 (7.7%)                | 33 (42.9%)              |          | 1 (5.6%)                | 19 (55.9%)              |          |
| Intratumor necrosis/hemorrhage   |                         |                         |          |                         |                         |          |
| Present                          | 33 (84.6%)              | 45 (58.4%)              | 0.005    | 17 (94.4%)              | 19 (55.9%)              | 0.011    |
| Absent                           | 6 (15.4%)               | 32 (41.6%)              |          | 1 (5.6%)                | 15 (44.1%)              |          |

HBV, hepatitis B virus; AFP, serum alpha-fetoprotein; ALT, alanine aminotransferase; AST, aspartate amino transferase; TBIL, total bilirubin; ALB, albumin; PIVKA-II, protein induced by vitamin K absence or antagonist-II; PT, prothrombin time; INR, international normalized ratio; MVI, microvascular invasion.

### Radiomics feature extraction, selection, and signature building

A total of 7045 radiomic features were extracted from each segmented lesion using the SlicerRadiomics plugin in 3DSlicer. Python (version 2.7.18) was used for radiological feature selection. These features included shape, first-order histogram features and texture features. The 1856 features with intra- and inter-correlation coefficients (ICCs) values less than 0.8 were firstly excluded, and the remaining features were initially screened by SelectKBest. The remaining features were selected by least absolute shrinkage and selection operator (LASSO) algorithms. The features of the LASSO regression result in which the corresponding coefficients with non-zero were retained. 10-fold cross-validation was performed to select the optimal  $\alpha$  value, and the coefficients of the corresponding radiomics features were obtained at the same time. The radiomic feature score (rad-score) reflecting the MVI was calculated for each patient by using a linear combination of the selected features weighted with the respective coefficients.

### Model construction, evaluation, and comparison

All variables (laboratory tests, miRNA, radiological characteristics, and radiation scores) were first screened by univariate analysis, and then independent risk factors for MVI were determined by stepwise backward regression with the principle of minimum AIC (Akaike information criterion) value by multivariate logistic regression analysis. All the independent risk factors were used separately to build the corresponding prediction models and construct the nomogram. The ROC curves were plotted, the discriminant efficiency of MVI predictions was quantified using AUC, and multiple comparisons between different models were carried out by Delong test. The 95% CI of AUCs, sensitivity, specificity, and accuracy were also calculated. F1-score was used to evaluate a binary classification model with unbalanced data samples. The clinical utility of the nomogram was evaluated using decision curve analysis, which quantifies the net benefit to the overall cohort at different threshold probabilities (20). The process of the present study is illustrated in Figure 2.

### Statistical analysis

Statistical analysis was performed using R software (version 3.4.1). Continuous variables were expressed as mean  $\pm$  standard

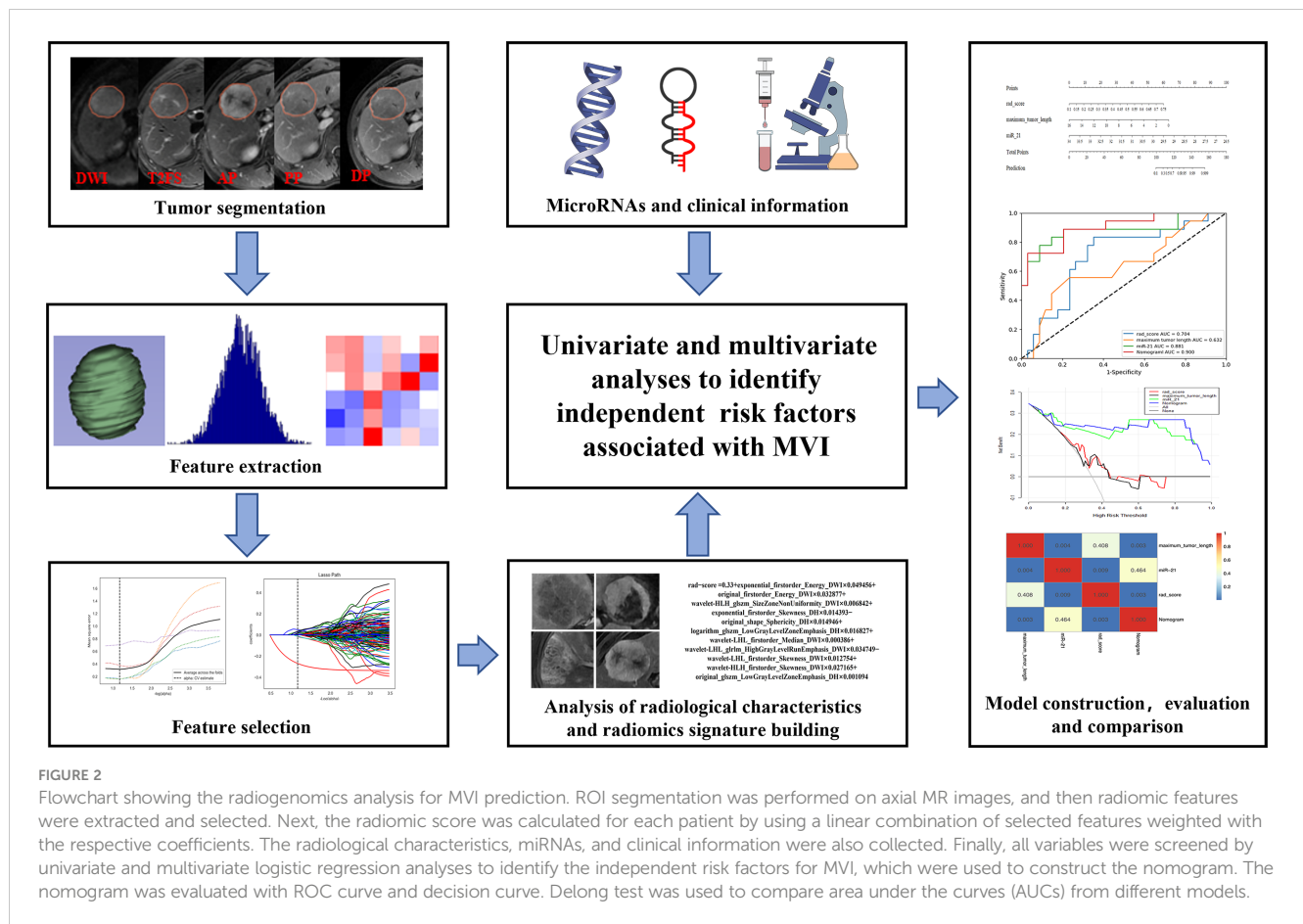
deviation. The categorical variables were presented as percentages. Kolmogorov–Smirnov tests were used to evaluate the distribution's normality. For identification of variables that differed significantly between the training and validation cohorts, Student's *t* test was used to compare the quantitative data, and Chi-square test or Fisher's exact test was used to compare the qualitative data. The reproducibility of the feature extraction was assessed by calculating the intra- and inter-correlation coefficients (ICCs), and ICCs > 0.80 were considered to have good reproducibility. The Hosmer–Lemeshow's goodness-of-fit test was used to evaluate whether the model's predicted probabilities fitted the actual probabilities. The sensitivity, specificity, and accuracy were calculated by confusion matrix in accordance with the cutoff value that maximized the Youden index. Statistical significance was set at *P* < 0.05.

## Results

### Clinico-radiological characteristics and MVI prediction factors

A comparison of the clinico-radiological characteristics is shown in Table 1. Among the 168 patients with HCC, MVI was diagnosed in the resected tissue of 57 patients. The comparison between the training and validation cohorts was not statistically different in terms of age, gender, AFP, and other clinical indicators (*P* = 0.144–0.941). All five miRNAs significantly differed between MVI<sup>+</sup> and MVI<sup>−</sup> in the training and validation cohorts (*P* < 0.05). The patients with MVI<sup>+</sup> and MVI<sup>−</sup> also showed significantly different imaging characteristics (tumor margin, radiologic capsule, arterial peritumoral enhancement, and intratumor necrosis/hemorrhage) (*P* < 0.05). No significant differences were found in the tumor number and enhancement pattern between MVI<sup>+</sup> and MVI<sup>−</sup> in either the training cohort or the validation cohort (*P* = 0.530–1.000).

The univariate analysis showed that five imaging features (maximum tumor length, tumor margin, radiologic capsule, peri-arterial tumor enhancement, and presence of hemorrhage and necrosis) and the five kinds of miRNAs (miR-21, miR-26a, miR-27a, miR-122, and miR-223) were significantly associated with MVI (*P* < 0.05). In the multivariate analysis, maximum tumor length and miR-21 were found to be independent predictors of MVI. The specific information is shown in Table 2.



## Feature selection and radiomics signature building

Radiomics features were downsampled by SelectKBest and LASSO, resulting in a final selection of 11 features, all of which were derived from DWI and DP sequences. The ICCs ranged from 0.856 to 0.989 for the intra-observers and from 0.843 to 0.982 for the inter-observers. These values demonstrated the high reliability of the measurements taken by the observers. A linear combination of the selected features, weighted by their respective logistic regression coefficients, was used to generate the rad-score (risk score reflecting the probability of MVI). This score was used to calculate each selected VOI as follows:

$$\begin{aligned} \text{rad-score} = & 0.33 + \text{exponential\_firstorder\_Energy\_DWI} \times 0.049456 + \text{original\_firstorder\_Energy\_DWI} \times 0.032877 + \\ & \text{wavelet-HLH\_glszm\_SizeZoneNonUniformity\_DWI} \times 0.006842 + \text{exponential\_firstorder\_Skewness\_DH} \times 0.014393 - \\ & \text{original\_shape\_Sphericity\_DH} \times 0.014946 + \\ & \text{logarithm\_glszm\_LowGrayLevelZoneEmphasis\_DH} \times 0.016827 + \\ & \text{wavelet-LHL\_firstorder\_Median\_DWI} \times 0.000386 + \\ & \text{wavelet-LHL\_glrlm\_HighGrayLevelRunEmphasis\_DWI} \times 0.034749 - \\ & \text{wavelet-LHL\_firstorder\_Skewness\_DWI} \times 0.012754 + \\ & \text{wavelet-HLH\_firstorder\_Skewness\_DWI} \times 0.027165 + \\ & \text{original\_glszm\_LowGrayLevelZoneEmphasis\_DH} \times 0.001094 \end{aligned}$$

The univariate and multifactorial regression analyses showed that the rad-score is an independent risk factor for MVI (Table 2).

## Model construction and evaluation

The independent predictive factors of MVI, which were maximum tumor length, miR-21, and rad-score, were identified by univariate and multivariate logistic regression methods. The MVI prediction model incorporated these three independent risk factors to develop a nomogram prediction model (Figure 3).

The three single-factor models of tumor maximum length, miR-21, and rad-score reached AUC values of 0.658 (95% CI: 0.551–0.764), 0.907 (95% CI: 0.866–0.949), and 0.836 (95% CI: 0.763–0.909) in the training cohort, respectively, and 0.632 (95% CI: 0.465–0.799), 0.881 (95% CI: 0.763–0.998), and 0.704 (95% CI: 0.551–0.857) in the validation cohort, respectively. The nomogram model had an AUC of 0.900 (95% CI: 0.808–0.992) in the validation cohort, with sensitivity, specificity, accuracy, and F1-score of 0.970, 0.722, 0.884, and 0.916, respectively (Figures 4A, B, Table 3). The Hosmer–Lemeshow’s goodness-of-fit test evaluated the model performance at  $P = 0.55 > 0.05$ , indicating that the actual value of the prediction model fitted well with the predicted value. The decision curve showed the clinical usefulness of the different



TABLE 2 Univariate and multivariate analysis to identify risk factors associated with MVI in the training cohort.

| Variables                        | Univariate analysis |          | Multivariate analysis |          |
|----------------------------------|---------------------|----------|-----------------------|----------|
|                                  | OR (95% CI)         | <i>P</i> | OR (95% CI)           | <i>P</i> |
| Sex                              | 0.95 (0.74–1.23)    | 0.726    | NA                    | NA       |
| Age                              | 1.00 (0.99–1.01)    | 0.769    | NA                    | NA       |
| HBV                              | 1.15 (0.75–1.77)    | 0.514    | NA                    | NA       |
| AFP                              | 1.00 (0.99–1.01)    | 0.198    | NA                    | NA       |
| PIVKA-II                         | 1.00 (0.99–1.00)    | 0.906    | NA                    | NA       |
| ALT                              | 0.99 (0.99–1.00)    | 0.318    | NA                    | NA       |
| AST                              | 0.99 (0.99–1.00)    | 0.364    | NA                    | NA       |
| TB                               | 0.98 (0.99–1.02)    | 0.379    | NA                    | NA       |
| ALB                              | 1.01 (0.99–1.13)    | 0.641    | NA                    | NA       |
| PT                               | 0.98 (0.94–1.06)    | 0.941    | NA                    | NA       |
| INR                              | 0.96 (0.87–1.05)    | 0.382    | NA                    | NA       |
| MiRNA-21                         | 0.81 (0.76–0.85)    | < 0.001  | 0.73 (0.66–0.78)      | < 0.001  |
| MiRNA-26a                        | 1.12 (1.08–1.17)    | < 0.001  | 1.05 (0.99–1.11)      | 0.065    |
| MiRNA-27a                        | 0.89 (0.84–0.94)    | < 0.001  | 0.95 (0.81–1.11)      | 0.518    |
| MiRNA-122                        | 0.88 (0.83–0.94)    | 0.002    | 1.04 (0.74–1.44)      | 0.840    |
| MiRNA-223                        | 1.11 (1.04–1.19)    | < 0.001  | 0.86 (0.68–1.09)      | 0.220    |
| Maximum tumor length             | 1.05 (1.02–1.07)    | < 0.001  | 0.90 (0.87–0.93)      | < 0.001  |
| Tumor margin                     | 1.33 (1.09–1.07)    | < 0.001  | 0.99 (0.83–1.18)      | 0.478    |
| Number                           | 0.91 (0.68–1.21)    | 0.509    | NA                    | NA       |
| Enhancement pattern              | 0.84 (0.60–1.18)    | 0.314    | NA                    | NA       |
| Radiologic capsule               | 0.73 (0.62–0.86)    | < 0.001  | 0.92 (0.79–1.07)      | 0.788    |
| Arterial peritumoral enhancement | 1.44 (1.21–1.71)    | < 0.001  | 1.10 (0.93–1.29)      | 0.329    |
| Intratumor necrosis/hemorrhage   | 1.30 (1.09–1.56)    | < 0.001  | 0.93 (0.81–1.08)      | 0.915    |
| Rad-score                        | 9.33 (4.65–18.74)   | < 0.001  | 7.92 (3.78–16.6)      | < 0.001  |

HBV, hepatitis B virus; AFP, serum alpha-fetoprotein; ALT, alanine aminotransferase; AST, aspartate amino transferase; TBIL, total bilirubin; ALB, albumin; PIVKA-II, protein induced by vitamin K absence or antagonist-II; PT, prothrombin time; INR, international normalized ratio; OR, odds ratio; CI, confidence interval. NA, not available.

models (Figures 4C, D). The prediction performance of the nomogram model was satisfactory in the validation cohort, with the decision curve shown in Figure 4D. The net benefit of predicting the decision curve for the nomogram and miR-21 model was higher than that for other models when the threshold probability was > 0%. This finding suggested that the nomogram and miR-21 models could achieve satisfactory net clinical benefits.

Model comparison

Among the three single-factor models of maximum tumor length, miR-21, and rad-score, the miR-21 model performed best, and the differences with the other two one-factor models were all statistically significant in the validation cohort (miR-21 vs. maximum tumor length: AUC of 0.881 vs. 0.632, *P* = 0.004; miR-21 vs. rad-score: AUC of 0.881 vs. 0.704, *P* = 0.009). The nomogram

prediction model outperformed the miR-21 model (AUC of 0.900 vs. 0.881; *P* = 0.464), the rad-score model (AUC of 0.900 vs. 0.704; *P* = 0.003), and the maximum tumor length model (AUC of 0.900 vs. 0.632; *P* = 0.003) in the validation cohort. However, no statistical difference was found between the nomogram model and the miR-21 model (*P* = 0.464), as detailed in Table 3 and Figure 5.

Discussion

In this study, we have successfully developed and rigorously validated a multi-omics nomogram prediction model, which integrates MRI-derived radiomics, radiological features, and miRNA-based genomics. The resulting radiogenomic nomogram has demonstrated excellent performance in accurately predicting microvascular invasion in HCC, thus providing a non-invasive yet reliable clinical method for preoperative prediction.

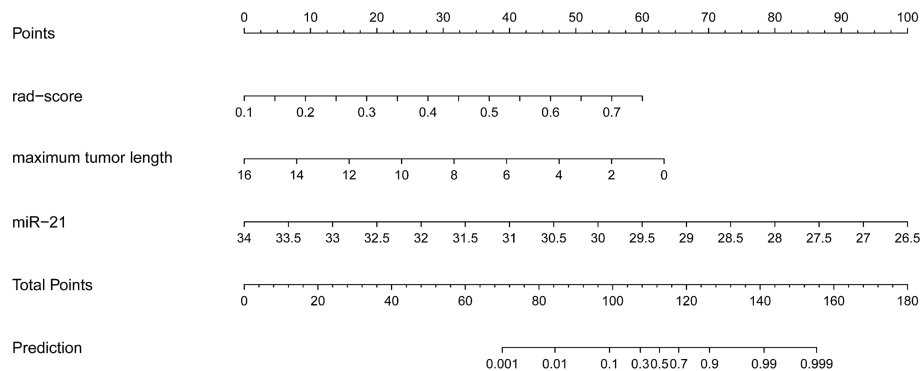


FIGURE 3

Radiogenomics nomogram for predicting MVI in hepatocellular carcinoma. (1) Factors in the prediction model: maximum tumor length, rad-score, and miRNA-21, the scale on the line segment corresponding to each factor represents the range of values of the factor, and the length of the line segment represents the contribution of the factor to the probability of occurrence of the outcome variable. (2) Points and total points: individual points represent the individual scores corresponding to each factor at different ranges of values, and the total points represents the sum of the individual scores corresponding to all the factors at different ranges of values. (3) Prediction: the scale value corresponding to the total points indicates the risk of microvascular invasion in patients with hepatocellular carcinoma.

The expression of miR-21, miR-27a, and miR-122 in the MVI<sup>+</sup> group was upregulated compared with that in the MVI<sup>-</sup> group, whereas the expression of miR-26a and miR-223 was downregulated. These differences were all statistically significant. By contrast, the differences in conventional laboratory indicators, such as AFP, were not statistically significant between the MVI<sup>+</sup> group and the MVI<sup>-</sup> group, indicating that the miRNAs extracted in the plasma of patients with HCC were more meaningful in suggesting MVI than the conventional laboratory indicators. Further univariate and multivariate analyses showed that miR-21 had better correlations than other clinical laboratory tests. Studies have shown that miR-21 is one of the most expressed miRNAs in liver diseases, such as nonalcoholic fatty liver disease (21). Ladeiro et al. (22) found that miR-21 was significantly overexpressed in HCC, as compared to benign tumor or non-neoplastic liver tissue. The maladjusted expression of miRNA could be used as a biomarker, and it could be detected in the plasma of patients. Studies have shown that miR-21 plays a role in promoting HCC growth invasion, distant metastasis, and other links (21). Furthermore, this study confirmed that among other miRNAs detected in plasma, only miR-21 emerged as an independent risk factor for MVI in HCC. The miR-21 model surpassed the rad score and maximum tumor length models in predicting MVI, and the differences were statistically significant in the validation cohort ( $P < 0.05$ ). These findings indicate that the miR-21 model performs well in preoperative prediction of MVI in HCC. Conventional clinical laboratory indicators, including tumor markers such as AFP, were all excluded in the univariate analysis due to their relatively poor correlation compared to other variables.

On the basis of the morphological characteristics of MRI, previous meta-analysis studies have found that some of MRI signs were significantly associated with MVI, including larger tumors ( $> 5$  cm), rim arterial enhancement, arterial peritumoral enhancement, non-smooth tumor margin, and multiple lesions (7). The present study incorporated these MR morphological features into the analysis and found that only the maximum tumor length is

an independent predictor of MVI in HCC. Increased tumor volume led to increased contact between the tumor and adjacent normal liver tissues, thereby promoting the formation of microvessels.

A total of 7045 features were extracted from five different MRI sequences, and then 11 radiomic features were screened out by SelectKBest and LASSO. Interestingly, these features were all derived from DWI and DP sequences. A previous study (23) found that primary radiomic signatures extracted from delayed-phase sequences were associated with MVI. Zhang et al. (24) compared the performance of different MRI sequences to predict MVI and found the performance of DP to be the best ( $AUC = 0.806$ ). The present study also found the value of DP to be the best in predicting MVI, which could be explained by the fact that tumors continue to release a large number of angiogenic factors that promote tumor angiogenesis and change tumor perfusion, resulting in the differences between MVI<sup>-</sup> and MVI<sup>+</sup> being more easily shown in DP sequences (25). DWI also has a satisfactory performance, as confirmed by some previous studies (26, 27). This finding could be explained that the minimum value of the apparent diffusion coefficient of DWI could reflect the densest tumor, the most abundant neovascularization, and the most active tumor proliferation. The hepatobiliary phase of specific contrast agents has been reported to make an important contribution to suggesting MVI (7). It is important to note that when hepatocyte-specific agents are used, DWI is usually scanned after contrast, and the ability and contribution of DWI may differ, which needs to be proven by further research.

After rigorous data analysis and model training, we have successfully developed a nomogram model that comprehensively incorporates various risk factors. In stringent tests using training and validation datasets, the model exhibited outstanding predictive performance, with AUC values reaching 0.942 and 0.900, significantly surpassing other single-factor models. The nomogram model demonstrated satisfactory performance in predicting microvascular invasion in HCC. Looking back at previous studies, although nomogram models based on radiomics

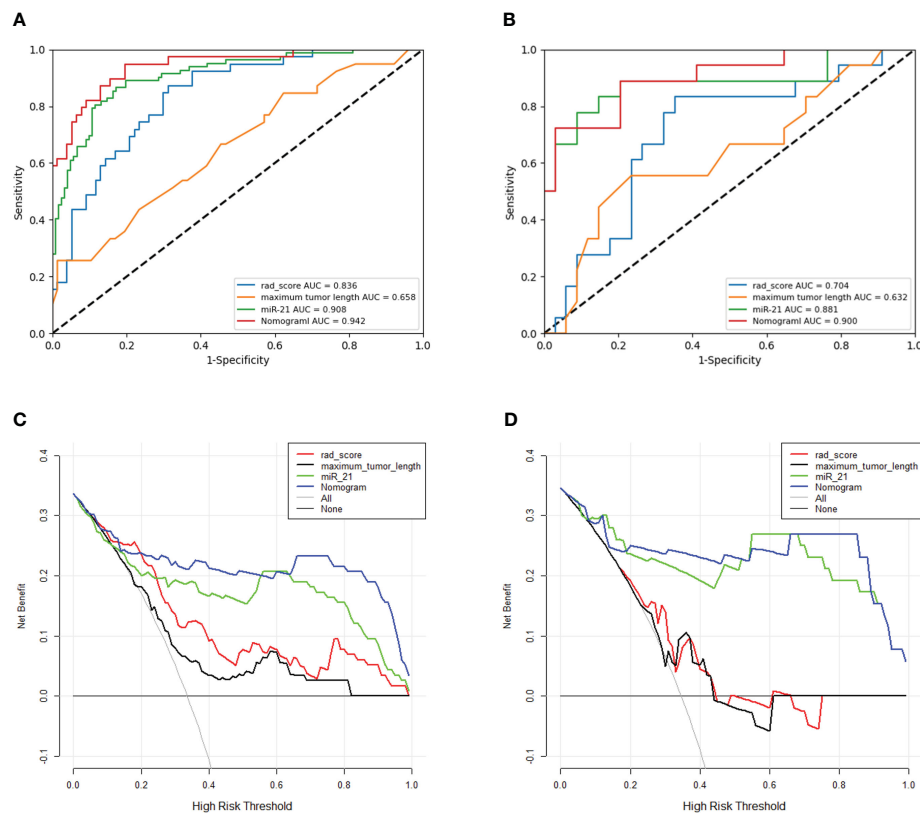


FIGURE 4

(A, B) ROC curves for different models in the training and validation cohorts. The ROC graph is a curve that reflects the relationship between sensitivity and specificity. According to the position of the curve, the whole graph is divided into two parts, the area below the curve is called AUC (Area Under Curve), which is used to indicate the prediction accuracy, the higher the value of AUC, that is, the larger the area under the curve, indicating that the prediction accuracy is higher. The closer the curve is to the upper left corner, the higher the prediction accuracy. Panels A, B represent the prediction values of different models for the training cohort and validation cohort, respectively. (C, D) Clinical utility is evaluated in terms of Decision curve analysis (DCA), which reflects the ability of a model to benefit patients by influencing clinical decisions. A good model should have a high net benefit value at the threshold required by its clinical question. The net benefit of predicting the decision curve for the nomogram and miR-21 model was higher than that for other models when the threshold probability was > 0%. This finding suggested that the nomogram and miR-21 models could achieve satisfactory net clinical benefits.

and clinical factors have shown promising predictive capabilities, most of them still have limitations in performance, with AUC values generally ranging from 0.801 to 0.861 (28–30). Notably, these models often involve numerous risk factors, whereas our nomogram model incorporates only three independent risk factors, highlighting the advantages of multi-omics approaches in data processing and model development.

In previous explorations, such as the study by Banerjee et al. (28), they delved into the radiogenomics of MVI in liver cancer, innovatively developing a novel imaging biomarker called radiogenomic venous invasion (RVI) by combining venous invasion genes in hepatocellular carcinoma with dynamic contrast-enhanced CT. This achievement has achieved significant results in predicting MVI and prognosis. Similarly, Taouli et al. (29) also conducted in-depth research on the imaging characteristics and genomic data of hepatocellular carcinoma, successfully identifying imaging features related to aggressive hepatocellular carcinoma genes through a combination of preoperative CT or MR examinations and transcriptomic analysis.

However, our study adopted a more unique and precise approach. We directly used miRNAs closely related to liver

cancer as variables in logistic regression analysis, screening out miRNAs species independently associated with MVI in liver cancer through rigorous univariate and multivariate analysis. Subsequently, we combined these crucial miRNAs with radiomics and clinical radiological features to construct an efficient and accurate prediction model, achieving satisfactory preoperative prediction of MVI. This achievement provides a powerful tool or method for the accurate prediction of microvascular invasion in hepatocellular carcinoma, guiding clinical decision-making, optimizing treatment plans, and ultimately improving the survival rate and quality of life of patients.

## Limitations

This study still has some limitations. First, it is a small, single-center study. Therefore, the results should be complemented by further validation from larger queues at other centers. Second, miRNAs have many types, and only a small number was detected in this study. Secondly, there are many types of miRNAs, only a small amount was detected in this study, although this part of

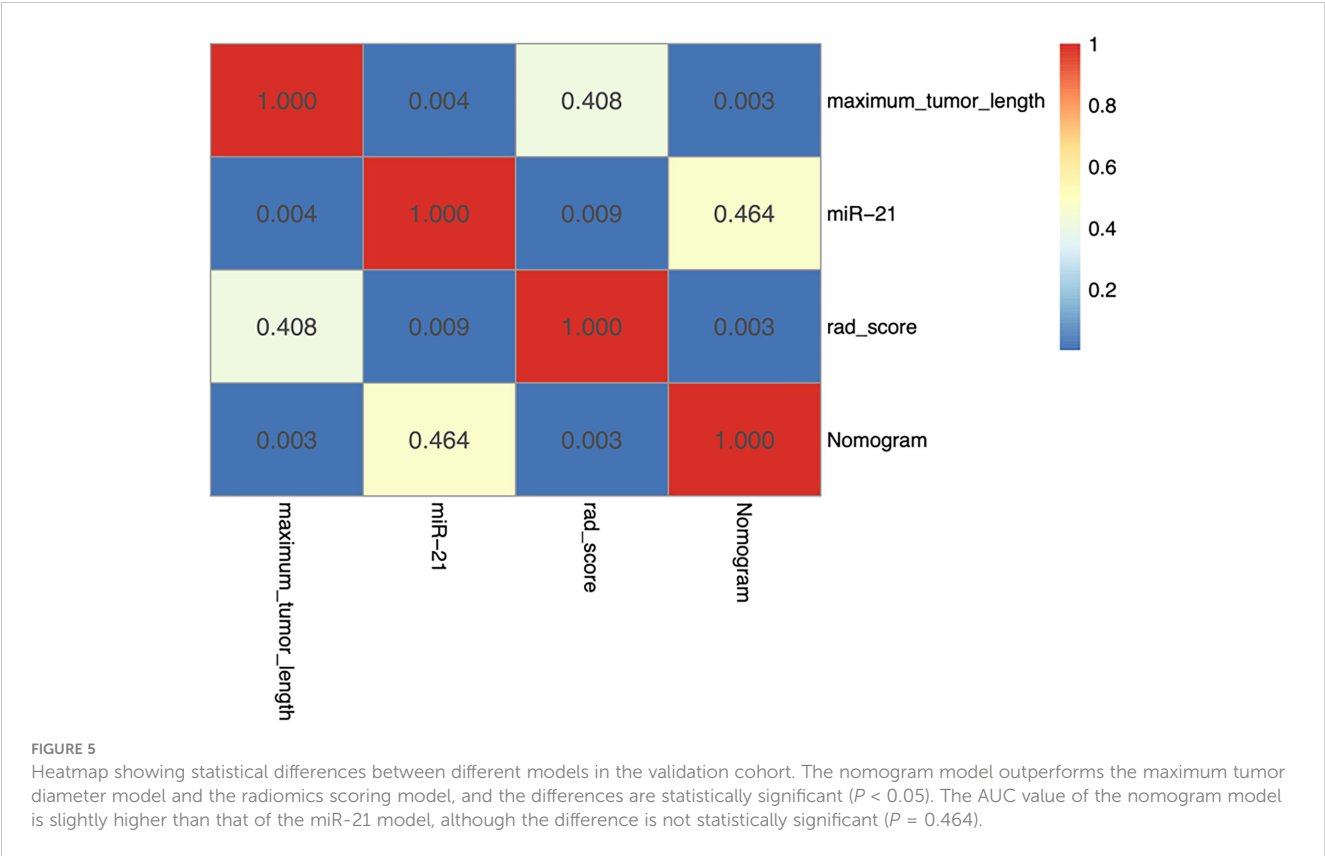
TABLE 3 Performance of different MVI prediction models.

| Models               | AUC                 | Sensitivity | Specificity | Accuracy | F1-score |
|----------------------|---------------------|-------------|-------------|----------|----------|
| Maximum tumor length |                     |             |             |          |          |
| Training cohort      | 0.658 (0.551–0.764) | 0.987       | 0.256       | 0.698    | 0.835    |
| Validation cohort    | 0.632 (0.465–0.799) | 0.764       | 0.555       | 0.692    | 0.764    |
| Rad-score            |                     |             |             |          |          |
| Training cohort      | 0.836 (0.763–0.909) | 0.688       | 0.871       | 0.750    | 0.785    |
| Validation cohort    | 0.704 (0.551–0.857) | 0.647       | 0.833       | 0.711    | 0.746    |
| MiRNA-21             |                     |             |             |          |          |
| Training cohort      | 0.907 (0.866–0.949) | 0.827       | 0.865       | 0.843    | 0.875    |
| Validation cohort    | 0.881 (0.763–0.998) | 0.911       | 0.777       | 0.865    | 0.898    |
| Nomogram             |                     |             |             |          |          |
| Training cohort      | 0.942 (0.899–0.985) | 0.805       | 0.948       | 0.853    | 0.832    |
| Validation cohort    | 0.900 (0.808–0.992) | 0.970       | 0.722       | 0.884    | 0.916    |

AUC, area under the ROC curve.

miRNAs has been shown to be associated with HCC caused by hepatitis B virus (HBV) (17), and the vast majority of patients we included are accompanied by HBV infection, but the effect of this data on HCC caused by non-HBV is unknown, so the results may only be valuable for HBV-associated HCC, in addition, whether there is genomic data with better performance than miR-21 needs to be further explored and verified. Third, miRNA is still a developing biomarker and is reported to have low reproducibility (30). Although we strictly follow standard procedures in the process of extracting miRNA, miRNA data stability is susceptible to a variety of factors such as limited

amount of analyte before analysis, cell contamination, risk of inhibition, etc., which may introduce some bias into the final result. It is believed that with the development of liquid biopsy technology, the reproducibility and stability of miRNA data will be improved, so as to be used for robust clinical prediction. Fourth, MVI involves the tumor edge, but only the internal characteristics of the tumor were analyzed, and the ROI outside the tumor, especially around the tumor, was not expanded. In some studies (10, 31, 32), radiological features were extracted by expanding the ROI, achieving good results. This method is also a part of the follow-up research that needs to be further improved.



## Conclusion

The radiogenomic nomogram exhibited promising preoperative predictive capabilities and clinical decision-making implications in forecasting microvascular invasion (MVI) in hepatocellular carcinoma (HCC). This model holds the potential to emerge as a biomarker for MVI in HCC in the future, though its efficacy necessitates further validation through extensive studies encompassing larger sample sizes from multiple centers.

## Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: Not publicly available due to privacy of laboratory data. Requests to access these datasets should be directed to QZ, [wfcyzhangqian@126.com](mailto:wfcyzhangqian@126.com).

## Ethics statement

The studies involving humans were approved by Institutional Review Board of Yantai Yuhuangding Hospital, Affiliated Hospital of Qingdao University. The studies were conducted in accordance with the local legislation and institutional requirements. The human samples used in this study were acquired from primarily isolated as part of your previous study for which ethical approval was obtained. 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

GH: Formal Analysis, Investigation, Methodology, Writing – original draft. JQ: Data curation, Writing – review & editing. JG: Data curation, Writing – review & editing. YC: Data curation, Visualization, Writing – review & editing. FW: Data curation, Formal Analysis, Writing – review & editing. HZ: Writing – review & editing. HZ: Conceptualization, Data curation, Methodology, Writing – review & editing. XW: Data curation,

Supervision, Writing – review & editing. HM: Investigation, Writing – review & editing. HX: Formal Analysis, Project administration, Writing – review & editing. CX: Supervision, Writing – review & editing. NL: Writing – review & editing, Software, Supervision. QZ: Funding acquisition, 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 National Natural Science Foundation of China (82001775) and the Natural Science Foundation of Shandong Province of China (ZR2021MH120) and Yantai Science and Technology Innovation Development Plan Project (2023YTYD06000894).

## 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/fonc.2024.1371432/full#supplementary-material>

## References

1. Sung H, Ferlay J, Siegel RL, 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
2. Villanueva A. Hepatocellular carcinoma. *N Engl J Med.* (2019) 380:1450–62. doi: 10.1056/NEJMr1713263
3. Kim AY, Sinn DH, Jeong WK, et al. Hepatobiliary MRI as novel selection criteria in liver transplantation for hepatocellular carcinoma. *J Hepatol.* (2018) 68:1144–52. doi: 10.1016/j.jhep.2018.01.024
4. Huang M, Liao B, Xu P, et al. Prediction of microvascular invasion in hepatocellular carcinoma: Preoperative Gd-EOB-DTPA-Dynamic enhanced MRI and histopathological correlation. *Contrast Media Mol Imaging.* (2018) 2018:9674565. doi: 10.1155/2018/9674565
5. Lee S, Kim SH, Lee JE, et al. Preoperative gadoxetic acid-enhanced MRI for predicting microvascular invasion in patients with single hepatocellular carcinoma. *J Hepatol.* (2017) 67:526–34. doi: 10.1016/j.jhep.2017.04.024
6. Ryu T, Takami Y, Wada Y, et al. A clinical scoring system for predicting microvascular invasion in patients with hepatocellular carcinoma within the milan criteria. *J Gastrointest Surg.* (2019) 23:779–87. doi: 10.1007/s11605-019-04134-y
7. Hong SB, Choi SH, Kim SY, et al. MRI features for predicting microvascular invasion of hepatocellular carcinoma: A systematic review and Meta-Analysis. *Liver Cancer.* (2021) 10:94–106. doi: 10.1159/000513704
8. Lambin P, Rios-Velazquez E, Leijenaar R, 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



9. Yao Z, Dong Y, Wu G, et al. Preoperative diagnosis and prediction of hepatocellular carcinoma: Radiomics analysis based on multi-modal ultrasound images. *BMC Cancer*. (2018) 18:1089. doi: 10.1186/s12885-018-5003-4
10. Xu X, Zhang HL, Liu QP, et al. Radiomic analysis of contrast-enhanced CT predicts microvascular invasion and outcome in hepatocellular carcinoma. *J Hepatol*. (2019) 70:1133–44. doi: 10.1016/j.jhep.2019.02.023
11. Harding-Theobald E, Louissaint J, Maraj B, et al. Systematic review: Radiomics for the diagnosis and prognosis of hepatocellular carcinoma. *Aliment Pharmacol Ther*. (2021) 54:890–901. doi: 10.1111/apt.16563
12. Oura K, Morishita A, Masaki T. Molecular and functional roles of MicroRNAs in the progression of hepatocellular Carcinoma-A review. *Int J Mol Sci*. (2020) 21. doi: 10.3390/ijms21218362
13. Esquela-Kerscher A, Slack FJ. Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer*. (2006) 6:259–69. doi: 10.1038/nrc1840
14. Bartel DP. MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell*. (2004) 116:281–97. doi: 10.1016/s0092-8674(04)00045-5
15. Maggio I, Franceschi E, Gatto L, et al. Radiomics, mirnomics, and radiomirRNomics in glioblastoma: Defining tumor biology from shadow to light. *Expert Rev Anticancer Ther*. (2021) 21:1265–72. doi: 10.1080/14737140.2021.1971518
16. Lin P, Wen DY, Chen L, et al. A radiogenomics signature for predicting the clinical outcome of bladder urothelial carcinoma. *Eur Radiol*. (2020) 30:547–57. doi: 10.1007/s00330-019-06371-w
17. Zhou J, Yu L, Gao X, et al. Plasma microRNA panel to diagnose hepatitis B virus-related hepatocellular carcinoma. *J Clin Oncol*. (2011) 29:4781–8. doi: 10.1200/JCO.2011.38.2697
18. Chen L, Chen S, Zhou Q, et al. Microvascular invasion status and its survival impact in hepatocellular carcinoma depend on tissue sampling protocol. *Ann Surg Oncol*. (2021) 28:6747–57. doi: 10.1245/s10434-021-09673-w
19. Chernyak V, Fowler KJ, Kamaya A, et al. Liver imaging reporting and data system (LI-RADS) version 2018: Imaging of hepatocellular carcinoma in At-Risk patients. *Radiology*. (2018) 289:816–30. doi: 10.1148/radiol.2018181494
20. Vickers AJ, Elkin EB. Decision curve analysis: A novel method for evaluating prediction models. *Med Decis Making*. (2006) 26:565–74. doi: 10.1177/0272989X06295361
21. Lai CY, Yeh KY, Lin CY, et al. Correction: Lai et al. MicroRNA-21 Plays Multiple Oncometabolic Roles in the Process of NAFLD-Related Hepatocellular Carcinoma via PI3K/AKT, TGF-beta, and STAT3 Signaling. *Cancers* 2021. (2022) 13:940. doi: 10.3390/cancers14020372
22. Ladeiro Y, Couchy G, Balabaud C, et al. MicroRNA profiling in hepatocellular tumors is associated with clinical features and oncogene/tumor suppressor gene mutations. *Hepatology*. (2008) 47:1955–63. doi: 10.1002/hep.22256
23. Wilson GC, Cannella R, Fiorentini G, et al. Texture analysis on preoperative contrast-enhanced magnetic resonance imaging identifies microvascular invasion in hepatocellular carcinoma. *HPB (Oxford)*. (2020) 22:1622–30. doi: 10.1016/j.hpb.2020.03.001
24. Zhang Y, Shu Z, Ye Q, et al. Preoperative prediction of microvascular invasion in hepatocellular carcinoma via Multi-Parametric MRI radiomics. *Front Oncol*. (2021) 11:633596. doi: 10.3389/fonc.2021.633596
25. Lee Y, Lee SS, Kim N, et al. Intravoxel incoherent motion diffusion-weighted MR imaging of the liver: Effect of triggering methods on regional variability and measurement repeatability of quantitative parameters. *Radiology*. (2015) 274:405–15. doi: 10.1148/radiol.14140759
26. Shirabe K, Kajiyama K, Abe T, et al. Predictors of microscopic portal vein invasion by hepatocellular carcinoma: Measurement of portal perfusion defect area ratio. *J Gastroenterol Hepatol*. (2009) 24:1431–6. doi: 10.1111/j.1440-1746.2009.05847.x
27. Lee S, Kim SH, Hwang JA, et al. Pre-operative ADC predicts early recurrence of HCC after curative resection. *Eur Radiol*. (2019) 29:1003–12. doi: 10.1007/s00330-018-5642-5
28. Banerjee S, Wang DS, Kim HJ, et al. A computed tomography radiogenomic biomarker predicts microvascular invasion and clinical outcomes in hepatocellular carcinoma. *Hepatology*. (2015) 62:792–800. doi: 10.1002/hep.27877
29. Taouli B, Hoshida Y, Kakite S, et al. Imaging-based surrogate markers of transcriptome subclasses and signatures in hepatocellular carcinoma: Preliminary results. *Eur Radiol*. (2017) 27:4472–81. doi: 10.1007/s00330-017-4844-6
30. Valihrach L, Androvic P, Kubista M. Circulating miRNA analysis for cancer diagnostics and therapy. *Mol Aspects Med*. (2020) 72:100825. doi: 10.1016/j.mam.2019.10.002
31. Yang L, Gu D, Wei J, et al. A radiomics nomogram for preoperative prediction of microvascular invasion in hepatocellular carcinoma. *Liver Cancer*. (2019) 8:373–86. doi: 10.1159/000494099
32. Chong HH, Yang L, Sheng RF, et al. Multi-scale and multi-parametric radiomics of gadoxetate disodium-enhanced MRI predicts microvascular invasion and outcome in patients with solitary hepatocellular carcinoma  $\leq$  5 cm. *Eur Radiol*. (2021) 31:4824–38. doi: 10.1007/s00330-020-07601-2



## OPEN ACCESS

## EDITED BY

Yuxiang Zhou,  
Mayo Clinic Arizona, United States

## REVIEWED BY

Fenglin Dong,  
The First Affiliated Hospital of Soochow  
University, China  
Xiao-Wan Bo,  
Tongji University, China

## \*CORRESPONDENCE

Chun-Wei Liu  
✉ liuchunwei008@163.com  
Xi Wei  
✉ weixi@tmu.edu.cn

RECEIVED 22 September 2023

ACCEPTED 08 July 2024

PUBLISHED 23 July 2024

## CITATION

Yang F, Zhang D, Zhao L-H, Mao Y-R, Mu J,  
Wang H-L, Pang L, Yang S-Q, Wei X and  
Liu C-W (2024) Prediction of clear cell renal  
cell carcinoma  $\leq 4$ cm: visual assessment of  
ultrasound characteristics versus  
ultrasonographic radiomics analysis.  
*Front. Oncol.* 14:1298710.  
doi: 10.3389/fonc.2024.1298710

## COPYRIGHT

© 2024 Yang, Zhang, Zhao, Mao, Mu, Wang,  
Pang, Yang, Wei and Liu. This is an open-  
access article distributed under the terms of  
the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/)  
(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.

# Prediction of clear cell renal cell carcinoma $\leq 4$ cm: visual assessment of ultrasound characteristics versus ultrasonographic radiomics analysis

Fan Yang<sup>1,2</sup>, Dai Zhang<sup>1,2</sup>, Li-Hui Zhao<sup>1,2</sup>, Yi-Ran Mao<sup>1,2</sup>,  
Jie Mu<sup>1,2</sup>, Hai-Ling Wang<sup>1,2</sup>, Liang Pang<sup>3</sup>, Shi-Qiang Yang<sup>4</sup>,  
Xi Wei<sup>1,2\*</sup> and Chun-Wei Liu<sup>5\*</sup>

<sup>1</sup>Department of Ultrasound Diagnosis and Treatment, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Tianjin, China, <sup>2</sup>Key Laboratory of Cancer Prevention and Therapy, Tianjin's Clinical Research Center for Cancer, Tianjin Medical University, Tianjin, China, <sup>3</sup>Department of Urology, Tianjin Occupational Diseases Prevention and Therapeutic Hospital, Tianjin, China, <sup>4</sup>Department of Urology, Tianjin First Central Hospital, Tianjin, China, <sup>5</sup>Department of Cardiology, Tianjin Chest Hospital, Tianjin University, Tianjin, China

**Objective:** To investigate the diagnostic efficacy of the clinical ultrasound imaging model, ultrasonographic radiomics model, and comprehensive model based on ultrasonographic radiomics for the differentiation of small clear cell Renal Cell Carcinoma (ccRCC) and Renal Angiomyolipoma (RAML).

**Methods:** The clinical, ultrasound, and contrast-enhanced CT(CECT) imaging data of 302 small renal tumors (maximum diameter  $\leq 4$ cm) patients in Tianjin Medical University Cancer Institute and Hospital from June 2018 to June 2022 were retrospectively analyzed, with 182 patients of ccRCC and 120 patients of RAML. The ultrasound images of the largest diameter of renal tumors were manually segmented by ITK-SNAP software, and Pyradiomics (v3.0.1) module in Python 3.8.7 was applied to extract ultrasonographic radiomics features from ROI segmented images. The patients were randomly divided into training and internal validation cohorts in the ratio of 7:3. The Random Forest algorithm of the Sklearn module was applied to construct the clinical ultrasound imaging model, ultrasonographic radiomics model, and comprehensive model. The efficacy of the prediction models was verified in an independent external validation cohort consisting of 69 patients, from 230 small renal tumor patients in two different institutions. The Delong test compared the predictive ability of three models and CECT. Calibration Curve and clinical Decision Curve Analysis were applied to evaluate the model and determine the net benefit to patients.

**Abbreviations:** ccRCC, clear cell Renal Cell Carcinoma; RAML, Renal Angiomyolipoma; CECT, contrast-enhanced CT; ROI, region of interest; AUC, Area under the ROC curve; GLCM, gray level cooccurrence matrix; GLRLM, gray-level run-length matrix; GLSZM, gray level size zone matrix; GLDM, gray level dependence matrix; ICCs, Interclass and intraclass correlation coefficients; MRMR, max-relevance and min-redundancy; VIF, Variance Inflation Factor; RFA, Random forest algorithm.

**Results:** 491 ultrasonographic radiomics features were extracted from 302 small renal tumor patients, and 9 ultrasonographic radiomics features were finally retained for modeling after regression and dimensionality reduction. In the internal validation cohort, the area under the curve (AUC), sensitivity, specificity, and accuracy of the clinical ultrasound imaging model, ultrasonographic radiomics model, comprehensive model, and CECT were 0.75, 76.7%, 60.0%, 70.0%; 0.80, 85.6%, 61.7%, 76.0%; 0.88, 90.6%, 76.7%, 85.0% and 0.90, 92.6%, 88.9%, 91.1%, respectively. In the external validation cohort, AUC, sensitivity, specificity, and accuracy of the three models and CECT were 0.73, 67.5%, 69.1%, 68.3%; 0.89, 86.7%, 80.0%, 83.5%; 0.90, 85.0%, 85.5%, 85.2% and 0.91, 94.6%, 88.3%, 91.3%, respectively. The DeLong test showed no significant difference between the clinical ultrasound imaging model and the ultrasonographic radiomics model ( $Z=-1.287$ ,  $P=0.198$ ). The comprehensive model showed superior diagnostic performance than the ultrasonographic radiomics model ( $Z=4.394$ ,  $P<0.001$ ) and the clinical ultrasound imaging model ( $Z=4.732$ ,  $P<0.001$ ). Moreover, there was no significant difference in AUC between the comprehensive model and CECT ( $Z=-0.252$ ,  $P=0.801$ ). Both in the internal and external validation cohort, the Calibration Curve and Decision Curve Analysis showed a better performance of the comprehensive model.

**Conclusion:** It is feasible to construct an ultrasonographic radiomics model for distinguishing small ccRCC and RAML based on ultrasound images, and the diagnostic performance of the comprehensive model is superior to the clinical ultrasound imaging model and ultrasonographic radiomics model, similar to that of CECT.

#### KEYWORDS

small renal tumor, clear cell renal cell carcinoma, ultrasound, radiomics, renal angiomyolipoma

## Introduction

With the improvement of imaging techniques, the incidence of renal cell carcinoma (RCC) has been steadily increasing at a rate of 2%-4% every year (1), among which the proportion of patients diagnosed with small RCC (diameter  $\leq 4$  cm) has been constantly increasing (2). Recently, small renal tumors have become a hot topic in research. About 20-30% of small renal tumors are benign, and renal angiomyolipoma (RAML) is the most common pathology type (3). Clear cell renal cell carcinoma (ccRCC) is the most common pathology type of RCC. There's a lack of typical malignant ultrasound features (necrosis or tumor embolism) in small renal tumors less than 4cm, which makes it difficult to discriminate small ccRCC from RAML. The low-fat content results in hypoechogenicity on ultrasound in fat-poor renal angiomyolipoma (fpRAML), similar to small renal carcinomas. Transabdominal ultrasonography is a common method in preoperative imaging examination of renal tumors, but the small renal tumors may be ignored due to the above sonographic characteristics.

In clinical practice, the identification of ccRCC and RAML is mainly based on contrast-enhanced CT (CECT). However, CECT is

an invasive procedure involving intravenous contrast injection, making it unsuitable for patients with renal dysfunction or iodine contrast allergies. Ultrasonography examination is widely used in clinical practice and is inexpensive, feasible, and reproducible. Moreover, the small ccRCC may be atypical on CECT, and their presentation could be easily confused with RAML, especially low-fat RAML (4). In case of an unclear diagnosis, a repeated examination is required during the follow-up period. So, repeated ultrasound examinations may be more acceptable to these patients. Improvement of the ultrasound diagnostic capability in clinical physical examination will be beneficial.

With the development of artificial intelligence, radiomics prediction models have gained attention in cancer diagnosis (5, 6). Radiomics can extract inaccessible feature data from medical images with high throughput and has great application prospects in predicting the biological behavior of tumors (7, 8). In recent years, few studies have been reported on ultrasonographic radiomics to identify small ccRCC. It is unclear whether the diagnostic performance could be improved using ultrasonographic radiomics in these patients. In this current study, we investigated the feasibility of ultrasonographic radiomics to discriminate ccRCC and RAML by

constructing a clinical ultrasound imaging model, ultrasonographic radiomics model, and comprehensive model. We also compared the diagnostic efficacy between the above models and CECT.

## Materials and methods

### Study population

This retrospective study was approved by the ethics committee of Tianjin Medical University Cancer Institute and Hospital (bc2023079). From June 2018 to June 2022, 385 small renal tumor patients with a histological examination at Tianjin Medical University Cancer Institute (institution 1) were retrospectively recruited to construct training and internal validation cohorts of the model. Another 230 small renal tumor patients with a histological examination from Tianjin First Central Hospital, and Tianjin Occupational Diseases Precaution and Therapeutic Hospital (institutions 2 and 3) were retrospectively recruited, constituting an independent external validation cohort to verify the efficacy of the prediction model. The inclusion criteria were as follows: (1) patients performed an ultrasonic examination and CECT within 2 weeks before the operation, and the images of the tumor's largest diameter were clear, (2) the diagnosis of ccRCC or RAML was confirmed by postoperative pathology, (3) patients had no previous history of other malignancies, (4) maximum diameter of renal tumor  $\leq 4$  cm. The excluding standards were as follows: (1) there were significant artifacts in the ultrasound or CT images, (2) tumor components were predominantly cystic (the solid component was less than 25%) (9), and (3) incomplete clinical information on patients. As a result, 302 patients with 302 small renal tumors were finally enrolled in our study to construct ultrasonographic radiomics models and internal validation; and 69 patients for external validation (Figure 1). Patients in institution 1 were divided into the ccRCC group ( $n = 182$ ; 107 men and 75 women; mean age  $56.85 \pm 10.71$  years) and RAML group ( $n = 120$ ; 57 men and 63 women; mean age  $53.64 \pm 12.23$  years). The mean age of the external validation cohort was  $55.72 \pm 14.58$  years (38 men and 31 women).

### Ultrasonography and CT scanning methods and image analysis

Color Doppler ultrasonic diagnostic apparatus of PHILIPS EPIQ5, Toshiba Aplio 500, and 800 were used. Transabdominal ultrasound was performed using a convex array probe with 1–6 MHz. Patients should be fasting for 8–12 hours to show the largest section of tumors clearly. We performed a multisectional examination of both kidneys in the supine, lateral, or prone position. Ultrasonographic features of renal tumors were recorded, including tumor location, maximum diameter, tumor boundaries, echo pattern, presence of calcifications, necrotic cystic degeneration, and blood flow signals.

Preoperative CECT was performed on multiple scanners: Siemens Somatom Definition, GE HiSpeed 16, and Philips Brilliance 64. Acquisition parameters were as follows: tube voltage, 120–140 kV; automated varied milliampere-second settings; collimation width, 1.5 mm. CT and CECT features included fat

density assessment, peak enhancement degree, homogeneity of enhancement, and the velocity of contrast in and out.

Ultrasonographic and CECT imaging were independently assessed by two sonographers and two radiologists (all with more than 10 years of experience). They were blinded to the pathology results. When the diagnostic results were inconsistent, they reached a consensus through discussion. Clinical information of these patients was recorded, including sex, age, and clinical symptoms.

### Segmentation and pre-processing of ultrasound images

The framework of this study is illustrated in Figure 2. Ultrasound images of the largest renal tumor cross-section were imported into ITK-SNAP software (version v 3.8.0, [www.itksnap.org](http://www.itksnap.org)), and the tumor edges were manually outlined as the region of interest (ROI) by sonographer A with more than 10 years' experience (Figure 3). Two weeks later, sonographer A and sonographer B (with 5 years of experience) made ROI outlining from 50 renal tumor images randomly, to assess intra- and inter-observer correlation coefficients (ICCs).

### Radiomics features extraction

Z-Score was performed to standardize the data of different orders before radiomics feature extraction. The Pyradiomics module of Python 3.8.7 (v3.0.1) was used to extract the radiomics features of ROIs, including shape-based features, first-order features, and structural texture features. The structural texture features included a gray level cooccurrence matrix (GLCM), gray level run-length matrix (GLRLM), gray level size zone matrix (GLSZM), and gray level dependence matrix (GLDM). ICCs were used to evaluate the agreement of feature extraction. The intra-observer ICC was calculated based on two feature extractions by sonographer A, and inter-observer ICC was calculated based on the features extracted firstly by sonographer A and subsequently by sonographer B. Features with better consistency ( $ICC > 0.9$ ) were retained. The maximum relevance-minimum redundancy features were obtained by max-relevance and min-redundancy (MRMR) algorithm filtering. The covariance between ultrasound features was assessed by calculating the Variance Inflation Factor (VIF). Finally, the high-stability radiomics features were subjected to Spearman's correlation analysis, with a correlation coefficient threshold of 0.7.

### Construction of ultrasonographic radiomics model

The patients were randomly divided into training and internal validation cohorts in the ratio of 7:3. The Random forest algorithm (RFA) of the Sklearn module (Python 3.8.7) was applied to construct the clinical ultrasound imaging model, ultrasonographic radiomics model, and comprehensive model, to predict diagnostic efficacy for small ccRCCs. Both feature extraction and model

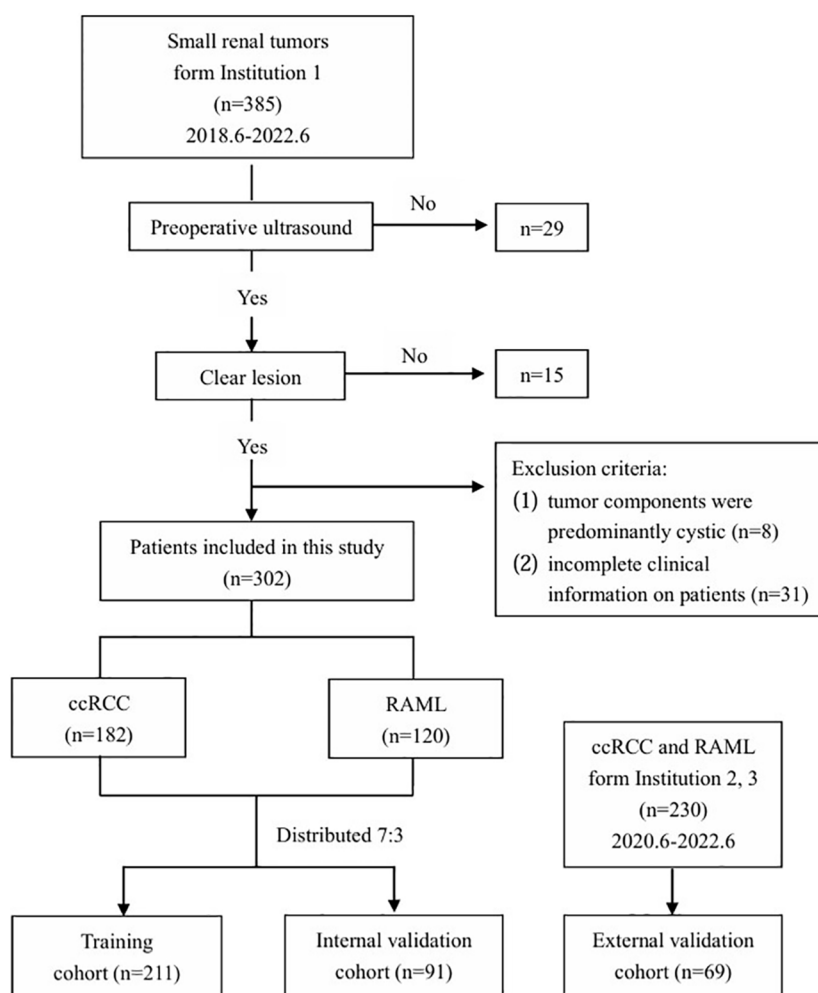


FIGURE 1  
Flowchart of inclusion and exclusion of the study population.

construction used ten-fold cross-validation and parameter tuning to optimize the predictive performance.

## Statistical analysis

All data were analyzed using the SPSS Statistics software version 23.0 (IBM, Armonk, NY, USA), Python3.8.7 and R software version 4.2.2. All categorical variables were expressed as numbers(n) and percentages, and continuous variables were expressed as mean value  $\pm$  standard deviation (SD) or median  $\pm$  inter-quartile range (IQR).  $\chi^2$  test was used to compare the clinical ultrasound characteristics between patients in ccRCC and RAML groups. The diagnostic ability of the ultrasound imaging model, ultrasonographic radiomics model, comprehensive model, and CECT for ccRCC were assessed by the receiver operating characteristic (ROC) curve, and the area under the curve (AUC), sensitivity, specificity, and accuracy of three models and CECT were calculated separately. The AUC values of different models and CECT were compared by the Delong test. The performance of the

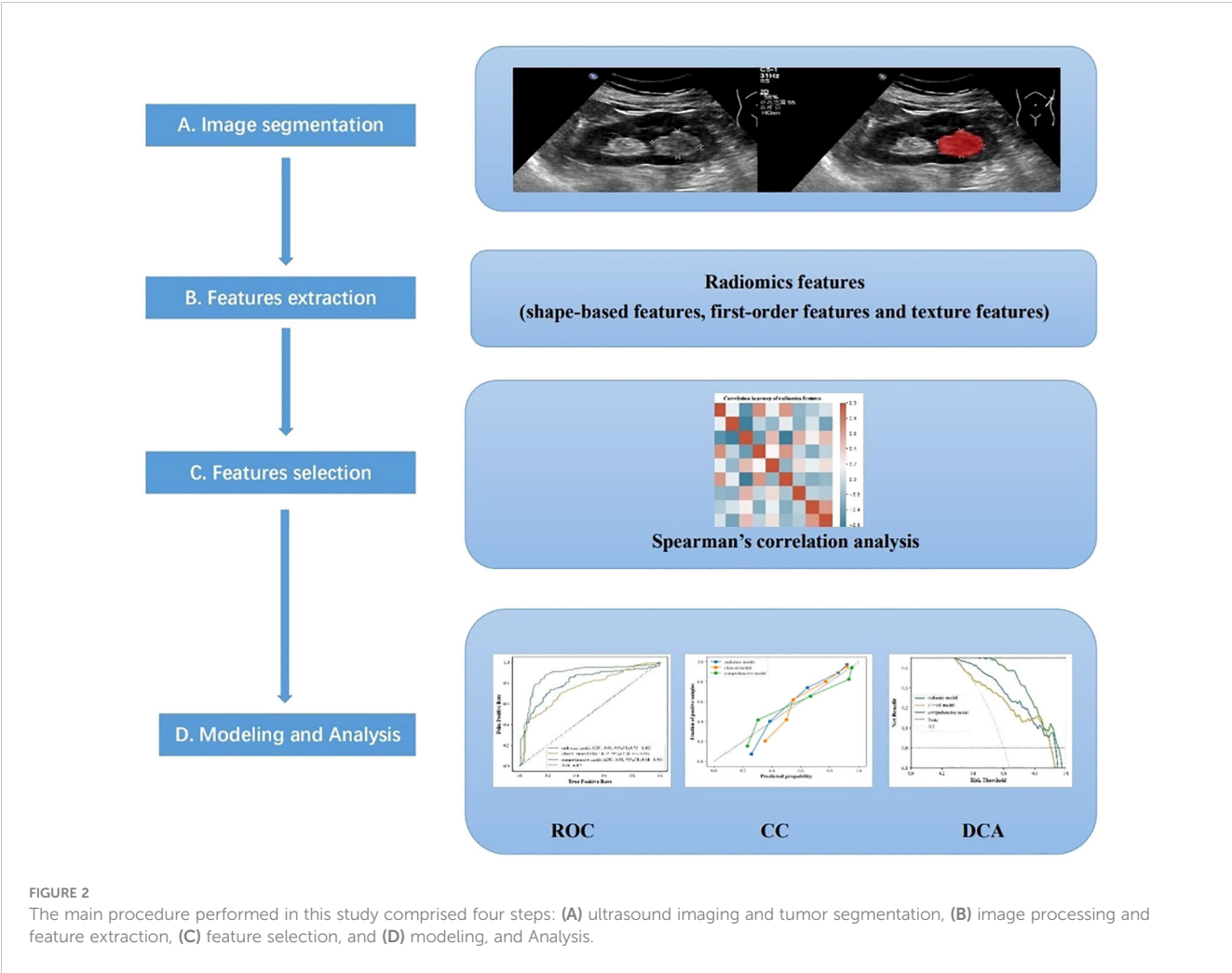
three models was evaluated by the Calibration Curve. Clinical Decision Curve Analysis was also applied to determine the net benefit of patients.  $P < 0.05$  was considered statistically significant.

## Results

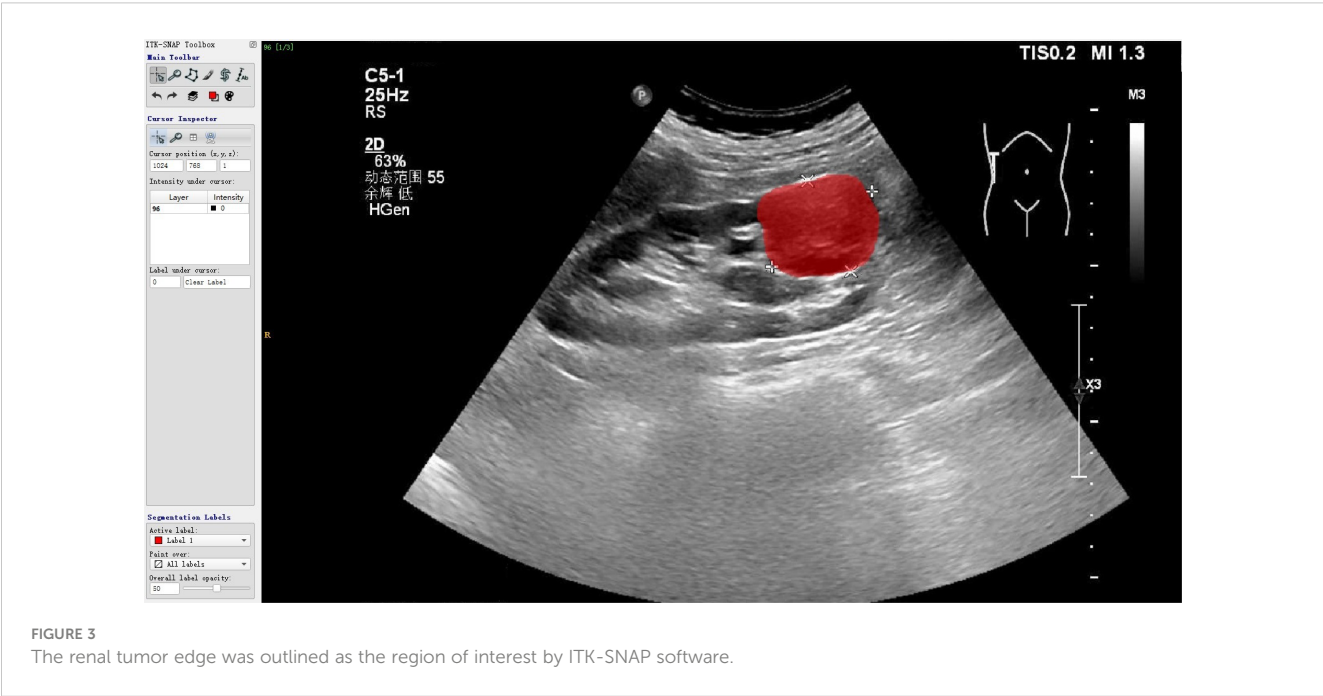
### Comparison of clinical ultrasound and CECT characteristics

In this study, 302 small renal tumor patients, comprising 182 with ccRCC and 120 with RAML, were enrolled as the training and internal validation cohorts. There were significant differences in the distribution of gender, clinical symptoms, echo pattern, necrotic cystic degeneration, blood flow signals, CT presence of bulk fat, and homogeneity of enhancement, but no significant differences in age, location, tumor boundaries, calcification, peak enhancement degree, existent of fast-in and fast-out between the two groups ( $P < 0.05$ , Table 1; Figures 4–6).





**FIGURE 2** The main procedure performed in this study comprised four steps: (A) ultrasound imaging and tumor segmentation, (B) image processing and feature extraction, (C) feature selection, and (D) modeling, and Analysis.



**FIGURE 3** The renal tumor edge was outlined as the region of interest by ITH-SNAP software.

TABLE 1 Clinical ultrasound and CECT characteristics of 302 small renal tumor patients.

|                              | ccRCC (n=182) | RAML (n=120) | $\chi^2$ | P     |
|------------------------------|---------------|--------------|----------|-------|
| Gender                       |               |              |          |       |
| Male                         | 105 (57.7%)   | 57 (47.5%)   | 3.021    | 0.082 |
| Female                       | 77 (42.3%)    | 63 (52.5%)   |          |       |
| Age                          |               |              |          |       |
| <50 years old                | 43 (23.6%)    | 34 (28.3%)   | 0.843    | 0.358 |
| ≥50 years old                | 139 (76.4%)   | 86 (71.7%)   |          |       |
| Clinical symptoms            |               |              |          |       |
| Hematuria                    | 29 (15.9%)    | 8 (6.7%)     | 5.777    | 0.016 |
| No hematuria                 | 153 (84.1%)   | 112 (93.3%)  |          |       |
| Ultrasound characteristics   |               |              |          |       |
| Location                     |               |              |          |       |
| Left kidney                  | 101 (55.5%)   | 59 (49.2%)   | 1.162    | 0.281 |
| Right kidney                 | 81 (44.5%)    | 61 (50.8%)   |          |       |
| Tumor boundaries             |               |              |          |       |
| Clear                        | 140 (76.9%)   | 95 (79.2%)   | 0.211    | 0.646 |
| Unclear                      | 42 (23.1%)    | 25 (20.8%)   |          |       |
| Echo pattern                 |               |              |          |       |
| Hypoechoic                   | 87 (47.8%)    | 31 (25.8%)   | 16.566   | 0.000 |
| Isoechoic                    | 39 (21.4%)    | 28 (23.3%)   |          |       |
| Hyperechoic                  | 56 (30.8%)    | 61 (50.9%)   |          |       |
| Calcification                |               |              |          |       |
| Existent                     | 38 (20.9%)    | 15 (12.5%)   | 3.509    | 0.061 |
| Non-existent                 | 144 (79.1%)   | 105 (87.5%)  |          |       |
| Necrotic cystic degeneration |               |              |          |       |
| Existent                     | 44 (24.2%)    | 9 (7.5%)     | 13.898   | 0.000 |
| Non-existent                 | 138 (75.8%)   | 111 (92.5%)  |          |       |
| blood flow signals           |               |              |          |       |
| Existent                     | 102 (56.0%)   | 51 (42.5%)   | 5.307    | 0.021 |
| Non-existent                 | 80 (44.0%)    | 69 (57.5%)   |          |       |
| CT and CECT characteristics  |               |              |          |       |
| Presence of bulk fat         |               |              |          |       |
| Existent                     | 73 (40.1%)    | 76 (63.3%)   | 15.604   | 0.000 |
| Non-existent                 | 109 (59.9%)   | 44 (36.7%)   |          |       |
| Peak enhancement degree      |               |              |          |       |
| hyper- enhancement           | 116 (63.7%)   | 87 (72.5%)   | 2.521    | 0.112 |
| iso-/hypo-enhancement        | 66 (36.3%)    | 33 (27.5%)   |          |       |

(Continued)

TABLE 1 Continued

|                            | ccRCC (n=182) | RAML (n=120) | $\chi^2$ | P     |
|----------------------------|---------------|--------------|----------|-------|
| Homogeneity of enhancement |               |              |          |       |
| homogeneous                | 78 (42.9%)    | 69 (57.5%)   | 6.207    | 0.013 |
| inhomogeneous              | 104 (57.1%)   | 51 (42.5%)   |          |       |
| Fast-in and fast-out       |               |              |          |       |
| Existent                   | 103 (56.6%)   | 62 (51.7%)   | 0.708    | 0.400 |
| Non-existent               | 79 (43.4%)    | 58 (48.3%)   |          |       |

P-values indicate comparisons between ccRCC and RAML groups.  
ccRCC, clear cell Renal Cell Carcinoma; RAML, Renal Angiomyolipoma; CECT, contrast-enhanced CT.

Selection of ultrasonographic radiomics features

The Pyradiomics software package extracted 491 ultrasonographic radiomics features. 9 ultrasonographic radiomics features were finally retained after regression dimensionality reduction processing, including 3 Shape, 1 GLRLM, 3 GLSZM, and 2 GLDM features. Spearman correlation heatmap of radiomics features is shown in Figure 7.

Diagnostic efficacy of predictive models

In the internal validation cohort, AUC, sensitivity, specificity, and accuracy of the clinical ultrasound imaging model, ultrasonographic radiomics model, comprehensive model, and CECT for ccRCC diagnostic prediction were 0.75, 76.7%, 60.0%, 70.0%; 0.80, 85.6%, 61.7%, 76.0%; 0.88, 90.6%, 76.7%, 85.0% and 0.90, 92.6%, 88.9%, 91.1%, respectively. In the external validation cohort, AUC, sensitivity, specificity, and accuracy of the three models and CECT were 0.73, 67.5%, 69.1%, 68.3%; 0.89, 86.7%, 80.0%, 83.5%; 0.90, 85.0%, 85.5%, 85.2% and 0.91, 94.6%, 88.3%, 91.3%, respectively

(Figure 8; Table 2). In the internal validation cohort, the DeLong test demonstrated no significant difference in AUC between the clinical ultrasound imaging model and ultrasonographic radiomics model ( $Z=-1.287$ ,  $P=0.198$ ), whereas the comprehensive model was superior to the ultrasonographic radiomics model ( $Z=4.394$ ,  $P<0.001$ ) and clinical ultrasound imaging model ( $Z=4.732$ ,  $P<0.001$ ). Moreover, there was no significant difference in AUC between the comprehensive model and CECT ( $Z=-0.252$ ,  $P=0.801$ ). The Calibration curve indicated a better performance of the comprehensive model (Figure 9), while Decision Curve Analysis showed a superior clinical utility of the comprehensive model (Figure 10).

Discussion

The clinical symptoms are usually atypical in patients with small ccRCC. Many patients are discovered incidentally during radiologic examinations (10, 11). It is also more complex considering clinical decision-making (12–14), including a variety of interventions available for these patients: renal tumor biopsy, partial nephrectomy, radical nephrectomy, thermal

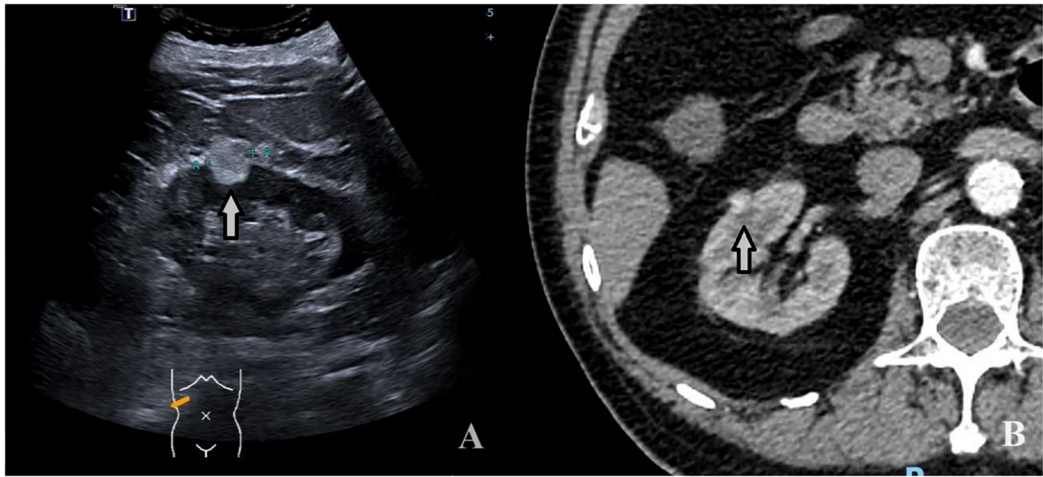


FIGURE 4  
A 74-year-old man with a 1.9 × 1.4cm RAML. (A) Ultrasound demonstrated a mildly hyperechoic mass located in the middle pole of the right kidney (arrow). (B) CECT: the mass showed inhomogeneous hyperenhancement (arrow).

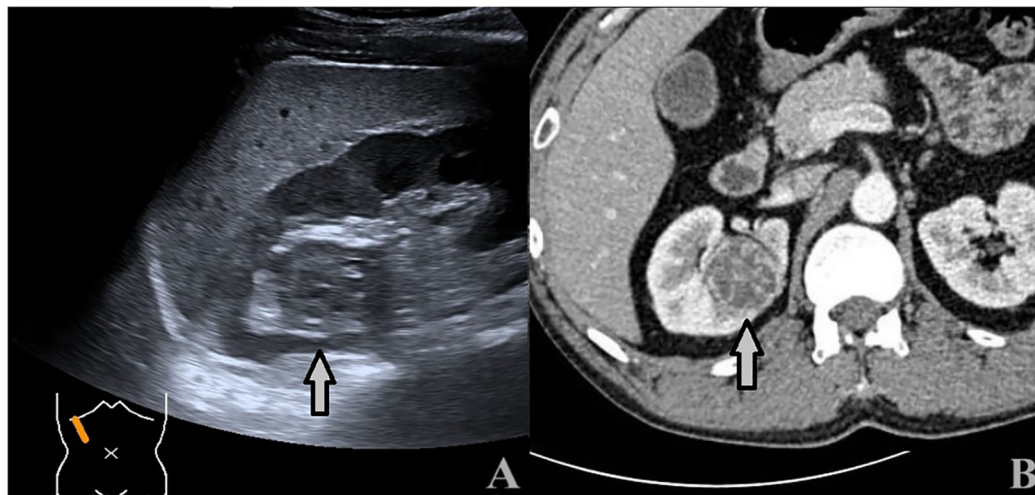


FIGURE 5

A 43-year-old man with a  $3.5 \times 3.4$  cm ccRCC. (A) Ultrasound showed a heterogeneous hypoechoic mass with intratumoral fluid areas located at the upper pole of the right kidney (arrows). (B) CECT: the tumor showed inhomogeneous hyperenhancement with many unenhanced areas (arrows).

ablation, and follow-up monitoring (6, 15, 16). Therefore, the evaluation of imaging features of renal tumors has gradually evolved from morphological presentation to criteria based on histological features and molecular typing features (17, 18). Radiomics can quantitatively assess the heterogeneity of tumors and can be applied to differentiate renal carcinoma from RAML, assess the biological behavior of tumors, and predict the risk of recurrence or survival (19–21). Several studies have reported radiomics models established by CT or MR images, demonstrating their utility in identifying benign and malignant renal tumors and predicting pathological grading (22–25). However, reports on the establishment of radiomics models of ultrasound, the most convenient imaging examination for screening renal tumors, are rather rare.

In this study, the sensitivity of the internal validation cohort of the clinical ultrasound imaging model for the prediction of ccRCC

was 76.7% and the accuracy was 70.0%. Although there were statistically significant differences in gender, clinical symptoms, echo pattern, necrotic cystic degeneration, and blood flow signals between patients in the ccRCC group and RAML group, 30.0% of small renal carcinomas were still misdiagnosed in this model. The reason may be that patients with small ccRCC do not have the specificity of clinical presentation and have a complex and varied pathohistological structure. Whereas ccRCC and fpRAML may exhibit some similar ultrasound characteristics (26, 27). The comprehensive model showed superior ability in predicting ccRCC, with 91% of sensitivity and 77% of specificity. The model extracted 491 ultrasonographic radiomics features, which were processed by regression dimensionality reduction, and finally retained 9 stable ultrasonographic radiomics features. Among them, Shape features described the morphological information of renal tumors. Major axis length, Mesh volume, and Sphericity

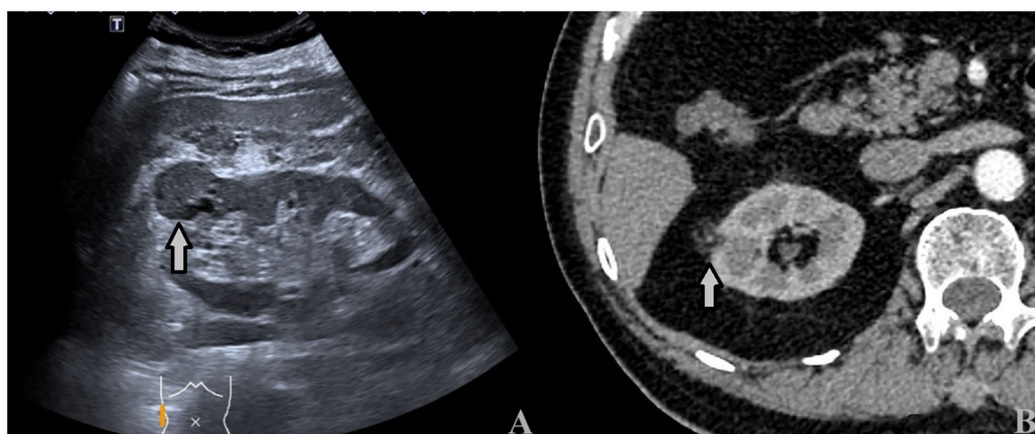
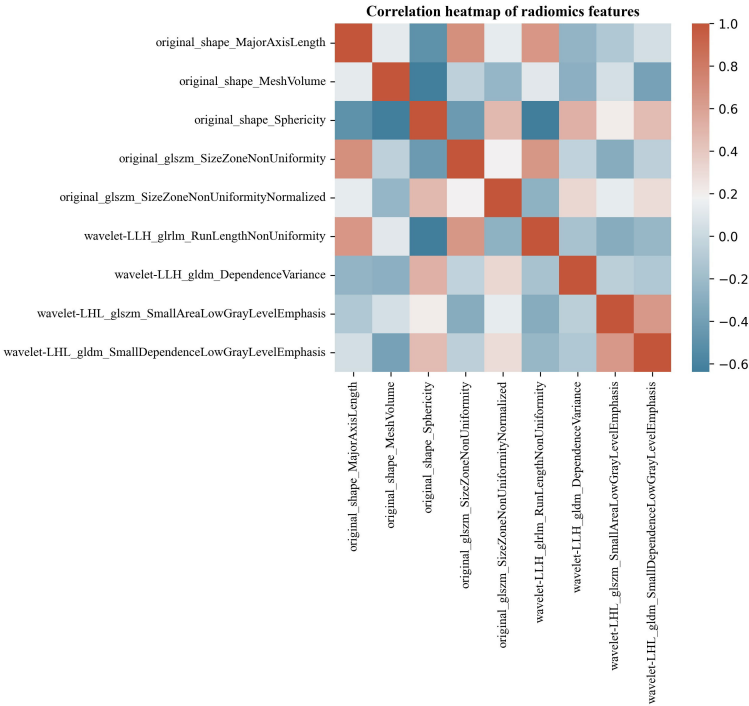


FIGURE 6

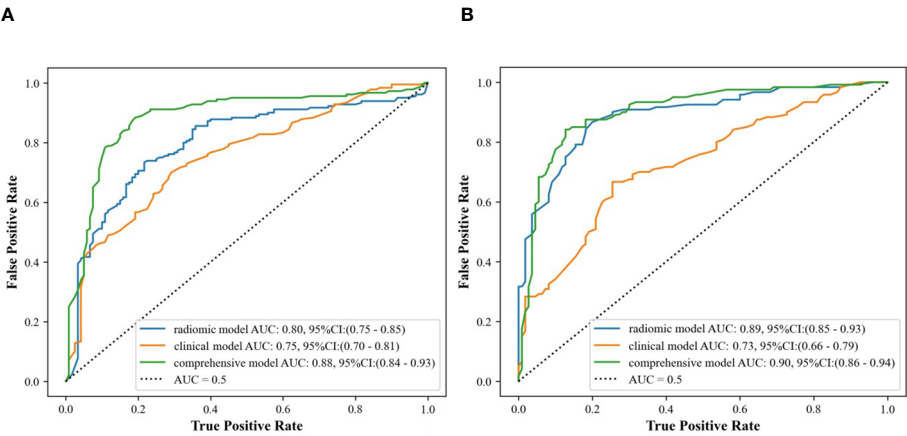
A 64-year-old man with a  $3.2 \times 2.8$  cm ccRCC. (A) Ultrasound showed a heterogeneous hypoechoic mass with intratumoral fluid areas located at the upper pole of the right kidney (arrows). (B) CECT: the tumor showed inhomogeneous hyperenhancement with many unenhanced areas (arrows).



**FIGURE 7**  
Spearman correlation heatmap of renal tumor ultrasonographic radiomics features. The color indicates a correlation: the darker the color, the higher the correlation (red indicates a positive correlation, and blue indicates a negative correlation).

described the similarity of renal tumor morphology to the standard sphere. Renal tumors in the RAML group had a smaller long-axis length and were closer to spherical than those in the ccRCC group. GLSZM was a count of the number of groups of interconnected neighboring pixels or voxels with the same gray level form the basis for the matrix (28). GLRLM provided information about the spatial distribution of runs of consecutive pixels with the same gray level, assessing the percentage of pixels or voxels within the ROI that are part of the runs and therefore reflect graininess (29). GLDM was also a count matrix that holds

information about the number of “dependent” pixels and the number of occurrences of all pixels in the image. All the above three features belonged to texture features, which suggested that the tumors in the ccRCC group had poor texture consistency and a significant effect of non-periodic or speckled texture in ultrasound images compared to those in the RAML group. These findings indicated higher tumor heterogeneity in ccRCC. Compared with the RAML group, tumors in the ccRCC group had more irregular morphology, wider image signal distribution, and rougher texture features. So, combining ultrasonographic radiomics features and

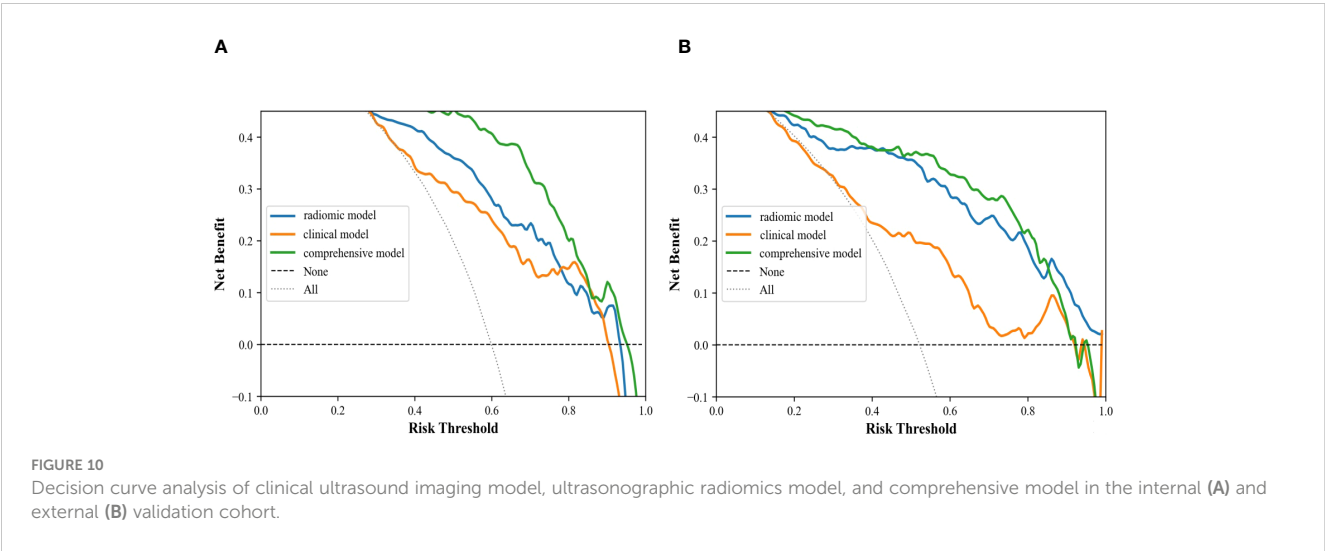
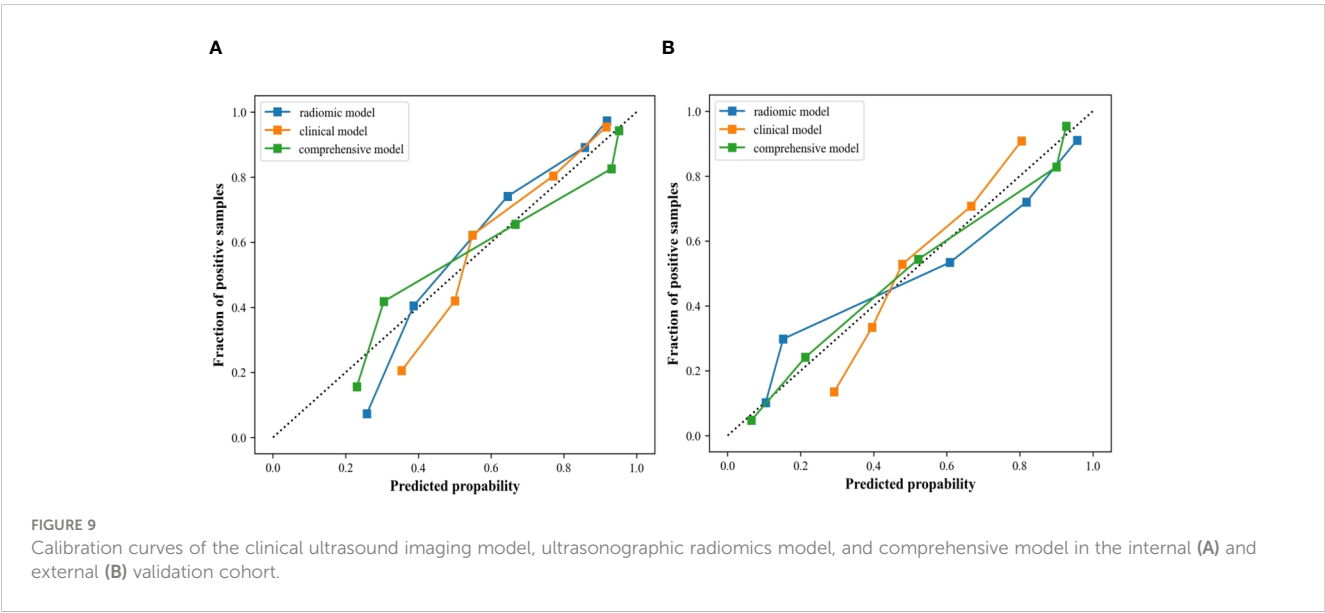


**FIGURE 8**  
ROC curves of clinical ultrasound imaging model, ultrasonographic radiomics model, and comprehensive model in the internal (A) and external (B) validation cohort.



TABLE 2 Comparison of the predictive efficacy of clinical ultrasound imaging model, ultrasonographic radiomics model, comprehensive model, and CECT in the internal and external validation cohorts.

|                                   | Model                             | AUC (95%CI)      | Sensitivity | Specificity | Accuracy |
|-----------------------------------|-----------------------------------|------------------|-------------|-------------|----------|
| Internal validation cohort (n=90) | Clinical ultrasound imaging model | 0.75 (0.70-0.81) | 76.7%       | 60.0%       | 70.0%    |
|                                   | Ultrasonographic radiomics model  | 0.80 (0.75-0.85) | 85.6%       | 61.7%       | 76.0%    |
|                                   | Comprehensive model               | 0.88 (0.84-0.93) | 90.6%       | 76.7%       | 85.0%    |
|                                   | CECT                              | 0.90 (0.84-0.98) | 92.6%       | 88.9%       | 91.1%    |
| External validation cohort (n=69) | Clinical ultrasound imaging model | 0.73 (0.66-0.79) | 67.5%       | 69.1%       | 68.3%    |
|                                   | Ultrasonographic radiomics model  | 0.89 (0.85-0.93) | 86.7%       | 80.0%       | 83.5%    |
|                                   | Comprehensive model               | 0.90 (0.86-0.94) | 85.0%       | 85.5%       | 85.2%    |
|                                   | CECT                              | 0.91 (0.87-0.95) | 94.6%       | 88.3%       | 91.3%    |



clinical ultrasound imaging features, the comprehensive model showed better diagnostic efficacy. The comprehensive model improved the sensitivity and accuracy of ccRCC prediction to 90.6% and 85.0%, which was similar to previous studies (20, 30–32). Our study suggested that ultrasonographic radiomics features could compensate for the shortcomings of clinical ultrasound imaging features and improve the predictive efficacy of small ccRCC. The Calibration Curve and Decision Curve Analysis of the three models also validated that the comprehensive model had a higher net benefit and a better performance in predicting patients with small ccRCC.

Our study had several improvements compared with the previous radiomics studies. Firstly, we compared the diagnostic efficacy between ultrasonographic radiomics models and CECT. Both of these methods have high diagnostic efficiency and there was no significant difference between the comprehensive model and CECT. Moreover, an external validation cohort was used to assess the diagnostic performance of different models. The AUC of the comprehensive model was 0.90 in the external validation cohort, demonstrating a good predictive ability and robustness on new data. Thus, the comprehensive model based on ultrasonographic radiomics and clinical ultrasound imaging features could provide a convenient, inexpensive, and radiation-free examination for small ccRCC patients.

In this study, we applied a “multivariate filtering” feature selection method, the MRMR algorithm, to maximize the correlation between the imaging features and the prediction target as far as possible. Meanwhile, the correlation between the individual features was minimized as far as possible, with the help of high computational speed and high discriminative power. Features were selected from multiple perspectives to minimize information loss in our study, thus avoiding overfitting or underfitting of the predictive model. Moreover, we used the same ratio to divide the training and validation cohort in both ccRCC and RAML, to ensure the stability of the prediction results. Finally, we chose random forests to build the model classifiers to ensure high overfitting resistance and stability.

There are several limitations in this study. Firstly, the cases in this retrospective study are only from three medical institutions, and the results of the study may be subject to selection bias. Secondly, the ultrasound and CECT images in this study are from different diagnostic apparatuses, and there may be heterogeneity in the study images. In addition, the manual segmentation of outlining the ROI may reduce the reproducibility of this study. In the future, we will verify the stability of the results through multicenter prospective studies.

In conclusion, It is feasible to establish a diagnostic prediction model by ultrasonographic radiomics features in ccRCC and RAML with a maximum diameter of  $\leq 4$  cm, and we find that ultrasonographic radiomics features have great potential in identifying tumor heterogeneity in these patients. The comprehensive model showed a superior diagnostic ability in identifying ccRCC, which was similar to that of CECT, providing valuable information for clinicians to make personalized treatment decisions.

## 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 Tianjin Medical University Cancer Institute and Hospital. 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

FY: Writing – original draft, Conceptualization. DZ: Data curation, Writing – original draft. L-HZ: Data curation, Writing – original draft. Y-RM: Formal analysis, Writing – original draft. JM: Methodology, Writing – original draft. H-LW: Resources, Writing – review & editing. XW: Validation, Conceptualization, Writing – review & editing. C-WL: Conceptualization, Methodology, Visualization, Writing – review & editing. LP: Data curation. S-QY: Data curation.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by National Natural Science Foundation of China (82272008), Tianjin Key Medical Discipline(Specialty) Construction Project (TJYXZDXK-009A), and the Tianjin Health Research Project (ZD20018).

## 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. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. *CA Cancer J Clin.* (2018) 68:7–30. doi: 10.3322/caac.21442
2. Kadomoto S, Makino T, Izumi K, Mizokami A. An important step in establishing a treatment strategy for small renal masses of clear cell renal cell carcinoma is based on the significance of adverse histopathologic features on tumor needle biopsy. *Ann Transl Med.* (2019) 7:S374. doi: 10.21037/atm
3. Finelli A, Ismaila N, Bro B, Durack J, Eggen S, Evans A, et al. Management of small renal masses: american society of clinical oncology clinical practice guideline. *J Clin Oncol.* (2017) 35:668–80. doi: 10.1200/JCO.2016.69.9645
4. Fang L, Bai K, Chen Y, Zhan J, Zhang Y, Qiu Z, et al. A comparative study of contrast-enhanced ultrasound and contrast-enhanced CT for the detection and characterization of renal masses. *Biosci Trends.* (2021) 15:24–32. doi: 10.5582/bst.2021.01026
5. Roussel E, Capitanio U, Kutikov A, Oosterwijk E, Pedrosa I, Rowe SP, et al. Novel imaging methods for renal mass characterization: A collaborative review. *Eur Urol.* (2022) 81:476–88. doi: 10.1016/j.eururo.2022.01.040
6. Carlo MI, Attalla K, Mazaheri Y, Gupta S, Yildirim O, Murray SJ, et al. Phase II study of neoadjuvant nivolumab in patients with locally advanced clear cell renal cell carcinoma undergoing nephrectomy. *Eur Urol.* (2022) 81:570–73. doi: 10.1016/j.eururo.2022.01.043
7. 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
8. Liang L, Zhi X, Sun Y, Li H, Wang J, Xu J, et al. A nomogram based on a multiparametric ultrasound radiomics model for discrimination between Malignant and benign prostate lesions. *Front Oncol.* (2021) 11:610785. doi: 10.3389/fonc.2021.610785
9. Silverman SG, Pedrosa I, Ellis JH, Hindman NM, Schieda N, Smith AD, et al. Bosniak classification of cystic renal masses, version 2019: an update proposal and needs assessment. *Radiology.* (2019) 292:475–88. doi: 10.1148/radiol.2019182646
10. Sun M, Thuret R, Abdollah F, Lughezzani G, Schmitges J, Tian Z, et al. Age-adjusted incidence, mortality, and survival rates of stage-specific renal cell carcinoma in North America: a trend analysis. *Eur Urol.* (2011) 59:135–41. doi: 10.1016/j.eururo.2010.10.029
11. Makino T, Kadomoto S, Izumi K, Mizokami A. Epidemiology and prevention of renal cell carcinoma. *Cancers (Basel).* (2022) 14(16):4059. doi: 10.3390/cancers14164059
12. Barata PC, Rini BI. Treatment of renal cell carcinoma: Current status and future directions. *CA Cancer J Clin.* (2017) 67:507–24. doi: 10.3322/caac.21411
13. Au L, Hatipoglu E, Robert de Massy M, Litchfield K, Beattie G, Rowan A, et al. Determinants of anti-PD-1 response and resistance in clear cell renal cell carcinoma. *Cancer Cell.* (2021) 39(11):1497–518.e11. doi: 10.1016/j.ccell.2021.10.001
14. Kang SK, Bjurlin MA, Huang WC. Management of small kidney tumors in 2019. *JAMA.* (2019) 321:1622–23. doi: 10.1001/jama.2019.1672
15. Sundaram M, Song Y, Rogerio JW, Zhang S, Bhattacharya R, Adejoro O, et al. Clinical and economic burdens of recurrence following nephrectomy for intermediate high- or high-risk renal cell carcinoma: A retrospective analysis of Surveillance, Epidemiology, and End Results-Medicare data. *J Manag Care Spec Pharm.* (2022) 28:1149–60. doi: 10.18553/jmcp.2022.22133
16. Bahadoram S, Davoodi M, Hassanzadeh S, Bahadoram M, Barahman M, Mafakher L. Renal cell carcinoma: an overview of the epidemiology, diagnosis, and treatment. *G Ital Nefrol.* (2022) 39(3):2022-vol3.
17. Dobyns AC, Coutelle NA, Suthumpong CY, Rodriguez PE, Castro G, Varella MH. Race/ethnicity and advanced stage of renal cell carcinoma in adults: results from surveillance, epidemiology, and end results program 2007–2015. *Eur J Cancer Prev.* (2022) 31:172–77. doi: 10.1097/CEJ.0000000000000689
18. Zhou Z, Qian X, Hu J, Ma X, Zhou S, Dai Y, et al. CT-based peritumoral radiomics signatures for Malignancy grading of clear cell renal cell carcinoma. *Abdom Radiol (NY).* (2021) 46:2690–98. doi: 10.1007/s00261-020-02890-z
19. Wang J, Tang J, Liu X, He D. A web-based prognostic nomogram for the cancer specific survival of elderly patients with T1-T3N0M0 renal pelvic transitional cell carcinoma based on the surveillance, epidemiology, and end results database. *BMC Urol.* (2022) 22:78. doi: 10.1186/s12894-022-01028-1
20. Zhang H, Yin F, Chen M, Yang L, Qi A, Cui W, et al. Development and Validation of a CT-Based Radiomics Nomogram for Predicting Postoperative Progression-Free Survival in Stage I-III Renal Cell Carcinoma. *Front Oncol.* (2022) 11:742547. doi: 10.3389/fonc.2021.742547
21. Zhao Y, Liu G, Sun Q, Zhai G, Wu G, Li ZC. Validation of CT radiomics for prediction of distant metastasis after surgical resection in patients with clear cell renal cell carcinoma: exploring the underlying signaling pathways. *Eur Radiol.* (2021) 31:5032–40. doi: 10.1007/s00330-020-07590-2
22. Moldovanu CG, Boca B, Lebovici A, Tamas-Szora A, Feier DS, Crisan N, et al. Preoperative predicting the WHO/ISUP nuclear grade of clear cell renal cell carcinoma by computed tomography-based radiomics features. *J Pers Med.* (2020) 11(1):8. doi: 10.3390/jpm11010008
23. Lin F, Ma C, Xu J, Lei Y, Li Q, Lan Y, et al. A CT-based deep learning model for predicting the nuclear grade of clear cell renal cell carcinoma. *Eur J Radiol.* (2020) 129:109079. doi: 10.1016/j.ejrad.2020.109079
24. Luo S, Wei R, Lu S, Lai S, Wu J, Wu Z, et al. Fuhrman nuclear grade prediction of clear cell renal cell carcinoma: influence of volume of interest delineation strategies on machine learning-based dynamic enhanced CT radiomics analysis. *Eur Radiol.* (2022) 32:2340–50. doi: 10.1007/s00330-021-08322-w
25. You MW, Kim N, Choi HJ. The value of quantitative CT texture analysis in differentiation of angiomyolipoma without visible fat from clear cell renal cell carcinoma on four-phase contrast-enhanced CT images. *Clin Radiol.* (2019) 74:547–54. doi: 10.1016/j.crad.2019.02.018
26. Li Q, Liu YJ, Dong D, Bai X, Huang QB, Guo AT, et al. Multiparametric MRI radiomic model for preoperative predicting WHO/ISUP nuclear grade of clear cell renal cell carcinoma. *J Magn Reson Imaging.* (2020) 52:1557–66. doi: 10.1002/jmri.27182
27. Yang G, Nie P, Yan L, Zhang M, Wang Y, Zhao L, et al. The radiomics-based tumor heterogeneity adds incremental value to the existing prognostic models for predicting outcome in localized clear cell renal cell carcinoma: a multicenter study. *Eur J Nucl Med Mol Imaging.* (2022) 49:2949–59. doi: 10.1007/s00259-022-05773-1
28. Thibault G, Angulo J, Meyer F. Advanced statistical matrices for texture characterization: application to cell classification. *IEEE Trans BioMed Eng.* (2014) 61:630–7. doi: 10.1109/TBME.2013.2284600
29. Mayerhoefer ME, Materka A, Langs G, Haggstrom I, Szczypinski P, Gibbs P, et al. Introduction to radiomics. *J Nucl Med.* (2020) 61:488–95. doi: 10.2967/jnumed.118.222893
30. Demirjian NL, Varghese BA, Cen SY, Hwang DH, Aron M, Siddiqui I, et al. CT-based radiomics stratification of tumor grade and TNM stage of clear cell renal cell carcinoma. *Eur Radiol.* (2022) 32:2552–63. doi: 10.1007/s00330-021-08344-4
31. Matsumoto S, Arita Y, Yoshida S, Fukushima H, Kimura K, Yamada I, et al. Utility of radiomics features of diffusion-weighted magnetic resonance imaging for differentiation of fat-poor angiomyolipoma from clear cell renal cell carcinoma: model development and external validation. *Abdom Radiol (NY).* (2022) 47:2178–86. doi: 10.1007/s00261-022-03486-5
32. Yang L, Gao L, Arefan D, Tan Y, Dan H, Zhang J. A CT-based radiomics model for predicting renal capsule invasion in renal cell carcinoma. *BMC Med Imaging.* (2022) 22:15. doi: 10.1186/s12880-022-00741-5



## OPEN ACCESS

## EDITED BY

Sharon R. Pine,  
University of Colorado Anschutz Medical  
Campus, United States

## REVIEWED BY

Minghui Liu,  
University of Electronic Science and  
Technology of China, China  
Jia-Ren Lin,  
Harvard Medical School, United States

## \*CORRESPONDENCE

Ling Qi

✉ qiling1718@gzhmu.edu.cn

Guiying Zhang

✉ guiyangzhang@gzhmu.edu.cn

<sup>†</sup>These authors have contributed equally to  
this work

RECEIVED 30 November 2023

ACCEPTED 15 July 2024

PUBLISHED 29 July 2024

## CITATION

Li C, Che S, Gong H, Ding Y, Luo Y, Xi J, Qi L  
and Zhang G (2024) PI-YOLO: dynamic  
sparse attention and lightweight  
convolutional based YOLO for vessel  
detection in pathological images.  
*Front. Oncol.* 14:1347123.  
doi: 10.3389/fonc.2024.1347123

## COPYRIGHT

© 2024 Li, Che, Gong, Ding, Luo, Xi, Qi 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.

# PI-YOLO: dynamic sparse attention and lightweight convolutional based YOLO for vessel detection in pathological images

Cong Li<sup>1,2†</sup>, Shuanlong Che<sup>3†</sup>, Haotian Gong<sup>4</sup>, Youde Ding<sup>1</sup>,  
Yizhou Luo<sup>2</sup>, Jianing Xi<sup>2</sup>, Ling Qi<sup>1,5\*</sup> and Guiying Zhang<sup>1,2\*</sup>

<sup>1</sup>The Affiliated Qingyuan Hospital (Qingyuan Peoples's Hospital), Guangzhou Medical University, Qingyuan, China, <sup>2</sup>School of Biomedical Engineering, Guangzhou Medical University, Guangzhou, China, <sup>3</sup>Department of Pathology, Guangzhou KingMed Center for Clinical Laboratory, Guangzhou, China, <sup>4</sup>School of Health Management, Guangzhou Medical University, Guangzhou, China, <sup>5</sup>Division of Gastroenterology, Institute of Digestive Disease, the Affiliated Qingyuan Hospital (Qingyuan Peoples's Hospital), Guangzhou Medical University, Qingyuan, China

Vessel density within tumor tissues strongly correlates with tumor proliferation and serves as a critical marker for tumor grading. Recognition of vessel density by pathologists is subject to a strong inter-rater bias, thus limiting its prognostic value. There are many challenges in the task of object detection in pathological images, including complex image backgrounds, dense distribution of small targets, and insignificant differences between the features of the target to be detected and the image background. To address these problems and thus help physicians quantify blood vessels in pathology images, we propose Pathological Images-YOLO (PI-YOLO), an enhanced detection network based on YOLOv7. PI-YOLO incorporates the BiFormer attention mechanism, enhancing global feature extraction and accelerating processing for regions with subtle differences. Additionally, it introduces the CARAFE upsampling module, which optimizes feature utilization and information retention for small targets. Furthermore, the GSConv module improves the ELAN module, reducing model parameters and enhancing inference speed while preserving detection accuracy. Experimental results show that our proposed PI-YOLO network has higher detection accuracy compared to Faster-RCNN, SSD, RetinaNet, YOLOv5 network, and the latest YOLOv7 network, with a mAP value of 87.48%, which is 2.83% higher than the original model. We also validated the performance of this network on the ICPR 2012 mitotic dataset with an F1 value of 0.8678, outperforming other methods, demonstrating the advantages of our network in the task of target detection in complex pathology images.

## KEYWORDS

pathological images, blood vessel, deep learning, object detection, attention mechanism

# 1 Introduction

The growth of most tumors is highly correlated with new blood vessels (1). Rapid tumor cell proliferation often results in hypoxia and malnutrition, prompting the formation of new blood vessels to fulfill the increased metabolic demands of tumors (2). According to the tumor angiogenesis switch hypothesis, when tumors grow to a diameter of 1–2 mm, they frequently trigger the expression of angiogenesis-related factors, initiating the formation of a new vascular network that promotes tumor growth and development. Blocking angiogenesis and tumor growth is an effective approach to treating tumors, such as colorectal cancer, lung cancer, and breast cancer (3). Further studies have revealed that quantitative analysis of blood vessels in tumors can help physicians determine tumor grade and predict patient prognosis (4). This, in turn, supports the development of more rational and effective treatment strategies. Therefore, there is a pressing need for a rapid and precise method to detect blood vessels within tumors.

In the past, there were three main methods for detecting blood vessels within tumors. However, none of these methods employed computer-based automatic detection due to issues such as limitations in imaging equipment. The first method involves the utilization of immunohistochemistry technology to selectively label specific antibodies targeting vascular endothelial cells, such as F8-RA, CD31, CD34, CD105 (5). Researchers then count the positive cells per unit area under a microscope. This method is one of the earliest approaches used for quantitative analysis of tumor vasculature and currently stands as the gold standard for such analysis (6). However, it requires manual selection of the area with the highest vessel density for counting, making it susceptible to subjective influences. The second method entails the use of target-enhanced ultrasound imaging of molecular markers that are overexpressed during angiogenesis (7), enabling indirect quantitative analysis of blood vessels. This approach has advantages such as low detection costs and real-time imaging capabilities but is limited by low detection sensitivity and limited penetration. The third method involves the targeted introduction of magnetic contrast agents into the tumor region, followed by high-resolution imaging of blood vessels within the tumor using MRI technology (8). This method, while capable of producing detailed images, demands sophisticated equipment and longer imaging times, thus limiting its clinical applicability.

In recent years, the field of histopathology has achieved significant advances through electron microscopic imaging, enabling pathologists to perform high-resolution tumor vascularization through digitized whole slide images (WSIs) (9). In addition, rapid advances in artificial intelligence technologies, particularly deep learning, have provided powerful tools for automated tissue section analysis, promising to provide more accurate and consistent results than traditional manual evaluations and to reduce the workload of pathologists. Artificial intelligence algorithms have been developed to identify and quantify vascular features such as density, morphology, and spatial distribution, which are often challenging for human observers (10). Studies have demonstrated the feasibility and efficacy of AI for vascular detection in histological sections of a wide range of malignancies, helping to improve the accuracy of lymphovascular invasion

detection, predict lymph node metastasis, and identify new morphological features with prognostic value (11). However, implementing AI-based vascular testing in clinical practice still faces a number of challenges, including the need for larger and more diverse datasets, and optimizing algorithms for better and faster testing performance so that testing models can be integrated with existing pathology workflows (12).

In response to challenges posed by small target proportions, complex image backgrounds, and subtle feature differences in pathology images, we propose a YOLOv7-based detection network for object detection in pathology images (13). Our approach also prioritizes meeting the speed requirements of clinical applications. The model fuses the BiFormer (14) attention mechanism, the lightweight generalized upsampling operator CARAFE (15) and a new lightweight convolutional technique GSConv (16) into the YOLOv7 model. The proposed model significantly enhances the accuracy of blood vessel detection in pathology images and offers an effective solution for target detection in pathology images.

The contributions of this paper are as follows:

1. This article proposes an improved object detection network model for pathological images based on YOLOv7. We fused the BiFormer attention mechanism, the CARAFE upsampling operator, and GSConv into the YOLOv7 model. This fusion concept effectively enhances detection accuracy and accelerates the blood vessel detection process in pathology images, offering an efficient solution for the task of target detection in pathology images.
2. On the Blood vessel detection dataset, PI-YOLO achieves a mean Average Precision (mAP) value of 87.48%, which is 2.83% higher than the original model. On the ICPR2012 Mitosis detection dataset, the F1 score reaches 0.8678. PI-YOLO outperforms other methods on both datasets, demonstrating superior detection accuracy and faster inference speed (17).
3. Extensive comparative and ablation experiments have provided both quantitative and qualitative verification of this model's superiority in vascular detection tasks within pathological images from various perspectives. The outcomes of this study are anticipated to be valuable for researchers in the fields of anti-angiogenic therapy for tumors and tumor prognosis prediction.

# 2 Related work

At present, classical object detection networks can be broadly categorized into two groups: anchor-based and anchor-free. The key distinction lies in the fact that anchor-based methods require the prior definition of anchor boxes, whereas anchor-free methods do not necessitate this step. One-stage anchor-based approaches, exemplified by YOLOv3 (18) and RetinaNet (19), are capable of directly performing regression and classification tasks for bounding boxes. These methods produce outputs in the form of regression parameters (anchor offsets) and category confidences. On the other hand, the mainstream two-stage anchor-based methods, such as Faster RCNN (20) and Mask RCNN (21), initially generate proposals and subsequently conduct regression and classification tasks for the bounding boxes. Similarly, a variety of anchor-free techniques have been developed, including CornerNet (22) and FSAP



(23). Among these, CornerNet is a classic example of the keypoint detection network, while FSAF incorporates a feature selection anchor-free module to achieve anchor-free object detection. While these conventional networks have delivered promising results in the context of natural images, their performance will be constrained when applied to the unique characteristics of pathological images during the detection process.

Pathological diagnosis, as the gold standard for cancer diagnosis, provides comprehensive information about tumors. In recent years, deep learning methods have been widely applied in the detection and segmentation of micro vessels in pathological images. Traditional methods rely on immunohistochemistry (IHC) staining and manual counting, which are not only time-consuming and labor-intensive but also highly subjective. To address these issues, Yi et al. (24) developed an automated detection method based on fully convolutional networks (FCNs). This method leverages deep learning to achieve end-to-end image training and pixel-level prediction, significantly improving detection efficiency. However, limitations such as small dataset sizes and high false-positive rates remain significant drawbacks. To further enhance detection accuracy and reliability, Fraz et al. (25) proposed a method for micro vessel segmentation in H&E-stained histological images. This method incorporates an uncertainty prediction mechanism that generates uncertainty maps by introducing random transformations during testing, highlighting areas where the network's predictions are uncertain, thus improving segmentation confidence. Additionally, they developed a novel Feature Attention-Based Network (FABnet) (26) for the simultaneous segmentation of micro vessels and nerves. FABnet combines feature attention modules and uncertainty prediction mechanisms to focus on salient features and perform multi-scale feature extraction, achieving more precise segmentation. Despite significant progress in accuracy and reliability, the complexity of the network architecture and the need for multiple random transformations increase computational costs. Furthermore, the study primarily focuses on oral squamous cell carcinoma datasets, lacking extensive validation across other cancer types. Additionally, Generative Adversarial Networks (GANs) have been introduced into vascular detection. Atzori et al. (27) employed GANs to generate synthetic ERG-stained images, reducing dependency on IHC staining. Although GANs have shown impressive results in improving image quality and accuracy, issues such as variability in staining quality and limited training dataset sizes persist. All these methods are based on segmentation approaches, which involve pixel-level classification to distinguish blood vessel boundaries from the background. While accuracy has been continuously improving, the complexity of these models often results in slower processing speeds, limiting their clinical practicality.

## 3 Materials and methods

### 3.1 Datasets

In this paper, two datasets are used for experiments. The experiments on blood vessel detection in pathology images were

performed on the blood vessel detection dataset we created, and the comparison experiments on other detection tasks were performed on the ICPR 2012 mitosis detection dataset (17).

#### 3.1.1 Blood vessel detection dataset

Blood vessels exist in different tumor tissues. To make the detection model applicable to various tumor tissues, we collect 36 WSIs from two institutions: TCGA (28) and the Guangzhou Kingmed Center for Clinical Laboratory. The dataset comprises twenty WSIs of breast cancer, eight WSIs of lung cancer, and eight WSIs of colon cancer. Due to the typically large size of WSIs, we divide them into patches measuring  $512 \times 512$  pixels to facilitate physician annotation and model training. Next, we use a pre-trained classification model to screen out patches with blood vessels. We select a total of 2000 patches containing blood vessels. These patches are annotated by two experienced pathologists (with more than five years of experience in pathology), and then reviewed by expert pathologists (with more than ten years of experience in pathology) after the annotation is completed.

The annotated dataset contains a total of 2000 images, including 4526 blood vessels. They were divided into training set, validation set and test set according to the ratio of 7:2:1. The training set contains 1400 images, including 3445 blood vessels. The validation set consists of 400 images, which include 681 blood vessels. The test set comprises the remaining 200 images, containing 400 blood vessels. All datasets are stored in PNG format. Table 1 shows the division of the datasets.

#### 3.1.2 ICPR 2012 mitosis detection dataset

The ICPR 2012 mitosis detection dataset was introduced in the ICPR 2012 competition, making it the first publicly available mitosis detection dataset. This dataset consists of five H&E-stained breast cancer biopsy slides. In each slide, a pathologist selects 10 high-power fields (HPF) at 40x magnification, resulting in a total of 50 HPFs in the dataset, which collectively contain more than 300 cells undergoing mitosis. The slides were scanned using various equipment, including an Aperio XT scanner (A scanner), a Hamamatsu Nano Zoomer scanner (H scanner), and a 10-band multispectral microscope. Each HPF's mitotic cells were annotated by a pathologist. For our experiments, we focused on data obtained from the A scanner, which includes 50 RGB images. Out of these, 35 were allocated for training, and the remaining 15 were designated

TABLE 1 The partitioning of the dataset.

|         | Name           | Proportion | Number of Pictures | Number of Blood vessels |
|---------|----------------|------------|--------------------|-------------------------|
|         | training set   | 70%        | 1400               | 3445                    |
| dataset | validation set | 20%        | 400                | 681                     |
|         | test set       | 10%        | 200                | 400                     |
| total   |                | 100%       | 2000               | 4526                    |

for testing. Since the image size of  $2084 \times 2084$  pixels isn't suitable for direct model training, we subdivided it into patches measuring  $521 \times 521$  pixels, resulting in 243 images that contain a total of 323 mitotic cells. Of these, 195 were used for the training set, and the remaining 48 were assigned to the test set.

## 3.2 The proposed PI-YOLO model

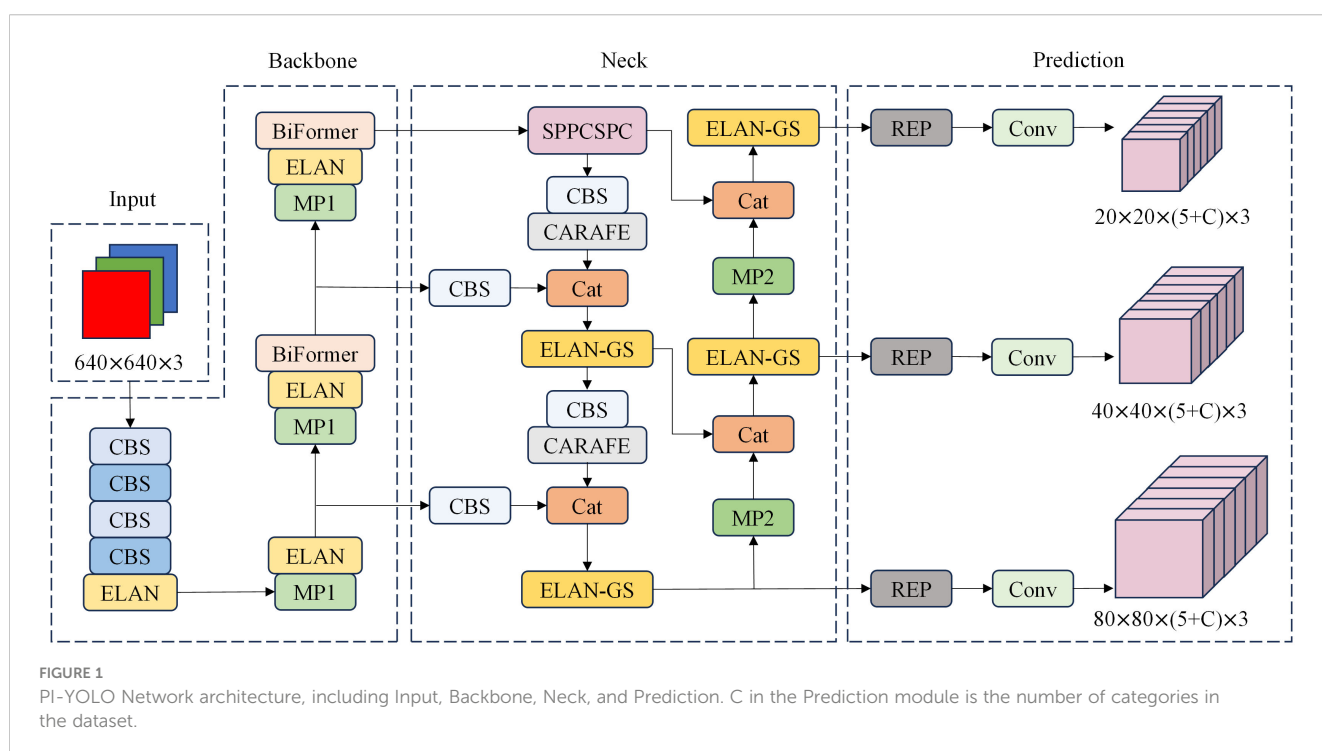
Since the background in pathology images typically occupies the majority of the image area, and the target to be detected is often small, it is essential to employ a network structure with global modeling capabilities as the backbone. YOLOv7's backbone network consists of convolutional stacks. The feature maps in the shallow layers of the network have a limited receptive field, allowing them to capture only local features of the original image. Conversely, the convolutional layers in the deeper network possess a larger receptive field but require more complex computations. BiFormer, a multi-head self-attention mechanism, provides a broader field of view in the shallow layers of the network, enabling it to capture global features. Its key idea lies in filtering out most of the irrelevant key-value pairs at the coarse region level, retaining only a small portion of the routing region. This significantly reduces computational demands while maintaining accuracy. In this paper, we propose a target detection method for pathological images by incorporating the BiFormer attention mechanism, CARAFE upsampling module, GSConv module, and the YOLOv7 network. We have enhanced, optimized, and reconstructed the method. The overall network framework, shown in Figure 1, consists of four main components: Input, Backbone, Neck, and Prediction.

### 3.2.1 Input layer

In the input layer, each training sample undergoes an initial Mosaic data augmentation process. This process involves the following steps: First, four different images are randomly selected from the dataset. Each selected image is then individually flipped, and its color gamut is adjusted. After these adjustments, the images are randomly cropped. Next, these four augmented images are combined into a single new image, forming a new training sample. This Mosaic augmentation technique enriches the background variations and generalizes the features used for detection. Furthermore, the locations of the detection targets in these new composite images are adaptively adjusted according to their original positions in the selected images. By incorporating diverse backgrounds, the model's ability to detect targets in complex environments is enhanced through training with these augmented samples.

### 3.2.2 Backbone

The backbone network is a critical component for feature extraction in our model. The original YOLOv7 backbone consists of 50 modules, which include CBS modules, ELAN modules, and MP1 modules. Specifically, there are four ELAN modules in the network, as shown in Figure 2. Each ELAN module is composed of six CBS modules. To enhance the feature extraction capability of the backbone network, we have introduced the BiFormer attention mechanism after the last CBS module of the last two ELAN modules. The BiFormer attention mechanism is characterized by dynamic sparse attention with a two-layer routing process. Its core concept involves filtering out the least relevant key-value pairs at the coarse area level. This process is carried out by constructing and pruning an area-level directed graph. Subsequently, fine-grained



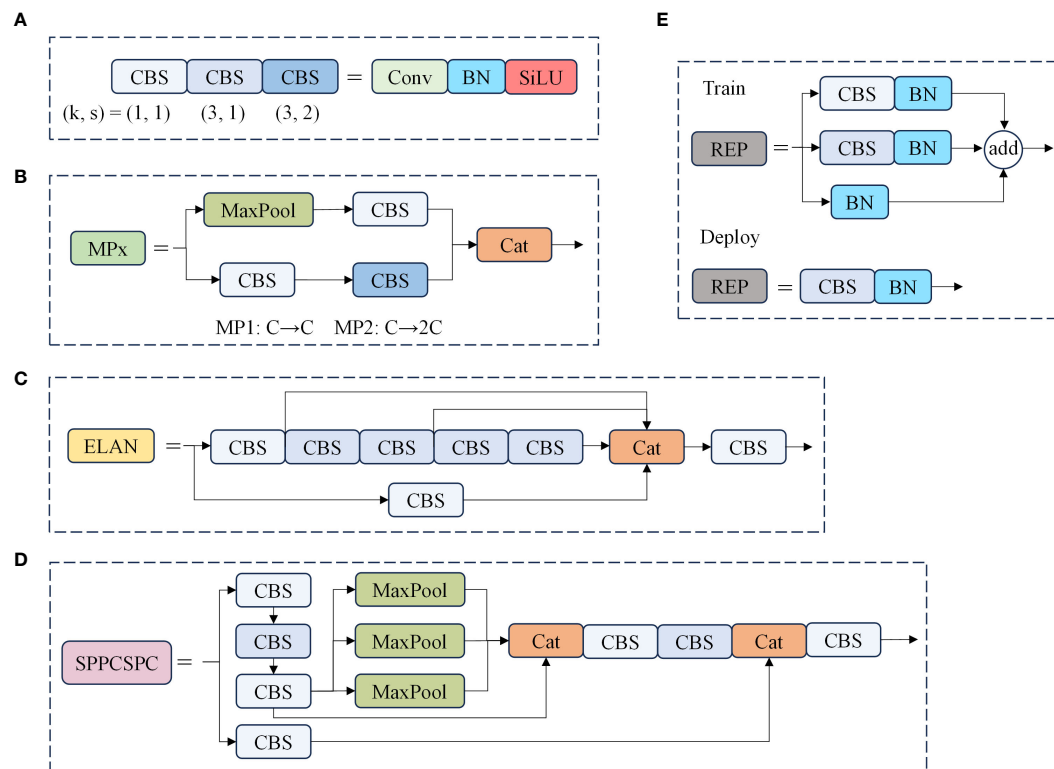


FIGURE 2

Structure diagram of the model part of the module. In this diagram, (A) illustrates the combination of different convolution modules, where “k” represents the convolution kernel size, and “s” signifies the convolution step size. (B) outlines the essential configuration of the MP module, while (C) provides an overview of the core structure of the ELAN module, (D) presents the layout of the SPPCSPC module, and (E) describes the architecture of the REP module.

token-to-token attention is applied within the union of the routed areas. The incorporation of the BiFormer attention mechanism enables dynamic query-aware sparsity, allowing for more flexible computational allocation and content awareness. This mechanism not only preserves dependencies and location information across different spatial regions but also significantly reduces computational costs. The workflow of the enhanced backbone network can be summarized as follows: Initially, input images pass through a series of CBS modules for basic feature extraction. These features are then fed into the ELAN modules, where the BiFormer attention mechanism is applied to enhance the relevant feature maps. Finally, the refined features proceed to subsequent network layers for further processing and prediction tasks. This structured approach ensures that the backbone network effectively captures and utilizes critical spatial information, ultimately improving the model’s overall performance in object detection tasks.

### 3.2.3 Neck and prediction layer

The neck network serves to disperse the multi-scale output learned from the backbone network into multiple feature mappings, and then integrates the learned multi-scale information. This is to enhance the model’s ability to capture diverse information and improve target detection performance. As shown in Figure 1, the neck network adopts a PAFPN structure, which combines enhanced components from FPN (29) and PANet (30) for feature extraction

and fusion. In place of the original upsampling module, we introduce a lightweight generalized upsampling operator called CARAFE within the neck network. This operator expands the receptive field without significantly increasing computational demands or introducing excess parameters. It efficiently leverages semantically relevant content from the feature map for upsampling. Additionally, we employ the lightweight convolutional block GSConv to enhance the ELAN module, reducing model parameters, computational complexity, and size while preserving rich features. Finally, after the input image undergoes two rounds of feature extraction via the backbone and neck networks, the feature information is amalgamated using repconv and transformed into the final prediction information to generate the model’s prediction results.

## 3.3 Attention for PI-YOLO

Due to the intricate backgrounds and a high prevalence of small objects in pathological images, numerous detection models struggle to effectively filter out background information. To shift the focus of the detection model towards the essential information within the input features while minimizing the influence of background data, we incorporate a dynamic sparse attention mechanism known as BiFormer into the backbone network of the model. This BiFormer

attention mechanism, as utilized in this study, can be delineated into two distinct phases.

The first phase initiates with coarse-grained attention, emphasizing sparsity control, while the second phase performs fine-grained attention based on the outcomes of the sparse attention from the first phase. In the initial phase, the image is partitioned into multiple coarse-grained blocks, upon which self-attention is applied. This process computes correlations between every two coarse-grained blocks using  $Q$  and  $K$ , resulting in a relational matrix. Subsequently, this matrix is sparsified, retaining only the top- $k$  elements with the highest values, signifying pairs of blocks that require further attention. In the subsequent phase, building upon the sparse coarse-grained matrix from the first stage, additional fine-grained self-attention is conducted. Each patch exclusively engages in attention computations with patches residing within other coarse-grained blocks that are associated with the coarse-grained block it occupies in the first stage. The implementation details are as follows:

BiFormer is built using Bi-Level Routing Attention (BRA) as the basic building block. The implementation details of BRA are as follows: given a 2D input feature map  $X \in R^{H \times W \times C}$ , it is first divided into  $S \times S$  non-overlapping regions to obtain  $Q, K, V$ , the related equation is as in (1):

$$Q = X^r W^q, K = X^r W^k, V = X^r W^v \quad (1)$$

where  $W^q, W^k, W^v \in R^{c \times c}$  are projection weights for the query, key, value, respectively.

Then the mean of  $Q$  and  $K$  is calculated to obtain the corresponding  $Q^r, K^r \in R^{S^2 \times C}$ , and then the affinity adjacency matrix  $A^r \in R^{S^2 \times S^2}$  between regions is obtained using transpose multiplication, the related equation is as in (2):

$$A^r = Q^r (K^r)^T \quad (2)$$

Then use the *topk* operator to keep the  $k$  regions with the closest relationships to obtain the region routing index matrix  $I^r$ , the related equation is as in (3):

$$I^r = \text{topkIndex}(A^r) \quad (3)$$

After obtaining  $I^r$ , fine-grained Token-to-token attention can be applied, as shown in the Figure 3.

First, collect all the routing regions indexed by all elements in  $I^r$  and collect all their  $K$  and  $V$  to obtain  $K^g, V^g \in R^{S^2 \times \frac{kHW}{S^2} \times C}$ , the related equation is as in (4):

$$K^g = \text{gather}(K, I^r), V^g = \text{gather}(V, I^r) \quad (4)$$

Then apply  $K^g$  and  $V^g$ , which are the gathered key and value tensor. Next, apply attention to the gathered key-value pairs as follows (5):

$$O = \text{Attention}(Q, K^g, V^g) + \text{LCE}(V) \quad (5)$$

Here, a local context enhancement term  $\text{LCE}(V)$  is introduced, as described in (31). The function  $\text{LCE}(\cdot)$  is parameterized using deep separable convolution, and set the convolution kernel size to 5. It follows the design of most vision transformer architectures, which also use a four-stage pyramid structure, i.e., downsampling by a factor of 32, as shown in Figure 4.

### 3.4 CARAFE for PI-YOLO

The up-sampling method adopted by YOLOv7 in the feature fusion part is nearest neighbor interpolation up-sampling, which determines the up-sampling kernel only by the spatial location of pixel points, and does not utilize the semantic information of the feature map, ignores the possible influence of surrounding feature points, and the perceptual field is small, and the quality of the image

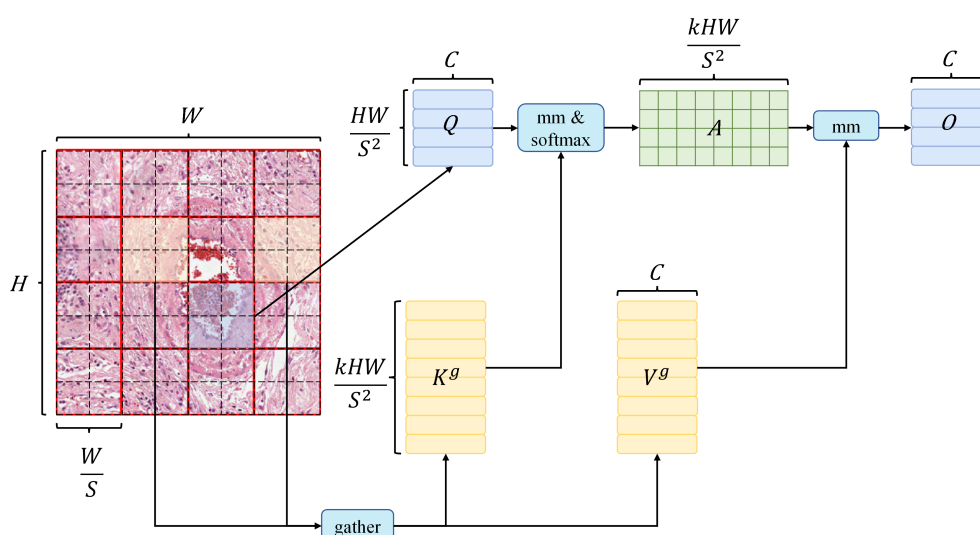
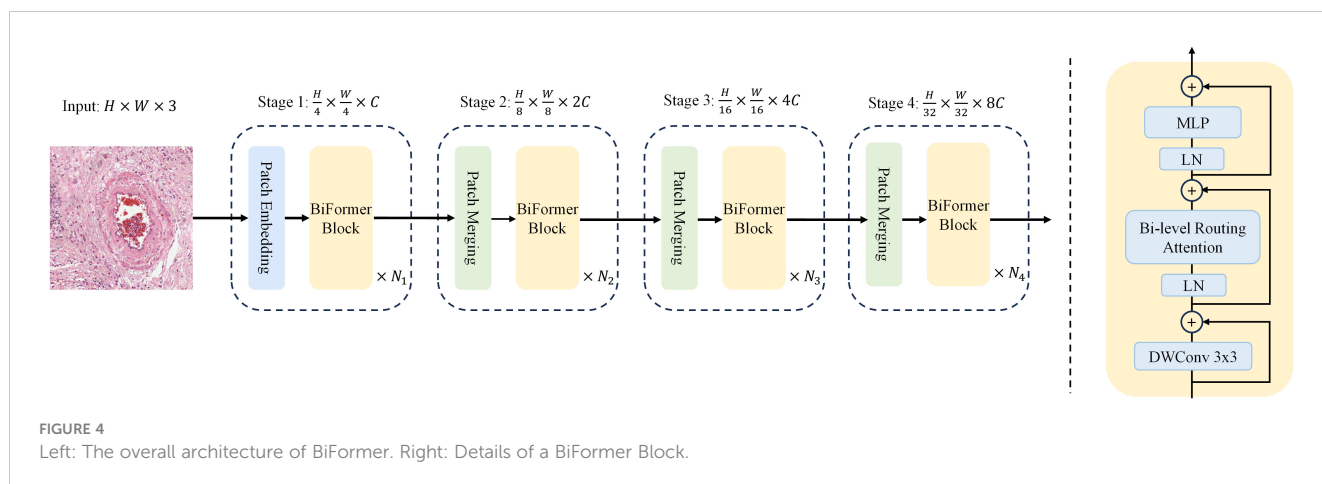


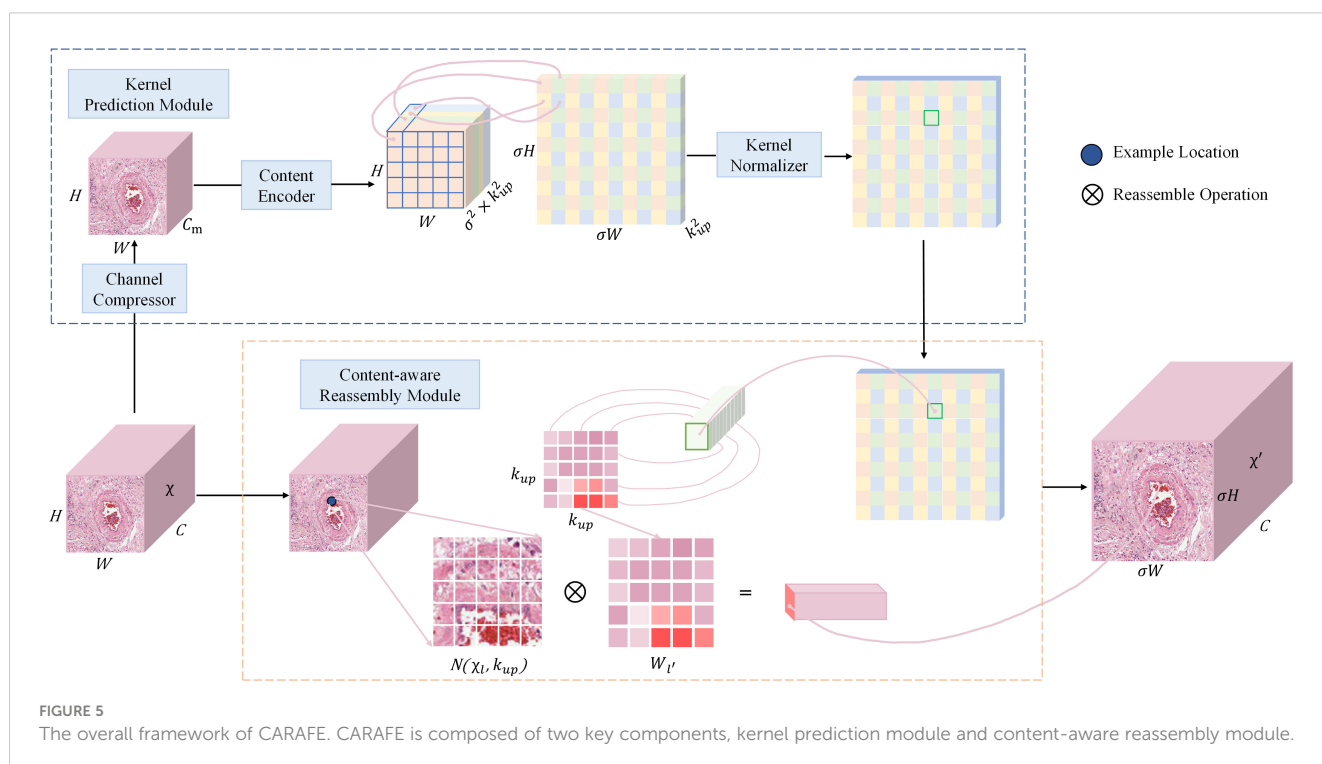
FIGURE 3  
The structure diagram of the BiFormer dynamic attention mechanism.



after up-sampling is not high. In this paper, the CARAFE lightweight up-sampling operator with a large perceptual field is used to improve the neck, which can make good use of the semantic information of the feature map.

CARAFE is divided into a kernel prediction module and a content-aware reassembly module. The kernel prediction module is responsible for generating the up-sampling reassembly kernel, which predicts the attention weights for each up-sampling location based on the mapping relationship between the down-sampled feature map and the up-sampled location. These weights are crucial for maintaining spatial details and contextual information during the feature reassembly process. The content-aware reassembly module focuses on retaining as much spatial information as possible during the up-sampling process to better preserve the accuracy of object boundaries. The structure of CARAFE is shown in Figure 5.

The overall sampling process of CARAFE is as follows. Firstly, for the input feature map  $\chi$  of shape  $H \times W \times C$ , a  $1 \times 1$  convolutional layer is used to compress the input channel from  $C$  to  $C_m$  in order to reduce the parameter and computational cost of the model. Next, a reorganization upsampling kernel of shape  $H \times W \times C_m$  is obtained based on the feature map of shape  $H \times W \times \sigma^2 \times k_{up}^2$  by using a convolutional layer as a content encoder for predicting the upsampling kernel, where  $\sigma$  is the upsampling multiplicity and  $k_{up}$  is the size of the receptive field region for the feature recombination process. Then the channel is expanded in the spatial dimension to obtain the upsampling kernel of shape  $\sigma H \times \sigma W \times k_{up}^2$ , and finally the upsampling kernel is normalized so that its convolutional kernel weights sum to 1. In the content-aware reassembly module, for each position in the output feature map, it is mapped back to the input feature map by taking the region centered on the  $k_{up} \times k_{up}$  region centered on it, and perform dot product





with the predicted upsampling kernel at that point to get the output value. Different channels at the same location share the same upsampling kernel, and finally the  $\sigma H \times \sigma W \times C$  upsampled feature map  $\chi'$  is obtained.

The CARAFE upsampling module enhances the ability of the neck network for image feature extraction and fusion, thus effectively addressing the challenges posed by the presence of a large number of backgrounds and densely distributed small targets in pathology images.

### 3.5 GSConv for PI-YOLO

Standard Convolution (SConv) operates on all three channels simultaneously, where the number of convolution kernels and channels matches the number of output and input channels, respectively. Consequently, employing an excessive number of standard convolution kernels results in an accumulation of parameters. Utilizing SConv for image feature extraction leads to a proliferation of parameters and feature redundancy, particularly in deeper layers. The Ghost Conv model module, proposed by Han K et al. (32), efficiently extracts valuable features while reducing parameters and computational overhead. It operates in two steps: initially involving a limited number of convolutional and linear transformation operations, followed by the integration of feature maps generated from these two operations, which are then output.

Ghost Conv is predominantly employed in the realm of lightweighting computer vision models due to its impressive performance. However, the Ghost Conv module does encounter a challenge in that it loses a significant amount of channel information during its second step of operation. To address this limitation, Li H et al. (16) introduced the GSConv module, as illustrated in Figure 6. The GSConv module is designed to mitigate this issue. Its final blending operation effectively disrupts channel information uniformly, enhances semantic information extraction, strengthens the fusion of feature data, and ultimately improves the representation of image features.

When the network conducts feature fusion at the Neck layer, it continuously propagates semantic information downward. However, this propagation can result in the loss of some semantic information, which may affect the final prediction, especially when the height and width of the feature map, as well as the number of channels, undergo continuous compression and expansion. In this paper, we introduce the GSConv module into the ELAN module of the network's neck layer, replacing the standard convolution. This adjustment not only reduces the model's parameter count and computational load but also maximizes the sampling effect. The structure of the GSConv module is illustrated in Figure 7. Specifically, the four convolutions preceding the Concat layer make use of the GSConv module. This modification reduces the model's parameter count while ensuring detection accuracy.

## 4 Experiments and results

### 4.1 Experimental environment and hyperparameter settings

All experimental data in this article were measured in the same environment. The hardware environment adopts Intel (R) Xeon (R) Gold 5218 @ 2.30GHz CPU, 64GB RAM, and NVIDIA GeForce RTX TITAN graphics card. The system environment is Linux version 5.13.0-30 generic. Python version 3.10, PyTorch version 1.13.0, CUDA version 11.7.

In the experimental models presented in this paper, we explored various hyperparameter configurations and found that the best results were obtained when using the default hyperparameters of the original YOLOv7. The relevant parameters used in the experiments are listed in Table 2. The gradient descent optimizer employed for updating the convolutional kernel parameters is Adam, with a momentum parameter of 0.937. During the training process, the learning rate is updated using a step-wise method, with a maximum learning rate of 0.001 and a training batch size of 24. The training duration spans 200 epochs, and it's

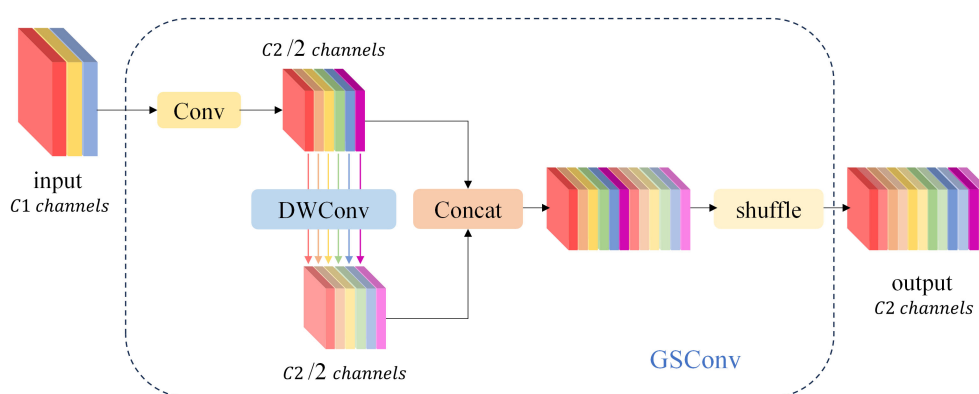
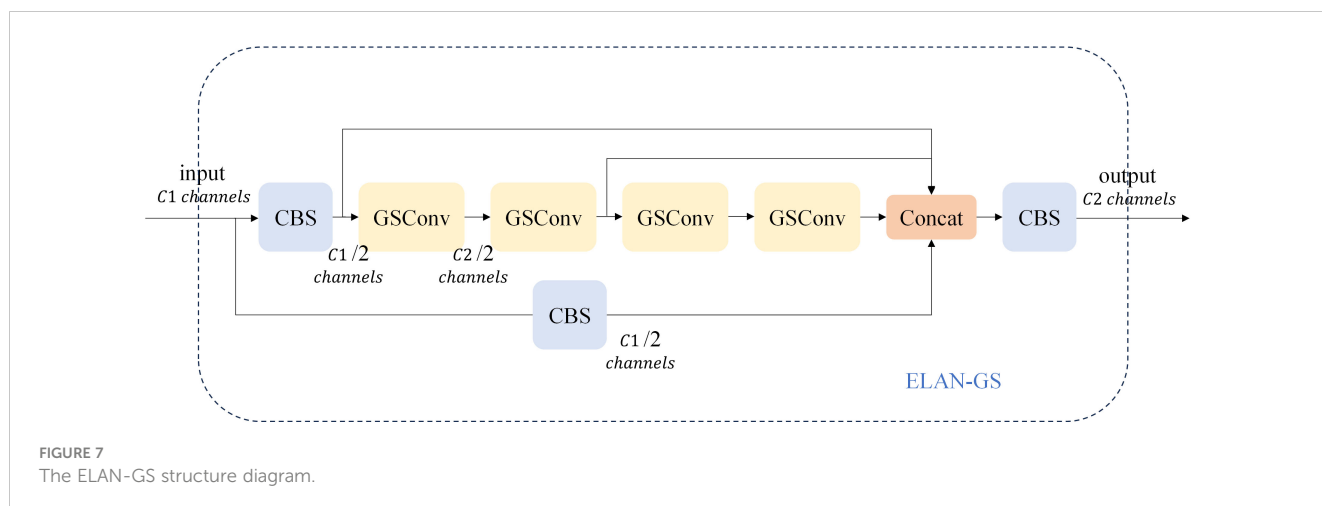


FIGURE 6

The structure of the GSConv module. The "Conv" box consists of three layers: a convolutional-2D layer, a batch normalization-2D layer, and an activation layer.



worth noting that all experiments were conducted without pre-training weights. The entire network model was trained from scratch.

## 4.2 Evaluation indicators

In this study, we utilize seven evaluation metrics to assess the model's performance:

**Precision:** This metric represents the ratio of correctly predicted positive instances (TP) to the total recognized objects and is calculated as shown in Equation (6). **Recall:** Recall signifies the ratio of correctly recognized objects to the total number of objects and is calculated using Equation (7). **F1 Score:** The F1 score is the harmonic mean of Precision and Recall, computed as indicated in Equation (8). **Average Precision (AP):** AP is the average of precision values at different recall points, quantified by the area under the Precision-Recall (PR) curve. A higher AP value indicates greater model precision, with the calculation formula shown in Equation (9). **Mean Average Precision (mAP):** mAP represents the average AP across all categories. A higher mAP value signifies a superior model with increased target recognition accuracy, with the formula outlined in Equation (10). **Frames Per Second (FPS):** FPS indicates the number of images processed per second and serves as an indicator of detection speed. A higher value implies faster model inference. **Giga Floating-point Operations Per Second (GFLOPS):** GFLOPS quantifies the computational complexity of the model, reflecting the number of computations required. Additionally, the term "Params" refers to the total number of trainable parameters in the model, serving as an indicator of the model's size and training requirements.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (7)$$

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

$$AP = \int_0^1 P(R) dR \quad (9)$$

$$mAP = \frac{\sum_1^N \int_0^1 P(R) dR}{N} \quad (10)$$

TP represents the count of positive samples correctly predicted by the model, whereas FP represents the count of negative samples predicted as positive by the model. FN represents the count of positive samples that the model incorrectly predicts as negative. In this context, P represents the class accuracy, R represents the class recall rate, and N represents the total number of classes. Given that the dataset contains only one type of blood vessel,  $N = 1$ .

## 4.3 Attention mechanism compatibility experiment

We chose to incorporate the BiFormer attention mechanism into our model. To assess its compatibility with the model, we conducted comparisons with models that lacked a fused attention mechanism, as well as models that integrated the fused SENet, ECA, CA, and CBAM attention mechanisms, respectively (33–36). In our qualitative analysis experiments, we employed a visualization technique commonly used in deep learning, known as Grad-CAM (37), to illustrate differences in the regions of interest within the model after integrating various attention mechanisms. This method

TABLE 2 Experiment-related hyperparameter settings.

| Hyperparameter | Epoch | Batch_size | Max_learning_rate | Optimizer | Momentum | Lr decay |
|----------------|-------|------------|-------------------|-----------|----------|----------|
| Value          | 200   | 24         | 0.001             | Adam      | 0.937    | Step     |

offers insights into the model's focus on different regions and helps explain variations in model performance. The importance of features is depicted using a color scale, with increasing importance denoted by a transition from blue to red hues.

As illustrated in Figure 8. Heatmap, which displays the heatmaps generated by different attention mechanisms, we conducted experiments using images from various sources. When compared with the visualization results of other attention mechanisms, it becomes evident that the heatmap produced by the BiFormer Attention Mechanism exhibits a larger overall coverage area. This suggests that the model focuses on a broader region of interest at the target location, resulting in more comprehensive feature extraction of the targets. This, in turn, facilitates the detection of small targets. Additionally, the red area in the heatmap is also more extensive, indicating enhanced extraction of effective target feature information. The model allocates greater attention to the pertinent target information. The experimental outcomes reveal that the integration of the BiFormer attention mechanism compels the model to prioritize the feature information of the target to be recognized. It also suppresses the influence of target features that may be less conspicuous due to the complexity of the background in pathology images. In comparison with other attention mechanisms, the BiFormer mechanism exhibits superior performance.

We performed a quantitative analysis of the experiment using the mAP evaluation criterion. We introduced changes only to the attention mechanism module, subsequently measuring the mAP values for each model. This allowed us to compare the mAP values among different models, assessing the compatibility between various attention mechanisms and the models. The comparative experimental data is presented in Table 3. The results indicate that

the model equipped with the BiFormer attention mechanism achieved higher detection accuracy compared to the model without any attention mechanism, as well as models incorporating the SE, ECA, CA, and CBAM attention mechanisms. Specifically, the model incorporating the ECA attention mechanism experienced a 0.28% reduction in detection accuracy. On the other hand, the models incorporating the SE attention mechanism, CA attention mechanism, and CBAM attention mechanism demonstrated improvements in detection accuracy by 0.56%, 1.02%, and 0.09%, respectively. Notably, the model's detection accuracy was enhanced by 1.48% with the inclusion of the fused BiFormer attention mechanism. These results indicate that, in comparison with the original YOLOv7 model, our model is better suited for handling pathology images.

## 4.4 Ablation experiment

This portion of the experiment investigates the impacts of the three improvement methods on the network model. The plotted data is presented in Table 4. We conducted eight sets of experiments with different modules added, and compared them with the original YOLOv7 model using metrics such as mAP, F1, Params, and FPS. For clarity and convenience, we have designated the network with the BiFormer attention module as “YOLOv7+BiFormer”, the network with the CARAFE upsampling module as “YOLOv7+CARAFE”, and the network with the GSConv convolution as “YOLOv7+GSConv”, and so forth.

As shown in Table 4, the incorporation of the BiFormer attention module, CARAFE upsampling module, and GSConv

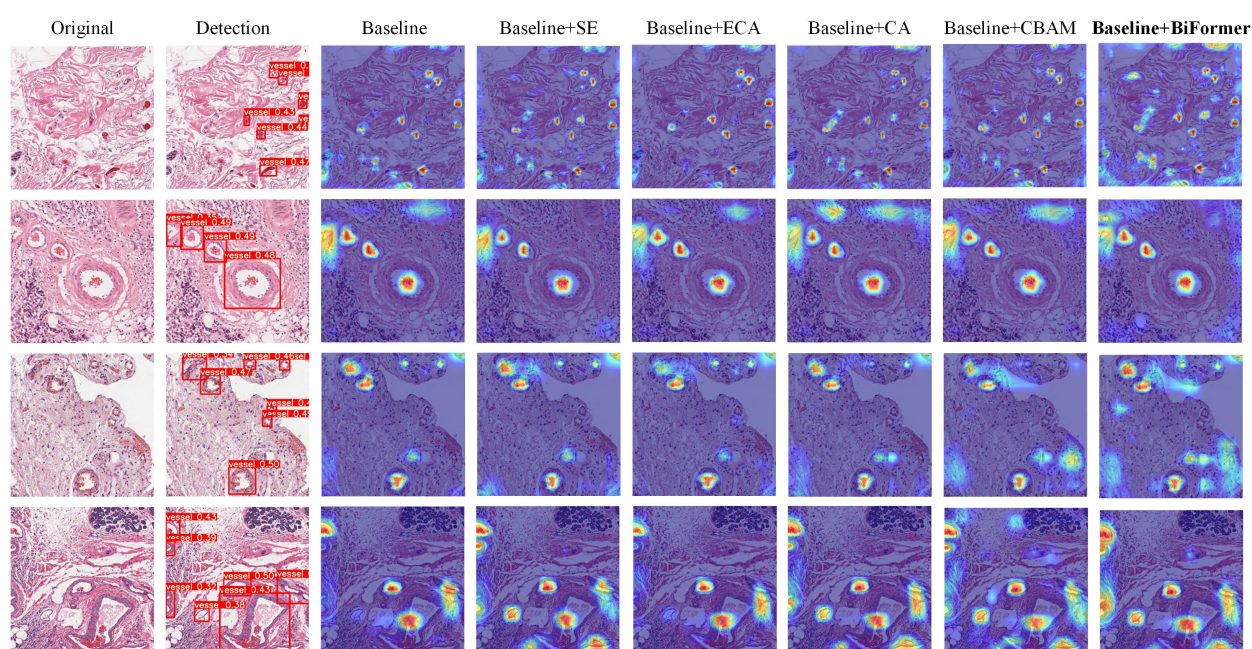


FIGURE 8

Heatmaps for various attention mechanisms. “Original” displays the dataset image. “Detection” presents the detection results of Baseline+BiFormer, while “Baseline” exhibits the heatmap of YOLOv7. “Baseline+XX” showcases the heatmap of YOLOv7 integrated with the XX attention mechanism (where XX represents SE, ECA, CA, CBAM, and BiFormer).

TABLE 3 mAP measurements for different attention mechanisms.

| Model  | Attention | Input shape | mAP(%) |
|--------|-----------|-------------|--------|
| YOLOv7 | –         | 640×640     | 84.65  |
| YOLOv7 | SENet     | 640×640     | 85.21  |
| YOLOv7 | ECA       | 640×640     | 84.37  |
| YOLOv7 | CA        | 640×640     | 85.67  |
| YOLOv7 | CBAM      | 640×640     | 84.74  |
| YOLOv7 | BiFormer  | 640×640     | 86.13  |

convolution into YOLOv7 leads to a slight improvement in the network’s detection accuracy. Specifically, these improvements are 1.48%, 0.56%, and 0.94% higher than the YOLOv7 model, respectively. This suggests that the integration of the BiFormer attention module directs the model’s attention more effectively toward the feature information of the detection target, enhancing the quality of feature mapping and significantly improving overall accuracy. However, it’s worth noting that the BiFormer attention mechanism increases the model’s complexity and reduces network inference speed. Additionally, we observed that the model’s inference speed can be substantially increased to 73.52 FPS after incorporating the CARAFE upsampling module into the network, which represents a 13.2% improvement over the original version. Furthermore, the fusion of the GSConv convolution module results in a reduction of the model’s parameters to 33.80M, a 7.3% decrease compared to the original version.

Moreover, when combining these modules in pairs, it becomes evident from the table that the combination of BiFormer + GSConv modules exhibits the most substantial improvement in model accuracy. The combination of CARAFE + GSConv modules enhances the model’s inference speed to 68.49 FPS. It is important to note that the introduction of the BiFormer attention module increases both the number of parameters and the inference time of the model. Nevertheless, we assert that this combination of

TABLE 4 The impact of the fusion of different modules of the model on the metrics.

| Methods                         | mAP (%) | F1(%) | Params (M) | FPS   |
|---------------------------------|---------|-------|------------|-------|
| YOLOv7                          | 84.65   | 82.45 | 36.48      | 64.93 |
| YOLOv7+BiFormer                 | 86.13   | 83.36 | 37.01      | 60.67 |
| YOLOv7+ CARAFE                  | 85.21   | 82.48 | 36.72      | 73.52 |
| YOLOv7+ GSConv                  | 85.59   | 83.12 | 33.80      | 66.89 |
| YOLOv7+ BiFormer + CARAFE       | 84.12   | 81.69 | 37.26      | 64.88 |
| YOLOv7+ BiFormer + GSConv       | 86.48   | 83.45 | 34.33      | 61.35 |
| YOLOv7+ CARAFE + GSConv         | 85.64   | 81.67 | 34.45      | 68.49 |
| YOLOv7+BiFormer+ CARAFE+ GSConv | 87.48   | 85.18 | 34.90      | 65.39 |

three modules is well-suited. By leveraging the CARAFE module and GSConv module for accelerated inference and lightweight deployment, the incorporation of the BiFormer attention module yields a qualitative improvement in detection accuracy. In summary, our approach demonstrates improvements in both detection accuracy and speed compared to the original model, marking a significant enhancement.

#### 4.5 Comparative experiments with other mainstream algorithms

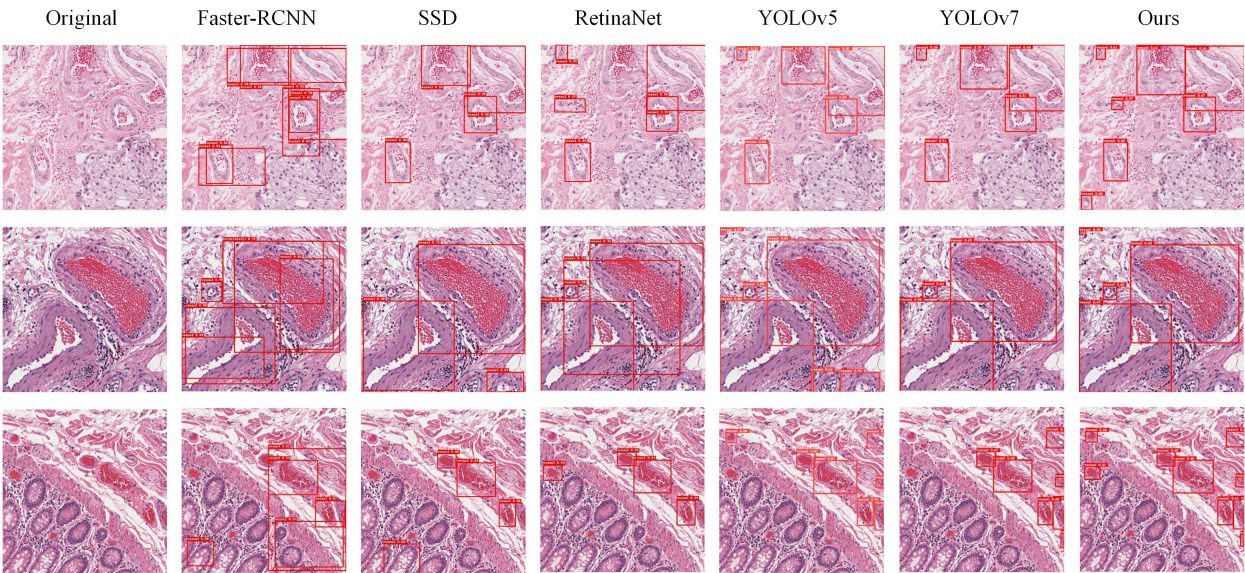
Our proposed PI-YOLO algorithm demonstrates strong feature extraction capabilities in complex pathological image scenes and achieves fast detection speeds, making it a high-performing solution for pathological image object detection. To validate the superiority of our proposed algorithm in the context of pathological images, we conducted comparisons with mainstream object detection algorithms, including Faster RCNN (38), SSD (39), RetinaNet (19), YOLOv5 (40), and YOLOv7 (13). Utilizing the same vascular dataset and training methodology, we performed both qualitative and quantitative analyses to assess the respective advantages of these algorithm models.

In our experiments, we conducted a qualitative analysis of the algorithms’ performance by examining the detection result plots of different models. Figure 9 displays the detection results of Faster-RCNN, SSD, RetinaNet, YOLOv5, YOLOv7, and PI-YOLO. From the visual results, it becomes evident that RetinaNet’s detection performance is superior to that of Faster-RCNN and SSD, and it is on par with YOLOv5. However, the number of detected targets in RetinaNet is generally lower than that in YOLOv5, and there are instances of target misclassifications. The PI-YOLO algorithm introduced in this paper exhibits better target recognition capabilities compared to Faster-RCNN, SSD, and RetinaNet. It also demonstrates fewer misclassifications and identifies a greater number of small vessels compared to YOLOv5.

To facilitate a comprehensive evaluation of the detection performance of the PI-YOLO algorithm, we conducted a comparative analysis between PI-YOLO and the leading detection algorithm in the field, YOLOv7. Figure 10. presents the detection results of both YOLOv7 and PI-YOLO on pathological sample images featuring small blood vessels with unclear edges. These vessels are highlighted with green bounding boxes. As observed, due to the indistinct edges of the small vessels, YOLOv7 struggles to distinguish them from the background, resulting in missed detections and false negatives. In contrast, PI-YOLO accurately identifies and delineates these small vessels. This improvement is attributed to the integration of the BiFormer attention mechanism, which enhances feature extraction, particularly for small and inconspicuous targets.

In our experimental assessment, we quantitatively compared and analyzed each model, employing metrics such as mAP, F1 score, Params, GFLOPS, and FPS. The outcomes of these comparative measurements are presented in Table 5. The data indicate that our PI-YOLO algorithm achieves an mAP of 87.48%, surpassing currently mainstream object detection

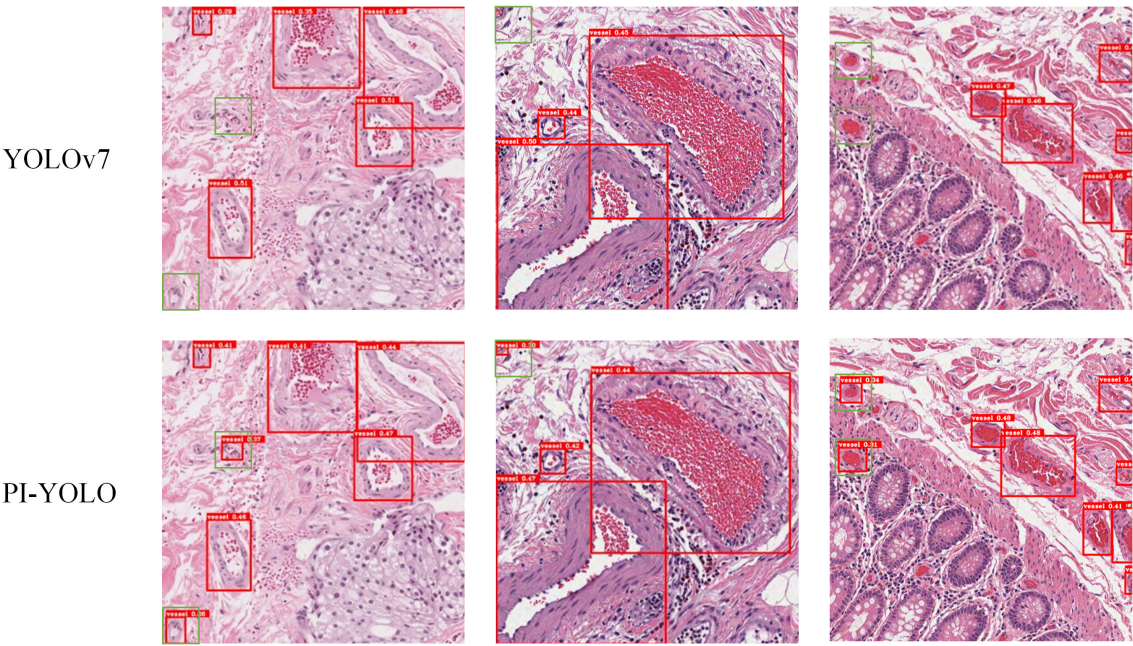




**FIGURE 9**  
Images displaying the detection results of six models. The first column, “Original,” represents the original images from the dataset. The second column shows the detection images from Faster-RCNN, the third column from SSD, the fourth column from RetinaNet, the fifth column from YOLOv5, the sixth column from YOLOv7, and the seventh column from PI-YOLO.

algorithms. In terms of detection speed, the integration of the CARAFE up-sampling module and the GSConv module has effectively reduced the model’s parameter count and increased computational speed, thereby maintaining commendable real-time performance. Notably, our enhanced PI-YOLO algorithm exhibits a significant improvement in accuracy by 18.94% and

performance by 90.97% compared to the widely used two-stage object detection algorithm, Faster-RCNN-ResNet. In contrast, when compared with the commonly adopted single-stage object detection algorithm YOLOv5, our PI-YOLO algorithm shows a 4.79% increase in mAP, although the detection speed is slightly reduced by 2.91%. Additionally, compared to the YOLOv7



**FIGURE 10**  
The detection effect of YOLOv7 and the detection effect of PI-YOLO.



TABLE 5 Performance metric values of mainstream target detection algorithms on the dataset.

| Model              | mAP (%) | F1(%) | FPS    | GFLOPS (G) | params (M) |
|--------------------|---------|-------|--------|------------|------------|
| Faster-RCNN-ResNet | 68.54   | 64.37 | 34.24  | 416.52     | 127.35     |
| SSD                | 75.23   | 73.36 | 107.64 | 215.37     | 23.72      |
| RetinaNet          | 80.34   | 74.47 | 44.69  | 120.43     | 35.56      |
| YOLOv5             | 82.69   | 80.15 | 67.35  | 115.32     | 45.53      |
| YOLOv7             | 84.65   | 82.45 | 64.93  | 103.23     | 36.48      |
| PI-YOLO            | 87.48   | 85.18 | 65.39  | 119.70     | 34.90      |

algorithm, our improved PI-YOLO algorithm shows an increase of 2.83% in mAP and a modest increase of 0.71% in detection speed, while the model size has been reduced by 1.58 M.

#### 4.6 Comparative experiments on other detection tasks in pathological images

To demonstrate the superior performance of the PI-YOLO algorithm in pathology image detection, we conducted experiments using the ICPR 2012 mitotic target detection dataset. The experimental results are presented in Figure 11. We measured the Precision, Recall, and F1 values of the model and compared them with the current state-of-the-art mitosis detection methods using different metrics. The comparison results are summarized in Table 6. While PI-YOLO falls within the middle range in terms of Precision, it achieves the highest Recall value among all methods, leading to the highest F1 score as well. These results indicate that the PI-YOLO algorithm excels in feature extraction, particularly in the context of complex pathology images from various sources and tissues.

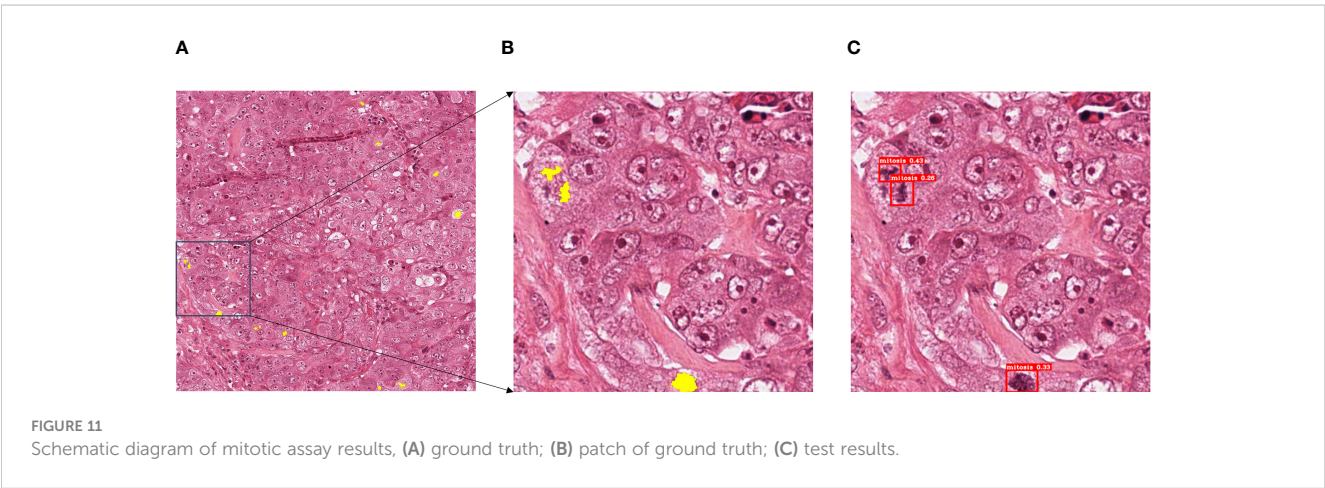
In summary, the proposed model achieves the highest detection accuracy among current mainstream detection algorithms and also maintains good detection and inference speed. The network

demonstrates significant advantages in pathology image object detection tasks.

#### 5 Conclusion

In this paper, we introduce the PI-YOLO target detection model to achieve automated blood vessel detection in pathology images using deep learning techniques. Our research addresses the challenges presented by pathology images, which include a high proportion of small targets, complex image backgrounds, dense target distribution, and subtle feature differences between the target and the background. Our model incorporates the BiFormer attention mechanism, which effectively reduces information loss during feature extraction while capturing long-range contextual dependencies. This not only saves computational resources but also enhances the overall feature extraction capabilities of the network. The integration of this attention mechanism into YOLOv7 results in improved detection accuracy for pathology images. Furthermore, by replacing the upsampling module and implementing GSConv convolution, we maintain detection accuracy while reducing model parameters and enhancing inference speed. These components, when integrated into YOLOv7, yield the enhanced PI-YOLO model. This model demonstrates superior performance in pathology image detection tasks, achieving a remarkable mAP value of 87.48%. It partially mitigates the challenges posed by complex backgrounds in pathology images. Moreover, automating blood vessel detection in pathology images significantly assists researchers in the study of anti-tumor vascular therapy, offering substantial medical value.

However, although our method is highly effective in vascular detection tasks, it currently lacks the capability to differentiate among various types of blood vessels, such as arterial, venous, and capillary. This limitation affects its specificity in tumor studies where such distinctions are crucial. Deploying PI-YOLO in clinical settings presents several challenges, including the need for high computational resources, seamless integration into existing diagnostic workflows without disruption, and robustness against variability in pathology image data due to differing laboratory



**TABLE 6** Performance index values of different methods on the ICPR 2012 mitosis detection dataset.

| Methods                | Dataset            | Precision | Recall | F1     |
|------------------------|--------------------|-----------|--------|--------|
| H. Chen et al. (41)    | MITOS-12-scanner A | 0.8040    | 0.7720 | 0.7880 |
| M. Ma et al. (42)      | MITOS-12-scanner A | 0.7760    | 0.7870 | 0.7810 |
| C. Li et al. (43)      | MITOS-12-scanner A | 0.8540    | 0.8120 | 0.8320 |
| H. Lei et al. (44)     | MITOS-12-scanner A | 0.9200    | 0.7920 | 0.8510 |
| T. Mahmood et al. (45) | MITOS-12-scanner A | 0.8760    | 0.8410 | 0.8580 |
| M. Sebai et al. (46)   | MITOS-12-scanner A | 0.9210    | 0.8110 | 0.8630 |
| ours                   | MITOS-12-scanner A | 0.8589    | 0.8769 | 0.8678 |

standards and imaging equipment. To overcome these challenges and improve the model, our future research will focus on developing methods to accurately distinguish between different blood vessel types to enhance clinical relevance in tumor analysis. We aim to optimize the detection speed of our models for real-time clinical use and expand our dataset to include a more diverse range of pathology images, thereby improving the model’s generalizability and robustness. Additionally, we plan to explore deployment on embedded devices to provide on-site assistance to medical professionals, facilitating quicker and more accurate diagnosis and treatment decisions. These steps will pave the way for the successful implementation of PI-YOLO in practical medical applications, ultimately benefiting patient care.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: <https://portal.gdc.cancer.gov/> and <https://ipal.cnr.fr/ICPR2012/>. The private dataset is not available because of privacy regulations.

Ethics statement

Ethical approval was not required for the studies on humans in accordance with the local legislation and institutional

requirements because only commercially available established cell lines were used.

Author contributions

CL: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. SC: Data curation, Resources, Validation, Writing – review & editing. HG: Investigation, Software, Visualization, Writing – review & editing. YD: Software, Writing – review & editing. YL: Writing – review & editing. JX: Writing – original draft, Writing – review & editing. LQ: Data curation, Writing – review & editing. GZ: Funding acquisition, Methodology, Project administration, Resources, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the Natural Science Foundation of Guangdong Province, grant number: 2020A1515110501; Key scientific research project of ordinary universities in Guangdong Province, grant number: 2023KCXTD026.

Conflict of interest

Author SC is employed by Guangzhou KingMed Center for Clinical Laboratory 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. Zhang M, Liu J, Liu G, Xing Z, Jia Z, Li J, et al. Anti-vascular endothelial growth factor therapy in breast cancer: Molecular pathway, potential targets, and current treatment strategies. *Cancer Lett* (2021) 520:422–33. doi: 10.1016/j.canlet.2021.08.005

2. Gillies RJ, Brown JS, Anderson ARA, Gatenby RA. Eco-evolutionary causes and consequences of temporal changes in intratumoural blood flow. *Nat Rev Cancer* (2018) 18(9):576–85. doi: 10.1038/s41568-018-0030-7

3. Folkman J. Tumor angiogenesis: therapeutic implications. *N Engl J Med* (1971) 285(21):1182–6. doi: 10.1056/NEJM197111182852108

4. Wei L, Zhu S, Li M, Li F, Wei F, Liu J, et al. High indoleamine 2, 3-dioxygenase is correlated with micro vessel density and worse prognosis in breast cancer. *Front Immunol* (2018) 9:724. doi: 10.3389/fimmu.2018.00724

5. Brekken RA, Li C, Kumar S. MINI REVIEW-Strategies for vascular targeting in tumors. *Int J Cancer* (2002) 100(2):123–30. doi: 10.1002/ijc.10462
6. Weidner N, Semple JP, Welch WR, Folkman J. Tumor angiogenesis and metastasis—correlation in invasive breast carcinoma. *N Engl J Med* (1991) 324(1):1–8. doi: 10.1056/NEJM199101033240101
7. Folkman J. Angiogenesis. *Angiogenesis. Annu. Rev. Med* (2006) 57:1–18. doi: 10.1146/annurev.med.57.121304.131306
8. Sipkins DA, Cheresch DA, Kazemi MR, Nevin LM, Bednarski MD, Li KCP, et al. Detection of tumor angiogenesis in vivo by  $\alpha v\beta 3$ -targeted magnetic resonance imaging. *Nat med* (1998) 4(5):623–6. doi: 10.1038/nm0598-623
9. Lugano R, Ramachandran M, Dimberg A. Tumor angiogenesis: causes, consequences, challenges and opportunities. *Cell Mol Life Sci* (2020) 77:1745–70. doi: 10.1007/s00018-019-03351-7
10. Esteves M, Monteiro MP, Duarte JA. The effects of vascularization on tumor development: A systematic review and meta-analysis of pre-clinical studies. *Crit Rev Oncol Hematol* (2021) 159:103245. doi: 10.1016/j.critrevonc.2021.103245
11. Teleanu RI, Chircov C, Grumezescu AM, Teleanu DM. Tumor angiogenesis and anti-angiogenic strategies for cancer treatment. *J Clin Med* (2019) 9(1):84. doi: 10.3390/jcm9010084
12. Raza SEA, Cheung L, Shaban M, Graham S, Epstein D, Pelengaris S, et al. Micro-Net: A unified model for segmentation of various objects in microscopy images. *Med Image Anal* (2019) 52:160–73. doi: 10.1016/j.media.2018.12.003
13. Wang C-Y, Bochkovskiy A, Liao H-YM. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. (2023), 7464–75. doi: 10.1109/CVPR52729.2023.00721
14. Zhu L, Wang X, Ke Z, Zhang W, Lau RWH. BiFormer: vision transformer with bi-level routing attention. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. (2023), 10323–33. doi: 10.1109/CVPR52729.2023.00995
15. Wang J, Chen K, Xu R, Liu Z, Loy CC, Lin D, et al. Carafe: Content-aware reassembly of features. In: *Proceedings of the IEEE/CVF international conference on computer vision*. (2019), 3007–16. doi: 10.1109/ICCV43118.2019
16. Li H, Li J, Wei H, Liu Z, Zhan Z, Ren Q. Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles. *arXiv preprint* (2022) arXiv:2206.02424. doi: 10.1007/s11554-024-01436-6
17. Ludovic R, Daniel R, Nicolas L, Maria K, Humayun I, Jacques K, et al. Mitosis detection in breast cancer histological images An ICPR 2012 contest. *J Pathol Inform* (2013) 4(1):8. doi: 10.4103/2153-3539.112693
18. Redmon J, Farhadi A. Yolov3: An incremental improvement. *arXiv preprint* (2018) arXiv:1804.02767. doi: 10.48550/arXiv.1804.02767
19. Lin T-Y, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. In: *Proceedings of the IEEE international conference on computer vision* (2017), 2980–8. doi: 10.1109/ICCV.2017.324
20. Ren S, He K, Girshick R, Sun J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv Neural Inf Process Syst* (2015) 28. doi: 10.1109/TPAMI.2016.2577031
21. He K, Gkioxari G, Dollár P, Girshick R. Mask r-cnn. In: *Proceedings of the IEEE international conference on computer vision* (2017), 2961–9. doi: 10.1109/ICCV.2017.322
22. Law H, Deng J. Cornernet: Detecting objects as paired keypoints. In: *Proceedings of the European conference on computer vision (ECCV)* (2018), 734–50.
23. Zhu C, He Y, Savvides M. Feature selective anchor-free module for single-shot object detection. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2019), 840–9. doi: 10.1109/CVPR41558.2019
24. Yi F, Yang L, Wang S, Guo L, Huang C, Xie Y, et al. Microvessel prediction in H&E Stained Pathology Images using fully convolutional neural networks. *BMC bioinform* (2018) 19:1–9. doi: 10.1186/s12859-018-2055-z
25. Fraz M, Shaban M, Graham S, Khurram SA, Rajpoot NM. Uncertainty driven pooling network for micro vessel segmentation in routine histology images. In: *Computational Pathology and Ophthalmic Medical Image Analysis: First International Workshop, COMPAY 2018, and 5th International Workshop, OMIA 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16–20, 2018, Proceedings 5*. Springer International Publishing. (2018), 156–64.
26. Fraz MM, Khurram SA, Graham S, Shaban M, Hassan M, Loya A, et al. FABnet: feature attention-based network for simultaneous segmentation of micro vessels and nerves in routine histology images of oral cancer. *Neural Comput Appl* (2020) 32:9915–28. doi: 10.1007/s00521-019-04516-y
27. Hamidinekoo A, Kelsey A, Trahearn N, Selfe J, Shipley J, Yuan Y. Automated quantification of blood microvessels in hematoxylin and eosin whole slide images. In: *MICCAI Workshop on Computational Pathology*, PMLR. (2021), 94–104.
28. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, et al. An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell* (2018) 173(2):400–16.e11. doi: 10.1016/j.cell.2018.02.052
29. Lin T-Y, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), 2117–25. doi: 10.1109/CVPR.2017.106
30. Liu S, Qi L, Qin H, Shi J, Jia J. Path aggregation network for instance segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018), 8759–68. doi: 10.1109/CVPR.2018.00913
31. Ren S, Zhou D, He S, Feng J, Wang X. Shunted self-attention via multi-scale token aggregation. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2022), 10853–62. doi: 10.1109/CVPR52688.2022.01058
32. Han K, Wang Y, Tian Q, Guo J, Xu C, Xu C. Ghostnet: More features from cheap operations. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2020), 1580–9. doi: 10.1109/CVPR42600.2020
33. Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018), 7132–41. doi: 10.1109/CVPR.2018.00745
34. Wang Q, Wu B, Zhu P, Li P, Zuo W, Hu Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2020), 11534–42. doi: 10.1109/CVPR42600.2020
35. Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2021), 13713–22. doi: 10.1109/CVPR46437.2021.01350
36. Woo S, Park J, Lee J-Y, Kweon IS. Cbam: Convolutional block attention module. In: *Proceedings of the European conference on computer vision (ECCV)* (2018), 3–19.
37. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: *Proceedings of the IEEE international conference on computer vision* (2017), 618–26. doi: 10.1109/ICCV.2017.74
38. Girshick R. Fast r-cnn. In: *Proceedings of the IEEE international conference on computer vision* (2015), 1440–8. doi: 10.1109/ICCV.2015.169
39. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C-Y, et al. Ssd: Single shot multibox detector. In: *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016. Proceedings, Part I*. Springer International Publishing (2016) 14:21–37.
40. Jocher G, Chaurasia A, Stoken A, Borovec J, Kwon Y, Michael K, et al. ultralytics/yolov5: v7. 0-yolov5 sota realtime instance segmentation. *Zenodo* (2022). doi: 10.5281/zenodo.3908559
41. Chen H, Dou Q, Wang X, Qin J, Heng P. Mitosis detection in breast cancer histology images via deep cascaded networks. In: *Proceedings of the AAAI conference on artificial intelligence* (2016) 30. doi: 10.1609/aaai.v30i1.10140
42. Ma M, Shi Y, Li W, Gao Y, Xu J. A novel two-stage deep method for mitosis detection in breast cancer histology images. In: *2018 24th International Conference on Pattern Recognition (ICPR)*, IEEE (2018), 3892–7. doi: 10.1109/ICPR.2018.8546192
43. Li C, Wang X, Liu W, Latecki LJ. DeepMitosis: Mitosis detection via deep detection, verification and segmentation networks. *Med image anal* (2018) 45:121–33. doi: 10.1016/j.media.2017.12.002
44. Lei H, Liu S, Xie H, Kuo JY, Lei B. An improved object detection method for mitosis detection. In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE (2019), 130–3. doi: 10.1109/EMBC43219.2019
45. Mahmood T, Arsalan M, Owais M, Lee MB, Park KR. Artificial intelligence-based mitosis detection in breast cancer histopathology images using faster R-CNN and deep CNNs. *J Clin Med* (2020) 9(3):749. doi: 10.3390/jcm9030749
46. Sebai M, Wang T, Al-Fadhli SA. PartMitosis: a partially supervised deep learning framework for mitosis detection in breast cancer histopathology images. *IEEE Access* (2020) 8:45133–47. doi: 10.1109/Access.6287639



## OPEN ACCESS

## EDITED BY

Poonam Yadav,  
Northwestern University, United States

## REVIEWED BY

Rahul Joshi,  
Symbiosis International University, India  
Maofeng Wang,  
Affiliated Dongyang Hospital of Wenzhou  
Medical University, China

## \*CORRESPONDENCE

Qing Feng  
✉ 245268195@qq.com  
Tao Li  
✉ li966511@163.com

<sup>†</sup>These authors have contributed  
equally to this work and share  
first authorship

RECEIVED 06 February 2024

ACCEPTED 06 August 2024

PUBLISHED 29 August 2024

## CITATION

Lan H, Wei C, Xu F, Yang E, Lu D, Feng Q and  
Li T (2024) 2.5D peritumoural radiomics  
predicts postoperative recurrence in  
stage I lung adenocarcinoma.  
*Front. Oncol.* 14:1382815.  
doi: 10.3389/fonc.2024.1382815

## COPYRIGHT

© 2024 Lan, Wei, Xu, Yang, Lu, 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.

# 2.5D peritumoural radiomics predicts postoperative recurrence in stage I lung adenocarcinoma

Haimei Lan<sup>1†</sup>, Chaosheng Wei<sup>1†</sup>, Fengming Xu<sup>1</sup>, Eqing Yang<sup>1</sup>,  
Dayu Lu<sup>2</sup>, Qing Feng<sup>1\*</sup> and Tao Li<sup>1\*</sup>

<sup>1</sup>Department of Radiology, Liuzhou Workers Hospital, Liuzhou, Guangxi, China, <sup>2</sup>Department of Radiology, Longtan Hospital, Liuzhou, Guangxi, China

**Objective:** Radiomics can non-invasively predict the prognosis of a tumour by applying advanced imaging feature algorithms. The aim of this study was to predict the chance of postoperative recurrence by modelling tumour radiomics and peritumour radiomics and clinical features in patients with stage I lung adenocarcinoma (LUAD).

**Materials and methods:** Retrospective analysis of 190 patients with postoperative pathologically confirmed stage I LUAD from centre 1, who were divided into training cohort and internal validation cohort, with centre 2 added as external validation cohort. To develop a combined radiation-clinical omics model nomogram incorporating clinical features based on images from low-dose lung cancer screening CT plain for predicting postoperative recurrence and to evaluate the performance of the nomogram in the training cohort, internal validation cohort and external validation cohort.

**Results:** A total of 190 patients were included in the model in centre 1 and randomised into a training cohort of 133 and an internal validation cohort of 57 in a ratio of 7:3, and 39 were included in centre 2 as an external validation cohort. In the training cohort (AUC=0.865, 95% CI 0.824-0.906), internal validation cohort (AUC=0.902, 95% CI 0.851-0.953) and external validation cohort (AUC=0.830, 95% CI 0.751-0.908), the combined radiation-clinical omics model had a good predictive ability. The combined model performed significantly better than the conventional single-modality models (clinical model, radiomic model), and the calibration curve and decision curve analysis (DCA) showed high accuracy and clinical utility of the nomogram.

**Abbreviations:** 2.5D, 2.5-dimension; 3D, Three-dimension; AUC, Area under the receiver operating characteristic curve; NSCLC, Non-small cell lung cancer; ROC, Receiver operating characteristic curve; ROI, Region of interest; SCLC, Small cell lung cancer; LUAD, Lung adenocarcinoma; LUSC, Lung squamous carcinoma; ICC, Intraclass correlation coefficient; TNM, Tumor, node, and metastasis; PPV, Positive predictive value; NPV, Negative predictive value.



**Conclusion:** The combined preoperative radiation-clinical omics model provides good predictive value for postoperative recurrence in stage I LUAD and combines the model's superiority in both internal and external validation cohorts, demonstrating its potential to aid in postoperative treatment strategies.

#### KEYWORDS

radiomics, lung adenocarcinoma, postoperative recurrence, nomogram, peritumoral regions

## 1 Introduction

Lung cancer is a significant contributor to global cancer mortality (1). Lung cancer is classified into two main groups (2, 3): non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC), of which about 85% of patients belong to NSCLC, which includes lung adenocarcinoma (LUAD), lung squamous carcinoma (LUSC), and other histological subtypes. In NSCLC patients, LUAD accounts for the largest proportion. With the development of low-dose computed tomography (LDCT) lung cancer screening, a large number of patients with early-stage NSCLC have been screened, and in particular, a considerable number of patients with stage I LUAD have been screened (4, 5), for which surgical resection is the preferred treatment (6). However, studies have found that the risk of recurrence remains high, even with a 20–50% recurrence rate for completely resected stage I LUAD (7). Therefore, assessment of postoperative recurrence is crucial for the prognosis of stage I LUAD.

Currently, most studies have focused on assessing benign and malignant tumours (8), disregarding the prognostic impact of subtle changes in the peritumoural microenvironment (9, 10). Furthermore, studies on the prognosis of LUAD have primarily concentrated on evaluating the prognosis of intermediate and advanced lung cancer based on genes and treatment regimens (11–14), while neglecting the impact of certain clinical factors such as immunohistochemistry and density on the prognosis. It is important to note that due to the heterogeneity of tumours (15, 16), even at the same stage, the prognosis can vary significantly. Moreover, most of the previous studies have been on two-dimension (2D) and three-dimension (3D) prognostic models (17, 18), and nowadays some scholars have started to study 2.5-dimension (2.5D) models (19) as well. Through the peritumoural radiomics prognostic study of stage I LUAD (20, 21), this study not only makes up for the shortcomings of previous studies, but also develops a new 2.5D peritumoural radiation-clinical omics model.

Compared with previous 2D or 3D radiomics features, the method is newer and more effective in studying the prognosis of LUAD.

## 2 Materials and methods

### 2.1 Patient selection and follow-up

This retrospective study was approved by two institutional review boards of the Guangxi Zhuang Autonomous Region (NO.LW2024009), exempting patients from informed consent. We collected medical records of all patients with stage I LUAD who underwent surgical resection and were pathologically confirmed between January 2010 and December 2018 at the centre 1. The inclusion criteria (1): underwent surgical complete resection of the lung lesion (2); postoperative pathological diagnosis of invasive stage I lung adenocarcinoma (3); CT examination within 2 weeks before surgery. The exclusion criteria (1): the presence of multiple primary cancers or other malignancies in the lungs (2); preoperative neoadjuvant therapy (3); failure to complete postoperative follow-up (4); CT image artefacts that severely impaired the visualisation of the tumour (5); absence of low-dose lung cancer screening CT plain images prior to surgery.

A total of 190 patients with stage I LUAD were included in centre 1 and randomised into two cohorts in a ratio of 7:3. The training cohort consisted of 133 patients, while the internal validation cohort had 57 patients. Additionally, 39 patients with stage I LUAD in centre 2 were collected as the external testing cohort from January 2016 to December 2018, following the same inclusion and exclusion criteria. A postoperative follow-up was conducted, including computed tomography (CT) and/or magnetic resonance imaging (MRI), PET-CT. Recurrence was defined as local recurrence and distant metastasis, as per relevant studies. Local recurrence included recurrence in N1 lymph nodes, N2 lymph nodes, mediastinum, primary lung or pleura. Distant metastases



included metastases to the adrenal gland, kidney, bone, brain, liver, contralateral lung, skin or N3 (22).

## 2.2 Clinical characteristic

Basic patient information and clinical variables including age, sex, white blood cell (WBC), neutrophils (NEU), C-reactive protein (C-RP), carcinoembryonic antigen (CEA), cytokeratin 19 fragment assay (CYFRA21-1), neuron-specific enolase assay (NSE), carbohydrate antigen (CA) 125, CA153, squamous cell carcinoma-associated antigen (SCCA), CA50, CA242, CA724, Ki-67, location of the tumour, distance from the pleura, T-stage, and type of nodule.

We divided the age into two groups: less than 65 years old and greater than or equal to 65 years old; T-stage was determined by experienced radiologists from preoperative CT images, based on the 9th edition of the TNM staging system for lung cancer, and was divided into T1a, T1b, and T1c; the division of the content of Ki67 is still controversial, and we used less than 10% for low expression and greater than or equal to 10% for high expression; and the type of nodules of stage I LUAD that we included showed mixed ground glass nodules (mGGN) and solid nodules (SN).

## 2.3 Procedure

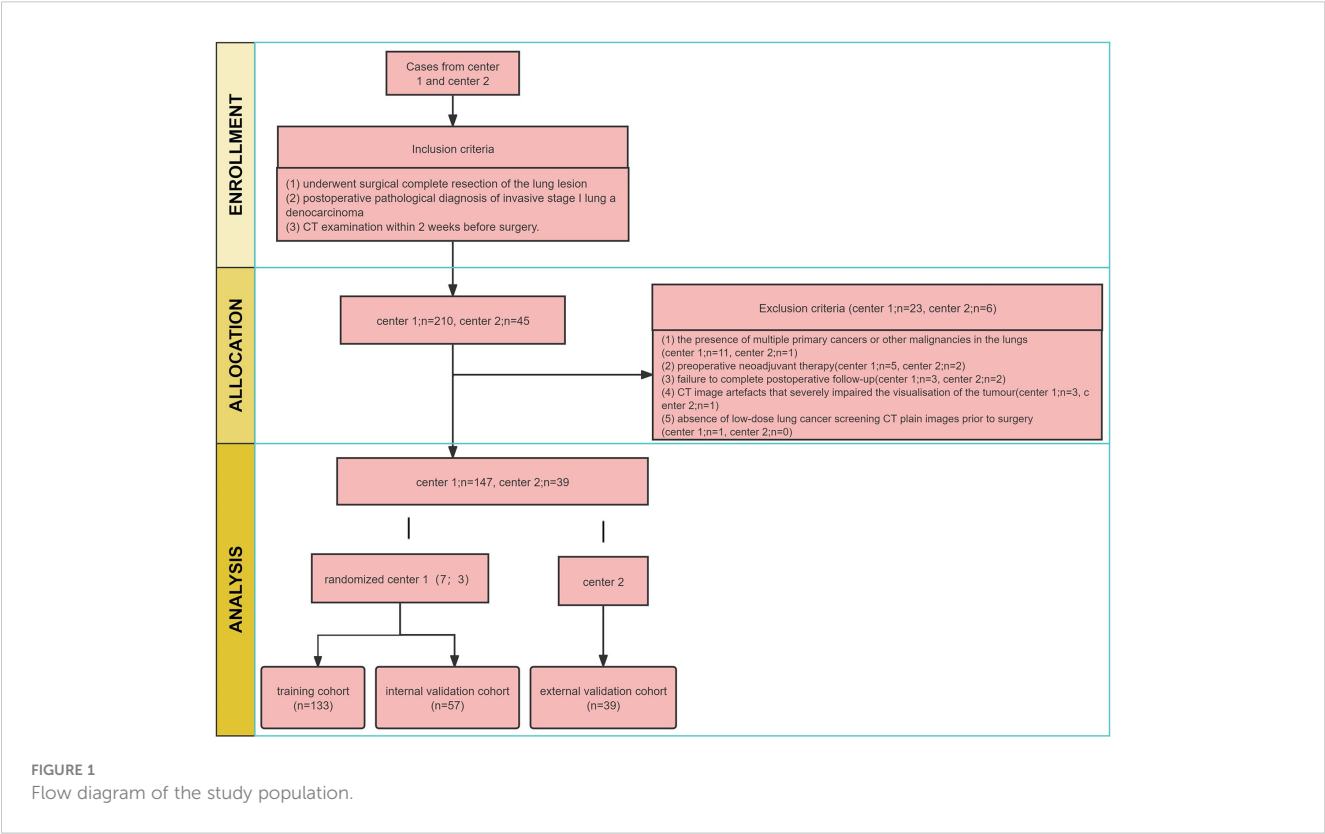
The study workflow is summarized in Figure 1, and the radiomics modelling pipeline in Figure 2.

## 2.4 CT image acquisition

The scanning machine at both hospitals was SIEMENS SOMATOM Definition Flash (Stellar) with the same lung scanning parameters. All CT scans were performed from the tip of the lungs to the base of the lungs, and the parameters of the scan reconstruction were: Tube voltage=120kV, Effective power of tube=30mAs, Detector collimation=128 × 0.625mm, Matrix=512×512, Slice thickness=0.625mm, CDTIvol=2.03mGy.

## 2.5 Radiomics feature extraction and feature selection

The DICOM format images of the patients were downloaded from the Picture Archiving and Communication Systems (PACS) and imported into the Darwin Intelligent Science Research Platform. The process of tumour region segmentation and radiomics feature extraction involves the following steps (1): Modal settings: the modal parameters for each patient were set to tumour body, peritumoural 3mm, peritumoural 6mm, peritumoural 9mm, peritumoural 12mm and peritumoural 15mm, and the window widths and window positions were uniformly set to 1200 and -600 (2); 2.5D region of interest (ROI) segmentation: The ROI was manually delineated on the CT images by two radiologists with 10 years of experience. For each CT image, the radiologist selected the largest section of the tumour on the Darwin Intelligent Science



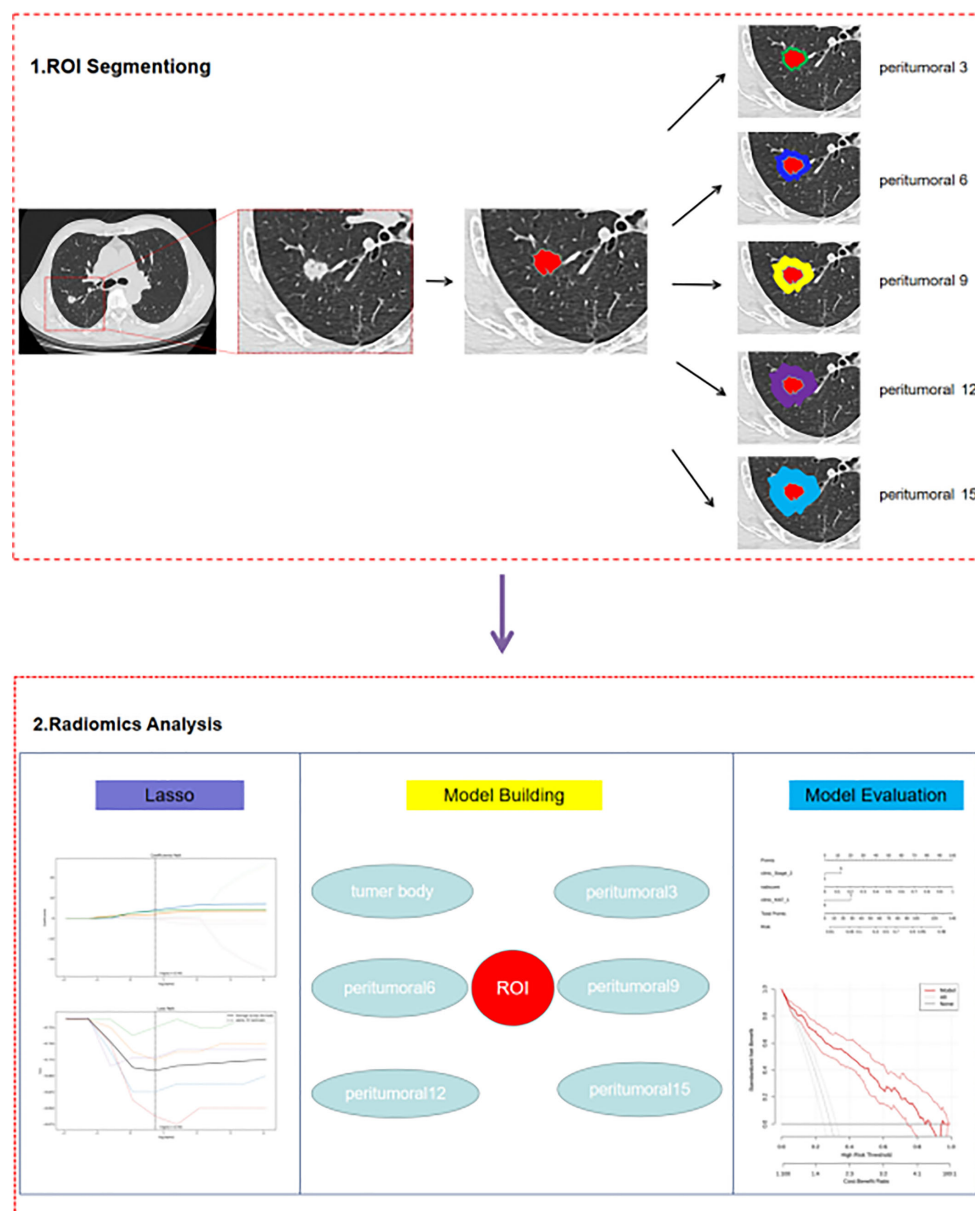


FIGURE 2  
The radiomics modelling pipeline.

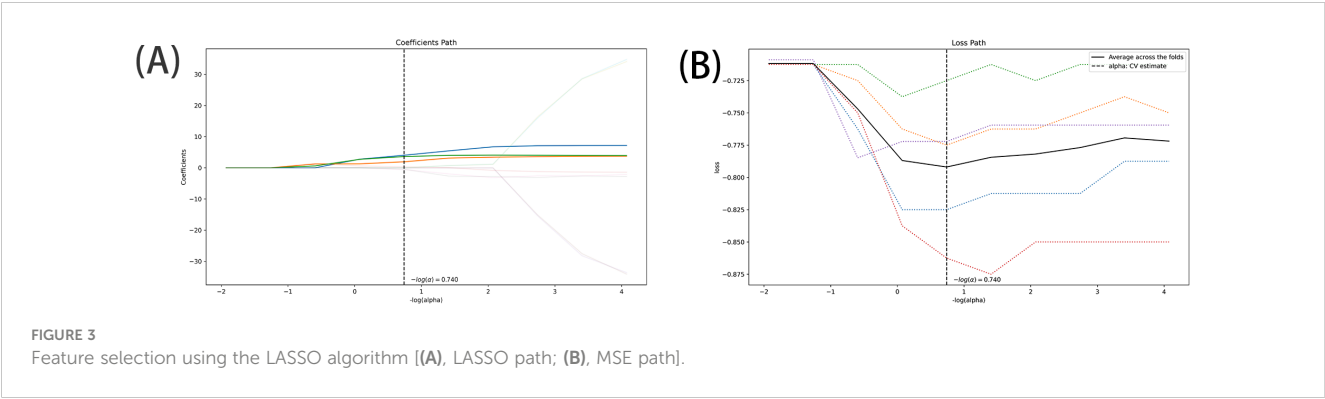
research platform (23) to draw a ROI, and then selected the forward and backward angles of 45° on this section to draw two ROIs. These three ROIs were then merged to create a 2.5D ROI for each tumour. In outlining ROIs, we exclude pleural walls, thick bronchial tubes, and blood vessels (3); A total of 1125 radiomics features were extracted using the Darwin Intelligent Science Research Platform (4); A minimum-maximum normalised, optimal feature filter was used to assess the linear correlation between each feature and the lesion category labels, and the 40 most relevant features were filtered out of 1125 features. The least absolute shrinkage and selection operator (LASSO) algorithm was used to select the most relevant features from 40 features (Figure 3). Finally, a total of 10 features most relevant to recurrence after surgery for stage I LUAD were selected and used to construct a prediction model (Figure 4).

## 2.6 Intra-observer and inter-observer consistency

We used intraclass correlation coefficient (ICC) to assess intra- and inter-observer correlation coefficients. A total of 49 patients were randomly selected from the training set, and ROI segmentation was independently performed by two physicians. We considered these features to be stable when the ICC value was greater than 0.80.

## 2.7 Model construction and validation

To predict postoperative recurrence of stage I LUAD, we performed univariate and multifactorial logistic regression (LR)



analyses to select statistically significant clinical characteristics ( $P < 0.05$ ) for clinical modelling. Clinical characteristics that were statistically significant for postoperative recurrence of stage I LUAD were retained in the univariate logistic regression analysis ( $p < 0.05$ ). Variables with  $p < 0.05$  in the multifactorial analysis were identified as independent predictors associated with postoperative recurrence and were included in the construction of the clinical model. Combining the radiomics model with the clinical model to create a joint model with different parameters. Receiver operating characteristic (ROC) curves were plotted, and area under the receiver operating characteristic curve (AUC) was calculated to assess the efficacy of each model. We compared thirteen predictive models, including six parameter radiomics models, clinical model, and six radiation-clinical omics models incorporating clinical factors. The best models were then selected from these to draw nomogram, and we used deLong tests, calibration curves, and decision curve analyses (DCA) to test the accuracy and clinical utility of the nomogram.

## 2.8 Statistical analysis

SPSS 24.0 was used for statistical analysis. Continuous variables were presented as mean  $\pm$  standard deviation and compared using independent samples t-test. Categorical variables were presented as percentage counts and compared using chi-square test. The model's

goodness of fit was assessed using the Hosmer-Lemeshow test, which showed no statistically significant difference ( $P > 0.05$ ), indicating good model fit. To comprehensively evaluate the predictive efficacy of different models, we used ROC curve, AUC, accuracy, sensitivity, specificity, Positive predictive value (PPV), and Negative predictive value (NPV). All statistical tests were two-sided with a significance level of  $p < 0.05$ .

## 3 Results

### 3.1 Patient clinical baseline characteristics

Table 1 lists and compares the clinical baseline characteristics of the analysed patients.

### 3.2 Establishment of clinical models

Logistic regression analysis was used to assess 19 possible risk factors. Univariate and multifactorial logistic regression analyses were performed on clinical indicators in training cohort of 133 patients with postoperative recurrence of stage I LUAD (Table 2). Univariate logistic regression analysis showed that T1c in T-stage, CEA, NSE,  $\geq 10\%$  in Ki67, and SN in nodal type were statistically significant for postoperative recurrence of stage ILUAD. For

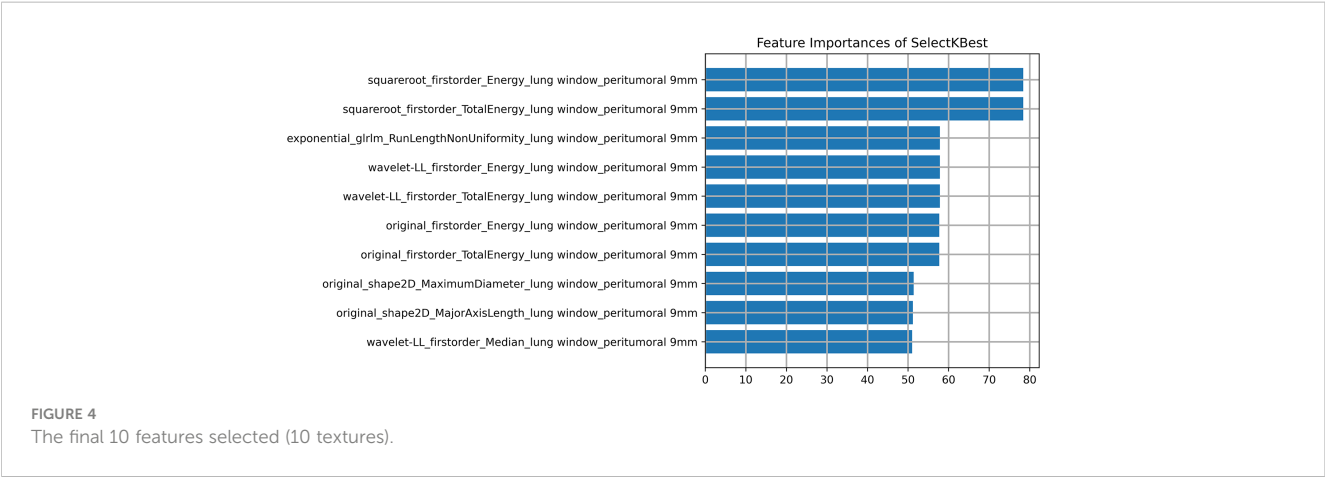


TABLE 1 Clinical baseline characteristics.

|                                  | Training Set<br>(n1 = 133) | Internal Test Set<br>(n2 = 57) | p value<br>(n1 VS n2) | External Test Set<br>(n3 = 39) | p value<br>(n1 VS n3) |
|----------------------------------|----------------------------|--------------------------------|-----------------------|--------------------------------|-----------------------|
| Age                              |                            |                                | 0.838                 |                                | 0.571                 |
| <65                              | 88(66.2)                   | 43 (75.4)                      |                       | 25 (64.1)                      |                       |
| ≥65                              | 45(33.8)                   | 14 (24.6)                      |                       | 14 (35.9)                      |                       |
| Sex                              |                            |                                | 0.777                 |                                | 0.705                 |
| Male                             | 66 (49.6)                  | 27 (47.4)                      |                       | 18 (46.2)                      |                       |
| Female                           | 67 (50.4)                  | 30 (52.6)                      |                       | 21 (53.8)                      |                       |
| Stage                            |                            |                                | 0.213                 |                                | 0.371                 |
| T1a                              | 8 (6.0)                    | 7 (12.2)                       |                       | 4 (10.2)                       |                       |
| T1b                              | 77 (57.9)                  | 38 (66.7)                      |                       | 23 (59.0)                      |                       |
| T1c                              | 48 (36.1)                  | 12 (21.1)                      |                       | 12 (30.8)                      |                       |
| WBC (Mean ± SD)                  | 6.97 ± 2.34                | 6.93 ± 2.06                    | 0.911                 | 6.77 ± 1.87                    | 0.707                 |
| NEU (Mean ± SD)                  | 4.30 ± 2.04                | 4.14 ± 1.48                    | 0.604                 | 3.99 ± 1.53                    | 0.441                 |
| C-RP (Mean ± SD)                 | 6.08 ± 14.58               | 2.66 ± 3.63                    | 0.093                 | 4.56 ± 14.98                   | 0.585                 |
| CEA (Mean ± SD)                  | 5.38 ± 7.90                | 3.82 ± 4.14                    | 0.082                 | 7.02 ± 18.99                   | 0.431                 |
| CYFRA21-1 (Mean ± SD)            | 3.09 ± 1.74                | 3.55 ± 2.28                    | 0.175                 | 2.97 ± 1.86                    | 0.776                 |
| NSE (Mean ± SD)                  | 13.26 ± 3.69               | 13.93 ± 6.19                   | 0.392                 | 13.18 ± 4.74                   | 0.792                 |
| CA125 (Mean ± SD)                | 15.29 ± 12.00              | 12.03 ± 16.85                  | 0.156                 | 14.09 ± 15.03                  | 0.678                 |
| CA15-3 (Mean ± SD)               | 15.68 ± 15.77              | 13.17 ± 9.26                   | 0.294                 | 19.78 ± 16.23                  | 0.318                 |
| SCCA (Mean ± SD)                 | 1.19 ± 0.79                | 1.42 ± 0.86                    | 0.088                 | 1.55 ± 0.79                    | 0.335                 |
| CA50 (Mean ± SD)                 | 9.56 ± 14.43               | 23.93 ± 71.84                  | 0.163                 | 8.32 ± 9.80                    | 0.568                 |
| CA242 (Mean ± SD)                | 6.59 ± 6.10                | 12.56 ± 34.26                  | 0.222                 | 5.09 ± 4.91                    | 0.205                 |
| CA72-4 (Mean ± SD)               | 4.42 ± 9.16                | 3.82 ± 7.78                    | 0.682                 | 5.04 ± 13.42                   | 0.750                 |
| Ki67                             |                            |                                | 0.078                 |                                | 0.095                 |
| <10%                             | 68 (51.1)                  | 26 (45.6)                      |                       | 14 (35.9)                      |                       |
| ≥10%                             | 65 (48.9)                  | 31 (54.4)                      |                       | 25 (64.1)                      |                       |
| Location                         |                            |                                | 0.854                 |                                | 0.580                 |
| Left superior lobar              | 37 (27.8)                  | 15 (26.3)                      |                       | 14 (35.9)                      |                       |
| Right superior lobar             | 45 (33.8)                  | 20 (35.1)                      |                       | 10 (25.7)                      |                       |
| Right middle lobar               | 8 (6.0)                    | 4 (7.0)                        |                       | 7 (17.9)                       |                       |
| Right inferior lobar             | 27 (20.3)                  | 14 (24.6)                      |                       | 2 (5.1)                        |                       |
| Left inferior lobar              | 16 (12.0)                  | 4 (7.0)                        |                       | 6 (15.4)                       |                       |
| Distance from pleura (Mean ± SD) | 1.45 ± 0.73                | 1.50 ± 0.65                    | 0.640                 | 1.50 ± 0.65                    | 0.402                 |
| Nodule type                      |                            |                                | 0.920                 |                                | 0.079                 |
| mGGN                             | 43 (32.3)                  | 18 (31.6)                      |                       | 18 (9.5)                       |                       |
| SN                               | 90 (67.7)                  | 39 (68.4)                      |                       | 39 (20.5)                      |                       |
| Recurrence                       |                            |                                | 0.862                 |                                | 0.628                 |
| Yes                              | 39 (29.3)                  | 16 (28.1)                      |                       | 11 (28.2)                      |                       |
| No                               | 94 (70.7)                  | 41 (71.9)                      |                       | 28 (71.8)                      |                       |

TABLE 2 Univariate and multivariate analysis.

|                      | N(100%)  | OR(95%CI)            | p value      | OR(95%CI)             | p value      |
|----------------------|----------|----------------------|--------------|-----------------------|--------------|
| Age                  |          |                      |              |                       |              |
| <65                  | 88(66.2) | 1.000                |              |                       |              |
| ≥65                  | 45(33.8) | 0.969(0.439-2.136)   | 0.937        |                       |              |
| Sex                  |          |                      |              |                       |              |
| Male                 | 66(49.6) | 1.000                |              |                       |              |
| Female               | 67(50.4) | 1.270(0.601-2.685)   | 0.531        |                       |              |
| Stage                |          |                      |              |                       |              |
| T1a                  | 8(6.0)   | 1.000                |              |                       |              |
| T1b                  | 77(7.9)  | 3.500(0.437-28.004)  | 0.238        | 4.092(0.497-33.707)   | 0.190        |
| T1c                  | 48(36.1) | 4.549(2.135-9.695)   | <b>0.000</b> | 14.237(1.704-118.970) | <b>0.014</b> |
| WBC                  |          | 1.123(0.962-1.311)   | 0.141        | –                     | –            |
| NEU                  |          | 1.165(0.974-1.392)   | 0.095        | –                     | –            |
| C-RP                 |          | 1.013(0.989-1.039)   | 0.293        | –                     | –            |
| CEA                  |          | 1.095(1.020-1.175)   | <b>0.012</b> | –                     | –            |
| CYFRA21-1            |          | 1.182(0.943-1.481)   | 0.147        | –                     | –            |
| NSE                  |          | 1.126(1.005-1.263)   | <b>0.041</b> | 1.215(1.032-1.430)    | <b>0.020</b> |
| CA125                |          | 1.020(0.988-1.052)   | 0.223        | –                     | –            |
| CA153                |          | 1.013(0.989-1.038)   | 0.304        | –                     | –            |
| SCCA                 |          | 0.752(0.428-1.320)   | 0.320        | –                     | –            |
| CA50                 |          | 1.010(0.985-1.036)   | 0.430        | –                     | –            |
| CA242                |          | 1.054(0.991-1.122)   | 0.097        | –                     | –            |
| CA724                |          | 1.028(0.984-1.073)   | 0.216        | –                     | –            |
| Ki67                 |          |                      |              |                       |              |
| <10%                 | 68(51.1) | 1.000                |              |                       |              |
| ≥10%                 | 65(48.9) | 10.656(4.044-28.078) | <b>0.000</b> | 0.081(0.020-0.322)    | <b>0.000</b> |
| Location             |          |                      |              |                       |              |
| Left superior lobar  | 37(27.8) | 1.000                |              | –                     | –            |
| Right superior lobar | 45(33.8) | 0.758(0.292-1.967)   | 0.568        | –                     | –            |
| Right middle lobar   | 8(6.0)   | 1.250(0.255-6.119)   | 0.783        | –                     | –            |
| Right inferior lobar | 27(20.3) | 0.595(0.191-1.859)   | 0.372        | –                     | –            |
| Left inferior lobar  | 16(12.0) | 1.250(0.368-4.251)   | 0.721        | –                     | –            |
| Distance from pleura |          | 0.822(0.470-1.439)   | 0.492        | –                     | –            |
| Nodule type          |          |                      |              |                       |              |
| mGGN                 | 43(32.3) | 1.000                |              |                       |              |
| SN                   | 90(67.7) | 6.205(2.039-18.881)  | <b>0.001</b> | 4.541(1.716-12.014)   | <b>0.002</b> |

mGGN, Mmixed ground-glass nodule; SN, Solid nodules; SD, Standard deviation. Bolded indicators are meaningful. Values in bold indicate statistical significance.



TABLE 3 Diagnostic effectiveness of different models.

|                         | AUC (95%CI)         | Accuracy | Sensitivity | Specificity | PPV   | NPV   |
|-------------------------|---------------------|----------|-------------|-------------|-------|-------|
| Training Set            |                     |          |             |             |       |       |
| Clinical model          | 0.772 (0.723-0.820) | 0.707    | 0.791       | 0.673       | 0.495 | 0.888 |
| Tumor body              | 0.762 (0.711-0.813) | 0.659    | 0.809       | 0.599       | 0.449 | 0.885 |
| Peritumoral 3mm         | 0.763 (0.711-0.815) | 0.632    | 0.852       | 0.542       | 0.430 | 0.901 |
| Peritumoral 6mm         | 0.708 (0.651-0.766) | 0.717    | 0.548       | 0.785       | 0.508 | 0.811 |
| Peritumoral 9mm         | 0.785 (0.734-0.837) | 0.724    | 0.757       | 0.711       | 0.515 | 0.878 |
| Peritumoral 12mm        | 0.677 (0.616-0.739) | 0.654    | 0.722       | 0.627       | 0.439 | 0.848 |
| Peritumoral 15mm        | 0.791 (0.744-0.838) | 0.714    | 0.687       | 0.725       | 0.503 | 0.851 |
| Tumor body+Clinic       | 0.855 (0.817-0.893) | 0.757    | 0.896       | 0.701       | 0.548 | 0.943 |
| Peritumoral 3mm+Clinic  | 0.861 (0.823-0.899) | 0.799    | 0.765       | 0.813       | 0.624 | 0.895 |
| Peritumoral 6mm+Clinic  | 0.836 (0.794-0.878) | 0.779    | 0.783       | 0.778       | 0.588 | 0.898 |
| Peritumoral 9mm+Clinic  | 0.865 (0.824-0.906) | 0.832    | 0.730       | 0.873       | 0.700 | 0.889 |
| Peritumoral 12mm+Clinic | 0.851 (0.810-0.892) | 0.820    | 0.696       | 0.870       | 0.684 | 0.876 |
| Peritumoral 15mm+Clinic | 0.855 (0.816-0.895) | 0.767    | 0.809       | 0.750       | 0.567 | 0.906 |
| Internal Test Set       |                     |          |             |             |       |       |
| Clinical model          | 0.779 (0.703-0.855) | 0.737    | 0.760       | 0.727       | 0.535 | 0.880 |
| Peritumoral 9mm         | 0.815 (0.742-0.888) | 0.813    | 0.640       | 0.884       | 0.696 | 0.856 |
| Peritumoral 9mm+Clinic  | 0.902 (0.851-0.953) | 0.871    | 0.720       | 0.934       | 0.818 | 0.890 |
| External Test Set       |                     |          |             |             |       |       |
| Clinical model          | 0.773 (0.732-0.814) | 0.721    | 0.764       | 0.704       | 0.512 | 0.880 |
| Peritumoral 9mm         | 0.712 (0.603-0.820) | 0.795    | 0.424       | 0.940       | 0.737 | 0.806 |
| Peritumoral 9mm+ Clinic | 0.830 (0.751-0.908) | 0.821    | 0.727       | 0.857       | 0.667 | 0.889 |

statistically significant clinical characteristics, multifactorial logistic regression analysis was used, which showed that NSE,  $\geq 10\%$  in Ki67, T-stage in T1c and SN in nodule type were independent risk factors for postoperative recurrence and could be used to establish clinical models.

### 3.3 Performance and comparison of models

In this study, we developed 13 models, including the radiomics models with 6 parameters (tumour body, peritumoral 3mm, peritumoral 6mm, peritumoral 9mm, peritumoral 12 mm, peritumoral 15mm), the clinical model, and the six-parameter radiation-clinical omics models that incorporates clinical factors, and evaluated the performance of all the models. Table 3 displays the AUC, accuracy, sensitivity, specificity, PPV, and NPV of various models. In the training cohort, the peritumoral 9mm model (AUC= 0.785) outperformed the clinical model (AUC= 0.772) in terms of postoperative recurrence. When clinical features were added to the peritumoral 9mm model, the combined radiation-clinical omics model's AUC significantly improved in the training

cohort (0.865), internal validation cohort (0.902), and external validation cohort (0.830) ( $p<0.001$ ). Figure 5 shows the ROC curves for the peritumoral 9mm model, the clinical model, and the combined radiation-clinical omics models in the training cohort, internal validation cohort, and external validation cohort. In order to develop a clinically applicable and more accurate model for predicting postoperative recurrence in stage ILUAD, we used the LR algorithm to construct a peritumoral 9mm radiomics nomogram incorporating some of the independent risk factors (Figure 6).

The DeLong test showed that the AUC values of the nomogram were significantly different from those of the other models in the training cohort ( $P < 0.05$ ). The combined radiation-clinical omics model's ROC curves were significantly better than those of the radiomics and clinical models. The calibration curves of the training cohort, and the internal validation cohort in the joint model showed significant agreement in predicting postoperative recurrence in stage ILUAD (Figure 7). The DCA of the training cohort, and the internal validation cohort, showed that the nomogram of the combined radiation-clinical omics model had a good net clinical benefit (Figure 8), suggesting that it is a reliable clinical tool for predicting recurrence after surgery for stage ILUAD.

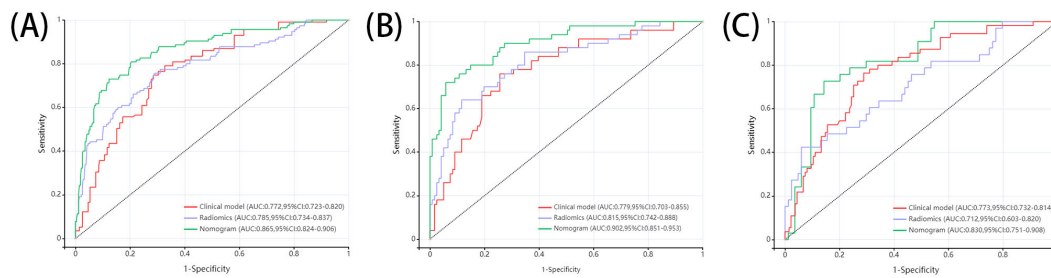


FIGURE 5

ROC curves of peritumoral 9mm model, clinical model, and combined radiation-clinical omics model in training cohort (A), internal validation cohort (B), and external validation cohort (C).

## 4 Discussion

In recent years, much attention has been paid to an emerging technology, radiomics, which automatically extracts a large number of imaging features from medical imaging data in a high-throughput manner; it appears to offer an almost unlimited range of imaging biomarkers, and shows great potential in oncology for detecting, diagnosing, evaluating prognosis, and predicting response to treatment (24–26). Furthermore, an increasing number of scholars are conducting radiomics studies on the interstitium of peripheral lung cancer, which refers to the tissue surrounding the primary tumour, and achieving favourable outcomes (17, 27, 28). This demonstrates the importance of the peritumoural region in radiomics analysis (29).

Tumour radiomics is widely used for prognostic prediction in LUAD (27). However, few studies have applied peritumoural imaging features to aid in the prediction of stage LUAD, and the selection of the peritumoural region remains controversial. Previous studies have defined the peritumoural region as ranging from 1.5 to 20 mm (8, 30, 31). Wu et al. concluded that peritumour radiomic features based on CT images are reliable for predicting the prognosis of non-small cell carcinoma (28). The study also noted that the peritumoural region should ideally extend 15 mm, 20 mm

or 30 mm from the tumour border. Chen et al. measured the bulk tumour volume as well as the bulk tumour volume in the peritumoural 3mm, peritumoural 6mm and peritumoural 9mm regions by extracting the radiomic feature regions (18), and finally constructed the bulk tumour volume of peritumoural 9mm region based on the extraction of the radiomics features had the highest AUC (training set = 0.82, internal validation = 0.75, external validation = 0.67). Liu et al. conducted another study where they extracted radiomics features from intratumoural to peritumoural 3mm, peritumoural 3mm and peritumoural 6mm regions (17). The study demonstrated that features from the intratumoural 3mm to peritumoural 3mm region had higher predictive performance. In a study using radiomics to predict early recurrence, Wang et al. selected 2.1 mm, 4.2 mm, and 8.4 mm as the peritumoural regions, extracted 2D and 3D deep learning image features, and constructed a radiomics model via an air cavity diffusion model, which resulted in good performance in both internal validation cohort and external validation cohort, demonstrating its potential for assisting in post-surgical treatment strategies (7). Wang et al. investigated 8 models of tumour perimeter 5mm, 10mm, 15mm, 20mm as well as tumour-perimeter 5mm, tumour-perimeter 10mm, tumour-perimeter 15mm, tumour-perimeter 20mm, and found that nomogram based on the combined model of tumour-

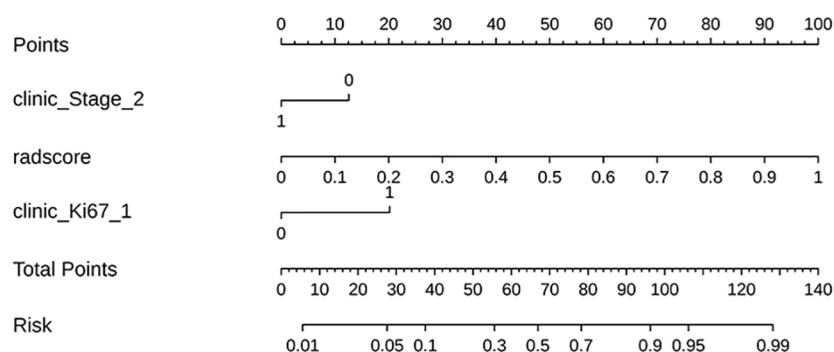


FIGURE 6

The radiomics nomogram incorporating some of the independent risk factors; the peritumoural 9mm model's radscore =  $+4.611 \times \text{wavelet-LL\_first order\_Median\_lung window\_peritumoural 9mm} + 4.114 \times \text{original\_shape 2.5D\_MaximumDiameter\_lung window\_peritumoural 9mm} - 4.119$ .

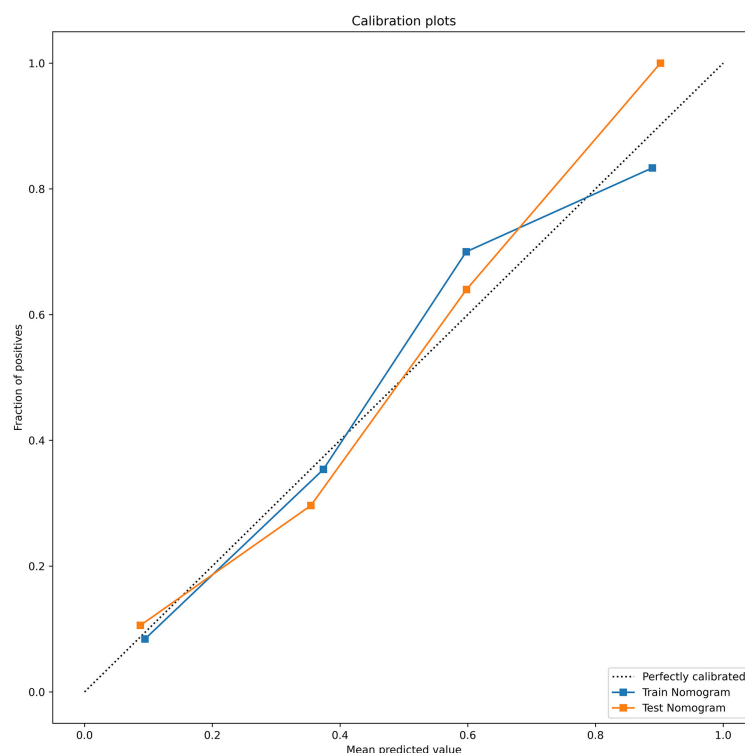


FIGURE 7

The calibration curves of combined radiation-clinical omics model for training cohort (blue dashed line) and internal validation cohort (orange dotted line).

perimeter 10mm and clinical features had a high predictive efficiency for STAS status in NSCLC patients (32). It can be seen that the researchers chose different peritumoural regions, but the best performing peritumoural features essentially consisted of features in the 3-9mm peritumoural regions. In addition, it was also found in previous studies that only intratumoural features were used to predict the prognosis of LUAD (33–36), whereas in this study, the use of peritumoural features performed well.

Based on these previous studies, we selected peritumoural 3mm, peritumoural 6mm, peritumoural 9mm, peritumoural 12mm and peritumoural 15mm as peritumoural regions, but unlike them, we used low-dose lung cancer screening CT plain images and performed 2.5D radiological feature extraction. In this study, we found that the combined peritumoural 9mm radiation-clinical omics model had the highest diagnostic efficacy (AUC=0.865) compared to the tumour and the rest of the peritumour models,

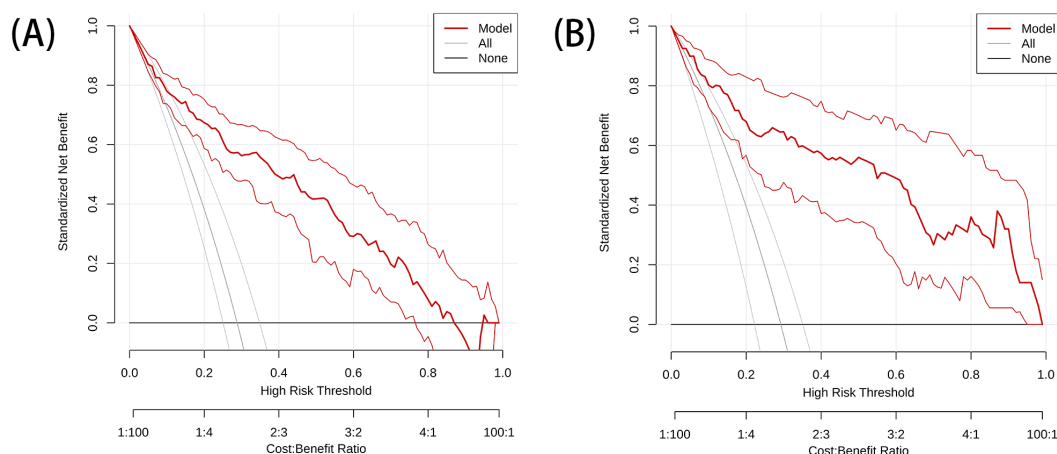


FIGURE 8

The DCA of training cohort (A) and internal validation cohort (B).

with good AUC and sensitivity, specificity, NPV, and PPV in both the internal validation cohort and the external validation cohort, and that it outperformed the conventional unimodal model. By combining a peritumoural 9mm radiomics model with clinical factors, we have also created a visual nomogram with high predictive power and net benefit in the evaluation of recurrence after surgery for stage ILUAD. Our study provides a new approach to prognostic assessment, helps to adjust the treatment plan for patients with stage ILUAD, and enables AI-personalised management of the prognosis of these patients.

Multifactorial logistic regression analysis identified T-stage, neuron-specific enolase assay, Ki67 and nodule type as independent predictors of recurrence after surgery for stage ILUAD, which can be used for clinical modelling. Higher clinical stage, Ki67 percentage, and percentage of nodal solid component imply higher proliferation and invasiveness of tumour cells and higher risk of postoperative recurrence, which is consistent with previous reports (37–39). In addition, multifactorial logistic regression showed that neuron-specific enolase assay and nodule type were also independent predictors of postoperative recurrence, but the clinical-omics features were not significant; therefore, we developed a nomogram combining some of the independent predictors in combination with peritumoural 9mm radiomic features to predict the probability of recurrence in patients with stage ILUAD. In clinical practice, the patient's clinical information and radiological score (radscore) are added to the nomogram to obtain multiple probability scales, and then the total score of the nomogram is calculated, which shows the probability of recurrence. Notably, there was a significant improvement in the AUC of the nomogram compared to a single radiomics and clinical model. It can gain valuable treatment time for patients with stage ILUAD that may recur, and it can help to develop a more rational and effective treatment plan. When it is known that a patient has a high probability of recurrence after surgery, some adjuvant treatments such as chemoradiotherapy or targeted drugs can be taken to reduce the chance of recurrence.

In addition, DeLong test of AUC for each model showed that in the training cohort, the AUC values for the nomogram were significantly different from those of the peritumoural 9mm radiomics model and the clinical model ( $P < 0.05$ ). The results of the study showed that the combined radiation-clinical omics model performed better than the single model, and that clinical parameters also play an important role in predicting postoperative recurrence for stage ILUAD.

The different models constructed in this study not only provide intratumoural and peritumoural biological information, but also give some guidance for clinical treatment. Furthermore, by comparing the diagnostic performance of the different peritumour models, the peritumoural 9mm model had the best predictive performance overall, possibly due to the higher reproducibility of radiomics features the further away from the intratumour area. This finding may be related to the presence of homogeneous lung parenchyma in the distal peritumoural area (31). Thus, in our study, the peritumoural 9mm model showed better predictive performance than the other models. According to the

recommendations of the NCCN guidelines for NSCLC 2024, 4th edition, for most patients with NSCLC, the margin requirement is to ensure that the lung parenchyma margin distance is  $\geq 2$  cm or  $\geq$  the size of the tumour nodule (40), and it was found that the peritumoural region was often extended from the tumour border to 15 mm, 20 mm, or 30 mm (30, 41, 42). However, in our study, when extending to 20 mm peritumour, we found it difficult to avoid thick blood vessels and bronchioles, and complex extrapulmonary tissues, so we only extended to 15 mm peritumour.

## 5 Conclusions and limitations

This study has several limitations. Firstly, it is a retrospective study and there may be recurrent cases in the 2018 cases so far. Secondly, the sample size in this study was small and the predictive efficiency of the external validation cohort may be erroneous, and due to the small sample size, we could not perform survival analysis, and more large sample studies are needed for further validation in the future.

In summary, the combined 2.5D peritumoural 9mm radiation-clinical omics model is more accurate than the tumour and the rest of the peritumoural model in predicting the prognosis of clinical stage ILUAD, and may serve as an effective non-invasive predictive tool, which may provide value in decision-making and defining personalised treatments. However, since most of the studies were conducted retrospectively, further prospective, multicentre and biologically relevant studies based on prospective, multicentre and biologically relevant studies should be carried out in order to facilitate its 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

The studies involving humans were approved by Liuzhou Workers Hospital, Guangxi Zhuang Autonomous Region, China. The studies were conducted in accordance with the local legislation and institutional requirements. The human samples used in this study were acquired from primarily isolated as part of your previous study for which ethical approval was obtained. 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

HL: Data curation, Formal analysis, Methodology, Writing – original draft. CW: Data curation, Formal analysis, Investigation,

Writing – original draft. FX: Methodology, Software, Supervision, Writing – original draft. EY: Data curation, Investigation, Methodology, Software, Writing – original draft. DL: Investigation, Software, Supervision, Writing – original draft. QF: Resources, Validation, Visualization, Writing – review & editing. TL: Project administration, Resources, Validation, Visualization, 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 Liuzhou Science and Technology Plan Project (Key Research and Development Plan, grant number 2019BJ10607); Liuzhou Science and Technology Bureau project (Scientific and Technological Research and New Product Reagents, grant number 2021CBC0128); and the Self-funded Research Project of Health and Family Planning Commission of Guangxi Zhuang Autonomous Region (Planned Project, grant number Z20180488).

## References

1. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. *CA: A Cancer J Clin.* (2023) 73:17–48. doi: 10.3322/caac.21763
2. Thai AA, Solomon BJ, Sequist LV, Gainor JF, Heist RS. Lung cancer. *Lancet.* (2021) 398:535–54. doi: 10.1016/s0140-6736(21)00312-3
3. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. *CA: A Cancer J Clin.* (2024) 74:12–49. doi: 10.3322/caac.21820
4. Mouabbi JA, Hadid TH, Uh E. Effectiveness of low-dose ct scan for lung cancer screening in the community setting. *J Clin Oncol.* (2019) 37:1543–. doi: 10.1200/jco.2019.37.15\_suppl.1543
5. Jacobs CD, Jafari ME. Early results of lung cancer screening and radiation dose assessment by low-dose ct at a community hospital. *Clin Lung Cancer.* (2017) 18:e327–e31. doi: 10.1016/j.clcc.2017.01.011
6. Chansky K, Detterbeck FC, Nicholson AG, Rusch VW, Vallières E, Groome P, et al. The iaslc lung cancer staging project: external validation of the revision of the tmn stage groupings in the eighth edition of the tmn classification of lung cancer. *J Thorac Oncol.* (2017) 12:1109–21. doi: 10.1016/j.jtho.2017.04.011
7. Wang Y, Ding Y, Liu X, Li X, Jia X, Li J, et al. Preoperative ct-based radiomics combined with tumour spread through air spaces can accurately predict early recurrence of stage I lung adenocarcinoma: A multicentre retrospective cohort study. *Cancer Imaging.* (2023) 23:83. doi: 10.1186/s40644-023-00605-3
8. Beig N, Khorrami M, Alilou M, Prasanna P, Braman N, Orooji M, et al. Perinodular and intranodular radiomic features on lung ct images distinguish adenocarcinomas from granulomas. *Radiology.* (2019) 290:783–92. doi: 10.1148/radiol.2018180910
9. Gu Y, She Y, Xie D, Dai C, Ren Y, Fan Z, et al. A texture analysis-based prediction model for lymph node metastasis in stage ia lung adenocarcinoma. *Ann Thorac Surg.* (2018) 106:214–20. doi: 10.1016/j.athoracsur.2018.02.026
10. Yang H, Wang L, Shao G, Dong B, Wang F, Wei Y, et al. A combined predictive model based on radiomics features and clinical factors for disease progression in early-stage non-small cell lung cancer treated with stereotactic ablative radiotherapy. *Front Oncol.* (2022) 12:967360. doi: 10.3389/fonc.2022.967360
11. Tan AC, Tan DSW. Targeted therapies for lung cancer patients with oncogenic driver molecular alterations. *J Clin Oncol.* (2022) 40:611–25. doi: 10.1200/jco.21.01626
12. Zhang X, Lu B, Yang X, Lan D, Lin S, Zhou Z, et al. Prognostic analysis and risk stratification of lung adenocarcinoma undergoing egfr-tki therapy with time-serial ct-based radiomics signature. *Eur Radiol.* (2022) 33:825–35. doi: 10.1007/s00330-022-09123-5
13. Gong J, Bao X, Wang T, Liu J, Peng W, Shi J, et al. A short-term follow-up ct based radiomics approach to predict response to immunotherapy in advanced non-small-cell lung cancer. *Oncol Immunology.* (2022) 11:2028962. doi: 10.1080/2162402x.2022.2028962

## Acknowledgments

The authors wish to thank Dr. Hao Liu for providing the Darwin Intelligent Science Research Platform and guiding the operation.

## 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.

14. Nardone V, Boldrini L, Grassi R, Franceschini D, Morelli I, Becherini C, et al. Radiomics in the setting of neoadjuvant radiotherapy: A new approach for tailored treatment. *Cancers.* (2021) 13:3590. doi: 10.3390/cancers13143590
15. Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature.* (2013) 501:338–45. doi: 10.1038/nature12625
16. Rami-Porta R, Asamura H, Travis WD, Rusch VW. Lung cancer — Major changes in the american joint committee on cancer eighth edition cancer staging manual. *CA: A Cancer J Clin.* (2017) 67:138–55. doi: 10.3322/caac.21390
17. Liu K, Li K, Wu T, Liang M, Zhong Y, Yu X, et al. Improving the accuracy of prognosis for clinical stage I solid lung adenocarcinoma by radiomics models covering tumor per se and peritumoral changes on ct. *Eur Radiol.* (2021) 32:1065–77. doi: 10.1007/s00330-021-08194-0
18. Chen Q, Shao J, Xue T, Peng H, Li M, Duan S, et al. Intratumoral and peritumoral radiomics nomograms for the preoperative prediction of lymphovascular invasion and overall survival in non-small cell lung cancer. *Eur Radiol.* (2022) 33:947–58. doi: 10.1007/s00330-022-09109-3
19. Kim H, Lee D, Cho WS, Lee JC, Goo JM, Kim HC, et al. Ct-based deep learning model to differentiate invasive pulmonary adenocarcinomas appearing as subsolid nodules among surgical candidates: comparison of the diagnostic performance with a size-based logistic model and radiologists. *Eur Radiol.* (2020) 30:3295–305. doi: 10.1007/s00330-019-06628-4
20. Liu Z, Wang S, Dong D, Wei J, Fang C, Zhou X, 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. Khorrami M, Bera K, Leo P, Vaidya P, Patil P, Thawani R, et al. Stable and discriminating radiomic predictor of recurrence in early stage non-small cell lung cancer: multi-site study. *Lung Cancer.* (2020) 142:90–7. doi: 10.1016/j.lungcan.2020.02.018
22. Potter AL, Costantino CL, Suliman RA, Haridas CS, Senthil P, Kumar A, et al. Recurrence after complete resection for non-small cell lung cancer in the national lung screening trial. *Ann Thorac Surg.* (2023) 116:684–92. doi: 10.1016/j.athoracsur.2023.06.004
23. Lufan C, Wenjing Z, Richeng W, Sai F, Hao L, Jing Y, et al. Darwin: A highly flexible platform for imaging research in radiology. *arXiv - CS - Comput Vision Pattern Recognition.* (2020). doi: arxiv-2009.00908
24. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun.* (2014) 5:4006. doi: 10.1038/ncomms5006
25. Limkin EJ, Sun R, Derle L, Zacharakis EI, Robert C, Reuzé S, et al. Promises and challenges for the implementation of computational medical imaging (Radiomics) in oncology. *Ann Oncol.* (2017) 28:1191–206. doi: 10.1093/annonc/mdx034



26. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. (2015) 278:563–77. doi: 10.1148/radiol.2015151169
27. Choe J, Lee SM, Do K-H, Kim S, Choi S, Lee J-G, et al. Outcome prediction in resectable lung adenocarcinoma patients: value of ct radiomics. *Eur Radiol*. (2020) 30:4952–63. doi: 10.1007/s00330-020-06872-z
28. Wu L, Lou X, Kong N, Xu M, Gao C. Can quantitative peritumoral ct radiomics features predict the prognosis of patients with non-small cell lung cancer? A systematic review. *Eur Radiol*. (2022) 33:2105–17. doi: 10.1007/s00330-022-09174-8
29. Zhuo Y, Feng M, Yang S, Zhou L, Ge D, Lu S, et al. Radiomics nomograms of tumors and peritumoral regions for the preoperative prediction of spread through air spaces in lung adenocarcinoma. *Trans Oncol*. (2020) 13:100820. doi: 10.1016/j.tranon.2020.100820
30. Tunali I, Hall LO, Napel S, Cherezov D, Guvenis A, Gillies RJ, et al. Stability and reproducibility of computed tomography radiomic features extracted from peritumoral regions of lung cancer lesions. *Med Phys*. (2019) 46:5075–85. doi: 10.1002/mp.13808
31. Lee H-S, Dou TH, Coroller TP, van Griethuysen JJM, Mak RH, Aerts HJWL. Peritumoral radiomics features predict distant metastasis in locally advanced nscL. *PLoS One*. (2018) 13:e0206108. doi: 10.1371/journal.pone.0206108
32. Wang Y, Lyu D, Hu L, Wu J, Duan S, Zhou T, et al. Ct-based intratumoral and peritumoral radiomics nomograms for the preoperative prediction of spread through air spaces in clinical stage ia non-small cell lung cancer. *J Imaging Inf Med*. (2024) 37:520–35. doi: 10.1007/s10278-023-00939-1
33. Wang X, Zhao X, Li Q, Xia W, Peng Z, Zhang R, et al. Can peritumoral radiomics increase the efficiency of the prediction for lymph node metastasis in clinical stage T1 lung adenocarcinoma on ct? *Eur Radiol*. (2019) 29:6049–58. doi: 10.1007/s00330-019-06084-0
34. Coroller TP, Agrawal V, Narayan V, Hou Y, Grossmann P, Lee SW, et al. Radiomic phenotype features predict pathological response in non-small cell lung cancer. *Radiotherapy Oncol*. (2016) 119:480–6. doi: 10.1016/j.radonc.2016.04.004
35. Traverso A, Wee L, Dekker A, Gillies R. Repeatability and reproducibility of radiomic features: A systematic review. *Int J Radiat OncologyBiologyPhysics*. (2018) 102:1143–58. doi: 10.1016/j.ijrobp.2018.05.053
36. Muñoz-Barrutia A, Grove O, Berglund AE, Schabath MB, Aerts HJWL, Dekker A, et al. Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma. *PLoS One*. (2015) 10:e0248541. doi: 10.1371/journal.pone.0118261
37. Zhang L, Liu J, Yang D, Ni Z, Lu X, Liu Y, et al. A nomogram based on consolidation tumor ratio combined with solid or micropapillary patterns for postoperative recurrence in pathological stage ia lung adenocarcinoma. *Diagnostics*. (2023) 13:2376. doi: 10.3390/diagnostics13142376
38. Huang Y, Liu Z, He L, Chen X, Pan D, Ma Z, et al. Radiomics signature: A potential biomarker for the prediction of disease-free survival in early-stage (I or II) non-small cell lung cancer. *Radiology*. (2016) 281:947–57. doi: 10.1148/radiol.2016152234
39. Peng H, Tan X, Wang Y, Dai L, Liang G, Guo J, et al. Clinical significance of ki67 and circulating tumor cells with an epithelial-mesenchymal transition phenotype in non-small cell lung cancer. *Am J Trans Res*. (2020) 14:4085. doi: 10.1038/s41598-023-48307-x
40. Riely GJ, Wood DE, Ettinger DS, Aisner DL, Akerley W, Bauman JR, et al. Non-small cell lung cancer, version 4.2024. *J Natl Compr Cancer Network*. (2024) 22:249–74. doi: 10.6004/jnccn.2204.0023
41. Chellappan S, Alwithenani A, Bethune D, Castonguay M, Drucker A, Flowerdew G, et al. Profiling non-small cell lung cancer reveals that pd-L1 is associated with wild type egfr and vascular invasion, and immunohistochemistry quantification of pd-L1 correlates weakly with rt-qpcr. *PLoS One*. (2021) 16:e0251080. doi: 10.1371/journal.pone.0251080
42. Li C, Tian Y, Shen Y, Wen B, He Y. Utility of volumetric metabolic parameters on preoperative fdg pet/ct for predicting tumor lymphovascular invasion in non-small cell lung cancer. *Am J Roentgenology*. (2021) 217:1433–43. doi: 10.2214/ajr.21.25814



## OPEN ACCESS

## EDITED BY

Zhimin Li,  
Northwestern University, United States

## REVIEWED BY

Sikandar Shaikh,  
Shadan Hospital and Institute of Medical  
Sciences, India  
Zhijiang Han,  
Hangzhou First People's Hospital, China  
Kai Deng,  
The First Affiliated Hospital of Shandong First  
Medical University, China

## \*CORRESPONDENCE

Chuan yu Zhang  
✉ zhangchuanyu0926@126.com  
Qing Fu  
✉ xiaoqing619@163.com

RECEIVED 16 November 2023

ACCEPTED 26 August 2024

PUBLISHED 11 September 2024

## CITATION

Liu Sh, Nie P, Liu Sl, Hao D, Zhang J, Sun R,  
Yang Zt, Zhang Cy and Fu Q (2024)  
Differentiation of pheochromocytoma and  
adrenal lipoid adenoma by radiomics: are  
enhanced CT scanning images necessary?  
*Front. Oncol.* 14:1339671.  
doi: 10.3389/fonc.2024.1339671

## COPYRIGHT

© 2024 Liu, Nie, Liu, Hao, Zhang, Sun, Yang,  
Zhang and Fu. 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.

# Differentiation of pheochromocytoma and adrenal lipoid adenoma by radiomics: are enhanced CT scanning images necessary?

Shi he Liu<sup>1</sup>, Pei Nie<sup>1</sup>, Shun li Liu<sup>1</sup>, Dapeng Hao<sup>1</sup>, Juntao Zhang<sup>2</sup>,  
Rui Sun<sup>1</sup>, Zhi tao Yang<sup>1</sup>, Chuan yu Zhang<sup>1\*</sup> and Qing Fu<sup>1\*</sup>

<sup>1</sup>Department of Radiology, The Affiliated Hospital of Qingdao University, Qingdao, China,

<sup>2</sup>GE Healthcare, PDX GMS Advanced Analytics, Shanghai, China

**Purpose:** To establish various radiomics models based on conventional CT scan images and enhanced CT images, explore their value in the classification of pheochromocytoma (PHEO) and lipid-poor adrenal adenoma (LPA) and screen the most parsimonious and efficient model

**Methods:** The clinical and imaging data of 332 patients (352 lesions) with PHEO or LPA confirmed by surgical pathology in the Affiliated Hospital of Qingdao University were retrospectively analyzed. The region of interest (ROI) on conventional and enhanced CT images was delineated using ITK-SNAP software. Different radiomics signatures were constructed from the radiomics features extracted from conventional and enhanced CT images, and a radiomics score (Rad score) was calculated. A clinical model was established using demographic features and CT findings, while radiomics nomograms were established using multiple logistic regression analysis. The predictive efficiency of different models was evaluated using the area under curve (AUC) and receiver operating characteristic (ROC) curve. The Delong test was used to evaluate whether there were statistical differences in predictive efficiency between different models.

**Results:** The radiomics signature based on conventional CT images showed AUCs of 0.97 (training cohort, 95% CI: 0.95~1.00) and 0.97 (validation cohort, 95% CI: 0.92~1.00). The AUCs of the nomogram model based on conventional scan CT images and enhanced CT images in the training cohort and the validation cohort were 0.97 (95% CI: 0.95~1.00) and 0.97 (95% CI: 0.94~1.00) and 0.98 (95% CI: 0.97~1.00) and 0.97 (95% CI: 0.94~1.00), respectively. The prediction efficiency of models based on enhanced CT images was slightly higher than that of models based on conventional CT images, but these differences were statistically insignificant ( $P > 0.05$ ).

**Conclusions:** CT-based radiomics signatures and radiomics nomograms can be used to predict and identify PHEO and LPA. The model established based on conventional CT images has great identification and prediction efficiency, and it can also enable patients to avoid harm from radiation and contrast agents caused by the need for further enhancement scanning in traditional image examinations.

#### KEYWORDS

adrenal adenoma, pheochromocytoma, CT, radiomics, classification

## Introduction

Adrenal adenoma accounts for 75% to 80% of all benign adrenal tumors and is the most common adrenal tumor (1). The clinical symptoms are often nonspecific. Depending on the lipid content of the tumor, a CT value of 10 HU is taken as the boundary. If the average CT value within the lesion is lower than 10HU, it indicates that the lesion is an adrenal adenoma rich in lipids. If the CT value is higher than 10HU, it indicates that the lesion is an adrenal adenoma lacking in lipids (LPA) (2–5). A pheochromocytoma (PHEO), which originates from the adrenal medulla, can secrete catecholamines and cause hypertension and may lead to neuropathy and heart disease (6). When the clinical and imaging manifestations of PHEO and adrenal adenoma are not typical, the classification of the two diseases is difficult, and the misdiagnosis rate is high before surgery (7–9). Moreover, there are significant differences in preoperative preparation, surgical approach and prognosis between the two adrenal adenomas (10). Therefore, it is very important to correctly distinguish these two diseases before surgery.

The aim of our research is to develop different radiomics models based on conventional CT scan images and enhanced CT images to identify PHEOs and LPAs and to compare the predictive efficacy of various models to screen the most parsimonious and efficient model.

## Materials and methods

### Patients

The imaging and clinical data of 167 patients (168 lesions) with LPA and 165 patients (184 lesions) with PHEO confirmed by surgical pathology in the Affiliated Hospital of Qingdao University from January 2016 to December 2021 were retrospectively collected (Figure 1). The inclusion criteria were as follows: (1) Both conventional CT scans and dynamic enhanced CT scans were performed before surgery. (2) The tumor lesions were confirmed by surgery and complete pathological data. The exclusion criteria were as follows: (1) The patient was

accompanied by other primary malignant tumors during the same period; (2) The average CT number of adrenal adenoma was less than 10HU; (3) The quality of the image could not meet the requirements of analysis; and (4) With incomplete clinical data. Using a stratified random sampling method, patients were divided into a training cohort (n=232) and a validation cohort (n=100) in a 7:3 ratio.

### Image acquisition and segmentation of lesions

All CT scans were performed on one of the following devices: GE Discovery CT 750 HD (GE Healthcare, USA), SOMATOM Definition AS (Siemens Medical Systems, Germany) and Brilliance iCT (Philips Healthcare, Netherlands). The acquisition and reconstruction parameters are shown in Table 1. The slice thickness of the conventional CT was set at 5 mm. During the enhanced scan, 50 ml iohexol (300 mg/mL) was injected into the elbow vein with a flow rate of 2.5–3.0 mL/s. The arterial and venous phase images were collected at 25 s and 65 s after the injection of contrast agent. The slice thickness of the reconstructed image was set to 1 mm.

Using image segmentation software (ITK-SNAP, <http://www.itksnap.org>, Version: 3.8.0, USA), we manually delineated the region of interest (ROI) on the axial image that displayed the largest cross-sectional area of the lesion on the conventional CT scan. On the axial section images of the arterial phase and venous phase (with a thickness of 1 mm) of the dynamic enhanced CT scan, drew ROI layer by layer along the edge of the lesion, and then apply automatic fusion software to generate 3D ROI of the lesion. All ROI delineation was completed by 2 radiologists (Doctor QF and Doctor SLL) with more than 10 years of chest CT diagnosis experience. Dr. QF outlined the ROI and performed feature extraction. After 1 week, the second ROI mapping and feature extraction were performed to evaluate the internal consistency of the measurers. Dr. SLL only performed ROI placement and feature extraction once. This approach was used to evaluate the inter- and intra- class correlation coefficients (ICCs). An ICC > 0.75 was regarded as satisfactory inter- and intra-reader reproducibility.

Retrospective, with PHEO or adenoma confirmed by surgical pathology from January 2016 to December 2021 (n=2145)

Exclusion criteria:

- 1) Without conventional CT scan or enhanced CT examination (n=354)
- 2) Adrenal adenomas with density less than 10HU(n=902)
- 3) With poor image quality or without a thin layer image (n=97)
- 4) Incomplete clinical data (n=407)
- 5) With other primary malignancies in the same period (n=53)

Cases meeting the inclusion criteria (n=332)

FIGURE 1

Flow diagram of the patient selection.

## Image standardization, feature extraction and development of the radiomics signature

Figure 2 shows the workflow of radiomics implementation. Before extracting the radiomics features, the original images were normalized through z score transformation, and the average intensity range for each imaging mode for all subjects was 0, with a standard deviation of 1. We used a two-step feature selection method to reduce curse of dimensionality, minimize overfitting, and determine the most effective feature for distinguishing PHEO and LPA. Firstly, a single factor analysis of variance (ANOVA) was performed on all features with ICC scores > 0.75, selecting statistically significant features for the training cohort. Secondly, the selected features are included in the Least Absolute Shrinkage and Selection Operator (LASSO) regression algorithm to determine the best feature (with non zero coefficients) to distinguish between PHEO and LPA (Figures 3A–C). Tuning regularization parameters that control regularization intensity were selected by using a minimum standard of 10 times cross validation  $\lambda$ . Then, the final selected feature with a nonzero coefficient was used to construct a radiomics signature. Features weighted by their corresponding nonzero coefficients were screened using a linear combination of

selected values. Then, the Rad scores of each patient in the training cohort and external validation cohort were calculated (Figure 4).

## Development of the radiomics nomogram and assessment of the performance of different models

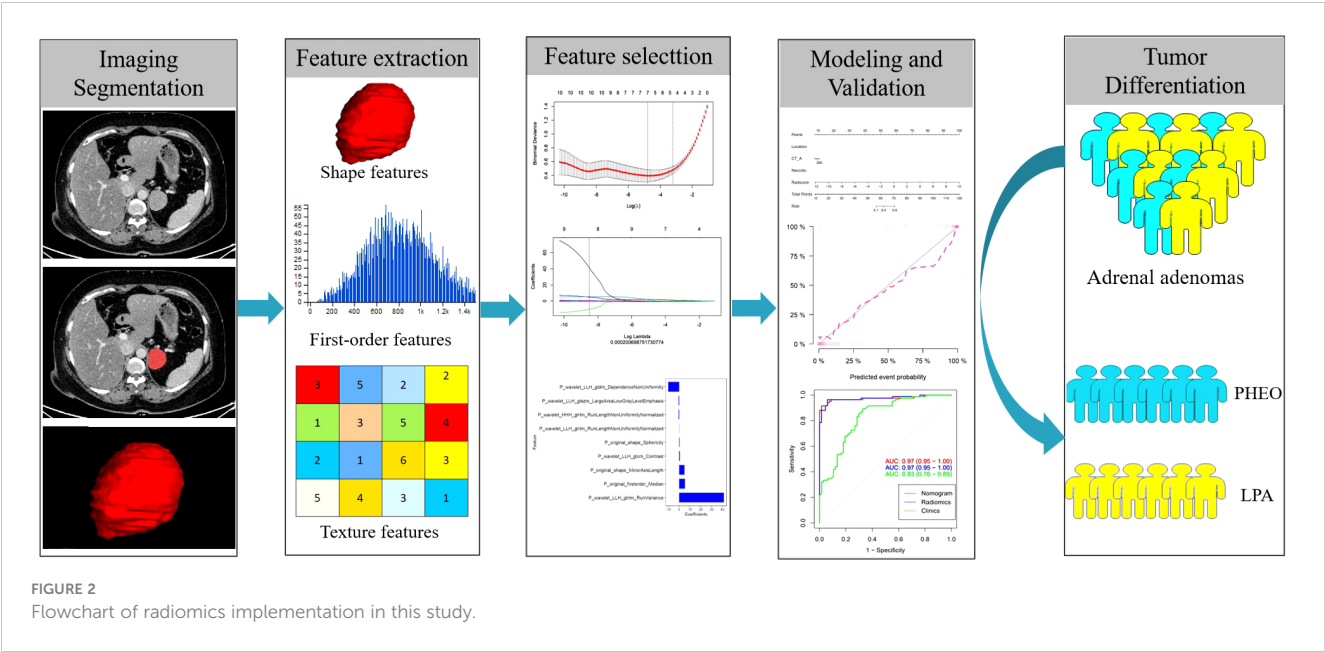
Integrate independent clinical factors and Rad scores developed on the training cohort into the radiomics nomogram using multivariate logistic regression. Then, the Rad score and independent clinical factors were used to calculate the radiomics nomogram score (Nomo score) for each patient in the training and validation cohort. Calibration curves for two groups of patients were graphically generated to evaluate the performance of the radiomics nomogram. The AUC, accuracy, specificity, and sensitivity were used to evaluate the effectiveness of different models. The calibration efficiency of the nomogram was evaluated using a calibration curve, and the analysis fitting was performed using the Hosmer Lime test, which was used to evaluate the calibration ability of the nomogram. Decision curve analysis (DCA) was used to evaluate the clinical application value of the prediction model. The DeLong test was used to evaluate the difference in prediction efficiency between different models.

TABLE 1 The scanning parameters and reconstruction parameters of these three CT scanners.

| Parameters                          | Discovery 750 HD | SOMATOM Definition AS          | Brilliance iCT     |
|-------------------------------------|------------------|--------------------------------|--------------------|
| Scan parameters                     | 120kVp, Smart mA | Care kV, Care Dose 4D, Ref mAs | 120kVp, Dose Right |
| Pitch                               | 0.984            | 0.6                            | 0.984              |
| Reconstruction slice thickness (mm) | 1.25             | 1                              | 1                  |
| Reconstruction kernel               | soft             | B30f                           | standard           |

## Statistical analysis

Statistical analysis was conducted using R software (version 4.2.0, <https://www.R-project.org>). Qualitative data analysis was conducted using Fisher's exact test or chi-square test, and quantitative data analysis was conducted using independent sample t-test. Delong test was used for comparing the predictive value of different models. The following software packages were used in our study: use the "glmnet (R)" software package for LASSO regression based on multivariate binary logistic regression. The ROC curve was plotted using the software package 'Partial Subject Operating Characteristics (pROC [R])'. Use the 'Regression



Modeling Strategy (rms [R])’ software package for nomogram development and calibration curves. The significance level is set at  $p < 0.05$ .

## Results

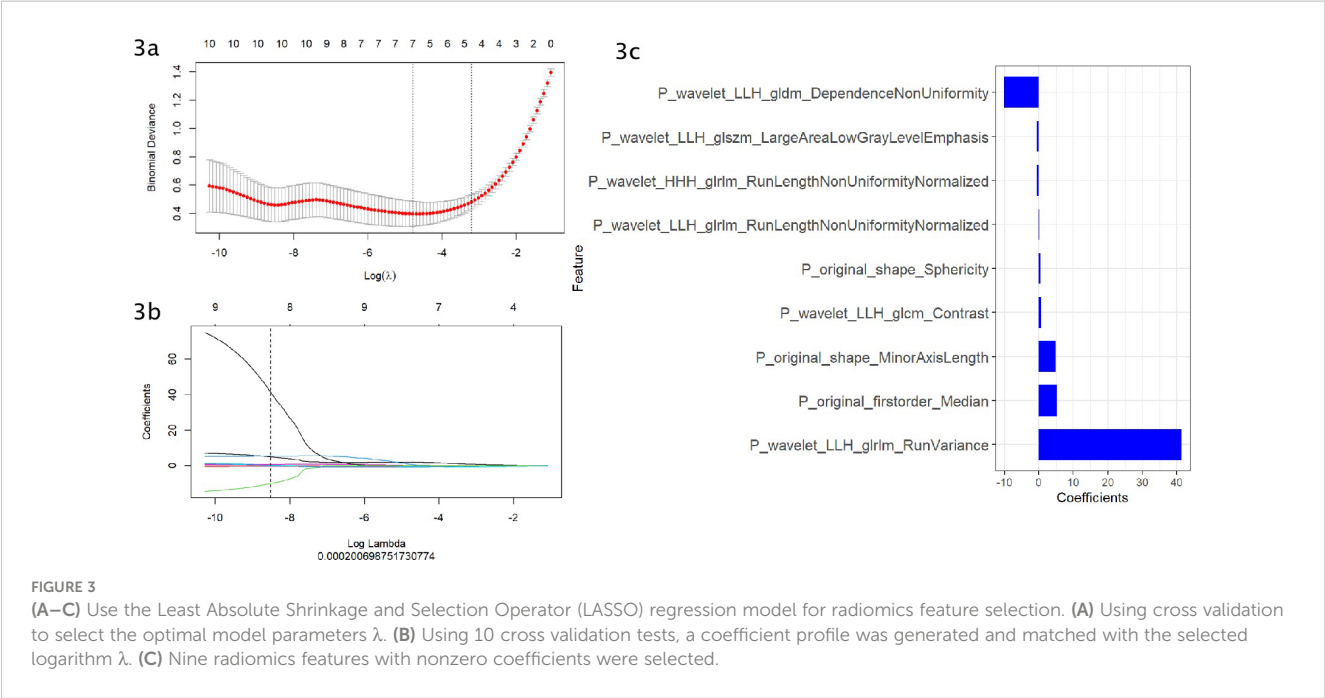
### Clinical factors of the patients

Comparison of clinical data and CT image characteristics of all patients with PHEO and LPA showed statistically significant differences ( $P < 0.01$ ) in lesion location, maximum lesion

diameter, necrosis, edge, CT values and perfusion values, as shown in Table 2.

### Validation of the models

The AUC of the clinical model was 0.83 (95% CI: 0.76-0.89) in the training cohort and 0.83 (95% CI: 0.72-0.94) in the validation cohort. Clinically relevant factors of lesion location, CT values (arterial phase CT values), and necrosis were independent predictors for classifying PHEO and adrenal LPA, and these factors were integrated with the radscore to create a nomogram,





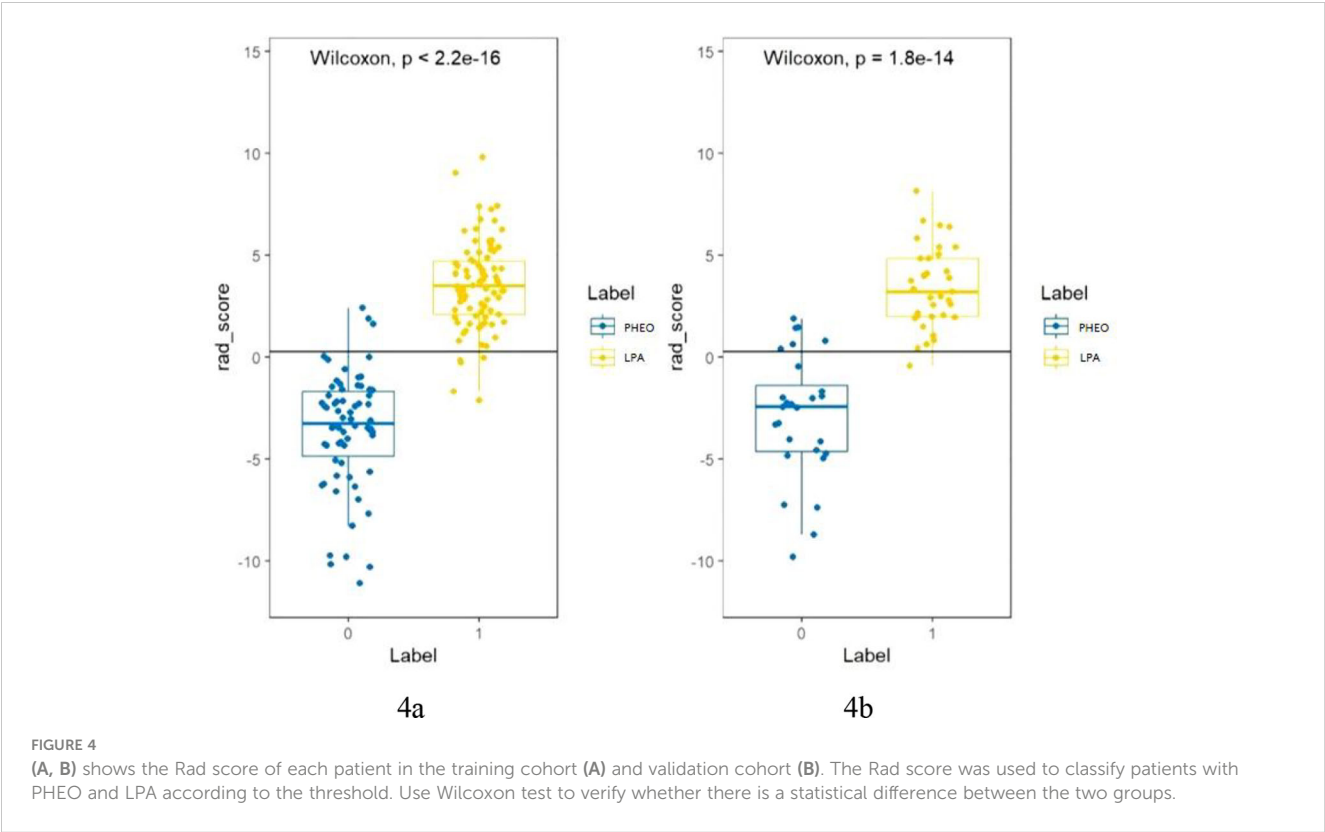


FIGURE 4 (A, B) shows the Rad score of each patient in the training cohort (A) and validation cohort (B). The Rad score was used to classify patients with PHEO and LPA according to the threshold. Use Wilcoxon test to verify whether there is a statistical difference between the two groups.

TABLE 2 Clinical factors of the patients.

| Clinical factors                   |           | PHEO(n=165, 184 lesions) | LPA (n=167, 168 lesions) | p-value | $\chi^2$ or t |
|------------------------------------|-----------|--------------------------|--------------------------|---------|---------------|
| Gender                             | Male      | 73(44.2%)                | 62(37.1%)                | 0.19    | 1.74*         |
|                                    | Female    | 92(55.8%)                | 105(62.9%)               |         |               |
| Age, year                          |           | 51.3 $\pm$ 12.6          | 51.2 $\pm$ 12.7          | 0.95    | 0.07          |
| Location                           | left      | 83(50.3%)                | 102(61.1%)               | <0.01   | 16.17*        |
|                                    | right     | 65(39.4%)                | 64(38.3%)                |         |               |
|                                    | bilateral | 17(10.3%)                | 1(0.6%)                  |         |               |
| Maximum diameter(mm)               |           | 49.4 $\pm$ 28.3          | 23.5 $\pm$ 11.8          | <0.01   | 11.02         |
| Hypertension                       | positive  | 88(53.3%)                | 97(58.1%)                | 0.66    | 0.19*         |
|                                    | negative  | 77(46.7%)                | 70(41.9%)                |         |               |
| Necrotic                           | positive  | 138(75.0%)               | 47(28.0%)                | <0.01   | 77.88*        |
|                                    | negative  | 46(25.0%)                | 121(72.0%)               |         |               |
| edge                               | positive  | 115(62.5%)               | 154(91.7%)               | <0.01   | 41.46*        |
|                                    | negative  | 69(37.5%)                | 14(8.3%)                 |         |               |
| Arterial phase CT value(CT_A) (Hu) |           | 123.2 $\pm$ 51.0         | 71.8 $\pm$ 25.5          | <0.01   | 11.79         |
| Conventional CT value (Hu)         |           | 41.4 $\pm$ 7.1           | 27.1 $\pm$ 9.3           | <0.01   | 16.32         |
| Perfusion value (Hu)               |           | 81.8 $\pm$ 52.4          | 44.6 $\pm$ 22.2          | <0.01   | 8.50          |

\*  $\chi^2$  test, Perfusion value=Arterial phase CT value-Conventional CT value.

as shown in Figure 5. The predictive effectiveness of the clinic model, radiomics signatures and radiomics nomogram models established based on conventional CT images and dynamic enhanced CT images (Mixed images of arterial and venous phases) are shown in Table 3. We compared the predictive value of different models using Delong test, and the results showed that the prediction efficiency of the model based on enhanced CT images was slightly higher than that based on conventional CT images, but the difference was not statistically significant ( $p>0.05$ ) (Table 3). Figure 6 shows that in the validation cohort, the predictive ability of the radiomics nomogram (AUC=0.97, 95% CI: 0.94-1.00) and radiomics signature (AUC=0.97, 95% CI: 0.92-1.00) based on conventional CT images was better than that of the clinical model (AUC=0.83, 95% CI: 0.72-0.94). Figure 7 shows the DCAs of the radiomics nomogram and radiomics signature.

Both radiomics models and clinical models can predict the classification of PHEO and LPA. In the training and validation cohorts, the predictive ability of the nomogram (red) (AUC=0.97) and the radiomics signature (blue) (AUC=0.97) were better than that of the clinical model (green) (AUC=0.83).

Net income is displayed on the y-axis. The dark blue line represents the nomogram model model that combined clinical features and radcores. The red line indicates that all patients are assumed to have PHEO, the light blue line indicates that none of the patients have PHEO, the yellow line represents the results of the clinical model, and the gray line represents the results of radiomics signature. It can be seen that the radiomics signature model and the nomogram model have higher net income.

Discussion

Adrenal adenoma is very common in clinical work, and CT has high specificity and sensitivity in the diagnosis of adrenal adenoma when its CT value on conventional scan is below 10 Hu due to its rich lipid component. However, some adrenal adenomas lacking lipids (called LPAs) are difficult to diagnose correctly (11, 12). PHEOs can secrete catecholamines. The typical clinical manifestation is hypertension, as well as headache and palpitations associated with hypertension, but in practice, approximately 10% to 20% of patients

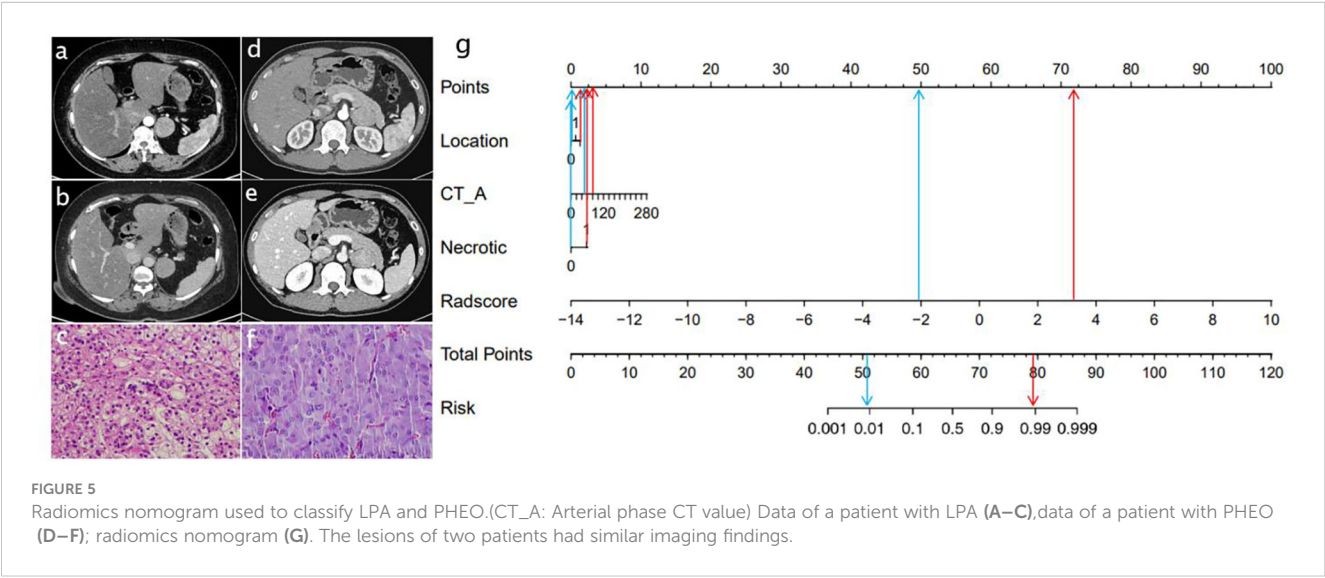


TABLE 3 Comparing the predictive value of different models using Delong test.

| Models                              | Training cohort  |          |             |             | p-value   | Validation cohort |          |             |             | p-value   |
|-------------------------------------|------------------|----------|-------------|-------------|-----------|-------------------|----------|-------------|-------------|-----------|
|                                     | AUC (95% CI)     | Accuracy | Sensitivity | Specificity |           | AUC (95% CI)      | Accuracy | Sensitivity | Specificity |           |
| CT-conventional radiomics signature | 0.97 (0.95~1.00) | 0.95     | 0.96        | 0.94        | Reference | 0.97 (0.92~1.00)  | 0.92     | 0.82        | 0.86        | Reference |
| Clinical model                      | 0.83 (0.76~0.89) | 0.79     | 0.89        | 0.67        | <0.001    | 0.83 (0.72~0.94)  | 0.68     | 0.91        | 0.39        | 0.03      |
| Enhanced CT radiomics signature     | 0.98 (0.97~1.00) | 0.95     | 0.94        | 0.96        | 0.94      | 0.98 (0.95~1.00)  | 0.89     | 0.97        | 0.79        | 0.48      |
| CT- conventional radiomics nomogram | 0.97 (0.95~1.00) | 0.95     | 0.92        | 0.99        | 0.98      | 0.97 (0.94~1.00)  | 0.91     | 0.87        | 0.96        | 0.74      |
| Enhanced CT radiomics nomogram      | 0.98 (0.97~1.00) | 0.96     | 0.98        | 0.94        | 0.87      | 0.97 (0.94~1.00)  | 0.89     | 0.85        | 0.96        | 0.81      |

p-value: Comparing the predictive value of different subgroup models using Delong test.

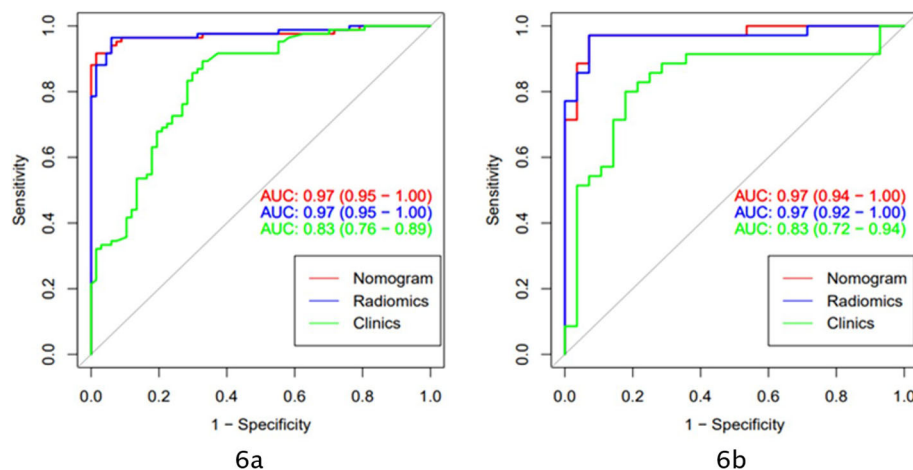


FIGURE 6

The models based on clinical and conventional CT scan images applied AUC to evaluate the prediction ability of different models (A: ROC curve of the training cohort; B: ROC curve of the validation cohort).

have no clinical manifestations or atypical manifestations (13, 14). Adrenal LPA and PHEO are both blood-rich tumors and have many similarities in CT presentation that make them difficult to differentiate (15–18). However, the surgical risk of PHEO is high, and the literature reports that adequate preoperative preparation could reduce the surgical mortality of PHEO from 30%-40% to 0-5.5% (19–21). Therefore, it is necessary to make an accurate clinical diagnosis of patients before surgery.

In the past, the relative and absolute enhancement washout rates were often used to characterize adrenal tumors, but Park said that it was difficult to identify PHEO and LPA using the enhancement washout rate (22, 23). In studying abdominal energy spectrum CT, Marin et al. found that lipid-rich components showed a certain pattern of CT value changes in a certain energy range with high specificity but had little diagnostic significance for lipid-poor components (24).

More and more radiomics analysis was being applied to medical imaging research (25). Radiomics can help clinicians make accurate diagnoses by exploring the connection between images and pathology and clinics (26–28). In addition, radiomics

characteristics may be important predictive factors for cancer differential diagnosis, treatment response, and survival prediction (29, 30). Xiaoping Yi et al (31) found that non enhanced CT quantitative texture analysis based on machine learning may be a reliable quantitative method for distinguishing PHEO from LPA. However, the sample size of this study was relatively small, and no model based on enhanced scanning 3D data had been established for comparison. Therefore, our study is the first to establish multiple imaging radiomics models based on conventional CT and enhanced CT images to predict LPA and PHEO, and we also compared the predictive performance of different models. The AUCs of the radiomics signature based on conventional CT images were 0.97 in the training cohort and 0.97 in the validation cohort. In the validation cohort, the AUCs of the radiomics signature and radiomics nomogram based on enhanced CT images were 0.98 and 0.97, respectively. Both models showed good predictive ability, better than the predictive performance of the clinical model. These results are also superior to the findings of Xiaoping Yi. The radiomics nomogram based on conventional CT images also

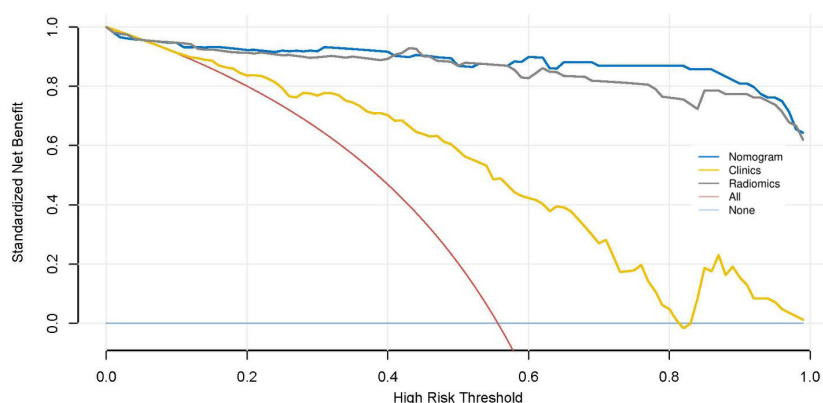


FIGURE 7

Decision curve analysis (DCA) of different models.

yielded satisfactory results. The Delong test results showed that the prediction efficiency of the models based on enhanced CT images was slightly higher than that of the models based on conventional CT images, but the difference was not statistically significant ( $P > 0.05$ ). CT scanning can cause ionizing radiation damage, and dynamic enhanced scanning not only increases radiation exposure but also produces harmful effects such as contrast agent allergy and contrast agent nephrotoxicity (32–35). At present, radiological examinations should strictly follow the principle of “As Low As Reasonably Achievable” (32, 33, 36–38). The model based on conventional CT scans can effectively distinguish adrenal LPAs from PHEOs, and the radiation and contrast hazards associated with further enhancement scans can be avoided.

In this study, radiomic features were selected to construct radiomics signature model for classifying PHEO and adrenal LPA, including: P\_wavelet\_LLH\_gldm\_Dependence Non Uniformity, P\_wavelet\_LLH\_glszm\_Large Area Low Gray Level Emphasis, P\_wavelet\_HHH\_glrlm\_Run Length Non Uniformity Normalized, P\_wavelet\_LLH\_glrlm\_Run Length Non Uniformity Normalized, P\_original\_shape\_Sphericity, P\_wavelet\_LLH\_gldm\_Contrast, P\_original\_shape\_Minor Axis Length, P\_original\_firstorder\_Median, P\_wavelet\_LLH\_glrlm\_Run Variance, among which 1 first order feature, 3 glrlm features, 1 glszm feature, 1 glcm feature, 1 gldm feature and 2 original shape features were included. A mix of first-order, texture and wavelet features seemed to be important for classifying PHEO and adrenal LPA. In our study, we used filters to extract radiomics features from the original images. Among the 9 independent imaging features ultimately selected, there are 6 wavelet features. Wavelet features can comprehensively analyze changes in spatial frequency. These features can provide detailed analysis of texture changes. Wavelet features can also quantify the heterogeneity of tumors in various directions through different spatial scales, so it is believed that wavelet features may help us understand the pathophysiology and morphology of tumors (39). Previous studies had revealed the potential value of wavelet features in histological subtype prediction and prognostic assessment (40, 41). Our results show that wavelet features also have significant capabilities in the prediction models of PHEO and LPA. First order features can reflect the grayscale distribution of tumors and are obtained by calculating the grayscale values of tumors, usually representing low dimensional information that is easy to perceive visually. In addition, our model also includes two original shape features, which respectively suggest that the short axis length and sphericity of the tumor may have value in distinguishing PHEO and LPA.

Nevertheless, our research has some limitations: (1) there may be problems of selection bias and information bias in retrospective studies; (2) different CT machines reduce the consistency of image comparison to a certain extent; and (3) future multicenter and prospective trials are needed to verify the results of this study.

In conclusion, the CT-based radiomics signature and radiomics nomogram in our research have good predictive efficacy in identifying PHEO and adrenal LPA. The model based on conventional CT scans can identify both diseases while avoiding the radiation and contrast hazards caused by dynamic enhancement scans.

## 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/s.

## Ethics statement

The studies were reviewed and approved by The Institutional Review Board of the Affiliated Hospital of Qingdao University. Written informed consent was waived by the Institutional Review Board.

## Author contributions

ShL: Conceptualization, Formal analysis, Methodology, Writing – original draft. PN: Data curation, Formal analysis, Methodology, Writing – review & editing. SIL: Data curation, Formal analysis, Writing – review & editing. DH: Conceptualization, Methodology, Writing – review & editing. JZ: Formal analysis, Writing – review & editing. RS: Data curation, Writing – review & editing. ZY: Data curation, Writing – review & editing. CZ: Conceptualization, Funding acquisition, Writing – review & editing. QF: Conceptualization, Funding acquisition, Methodology, Writing – original draft.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study has received funding by the Postdoctoral Science Foundation of China (2021M701811).

## Conflict of interest

Author JZ was employed by company GE Healthcare.

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

- Northcutt BG, Trakhtenbroit MA, Gomez EN, Fishman EK, Johnson PT. Adrenal adenoma and PHEO: comparison of multidetector CT venous enhancement levels and washout characteristics. *J Comput Assisted Tomography*. (2016) 40:194–200. doi: 10.1097/RCT.0000000000000343
- Lam KY. Update on adrenal tumours in 2017 world health organization (WHO) of endocrine tumours. *Endocr Pathol*. (2017) 28:213–27. doi: 10.1007/s12022-017-9484-5
- Mileto A, Nelson RC, Marin D, Roy Choudhury K, Ho LM. Dual-energy multidetector CT for the characterization of incidental adrenal nodules: diagnostic performance of contrast-enhanced material density analysis. *Radiology*. (2015) 274:445–54. doi: 10.1148/radiol.14140876
- Kamiyama T, Fukukura Y, Yoneyama T, Takumi K, Nakajo M. Distinguishing adrenal adenomas from nonadenomas: combined use of diagnostic parameters of unenhanced and short 5-minute dynamic enhanced CT protocol. *Radiology*. (2009) 250:474–81. doi: 10.1148/radiol.2502080302
- Schieda N, Siegelman ES. Update on CT and MRI of adrenal nodules. *AJR Am J Roentgenol*. (2017) 208(6):1206–17. doi: 10.2214/AJR.16.17758
- Lafont M, Fagour C, Haissaguerre M, Darancette G, Wagner T, Corcuiff JB, et al. Per-operative hemodynamic instability in normotensive patients with incidentally discovered pheochromocytomas. *J Clin Endocrinol Metab*. (2015) 2:jc20142998. doi: 10.1210/jc.2014-2998
- Umanodan T, Fukukura Y, Kumagae Y, Shindo T, Nakajo M, Takumi K, et al. ADC histogram analysis for adrenal tumor histogram analysis of apparent diffusion coefficient in differentiating adrenal adenoma from pheochromocytoma. *J Magnetic Resonance Imaging*. (2017) 45:1195–203. doi: 10.1002/jmri.25452
- Park BK, Kim B, Ko K, Jeong SY, Kwon GY. Adrenal masses falsely diagnosed as adenomas on unenhanced and delayed contrast-enhanced computed tomography: Pathological correlation. *Eur Radiol*. (2006) 16:642–47. doi: 10.1007/s00330-005-0017-0
- Zhang YM, Lei PF, Chen MN, Lv XF, Ling YH, Cai PQ, et al. CT findings of adrenal schwannoma. *Clin Radiol*. (2016) 71:464–70. doi: 10.1016/j.crad.2016.01.010
- Garrett RW, Nepute JC, Hayek M E, Hayek ME, Albert SG. Adrenal incidentalomas: clinical controversies and modified recommendations. *AJR Am J Roentgenol*. (2016) 206:1170–8. doi: 10.2214/AJR.15.15475
- Zhu M, Qu J, Han Z. Evaluate the efficacy of minimum attenuation value in differentiation of adrenal adenomas from nonadenomas on unenhanced CT. *Clin Imaging*. (2016) 40:86–9. doi: 10.1016/j.clinimag.2015.09.006
- Blake MA, Krishnamoorthy SK, Boland GW, Sweeney AT, Pitman MB, Harisinghani M, et al. Low-density pheochromocytoma on CT: a mimicker of adrenal adenoma. *AJR Am J Roentgenol*. (2003) 181:1663–8. doi: 10.2214/ajr.181.6.1811663
- Reisch N, Peczkowska M, Januszewicz A, Neumann HP. Pheochromocytoma: presentation, diagnosis and treatment. *J Hypertension*. (2006) 24:2331–39. doi: 10.1097/01.hjh.0000251887.01885.54
- Tsirlin A, Oo Y, Sharma R, Kansara A, Gliwa A, Banerji MA. Pheochromocytoma: a review. *Maturitas*. (2014) 77:229–38. doi: 10.1016/j.maturitas.2013.12.009
- Adam SZ, Nikolaidis P, Horowitz JM, Gabriel H, Hammond NA, Patel T, et al. Chemical shift MR imaging of the adrenal gland: principles, pitfalls, and applications. *Radiographics*. (2016) 36:414–32. doi: 10.1148/rg.2016150139
- Choi YA, Kim CK, Park BK, Kim B. Evaluation of adrenal metastases from renal cell carcinoma and hepatocellular carcinoma: use of delayed contrast-enhanced CT. *Radiology*. (2013) 266:514–20. doi: 10.1148/radiol.12120110
- Blake MA, Kalra MK, Maher MM, Sahani DV, Sweeney AT, Mueller PR, et al. Pheochromocytoma: an imaging chameleon. *Radiographics*. (2004) 24 Suppl 1:S87–99. doi: 10.1148/rg.24si045506
- Jacques AE, Sahdev A, Sandrasagara M, Goldstein R, Berner D, Rockall AG, et al. Adrenal pheochromocytoma: correlation of MRI appearances with histology and function. *Eur Radiol*. (2008) 18:2885–92. doi: 10.1007/s00330-008-1073-z
- Shuichiro K, Yoshinori K. Hypertensive attack induced by dexamethasone during induction of anesthesia in a patient with an PHEO: a case report. *JA Clin Rep*. (2022) 8:59. doi: 10.1186/s40981-022-00547-y
- Kaman L, Behera A, Singh R, Katariya RN. Surgical management of pheochromocytoma. *Asian J Surg*. (2002) 25:139–44. doi: 10.1016/S1015-9584(09)60162-7
- Pacak K. Preoperative management of the pheochromocytoma patient. *J Clin Endocrinol Metab*. (2007) 92:4069–79. doi: 10.1210/jc.2007-1720
- Park BK, Kim CK, Kwon GY, Kim JH. Re-evaluation of PHEOs on delayed contrast-enhanced CT: washout enhancement and other imaging features. *Eur Radiol*. (2007) 17:2804–09. doi: 10.1007/s00330-007-0695-x
- Woo S, Suh CH, Kim SY, Cho JY, Kim SH. Pheochromocytoma as a frequent false-positive in adrenal washout CT: A systematic review and meta-analysis. *Eur Radiol*. (2018) 28:1027–36. doi: 10.1007/s00330-017-5076-5
- Marin D, Boll DT, Mileto A, Nelson RC. State of the art: dual-energy CT of the abdomen. *Radiology*. (2014) 271:327–42. doi: 10.1148/radiol.14131480
- Limkin EJ, Sun R, Dercle L, Zacharakis EI, Robert C, Reuzé S, et al. Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology. *Ann Oncol*. (2017) 28:1191–206. doi: 10.1093/annonc/mdx034
- 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
- 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
- Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*. (2014) 5:4006. doi: 10.1038/ncomms5006
- Huang Y, Liu Z, He L, Chen X, Pan D, Ma Z, et al. Radiomics signature: A potential biomarker for the prediction of disease-free survival in early-stage (I or II) non-small cell lung cancer. *Radiology*. (2016) 281:947–57. doi: 10.1148/radiol.2016152234
- Bogowicz M, Riesterer O, Ikenberg K, Stieb S, Moch H, Studer G, et al. Computed tomography radiomics predicts HPV status and local tumor control after definitive radiochemotherapy in head and neck squamous cell carcinoma. *Int J Radiat OncologyBiologyPhys*. (2017) 99:921–8. doi: 10.1016/j.ijrobp.2017.06.002
- Yi X, Guan X, Zhang Y, Liu L, Long X, Yin H, et al. Radiomics improves efficiency for differentiating subclinical PHEO from LPA: a predictive, preventive and personalized medical approach in adrenal incidentalomas. *EPMA J*. (2018) 9:421–9. doi: 10.1007/s13167-018-0149-3
- Sodickson A, Baeyens PF, Andriole KP, Prevedello LM, Nawfel RD, Hanson R. Recurrent CT, cumulative radiation exposure, and associated radiation-induced cancer risks from CT of adults. *Radiology*. (2009) 251:175–84. doi: 10.1148/radiol.2511081296
- Pannu N, Wiebe N, Tonelli M. Prophylaxis strategies for contrast-induced nephropathy. *JAMA J Am Med Assoc*. (2006) 295:2765–79. doi: 10.1001/jama.295.23.2765
- Persson PB, Hansell P, Liss P. Pathophysiology of contrast medium-induced nephropathy. *Kidney Int*. (2005) 68:14–22. doi: 10.1111/j.1523-1755.2005.00377.x
- Seeliger E, Sendeski M, Rihal CS, Persson PB. Contrast-induced kidney injury: mechanisms, risk factors, and prevention. *Eur Heart J*. (2012) 33(16):2007–15. doi: 10.1093/eurheartj/ehr494
- Oakley PA, Harrison DE. Death of the ALARA radiation protection principle as used in the medical sector. *Dose Response*. (2020) 18:710596823. doi: 10.1177/1559325820921641
- Gupta RV, Kalra MK, Ebrahimian S, Kaviani P, Primak A, Bizzo B, et al. Complex relationship between artificial intelligence and CT radiation dose. *Acad Radiol*. (2022) 29:1709–19. doi: 10.1016/j.acra.2021.10.024
- Brix G, Nagel HD, Stamm G, Veit R, Lechel U, Griebel J, et al. Radiation exposure in multi-slice versus single-slice spiral CT: results of a nationwide survey. *Eur Radiol*. (2003) 13:1979–91. doi: 10.1007/s00330-003-1883-y
- Hu S, Xu C, Guan W, Tang Y, Liu Y. Texture feature extraction based on wavelet transform and gray-level co-occurrence matrices applied to osteosarcoma diagnosis. *Bio-med Mater Eng*. (2014) 24:129–43. doi: 10.3233/BME-130793
- Huynh E, Coroller TP, Narayan V, Agrawal V, Hou Y, Romano J, et al. CT-based radiomic analysis of stereotactic body radiation therapy patients with lung cancer. *Radiother Oncol*. (2016) 120(2):258–66. doi: 10.1016/j.radonc.2016.05.024
- Wu W, Parmar C, Grossmann P, Quackenbush J, Lambin P, Bussink J, et al. Exploratory study to identify radiomics classifiers for lung cancer histology. *Front Oncol*. (2016) 6:71. doi: 10.3389/fonc.2016.00071





## OPEN ACCESS

## EDITED BY

Sharon R. Pine,  
University of Colorado Anschutz Medical  
Campus, United States

## REVIEWED BY

Paulina Cegla,  
Greater Poland Cancer Center (GPCC),  
Poland  
Maryam Haghsomar,  
Northwestern Medicine, United States

## \*CORRESPONDENCE

Ruozheng Wang  
✉ wrz8526@vip.163.com

RECEIVED 29 November 2023

ACCEPTED 27 August 2024

PUBLISHED 17 September 2024

## CITATION

Liu H, Lao M, Zhang Y, Chang C, Yin Y and  
Wang R (2024) Radiomics-based machine  
learning models for differentiating  
pathological subtypes in cervical  
cancer: a multicenter study.  
*Front. Oncol.* 14:1346336.  
doi: 10.3389/fonc.2024.1346336

## COPYRIGHT

© 2024 Liu, Lao, Zhang, Chang, Yin 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.

# Radiomics-based machine learning models for differentiating pathological subtypes in cervical cancer: a multicenter study

Huiling Liu<sup>1,2</sup>, Mi Lao<sup>3</sup>, Yalin Zhang<sup>1</sup>, Cheng Chang<sup>4</sup>,  
Yong Yin<sup>5</sup> and Ruozheng Wang<sup>1,6,7\*</sup>

<sup>1</sup>Department of Radiation Oncology, The Third Affiliated Teaching Hospital of Xinjiang Medical University, Affiliated Cancer Hospital, Urumuqi, China, <sup>2</sup>Department of Radiation Oncology, Binzhou People's Hospital, Binzhou, China, <sup>3</sup>Department of Cardiology, Binzhou People's Hospital, Binzhou, China, <sup>4</sup>Department of Nuclear Medicine, The Third Affiliated Teaching Hospital of Xinjiang Medical University, Affiliated Cancer Hospital, Urumuqi, China, <sup>5</sup>Department of Radiation Oncology, Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences, Jinan, China, <sup>6</sup>Key Laboratory of Oncology of Xinjiang Uyghur Autonomous Region, Urumuqi, China, <sup>7</sup>Clinical Key Specialty of Radiotherapy of Xinjiang Uygur Autonomous Region, Urumuqi, China

**Purpose:** This study was designed to determine the diagnostic performance of fluorine-18-fluorodeoxyglucose (<sup>18</sup>F-FDG) positron emission tomography (PET)/computed tomography (CT) radiomics-based machine learning (ML) in the classification of cervical adenocarcinoma (AC) and squamous cell carcinoma (SCC).

**Methods:** Pretreatment <sup>18</sup>F-FDG PET/CT data were retrospectively collected from patients who were diagnosed with locally advanced cervical cancer at two centers. Radiomics features were extracted and selected by the Pearson correlation coefficient and least absolute shrinkage and selection operator regression analysis. Six ML algorithms were then applied to establish models, and the best-performing classifier was selected based on accuracy, sensitivity, specificity, and area under the curve (AUC). The performance of different model was assessed and compared using the DeLong test.

**Results:** A total of 227 patients with locally advanced cervical cancer were enrolled in this study (N=136 for the training cohort, N=59 for the internal validation cohort, and N=32 for the external validation cohort). The PET radiomics model constructed based on the lightGBM algorithm had an accuracy of 0.915 and an AUC of 0.851 (95% confidence interval [CI], 0.715-0.986) in the internal validation cohort, which were higher than those of the CT radiomics model (accuracy: 0.661; AUC: 0.513 [95% CI, 0.339-0.688]). The DeLong test revealed no significant difference in AUC between the combined radiomics model and the PET radiomics model in either the training cohort (z=0.940, P=0.347) or the internal validation cohort (z=0.285, P=0.776). In the external validation cohort, the lightGBM-based PET radiomics model achieved good discrimination between SCC and AC (AUC = 0.730).

**Conclusions:** The lightGBM-based PET radiomics model had great potential to predict the fine histological subtypes of locally advanced cervical cancer and might serve as a promising noninvasive approach for the diagnosis and management of locally advanced cervical cancer.

#### KEYWORDS

locally advanced cervical cancer, positron emission tomography, PET, radiomics, adenocarcinoma, AC, squamous cell carcinoma, SCC

## 1 Introduction

Cervical cancer is the fourth most common female cancer worldwide (1). In 2016, there were approximately 34,000 cervical cancer-related deaths in Chinese women (2). Squamous cell carcinoma (SCC) and adenocarcinoma (AC) are the main pathological subtypes of cervical cancer, accounting for 70-75% and 10-25% respectively. The incidence of AC has been observed to increase in recent decades (3). Patients with locally advanced cervical cancer who receive radiation therapy or concurrent chemoradiotherapy have a worse prognosis for AC compared to SCC, highlighting the need for alternative treatment options specifically for AC cases (4). High intratumor heterogeneity exhibited a significantly poor clinical outcome (5). Therefore, it is essential to uncover the differences between AC and SCC from multiple perspectives, explore the underlying reasons for these differences, and develop personalized treatment strategies and plans, as this holds considerable importance.

Pathological diagnosis is considered the gold standard for the detection of cervical cancer, with cervical cytology and cervical biopsy being the primary recommended methods (6). However, AC may sometimes result in cytological false-negatives (7). Biopsy is an invasive procedure associated with risks of bleeding and infection. Point-to-point biopsy performed on larger tumors only evaluates a small portion of the sample, resulting in sampling bias and an inability to comprehensively assess tumor heterogeneity (8–10). In addition, radiologists find it challenging to differentiate AC from SCC based on conventional imaging modalities such as magnetic resonance imaging (MRI), positron emission tomography (PET)/computed tomography (CT), ultrasound, etc., and the interobserver agreement is typically low (11, 12).

Radiomics is a rapidly growing field of research that utilizes medical images to extract quantitative features, converting them into high-dimensional data for analysis and exploration. This technique enhances our understanding of diseases and provides valuable support for clinical decision-making (13, 14). Malignant tumors exhibit considerable spatial variation within the tumor at the morphological and histopathological levels, including cellularity, vascularization, extracellular matrix, and necrotic components (15, 16). As a noninvasive tool, radiomics can quantify intratumoral heterogeneity and is widely used in

diagnosis, treatment response evaluation, and survival prediction (17). Among conventional imaging modalities, MRI has better fine exquisite soft tissue resolution than CT and PET, so it has long been considered as the preferred imaging method of choice for the evaluation of local tumor extension in primary cervical cancer. Unsurprisingly, numerous radiomics studies aimed at identifying the pathological subtypes of cervical cancer are primarily founded on pretreatment MRI (18). To the best of our knowledge, there are no radiomics studies that delineate the primary cervical tumor on CT images, which might be related to the inability to clearly define the boundary of the primary cervical tumor on CT images. In addition, only two published studies have preliminarily evaluated PET radiomics in cervical cancer to discriminate between AC and SCC (19, 20). Nevertheless, both of the studies were single-center ones, extracted too few radiomic features, and the methods for radiomics feature selection and model construction were simplistic. Even in one of the studies, only 83 patients were included.

Previous studies have shown that radiomic features based on CT or PET images can achieve the differentiation of pathological types of lung cancer (21). The application of radiomics methods for diagnosis and tumor characterization might be a potential supplement for omics datasets, or an alternative for pathological diagnosis, particularly for patients who are at an advanced stage, inoperable, or unable to undergo biopsies. A more extensive and comprehensive study is required to investigate the value of PET/CT imaging in differentiating the subtypes of cervical cancer. Therefore, the aim of this study was to develop and validate an optimal machine learning (ML) model based on pretherapeutic fluorine-18-fluorodeoxyglucose ( $^{18}\text{F}$ -FDG) PET/CT for differentiating between SCC and AC in cervical cancer.

## 2 Materials and methods

### 2.1 Study design

This retrospective study was conducted in accordance with the Declaration of Helsinki. Ethical approval was obtained from the Institutional Review Board of the Affiliated Cancer Hospital of Shandong First Medical University (No. SDTHEC2023006030) and the Affiliated Cancer Hospital of Xinjiang Medical University (No.

K-2022006), and the requirement for written informed consent was waived. The workflow of our study is shown in Figure 1.

## 2.2 Patient cohort

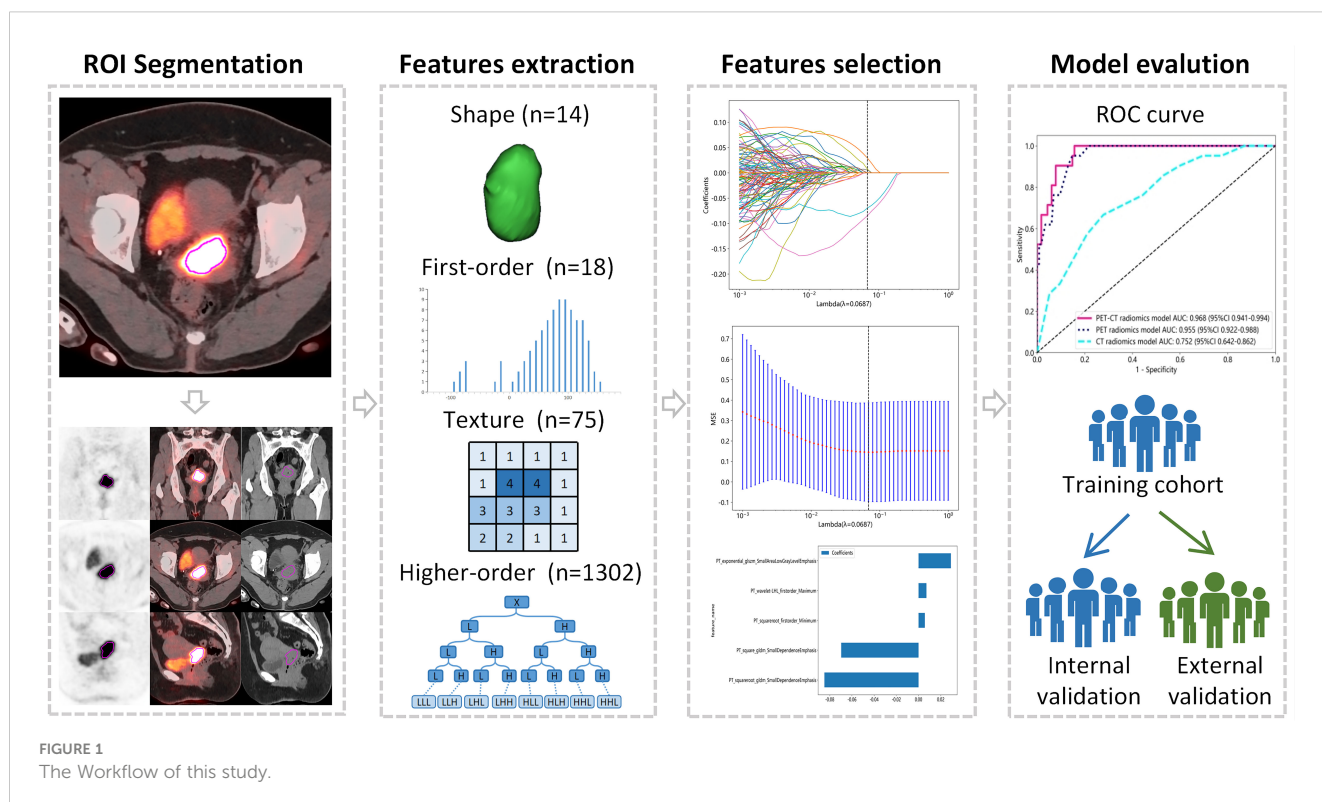
The study included patients with a diagnosis of cervical cancer between September 2015 and February 2022. The inclusion criteria were as follows: (1) pathologically confirmed cervical cancer with the 2018 International Federation of Gynecology and Obstetrics (FIGO) stage IB-IVA; (2) underwent  $^{18}\text{F}$ -FDG PET/CT; and (3) complete clinical data retrievable from the electronic medical records. Exclusion criteria included: (1) a history of any previous anticancer treatment; (2) pathological types other than SCC and AC; (3) patients with a diagnosis of other unrelated malignant tumors; (4) presence of extensive abdominal metastasis; (5) poor PET/CT image quality; and (6) primary maximal tumor diameter less than 1.0 cm.

All patients were initially confirmed by hematoxylin-eosin (HE) staining, and the poorly differentiated patients whose subtypes could not be affirmed were further confirmed by immunohistochemistry (IHC) staining. Ultimately, based on the pathological reports of biopsy specimens, a total of 195 patients were recruited in the Center 1 (the Affiliated Cancer Hospital of Shandong First Medical University), among which 164 were confirmed by HE staining and 31 were confirmed by IHC. The Center 2 (the Affiliated Cancer Hospital of Xinjiang Medical University) recruited 32 patients, of whom 23 were confirmed by HE staining and 9 were confirmed by IHC. The patients recruited in the Center 1 were randomly allocated to the training cohort ( $n =$

136) and the internal validation cohort ( $n = 59$ ) in a 7:3 ratio, while the Center 2 serves as the external validation cohort. Figure 2 illustrates a flow chart outlining the process of patient selection. The clinical information of the patients, including age, pathology, maximal tumor diameter (MTD) on PET/CT images, menopausal status, lymph node metastasis (LNM), and red blood cell count, was collected from electronic medical records.

## 2.3 PET/CT acquisition

All enrolled patients underwent  $^{18}\text{F}$ -FDG PET/CT with a standardized scan setup and parameters before treatment. Patients enrolled in Center 1 were scanned with the Philips Gemini TF PET/CT scanner (Philips Medical Systems, Holland), and the  $^{18}\text{F}$ -FDG was generated by the MINitrac cyclotron from GE Healthcare. At Center 2, the Philips Ingenuity TF (Philips Medical Systems, Holland) was used, and  $^{18}\text{F}$ -FDG was generated by the Sumitomo Heavy Industries HM-10 cyclotron. The radiochemical purity was above 95%. All patients fasted for at least 6 h, and their peripheral blood glucose levels were confirmed to be  $\leq 150$  mg/dL before  $^{18}\text{F}$ -FDG injection.  $^{18}\text{F}$ -FDG was intravenously administered at 3.7–4.4 MBq/kg body weight. The key scanning parameters were as follows: tube voltage of 120–130 KV; tube current of 150–300 mA. PET images were reconstructed using ordered-subset expectation maximization. Reconstruction using standard convolution kernel with 1.5 mm layer thickness (median 1.5 mm; range 1.0–3.0 mm). Each CT image was reconstructed in a  $512 \times 512$  pixels image matrix and each PET image was reconstructed in a  $144 \times 144$  pixel image matrix. To eliminate image differences between images acquired by different scanners, all



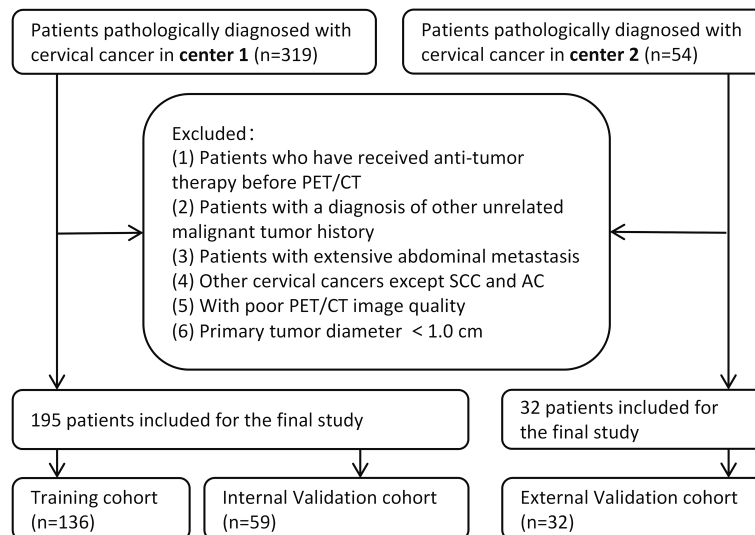


FIGURE 2  
Flow chart of patients selection.

images were resampled to the same image spacing of 1 mm×1 mm ×1 mm.

## 2.4 Tumor segmentation

PET images were attenuated, corrected, reconstructed in multiple layers, and then fused with noncontrast-enhanced low-dose CT images. The resulting images were imported into MIM Maestro version 7.1.7 (MIM Software Inc., Cleveland, OH, USA). The regions of interest (ROIs) were delineated using a fixed threshold value at 42% of the maximum standardized uptake value (SUVmax) of the primary tumor. Regions corresponding to the bladder were manually excluded from the analysis. For the obtained ROIs, various parameters, such as metabolic active tumor volume (MTV), mean standardized uptake value (SUVmean), total lesion glycolysis (TLG), and SUVmax, were calculated using MIM Software. The contoured ROIs were then transferred to PET and CT images using rigid registration. Another experienced oncologist carefully reviewed and modified the transferred results on a slice-by-slice basis. Figures 3, 4 show a set of representative PET/CT images from a 53-year-old woman with SCC and a 41-year-old woman with AC, respectively. The ROI, labeled in red, was segmented in each slice of the axial, sagittal, and coronal views of the PET, CT, and fusion images.

## 2.5 Feature extraction and normalization

A total of 1409 PET and 1409 CT radiomics features were extracted from each segmented ROI using AccuContour software version 3.2 (Manteia Medical Technologies Co. Ltd., Xiamen, China), which is a commercial software application that allows for standardized preprocessing of medical imaging data. The

radiomics features based on the original images included shape features, first-order intensity histogram features, gray-level cooccurrence matrix (GLCM) features, gray-level run-length matrix (GLRLM) features, gray-level size zone matrix (GLSZM) features, neighboring gray-tone difference matrices (NGTDM), and gray-level dependence matrix (GLDM) features.

## 2.6 Feature selection and model development

All features were standardized to Z scores with the mean and standard deviation. The Pearson correlation coefficient (PCC) for each feature pair was calculated to evaluate their similarity, and if the PCC value exceeded 0.9, one of the features was randomly eliminated. After this process, the dimension of the feature space was reduced, and features were independent of each other. Then, least absolute shrinkage and selection operator (LASSO) regression analysis with 10-fold cross-validation was employed to select the effective radiomics features. Clinical features were selected using logistic regression analysis. Separate models with good prediction performance were built to differentiate pathological subtypes in locally advanced cervical cancer. Ultimately, the predictive performance of the models was assessed using the receiver operating characteristic (ROC) curve, decision curve analysis (DCA), and calibration curve.

## 2.7 Statistical analysis

Quantitative data that followed a normal distribution are presented as the mean ± standard deviation (s), while qualitative data are expressed as frequencies (percentages). The patient characteristics between the training and validation cohorts were



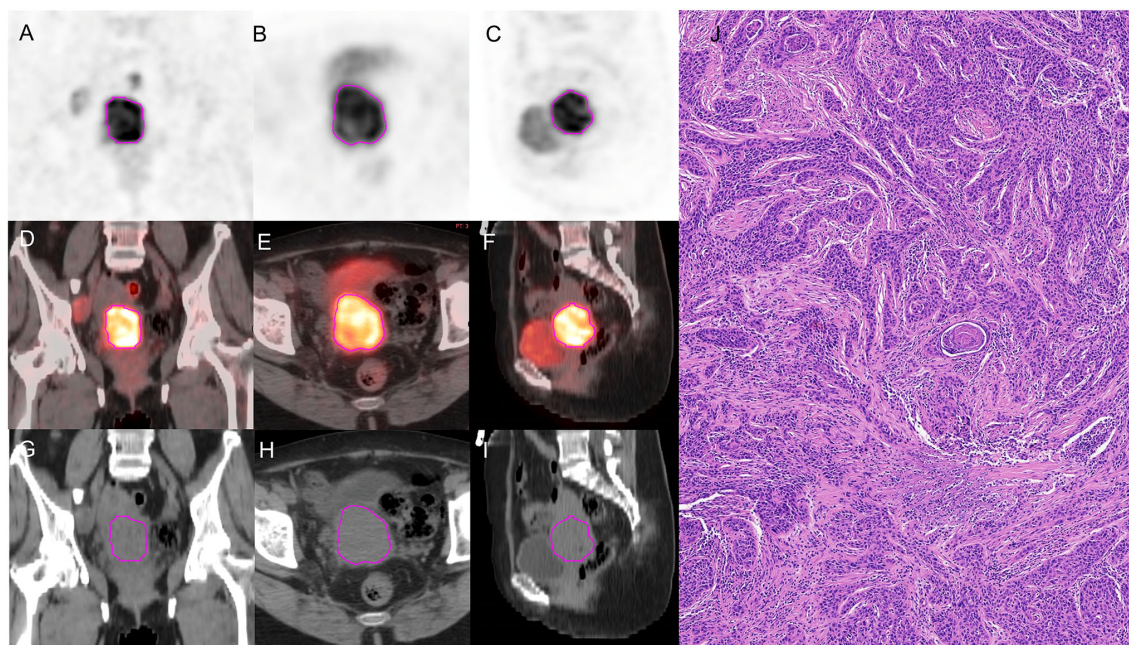


FIGURE 3

A 53-year-old woman diagnosed with SCC. (A–C) The ROI, labeled in red, was segmented on the coronal, axial, sagittal PET images. (D–F) The ROI, labeled in red, was segmented on the coronal, axial, sagittal fusion images. (G–I) The ROI, labeled in red, was segmented on the coronal, axial, sagittal CT images. (J) pathological examination confirmed SCC.

compared using various statistical tests, such as the Pearson Chi-square test, Fisher's exact test, Student's *t* test, and Mann–Whitney *U* test. Clinical features were selected using univariate and multivariate logistic regression analyses. Six ML classifiers, including logistic

regression (LR), naive Bayes (NB), support vector machine (SVM), k-nearest neighbors (KNN), light gradient boosting machine (lightGBM), and multilayer perceptron neural network (MLP), were used to build a model to differentiate pathological subtypes.

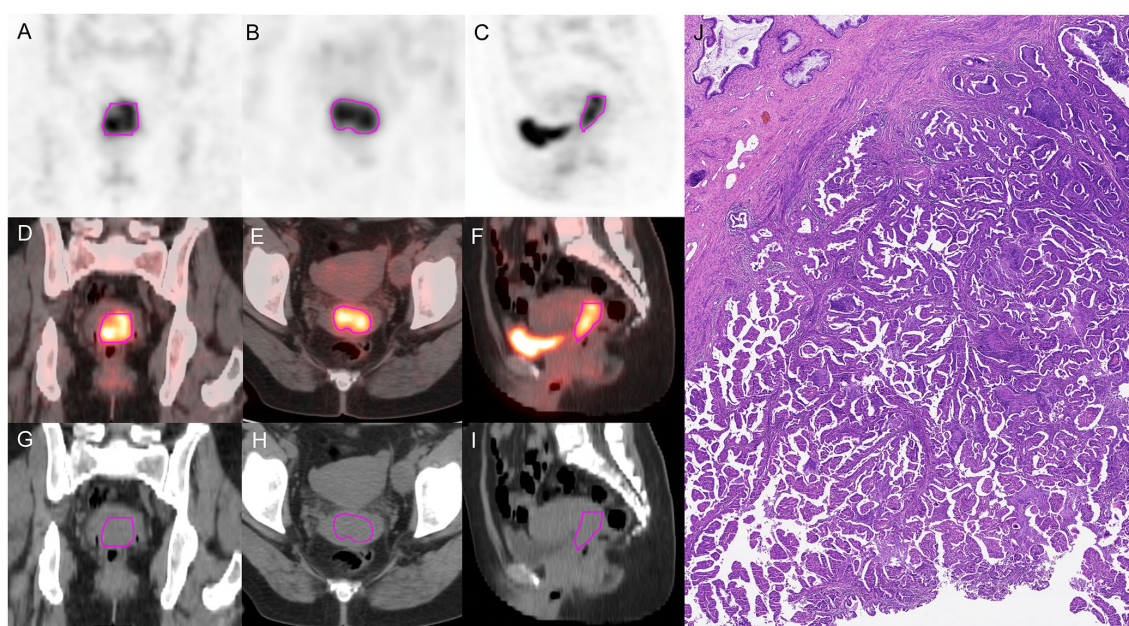


FIGURE 4

A 41-year-old woman diagnosed with AC. (A–C) The ROI, labeled in red, was segmented on the coronal, axial, sagittal PET images. (D–F) The ROI, labeled in red, was segmented on the coronal, axial, sagittal fusion images. (G–I) The ROI, labeled in red, was segmented on the coronal, axial, sagittal CT images. (J) Pathological examination confirmed AC.



The optimal ML model was selected based on its AUC, accuracy (ACC), sensitivity (SEN), and specificity (SPE). The AUC values were compared between different models using the DeLong test. The data analyses were performed using SPSS software (Version 25.0, IBM Corp., Armonk, NY, USA) and R software (Version 3.4.0, R Foundation for Statistical Computing, Vienna, Austria). A two-sided  $p$ -value $<0.05$  was considered statistically significant.

### 3 Results

#### 3.1 Clinical characteristics and PET metabolic parameters

Table 1 presents the clinical characteristics and PET metabolic parameters of 227 patients with locally advanced cervical cancer.

TABLE 1 Comparison of Clinical characteristics and PET metabolic parameters between SCC and AC in the training, internal validation and external validation cohorts.

|                   | Training (N = 136)        |                           | $p$ Value | Internal validation (N = 59) |                          | $p$ Value | External validation (N=32) |                        | $p$ Value |
|-------------------|---------------------------|---------------------------|-----------|------------------------------|--------------------------|-----------|----------------------------|------------------------|-----------|
|                   | SCC (N=115)               | AC (N=21)                 |           | SCC (N=45)                   | AC (N=14)                |           | SCC (N=29)                 | AC (N=3)               |           |
| Age (years)       | 53.91 $\pm$ 9.37          | 56.33 $\pm$ 12.86         | 0.418     | 52.67 $\pm$ 11.41            | 55.28 $\pm$ 12.22        | 0.464     | 52.24 $\pm$ 11.27          | 51.33 $\pm$ 4.93       | 0.892     |
| Abortion          |                           |                           |           |                              |                          | 0.105     |                            |                        |           |
| NO                | 64(55.7%)                 | 13(61.9%)                 | 0.595     | 24(53.33%)                   | 4(28.57%)                |           | 11(37.93%)                 | 1(33.33%)              | 0.876     |
| YES               | 51(44.3%)                 | 8(38.1%)                  |           | 21(46.67%)                   | 10(71.43%)               |           | 18(62.07%)                 | 2(66.67%)              |           |
| MTD (cm)          |                           |                           |           | 5.22 $\pm$ 1.67              |                          | 0.29      | 4.48 $\pm$ 1.48            | 3.27 $\pm$ 0.25        | 0.173     |
| LNM               |                           |                           | 0.430     |                              |                          | 0.516     |                            |                        | 0.909     |
| NO                | 39(33.9%)                 | 9(42.9%)                  |           | 13(28.89%)                   | 6(42.86%)                |           | 13(44.83%)                 | 2(66.67%)              |           |
| YES               | 76(66.1%)                 | 12(57.1%)                 |           | 32(71.11%)                   | 8(57.14%)                |           | 16(55.17%)                 | 1(33.33%)              |           |
| Para-aortic LNM   |                           |                           | 0.350     |                              |                          | 0.759     |                            |                        | 0.476     |
| NO                | 88(76.5%)                 | 18(85.7%)                 |           | 37(82.22%)                   | 11(78.57%)               |           | 24(82.76%)                 | 2(66.67%)              |           |
| YES               | 27(23.5%)                 | 3(14.3%)                  |           | 8(17.78%)                    | 3(21.43%)                |           | 5(17.24%)                  | 1(33.33%)              |           |
| Menopause         |                           |                           | 0.929     |                              |                          | 0.849     |                            |                        | 0.819     |
| NO                | 45(39.1%)                 | 8(38.1%)                  |           | 18(40.0%)                    | 6(42.9%)                 |           | 17(58.62%)                 | 1(33.33%)              |           |
| YES               | 70(60.9%)                 | 13(61.9%)                 |           | 27(60.0%)                    | 8(57.1%)                 |           | 12(41.38%)                 | 2(66.67%)              |           |
| SUVmax (SUVbw)    | 15.59 $\pm$ 5.88          | 16.38 $\pm$ 8.01          | 0.672     | 16.20 $\pm$ 5.82             | 17.27 $\pm$ 5.90         | 0.010     | 15.30 $\pm$ 7.28           | 9.85 $\pm$ 6.81        | 0.225     |
| MTV (ml)          | 27.48<br>(15.54,54.08)    | 34.9<br>(16.85,57.56)     | 0.555     | 35.12<br>(22.20,73.87)       | 27.44<br>(9.21,42.97)    | <0.001    | 11.01<br>(6.10,25.15)      | 6.56<br>(5.27,13.24)   | 0.580     |
| SUVmean (SUVbw)   | 9.23 $\pm$ 3.52           | 9.35 $\pm$ 4.43           | 0.908     | 10.21 $\pm$ 3.44             | 9.64 $\pm$ 2.73          | 0.013     | 9.28 $\pm$ 4.62            | 6.08 $\pm$ 4.42        | 0.261     |
| TLG (SUVbw*ml)    | 231.53<br>(117.85,510.18) | 305.64<br>(124.16,693.92) | 0.671     | 405.52<br>(171.37,775.63)    | 189.45<br>(74.06,294.13) | 0.012     | 101.02<br>(50.60,177.68)   | 57.04<br>(36.96,64.96) | 0.164     |
| WBC count         | 6.85 $\pm$ 2.43           | 6.56 $\pm$ 3.84           | 0.644     | 7.54 $\pm$ 3.22              | 6.28 $\pm$ 1.91          | 0.171     | 6.17 $\pm$ 1.85            | 6.55 $\pm$ 0.62        | 0.730     |
| RBC count         | 4.12 $\pm$ 0.50           | 4.02 $\pm$ 0.43           | 0.384     | 4.19 $\pm$ 0.65              | 4.32 $\pm$ 0.45          | 0.494     | 4.22 $\pm$ 0.61            | 4.45 $\pm$ 0.39        | 0.529     |
| Plt count         | 291.76 $\pm$ 94.11        | 285.67 $\pm$ 101.24       | 0.788     | 326.24 $\pm$ 103.28          | 313.57<br>$\pm$ 142.57   | 0.716     | 233.25 $\pm$ 62.74         | 225.00<br>$\pm$ 37.51  | 0.826     |
| lymphocyte count  | 1.68 $\pm$ 0.57           | 1.56 $\pm$ 0.39           | 0.355     | 1.63 $\pm$ 0.59              | 1.65 $\pm$ 0.40          | 0.916     | 3.00 $\pm$ 5.79            | 1.81 $\pm$ 0.10        | 0.728     |
| neutrophile count | 4.52 $\pm$ 2.16           | 4.52 $\pm$ 3.60           | 0.995     | 5.19 $\pm$ 2.56              | 4.01 $\pm$ 1.61          | 0.13      | 5.57 $\pm$ 10.59           | 4.22 $\pm$ 0.66        | 0.830     |
| Hb count          | 120.81 $\pm$ 16.76        | 116.71 $\pm$ 14.91        | 0.298     | 117.98 $\pm$ 23.78           | 116.43 $\pm$ 24.42       | 0.833     | 126.17 $\pm$ 18.74         | 131.33 $\pm$ 1.16      | 0.642     |

SCC, squamous cell carcinoma; AC, adenocarcinoma; MTD, maximal tumor diameter; LNM, lymph node metastasis; SUVmax, maximum standardized uptake value; SUVmean, mean standardized uptake value; MTV, metabolic active tumor volume; TLG, total lesion glycolysis; WBC, white blood cell; RBC, red blood cell; WBC, white blood cell; Plt, blood platelet; Hb, hemoglobin.

The comparison between SCC and AC in three groups are shown in [Supplementary Table S1](#). The results of the univariate logistic regression analysis are provided in [Table 2](#). None of the clinical features or PET metabolic parameters showed significant differentiation ability for the pathological subtypes.

### 3.2 Radiomics features extraction and selection

A total of 2818 radiomic features were extracted from the ROIs of CT and PET images. Among them, a total of 391 and 242 radiomics features were selected from the CT and PET images, respectively, based on the PCC. Subsequently, LASSO regression analysis was performed to select one CT radiomics feature ([Figures 5A, C](#)) and five PET radiomic features ([Figures 5B, D, 6](#)). Then, [Table 3](#) displays the final PET and CT radiomic features. The quantitative differences in PET radiomic features between cervical SCC and AC are shown in [Supplementary Table S2](#).

### 3.3 Radiomics model development and evaluation

[Table 4](#) presents a summary of the prediction performance in distinguishing between AC and SCC using various ML classifiers in the training and internal validation cohorts. The LightGBM model exhibited superior performance in terms of AUC, ACC, SEN, SPE compared to the other ML models, and was consequently employed as the ML algorithm for differentiating the described pathological subtypes.

[Figure 7](#) illustrates the ROC curves of the CT radiomics model, PET radiomics model, and combined model. In the training cohort, the best differentiation performance was demonstrated by the combined radiomics model (AUC=0.968), followed by the PET radiomics model (AUC=0.955), while the differentiation performance of the CT radiomics model was average (AUC=0.752). The DeLong test indicated that there was no statistically significant difference between the combined radiomics model and the PET radiomics model ( $z=0.940$ ,  $p\text{-value}=0.347$ ). Nevertheless, both the combined radiomics model and the PET radiomics model significantly outperformed the CT radiomics model ( $z=3.291$ ,  $p\text{-value}<0.001$ ). In the internal validation cohort, the PET radiomics model had the best differentiation effectiveness (AUC=0.851), followed by the combined radiomics model (AUC=0.842), while the differentiation performance of the CT radiomics model was poor (AUC=0.513). The DeLong test showed no statistically significant difference between the combined radiomics model and the PET radiomics model ( $z=0.285$ ,  $p\text{-value}=0.776$ ). However, both the combined radiomics model and the PET radiomics model were significantly better than the CT radiomics model ( $z=2.807$ ,  $p\text{-value}=0.005$  and  $z=2.697$ ,  $p\text{-value}=0.007$ , respectively). In the external validation cohort, the DeLong test showed no statistically significant difference between the combined radiomics model and the PET radiomics model ( $z=0.272$ ,  $p\text{-value}=0.809$ ).

**TABLE 2** Univariate logistic regression analysis of clinical and PET metabolic parameters to differentiate pathological subtypes in the training cohort.

|                   | Univariate logistic analysis |             |         |
|-------------------|------------------------------|-------------|---------|
|                   | OR                           | 95% CI      | p-Value |
| Age (years)       | 1.022                        | 0.986-1.060 | 0.228   |
| Abortion          | 0.915                        | 0.647-1.295 | 0.617   |
| MTD (cm)          | 0.859                        | 0.680-1.087 | 0.206   |
| LNM               | 0.642                        | 0.304-1.355 | 0.245   |
| Para-aortic LNM   | 0.739                        | 0.284-1.921 | 0.535   |
| Menopause         | 0.974                        | 0.462-2.056 | 0.945   |
| SUVmax (SUVbw)    | 0.968                        | 0.909-1.031 | 0.316   |
| MTV (ml)          | 0.996                        | 0.986-1.006 | 0.414   |
| SUVmean (SUVbw)   | 0.933                        | 0.837-1.040 | 0.211   |
| TLG (SUVbw*ml)    | 1.000                        | 0.999-1.001 | 0.385   |
| WBC count         | 0.911                        | 0.778-1.067 | 0.249   |
| RBC count         | 0.995                        | 0.496-1.997 | 0.989   |
| Plt count         | 1.000                        | 0.996-1.003 | 0.806   |
| lymphocyte count  | 0.780                        | 0.386-1.575 | 0.488   |
| neutrophile count | 0.929                        | 0.780-1.107 | 0.410   |
| Hb count          | 0.991                        | 0.972-1.010 | 0.755   |

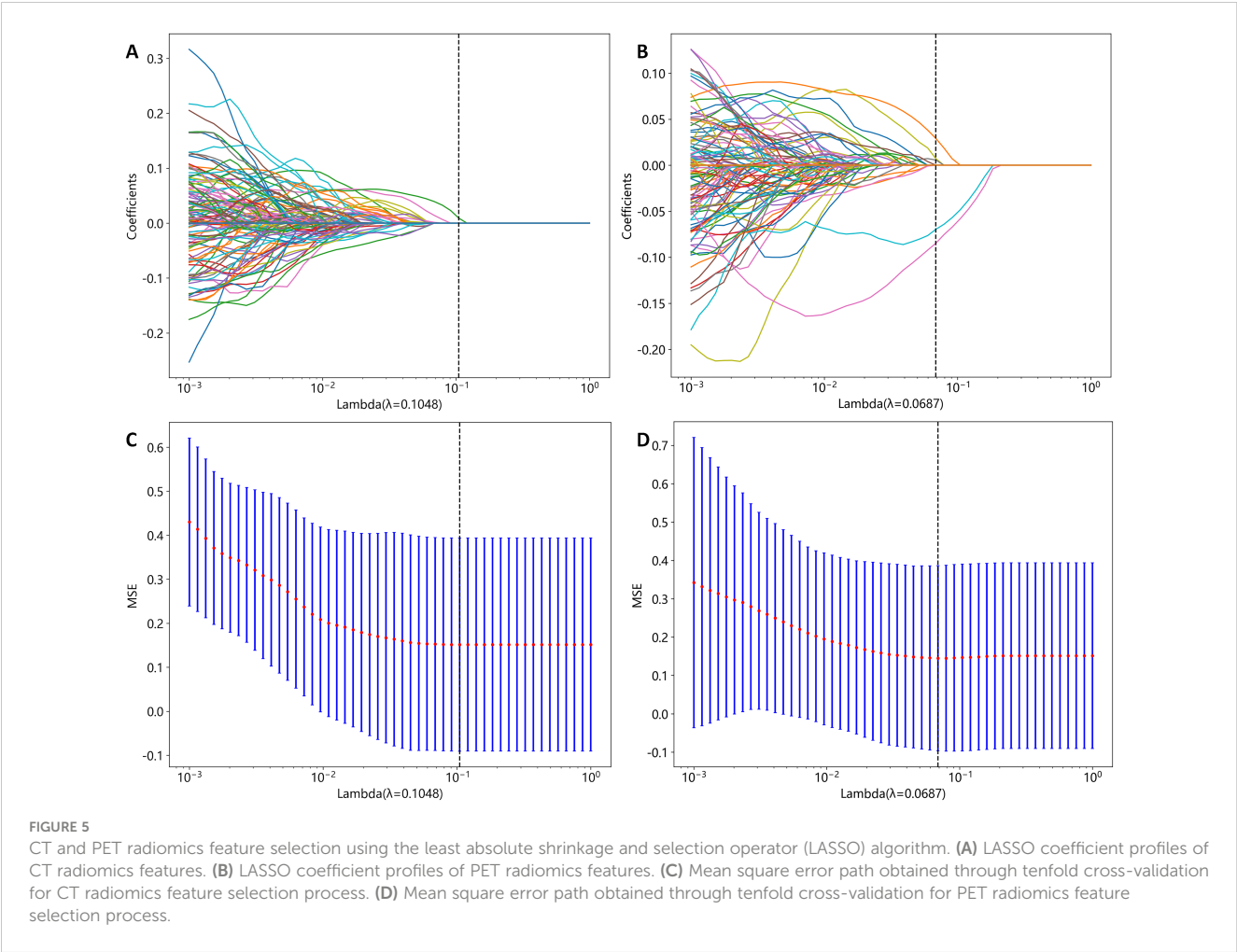
MTD, maximal tumor diameter; LNM, lymph node metastasis; MTV, metabolic active tumor volume; SUV, standardized uptake value; TLG, total lesion glycolysis; WBC, white blood cell; RBC, red blood cell; WBC, white blood cell; Plt, blood platelet; Hb, hemoglobin.

The DCA results showed that the PET radiomics model performed better and provided a higher clinical application value in differentiating pathological subtypes than CT radiomics mode and PET-CT radiomics model ([Supplementary Figure S1](#)). The calibration curves for the training cohort, internal validation cohort and external validation cohort ([Supplementary Figures S2](#)), assessed using the Hosmer-Lemeshow test, showed no significant differences in both the training cohort ( $P=0.129$ ), internal validation cohort ( $P=0.351$ ) and external validation cohort ( $P=0.529$ ). This suggests good consistency between the actual and predicted risks.

## 4 Discussion

In this study, we successfully developed six ML models based on PET and CT images, among which the lightGBM model based on PET radiomics features performed excellently in distinguishing AC and SCC.

Previous literature has indicated that CT radiomics features exhibit better predictive performance than PET radiomics features in predicting survival, and CT radiomics features are also more abundant than PET features ([22, 23](#)). Nevertheless, with regard to distinguishing pathological subtypes, Kirienko et al. ([24](#)) discovered that PET radiomics features had a greater ability to discriminate between primary and metastatic pulmonary lesions than CT



radiomics features. Further studies conducted by Hyun et al. (25) and Han et al. (21) demonstrated that a PET/CT-based machine learning method was able to make a distinction between AC and SCC in patients with non-small cell lung cancer. This study indicates that in distinguishing SCC and AC, the selected PET radiomics features are substantially more numerous than CT radiomics features, and the performance of the PET radiomics model is notably superior to that of the CT radiomics model. Furthermore, the Delong test showed that although there was a slight improvement in performance when combining PET radiomics features with CT radiomics features, the increase in AUC value did not reach statistical significance (the p values of

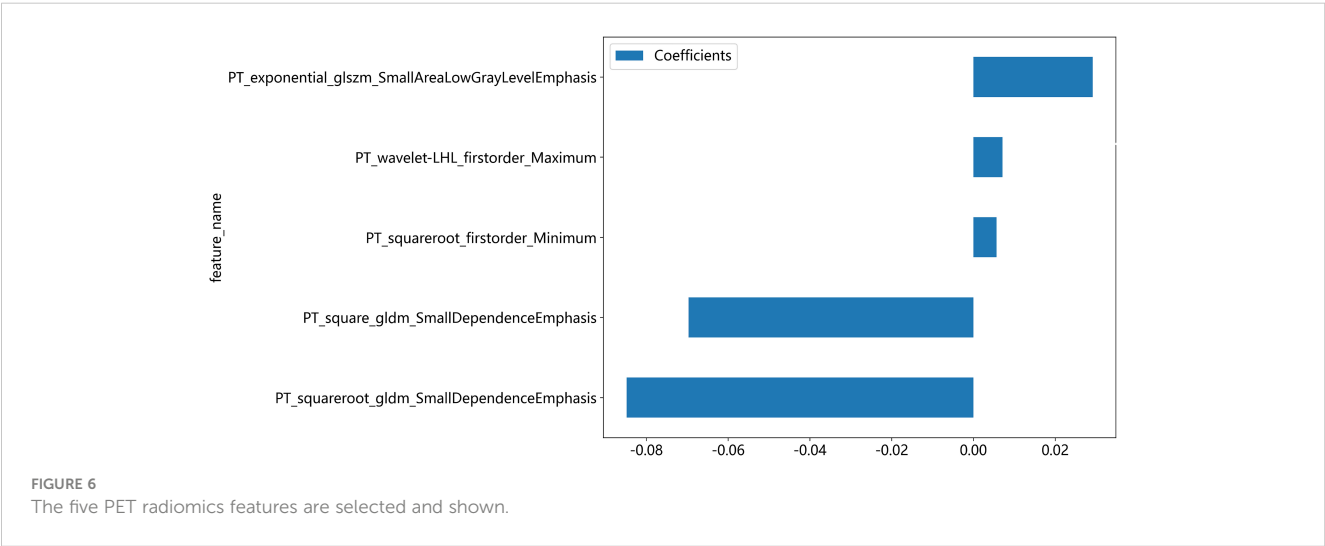


TABLE 3 The final PET and CT radiomics features used for models.

| Image | Filter        | Feature class | Feature                            |
|-------|---------------|---------------|------------------------------------|
| PET   | Exponential   | GLSZM         | Small area low gray level emphasis |
|       | Wavelet (LHL) | First-order   | Maximum                            |
|       | Square        | GLDM          | Small dependence emphasis          |
|       | Squareroot    | First-order   | Minimum                            |
|       | Squareroot    | GLDM          | Small dependence emphasis          |
| CT    | Wavelet (HLL) | NGTDM         | Busyness                           |

the training and internal validation cohorts were 0.347 and 0.776, respectively). In a retrospective study, Shen et al. (19) first found that short-zone emphasis (SZE) from GLSZM was the only PET-based radiomics feature that showed quantitative differences between SCC and non-SCC in cervical cancer. Tsujikawa et al. (20) reported that the correlation from normalized gray-level co-occurrence matrix (NGLCM) was the only feature extracted from <sup>18</sup>F-FDG PET that showed significant differences between cervical SCC and non-SCC. The previous two published studies extracted merely 18 or 76 features from the original images. In contrast, our study extracted 2818 features from the original images as well as the converted images. This might be the cause for which the previous study could select only one meaningful feature, while our study selected five. In summary, the findings of this series of PET/CT radiomic studies highlights the importance of functional imaging-based radiomics research in differentiating tumor pathological subtypes. This may be related to the FDG uptake heterogeneity

between AC and SCC, which is consistent with the identification of the pathological subtypes of lung cancer based on PET/CT (21).

MRI techniques also introduced various functional sequences, including apparent diffusion coefficient (ADC), dynamic contrast-enhanced imaging, and perfusion-weighted imaging (26). Wang et al. (18) achieved good differentiation between SCC and AC using a multiparameter MRI radiomics model based on ADC, enhanced T1-weighted imaging, and other anatomical and functional sequences. Although the differentiation performance of the multiparametric MRI-based radiomics model was the highest among the published MRI-based radiomics studies, its differentiation performance (AUC = 0.89) was lower than that of the pure PET-based radiomics model constructed in our study (AUC = 0.955). These findings demonstrate the advantages of PET radiomics features over multiparametric MRI radiomics features to a certain extent. PET-based radiomics can not only reveal the intratumoral heterogeneity of imaging structures between AC and SCC but also demonstrate the heterogeneity of tumor cell metabolism. Moreover, radiomics features are based on manually segmented ROIs in five MRI sequences, which not only requires a substantial amount of work but also increases the instability of the features.

The radiomics features selected in our study are all derived from processed images, which may reveal greater tumor heterogeneity differences between SCC and AC compared to the original images, showcasing the advantage of radiomics. Among these radiomics features selected in our study, the firstorder\_Maximum and firstorder\_Minimum represent the maximum and minimum gray level intensities, respectively. SCC exhibits significantly higher values than AC, indicating that SCC has a stronger FDG uptake than AC. Campos-Parra et al. found that compared to AC, SCC exhibits higher activation levels of key cancer pathways, such as IL-17, JAK/STAT, and Ras signaling (27). high-risk human papilloma virus (HPV) -16

TABLE 4 Performance of machine learning classifiers for differentiating pathological subtypes in the training and internal validation cohort.

| ML       | DS | PET radiomics model |               |       |       |       | CT radiomics model |               |       |       |       |
|----------|----|---------------------|---------------|-------|-------|-------|--------------------|---------------|-------|-------|-------|
|          |    | AUC                 | 95% CI        | ACC   | SEN   | SPE   | AUC                | 95% CI        | ACC   | SEN   | SPE   |
| LR       | T  | 0.916               | 0.852 - 0.979 | 0.919 | 0.714 | 0.957 | 0.597              | 0.441 - 0.753 | 0.779 | 0.429 | 0.843 |
|          | V  | 0.779               | 0.631 - 0.928 | 0.814 | 0.571 | 0.889 | 0.521              | 0.330 - 0.711 | 0.746 | 0.286 | 0.909 |
| NB       | T  | 0.848               | 0.739 - 0.957 | 0.919 | 0.667 | 0.965 | 0.684              | 0.549 - 0.820 | 0.603 | 0.762 | 0.574 |
|          | V  | 0.719               | 0.517 - 0.921 | 0.847 | 0.643 | 0.911 | 0.524              | 0.334 - 0.712 | 0.746 | 0.286 | 0.909 |
| SVM      | T  | 0.941               | 0.885 - 0.998 | 0.941 | 0.857 | 0.957 | 0.612              | 0.465 - 0.760 | 0.632 | 0.619 | 0.635 |
|          | V  | 0.811               | 0.647 - 0.975 | 0.864 | 0.786 | 0.889 | 0.484              | 0.287 - 0.681 | 0.780 | 0.214 | 0.977 |
| KNN      | T  | 0.96                | 0.931 - 0.989 | 0.824 | 1.000 | 0.791 | 0.802              | 0.735 - 0.870 | 0.559 | 1.000 | 0.478 |
|          | V  | 0.700               | 0.535 - 0.865 | 0.847 | 0.357 | 1.000 | 0.417              | 0.253 - 0.582 | 0.763 | 0.071 | 1.000 |
| LightGBM | T  | 0.955               | 0.922 - 0.988 | 0.868 | 0.952 | 0.852 | 0.752              | 0.642 - 0.862 | 0.713 | 0.667 | 0.761 |
|          | V  | 0.851               | 0.715 - 0.986 | 0.915 | 0.643 | 1.000 | 0.513              | 0.339 - 0.688 | 0.661 | 0.286 | 0.814 |
| MLP      | T  | 0.930               | 0.877 - 0.984 | 0.809 | 0.905 | 0.791 | 0.597              | 0.440 - 0.753 | 0.779 | 0.429 | 0.843 |
|          | V  | 0.816               | 0.667 - 0.965 | 0.847 | 0.643 | 0.911 | 0.521              | 0.330 - 0.711 | 0.746 | 0.286 | 0.909 |

ML, machine learning; DS, data set; PET, positron emission tomography; CT, computed tomography; AUC, area under the curve; CI, confidence interval; ACC, Accuracy; SEN, Sensitivity; SPE, Specificity; LR, logistic regression; T, training cohort; V, internal validation cohort; NB, Naive Bayes; SVM, support vector machine; KNN, k-nearest neighbors; lightGBM, light gradient boosting machine; MLP, multilayer perceptron neural network.

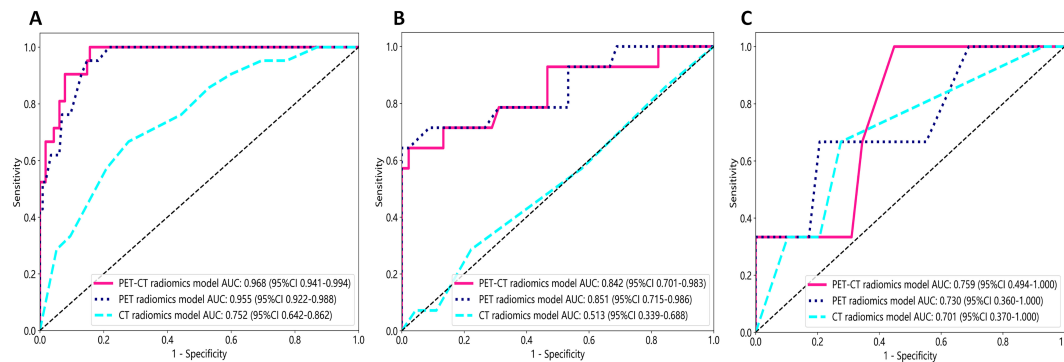


FIGURE 7

The receiver operating characteristic (ROC) curves of all three radiomics models were used to differentiate pathological subtypes in the training cohort (A), internal validation cohort (B), and external validation cohort (C).

infection is more common in SCC, while HPV-18 and HPV-45 are more frequently observed in AC (27, 28). Priego-Hernández et al. discovered that cervical cancer and HPV-16-positive cell lines have increased expression of HIF-1 $\alpha$  and glucose metabolism-related genes (GLUT1, LDHA, CAIX, MCT4, and BSG) (29). Furthermore, there are significant variations in the expression of glucose metabolism-related genes between SCC and AC (30). Choi et al. demonstrated that tumor FDG uptake is associated with glucose transporters (Glut-1 and Glut-3), with SCC exhibiting higher expression intensity and proportion of Glut-1 compared to AC. Consequently, SCC demonstrates higher SUVmax and stronger FDG uptake capacity (31). Small dependence emphasis (SDE) from GLDM and small area low gray level emphasis (SALGLE) from GLSZM represent tumor heterogeneity, with higher values indicating more significant heterogeneity. In our study, GLDM\_SDE and GLSZM\_SALGLE features were significantly higher in SCC compared to AC, indicating that the intratumoral metabolic heterogeneity based on PET imaging in SCC is significantly higher than that in AC. This may be related to the previously mentioned metabolic and histomorphological differences between SCC and AC. The tissue structure of SCC is tight, with small gaps between tumor cells, wrapped in several matrix structures, forming cancer nests. In contrast, the tissue structure of AC is more loose, characterized by glandular differentiation. Therefore, the differential expression of pathogenic molecular mechanisms, especially glucose metabolism genes, determines the metabolic differences of tumor cells, while cell arrangement and tissue morphology determine the spatial heterogeneity of tumor cells. The tumor heterogeneity revealed by PET images manifests these metabolic differences and spatial heterogeneity of tumor cells. These findings require further validation with a larger-scale patient or in combination with pathomics.

In this study, we employed six ML algorithms to develop models for distinguishing SCC and AC. Among the algorithms, the radiomics model constructed by the LightGBM algorithm exhibited excellent differentiation performance, accuracy, sensitivity, and specificity with a relatively balanced performance. This finding is consistent with a similar study conducted by Lam et al., who investigated the correlation between radiomics features

and tumor mutation burden in glioma based on MRI images using LR, SVM, and six other ML algorithms (32). They found that the radiomics model constructed by the LightGBM algorithm also demonstrated the best discriminative performance with relatively balanced sensitivity and specificity. Furthermore, researchers have successfully achieved good discriminative performance in distinguishing low-grade and high-grade meningiomas using the LightGBM algorithm for both radiomics and deep learning models (33). Similarly, Chang et al. constructed LightGBM and convolutional neural network (CNN) models based on noncontrast CT and enhanced images to differentiate thymic epithelial tumors from other anterior mediastinal tumors (34). The results demonstrated that the LightGBM model outperformed the CNN model in both the noncontrast CT dataset and the enhanced CT dataset. The LightGBM algorithm, which is based on the gradient boosting decision tree model, optimizes the search for optimal split points and the tree growth process. It supports efficient parallel training and possesses advantages such as faster training speed, lower memory consumption, better accuracy, and quick processing of massive data, making it widely applicable. Therefore, ML can better handle complex nonlinear relationships in large-scale datasets and holds great potential for clinical applications (35). However, it is important to acknowledge that ML models and algorithms also have limitations, including overfitting and lack of interpretability. Overfitting can undermine predictive performance, while the lack of interpretability can hinder the use of ML (36). Hence, it is essential to prioritize the future optimization of ML algorithms and conduct independent validations to verify their performance.

There were several limitations in this study. Firstly, it was a retrospective and preliminary study, carrying a potential selection bias despite the use of strict inclusion and exclusion criteria. Secondly, HPV status and histological differentiation were not available for some patients when retrieving the electronic medical record system, and we were unable to further explore their impact on pathological subtypes. Lastly, the sample size of AC in this study is relatively small, but this is consistent with the epidemiology of cervical cancer. To improve the generalizability of the model, it is necessary to investigate a larger sample size from multiple centers in future research.



## 5 Conclusion

The lightGBM-based PET radiomics model effectively identified pathological subtypes in patients with locally advanced cervical cancer and may help clinicians in their daily decision-making process.

## 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 humans were approved by the Affiliated Cancer Hospital of Shandong First Medical University and the Affiliated Cancer Hospital of Xinjiang 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 This study was retrospective study. Ethical approval was obtained from the Institutional Review Board of the Affiliated Cancer Hospital of Shandong First Medical University (No. SDTHEC2023006030) and the Affiliated Cancer Hospital of Xinjiang Medical University (No. K-2022006), and the requirement for written informed consent was waived.

## Author contributions

HL: Conceptualization, Data curation, Formal analysis, Writing – original draft, Methodology, Validation. LM: Formal analysis, Software, Writing – review & editing. CC: Data curation, Supervision, Writing – review & editing. YZ: Investigation, Writing – review & editing. YY: Writing – review & editing, Conceptualization, Data curation, Methodology. RW: Conceptualization, Funding acquisition, Project administration, Writing – review & editing, Methodology.

## References

1. 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
2. Zheng R, Zhang S, Zeng H, Wang S, Sun K, Chen R, et al. Cancer incidence and mortality in China, 2016. *J Natl Cancer Center.* (2022) 2:1–9. doi: 10.1016/j.jncc.2022.02.002
3. Smith AJB, Beavis AL, Rositch AF, Levinson K. Disparities in diagnosis and treatment of cervical adenocarcinoma compared with squamous cell carcinoma: an analysis of the national cancer database, 2004–2017. *J lower genital tract Dis.* (2023) 27:29–34. doi: 10.1097/igt.0000000000000702
4. Hu K, Wang W, Liu X, Meng Q, Zhang F. Comparison of treatment outcomes between squamous cell carcinoma and adenocarcinoma of cervix after definitive radiotherapy or concurrent chemoradiotherapy. *Radiat Oncol (London England).* (2018) 13:249. doi: 10.1186/s13014-018-1197-5
5. Su GH, Xiao Y, You C, Zheng RC, Zhao S, Sun SY, et al. Radiogenomic-based multiomic analysis reveals imaging intratumor heterogeneity phenotypes and therapeutic targets. *Sci Adv.* (2023) 9:eadf0837. doi: 10.1126/sciadv.adf0837
6. Rajaram S, Gupta B. Screening for cervical cancer: Choices & dilemmas. *Indian J Med Res.* (2021) 154:210–20. doi: 10.4103/ijmr.IJMR\_857\_20
7. Sasieni P, Castanon A, Cuzick J. Screening and adenocarcinoma of the cervix. *Int J Cancer.* (2009) 125:525–9. doi: 10.1002/ijc.24410

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was funded by the Key Research and Development Program of Xinjiang Uygur Autonomous Region of China (Grant No. 2022B03019-5), the Special Funds Project of Central Guidance on Local Science and Technology Development (Grant No. ZYYD2022B18), the State Key Laboratory of Pathogenesis, Prevention and Treatment of High Incidence Diseases in Central Asian Fund (Grant No. SKL-HIDCA-2020-GJ4) and Tianshan Talent Program-Science and Technology Innovation Team (Grant No. 2023TSYCTD0015).

## Acknowledgments

The authors thank all their coworkers involved in the study for their support and assistance.

## 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/fonc.2024.1346336/full#supplementary-material>

8. Fan A, Zhang L, Wang C, Wang Y, Han C, Xue F. Analysis of clinical factors correlated with the accuracy of colposcopically directed biopsy. *Arch gynecol obstet.* (2017) 296:965–72. doi: 10.1007/s00404-017-4500-z
9. Kim SI, Kim SJ, Suh DH, Kim K, No JH, Kim YB. Pathologic discrepancies between colposcopy-directed biopsy and loop electrosurgical excision procedure of the uterine cervix in women with cytologic high-grade squamous intraepithelial lesions. *J gynecol Oncol.* (2020) 31:e13. doi: 10.3802/jgo.2020.31.e13
10. Nassiri S, Aminimoghaddam S, Sadaghian MR, Nikandish M, Jamshidnezhad N, Saffarieh E. Evaluation of the diagnostic accuracy of the cervical biopsy under colposcopic vision. *Eur J Trans myol.* (2022) 32:10670. doi: 10.4081/ejtm.2022.10670
11. Winfield JM, Orton MR, Collins DJ, Ind TE, Attygalle A, Hazell S, et al. Separation of type and grade in cervical tumours using non-mono-exponential models of diffusion-weighted MRI. *Eur Radiol.* (2017) 27:627–36. doi: 10.1007/s00330-016-4417-0
12. Haldorsen IS, Lura N, Blaakaer 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. doi: 10.1007/s11912-019-0824-0
13. McCague C, Ramllee S, Reinius M, Selby I, Hulse D, Piyatissa P, et al. Introduction to radiomics for a clinical audience. *Clin Radiol.* (2023) 78:83–98. doi: 10.1016/j.crad.2022.08.149
14. Litvin AA, Burkin DA, Kropinov AA, Paramzin FN. Radiomics and digital image texture analysis in oncology (Review). *Sovremennyye tekhnologii v meditsine.* (2021) 13:97–104. doi: 10.17691/stm2021.13.2.11
15. Davnall F, Yip CS, Ljungqvist G, Selmi M, Ng F, Sanghera B, et al. Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice? *Insights into Imaging.* (2012) 3:573–89. doi: 10.1007/s13244-012-0196-6
16. Patkulkar PA, Subbalakshmi AR, Jolly MK, Sinharay S. Mapping spatiotemporal heterogeneity in tumor profiles by integrating high-throughput imaging and omics analysis. *ACS omega.* (2023) 8:6126–38. doi: 10.1021/acsomega.2c06659
17. Liu Z, Wang S, Dong D, Wei J, Fang C, Zhou X, 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
18. Wang W, Jiao Y, Zhang L, Fu C, Zhu X, Wang Q, et al. Multiparametric MRI-based radiomics analysis: differentiation of subtypes of cervical cancer in the early stage. *Acta radiol (Stockholm Sweden: 1987).* (2022) 63:847–56. doi: 10.1177/02841851211014188
19. Shen WC, Chen SW, Liang JA, Hsieh TC, Yen KY, Kao CH. [18] Fluorodeoxyglucose positron emission tomography for the textural features of cervical cancer associated with lymph node metastasis and histological type. *Eur J Nucl Med Mol Imaging.* (2017) 44:1721–31. doi: 10.1007/s00259-017-3697-1
20. Tsujikawa T, Rahman T, Yamamoto M, Yamada S, Tsuyoshi H, Kiyono Y, et al. <sup>18</sup>F-FDG PET radiomics approaches: comparing and clustering features in cervical cancer. *Ann Nucl Med.* (2017) 31:678–85. doi: 10.1007/s12149-017-1199-7
21. Han Y, Ma Y, Wu Z, Zhang F, Zheng D, Liu X, et al. Histologic subtype classification of non-small cell lung cancer using PET/CT images. *Eur J Nucl Med Mol Imaging.* (2021) 48:350–60. doi: 10.1007/s00259-020-04771-5
22. Liu S, Li R, Liu Q, Sun D, Yang H, Pan H, et al. Radiomics model of <sup>18</sup>F-FDG PET/CT imaging for predicting disease-free survival of early-stage uterine cervical squamous cancer. *Cancer biomark: section A Dis Markers.* (2022) 33:249–59. doi: 10.3233/cbm-210201
23. Kirienko M, Cozzi L, Antunovic L, Lozza L, Fogliata A, Voulaz E, et al. Prediction of disease-free survival by the PET/CT radiomic signature in non-small cell lung cancer patients undergoing surgery. *Eur J Nucl Med Mol Imaging.* (2018) 45:207–17. doi: 10.1007/s00259-017-3837-7
24. Kirienko M, Cozzi L, Rossi A, Voulaz E, Antunovic L, Fogliata A, et al. Ability of FDG PET and CT radiomics features to differentiate between primary and metastatic lung lesions. *Eur J Nucl Med Mol Imaging.* (2018) 45:1649–60. doi: 10.1007/s00259-018-3987-2
25. Hyun SH, Ahn MS, Koh YW, Lee SJ. A machine-learning approach using PET-based radiomics to predict the histological subtypes of lung cancer. *Clin Nucl Med.* (2019) 44:956–60. doi: 10.1097/RLU.00000000000002810
26. Matani H, Patel AK, Horne ZD, Beriwal S. Utilization of functional MRI in the diagnosis and management of cervical cancer. *Front Oncol.* (2022) 12:1030967. doi: 10.3389/fonc.2022.1030967
27. Campos-Parra AD, Pérez-Quintanilla M, Martínez-Gutiérrez AD, Pérez-Montiel D, Coronel-Martínez J, Millán-Catalán O, et al. Molecular differences between squamous cell carcinoma and adenocarcinoma cervical cancer subtypes: potential prognostic biomarkers. *Curr Oncol (Toronto Ont).* (2022) 29:4689–702. doi: 10.3390/curroncol29070372
28. Wang WP, An JS, Yao HW, Li N, Zhang YY, Ge L, et al. Prevalence and attribution of high-risk HPV in different histological types of cervical cancer. *Zhonghua fu chan ke za zhi.* (2019) 54:293–300. doi: 10.3760/cma.j.issn.0529-567x.2019.05.002
29. Priego-Hernández VD, Arizmendi-Izazaga A, Soto-Flores DG, Santiago-Ramón N, Feria-Valadez MD, Navarro-Tito N, et al. Expression of HIF-1 $\alpha$  and genes involved in glucose metabolism is increased in cervical cancer and HPV-16-positive cell lines. *Pathog (Basel Switzerland).* (2022) 12:33. doi: 10.3390/pathogens12010033
30. Martínez-Morales P, Morán Cruz I, Roa-de la Cruz L, Maycotte P, Reyes Salinas JS, Vazquez Zamora VJ, et al. Hallmarks of glycogene expression and glycosylation pathways in squamous and adenocarcinoma cervical cancer. *PeerJ.* (2021) 9:e12081. doi: 10.7717/peerj.12081
31. Choi WH, Yoo IR, O JH, Kim TJ, Lee KY, Kim YK. Is the Glut expression related to FDG uptake in PET/CT of non-small cell lung cancer patients? *Technol Health care: Off J Eur Soc Eng Med.* (2015) 23 Suppl 2:S311–8. doi: 10.3233/thc-150967
32. Lam LHT, Chu NT, Tran TO, Do DT, Le NQK. A radiomics-based machine learning model for prediction of tumor mutational burden in lower-grade gliomas. *Cancers.* (2022) 14:3492. doi: 10.3390/cancers14143492
33. Yang L, Xu P, Zhang Y, Cui N, Wang M, Peng M, et al. A deep learning radiomics model may help to improve the prediction performance of preoperative grading in meningioma. *Neuroradiology.* (2022) 64:1373–82. doi: 10.1007/s00234-022-02894-0
34. Chang CC, Tang EK, Wei YF, Lin CY, Wu FZ, Wu MT, et al. Clinical radiomics-based machine learning versus three-dimension convolutional neural network analysis for differentiation of thymic epithelial tumors from other prevascular mediastinal tumors on chest computed tomography scan. *Front Oncol.* (2023) 13:1105100. doi: 10.3389/fonc.2023.1105100
35. Luo W. Predicting cervical cancer outcomes: statistics, images, and machine learning. *Front Artif Intell.* (2021) 4:627369. doi: 10.3389/frai.2021.627369
36. Luo Y, Chen S, Valdes G. Machine learning for radiation outcome modeling and prediction. *Med Phys.* (2020) 47:e178–e84. doi: 10.1002/mp.13570

# Frontiers in Oncology

Advances knowledge of carcinogenesis and tumor progression for better treatment and management

The third most-cited oncology journal, which highlights research in carcinogenesis and tumor progression, bridging the gap between basic research and applications to improve diagnosis, therapeutics and management strategies.

## Discover the latest Research Topics

See more →

### Frontiers

Avenue du Tribunal-Fédéral 34  
1005 Lausanne, Switzerland  
[frontiersin.org](https://frontiersin.org)

### Contact us

+41 (0)21 510 17 00  
[frontiersin.org/about/contact](https://frontiersin.org/about/contact)

