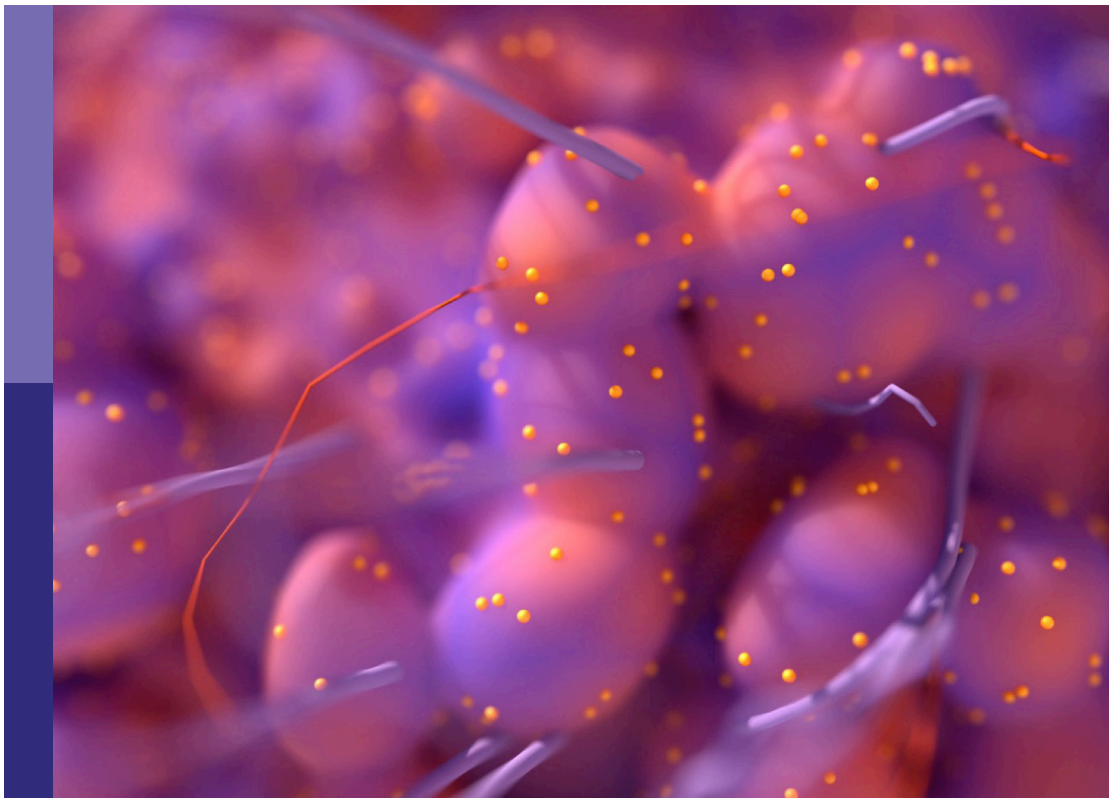
frontiers | Research Topics

Artificial intelligence in digital pathology image analysis, 2nd edition

Edited by
Min Tang, Jialiang Yang and Li Xiao

Published in
Frontiers in Oncology
Frontiers in Bioinformatics



ISSN 1664-8714
ISBN 978-2-8325-5502-6
DOI 10.3389/978-2-8325-5502-6

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

Artificial intelligence in digital pathology image analysis, 2nd edition

Topic editors
Min Tang — Jiangsu University, China
Jialiang Yang — Geneis (Beijing) Co. Ltd, China
Li Xiao — University of Science and Technology of China, China

Citation
Tang, M., Yang, J., Xiao, L., eds. (2024). *Artificial intelligence in digital pathology image analysis, 2nd edition*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-5502-6

Publisher's note: This is a 2nd edition due to an article retraction.

Table of contents

frontiers | Frontiers in Bioinformatics

TYPE Editorial
PUBLISHED 13 April 2023
DOI 10.3389/fbinf.2023.1007986

OPEN ACCESS

EDITED AND REVIEWED BY
Kevin Eliceiri,
University of Wisconsin-Madison,
United States

*CORRESPONDENCE
Min Tang,
mt3138@ujs.edu.cn

RECEIVED 31 July 2022
ACCEPTED 04 April 2023
PUBLISHED 13 April 2023

CITATION
Liu Y, Liu X, Zhang H, Liu J, Shan C, Guo Y, Gong X and Tang M (2023), Editorial: Artificial intelligence in digital pathology image analysis.
Front. Bioinform. 3:1007986.
doi: 10.3389/fbinf.2023.1007986

Editorial: Artificial intelligence in digital pathology image analysis

Yi Liu[1,2], Xiaoyan Liu[1], Hantao Zhang[1], Junlin Liu[1], Chaofan Shan[1], Yinglu Guo[1], Xun Gong[3] and Min Tang[1]*

[1]School of Life Sciences, Jiangsu University, Zhenjiang, Jiangsu, China, [2]Institute of Animal Science, Jiangsu Academy of Agricultural Sciences, Nanjing, China, [3]Affiliated Hospital of Jiangsu University, Zhenjiang, China

KEYWORDS

artificial intelligence, digital pathology, disease diagnosis, interpretable deep learning, multi-modality, precision medicine

Editorial on the Research Topic
Artificial intelligence in digital pathology image analysis

In the 21st century, cancer is the top cause of death in hospitals and the key limitation of life expectancy in most countries (Luo et al.). The analysis of medical images including histopathological slides, radiological images such as magnetic resonance imaging (MRI) and CT, and ultrasound images, etc. is an essential tool in cancer research, disease diagnosis and treatment. Moreover, the availability of faster networks and cheaper storage solutions make these images easier to manage and share, leading to the emergence of digital pathology images, for example, whole slide imaging (WSI). However, extracting important information from these images for clinical use requires a big effort from pathologist and is also error-prone due to inexperience and fatigue. Recently, Artificial intelligence (AI) such as deep learning (DL) shows clear potential to mine image features from medical images, better quantitative model disease appearance and hence possibly improve prediction of disease aggressiveness and patient outcome. The application of AI not only reduces the burden on pathologists but also saves high costs and time, thus attracting great attention.

In this editorial, we presented an account of how AI has greatly facilitated digital pathology image analysis as well as other medical image analysis. This editorial is based on 11 research articles, 1 regular review and a methods article, shedding light on the power of AI to analyze medical images including but not limited to magnetic resonance imaging (MRI), CT images, and digital pathology images, primarily WSI.

Five research articles use machine learning to construct prediction models based on pathological and radiological images. Wang D. et al. developed an automated machine-learning framework for predicting IDH1 mutation status in glioma. In their framework, a random forest algorithm is applied to select relevant features in regions of interest (ROIs) extracted from high-resolution pathology slides and multi-sequence MRI scans. The model integrating histopathological and radiological information can predict glioma IDH genotype with greater accuracy and reliability (Wang D. et al.). Zhao F. et al. used random forest combined with hyperparameter tuning for feature selection and radiomics prediction modeling to distinguish invasive adenocarcinoma (IAC) and minimal invasive adenocarcinoma (MIA) presenting as ground-glass nodules (GGNs). The result of ROC curve showed that their model effectively distinguished IAC from MIA presenting as GGNs and represented a non-invasive, low-cost, rapid, and reproducible preoperative prediction method for clinical application (Zhao F. et al.). In Wang X. et al.'s study, Mann–Whitney *U*

test and least absolute shrinkage and selection operator (LASSO) were applied for feature preselection and radiomic signature construction based on CT. SVM-linear models were trained by incorporating the radiomic signature with clinical characteristics. Importantly, they chose the optimal model to build a nomogram which could be useful to preoperatively predict histologic grade in pancreatic neuroendocrine tumors (Wang X. et al.). Hu et al. also utilized LASSO to select radiomics signatures. The Logistic algorithm and a combinatorial modeling approach were used to establish unimodal radiomics models and multimodal radiomics models respectively based on tumors and peritumors extracted from enhanced MRI images. The radiomics signatures of the dual regions for tumor and peritumor were fund to be of significance to predicty microvascular invasion risk grades in hepatocellular carcinoma preoperatively (Hu et al.). Furthermore, through multivariate logistic regression analysis and statistical analysis, Luo et al. found that shear wave elastography (SWE) examination, a newly emerging elastography technique which can display tissue stiffness in a quantified form to obtain the biological information of the primary lesion, can be used as a routine auxiliary method of conventional ultrasonic examination for axillary node metastasis and the elastic modulus values of SWE had no significant correlation with the molecular types of breast cancer (Luo et al.).

Several research articles focus on the application of deep-learning in image analysis for detection, feature extraction, and tissue classification. Based on whole slide imaging, Tao et al. used deep learning (DL) including AlexNet, VGG-16, Inception V3, DenseNet-121, ResNet-50, and MnasNet to classify bone tumors histopathologically in terms of aggressiveness. The results showed that DL can effectively classify bone tumors similar to senior pathologists, which is promising and would help expedite the future application of DL-assisted histopathological diagnosis for bone tumors (Tao et al.). The most of traditional DL models were also applied in Guo et al.'s study for automatically detecting circulating tumor cell (CTC) which is a critical biomarker for cancer diagnosis and prognosis based on immunofluorescence *in situ* hybridization (imFISH) images. Additionally, they used transfer learning to improve the prediction performance and save computing resources. Both DL and transfer learning detected CTCs with high sensitivity (Guo et al.). Xu Y. et al. evaluated DL model for predicting tumor invasiveness of ground-glass nodules (GGNs) through analyzing time series CT images (baseline CT and 3-month follow-up CT images). They also evaluated the effect of different ROIs on prediction. The DL model integrating full ROIs that contain both tumor and peritumor regions from serial CT images showed improved predictive performance, which could benefit the clinical management of GGNs (Xu Y. et al.). Moreover, Berberine was found to Suppress stemness and tumorigenicity of colorectal cancer stem-like cells by inhibiting m6A methylation in Zhao Z. et al.'s study by experiment (Zhao Z. et al.).

Notably, four articles proposed novel DL models to better analyze medical images based on existing algorithm. Luan et al. proposed a neural network (S-Net) which obtained more semantic information with the introduction of an attention mechanism and long jump connection, thus effectively improving the effect of liver tumors' automated segmentation from CT images (Luan et al.). Xu F. et al. built a DL core-needle biopsy (DL-CNB) model on the attention-based multiple instance-learning frameworks to predict axillary lymph node metastasis in early breast cancer utilizing the DL features, which were extracted from the cancer areas of WSIs of breast CNB specimens annotated by two pathologists. And the interpretation of DL-CNB model showed that the top signatures most predictive of ALN metastasis were characterized by the nucleus features (Xu F. et al.). Based on the Residual Network (ResNet) model, Liu et al. proposed a deep learning method, DeepHE. On images of tissue, DeepHE can efficiently identify and analyze characteristics of tumor cells to predict the Tumor mutational burden (TMB) avoiding whole-exome sequencing which is a standard but costly and inefficient method to measure TMB (Liu et al.). Yang et al. proposed a novel deep learning framework to predict the benign and malignant thyroid nodules accurately. They first trained a ResNet18 model by an ultrasound image dataset. Gradient-weighted Class Activation Mapping (Grad-CAM) was then proposed to highlight sensitive regions, extracting the sensitive regions and analyzing their shape features. Shape features of the sensitive regions were helpful in diagnosis to a great extent (Yang et al.).

Finally, the review article summarized process and key steps of current pathological image processing including image preprocessing, image segmentation, feature extraction and model construction, to help researchers choose more suitable medical image processing methods and predict cancer-related biomarkers more accurately (Xie et al.).

In conclusion, the articles in this Research Topic show how AI is applied to analyze medical images including digital pathology images. For fully utilizing data from various pathological and radiological images to gain insights into disease, the applications of AI are becoming more relevant every day.

Author contributions

All authors have made a direct and intellectual contribution to this topic and approved the article for publication.

Funding

This work was supported by grants from Jiangsu University (19JDG039) and the National Natural Science Foundation of China (32002235).

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

frontiers in Oncology

ORIGINAL RESEARCH
published: 09 August 2021
doi: 10.3389/fonc.2021.680807

Adaptive Attention Convolutional Neural Network for Liver Tumor Segmentation

Shunyao Luan[1†], *Xudong Xue*[2†], *Yi Ding*[2], *Wei Wei*[2*] *and Benpeng Zhu*[1*]

[1] Department of Optoelectronic Engineering, Huazhong University of Science and Technology, Wuhan, China, [2] Oncology Radiotherapy Department, Hubei Cancer Hospital, Wuhan, China

Purpose: Accurate segmentation of liver and liver tumors is critical for radiotherapy. Liver tumor segmentation, however, remains a difficult and relevant problem in the field of medical image processing because of the various factors like complex and variable location, size, and shape of liver tumors, low contrast between tumors and normal tissues, and blurred or difficult-to-define lesion boundaries. In this paper, we proposed a neural network (S-Net) that can incorporate attention mechanisms to end-to-end segmentation of liver tumors from CT images.

Methods: First, this study adopted a classical coding-decoding structure to realize end-to-end segmentation. Next, we introduced an attention mechanism between the contraction path and the expansion path so that the network could encode a longer range of semantic information in the local features and find the corresponding relationship between different channels. Then, we introduced long-hop connections between the layers of the contraction path and the expansion path, so that the semantic information extracted in both paths could be fused. Finally, the application of closed operation was used to dissipate the narrow interruptions and long, thin divide. This eliminated small cavities and produced a noise reduction effect.

Results: In this paper, we used the MICCAI 2017 liver tumor segmentation (LiTS) challenge dataset, 3DIRCADb dataset and doctors' manual contours of Hubei Cancer Hospital dataset to test the network architecture. We calculated the Dice Global (DG) score, Dice per Case (DC) score, volumetric overlap error (VOE), average symmetric surface distance (ASSD), and root mean square error (RMSE) to evaluate the accuracy of the architecture for liver tumor segmentation. The segmentation DG for tumor was found to be 0.7555, DC was 0.613, VOE was 0.413, ASSD was 1.186 and RMSE was 1.804. For a small tumor, DG was 0.3246 and DC was 0.3082. For a large tumor, DG was 0.7819 and DC was 0.7632.

Conclusion: S-Net obtained more semantic information with the introduction of an attention mechanism and long jump connection. Experimental results showed that this method effectively improved the effect of tumor recognition in CT images and could be applied to assist doctors in clinical treatment.

Keywords: liver tumor, automatic segmentation, attention mechanism, CT images, deep learning

OPEN ACCESS

Edited by:
Min Tang,
Jiangsu University, China

Reviewed by:
Youwen Zhang,
University of South Carolina, United States
Juan Ye,
National Institutes of Health (NIH), United States

****Correspondence:***
Wei Wei
ww_hbch@163.com
Benpeng Zhu
benpengzhu@hust.edu.cn

[†]These authors share first authorship

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *25 March 2021*
Accepted: *12 July 2021*
Published: *09 August 2021*

Citation:
Luan S, Xue X, Ding Y, Wei W and Zhu B (2021) Adaptive Attention Convolutional Neural Network for Liver Tumor Segmentation. Front. Oncol. 11:680807. doi: 10.3389/fonc.2021.680807

INTRODUCTION

Currently, liver cancer is the fifth most common malignancy and the second-leading cause of cancer-related death worldwide (1, 2). An accurate contour of the location, volume, and shape of liver tumors can help radiotherapists develop precise treatment plans. At the present time, there are several barriers to automated segmentation of liver tumors. Lesion tissue is often uniformly gray in color, which hinders automatic segmentation. Some lesions do not have clear boundaries, which limits the performance of edge segmentation methods. The specificity of lesions exists in different samples of tumors, which vary in location, size, shape, and volume. This presents further challenges to the process of segmentation. On account of these variables, automatic segmentation of tumors from the liver is a difficult task.

To solve these problems, researchers have proposed different segmentation methods, including the regional growth method, deformation model method, intensity threshold method, and the watershed algorithm. Each of these methods has individual strengths and limitations (1–9). When compared with traditional segmentation methods, fully convolutional neural networks (FCNs) have shown powerful efficacy in segmenting liver tumors. Many researchers have introduced deep learning into the liver tumor segmentation problem and found positive results.

Since U-Net was proposed by Ronneberge (10) in 2015, it has become the most common convolutional neural network architecture in medical image segmentation. Because of this finding, more U-Net derived networks were developed like the H-DenseUnet proposed by Li et al. (11) This combines U-Net with DenseNet (12) to explore the intra- and inter-slice features. U-Net++ proposed by Zhou (13) uses full-scale hopping connectivity and deep supervision to fuse high-level semantic information with low-level semantic information from feature maps at different scales and to learn hierarchical representations from multiscale aggregated feature maps. The coding-decoding network, proposed by Ginneken et al. (14), improved the accuracy of liver tumor sketching followed by shape-based post-processing to refine liver tumor margins. Roth proposed a two-stage coarse-to-fine 3D FCN. Roth HR et al. (15) proposed a two-dimensional (2D) FCN that fused three orthogonal planes to generate voxel predictions by averaging the probabilities of the three different planes.

Along with the development of different architectures of convolutional neural networks, some special modules have been proposed like the integrated attention gate (attention U-Net) by Oktay et al. (16) This network suppresses the task-irrelevant part and enhances the learning of the task-relevant part. This greatly improved the performance of semantic segmentation. Fu et al. (17) proposed DANet, a dual-attention mechanism that used network fusing channel attention and location attention to infer attention concentrated regions from two specific and mutually independent dimensions. This improved the segmentation accuracy of the model. Woo et al. (18) proposed a network called Convolutional Block Attention Module (CBAM) fusing spatial attention mechanism and channel attention mechanism. The overall architecture of the attention mechanism (19–21) is light and easy to integrate into neural networks and engage in model training end-to-end.

Although the existing algorithms have made significant achievements in liver tumor segmentation, some networks still have large and cumbersome structures. Other networks do not effectively fuse the spatial feature information captured in the down-sampling phase with the up-sampling phase. This leads to disregarding the spatial architecture of the network. To address these problems, this study proposes a small, lightweight, end-to-end, convolutional neural network of S-Net with the fusion of spatial features and attention mechanisms. The contributions of this paper include the following:

1. Proposing a pre-processing means of pixel point-to-point flipping to improve the accuracy.
2. Using small convolutional kernels and multiple batch processing to extract smaller semantic information.
3. Using a long-hop connection between the encoder and decoder to fuse spatial features and high-level semantic features.
4. Introducing attention mechanisms in neural networks to encode a longer range of semantic information in local features and to find correspondences between different channels.

Throughout this paper, researchers used the LITS dataset, as shown in **Figure 1**. **Figure 1A** shows the original image sample with the liver region in the red dashed line, and **Figure 1B** shows the processed liver image sample with the tumor region in the shaded part in the red dashed line. The white image in **Figure 1C** is the tumor label of the original sample.

This paper is structured as follows: Section I introduces the work on liver tumor segmentation, Section II describes the research methods of this paper, Section III gives the experimental results, and Section IV presents the conclusion and summary.

MATERIALS AND METHODS

Algorithm Flow

The main process of the algorithm in this study included three stages: pre-processing, tumor segmentation, and post-processing **Figure 2**.

In the pre-processing stage, we discussed basic processing means, such as image scaling deformation, grayscale floating, pixel normalization to eliminate overfitting, pixel flipping to change the image grayscale value, and point-to-point to flip the pixel grayscale. Image scaling deformation includes the rotation, mirroring, translation, and affine transformation of each layer of the CT image with its corresponding contour outline. Image grayscale float multiplies the grayscale values of all pixel points on the image by a random number between 0.8 and 1.2, and then superimposes a random number between -0.2 and 0.2. The pixel point-to-point flip first divides each pixel point of the foreground

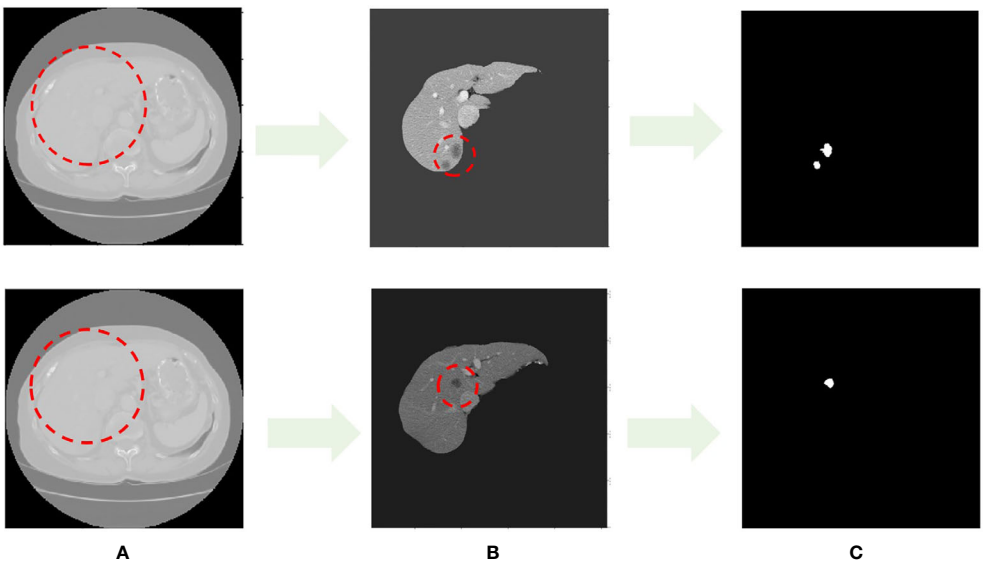


FIGURE 1 | CT image sample with liver and liver tumors. **(A)** Images in LITS dataset. **(B)** Liver image. **(C)** Tumor label.

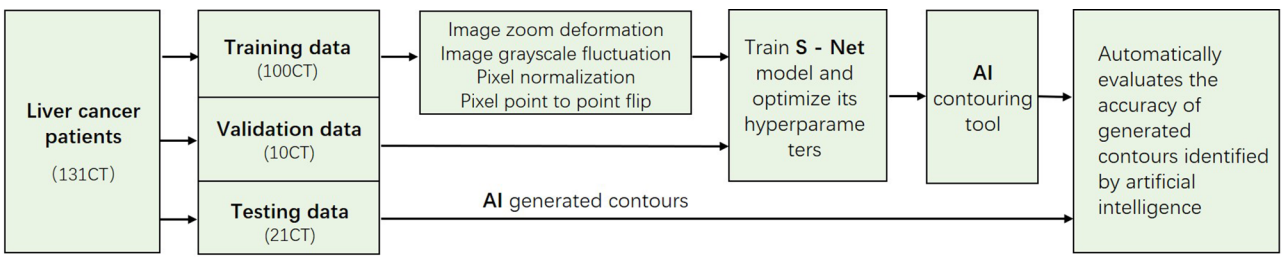


FIGURE 2 | The 2017 LiTS public dataset with a total sample of 131 patients. In this study, we adaptively partitioned the entire sample into 100 training samples, 10 validation samples, and 21 test samples. The training samples are pre-processed, and the test samples are used to evaluate the segmentation effect of the network model.

image by 255 to obtain a new pixel point and then subtracts the new pixel point by the value 1. This value is multiplied by 255 to achieve the function of the grayscale flip. The image after the point-to-point flip is shown in **Figure 3**.

In 2D sections of some samples, the overlap of Hounsfield unit (HU) values between the liver and tumor leads to poor training and makes the network model misleading, especially during the learning process. As a result of this, we used the critical threshold method throughout this study to remove the sample cuts with low contrast to increase the learning ability of the network. **Figure 4** shows the two cuts with strong and low contrast, as well as the HU diagram. In the post-processing stage, we performed noise reduction through the closed operation.

S-Net Network Architecture

This study proposed a novel convolutional neural network of S-Net based on 2D U-Net, as displayed in **Figure 5**. The architecture introduced an attention mechanism based on U-Net while using a typical encoding and decoding structure. In this structure, the left path is the contraction path (encoder) from top to bottom and the right path is the expansion path (decoder) from bottom to top. Because the target area of some samples was small, it was difficult to extract the semantic information in them. Therefore, we used small convolutional kernels and multiple batch processing for training. To extract deeper semantic information, the number of convolutional kernel channels of the contraction path was gradually increased. The feature map size gradually decreased in the down-sampling phase by reason of the pooling layer. In the up-sampling stage, the pooling layer was changed to an up-sampling layer because of the expansion path. This helped to recover the resolution of the original image. In addition, the number of convolutional kernel channels was gradually reduced to achieve end-to-end segmentation. At the intersection of

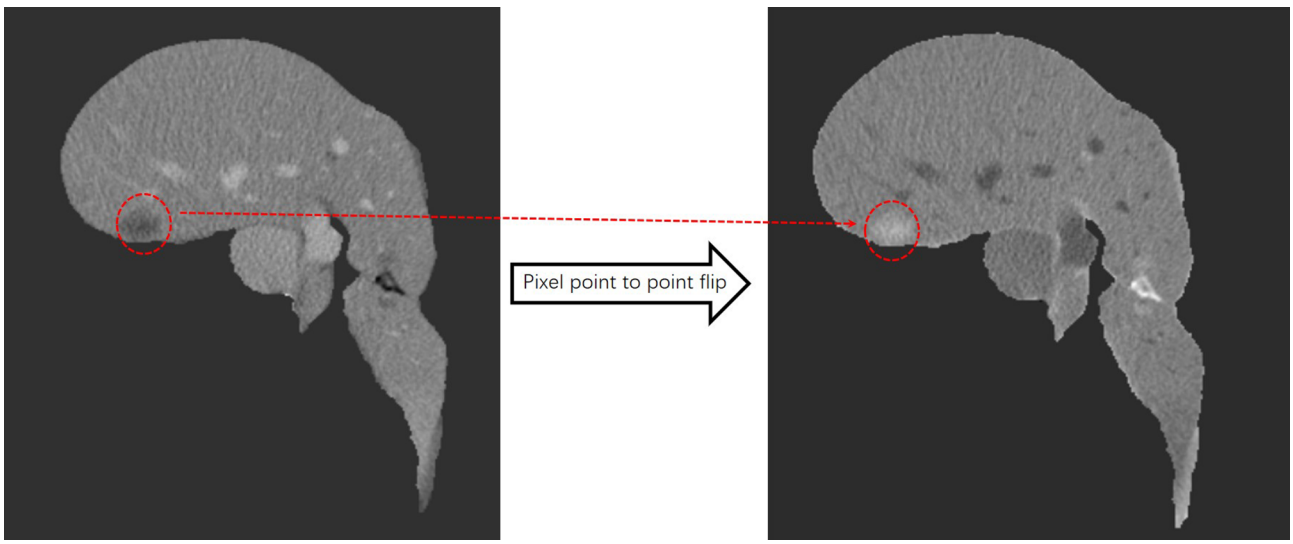


FIGURE 3 | After pixel point-to-point flipping, the liver turns gray, the tumor turns white, and the area outside the liver is black.

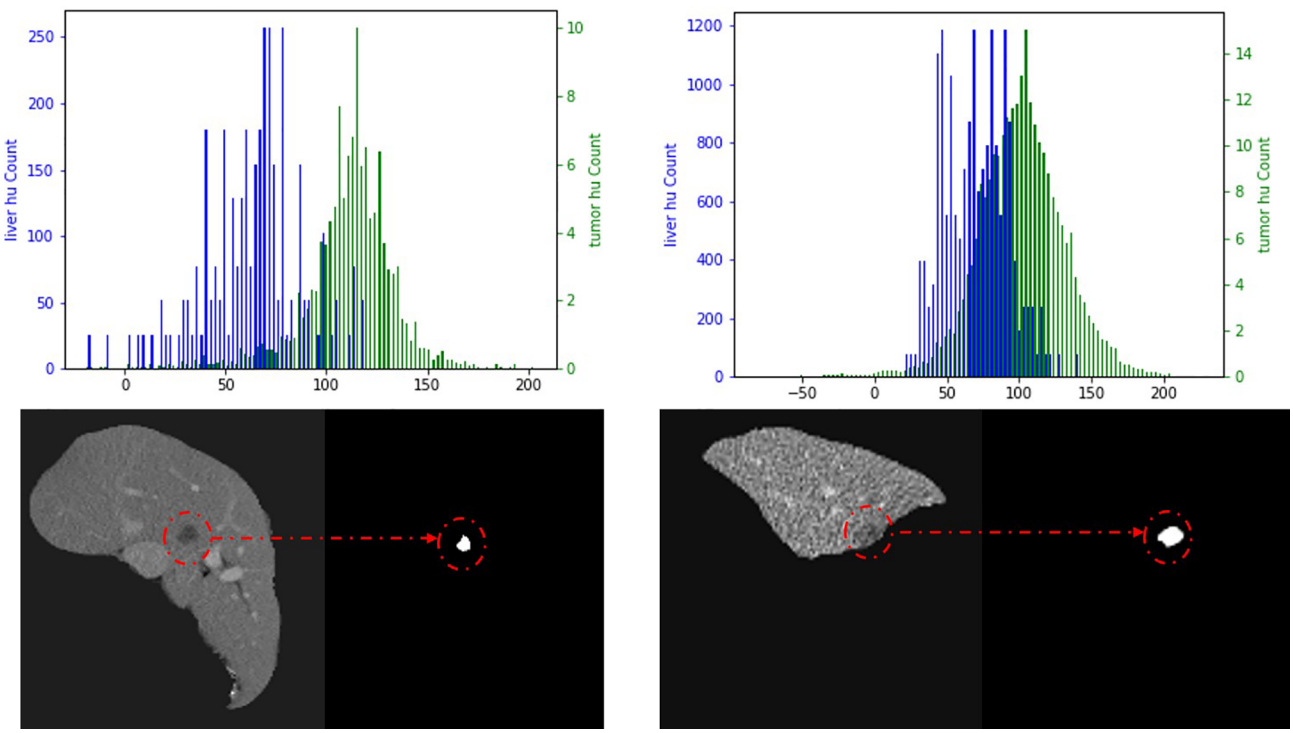


FIGURE 4 | The left panel is a high-contrast section and the right panel is a low-contrast section. Blue is the liver HU value and green is the tumor HU value.

contraction and expansion paths, we introduced a spatial attention mechanism and a channel attention mechanism to enable the network to encode longer-range semantic information in local features and to find correspondences between different channels. We introduced the long-hop connection between the layers of the contraction path and the expansion path so that the semantic information extracted in the contraction path was fused with the semantic information extracted in the expansion path.

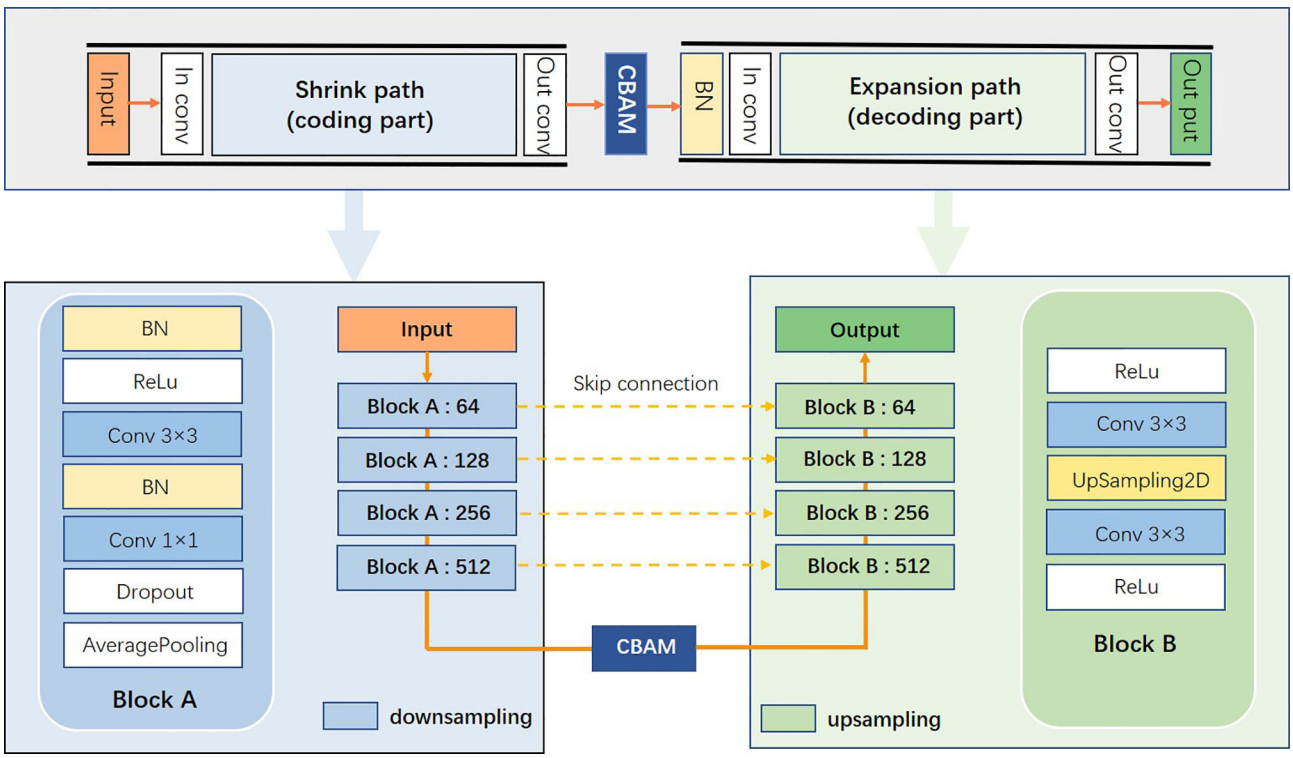


FIGURE 5 | S-Net network architecture diagram.

The Convolutional Attention Module is a simple and effective attention module for feed-forward convolutional neural networks. The overall architecture is shown in **Figure 6**. The attention module inferred attentional regions along two specific and mutually independent dimensions, multiplied the channel attention mechanism with the spatial attention mechanism, and adaptively optimized the local features. Because the attention mechanism architecture was small and lightweight, it could be seamlessly integrated into any network architecture and could be trained end-to-end along with neural networks.

We used the channel attention module illustrated in **Figure 7** to find the dependencies between different channels and to

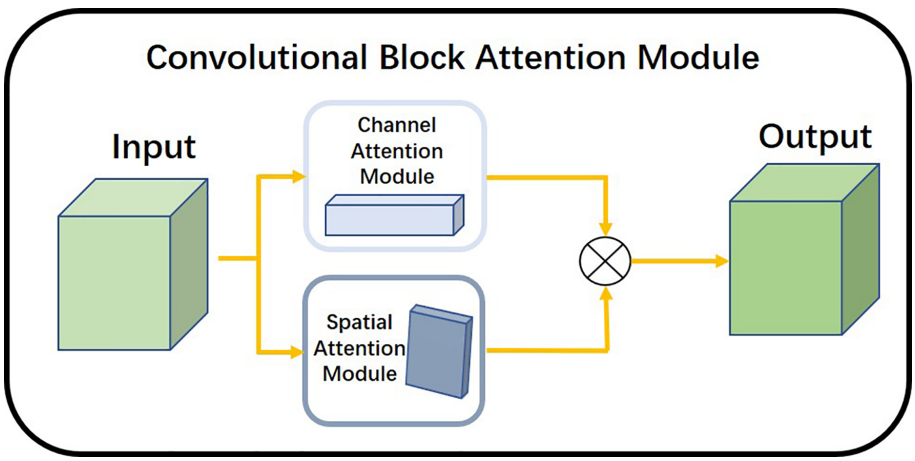


FIGURE 6 | Overall architecture of the attention mechanism.

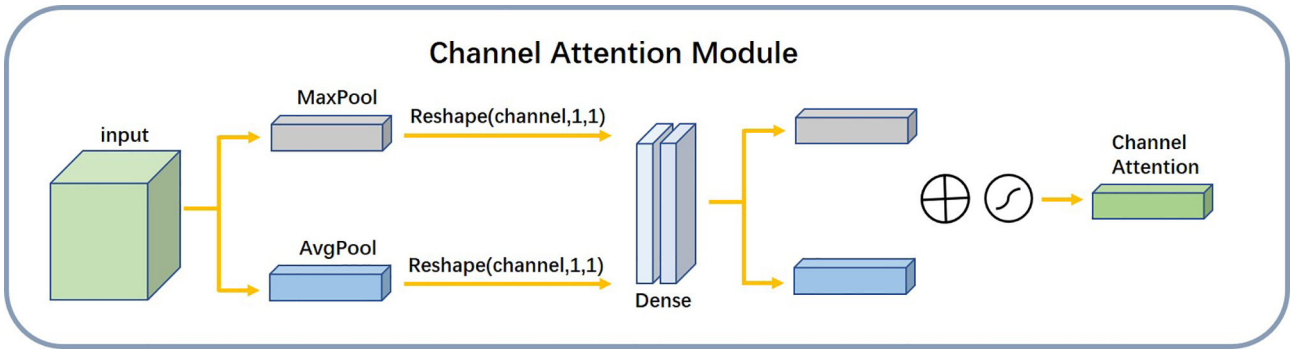


FIGURE 7 | Structure diagram of channel attention module.

enhance the dependent features. It focused mainly on the region of interest of the input image and compressed the spatial dimension of the input feature map. The module used the average pooling layer F^c_{avg} and the maximum pooling layer F^c_{max} to extract semantic information between channels. The shared network consisted of multiple layers of perceptrons. The workflow of the module is described as follows:

Maximum Pooling and Average Pooling Feature Maps: We used the maximum pooling layer to select the maximum value of the image region as the pooled value of the region. This eliminated nonextreme values and reduced the complexity of the upper-layer calculation. In addition, this layer could achieve translation invariance. The average pooling layer calculated the average value of the image region as the pooled value of the region. This could fade the combination of the relative positions between different features. Passing the pooled output through multiple layers of perceptrons: The multilayer perceptron played the role of a dimensional transformer. It converted high-dimensional information into low-dimensional information while preserving useful information. After a SoftMax function: The attentional mechanism was nonequivalent to the input of the overall sample. The channel attention module analyzed the weight of all input channels of a certain feature map and automatically selected the channels that need to be emphasized. The model set higher weights where necessary and smaller weights for those channels that were not emphasized. We used the SoftMax function to generate the probability of the importance of each channel.

The spatial attention module shown in **Figure 8** allowed the network to encode a longer range of semantic information in local features. Unlike the channel attention module, spatial attention focuses on the "where" as the most informative piece of information and complements channel attention. We applied the average pooling and maximum pooling operations along the channel axes and concatenated the operations to produce valid feature descriptors. With these descriptors, the channel information was merged to produce two feature maps: F^c_{avg}, F^S_{max}. The workflow of the spatial attention module is described as follows. maximum pooling and average pooling were done along the channel level to compress the image of N channels of

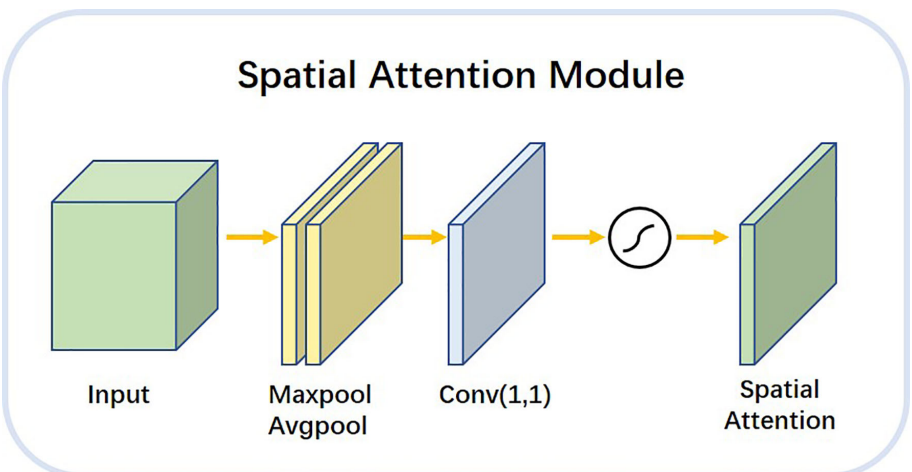


FIGURE 8 | Architecture diagram of spatial attention module.

HxW into a single channel of HxW. The 1x1 convolution layer was a linear combination of each pixel on different channels that retained the original planar structure of the feature map. It only changed the number of channels, thus achieving both up and down dimensional functions. The final attention map was normalized by the SoftMax function.

RESULTS

Evaluation Index and Experimental Result

To demonstrate the effectiveness of the algorithm, we conducted experiments on the LITS dataset, in which all image data was collected from academic and clinical institutions worldwide, including the data of 131 liver cancer patients. The number of layers per CT scan varied between 42 and 1026 for each sample, whereas the pixel size of each CT layer was 512 × 512 pixels. The number of liver tumors in each sample ranged from 0 to 75 and the size of tumors ranged from $38mm^3$ to 349 mm^3.

To evaluate the effectiveness of S-Net, researchers calculated the overlap measure according to the evaluation of LITS dataset, including Dice Global (DG) score, Dice per Case (DC) score, volumetric overlap error (VOE), average symmetric surface distance (ASSD), and root mean square error (RMSE). The Dice Global (DG) score is applied across all cases if they combine in a single volume, while the Dice per Case (DC) score refers to an average Dice score per volume. The mask labels provided by the LiTS dataset, 3DIRCADb dataset and doctors' manual contours of Hubei Cancer Hospital dataset are defined as the gold standard. The Dice score can be formulated as:

$$Dice(A, B) = \frac{2|A \cap B|}{|A| + |B|}$$

In this formula, A represents predicted results while B represents true annotations. The loss function, Loss, was calculated using the formula:

$$\text{Loss} = 1 - Dice(A, B)$$

The DC, DG, VOE, ASSD and RMSE for this study were found to be 0.613, 0.755, 0.413, 1.186, 1.703, respectively. These values of this paper's network architecture and three different network architectures (U-Net, DenseNet, and ResNet) were measured separately using the LiTS dataset as shown in **Table 1**. The models in the experiments all used 2D convolutional neural networks. When compared with the U-Net network, the DC value is improved by 0.099, the DG value was improved by 0.112, the VOE value is declined by 0.12, the RMSE is declined by 0.574. When compared with the three networks, these values were also found to show significant improvement. These values of different networks were also improved after applying the post-processing method.

Experimental Details and Parameter Settings

In this study, we adaptively partitioned all samples into 100 training samples, 10 validation samples, and tested our trained model from the LITS dataset on 3DIRCADb dataset and Hubei Cancer Hospital dataset. The 3DIRCADb datasets are composed of 20 CT scans, where 15 cases have hepatic tumors in the liver. The Hubei Cancer Hospital datasets included 20 enhanced CT scans of hepatic carcinoma with contrast from radiology department of Hubei Cancer Hospital. The auto-delineation results of this study used by Hubei Cancer Hospital dataset are shown in **Figure 11**. The training samples were pre-processed and the test samples were used to evaluate the segmentation effect of the network architecture.

We built the network architecture using Keras, with an NVIDIA Tesla P100 graphics card, and trained the network using a momentum gradient optimizer. We found initial learning rate to be 0.01, When the loss rate of the verification set did not decrease in three cycles, the learning rate was automatically reduced. A total of 200 cycles were trained. The set of weights with the highest Dice coefficients on the validation set was saved as the set of weights used for the testing phase. The activation function was a linear correction unit (RELU). The Dice and Loss values for training and validation are shown in **Figure 9**.

Auto-Delineation Results

Figure 10 shows the 2D visualizations of the auto-segmented contours for selected two CT scans from LiTS dataset. Green lines represent predicted results by S-Net model, while the red ones are gold standard. As can be seen from the figures, the auto-

TABLE 1 | Liver tumor segmentation results by S-Net, and U-Net, DenseNet and ResNet.

Model	Dice per case (DC)	Dice global (DG)	volumetric overlap error (VOE)	average symmetric surface distance (ASSD)	root mean square error (RMSE)
U-Net (10)	0.514	0.643	0.533	1.763	2.378
U-Net and post-processing	0.525	0.651	0.527	1.759	2.372
DenseNet (22)	0.504	0.612	0.557	1.862	2.765
DenseNet and post-processing	0.513	0.623	0.546	1.831	2.755
ResNet (23)	0.503	0.631	0.431	1.279	2.076
ResNet and post-processing	0.512	0.642	0.427	1.254	2.059
S-Net	0.613	0.755	0.413	1.186	1.804
S-Net and post-processing	**0.647**	**0.761**	**0.409**	**1.177**	**1.801**

Compared with U-Net, DenseNet and ResNet, the S-Net proposed in this paper has a higher DC value and DG value, lower VOE value, RVD value, ASSD value and RMSE value. The meaning of the bold values are best results.

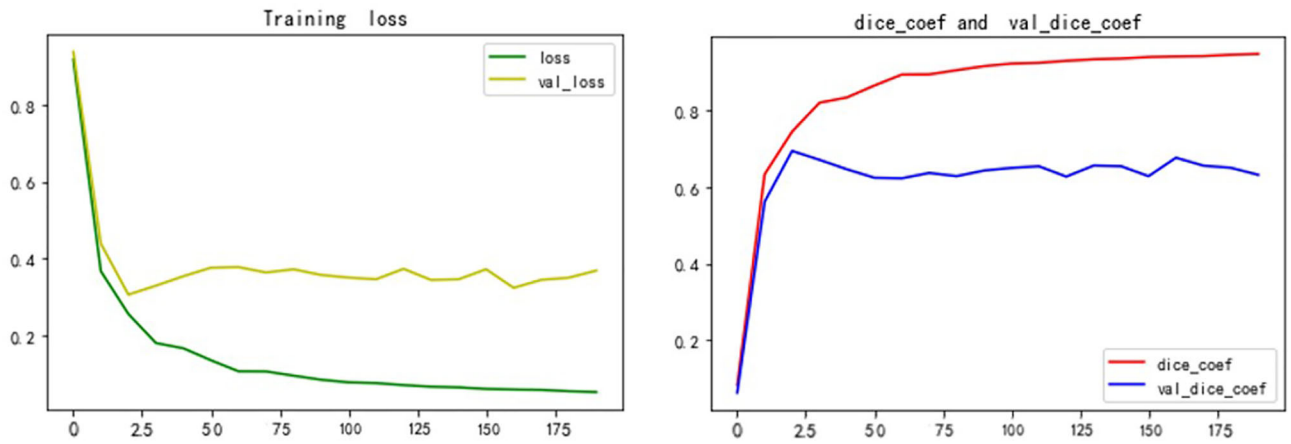


FIGURE 9 | Dice and Loss values for training and verification.

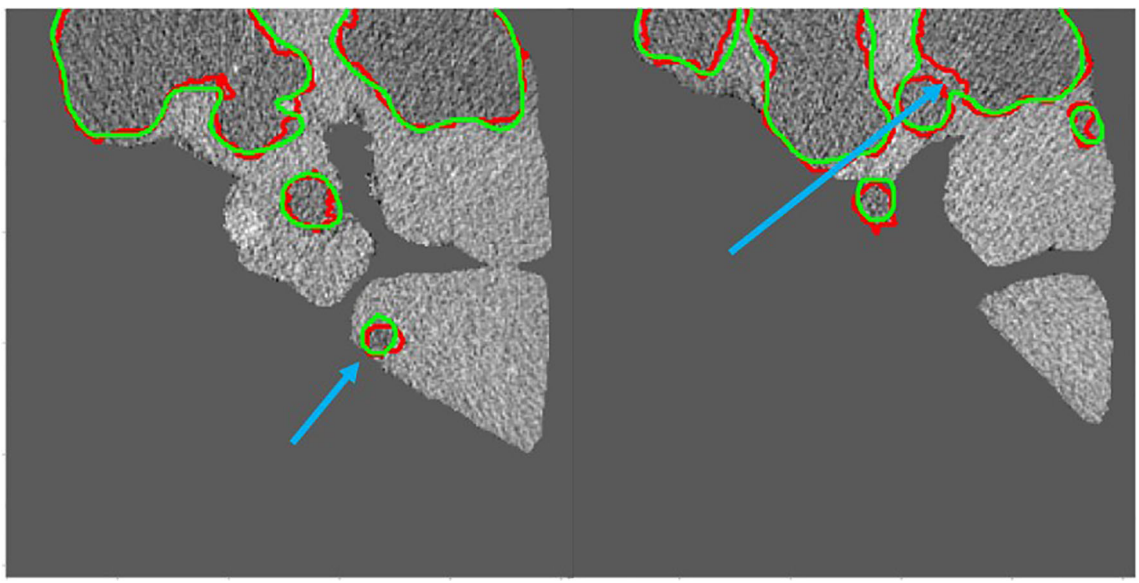


FIGURE 10 | Green contour shows the predicted results by S-Net models, while red shows the gold standard. The small and multiple lesions, marked by the blue arrows.

segmentations were close to the gold standard delineations, especially for the large liver lesions. The small and multiple lesions, marked by the blue arrows in figures, the overlapped regions slightly reduced. Therefore, the S-Net model could well segment the liver and liver tumors, but as for small and multiple tumors, it still needs more attention to enhance.

DISCUSSION

In this study, researchers proposed a novel convolutional neural network called S-Net to auto-segmentate liver tumors. The evaluation metric DC, DG, VOE, ASSD and RMSE were found to be 0.613, 0.755, 0.413, 1.186 and 1.804 respectively. The novel network S-Net was able to outperform other networks like U-Net, DenseNet, and ResNet. In addition, we proposed a pre-processing method of pixel point-to-point flipping, which improved the contrast of the HU values of CT sections, made the network learn useful information more easily. Unlike existing FCN network architectures, this architecture had the following two features:

1. Ability to add spatial attention mechanism and channel attention mechanism between encoder and decoder. The

spatial attention mechanism allowed the network to encode a longer range of semantic information in local features, whereas the channel attention mechanism found correspondences between different channels. In practice, the attention mechanism allowed the network to fully focus on the learning area, which greatly improved the accuracy of segmentation.

2. Long-hop connections increased the fusion rate of spatial feature information in the network, which could aid in the transfer of different spatial feature information from layer to layer.

On the basis of the 2017 LITS dataset, which tested the learning ability of the network, we concluded that the S-Net used in this study improved DC and DC values when compared with the U-Net network. DG and DC values were improved by 0.112 in and 0.117 in, respectively. When compared with cutting-edge algorithms like the cascaded FCN architecture proposed by Bellver et al. (24), these values improved by 0.015 and 0.023, respectively. When compared with the convolutional neural network incorporating spatial feature information proposed by Liu et al. (21), the DC value was improved by 0.021. The proposed network has a higher Dice value compared with Kaluva (25) and Pandey (26) et al., who added a residual structure to the conventional U-shaped structure to segment the tumor. The main reason for the improvement of segmentation accuracy is that this architecture adopted a canonical code-decode structure. It also integrated an attention mechanism based on the original U-Net network and used small, convoluted kernels to extract small amounts of semantic information. This architecture used a modular approach to gradually increase the number of convoluted kernel channels that extract deeper semantic information, thus improving the accuracy of segmentation. These results are shown in **Table 2**.

We mainly selected the results based on 2D model (14, 25–27, 29), except for some 3D model results by Li and Liu (11, 28). The proposed S-Net with the fusion of spatial features and attention mechanisms, outperformed than other 2D models. Indeed, the

TABLE 2 | Liver tumor segmentation results compared with other methods on the LiTS test dataset.

Model	Dice per case (DC)	Dice global (DG)	volumetric overlap error (VOE)	average symmetric surface distance (ASSD)	root mean square error (RMSE)
Chlebus et al. (14)	0.580	—	—	—	—
Song et al. (27)	0.569	0.751	0.437	1.702	—
Kaluva et al. (25)	0.492	0.625	**0.411**	1.441	
Pandey et al. (26)	0.587	—	—	—	—
Bi et al. (26)	0.500	—	—	—	—
T Liu al (28).	0.592	0.764	0.416	1.585	—
Li et al. (11)	**0.722**	**0.824**	0.497	**0.529**	**1.111**
Our S-Net	0.613	0.755	0.413	1.186	1.804

The DC, DG, Precision of other methods are obtained from the LITS leaderboard. The S-Net achieves much better performance in the precision score of liver tumors. The bold digitals denote the best results.

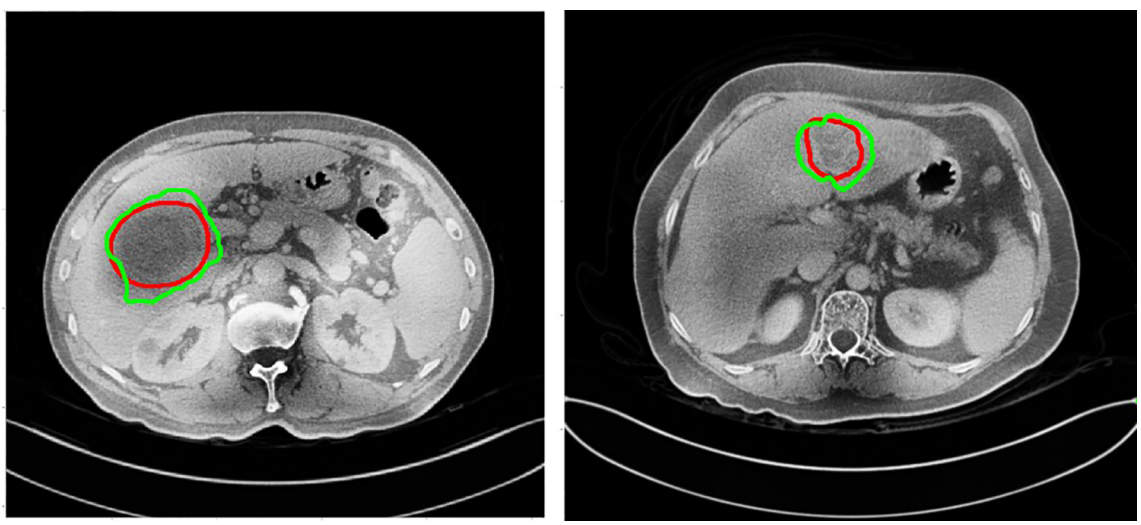


FIGURE 11 | The auto-delineation results of this study used by Hubei Cancer Hospital dataset. Green contour shows automatic results, while red shows the gold standard.

LiTS Leaderboard currently shows many higher DC and DG scores than those in the manuscript, especially the highest DC and DG by user 'liver_seg' were 0.7990 and 0.8500 respectively. Although some of the results are very high, there are still some reasons why we do not choose to repeat them. Firstly, it remains difficult to give recommendations about the exact network design, since the number and order of CNN layers and other hyperparameters were rough ideas instead of strict, proven guidelines. Secondly, the use of 3D architectures outperformed than the 2D ones, but in clinical practice they were not widely implemented due to memory constraints. Thirdly, more standard contour datasets will improve the segmentation accuracy, because state-of-art methods highly benefit from larger training datasets.

To demonstrate the applicability of our method in clinical practice, in this study we tested our trained model from the LiTS datasets on the new 3DIRCADb and Hubei Cancer Hospital datasets as shown in **Table 3**. The auto-delineation results of this study used by Hubei Cancer Hospital dataset are shown in **Figure 11**. And it achieved the slightly decreased results on tumor segmentation, with 0.578 on DC, 0.706 on DG, 0.576 on VOE and 1.673 on ASSD in 3DIRCADb dataset, with 0.527 on DC, 0.654 on DG, 0.594 on VOE and 1.862 on ASSD in Hubei Cancer Hospital dataset. The auto-segmentation results of 3DIRCADb dataset and Hubei Cancer Hospital dataset indicated the robustness of S-Net. Although the evaluation metric slightly decreased, it can effectively improve the effect of tumor recognition in CT images and could be applied to assist doctors in clinical treatment. If more datasets from different clinical centers are added for training, it is believed that the accuracy could be further increased.

A threshold of 0.2 was used to distinguish large tumors from small ones. We calculated tumor size by aggregating the tumor voxels in each real CT image. We did this to further understand the performance metrics of the network and to analyze the accuracy of the S-Net architecture in identifying the tumor size of different patients. The voxel values of the 21 CT tumors in the test set are shown in **Figure 12**. We found the tumor sizes in this data set to be widely variable. To facilitate experimental development, the data set was partitioned into a large tumor group and a small tumor group. This was determined by the orange line. We have used Baseline values to shows the effectiveness of our network in small and large tumors, Baseline is the 2D U-Net **Table 4**. The DC value and DG value of the two sample groups were tested separately. These results are shown in **Table 4**. From **Table 4**, It can be clearly observed that the large tumor achieves 0.0469 (Dice per case) accuracy improvements while the score for the small-tumor group is slightly advanced, with 0.0353 (Dice per case). From the comparison, we claim that the main reasons for the improve in Dice value is by adopting adaptive attention convolutional neural network, which can notice different dimension of semantic information. As a result, the accuracy of segment large tumor will be improved considerably. But semantic information of small tumors are more difficult to extract, so segmentation for small tumors have limited improvement. This is because many small lesions only occupying a few voxels, and it's difficult to distinguish the surrounding pixels in the lesion border. In addition, the difference in the HU value of liver and tumor may affect the segmentation accuracy.

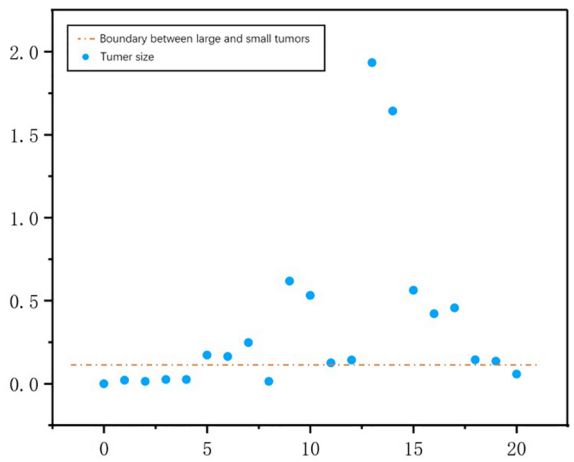


FIGURE 12 | The number of tumor voxels per patient in the test dataset. The orange line was used to distinguish between a large tumor group and a small tumor group.

TABLE 3 | Tested our trained model from the LITS dataset on 3D-IRCADb dataset and Hubei Cancer Hospital dataset.

Dataset	Dice per case (DC)	Dice global (DG)	volumetric overlap error (VOE)	average symmetric surface (ASSD)
3D-IRCADb	0.578	0.706	0.576	1.673
Hubei Cancer Hospital	0.527	0.654	0.594	1.862

TABLE 4 | The effectiveness of our network to small and large tumors segmentations.

	Small tumors		large tumors	
	Dice per case	Dice global	Dice per case	Dice global
Baseline	0.2613	0.2917	0.7279	0.7548
S-Net	0.3082	0.3246	0.7632	0.7819

As can be seen from the table. The segmentation accuracy of large tumors are better than that of small tumors. Baseline is the 2D U-Net with trained model.

To tackle with the small and multiple liver tumor segmentation problem, several methods were proposed. Li et al. used perceptual generative adversarial networks (GANs) to generate super-resolved representation for small object by revealing the intrinsic structural correlations between small and large objects (30). Kamnitsas et al. proposed a multi-scale 3D CNN with fully connected CRF for small brain lesion segmentation (31). Some post-processing methods utilize a custom criteria of removing lesions as noise if they have large variation between adjacent slices, because the size of lesions usually increase/decrease gradually with image slices up-and-down (32). A new loss function combined with Dice score and focal loss was better for segmenting small-volume structures such as optic nerves and chiasm (33). In summary, GANs, multi-scale representation, new loss function and custom post-processing methods may be the potential solution to overcome this challenging problem (34). This should be explored further in future studies.

CONCLUSION

In this study, an automatic CT image segmentation method based on S-Net network architecture used to automatically segment liver tumors from CT images was proposed. This study focused on the attention mechanism and the fusion of semantic information at different spatial dimensions. In this research, experiments based on the LITS dataset demonstrated that the methods discussed in this paper could improve the effect of automatic tumor segmentation in CT images. A drawback is that this algorithm was not effective in segmenting small tumors and multiple tumors. Future research will focus on the problem of case segmentation of small tumors and multiple tumors along with application of deep learning to clinical adjuvant therapy.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: Figshare. DOI: https://doi.org/10.6084/m9.figshare.14926695.v1

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

WW proposed the research idea. SL and XX performed the coding algorithms and manuscript drafting work. YD helped with the data analysis. WW and BZ gave useful discussions and provided the research platform. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the National Natural Science Foundation of China: 12075095 and 11704108, the Health Commission of Hubei Province scientific research project WJ2021M192.

ACKNOWLEDGMENTS

We thank LetPub (www.letpub.com) for its linguistic assistance during the preparation of this manuscript.

REFERENCES

1. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of Worldwide Burden of Cancer in 2008: GLOBOCAN 2008. *Int J Cancer* (2010) 127(12):2893–917. doi: 10.1002/ijc.25516
2. Lu R, Marziliano P, Thng CH. Liver Tumor Volume Estimation by Semi-Automatic Segmentation Method. In: *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference* (2006). p. 3296–9. doi: 10.1109/IEMBS.2005.1617181
3. Cootes TF, Taylor CJ, Cooper DH, Graham J. Active Shape Models-Their Training and Application. *Comput Vision Image Understanding* (1995) 61 (1):38–59. doi: 10.1006/cviu.1995.1004
4. Zhang X, Tian J, Deng K, Wu Y, Li X. Automatic Liver Segmentation Using a Statistical Shape Model With Optimal Surface Detection. *IEEE Trans Biomed Eng* (2010) 57(10):2622–6. doi: 10.1109/TBME.2010.2056369
5. Heimann T, Münzing S, Meinzer HP, Wolf I. A Shape-Guided Deformable Model With Evolutionary Algorithm Initialization for 3D Soft Tissue Segmentation. In: *Biennial International Conference on Information Processing in Medical Imaging*. Berlin, Heidelberg: Springer (2007). p. 1–12.
6. Tomoshige S, Oost E, Shimizu A, Watanabe H, Nawano S. A Conditional Statistical Shape Model With Integrated Error Estimation of the Conditions; Application to Liver Segmentation in non-Contrast CT Images. *Med Image Anal* (2014) 18(1):130–43. doi: 10.1016/j.media.2013.10.003
7. Wu W, Zhou Z, Wu S, Zhang Y. Automatic Liver Segmentation on Volumetric CT Images Using Supervoxel-Based Graph Cuts. *Comput Math Methods Med* (2016). doi: 10.1155/2016/9093721
8. Zheng S, Fang B, Li L, Gao M, Wang Y. A Variational Approach to Liver Segmentation Using Statistics From Multiple Sources. *Phys Med Biol* (2018) 63(2):025024. doi: 10.1088/1361-6560/aaa360
9. Conze PH, Noblet V, Rousseau F, Heitz F, de Blasi V, Memeo R, et al. Scale-Adaptive Supervoxel-Based Random Forests for Liver Tumor Segmentation in Dynamic Contrast-Enhanced CT Scans. *Int J Comput Assisted Radiol Surg* (2017) 12(2):223–33. doi: 10.1007/s11548-016-1493-1
10. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham: Springer (2015). p. 234–41.
11. Li X, Chen H, Qi X, Dou Q, Fu CW, Heng PA. H-DenseUNet: Hybrid Densely Connected UNet for Liver and Tumor Segmentation From CT Volumes. *IEEE Trans Med Imaging* (2018) 37(12):2663–74. doi: 10.1109/TMI.2018.2845918
12. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017). p. 2261–69. doi: 10.1109/CVPR.2017.243
13. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. Unet++: A Nested U-Net Architecture for Medical Image Segmentation. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham: Springer (2018). p. 3–11. doi: 10.1007/978-3-319-67558-9

14. Chlebus G, Schenk A, Moltz JH, van Ginneken B, Hahn HK, Meine H. Automatic Liver Tumor Segmentation in CT With Fully Convolutional Neural Networks and Object-Based Postprocessing. *Sci Rep* (2018) 8(1):1–7. doi: 10.1038/s41598-018-33860-7
15. Roth HR, Oda H, Zhou X, Shimizu N, Yang Y, Hayashi Y, et al. An Application of Cascaded 3D Fully Convolutional Networks for Medical Image Segmentation. *Computerized Med Imaging Graphics* (2018) 66:90–9. doi: 10.1016/j.compmedimag.2018.03.001
16. Schlemper J, Oktay O, Schaap M, Heinrich M, Kainz B, Glocker B, et al. Attention Gated Networks: Learning to Leverage Salient Regions in Medical Images. *Med image Anal* (2019) 53:197–207. doi: 10.1016/j.media.2019.01.012
17. Fu J, Liu J, Jiang J, Li Y, Bao Y, Lu H. Scene Segmentation With Dual Relation-Aware Attention Network. In: *IEEE Transactions on Neural Networks and Learning Systems*. (2020) 32(6):2547–60. doi: 10.1109/TNNLS.2020.3006524
18. Woo S, Park J, Lee JY, Kweon IS. Cbam: Convolutional Block Attention Module. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. (2018), 3–19. doi: 10.1007/978-3-030-01234-2_1
19. Lee H, Park J, Hwang JY. Channel Attention Module With Multiscale Grid Average Pooling for Breast Cancer Segmentation in an Ultrasound Image. *IEEE Trans Ultrasonics Ferroelectrics Frequency Control*. (2020) 67(7):1344–53. doi: 10.1109/TUFFC.2020.2972573
20. Wang C, Li B, Jiao B. Fault Diagnosis of Rolling Bearing Based on Convolutional Neural Network of Convolutional Block Attention Module. *J Phys: Conf Series IOP Publishing* (2021) 1732(1):012045. doi: 10.1088/1742-6596/1732/1/012045
21. Chen B, Zhang Z, Liu N, Tan Y, Liu X, Chen T. Spatiotemporal Convolutional Neural Network With Convolutional Block Attention Module for Micro-Expression Recognition. *Information* (2020) 11(8):380. doi: 10.3390/info11080380
22. Jégou S, Drozdzal M, Vazquez D, Romero A, Bengio Y. The One Hundred Layers Tiramisu: Fully Convolutional Densenets for Semantic Segmentation. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. (2017). p. 1175–83. doi: 10.1109/CVPRW.2017.156
23. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. (2016). p. 770–8. doi: 10.1109/CVPR.2016.90
24. Bellver M, Maninis KK, Pont-Tuset J, Giró-i-Nieto, X, Torres J, Van Gool L. Detection-Aided Liver Lesion Segmentation Using Deep Learning. *arXiv preprint arXiv* (2017) arXiv:1711.11069.
25. Kaluva KC, Khened M, Kori A, Krishnamurthi G. 2D-Densely Connected Convolution Neural Networks for Automatic Liver and Tumor Segmentation. *arXiv preprint arXiv* (2018) arXiv:1802.02182.
26. Pandey RK, Vasan A, Ramakrishnan AG. Segmentation of Liver Lesions With Reduced Complexity Deep Models. *arXiv preprint arXiv* (2018) 1805.09233.
27. Song LI, Geoffrey KF, Kaijian HE. Bottleneck Feature Supervised U-Net for Pixel-Wise Liver and Tumor Segmentation. *Expert Syst Appl* (2020) 145:113131. doi: 10.1016/j.eswa.2019.113131
28. Liu T, Liu J, Ma Y, He J, Han J, Ding X, et al. Spatial Feature Fusion Convolutional Network for Liver and Liver Tumor Segmentation From CT Images. *Med Phys* (2021) 48(1):264–72. doi: 10.1002/mp.14585
29. Bi L, Kim J, Kumar A, Feng D. Automatic Liver Lesion Detection Using Cascaded Deep Residual Networks. *arXiv preprint arXiv* (2017) arXiv:1704.02703.
30. Li J, Liang X, Wei Y, Xu T, Feng J, Yan S, et al. Perceptual Generative Adversarial Networks for Small Object Detection. In. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. (2017) p. 1222–30. doi: 10.1109/CVPR.2017.211
31. Kamnitsas K, Ledig C, Newcombe VF, Simpson JP, Kane AD, Menon DK, et al. Efficient Multi-Scale 3D CNN With Fully Connected CRF for Accurate Brain Lesion Segmentation. *Med Image Anal* (2017) 36:61–78. doi: 10.1016/j.media.2016.10.004
32. Bilic P, Christ PF, Vorontsov E, Chlebus G, Chen H, Dou Q, et al. The Liver Tumor Segmentation Benchmark (Lits). *arXiv preprint arXiv* (2019) arXiv:1901.04056.
33. Zhu W, Huang Y, Zeng L, Chen X, Liu Y, Qian Z, et al. AnatomyNet: Deep Learning for Fast and Fully Automated Whole-Volume Segmentation of Head and Neck Anatomy. *Med Phys* (2019) 46(2):576–89. doi: 10.1002/mp.13300
34. Du X, Li J, Niu G, Yuan JH, Xue KH, Xia M, et al. Lead Halide Perovskite for Efficient Optoacoustic Conversion and Application Toward High-Resolution Ultrasound Imaging. *Nat Commun* (2021) 12:3348. doi: 10.1038/s41467-021-23788-4

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.

frontiers in Oncology

ORIGINAL RESEARCH
published: 10 September 2021
doi: 10.3389/fonc.2021.725599

Consecutive Serial Non-Contrast CT Scan-Based Deep Learning Model Facilitates the Prediction of Tumor Invasiveness of Ground-Glass Nodules

*Yao Xu[1], Yu Li[2], Hongkun Yin[3], Wen Tang[3] and Guohua Fan[1]**

[1] Department of Radiology, Second Affiliated Hospital of Soochow University, Suzhou, China, [2] Department of Radiology, Dushuhu Public Hospital Affiliated of Soochow University, Suzhou, China, [3] Department of Advanced Research, Infervision Medical Technology Co. Ltd, Beijing, China

OPEN ACCESS

Edited by:
Min Tang,
Jiangsu University, China

Reviewed by:
Pritam Mukherjee,
Stanford University, United States
Xiuting Liu,
Washington University in St. Louis, United States
Zhixue Cheng,
Third Affiliated Hospital of Sun Yat-Sen University, China

***Correspondence:**
Guohua Fan
fangh22@126.com

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *15 June 2021*
Accepted: *19 August 2021*
Published: *10 September 2021*

Citation:
Xu Y, Li Y, Yin H, Tang W and Fan G (2021) Consecutive Serial Non-Contrast CT Scan-Based Deep Learning Model Facilitates the Prediction of Tumor Invasiveness of Ground-Glass Nodules. Front. Oncol. 11:725599. doi: 10.3389/fonc.2021.725599

Introduction: Tumors are continuously evolving biological systems which can be monitored by medical imaging. Previous studies only focus on single timepoint images, whether the performance could be further improved by using serial noncontrast CT imaging obtained during nodule follow-up management remains unclear. In this study, we evaluated DL model for predicting tumor invasiveness of GGNs through analyzing time series CT images

Methods: A total of 168 pathologically confirmed GGN cases (48 noninvasive lesions and 120 invasive lesions) were retrospectively collected and randomly assigned to the development dataset (n = 123) and independent testing dataset (n = 45). All patients underwent consecutive noncontrast CT examinations, and the baseline CT and 3-month follow-up CT images were collected. The gross region of interest (ROI) patches containing only tumor region and the full ROI patches including both tumor and peritumor regions were cropped from CT images. A baseline model was built on the image features and demographic features. Four DL models were proposed: two single-DL model using gross ROI (model 1) or full ROI patches (model 3) from baseline CT images, and two serial-DL models using gross ROI (model 2) or full ROI patches (model 4) from consecutive CT images (baseline scan and 3-month follow-up scan). In addition, a combined model integrating serial full ROI patches and clinical information was also constructed. The performance of these predictive models was assessed with respect to discrimination and clinical usefulness.

Results: The area under the curve (AUC) of the baseline model, models 1, 2, 3, and 4 were 0.562 [(95% confidence interval (C)], 0.406~0.710), 0.693 (95% CI, 0.538–0.822), 0.787 (95% CI, 0.639–0.895), 0.727 (95% CI, 0.573–0.849), and 0.811 (95% CI, 0.667–0.912) in the independent testing dataset, respectively. The results indicated that the peritumor region had potential to contribute to tumor invasiveness prediction, and the model performance was further improved by integrating imaging scans at multiple timepoints. Furthermore, the combined model showed best discrimination ability,

with AUC, sensitivity, specificity, and accuracy achieving 0.831 (95% CI, 0.690–0.926), 86.7%, 73.3%, and 82.2%, respectively.

Conclusion: The DL model integrating full ROIs from serial CT images shows improved predictive performance in differentiating noninvasive from invasive GGNs than the model using only baseline CT images, which could benefit the clinical management of GGNs.

Keywords: ground-glass nodules (GGNs), deep learning - artificial neural network (DL-ANN), computed tomography, follow-up, convolutional neural network

INTRODUCTION

Lung cancer is one of the most fatal cancers worldwide, and pulmonary nodules may represent early lung cancers. The National Lung Screening Trial and Dutch-Belgian Randomized Lung Cancer Screening Trial demonstrated that early lung cancer screening programs using low-dose computed tomography (CT) of the chest should be implemented globally (1). With the increasing use of CT screening for lung nodules—in particular, the extensive application of high-resolution CT (HRCT)—ground-glass nodules (GGNs) are observed increasingly often. As HRCT scans of the chest are gradually emerging as part of a routine physical examination, reasonable and necessary management of screening-detected and incidentally detected indeterminate pulmonary nodules is warranted.

Persistent lung GGNs may represent noninvasive or invasive adenocarcinoma. The prognosis of noninvasive lesions (atypical adenomatous hyperplasia (AAH) and adenocarcinoma *in situ* (AIS)) is quite different from that of invasive lesions (minimally invasive adenocarcinoma (MIA) and invasive adenocarcinoma (IAC)). Moreover, different pathological subtypes of GGNs vary with respect to the surgical approaches and clinical nodule management strategies required. In general, conservative nodule management is appropriate for noninvasive GGNs, whereas invasive GGNs are suitable for surgical resection. The overall survival rate after surgery for noninvasive GGN patients can be close to 100%, and there is a promising 5-year survival rate of 80%–95% in the case of invasive GGNs. However, they are characterized by very slow growth, so regular follow-up management and a wait-and-see policy are advocated by many experts (2). The Fleischner Society Guidelines for management of incidental subsolid nodules was published in 2017 and recommended follow-up intervals ranging from 3 months to several years (3). Guidelines for GGN management mainly include qualitative characteristics and patient history (e.g., smoking history and cancer history). Follow-up has a crucial role in clinical decision-making and assessment of surgical indication, and is increasingly recommended by thoracic and pulmonary guidance.

Most previous CT-based quantitative studies have used single screening images to estimate the invasiveness of GGNs on the basis of size, density, and mass volume (4, 5). However, there are known limitations to this approach owing to inter- and intraobserver variability in morphological features of GGNs. Nevertheless, in short-term follow-up, it is difficult to evaluate the invasion characteristics of GGNs based on morphological characteristics such as volume-doubling time and mass volume (6). In recent years, artificial intelligence has achieved great progress in the automatic quantitative image characterization of medical images; in particular, deep learning (DL) algorithms have proved to be versatile and efficient (1, 7). The clinical application of AI is currently extensive. The image-based AI can be used to distinguish the tissues of a COVID-19 patient and assessed the severity of their infection. Additionally, DL model can exhibit superior performance to that of general physicians and general orthopedic surgeons on shoulder radiographs in fracture datasets (8, 9). Owing to its favorable performance, DL has been widely used for the early detection, molecular subtype diagnosis, and prognosis prediction of GGNs. Several previous studies have reported the application of DL in predicting tumor invasiveness of GGNs. However, those studies only focused on the development of CT imaging biomarkers from a single timepoint, none of them had used serial CT images including the follow-up scans (10, 11). Since the evolution of tumor invasiveness is a dynamic biological progression with stem cell and vascular contributions, CT scan at a single time point might not capture the tumor phenotype completely (12–15). Incorporation of serial CT images from routine follow-up exanimations could be beneficial to track the phenotypic changes of GGNs and achieve more accurate diagnosis. There are few reports on using medical images from multiple time points for diagnosis (13, 16, 17). To the best of our knowledge, only two studies have investigated the use of serial CT images with DL algorithms for diagnosis of nodule malignancy or prognostic prediction in lung cancer patients (13, 18). Therefore, it is unclear whether the performance of DL models for predicting tumor invasiveness of GGNs could be further improved by using serial CT imaging.

In the current study, we aimed to analyze the characteristics of lesions based on combined baseline scans and follow-up scan images *via* artificial intelligence. We used DL methods—specifically, convolutional neural networks (CNNs) and recurrent neural networks (RNNs)—to predict early-stage lung adenocarcinoma presenting as GGNs by incorporating baseline and 3-month follow-up CT images. Our results have potential implications for the use of DL-based analysis of routine follow-up CT scans in patients with GGNs, as DL can be applied to predict tumor invasiveness noninvasively and is beneficial in precision medicine as well as clinical therapy.

METHODS

Study Population

This study was approved by the institutional ethics committee of our hospital, and the informed consent requirement was waived. Data for 1724 patients who underwent CT examinations and were diagnosed as having GGNs between December 2015 and January 2020 were retrospectively retrieved from the picture archiving and communication system (PACS). The exclusion criteria were as follows: (1) the patient did not undergo biopsy or surgery in our hospital, and tumor invasiveness status was not available (n = 1089); (2) lack of consecutive series of CT scans (n = 297); (3) incomplete reconstructed thin-slice images (≤ 1.5 mm) or low image quality (n = 30); (4) patient had received any previous treatment before CT scan (n = 128); and (5) pleural or mediastinal adhesion was present and GGNs were difficult to label on CT images (n = 12). Finally, a total of 168 patients with 168 GGNs [13 atypical adenomatous hyperplasia (AAH), 107 minimally invasive adenocarcinoma (MIA), 35 adenocarcinoma *in situ* (AIS), and 13 invasive adenocarcinoma (IAC)] were enrolled, and two consecutive CT scans within about 3 months (82–109 days, median 93 days) for each patient were used in this study. The invasiveness of GGNs was later confirmed through pathological analysis. AAH and AIS were classified as noninvasive lesions, and MIA and IAC were classified as invasive lesions. The study workflow is depicted in **Figure 1**.

CT Image Acquisition

All patients underwent CT scanning at our hospital with the 750 HD CT (Discovery, GE Healthcare, North Richland Hills, TX, USA) scanner or the 256 multidetector row scanner (Brilliance iCT, Philips Healthcare, Cleveland, OH, USA). Scan parameters were as follows: 0.625-mm slice thickness and 1.25-mm slice spacing; 120 kV voltage; automatic tube-current modulation with a mean tube current of 100 mA; 5 mm collimator, and a 512 × 512 matrix. All the thin-slice CT images were reviewed by a thoracic radiologist with more than 10 years of experience in chest CT for image qualitative evaluation.

Annotation and Pretreatment of Tumor Regions

The GGNs in CT images were manually labeled by a radiologist with 5 years of experiences in chest imaging using the ITK-SNAP software (version 3.8.0, http://www.itksnap.org). The GGNs were

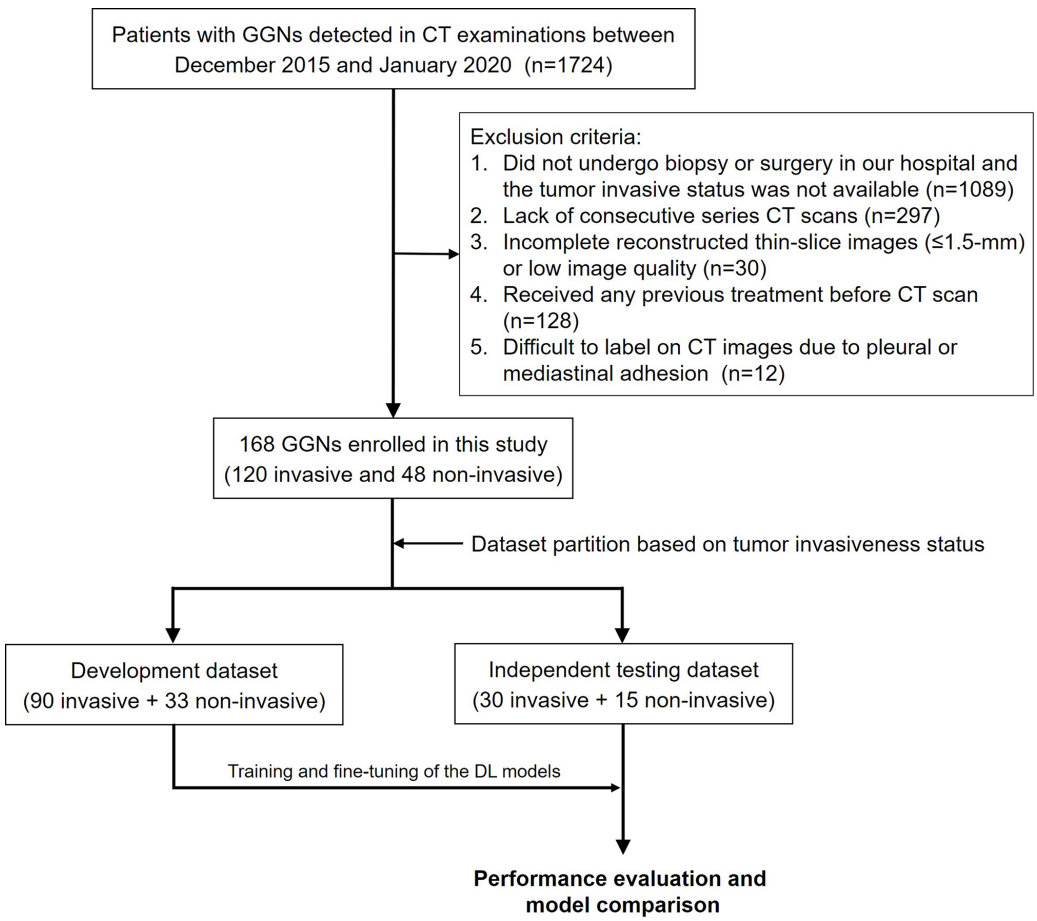


FIGURE 1 | Patient enrollment and study design.

segmented at the lung window set (Window Width 1500 Hu, Window Level -450 Hu) by carefully drawing a region of interest (ROI) that traced the edge of the GGN on all axial images until the entire GGN was covered. The annotation results were further checked by another senior radiologist with 10 years of experience. When the boundary of the GGN was uncertain, an expert radiologist with more than 20 years of experience in lung cancer diagnosis was consulted for the final decision. All radiologists were blinded to the pathological results.

Before the development of the DL models, a resampling approach was used for data pretreatment. The CT images were rescaled to a uniform size with 1-mm isotropic voxel spacing, then each manually labeled ROI was transformed and defined as follows. (i) A 64 × 64 × 64-pixel three-dimensional (3D) patch containing the nodule region which was cropped from each CT image, and the size was determined based on the largest ROI. (ii) The pathologically identified label of tumor invasiveness. As previous studies have indicated that the peritumoral region provides valuable insight for determining the prognosis of lung cancer (19, 20), both the gross ROI patch and the full ROI patch containing perinodular regions were automatically generated, as the nonlesion area of the 3D patch was left padding zero or reserving perinodular imaging (**Figure 2**).

Development of the Baseline Model

A baseline model was constructed by using the image and demographic features including tumor size, location, age, gender, cancer history, and smoking history. The logistic regression (LR) analysis was used as the classifier. The baseline model was built in the development dataset and validated in the independent testing dataset.

Construction of the DL Models

Fifteen noninvasive lesions and 30 invasive lesions were firstly randomly selected to serve as the independent testing dataset, and the remaining samples (33 noninvasive and 90 invasive) were used as the development dataset. Owing to the limited amount of training data, data augmentation techniques including flipping (perpendicular to the *x*- and *y*-axis), random shifting (15% towards the eight vertexes of the 3D patch), random rotation (90°, 180°, and 270° perpendicular to the *z*-axis), mirroring, and random brightness contrast (80%, 90%, 110%, and 120%) were used in the development of the neural networks. After data augmentation, the sample size increased to 19 times that of the original, yielding a total of 2,337 samples in the development dataset.

We employed a modified 3D ResNeXt34 as the backbone network of the DL models, as the 3 × 3 2D convolution filters were replaced by 3 × 3 × 3 3D convolution filters (21). Transfer learning approach was applied to improve the robustness and generalization of the DL models (22). To pretrain the modified 3D ResNeXt34 network in this study, a total of 178 pulmonary nodules were manually labeled on the TCGA-LUAD, CPTAC-LUAD, TCGA-LUSC, and CPTAC-LSCC datasets which were downloaded from The Cancer Imaging Archive (TCIA) database. For the prediction of tumor invasiveness based on single or serial CT images, two kinds of DL models were designed: the single-DL model using only the baseline CT images as inputs and the serial-DL model using two consecutive series (baseline and follow-up) CT images as inputs. For the single-DL model, the 3D ResNeXt34-based CNN with a fully connected layer was used to extract high-dimensional features from the imaging data, followed by a softmax output layer to predict the probability of tumor invasiveness. The neural architecture for the serial-DL model included ResNeXt34-based CNN merged with a RNN (**Figure 3A**). In brief, two weight-sharing 3D ResNeXt34 networks were used for feature extraction from two consecutive series CT images, and the outputs of each CNN model were fed into the RNN with a long short-term

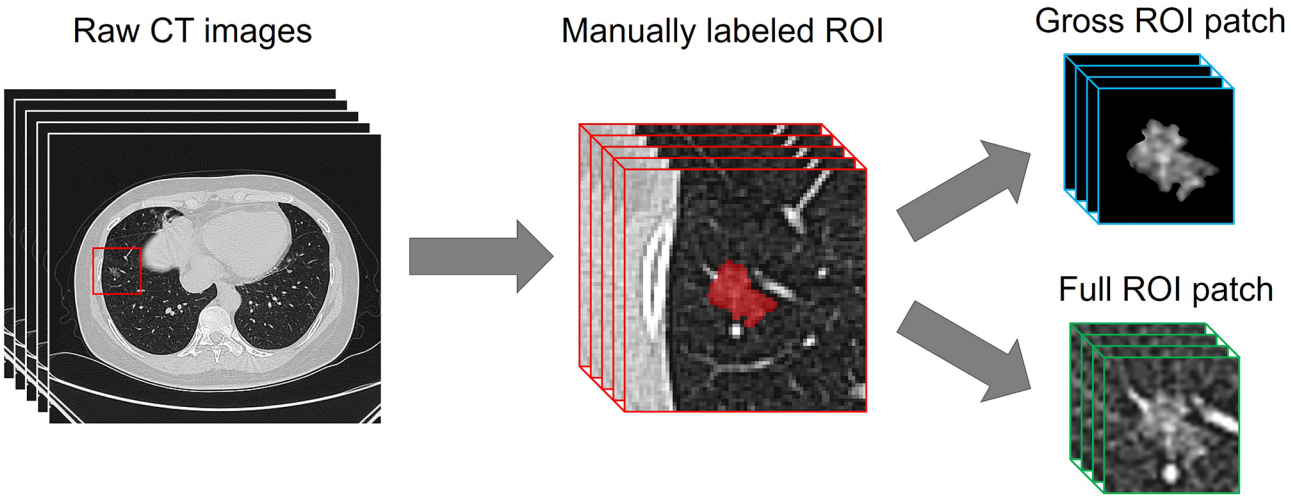


FIGURE 2 | Examples of the automatedly generated gross ROI patch and full ROI patch.

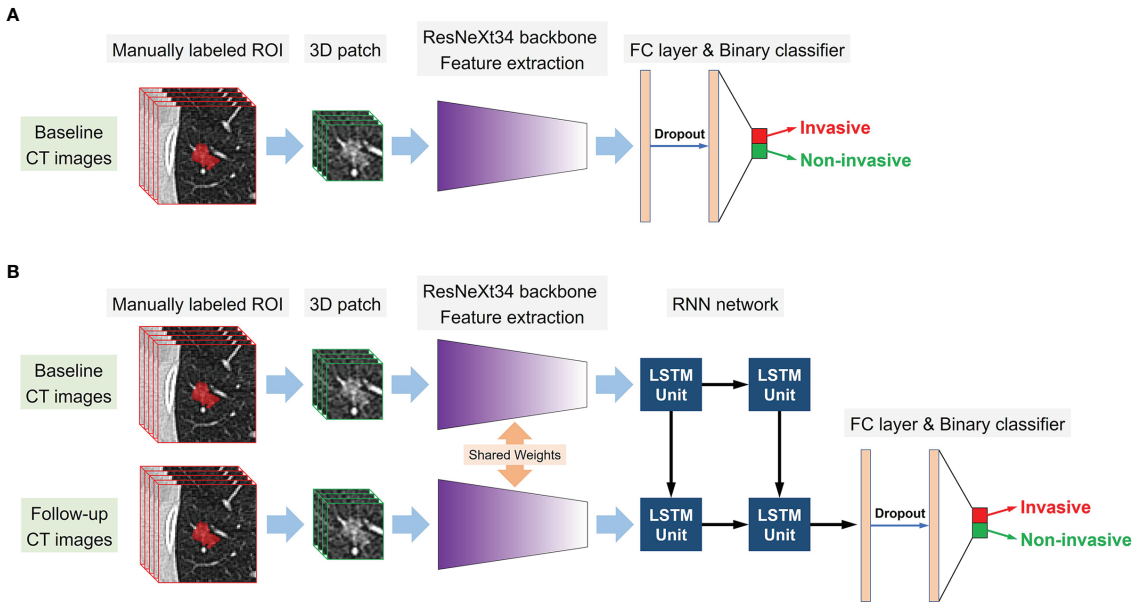


FIGURE 3 | Conceptual architecture of the single-DL model using only baseline CT images **(A)** and the serial-DL model integrating serial CT images at multiple timepoints **(B)**.

memory (LSTM) architecture as time-varying inputs (**Figure 3B**). Therefore, four DL models were designed in our study: a single-DL model using gross ROI patches as inputs (model 1), a serial-DL model using gross ROI patches as inputs (model 2), a single-DL model using full ROI patches as inputs (model 3), and a serial-DL model using full ROI patches as inputs (model 4).

The proposed single-DL models were trained based on the binary cross-entropy loss function; the weights of hidden layers were randomly initialized by Xavier, and the initial learning rate was set to 0.0001. Adam was used as the optimizer in the training stage owing to its fast convergence and weight-dependent learning rate, and the beta 1 and beta 2 parameters were 0.9 and 0.99, respectively. In addition, a weighted oversampling technique was used to train the model; only resampled minibatches with a noninvasive/invasive ratio of 1:1 was selected for training. The minibatch size was 24, and the dropout rate was set to 0.5; other parameters were set to their default values. As for the serial-DL models, there were three cells in the LSTM unit with the dropout rate set to 0.8, and each cell contained 512 features. The training was stopped when the loss of function was stable (23). Since deep learning models based on small sample size could be subject to obvious overfitting after a certain number of epochs, the early stopping method was used to halt parameter iteration for the model. The number of modeling epochs was set between 60 and 100 in this study. The changes in model efficiency (AUC) and cross-entropy loss function index corresponding to each epoch in the training process of the DL models were shown in **Figure 4**.

The code of these DL models was open sourced at https://github.com/TangWen920812/3d-resnext-lstm.

Supervised training was performed on a computer with a Core i7-6700 K central processing unit (Intel, Santa Clara, CA, USA), 32 GB memory, and a GeForce GTX 1080 graphics processing unit (NVIDIA, Santa Clara, CA, USA). Python 3.6.8 (https://www.python.org) and the Mxnet 1.5.0 (https://mxnet.incubator.apache.org) framework for neural networks were used to construct the DL models. The development and independent of the DL models were performed with InferScholar platform version 3.3 (InferVision, China).

Development of the Combined Model

To integrate both serial CT images and clinical information, a combined model was constructed by incorporating the following candidate variables: age, gender, GGN size, GGN location, cancer history, smoking history, and the invasiveness probability calculated by the serial-DL model. The combined model was developed in the development dataset by using linear support vector machine (linear SVM) classifier, and the prediction value was calculated using following formula:

$$\begin{aligned}\text{Prediction value} = {} & -0.4061 * \text{gender (male} = 1, \text{female} = -1) \\ & -0.0316 * \text{age} + 0.5521 * \text{GGN size} + 0.6422 * \text{RUL location (yes} = 1, \textit{no} = 0) \\ & - 0.5790 * \text{RML location (yes} = 1, \text{no} = 0) - 1.9971 * \text{RLL location (yes} = 1, \text{no} = 0) \\ & +1.1625 * \text{LUL location (yes} = 1, \text{no} = 0) + 0.7715 * \text{LLL location (yes} = 1, \text{no} = 0) \\ & +1.0314 * \text{cancer history (yes} = 1, \text{no} = 0) + 0.4313 * \text{smoking history (yes} = 1, \text{n} = 0) \\ & +39.818 * \text{invasiveness probability} - 14.6215 \text{ (LUL left upper lobe; LLL, left lower lobe;} \\ & \text{RUL, right upper lobe; RML, right middle lobe ; RLL, right lower lobe)}\end{aligned}$$

In addition, the decision curve analysis (DCA) was applied to assess the clinical usefulness of the combined model as well as the deep learning models on the independent testing dataset.

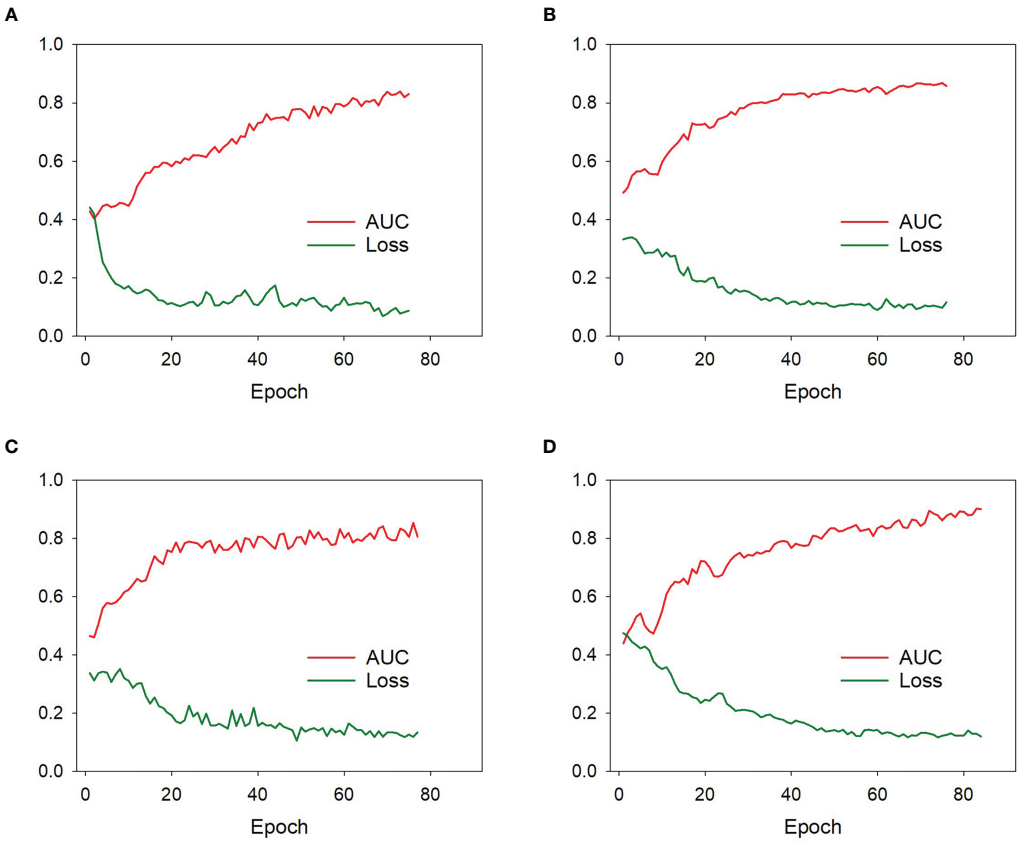


FIGURE 4 | The model efficiency (AUC) and cross-entropy loss function corresponding to each epoch of the model 1 **(A)**, model 2 **(B)**, model 3 **(C)**, and model 4 **(D)** during training process. The model efficiency corresponding to each epoch gradually increased while the model loss function decreased and eventually stabilized.

Statistical Analysis

In order to evaluate the performance of the DL models for the discrimination of noninvasive from invasive lesions, a receiver operating characteristic (ROC) curve was plotted for the calculation of sensitivity and specificity, and the area under the curve (AUC) was determined. The sensitivity, specificity, and accuracy were calculated under optimal threshold according to the maximum Youden index (24). Delong's test was used to compare the differences between two or more AUCs of different models. The association between categorical variables was assessed by Chi-square test or Fisher's exact test, and the Mann–Whitney *U* test was performed to evaluate the differences among variables with a continuous distribution. The DCA curve was plotted using the "rmda" package. All analyses were performed using Prism 5 for Windows (version 5.01), and a two-sided *p*-value <0.05 was considered statistically significant.

RESULTS

Patient Characteristics

The clinicopathologic characteristics of the enrolled patients in the independent and datasets are summarized in **Table 1**. There were no significant differences in gender, age, cancer history, or smoking history between patients with noninvasive nodules and those with invasive nodules in either the development or the independent testing dataset (all $p > 0.05$). The prevalence of GGN showed a greater tendency to occur in the upper lobe in the noninvasive group compared with the invasive group (70.8% [34/48] in the noninvasive group *vs.* 47.5% [57/120] in the invasive group, $p = 0.03$). The average GGN in the invasive group was larger than that of the noninvasive group across all patients ($p = 0.02$); however, the differences were not significant in the development dataset ($p = 0.08$) or the independent testing dataset ($p = 0.10$).

Performance of Different Predictive Models in the Independent Testing Dataset

In the independent testing dataset, the baseline model showed limited discrimination capability with the AUC, accuracy, sensitivity, and specificity achieving 0.562 (95% CI, 0.406–0.710), 64.4%, 67.7%, and 60.0%, respectively. The result indicated that an effective diagnosis of tumor invasion in GGNs was not possible when only using the clinical variables.

TABLE 1 | The clinicopathologic characteristics of enrolled patients.

	All patients			Development dataset			Independent dataset		
	Noninvasive	Invasive	*p*-Value	Noninvasive	Invasive	*p*-Value	Noninvasive	Invasive	*p*-Value
Gender			0.67			0.19			0.27
Male	14	39		7	30		7	9	
Female	34	81		26	60		8	21	
Age (years)			0.16			0.11			0.83
Mean	46.8	49.8		45.7	49.7		49.1	50.0	
SD	10.4	13.0		11.3	12.4		8.0	14.8	
GGN size (mm)			0.02			0.08			0.10
Mean	7.6	9.1		7.8	9.0		7.3	9.3	
SD	2.3	3.9		2.4	3.8		1.9	4.3	
GGN location			0.06			0.05			0.53
LUL	18	30		11	23		7	7	
LLL	3	19		2	13		1	6	
RUL	16	27		13	20		3	7	
RML	2	17		0	13		2	4	
RLL	9	27		7	21		2	6	
Cancer history			0.44			0.52			0.71
Yes	4	15		3	12		1	3	
No	44	105		30	78		14	27	
Smoking history			0.69			0.80			0.36
Yes	13	29		7	21		6	8	
No	35	91		26	69		9	22	

GGN, ground-glass nodules; LUL, left upper lobe; LLL, left lower lobe; RUL, right upper lobe; RML, right middle lobe; RLL, right lower lobe.

The accuracy of DL models was 66.6% (model 1), 71.1% (model 2), 75.6% (model 3), and 84.4% (model 4) in the independent testing dataset. The AUCs of models 1 and 2 (using gross ROI patches as inputs) were 0.693 (95% CI, 0.538–0.822) and 0.787 (95% CI, 0.639–0.895), respectively (**Figures 5A, B**), whereas that of models 3 and 4 (using full ROI patches as inputs) were 0.727 (95% CI, 0.573–0.849) and 0.811 (95% CI, 0.667–0.912), respectively (**Figures 5C, D**). The serial-DL models and the combined model outperformed the baseline model with significant differences (Delong's test, p = 0.046, 0.022, and 0.024 for model 2, model 4, and combined model, respectively). The full ROI patch-based DL models showed an increased performance tendency than the gross ROI patch-based models; however, the differences were not statistically significant (model 3 *vs.* model 1, p = 0.753, model 4 *vs.* model 2, p = 0.796). In addition, the AUC of serial CT image-based DL models was also higher than that of single CT image-based DL models (model 2 *vs.* model 1, p = 0.187, model 4 *vs.* model 3, p = 0.383). The accuracy of the serial-DL model using full ROI was significantly higher than that of the single-DL model using gross ROI (model 4 *vs.* model 1, 84.4% *vs.* 66.6%, Chi-square test p = 0.049), indicating that the spatial pattern of perinodular regions and incorporation of serial CT images could facilitate the prediction of tumor invasiveness in patients with GGNs.

Evaluation of the Combined Model

The combined model showed the best performance with AUC, sensitivity, specificity, and accuracy achieving 0.831 (95% CI, 0.690–0.926), 86.7%, 73.3%, and 82.2%, in the independent testing dataset, respectively (**Figure 6A**). The details of the predictive performance of the baseline model, the DL models, and the combined model are summarized in **Table 2**. DCA was used to evaluate the clinical usefulness of the different predictive models. The results showed that the combined model had a slightly higher overall net benefit compared with the DL models across the majority of probability threshold (**Figure 6B**).

DISCUSSION

In this study, we developed novel DL models to detect the invasiveness of GGNs based on consecutive CT thin-scanned images (baseline and 3-month follow-up scans). We found that the peritumoral region could contribute to invasiveness prediction. Notably, integrating consecutive serial CT images further improved the performance of DL models for predicting tumor invasiveness of GGNs. In addition, the combination of clinical variables and risk probability calculated by DL model showed favorable capability in distinguishing noninvasive GGNs from invasive GGNs.

Previous studies have reported that the size of GGNs is a critical risk factor for potential invasiveness (3–5, 21). Lee et al. found that the optimal cutoff size for preinvasive lesions was less than 10 mm (sensitivity, 53.33%; specificity, 100%) in a pure GGN cohort; this could be used as a selection criterion to identify patients suitable for sublobar resection (25). In addition, Kim showed that 8 mm was the optimal cutoff value for discrimination of noninvasive GGNs from invasive GGNs (26). In short, these results combined with those of previous studies indicate that the clinical feature of size is indeed highly correlated with the invasiveness of GGNs (27). Notably, we also observed significant differences in size between the preinvasive cohort and the invasive cohort in the all-patient dataset ($p < 0.05$). However, there were no such significant differences in either the

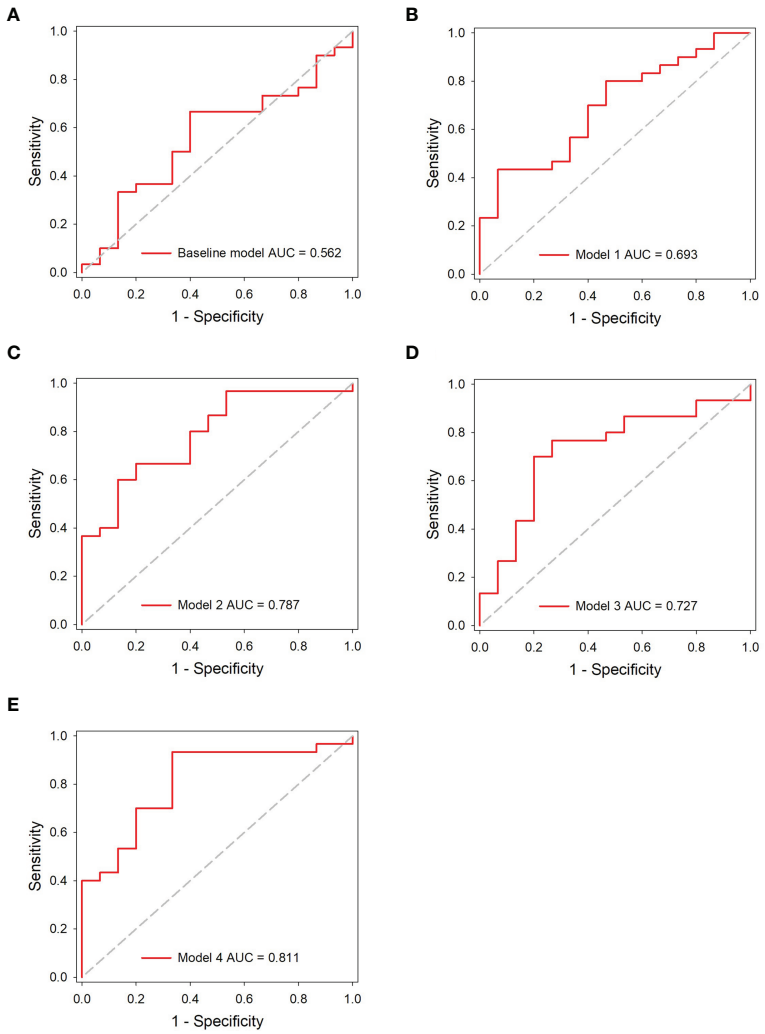


FIGURE 5 | ROC analysis of the predictive models in the independent testing dataset. **(A)** The baseline model. **(B–E)** The DL models.

independent testing dataset or the development dataset, although the average GGN size of the invasive group was larger than that of the noninvasive group in all patients.

The largest lung cancer screening trial in Europe showed that malignant tumors were localized predominantly in the periphery of lungs and the right upper lobe (27). Interestingly, in our study, the GGNs in the noninvasive group showed a greater tendency to occur in the upper lobe compared with the invasive group (noninvasive group 34/48 *vs.* invasive group 57/120). This bias is probably introduced in our study because of the difference between tumor malignancy and tumor invasiveness.

However, previous studies had certain limitations. First, most studies only considered whether clinical characteristics such as smoking *vs.* nonsmoking or tumor history (yes *vs.* no) were related to GGN growth. Few studies have evaluated the weights of CT images and clinical information for predicting GGN invasiveness with specific numerical formulas, which could be better verified and recognized by radiologists. Second, the most of reported DL algorithms were applied to lung nodule classification as benign or malignant, and they focus on a single scan for the model input.

Most of the previous studies only considered the relation between the quantitative radiographic characteristic and pathologic classification that are limited at the single time-point (19, 21, 28). In our study, we applied different DL algorithms to predict the invasiveness of GGNs. The single-DL models and serial-DL models that are based on whether they used single or serial CT images were proposed, and their performance was compared. This is a different approach from the current method of predicting malignant nodules based on a single CT scan (26, 28). We combined baseline and 3-month follow-up continuous CT scan images to obtain the original features and changed features that maximized the risk of tumor invasion. Recently, numerous studies and guidelines have advocated that early follow-up of patients with GGN should replace unnecessary surgical resection (3, 5, 19). However, subtle

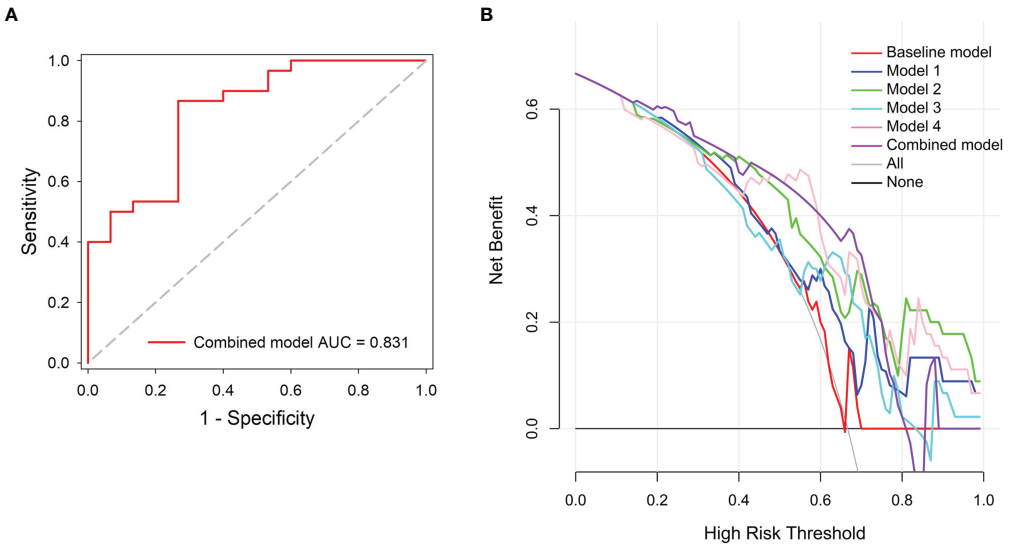


FIGURE 6 | Performance evaluation of the combined model. **(A)** ROC analysis. **(B)** Decision curve analysis for the predictive models; the combined model had higher net benefit compared with the other models across majority range of threshold probabilities.

TABLE 2 | Performance comparison of the predictive models in the independent dataset.

Models	AUC (95% CI)	p-Value	Cut-off threshold	Accuracy	Sensitivity	Specificity
Baseline	0.562 (0.406~0.710)	Reference	0.5396	64.4% (29/45)	66.7% (20/30)	60.0% (9/15)
Model 1	0.693 (0.538~0.822)	0.314	0.5239	66.7% (30/45)	70.0% (21/30)	60.0% (9/15)
Model 2	0.787 (0.639~0.895)	0.046	0.5248	71.1% (32/45)	66.7% (20/30)	80.0% (12/15)
Model 3	0.727 (0.573~0.849)	0.197	0.4918	75.6% (34/45)	76.7% (23/30)	73.3% (11/15)
Model 4	0.811 (0.667~0.912)	0.022	0.4685	84.4% (38/45)	93.3% (28/30)	66.7% (10/15)
Combined	0.831 (0.690~0.926)	0.024	0.6570	82.2% (37/45)	86.7% (26/30)	73.3% (11/15)

AUC, area under the curve.

changes between the short-term follow-up images and the baseline CT images are often invisible to mostly radiologists, hence the need to evaluate invasiveness using DL algorithms. Previous studies have shown that an increase of 2 mm or more in diameter indicates that a GGN is growing; this change is often related to the malignant characteristics of the nodule (26, 29). Qi et al. showed that compared with the 2D diameter, a 20% increase in volume can reflect the growth of GGNs with greater sensitivity and accuracy (30). In addition, the development of solid proportions is considered strong evidence for clinical management of part-solid nodules (31). Recently, increasing numbers of studies have shown that high-throughput extraction of details that are not obvious or visible to the human eye, using radiomics and artificial intelligence, has great advantages and promising applications (11, 12, 32–34). Therefore, considering the results of the above studies, it may be reasonable to infer that the DL model incorporating both follow-up and baseline CT scans could better predict the invasiveness of nodules, enabling GGNs to be managed more rationally and avoiding unnecessary surgical resection.

Several studies have reported that radiomics and DL algorithms could be used to detect the invasiveness of GGNs. However, most of these studies just focused on the nodules themselves, few had investigated the contribution of the microgrowth environment to the prediction of tumor invasiveness. Wu suggested that there were differences between ICA and MIA/AIS in the radiomic features of cluster prominence and the gray level run-length matrix in the surrounding area of the tumor (35). Wang and Beig et al. found that clinical interpretation of peritumoral radiomics features could be used to differentiate adenocarcinoma from granuloma, predict the characteristics of lymph node metastasis, and evaluate recurrence rates after surgery (36, 37). Those reports indicated that the peritumor regions of GGNs could be used for diagnosis of invasiveness.

In our study, the proposed DL algorithms could be categorized into two groups (gross ROI-DL and full ROI-DL) depending on the size of the extracted GGN ROI range. The full ROI patch containing both the nodules and perinodular regions provided information on the nodules themselves as well as their microgrowth environment. As expected, the full ROI group (models 3 and 4) achieved higher AUC values for predicting tumor invasiveness than the gross ROI group (models 1 and 2). These findings suggested that the spatial pattern of perinodular regions could also have a role in tumor invasiveness prediction. In addition, the results were similar to those of previous

radiomics studies, in which combining texture features extracted from both intranodular and perinodular regions led to better performance compared with the single intranodular-based approach (35, 37). Unfortunately, there are currently no relevant authoritative studies or guidance on the specific size of the perinodular region (38). Furthermore, the combined model that integrated the serial CT images and clinical information involved calculations using a specific formula for prediction. DCA demonstrated that the combined model had a moderate increase in overall net benefit compared with the DL models (models 1–4).

To our knowledge, this is the first study combining the serial CT imaging to corroborate the quantitative predictive relationship between clinical-radiological characteristics and invasiveness. Most previous studies focused on the time-to-growth characteristics of tumors or the effectiveness of computer-aided diagnosis. Yoshihisa reported that smoking history and initial lesion diameter were strongly related to GGN growth (18). Matsuguma et al. analyzed the growth of 174 subsolid GGNs during the follow-up period and found that history of lung cancer was a significant predictive factor in GGN growth (39). In recent years, various DL models have been widely used to evaluate and detect changes in GGNs and have shown excellent performance compared with radiologists (40, 41). Zhao et al. reported a DL system based on 3D CNNs, and multitask learning, which achieved better classification performance than senior and junior doctors in pathological labeling of GGNs (41). Moreover, Ding et al. applied two models for distinguishing degree of nodule invasiveness, the lung DL model and dense model; both modes showed high performance in terms of AUC (0.88 and 0.86, respectively), especially the lung DL model (42).

Our study also had several limitations. First, it was a retrospective single-center study and the number of GGNs used for model development was limited; a prospective multicenter study with a larger sample size will be required in the future. Second, the ROIs were mainly manually segmented. Automatic detection and segmentation of GGNs will be considered in our future research. Third, whether incorporating more time-point serial CT images (e.g., 6-month follow-up and 12-month follow-up) could further benefit DL models in predicting tumor invasiveness still needs investigation.

In conclusion, integration of consecutive serial CT images improves the predictive efficacy of DL models in differentiating noninvasive GGNs from invasive GGNs, and the performance could be further improved by incorporating clinical information. The proposed DL models in this study show favorable performance and might have the potential to assist clinicians in tailoring precise therapy.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

This study was approved by the Institutional Ethics Committee of the Second Affiliated Hospital of Soochow University, and the informed consent requirement was waived.

AUTHOR CONTRIBUTIONS

HY and WT performed the deep learning model, analyzed the data. All authors contributed to the article and approved the submitted version.

REFERENCES

1. Singh R, Kalra MK, Homayounieh F, Nitiwarangkul C, McDermott S, Little BP, et al. Artificial Intelligence-Based Vessel Suppression for Detection of Sub-Solid Nodules in Lung Cancer Screening Computed Tomography. *Quant Imaging Med Surg* (2021) 11(4):1134–43. doi: 10.21037/qims-20-630
2. Huang CY, Huang CC, Huang WM, Liang CH, Wu FZ. Letter to the Editor Regarding "Long-Term Follow-Up of Ground-Glass Nodules After 5 Years of Stability." by Lee Et al., J Thorac Oncol 2019;14:1370-7. *Heart Lung Circ* (2020) 29:e254–7. doi: 10.1016/j.hlc.2020.02.010
3. MacMahon H, Naidich DP, Goo JM, Lee KS, Leung ANC, Mayo JR, et al. Guidelines for Management of Incidental Pulmonary Nodules Detected on CT Images: From the Fleischner Society 2017. *Radiology* (2017) 284(1):228–43. doi: 10.1148/radiol.2017161659
4. Xiong Z, Jiang Y, Che S, Zhao W, Guo Y, Li G, et al. Use of CT Radiomics to Differentiate Minimally Invasive Adenocarcinomas and Invasive Adenocarcinomas Presenting as Pure Ground-Glass Nodules Larger Than 10 Mm. *Eur J Radiol* (2021) 141:109772. doi: 10.1016/j.ejrad.2021.109772
5. Hong JH, Park S, Kim H, Goo JM, Park IK, Kang CH, et al. Volume and Mass Doubling Time of Lung Adenocarcinoma According to WHO Histologic Classification. *Korean J Radiol* (2021) 22(3):464–75. doi: 10.3348/kjr.2020.0592
6. Peng M, Li Z, Hu H, Liu S, Xu B, Zhu W, et al. Pulmonary Ground-Glass Nodules Diagnosis: Mean Change Rate of Peak CT Number as a Discriminative Factor of Pathology During a Follow-Up. *Br J Radiol* (2016) 89(1058):20150556. doi: 10.1259/bjr.20150556
7. Nam JG, Park S, Hwang EJ, Lee JH, Jin KN, Lim KY, et al. Development and Validation of Deep Learning-Based Automatic Detection Algorithm for Malignant Pulmonary Nodules on Chest Radiographs. *Radiology* (2019) 290(1):218–28. doi: 10.1148/radiol.2018180237
8. Shi W, Peng X, Liu T, Cheng Z, Lu H, Yang S, et al. A Deep Learning-Based Quantitative Computed Tomography Model for Predicting the Severity of COVID-19: A Retrospective Study of 196 Patients. *Ann Transl Med* (2021) 9 (3):216. doi: 10.21037/atm-20-2464
9. Kalmet PHS, Sanduleanu S, Primakov S, Wu G, Jochems A, Refaee T, et al. Deep Learning in Fracture Detection: A Narrative Review. *Acta Orthop* (2020) 91(2):215–20. doi: 10.1080/17453674.2019.1711323
10. Yu Y, Wang N, Huang N, Liu X, Zheng Y, Fu Y, et al. Determining the Invasiveness of Ground-Glass Nodules Using a 3D Multi-Task Network. *Eur Radiol* (2021) 31(9):7162–71. doi: 10.1007/s00330-021-07794-0
11. Wang X, Li Q, Cai J, Wang W, Xu P, Zhang Y, et al. Predicting the Invasiveness of Lung Adenocarcinomas Appearing as Ground-Glass Nodule on CT Scan Using Multi-Task Learning and Deep Radiomics. *Transl Lung Cancer Res* (2020) 9(4):1397–406. doi: 10.21037/tlcr-20-370
12. Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, Veeriah S, et al. Tracking the Evolution of Non-Small-Cell Lung Cancer. *N Engl J Med* (2017) 376(22):2109–21. doi: 10.1056/NEJMoa1616288

13. Xu Y, Hosny A, Zeleznik R, Parmar C, Coroller T, Franco I, et al. Deep Learning Predicts Lung Cancer Treatment Response From Serial Medical Imaging. *Clin Cancer Res* (2019) 25(11):3266–75. doi: 10.1158/1078-0432.CCR-18-2495
14. Ather S, Kadir T, Gleeson F. Artificial Intelligence and Radiomics in Pulmonary Nodule Management: Current Status and Future Applications. *Clin Radiol* (2020) 75(1):13–9. doi: 10.1016/j.crad.2019.04.017
15. Yang Y, Feng X, Chi W, Li Z, Duan W, Liu H, et al. Deep Learning Aided Decision Support for Pulmonary Nodules Diagnosing: A Review. *J Thorac Dis* (2018) 10(7):S867–75. doi: 10.21037/jtd.2018.02.57
16. Sakamoto R, Yakami M, Fujimoto K, Nakagomi K, Kubo T, Emoto Y, et al. Temporal Subtraction of Serial CT Images With Large Deformation Diffeomorphic Metric Mapping in the Identification of Bone Metastases. *Radiology* (2017) 285(2):629–39. doi: 10.1148/radiol.2017161942
17. Chung MP, Nam BD, Lee KS, Han J, Park JS, Hwang JH, et al. Serial Chest CT in Cryptogenic Organizing Pneumonia: Evolutional Changes and Prognostic Determinants. *Respirology* (2018) 23(3):325–30. doi: 10.1111/resp.13188
18. Kobayashi Y, Sakao Y, Deshpande GA, Fukui T, Mizuno T, Kuroda H, et al. The Association Between Baseline Clinical-Radiological Characteristics and Growth of Pulmonary Nodules With Ground-Glass Opacity. *Lung Cancer* (2014) 83(1):61–6. doi: 10.1016/j.lungcan.2013.10.017
19. Zhang G, Yang Z, Gong L, Jiang S, Wang L. Classification of Benign and Malignant Lung Nodules From CT Images Based on Hybrid Features. *Phys Med Biol* (2019) 64(12):125011. doi: 10.1088/1361-6560/ab2544
20. Duan C, Cao Y, Zhou L, Tan MT, Chen P. A Novel Nonparametric Confidence Interval for Differences of Proportions for Correlated Binary Data. *Stat Methods Med Res* (2018) 27(8):2249–63. doi: 10.1177/0962280216679040
21. Eguchi T, Adusumilli PS. Risk Stratification for Lung Nodules: Size Isn't Everything. *J Thorac Cardiovasc Surg* (2017) 153(6):1557–62. doi: 10.1016/j.jtcvs.2016.12.068
22. Hendrycks D, Lee K, Mazeika M. Using Pre-Training Can Improve Model Robustness and Uncertainty. *PMLR* (2019) 97:2712–21.
23. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. *arXiv* (2014) 1412.6980v9.
24. WJ YOUDEN. Index for Rating Diagnostic Tests. *Cancer* (1950) 3(1):32–5. doi: 10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3
25. Lee SM, Park CM, Goo JM, Lee HJ, Wi JY, Kang CH. Invasive Pulmonary Adenocarcinomas *Versus* Preinvasive Lesions Appearing as Ground-Glass Nodules: Differentiation by Using CT Features. *Radiology* (2013) 268(1):265–73. doi: 10.1148/radiol.13120949
26. Kim HK, Choi YS, Kim K, Shim YM, Jeong SY, Lee KS, et al. Management of Ground-Glass Opacity Lesions Detected in Patients With Otherwise Operable Non-Small Cell Lung Cancer. *J Thorac Oncol* (2009) 4(10):1242–6. doi: 10.1097/JTO.0b013e3181b3fee3
27. Horeweg N, van der Aalst CM, Thunnissen E, Nackaerts K, Weenink C, Groen HJ, et al. Characteristics of Lung Cancers Detected by Computer Tomography Screening in the Randomized NELSON Trial. *Am J Respir Crit Care Med* (2013) 187(8):848–54. doi: 10.1164/rccm.201209-1651OC
28. Song SH, Park H, Lee G, Lee HY, Sohn I, Kim HS, et al. Imaging Phenotyping Using Radiomics to Predict Micropapillary Pattern Within Lung Adenocarcinoma. *J Thorac Oncol* (2017) 12(4):624–32. doi: 10.1016/j.jtho.2016.11.2230
29. Cho J, Kim ES, Kim SJ, Lee YJ, Park JS, Cho YJ, et al. Long-Term Follow-Up of Small Pulmonary Ground-Glass Nodules Stable for 3 Years: Implications of the Proper Follow-Up Period and Risk Factors for Subsequent Growth. *J Thorac Oncol* (2016) 11(9):1453–9. doi: 10.1016/j.jtho.2016.05.026
30. Qi LL, Wu BT, Tang W, Zhou LN, Huang Y, Zhao SJ, et al. Long-Term Follow-Up of Persistent Pulmonary Pure Ground-Glass Nodules With Deep Learning-Assisted Nodule Segmentation. *Eur Radiol* (2020) 30(2):744–55. doi: 10.1007/s00330-019-06344-z
31. Lee SW, Leem CS, Kim TJ, Lee KW, Chung JH, Jheon S, et al. The Long-Term Course of Ground-Glass Opacities Detected on Thin-Section Computed Tomography. *Resp Med* (2013) 107:904–10. doi: 10.1016/j.rmed.2013.02.014
32. Dou Q, Chen H, Yu L, Qin J, Heng PA. Multilevel Contextual 3-D CNNs for False Positive Reduction in Pulmonary Nodule Detection. *IEEE Trans BioMed Eng* (2017) 64(7):1558–67. doi: 10.1109/TBME.2016.2613502
33. Setio AA, Ciompi F, Litjens G, Gerke P, Jacobs C, van Riel SJ, et al. Pulmonary Nodule Detection in CT Images: False Positive Reduction Using Multi-View Convolutional Networks. *IEEE Trans Med Imaging* (2016) 35(5):1160–9. doi: 10.1109/TMI.2016.2536809
34. Cui S, Ming S, Lin Y, Chen F, Shen Q, Li H, et al. Development and Clinical Application of Deep Learning Model for Lung Nodules Screening on CT Images. *Sci Rep* (2020) 10(1):13657. doi: 10.1038/s41598-020-70629-3
35. Wu L, Gao C, Xiang P, Zheng S, Pang P, Xu M, et al. CT-Imaging Based Analysis of Invasive Lung Adenocarcinoma Presenting as Ground Glass Nodules Using Peri- and Intra-Nodular Radiomic Features. *Front Oncol* (2020) 27(10):838. doi: 10.3389/fonc.2020.00838
36. 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
37. 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 (3):783–92. doi: 10.1148/radiol.2018180910
38. Sun R, Limkin EJ, Vakalopoulou M, Dercle L, Champiat S, Han SR, et al. A Radiomics Approach to Assess Tumour-Infiltrating CD8 Cells and Response to Anti-PD-1 or Anti-PD-L1 Immunotherapy: An Imaging Biomarker, Retrospective Multicohort Study. *Lancet Oncol* (2018) 19(9):1180–91. doi: 10.1016/S1470-2045(18)30413-3
39. Matsuguma H, Mori K, Nakahara R, Suzuki H, Kasai T, Kamiyama Y, et al. Characteristics of Subsolid Pulmonary Nodules Showing Growth During Follow-Up With CT Scanning. *Chest* (2013) 143(2):436–43. doi: 10.1378/chest.11-3306
40. Gong J, Liu J, Hao W, Nie S, Zheng B, Wang S, et al. A Deep Residual Learning Network for Predicting Lung Adenocarcinoma Manifesting as Ground-Glass Nodule on CT Images. *Eur Radiol* (2020) 30(4):1847–55. doi: 10.1007/s00330-019-06533-w
41. Zhao W, Yang J, Sun Y, Li C, Wu W, Jin L, et al. 3d Deep Learning From CT Scans Predicts Tumor Invasiveness of Subcentimeter Pulmonary Adenocarcinomas. *Cancer Res* (2018) 78(24):6881–9. doi: 10.1158/0008-5472.CAN-18-0696
42. Ding H, Xia W, Zhang L, Mao Q, Cao B, Zhao Y, et al. CT-Based Deep Learning Model for Invasiveness Classification and Micropapillary Pattern Prediction Within Lung Adenocarcinoma. *Front Oncol* (2020) 22(10):1186. doi: 10.3389/fonc.2020.01186

Conflict of Interest: HY and WT were employed by Beijing Infervision Technology Co., Ltd.

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

frontiers in Oncology

ORIGINAL RESEARCH
published: 06 October 2021
doi: 10.3389/fonc.2021.735739

Qualitative Histopathological Classification of Primary Bone Tumors Using Deep Learning: A Pilot Study

Yuzhang Tao[1], Xiao Huang[1], Yiwen Tan[2,3], Hongwei Wang[1], Weiqian Jiang[1], Yu Chen[1], Chenglong Wang[4], Jing Luo[2], Zhi Liu[5], Kangrong Gao[5], Wu Yang[1], Minkang Guo[1], Boyu Tang[1], Aiguo Zhou[1], Mengli Yao[6], Tingmei Chen[6], Youde Cao[2], Chengsi Luo[7] and Jian Zhang[1]**

[1] Department of Orthopaedics, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China, [2] Department of Pathology, College of Basic Medicine, Chongqing Medical University, Chongqing, China, [3] Department of Pathology, The Second Affiliated Hospital of Chongqing Medical University, Chongqing, China, [4] Department of Pathology, Chongqing Hospital of Traditional Chinese Medicine, Chongqing, China, [5] Research and Development Department, Chongqing Defang Information Technology Co., Ltd, Chongqing, China, [6] Key Laboratory of Clinical Laboratory Diagnostics (Ministry of Education), College of Laboratory Medicine, Chongqing Medical University, Chongqing, China, [7] School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, China

OPEN ACCESS

Edited by:
Jialiang Yang,
Geneis (Beijing) Co. Ltd., China

Reviewed by:
Sikandar Shaikh,
Shadan Hospital and Institute of Medical Sciences, India
Massimo Salvi,
Polytechnic University of Turin, Italy

***Correspondence:**
Jian Zhang
zhangjian@hospital.cqmu.edu.cn
Chengsi Luo
luosicheng945@gmail.com

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 03 July 2021
Accepted: 13 September 2021
Published: 06 October 2021

Citation:
Tao Y, Huang X, Tan Y, Wang H, Jiang W, Chen Y, Wang C, Luo J, Liu Z, Gao K, Yang W, Guo M, Tang B, Zhou A, Yao M, Chen T, Cao Y, Luo C and Zhang J (2021) Qualitative Histopathological Classification of Primary Bone Tumors Using Deep Learning: A Pilot Study. Front. Oncol. 11:735739. doi: 10.3389/fonc.2021.735739

Background: Histopathological diagnosis of bone tumors is challenging for pathologists. We aim to classify bone tumors histopathologically in terms of aggressiveness using deep learning (DL) and compare performance with pathologists.

Methods: A total of 427 pathological slides of bone tumors were produced and scanned as whole slide imaging (WSI). Tumor area of WSI was annotated by pathologists and cropped into 716,838 image patches of 256 × 256 pixels for training. After six DL models were trained and validated in patch level, performance was evaluated on testing dataset for binary classification (benign *vs*. non-benign) and ternary classification (benign *vs*. intermediate *vs*. malignant) in patch-level and slide-level prediction. The performance of four pathologists with different experiences was compared to the best-performing models. The gradient-weighted class activation mapping was used to visualize patch's important area.

Results: VGG-16 and Inception V3 performed better than other models in patch-level binary and ternary classification. For slide-level prediction, VGG-16 and Inception V3 had area under curve of 0.962 and 0.971 for binary classification and Cohen's kappa score (CKS) of 0.731 and 0.802 for ternary classification. The senior pathologist had CKS of 0.685 comparable to both models ($p = 0.688$ and $p = 0.287$) while attending and junior pathologists showed lower CKS than the best model (each $p < 0.05$). Visualization showed that the DL model depended on pathological features to make predictions.

Conclusion: DL can effectively classify bone tumors histopathologically in terms of aggressiveness with performance similar to senior pathologists. Our results are

promising and would help expedite the future application of DL-assisted histopathological diagnosis for bone tumors.

Keywords: primary bone tumors, deep learning, histopathological classification, convolutional neural network (CNN), diagnosis

1 INTRODUCTION

Primary bone tumors are a variety of neoplasms formed from the bone tissue (1). Although the incidence is relatively low, primary bones and joints' malignancy is ranked the third and fourth leading cause of death for males and females under 20 years of age in the United States (2). The biological behavior of bone tumors varies greatly among different classes (3). However, their clinical management is mainly determined by the extent of the tumor's aggressiveness, which is usually graded as benign, intermediate, and malignant (4). While the bone tumor's clinical characteristics and radiological information may help physicians reach an initial diagnosis, histopathological assessment of biopsy tissue remains decisive in determining the bone tumor's biological nature and confirming its aggressiveness (5). Therefore, an accurate and reliable histopathological differentiation is imperative to ensure a satisfactory patient outcome.

Unlike tumors of epithelial origin that are more prevalent, pathologists' experience in diagnosing bone tumors usually lacks due to the relatively low incidence and various histological morphology. Additionally, some bone tumors of different kinds may share similar histologic morphology because of mesenchymal origin, thus introducing confounding factors in classification. Moreover, the pathologist's prediction of bone tumor's histopathological classification, which is prone to subjectivity, could not be adequately quantified for the moment.

Considering the drawbacks of traditional histopathological analysis mentioned above, diagnostic approaches based on artificial intelligence gradually come into existence, along with the accelerated development of computational power and deep learning (DL) (6). The convolutional neural network (CNN), a network composed of deep layers, can be trained to extract specific features from an image dataset to output a quantitative probability and build a classifier (7). In addition, the emergence of whole slide imaging (WSI) enables slides digitalized as macro data without information loss (8), which is suitable for neural networks to process and learn. Utilizing WSI over the last few years, the CNN has been verified efficient in the histopathological classification of numerous tumors of epithelial origin, such as breast cancer (9), lung cancer (10), gastric cancer (11), prostate cancer (12), and nasopharyngeal cancer (13). In comparison to tumors of epithelial origin, bone tumors are mostly of mesenchymal origin, showing remarkably different and diverse microscopic morphology. However, there lacks relevant evidence regarding the performance of DL-based histopathological classification for bone tumors so far.

Accurate DL-assisted differentiation of primary bone tumors microscopically and qualitatively as benign, intermediate, and malignant would not only compensate for the limited experience and biased interpretation of physicians, but also provide a quantitative approach to assess the biological nature of bone tumors, potentially leading to a better treatment decision. In this study, we evaluate the feasibility of using DL in qualitative histopathological differentiation of primary bone tumors and compare the performance of the best model with pathologists of different levels of experience.

2 MATERIALS AND METHODS

2.1 Specimen Information

According to the 1964 Helsinki declaration and its later amendments, this study was approved by the ethics committee of the First Affiliated Hospital of Chongqing Medical University (No. 2020-287). After ensuring that informed consents were obtained from relevant patients, all specimens of primary bone tumor resected in the hospital between July 2014 and October 2020 were retrieved from the Department of Pathology, Chongqing Medical University. Based on the histopathological, clinical, and radiological information, the collected samples' diagnoses were confirmed by at least one senior pathologist in accordance with the 2013 World Health Organization (WHO) classification (4). A total of 458 specimens were finally determined and classified into three groups, in which 206 were benign, 96 were intermediate, 156 were malignant.

2.2 Data Preparation

2.2.1 Section and Staining

The collected paraffin-embedded specimens were sectioned and stained under a standardized protocol, producing one corresponding hematoxylin and eosin (H&E) slide for each specimen. All slides were de-identified and only labeled with diagnosis. The quality control of all slides was done by a senior pathologist, and 31 slides (8 benign, 10 intermediate, and 13 malignant cases) were excluded from the study. The remaining 427 slides were finally chosen for scanning. **Supplementary Table S1** shows the detailed number of cases with definitive diagnoses in each group. The average age of the included cases was 38.06 years (from 7 to 89 years), while males and females accounted for 53.62% and 46.37%, respectively.

2.2.2 WSI Scanning and Storage

The selected slides were scanned using a digital slide scanner (Chongqing Defang Information Technology Co., Ltd, Chongqing, China) to produce ultra-high-resolution whole slide images at the default 40× objective magnification (**Figures 1A–C**), which then were stored as svs format. The average memory size of all WSIs was 5.76 GB, and the width and height of WSIs were at least 149,520 and 150,420 pixels.

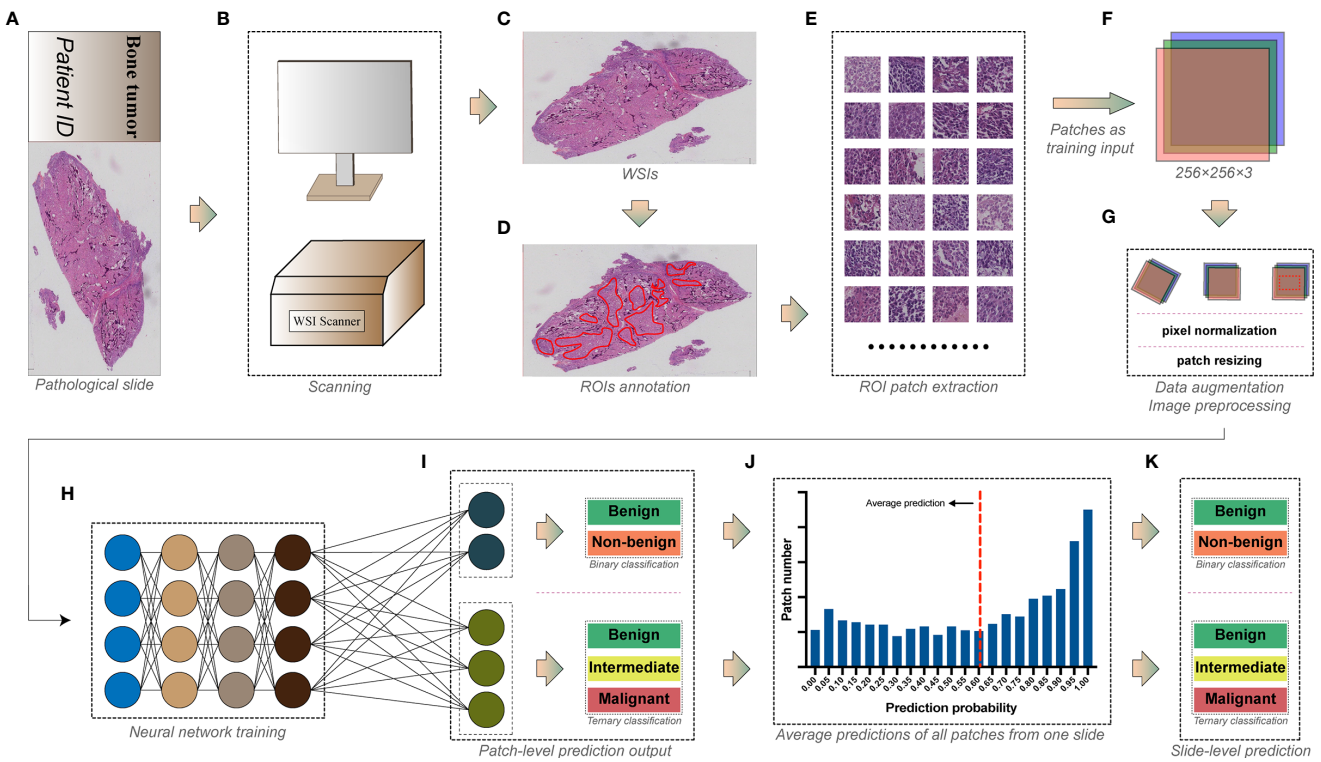


FIGURE 1 | The workflow for deep learning training. **(A)** The pathological slide prepared from bone tumor specimens. **(B)** Whole slide imaging (WSI) scanning. **(C)** Whole slide images with ultra-high resolution. **(D)** Tumor areas annotated by pathologists as regions of interest (ROIs). **(E)** Regions of interest were exported as numerous square image patches of the same size. **(F)** Three-channel RGB image patches of 256 × 256 pixels were used as training input. **(G)** Data augmentation and image preprocessing before training. **(H)** Image patches in the training dataset were fed into the convolutional neural network for training. **(I)** Models were trained for the patch-level binary and ternary classification. **(J)** Predictions of all patches from one slide were averaged to obtain **(K)** the slide-level binary and ternary prediction.

2.2.3 Annotation

WSIs were analyzed by pathologists using Qupath (14) (version 0.2.3, Queens University). Areas constituted of tumor-related cells and structures were considered as viable tumor areas, while other non-specific normal connective tissues and white space were regarded as non-tumor areas. Two junior pathologists examined all WSIs under 1× to 40× objective magnification before determining and annotating viable tumor areas as regions of interest (ROIs) using Qupath built-in annotation tools (**Figure 1D**). WSIs were subsequently rechecked by another senior pathologist to ensure the accuracy of annotation.

2.2.4 Dataset Allocation

All WSIs under each group were randomly split into training, validation, and testing datasets in a proportion of 70:15:15. Slide dataset information for each group is shown in **Supplementary Table S2**.

2.2.5 Image Patch Extraction

WSIs are images with more than hundreds of millions of pixels (8), which are too huge to be used as input in training DL models. Moreover, the discriminative information of histopathology is usually retained at the cellular level (15). Therefore, ROIs of WSIs are usually cropped into plenty of image patches with fixed dimensions (typically from 32 × 32 to 10,000 × 10,000 pixels, where 256 × 256 is the most widely used) as input, making the training possible and efficient (16). As a result, we used the Qupath script editor to continuously crop the annotated viable tumor areas into square image patches of 256 × 256 pixels without overlapping (**Figure 1E**). In this study, we used a down-sampling factor of four when cropping the ROIs because the image patch of 256 × 256 pixels generated from the original 40× scanning magnification was insufficient to include a satisfactory tumor area. Image patches with background constituting more than 50% of their areas were abandoned. The cropped patches share the same group label as the slides from which they were generated. A total of 716,838 patches were finally generated, and detailed information of image patches for each group is shown in **Supplementary Table S3**.

2.3 Network Training and Performance Evaluation

Several widely tested convolutional neural network architectures, including AlexNet (17), VGG-16 (18), Inception V3 (19),

DenseNet-121 (20), ResNet-50 (21), and MnasNet (22) were chosen for training the patch-level classification. All image patches extracted were saved in 8-bit JPEG format (**Figure 1F**). We performed the data augmentation and preprocessing by random rotation, random horizontal flip, and normalization of the original image (**Figure 1G**). The angle of random rotation ranged from −45° to 45°. The probability of images being flipped was 0.5. Pixel values of three-channel images were normalized by scaling their values into the range from zero to one, then subtracting [0.485, 0.456, 0.406] and dividing by [0.229, 0.224, 0.225] channel-wise. Random resized cropping was used such that a crop of random size (0.08 to 1.0) of the original size and a random aspect ratio (of 3/4 to 4/3) of the original aspect ratio was made, finally resizing the image to a given size (224 × 224 or 299 × 299, according to the model's pre-trained dataset, shown in **Supplementary Figures S1–S6**) as training input.

All models were pre-trained on the ImageNet dataset to initialize kernel weights. Stochastic gradient descent (SGD) with a categorical cross-entropy loss was implemented to update the model's weights, accompanied by a cyclic learning rate (23) (cLR) oscillating between 10^{-4} and 10^{-6} every quarter epoch. The batch size of 64 was set for training. Models were trained on patch level (**Figures 1H, I**) for binary classification (benign *vs*. non-benign) and a ternary classification (benign *vs*. intermediate *vs*. malignant). The architecture and specific hyper-parameters of each model are shown in **Supplementary Figures S1–S6**.

The model's generalizability for each epoch during training was evaluated with validation dataset using loss and accuracy for binary classification or using loss, accuracy, and the Cohen's kappa score (CKS) for ternary classification. All models were trained for 30 epochs, and parameters of the epoch with the highest validation accuracy (binary task) or CKS (ternary task) were used to predict the classification of the testing dataset. We compared the patch-level diagnostic metrics on the testing dataset between different models and determined the best architecture, which was then used to predict the slide-level classification.

The model's predictions of all image patches generated from one slide WSI were averaged to produce a slide-level prediction (**Figures 1J, K**). Then, the true label of the slide was used to assess the model's slide-level classification performance on the testing dataset.

Metrics of performance for binary classification included accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), F1-score, the receiver operating characteristic (ROC) curve, and the area under the curve (AUC), whereas accuracy, the Cohen's kappa score, precision, recall, and F1-score were used to evaluate the model's ternary classification performance.

2.4 Experiment Setup

Our DL experiments were performed on a server with 4× NVIDIA GeForce RTX 2080 Ti graphics processing units (11 GB of memory for each). We developed the relevant DL algorithms with Python 3.6 and PyTorch 1.7.1 on an Ubuntu platform.

2.5 Evaluation of Pathologist's Performance

All slides in the testing dataset were read by one senior pathologist (pathologist #1, with more than 25 years of experience), two attending pathologists (pathologist #2 and #3, with more than 10 years of experience), and one resident pathologist (pathologist #4, with less than 5 years of experience) without knowing any slide's information beforehand. Then, they labeled each slide as benign, intermediate, or malignant according to their own interpretations. Their predictions of all slides were recorded and compared with the slides' corresponding ground-truth labels to calculate the pathologist's diagnostic performance. Metrics used in the model's slide-level performance evaluation were analyzed for pathologists and finally compared between model and human.

2.6 Model Visualization and Case Review

Gradient-weighted class activation mapping (Grad-CAM) is an approach that uses the gradients flowing into the last convolutional layer to create a map localizing and highlighting the important regions relevant to model prediction in an image (24). Therefore, we used Grad-CAM to visualize the important regions associated with discriminative histopathological features that the DL model relies on, thus revealing the underlying mechanism of the model's prediction. In the slide-level classification, we identified the slide cases that the model, or pathologist, or both wrongly classified. Then, the senior pathologist was asked to determine the potential causes of such misclassifications by reviewing the representative image patches of the corresponding slide, along with the model visualization.

2.7 Statistical Analysis

Data used in this study were analyzed with SPSS software (version 26.0; IBM, Chicago, IL) and SAS 9.4 (SAS Institute, Cary, NC, USA). The metrics of performance for slide-level binary classification between models and pathologists were compared using McNemar's test. The 95% confidence intervals (CIs) of AUCs were calculated and compared between groups using the Delong methods (25), and the 95% CIs of the Cohen's kappa scores were acquired by the bootstrap method (26) with 10,000 replications and compared between the model and the pathologist using the permutation test with 10,000 iterations. The AUC in different ranges represented the following predictive performance: poor ($0.5 \leq AUC < 0.7$), fair ($0.7 \leq AUC < 0.8$), good ($0.8 \leq AUC < 0.9$), and excellent ($0.9 \leq AUC$). We characterize the Cohen's kappa score of 0–0.20, 0.21–0.41, 0.41–0.60, 0.61–0.80, and 0.81–1 as slight, fair, moderate, substantial, excellent agreement with the ground truth label, respectively. A *p*-value of less than 0.05 was considered statistically significant.

3 RESULT

3.1 Patch-Level Performance of Models

3.1.1 Binary Classification

All generated patches from the training dataset were fed into six pre-trained CNN models to build a binary classifier (benign *vs*.

non-benign). The learning curves for 30 epochs of all models are shown in **Figure 2**. The validation loss of most models reached the lowest level in the first 15 epochs before rising slowly afterwards, indicating that the models gained the high level of generalizability in the initial training process.

After determining models' best-performing parameters using validation accuracy, we assessed the performance of models' patch-level binary classification on the testing dataset. **Figure 3** depicts the ROC curves for each model, where the VGG-16 showed the best predictive value with an AUC of 0.940 (95% CI, 0.939–0.941), while the AlexNet had the smallest AUC of 0.902 (95% CI, 0.939–0.941) among six models. For other diagnostic metrics, the VGG-16 also had the highest accuracy (85.96%), sensitivity (83.66%), NPV (77.78%), and F1-score (87.91%) compared with other network architectures, whereas the Inception V3 showed the greatest specificity (91.34%) and PPV (93.56%). The detailed information of performance metrics for patch-level binary classification is demonstrated in **Table 1**. Therefore, we chose the VGG-16 and the Inception V3 for slide-level binary prediction.

3.1.2 Ternary Classification

Similar to the training for binary classification, six models were fed with patches that were labeled as benign, or intermediate, or malignant to train a ternary classifier. However, we utilized the CKS, rather than accuracy, to decide the best parameters in the training process because of the relative imbalance of the patch number between each class in ternary classification. **Figure 4** illustrates the learning curves for 30 epochs of each model. As the epoch increased, the validation loss of VGG-16 and Inception V3 was fairly stable at a low level compared with the other four models, showing less chance of overfitting for these two models.

For the model's performance in patch-level ternary classification on testing dataset, the VGG-16 triumphed over others on accuracy (74.78%), CKS (0.601, 95% CI 0.597–0.605), weighted average recall (0.75), and weighted average F1-score (0.75), while sharing the highest weighted average precision (0.79) with the Inception V3. **Table 2** summarizes the performance metrics of six models, and the detailed classification report for each class is shown in **Supplementary Table S4**. As a result, the VGG-16 and the Inception V3 were finally selected for slide-level ternary prediction.

3.2 Slide-Level Performance of Models and Pathologists

3.2.1 Binary Classification

The predictive probabilities of all patches generated from one slide were averaged to obtain the model's slide-level prediction.

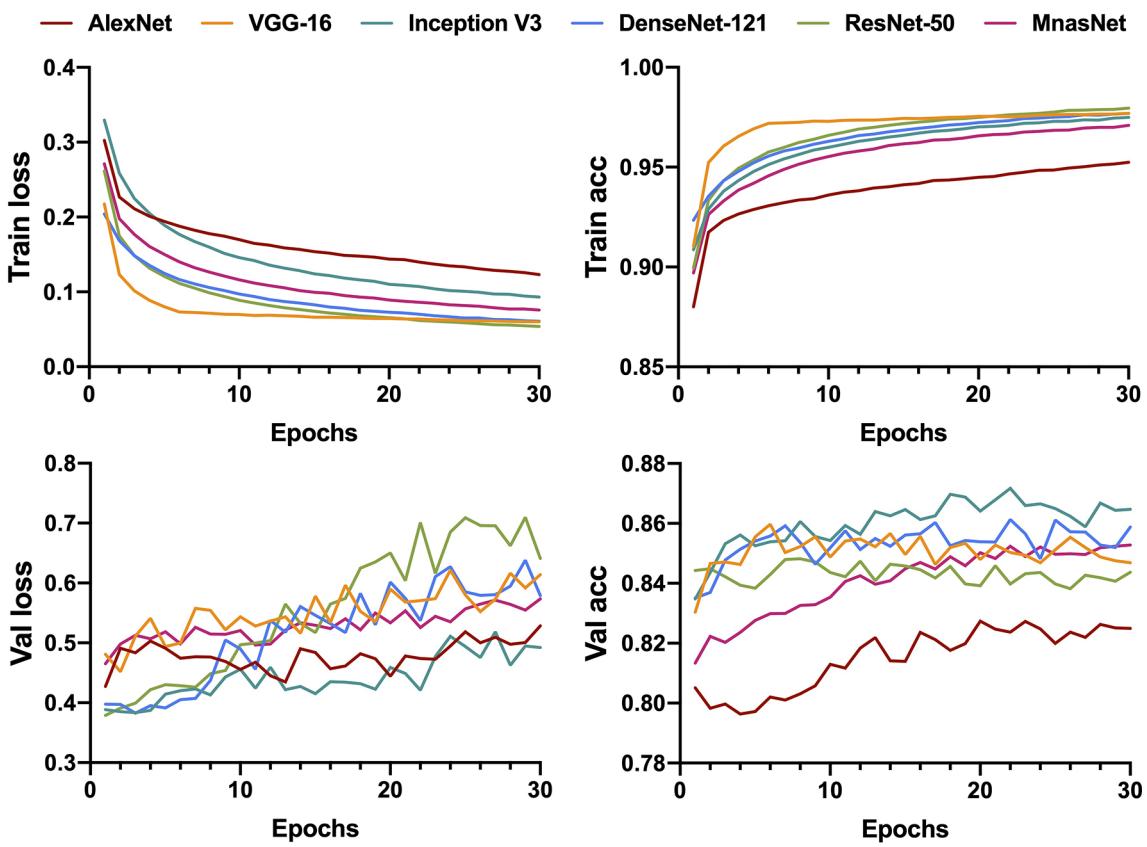


FIGURE 2 | Learning curves for patch-level binary classification of six models showing the loss and accuracy in training and validation. Train, training; Val, validation; acc, accuracy.

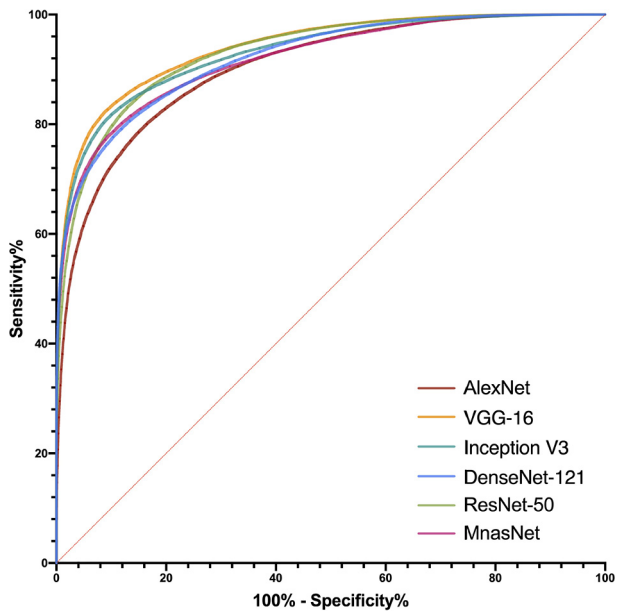


FIGURE 3 | Receiver operating characteristic curves in the patch-level binary classification of each model.

For the differentiation of benign from non-benign bone tumors on the testing dataset, the VGG-16 and the Inception V3 both showed excellent predictive capability on slide-level with the AUC of 0.962 (95% CI, 0.882–0.994) and 0.971 (95% CI, 0.897–0.997), respectively. In addition, there was no statistically significant difference between the AUCs of both models (p = 0.304). The ROC curves for slide-level binary classification of models are demonstrated in **Figure 5**, along with the results of pathologists' assessments. **Table 3** summarizes the detailed performance metrics for models and pathologists. Among models and pathologists, the VGG-16 had the highest accuracy (90.77%) and F1-score (90.91%), and the Inception V3 showed the greatest specificity (100.00%) and PPV (100.00%). Senior pathologist #1 had the best accuracy (84.62%) among pathologists, while owning better sensitivity (91.43%) and NPV (88.46%) compared with models. However, the heterogeneity of predictive performance among pathologists was significant that their sensitivities and specificity ranged from 57.14% and 76.67% to 91.43% and 93.33%, respectively. The p-values for comparison of accuracy, sensitivity, and specificity between VGG-16 and pathologists are shown in **Table 3**.

3.2.2 Ternary Classification

Slide-level ternary classification performances of models and pathologists are outlined in **Table 4**. The Inception V3 had the greatest value in each metric. Both the VGG-16 and the Inception V3 showed substantial predictive value with the CKS of 0.731 (95% CI, 0.573–0.860) and 0.802 (95% CI 0.662–0.920), whereas pathologists of all levels had the CKSs of less than 0.7. However, after pairwise comparison of CKS, we found that there were no significant differences between the VGG-16 and the Inception V3 (p = 0.182), the VGG-16 and pathologist #1 (p = 0.689), and the Inception V3 and pathologist #1 (p = 0.288). The CKSs of pathologists #2–4 were significantly lower than the Inception V3 (see **Table 4** for the detailed p-values). The detailed classification report for each class is shown in **Supplementary Table S5**.

3.3 Model Visualization

We located the slides that were correctly classified by both models (VGG-16 and Inception V3) and the senior pathologist #1 in binary and ternary classification, then chose the representative patches of selected slides for Grad-CAM visualization. The model VGG-16 was used for visualization because it showed the best binary and ternary patch-level predictive performances.

Figure 6 illustrates the heatmaps of Grad-CAM results for binary classification. For most benign cases, the model identified the widespread stromal area without cells (**Figure 6A**) or stromal areas with scattered benign cells (**Figure 6C**) as essential regions for benign prediction. In some particular cases of benign tumors that share highly similar "dense-cell" microscopic morphology with non-benign tumors, the model effectively differentiated the confusing area as the benign region (**Figure 6B**). Visualization for non-benign patches showed that the different arrangements of atypical cells were deemed by the model as discriminative features for non-benign prediction (**Figures 6D–F**).

Visualization of representative patches for ternary classification is shown in **Figure 7**. The mechanism for benign prediction of ternary classification (**Figures 7A–D**) was similar to that of binary classification. Intriguingly, the model could accurately identify the specific structures, such as giant cells (**Figures 7E–G**) and chondroblasts (**Figure 7H**), as important regions for intermediate classification. Furthermore, the highly

TABLE 1 | Performance of patch-level binary classification on testing dataset.

	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	F1-score (%)	AUC (95% CI)
AlexNet	81.13	78.77	84.83	89.05	71.85	83.59	0.902 (0.900, 0.904)
VGG-16	**85.96**	**83.66**	89.58	92.63	**77.78**	**87.91**	**0.940** (0.939, 0.941)
Inception V3	84.62	80.32	**91.34**	**93.56**	74.78	86.44	0.930 (0.929, 0.932)
DenseNet-121	83.02	81.02	86.16	90.16	74.35	85.34	0.922 (0.920, 0.923)
ResNet-50	84.73	83.58	86.54	90.67	77.09	86.98	0.930 (0.929, 0.932)
MnasNet	83.32	80.68	87.46	90.97	74.31	85.52	0.918 (0.916, 0.919)

Metric with the greatest value among different models is bolded. PPV, positive predictive value; NPV, negative predictive value; AUC, area under the curve; CI, confidence interval.

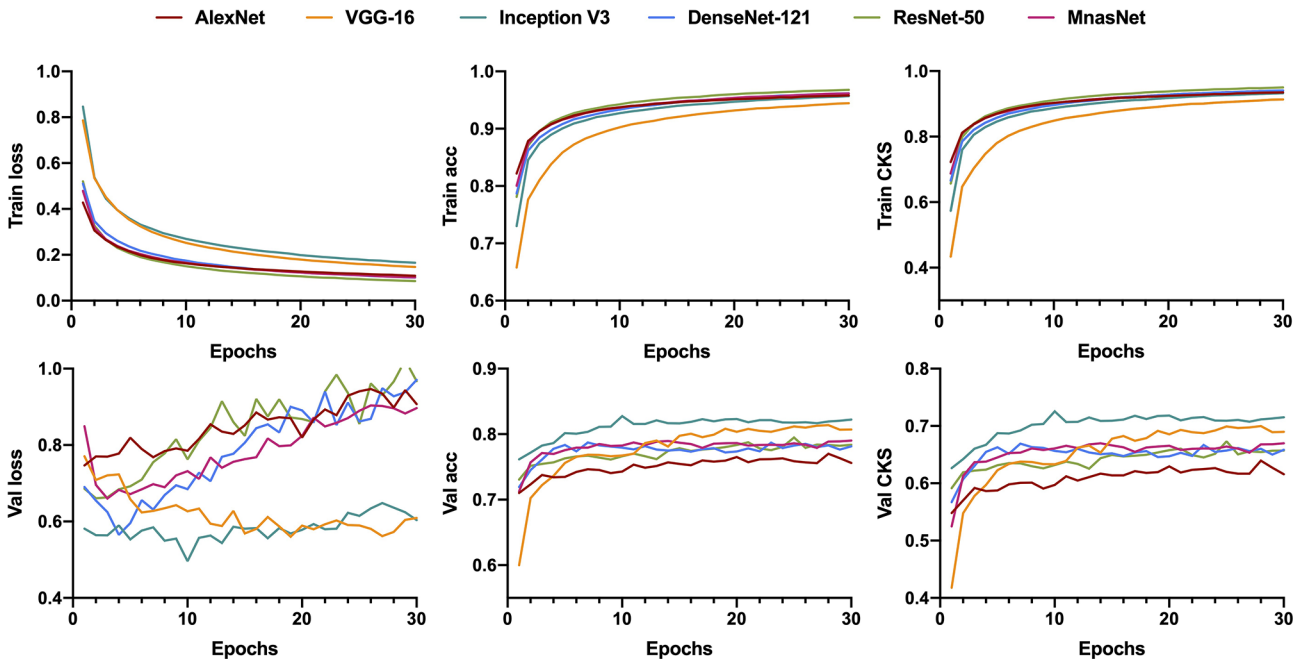


FIGURE 4 | Learning curves for patch-level ternary classification of six models showing the loss, accuracy, and the Cohen's kappa score in training and validation. Train, training; Val, validation; acc, accuracy; CKS, the Cohen's kappa score.

dense organization of atypical cells (**Figures 7I, J**) and the combination of stroma and scattered malignant cells (**Figures 7K, L**) were regarded by the model as morphological features for malignant prediction.

3.4 Case Review

We examined the slides that pathologist #1 correctly predicted, whereas both models wrongly classified. Interestingly, all six malignant slides that were classified as benign by models belonged to chondrosarcoma. After visualizing some of the patches for these six slides, we found that the model could favorably recognize atypical cells for chondrosarcoma in the patch level (**Figures 8B, C**). However, there were numerous patches of the normal interterritorial matrix (**Figures 8A, D**), which were unintentionally cropped by pathologists as ROI, for one chondrosarcoma slide. This kind of patch-level annotation noise was remarkable in chondrosarcoma, causing the number of noise patches to overcome that of the true malignant patches in the averaging process of slide-level prediction. In addition, both models classified one malignant slide as intermediate, and this slide turned out to be a malignant giant cell tumor (GCT). The mechanism of such erroneous slide-level prediction was also associated with the annotation noise (similar to that of chondrosarcoma), although the patch-level discriminative features were successfully identified by the model (**Figure 8F**). This malignant GCT was mainly composed of the normal GCT area that was characteristic of giant cells (**Figure 8E**), whereas malignant cells only constituted a small part of the annotated ROI.

Figure 9 depicts the representative patches of slides that pathologist #1 incorrectly classified but both models correctly predicted. For the benign bone tumor that has seemingly malignant microscopic structures, the model could effectively differentiate the associated patches as benign classification (**Figure 9A**). In addition, the model also showed favorable performance in identifying specific features of the intermediate

TABLE 2 | Performance of patch-level ternary classification on testing dataset.

	Accuracy (%)	Cohen's kappa score (95% CI)	WA precision	WA recall	WAF1-score
AlexNet	68.62	0.505 (0.501, 0.509)	0.73	0.69	0.69
VGG-16	**74.78**	**0.601** (0.597, 0.605)	**0.79**	**0.75**	**0.75**
Inception V3	74.17	0.591 (0.587, 0.595)	**0.79**	0.74	0.74
DenseNet-121	72.48	0.570 (0.566, 0.574)	0.78	0.72	0.73
ResNet-50	70.02	0.527 (0.523, 0.531)	0.74	0.70	0.70
MnasNet	73.39	0.575 (0.571, 0.579)	0.76	0.73	0.73

Metric with the greatest value among different models is bolded. CI, confidence interval; WA, weighted average.

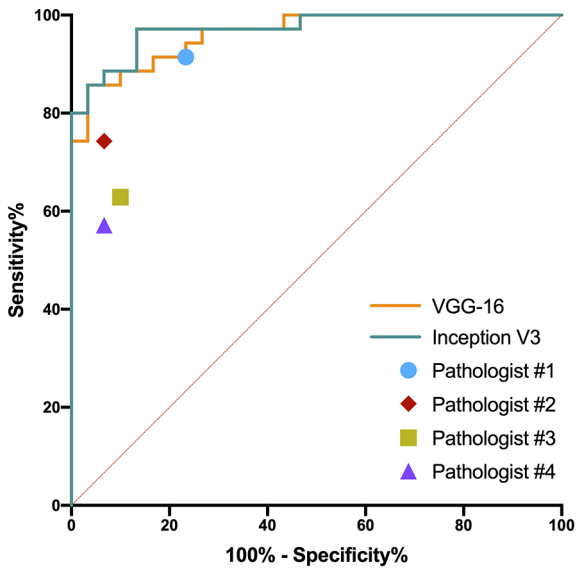


FIGURE 5 | Receiver operating characteristic curves in the slide-level binary classification of VGG-16 and Inception V3, along with results of pathologists' assessments.

slides that the pathologist was unsure of diagnosing solely based on the microscopic assessment (**Figures 9B–D**). Furthermore, for patches of malignant slides that share similar cell arrangements with intermediate cases, the model could easily and correctly distinguish the corresponding area with high predictive probability (**Figures 9E, F**).

4 DISCUSSION

In this preliminary study, we found that several widely proved DL models trained with limited pathological slides could effectively classify bone tumors histopathologically in terms of aggressiveness. The VGG-16 and Inception V3, which defeated other models in patch-level performance, showed comparable diagnostic abilities with the senior pathologist and triumphed over attending and resident pathologists in slide-level predictive performance. Moreover, we discovered that the DL model could extract specific visual features of each classification and relied on them to make favorable predictions.

In the conventional clinical setting, a patient who is suspected of bone tumor usually undergoes clinical and radiological examinations for an initial assessment. However, many cases are challenging for physicians to give definitive or qualitative diagnoses solely based on patient history or plain radiographs. Therefore, a tissue biopsy is needed under such circumstances to determine the tumor's biological nature, thus directing more appropriate treatment (5). The qualitative classification for bone tumor after biopsy is usually divided into benign, intermediate, and malignant tumor according to the aggressiveness evaluated under microscopy. Intermediate and malignant cases can be grouped as non-benign tumors because they normally require subsequent interventions. Patients diagnosed with benign bone tumors are generally requested for regular follow-up after one-stage resection during biopsy surgery. In comparison, secondary surgery that includes extensive resection and structural fixation is commonly required for intermediate bone tumors owing to the moderate recurrence rate and the local aggressiveness. For malignant bone tumors, extensive resection with implant support

TABLE 3 | Performance of slide-level binary classification on testing dataset.

	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	F1-score (%)	AUC (95% CI)
VGG-16	**90.77**	85.71	96.67	96.77	85.29	**90.91**	0.962 (0.882, 0.994)
Inception V3	87.69	77.14	**100.00**	**100.00**	78.95	87.10	**0.971** (0.897, 0.997)
Pathologist #1	84.62^{a1}	**91.43**^{b1}	76.67^{c1}	82.05	**88.46**	86.49	–
Pathologist #2	83.08^{a2}	74.29^{b2}	93.33^{c2}	92.86	75.68	82.54	–
Pathologist #3	75.38^{a3}	62.86^{b3}	90.00^{c3}	88.00	67.50	73.33	–
Pathologist #4	73.85^{a4}	57.14^{b4}	93.33^{c4}	90.91	65.12	70.18	–

Metric with the greatest value among different groups is bolded. $^{a1\text{-}4,b1\text{-}4,c1\text{-}4}$*indicate the p-values compared with the VGG-16 in accuracy, sensitivity, and specificity, respectively. a1, p = 0.317; a2, p = 0.096; a3, p = 0.012; a4, p = 0.008; b1, p = 0.480; b2, p = 0.103; b3, p = 0.021; b4, p = 0.008; c1, p = 0.034; c2, p = 0.564; c3, p = 0.317; c4, p = 0.564. PPV, positive predictive value; NPV, negative predictive value; AUC, area under the curve; CI, confidence interval.*

TABLE 4 | Performance of slide-level ternary classification on testing dataset.

	Accuracy (%)	Cohen's kappa score (95% CI)	WA precision	WA recall	WAF1-score
VGG-16	83.10	0.732 (0.574, 0.867)	0.83	0.83	0.82
Inception V3	**87.70**	**0.803** (0.664, 0.922)	**0.90**	**0.88**	**0.87**
Pathologist #1	80.00	0.686 (0.526, 0.829) ^{a1,b1}	0.81	0.80	0.80
Pathologist #2	72.31	0.543 (0.376, 0.703) ^{a2,b2}	0.70	0.72	0.70
Pathologist #3	69.23	0.490 (0.307, 0.664) ^{a3,b4}	0.70	0.69	0.68
Pathologist #4	70.77	0.507 (0.335, 0.679) ^{a4,b4}	0.74	0.71	0.69

Metric with the greatest value among different groups is bolded. $^{a1\text{-}4}$*indicates the p-value compared with the VGG-16 and* $^{b1\text{-}4}$*indicates the p-value compared with the Inception V3. a1, p = 0.689; a2, p = 0.060; a3, p = 0.036; a4, p = 0.048; b1, p = 0.288; b2, p = 0.004; b3, p = 0.002; b4, p = 0.003. CI, confidence interval; WA, weighted average.*

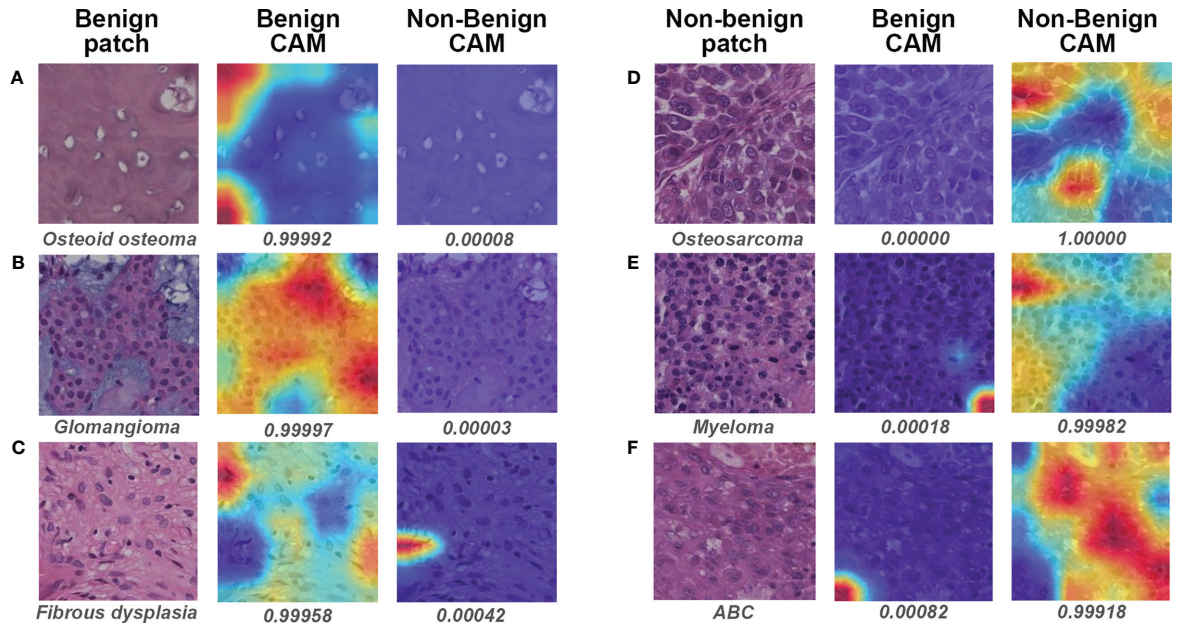


FIGURE 6 | Gradient-weighted class activation mapping (Grad-CAM) in binary classification for representative patches of slides correctly classified by both models and pathologist #1. The specific classification is shown under the original patch, and the predictive probability for CAM of each class is shown below the corresponding Grad-CAM heatmap. **(A–C)** show the representative patches of benign bone tumors, whereas **(D–F)** show the representative patches of non-benign bone tumors. CAM, class activation mapping; ABC, aneurysmal bone cyst.

is combined with chemotherapy or radiotherapy for cases without metastasis, whereas palliative therapy is needed for metastatic cases (1). Given that it is difficult for general pathologists to accurately classify bone tumors histopathologically because of the low incidence and tumor heterogeneity, the expected goal of the current study was to build a DL-based classifier that reaches the diagnostic level of pathologists from the academic medical center.

The focus of AI-related research for bone tumor diagnosis is mainly on the radiographic analysis for the moment (27, 28). Bao et al. (29) have incorporated various features from radiographic observations and demographic information to build a naïve Bayesian-based model for ranking and classifying a wide range of bone tumor diagnoses. Yu et al. (30) have established a DL algorithm to classify bone tumors in terms of aggressiveness on plain radiographs, finding the model has the ROC curve AUC of 0.877 for binary classification (benign *vs.* non-benign) and the CKS of 0.560 for ternary classification on testing dataset. However, the radiological information is relatively limited for AI models to train and learn because only several radiographic images can be obtained from one patient diagnosed with the bone tumor. The bone tumor's morphological information presented on the radiograph can be inconsistent due to the variabilities of radiation intensity, patient position, and film magnification. Therefore, DL models may not grasp sufficient discriminative features only from limited radiographs from one patient, whereas a much more sample size per class is needed to control the overfitting for a DL model with more than hundreds of thousands of parameters (31). In contrast, the current study used WSI to scan almost all cell-level image data from the pathological slide, which was then exported as hundreds of thousands of image patches for training the DL model. Six selected models showed satisfactory learning curves, but models' performances in differentiating the bone tumor's pathological data are different from that in distinguishing the ImageNet dataset, where the VGG-16 and Inception V3 showed better results in our dataset. The best models trained with more than 400,000 histopathological image patches in this study also showed relatively higher patient-level (slide-level) predictive capabilities in binary and ternary classification for bone tumors compared with that of models trained with limited radiological data (30), reaching the diagnostic level of senior pathologists while outperforming attending and junior pathologists.

The tumor area annotated by pathologists instead of the whole tissue area on the pathological slide was used as ROI for patch extraction in this study because of the following three reasons. First, the histopathological component of bone tumors is usually mixed with various normal connective tissues (bone, cartilage, vascular tissue, fibrous tissue, and muscular tissue), which would introduce a lot of noise data for DL models if the normal area is used as training input. Second, the atypical tumor areas that exclude normal connective tissue areas are relatively easy for general pathologists to identify. Third, given the huge morphological heterogeneity (various origins, such as osteogenic, chondrogenic, and fibrous) among bone tumors from each qualitative class and the moderate slide sample size of this study, it is impractical to automate the ROI selection process

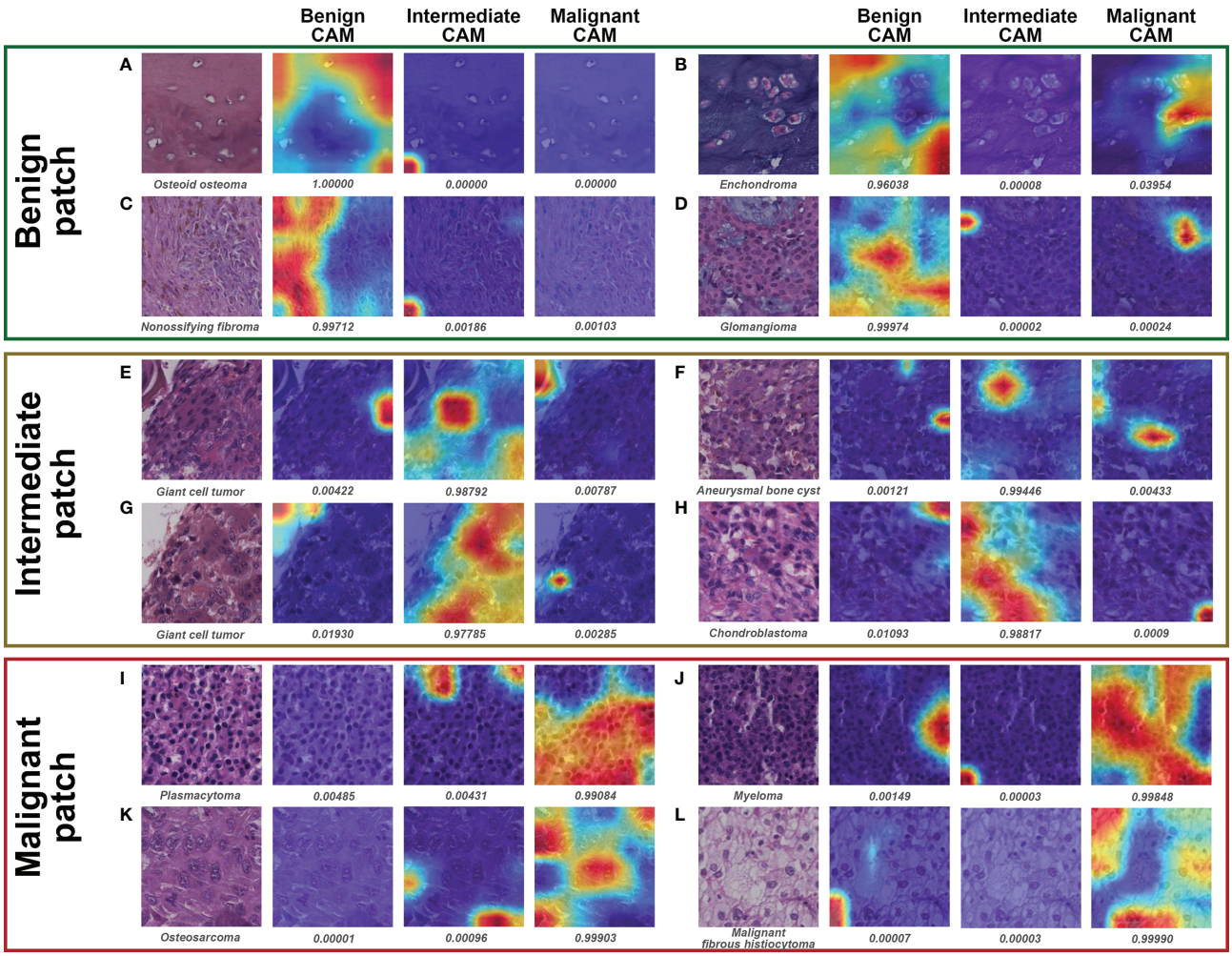


FIGURE 7 | Grad-CAM in ternary classification for representative patches of slides correctly classified by both models and pathologist #1. The specific classification is shown under the original patch, and the predictive probability for CAM of each class is shown below the corresponding Grad-CAM heatmap. **(A–D)**, **(E–H)**, and **(I–L)** show the representative patches of benign, intermediate, and malignant bone tumors, respectively.

using algorithms for the moment. Some slides with suboptimal stain quality were excluded before WSI scanning in this study, making the pipeline workflow not clinically applicable enough. An updated algorithm including stain normalization and defective slide detection should be integrated into the model in the future.

The patch-level Grad-CAM visualization showed that the DL model could overcome the interference of histomorphological heterogeneity among the same class, well-differentiating bone tumors in terms of aggressiveness according to the diagnostic feature. We speculate that the under-differentiated and non-specific abstract features in non-benign bone tumors could be effectively extracted and learned by the DL model to make the correct patch-level binary prediction. However, for malignant bone tumors with atypical cell components accounting for a small proportion of the whole slide, the DL model gave wrong slide-level predictions because noise patches (patches of benign structure) unintentionally produced in the ROI selection process was used for slide-level probability calculation. Such label noise is inevitable when the pathologist annotates an ROI area, and many weakly supervised approaches have been attempted to address this issue and reduce annotation workload (10, 32). Therefore, given the histopathological diversity for bone tumors of the same qualitative classification, future studies with more sample sizes that have numerous cases of different origins in each class are needed to build an annotation-free DL classifier with high performance.

Most of the current DL-based histopathological diagnosis system has been built as the assistant role for human pathologists (11, 33) because of the related ethical issues of entirely relying on DL models (8). It is usually devastating for pathologists to miss the diagnosis of a non-benign bone tumor in adolescents that

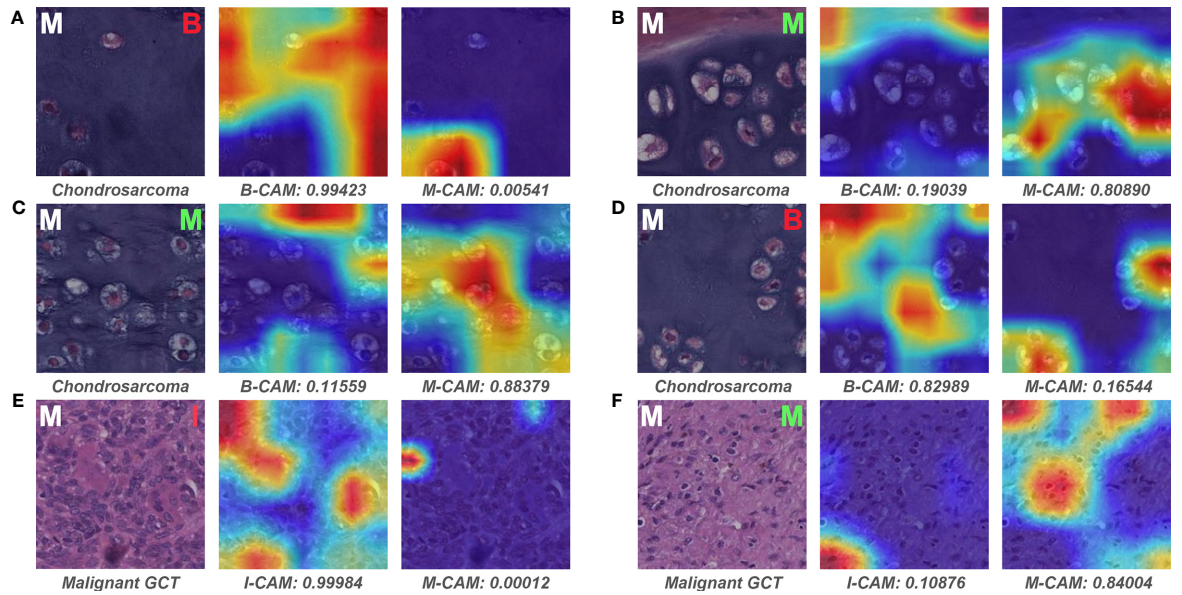


FIGURE 8 | Representative patches of slides wrongly classified by models and the associated Grad-CAM results. The ground truth label of the original patch is displayed on the upper left in white, and the predictive classification of the model is presented on the upper right in red (false prediction) or green (correct prediction). The specific classification is shown under the original patch, and the predictive probability for CAM of each class is shown below the corresponding Grad-CAM heatmap. **(A–F)** show the representative patches of chondrosarcoma and malignant giant cell tumor, respectively. B, benign; I, intermediate; M, malignant; CAM, class activation mapping; GCT, giant cell tumor.

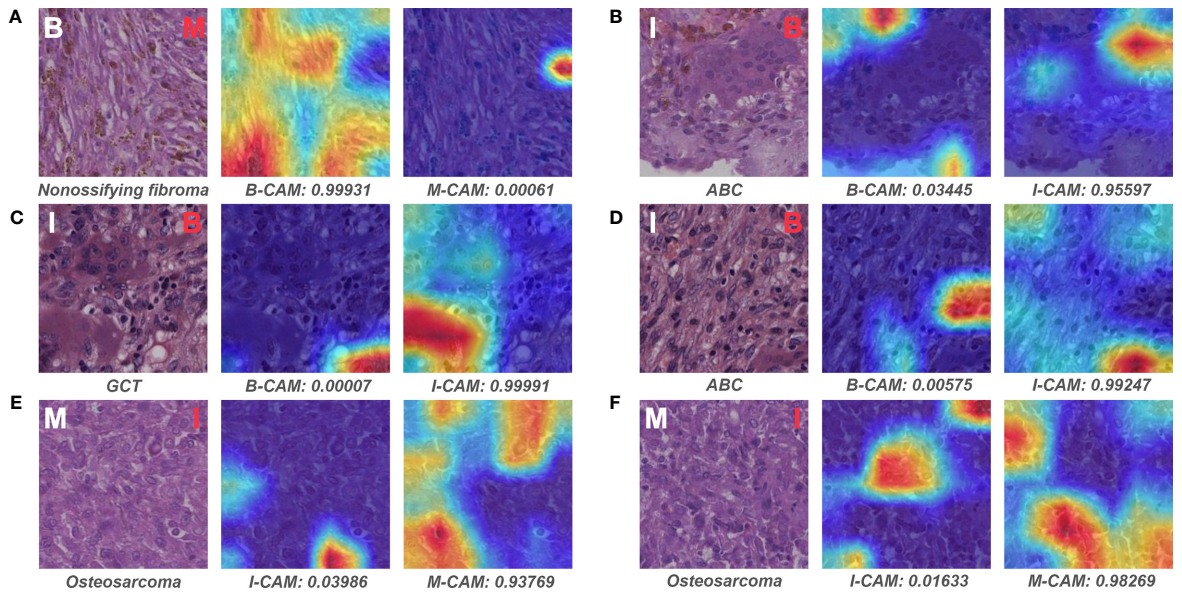


FIGURE 9 | Representative patches of slides wrongly classified by pathologist #1 and the associated Grad-CAM results. The ground truth label of the original patch is displayed on the upper left in white, and the predictive classification of pathologist #1 is presented on the upper right in red (false prediction). The specific classification is shown under the original patch, and the predictive probability for CAM of each class is shown below the corresponding Grad-CAM heatmap. **(A–F)** show the representative patches of benign, intermediate, and malignant bone tumors, respectively. B, benign; I, intermediate; M, malignant; CAM, class activation mapping; ABC, aneurysmal bone cyst; GCT, giant cell tumor.

could have been properly managed. Therefore, a screening tool with high sensitivity would assist inexperienced pathologists in general hospitals to confidently exclude non-benign bone tumors and refer suspected aggressive cases to specialized hospitals for further treatment. The best-performing DL model in this study showed a comparable sensitivity and a higher specificity compared with the senior pathologist in slide-level prediction, indicating the promising value of DL in screening non-benign bone tumors in the future. Besides histopathological features, pathologists typically use radiological and demographic findings as references to reach the final clinical diagnosis of the bone tumor. However, when given the pathological information alone, the evaluation results among pathologists seemed not consistent in this study, which shows that the human's classification of bone tumors may be unreliable solely based on the histopathological assessment. Later DL-related research should focus on combining clinical, radiological, and histopathological data of bone tumors, along with cutting-edge approaches like the ensemble model (34), to raise the sensitivity to near 100% while maintaining the high specificity of the model.

To our knowledge, this is the first study that verifies the feasibility of using the DL-based model to classify bone tumors histopathologically in terms of aggressiveness. In contrast, previous related works only concentrated on the histologic analysis of specific diagnoses of bone tumors and had small numbers of WSI slides (35–37). Considering the low prevalence of bone tumors and the relative difficulty to obtain biopsy tissues compared with radiographs, the sample size of more than 700,000 patches generated from 427 slides was fairly adequate to train a DL model. With the help of the Grad-CAM, we found that the model could easily differentiate some cases that were confusing for pathologists. The visualization also helped us partly interpret the DL underlying mechanism of classifying bone tumors, which was deemed a black box that was hard to explain before. These results would provide a theoretical basis for the future application of DL-assisted histopathological diagnosis for bone tumors.

There exist several limitations in our study. First, this is a single-center study with a moderate number of pathological slides. The variety in the process of slide preparation and WSI scanning from different institutions may have an impact on the image quality of training input. Therefore, the model's generalizability might be partially limited by the training dataset of this study, and a multi-center research should assist in achieving a more robust result in the future. Second, the label noise (wrongly labeled patches) generated from manually annotated ROIs would introduce information bias to some degree for DL training, although such bias could be mostly compensated in the averaging process of the slide-level prediction. Third, due to the retrospective nature of the data acquisition, the number of slides in each classification was not well balanced, thus bringing in selection bias for training and evaluation of the model. However, we used the average metric weighted by each class to minimize this kind of bias. Moreover, the subjectivity of pathologists in determining tumor areas would also result in selection bias, and future studies are needed to address this problem with weakly supervised or unsupervised DL models. Fourth, there were few specific cases with the rare incidence in each qualitative classification (for example, fibrous histiocytoma and fibrosarcoma). The DL model may not be trained well to extract and learn morphological features specific to these rare cases based on the limited number of representative patches. Future studies should include more data about rare cases to make the model more generalizable. Fifth, there exist some non-neoplastic lesions mimicking bone tumors radiographically or histopathologically, such as osteomyelitis and osteonecrosis. Our DL models were only trained on the neoplastic lesions, leading to inapplicability to differentiate such non-neoplastic lesions, although these kinds of tumor mimics are relatively easy for pathologists to distinguish from neoplastic lesions based on the laboratory test and the histological absence of neoplastic cells. Finally, histopathological results of bone tumors are more often combined with clinical and radiological features of patients for pathologists to predict the clinical classification, whereas we only focused on the histopathological side in this study. In order to make the model's underlying prediction mechanism closer to the human being, later research should consider integrating multiple levels of data to train a comprehensive DL model.

In summary, the present study shows that the DL model can effectively classify primary bone tumors histopathologically in terms of aggressiveness, reaching the predictive performance similar to the senior pathologist while higher than attending and resident pathologists. These results are promising and would help expedite the future application of DL-assisted histopathological diagnosis for primary bone tumors.

DATA AVAILABILITY STATEMENT

The original data 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 human participants were reviewed and approved by the Ethics Committee of the First Affiliated Hospital of Chongqing Medical University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

YuT, XH, and JZ conceived the study. YuT, XH, CL, and JZ contributed to developing the methodology. YuT, HW, WJ, YCh, WY, MG, BT, and MY performed the data collection. CL developed the algorithms for deep learning training. CL, ZL, and KG provided the experimental and computational resources. YiT, CW, JL, and

YCa performed the histopathological evaluation. YuT, XH, and CL analyzed the data. TC, AZ, YCa, and JZ supervised the execution of the study. YuT wrote the original draft. CL, JZ, and TC edited and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Chongqing Science and Health Joint Medical Research Project from the Chongqing Municipal Health Commission [No. 2020MSXM002]; the Chongqing Postgraduate Research and Innovation Project from the Chongqing Municipal Education Commission [No. CYB20142]; and the General Project of Technology Innovation and Application Development from the Chongqing Science and Technology Commission [No. cstc2019jscx-msxm0156]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

The authors would like to thank Prof. Bin Peng for his support of statistical analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.735739/full#supplementary-material

REFERENCES

1. Santini-Araujo E, Kalil RK, Bertoni F, Park Y-K. *Tumors and Tumor-Like Lesions of Bone*. London, UK: Springer Nature (2020).
2. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. *CA Cancer J Clin* (2021) 71:7–33. doi: 10.3322/caac.21654
3. Mangham DC, Athanasou NA. Guidelines for Histopathological Specimen Examination and Diagnostic Reporting of Primary Bone Tumours. *Clin Sarcoma Res* (2011) 1:1–13. doi: 10.1186/2045-3329-1-6
4. Fletcher CDM, Bridge JA, Hogendoorn PCW, Mertens FWorld Health Organization and International Agency for Research on Cancer. *WHO Classif Tumours Soft Tissue Bone*. Lyon: IARC Press (2013).
5. Franchi A. Epidemiology and Classification of Bone Tumors. *Clin Cases Miner Bone Metab* (2012) 9:92–5.
6. Salto-Tellez M, Maxwell P, Hamilton P. Artificial Intelligence-the Third Revolution in Pathology. *Histopathology* (2019) 74:372–6. doi: 10.1111/his.13760
7. Albawi S, Mohammed TA, Al-Zawi S. Understanding of a Convolutional Neural Network. In: *2017 International Conference on Engineering and Technology (ICET)*. IEEE (2017). p. 1–6.
8. Niazi MKK, Parwani AV, Gurcan MN. Digital Pathology and Artificial Intelligence. *Lancet Oncol* (2019) 20:e253–61. doi: 10.1016/S1470-2045(19)30154-8
9. Duggento A, Conti A, Mauriello A, Guerrisi M, Toschi N. Deep Computational Pathology in Breast Cancer. *Semin Cancer Biol* (2020) 72:226–37. doi: 10.1016/j.semcancer.2020.08.006
10. Chen CL, Chen CC, Yu WH, Chen SH, Chang YC, Hsu TI, et al. An Annotation-Free Whole-Slide Training Approach to Pathological Classification of Lung Cancer Types Using Deep Learning. *Nat Commun* (2021) 12:1193. doi: 10.1038/s41467-021-21467-y
11. Song Z, Zou S, Zhou W, Huang Y, Shao L, Yuan J, et al. Clinically Applicable Histopathological Diagnosis System for Gastric Cancer Detection Using Deep Learning. *Nat Commun* (2020) 11:4294. doi: 10.1038/s41467-020-18147-8
12. Kott O, Linsley D, Amin A, Karagounis A, Jeffers C, Golijanin D, et al. Development of a Deep Learning Algorithm for the Histopathologic Diagnosis and Gleason Grading of Prostate Cancer Biopsies: A Pilot Study. *Eur Urol Focus* (2021) 7:347–51. doi: 10.1016/j.euf.2019.11.003
13. Chuang WY, Chang SH, Yu WH, Yang CK, Yeh CJ, Ueng SH, et al. Successful Identification of Nasopharyngeal Carcinoma in Nasopharyngeal Biopsies Using Deep Learning. *Cancers (Basel)* (2020) 12:507. doi: 10.3390/cancers12020507
14. Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, et al. QuPath: Open Source Software for Digital Pathology Image Analysis. *Sci Rep* (2017) 7:1–7. doi: 10.1038/s41598-017-17204-5
15. Hou L, Samaras D, Kurc TM, Gao Y, Davis JE, Saltz JH. Patch-Based Convolutional Neural Network for Whole Slide Tissue Image Classification. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE (2016). p. 2424–33.
16. Dimitriou N, Arandjelović O, Caie PD. Deep Learning for Whole Slide Image Analysis: An Overview. *Front Med* (2019) 6:264. doi: 10.3389/fmed.2019.00264
17. Krizhevsky A, Sutskever I, Hinton GE. Imagenet Classification With Deep Convolutional Neural Networks. *Adv Neural Inf Process Syst* (2012) 25:1097–105. doi: 10.1145/3065386
18. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv [Preprint]* (2014). Available at: https://arxiv.org/abs/1409.1556
19. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE (2016). p. 2818–26.
20. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE (2017). p. 4700–8.
21. 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*. IEEE (2016). p. 770–8.
22. Tan M, Chen B, Pang R, Vasudevan V, Sandler M, Howard A, et al. Mnasnet: Platform-Aware Neural Architecture Search for Mobile. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE (2019). p. 2820–8.
23. Smith LN. Cyclical Learning Rates for Training Neural Networks. In: *2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE (2017). p. 464–72.
24. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-Cam: Visual Explanations From Deep Networks *via* Gradient-Based Localization. In: *Proceedings of the IEEE International Conference on Computer Vision*. IEEE (2017). p. 618–26.
25. 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
26. Vanbelle S, Albert A. A Bootstrap Method for Comparing Correlated Kappa Coefficients. *J Stat Comput Simul* (2008) 78:1009–15. doi: 10.1080/00949650701410249
27. Li MD, Ahmed SR, Choy E, Lozano-Calderon SA, Kalpathy-Cramer J, Chang CY. Artificial Intelligence Applied to Musculoskeletal Oncology: A Systematic Review. *Skeletal Radiol* (2021). doi: 10.1007/s00256-021-03820-w
28. Vogrin M, Trojner T, Kelc R. Artificial Intelligence in Musculoskeletal Oncological Radiology. *Radiol Oncol* (2020) 55:1–6. doi: 10.2478/raon-2020-0068
29. Do BH, Langlotz C, Beaulieu CF. Bone Tumor Diagnosis Using a Naïve Bayesian Model of Demographic and Radiographic Features. *J Digit Imaging* (2017) 30:640–7. doi: 10.1007/s10278-017-0001-7

30. He Y, Pan I, Bao B, Halsey K, Chang M, Liu H, et al. Deep Learning-Based Classification of Primary Bone Tumors on Radiographs: A Preliminary Study. *EBioMedicine* (2020) 62:103121. doi: 10.1016/j.ebiom.2020.103121
31. Benkendorf DJ, Hawkins CP. Effects of Sample Size and Network Depth on a Deep Learning Approach to Species Distribution Modeling. *Ecol Inform* (2020) 60:101137. doi: 10.1016/j.ecoinf.2020.101137
32. 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
33. Wang S, Yang DM, Rong R, Zhan X, Fujimoto J, Liu H, et al. Artificial Intelligence in Lung Cancer Pathology Image Analysis. *Cancers (Basel)* (2019) 11:1–16. doi: 10.3390/cancers11111673
34. Salvi M, Molinari F, Iussich S, Muscatello LV, Pazzini L, Benali S, et al. Histopathological Classification of Canine Cutaneous Round Cell Tumors Using Deep Learning: A Multi-Center Study. *Front Vet Sci* (2021) 8:640944. doi: 10.3389/fvets.2021.640944
35. Mishra R, Daescu O, Leavey P, Rakheja D, Sengupta A. Convolutional Neural Network for Histopathological Analysis of Osteosarcoma. *J Comput Biol* (2018) 25:313–25. doi: 10.1089/cmb.2017.0153
36. Arunachalam HB, Mishra R, Daescu O, Cederberg K, Rakheja D, Sengupta A, et al. Viable and Necrotic Tumor Assessment From Whole Slide Images of Osteosarcoma Using Machine-Learning and Deep-Learning Models. *PloS One* (2019) 14:1–19. doi: 10.1371/journal.pone.0210706
37. Fu Y, Xue P, Ji H, Cui W, Dong E. Deep Model With Siamese Network for Viable and Necrotic Tumor Regions Assessment in Osteosarcoma. *Med Phys* (2020) 47:4895–905. doi: 10.1002/mp.14397

Conflict of Interest: Authors ZL and KG are employed by Chongqing Defang Information Technology Co., Ltd.

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

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

frontiers in Oncology

ORIGINAL RESEARCH
published: 14 October 2021
doi: 10.3389/fonc.2021.759007

Predicting Axillary Lymph Node Metastasis in Early Breast Cancer Using Deep Learning on Primary Tumor Biopsy Slides

Feng Xu[1†], Chuang Zhu[2*†], Wenqi Tang[2†], Ying Wang[3†], Yu Zhang[2], Jie Li[1], Hongchuan Jiang[1], Zhongyue Shi[3], Jun Liu[2] and Mulan Jin[3*]*

[1] Department of Breast Surgery, Beijing Chao-Yang Hospital, Beijing, China, [2] School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing, China, [3] Department of Pathology, Beijing Chao-Yang Hospital, Beijing, China

Objectives: To develop and validate a deep learning (DL)-based primary tumor biopsy signature for predicting axillary lymph node (ALN) metastasis preoperatively in early breast cancer (EBC) patients with clinically negative ALN.

Methods: A total of 1,058 EBC patients with pathologically confirmed ALN status were enrolled from May 2010 to August 2020. A DL core-needle biopsy (DL-CNB) model was built on the attention-based multiple instance-learning (AMIL) framework to predict ALN status utilizing the DL features, which were extracted from the cancer areas of digitized whole-slide images (WSIs) of breast CNB specimens annotated by two pathologists. Accuracy, sensitivity, specificity, receiver operating characteristic (ROC) curves, and areas under the ROC curve (AUCs) were analyzed to evaluate our model.

Results: The best-performing DL-CNB model with VGG16_BN as the feature extractor achieved an AUC of 0.816 (95% confidence interval (CI): 0.758, 0.865) in predicting positive ALN metastasis in the independent test cohort. Furthermore, our model incorporating the clinical data, which was called DL-CNB+C, yielded the best accuracy of 0.831 (95%CI: 0.775, 0.878), especially for patients younger than 50 years (AUC: 0.918, 95%CI: 0.825, 0.971). The interpretation of DL-CNB model showed that the top signatures most predictive of ALN metastasis were characterized by the nucleus features including density (p = 0.015), circumference (p = 0.009), circularity (p = 0.010), and orientation (p = 0.012).

Conclusion: Our study provides a novel DL-based biomarker on primary tumor CNB slides to predict the metastatic status of ALN preoperatively for patients with EBC.

Keywords: deep learning, axillary lymph node metastasis, breast cancer, core-needle biopsy, whole-slide images

OPEN ACCESS

Edited by:
Jialiang Yang,
Geneis (Beijing) Co. Ltd, China

Reviewed by:
Shuhao Wang,
Tsinghua University, China
Zeyu Xing,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

***Correspondence:**
Feng Xu
drxufeng@mail.ccmu.edu.cn
Chuang Zhu
czhu@bupt.edu.cn
Mulan Jin
kinmokuran@163.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *15 August 2021*
Accepted: *21 September 2021*
Published: *14 October 2021*

Citation:
Xu F, Zhu C, Tang W, Wang Y, Zhang Y, Li J, Jiang H, Shi Z, Liu J and Jin M (2021) Predicting Axillary Lymph Node Metastasis in Early Breast Cancer Using Deep Learning on Primary Tumor Biopsy Slides. Front. Oncol. 11:759007. doi: 10.3389/fonc.2021.759007

INTRODUCTION

Breast cancer (BC) has become the greatest threat to women's health worldwide (1). Clinically, identification of axillary lymph node (ALN) metastasis is important for evaluating the prognosis and guiding the treatment for BC patients (2). Sentinel lymph node biopsy (SLNB) has gradually replaced ALN dissection (ALND) to identify ALN status, especially for early BC (EBC) patients with clinically negative lymph nodes. Although SLNB had the advantage of less invasiveness than ALND, SLNB still caused some complications such as lymphedema, axillary seroma, paraesthesia, and impaired shoulder function (3, 4). Moreover, SLNB has been considered a controversial procedure, owing to the availability of radionuclide tracers and the surgeon's experience (5, 6). In fact, SLNB can be avoided if there are some reliable methods of preoperative prediction of ALN status for EBC patients.

Several studies intended to predict the ALN status by clinicopathological data and genetic testing score (7, 8). However, due to the relatively poor predictive values and high genetic testing costs, these methods are often limited. Recently, deep learning (DL) can perform high-throughput feature extraction on medical images and analyze the correlation between primary tumor features and ALN metastasis information. In a previous study, deep features extracted from conventional ultrasound and shear wave elastography (SWE) were used to predict ALN metastasis, presenting an area under the curve (AUC) of 0.796 in the test set (9). Nevertheless, SWE has not been integrated into routine clinical breast examinations in many hospitals. Another recent study demonstrated that the DL model based on diffusion-weighted imaging–magnetic resonance imaging (DWI-MRI) database of 172 patients achieved an AUC of 0.852 for preoperative prediction of ALN metastasis (10), but the small sample size enrolled could not be representative.

Currently, DL has enabled rapid advances in computational pathology (11, 12). For example, DL methods have been applied to segment and classify glomeruli with different staining and various pathologic changes, thus achieving the automatic analysis of renal biopsies (13, 14); meanwhile, DL-based automatic colonoscopy tissue segmentation and classification have shown promise for colorectal cancer detection (15, 16); besides, the analysis of gastric carcinoma and precancerous status can also benefit from DL schemes (17, 18). More recently, for the ALN metastasis detection, it is reported that DL algorithms on digital lymph node pathology images achieved better diagnostic efficiency of ALN metastasis than pathologists (19, 20). In particular, the assistance of algorithm significantly increases the sensitivity of detection for ALN micro-metastases (21). In addition to diagnosis, several previous studies indicated that deep features based on whole-slide images (WSIs) of postoperative tumor samples potentially improved the prediction performance of lymph node metastasis in a variety of cancers (20, 22). So far, there is no relevant research on preoperatively predicting ALN metastasis based on WSIs of primary BC samples. In this study, we investigated a clinical data set of EBC patients treated by preoperative core-needle biopsy (CNB) to determine whether DL models based on primary tumor biopsy slides could help to refine the prediction of ALN metastasis.

PATIENTS AND METHODS

Patients

On approval by the Institutional Ethical Committees of Beijing Chaoyang Hospital affiliated to Capital Medical University, we retrospectively analyzed data from EBC patients with clinically negative ALN from May 2010 to August 2020. Written consent was obtained from all patients and their families.

The detailed inclusion criteria were as follows: 1) patients with CNB pathologically confirmed primary invasive BC; 2) patients who underwent breast surgery with SLNB or ALND; 3) baseline clinicopathological data including age, tumor size, tumor type, ER/PR/HER-2 status, and the number of ALN metastasis were comprehensive; 4) complete concordance of molecular status was found between CNB and excision specimens; 5) no history of preoperative radiotherapy and chemotherapy; and 6) adequate volume of biopsy materials with three or more cores for each patient.

The exclusion criteria included the following: 1) patients with physically positive or imaging-positive ALN; 2) missing postoperative pathology information; 3) missing wax blocks and hematoxylin and eosin (H&E) slices; and 4) low-quality H&E slices or WSIs. The patient recruitment workflow is shown in **Figure 1**.

Deep Learning Model Development

To avoid the inter-observer heterogeneity, all available tumor regions in each CNB slide were examined and annotated by two independent and experienced pathologists blinded to all patient-related information. A WSI was classified into positive (N(+)) or negative (N0) using the proposed DL CNB (DL-CNB) model. Our DL-CNB model was constructed with the attention-based multiple-instance learning (MIL) approach (23). In MIL, each training sample was called a bag, which consisted of multiple instances (24–26) (each instance corresponds to an image patch of size 256 × 256 pixels). Different from the general fully supervised problem where each sample had a label, only the label of bags was available in MIL, and the goal of MIL was to predict the bag label by considering all included instances comprehensively. The whole algorithm pipeline comprised the following five steps:

(1) Training data preparation (**Figure 2A**). For each raw WSI, amounts of non-overlapping square patches were first cropped from the selected tumor regions. Then each WSI could be represented as a bag with N randomly selected patches. To increase the training samples, M bags were built for each WSI. All M bags were labeled as positive if the slide is an ALN metastasis case, and vice versa. Note that we could add the clinical information of the slide to all the M constructed bags to involve more useful information for predicting, and in this situation, the developed model was called DL-CNB+C.

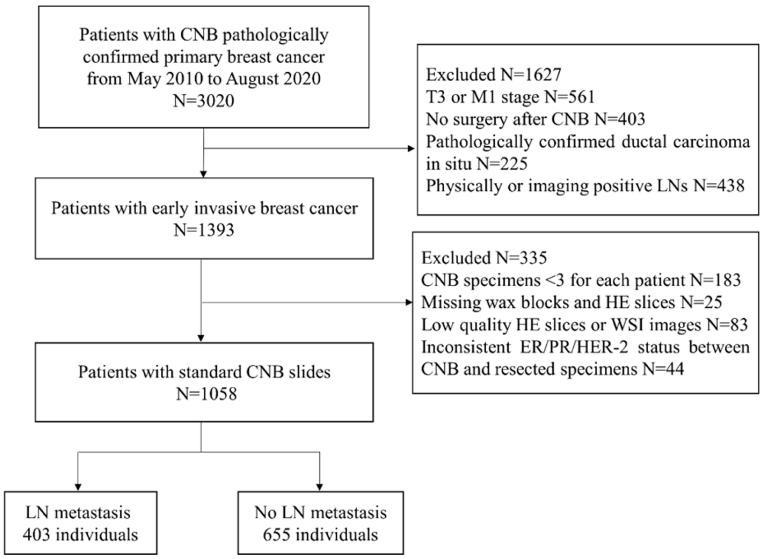


FIGURE 1 | Patient recruitment workflow.

(2) Feature extraction (left part of **Figure 2B**). *N* feature vectors were extracted for the *N* image instances in each bag by using a convolutional neural network (CNN) model. The performances of AlexNet (27), VGG16 (28) with batch norm (VGG16_BN), ResNet50 (29), DenseNet121 (30), and Inception-v3 (31) were compared to find the best feature extractor. At this stage, the clinical data were also preprocessed for feature extraction. Concretely, the numerical properties in clinical data were standardizing by removing the mean and scaling to unit variance, thus eliminating the effect of data range and scale; furthermore, considering that there was no natural ordinal relationship between different values of the category attributes, the categorical properties in clinical data were encoded as the one-hot vectors, which could express different values equally.

(3) MIL (right part of **Figure 2B**). The extracted *N* feature vectors of image instances were first processed by the max-pooling (32–34) and reshaping and then were passed to a two-layer fully connected (FC) layer. The *N* weight factors for the instances in the bag were thus obtained and then were further multiplied to the original feature vectors (23) to adaptively adjust the effect of instance features. Finally, the weighted image feature vectors and the clinical features were fused by concatenation; due to the large difference of dimensions between image features and clinical features, the clinical features were copied 10 times for expansion. Then, the fused features were fed into the classifier, and the outputs and the ground truth labels were used to calculate the cross-entropy loss.

(4) Model training and testing. We randomly divided the WSIs into training cohort and independent test cohort with the ratio of 4:1 and randomly selected 25% of the training cohort as the validation cohort. We used Adam optimizer with learning rate 1e−4 to update the model parameters and weight decay 1e−3 for regularization. In the training phase, we used the cosine annealing warm restarts strategy to adjust the learning rate (35). In the testing phase, the ALN status is predicted by aggregating the model outputs of all bags from the same slide (**Figure 2C**).

The DL models are available at: https://github.com/bupt-ai-cz/BALNMP.

Visualization of Salient Regions From Deep Learning Core-Needle Biopsy Model

We visualized the important regions that were more associated with metastatic status. After the processing of attention-based MIL pooling, the weights of different patches can be obtained, and the corresponding feature maps were then weighted together in the following FC layers to conduct ALN status prediction. With the attention weights, we created a heat map to visualize the important salient regions in each WSI.

Interpretability of Deep Learning Core-Needle Biopsy Model With Nucleus Features

Interpretability of DL-CNB model with nucleus features was performed to study the contribution of different nucleus morphological characteristics in the prediction of lymph node metastasis (36, 37). Multiple specially designed nucleus features were firstly extracted for each WSI, and these features together formed a training bag. With the constructed feature bags, the proposed DL-CNB model was re-trained. The weights of different features (instances) can be obtained based on the attention-based MIL pooling, and thus the contribution of different features was yielded. The specific process is described in **Figure 3**.

Statistical Analysis

The logistic regression was used to predict ALN status by clinical data only model. The clinical difference of N0 and N(+) was compared by using the Mann–Whitney U test and chi-square

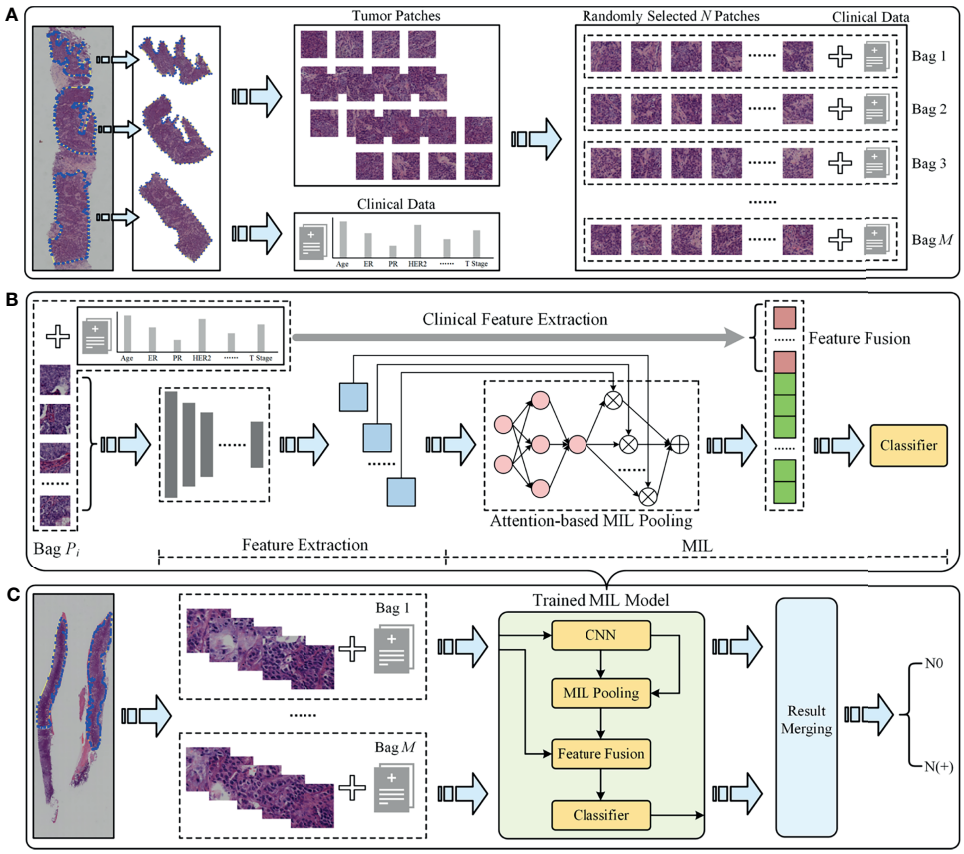


FIGURE 2 | The overall pipeline of the deep learning core-needle biopsy incorporating the clinical data (DL-CNB+C) model to predict axillary lymph node (ALN) status between N0 and N(+). **(A)** Multiple training bags were built based on clinical data and the cropped patches from the selected tumor regions of each core-needle biopsy (CNB) whole-slide image (WSI). **(B)** DL-CNB+C model training process included two phases of feature extraction and multiple-instance learning (MIL), and finally the weighted features fused with clinical features were used to predict classification probabilities and calculate the cross-entropy loss. **(C)** The predicted probabilities of each bag from a raw CNB WSI were merged to guide the final ALN status classification between N0 and N(+).

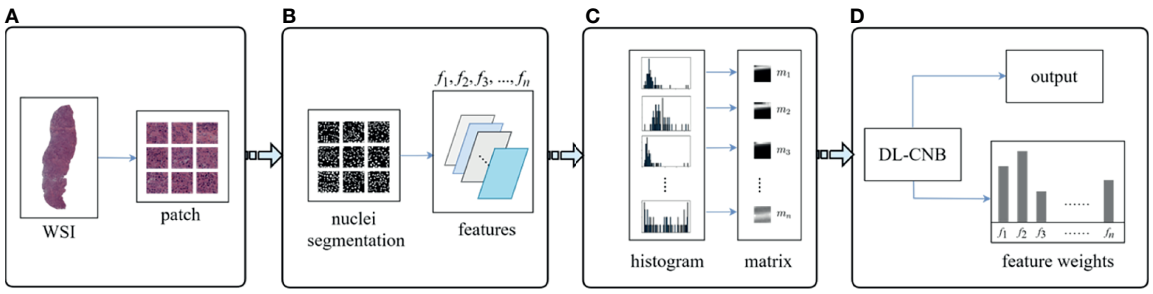


FIGURE 3 | Overview on interpretability methods of deep learning core-needle biopsy (DL-CNB) model based on nucleus morphometric features. **(A)** The selected tumor regions of each whole-slide image (WSI) was cropped into patches. **(B)** For each patch, we processed nucleus segmentation (a weakly supervised segmentation framework was applied to obtain the nucleus), defined multiple nucleus morphometric features (such as major axis, minor axis, area, orientation, circumference, density, circularity, and rectangularity, which are denoted as f_1, f_2, f_3, …, f_n), and extracted n feature parameters correspondingly. **(C)** All n kinds of feature parameters from a WSI were quantized into n distribution histograms and saved to n feature matrices (m_1, m_2, m_3, …, m_n). **(D)** The matrices from a WSI were considered as instances of a bag and served as the input of DL-CNB model; the re-trained DL-CNB model could generate scores of features (instances) in the bag, which represented the weight of each feature in pathological diagnosis.

test. The AUCs of different methods were compared by using Delong et al. (38). The other measurements like accuracy (ACC), sensitivity (SENS), specificity (SPEC), positive predictive value (PPV), and negative predictive value (NPV) were also used to estimate the model performance. All the statistics were two-sided, and a *p*-value less than 0.05 was considered statistically significant. All statistical analyses were performed by MedCalc software (V 19.6.1; 2020 MedCalc Software bvba, Mariakerke, Belgium), Python 3.7, and SPSS 24.0 (IBM, Armonk, NY, USA).

RESULTS

Clinical Characteristics

A total of 1,058 patients with EBC were enrolled for analysis. Among them, 957 (90.5%) patients had invasive ductal carcinomas, and 101 (9.5%) patients had invasive lobular carcinomas. There were 840 patients in the training cohort and 218 patients in the independent test cohort after all WSIs were randomly divided by using N0 as the negative reference standard and others as the positive. The average patient age was 57.6 years (range, 26–90 years) for the training and validation sets and 56.7 years (range, 22–87 years) for the test set. The mean ultrasound tumor size was 2.23 cm (range, 0.5–4.5 cm). A total of 556 patients (52.6%) had T1 tumors, while 502 patients (47.4%) had T2 tumors. According to the results of SLNB or ALND, positive lymph nodes were found in 403 patients. Among them, 210 patients (52.1%) had one or two positive lymph nodes ($N_+(1-2)$), and 193 patients (47.9%) had three or more positive lymph nodes ($N_+(\geq 3)$). As shown in **Table 1**, there was no significant difference between the detailed characteristics of the training and independent test cohorts (all $p > 0.05$).

Convolutional Neural Network Model Selection

The detailed results are summarized in **Supplementary Table 1**. Based on the overall analysis, VGG16_BN model pre-trained on ImageNet (39) provided the best performance in the validation cohort and the independent test cohort (AUC: 0.808, 0.816), compared with AlexNet (AUC: 0.764, 0.780), ResNet50 (AUC: 0.644, 0.607), DenseNet121 (AUC: 0.714, 0.739), and Inception-v3 (AUC: 0.753, 0.762). Furthermore, considering other metrics, VGG16_BN achieved the best ACC, SPEC, and PPV in the independent test cohort. VGG16_BN consisted of (convolution layer, batch normalization layer, and Rectified Linear Unit (ReLU)) as the basic block where ReLU played a role of activation function to provide the non-linear capability; and max-pooling layers were inserted between basic blocks for down-sampling; besides, there was an adaptive average pooling layer at the end of VGG16_BN for obtaining features with a fixed size. The details of VGG16_BN are described in **Supplementary Table 2**.

Predictive Value of Deep Learning Core-Needle Biopsy Incorporating the Clinical Data Model Between N0 and N(+)

In the training cohort, DL-CNB+C achieved an AUC of 0.878, while DL-CNB and classification by clinical data only model achieved AUCs of 0.901 and 0.661, respectively. And in the validation cohort, the DL-CNB+C model achieved an AUC of 0.823, which was higher than an AUC of 0.808 obtained by DL-CNB only and an AUC of 0.709 obtained by classification by clinical data.

In the independent test cohort, the DL-CNB+C model still achieved the highest AUC of 0.831, which was better than the AUC of DL-CNB only (AUC: 0.816, $p = 0.453$) and classification

TABLE 1 | Patient and tumor characteristics.

Characteristics		All patients	Training	Test	*p*
Number		1,058	840 (80%)	218 (20%)	
Age, mean ± SD, years		57.58 ± 12.523	57.80 ± 12.481	56.72 ± 12.674	0.344
Tumor size, mean ± SD, cm		2.234 ± 0.8623	2.228 ± 0.8516	2.256 ± 0.9040	0.898
Number of LNM, mean ± SD		1.20 ± 2.081	1.20 ± 2.095	1.20 ± 2.033	0.847
Tumor type					0.812
	Invasive ductal carcinoma	957	760 (90.5%)	197 (90.4%)	
	Invasive lobular carcinoma	101	80 (9.5%)	21 (9.6%)	
T stage					0.327
	T1	556	435 (51.8%)	121 (55.5%)	
	T2	502	405 (48.2%)	97 (44.5%)	
ER					0.333
	Positive	831	665 (79.2%)	166 (76.1%)	
	Negative	227	175 (20.8%)	52 (23.9%)	
PR					0.312
	Positive	790	633 (75.4%)	157 (72.0%)	
	Negative	268	207 (24.6%)	61 (28.0%)	
HER-2					0.613
	Positive	277	217 (25.8%)	60 (27.5%)	
	Negative	781	623 (74.2%)	158 (72.5%)	
LNM					0.880
	Yes	403	521 (62.0%)	134 (61.5%)	
	No	655	319 (38.0%)	84 (38.5%)	

Qualitative variables are in n (%), and quantitative variables are in mean ± SD, when appropriate.
SD, standard deviation; ER, estrogen receptor; PR, progesterone receptor; HER-2, human epidermal growth factor receptor-2; LNM, lymph node metastasis.

by clinical data only (AUC: 0.613, $p < 0.0001$). The ACC, SENS, and NPV of DL-CNB+C were also better than those of other methods. The detailed statistical results are summarized in **Table 2**, and its corresponding receiver operating characteristics (ROCs) are shown in **Figure 4**.

We further divided N(+) into low metastatic potential (N_+ (1 – 2)) and high metastatic potential ($N_+(\geq 3)$) according to the number of ALN metastasis. Adopting N0 as the negative reference standard, the combined model showed better discriminating ability between N0 and $N_+(1 - 2)$ (AUC: 0.878) and between N0 and $N_+(\geq 3)$ (AUC: 0.838).

The detailed statistical results are summarized in **Supplementary Tables 3, 4**, and the corresponding ROCs are shown in **Supplementary Figures 1, 2**.

Predictive Value of Deep Learning Core-Needle Biopsy Incorporating the Clinical Data Model Among N0, $N_+(1 - 2)$, and $N_+(\geq 3)$

The overall AUC of multi-classification in the independent test cohort based on DL-CNB+C model was 0.791; there existed the highest precision and recall of 0.747 and 0.947, respectively, in N0; there existed the precision and recall of 0.556 and 0.400 in N_+ (1 – 2); and there existed the precision and recall of 0.375 and 0.162 in $N_+(\geq 3)$. The confusion matrix under the classification threshold of 0.5 is shown in **Figure 5**. According to the results, the model performed well in differentiating the N0 group while showing poor diagnostic efficacy in the other two groups.

TABLE 2 | The performance in prediction of ALN status (N0 vs. N(+)).

Methods		AUC	ACC (%)	SENS (%)	SPEC (%)	PPV (%)	NPV (%)
Clinical data only	T	0.661 [0.622, 0.698]	64.13 [60.24, 67.88]	64.58 [58.17, 70.63]	63.85 [58.86, 68.62]	52.36 [48.32, 56.38]	74.55 [70.85, 77.92]
	V	0.709 [0.643, 0.770]	67.62 [60.84, 73.90]	65.82 [54.29, 76.13]	68.70 [60.02, 76.52]	55.91 [48.46, 63.11]	76.92 [70.62, 82.22]
	I–T	0.613[a,b] [0.545, 0.678]	61.93 [55.12, 68.40]	50.00 [38.89, 61.11]	69.40 [60.86, 77.07]	50.60 [42.34, 58.83]	68.89 [63.49, 73.82]
DL-CNB model	T	0.901 [0.875, 0.923]	80.32 [76.99, 83.35]	94.17 [90.41, 96.77]	71.79 [67.05, 76.21]	67.26 [63.61, 70.71]	95.24 [92.30, 97.09]
	V	0.808 [0.748, 0.859]	72.86 [66.31, 78.75]	77.22 [66.40, 85.90]	70.23 [61.62, 77.90]	61.00 [53.95, 67.62]	83.64 [77.04, 88.62]
	I–T	0.816[c] [0.758, 0.865]	74.77 [68.46, 80.39]	80.95 [70.92, 88.70]	70.90 [62.43, 78.42]	63.55 [56.76, 69.84]	85.59 [79.04, 90.34]
DL-CNB+C model	T	0.878 [0.622, 0.698]	76.51 [73.00, 79.77]	93.33 [89.40, 96.14]	66.15 [61.22, 70.84]	62.92 [59.53, 66.19]	94.16 [90.90, 96.30]
	V	0.823 [0.765, 0.872]	75.71 [69.34, 81.35]	74.68 [63.64, 83.80]	76.34 [68.12, 83.32]	65.56 [57.69, 72.65]	83.33 [77.19, 88.08]
	I–T	0.831 [0.775, 0.878]	75.69 [69.44, 81.23]	89.29 [80.63, 94.98]	67.16 [58.53, 75.03]	63.03 [56.96, 68.71]	90.91 [84.21, 94.94]

95% confidence intervals are included in brackets.

AUC, area under the receiver operating characteristic curve; ACC, accuracy; SENS, sensitivity; SPEC, specificity; PPV, positive predictive value; NPV, negative predictive value; T, training cohort (n = 630); V, validation cohort (n = 210); I–T, independent test cohort (n = 218); ALN, axillary lymph node; DL-CNB+C, deep learning core-needle biopsy incorporating the clinical data.

[a]Indicates p < 0.0001, Delong et al. in comparison with DL-CNB model in independent test cohort.

[b]Indicates p < 0.0001, Delong et al. in comparison with DL-CNB+C model in independent test cohort.

[c]Indicates p = 0.4532, Delong et al. in comparison with DL-CNB+C model in independent test cohort.

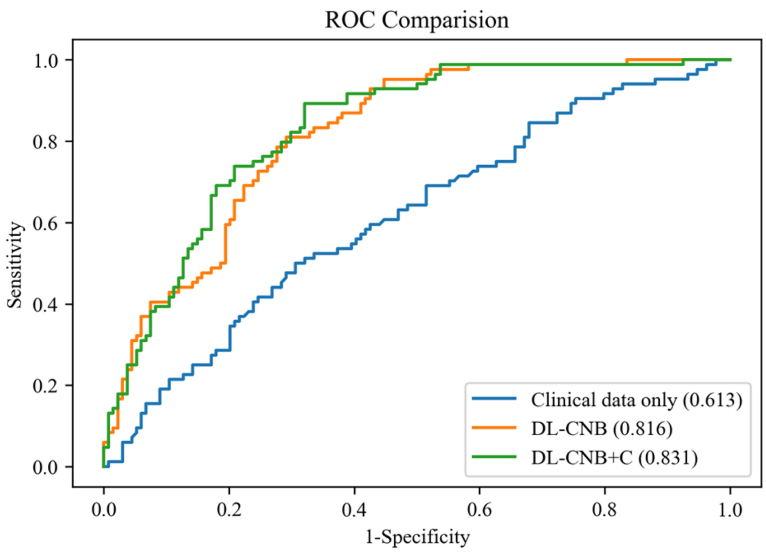


FIGURE 4 | Comparison of receiver operating characteristic (ROC) curves between different models for predicting disease-free axilla (N0) and heavy metastatic burden of axillary disease (N(+)). Numbers in parentheses are areas under the receiver operating characteristic curve (AUCs).

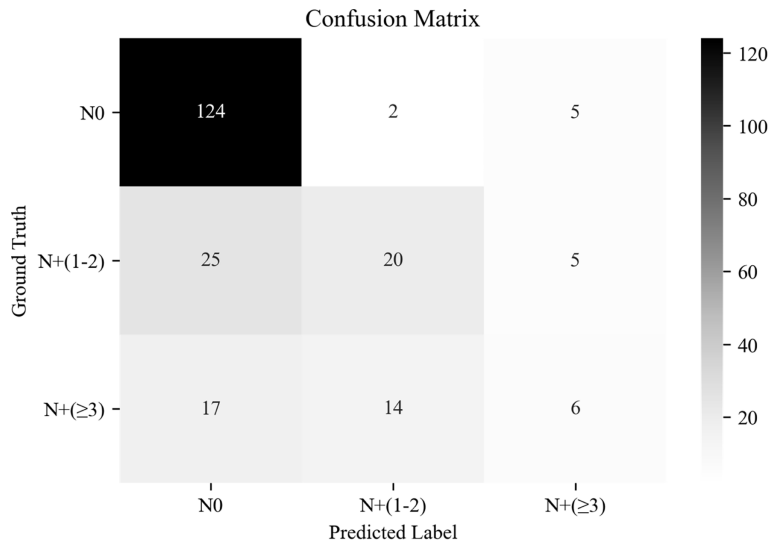


FIGURE 5 | The confusion matrix of predicting axillary lymph node (ALN) status between disease-free axilla (N0), low metastatic burden of axillary disease (N+(1 – 2)), and heavy metastatic burden of axillary disease (N+(≥3)).

Subgroup Analysis of Deep Learning Core-Needle Biopsy Incorporating the Clinical Data Model

Furthermore, we analyzed the measurement results of the different subgroups in the independent test cohort of predicting ALN status between N0 and N(+) by the DL-CNB+C model. The detailed statistical results are summarized in **Supplementary Table 5**. In the independent test cohort, compared with an AUC of 0.794 (95%CI: 0.720, 0.855) in the subgroup of age >50, there existed better performance in the subgroup of age ≤50 with an AUC of 0.918 (95%CI: 0.825, 0.971, $p = 0.015$). There were no significant differences regarding other subgroups of ER(+) vs. ER(−) ($p = 0.125$), PR(+) vs. PR(−) ($p = 0.659$), HER-2(+) vs. HER-2(−) ($p = 0.524$), and T1 vs. T2 stage ($p = 0.743$) between N0 and N(+).

Interpretability of Deep Learning Core-Needle Biopsy Model

To investigate the interpretability of the DL-CNB, we conducted two studies for digging the correlation factors of ALN status prediction. In the first study, we adopted the attention-based MIL pooling to find the important regions that contributing to the prediction. The heat map in **Figure 6A** highlights the red patches as the important regions. Although the obtained important areas can provide some clues to the diagnosis of DL-CNB model, it is not clear that the model makes decisions based on what features of the tumor area.

In the second study, we specially designed and extracted multiple nucleus features for each WSI. The weights of different features were then obtained based on the same attention-based MIL pooling in our DL-CNB. The weights highlighted the nucleus features that were most relevant to the ALN status prediction of each WSI. We found that the WSI of N(+) group had higher nuclear density ($p = 0.015$) and orientation ($p = 0.012$) but lower circumference ($p = 0.009$), circularity ($p = 0.010$), and area ($p = 0.024$) compared with N0 group (**Figures 6B, C**). There were no significant differences in other nucleus features including major axis ($p = 0.083$), minor axis ($p = 0.065$), and rectangularity ($p = 0.149$) between N0 and N(+).

DISCUSSION

In most previous studies, DL signatures of ALN metastases were based on medical images such as ultrasound, CT, and MRI (10, 40, 41). However, since many patients had undergone CNB at the time of imaging examination, and the reactive changes such as needle path in the tumor would result in the predictive inaccuracy of imaging information. This study focused on preoperative CNB WSI, which also played an important role in BC management and has been increasingly performed in clinical practice. Preoperative CNB can provide not only the histopathological diagnosis of BC but also the molecular status including ER/PR/HER-2 status, which is associated with ALN metastasis (42). Otherwise, the morphological features of tumor cells can be visualized on CNB WSI. Therefore, primary tumor biopsy WSI as a complementary imaging tool has the potential for ALN metastasis prediction. To the best of our knowledge, this is the first study to apply the DL-based histopathological features extracted from primary tumor WSIs for ALN prediction analysis.

Here, the best-performing DL-CNB model yielded satisfactory predictions with an AUC of 0.816, a SENS of 81.0%, and a SPEC of 70.9% on the test set, which had superior predictive capability as compared with clinical data

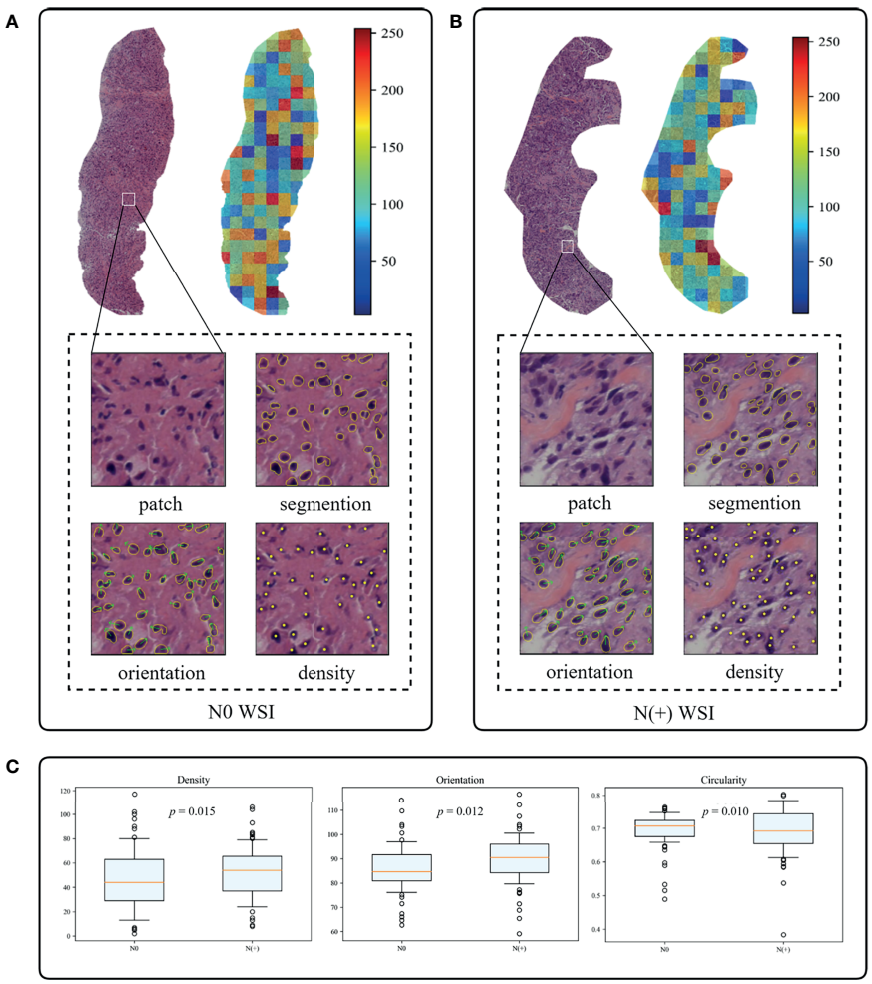


FIGURE 6 | The interpretability of the deep learning core-needle biopsy (DL-CNB) model of two patients. **(A, B)** The heat maps and nuclear segmentation from core-needle biopsy (CNB) whole-slide images (WSIs) of the N0 and the N(+) separately, and the red regions show greater contribution to the final classification. **(C)** The statistical analysis of three nuclear characteristics most relevant to diagnosis of all patients.

alone. Furthermore, unlike other combined models incorporating clinical data (7, 9), the DL-CNB+C model slightly improved the ACC to 0.831, which showed that our results were mainly derived from the contribution of DL-CNB model. In addition, during the subgroup analysis stratified by patient's age, our DL-CNB+C model achieved an AUC of 0.918 for patients younger than 50 years, indicating that age was the critical factor in predicting ALN status. Regarding the number of ALN metastasis, the DL-CNB+C model showed better discriminating ability between N0 and $N_+(1-2)$, and between N0 and $N_+(\geq 3)$. However, the unfavorable discriminating ability was found between $N_+(1-2)$ and $N_+(\geq 3)$. This was consistent with the study of Zheng et al. (9), who also reported poor efficacy between $N_+(1-2)$ and $N_+(\geq 3)$, utilizing the DL radiomics model. In the future, further exploration of ALN staging prediction is needed.

Indeed, computer-assisted histopathological analysis can provide a more practical and objective output (43). For example, different molecular subtypes (44) and Oncotype DX risk score (45) occurring in BC could be directly predicted from the H&E slides. On the one hand, our DL model can provide significant information for risk stratification and axillary staging, thereby avoiding axillary surgery and reducing the complication and hospitalization costs. On the other hand, our results also highlight the development of algorithms based on artificial intelligence, which will reduce the labor intensity of pathologists. Similar approaches may be used to the pathology of other organs.

In our study, we are first to quantitatively assess the role of nuclear disorder in predicting ALN metastasis in BC. Our finding is consistent with several recent studies that demonstrate the powerful predictive effect of nuclear disorder on patient survival (46, 47). Interestingly, the top predictive signatures that distinguished N0 from N(+) were characterized by the nucleus features including density, circumference,

circularity, and orientation. We found that the WSI of N(+) had higher nuclear density and polarity but lower circularity, which was understandable since in the tumors with ALN metastasis, tumor cells became poorly differentiated as a result of rapid cell growth, encouraging the nuclei in these structures to form highly clustered and consistently metastatic patterns. Our results showed that nuanced patterns of nucleus density and orientation of tumor cells are important determinants of ALN metastasis.

There are some limitations in our study. First, the selection of regions of interest within each CNB slide required pathologist guidance. Future studies will explore more advanced methods for automatic segmentation of tumor regions. Second, this is a retrospective study, and prospective validation of our model in a large multicenter cohort of EBC patients is necessary to assess the clinical applicability of the biomarker. Third, recent evidence indicated that a set of features related to tumor-infiltrating lymphocytes (TILs) was found to be associated with positive LNs in bladder cancer (22). However, due to few TILs on breast CNB slides, we only selected sufficient tumor cells for the identification of salient regions rather than whole slides. Finally, we only chose H&E stained images of CNB samples. The clinical utility of immunochemical stained images remains to be established as an interesting attempt.

CONCLUSION

In brief, we demonstrated that a novel DL-based biomarker on primary tumor CNB slides predicted ALN metastasis preoperatively for EBC patients with clinically negative ALN, especially for younger patients. Our methods could help to avoid unnecessary axillary surgery based on the widely collected H&E-stained histopathology slides, thereby contributing to precision oncology treatment.

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

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

FX, CZ, JiL, YW, and MJ designed the study. CZ, WT, YZ, and JiL trained the model. FX, YW, ZS, JuL, and HJ collected the data. FX, WT, YZ, CZ, YW, MJ, and JuL analyzed and interpreted the data. FX, CZ, WT, YZ, and MJ prepared the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The work was supported by National Natural Science Foundation of China [No. 8197101438].

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.759007/full#supplementary-material

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2019. *CA Cancer J Clin* (2019) 69(1):7–34. doi: 10.3322/caac.21551
2. Ahmed M, Purushotham AD, Douek M. Novel Techniques for Sentinel Lymph Node Biopsy in Breast Cancer: A Systematic Review. *Lancet Oncol* (2014) 15(8):e351–62. doi: 10.1016/S1470-2045(13)70590-4
3. Kootstra J, Hoekstra-Weebers JEHM, Rietman H, de Vries J, Baas P, Geertzen JHB, et al. Quality of Life After Sentinel Lymph Node Biopsy or Axillary Lymph Node Dissection in Stage I/II Breast Cancer Patients: A Prospective Longitudinal Study. *Ann Surg Oncol* (2008) 15(9):2533–41. doi: 10.1245/s10434-008-9996-9
4. Wilke LG, McCall LM, Posther KE, Whitworth PW, Reintgen DS, Leitch AM, et al. Surgical Complications Associated With Sentinel Lymph Node Biopsy: Results From a Prospective International Cooperative Group Trial. *Ann Surg Oncol* (2006) 13(4):491–500. doi: 10.1245/ASO.2006.05.013
5. Manca G, Rubello D, Tardelli E, Giammarile F, Mazzarri S, Boni G, et al. Sentinel Lymph Node Biopsy in Breast Cancer: Indications, Contraindications, and Controversies. *Clin Nucl Med* (2016) 41(2):126–33. doi: 10.1097/RLU.0000000000000985
6. Hindié E, Groheux D, Brenot-Rossi I, Rubello D, Moretti J-L, Espié M. The Sentinel Node Procedure in Breast Cancer: Nuclear Medicine as the Starting Point. *J Nucl Med* (2011) 52(3):405–14. doi: 10.2967/jnumed.110.081711
7. Dihge L, Vallon-Christersson J, Hegardt C, Saal LH, Häkkinen J, Larsson C, et al. Prediction of Lymph Node Metastasis in Breast Cancer by Gene Expression and Clinicopathological Models: Development and Validation Within a Population-Based Cohort. *Clin Cancer Res* (2019) 25(21):6368–81. doi: 10.1158/1078-0432.CCR-19-0075
8. Shiino S, Matsuzaki J, Shimomura A, Kawauchi J, Takizawa S, Sakamoto H, et al. Serum Mirna–Based Prediction of Axillary Lymph Node Metastasis in Breast Cancer. *Clin Cancer Res* (2019) 25(6):1817–27. doi: 10.1158/1078-0432.CCR-18-1414
9. 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(1):1236. doi: 10.1038/s41467-020-15027-z
10. Luo J, Ning Z, Zhang S, Feng Q, Zhang Y. Bag of Deep Features for Preoperative Prediction of Sentinel Lymph Node Metastasis in Breast Cancer. *Phys Med Biol* (2018) 63(24):245014. doi: 10.1088/1361-6560/aaf241
11. 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 (8):1301–9. doi: 10.1038/s41591-019-0508-1
12. Gu F, Burlutskiy N, Andersson M, Wilén LK. Multi-Resolution Networks for Semantic Segmentation in Whole Slide Images. In: *Computational Pathology and Ophthalmic Medical Image Analysis; 2018*. Cham, Switzerland: Springer International Publishing (2018). p. 11–8. doi: 10.1007/978-3-030-00949-6_2

13. Mei K, Zhu C, Jiang L, Liu J, Qiao Y. Cross-Stained Segmentation From Renal Biopsy Images Using Multi-Level Adversarial Learning. In: *IEEE International Conference on Acoustics, Speech and Signal Processing; 2020*. IEEE (2020). p. 1424–8. doi: 10.1109/ICASSP40776.2020.9054505
14. Jiang L, Chen W, Dong B, Mei K, Zhu C, Liu J, et al. A Deep Learning-Based Approach for Glomeruli Instance Segmentation From Multistained Renal Biopsy Pathologic Images. *Am J Pathol* (2021) 191(8):1431–41. doi: 10.1016/j.ajpath.2021.05.004
15. Zhu C, Mei K, Peng T, Luo Y, Liu J, Wang Y, et al. Multi-Level Colonoscopy Malignant Tissue Detection With Adversarial CAC-Unet. *Neurocomputing* (2021) 438:165–83. doi: 10.1016/j.neucom.2020.04.154
16. Feng R, Liu X, Chen J, Chen DZ, Gao H, Wu J. A Deep Learning Approach for Colonoscopy Pathology WSI Analysis: Accurate Segmentation and Classification. *IEEE J BioMed Health Inform* (2020). doi: 10.1109/JBHI.2020.3040269
17. Iizuka O, Kanavati F, Kato K, Rambeau M, Arihiro K, Tsuneki M. Deep Learning Models for Histopathological Classification of Gastric and Colonic Epithelial Tumours. *Sci Rep* (2020) 10(1):1504. doi: 10.1038/s41598-020-58467-9
18. Song Z, Zou S, Zhou W, Huang Y, Shao L, Yuan J, et al. Clinically Applicable Histopathological Diagnosis System for Gastric Cancer Detection Using Deep Learning. *Nat Commun* (2020) 11(1):4294. doi: 10.1038/s41467-020-18147-8
19. Hu Y, Su F, Dong K, Wang X, Zhao X, Jiang Y, et al. Deep Learning System for Lymph Node Quantification and Metastatic Cancer Identification From Whole-Slide Pathology Images. *Gastric Cancer* (2021) 24(4):868–77. doi: 10.1007/s10120-021-01158-9
20. Zhao Y, Yang F, Fang Y, Liu H, Zhou N, Zhang J, et al. Predicting Lymph Node Metastasis Using Histopathological Images Based on Multiple Instance Learning With Deep Graph Convolution. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2020*. IEEE (2020). p. 4837–46. doi: 10.1109/CVPR42600.2020.00489
21. Steiner DF, MacDonald R, Liu Y, Truszkowski P, Hipp JD, Gammage C, et al. Impact of Deep Learning Assistance on the Histopathologic Review of Lymph Nodes for Metastatic Breast Cancer. *Am J Surg Pathol* (2018) 42(12):1636–46. doi: 10.1097/PAS.0000000000001151
22. Harmon SA, Sanford TH, Brown GT, Yang C, Mehralivand S, Jacob JM, et al. Multiresolution Application of Artificial Intelligence in Digital Pathology for Prediction of Positive Lymph Nodes From Primary Tumors in Bladder Cancer. *JCO Clin Cancer Inform* (2020) 4:367–82. doi: 10.1200/CCI.19.00155
23. Ilse M, Tomczak J, Welling M. Attention-Based Deep Multiple Instance Learning. In: *Proceedings of the International Conference on Machine Learning* PMLR (2018). p. 2127–36.
24. Das K, Conjeti S, Roy AG, Chatterjee J, Sheet D. Multiple Instance Learning of Deep Convolutional Neural Networks for Breast Histopathology Whole Slide Classification. In: *IEEE International Symposium on Biomedical Imaging; 2018*. IEEE (2018). p. 578–81. doi: 10.1109/ISBI.2018.8363642
25. Sudharshan PJ, Petitjean C, Spanhol F, Oliveira LE, Heutte L, Honeine P. Multiple Instance Learning for Histopathological Breast Cancer Image Classification. *Expert Syst Appl* (2019) 117:103–11. doi: 10.1016/j.eswa.2018.09.049
26. Couture HD, Marron JS, Perou CM, Troester MA, Niethammer M. Multiple Instance Learning for Heterogeneous Images: Training a CNN for Histopathology. In: *Medical Image Computing and Computer Assisted Intervention; 2018*. Cham, Switzerland: Springer International Publishing (2018). p. 254–62. doi: 10.1007/978-3-030-00934-2_29
27. Krizhevsky A, Sutskever I, Hinton GE. Imagenet Classification With Deep Convolutional Neural Networks. *Commun ACM* (2017) 60(6):84–90. doi: 10.1145/3065386
28. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In: *International Conference on Learning Representations; 2015*. ICLR (2015).
29. 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*. IEEE (2016). p. 770–8. doi: 10.1109/CVPR.2016.90
30. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2017*. IEEE (2017). p. 4700–8. doi: 10.1109/CVPR.2017.243
31. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the Inception Architecture for Computer Vision. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016*. IEEE (2016). p. 2818–26. doi: 10.1109/CVPR.2016.308
32. Feng J, Zhou Z-H. Deep Miml Network. In: *Proceedings of the AAAI Conference on Artificial Intelligence; 2017*. AAAI Press (2017). p. 1884–90.
33. Pinheiro PO, Collobert R. From Image-Level to Pixel-Level Labeling With Convolutional Networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2015* IEEE (2015). p. 1713–21. doi: 10.1109/CVPR.2015.7298780
34. Zhu W, Lou Q, Vang YS, Xie X. *Deep Multi-Instance Networks With Sparse Label Assignment for Whole Mammogram Classification Medical Image Computing and Computer Assisted Intervention; 2017*. Cham, Switzerland: Springer International Publishing (2017) p. 603–11. doi: 10.1007/978-3-319-66179-7_69
35. Loshchilov I, Hutter F. SGDR: Stochastic Gradient Descent With Warm Restarts. In: *International Conference on Learning Representations; 2017*. ICLR (2017).
36. Mueller JL, Gallagher JE, Chitalia R, Krieger M, Erkanli A, Willett RM, et al. Rapid Staining and Imaging of Subnuclear Features to Differentiate Between Malignant and Benign Breast Tissues at a Point-of-Care Setting. *J Cancer Res Clin Oncol* (2016) 142(7):1475–86. doi: 10.1007/s00432-016-2165-9
37. Radhakrishnan A, Damodaran K, Soylemezoglu AC, Uhler C, Shivashankar GV. Machine Learning for Nuclear Mechano-Morphometric Biomarkers in Cancer Diagnosis. *Sci Rep* (2017) 7(1):17946. doi: 10.1038/s41598-017-17858-1
38. 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(3):837–45. doi: 10.1038/s41591-019-0508-1
39. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. Imagenet: A Large-Scale Hierarchical Image Database. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2009*. IEEE (2009). p. 248–55.
40. Zhou L-Q, Wu X-L, Huang S-Y, Wu G-G, Ye H-R, Wei Q, et al. Lymph Node Metastasis Prediction From Primary Breast Cancer US Images Using Deep Learning. *Radiology* (2020) 294(1):19–28. doi: 10.1148/radiol.2019190372
41. Yang X, Wu L, Ye W, Zhao K, Wang Y, Liu W, et al. Deep Learning Signature Based on Staging CT for Preoperative Prediction of Sentinel Lymph Node Metastasis in Breast Cancer. *Acad Radiol* (2020) 27(9):1226–33. doi: 10.1016/j.acra.2019.11.007
42. Calhoun KE, Anderson BO. Needle Biopsy for Breast Cancer Diagnosis: A Quality Metric for Breast Surgical Practice. *J Clin Oncol* (2014) 32(21):2191–2. doi: 10.1200/JCO.2014.55.6324
43. Acs B, Rantalainen M, Hartman J. Artificial Intelligence as the Next Step Towards Precision Pathology. *J Intern Med* (2020) 288(1):62–81. doi: 10.1111/joim.13030
44. Jaber MI, Song B, Taylor C, Vaske CJ, Benz SC, Rabizadeh S, et al. A Deep Learning Image-Based Intrinsic Molecular Subtype Classifier of Breast Tumors Reveals Tumor Heterogeneity That may Affect Survival. *Breast Cancer Res* (2020) 22(1):12. doi: 10.1186/s13058-020-1248-3
45. Whitney J, Corredor G, Janowczyk A, Ganesan S, Doyle S, Tomaszewski J, et al. Quantitative Nuclear Histomorphometry Predicts Oncotype DX Risk Categories for Early Stage ER+ Breast Cancer. *BMC Cancer* (2018) 18(1):610. doi: 10.1186/s12885-018-4448-9
46. Lu C, Romo-Bucheli D, Wang X, Janowczyk A, Ganesan S, Gilmore H, et al. Nuclear Shape and Orientation Features From H&E Images Predict Survival in Early-Stage Estrogen Receptor-Positive Breast Cancers. *Lab Invest* (2018) 98(11):1438–48. doi: 10.1038/s41374-018-0095-7
47. Lee G, Veltri RW, Zhu G, Ali S, Epstein JI, Madabhushi A. Nuclear Shape and Architecture in Benign Fields Predict Biochemical Recurrence in Prostate Cancer Patients Following Radical Prostatectomy: Preliminary Findings. *Eur Urol Focus* (2017) 3(4):457–66. doi: 10.1016/j.euf.2016.05.009

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.

this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Bioinformatics

ORIGINAL RESEARCH
published: 26 October 2021
doi: 10.3389/fbinf.2021.718697

Automated Machine-Learning Framework Integrating Histopathological and Radiological Information for Predicting IDH1 Mutation Status in Glioma

Dingqian Wang[1], Cuicui Liu[2], Xiuying Wang[1], Xuejun Liu[3], Chuanjin Lan[4,5], Peng Zhao[4,5], William C. Cho[6], Manuel B. Graeber[7] and Yingchao Liu[4,5]**

[1]School of Computer Science, The University of Sydney, Sydney, NSW, Australia, [2]Department of Neurology, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, China, [3]Department of Radiology, Hospital Affiliated to Qingdao University, Qingdao, China, [4]Department of Neurosurgery, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, China, [5]Department of Neurosurgery, Shandong Provincial Hospital Affiliated to Shandong University, Jinan, China, [6]Department of Clinical Oncology, Queen Elizabeth Hospital, Kowloon, Hong Kong, SAR China, [7]Ken Parker Brain Tumor Research Laboratories, Brain and Mind Centre, Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, Australia

OPEN ACCESS

Edited by:
Li Xiao,
University of Science and Technology of China, China

Reviewed by:
Xingcan Hu,
University of Science and Technology of China, China
Gongning Luo,
Harbin Institute of Technology, China

***Correspondence:**
Xiuying Wang
xiu.wang@sydney.edu.au
Yingchao Liu
Yingchaoliu@email.sdu.edu.cn

Specialty section:
This article was submitted to Computational BioImaging, a section of the journal Frontiers in Bioinformatics

Received: *01 June 2021*
Accepted: *28 September 2021*
Published: *26 October 2021*

Citation:
Wang D, Liu C, Wang X, Liu X, Lan C, Zhao P, Cho WC, Graeber MB and Liu Y (2021) Automated Machine-Learning Framework Integrating Histopathological and Radiological Information for Predicting IDH1 Mutation Status in Glioma. Front. Bioinform. 1:718697. doi: 10.3389/fbinf.2021.718697

Diffuse gliomas are the most common malignant primary brain tumors. Identification of isocitrate dehydrogenase 1 (IDH1) mutations aids the diagnostic classification of these tumors and the prediction of their clinical outcomes. While histology continues to play a key role in frozen section diagnosis, as a diagnostic reference and as a method for monitoring disease progression, recent research has demonstrated the ability of multi-parametric magnetic resonance imaging (MRI) sequences for predicting IDH genotypes. In this paper, we aim to improve the prediction accuracy of IDH1 genotypes by integrating multi-modal imaging information from digitized histopathological data derived from routine histological slide scans and the MRI sequences including T1-contrast (T1) and Fluid-attenuated inversion recovery imaging (T2-FLAIR). In this research, we have established an automated framework to process, analyze and integrate the histopathological and radiological information from high-resolution pathology slides and multi-sequence MRI scans. Our machine-learning framework comprehensively computed multi-level information including molecular level, cellular level, and texture level information to reflect predictive IDH genotypes. Firstly, an automated pre-processing was developed to select the regions of interest (ROIs) from pathology slides. Secondly, to interactively fuse the multimodal complementary information, comprehensive feature information was extracted from the pathology ROIs and segmented tumor regions (enhanced tumor, edema and non-enhanced tumor) from MRI sequences. Thirdly, a Random Forest (RF)-

Abbreviations: TCGA, The Cancer Genome Atlas; TCIA, The Cancer Imaging Archive; IDH, Isocitrate dehydrogenase; IDH-wt, Isocitrate dehydrogenase wild type; IDH-m, Isocitrate dehydrogenase mutation; MRI, Magnetic Resonance Imaging; T1C, T1-weighted post-Gd contras; FLAIR, Fluid-attenuated inversion recovery; GLCM, Grey Level Co-occurrence Matrix; GLRLM, Grey Level Length Matrix; GLSZM, Grey Level Size Zone Matrix; GLDM, Grey level Dependence Matrix; LoG, Laplacian of Gaussian; ROI, Region of Interest; VOI, Volume of Interest; LIME, Local Interpretable Model-agnostic Explanations; ROC, Receiver Operating Characteristic; Label 1, the area of edema and non-enhancing tumor; Label 2 the area of enhanced tumor.

based algorithm was employed to identify and quantitatively characterize histopathological and radiological imaging origins, respectively. Finally, we integrated multi-modal imaging features with a machine-learning algorithm and tested the performance of the framework for IDH1 genotyping, we also provided visual and statistical explanation to support the understanding on prediction outcomes. The training and testing experiments on 217 pathologically verified IDH1 genotyped glioma cases from multi-resource validated that our fully automated machine-learning model predicted IDH1 genotypes with greater accuracy and reliability than models that were based on radiological imaging data only. The accuracy of IDH1 genotype prediction was 0.90 compared to 0.82 for radiomic result. Thus, the integration of multi-parametric imaging features for automated analysis of cross-modal biomedical data improved the prediction accuracy of glioma IDH1 genotypes.

Keywords: digital histological slides, isocitrate dehydrogenase 1 mutations, machine-learning, magnetic resonance imaging, multimodal integration, imaging information analysis

INTRODUCTION

The current WHO classification of CNS tumors not only considers histopathological phenotypes but also molecular genetic parameters, e.g., DNA methylome profiling (Louis et al., 2021; Lopes 2017; Chang et al., 2018; Lee et al., 2019). IDH mutations in glioma have been found to be associated with better outcomes and are therefore of great relevance in the clinical assessment of glioma patients (Louis et al., 2016b). Recently, some attempts have been made to use radiological images for the pre-surgical prediction of IDH1 genotypes (Gillies et al., 2015; Kesler et al., 2019; Lee et al., 2019; Tatekawa et al., 2021).

Pathological and radiological imaging results are increasingly available in digitized format (Nance et al., 2013; Farahani and Pantanowitz 2015; Griffin and Treanor 2017). It has become apparent that fully utilizing the data of digital radiology and pathology images through machine-learning can facilitate the identification of biomarkers that reflect information on the basic biology and physiology of various malignancies (Deo 2015; Gillies et al., 2016). Although tumor diagnoses increasingly consider molecular genetics markers, histology continues to play a key role in frozen section diagnosis, as a diagnostic reference and as a method for monitoring disease progression. In addition, compared to the indirect visualization of disease phenotypes by means of imaging, histology provides direct information at high resolution (Kinjo et al., 2008; Missbach-Guentner et al., 2018; Vågberg et al., 2018; Rathore et al., 2020). Competing with the computer-aid image technique used in radiology and pathology, clinical practice demands professional knowledge and long-term training to obtain useful information from the image with the naked eye for diagnosis and evaluation (Harezlak and Kasprowski 2018).

Computerized image analysis can reduce subjective inter-observer bias that is known to limit all human observation including in histopathology (Emblem et al., 2014; Zhang et al., 2016; Choi et al., 2019). Machine-learning algorithms are already widely used in glioma research, and most are based on the analysis of features extracted from MRIs (Ellingson et al., 2011; Zacharaki et al., 2012; Emblem et al., 2014; Macyszyn et al., 2015). Zhou et al. (2017) have recently demonstrated the ability to predict IDH genotypes in cases of primary grade II and III glioma using clinical and pathological variables and textual features extracted from regions of interest (ROI) in four sequences of MRI, including T1W, T2W, T1CE and T2-FLAIR, achieving an accuracy of 0.86. Compared to the results obtained by Zhou and colleagues, Eichinger et al. (2017) were able to increase the accuracy of IDH genotyping to 92% by designing an algorithm based on feature extraction of local binary pattern, which represent texture features extracted from multimodal MRI data. In addition to traditional machine-learning techniques, the method employed by Xing et al. (2017) classifies IDH mutations and IDH wild type (IDH-wt) using conventional machine-learning algorithms in order to extract deep features from four sequences of MRI (T1W, T2W, T1CE, T2-FLAIR). Zhang et al. (2016) aimed to distinguish the presence of an IDH mutation and IDH-wt in primary grade III and IV gliomas by means of additional features (intensity, texture and shape features) extracted from multimodality MRIs (T1, T1CE, T2, T2-FLAIR and DWI), achieving an accuracy of 0.883 using the Random Forest algorithm.

Recently, the improvements in deep learning is capable to overcome the previous challenges by learning high-dimensional representations of imaging data. Novel, fully automated postprocessing analyses of standard and advanced MR images are clearly rapidly approaching. These fully automated analyses are especially appealing because they provide unbiased evaluations independent of operator, training or experience. Fully automated postprocessing with deep learning analyses of standard and advanced MR images have achieved high accuracy even at 92.8% accuracy, 93.1% specificity, and 92.6% sensitivity (Choi et al., 2019). Although they can be very powerful for the prediction of IDH for glioma, deep CNN models are vulnerable to overfitting to their given training dataset and inherent difficult for interpretation which is the most crucial for decision support system. Comprehensive understanding of the mechanism of deep and machine-learning is necessary, however, to best develop and then apply these algorithms to clinical practice we need to avoid

TABLE 1 | Patient characterizes.

	Shandong Provincial Hospital	TCGA[a]
Grade III (*n*; %)	41	25
IDH-mutated in Grade III (*n*; % column)	20, 48.9%	17, 68%
Grade IV (*n*; %)	85	66
IDH-mutated in Grade IV (*n*; % column)	20; 23.5%	12, 18.2%
Age (years; mean; range)	49; [5, 79]	53; [18, 81]
Sex (n male; % column)	55; 43.7%	55; 60.4%

[a]*TCGA, The Cancer Genome Atlas.*

their potential pitfalls. It is unlikely to replace tissue sampling for now; therefore, the continued improvement in model performance and consistency across diverse imaging modality brings us closer to the precise molecular diagnosis (Gutman and Young 2021).

In this study, we introduce an improved approach to IDH prediction, which integrates radiological and histopathological data analyses in a single combined framework. Radiological data analysis in this context refers to the extraction and analysis of high-throughput features from tomographic images (MR images) whereas histopathological data analysis refers to the features extraction and analysis from whole slide images. We envision that this model could set a pathway for the non-invasive evaluation of IDH mutation in gliomas and may provide a quantitative result analysis for the researchers. Compared with deep learning-based method, we aim to provide doctors with an intuitive, interpretable, and cost-effective mechanism through machine-learning based method to support the decision on IDH status prediction.

MATERIALS AND METHODS

Patient Enrolment

The imaging data of 217 subjects that had been diagnosed with glioma were collected from two different sources. 126 cases were from Shandong Provincial Hospital that is affiliated with Shandong University, comprising 41 histological grade III cases and 85 histological grade IV cases. The remaining 91 cases were retrieved from The Cancer Imaging Archive (TCIA), comprising 25 histological grade III cases and 66 histological grade IV cases (**Table 1**). The criteria for image acquisition in this study are as follows: I) available histology, age at diagnosis, sex, and IDH status; II) MR imaging data, including post-contrast T1-weighted images (T1CE), and T2-FLAIR, and III) histopathological images.

Dataset

Histopathological Images

Shandong provincial hospital's cohort: cases were diagnosed according to WHO criteria (Louis et al., 2021). Paraffin-embedded tissue samples were cut into 3 μm thick slides and stained with H&E stain for all patients in this cohort. All H&E stained images were scanned on a Leica SCN400 slide scanner (Leica Biosystems, Nussloch, Germany) with multi-resolution varying from 20× to 40× for analysis.

Genomic DNA was isolated from formalin-fixed paraffin-embedded glioma tissues. DNA was extracted using the QIAamp DNA Micro kit (Qiagen GmbH, Hilden, Germany) as previously described (Perizzolo et al., 2012).

Expression of IDH-R132H mutant was firstly analyzed by IHC as previously described (Reyes-Botero et al., 2014). For IDH R132H–negative tumors, multiple-gene Sanger sequencing was performed to identify alternative IDH mutations (Sanson et al., 2009). IDH status was defined according to the absence of IDH-R132H immunopositivity and/or mutations in IDH1 and IDH2 genes identified by sequencing.

The Cancer Genome Atlas (TCGA) cohort: Digital pathology slides of diagnosed diffuse gliomas were downloaded from TCGA Data Portal (http://cancergenome.nih.gov.) including information on IDH status, and the corresponding MRI images were acquired from the Cancer Imaging Archive (TCIA) Data Portal (https://www.cancerimagingarchive.net).

Multimodal MR Images

All patients were imaged in the supine position with a 3.0-T MRI machine (Magnetom, Skyra; Siemens Healthcare, Erlangen, Germany) using a transmit/receive quadrature 20-channel head-and-neck coil. The imaging protocol was the same for all patients.

T1-contrast: TR, 1820 ms; TE, 13 ms; slice number, 19; FOV, 230 mm; slice thickness, 5 mm; distance factor, 30%; FA, 150 deg; inversion time (TI), 825 ms; voxel size, 0.4 × 0.4 × 5.0 mm; accelerate factor, 2; bandwidth, 260 Hz/Px; echo spacing, 13 ms.

Fluid-attenuated inversion recovery imaging (T2-FLAIR): TR, 8,000 ms, TE: 81 ms, slice number: 19, FOV, 220 mm; slice thickness, 5 mm; distance factor, 30%; FA, 150 deg; inversion time (TI), 2,370 ms; voxel size, 0.7 mm × 0.7 mm × 5.0 mm; accelerate factor, 2; bandwidth, 289 Hz/Px; echo spacing, 9.02 ms.

All MRI sequences of each patient from our own datasets and from TCIA have the same imaging scale, position, slice anatomy and slice thickness.

COMPUTER ANALYSIS

An automated framework was designed to predict IDH genotype, consisting of the following steps, which were carried out in sequence: I) automated image pre-processing to select the regions of interest (ROIs), II) feature extraction, III) feature selection, and IV) automated IDH genotype prediction and results interpretation (**Figure 1**).

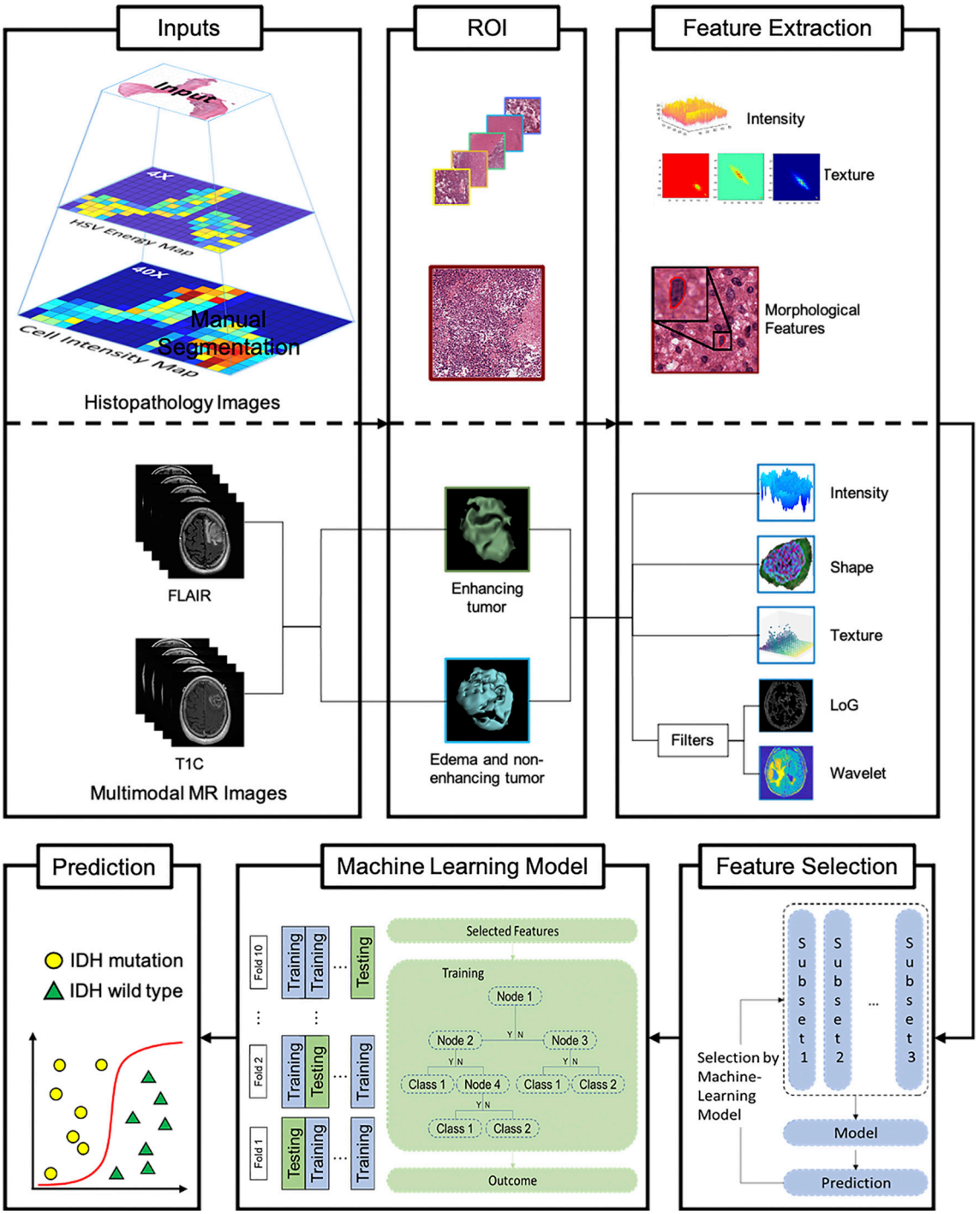


FIGURE 1 | Machine-learning framework for automated prediction of IDH glioblastoma genotypes. Histopathology and multimodal MR images are used as input (left column). Representative regions of interest (ROIs) are extracted (middle column) followed by histopathological and radiomics feature selection (right column). Subsequently, a Random Forest model-based Recursive Feature Elimination (RF-RFE)-algorithm is applied to select relevant while reducing redundant features. Following 10-fold cross-validation, the automated machine-learning model for glioma IDH genotype prediction is established. Abbreviations: GCLM, Grey Level Co-occurrence Matrix features; GLRLM, Grey Level Length Matrix features; GLSZM, Grey Level Size Zone Matrix features; NGTDM, Neighboring Gray Tone Difference Matrix features; GLDM, Grey level Dependence Matrix features, mass of features, which may have redundant information, then selected and processed to improve the predictive power of the machine-learning model.

Automatic Extraction of Region of Interest

Our computational algorithm used for the analysis of histological images approaches the region of interest at two different levels. First, one tile with the highest cell density (5,120 * 5,120) (Sertel et al., 2009; Mobadersany et al., 2018) is extracted employing the watershed nuclei detection algorithm (Al-Kofahi et al., 2010; Kumar et al., 2017; Wang et al., 2018). Then, based on Hue, Saturation, and Value of Brightness (HSV channel), five tiles representing the whole image (Al-Kofahi et al., 2010; Kumar et al., 2017) at 40X resolution are identified. Third, based on the HSV channel, five tiles representing the entire image at 4X resolution are identified.

For the analysis of radiological images, we have segmented edema and non-enhanced tumors from T2-FLAIR image. In addition, T1CE images were used for enhancing tumor volume segmentation. The lesions were separated into three parts, enhancing tumor, tumor necrosis and peritumoral edema. The process of tumor segmentation was performed manually using the ITK-SNAP software (version 3.6.0; www.itksnap.org). First, all MRI sequences were retrieved from the Picture Archiving and Communication System (PACS). Then we applied N4 bias field correction to remove the presence of low frequency intensity non-uniformity. Inter-modality co-registration with different 2D MRI sequences was achieved by means of ITK-SNAP. Using this method, ROIs of enhancing tumor were delineated on post-contrast T1WI images by a semi-automatic method, in which only the enhancing area was selected. Tumor necrosis was defined as the non-enhancing area within enhancing area on post-contrast T1WI. ROIs of peritumoral edema were delineated on T2-FLAIR, which was defined as the high-signal region beyond the enhancing area. The process was performed by a consultant neuro-radiologist. Finally, the ROIs were registered on each slice of each 2D MRI sequence.

Feature Extraction

In this step, we extracted quantitative features from histopathology images and MRIs. In case of the histopathology images, we extracted two types, visual features and sub-visual features, at two different resolutions. The visual features quantitatively describe the morphology of nuclei such as the mean area occupied and the pattern of staining. Sub-visual features are derived from a high-throughput intensity and texture matrix, which reflects the intensity distribution at the single pixel level.

In the case of MRIs, we obtained shape features from the volume of interest (VOI) reflecting tumor area and volume. Subsequently, we extracted first-order, second-order and high order features from the ROIs. The shape-based features describe the three-dimensional (3D) properties of the tumor, such as tumor volume, sphericity, and 3D diameter. First-order statistical features reflect the distribution of voxel intensities within the tumor area, including energy and entropy. Second-order statistical features were obtained from the relationships between adjacent voxels (Balagurunathan et al., 2014) to describe the second-order joint probability function of the tumor region as a gray-level co-occurrence matrix (GLCM) and Gray-level run-length matrix (GLRLM), respectively, which reflect intra-tumoral heterogeneity. High-order features were calculated with the help of different filters such as the wavelet transform.

Feature Selection

Although a large number of image features can be used to construct a model that fully reflects the characteristics of gliomas, removing redundant information can improve the efficacy of the model for glioma genotyping (Guyon and Elisseeff 2003). In order to reduce the amount of redundant information inherent to quantitative features, we built a Random Forest algorithm enhanced by a recursive feature elimination (RF-RFE) procedure in order to identify the relevant and important characteristics before implementation in a classification model all (Saeys et al., 2007). As shown in **Figure 1**, the feature with the lowest importance for classification calculated by the algorithm will be eliminated.

Modeling and Validation

We are proposing a binary classification model to differentiate patients with an IDH mutation from wild type high-grade gliomas (HGGs) based on clinical features (age and sex), digital histopathological image features and MRI Radiomics features.

The Random Forest algorithm employed in this study is widely used in medical imaging analysis. The corresponding model is able to accommodate a very large set of features. All machine-learning methods were implemented with the Statistics and Machine-Learning package on Python 3.6.

As discussed by Guyon and Elisseeff (2003), although a large number of image features can be used to construct a model to better reflects the characteristics of gliomas, the model may face over-fitting problems, and therefore redundant information needs to be carefully removed to improve the efficacy of the model for genotyping gliomas. In order to ensure the stability and efficiency of the selected features, the 10-Fold cross-validation is nested with the RF training model to select a valuable feature set. The random forest algorithm enhanced by the recursive feature elimination (RF-RFE) process is used to identify relevant and important features before all implementations in the classification model (Saeys et al., 2007).

RESULTS

Feature Extraction and Selection

We extracted a total of 22 morphological features, which were identified in the glioma cases studied (**Supplementary Material I**. Extraction of histopathologic features), including nuclear shape and staining intensity (**Figure 1**). In addition, we extracted 171 sub-visual features (Wang et al., 2018) from the high resolution digital histopathology images, including intensity features and GLCM features.

As for results of IDH status prediction for HGGs, the histopathological features extracted from histopathology images, which reached an accuracy of 0.81 ± 0.03 with 10-fold

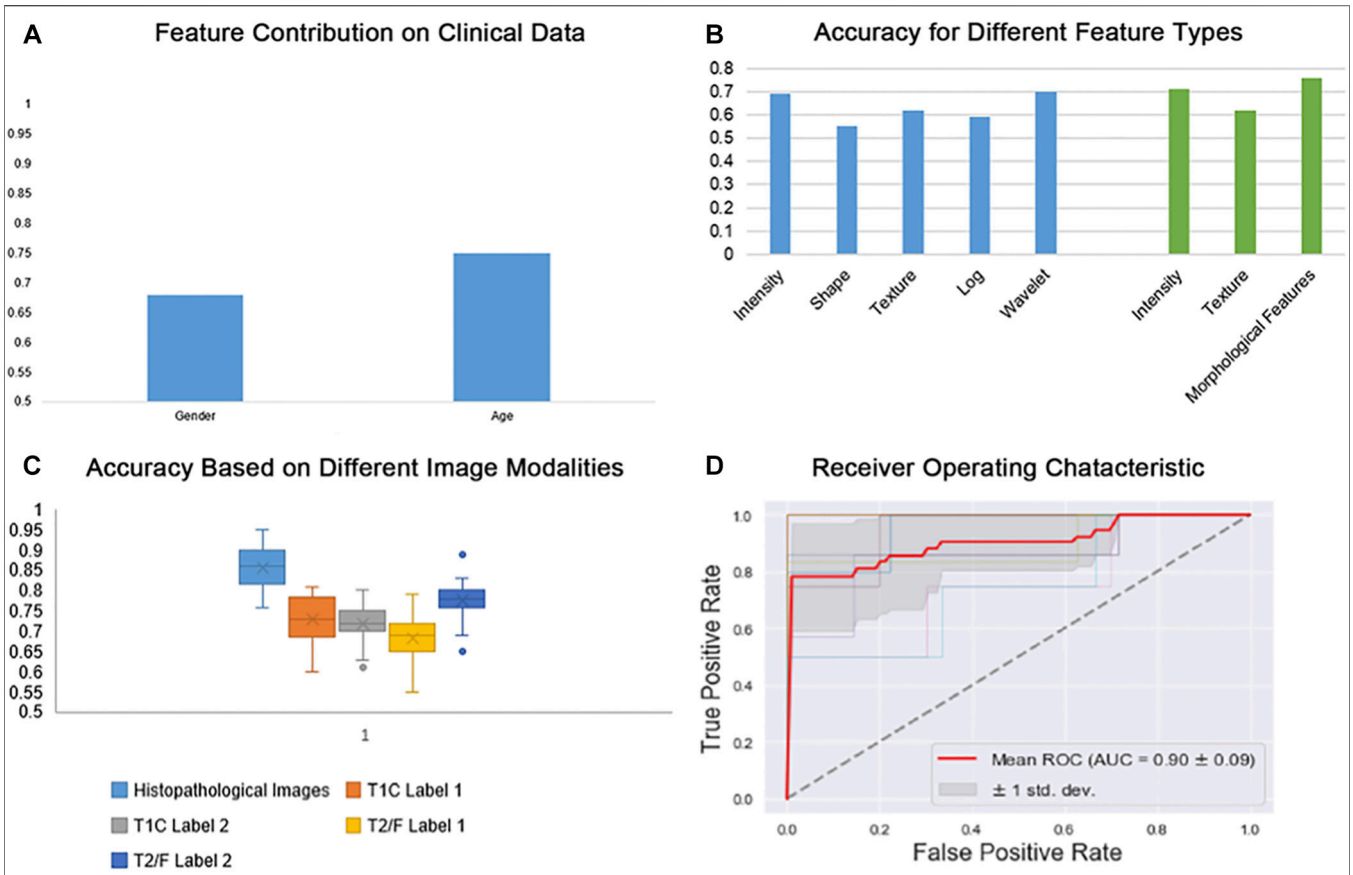


FIGURE 2 | Random forest classifier scores for IDH genotype prediction. **(A)** Prediction results based on Clinical data (age and gender); **(B)** Prediction results based on different feature types; **(C)** Prediction results based on different image modalities; **(D)** Receiver Operating Characteristic (ROC) for IDH genotype prediction across multi-parametric medical images.

cross validation. Regarding multimodal MRIs (**Supplementary Material II**. Radiomics Features Extraction), 1,132 features were extracted from the individual patients' different MR image sequences, including 234 first order features, 14 shape-based features, 286 grey level co-occurrence matrix features, 208 grey level run length matrix features, 208 grey level size zone matrix features, and 182 grey level dependence matrix features (**Figure 1**). The area under the curve (AUC) for features extracted from different histopathological grade of tumors was 0.90 ± 0.09.

Comparison of Performances When Using Different Modalities and Feature Types

In order to assess the differential relevance of the modalities tested (T1CE, FLAIR and digital pathology images) in the prediction of IDH genotype, a Random Forest machine-learning model with 10-fold cross-validation was established. In general, scans of histopathological images yielded more accurate results in the IDH genotype prediction than other image types (**Figure 2C**). Considering quantitative features obtained from the different modalities, our morphologically defined visual features also showed high accuracy and stability (**Figure 2B**). With the multiparameter imaging features minded from different modalities images, our quantitative and objective analysis platform achieved high diagnostic accuracy (0.90 ± 0.05). On the other hand, the mined multiparametric features were achieved different accuracy in corresponding image modality, including 0.86 (±0.03) in the Digital Histopathological Images, 0.73 (±0.06) in T1CE (edema and non-enhanced tumor), 0.72 (±0.04) in T1CE (enhanced tumor), 0.68 (±0.05) in T2-FLAIR (edema and non-enhanced tumor) and 0.78 (±0.04) in T2-FLAIR (enhanced tumor).

For IDH genotype prediction, optimal features were selected from the different modalities of medical images, including seven features from the digital histopathological images, four features form the T1CE images and five from the T2-FLAIR images. As shown in **Figure 2B**, GLRLM, Shape-based and GLCM features had the greatest power in predicting glioma IDH status. Age, counts of nuclei and first-order features were the most important factors that contributed to this result. Top-performers within different groups of image features contributed to IDH status prediction as summarized in **Table 2**.

The accuracy of IDH status prediction was as high as 0.88 ± 0.03 when multi-parametric features were extracted from

TABLE 2 | TOP-performing features in IDH status prediction by means of univariate analysis.

Types	Mask	Feature name	Feature description	Accuracy
Clinical	N/A	Age	Age at diagnose	0.74
Intensity	T1C-edema	Uniformity	Formula $F_u = \sum_{i=1}^{N_g} p(i)^2$ Where $p_{(i)}$ refers to the features calculated form N_g discrete pixel levels Measuring the sum of the square of image VOI pixel value	0.69
Shape	FLAIR-edema	Flatness	Formula $F_{flatness} = \frac{\lambda_{least}}{\lambda_{major}}$ Where λ_{major} and λ_{least} refer to the length of the maximum and minimum principal component axes, respectively Measuring the relationship between the largest and smallest principal components in the VOI shape	0.67
Texture	T1C-tumor	Wavelet-LLL_glrlm_LRLGLE	Formula $F_L = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\frac{P(i,j\|\theta)}{j^2}}{N_r(\theta)}$ Where N_g refers to the gray level distribution within the VOI, N_r refers to the maximal length within the VOI, $P(i,j\|\theta)$ refers to the run length matrix for an arbitrary direction θ, $N_r(\theta)$ is the number of runs in the image along θ This feature quantitative describes the joint distribution of long-run lengths with lower gray level values after a wavelet filter	0.72
Wavelet	T1C-tumor	wavelet-HHH-glcm-MP	Formula $F_{MP} = \max(p(i,j))$ Where $p(i,j)$ is the normalized co-occurrence matrix Quantify the occurrences of the most predominant pair of neighboring intensity values through a Gray Level Co-occurrence Matrix after an image filter by a high-frequency wavelet.	0.69
LoG	FLAIR-tumor	Log-glszm- SALGLE	Formula $F_{SALGLE} = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_s}\frac{P(i,j)}{j^2}}{N_z}$ Where N_g refers to the distribution values within the VOI, N_s refers to the zone sizes quantity within the VOI., N_z refers to refers to the zones quantity within the VOI, and $P(i,j)$ is the size zone matrix Quantify the proportion in the mask of VOI by quantify the Gray Level Size Zone joint distribution of smaller size zones with lower gray level values after the LoG filter	0.65
Morphology	Tile	Cell counts	Quantitative describes the cell intensity in the ROI.	0.78

MP, Maximum Probability; SALGLE, Small Area Low Gray Level Emphasis; LRLGLE, Long Run Low Gray Level Emphasis; HHH, high, high and high frequency.

TABLE 3 | Prediction of IDH genotype based on high grade gliomas.

Image modalities	Accuracy
Digital Histopathological Images	0.86 (± 0.03)
T1CE (edema and non-enhanced tumor)	0.73 (± 0.06)
T1CE (enhanced tumor)	0.72 (± 0.04)
T2-FLAIR (edema and non-enhanced tumor)	0.68 (± 0.05)
T2-FLAIR (enhanced tumor)	0.78 (± 0.04)
Multi-modal Image Data	0.90 (± 0.05)

different histopathological and radiomics images through our implementation of the Random Forest algorithm. **Table 3** shows the features that play important roles in our classification model.

Quantitative Isocitrate Dehydrogenase Status Prediction and Results Interpretation

LIME (Local Interpretable Model-agnostic Explanations) is a tool for facilitating local model interpretability. The technique perturbs the input data to understand how the predictions are affected. **Figures 3**, **4** illustrate two representative cases from visual analysis and the machine-learning model. The first case is an IDH-wt patient (**Figure 3**), who is 43 years old (age at diagnosis), female with a histopathological grade IV glioma. The second one is a patient with an IDH mutation (**Figure 4**), who is 22 years old (age at diagnosis), male with histopathological grade III glioma.

Machine-learning models taking into account the extracted features' different contributions, then quantitatively predict the results by combining these different features according to their corresponding contributions. During model training, the LIME model can generate weighted coefficients to illustrate the contributions made by different features. The predictive ability of LIME algorithm made the most important contribution to achieve a higher weight value. Positive weights reflect the increase in the corresponding features may make a positive contribution to the IDH status prediction. In contrast, negative weights would have negative predictive power. As shown in **Figures 3**, **4**, the feature contribution for IDH genotyping for two representative cases have been listed, which are derived from the LIME model to obtain the linear combination of feature values and weights.

DISCUSSION AND CONCLUSION

Discussion

Determination of IDH status has become a standard for glioma diagnosis as it helps to guide clinical decision-making. In this study, we have developed a Random Forest algorithm-based

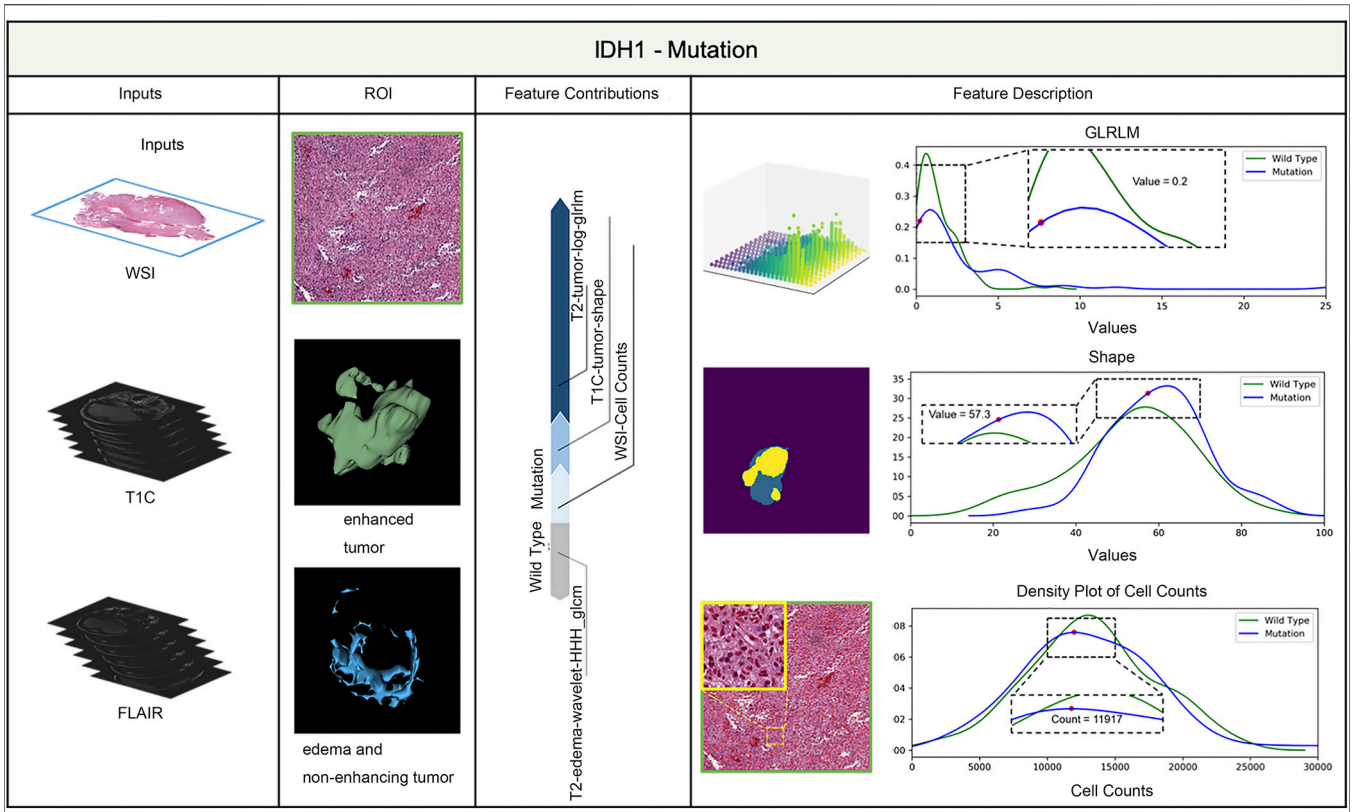


FIGURE 3 | IDH status prediction result explanation of a representative case with the LIME algorithm for the RF model (IDH-Wild type case).

genotype classifier that allows the prediction of IDH mutation status in glioma patients from pre-surgical MRI scans (Zhang et al., 2016) with improved accuracy. The Random Forest algorithm-based genotype classifier aims to employ the machine-learning algorithm to do the IDH genotype and a stable and efficient prediction result of IDH genotype. In this situation, the Random Forest machine-learning model with 10-fold cross-validation was implement into this experiment. To be more specific, we apply the RF algorithm into the experiment due to the following advantages: I) in specific experiments, training can be highly parallelized and run efficiently on large data sets; II) since the partition features of decision tree nodes can be selected randomly, which leads to the input samples with high-dimensional features can be processed without dimensionality reduction; III) the algorithm is able to calculate the importance of each feature to the prediction result and IV) due to the adoption of random sampling and random feature selection, the model has small variance at the training location and strong induction ability. We adopt the RF into the experiment, due to the advantages the model have which match the height of our datasets.

In order to improve the accuracy of IDH phenotype prediction, visual and sub-visual features extracted from digital histopathological images and quantitative radiomics feature extracted from different multimodality MRIs were implemented into our Random-Forest-Recursive Feature Elimination (RF-RFE) feature selection model to identify optimal criteria for further analysis. In this task, the "visual features" refer to not only the basic features including the color and appearance of nuclear staining, but also non-basic features including different directions. On the other hand, the "sub-visual features" allude to the computerized high-throughput first-order and second-order features, which includes intensity and texture information. In this experiment, features were extracted from different conduits of the H&E images, which aims to improve the prediction accuracy of the IDH phenotype.

Our novel integrated approach, which combined multi-parametric biomedical imaging features, was found to be a more accurate predictor of IDH genotype than either radiomics or histopathological feature recognition alone. Multi-parametric biomedical imaging characterizes tumor properties at different biological levels, it meets the need to understand correlations between image features, genomics, and clinical outcomes.

Specifically, the IDH predictive performance of histopathological images was found to be superior to T2-FLAIR and T1CE (0.86 vs. 0.71, 0.75). Among the leading histomorphometrical features, the mean cell area and the mean cell axis were most significant. These top identified features mirror the fact that gliomas with an IDH mutation have a more coherent nuclear architecture, i.e., they are

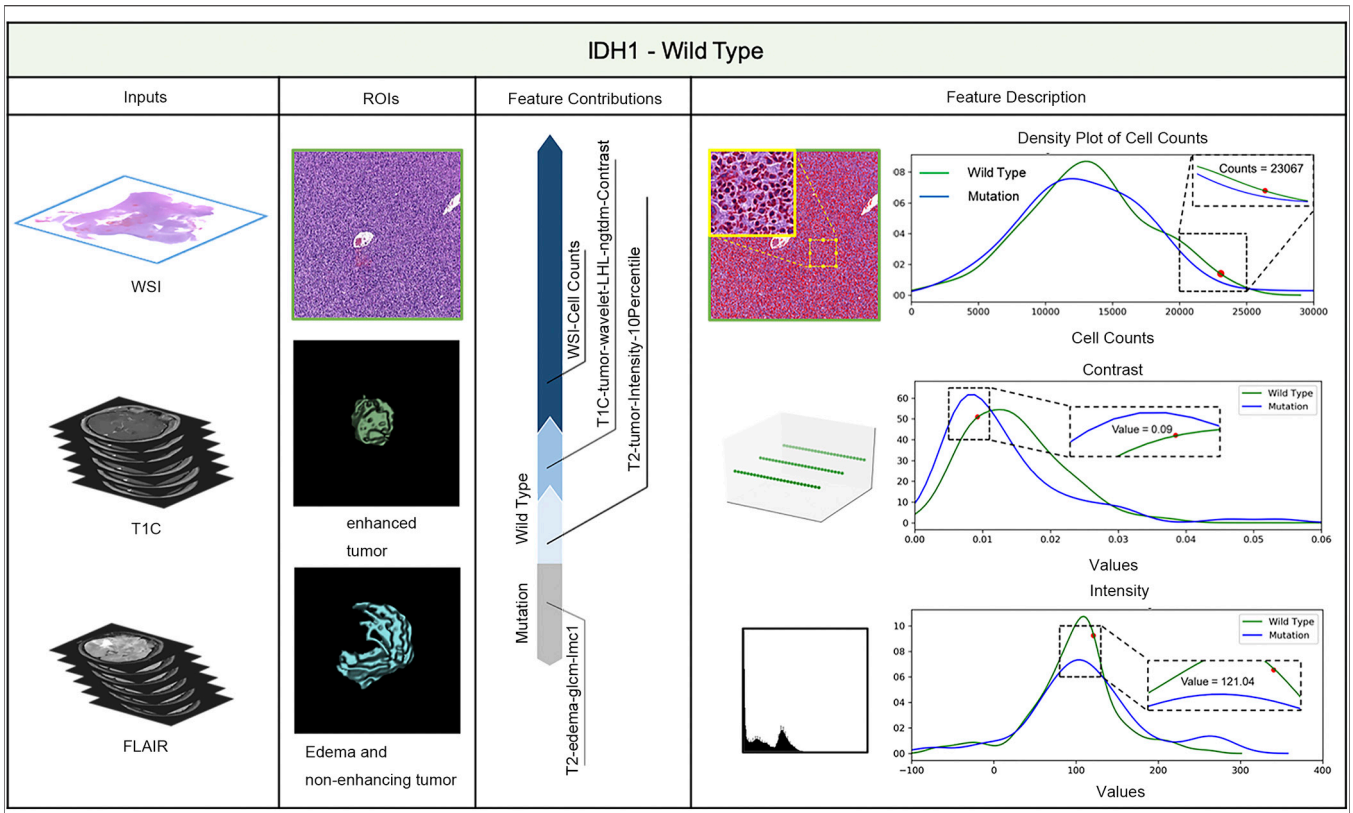


FIGURE 4 | IDH status prediction result explanation of a representative case with the LIME algorithm for the RF model (IDH-Mutation case).

morphologically less atypical than IDH wild type, which is associated with a higher risk of recurrence.

Our IDH genotype prediction achieved high accuracy for mainly two reasons: First, we integrated MRIs, digital histopathological images and clinical information for IDH prediction. Second, we used the selected features to significantly reduce the number of parameters in the model to avoid overfitting while making our model more robust. To the best of our knowledge, this is the first study to integrate MRI and digital pathology images in a computerized model for predicting IDH genotype. It is worth noting that T1CE and T2-FLAIR images conferred a higher predictive value than other MR sequences.

Conclusion

In conclusion, our work is a step towards a more effective use of radiomic and histopathological data. It should be particularly helpful for retrospective studies on gliomas where imaging results are available but also to point of care that do not have timely access to a molecular genetics laboratory. To sum up, our results i) demonstrate that machine-learning is capable of indirectly identifying genetic information within structural MR images and histopathological datasets, ii) suggest a complementary method for the IDH genotyping of gliomas suitable for patient screening, and iii) demonstrate the potential for algorithmic tools to support clinical decision-making. Taken together, it is expected that the integration of multimodal biomedical data analysis will become more popular in oncology research and practice as technology evolves, with significant potential for the future clinical management of brain tumor patients.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

This study was approved by the institutional review board (IRB) of Shandong Provincial Hospital affiliated to Shandong First Medical University (SFMU) and conducted in accordance with ethical guidelines, including the Declaration of Helsinki and Belmont Report. Informed consent was waived for the TCIA given the use of existing public datasets; and for the 126 patients from Shandong First Medical University (SFMU), written informed consents to participate in this study was provided by the patient/next of kin.

AUTHOR CONTRIBUTIONS

Conceptualization, XW and YL; Data curation, DW, CLi, XL, CLa, and PZ; Formal analysis, XW, XL, and MG; Funding acquisition, YL; Investigation, DW, XW, CLi, XL, CLa, PZ, WCC and YL; Methodology, DW and XW; Project administration, XW and YL; Resources, CLi, XL, CLa, and PZ; Software, DW and XL; Supervision, XW and YL; Validation, DW; Visualization, DW; Writing – original draft, DW, XW, MG, and YL; Writing – review and editing, XW, WCC, MG, and YL. All authors have read and agreed to the published version of the manuscript.

FUNDING

This work was supported in part to the Natural Science Foundation of Shandong Province (ZR2019QF007) and the Taishan Scholars Program (No. tsqn20161070).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbinf.2021.718697/full#supplementary-material

REFERENCES

Al-Kofahi, Y., Lassoued, W., Lee, W., and Roysam, B. (2010). Improved Automatic Detection and Segmentation of Cell Nuclei in Histopathology Images. *IEEE Trans. Biomed. Eng.* 57 (4), 841–852. doi:10.1109/TBME.2009.2035102

Balagurunathan, Y., Kumar, V., Gu, Y., Kim, J., Wang, H., Liu, Y., et al. (2014). Test-retest Reproducibility Analysis of Lung CT Image Features. *J. Digit Imaging* 27 (6), 805–823. doi:10.1007/s10278-014-9716-x

Chang, P., Grinband, J., Weinberg, B. D., Bardis, M., Khy, M., Cadena, G., et al. (2018). Deep-learning Convolutional Neural Networks Accurately Classify Genetic Mutations in Gliomas. *AJNR Am. J. Neuroradiol.* 39 (7), 1201–1207. doi:10.3174/ajnr.A5667

Choi, K. S., Choi, S. H., and Jeong, B. (2019). Prediction of IDH Genotype in Gliomas with Dynamic Susceptibility Contrast Perfusion MR Imaging Using an Explainable Recurrent Neural Network. *Neuro-oncology* 21, 1197–1209. doi:10.1093/neuonc/noz095

Deo, R. C. (2015). Machine Learning in Medicine. *Circulation* 132 (20), 1920–1930. doi:10.1161/CIRCULATIONAHA.115.001593

Eichinger, P., Alberts, E., Delbridge, C., Trebeschi, S., Valentinitsch, A., Bette, S., et al. (2017). Diffusion Tensor Image Features Predict IDH Genotype in Newly Diagnosed WHO Grade II/III Gliomas. *Sci. Rep.* 7 (1), 13396. doi:10.1038/s41598-017-13679-4

Ellingson, B. M., Malkin, M. G., Rand, S. D., LaViolette, P. S., Connelly, J. M., Mueller, W. M., et al. (2011). Volumetric Analysis of Functional Diffusion Maps Is a Predictive Imaging Biomarker for Cytotoxic and Anti-angiogenic Treatments in Malignant Gliomas. *J. Neurooncol.* 102 (1), 95–103. doi:10.1007/s11060-010-0293-7

Emblem, K. E., Pinho, M. C., Zollner, F. G., Due-Tonnessen, P., Hald, J. K., Schad, L. R., et al. (2014). A Generic Support Vector Machine Model for Preoperative Glioma Survival Associations. *Radiology* 275 (1), 228–234. doi:10.1148/radiol.14140770

Farahani, N., Anil, V. P., and Pantanowitz, L. (2015). Whole Slide Imaging in Pathology: Advantages, Limitations, and Emerging Perspectives. *Pathol. Lab. Med. Int.* 7 (23-33), 4321. doi:10.2147/PLMI.S59826

Gillies, R. J., Kinahan, P. E., and Hricak, H. (2015). Radiomics: Images Are More Than Pictures, They Are Data. *Radiology* 278 (2), 563–577. doi:10.1148/radiol.2015151169

Gillies, R. J., Kinahan, P. E., and Hricak, H. (2016). Radiomics: Images Are More Than Pictures, They Are Data. *Radiology* 278 (2), 563–577. doi:10.1148/radiol.2015151169

Griffin, J., and Treanor, D. (2017). Digital Pathology in Clinical Use: where Are We Now and what Is Holding Us Back?. *Histopathology* 70 (1), 134–145. doi:10.1111/his.12993

Gutman, D. C., and Young, R. J. (2021). *IDH Glioma Radiogenomics in the Era of Deep Learning*. Oxford University Press US.

Guyon, I., and Elisseeff, A. (2003). An Introduction to Variable and Feature Selection. *J. machine Learn. Res.* 3 (Mar), 1157–1182. doi:10.1162/153244303322753616

Harezlak, K., and Kasprowski, P. (2018). Application of Eye Tracking in Medicine: A Survey, Research Issues and Challenges. *Comput. Med. Imaging Graph* 65, 176–190. doi:10.1016/j.compmedimag.2017.04.006

Kesler, S. R., Harrison, R. A., Petersen, M. L., Rao, V., Dyson, H., Alfaro-Munoz, K., et al. (2019). Pre-surgical Connectome Features Predict IDH Status in Diffuse Gliomas. *Oncotarget* 10 (60), 6484–6493. doi:10.18632/oncotarget.27301

Kinjo, S., Hirato, J., and Nakazato, Y. (2008). Low Grade Diffuse Gliomas: Shared Cellular Composition and Morphometric Differences. *Neuropathology* 28 (5), 455–465. doi:10.1111/j.1440-1789.2008.00897.x

Kumar, N., Verma, R., Sharma, S., Bhargava, S., Vahadane, A., and Sethi, A. (2017). A Dataset and a Technique for Generalized Nuclear Segmentation for Computational Pathology. *IEEE Trans. Med. Imaging* 36 (7), 1550–1560. doi:10.1109/TMI.2017.2677499

Lee, M. H., Kim, J., Kim, S. T., Shin, H. M., You, H. J., Choi, J. W., et al. (2019). Prediction of IDH1 Mutation Status in Glioblastoma Using Machine Learning Technique Based on Quantitative Radiomic Data. *World Neurosurg.* 125, e688–e696. doi:10.1016/j.wneu.2019.01.157

Lopes, M. B. S. (2017). The 2017 World Health Organization Classification of Tumors of the Pituitary Gland: a Summary. *Acta Neuropathol.* 134 (4), 521–535. doi:10.1007/s00401-017-1769-8

Louis, D. N., Ohgaki, H., Wiestler, O. D., and Cavenee, W. K. (2016b). *World Health Organization Histological Classification of Tumours of the Central Nervous System*. Lyon, France: International Agency for Research on Cancer (IARC).

Louis, D. N., Perry, A., Reifenberger, G., Von Deimling, A., Figarella-Branger, D., Cavenee, W. K., et al. (2016a). The 2016 World Health Organization Classification of Tumors of the central Nervous System: a Summary. *Acta Neuropathol.* 131 (6), 803–820. doi:10.1007/s00401-016-1545-1

Louis, D. N., Perry, A., Wesseling, P., Brat, D. J., Cree, I. A., Figarella-Branger, D., et al. (2021). The 2021 WHO Classification of Tumors of the Central Nervous System: A Summary. *Neuro-Oncology* 23 (8), 1231–1251. doi:10.1093/neuonc/noab106

Macyszyn, L., Akbari, H., Pisapia, J. M., Da, X., Attiah, M., Pigrish, V., et al. (2015). Imaging Patterns Predict Patient Survival and Molecular Subtype in Glioblastoma via Machine Learning Techniques. *Neuro Oncol.* 18 (3), 417–425. doi:10.1093/neuonc/nov127

Missbach-Guentner, J., Pinkert-Leetsch, D., Dullin, C., Ufartes, R., Hornung, D., Tampe, B., et al. (2018). 3D Virtual Histology of Murine Kidneys -high Resolution Visualization of Pathological Alterations by Micro Computed Tomography. *Sci. Rep.* 8 (1), 1407–1414. doi:10.1038/s41598-018-19773-5

Mobadersany, P., Yousefi, S., Mohamed, A., Gutman, D. A., Barnholtz-Sloan, J. S., José, E. V., et al. (2018). "Predicting Cancer Outcomes from Histology and Genomics Using Convolutional Networks," in Proceedings of the National Academy of Sciences: 201717139 (Springer). doi:10.1073/pnas.1717139115

Nance, J., Meenan, C., Nagy, P. G., and Nagy, P. G. (2013). The Future of the Radiology Information System. *AJR Am. J. Roentgenol.* 200 (5), 1064–1070. doi:10.2214/AJR.12.10326

Perizzolo, M., Winkfein, B., Hui, S., Krulicki, W., Chan, J. A., and Demetrick, D. J. (2012). IDH Mutation Detection in Formalin-Fixed Paraffin-Embedded

Gliomas Using Multiplex PCR and Single-Base Extension. *Brain Pathol.* 22 (5), 619–624. doi:10.1111/j.1750-3639.2012.00579.x

Rathore, S., Niazi, T., Iftikhar, M. A., and Chaddad, A. (2020). Glioma Grading via Analysis of Digital Pathology Images Using Machine Learning. *Cancers (Basel)* 12 (3), 578. doi:10.3390/cancers12030578

Reyes-Botero, G., Giry, M., Mokhtari, K., Labussière, M., Idbaih, A., Delattre, J. Y., et al. (2014). Molecular Analysis of Diffuse Intrinsic Brainstem Gliomas in Adults. *J. Neurooncol.* 116 (2), 405–411. doi:10.1007/s11060-013-1312-2

Saeys, Y., Inza, I., and Larranaga, P. (2007). A Review of Feature Selection Techniques in Bioinformatics. *bioinformatics* 23 (19), 2507–2517. doi:10.1093/bioinformatics/btm344

Sanson, M., Marie, Y., Paris, S., Idbaih, A., Laffaire, J., Ducray, F., et al. (2009). Isocitrate Dehydrogenase 1 Codon 132 Mutation Is an Important Prognostic Biomarker in Gliomas. *J. Clin. Oncol.* 27 (25), 4150–4154. doi:10.1200/JCO.2009.21.9832

Sertel, O., Kong, J., Shimada, H., Catalyurek, U. V., Saltz, J. H., and Gurcan, M. N. (2009). Computer-aided Prognosis of Neuroblastoma on Whole-Slide Images: Classification of Stromal Development. *Pattern Recognit.* 42 (6), 1093–1103. doi:10.1016/j.patcog.2008.08.027

Tatekawa, H., Hagiwara, A., Uetani, H., Bahri, S., Raymond, C., Lai, A., et al. (2021). Differentiating IDH Status in Human Gliomas Using Machine Learning and Multiparametric MR/PET. *Cancer Imaging* 21 (1), 27–10. doi:10.1186/s40644-021-00396-5

Vågberg, W., Persson, J., Szekely, L., and Hertz, H. M. (2018). Cellular-resolution 3D Virtual Histology of Human Coronary Arteries Using X-ray Phase Tomography. *Sci. Rep.* 8 (1), 11014–11017. doi:10.1038/s41598-018-29344-3

Wang, X., Wang, D., Yao, Z., Xin, B., Wang, B., Lan, C., et al. (2018). Machine Learning Models for Multiparametric Glioma Grading with Quantitative Result Interpretations. *Front. Neurosci.* 12, 1. doi:10.3389/fnhum.2018.00164

Xing, Z., Yang, X., She, D., Lin, Y., Zhang, Y., and Cao, D. (2017). Noninvasive Assessment of IDH Mutational Status in World Health Organization Grade II and III Astrocytomas Using DWI and DSC-PWI Combined with Conventional MR Imaging. *AJNR Am. J. Neuroradiol.* 38 (6), 1138–1144. doi:10.3174/ajnr.A5171

Zacharaki, E. I., Morita, N., Bhatt, P., O'rourke, D. M., Melhem, E. R., and Davatzikos, C. (2012). Survival Analysis of Patients with High-Grade Gliomas Based on Data Mining of Imaging Variables. *AJNR Am. J. Neuroradiol.* 33 (6), 1065–1071. doi:10.3174/ajnr.A2939

Zhang, B., Chang, K., Ramkissoon, S., Tanguturi, S., Bi, W. L., Reardon, D. A., et al. (2016). Multimodal MRI Features Predict Isocitrate Dehydrogenase Genotype in High-Grade Gliomas. *Neuro Oncol.* 19 (1), 109–117. doi:10.1093/neuonc/now121

Zhou, H., Vallières, M., Bai, H. X., Su, C., Tang, H., Oldridge, D., et al. (2017). MRI Features Predict Survival and Molecular Markers in Diffuse Lower-Grade Gliomas. *Neuro Oncol.* 19 (6), 862–870. doi:10.1093/neuonc/now256

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.

frontiers in Oncology

REVIEW
published: 10 November 2021
doi: 10.3389/fonc.2021.763527

Evaluating Cancer-Related Biomarkers Based on Pathological Images: A Systematic Review

Xiaoliang Xie[1,2†], Xulin Wang[3†], Yuebin Liang[4,5], Jingya Yang[4,5,6], Yan Wu[4,5], Li Li[7], Xin Sun[8], Pingping Bing[9], Binsheng He[9], Geng Tian[4,5,10] and Xiaoli Shi[4,5*]*

[1] Department of Colorectal Surgery, General Hospital of Ningxia Medical University, Yinchuan, China, [2] College of Clinical Medicine, Ningxia Medical University, Yinchuan, China, [3] Department of Oncology Surgery, Central Hospital of Jia Mu Si City, Jia Mu Si, China, [4] Geneis Beijing Co., Ltd., Beijing, China, [5] Qingdao Geneis Institute of Big Data Mining and Precision Medicine, Qingdao, China, [6] School of Electrical and Information Engineering, Anhui University of Technology, Ma'anshan, China, [7] Beijing Shanghe Jiye Biotech Co., Ltd., Bejing, China, [8] Department of Medical Affairs, Central Hospital of Jia Mu Si City, Jia Mu Si, China, [9] Academician Workstation, Changsha Medical University, Changsha, China, [10] IBMC-BGI Center, The Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital), Institute of Basic Medicine and Cancer (IBMC), Chinese Academy of Sciences, Hangzhou, China

OPEN ACCESS

Edited by:
Min Tang,
Jiangsu University, China

Reviewed by:
Min Chen,
Hunan Institute of Technology, China
Li Peng,
Hunan University of Science and Technology, China

***Correspondence:**
Geng Tian
tiang@geneis.cn
Xiaoli Shi
shixl@geneis.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *24 August 2021*
Accepted: *18 October 2021*
Published: *10 November 2021*

Citation:
Xie X, Wang X, Liang Y, Yang J, Wu Y, Li L, Sun X, Bing P, He B, Tian G and Shi X (2021) Evaluating Cancer-Related Biomarkers Based on Pathological Images: A Systematic Review. Front. Oncol. 11:763527. doi: 10.3389/fonc.2021.763527

Many diseases are accompanied by changes in certain biochemical indicators called biomarkers in cells or tissues. A variety of biomarkers, including proteins, nucleic acids, antibodies, and peptides, have been identified. Tumor biomarkers have been widely used in cancer risk assessment, early screening, diagnosis, prognosis, treatment, and progression monitoring. For example, the number of circulating tumor cell (CTC) is a prognostic indicator of breast cancer overall survival, and tumor mutation burden (TMB) can be used to predict the efficacy of immune checkpoint inhibitors. Currently, clinical methods such as polymerase chain reaction (PCR) and next generation sequencing (NGS) are mainly adopted to evaluate these biomarkers, which are time-consuming and expansive. Pathological image analysis is an essential tool in medical research, disease diagnosis and treatment, functioning by extracting important physiological and pathological information or knowledge from medical images. Recently, deep learning-based analysis on pathological images and morphology to predict tumor biomarkers has attracted great attention from both medical image and machine learning communities, as this combination not only reduces the burden on pathologists but also saves high costs and time. Therefore, it is necessary to summarize the current process of processing pathological images and key steps and methods used in each process, including: (1) pre-processing of pathological images, (2) image segmentation, (3) feature extraction, and (4) feature model construction. This will help people choose better and more appropriate medical image processing methods when predicting tumor biomarkers.

Keywords: histopathological image analysis, cancer biomarker, deep learning, color normalization, feature extraction

INTRODUCTION

Biomarkers are critical in cancer diagnosis, treatment, and prognosis. They can be used for patient's evaluation in a variety of clinical settings, such as risk assessment, early diagnosis, drug effect evaluation, and prognosis prediction (1–3). With the development of immunology, molecular biology and genomics, studies of cancer biomarkers have attracted a lot of attention in recent years (4). Currently, biomarker identification usually employs technologies such as PCR, NGS and gene expression arrays (5). However, the data generated by these technologies need to be analyzed and interpreted manually. In addition, this kind of test usually costs a lot of money. For example, the test of tumor mutation burden (TMB) usually costs more than one thousand dollars. Thus, it will be of great value to develop a more intelligent and economical method in tumor biomarker identification (6).

Pathological image analysis is used to solve problems related to medical images which were applied in biomedical research and diagnosis. Its main objective is to extract clinically relevant physiological and pathological information or knowledge from images, and its main research direction is image segmentation, classification, and retrieval (7). With the rapid development and popularization of medical imaging technology, the amount of medical image data is growing rapidly. It will provide important and beneficial support for nursing and medical research to extract useful knowledge and information automatically from massive medical image data for clinical diagnosis and treatment (8). Recently, researchers have paid much attention to the analysis and study of tumor patients through pathological images and morphology (9). Mobadersany (10) proposed that the morphological characteristics of tumor tissue images could reflect the genetic and molecular characteristics and predict the degree of tumor deterioration, and the deep learning method could be used to integrate the morphological characteristics of tumor tissue images and genomics to predict the survival rate of glioma patients. Xu (11) proposed a method based on deep tissue network to automatically distinguish 10 tissue components in the colorectal full-scan tissue image. Yu (12) for the first time constructed the recurrence risk prediction model of LUAD and LUSC by automatically extracting morphological features from the full-scan histopathological images of lung cancer to provide prognostic information for patients. Vaidya (13) proposed to combine radiology and pathology to predict the risk of early lung cancer recurrence, with an accuracy rate of 70%. Wu (14) and others constructed a deep convolutional neural network framework to evaluate the risk of lung cancer recurrence and metastasis from histopathology images, with the area under the receiver operating characteristic (ROC) curve (AUC) in the test dataset of 0.79. Jain and Massoud explored a machine learning algorithm named Image2TMB to predict TMB from lung adenocarcinoma histopathological images. Its average precision was 0.89 and achieved predictive level of a panel of ~100 genes. Microsatellite instability (MSI) was another immunotherapy biomarker (15) which requires additional immunohistochemical or genetic analyses in clinical practice (16). Kather et al. developed a deep residual learning method that can predict MSI status directly from hematoxylin and eosin (H&E) stained histology slides (17). These findings suggest that inferring genomic features from histopathological images is possible and analyzing histopathological images is important for studying cancer treatments, mutated gene expression status, cancer prognosis and risk of recurrence.

However, full-scan histopathological images are highly complex, with large image size and about 2 GB of storage space after compression. It is a big challenge for hardware and image analysis algorithm to use computer to process image directly in this kind of high resolution and large size image. At the same time, the histopathological structure types in the images are disordered, and the histological morphology is very different, so it is difficult to describe with fixed features. All these factors bring great difficulty to the processing of full scan histopathological images. Based on the above problems, this paper summarizes the whole process and key steps of current pathological image processing, including image preprocessing, image segmentation, feature extraction and model construction, to help researchers choose more suitable medical image processing methods and predict biomarkers more accurately. We summarized the overall flow of pathological image processing in **Figure 1**.

IMAGE PREPROCESSING

The biggest obstacle to histopathological image analysis is the difference in image morphology due to high heterogeneity of the disease itself. At the same time, improper tissue treatment or staining during the slice preparation will result in morphological changes of cells and tissues, making it difficult to identify its original structure. In addition, the background noise and the lack of contrast caused by the different light source conditions were also important factors. Proper preprocessing method can correct images by eliminating irrelevant information, and filter out interference and noise, which can improve the detectability of target information and simplify the calculation to the maximum extent.

Common preprocessing methods such as using spatial filtering techniques to enhance the main structure in the image, image enhancement can improve the contrast between the region of interest and the background, and color normalization can reduce the effect of staining batches (18, 19). Among these, color normalization is the most commonly used image preprocessing methods for evaluating cancer-related biomarkers based on histopathological images.

Color Normalization

In response to the problem of color change, Reinhard (20) and others proposed a method of color normalization, that is, in the lαβ color space, the mean and standard deviation of each channel in the image are compared with the target by a set of linear transformations. Then, match the mean and standard deviation. However, if multiple patches are used, the assumption of a unimodal distribution of pixels in each channel of the lαβ color space is not valid. Therefore, this may cause the background area to be mapped as a colored area and

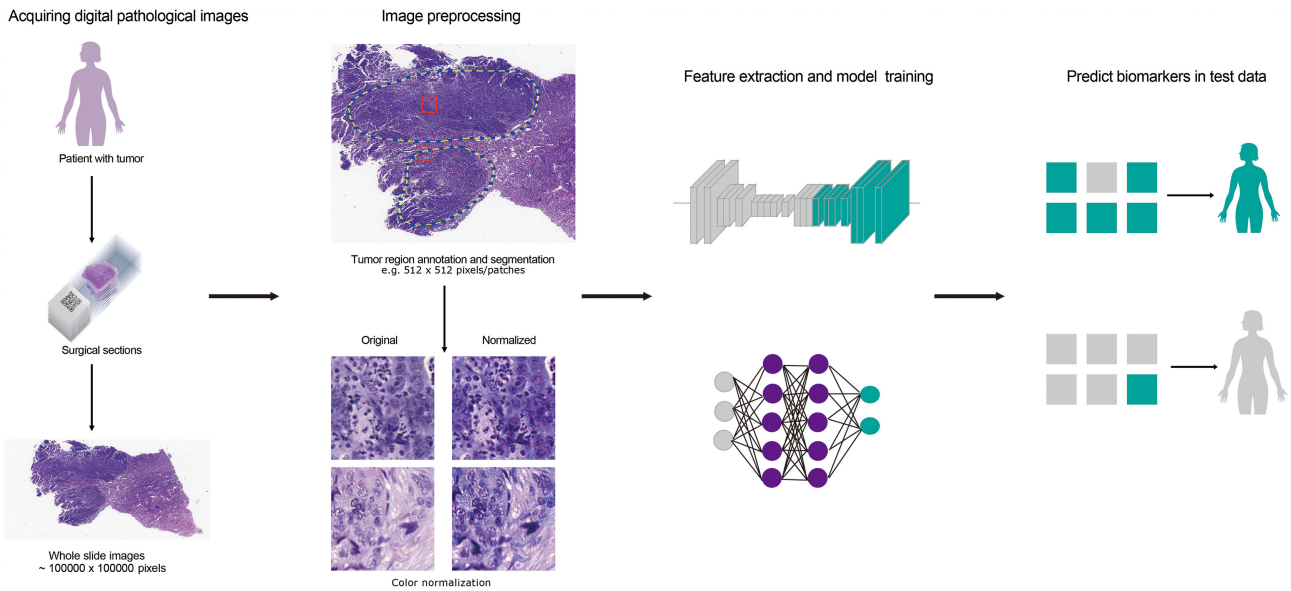


FIGURE 1 | The flow chart for predicting cancer-related biomarkers based on digital pathological images. Firstly, H&E stained histology slides of patients were obtained and whole slide images (WSIs) was obtained after scanning. Secondly, tumor regions were annotated by pathologists or through CNN model. Then, the regions were segmented to patches and color-normalized. Thirdly, feature extraction and model training were carried out according to biomarker labels. Finally, biomarker prediction was implemented in test dataset.

the foreground to be incorrectly mapped. As shown in **Table 1** below, some methods of color normalization were summarized.

IMAGE SEGMENTATION

Medical image segmentation is a complex and critical step in the field of medical image processing and analysis. The purpose of this process is to segment certain parts of the medical image with specific meaning, extract relevant features, and then provide reliable information for clinical diagnosis and pathological research. The two most common types of medical image segmentation are tissue segmentation and cell segmentation.

Tissue Segmentation

Pathologists have identified that degree of structural differentiation of the tissue is one of the earliest prognostic factors for breast cancer patients. Cancer destroys the ability of the nucleus to communicate with each other and causes it to organize itself into structures such as tubules, thereby making the tubules lack of indicators of advanced malignant tumors. Tubules are usually round or oval in structure and consist of a lumen surrounded by a layer of epithelial cells. The main challenge of tubule segmentation is that it has a similar appearance to other structures, such as the tearing of adipose tissue formed during tissue preparation, and the outer layer of well-arranged epithelial cell with nuclei missing.

TABLE 1 | A summary of color normalization methods.

Authors	Methods	Characteristics	References
Magee	A method based on supervised pixel classification	Estimate the color of the coloring.	(21)
Macenko	A method based on singular value decomposition (SVD)	Direct estimation matrix.	(22)
Niethammer	An improved method based on singular value decomposition (SVD)	By expanding (22), a priori estimation of staining matrix is used to improve the stability of each staining.	(23)
Khan	Nonlinear mapping based on source image to target image	An improvement is proposed on the method of (21), using the representation method of color deconvolution.	(18)
Vahadane	A technique of dye separation and color normalization (SPCN)	It does a good job of maintaining the quality of biological structure and the number of stains.	(24)
Ramakrishnan	The improved SPCN	In the SPCN technology, some improvements are proposed for the occasional errors in estimating color bases, which lead to artifacts.	(25)

For glandular segmentation, most of the early attempts used hand-made features for segmentation. Wu (26) identified the initial seed region based on large cavity regions and extended the seed to the surrounding epithelial nuclear chain. Farjam (27) proposed using a variance filter to compute cluster texture features for segmentation. However, robust segmentation requires more domain knowledge, and calculating texture features only using the variance filter may not provide sufficient information for the local structure of the organization. Naik (28) used a Bayesian classifier to detect the lumen region, and then used the kernel-based level set to stop the curve and refine it. Although this method has been reported to work well in benign cases, it may fail in malignant cases with fairly complex glands. Nguyen (29) used color space analysis to group the nucleus, cytoplasm, and lumen, and increased the lumen area to achieve segmentation under constraints. Gunduz-Demir (30) represented each tissue component as a disc and connected nearby discs with an edge to construct a graph. They performed area growth on a cavity disc bounded by a line connected to the nuclear disc. Nosrati, Hamarneh (31) and Cohen (32) first divided the tissue area into different components, and then used a constrained level set algorithm to segment the glands. Sirinukunwattana (33) identified epithelial superpixels and used the epithelial region as the vertex of a polygon, which approximated the boundary of a gland. Most of the methods discussed above first distinguish tissue regions and then use region growth or level sets to segment glandular regions. Recently, a slightly different approach that first used background information to identify potential epithelial regions, and then used multi-resolution cell localization descriptors to identify connected epithelial cells to segment glands was proposed by Li (34).

Cell Segmentation

The morphology of cells in histopathological images provides important information for the diagnosis and prognosis of cancer. Researchers at home and abroad have tried a variety of algorithms to solve the problem of cell segmentation in H&E images (34–36). The algorithms generally divided into two categories, one is to detect single cells accurately and the other is to segment cells. The algorithms in **Table 2** is commonly used to detect the appropriate seed point or contour of the nucleus.

The other type detects the candidate area of the cell and then divides it into individual nuclei. The first step in morphological analysis of a cell is the segmentation of individual nuclei, which is usually performed manually in current clinical practice. However, due to the large volume of histopathological images and complex cell structures, manual examination is a time-consuming and labor-intensive task. It is necessary to study computerized methods to reduce the workload of pathologists and improve the analysis efficiency (45). Nuclear segmentation tasks still have some major challenges. First, different types of organs or cells are highly heterogeneous in appearance. Therefore, the method based on prior knowledge of geometric features cannot be directly applied to different images. Second, some other structures, such as the cytoplasm and cell matrix, may have similar characteristics to the nucleus, making it difficult to distinguish the nucleus from the background. Third, the cells are often stacked together. In order to find the exact location and boundary of each nucleus, it is usually necessary to perform the next step to separate the clustered or overlapped nuclei.

In view of the importance of nuclear distribution and morphology, the task of using computer algorithms for accurate nuclear segmentation provides a logical starting point for computer-aided tissue image analysis. The precise segmentation of the nucleus can not only perform deeper level feature extraction and classification in the nucleus, but also serve as a relatively simple distribution of basal cells and acellular cells. Many techniques have been applied to the task of nuclear segmentation, but in some cases they have only achieved partial success. For example, the intensity threshold method usually fails due to image noise and nucleus clustering. Label-based watershed segmentation requires accurate parameter selection, while the computational cost of active contours and deformable models is too high (24, 42, 46–50). Machine learning-based kernel segmentation methods are generally better at meeting these challenges because they can learn to recognize changes in nuclear morphology and staining patterns. More precisely, convolutional neural networks (CNNs) have recently demonstrated their latest performance in kernel segmentation (51, 52). Ciregan (53) applied deep CNN to the automatic detection of mitotic cells in breast cancer histological images. Using the original intensity of the test image, CNN provides a probability map where each pixel value is the probability of the mitotic cell centroid. Then using the disk to check the probability map for smoothing, and non-maximum suppression to get the final centroid detection. Xing (54) and others respectively learned three different CNN models corresponding to pathological images of brain tumors, pancreatic neuroendocrine tumors, and breast cancer, and

TABLE 2 | A summary of methods on segmentation after detection of individual cells.

Methods	Characteristics	References
Based on different voting rules	Simple and suitable for segmentation of most images	(11, 37–39)
Based on Laplace operator and gaussian filter	Accurately detect the edge of the cell	(40)
Based on H-minima transformation	Effectively restrain oversegmentation and reduce undersegmentation	(41)
Based on Morphologic manipulation	Could output an image by acting a structure element on the input image	(42, 43)
Based on back propagation with MRF	Good at dealing with the problems of image local volume and artifacts	(34)
Based on the active contour model	Could convert pixels to a distance field	(43)
Based on the level set	A numerical method based on the theory of geometric active contour model	(37, 44)

applied them to automatic nuclear detection. Liu and Yang (55) did not use simple non-maximum suppression to refine the detection, but converted the detection problems of pancreatic neuroendocrine and lung cancer cell nuclear into optimization problems. Xing (47), Sirinukunwattana (51) and Song (55) have proposed some advanced techniques in nuclear detection and segmentation, which estimate the probability of nuclear and non-nuclear regions (both types) based on the learned nuclear phenomena graphs and rely on complex post-processing methods to obtain the final core shape and the separation between contacting nuclei. Song et al. used a graph partitioning method (55) and Xing et al. used a kernel mapping distance transformation, followed by H-minima thresholding and region growth (47). Although different methods have been developed for the problem of overlapping and clustering nuclei in many literatures, and have achieved varying degrees of success, this problem has not been completely solved, as there is a large amount of overlap contact specimens of nuclei.

In addition, a special type of nucleus, mitosis, has attracted much attention in the field of image analysis. Mainly because the mitotic index is used to evaluate the cell proliferation rate of cancer cells, it could predict the prognosis of invasive breast cancer well, but its evaluation process is extremely time-consuming (56). On the H&E image, mitosis has specific morphological features: dense nuclear staining, enlarged nuclei, less clear nuclear membrane, and burr-like edges. Researchers such as Belien (57) proposed image processing technology for semi-automatic segmentation of mitotic images in the 1990s. Due to the limitations of the image quality and machine learning algorithms at the time, the algorithm proposed by Belien et al. (57) required fourgen staining to display chromosomes, and the false positive rate is 19-42%. With the digitization of pathological images, two H&E tissue datasets of breast cancer have been published internationally, and pathological experts have annotated mitotic images in the images, which has greatly promoted the development of algorithms in mitotic image segmentation. Then, the International Conference on Pattern Recognition (ICPR) (58) held a competition for mitotic detection in breast cancer tissue images in 2012, providing different types of images, allowing participants to analyze classic images of H&E stained sections, and use 10 bands multispectral microscope images, which may be more discriminatory for detecting mitosis. Deep learning maximizing CNN significantly outperforms other manual feature-based methods and paves the way for future use of CNNs (53).

The biggest challenge for mitosis detection is that apoptosis, necrosis or squeezed nuclei and lymphocyte nuclei have similar morphology to mitosis, which is difficult for even experienced pathologists to identify. In addition, pathologists need to observe suspicious split images on multiple focal planes, while currently digital images are single focal plane imaging. Although some scanners can acquire multifocal plane images, their storage capacity is large and cannot be widely used. We expect that in the future, as storage costs decrease and new image compression technologies emerge, this limitation will be eliminated (59). Therefore, the automatic segmentation of mitotic images in H&E images at this stage is more challenging than general nuclear segmentation and is far from being applicable to pathological work.

MODEL CONSTRUCTION

After the ideal segmentation results were obtained from the tissue segmentation and nuclear segmentation modules, the morphological features of histopathological images were extracted, and the correlation between the morphological features and biomarkers of the full-scan histopathological images was found and the feature model was established.

Beck et al. constructed a computer pathologist system to extract 6,642 dimensional features from H&E histopathological images of breast cancer (60). Some of the features are based on the existing knowledge system, such as the formation degree of counting glandular tube after automatic segmentation (61) and automatic grading (62), but most of the features go beyond the existing descriptive semantics of pathology. Computer-aided diagnosis is also based on the prognosis of characteristic models, modeling based on object characteristics, and then estimating the prognosis of model parameters. Tutac (63) proposed a semi-automatic grading system based on knowledge model for the first time, which automatically detected and measured the three components of histological grading, namely nucleus, adenotuine and mitosis, through semantic retrieval. The consistency of the scoring results of this model was higher than that of manual evaluation. Dalle (64) further improved the above work based on multi-resolution method and Gaussian model function, realized automatic histological classification, and the automatic classification results were highly consistent with the manual evaluation results.

Pathology is morphology-based, but the classification and assessment of disease is not limited to morphology, and requires reference to immunological, molecular, and clinical characteristics of patients. Based on the genome, Wang (65) mined prognostic features in H&E histopathological images of triple negative breast cancer (TNBC), and selected 48 pairs of significantly correlated image features and gene clusters through the TNBC genome map and H&E images of 44 cases, among which 4 pairs were significantly correlated with prognosis. Basavanhally (66) showed that H&E morphological characteristics and IHC molecular characteristics can replace expensive Oncotype DX risk assessment for the invasiveness of ER negative breast cancer. Yuan (67) proposed a mathematical statistical model to evaluate the proportion of lymphocytes in TNBC tumors, and the results showed that lymphocytes were related to the survival of TNBC, and the image-based evaluation results were similar to the results of gene expression spectrum detection. According to the prognostic model theory of Steyerberg (61), we can further utilize the results of image characteristics and molecular characteristics, and construct a prediction model by integrating complementary prognostic factors, which can be used to comprehensively and accurately predict the prognosis of breast cancer. Currently, integrating information from different dimensions to construct multimodal

fusion models for predicting cancer biomarkers or prognosis of patients have been studied in several laboratories. The main process of building multimodal fusion models is shown in **Figure 2**. Making full use of multidimensional information for fusion modeling is of great help to improve the prediction accuracy, which will also be a direction of the development of digital pathology. Chen et al. used CNNs and GCNs to extract morphological features from digital histology images and SNNs to extract genomic signatures (68). Then they employed the Kronecker Product and a gating-based attention mechanism to fuse these deep features and further validated the approach on glioma and clear cell renal cell carcinoma (CCRCC) data from TCGA. Mobadersany et al. presented a novel method to predict outcomes of patients from histopathological images and proved that the accuracy was comparable to the traditional manual histological grading. To further improve performance, they combined histopathological images and genomic data to develop a comprehensive model called GSCNN. And its performance was significantly better than that of SCNN model and WHO paradigm based on genomic subtype and histological grading (69).

LIMITATIONS AND FUTURE WORK

Cancer histology contains rich phenotypic information and can reflect underlying molecular mechanisms and disease progression. A large number of studies have shown that deep learning of digital pathological images of tumor tissue samples can be used for cancer diagnosis, classification, drug efficacy evaluation and prognosis prediction. This method has the advantage of fast and low cost. In this work, we summarized the overall process and key steps of processing full-scan section images to help people choose better and more appropriate medical image processing methods when predicting tumor biomarkers.

However, the application of artificial intelligence (AI) technology in precision medicine has some limitations currently. Firstly, the diagnosis process of deep learning model is fuzzy and the interpretability is limited, and the lack of interpretability is unacceptable to the Medical Association (70–72). So this problem is an important obstacle to its verification and application in clinical practice. Heat map analysis provides an in-depth understanding of the histological patterns related to the prediction target, which is helpful for the interpretation of the deep learning model. Chen et al. had used this method to locate and interpret features in the study of multimodal fusion for predicting survival outcome of cancer patients (68). It can also be used as a practical tool to lead pathologists to discover the tissue regions related to biomarkers. For example, the presence of edema in glioma was not previously considered as an adverse marker by pathologists, but was detected as a recognition feature in the model of predicting cancer prognosis (69). Associated with this finding, the degree of edema may be correlated to the growth rate of cancer in previous study (73). Cao et al. verified the reliability of the deep learning model in two independent cohorts when predicting MSI with pathological images, and explained the interpretability of the model by exploring the correlation between pathological features and multi-omics signatures. This is also a method to promote clinicians to accept the application of AI in digital pathological images (74). It can be predicted that improving the interpretability of the model or establishing interpretable machine learning methods will be an important topic to be explored in the future.

Secondly, a substantive problem limiting its clinical application is the frequent workflow switching due to the limited integration of computer-aided pathological diagnosis in the current pathological workflow (70–72). Currently, the

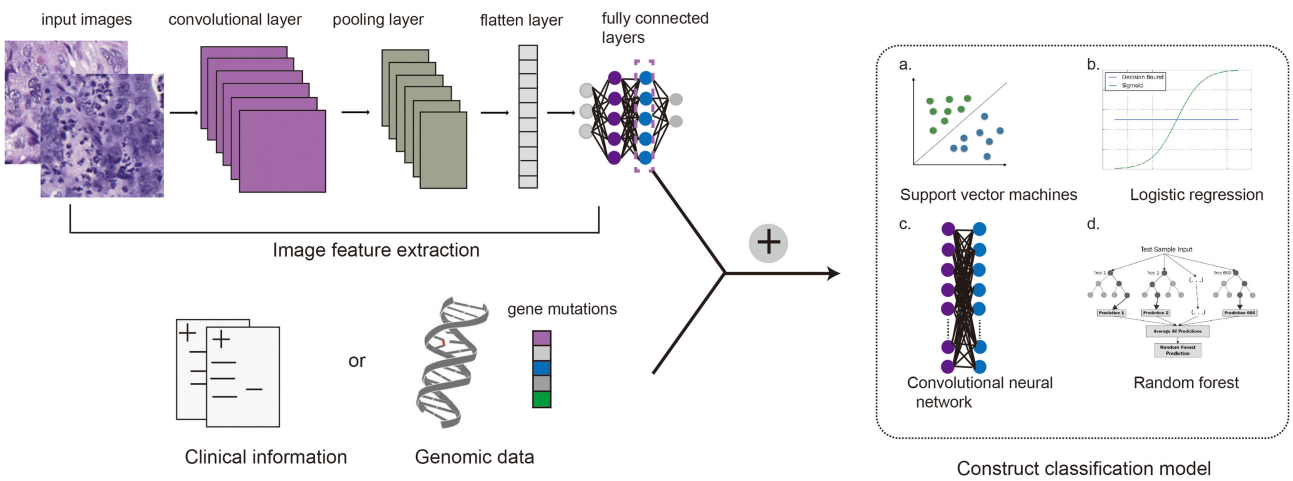


FIGURE 2 | Main process of constructing a compound framework by combining pathological images with genomic data or clinical information. Convolutional neural networks are commonly used to extract image features, and then genomic features or clinical information are integrated into the full connection layer. Support vector machine (a), logistic regression (b), convolutional neural network (c) or random forest (d) can be used to establish the final multimodal fusion model.

research on diagnosis and subtyping of cancer through digital pathological images is relatively mature. Some latest studies on predicting cancer prognosis, treatment response and disease progress monitoring through pathological images have been reported. Kather et al. developed a deep learning model that can directly predict microsatellite instability from H&E histological images of stomach and colorectal cancer and the AUC values ranged from 0.69 to 0.84 in independent validation datasets (17). Cao et al. explored an EPLA model with AUC of 0.8504 (95% CI: 0.7591-0.9323) in the external validation set (74). However, more histological images of patients are needed to optimize the model and improve accuracy. If a complete pathological diagnosis and prediction process through extensive analysis of various data can be established and verified clinically, it will contribute to the application of AI in precision medicine (71, 75).

Thirdly, it is difficult to unify the staining and imaging process of tissue section in different laboratories, which leads to a large number of variables in pathological images and further makes it difficult to establish models with high stability and good generalization performance. Just as molecular diagnosis relies on qualified samples and sequencing data, digital image analysis also requires strict control of sample quality, clear quality requirements for input files, and adequate training for pathologists. These requirements of digital pathological image analysis will also drive to improve the volume and accuracy of histomorphological evaluation. On the other hand, in order to promote clinical transformation, a roadmap and regulatory framework for the routine use of AI in pathology have been published (76).

Other literatures also list possible practical problems: slow implementation time of computer-aided pathology, insufficient clinical validation of computer-aided pathology, and limited impact on health economics (9, 71). The ability to overcome these limitations will determine the future of digital pathology.

AUTHOR CONTRIBUTIONS

GT and XShi designed the project. XX, XW, and XShi searched literatures and wrote the manuscript. YL, JY, YW, LL, XSun, PB, and BH revised the manuscript. All authors have approved the final version of the manuscript.

FUNDING

This study was supported by Natural Science Foundation of Hunan, China (Grant No. 2018JJ3570), Major Project for New Generation of AI (Grant No. 2018AAA0100400), the National Natural Science Foundation of Hunan (Grant Nos. 2018JJ2098), the National Natural Science Foundation of China (Grant No. 11571052, 11731012).

REFERENCES

1. Henry NL, Hayes DF. Cancer Biomarkers. *Mol Oncol* (2012) 6(2):140–6. doi: 10.1016/j.molonc.2012.01.010
2. Liu H, Qiu C, Wang B, Bing P, Tian G, Zhang X, et al. Evaluating DNA Methylation, Gene Expression, Somatic Mutation, and Their Combinations in Inferring Tumor Tissue-Of-Origin. *Front Cell Dev Biol* (2021) 9:619330. doi: 10.3389/fcell.2021.619330
3. He B, Lang J, Wang B, Liu X, Lu Q, He J, et al. TOOme: A Novel Computational Framework to Infer Cancer Tissue-of-Origin by Integrating Both Gene Mutation and Expression. *Front Bioeng Biotechnol* (2020) 8:394. doi: 10.3389/fbioe.2020.00394
4. Zhang Y, Huang H, Zhang D, Qiu J, Zhang J, Wang K, et al. A Review on Recent Computational Methods for Predicting Noncoding RNAs. *BioMed Res Int* (2017) 2017:9139504. doi: 10.1155/2017/9139504
5. Yang J, Hui Y, Zhang Y, Zhang M, Ji B, Tian G, et al. Application of Circulating Tumor DNA as a Biomarker for Non-Small Cell Lung Cancer. *Front Oncol* (2021) 11:725938. doi: 10.3389/fonc.2021.725938
6. Yang J, Peng S, Zhang B, Houten S, Schadt E, Zhu J, et al. Human Geroprotector Discovery by Targeting the Converging Subnetworks of Aging and Age-Related Diseases. *Geroscience* (2020) 42(1):353–72. doi: 10.1007/s11357-019-00106-x
7. Ma X, Xi B, Zhang Y, Zhu L, Sui X, Tian G, et al. A Machine Learning-Based Diagnosis of Thyroid Cancer Using Thyroid Nodules Ultrasound Images. *Curr Bioinf* (2020) 15(4):349–58. doi: 10.2174/1574893614666191017091959
8. Stålhammar G, Fuentes Martinez N, Lippert M, Tobin NP, Mølholm I, Kis L, et al. Digital Image Analysis Outperforms Manual Biomarker Assessment in Breast Cancer. *Modern Pathol* (2016) 29(4):318–29. doi: 10.1038/modpathol.2016.34
9. Acs B, Rantalainen M, Hartman J. Artificial Intelligence as the Next Step Towards Precision Pathology. *J Intern Med* (2020) 288(1):62–81. doi: 10.1111/joim.13030
10. Mobadersany P, Yousefi S, Amgad M, Gutman DA, Barnholtz-Sloan JS, Velázquez Vega JE, et al. Predicting Cancer Outcomes From Histology and Genomics Using Convolutional Networks. *Proc Natl Acad Sci USA* (2018) 115 (13):E2970–9. doi: 10.1073/pnas.1717139115
11. Xu J, Janowczyk A, Chandran S, Madabhushi A. A High-Throughput Active Contour Scheme for Segmentation of Histopathological Imagery. *Medical Image Analysis* (2011) 15(6):851–62. doi: 10.1016/j.media.2011.04.002
12. Yu KH, Zhang C, Berry GJ, Altman RB, Ré C, Rubin DL, et al. Predicting non-Small Cell Lung Cancer Prognosis by Fully Automated Microscopic Pathology Image Features. *Nat Commun* (2016) 7:12474. doi: 10.1038/ncomms12474
13. Vaidya P, Wang X, Bera K, Khunger A, Choi H, Patil P, et al. RaPtomics: Integrating Radiomic and Pathomic Features for Predicting Recurrence in Early Stage Lung Cancer. *Digital Pathol* (2018). doi: 10.1117/12.2296646
14. Wu Z, Wang L, Li C, Cai Y, Liang Y, Mo X, et al. DeepLRHE: A Deep Convolutional Neural Network Framework to Evaluate the Risk of Lung Cancer Recurrence and Metastasis From Histopathology Images. *Front Genet* (2020) 11:768. doi: 10.3389/fgene.2020.00768
15. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 Blockade in Tumors With Mismatch-Repair Deficiency. *N Engl J Med* (2015) 372(26):2509–20. doi: 10.1056/NEJMoa1500596
16. Kather JN, Halama N, Jaeger D. Genomics and Emerging Biomarkers for Immunotherapy of Colorectal Cancer. *Semin Cancer Biol* (2018) 52(Pt 2):189–97. doi: 10.1016/j.semcancer.2018.02.010
17. 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(7):1054–6. doi: 10.1038/s41591-019-0462-y
18. Khan AM, Rajpoot N, Treanor D, Magee D. A Nonlinear Mapping Approach to Stain Normalization in Digital Histopathology Images Using Image-Specific Color Deconvolution. *IEEE Trans Biomed Eng* (2014) 61(6):1729–38. doi: 10.1109/TBME.2014.2303294
19. Chi J, Eramian M. Enhancement of Textural Differences Based on Morphological Component Analysis. *IEEE Trans Image Process* (2015) 24 (9):2671–84. doi: 10.1109/TIP.2015.2427514
20. Erik Reinhard MA, Gooch B, Shirley P. Color Transfer Between Images. *IEEE Comput Graphics Appl* (2001) 21(5):34–41. doi: 10.1109/38.946629

21. Magee D, Treanor D, Crellin D, Shires M, Smith K, Mohee K, et al. Color Normalization in Digital Histopathology Images. *Comput Sci* (2009) p:100– 111.
22. Macenko M, Niethammer M, Marron J, Borland D, Woosley JT, Guan X, et al. A Method for Normalizing Histology Slides for Quantitative Analysis. In: *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. (2009). doi: 10.1109/ISBI.2009.5193250
23. Niethammer M, Borland D, Marron JS, Woosley JT, Thomas NE. Appearance Normalization of Histology Slides. (2010). doi: 10.1007/978-3-642-15948-0_8
24. Vahadane A, Sethi A. Towards Generalized Nuclear Segmentation in Histological Images. (2014). doi: 10.1109/BIBE.2013.6701556
25. Ramakrishnan G, Anand D, Sethi A. Fast GPU-Enabled Color Normalization for Digital Pathology. In: *2019 International Conference on Systems, Signals and Image Processing (IWSSIP)*. IEEE (2019).
26. Wu H-S, Xu R, Harpaz N, Burstein D, Gil J. Segmentation of Microscopic Images of Small Intestinal Glands With Directional 2-D Filters. *Analytical and quantitative cytology and histology* (2005) 27(5):291–300.
27. Farjam R, Soltanian-Zadeh H, Jafari-Khouzani K, Zoroofi RA. An Image Analysis Approach for Automatic Malignancy Determination of Prostate Pathological Images. *Cytomet Part B Clin Cytomet* (2007) 72B(4):227–40. doi: 10.1002/cyto.b.20162
28. Naik S, Doyle S, Agner S, Madabhushi A, Tomaszewski J. Automated Gland and Nuclei Segmentation for Grading Prostate and Breast Cancer Histopathology. In: *5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. (2008).
29. Nguyen K, Sarkar A, Jain AK. Structure and Context in Prostatic Gland Segmentation and Classification. In: Ayache N, Delingette H, Golland P, Mori K, editors. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012. MICCAI 2012. Lecture Notes in Computer Science, vol 7510. Berlin, Heidelberg: Springer (2012). doi: 10.1007/978-3-642-33415-3_15
30. Gunduz-Demir C, Kandemir M, Tosun AB, Sokmensuer C. Automatic Segmentation of Colon Glands Using Object-Graphs. *Medical Image Analysis* (2010) 14(1):1–12. doi: 10.1016/j.media.2009.09.001
31. Nosrati MS, Hamarneh G. Local Optimization Based Segmentation of Spatially-Recurring, Multi-Region Objects With Part Configuration Constraints. *IEEE Trans Med Imaging* (2014) 33(9). doi: 10.1109/TMI.2014.2323074
32. Cohen A RE, Shimshoni I, Sabo E. Memory Based Active Contour Algorithm Using Pixel-Level Classified Images for Colon Crypt Segmentation. *Comput Med Imaging Graphics* (2015) 2015:43:150–64. doi: 10.1016/j.compmedimag.2014.12.006
33. Sirinukunwattana K, Snead D, Rajpoot N. A Stochastic Polygons Model for Glandular Structures in Colon Histology Images. *IEEE Trans Med Imaging* (2015) 34(11):1–1. doi: 10.1109/TMI.2015.2433900
34. Paramanandam M, O'Byrne M, Ghosh B, Mammen JJ, Manipadam MT, Thamburaj R, et al. Automated Segmentation of Nuclei in Breast Cancer Histopathology Images. *PloS One* (2016) 11(9):e0162053. doi: 10.1371/journal.pone.0162053
35. Xu J, Lei X, Liu Q, Gilmore H, Wu J, Tang J, et al. Stacked Sparse Autoencoder (SSAE) for Nuclei Detection on Breast Cancer Histopathology Images. (2014). doi: 10.1109/ISBI.2014.6868041
36. Irshad H, Veillard A, Roux L, Racoceanu D. Methods for Nuclei Detection, Segmentation, and Classification in Digital Histopathology: A Review—Current Status and Future Potential. *IEEE Rev Biomed Eng* (2017) 7:97–114. doi: 10.1109/RBME.2013.2295804
37. Qi X, Xing F, Foran DJ, Yang L. Robust Segmentation of Overlapping Cells in Histopathology Specimens Using Parallel Seed Detection and Repulsive Level Set. *Biomed Eng IEEE Trans* (2011) 59(3):754–65. doi: 10.1109/TBME.2011.2179298
38. Lu C, Xu H, Xu J, Gilmore H, Mandal M, Madabhushi A. Multi-Pass Adaptive Voting for Nuclei Detection in Histopathological Images. *Sci Rep* (2016) 6 (1):33985. doi: 10.1038/srep33985
39. Parvin B, Yang Q, Han J, Chang H, Rydberg M, Barcellos-Hoff MH. Iterative Voting for Inference of Structural Saliency and Characterization of Subcellular Events. *IEEE Trans Image Process* (2007) 16(3):615–23. doi: 10.1109/TIP.2007.891154
40. Arteta C, Lempitsky V, Noble JA, Zisserman A. Learning to Detect Cells Using Non-Overlapping Extremal Regions. In: *Proceedings of the 15th International Conference on Medical Image Computing and Computer-Assisted Intervention - Volume Part I* (2012).
41. Cheng J, Rajapakse J. Segmentation of Clustered Nuclei With Shape Markers and Marking Function. *IEEE Transactions on Biomedical Engineering* (2009) 56(3):741–8.
42. Veta M, Van Diest PJ, Kornegoor R, Huisman A, Viergever MA, Pluim JP. Automatic Nuclei Segmentation in H&E Stained Breast Cancer Histopathology Images. *PloS One* (2013) 8(7):e70221. doi: 10.1371/journal.pone.0070221
43. Vink JP, Van Leeuwen M, Van Deurzen C, de Haan G. Efficient Nucleus Detector in Histopathology Images. *J Microscopy* (2013) 249(2):124–35. doi: 10.1111/jmi.12001
44. Ali S, Madabhushi A. An Integrated Region-, Boundary-, Shape-Based Active Contour for Multiple Object Overlap Resolution in Histological Imagery. *IEEE Trans Med Imaging* (2012) 31:1448–60. doi: 10.1109/TMI.2012.2190089
45. Veta M, Pluim JP, Van Diest PJ, Viergever MA. Breast Cancer Histopathology Image Analysis: A Review. *IEEE Trans Biomed Eng* (2014) 61(5):1400–11. doi: 10.1109/TBME.2014.2303852
46. Xing F, Yang L. Robust Nucleus/Cell Detection and Segmentation in Digital Pathology and Microscopy Images: A Comprehensive Review. *IEEE Rev BioMed Eng* (2016) 9:234–63.
47. Xing F, Xie Y, Yang L. An Automatic Learning-Based Framework for Robust Nucleus Segmentation. *IEEE Trans Med Imaging* (2016) 35(2):550–66. doi: 10.1109/TMI.2015.2481436
48. Yang X, Li H, Zhou X. Nuclei Segmentation Using Marker-Controlled Watershed, Tracking Using Mean-Shift, and Kalman Filter in Time-Lapse Microscopy. *IEEE Transactions on Circuits and Systems I: Regular Papers* (2006) 53(11):2405–14. doi: 10.1109/TCSI.2006.884469
49. Xue JH, Titterington DM. T-Tests, F-Tests and Otsu's Methods for Image Thresholding. *IEEE Trans Image Process A Publ IEEE Signal Process Soc* (2011) 20(8):2392–6.
50. Zhang C, Sun C, Pham TD. Segmentation of Clustered Nuclei Based on Concave Curve Expansion. *J Microscopy* (2013) 251(1):57–67. doi: 10.1111/jmi.12043
51. Sirinukunwattana K, Raza SEA, Tsang Y-W, Snead DR, Cree IA, Rajpoot NM. Locality Sensitive Deep Learning for Detection and Classification of Nuclei in Routine Colon Cancer Histology Images. *IEEE Trans Med Imaging* (2016) 35:1196–206.
52. Kumar N, Verma R, Sharma S, Bhargava S, Vahadane A, Sethi A. A Dataset and a Technique for Generalized Nuclear Segmentation for Computational Pathology. *IEEE Trans Med Imaging* (2017) 36:1550–60.
53. Ciresan DC, Giusti A, Gambardella LM, Schmidhuber J. Mitosis Detection in Breast Cancer Histology Images With Deep Neural Networks. *Med Image Comput Comput Assist Interv* (2013) 16(Pt 2):411–8. doi: 10.1007/978-3-642-40763-5_51
54. Liu F, Lin Y. A Novel Cell Detection Method Using Deep Convolutional Neural Network and Maximum-Weight Independent Set. In: *Deep Learning and Convolutional Neural Networks for Medical Image Computing*. Springer International Publishing (2017).
55. Song Y, Zhang L, Chen S, Ni D, Lei B, Wang T. Accurate Segmentation of Cervical Cytoplasm and Nuclei Based on Multiscale Convolutional Network and Graph Partitioning. *IEEE Trans Biomed Eng* 62(10):2421–33. doi: 10.1109/TBME.2015.2430895
56. Chang JM. Back to Basics: Traditional Nottingham Grade Mitotic Counts Alone Are Significant in Predicting Survival in Invasive Breast Carcinoma. *Ann Surg Oncol* (2015) 22:509–15. doi: 10.1245/s10434-015-4616-y
57. Belien J, Baak J, Van Diest P, Van Ginkel A. Counting Mitoses by Image Processing in Feulgen Stained Breast Cancer Sections: The Influence of Resolution. *Cytometry* (1997) 28(2):135–40. doi: 10.1002/(SICI)1097-0320(19970601)28:2<135::AID-CYTO6>3.0.CO;2-E
58. Roux L, Racoceanu D, Loménie N, Kulikova M, Irshad H, Klossa J, et al. Mitosis Detection in Breast Cancer Histological Images An ICPR 2012 Contest. *J Pathol Inf* (2013) 4(1):8. doi: 10.4103/2153-3539.112693
59. Zhang X, Dou H, Ju T, Xu J, Zhang S. Fusing Heterogeneous Features From Stacked Sparse Autoencoder for Histopathological Image Analysis. *IEEE J Biomed Health Inf* (2017) 20(5):1377–83.
60. Beck AH, Sangoi AR, Leung S, Marinelli RJ, Nielsen TO, Van De Vijver MJ, et al. Systematic Analysis of Breast Cancer Morphology Uncovers Stromal Features Associated With Survival. *Sci Trans Med* (2011) 3(108):108ra113–108ra113. doi: 10.1126/scitranslmed.3002564
61. Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. *PloS Med* (2018) 10(2):e1001381. doi: 10.1371/journal.pmed.1001381

62. Tizhoosh HR PL. Artificial Intelligence and Digital Pathology: Challenges and Opportunities. *J Pathol Inform* (2018) 9:38.
63. Tutac AE, Racoceanu D, Putti T, Xiong W, Leow W-K, Cretu V, et al. Knowledge-Guided Semantic Indexing of Breast Cancer Histopathology Images. In: *2008 International Conference on Biomedical Engineering and Informatics*. IEEE (2008).
64. Dalle JR, Leow WK, Racoceanu D, Tutac AE, Putti TC. Automatic Breast Cancer Grading of Histopathological Images. *Conf Proc IEEE Eng Med Biol Soc* (2008) 2008:3052–5. doi: 10.1109/IEMBS.2008.4649847
65. Wang C, Pécot T, Zynger DL, Machiraju R, Shapiro CL, Huang K. Research and Applications: Identifying Survival Associated Morphological Features of Triple Negative Breast Cancer Using Multiple Datasets. *J Am Med Inform Assoc* (2013) 20:680–7.
66. Basavanhally A, Ganesan S, Feldman M, Shih N, Madabhushi A. Multi-Field-Of-View Framework for Distinguishing Tumor Grade in ER+ Breast Cancer From Entire Histopathology Slides. *IEEE Trans Biomed Eng* (2013) 60 (8):2089–99. doi: 10.1109/TBME.2013.2245129
67. Yuan Y, Failmezger H, Rueda OM, Ali HR, Gräf S, Chin S-F, et al. Quantitative Image Analysis of Cellular Heterogeneity in Breast Tumors Complements Genomic Profiling. *Sci Trans Med* (2020) 4(157):157ra143–157ra143.
68. Chen RJ, Lu MY, Wang J, Williamson DF, Rodig SJ, Lindeman NI, et al. Pathomic Fusion: An Integrated Framework for Fusing Histopathology and Genomic Features for Cancer Diagnosis and Prognosis. *IEEE Trans Med Imaging* (2020). doi: 10.1109/TMI.2020.3021387
69. Mobadersany P, Yousefi S, Amgad M, Gutman DA, Barnholtz-Sloan JS, Velázquez Vega JE, et al. Predicting Cancer Outcomes From Histology and Genomics Using Convolutional Networks. *Proc Natl Acad Sci USA* (2018) 115 (13):E2970–9. doi: 10.1073/pnas.1717139115
70. Serag A I-MA, Qureshi H. Translational AI and Deep Learning in Diagnostic Pathology. *Front Med* (2019) 6:185. doi: 10.3389/fmed.2019.00185
71. Chang HY JC, Woo JI. Artificial Intelligence in Pathology. *J Pathol Transl Med* (2019) 53:1–12. doi: 10.4132/jptm.2018.12.16
72. Daneshjou R, He B, Ouyang D, Zou JY. How to Evaluate Deep Learning for Cancer Diagnostics - Factors and Recommendations. *Biochim Biophys Acta Rev Cancer* (2021) 1875(2):188515.
73. Pope WB, Sayre J, Perlina A, Villablanca JP, Mischel PS, Cloughesy TF. MR Imaging Correlates of Survival in Patients With High-Grade Gliomas. *AJNR Am J Neuroradiol* (2005) 26(10):2466–74.
74. Cao R, Yang F, Ma SC, Liu L, Zhao Y, Li Y, et al. Development and Interpretation of a Pathomics-Based Model for the Prediction of Microsatellite Instability in Colorectal Cancer. *Theranostics* (2020) 10 (24):11080–91. doi: 10.7150/thno.49864
75. Echle A, Rindtorff NT, Brinker TJ, Luedde T, Pearson AT, Kather JN. Deep Learning in Cancer Pathology: A New Generation of Clinical Biomarkers. *Br J Cancer* (2021) 124(4):686–96. doi: 10.1038/s41416-020-01122-x
76. Colling R, Pitman H, Oien K, Rajpoot N, Macklin P, Snead D, et al. Artificial Intelligence in Digital Pathology: A Roadmap to Routine Use in Clinical Practice. *J Pathol* (2019) 249(2):143–50. doi: 10.1002/path.5310

Conflict of Interest: Author XShi, GT, YL, JY and YW were employed by the company Geneis Beijing Co., Ltd. Author LL was employed by the company Beijing Shanghe Jiye Biotech Co., Ltd.

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

frontiers in Oncology

ORIGINAL RESEARCH
published: 16 February 2022
doi: 10.3389/fonc.2022.843879

Circulating Tumor Cell Identification Based on Deep Learning

*Zhifeng Guo, Xiaoxi Lin, Yan Hui, Jingchun Wang, Qiuli Zhang and Fanlong Kong**

Department of Oncology, Chifeng Municipal Hospital, Chifeng, China

As a major reason for tumor metastasis, circulating tumor cell (CTC) is one of the critical biomarkers for cancer diagnosis and prognosis. On the one hand, CTC count is closely related to the prognosis of tumor patients; on the other hand, as a simple blood test with the advantages of safety, low cost and repeatability, CTC test has an important reference value in determining clinical results and studying the mechanism of drug resistance. However, the determination of CTC usually requires a big effort from pathologist and is also error-prone due to inexperience and fatigue. In this study, we developed a novel convolutional neural network (CNN) method to automatically detect CTCs in patients' peripheral blood based on immunofluorescence *in situ* hybridization (imFISH) images. We collected the peripheral blood of 776 patients from Chifeng Municipal Hospital in China, and then used Cyttel to delete leukocytes and enrich CTCs. CTCs were identified by imFISH with CD45+, DAPI+ immunofluorescence staining and chromosome 8 centromeric probe (CEP8+). The sensitivity and specificity based on traditional CNN prediction were 95.3% and 91.7% respectively, and the sensitivity and specificity based on transfer learning were 97.2% and 94.0% respectively. The traditional CNN model and transfer learning method introduced in this paper can detect CTCs with high sensitivity, which has a certain clinical reference value for judging prognosis and diagnosing metastasis.

Keywords: circulating tumor cells, detection, count, convolutional neural network, transfer learning

OPEN ACCESS

Edited by:
Min Tang,
Jiangsu University, China

Reviewed by:
Jiasheng Yang,
Changsha Medical University, China
Lan Yu,
Inner Mongolia People's Hospital, China

***Correspondence:**
Fanlong Kong
dinosaurand@163.com

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *27 December 2021*
Accepted: *21 January 2022*
Published: *16 February 2022*

Citation:
Guo Z, Lin X, Hui Y, Wang J, Zhang Q and Kong F (2022) Circulating Tumor Cell Identification Based on Deep Learning. Front. Oncol. 12:843879. doi: 10.3389/fonc.2022.843879

INTRODUCTION

Circulating tumor cells (CTC) are all kinds of tumor cells in peripheral blood (1). Most of the CTCs undergo apoptosis or phagocytosis after entering the peripheral blood, while a minority of CTCs develop into metastasis and undergo for a period of dormancy, and lead to metastatic tumor (2, 3). Cancer recurrence and metastasis are the main causes of death in cancer patients (4, 5). A large number

of experiments on esophageal squamous cell carcinoma (6), breast cancer (7, 8), prostate cancer (9) and lung cancer (10) have proved that CTCs were closely related to the prognosis of patients with advanced cancer. As a simple blood test, CTCs detection has the advantages of high safety, low cost and repeatability, which is available at any time to evaluate the prognosis and recurrence risk of patients (11, 12). Many experiments used liquid biopsy to monitor the CTCs response in patients with malignant tumors to evaluate the therapeutic response (13–15). Many studies have shown that CTCs count is closely related to prognosis, which has an important reference value for determining clinical results and recurrence risk (16–18). The fluid biopsy can predict disease progression in real time to assess tumor heterogeneity, and it was possible to detect single CTCs or cell clusters (13, 19–21). Immune enrichment with multiparameter flow cytometric is the gold standard of CTCs detection (22), but this method was limited due to the lack of tumor-specific markers, in this case, multi-label immunofluorescence staining was essential. Epithelial cell adhesion molecule (EpCAM) was often used to detect cancer cells in the blood because it mediates contact with homotype cells in epithelial tissue (23–25). The methods of $CD45^+$, DNA fluorescence *in situ* hybridization (FISH) of the centromere of chromosome 8 probe ($CEP8^+$)/chromosome 17 centromere duplication ($CEP17^+$) have been widely used to identify CTCs (26, 27).

In recent years, rapid and automatic identification of CTCs is becoming more and more important, and the research on the automatic identification process of CTCs was also accelerating (28, 29), such as cell search system to obtain digital images (30), rare event imaging system (REIS) (31), microfluidic platform composed of multi-functional microfluidic chip and unique image processing algorithm (32). However, due to the heterogeneity of CTCs, these classification methods were often subjective. Therefore, under certain conditions, test results vary from examiner to examiner. The development of artificial intelligence (AI) has accelerated scientists' research on machine learning. Machine learning has been widely used in medical research because of its advantages of objectivity, rapidity, and overcoming noise (33–35), especially in medical images (36, 37). As the classical algorithms of machine learning, deep learning and convolutional neural network (CNN) have made outstanding contributions in promoting medical research (38–40). Anthimopoulos et al. proposed the first problem specific deep CNN for classification of interstitial lung diseases (ILD), the results showed that (classification performance~85.5%) CNN has potential in analyzing ILD (41). Poplin et al. used the Inception-v3 neural network structure to predict potential cardiovascular risk factors in retinal fundus images (42). Le et al. constructed a deep neural network to classify Rab protein molecules through two-dimensional CNN, which provided a valuable reference for biological modeling using deep neural network (43). At present, CNN has been widely used to promote biomedical image analysis and successfully applied in cancer diagnosis and tissue identification (44, 45). Compared with the traditional machine learning methods, CNN-based automatic image processing method has the advantage of eliminating the bias caused by personal subjectivity (46). Negative enrichment combined with immuno fluorescence *in situ* hybridization (imFISH) to detect CTCs has been proven to be feasible and clinically valuable (47, 48).

In this study, we applied deep learning to identify CTCs to reduce the subjective error. ImFISH was used to detect patients' CTCs, each image contains positive CTCs nucleus and negative control to segment the images of circulating tumor cells. CNN deep learning network was used to identify circulating tumor cells and count CTCs.

MATERIALS AND METHODS

A Framework for Identifying CTCs

The complete process of identifying CTCs based on CNN was shown in **Figure 1**. Specifically, the peripheral blood of 776 cancer patients in Chifeng Municipal Hospital was collected firstly. Then the blood samples were processed by the Cyttel method. Based on the principle of immunology and with the help of magnetic particle technology, CTCs were enriched by gradually removing the components of plasma, red blood cells and white blood cells, and CTCs were processed by imFISH (26). Finally, after preprocessing the images, stratified sampling was used to divide the data, 80% of the images were used for training, and a deep learning algorithm based on CNN was used to train the model. The remaining 20% were used as separate test set. The prediction performance of the model was evaluated with the results of a 5-fold cross validation (CV).

Samples Preparation

We conducted a retrospective study using plasma samples from the Chifeng Municipal Hospital. A total of 776 patients were enrolled from 2017 to 2019. Cancer types include lung cancer, liver cancer, gastrointestinal cancer, breast cancer, carcinoma of thyroid, NPC and others. After puncture for each patient, discarding the first 2ml blood sample to avoid skin epithelial cell pollution, then routinely collect 4ml peripheral venous blood, these samples were placed in a blood collection vessel containing acid citrate dextrose (ACD), gently inverted and mixed for 8 times before stored at room temperature, and CTCs were enriched within 24 hours after collection. The study was approved by the Ethics Committee of Chifeng Municipal Hospital.

Enrichment of CTCs

The detection method selected in this study was Cyttel (49). The collected blood was taken out and put into a centrifuge tube for centrifugation experiment. After centrifugation at 776 g for 5 minutes, the supernatant was discarded to retain the precipitation, washed the precipitate with CS1 buffer (Cyttel Biosciences Co., Ltd., Beijing, China), and then the red blood cells were fully dissolved with CS2 buffer. Added anti-$CD45^+$ monoclonal antibody binding beads and the mixture was shaken evenly for 20 minutes to fully bind with leukocytes. Another 3 ml of separation medium was added to the centrifuge tube and centrifuged at a gradient of 300 g for 5 minutes. Then the upper rare cell layer was then centrifuged at a gradient of 776 g for 5 minutes, resuspended with CS1 buffer, and

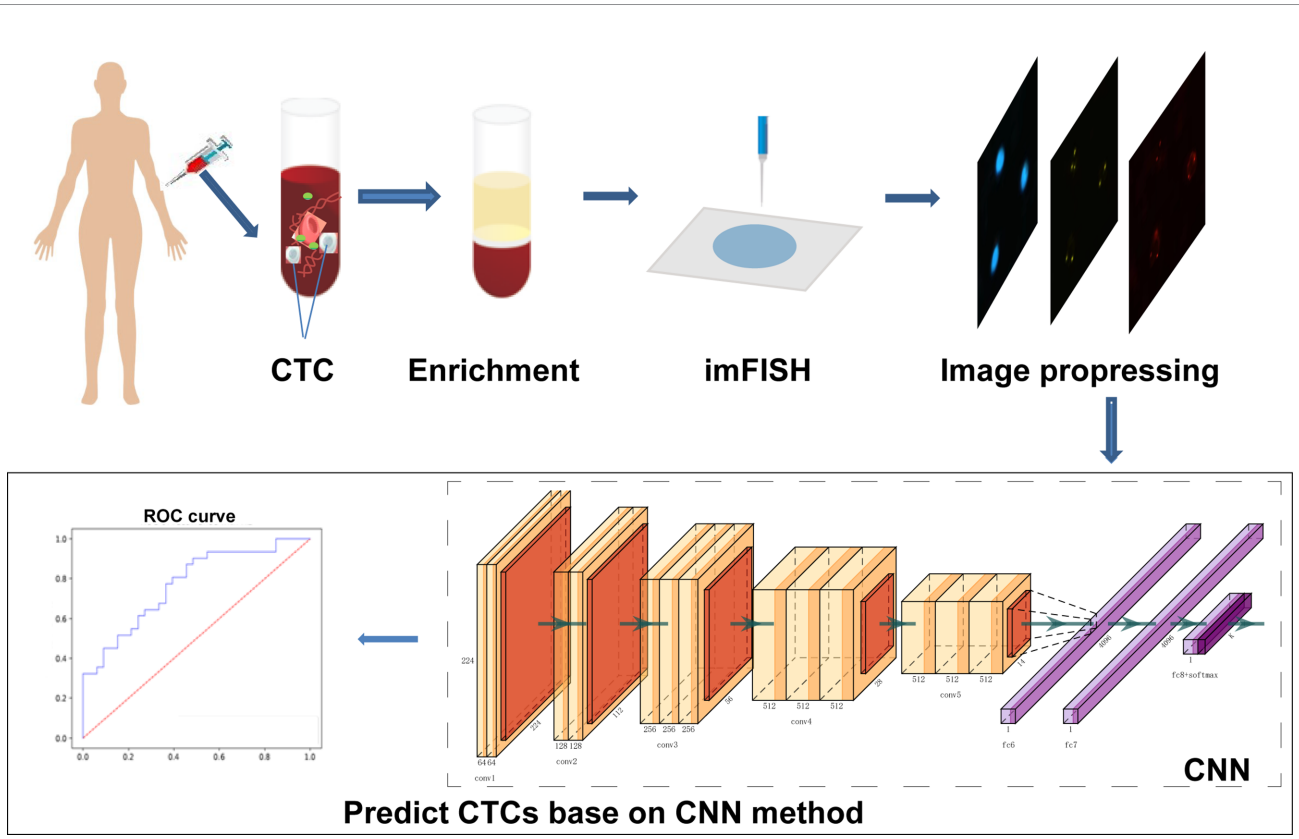


FIGURE 1 | The protocol of whole process.

the test tube was placed on a magnetic scaffold for 2 minutes. ImFISH identification was performed within 24 hours after coating, fixing, and drying.

imFISH Identification of CTCs

The samples were fixed with a fixative, dehydrated and dried at room temperature. The slide was coated with 10μl CEP8$^+$ antibody, sealed and hybridized at 37°C for 1.5 h. After hybridization, removed the cover slide and eluted the probe for 15 min. The slides were washed twice in 2×SSC. Then, the prepared CD45$^+$ fluorescent antibody was added to the sample area, and the slides were placed in a humid box and incubated in an oven at 33°C for 1 hour. After incubation, fluorescent antibody against CD45$^+$ was absorbed and 10 μl DAPI$^+$ was added to the specimen area. Then, CTCs were observed and counted under a fluorescence microscope.

Detection Standard of CTCs

Each cell was divided into three different color channels: blue, orange and red. Among them, the nucleus was shown blue in DAPI$^+$ (**Figure 2A**), and the centromere was shown in orange by CEP8$^+$ (**Figure 2B**), and the white blood cells were stained by CD45$^+$ immunofluorescence (**Figure 2C**). The interpretation criteria of CTCs count are: (1) eliminate the aggregation, superposition and interference of nuclei or impurities, (2) positive for DAPI$^+$, (3) negative for CD45$^+$, (4) CEP8$^+$ signal points >2. That is, cells are regarded as CTCs if they are CD45-/DAPI+/CEP8≥3 (50, 51).

Image Preprocessing

The Python package *openCV* was used to handle CTCs images (52), including color and shape conversions. To be specific, the DAPI$^+$ channel was first transformed into gray scale, and then the Gaussian filter was used to denoise. After extracting the gradient of the image, the regions with a high horizontal gradient and low vertical gradient were left, and the Gaussian filter was used to denoise. Then, the blurred image was binarization, that is, each pixel was replaced by the average value of the surrounding pixels in order to smooth and replace those regions with obvious intensity changes. Due to the lack of details in the contour of the obtained image, it may interfere with the subsequent contour detection, so it is necessary to be expanded and perform four morphological corrosion and expansion respectively. After the contour of the nuclear region was found, the minimum matrix coordinates of the contour were obtained, and the coordinates were mapped to the CEP8$^+$ channel and segmented.

Computational Identification of CTC

With the development of artificial intelligence, deep learning has been widely used in medical image processing. CNN is one of the

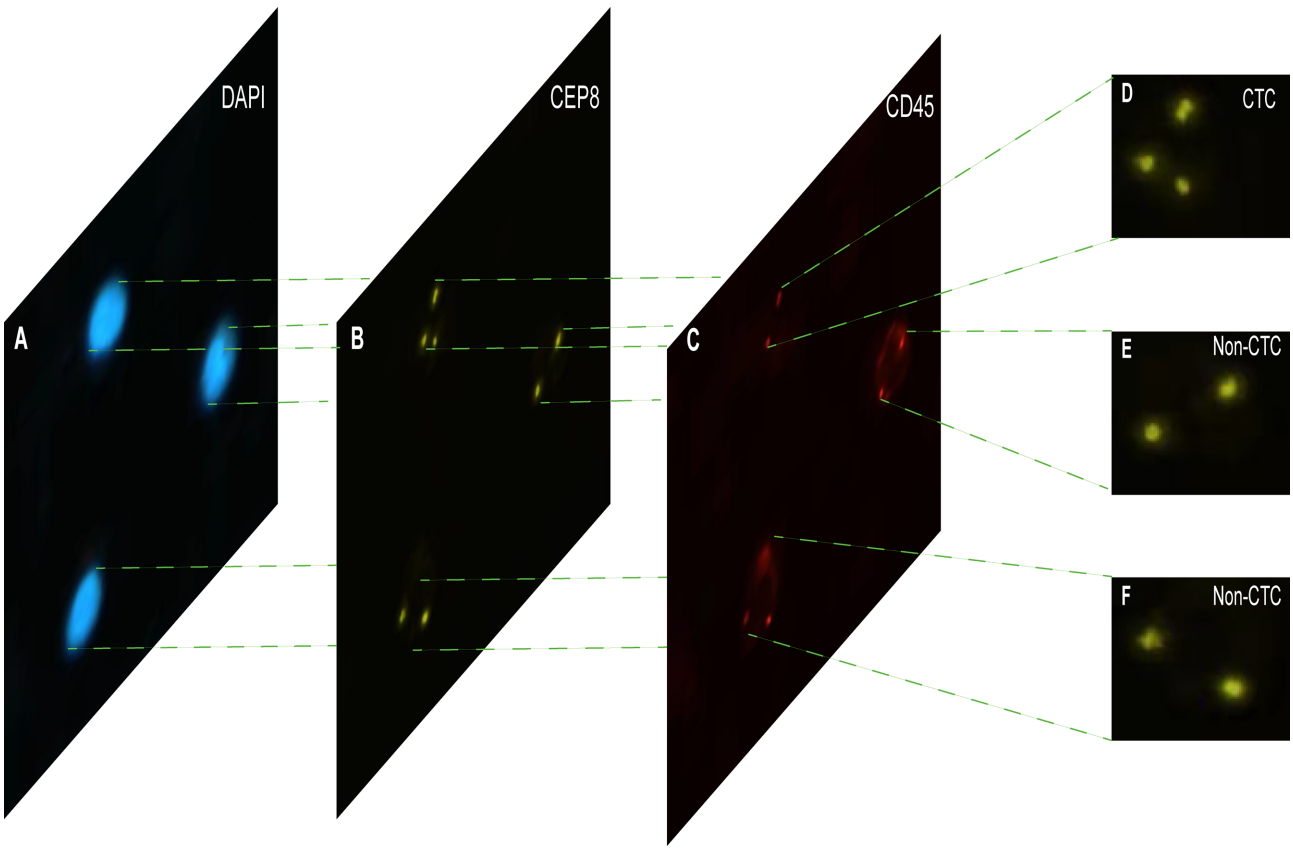


FIGURE 2 | The imFISH result and CTC count results. **(A–C)** The channels (DAPI, CEP8+, CD45+) of each image by imFISH. **(D)** The cell was regarded as CTC because the number of centromeres was 3 (>2). **(E, F)** The cell was regarded as non-CTC because the number of centromeres was 2.

representative algorithms of deep learning, which allows higher-level feature extraction and higher-level data prediction. After images were preprocessed, stratified sampling was used to divide the data, the model was trained by 5-fold CV, and the down sampling method was used for the training set to ensure the balance of positive samples and negative samples. CTCs in a single nucleus were identified through CNN. CNN includes input layer, hidden layer and output layer, the hidden layer includes layer1, layer2 and layer3, and each layer also includes convolution layer, excitation layer and pooling layer. After the images were fed into the input layer, it first enters the first intermediate hidden layer with convolution layer is composed of 32 5x5 convolution cores, and then fed to the pool layer through the ReLU excitation layer for dimensionality reduction. After dimensionality reduction, data was output from the first hidden layer to complete the feature extraction process. Then, all features are extracted through layer2 and layer3 hidden layers in turn. Finally, it enters the output layer and outputs the result of whether it is CTC or not. The CNN in this study involved VGG16, VGG19 (53), ResNet18, ResNet50 (54) and AlexNet (55).

These pre-training models have consumed huge time resources and computing resources when developing neural networks usually. In recent years, transfer learning has become a new learning framework to solve this problem (56, 57). CNN model is pre-trained using a large number of images, and the trained model is distributed by its inventors for adoption. Transfer learning relies on the pre-trained CNN model to realize the knowledge transfer of different but related tasks, that is, using the existing knowledge learned from the completed tasks to help complete the new tasks. If the knowledge transfer is successful, it will greatly improve the learning efficiency by avoiding expensive data labeling. Transfer learning is defined as follows: a given domain D consists of feature space X and edge probability distribution $P(X)$, a label space y and a prediction function f consist a task T. D_S and D_T represent the source domain and the target domain, respectively, may have different feature spaces or different edge probability distributions, that is, $X_S \neq X_T$ or $P_S(X) \neq P_T(X)$, in addition, task T_S and T_T are subject to different label spaces (58).

Statistical Analysis

Receiver operating characteristic (ROC) analysis was used to evaluate the performance of the model to identify CTC, and the area under the curve (AUC) at the 0.5 cut-off point was used to

measure the prediction accuracy. At the same time, the confusion matrix was used to observe the specificity and sensitivity. All of the analyses were performed using python version 3.6.9 and "sklearn" package version 0.23.2

RESULTS

Patient Characteristics

From Jan. 2017 to Jun. 2019, a total of 776 patients from Chifeng Municipal Hospital were included in this study. All the sample types were peripheral blood, and their ages ranged from 11 to 90 years old, with an average age of 65 years. Among the known cancer types, lung cancer was the most common (20.7%), followed by breast cancer and gastrointestinal cancer, and thyroid cancer was the least with only 2 cases. The clinical characteristic data of the enrolled patients were shown in **Table 1**.

Thousands of CTC and Non-CTC Images Were Segmented by openCV

The imFISH was performed on the samples from 776 patients. In order to avoid the influence of human factors, Python package *OpenCV* was used to process cell images. After the nuclear region contour was found in the blue channel ($DAPI^+$), the minimum matrix coordinates of the contour were obtained, mapped to the orange channel ($CEP8^+$) and segmented, and the number of centromeres was observed. If there were more than 2 centromeres, the cell was considered CTC (**Figure 2D**). Otherwise it was non-CTC (**Figures 2E, F**). Finally, we obtained 14166 images, including 694 CTC images and 13472 non-CTC images. The details of data were shown in **Table 2**, in original train set, the number of CTC and non-CTC were 555 and 10777, respectively, ensuring balanced positive and negative samples, we performed down sampling method, at last, the number of CTC and non-CTC after down sampling were 555, respectively.

TABLE 1 | Summary of the general clinical information of patients.

Characteristics	No. (%) of Participants
Age	
0-39	30 (3.9)
40-69	313 (40.3)
>70	103 (13.3)
Unknown	330 (42.5)
Gender	
Male	248 (42.0)
Female	199 (25.6)
Unknown	329 (42.4)
CTC number	9.9(0-318)
Cancer type	
Lung cancer	161 (20.7)
Liver cancer	18 (2.3)
Gastrointestinal cancer	107 (13.8)
Breast cancer	91 (11.7)
Carcinoma of thyroid	2 (0.3)
NPC	30 (3.9)
Other	367 (47.3)

The Computational Method Performed Well in Identifying CTC

The CNN method was used to identify the segmented cell images. The whole process was shown in **Figure 3**. Firstly, the hierarchical sampling method was adopted for all images, 80% of the data were used for training and 20% of the data were used for testing. The CNN-based methods were used to train the model, including VGG16, VGG19, ResNet18, ResNet50, and AlexNet. Specifically, the traditional CNN model and transfer learning model were respectively trained on the training set based on 5-fold CV. The transfer learning model relied on the pre-trained CNN model to realize the task of CTC recognition. The trained traditional CNN model and the transfer learning model were used to test sets and output the final results. The results of 5-fold CV based on the trainset were shown in **Figure 4A**, the best performance was based on ResNet18 with AUC was about 0.98, in which 90.26% of non-CTCs were successfully identified, 9.74%were incorrectly identified as CTCs, while only 5.41% of CTCs are misclassified (**Figure 4B**), the sensitivity and specificity were 95.3% and 91.7% respectively. After training the model, we used ResNet18 with the best performance on the test data set, its ROC curve was shown in **Figure 4C** with AUC was 0.988, and the confusion matrix also shown that ResNet18 performed well (**Figure 4D**). In addition, in order to improve the prediction performance and save computing resources, transfer learning was also used to train the model. The results of 5-fold CV based on transfer learning in train data set was shown in **Figure 5A**, the results of confusion matrix showed that the sensitivity and specificity of transfer learning were 97.2% and 94.0% respectively (**Figure 5B**). After training the

TABLE 2 | The number of images.

	Original			Down sampling	
	Train set	Test set	Total	Train set	Test set
No. of CTC	555	139	694	555	139
No. of Non-CTC	10777	2695	13472	555	2695
Total	11332	2834	14166	1110	2834

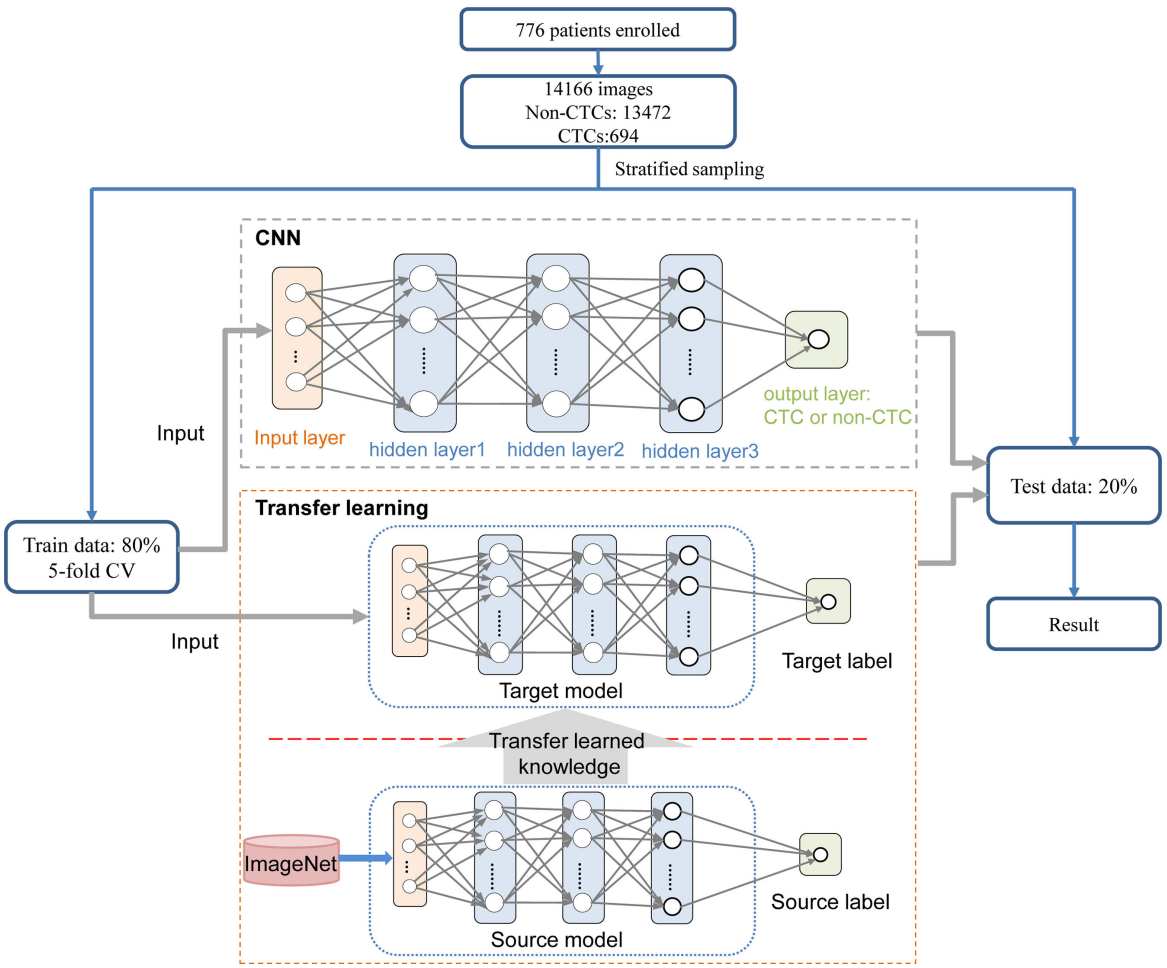


FIGURE 3 | Process for identifying CTC. On the training set, the traditional CNN model and transfer learning model were respectively trained based on 5-fold CV. The transfer learning model relied on the pre-trained CNN model to realize the task of CTC recognition. The trained traditional CNN model and the transfer learning model were used to test sets and output the final results.

model with transfer learning, we used VGG16 with the best performance on the test data set, its AUC was 0.988 (**Figure 5C**), and the sensitivity and specificity were 98.6% and 96.5% respectively (**Figure 5D**).

The experimental results showed that the deep learning method based on CNN can accurately identify CTC and provide a powerful reference for the prognosis of patients. In addition, we also summarized some samples that were discriminated incorrectly, such as samples that were originally non-CTC but were predicted to be CTC (**Figure 6A**), and samples that predicted CTC to be non-CTC (**Figure 6B**). The reason for the misjudgment first considers the artificial or instrumental noise in the process of negative enrichment techniques. Secondly, the exposure during the photographing process after imFISH processing resulted in us not getting the original film data. The third is that the centromere was not completely located in the nucleus due to the platform.

DISCUSSION

More and more evidence showed that CTCs are closely related to the prognosis of patients with advanced cancer. It has an important reference value for determining the clinical results and recurrence risk. In recent years, blood testing has been widely used to monitor the CTC response of patients with malignant tumors and evaluate the prognosis and recurrence risk of patients since its high safety and low cost. It reduces errors caused by manually setting interpretation standards and save time and labor costs. The importance of CTC rapid automatic recognition is increasing, and the research of the automatic recognition process of CTCs is also accelerating. Deep learning has been proved to be suitable for detecting CTC due to its high sensitivity and specificity in CTC counting. In addition, image interpretation using machine learning can capture important image features.

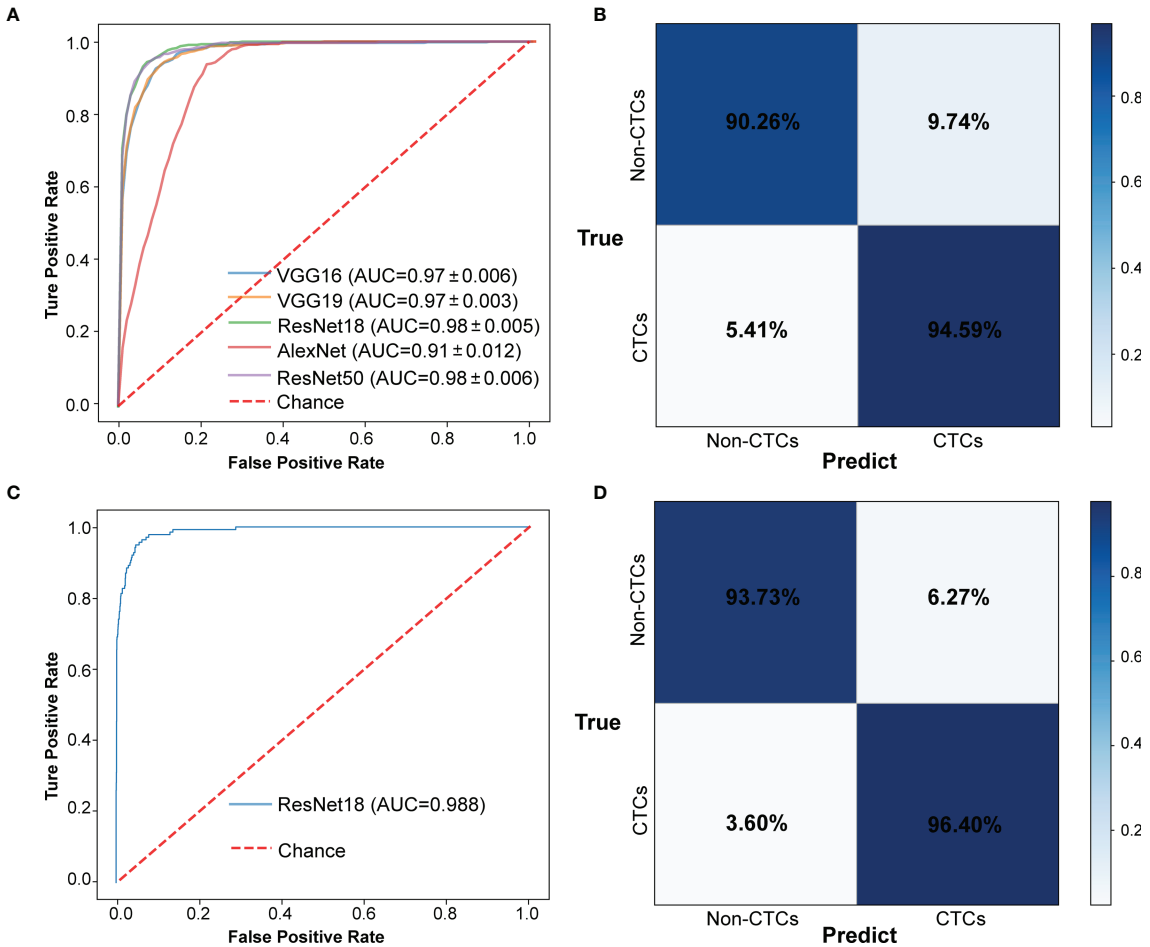


FIGURE 4 | The results of CTCs identify based on traditional CNN. **(A)** ROC curve of 5-fold CV in train data set. **(B)** Confusion matrix of 5-fold CV in train data set. **(C)** ROC curve in test data set. **(D)** Confusion matrix in test data set.

In this study, we developed a CTCs recognition method based on deep learning. After collecting the blood samples from Chifeng Municipal Hospital, we conducted CTCs enrichment and imFISH experiments on the samples, and screened the fluorescent images according to the image quality. A total of 14166 images were used for downstream analysis, including 694 CTC images and 13472 non-CTC images. 80% of the images were used for training models and 20% for test. In order to reduce the error caused by manual intervention, we used machines instead of manual screening. Firstly, images were segmented by using the Python package *openCV*, and the coordinate information of the nucleus was obtained after image preprocessing. Then, we used CNN models such as VGG16, VGG19, ResNet18, ResNet50 and AlexNet to identify CTCs. The results of 5-fold CV showed that their AUC reached 0.98, and the sensitivity and specificity were 95.3% and 91.7%, respectively. In order to overcome the shortcomings of consuming a lot of time and computing resources when developing neural networks, transfer learning was used to train the model. Finally, the AUC was improved to 0.99, and the recognition sensitivity and specificity also reached to 97.2% and 94.0% based on transfer learning.

The method of transfer learning was proposed, which can carry out image interpretation, capture important image features, reduce the errors caused by subjective factors in manual interpretation, and save computing time and computing resources. In the process of 5-fold CV, the down-sampling method was used to overcome the serious imbalance between positive samples and negative samples, and the 5-fold CV results of transfer learning shown higher sensitivity and specificity. Nevertheless, this study still has some limitations. The CTC images contained in the enrollment data do not cover the whole film, but focus on a CTC positive area under the microscope. Due to quality issues, some images in the enrollment data are abandoned. How to expand the image scope is the focus of attention in the future.

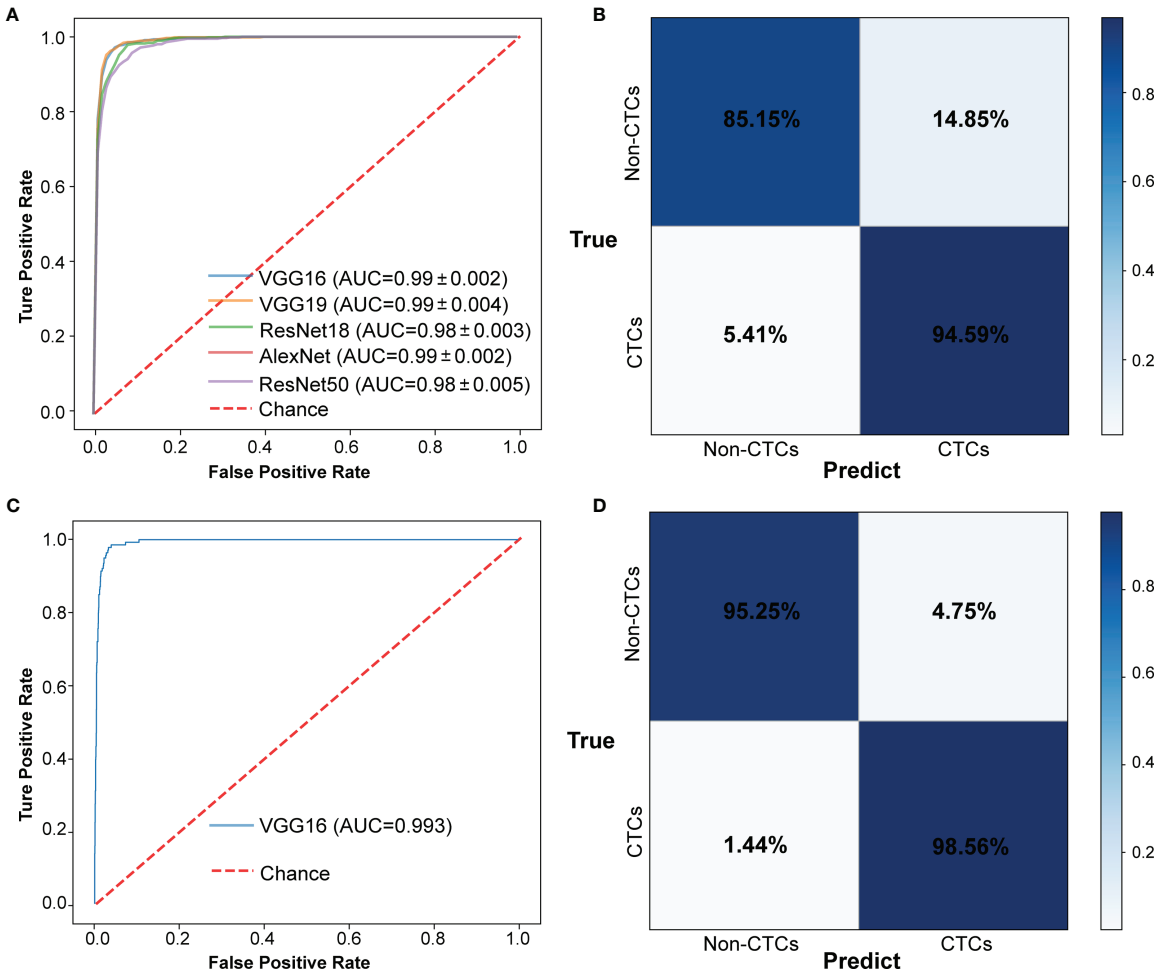


FIGURE 5 | The results of CTCs identify based on transfer learning. **(A)** ROC curve of 5-fold CV in train data set. **(B)** Confusion matrix of 5-fold CV in train data set. **(C)** ROC curve in test data set. **(D)** Confusion matrix in test data set.

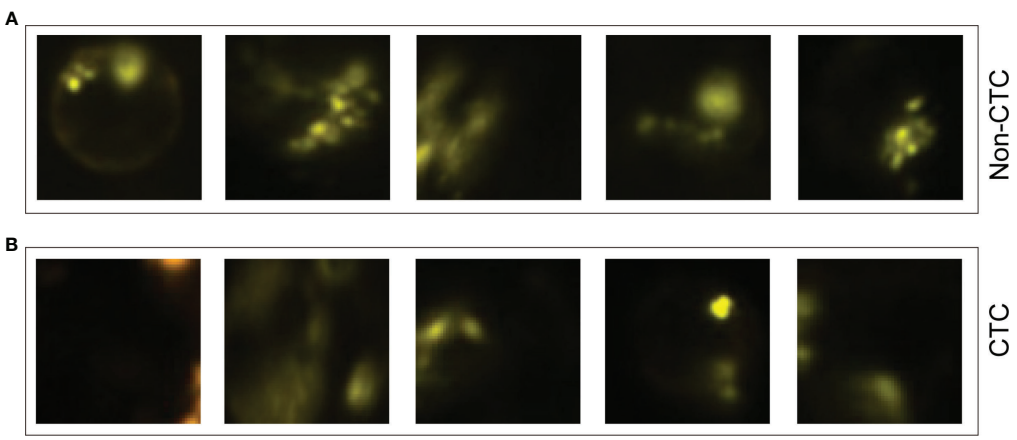


FIGURE 6 | Some misjudged images. **(A)** Non-CTC images, but they were identified as CTCs. **(B)** CTC images, but they were identified as non-CTCs.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://github.com/bensteven2/CTC_project.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of Chifeng Municipal Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

FK designed the study. ZG collected, analyzed and interpreted the data, and wrote the manuscript. XL and YH performed the experiment. JW and QZ reviewed and modified the manuscript. All authors approved the final version of the manuscript.

FUNDING

This work was supported by the 2020 Chifeng Natural Science Research project (grant number SZR157).

REFERENCES

1. Masuda T, Hayashi N, Iguchi T, Ito S, Eguchi H, Mimori K. Clinical and Biological Significance of Circulating Tumor Cells in Cancer. *Mol Oncol* (2016) 10(3):408–17. doi: 10.1016/j.molonc.2016.01.010
2. Arko D, Lim AR, Ghajar CM. Circulating and Disseminated Tumor Cells: Harbingers or Initiators of Metastasis? *Mol Oncol* (2017) 11:40–61. doi: 10.1002/1878-0261.12022
3. Liu H, Qiu C, Wang B, Bing P, Tian G, Zhang X, et al. Evaluating DNA Methylation, Gene Expression, Somatic Mutation, and Their Combinations in Inferring Tumor Tissue-of-Origin. *Front Cell Dev Biol* (2021) 9:619330. doi: 10.3389/fcell.2021.619330
4. He B, Dai C, Lang J, Bing P, Tian G, Wang B, et al. A Machine Learning Framework to trace Tumor Tissue-Of-Origin of 13 Types of Cancer Based on DNA Somatic Mutation. *Biochim Biophys Acta Mol Basis Dis* (2020) 1866:165916. doi: 10.1016/j.bbadis.2020.165916
5. He B, Lang J, Wang B, Liu X, Lu Q, He J, et al. TOOme: A Novel Computational Framework to Infer Cancer Tissue-of-Origin by Integrating Both Gene Mutation and Expression. *Front Bioeng Biotechnol* (2020) 8:394. doi: 10.3389/fbioe.2020.00394
6. Qiao Y, Li J, Shi C, Wang W, Qu X, Ming X, et al. Prognostic Value of Circulating Tumor Cells in the Peripheral Blood of Patients With Esophageal Squamous Cell Carcinoma. *Oncotargets Ther* (2017) 10:1363–73. doi: 10.2147/OTT.S129004
7. Giuliano M, Giordano A, Jackson S, Hess KR, Giorgi UD, Mego M, et al. Circulating Tumor Cells as Prognostic and Predictive Markers In Metastatic Breast Cancer Patients Receiving First-Line Systemic Treatment. *Breast Cancer Res* (2011) 13:1–9. doi: 10.1186/bcr2907
8. Pierga JY, Hajage D, Bachelot T, Delaloge S, Brain E, Campone M, et al. High Independent Prognostic and Predictive Value of Circulating Tumor Cells Compared With Serum Tumor Markers in a Large Prospective Trial in First-Line Chemotherapy for Metastatic Breast Cancer Patients. *Ann Oncol* (2011) 23:618–24. doi: 10.1093/annonc/mdr263
9. Stott SL, Lee RJ, Nagrath S, Yu M, Maheswaran S. Isolation and Characterization of Circulating Tumor Cells from Patients with Localized and Metastatic Prostate Cancer. *Sci Trans Med* (2010) 2:25ra23. doi: 10.1126/scitranslmed.3000403
10. Maheswaran S, Sequist LV, Nagrath S, Ulkus L, Brannigan B, Collura CV, et al. Detection of Mutations in EGFR in Circulating Lung-Cancer Cells. *N Engl J Med* (2008) 359:366–77. doi: 10.1056/NEJMoa0800668
11. Lindsay CR, Faugeroux V, Michiels S, Pailler E, Facchinetti F, Ou D, et al. A Prospective Examination of Circulating Tumor Cell Profiles in Non-Small-Cell Lung Cancer Molecular Subgroups. *Ann Oncol* (2017) 28(7):1523–31. doi: 10.1093/annonc/mdx156
12. Praharaj PP, Bhutia SK, Nagrath S, Bitting RL, Deep G. Circulating Tumor Cell-Derived Organoids: Current Challenges and Promises in Medical Research and Precision Medicine. *Biochim Biophys Acta (BBA) - Rev Cancer* (2018) 1869(2):117–27. doi: 10.1016/j.bbcan.2017.12.005
13. Yan M, Wang L, Song Y, Zhang K. Monitoring Non-Operative Treatment for Advanced Esophageal Cancer by Dynamic Ctcs Count: A Case Report. *J Clin Case Rep* (2017) 9(4):10001227. doi: 10.4172/2165-7920.10001227
14. Huebner H, Fasching PA, Gumbrecht W, Jud S, Rauh C, Matzas M, et al. Filtration Based Assessment of CTCs and CellSearch based Assessment are Both Powerful Predictors of Prognosis for metastatic Breast Cancer Patients. *BMC Cancer* (2018) 18:204. doi: 10.1186/s12885-018-4115-1
15. Yang J, Hui Y, Zhang Y, Zhang M, Ji B, Tian G, et al. Application of Circulating Tumor DNA as a Biomarker for Non-Small Cell Lung Cancer. *Front Oncol* (2021) 11:725938. doi: 10.3389/fonc.2021.725938
16. Banys P, Schneck H, Blassl C, Schultz S, Meier S, Niederacher D, et al. Prognostic Relevance of Circulating Tumor Cells in Molecular Subtypes of Breast Cancer. *Geburtshilfe Frauenheilkd* (2015) 75(3):232–7. doi: 10.1055/s-0035-1545788
17. Jingsi D, Zhu D, Tang X, Qiu X, Bingjie D. Detection of Circulating Tumor Cell Molecular Subtype in Pulmonary Vein Predicting Prognosis of Stage I-III Non-small Cell Lung Cancer Patients. *Front Oncol* (2019) 9:1139–9. doi: 10.3389/fonc.2019.01139
18. Cristofanilli M, Pierga JY, Reuben J, Rademaker A, Davis AA, Peeters DJ, et al. The Clinical Use of Circulating Tumor Cells (CTCs) Enumeration for Staging of Metastatic Breast Cancer (MBC): International Expert Consensus Paper. *Crit Rev Oncology/Hematol* (2019) 134:39–45. doi: 10.1016/j.critrevonc.2018.12.004
19. Wan J, Massie C, Garcia-Corbacho J, Mouliere F, Brenton JD, Caldas C, et al. Liquid Biopsies Come of Age: Towards Implementation of Circulating Tumour DNA. *Nat Rev Cancer* (2017) 17:223. doi: 10.1038/nrc.2017.7
20. Marcuello M, Vymetalkova V, Rui P, Duran-Sanchon S, Gironella M. Circulating Biomarkers For Early Detection and Clinical Management of Colorectal Cancer. *Mol Aspects Med* (2019) 69:107–22. doi: 10.1016/j.mam.2019.06.002
21. Dba A, Lc A, Mza D, Tmmab C, Esg A. Liquid Biopsy in Ovarian Cancer Using Circulating Tumor DNA and Cells: Ready for Prime Time? - ScienceDirect. *Cancer Lett* (2020) 468:59–71. doi: 10.1016/j.canlet.2019.10.014
22. Racila E, Euhus D, Weiss AJ, Rao C, Mcconnell J, Terstappen L, et al. Detection and Characterization of Carcinoma Cells in the Blood. *Proc Natl Acad Sci* (1998) 95:4589–94. doi: 10.1073/pnas.95.8.4589
23. Grover PK, Cummins AG, Price TJ, Roberts-Thomson IC, Hardingham JE. Circulating Tumour Cells: The Evolving Concept and the Inadequacy of their Enrichment by EpCAM-Based Methodology for Basic and Clinical Cancer Research. *Ann Oncol Off J Eur Soc Med Oncol* (2014) 25:1506. doi: 10.1093/annonc/mdu018
24. Tellez GM, Rodriguez CL, Antoine C, Benjamin O, Dominique H. Circulating Tumor Cells: A Review of Non-EpCAM-Based Approaches for Cell Enrichment and Isolation. *Clin Chem* (2016) 62(4):571–81. doi: 10.1373/clinchem.2015.249706
25. Keller L, Werner S, Pantel K. Biology and Clinical Relevance of EpCAM. *Other* (2019) 3:165–80. doi: 10.15698/cst2019.06.188
26. Koudelakova V, Trojanec R, Vrbkova J, Donevska S, Bouchalova K, Kolar Z, et al. Frequency of Chromosome 17 Polysomy in Relation to CEP17 Copy Number in a Large Breast Cancer Cohort. *Genes Chromosomes Cancer* (2016) 55:409–17. doi: 10.1002/gcc.22337
27. Xiangqi L, Zhang Z, Yijie B, Chao Z, Shan B. Circulating Tumor Cells Detection in Neuroblastoma Patients by EpCAM-Independent Enrichment and Immunostaining-Fluorescence In Situ Hybridization. *Ebiomedicine* (2018) 35:244–50. doi: 10.1016/j.ebiom.2018.08.005

28. Nagrath S, Sequist LV, Maheswaran S, Bell DW, Irimia D, Ulkus L, et al. Isolation of Rare Circulating Tumour Cells in Cancer Patients by Microchip Technology. *Nature* (2007) 450:1235–9. doi: 10.1038/nature06385
29. Yang C, Zhang N, Wang S, Shi D, Zhang C, Kan L, et al. Wedge-Shaped Microfluidic Chip for Circulating Tumor Cells Isolation and Its Clinical Significance in Gastric Cancer. *J Trans Med* (2018) 16:1–12. doi: 10.1186/s12967-018-1521-8
30. Ligthart ST, Coumans F, Attard G, Cassidy AM, Bono J, Terstappen L. Unbiased and Automated Identification of a Circulating Tumour Cell Definition That Associates with Overall Survival. *PloS One* (2011) 6:e27419. doi: 10.1371/journal.pone.0027419
31. Kraeft S,K. Reliable and Sensitive Identification of Occult Tumor Cells Using the Improved Rare Event Imaging System. *Clin Cancer Res* (2004) 10:3020–8. doi: 10.1158/1078-0432.CCR-03-0361
32. Zhou M, Hui Z, Wang Z, Li R, Liu X, Zhang W, et al. Precisely Enumerating Circulating Tumor Cells Utilizing a Multi-Functional Microfluidic Chip and Unique Image Interpretation Algorithm. *Theranostics* (2017) 7:4710–21. doi: 10.7150/thno.20440
33. Liu C, Wei D, Xiang J, Ren F, Huang L, Lang J, et al. An Improved Anticancer Drug-Response Prediction Based on an Ensemble Method Integrating Matrix Completion and Ridge Regression. *Mol Ther Nucleic Acids* (2020) 21:676–86. doi: 10.1016/j.omtn.2020.07.003
34. Tang X, Cai L, Meng Y, Xu J, Lu C, Yang J. Indicator Regularized Non-Negative Matrix Factorization Method-Based Drug Repurposing for COVID-19. *Front Immunol* (2020) 11:603615. doi: 10.3389/fimmu.2020.603615
35. Yang J, Peng S, Zhang B, Houten S, Schadt E, Zhu J, et al. Human Geroprotector Discovery by Targeting the Converging Subnetworks of Aging and Age-Related Diseases. *Geroscience* (2020) 42:353–72. doi: 10.1007/s11357-019-00106-x
36. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine Learning for Medical Imaging. *Radiographics* (2017) 37:160130. doi: 10.1148/rg.2017160130
37. Ma X, Xi B, Zhang Y, Zhu L, Yang J. A Machine Learning-based Diagnosis of Thyroid Cancer Using Thyroid Nodules Ultrasound Images. *Curr Bioinf* (2019) 14:349–58. doi: 10.2174/1574893614666191017091959
38. Lundervold AS, Lundervold A. An Overview of Deep Learning in Medical Imaging Focusing on MRI. *Z für Medizinische Physik* (2018) 29:102–27. doi: 10.1016/j.zemedi.2018.11.002
39. Mcbee MP, Awan OA, Colucci AT, Ghobadi CW, Auffermann WF. Deep Learning in Radiology. *Acad Radiol* (2018) 25:1472–80. doi: 10.1016/j.acra.2018.02.018
40. Maier A, Syben C, Lasser T, Riess C. A gentle Introduction to Deep Learning in Medical Image Processing. *Z für Medizinische Physik* (2019) 29:86–101. doi: 10.1016/j.zemedi.2018.12.003
41. Anthimopoulos M, Christodoulidis S, Ebner L, Christe A, Mougiakakou S. Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network. *IEEE Trans Med Imaging* (2016) 35:1207–16. doi: 10.1109/TMI.2016.2535865
42. Poplin R, Varadarajan AV, Blumer K, Liu Y, Mcconnell MV, Corrado GS, et al. Predicting Cardiovascular Risk Factors from Retinal Fundus Photographs using Deep Learning. *Nat Biomed Eng* (2017) 2(3):158–64. doi: 10.1038/s41551-018-0195-0
43. Le N, Quang-Thai H, Ou YY. Classifying the Molecular Functions of Rab GTPases in Membrane Trafficking Using Deep Convolutional Neural Networks. *Anal Biochem* (2018) 555:33–41. doi: 10.1016/j.ab.2018.06.011
44. Domínguez C, Heras J, Pascual V. IJ-OpenCV: Combining ImageJ and OpenCV for Processing Images In Biomedicine. *Comput Biol Med* (2017) 84:189–94. doi: 10.1016/j.compbiomed.2017.03.027
45. Le N, Huynh TT, Yapp E, Yeh HY. Identification of Clathrin Proteins by Incorporating Hyperparameter Optimization in Deep Learning and PSSM Profiles. *Comput Methods Programs Biomed* (2019) 177:81–8. doi: 10.1016/j.cmpb.2019.05.016
46. Xing F, Xie Y, Yang L. An Automatic Learning-Based Framework for Robust Nucleus Segmentation. *IEEE Trans Med Imaging* (2016) 35:550–66. doi: 10.1109/TMI.2015.2481436
47. Copie-Bergman C, Gaulard P, Leroy K, Briere J, Baia M, Jais JP, et al. Immuno-Fluorescence *in situ* Hybridization Index Predicts Survival in Patients With Diffuse Large B-Cell Lymphoma Treated With R-CHOP: a GELA Study. *J Clin Oncol* (2009) 27:5573–9. doi: 10.1200/JCO.2009.22.7058
48. Tamminga M, Groen HH, Hiltermann TJ. Investigating CTCs in NSCLC-a Reaction to the Study of Jia-Wei Wan: A Preliminary Study on the Relationship Between Circulating Tumor Cells Count and Clinical Features in Patients with Non-Small Cell Lung Cancer. *J Thorac Dis* (2016) 8(6):1032–6. doi: 10.21037/jtd.2016.04.17
49. Yu H, Ma L, Zhu Y, Li W, Ding L, Gao H. Significant Diagnostic Value of Circulating Tumour Cells in Colorectal Cancer. *Oncol Lett* (2020) 20:317–25. doi: 10.3892/ol.2020.11537
50. Li J, Wang Z, Chong T, Chen H, Li H, Li G, et al. Over-Expression of a Poor Prognostic Marker in Prostate Cancer: AQP5 Promotes Cells Growth and Local Invasion. *World J Surg Oncol* (2014) 12:284. doi: 10.1186/1477-7819-12-284
51. Qian Y, Wu Y, Yuan Z, Niu X, He Y, Peng J, et al. The Frequency of Circulating Tumour Cells and the Correlation with the Clinical Response to Standard Chemoradiotherapy in Locally Advanced Nasopharyngeal Carcinoma: A Prospective Study. *Cancer Manag Res* (2019) 11:10187–93. doi: 10.2147/CMAR.S222916
52. Bradski G. The OpenCV Library. *Dr Dobb's J: Software Tools Prof Programmer* (2000) 25:120–6. doi: 10.1002/1096-9128(200011)12:13<1317::AID-CPE536>3.0.CO;2-K
53. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *Comput Sci* (2014) 1409.1556.
54. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet Large Scale Visual Recognition Challenge. *Int J Comput Vision* (2015) 115:211–52. doi: 10.1007/s11263-015-0816-y
55. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification With Deep Convolutional Neural Networks. *Commun ACM* (2017) 60:84–90. doi: 10.1145/3065386
56. Pan SJ, Qiang Y. A Survey on Transfer Learning. *IEEE Trans Knowledge Data Eng* (2010) 22:1345–59. doi: 10.1109/TKDE.2009.191
57. Vilalta R, Giraud-Carrier C, Brazdil P, Soares C. *Inductive Transfer*. Springer US (2017). doi: 10.1007/978-1-4899-7687-1
58. Weiss K, Khoshgoftaar TM, Wang DD. A Survey of Transfer Learning. *J Big Data* (2016) 3:1–40. doi: 10.1186/s40537-016-0043-6

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.

frontiers in Oncology

ORIGINAL RESEARCH
published: 17 March 2022
doi: 10.3389/fonc.2022.846568

The Value of Shear Wave Elastography in the Diagnosis of Breast Cancer Axillary Lymph Node Metastasis and Its Correlation With Molecular Classification of Breast Masses

Changyun Luo[1†], *Li Lu*[2†], *Weifu Zhang*[3], *Xiangqi Li*[4], *Ping Zhou*[5] *and Zhangshen Ran*[1*]

[1] Regular Physical Examination Centre, The Second Affiliated Hospital of Shandong First Medical University, Taian, China, [2] Ultrasonography Department, The Second Affiliated Hospital of Shandong First Medical University, Taian, China, [3] Public Health Section, The Second Affiliated Hospital of Shandong First Medical University, Taian, China, [4] Breast Surgery, The Second Affiliated Hospital of Shandong First Medical University, Taian, China, [5] Liyang People's Hospital, Liyang, China

OPEN ACCESS

Edited by:
Min Tang,
Jiangsu University, China

Reviewed by:
Chen Li,
Free University of Berlin, Germany
Xun Gong,
Anhui Medical University, China

***Correspondence:**
Zhangshen Ran
ranzhangshen@163.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 31 December 2021
Accepted: 11 February 2022
Published: 17 March 2022

Citation:
Luo C, Lu L, Zhang W, Li X, Zhou P and Ran Z (2022) The Value of Shear Wave Elastography in the Diagnosis of Breast Cancer Axillary Lymph Node Metastasis and Its Correlation With Molecular Classification of Breast Masses. Front. Oncol. 12:846568. doi: 10.3389/fonc.2022.846568

Objective: To explore the diagnostic value of shear wave elastography examination (SWE) on axillary node metastasis (ANM) in breast cancer, this study aimed to evaluate the correlation between the SWE features and different molecular types of breast cancer, and to check the elastic modulus differences among the molecular types.

Methods: Breast cancer patients from November 2020 to December 2021 were subjected to both conventional ultrasonic examination (CUE) and SWE before ultrasound-guided percutaneous biopsy or axillary lymph node dissection (ALND). We used the pathological results as the gold standard to draw the receiver operating characteristic (ROC) curve.

Results: SWE outperforms CUE, but their conjunctive use is the best option. No significant correlation was found between the elastic modulus values and the molecular types of breast cancer.

Conclusion: SWE can be used as an routine auxiliary method of CUE for ANM.

Keywords: shear wave elastography, breast cancer, metastasis, molecular classification, diagnosis

INTRODUCTION

In the 21st century, cancer is the top cause of death in hospitals and the key limitation of life expectancy in most countries, whatever level their economics and social civilization (1, 2). The Global Cancer Statistics Report published by the United States Cancer Research Institute shows that breast cancer surpassed lung cancer in 2020, becoming the most common type of cancer among female patients and the main cause of death for female patients in 185 countries (3). According to statistics, 19.3 million people were diagnosis with cancer and nearly 10 million deaths worldwide in 2020 were because of cancer (4). Around 2.3 million new-onset cases are of female breast cancer

accounting for 11.7%, and breast cancer causes 0.69 million deaths which account for 6.9% of all global cancer deaths (5). In fact, 1 in every 18 women will develop breast cancer globally, and the clinical manifestations and prognosis of patients are different (6). Age, molecular subtype, and axillary lymph node status are considered to be independent factors affecting the prognosis of patients suffering from breast cancer (7). In addition, tumor-related factors such as pathological type, grade, and stage can also explain the higher mortality of breast cancer to a certain extent (8).

With the development of molecular biology, it has been recognized that breast cancer has large biological diversity and high heterogeneity, which result in different morphological subtypes, recurrence rate, targeted therapy strategies, and survival risks (9–11). Therefore, if the patients with breast cancer can be accurately classified, it should help to select individualized precision treatment and effectively predict the prognosis (12–14). According to immunohistochemical indexes such as estrogen receptor (ER), progesterone receptor (PR), proliferating cell nuclear antigen (Ki-67), and human epidermal growth factor receptor-2 (HER-2), clinicians determine the molecular subtype of breast cancer, namely luminal A, luminal B, Her-2-positive, and triple-negative breast cancer (TNBC) (15–18), where HER2-positive includes HR-negative and HR-positive (15, 19–23). The status of axillary lymph nodes is also an important factor influencing the prognosis of patients. According to reports, 70-80% of early non-metastatic breast cancer patients can be cured. Patients with advanced breast cancer and distant organ metastasis are considered to be incurable by existing therapies. The prognosis of patients with advanced breast cancer is poor, and the 5-year survival rate is only 20% (24). In addition, the axillary lymph node (ALN) is deemed as the first site to be metastasized by breast cancer through the lymphatic vessels (25).

Shear wave elastography examination (SWE) is a newly emerging elastography technique, which can display tissue stiffness in a quantified form to obtain the biological information of the primary lesion (26–30). At present, many studies have verified the diagnostic value of SWE for benign and malignant lesions in breasts (31–34). The technique has been widely employed to check the thyroid, pancreas, kidney, prostate, liver, and other organs while few studies about axillary node metastasis (ANM) and its application for the molecular classification of breast cancer were reported (35–39). Here, we applied SWE to assess the axillary node status of patients with breast cancer with a goal to explore its feasibility in the diagnosis of ANM, and to check the relationship of the SWE elastic modulus and the molecular types of breast cancer.

MATERIALS AND METHODS

Research Objects

After the pathological verification for breast cancer, 114 patients who had never received any treatment were recruited in the Second Affiliated Hospital of Shandong First Medical University from November 2020 to December 2021 (40). The mean age is 52.52 ± 9.03 (range, 31-75 years old), and the mean long diameter of the lymph node is 1.60 ± 0.70 (range, 0.5 ± 4.8 cm). All of the patients underwent conventional ultrasonic examination (CUE) before ultrasound-guided percutaneous biopsy on the axillary lode or axillary lymph node dissection (ALND). Some key indexes were scored (see **Table 1**) . This program was approved by the Medical Ethics Committee of the Second Affiliated Hospital of Shandong First Medical University.

Shear Wave Elastography Examination

The Toshiba Apio500 ultrasonic diagnostic machine equipped with high-frequency linear array probe PLY-805AT (2.0-12.0 MHZ) called shear wave was used for SWE. Based on the operations of Skerl et al., the parameters of SWE were set when ROI = 2 mm (41). Both of the transverse and the longitudinal sections of each breast mass and suspicious lymph node were measured three times to obtain average values.

Image Analysis

Two physicians who have more than 5 years of experience in breast and axillary lymph node diagnosis analyzed the image results. A score was evaluated based on the aspect ratio and the short axis diameter of the lymph node (42). Plus, the maximum value (Emax), average value (Emean), and minimum value (Emin) of Young's modulus were assessed by SWE, and a static image was kept (43, 44). Afterwards, univariate analysis was performed using ALN metastasis as a dependent variable and the CUE scores. The obtained indexes with statistical significance were extracted for multivariate logistic regression analysis with the Emean of SWE as independent variables. In the predictive model, ANMs were confirmed as benign or malignant lesions. Then, the predictive results were compared with the below pathological results to draw the gold standard receiver operating characteristic (ROC) curves and get the area under the curve (AUC) values.

The patients who underwent surgical treatment were classified into four groups according to their molecular classification results to check whether the elastic modulus values of SWE were statistically different between the groups, and explore its relationship with the classification strategy.

Pathological Examination

Ultrasound-guided axillary nodal puncture was accomplished in 62 patients and breast mass resection was done in 86 patients. Tissue biopsy including postoperative pathological section and immunohistochemical examination was implemented (45, 46).

TABLE 1 | Criteria and evaluation of CUE in the diagnosis of lymph node status.

Index	1 point	2 point
Aspect ratio	>2	<2
Short axis diameter	<7 mm	>7 mm
Lymphatic hilus	Yes	No
Cortical thickness	<3 mm	>3 mm
Blood flow type	Gate type	Peripheral type or mixed type

Statistical Analysis

SPSS 25.0 software was used to process the above data. The measured data were expressed as mean ± standard deviation, and the count data were expressed as rate (%). The chi-square test was used to compare the two-category data between the two groups, and the KAPPA test was used to compare the consistency of the diagnosis results of CUE, SWE, and their conjunctive usage with the pathological results. Multivariate logistic regression analysis was used to construct a prediction model of CUE combined with SWE to obtain the prediction probability.

RESULTS

Pathological Results

A total of 114 women with breast cancer were enrolled in this study, who were then divided into two groups: the ANM group with 58 cases and the non-metastasis group with 56 cases. The mean age, medical course, and ALN size in the former group were 49.59 ± 8.54 years, 11.0 ± 25.22 months, and 2.16 ± 2.63 cm. In the latter group, the same data were 51.48 ± 9.50 years, 5.70 ± 14.27 months, and 1.36 ± 0.51 cm.

Comparison of the Two Groups With/Without ANM

As **Table 2** shows, on one hand, the CUE comparison displayed significant differences in lymph nodal size, aspect ratio, and short axis diameter. However, the hilum structures, cortex thicknesses, and blood-flow types did not show any difference. On the other hand, the SWE comparison also showed significant differences in the Emax, Emean, and Emin values.

Comparison of the Consistency Between the Diagnosis Results of the Three Medical Means and the Pathological Outcomes

In the CUE assessment for ANM, as **Table 3** shows, compared with the pathological results, the accuracy, sensitivity, specificity, positive predictive value, and negative predictive value calculated from the exclusive usage of CUE in the diagnosis of malignant lymph nodes were 65.8%, 72.4%, 58.9%, 64.6%, and 67.3%, respectively. As a result, the whole consistency with the pathological results was 0.314.

In the SWE assessment for ANM, according to the literature, Emean>18.7 was set as the metastatic threshold of the lymph nodes. When the average stiffness of the lymph node was greater than 18.7 Kpa, we believed that the lymph node was more likely to be malignant. On the contrary, the lymph node was more likely to be benign. Compared with the pathological results, the sensitivity, specificity, and positive and negative predictive value of SWE for malignancy were 70.7%, 76.7%, 75.9%, and 71.6%, respectively, which resulted in a diagnostic accuracy of 73.6%, and the consistency with the pathological result was 0.474.

In the conjunctive use of CUE and SWE to assess ANM, whether metastasis exists was used as the dependent variable and the indicators observed by CUE were used as independent variables, and univariate analysis was performed. It was found that the aspect ratio and short axis diameter of lymph nodes observed by CUE were statistically significant with ANM. Then, taking the average elastic modulus of SWE, the aspect ratio of the lymph node, and the short axis diameter of the lymph node as the independent variables into the multivariate logistic regression, a predictive model was constructed. The results show that Emean and both the aspect ratio of the lymph node and the short axis diameter of the lymph node can be entered into the equation.

Based on the above results, three ROC curves were drawn and the corresponding AUC values were calculated (see **Figure 1**).The results showed that conjunctive use (AUC, 0.88) had the best predictive ability compared to exclusive use of CUE (AUC, 0.657) or SWE (AUC, 0.737).

The Best Cut-Off Value of SWE for the Diagnosis of ANM

Although many studies have shown that quantitative SWE can help diagnose breast diseases, the cut-off values used were different. In order to evaluate the optimal SWE parameters to quantify ANM, ROC curves for Emax, Emean, and Emin were also drawn. We suggest that when Emean=23.2 is used as the cut-off value, SWE is the best (see **Figure 2**).

TABLE 2 | Comparison of CUE and SWE elastic modulus between the two groups with/without ANM.

	No lymph node metastases	Lymph node metastases	T/X^2	p value
Lymph node size	1.36 ± 0.51	2.16 ± 2.63	-2.218	0.029
Aspect ratio >2	41 (73.2%)	13 (22.4%)		
<2	15 (26.8%)	45 (77.6%)	29.49	0
Short axis diameter <7	44	19	24.18	0
>7	12	39		
Lymphatic hilus	26	26	0.029	0.864
No lymphatic hilus	30	32		
Gate type	28	21	2.21	0.137
Not gate type	28	37		
Cortical thickness <3	11	13	0.13	0.717
>3	45	45		
Emax	24.68 ± 18.91	77.68 ± 48.06	-7.693	0
Emean	17.34 ± 14.13	58.33 ± 42.31	-6.887	0
Emin	13.35 ± 10.39	39.77 ± 36.79	-5.255	0

TABLE 3 | Comparison of diagnostic efficacy of CUE, SWE, and their conjunctive use in ANM.

	Sensitivity	Specificity	Positive predictive value	Negative predictive value	Accuracy	Kappa
CUE	72.40%	58.90%	64.60%	67.30%	65.80%	0.314
SWE	70.70%	76.70%	75.90%	71.60%	73.60%	0.474
CUE+SWE	79.30%	82.10%	82.10%	79.30%	80.70%	0.616

The SWE Elastic Modulus Value and Analysis of Variance Results of Different Molecular Types of Breast Cancer

According to the expression of ER, PR, HER-2, and Ki-67, 86 people who underwent surgery were divided into six groups (19, 47). Variance analysis on the elastic modulus values for each group was performed. We found that all of the comparisons of Emax, Emean, and Emin between the molecular types did not have statistical significance (see **Table 4**).

DISCUSSION

Worldwide, the incidence and mortality of breast cancer always takes the top cancer spot in female patients (3). Several reports show that the incidence of breast cancer has been increasing year by year in the past 5 years (48, 49), and the choice of treatment strategy is determined by the state of ALNs, which decide the final bill of the patients (50). Therefore, it is substantially important to correctly assess the status of ALNs in breast cancer patients before surgery.

Although CUE has high sensitivity in the diagnosis of breast cancer, some previous studies have found that the accuracy, sensitivity, and specificity of ultrasound in diagnosing ALN metastasis is not so high (51–53). The concept of SWE was first proposed by Sarvazyan et al. in 1998 (54). Its principle is to use acoustic radiation force pulses (ARFI) to apply pressure to the tissues to induce mechanical vibrations in the tissues. In the process, by collecting the shear echo signal reflected by the tissue vibration, the propagation speed in the tissue can by calculated and automatically converted into the elastic modulus value through the conversion system. Consequently, the hardness information of the tissue can be quantified (55). Nowadays, a large number of studies have shown that the advantages of SWE in the application of breast, thyroid, prostate, liver, and other organs, but no research has reported the assessment of ALN metastasis. Besides, there is no standard for the cut-off value of SWE in the diagnosis of metastatic lymph nodes.

In this study, the Emax, Emean, and Emin of the lymph node metastasis group were all higher than those of the lymph node non-metastasis group, and the difference was statistically significant ($p<0.05$). The average elasticity of benign and malignant lymph nodes was 17.34 Kpa and 58.33 Kpa, respectively. The average value we obtained was higher than some previous studies, which may be longer than the course of some breast cancer patients in our study. The tumor cells synthesize a large number of collagen fibers and lymphocytes in the tumor microenvironment during the process of metastasis. Infiltration changes increased the stiffness of the lymph nodes in this part of the patients, leading to a corresponding increase in the average stiffness of the lymph node metastasis group. In this study, Emean=18.7 kpa was selected as the critical value of metastatic lymph nodes. A preliminary exploration was carried out on shear wave elastography to assess lymph node metastasis. A total of 114 lymph nodes were examined by shear wave elastography, and 41 cases of metastatic lymph nodes were correctly diagnosed, which were benign. There were 43 cases of lymph nodes and 30 cases of

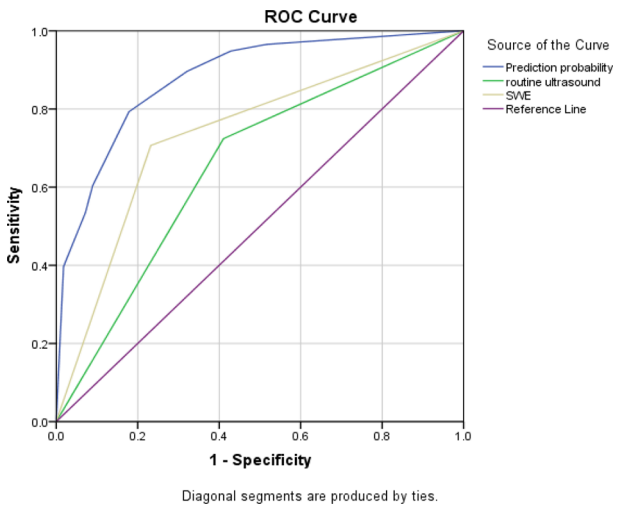


FIGURE 1 | Three ROC curves for the sensitivity and specificity of CUE, SWE, and their conjunctive use in ANM diagnosis.

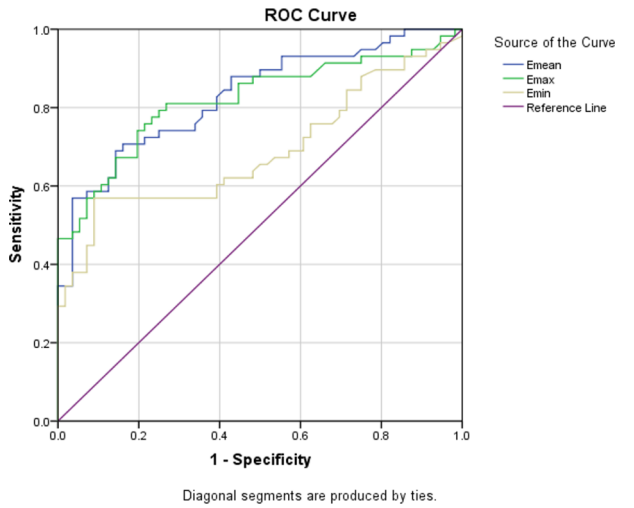


FIGURE 2 | Three ROC curves for the sensitivity and specificity of the Emean, Emax, and Emin in the application of ANM diagnosis.

TABLE 4 | Variance analysis of SWE elastic modulus and different molecular subtypes of breast cancer.

	Emax (Kpa)	Emean (Kpa)	Emin (Kpa)
Luminal A type	119.04 ± 17.63	88.07 ± 23.02	59.53 ± 24.19
Luminal B type	112.43 ± 14.62	80.97 ± 11.65	51.07 ± 10.29
HER-2 high expression	119.51 ± 21.60	89.19 ± 21.03	60.50 ± 20.82
Triple-negative	126.87 ± 16.45	98.14 ± 17.53	68.90 ± 16.53
F	1.372	1.714	1.719
P	0.257	0.171	0.169

misdiagnosis. Its specificity (76.7%) and accuracy (73.6%) were higher than those of conventional ultrasound, but its sensitivity (70.7%) was lower, so it could not be used as a substitute for conventional ultrasound. The reasons for the misdiagnosis included: 1) It may be because the volume of some lymph nodes is relatively small or the location is relatively deep. Affected by the anatomical structure of the axilla, the shear wave cannot spread well, resulting in a low measured elastic modulus value. 2) There may be liquefaction and necrosis in some malignant lymph nodes, and there are relatively few elastic and collagen fibers in them, so the measured elastic modulus value is not high.

Additionally, as we believe that the average stiffness of metastatic lymph nodes has the highest specificity, the conjunctive use of Emean and CUE can form complementary advantages, obtaining more objective information to determine which lymph nodes are suitable for biopsy. However, up to now, the optimal cut-off value of each parameter of SWE has not yet reached agreement. It may be affected by many factors, such as pre-compression, the machine model and the depth of the lesion, and the progression of the patient's disease (56, 57). Therefore, studies with a larger sample size involving multiple units should be considered. Plus, we tried to employ SWE to predict the molecular type of breast cancer with the elastic modulus values (58, 59). Unfortunately, no significant difference was found between the six different groups. This is in line with the conclusion drawn by previously published papers (60, 61). As a result, we do not recommend the implementation of the molecular classification of breast cancer *via* SWE at this stage. However, with the increase of clinical experience, doctors have gradually realized the value of SWE in the diagnosis of breast cancer axillary lymph node metastasis (62–65).

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 study was approved by the Medical Ethics Committee of the Second Affiliated Hospital of Shandong First Medical University.

AUTHOR CONTRIBUTIONS

ZSR designed the study. WZ and XL recruited the patients and summarized their medical records. CL and LL performed the data analysis and wrote the manuscript. PZ revised the manuscript. All the authors made a direct and intellectual contribution to this topic and approved the article for publication.

FUNDING

This work was supported by grants from the Shandong Province Traditional Chinese Medicine Science and Technology Development Plan (No. 2015-266); the Shandong First Medical University Academic Improvement Plan (No. 2019QL017); the Shandong Province Medicine and Health Science and Technology Development Plan (No. 202009020793); the Shandong Province Natural Science Foundation (No. ZR2020MH357); and the Tai'an Science and Technology Innovation Development Project (No. 2020NS205).

ACKNOWLEDGMENTS

We thank all the individuals who have helped us in this study. We acknowledge the valuable work of the many investigators whose published articles we were unable to cite owing to space limitations.

REFERENCES

1. Global Burden of Disease Cancer, C, Fitzmaurice C, Abate D, Abbasi N, Abbastabar H, Abd-Allah F, et al. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years for 29 Cancer Groupto 2017: A Systematic Analysis for the Global Burden of Disease Study. *JAMA Oncol* (2019) 5 (12):1749–68. doi: 10.1001/jamaoncol.2019.2996
2. Cao W, Chen HD, Yu YW, Li N, Chen WQ. Changing Profiles of Cancer Burden Worldwide and in China: A Secondary Analysis of the Global Cancer Statistics 2020. *Chin Med J (Engl)* (2021) 134(7):783–91. doi: 10.1097/CM9.0000000000001474
3. Coughlin SS. Epidemiology of Breast Cancer in Women. *Adv Exp Med Biol* (2019) 1152:9–29. doi: 10.1007/978-3-030-20301-6_2
4. Youlden DR, Cramb SM, Yip CH, Baade PD. Incidence and Mortality of Female Breast Cancer in the Asia-Pacific Region. *Cancer Biol Med* (2014) 11(2):101–15. doi: 10.7497/j.issn.2095-3941.2014.02.005
5. 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
6. Thorat MA, Balasubramanian R. Breast Cancer Prevention in High-Risk Women. *Best Pract Res Clin Obstet Gynaecol* (2020) 65:18–31. doi: 10.1016/j.bpobgyn.2019.11.006
7. Loibl S, Poortmans P, Morrow M, Denkert C, Curigliano G. Breast Cancer. *Lancet* (2021) 397(10286):1750–69. doi: 10.1016/s0140-6736(20)32381-3
8. Johansson ALV, Trewin CB, Hjerkind KV, Ellingjord-Dale M, Johannesen TB, Ursin G. Breast Cancer-Specific Survival by Clinical Subtype After 7 Years Follow-Up of Young and Elderly Women in a Nationwide Cohort. *Int J Cancer* (2019) 144(6):1251–61. doi: 10.1002/ijc.31950
9. Jin L, Han B, Siegel E, Cui Y, Giuliano A, Cui X. Breast Cancer Lung Metastasis: Molecular Biology and Therapeutic Implications. *Cancer Biol Ther* (2018) 19(10):858–68. doi: 10.1080/15384047.2018.1456599
10. Waks AG, Winer EP. Breast Cancer Treatment: A Review. *JAMA* (2019) 321(3):288–300. doi: 10.1001/jama.2018.19323
11. Khrouf S, Ksontini FL, Ayadi M, Rais HBA, Mezlini A. Breast Cancer Screening- a Dividing Controversy. *Tunis Med* (2020) 98(1):22–34.
12. Tsang JYS, Tse GM. Molecular Classification of Breast Cancer. *Adv Anat Pathol* (1998) 27(1):27–35. doi: 10.1097/PAP.0000000000000232
13. Cejalvo JM, Martinez de Duenas E, Galvan P, Garcia-Recio S, Burgues Gasion O, Pare L, et al. Intrinsic Subtypes and Gene Expression Profiles in Primary and Metastatic Breast Cancer. *Cancer Res* (2017) 77(9):2213–21. doi: 10.1158/0008-5472.CAN-16-2717
14. Curigliano G, Burstein HJ, Winer EP, Gnant M, Dubsky P, Loibl S, et al. De-Escalating and Escalating Treatments for Early-Stage Breast Cancer: The St. Gallen International Expert Consensus Conference on the Primary Therapy of Early Breast Cancer 2017. *Ann Oncol* (2017) 28(8):1700–12. doi: 10.1093/annonc/mdx308
15. Denkert C, Liedtke C, Tutt A, von Minckwitz G. Molecular Alterations in Triple-Negative Breast Cancer—the Road to New Treatment Strategies. *Lancet* (2017) 389(10087):2430–42. doi: 10.1016/s0140-6736(16)32454-0
16. Garrido-Castro AC, Lin NU, Polyak K. Insights Into Molecular Classifications of Triple-Negative Breast Cancer: Improving Patient Selection for Treatment. *Cancer Discov* (2019) 9(2):176–98. doi: 10.1158/2159-8290.CD-18-1177
17. Chang-Qing Y, Jie L, Shi-Qi Z, Kun Z, Zi-Qian G, Ran X, et al. Recent Treatment Progress of Triple Negative Breast Cancer. *Prog Biophys Mol Biol* (2020) 151:40–53. doi: 10.1016/j.pbiomolbio.2019.11.007
18. Won KA, Spruck C. Triplenegative Breast Cancer Therapy: Current and Future Perspectives (Review). *Int J Oncol* (2020) 57(6):1245–61. doi: 10.3892/ijo.2020.5135
19. Rakha EA, Green AR. Molecular Classification of Breast Cancer: What the Pathologist Needs to Know. *Pathology* (2017) 49(2):111–9. doi: 10.1016/j.pathol.2016.10.012
20. Wolff AC, Hammond MEH, Allison KH, Harvey BE, Mangu PB, Bartlett JMS, et al. Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer-American Society of Clinical Oncology:College of American Pathologists Clinical Practice Guideline Focused Update. *J Clin Oncol* (2018) 36(20):2105–22. doi: 10.1200/JCO.2018
21. Barzaman K, Karami J, Zarei Z, Hosseinzadeh A, Kazemi MH, Moradi-Kalbolandi S, et al. Breast Cancer: Biology, Biomarkers, and Treatments. *Int Immunopharmacol* (2020) 84:106535. doi: 10.1016/j.intimp.2020.106535
22. Kreutzfeldt J, Rozeboom B, Dey N, De P. The Trastuzumab Era- Current and Upcoming Targeted HER2+ Breast Cancer Therapies. *Am J Cancer Res* (2020) 10(4):1045–67.
23. Ruschoff J, Nagelmeier I, Jasani B, Stoss O. ISH-Based HER2 Diagnostics. *Pathologe* (2021) 42(Suppl 1):62–8. doi: 10.1007/s00292-020-00878-6
24. Harbeck N, Penault-Llorca F, Cortes J, Gnant M, Houssami N, Poortmans P, et al. Breast Cancer. *Nat Rev Dis Primers* (2019) 5(1):66. doi: 10.1038/s41572-019-0111-2
25. Chas M, Boivin L, Arbion F, Jourdan ML, Body G, Ouldamer L. Clinicopathologic Predictors of Lymph Node Metastasis in Breast Cancer Patients According to Molecular Subtype. *J Gynecol Obstet Hum Reprod* (2018) 47(1):9–15. doi: 10.1016/j.jogoh.2017.10.008
26. Sarvazyan AP, Skovoroda AR, Emelianov SY, Fowlkes JB, Pipe JG, Adler RS, et al. Biophysical Bases of Elasticity Imaging. *Acoustical Imaging* (1995) 21:223–40. doi: 10.1007/978-1-4615-1943-0_23
27. Bercoff J, Tanter M, Fink M. Supersonic Shear Imaging- A New Technique for Soft Tissue Elasticity Mapping. *IEEE Trans Ultrason Ferroelectr Freq Control* (2004) 51(4):396–409. doi: 10.1109/TUFFC.2004.1295425
28. Athanasiou A, Tardivon A, Tanter M, Sigal-Zafrani B, Bercoff J, Deffieux T, et al. Breast Lesions- Quantitative Elastography With Supersonic Shear Imaging—Preliminary Results. *Radiology* (2010) 256(1):297–303. doi: 10.1148/radiol.10090385
29. Evans A, Whelehan P, Thomson K, McLean D, Brauer K, Purdie C, et al. Quantitative Shear Wave Ultrasound Elastography- Initial Experience in Solid Breast Masses. *Breast Cancer Res* (2010) 12(6):R104. doi: 10.1186/bcr2787
30. Berg WA, Cosgrove DO, Doré CJ, Schäfer FKW, Svensson WE, Hooley RJ, et al. Shear-Wave Elastography Improves the Specifi City of Breast US- The BE1 Multinational Study of 939 Masses. *Radiology* (2012) 262(2):435–49. doi: 10.1148/radiol.11110640/-/DC1
31. Nightingale K, McAleavey S, Trahey G. Shear-Wave Generation Using Acoustic Radiation Force- In Vivo and Ex Vivo Results. *Ultrasound Med Biol* (2003) 29(12):1715–23. doi: 10.1016/S0301-5629(03)01080-9
32. Chang JM, Moon WK, Cho N, Yi A, Koo HR, Han W, et al. Clinical Application of Shear Wave Elastography (SWE) in the Diagnosis of Benign and Malignant Breast Diseases. *Breast Cancer Res Treat* (2011) 129(1):89–97. doi: 10.1007/s10549-011-1627-7
33. Herman J, Sedlackova Z, Furst T, Vachutka J, Salzman R, Vomacka J, et al. The Role of Ultrasound and Shear-Wave Elastography in Evaluation of Cervical Lymph Nodes. *BioMed Res Int* (2019) 2019:4318251. doi: 10.1155/2019/4318251
34. Gemici AA, Ozal ST, Hocaoglu E, Inci E. Relationship Between Shear Wave Elastography Findings and Histologic Prognostic Factors of Invasive Breast Cancer. *Ultrasound Q* (2020) 36(1):79–83. doi: 10.1097/RUQ.0000000000000471
35. Tozaki M, Fukuma E. Pattern Classification of ShearWave Elastography Images for Differential Diagnosis Between Benign and Malignant Solid Breast Masses. *Acta Radiol* (2011) 52(10):1069–75. doi: 10.1258/ar.2011.110276
36. Azizi G, Faust K, Mayo ML, Farrell J, Malchoff C. Diagnosis of Thyroid Nodule With New Ultrasound Imaging Modalities. *VideoEndocrinology* (2019) 7(1). doi: 10.1089/ve.2020.0173
37. Cirocchi R, Amabile MI, De Luca A, Frusone F, Tripodi D, Gentile P, et al. New Classifications of Axillary Lymph Nodes and Their Anatomical-Clinical Correlations in Breast Surgery. *World J Surg Oncol* (2021) 19(1):93. doi: 10.1186/s12957-021-02209-2
38. Liu M, Wang CB, Xie F, Peng Y, Wang SChinese Society of Breast, S. Clinical Practice Guidelines for Diagnosis and Treatment of Invasive Breast Cancer: Chinese Society of Breast Surgery (CSBrS) Practice Guidelines 2021. *Chin Med J (Engl)* (2021) 134(9):1009–13. doi: 10.1097/CM9.0000000000001498
39. Wen R, Dong J, Wang Y, Wang L. Combination of Color Doppler Ultrasound and CT for Diagnosing Breast Cancer. *Am J Transl Res* (2021) 13(9):10771–6.
40. Wang R, Zhu Y, Liu X, Liao X, He J, Niu L. The Clinicopathological Features and Survival Outcomes of Patients With Different Metastatic Sites in Stage IV Breast Cancer. *BMC Cancer* (2019) 19(1):1091. doi: 10.1186/s12885-019-6311-z
41. Skerl K, Vinnicombe S, Giannotti E, Thomson K, Evans A. Influence of Region of Interest Size and Ultrasound Lesion Size on the Performance of 2D Shear Wave Elastography (SWE) in Solid Breast Masses. *Clin Radiol* (2015) 70(12):1421–7. doi: 10.1016/j.crad.2015.08.010
42. Zhou J, Zhang B, Dong Y, Yu L, Gao T, Wang Z. Value on the Diagnosis of Axillary Lymph Node Metastasis in Breast Cancer by Color Doppler Ultrasound Combined With Computed Tomography. *J BUON* (2020) 25(4):1784–91.
43. Wang J, Ben Z, Gao S, Lyu S, Wei X. The Role of Ultrasound Elastography and Virtual Touch Tissue Imaging in the Personalized Prediction of Lymph Node Metastasis of Breast Cancer. *Gland Surg* (2021) 10(4):1460–9. doi: 10.21037/gs-21-199
44. Zhang Q, Agyekum EA, Zhu L, Yan L, Zhang L, Wang X, et al. Clinical Value of Three Combined Ultrasonography Modalities in Predicting the Risk of Metastasis to Axillary Lymph Nodes in Breast Invasive Ductal Carcinoma. *Front Oncol* (2021) 11:715097. doi: 10.3389/fonc.2021.715097
45. Wen X, Yu X, Tian Y, Liu Z, Cheng W, Li H, et al. Quantitative Shear Wave Elastography in Primary Invasive Breast Cancers, Based on Collagen-S100A4

Pathology, Indicates Axillary Lymph Node Metastasis. *Quant Imaging Med Surg* (2020) 10(3):624–33. doi: 10.21037/qims.2020.02.18
46. Leng X, Japaer R, Zhang H, Yeerlan M, Ma F, Ding J. Relationship of Shear Wave Elastography Anisotropy With Tumor Stem Cells and Epithelial-Mesenchymal Transition in Breast Cancer. *BMC Med Imaging* (2021) 21 (1):171. doi: 10.1186/s12880-021-00707-z
47. Yang H, Wang R, Zeng F, Zhao J, Peng S, Ma Y, et al. Impact of Molecular Subtypes on Metastatic Behavior and Overall Survival in Patients With Metastatic Breast Cancer: A Single-Center Study Combined With a Large Cohort Study Based on the Surveillance, Epidemiology and End Results Database. *Oncol Lett* (2020) 20(4):87. doi: 10.3892/ol.2020.11948
48. Cronin KA, Lake AJ, Scott S, Sherman RL, Noone AM, Howlader N, et al. Annual Report to the Nation on the Status of Cancer, Part I: National Cancer Statistics. *Cancer* (2018) 124(13):2785–800. doi: 10.1002/cncr.31551
49. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistic. *CA Cancer J Clin* (2021) 71(1):7–33. doi: 10.3322/caac.21654
50. Magnoni F, Galimberti V, Corso G, Intra M, Sacchini V, Veronesi P. Axillary Surgery in Breast Cancer: An Updated Historical Perspective. *Semin Oncol* (2020) 47(6):341–52. doi: 10.1053/j.seminoncol.2020.09.001
51. Zhao Q, Sun JW, Zhou H, Du LY, Wang XL, Tao L, et al. Pre-Operative Conventional Ultrasound and Sonoelastography Evaluation for Predicting Axillary Lymph Node Metastasis in Patients With Malignant Breast Lesions. *Ultrasound Med Biol* (2018) 44(12):2587–95. doi: 10.1016/j.ultrasmedbio.2018.07.017
52. Hotton J, Salleron J, Henrot P, Buhler J, Leufflen L, Rauch P, et al. Pre-Operative Axillary Ultrasound With Fine-Needle Aspiration Cytology Performance and Predictive Factors of False Negatives in Axillary Lymph Node Involvement in Early Breast Cancer. *Breast Cancer Res Treat* (2020) 183 (3):639–47. doi: 10.1007/s10549-020-05830-z
53. Du LW, Liu HL, Gong HY, Ling LJ, Wang S, Li CY, et al. Adding Contrast-Enhanced Ultrasound Markers to Conventional Axillary Ultrasound Improves Specificity for Predicting Axillary Lymph Node Metastasis in Patients With Breast Cancer. *Br J Radiol* (2021) 94(1118):20200874. doi: 10.1259/bjr.20200874
54. Sarvazyan AP, Rudenko OV, Swanson SD, Fowlkes JB, Emelianov SY. Shear Wave Elasticity Imaging- a New Ultrasonic Technology of Medical Diagnostics. *Ultrasound Med Biol* (1998) 24(9):1419–35. doi: 10.1016/S0301-5629(98)00110-0
55. Song EJ, Sohn YM, Seo M. Diagnostic Performances of Shear-Wave Elastography and B-Mode Ultrasound to Differentiate Benign and Malignant Breast Lesions: The Emphasis on the Cutoff Value of Qualitative and Quantitative Parameters. *Clin Imaging* (2018) 50:302–7. doi: 10.1016/j.clinimag.2018.05.007
56. Han P, Yang H, Liu M, Cheng L, Wang S, Tong F, et al. Lymph Node Predictive Model With in Vitro Ultrasound Features for Breast Cancer Lymph Node Metastasis. *Ultrasound Med Biol* (2020) 46(6):1395–402. doi: 10.1016/j.ultrasmedbio.2020.01.030
57. Hosonaga M, Saya H, Arima Y. Molecular and Cellular Mechanisms Underlying Brain Metastasis of Breast Cancer. *Cancer Metastasis Rev* (2020) 39(3):711–20. doi: 10.1007/s10555-020-09881-y
58. Wei X, Wang M, Wang X, Zheng X, Li Y, Pan Y, et al. Prediction of Cervical Lymph Node Metastases in Papillary Thyroid Microcarcinoma by Sonographic Features of the Primary Site. *Cancer Biol Med* (2019) 16 (3):587–94. doi: 10.20892/j.issn.2095-3941.2018.0310
59. Yang C, Dong J, Liu Z, Guo Q, Nie Y, Huang D, et al. Prediction of Metastasis in the Axillary Lymph Nodes of Patients With Breast Cancer: A Radiomics Method Based on Contrast-Enhanced Computed Tomography. *Front Oncol* (2021) 11:726240. doi: 10.3389/fonc.2021.726240
60. Au FW, Ghai S, Lu FI, Moshonov H, Crystal P. Quantitative Shear Wave Elastography: Correlation With Prognostic Histologic Features and Immunohistochemical Biomarkers of Breast Cancer. *Acad Radiol* (2015) 22 (3):269–77. doi: 10.1016/j.acra.2014.10.007
61. Suvannarerg V, Chitchumnong P, Apiwat W, Lertdamrongdej L, Tretipwanit N, Pisarnturakit P, et al. Diagnostic Performance of Qualitative and Quantitative Shear Wave Elastography in Differentiating Malignant From Benign Breast Masses, and Association With the Histological Prognostic Factors. *Quant Imaging Med Surg* (2019) 9(3):386–98. doi: 10.21037/qims.2019.03.04
62. Glechner A, Wockel A, Gartlehner G, Thaler K, Strobelberger M, Griebler U, et al. Sentinel Lymph Node Dissection Only Versus Complete Axillary Lymph Node Dissection in Early Invasive Breast Cancer: A Systematic Review and Meta-Analysis. *Eur J Cancer* (2013) 49(4):812–25. doi: 10.1016/j.ejca.2012.09.010
63. Wockel A, Albert US, Janni W, Scharl A, Kreienberg R, Stuber T. The Screening, Diagnosis, Treatment, and Follow-Up of Breast Cancer. *Dtsch Arztebl Int* (2018) 115(18):316–23. doi: 10.3238/arztebl.2018.0316
64. Cocco S, Leone A, Piezzo M, Caputo R, Di Lauro V, Di Rella F, et al. Targeting Autophagy in Breast Cancer. *Int J Mol Sci* (2020) 21(21). doi: 10.3390/ijms21217836
65. Miricescu D, Totan A, Stanescu S, Badoiu SC, Stefani C, Greabu M. PI3K/AKT/mTOR Signaling Pathway in Breast Cancer: From Molecular Landscape to Clinical Aspects. *Int J Mol Sci* (2020) 22(1). doi: 10.3390/ijms22010173

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.

frontiers in Oncology

ORIGINAL RESEARCH
published: 22 March 2022
doi: 10.3389/fonc.2022.853336

Preoperative Prediction of Microvascular Invasion Risk Grades in Hepatocellular Carcinoma Based on Tumor and Peritumor Dual-Region Radiomics Signatures

Fang Hu[1,2†], Yuhan Zhang[1†], Man Li[3], Chen Liu[1], Handan Zhang[1], Xiaoming Li[1], Sanyuan Liu[3], Xiaofei Hu[1] and Jian Wang[1*]*

[1] Department of Radiology, Southwest Hospital, Third Military Medical University (Army Medical University), Chongqing, China, [2] Department of Radiology, Tongliang District People's Hospital, Chongqing, China, [3] Department of Research and Development, Shanghai United Imaging Intelligence Co., Ltd., Shanghai, China

Objective: To predict preoperative microvascular invasion (MVI) risk grade by analyzing the radiomics signatures of tumors and peritumors on enhanced magnetic resonance imaging (MRI) images of hepatocellular carcinoma (HCC).

Methods: A total of 501 HCC patients (training cohort n = 402, testing cohort n = 99) who underwent preoperative Gd-EOB-DTPA-enhanced MRI and curative liver resection within a month were studied retrospectively. Radiomics signatures were selected using the least absolute shrinkage and selection operator (Lasso) algorithm. Unimodal radiomics models based on tumors and peritumors (10mm or 20mm) were established using the Logistic algorithm, using plain T1WI, arterial phase (AP), portal venous phase (PVP), and hepatobiliary phase (HBP) images. Multimodal radiomics models based on different regions of interest (ROIs) were established using a combinatorial modeling approach. Moreover, we merged radiomics signatures and clinico-radiological features to build unimodal and multimodal clinical radiomics models.

Results: In the testing cohort, the AUC of the dual-region (tumor & peritumor 20 mm) radiomics model and single-region (tumor) radiomics model were 0.741 vs 0.694, 0.733 vs 0.725, 0.667 vs 0.710, and 0.559 vs 0.677, respectively, according to AP, PVP, T1WI, and HBP images. The AUC of the final clinical radiomics model based on tumor and peritumoral 20mm incorporating radiomics features in AP&PVP&T1WI images for predicting MVI classification in the training and testing cohorts were 0.962 and 0.852, respectively.

Conclusion: The radiomics signatures of the dual regions for tumor and peritumor on AP and PVP images are of significance to predict MVI.

Keywords: hepatocellular carcinoma, microvascular invasion, radiomics, MRI, peritumor

OPEN ACCESS

Edited by:
Li Xiao,
University of Science and Technology of China, China

Reviewed by:
Shi-Ting Feng,
The First Affiliated Hospital of Sun Yat-sen University, China
Xuntao Yin,
Guangzhou Medical University, China

****Correspondence:***
Xiaofei Hu
harryzonetmmu@163.com
Jian Wang
wangjian@aifmri.com

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

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *12 January 2022*
Accepted: *23 February 2022*
Published: *22 March 2022*

Citation:
Hu F, Zhang YH, Li M, Liu C, Zhang HD, Li XM, Liu SY, Hu XF and Wang J (2022) Preoperative Prediction of Microvascular Invasion Risk Grades in Hepatocellular Carcinoma Based on Tumor and Peritumor Dual-Region Radiomics Signatures. Front. Oncol. 12:853336. doi: 10.3389/fonc.2022.853336

1 INTRODUCTION

Hepatocellular carcinoma (HCC) has a high recurrence rate, with a five-year recurrence rate of 70% and 35% after liver resection and liver transplantation, respectively (1). Several findings suggest that microvascular invasion (MVI) is essential in the prognosis of HCC patients (2–4). Microvascular invasion is the formation of nested clusters of cancer cells in the lumen of endothelium-covered vessels on a microscopic scale (5), and it can only be detected using pathological diagnostics. In recent years, MVI has attracted increasing attention from clinicians, and the more severe the degree of microvascular invasion, the earlier the recurrence and the shorter the overall survival time of patients (6, 7). Zhao et al. (8) identified significant differences in the prognosis of HCC patients with different MVI risk grades. Furthermore, the cumulative five-year postoperative survival and tumor-free survival rates in the high-risk MVI group were only 25.4% and 15.8%, respectively, significantly worse than the low-risk MVI and no-MVI groups. Predicting preoperative MVI risk grading could help clinicians in providing personalized treatments to patients with high-risk HCC. Furthermore, several studies have illustrated that among HCC patients presenting with MVI, the anatomical liver resection group has a higher recurrence-free survival rate than the non-anatomical liver resection group (9, 10). This clearly demonstrates the importance of preoperative MVI prediction for improved prognosis of HCC patients.

Several scholars have attempted to predict MVI using hematologic indicators, such as serum alpha-fetoprotein (AFP) (11) or imaging features such as peritumoral hypointensity in the hepatobiliary phase (12, 13), arterial peritumoral enhancement (14), and nonsmooth tumor margins (12, 14) to find a reliable and non-invasive method for preoperative diagnosis of MVI. Although the results showed a correlation between AFP or these imaging features and MVI, the criteria for determining the AFP threshold value have not been identified. These imaging features lacked objectivity and were greatly influenced by the knowledge base, diagnostic experience, and work status of radiologists.

In recent years, with the advent of radiomics technology, some scholars have been extracting signatures from CT or MRI images that are difficult to perceive with human eyes, and building models to preoperatively predict negative or positive hepatocellular carcinoma MVI using automatic algorithms. Their findings demonstrate that radiomics signatures on radiological images are promising for preoperative prediction of MVI (15–28). Furthermore, attempts have been made to predict MVI preoperatively using tumor and peritumor radiomics signatures. Microvascular invasion is mostly common in small portal vein branches inside paracancerous liver tissue (29). However, whether peritumor signatures are valuable in predicting MVI is controversial.

Feng et al. (15) used Gd-EOB -DTPA-enhanced MRI radiomics signatures to predict MVI and found that peritumor signatures are important in MVI prediction. They also realized that the dual-region (tumor and peritumor; 10 mm) radiomics model was superior to the tumor radiomics model. In contrast, Xu et al. (24) found that the dual-region (tumor and peritumor; 5 mm) radiomics model did not highlight any advantages in predicting MVI compared to the tumor-based radiomics model. This discrepancy may be due to inconsistent imaging methods and peritumor extent. Therefore, to explore the impact of dual-region radiomics signatures of the tumor and peritumor on MVI prediction, this study aimed to develop enhanced MRI radiomics models with different ROIs (including tumor, tumor & peritumor 10 mm, and tumor & peritumor 20 mm) for preoperative prediction of MVI risk grades.

2 METHODS

2.1 Patient Data Collection and Follow-Up

2.1.1 Inclusion and Exclusion Criteria

We retrospectively analyzed a total of 501 HCC patients who met the inclusive criteria from June 2017 to July 2020. All patients were randomly divided two cohorts (4:1): a training cohort (n = 402) and a testing cohort (n = 99). The inclusion criteria included: (i) Gd-EOB-DTPA-enhanced MRI within one month before surgery; (ii) pathologically confirmed HCC; and (iii) curative surgical resection or liver transplantation. The exclusion criteria included: (i) history of recurrent HCC or HCC combined with other primary tumors; (ii) poor image quality; (iii) MRI showed large vessel cancer thrombus; and (iv) history of preoperative anti-cancer treatment. This study was approved by the ethical review committee of the First Affiliated Hospital of the Army Medical University. Patients were exempted from providing informed consent.

2.1.2 Image Acquisition

Pre-scan preparation required the patients to fast and abstain from food and drinks for over six hours. Breathing training, which involved breath-holding in a calm state, was provided. A 3.0T MRI (magnetom trio, siemens healthcare, erlangen, Germany), 12-channel phased-array body coil, and high-pressure injector were used for image acquisition. The positioning image, in-phase and opposed-phase T1-weighted imaging (T1WI), and dynamic three-dimensional volumetric interpolated breath-hold examination (3D-VIBE) flat scan were obtained before MRI enhancement. Post-contrast dynamic 3D-VIBE was performed at the arterial phase (30 s), portal venous phase (70 s), transitional phase(3 min) and hepatobiliary phase (15min) after a rapid bolus injection of contrast agent (Primovist; Bayer Schering Pharma, Berlin, Germany) with a rate of 1 mL/s, followed by a 20 mL saline flush. T2-weighted images were obtained with a technique of half-Fourier acquisition single-shot fast spin-echo sequence. Diffusion-weighted imaging (DWI) adopts a breathing- triggered technique at b values of 0, 50, 400, and 800 s/mm^2, and the apparent diffusion coefficient (ADC) was calculated using a single exponential function with b values of 0 and 800 s/mm^2. Susceptibility weighted imaging (SWI) adopts high-resolution, 3D gradient echo and 3D fully flow-compensated sequence for scanning.

2.1.3 Clinical and Imaging Data

Information including gender, age, cirrhosis, hepatitis B surface antigen (HBS Ag), platelet count (PLT), serum albumin (ALB), alanine transarninase (ALT), aspertate aminotransferase (AST), alkaline phosphatase (ALP), serum total bilirubin (TBIL), serum α-fetoprotein (AFP), activated partial thromboplastin time (APTT), prothrombin time (PT), international normalized ratio (INR) and other relevant details were collected from the electronic medical record system and laboratory tests. Patients were evaluated and classified according to MRI manifestations and laboratory tests using liver function grading criteria (Child-Pugh).

All MRI image features were analyzed jointly by two radiologists with three and four years of diagnostic abdominal imaging experience, respectively. Then the results were reviewed by two senior doctors and cross-reviewed. If there is any dispute, the final decision will be made after discussion by two doctors. If there is any dispute, the final decision will be made after discussion by two doctors. During the image analysis, the four aforementioned radiologists did not refer to clinical laboratory tests, other imaging tests, or postoperative pathological diagnoses. The assessment of imaging features included the number of tumors, the maximum length of the tumor, satellite noduels, tumor morphology, tumor envelope integrity, intra-tumor hemorrhage, intra-tumor fat, arterial peritumor enhancement, and hepatobiliary peritumor hypointensty. The maximum length of the tumor is measured in the coronal, sagittal, or axial plane. The capsule is defined as the portal venous phase or delayed phase, with annular high enhancement around the lesion (30). The state of the capsule is divided into two types: intact, incomplete or absent. Satellite noduels mainly refers to the small tumor focus with a diameter ≤ 2cm within the range of the main tumor ≤ 2cm (5). The shape of the tumor was evaluated as round or irregular. Intratumoral hemorrhage defined as low signal intensity in SWI phase, intratumoral fat defined as high signal intensity in the in- phase and low signal intensity in the opposed-phase. Peritumoral enhancement in arterial phase defined as obvious crescent or patchy enhancement in arterial phase, but consistent with hepatic parenchyma enhancement in portal venous phase (31). Peritumoral hypointensity in the hepatobiliary phase is defined as patchy abnormal signal shadow around hepatobiliary tumor, and the signal intensity is lower than that of normal liver parenchyma (32).

2.1.4 Evaluation of Pathological MVI

Two pathologists assessed the MVI status of all HCC cases by examining the hematoxylin-eosin (HE) stained sections under a microscope. The Guidelines for the Standardized Pathological Diagnosis of Primary Liver Cancer (2015 edition) were used to grade MVI risk. The three risk levels of MVI included M0: no MVI detected; M1 (low-risk group): 0<the number of MVI ≤ 5 and MVI occurred in the proximal paracancerous liver(<1cm); M2 (high-risk group): the number of MVI > 5 MVI or MVI occurred in the distal paracancerous liver tissue area (> 1cm) (29).

2.1.5 Follow-Up Visits

The endpoint of this study, which was December 31, 2020, was considered the date of recurrence. The time from the first postoperative day to tumor recurrence or termination for follow-up observation was referred to as recurrence-free survival (RFS). Postoperative recurrence was mainly detected using CT, MRI, ultrasonography and other imaging examinations, combining with laboratory examinations, such as serum AFP at the same time. The time of recurrence was recorded after the diagnosis of recurrence.

2.2 Enhanced MRI Radiomics Analysis

Radiomics analyses were performed at uAl-Research-Portal (Shanghai United Imaging Intelligence Co., Ltd), a clinical research platform written in the Python programming language (version 3.7.3, https://www.python.org). The widely used software package Py Radiomics (https://pyradiomics.readthedocs.io/en/latest/html) is embedded in this platform. Enhanced MRI radiomics analysis included annotation of tumor lesions and peritumor extension, extraction and selection of radiomics signatures, and model building (**Figure 1**).

2.2.1 Labeling of Tumor Lesions and Peritumor Extension

Two physicians with three and four years of diagnostic abdominal MRI experience, respectively, selected plain T1WI, arterial phase (AP), portal venous phase (PVP), and hepatobiliary phase (HBP) sequences to label tumor lesions on a 3D slicer. The annotation results of all tumor lesions were validated by two radiologists with nine years of experience in diagnostic abdominal imaging. The annotation of 304 MRI liver data was completed for T1WI, resulting in a preliminary version of the automated liver segmentation model for T1WI. The T1WI liver segmentation was also extended to AP, PVP, and HBP images using the alignment and fine-tuning. Furthermore, the original tumor lesions were extended by 10 mm and 20 mm, respectively, in the uAl-Research-Portal (Shanghai United Imaging Intelligent Medical Technology Co., Ltd.). The extension beyond the boundary of the liver was adjusted by combining the results of the liver segmentation model (**Figure 2**).

2.2.2 Extraction and Selection of Radiomics Signatures

The images were imported into uAl-Research-Portal (Shanghai United Imaging Intelligent Medical Technology Co., Ltd.) and preprocessed by resampling all voxels of images to $1 \times 1 \times 1$ mm^3 using the 3D nearest neighbor interpolation method. Then, 2,600 radiomics signatures were extracted within the lesion annotation range, respectively from T1WI of different ROI (tumor edge, tumor and peritumor 10 mm, and tumor and peritumor 20 mm), AP, PVP, and HBP modalities. In the training cohort, signatures were selected using the least absolute shrinkage and selection operator (Lasso) algorithm. The results of multiple ROI signatures selection with the same modality were merged to eliminate the possible influence of distinct radiomics types of signatures after signatures selection. We selected 85, 185, 178, and 62 radiomics signatures as the most critical for MVI risk grading on T1WI, AP, PVP, and HBP, respectively.

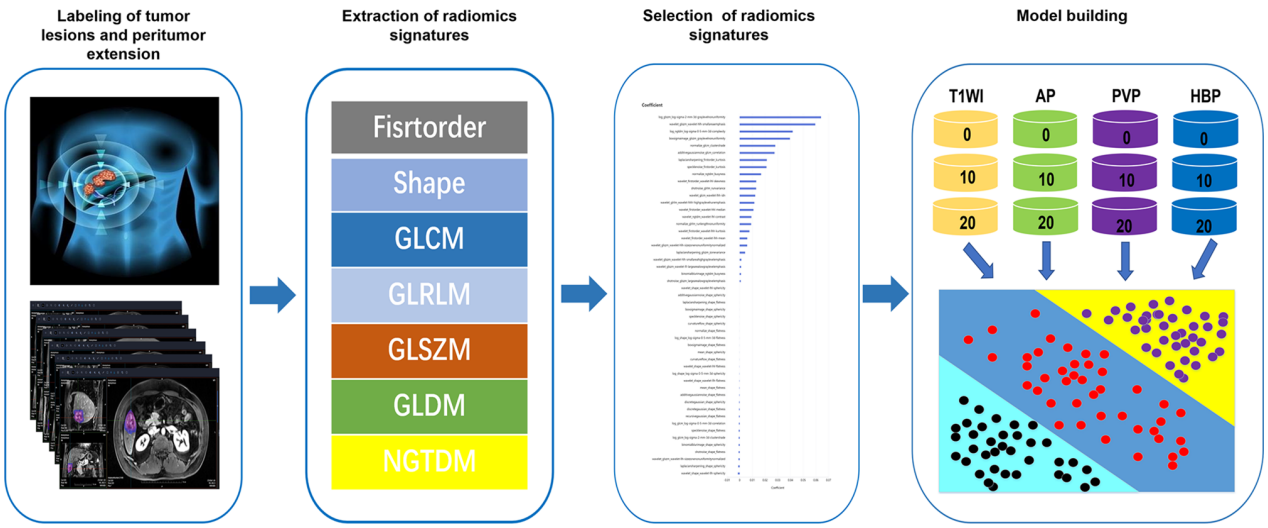


FIGURE 1 | Workflow of radiomics analysis. The radiomics workflow started with labeling of tumor lesions and peritumor extension in MR images. After that, radiomic features including firstorder, shape, gray-level co-occurrence matrix (GLCM), gray-Level run-length matrix (GLRLM), gray-level size zone matrix (GLSZM), gray level difference method (GLDM) and neighborhood gray-tone difference matrix (NGTDM) were extracted within the tumor and peritumor dual-region. Next, least absolute shrinkage and selection operator (LASSO) were used for the radiomic feature selection. Finally, unimodal radiomics models and multimodal radiomics models based on different regions of interest (ROIs) in the tumor and peritumor was developed.

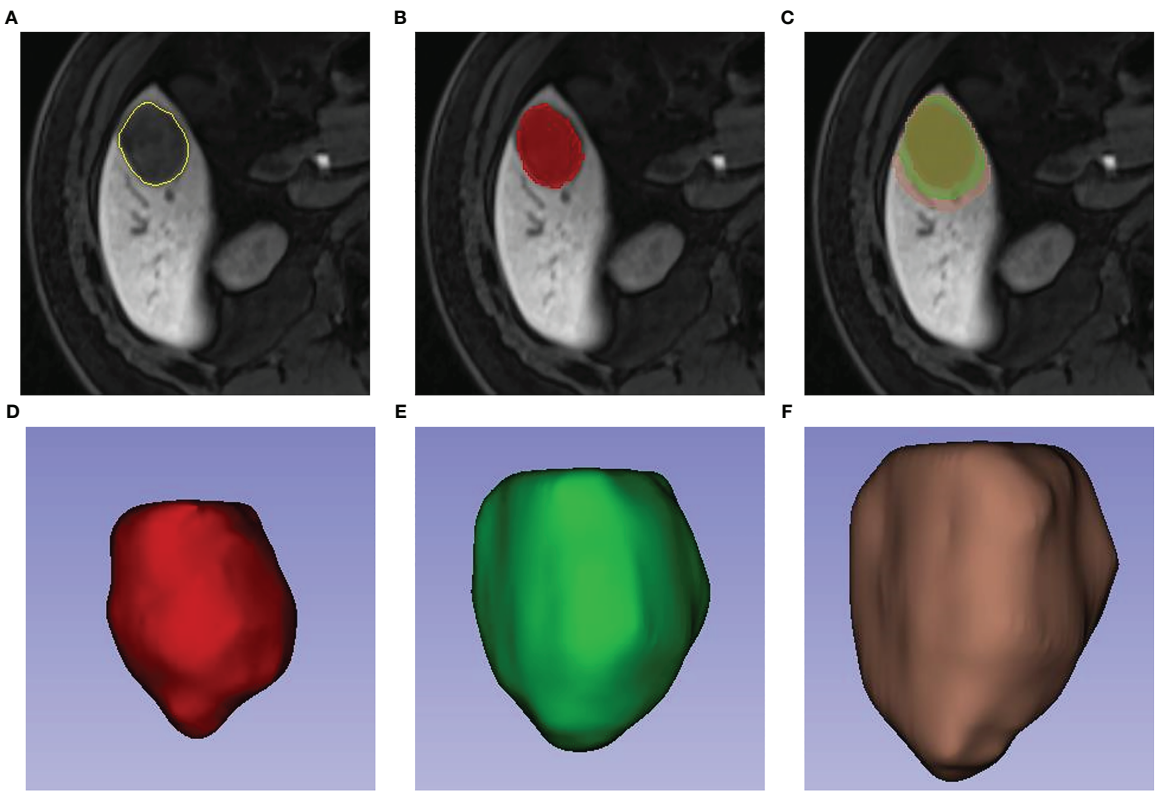


FIGURE 2 | Labeling of tumor lesions and peritumor extension. First, radiologists manually draw the volume(VOI) of the tumor **(A, B)**. On the bases of VOI entire of the tumor, a region with 10-mm and 20-mm distance to tumor surface were automatically reconstructed **(C)**. 3-dimensional view of the VOI entire, VOI (tumor & peritumor 10 mm) and VOI (tumor & peritumor 20 mm) respectively were showed in Picture **(D–F)**.

2.3 Model Building

2.3.1 Building Unimodal Radiomics Models Based on Different ROIs in the Tumor and Peritumor

Box-Cox transformation was performed on radiomics signatures after selecting T1WI, AP, PVP, and HBP modal features. According to above phases, Logistic algorithm was used to build unimodal radiomics models based on tumor, tumor and peritumor 10 mm, and tumor and peritumor 20 mm [T1WI (0\10\20), AP (0\10\20), PVP (0\10\20), HBP (0\10\20)] in the training cohort. In addition, their ability to predict MVI was tested in the testing cohort.

2.3.2 Building Multimodal Radiomics Models Based on Different ROIs in the Tumor and Peritumor

Tumor-based unimodal radiomics models were combined to create an unexpanded multimodal peritumor radiomics model. The prediction probabilities of each modality corresponding to the MVI category were summed to determine the final prediction sequence. The best tumor-based radiomics model for ROI was selected based on the AUC. The combination modeling method described above was then used to create multimodal radiomics models with different ROIs for tumor and peritumor (10 mm and 20 mm) accordingly.

2.3.3 Building Clinical Radiomic Model

The radiomic model with the best predictive performance for MVI risk grading was selected based on the AUC. We combined the essential clinical and radiological features selected using the Lasso algorithm with the corresponding unimodal radiomics signatures in the best radiomics model to build a unimodal clinical radiomics model. We also established a multimodal clinical radiomics model using the above-combined modeling approach.

2.4 Statistical Analysis

Statistical analysis was performed using R software. The rank sum test, one-way ANOVA, and chi-square test were used to analyze statistical differences between clinical indicators and radiological signals in M0, M1, and M2 groups. The ROC curves of different models were plotted, while AUC values were calculated using PyCharm software. Survival curves were plotted using Kaplan-Meier and tested using the two-sided log-rank test. A two-tailed p value less than 0.05 was considered statistically significant.

3 RESULTS

3.1 Clinical and Imaging Features of Patients

A total of 501 patients met the inclusive and exclusive criteria, with 252 (50.3%) patients were pathologically diagnosed as MVI negative and 249 (49.70%) patients were pathologically identified as MVI positive: 207 (41.32%) in group M1 and 42 (8.38%) in group M2. The three groups were statistically different in INR, AFP, Child-Pugh, number of nodes, shape, arterial peritumoral enhancement, peritumoral hypointensity in the hepatobiliary phase, tumor diameter, intratumoral hemorrhage, satellite foci, and envelope ($p < 0.05$), but not in the remaining clinical and radiological indices (**Table 1**). The differences of clinicoradiological characteristics in between training and testing datasets are listed in **Table 2**. A total of 24 clinical and radiological signatures were selected using the Lasso algorithm. Fifteen essential clinical and radiological features, including serum AFP level, Child-Pugh, cirrhosis, age, PT, PLT, shape, peritumoral hypointensity in the hepatobiliary phase, intratumoral hemorrhage, satellite foci, diameter, number of nodes, arterial peritumoral enhancement, envelope, and tumor diameter, were finally selected (**Figure 3**).

3.2 Building Models for Predicting MVI Risk Grading

3.2.1 Developing Unimodal Radiomics Models Based on Different ROIs in the Tumor and Peritumor

The prediction results of unimodal radiomics models based on different ROIs in the tumor and peritumor are shown in **Table 3**. According to T1WI and HBP images, the radiomics model with ROI based on tumor [T1WI (0), HBP (0)] had the best prediction results. The AUC and ACC in the testing cohort were 0.710 and 0.677 and 0.566 and 0.535, respectively. However, the radiomics model with ROI based on tumor and peritumor 20mm on AP and PVP images [AP (20),PVP (20)] performed better in predicting MVI risk grades in the testing cohort in the unimodal radiomics models using different ROIs, with AUC of 0.741 and 0.733 and ACC of 0.556 and 0.586, respectively.

3.2.2 Building Multimodal Radiomics Models Based on Different ROIs in the Tumor and Peritumor

The prediction results of multimodal radiomics models using different ROIs of the tumor and peritumor are presented in **Table 4**. The fusion radiomics model [T1WI (0) & PVP (0) & AP (0)] performed the best in the ROI-based multimodal radiomics model with AUC and ACC values of 0.758 and 0.616, respectively, in the testing cohort. In the corresponding dual-region radiomics model created by combined modeling, the ROI's tumor and peritumor (20 mm) based on multimodal radiomics model [T1WI (20) & PVP (20) & AP (20)] performed better in predicting MVI risk grading with AUC and ACC values of 0.778 and 0.636, respectively, in the testing cohort.

3.2.3 Comparison of Clinical Radiomics Models and Optimal Radiomics Models

The clinical radiomics model [T1WI (20) & AP (20) & PVP (20)] was more effective than the corresponding multimodal radiomics predictive model in the testing cohort (AUCs: 0.852 vs 0.778; ACCs: 0.747 vs. 0.636) (**Figure 4**).

3.3 Survival Analysis

As of December 31, 2020, 501 patients had completed tumor recurrence-free follow-up. The overall recurrence rate was 24.35% (122/501). The median RFS was 38 months for patients

TABLE 1 | Comparisons of clinicoradiological characteristics in different microvascular invasion grades.

Variable	M 0n =252	M1n =207	M2n =42	P-value
Age (years)	52.81 ± 10.56	50.99 ± 11.22	51.12 ± 11.37	0.183
Gender				0.344
male	210 (83.33%)	180 (86.96%)	38 (90.48%)	
female	42 (16.67%)	27 (13.04%)	4 (9.52%)	
PLT, 10^9/L	150.25 ± 74.36	157.60 ± 71.89	158.79 ± 91.23	0.528
ALT, IU/L	62.15 ± 90.40	63.92 ± 97.84	80.62 ± 110.65	0.507
AST, IU/L	56.57 ± 77.51	63.22 ± 92.43	65.45 ± 47.28	0.625
ALB, g/L	41.80 ± 5.04	41.97 ± 5.02	41.69 ± 8.20	0.922
TBIL, μmol	18.39 ± 20.98	18.08 ± 9.57	18.33 ± 6.67	0.979
ALP, IU/L	113.45 ± 96.19	109.72 ± 67.12	123.58 ± 86.38	0.615
APTT, sec	25.91 ± 19.23	28.09 ± 2.82	28.37 ± 7.00	0.202
PT, sec	13.81 ± 12.57	11.91 ± 5.17	12.15 ± 1.18	0.092
INR	1.01 ± 0.11	1.00 ± 0.08	1.05 ± 0.11	0.02
AFP				0.005
0, normal	112 (45.90%)	70 (34.48%)	10 (23.81%)	
1, abnormal	132 (54.10%)	133 (65.52%)	32 (76.19%)	
Hbs Ag				0.315
0, Hbs Ag(-)	38 (15.14%)	21 (10.34%)	5 (12.20%)	
1, Hbs Ag(+)	213 (84.86%)	182 (89.66%)	36 (87.80%)	
Child-Pugh				0.045
0, A	248 (98.41%)	203 (98.54%)	39 (92.86%)	
1, B	4 (1.59%)	3 (1.46%)	3 (7.14%)	
Cirrhosis				0.977
0, no	68 (26.98%)	56 (27.32%)	12 (28.57%)	
1, yes	184 (73.02%)	149 (72.68%)	30 (71.43%)	
No. of nodes				0.021
0, 1	242 (96.03%)	190 (91.79%)	36 (85.71%)	
1, ≥ 2	10 (3.97%)	17 (8.21%)	6 (14.29%)	
Shape				<0.001
0, circle	151 (59.92%)	52 (25.12%)	5 (11.90%)	
1, irregular	101 (40.08%)	155 (74.88%)	37 (88.10%)	
Arterial peritumoral enhancement				<0.001
0, absent	211 (83.73%)	136 (65.70%)	23 (54.76%)	
1, present	41 (16.27%)	71 (34.30%)	19 (45.24%)	
Peritumoral hypotensity on HBP				<0.001
0, absent	219 (86.90%)	145 (70.05%)	16 (38.10%)	
1, present	33 (13.10%)	62 (29.95%)	26 (61.90%)	
The maximum length				<0.001
0, ≤5cm	200 (79.37%)	125 (60.39%)	16 (38.10%)	
1, >5cm	52 (20.63%)	82 (39.61%)	26 (61.90%)	
Intratumoral hemorrhage				<0.001
0, absent	181 (71.83%)	113 (54.59%)	12 (28.57%)	
1, present	71 (28.17%)	94 (45.41%)	30 (71.43%)	
Intratumoral fat				0.204
0, absent	214 (84.92%)	180 (86.96%)	32 (76.19%)	
1, present	38 (15.08%)	27 (13.04%)	10 (23.81%)	
Satellite nodules				<0.001
0, absent	241 (96.02%)	186 (89.86%)	34 (80.95%)	
1, present	10 (3.98%)	21 (10.14%)	8 (19.05%)	
Capsule				<0.001
0, absence or incomplete	71 (28.17%)	136 (65.70%)	31 (73.81%)	
1, complete	181 (71.83%)	71 (34.30%)	11 (26.19%)	

Unless otherwise noted, data are shown as number of patients, with the percentage in parentheses. MVI, microvascular invasion.M 0= no MVI; M1=≤ 5 MVI, and occurred in the adjacent liver tissue area (≤ 1 cm); M2= > 5 MVI, or MVI occurred in the distant paracancerous liver tissue area (> 1cm).

PLT, platelet count; ALT, alanine transaminase; AST, aspertate aminotransferase; ALB, serum albumin; TBIL, serum total bilirubin; ALP, Alkaline phosphatase; APTT, activated partial thromboplastin time; PT, prothrombin time; INR, international normalized ratio; AFP, serum a-fetoprotein; Hbs Ag, hepatitis B surface antigen; HBP, hepatobiliary phase.

in the M0 group, 29 months in the M1 group, and nine months in the M2 group (log-rank test, $p < 0.001$). Similar results were observed in the prediction model: median RFS was 37 months for patients in M0, 27 months in M1, and eight months in M2 (log-rank test, $p < 0.001$) (**Figure 5**).

4 DISCUSSION

In this study, we analyzed the radiomics signatures of enhanced MRI for tumor and peritumor to determine the impact of dual-region radiomics signatures of tumor and peritumor on different

TABLE 2 | The differences of clinicoradiological characteristics in between training and testing datasets.

Group	taining	testing	P-value
N	402	99	
Age (years)	51.66 ± 10.89	52.93 ± 11.03	0.301
PLT, 10^9/L	156.62 ± 77.43	143.45 ± 62.58	0.117
ALT, IU/L	59.16 ± 87.60	85.86 ± 119.67	0.012
AST, IU/L	55.04 ± 73.86	80.45 ± 107.22	0.006
ALB, g/L	41.96 ± 5.11	41.48 ± 6.26	0.422
TBIL, μmol	17.42 ± 8.33	21.64 ± 32.26	0.020
ALP, IU/L	112.84 ± 91.99	112.46 ± 41.64	0.969
APTT, sec	27.61 ± 14.69	24.60 ± 10.11	0.054
PT, sec	13.17 ± 10.65	11.75 ± 1.17	0.186
INR	1.00 ± 0.09	1.04 ± 0.14	<0.001
Gender			0.146
0, male	348 (86.57%)	80 (80.81%)	
1, female	54 (13.43%)	19 (19.19%)	
AFP			0.761
0, normal	153 (38.93%)	39 (40.62%)	
1, abnormal	240 (61.07%)	57 (59.38%)	
Hbs Ag			0.912
0, Hbs Ag(-)	51 (12.85%)	13 (13.27%)	
1, Hbs Ag(+)	346 (87.15%)	85 (86.73%)	
Child-Pugh			0.414
0, A	394 (98.25%)	96 (96.97%)	
1, B	7 (1.75%)	3 (3.03%)	
Cirrhosis			0.611
0, no	107 (26.75%)	29 (29.29%)	
1,yes	293 (73.25%)	70 (70.71%)	
No. of nodes			0.254
0, 1	373 (92.79%)	95 (95.96%)	
1, ≥ 2	29 (7.21%)	4 (4.04%)	
The maximum length			0.739
0,≤5cm	275 (68.41%)	66 (66.67%)	
1, >5cm	127 (31.59%)	33 (33.33%)	
Shape			0.179
0, circle	161 (40.05%)	47 (47.47%)	
1, irregular	241 (59.95%)	52 (52.53%)	
Satellite nodules			0.048
0, absent	365 (91.02%)	96 (96.97%)	
1, present	36 (8.98%)	3 (3.03%)	
Capsule			0.496
0, absence or incomplete	194 (48.26%)	44 (44.44%)	
1, complete	208 (51.74%)	55 (55.56%)	
Intratumoral hemorrhage			0.137
0, absent	252 (62.69%)	54 (54.55%)	
1, present	150 (37.31%)	45 (45.45%)	
Intratumoral fat			0.796
0, absent	341 (84.83%)	85 (85.86%)	
1, present	61 (15.17%)	14 (14.14%)	
Arterial peritumoral enhancement			0.294
0, absent	301 (74.88%)	69 (69.70%)	
1, present	101 (25.12%)	30 (30.30%)	
Peritumoral hypotensity on HBP			0.034
0, absent	313 (77.86%)	67 (67.68%)	
1, present	89 (22.14%)	32 (32.32%)	
MVI grade			0.993
0	202 (50.25%)	50 (50.51%)	
1	166 (41.29%)	41 (41.41%)	
2	34 (8.46%)	8 (8.08%)	

MVI, microvascular invasion; M0: no MVI detected; M1 (low-risk group): 0<the number of MVI ≤ 5 and MVI occurred in the proximal paracancerous liver; M2 (high-risk group): the number of MVI > 5 MVI or MVI occurred in the distal paracancerous liver tissue area (> 1 cm). PLT, platelet count; ALT, alanine transarninase; AST, aspertate aminotransferase; ALB, serum albumin; TBIL, serum total bilirubin; ALP, alkaline phosphatase; APTT, activated partial thromboplastin time; PT, prothrombin time; INR, international normalized ratio; AFP, serum afetoprotein; Hbs Ag, hepatitis B surface antigen; HBP, hepatobiliary phase

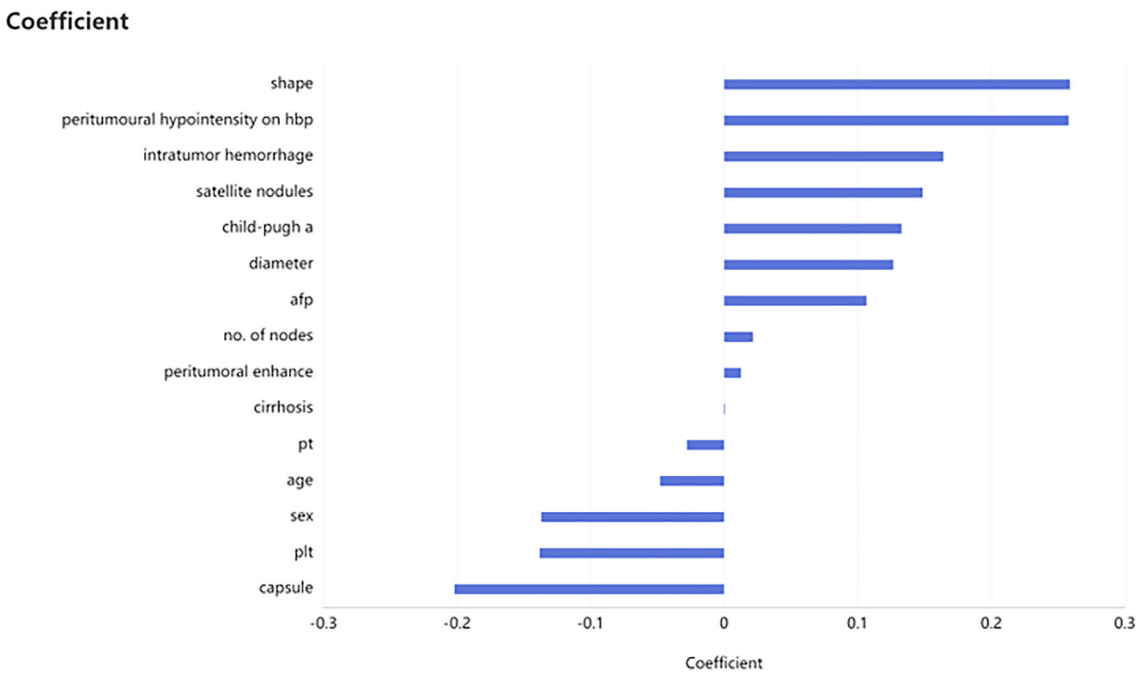


FIGURE 3 | Selection of clinicoradiological characteristics. Fifteen necessary clinicoradiological features were finally selected by using least absolute shrinkage and selection operator (LASSO), including serum AFP level, Child-Pugh, cirrhosis, age et al.

TABLE 3 | Unimodal radiomics models based on different ROIs.

Modality	ROI	Training group		Testing group	
		AUC	ACC	AUC	ACC
T1WI	Tumor	0.796	0.644	0.710	0.566
	Tumor & Margin (10)	0.803	0.632	0.604	0.535
	Tumor & Margin (20)	0.780	0.617	0.667	0.566
AP	Tumor	0.893	0.726	0.694	0.505
	Tumor & Margin (10)	0.909	0.776	0.718	0.596
	Tumor & Margin (20)	0.927	0.808	0.741	0.556
PVP	Tumor	0.907	0.741	0.725	0.545
	Tumor & Margin (10)	0.911	0.766	0.726	0.586
	Tumor & Margin (20)	0.921	0.769	0.733	0.586
HBP	Tumor	0.795	0.657	0.677	0.535
	Tumor & Margin (10)	0.781	0.639	0.636	0.505
	Tumor & Margin (20)	0.777	0.617	0.559	0.424

TABLE 4 | Multimodal radiomics models based on different ROIs.

Modality	ROI	Training group		Testing group	
		AUC	ACC	AUC	ACC
T1WI+AP+PVP	Tumor	0.939	0.806	0.758	0.616
	Tumor & Margin (10)	0.947	0.818	0.743	0.606
	Tumor & Margin (20)	0.953	0.838	0.778	0.636

phases of predicting MVI risk grading. We also explored the efficacy of a combined model for clinical factors, imaging features, and radiomics signatures in predicting MVI risk grading. Our results indicated that the impact of dual-region radiomics signatures for tumor and peritumor on predicting MVI risk grades varied at different phases. The established clinical multimodal radiomics model predicted MVI grades with 83.3% and 74.7% accuracy, in the training and testing cohorts, respectively.

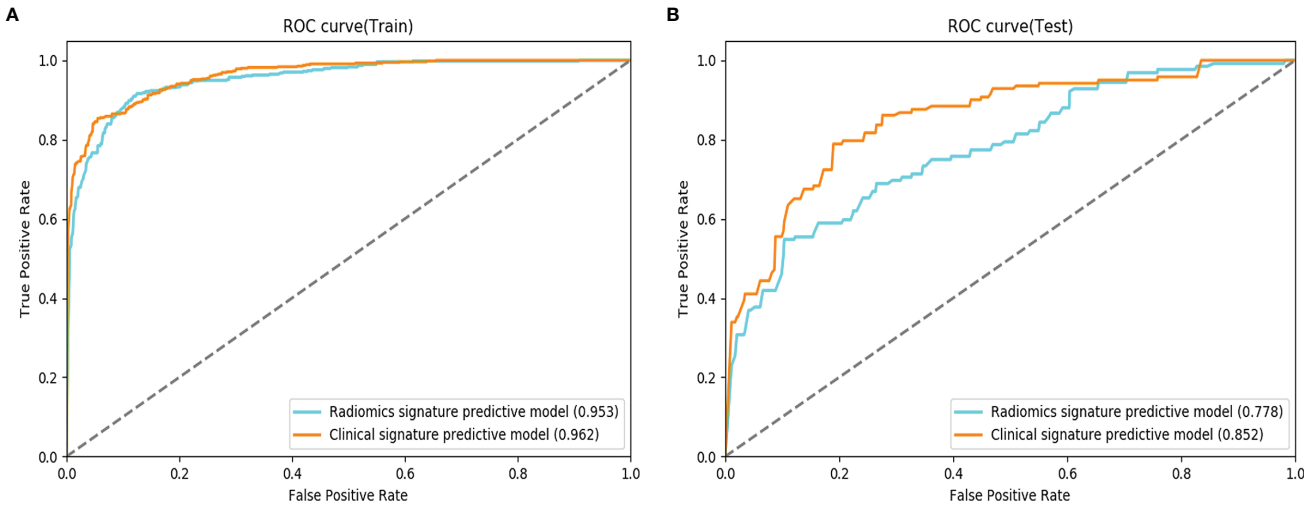


FIGURE 4 | Comparison of receiver operating characteristic (ROC) curves for the prediction of microvascular invasion. ROC curves of the radiomics signature predictive model and the clinical signature predictive model, which combines the fusion radiomics signature and clinicoradiological factors in the training **(A)** and testing **(B)** datasets.

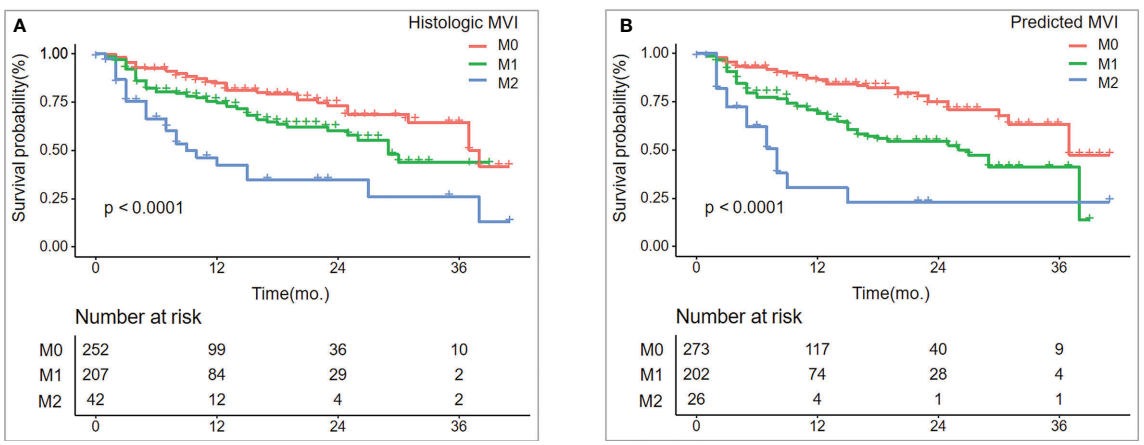


FIGURE 5 | Recurrence-free survival (RFS) between histologic MVI and predicted MVI.RFS **(A, B)** curves were scaled by histologic MVI status and final model-predicted MVI risk grades with Kaplan Meier analysis.

The radiomics model that used tumor and peritumor (10 mm or 20 mm) was superior to the radiomics model that was based only on the tumor in predicting MVI risk grades on arterial and portal images. Nebbia et al. (20) found the similar results. The dual-region radiomics signatures of the tumor and peritumor in the late arterial and portal phases were more beneficial in predicting MVI than single-region radiomics signatures. The occurrence of MVI is a complex biological process involving many factors. According to the related studies of Zhou (33) and Wan (34), the activation of epithelial-mesenchymal transformation (EMT) transcription may be an important pathogenic mechanism of MVI in HCC. When EMT transcription is activated, intercellular adhesion proteins such as E-Cadherin are down-regulated and EMT markers such as N-Cadherin and vimentin are increased, which may induce HCC dedifferentiation and increase tumor invasiveness, which leads to the occurrence of MVI (35). MVI is commonly found in the portal vein branches of the liver tissue adjacent to the tumor (29), which is related to the fact that portal vein is the main outflow vessel of liver cancer. When the tumor embolus invades the tiny portal vein, it will cause small branch occlusion and reduce the blood flow of the portal vein around the tumor, giving rise to compensatory peri-tumor hyperperfusion (30, 36). Peritumoral hemodynamic changes can lead to different imaging findings, such as abnormal peritumoral enhancement in arterial phase and low signal intensity in hepatobiliary phase, which are risk factors

for MVI. These shows that the peritumoral area of HCC plays an important role in the diagnosis of MVI. This may explain why the combination of tumor and peritumoral radiomics features on AP or PVP images are more conducive to the prediction of MVI risk grade. Feng et al. illustrated that the ROI based on hepatobiliary-phase radiomics model (tumor and peritumor; 10 mm) outperformed the radiomics model based only on the tumor in predicting negative or positive MVI in HCC (15). However, in this study, the ability of the dual-region radiomics model (ROI based on the tumor and the peritumor; 10 mm and 20 mm) to predict MVI risk grades was not better than that of the single region radiomics model based on tumor only. This is because the rate of Gd-EOB-DTPA uptake by hepatocytes is correlated with liver function. Thus, patients with impaired liver function should be reasonably delayed in hepatobiliary-phase scans (37).

This study also found that fusing multimodal radiomics signatures improved the ability of the radiomics model to predict MVI risk grades. Ma et al. (19) made similar conclusions using a multimodal radiomics model, which combined radiomics signatures from enhanced CT arterial, portal venous, and delayed phases. The multimodal radiomics model exhibited better performance than the corresponding unimodal radiomics model in predicting the presence or absence of MVI. This indicated that the inclusion of radiomics signatures from different modalities could improve MVI predictors for various aspects of the tumor. However, the predictive efficiency of the model did not always improve with the incorporation of more radiomics signatures of different modalities. Therefore, this study demonstrated that the number of modalities for radiomics analysis and the diagnostic performance of models are not positively correlated. According to Zhang et al., the AUC values for their radiomics model, which incorporated AP, PVP, and DP in the training and validation cohorts, were 0.784 and 0.82, respectively. However, the radiomics model fusing six sequences of T1WI, T2WI, DWI, AP, PVP, and DP had AUC values of 0.778 and 0.803 in the training and validation cohorts, respectively (26). Merging radiomics signatures for numerous models may exclude radiomics signatures that respond to different traits of the tumor and are meaningful for MVI prediction due to low correlation in the signature screening. Therefore, the performance of the model decreased. Accordingly, further research into the optimal modalities combination for imaging radiomics analysis in MVI prediction is recommended.

The importance of combining different aspects such as laboratory tests, imaging features, and radiomics signatures in predicting MVI was proven in this study. The model that incorporated clinical and radiological signatures such as serum AFP level, cirrhosis, PT, shape, hepatobiliary-phase peritumor hypointensty, and intra-tumoral hemorrhage had higher AUC values than the radiomics model, which corroborates the results of Xu's research. Xu et al. combined AST, AFP, tumor margins, growth pattern, envelope, peritumor enhancement, and radiomics score to create a nomogram with AUC values of 0.909 and 0.889 in the training and validation cohorts, respectively. In contrast, the radiomics model had AUC values of 0.841 and 0.819 in the training and validation cohorts, respectively (24). Yang et al. also found similar results (25). This suggests that combining the three aspects can improve the ability of models to predict MVI.

The final prediction model established in this study could provide effective prognostic stratification for HCC patients. Tanaka et al. found significant differences in the prognosis of HCC patients with different degrees of MVI progression (38). Therefore, preoperative prediction of MVI grades could more accurately assess the severity of MVI and prognosis of HCC patients than the prediction of presence or absence of MVI, providing clinicians with more beneficial information. In this study, the recurrence-free survival time predicted by the clinical multimodal radiomics model was significantly different among patients with different MVI gradings ($p < 0.001$). Moreover, HCC patients in the M2 group had a significantly shorter recurrence-free survival time than those in the M1 and M0 groups. This finding suggested that our model could assist clinicians in assessing the prognosis of HCC patients preoperatively and providing more personalized treatments.

5 LIMITATIONS

Although, our clinical radiomics model can be used as a preoperative predictor of MVI risk grading, it has the following limitations. First, all of the MRI images were generated by the same machine at the same hospital. Although this may reduce certain confounding effects, external validation, which could have generated more data from multiple centers, was missing. Future research should include multi-center data to perform independent external validation to confirm the predictive validity of the model. Second, a lack of consistency in the clarity of tumor boundaries on different simultaneous images may have resulted in less accurate tumor ROI segmentation. Therefore, the MRI scanning technique should be updated to obtain clearer tumor contours and establish an automatic segmentation model for HCC to reduce the segmentation discrepancy of ROI. Lastly, the unbalanced data volume between M0, M1, and M2 groups may have affected the predictive performance of the model. Thus, a more considerable amount of data is required to balance differences between the various groups.

6 CONCLUSION

In summary, the radiomics signatures of the dual regions for tumor and peritumor on different phases have diverse effects on the prediction of MVI risk grades. The radiomics signatures of the dual regions for tumor and peritumor on AP and PVP images are of merit to predict MVI. Our final preoperative prediction model can assist clinicians in the preoperative diagnosis of HCC for MVI risk grading and prognostic assessment.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by author JW at wangjian@aifmri.com, with reasonable request.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The ethical review committee of the First Affiliated Hospital of the Army Medical University. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

All authors made substantial contributions to the conception and design of the study, acquisition of data, analysis and interpretation of data, drafting the article, and revising it critically for important intellectual content. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the National Key Research and Development Program of China (No. 2016YFC0107101).

REFERENCES

1. De Angelis R, Sant M, Coleman MP, Francisci S, Baili P, Pierannunzio D, et al. Cancer Survival in Europe 1999-2007 by Country and Age: Results of EUROCA- RE-5-a Population-Based Study. *Lancet Oncol* (2014) 15:23–34. doi: 10.1016/s1470-2045(13)70546-1
2. Lim K-C, Chow PK-H, Allen JC, Chia G-S, Lim M, Cheow P-C, et al. Microvascular Invasion Is a Better Predictor of Tumor Recurrence and Overall Survival Following Surgical Resection for Hepatocellular Carcinoma Compared to the Milan Criteria. *Ann Surg* (2011) 254:108–13. doi: 10.1097/SLA.0b013e31821ad884
3. Mazzaferro V, Llovet JM, Miceli R, Bhoori S, Schiavo M, Mariani L, et al. Predicting Survival After Liver Transplantation in Patients With Hepatocellular Carcinoma Beyond the Milan Criteria: A Retrospective, Exploratory Analysis. *Lancet Oncol* (2009) 10:35–43. doi: 10.1056/nejm199603143341104
4. Rodriguez-Peralvarez M, Luong TV, Andreana L, Meyer T, Dhillon AP, Burroughs AK. A Systematic Review of Microvascular Invasion in Hepatocellular Carcinoma: Diagnostic and Prognostic Variability. *Ann Surg Oncol* (2013) 20:325–39. doi: 10.1245/s10434-012-2513-1
5. Roayaie S, Blume IN, Thung SN, Guido M, Fiel M-I, Hiotis S, et al. A System of Classifying Microvascular Invasion to Predict Outcome After Resection in Patients With Hepatocellular Carcinoma. *Gastroenterology* (2009) 137:850–5. doi: 10.1053/j.gastro.2009.06.003
6. Feng LH, Dong H, Lau WY, Yu H, Zhu YY, Zhao Y, et al. Novel Microvascular Invasion-Based Prognostic Nomograms to Predict Survival Outcomes in Patients After R0 Resection for Hepatocellular Carcinoma. *J Cancer Res Clin Oncol* (2017) 143:293–303. doi: 10.1007/s00432-016-2286-1
7. Sumie S, Nakashima O, Okuda K, Kuromatsu R, Kawaguchi A, Nakano M, et al. The Significance of Classifying Microvascular Invasion in Patients With Hepatocellular Carcinoma. *Ann Surg Oncol* (2014) 21:1002–9. doi: 10.1245/s10434-013-3376-9
8. Zhao H, Chen J, Yan X, Fu X, Sun S, Qiu Y. Prognostic Value of the Risk Classification of Microvascular Invasion in Patients With Hepatocellular Carcinoma. *Chin J Hepatobiliary Surg* (2019) 25(6):401–5. doi: 10.18632/oncotarget.12547
9. Cucchetti A, Qiao GL, Cescon M, Li J, Xia Y, Ercolani G, et al. Anatomic Versus Nonanatomic Resection in Cirrhotic Patients With Early Hepatocellular Carcinoma. *Surgery* (2014) 155:512–21. doi: 10.1016/j.surg.2013.10.009
10. Zhao H, Chen C, Gu S, Yan X, Jia W, Mao L, et al. Anatomical Versus Non-Anatomical Resection for Solitary Hepatocellular Carcinoma Without Macroscopic Vascular Invasion: A Propensity Score Matching Analysis. *J Gastroenterol Hepatol* (2017) 32:870–8. doi: 10.18632/oncotarget.12547
11. McHugh PP, Gilbert J, Vera S, Koch A, Ranjan D, Gedaly R. Alpha-Fetoprotein and Tumour Size Are Associated With Microvascular Invasion in Explanted Livers of Patients Undergoing Transplantation With Hepatocellular Carcinoma. *HPB (Oxf)* (2010) 12:56–61. doi: 10.1111/j.1477-2574.2009.00128.x
12. Lee S, Kim SH, Lee JE, Sinn DH, Park CK. 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
13. Kim S, Shin J, Kim D-Y, Choi GH, Kim M-J, Choi J-Y. Radiomics on Gadoxetic Acid-Enhanced Magnetic Resonance Imaging for Prediction of Postoperative Early and Late Recurrence of Single Hepatocellular Carcinoma. *Clin Cancer Res* (2019) 25:3847–55. doi: 10.1158/1078-0432.CCR-18-2861
14. Huang M, Liao B, Xu P, Cai H, Huang K, Dong Z, et al. Prediction of Microvascular Invasion in Hepatocellular Carcinoma: Preoperative Gd-EOB-DTPA-Dynamic Enhanced MRI and Histopathological Correlation. *Contrast Media Mol Imaging* (2018) 2018:1–9. doi: 10.1155/2018/9674565
15. 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 Radiol* (2019) 29:4648–59. doi: 10.1007/s00330-018-5935-8
16. Ji GW, Zhu FP, Xu Q, Wang K, Wu MY, Tang WW, 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
17. Jiang YQ, Cao SE, Cao S, Chen JN, Wang GY, Shi WQ, et al. Preoperative Identification of Microvascular Invasion in Hepatocellular Carcinoma by XGBoost and Deep Learning. *J Cancer Res Clin Oncol* (2020) 147(3):821–33. doi: 10.1007/s00432-020-03366-9
18. Lahan-Martins D, Perales SR, Gallani SK, da Costa LBE, Lago EAD, Boin I, et al. Microvascular Invasion in Hepatocellular Carcinoma: Is It Predictable With Quantitative Computed Tomography Parameters? *Radiol Bras* (2019) 52:287–92. doi: 10.1590/0100-3984.2018.0123
19. 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 Radiol* (2019) 29:3595–605. doi: 10.1007/s00330-018-5985-y
20. Nebbia G, Zhang Q, Arefan D, Zhao X, Wu S. Pre-Operative Microvascular Invasion Prediction Using Multi-Parametric Liver MRI Radiomics. *J Digit Imaging* (2020) 33(6):1376–86. doi: 10.1007/s10278-020-00353-x
21. Ni M, Zhou X, Lv Q, Li Z, Gao Y, Tan Y, et al. Radiomics Models for Diagnosing Microvascular Invasion in Hepatocellular Carcinoma: Which Model Is the Best Model? *Cancer Imaging* (2019) 19:60. doi: 10.1186/s40644-019-0249-x
22. Peng J, Zhang J, Zhang Q, Xu Y, Zhou J, Liu L. A Radiomics Nomogram for Preoperative Prediction of Microvascular Invasion Risk in Hepatitis B Virus-Related Hepatocellular Carcinoma. *Diagn Interv Radiol* (2018) 24:121–7. doi: 10.5152/dir.2018.17467
23. Wilson GC, Cannella R, Fiorentini G, Shen C, Borhani A, Furlan A, et al. Texture Analysis on Preoperative Contrast-Enhanced Magnetic Resonance Imaging Identifies Microvascular Invasion in Hepatocellular Carcinoma. *HPB (Oxf)* (2020) 22(11):1622–30. doi: 10.1016/j.hpb.2020.03.001
24. Xu X, Zhang HL, Liu QP, Sun SW, Zhang J, Zhu FP, 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
25. Yang L, Gu D, Wei J, Yang C, Rao S, Wang W, et al. A Radiomics Nomogram for Preoperative Prediction of Microvascular Invasion in Hepatocellular Carcinama. *Liver Cancer* (2019) 8:373–86. doi: 10.1159/000494099
26. Zhang R, Xu L, Wen X, Zhang J, Yang P, Zhang L, et al. A Nomogram Based on Bi-Regional Radiomics Features From Multimodal Magnetic Resonance Imaging for Preoperative Prediction of Microvascular Invasion in Hepatocellular Carcinoma. *Quant Imaging Med Surg* (2019) 9:1503–15. doi: 10.21037/qims.2019.09.07
27. Zhang X, Ruan S, Xiao W, Shao J, Tian W, Liu W, et al. Contrast-Enhanced CT Radiomics for Preoperative Evaluation of Microvascular Invasion in Hepatocellular Carcinoma: A Two-Center Study. *Clin Transl Med* (2020) 10 (2):e111. doi: 10.1002/ctm2.111
28. Zhu YJ, Feng B, Wang S, Wang LM, Wu JF, Ma XH, et al. Model-Based Three-Dimensional Texture Analysis of Contrast-Enhanced Magnetic Resonance Imaging as A Potential Tool for Preoperative Prediction of Microvascular Invasion in Hepatoce-Llular Carcinoma. *Oncol Lett* (2019) 18:720–32. doi: 10.3892/ol.2019.10378
29. Chinese Society of Liver Cancer (CSLC). Evidence-Based Practice Guidelines for Standardized Pathological Diagnosis of Primary Liver Cancer in China: 2015. *Zhonghua Gan Zang Bing Za Zhi* (2015) 23:321–7. doi: 10.13315/j.cnki.cjcep.2015.03.001
30. Choi JY, Lee JM, Sirlin CB. CT and MR Imaging Diagnosis and Staging of Hepatocellular Carcinoma: Part II. Extracellular Agents, Hepatobiliary Agents,and Ancillary Imaging Features. *Radiology* (2014) 273(1):30–50. doi: 10.1148/radiol.14132362
31. Kim H, Park MS, Choi JY, Park YN, Kim MJ, Kim KS, et al. Can Microvesselinvasion of Hepatocellular Carcinoma be Predicted by Pre-Operative MRI? *Eur Radiol* (2009) 19:1744–51. doi: 10.1007/s00330-009-1331-8
32. Kim KA, Kim MJ, Jeon HM, Kim KS, Choi JS, Ahn SH, et al. Prediction of Microvascular Invasion of Hepatocellular Carcinoma: Usefulness of Peritu-Moral Hypointensity Seen on Gadoxetate Disodium-Enhanced Hepatobiliary Phase Images. *J Magn Reson Imaging* (2012) 35:629–34. doi: 10.1002/jmri.22876
33. Zhou YM, Cao L, Li B, Zhang RX, Sui CJ, Yin ZF, et al. Clinicopathological Significance of ZEB1 Protein in Patients With Hepatocellular Carcinoma. *Ann Surg Oncol* (2012) 19(5):1700–6. doi: 10.1245/s10434-011-1772-6
34. Wan T, Zhang TW, Si XY, Zhou YM. Overexpression of EMT-Inducing Transcription Factors as a Potential Poor Prognostic Factor for Hepatocellular Carcinoma in Asian Populations: A Meta-Analysis. *Oncotarget* (2017) 8 (35):59500–8. doi: 10.18632/oncotarget.18352
35. Erstad Derek J, Tanabe Kenneth K. Prognostic and Therapeutic Implications of Microvascular Invasion in Hepatocellular Carcinoma. *Ann Surg Oncol* (2019) 26(5):1474–93. doi: 10.1245/s10434-019-07227-9
36. Matsui O, Kobayashi S, Sanada J, Kouda W, Ryu Y, Kozaka K, et al. Hepatocelluar Nodules in Liver Cirrhosis: Hemodynamic Evaluation (Angiography-Assisted CT) With Special Reference to Multi-Step Hepatocarcinogenesis. *Abdom Imaging* (2011) 36(3):264–72. doi: 10.1007/s00261-011-9685-1
37. Masahiro Okada M, Takamichi Murakami M, Ryohei Kuwatsuru M, Yuko Nakamura M, Hiroyoshi I, Satoshi G, et al. Biochemical and Clinical Predictive Approach and Time Point Analysis of Hepatobiliary Phase Liver Enhancement on Gd- EOB-DTPA–enhanced Mr Images: A Multicenter Study. *Radiology* (2016) 281(2):1–10. doi: 10.1148/radiol.2016151061
38. Tanaka S, Mogushi K, Yasen M, Noguchi N, Kudo A, Nakamura N, et al. Gene- Expression Phenotypes for Vascular Invasiveness of Hepatocellular Carcinomas. *Surgery* (2010) 147:405–14. doi: 10.1016/j.surg.2009.09.037

Conflict of Interest: Author ML and Author SL were employed by Shanghai United Imaging Intelligence Co., Ltd.

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

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

frontiers in Oncology

ORIGINAL RESEARCH
published: 31 March 2022
doi: 10.3389/fonc.2022.843376

Development and Validation of a Novel Radiomics-Based Nomogram With Machine Learning to Preoperatively Predict Histologic Grade in Pancreatic Neuroendocrine Tumors

Xing Wang[1†], Jia-Jun Qiu[2†], Chun-Lu Tan[1], Yong-Hua Chen[1], Qing-Quan Tan[1], Shu-Jie Ren[1], Fan Yang[3], Wen-Qing Yao[4], Dan Cao[5], Neng-Wen Ke[1] and Xu-Bao Liu[1*]*

[1] Department of Pancreatic Surgery, West China Hospital, Sichuan University, Chengdu, China, [2] Department of West China Biomedical Big Data Center, West China Hospital, Sichuan University, Chengdu, China, [3] Department of Radiology, West China Hospital, Sichuan University, Chengdu, China, [4] Department of Pathology, West China Hospital, Sichuan University, Chengdu, China, [5] Department of Oncology, West China Hospital, Sichuan University, Chengdu, China

Backgroud: Tumor grade is the determinant of the biological aggressiveness of pancreatic neuroendocrine tumors (PNETs) and the best current tool to help establish individualized therapeutic strategies. A noninvasive way to accurately predict the histology grade of PNETs preoperatively is urgently needed and extremely limited.

Methods: The models training and the construction of the radiomic signature were carried out separately in three-phase (plain, arterial, and venous) CT. Mann–Whitney *U* test and least absolute shrinkage and selection operator (LASSO) were applied for feature preselection and radiomic signature construction. SVM-linear models were trained by incorporating the radiomic signature with clinical characteristics. An optimal model was then chosen to build a nomogram.

Results: A total of 139 PNETs (including 83 in the training set and 56 in the independent validation set) were included in the present study. We build a model based on an eight-feature radiomic signature (group 1) to stratify PNET patients into grades 1 and 2/3 groups with an AUC of 0.911 (95% confidence intervals (CI), 0.908–0.914) and 0.837 (95% CI, 0.827–0.847) in the training and validation cohorts, respectively. The nomogram combining the radiomic signature of plain-phase CT with T stage and dilated main pancreatic duct (MPD)/bile duct (BD) (group 2) showed the best performance (training set: AUC = 0.919, 95% CI = 0.916–0.922; validation set: AUC = 0.875, 95% CI = 0.867–0.883).

Conclusions: Our developed nomogram that integrates radiomic signature with clinical characteristics could be useful in predicting grades 1 and 2/3 PNETs preoperatively with powerful capability.

Keywords: pancreas, pancreatic neuroendocrine tumor, tumor grade, radiomics, CT

OPEN ACCESS

Edited by:
Min Tang,
Jiangsu University, China

Reviewed by:
Yifeng Tao,
Carnegie Mellon University, United States
Sarbesh Pandeya,
Harvard Medical School, United States
Weinan Zhou,
University of Illinois at Urbana-Champaign, United States

***Correspondence:**
Xu-Bao Liu
xbliu@medmail.com.cn
Neng-Wen Ke
kenengwen@scu.edu.cn

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

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *25 December 2021*
Accepted: *01 March 2022*
Published: *31 March 2022*

Citation:
Wang X, Qiu J-J, Tan C-L, Chen Y-H, Tan Q-Q, Ren S-J, Yang F, Yao W-Q, Cao D, Ke N-W and Liu X-B (2022) Development and Validation of a Novel Radiomics-Based Nomogram With Machine Learning to Preoperatively Predict Histologic Grade in Pancreatic Neuroendocrine Tumors. Front. Oncol. 12:843376. doi: 10.3389/fonc.2022.843376

INTRODUCTION

Pancreatic neuroendocrine tumor (PNET) is a relatively rare pancreatic disorder thought to arise in hormone secretory cells of the islets of Langerhans (1) and ever known as islet cell tumor (2). It consists of about 3%–5% of all the pancreatic neoplasm but predominates human neuroendocrine tumors (3). Additionally, the incidence and prevalence of PNETs are steadily increasing in recent decades (4, 5).

PNETs are characterized by tumor heterogeneity (6), and of which the clinical behavior are relatively indolent but vary dramatically (7). Tumor grade is the crucial determinant of the biological aggressiveness of PNETs. Additionally, it is suggested to be associated with lymph node involvement (7), tumor recurrence (8), and overall prognosis (9). According to the 2010 World Health Organization (WHO) classification criteria (10), tumor grade is defined numerically by the proliferative indicator Ki-67, in which low-grade (grade 1 (G1)) tumors have a Ki-67 index from 0% to 2%, intermediate-grade (G2) tumors have a Ki-67 index from 3% to 20%, and high-grade (G3) tumors have a Ki-67 index greater than 20%. Surgery is thought to be the cornerstone of treatment of PNETs in each stage and the only potential way to cure local PNETs (5, 11). However, different surgical strategies could be applied for PNETs of grades 1 and 2/3. The last but not the least, for advanced PNETs, there are also other treatment options, e.g., somatostatin analog (SSA), targeted therapy, or chemotherapy, based on tumor grades. In short, the WHO tumor grading is the best current tool to predict prognosis, guide therapy selection, and aid surgical decision-making by stratification of PNETs.

Of note, tumor grade is always obtained according to postoperative pathology specimen. Although the preoperative endoscopic ultrasound (EUS)-guided fine-needle aspiration cytology (FNA) is proved to be efficient in diagnosing PNETs, the accuracy in differentiating tumor grade remains challenging, possibly due to limited tissue availability or missing the most mitotically active areas (hot pot) of the tumor. The research of Heidsma et al. showed that tumor grade differentiation could be accurately determined by FNA in only 20%–50% of patients (6, 12). Additionally, EUS-guided fine-needle biopsy (FNB) with thicker tissue biopsy needle was reported to have better performance in tumor grade differentiation, as more tumor tissues could be obtained (13). However, both of them were invasive procedures which largely depended on the operators' experience (13). Therefore, the effective method of preoperatively predicting the pathologic grade of PNETs is still imperatively needed to help establish individualized therapeutic strategies and aid surgical decision-making.

Several previous studies tried to identify the tumor grade of PNETs by computed tomography (CT), magnetic resonance imaging (MRI), and PET/CT (14–17). Although they provided a noninvasive way to preoperatively predict the aggressiveness of PNETs, the accuracy was limited, as the prediction of the frequently occurring heterogeneous tumor was mainly established based on visual observation rather than quantitative information. Recently, "radiomics" brings a new hope for this problem. It is a method that automatically extract a large number of quantitative features from medical images using data-characterization algorithms, and subsequently identify the most significant radiomic signatures through machine learning methods (18, 19). Therefore, we can realize cancer detection, prediction of clinical outcome, and treatment evaluation as reported previously (20, 21). Additionally, radiomics was reported to be successfully applied in differentiating pathologic grading in patients with clear cell renal cell carcinoma (22), colorectal adenocarcinoma (23), etc. Nevertheless, to the best of our knowledge, a noninvasive optimal combined model to incorporate imaging features with clinical characteristics (such as tumor size and tumor margin status) to predict the pathologic grade of PNETs is extremely limited.

Thus, this work attempted to establish a multimodal artificial intelligence (AI) model that integrates a radiomic signature based on plain CT images with clinical features for noninvasive and preoperative prediction of the pathologic grades of PNETs.

MATERIALS AND METHODS

Patients

This retrospective study was approved by the ethics committee of Sichuan University, and the signed informed consent was waived. From July 2008 to June 2018, patients with histologically confirmed PNETs who underwent surgical resection in our institution were retrospectively reviewed. The patients with a PNET that was too small to display clearly on CT, several patients with cystic PNET, and patients without preoperative CT scan were excluded at the present study. The final diagnosis of PNETs was made by specialized pathologists, including the diagnosis of the tumor grade basing on Ki-67 immunohistochemical staining data. Clinical data were obtained from the electronic medical records or external medical reports, including demographic characteristics and classification. Finally, 139 patients with complete data available were identified for analysis in the present study. Of these, 83 patients were taken randomly as the training set, and the other 56 patients were used for the independent validation set (also called test set, not the set in a crossvalidation approach). The training dataset and validation dataset had an even distribution in patient characteristics (**Table 1**). No significant difference was found in PNET pathologic grade and clinical characteristics (age, maximum diameter, and clinical stage of the tumor, etc.) between the training dataset and validation dataset.

CT Image Acquisition

All patients underwent an abdominal contrast-enhanced CT scan preoperatively. Contrast-enhanced CT scan was performed on three CT scanners including a 16-slice CT (Toshiba Medical Systems, Japan), a 64-, and a 256-slice CT (Philips Healthcare, Netherlands). CT scans used the same CT scanning parameters: tube voltage of 120 kVp, tube current of 125 to 300 mAs, pitch of 0.6 to 1.25 mm, slice thickness of 3 to 5 mm, and reconstruction interval of 3 to 5 mm.

TABLE 1 | Comparison of patient and lesion features between grades 1 and 2/3 groups in training and validation sets.

Features	Training set (*n* = 83)		*p*-value	Validation set (*n* = 56)		*p*-value
	Grade 2/3 (*n* = 55, %)	Grade 1 (*n* = 28, %)		Grade 2/3 (*n* = 37, %)	Grade 1 (*n* = 19, %)	
Age(range, years)[a]	49.7 (20–77)	49.2 (24–70)	0.883	52 (22–77)	53.2 (16–75)	0.799
Gender			0.141			
Women	27 (49.1)	9 (32.1)		10 (27)	5 (26.3)	0.955
Men	28 (50.9)	19 (67.9)		27 (73)	14 (73.7)	
Tumor size (range, mm)[a]	40.7 (12–150)	28.4 (10–80)	**0.028**	47.7 (12–180)	22.2 (12–42)	**<0.001**
T stage (T3–T4)	28 (50.9)	5 (17.9)	**0.004**	23 (62.2)	Nil	**<0.001**
Clinical TNM stage (IIB and above)	31 (56.4)	5 (17.9)	**0.001**	25 (67.6)	1 (5.3)	**<0.001**
Dilated MPD/BD[b]	19 (34.5)	3 (10.7)	**0.02**	14 (37.8)	Nil	**0.006**
Tumor margin			**0.013**			**0.034**
Well defined	30 (54.5)	23 (82.1)		24 (64.9)	18 (94.7)	
Poorly defined	25 (45.5)	5 (17.9)		13 (35.1)	1 (5.3)	
Tumor location			0.502			0.096
Head and neck	29 (52.7)	14 (50)		20 (54.1)	6 (31.6)	
Body and tail	26 (47.3)	13 (46.4)		14 (37.8)	13 (68.4)	
Multiple	Nil	1 (3.6)		3 (8.1)	Nil	
Pathology			0.175			0.080
Functional	12 (21.8)	10 (35.7)		9 (24.3)	9 (47.4)	
Nonfunctional	43 (78.2)	18 (64.3)		28 (75.7)	10 (52.6)	
Insulinoma			**0.047**			**0.006**
Yes	5 (9.1)	8 (28.6)		4 (10.8)	9 (47.4)	
No	50 (90.9)	20 (71.4)		33 (89.2)	10 (52.6)	

[a]*The values indicated are expressed as median (range).*

[b]*Dilated MPD/BD, dialated main pancreatic duct (MPD) or bile duct (BD). The clinical TNM stage and T stage of the tumor was determined preoperatively according to the American Joint Committee on Cancer TNM Staging System Manual, 7th edition.*

The bold values in this table are p-value less than 0.05, which means the features between grade 1 and 2/3 groups are significantly different.

Radiomic Analysis

We performed a radiomic analysis on preoperative CT images to evaluate the pathologic grades of PNETs. **Figure 1** illustrates the work flow of the radiomic analysis.

Step 1: Tumor regions were delineated and segmented into regions of interest (ROIs) from which texture features were extracted. We evaluated the CT images in plain, arterial, and portal venous phases, respectively.

Step 2: We used 10 texture analysis methods to extract features. The **Supplementary Material** described the methods in detail. A total of 1,133 features were extracted from a ROI (24–26).

Steps 3–4: Preselection was performed on the 1,133 features using the Mann–Whitney *U* test (*p*-value ≤ 0.25). We then combined the methods of least absolute shrinkage and selection operator (LASSO) and stepwise logistical regression to perform feature selection. Feature preselection and feature selection were both performed on the training set. A radiomic signature can be built based on the final selected features.

Step 5: We combined the radiomic signature and four clinical data to train SVM-linear models. Features were divided into three groups: radiomic signature (group 1), radiomic signature combining T stage and dilated main pancreatic duct (MPD)/bile duct (BD) (group 2), and radiomic signature combining T stage, dilated MPD/BD, clinical TNM stage, and tumor margin (group 3). The models training and the construction of the radiomic signature were carried out separately in three phases (plain, arterial, and venous). Thus, a total of 9 prediction tasks were performed.

Step 6: The independent validation dataset (*n* = 56) was tested on the 9 trained models. We chose an optimal model to construct a nomogram, and then used the nomogram to predict the pathologic grades of these 56 patients. A calibration curve and a goodness of fit to the ideal model are calculated to evaluate the nomogram.

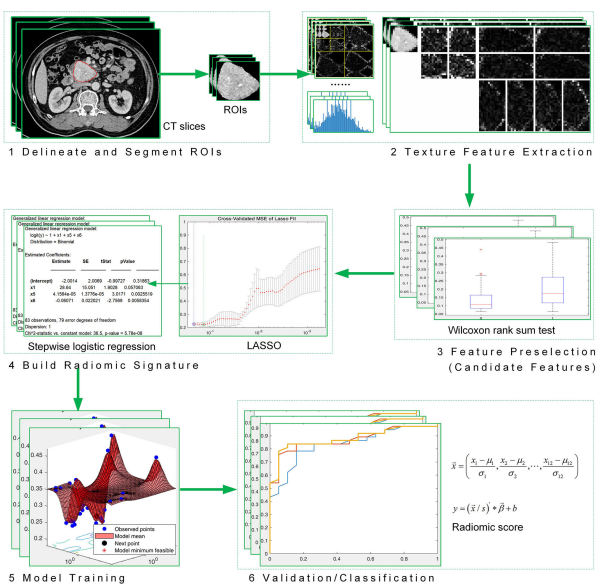


FIGURE 1 | Work flow of radiomic analysis.

RESULTS

The detailed distribution of clinical characteristics in the G1 group (grade 1) and G2/3 group (grade 2/3) is summarized in **Table 1**. The tumor size of PNET in grade 2/3 group was significantly larger than that in grade 1 group ($p = 0.028$). T stage (T3–T4), clinical TNM stage (IIB and above), Dilated MPD/BD, and poorly defined tumor margin were more frequently detected in patients with grade 2/3 PNETs than those with grade 1 ($p = 0.004$, $p = 0.001$, $p = 0.02$, and $p = 0.013$, respectively). The consistent results occurred both in the training and validation datasets.

As illustrated in **Figure 1**, this study aims to build a radiomic> signature and evaluate the ability of the signature to predict PNET grades. **Table 2** shows the features used to build the radiomic signature, that is, the result of feature selection on the training set. We also evaluated the performance of combining the radiomic signature and 4 clinical variables to predict PNET grades. The clinical variables are x_9 to x_{12} in **Table 2**.

The linear combination of x_1 to x_8 in **Table 2** expresses the radiomic signature y. The linear combination is shown in Equations (1) and (2).

$$y = \left(\overrightarrow{x}/s\right) * \overrightarrow{\beta} + b$$

$$\overrightarrow{x} = \left(\frac{x_1 - \mu_1}{\sigma_1}, \frac{x_2 - \mu_2}{\sigma_2}, \ldots, \frac{x_m - \mu_m}{\sigma_m}\right)$$

where y is the score of group 1 (or 2 or 3) for grade 1, $-y$ is the corresponding score for grade 2/3, $\overrightarrow{x}$ is an observation comprising the m predictors, s is the kernel scale, β is the bias term, and the vector β contains the coefficients that define an orthogonal vector to the hyperplane,, and μ_i and σ_i are the corresponding weighted mean and weighted standard deviation for the ith predictor (used for standardization). When predicting the result, we inputted $[-y, y]$ into the function softmax to obtain the probabilities that the observer belongs to the positive class (grade 2/3) and the negative class (grade 1).

Next, we trained prediction models based on the radiomic signature and the clinical variables to approximately calculate the value of each unknown variable in Equations (1) and (2). We then validated the performance of these models on the independent validation set. The training results and the validation results are shown in **Table 3**. Of note, compared with A (arterial) phase and V (venous) phase, P (plain) phase obtained the best prediction performance for each group in the validation set. What is more, for each phase, we calculated the receiver operating characteristics curves (ROCs) and compared the ROCs of validation using the DeLong's test method. **Figure 2** illustrates the ROC results. It demonstrates that the models based on the radiomic signature combined with clinical data (models based on groups 2 and 3) obtain better prediction results than the models based on the radiomic signature alone (models based on group 1). Additionally, although the indicator values based on group 3 show the highest performance, the model based on the radiomic signature combined with 4 clinical data in group 3 showed no significantly better prediction results in plain phase, compared with that combined with 2 clinical data in group 2 ($p < 0.629$). As can be seen from **Table 3**, the experiments based on plain phase obtained the best prediction performance than other phases. Thus, we also calculated the indicator values of accuracy, sensitivity, and specificity in the experiments of plain phase. **Table 4** and **Supplementary Figure S2** show the prediction results (on the validation set) as the threshold varied.

Above all, the model based on the radiomic signature of plain phase combined with 2 clinical data (T stage and Dilated MPD/BD) in group 2 obtained the best prediction performance. Although the prediction results of the model based on the radiomic signature combined with 4 clinical data in group 3 seemed a little better, there was no significant differences between the groups 3 and 2 models ($p < 0.629$). Consider the balance between the convenience and predictive power of the model, we established a novel nomogram to preoperatively

TABLE 2 | Radiomic signature and clinical data.

No.	Analysis method	Subband	Feature name
x_1	Fistogram		Variance
x_2	Wavelet-COM	H1	Maximal correlation coefficient
x_3	Wavelet-RLM	D1	Short-run low gray-level emphasis
x_4	Wavelet-COM	D2	Sum of squares
x_5	Contourlet-histogram	L2-1	1% percentile
x_6	Contourlet-COM	L2-2	Cluster shade
x_7	Contourlet-histogram	L2-3	99% percentile
x_8	Contourlet-histogram	L1-2	90% percentile
x_9			T stage
x_{10}			Dilated MPD/BD
x_{11}			Clinical TNM stage
x_{12}			Tumor margin

The radiomic signature are composed of x_1 to x_8. The clinical data in group 2 are composed of x_9 and x_{10}. The clinical data in group 3 are composed of x_9 to x_{12}. The number following A, H, V, or D represents the decomposition level. The clinical TNM and T stages of the tumor were determined preoperatively according to the American Joint Committee on Cancer TNM Staging System Manual, 7th edition.

COM, cooccurrence matrix; RLM, run-length matrix; A (in the wavelet transform), approximate; H (in the wavelet transform), horizontal; V (in the wavelet transform), vertical; D (in the wavelet transform), diagonal; Li–j, jth component in the ith decomposition in the contourlet transform; Dilated MPD/BD, dilated main pancreatic duct (MPD) or bile duct (BD).

TABLE 3 | Results of training and validation: plain (P), arterial (A), and venous (V); unless otherwise specified, the contents of parentheses are 95% confidence intervals.

Features	Training set (*n* = 83)	Validation set (*n* = 56)
Group 1		
P	0.911 (0.908–0.914)	**0.837 (0.827–0.847)**
A	0.913 (0.909–0.917)	0.710 (0.695–0.725)
V	0.874 (0.869–0.879)	0.625 (0.609–0.641)
Group 2		
P	0.919 (0.916–0.922)	**0.875 (0.867–0.883)**
A	0.895 (0.892–0.898)	0.783 (0.770–0.796)
V	0.900 (0.894–0.906)	0.742 (0.729–0.755)
Group 3		
P	0.895 (0.891–0.899)	**0.879 (0.869–0.889)**
A	0.892 (0.889–0.895)	0.828 (0.817–0.839)
V	0.902 (0.898–0.906)	0.797 (0.784–0.810)

group 1: radiomic signature; group 2: radiomic signature combining T stage and Dilated MPD/BD; group 3: radiomic signature combining T stage, Dilated MPD/BD, clinical TNM stage, and tumor margin. In the training, we used the fivefold crossvalidation technique to calculate the average AUC, then randomly performed 50 fivefold crossvalidations to calculate the average AUC and the 95% confidence intervals. In the independent validation, the bootstrap method based on sampling with replacement was used to calculate the average AUC and the 95% confidence intervals (based on 100 bootstraps). The sampling with replacement randomly sampled one sample at a time and drawn 56 times. The clinical TNM and T stages of the tumor were determined preoperatively according to the American Joint Committee on Cancer TNM Staging System Manual, 7th edition. Dilated MPD/BD, dilated main pancreatic duct (MPD) or bile duct (BD).

The bold values in this table showed the best performance in each group.

predict histologic grade in PNETs based on the radiomic signature of plain phase combined with 2 clinical data (T stage and Dilated MPD/BD) in group 2 (**Figure 3**).

Figure 3 shows that the nomogram achieves a goodness of fit of 0.868 to the ideal model. Correspondingly, the score for the radiomic signature based on plain phase is

$$y = \frac{x_1-217.2809}{6\times167.0185}\times 1.943 + \frac{x_2-0.9677}{6\times0.0548}\times 3.1606 + \frac{x_3-0.0956}{6\times0.0264}\times (-2.4079) + \frac{x_4-546.9485}{6\times143.0338}\times 2.7323 + \frac{x_5-1.0120}{6\times0.1098}\times 1.0809 + \frac{x_6-6.4072\times10^4}{6\times3.2953\times10^4}\times(-37455) + \frac{x_7-117.4458}{6\times17.8108}\times(-1.9601) + \frac{x_8-38.2048}{6\times17.6032}\times 2.7947 - 0.6808$$

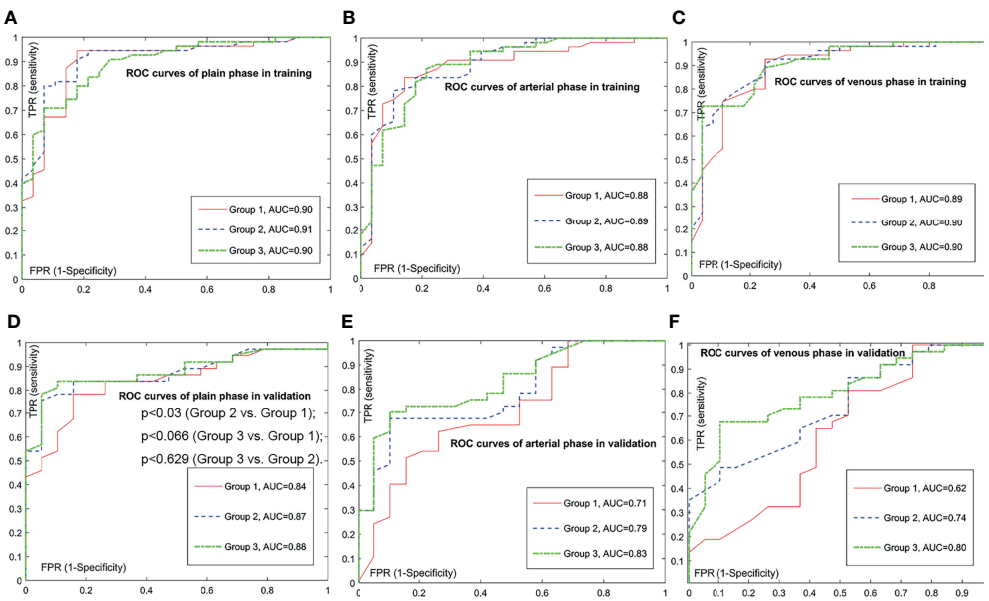


FIGURE 2 | Comparison of receiver operating characteristic (ROC) curves for prediction of the histologic grade. The positive class is grade 2/3; the negative class is grade 1. Subfigure **(A–C)** illustrate the training ROCs. Subfigures **(D–F)** illustrate the validation ROCs. In validations, we performed DeLong's tests to compare two ROC curves. In **(D)**, the DeLong's tests show that the *p*-value between the ROC curve of group 3 and the ROC curve of group 1 is less than 0.066, the *p*-value between the ROC curve of group 3 and the ROC curve of group 2 is less than 0.629, and the p-value between the ROC curve of group 2 and the ROC curve of group 1 is less than 0.030. In subfigure **(E)**, the DeLong's tests show that the p-value between the ROC curve of group 3 and the ROC curve of group 1 is less than 0.003, the p-value between the ROC curve of group 3 and the ROC curve of group 2 is less than 0.037, and the *p*-value between the ROC curve of group 2 and the ROC curve of group 1 is less than 0.013. In **(F)**, the DeLong's tests show that the *p*-value between the ROC curve of group 3 and the ROC curve of group 1 is less than 0.001, the *p*-value between the ROC curve of group 3 and the ROC curve of group *2* is less than 0.059, and the *p*-value between the ROC curve of group 2 and the ROC curve of group 1 is less than 0.003.

TABLE 4 | Validation results based on plain phase as the threshold varied: accuracy (ACC, %), sensitivity (SEN, %), and specificity (SPE, %).

Features	Threshold						
	0.5	**0.55**	**0.6**	**0.65**	**0.7**	**0.75**	**0.8**
Group 1							
ACC	75.0	75.0	76.8	78.6	78.6	**80.4**	76.8
SEN	83.8	83.8	83.8	81.1	78.4	**78.4**	73.0
SPE	57.9	57.9	63.2	73.7	78.9	**84.2**	84.2
Group 2							
ACC	75.0	78.6	82.1	**83.9**	80.4	80.4	82.1
SEN	83.8	83.8	83.8	**83.8**	78.4	78.4	75.7
SPE	57.9	68.4	78.9	**84.2**	84.2	84.2	94.7
Group 3							
ACC	78.6	76.8	80.4	**85.7**	85.7	83.9	80.4
SEN	86.5	83.8	83.8	**83.8**	83.8	81.1	73.0
SPE	63.2	63.2	73.7	**89.5**	89.5	89.5	94.7

group 1: radiomic signature; group 2: radiomic signature combining T stage and Dilated MPD/BD; group 3: radiomic signature combining T stage, Dilated MPD/BD, clinical TNM stage, and tumor margin. The clinical TNM and T stages of the tumor were determined preoperatively according to the American Joint Committee on Cancer TNM Staging System Manual, 7th edition. Dilated MPD/BD, dilated main pancreatic duct (MPD) or bile duct (BD).
The bold values in this table showed the best performance in each group.

The score for the radiomic signature and clinical data in group 2 based on plain-phase is

$$y = \frac{x_1-217.2809}{11.5\times167.0185}\times 3.1548 + \frac{x_2-0.9677}{11.5\times0.0548}\times 3.4514 + \frac{x_3-0.0956}{11.5\times0.0264}\times (-3.5448) + \frac{x_4-546.9485}{11.5\times143.0338}\times 3.6584 + \frac{x_5-1.0120}{11.5\times0.1098}\times 1.55 + \frac{x_6-6.4072\times10^4}{11.5\times3.2953\times10^4}\times(-5.6929) + \frac{x_7-117.4458}{11.5\times17.8108}\times(-2.5886) + \frac{x_8-38.2048}{11.5\times17.6032}\times 4.335 + \frac{x_{10}-2.3133}{11.5\times0.9097}\times(-2.6004) + \frac{x_{11}-0.2590}{11.5\times0.4373}\times(-1.4059) - 0.5775$$

The score for the radiomic signature and clinical data in group 3 based on plain phase is

$$y = \frac{x_1217.2809}{7\times167.0185}\times 1.463 + \frac{x_2-0.9677}{7\times0.0548}\times 1.2912 + \frac{x_3-0.0956}{7\times0.0264}\times(-1.5801) + \frac{x_4-546.9485}{7\times143.0338}\times 1.1559 + \frac{x_5-1.0120}{7\times0.1098}\times 1.1345 + \frac{x_6-6.4072\times10^4}{7\times3.2953\times10^4}\times(-2.3502) + \frac{x_7-117.4458}{7\times17.8108}\times(-0.7985) + \frac{x_8-38.2048}{7\times17.6032}\times 2.084 + \frac{x_9-2.4535}{7\times0.9269}\times(-1.3666) + \frac{x_{10}-2.3133}{7\times0.9097}\times(-1.2045) + \frac{x_{11}-0.2590}{7\times0.4373}\times(-0.5704) + \frac{x_{12}-0.3554}{7\times0.4784}\times(-1.1731) - 0.595$$

DISCUSSION

PNETs are relatively rare neoplasms, the incidence of which is about 4–5 individuals per 100,000 annually (27). Nevertheless, PNETs have been increasingly detected and diagnosed in recent decades and currently represent the second most common pancreatic tumor followed by pancreatic adenocarcinoma (28). Most of PNETs carry MEN1, ATRX, or DAXX gene mutations, while approximately 15% activate mammalian target of rapamycin (mTOR) signaling (29, 30). PNETs are heterogenous neoplasms, of which the prognosis varies widely. The current most important prognostic stratification factor is WHO tumor grade classification, which might optimize tailored therapeutic strategies. So far, tumor grade is obtained by postoperative pathology. The preoperative fine-needle aspiration (FNA) is invasive, and the accuracy in differentiating tumor grade remains challenging. In the present study, we establish a combined nomogram that integrates a radiomic signature based on plain CT images with clinical features for noninvasive and preoperative prediction of pathologic grades of PNETs with high accuracy.

Firstly, we build a model based on an eight-feature radiomic signature to stratify PNET patients into G1 and G2/3 groups with an AUC of 0.911 (95% CI, 0.908–0.914) and 0.837 (95% CI, 0.827–0.847) in the training and validation cohorts, respectively. Moreover, we identified some objective clinical features (including T stage and dilated main pancreatic duct/bile duct status) related to tumor grade. Interestingly, the predictive performance was further improved by combining the radiomic signature with the clinical features mentioned above as a combined nomogram, achieving an AUC of 0.919 (95% CI, 0.916–0.922) and 0.875 (95% CI, 0.867–0.883) in the training and validation cohorts, respectively.

Recent developments in radiomics attract much interests in tumor detection, subtype classification, therapeutic response assessment, prediction of clinical outcome and tumor monitoring, etc. Most of them were attempt to stratify the biological behavior and optimize tailored therapeutic strategies for these heterogenous tumors such as PNETs. Traditional radiographic assessment [including CT (15) and MRI (17)] which commonly relies on visual evaluation, was previously reported to predict the biological aggressiveness of PNETs. Moreover, (18)F-FDG-PET/CT and (68)Ga-DOTANOC-PET/CT were reported to be useful in predicting tumor grade (14). However, the results vary a lot and the accuracy remains challenging, as the prediction was mainly established based on visual observation rather than quantitative information.

Radiomics and artificial intelligence (AI) automatically extract high-throughput quantitative image data. Just as limited studies reported previously, it could be more useful for differentiating pathologic grading in patients with PNETs than routine CT image features alone (31, 32). Whereas, combining

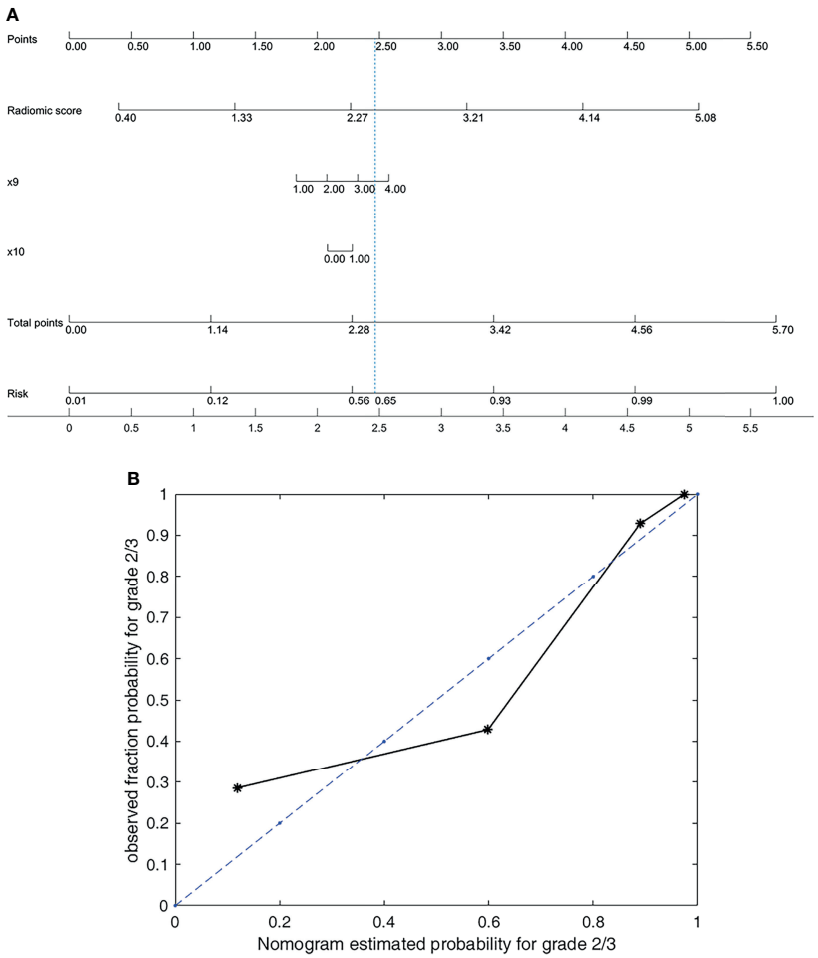


FIGURE 3 | Nomogram and its calibration curve based on group 2 for predicting grade 2/3. **(A)** Nomogram for group 2. **(B)** Calibration curve, where the diagonal dotted line is a perfect estimation by an ideal model. The predicted (estimated) probabilities of the validation set were sorted and divided into four groups based on quartiles to calculate the observed true probabilities. We calculated the goodness of fit to evaluate how well the solid line fits the dotted line. The goodness of fit is 0.8683, which indicates that the two lines fit well.

feature engineering and machine learning is a widely used scheme in radiomics-aided diagnosis (32). Deep learning features are highly versatile, their ability to solve specific problems is relatively weak (33). In contrast, building an interpretable AI model based on feature engineering is relatively easy. The output of the model is expected to be understood by physicians in clinical applications. Nomograms based on linear models intuitively illustrate what drives the recognition in machine learning. We build a nomogram based on the group 2 model in plain phase, as shown in **Figure 3**. Wan's research (34) investigated the performance of the combination of conventional handcrafted and learning-based features in disease recognition. For a specific research question, they emphasised that developing specific feature selection and model optimization approaches was necessary to achieve high accuracy and robustness. Consistent with this, the present paper proposed our optimized approaches according to the PNET grading issue (as illustrated in **Figure 1**).

As depicted in **Table 1**, our data showed that tumor size in grade 2/3 group was significantly larger than that in grade 1 group (p = 0.028). Dilated MPD/BD and poorly defined tumor margin were more frequently detected in patients with grade 2/3 PNETs than those with grade 1 (p = 0.02 and p = 0.013, respectively). Consistent results occurred in both training and validation sets in the present study. Moreover, research by Kim and colleague (15) identified the three indentical tumor CT features above as predictors of higher tumor grade of PNETs. Of note, the assessment of these features was relatively objective and the data can be automatically acquired in bulk. On the contrary, the data of tumor T stage, TNM stage, and diagnosis of insulinoma were obtained partly by doctors' experience, although these features were suggested to be significantly different between grade 2/3 PNETs and grade 1 group in our study (**Table 1**). Therefore, to improve the predictive performance, we establish a combined nomogram model that integrates radiomic signature with the former three relatively

objective clinical features (including tumor size, tumor margin, and dilated main pancreatic duct/bile duct) (**Table 2**). To our knowledge, our present comprehensive nomogram is the first study that integrates radiomic signature based on plain CT images with objective clinical features for noninvasive and preoperative prediction of pathologic grades for each PNET patients with high accuracy (both in training and independent validation set). Wenjie Liang and colleagues (35) reported a nomogram combining radiomic signature based on contrast-enhanced CT and clinical stage. Plain CT has lower cost and more convenience than contrast-enhanced CT. Also, accurate preoperative TNM staging of the tumor is difficult, as the preoperative assessment of "N" and "M" status remains challenging. Interestingly, Zhang's research (36) depicted impressive results based on enhanced CT radiomic features with 3D modeling.

As is known to us, a quite different therapeutic strategy could be applied for PNETs of grades 1 and 2/3. For clinical practice, the present combined nomogram may facilitate personalized treatment decisions for each patient with this heterogeneous tumor. It is noninvasive and could identify PNETs of grades 1 and 2/3 with high accuracy preoperatively. According to ENET guidelines in terms of PNET, NF-PNETs of less than 2 cm with grade 1 were optimized candidates for a "wait and see" policy. Moreover, parenchyma-sparing procedure such as enucleation could be an alternative for PNET with grade 1, while radical resection with formal lymphadenectomy was recommended for PNET with grade 2/3. In addition, the therapeutic strategies for the advanced PNETs of graded 1 and 2/3 varied dramatically (palliative surgery, somatostatin analog, targeted therapy, or chemotherapy). Our present combined model may facilitate tailored surgical decisions. Additionally, given the spatial and temporal heterogeneity of the specific tumor, the noninvasive model can be used repeatedly for tumor monitoring (especially for the patient initially recommended to wait and see) and to dynamically optimize therapeutic regimen for patients with advanced PNETs.

A major limitation of the present study was the relatively insufficient sample size. In addition, given that the G3 group was small (approximately 10% of PNETs in our series and as previously reported), our present nomogram model was established to stratify PNET patients into G1 and G2/3 groups. To better optimize personalized therapeutic strategies, a nomogram to separate G2 and G3 groups is further needed to be established based on lager samples. We are trying to collect more cases from multicenters and explore more appropriate methods to conduct further studies. Thirdly, we used single-layer CT image in this study, while 3D modeling may more comprehensively reflect the overall characteristics of the tumor, it is worth exploring whether it can obtain a more powerful predictive capability. On the other hand, manual tumor segmentation for 3D modeling was time consuming, and it was not applied for small tumors without thin-slice CT scans.

CONCLUSIONS

The developed combined nomogram that integrates radiomic signature based on plain CT images with clinical features (including T stage and dilated main pancreatic duct/bile duct status) can effectively predict the pathologic grades of PNETs preoperatively with powerful predictive capability. The noninvasive predictive model could assist clinicians to optimize tailored therapeutic strategies and facilitate surgical decision-making for each patient with PNETs in practice. It intuitively illustrates what drives the recognition in the prediction, which is potentially valuable in actual clinical applications and precision medicine in the future.

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

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements.

AUTHOR CONTRIBUTIONS

XW, J-JQ, DC, X-BL, and N-WK: study concept and design, data analysis and interpretation, and drafting of the manuscript. XW, J-JQ, C-LT, Q-QT, S-JR, and Y-HC: substantial contribution to data acquisition and interpretation and critical revision of the manuscript. XW and J-JQ: substantial contribution to data acquisition and data analysis. All authors: final approval of the manuscript and agreement with all the aspects of the work. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by research grants from the National Natural Science Foundation of China (82002579), the China Postdoctoral Science Foundation Funded Project (2019M663519), the Science and Technology Support Project of Sichuan Province (2020YFS0262), and the Post-Doctor Research Project, West China Hospital, Sichuan University (2019HXBH044).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.843376/full#supplementary-material

REFERENCES

1. Vortmeyer AO, Huang S, Lubensky I, Zhuang Z. Non-Islet Origin of Pancreatic Islet Cell Tumors. *J Clin Endocrinol Metab* (2004) 89(4):1934–8. doi: 10.1210/jc.2003-031575
2. Rindi G, Falconi M, Klersy C, Albarello L, Boninsegna L, Buchler MW, et al. TNM Staging of Neoplasms of the Endocrine Pancreas: Results From a Large International Cohort Study. *J Natl Cancer Institute* (2012) 104(10):764–77. doi: 10.1093/jnci/djs208
3. Fan JH, Zhang YQ, Shi SS, Chen YJ, Yuan XH, Jiang LM, et al. A Nation-Wide Retrospective Epidemiological Study of Gastroenteropancreatic Neuroendocrine Neoplasms in China. *Oncotarget* (2017) 8(42):71699–708. doi: 10.18632/oncotarget.17599
4. Dasari A, Shen C, Halperin D, Zhao B, Zhou S, Xu Y, et al. Trends in the Incidence, Prevalence, and Survival Outcomes in Patients With Neuroendocrine Tumors in the United States. *JAMA Oncol* (2017) 3 (10):1335–42. doi: 10.1001/jamaoncol.2017.0589
5. Cives M, Strosberg JR. Gastroenteropancreatic Neuroendocrine Tumors. *CA: Cancer J Clin* (2018) 68(6):471–87. doi: 10.3322/caac.21493
6. Yang Z, Tang LH, Klimstra DS. Effect of Tumor Heterogeneity on the Assessment of Ki67 Labeling Index in Well-Differentiated Neuroendocrine Tumors Metastatic to the Liver: Implications for Prognostic Stratification. *Am J Surg Pathol* (2011) 35(6):853–60. doi: 10.1097/PAS.0b013e31821a0696
7. Partelli S, Gaujoux S, Boninsegna L, Cherif R, Crippa S, Couvelard A, et al. Pattern and Clinical Predictors of Lymph Node Involvement in Nonfunctioning Pancreatic Neuroendocrine Tumors (NF-PanNETs). *JAMA Surg* (2013) 148(10):932–9. doi: 10.1001/jamasurg.2013.3376
8. Pulvirenti A, Javed AA, Landoni L, Jamieson NB, Chou JF, Miotto M, et al. Multi-Institutional Development and External Validation of a Nomogram to Predict Recurrence After Curative Resection of Pancreatic Neuroendocrine Tumors. *Ann Surg* (2019) 274(6):1051–7. doi: 10.1097/SLA.0000000000003579
9. Tan QQ, Wang X, Yang L, Chen YH, Tan CL, Ke NW, et al. Predicting Survival in Non-Functional Pancreatic Neuroendocrine Tumours. *ANZ J Surg* (2020) 90(10):2026–31. doi: 10.1111/ans.16072
10. Bosman FT, Carneiro F, Hruban RH, Theise ND. *WHO Classifcation of Tumours of the Digestive System*. Lyon: International Agency for Research on Cancer (2010).
11. Marchegiani G, Landoni L, Andrianello S, Masini G, Cingarlini S, D'Onofrio M, et al. Patterns of Recurrence After Resection for Pancreatic Neuroendocrine Tumors: Who, When, and Where? *Neuroendocrinology* (2019) 108(3):161–71. doi: 10.1159/000495774
12. Heidsma CM, Tsilimigras DI, Rocha F, Abbott DE, Fields R, Smith PM, et al. Clinical Relevance of Performing Endoscopic Ultrasound-Guided Fine-Needle Biopsy for Pancreatic Neuroendocrine Tumors Less Than 2 Cm. *J Surg Oncol* (2020) 122(7):1393–400. doi: 10.1002/jso.26158
13. Crinò SF, Ammendola S, Meneghetti A, Bernardoni L, Conti Bellocchi MC, Gabbrielli A, et al. Comparison Between EUS-Guided Fine-Needle Aspiration Cytology and EUS-Guided Fine-Needle Biopsy Histology for the Evaluation of Pancreatic Neuroendocrine Tumors. *Pancreatol Off J Int Assoc Pancreatol (IAP) [et al]* (2021) 21:443–50. doi: 10.1016/j.pan.2020.12.015
14. Majala S, Seppänen H, Kemppainen J, Sundström J, Schalin-Jäntti C, Gullichsen R, et al. Prediction of the Aggressiveness of Non-Functional Pancreatic Neuroendocrine Tumors Based on the Dual-Tracer PET/Ct. *EJNMMI Res* (2019) 9(1):116. doi: 10.1186/s13550-019-0585-7
15. Kim DW, Kim HJ, Kim KW, Byun JH, Song KB, Kim JH, et al. Neuroendocrine Neoplasms of the Pancreas at Dynamic Enhanced CT: Comparison Between Grade 3 Neuroendocrine Carcinoma and Grade 1/2 Neuroendocrine Tumour. *Eur Radiol* (2015) 25(5):1375–83. doi: 10.1007/s00330-014-3532-z
16. Jang KM, Kim SH, Lee SJ, Choi D. The Value of Gadoxetic Acid-Enhanced and Diffusion-Weighted MRI for Prediction of Grading of Pancreatic Neuroendocrine Tumors. *Acta radiologica (Stockholm Sweden 1987)* (2014) 55(2):140–8. doi: 10.1177/0284185113494982
17. Lotfalizadeh E, Ronot M, Wagner M, Cros J, Couvelard A, Vullierme MP, et al. Prediction of Pancreatic Neuroendocrine Tumour Grade With MR Imaging Features: Added Value of Diffusion-Weighted Imaging. *Eur Radiol* (2017) 27(4):1748–59. doi: 10.1007/s00330-016-4539-4
18. 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
19. 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/ncomms5644
20. 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 an Off J Am Assoc Cancer Res* (2017) 23(23):7253–62. doi: 10.1158/1078-0432.CCR-17-1038
21. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. *J Clin Oncol Off J Am Soc Clin Oncol* (2016) 34(18):2157–64. doi: 10.1200/JCO.2015.65.9128
22. Ding J, Xing Z, Jiang Z, Chen J, Pan L, Qiu J, et al. CT-Based Radiomic Model Predicts High Grade of Clear Cell Renal Cell Carcinoma. *Eur J Radiol* (2018) 103:51–6. doi: 10.1016/j.ejrad.2018.04.013
23. Huang X, Cheng Z, Huang Y, Liang C, He L, Ma Z, et al. CT-Based Radiomics Signature to Discriminate High-Grade From Low-Grade Colorectal Adenocarcinoma. *Acad Radiol* (2018) 25(10):1285–97. doi: 10.1016/j.acra.2018.01.020
24. Banik S, Rangayyan RM, Desautels JE. Measures of Angular Spread and Entropy for the Detection of Architectural Distortion in Prior Mammograms. *Int J Comput Assist Radiol Surg* (2013) 8(1):121–34. doi: 10.1007/s11548-012-0681-x
25. Szczypinski PM, Strzelecki M, Materka A, Klepaczko A. MaZda–a Software Package for Image Texture Analysis. *Comput Methods Programs BioMed* (2009) 94(1):66–76. doi: 10.1016/j.cmpb.2008.08.005
26. Yang X, Tridandapani S, Beitler JJ, Yu DS, Yoshida EJ, Curran WJ, et al. Ultrasound GLCM Texture Analysis of Radiation-Induced Parotid-Gland Injury in Head-and-Neck Cancer Radiotherapy: An *In Vivo* Study of Late Toxicity. *Med Phys* (2012) 39(9):5732–9. doi: 10.1118/1.4747526
27. Yao JC, Hassan M, Phan A, Dagohoy C, Leary C, Mares JE, et al. One Hundred Years After "Carcinoid": Epidemiology of and Prognostic Factors for Neuroendocrine Tumors in 35,825 Cases in the United States. *J Clin Oncol* (2008) 26(18):3063–72. doi: 10.1200/JCO.2007.15.4377
28. Pulvirenti A, Marchegiani G, Pea A, Allegrini V, Esposito A, Casetti L, et al. Clinical Implications of the 2016 International Study Group on Pancreatic Surgery Definition and Grading of Postoperative Pancreatic Fistula on 775 Consecutive Pancreatic Resections. *Ann Surg* (2018) 268(6):1069–75. doi: 10.1097/SLA.0000000000002362
29. Jiao Y, Shi C, Edil BH, de Wilde RF, Klimstra DS, Maitra A, et al. DAXX/ATRX, MEN1, and mTOR Pathway Genes Are Frequently Altered in Pancreatic Neuroendocrine Tumors. *Sci (New York NY)* (2011) 331 (6021):1199–203. doi: 10.1126/science.1200609
30. Scarpa A, Chang DK, Nones K, Corbo V, Patch AM, Bailey P, et al. Whole-Genome Landscape of Pancreatic Neuroendocrine Tumours. *Nature* (2017) 543(7643):65–71. doi: 10.1038/nature21063
31. Zhao Z, Bian Y, Jiang H, Fang X, Li J, Cao K, et al. CT-Radiomic Approach to Predict G1/2 Nonfunctional Pancreatic Neuroendocrine Tumor. *Acad Radiol* (2020) 27(12):e272–e81. doi: 10.1016/j.acra.2020.01.002
32. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial Intelligence in Cancer Imaging: Clinical Challenges and Applications. *CA: Cancer J Clin* (2019) 69(2):127–57. doi: 10.3322/caac.21552
33. Niethammer M, Kwitt R, Vialard FX. Metric Learning for Image Registration. *Proc IEEE Comput Soc Conf Comput Vision Pattern Recognition* (2019) 2019:8455–64. doi: 10.1109/CVPR.2019.00866
34. Wan Y, Yang P, Xu L, Yang J, Luo C, Wang J, et al. Radiomics Analysis Combining Unsupervised Learning and Handcrafted Features: A Multiple-Disease Study. *Med Phys* (2021) 48:7003–15. doi: 10.1002/mp.15199
35. Liang W, Yang P, Huang R, Xu L, Wang J, Liu W, et al. A Combined Nomogram Model to Preoperatively Predict Histologic Grade in Pancreatic Neuroendocrine Tumors. *Clin Cancer Res an Off J Am Assoc Cancer Res* (2019) 25(2):584–94. doi: 10.1158/1078-0432.CCR-18-1305

36. Zhang T, Zhang Y, Liu X, Xu H, Chen C, Zhou X, et al. Application of Radiomics Analysis Based on CT Combined With Machine Learning in Diagnostic of Pancreatic Neuroendocrine Tumors Patient's Pathological Grades. *Front Oncol* (2020) 10:521831. doi: 10.3389/fonc.2020.521831

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.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 12 May 2022
doi: 10.3389/fonc.2022.872503

Predictive Efficacy of a Radiomics Random Forest Model for Identifying Pathological Subtypes of Lung Adenocarcinoma Presenting as Ground-Glass Nodules

Fen-hua Zhao[1†], Hong-jie Fan[2†], Kang-fei Shan[1], Long Zhou[2], Zhen-zhu Pang[2], Chun-long Fu[1], Ze-bin Yang[1], Mei-kang Wu[1], Ji-hong Sun[2], Xiao-ming Yang[3] and Zhao-hui Huang[1*]*

[1] Department of Radiology, Affiliated Dongyang Hospital of Wenzhou Medical University, Dongyang, China, [2] Department of Radiology, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China, [3] Image-Guided Bio-Molecular Intervention Research, Department of Radiology, University of Washington School of Medicine, Seattle, WA, United States

OPEN ACCESS

Edited by:
Min Tang,
Jiangsu University, China

Reviewed by:
Weinan Zhou,
University of Illinois at Urbana-Champaign, United States
Pritam Mukherjee,
Stanford University, United States

***Correspondence:**
Zhao-hui Huang
hzh476@yeah.net
Xiao-ming Yang
xmyang@zju.edu.cn;
xmyang@uw.edu

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 09 February 2022
Accepted: 18 April 2022
Published: 12 May 2022

Citation:
Zhao F-h, Fan H-j, Shan K-f, Zhou L, Pang Z-z, Fu C-l, Yang Z-b, Wu M-k, Sun J-h, Yang X-m and Huang Z-h (2022) Predictive Efficacy of a Radiomics Random Forest Model for Identifying Pathological Subtypes of Lung Adenocarcinoma Presenting as Ground-Glass Nodules. Front. Oncol. 12:872503. doi: 10.3389/fonc.2022.872503

Purpose : To establish and verify the ability of a radiomics prediction model to distinguish invasive adenocarcinoma (IAC) and minimal invasive adenocarcinoma (MIA) presenting as ground-glass nodules (GGNs).

Methods: We retrospectively analyzed 118 lung GGN images and clinical data from 106 patients in our hospital from March 2016 to April 2019. All pathological classifications of lung GGN were confirmed as IAC or MIA by two pathologists. R language software (version 3.5.1) was used for the statistical analysis of the general clinical data. ITK-SNAP (version 3.6) and A.K. software (Analysis Kit, American GE Company) were used to manually outline the regions of interest of lung GGNs and collect three-dimensional radiomics features. Patients were randomly divided into training and verification groups (ratio, 7:3). Random forest combined with hyperparameter tuning was used for feature selection and prediction modeling. The receiver operating characteristic curve and the area under the curve (AUC) were used to evaluate model prediction efficacy. The calibration curve was used to evaluate the calibration effect.

Results: There was no significant difference between IAC and MIA in terms of age, gender, smoking history, tumor history, and lung GGN location in both the training and verification groups (P>0.05). For each lung GGN, the collected data included 396 three-dimensional radiomics features in six categories. Based on the training cohort, nine optimal radiomics features in three categories were finally screened out, and a prediction model was established. We found that the training group had a high diagnostic efficacy [accuracy, sensitivity, specificity, and AUC of the training group were 0.89 (95%CI, 0.73 - 0.99), 0.98 (95%CI, 0.78 - 1.00), 0.81 (95%CI, 0.59 - 1.00), and 0.97 (95%CI, 0.92-1.00), respectively; those of the validation group were 0.80 (95%CI, 0.58 - 0.93), 0.82 (95%CI, 0.55 - 1.00), 0.78 (95%CI, 0.57 - 1.00), and 0.92 (95%CI, 0.83 - 1.00), respectively]. The model calibration curve showed good consistency between the predicted and actual probabilities.

Conclusions: The radiomics prediction model established by combining random forest with hyperparameter tuning effectively distinguished IAC from MIA presenting as GGNs and represents a noninvasive, low-cost, rapid, and reproducible preoperative prediction method for clinical application.

Keywords: lung tumor, ground-glass nodules, radiomics, random forest, diagnosis

1 INTRODUCTION

Ground-glass nodule (GGN) refers to a nodular shadow with slightly increased density on high-resolution computed tomography (HRCT), in which the vascular and bronchial bundles are not covered (1, 2). With the popularization of HRCT and the extensive application of low-dose screening for lung cancer, the detection rate of lung GGN has been constantly increasing (3). Lung GGN is a characteristic but non-specific imaging manifestation. Theoretically, with any decrease in air content in the lung tissue, increase in cell density, and proliferation of columnar cells in the alveolar wall leading to a decrease in gas filling in the terminal saccules and alveoli, ground-glass opacities can appear before alveoli collapse completely. Research has shown that persistent lung GGNs are mostly attributed to precancerous lesions or early-stage lung adenocarcinoma (4). The 2011 International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society International Multidisciplinary Lung Adenocarcinoma Classification (5) and WHO (2021) Classification of Lung Tumors Pathology (6) divided lung adenocarcinoma into three categories: pre-invasive lesions, minimally invasive adenocarcinoma (MIA), and invasive adenocarcinoma (IAC), among which pre-invasive lesions include atypical adenomatous hyperplasia (AAH) and adenocarcinoma *in situ* (AIS). AAH, AIS, MIA, and IAC are a dynamic process of continuous progression involving multiple genes, and AAH and AIS can gradually develop into MIA and IAC (7). According to the literature, when a lung GGN was completely removed and the margin was negative, the 5-year disease-free survival of AIS and MIA was 100% or close to 100% (8), the 10-year disease-specific survival (DSS) was 100% or 97.3%, and the 10-year DSS of IAC was 74.8% or 80.2%. Thus, the prognosis of IAC was significantly worse than that of MIA and AIS (9). The difference in prognosis determines the difference in clinical diagnosis and treatment schemes. Although it is still controversial, most researchers believe that scheduled follow-up or sublobar resection (wedge resection or segmental resection) is suitable for pre-invasive lesions and MIA, which can preserve more lung tissue as well as reduce the mortality and morbidity related to surgery, while lobectomy should be performed for IAC (10, 11). Therefore, accurate preoperative differentiation between IAC and MIA+ pre-invasive lesions, especially IAC and MIA, will assist in determining the appropriate surgical methods and the judgment of prognosis (12).

Radiomics is a newly emerging field and was first proposed by Dutch scholars Lambin et al. It refers to extracting a large volume of data that are hard to observe with the naked human eye from medical images such as B-mode ultrasonography, CT, magnetic resonance imaging, and positron emission tomography; radiomics uses a data characterization algorithm to transform the medical image data into minable feature space data with high resolution (13). Quantifying the heterogeneity of tumors using radiomics analysis software can help to obtain more information. Moreover, radiomics is not affected by the inherent limitations of the professional level or subjective analysis and traditional image interpretation; it can help us to effectively carry out pathological classification, treatment plan formulation, and treatment outcome and prognosis evaluation, among other tasks. It is widely known that heterogeneity is a recognized malignant feature of tumors, which is related to their adverse biological behavior. The heterogeneity of tumors is related to various gene subtypes, growth expression, and neovascularization and tumor microenvironment factors, which lead to local differences in the proliferation, metabolic activity, apoptosis, and blood supply among different tumors (14). At present, radiomics has been gradually applied in differentiating benign from malignant pulmonary nodules (15), evaluating treatment outcomes of lung cancer (16), and predicting the recurrence and metastasis of lung cancer (17), among other tasks. There are few studies on predicting the pathological subtypes of GGN lung adenocarcinoma, and most of them were used to differentiate MIA/IAC from pre-invasive lesions (AAH/AIS). Weng et al. (18) attempted to differentiate IAC from MIA by combining morphology with omics; however, no research has been reported on differentiation between IAC and MIA by pure omics labeling models. In this study, random forest (19) combined with hyperparameter tuning was used to establish and verify the ability of a radiomics prediction model to distinguish IAC and MIA presenting as lung GGN and to evaluate the consistency between the probability predicted by the model and the actual probability.

2 MATERIALS AND METHODS

2.1 Research Subjects

The study design was approved by the appropriate ethics review board, and the requirement for obtaining informed patient consent was waived owing to the retrospective nature of the study.

In this study, we retrospectively analyzed data from patients treated in our hospital from March 2016 to April 2019. The inclusion criteria were as follows (6, 20, 21): (i) GGN on pre-operation chest HRCT scans; (ii) the images were scanned using

the same scanning protocol on the same CT machine; (iii) presence of lesions on at least five sections on HRCT axial images; (iv) all lesions were confirmed as lung IAC or MIA by pathology after surgical resection of specimens or percutaneous biopsy: because the pathological classification of lung adenocarcinoma would have been influenced by the subjective experience of the pathologists, the pathological classification was confirmed by two pathologists who had worked for 10 years after reaching a consensus according to the new 2015 classification criteria for lung adenocarcinoma; and (v) treatment-naive cases before HRCT. Cases that did not meet the diagnostic requirements where only a routine CT scan was performed or the respiratory motion artifact was too severe were excluded.

2.2 Examination Methods

A Brilliance 64-slice CT (Philips Medical Systems, Netherlands) machine was used for scanning. All patients received strict breathing training before scanning, adopted the head-first supine position, and adopted an end-inspiratory hold during scanning. The scanning scope covered all areas from the apices to the bottom of the lungs. Exposure conditions: 120 kV, 150 mA, collimation 0.625 mm × 64, pitch 0.64, scanning and reconstruction matrix both at 1024 × 1024, reconstruction slice thickness and interval both at 0.67 mm. Scanning image observation: mediastinal window (window position: 30–50 Hu; window width: 250–350 Hu); pulmonary window (window position: −450 to −600 Hu; window width: 1500–2000 Hu). The conventional CT scan images could not accurately identify lung GGNs; therefore, this study did not use them.

2.3 Image Analysis

First, all Digital Imaging and Communications in Medicine images of lung GGNs were imported into the A.K. (Analysis Kit) software developed by GE (USA) for pre-processing. Then, ITK-SNAP software (Version 3.6) was used to manually outline the regions of interest (ROIs) layer by layer along the inner edges of lung GGNs based on pixel points, and then they were fused and saved into three-dimensional (3D) images (**Figure 1**). All lung GGN images were outlined by a resident who had worked for 5 years and a deputy chief physician who had worked for 15 years. The intraclass correlation coefficient (ICC) was used for consistency analysis, and an ICC >0.8 indicated good consistency (22). The sketchers were blinded to the pathological results of the lung GGNs.

To acquire radiomics feature, the original images of lung GGNs after pre-processing and the corresponding ROI 3D images were imported into A.K. software in batches, and six types of radiomics features were quantitatively calculated: histogram, form factor, texture, gray level co-occurrence matrix (GLCM), run-length matrix (RLM), and gray level zone size matrix (GLSZM).

2.4 Statistical Methods

2.4.1 Statistical Analysis of the Clinical Data

Using R language software (Version 3.5.1), descriptive statistical analysis was carried out between the training and verification groups. The chi-square test was used for qualitative variables, and the t-test or rank sum test was used for continuous variables, with $P<0.05$ indicating that the difference was statistically significant. In addition, the Bootstrap method is used to estimate the confidence interval.

2.4.2 Screening of Radiomics Features and Construction of a Random Forest Prediction Model

In this study, random forest combined with hyperparameter tuning was used for prediction modeling. As a leader of ensemble learning methods, random forest trains decision tree models with partial data and partial features and then fuses these tree models, and finally, uses voting to solve classification problems. Random forest can directly deal with high-dimensional data, and there is no need for feature screening before modeling. Random forest hyperparameter tuning includes model framework parameters and decision tree parameters. For model framework parameters, the number of weak learners is mainly tuned, i.e., the number of decision trees, and the range set in this study was 50–500. Decision tree parameters tuning includes tree depth (3–10 layers) and the number of leaf nodes (10–50). Using random grid search and 10-fold cross-validation, the results of hyperparameter tuning in each iteration were evaluated by accuracy. Finally, we found the best random forest hyperparameters. Based on the training data, we found the following optimal hyperparameters in this study: the number of decision trees was 184, the tree depth was 5, and the number of leaf nodes was 20. The best random forest model was also utilized in the training data to assess the importance of features and feature selection.

Using this set of hyperparameter settings, prediction analysis of the training group and verification group based on the random forest algorithm was carried out again, and the receiver operating characteristic (ROC) curve was used to evaluate the prediction efficacy of the model. The calibration curve was used to evaluate the consistency between the probability predicted by the model and the actual probability.

3 RESULTS

3.1 Comparison of the General Clinical Data of Patients in the Training and Verification Group

A total of 118 lung GGNs [36 pure GGNs (pGGNs) and 82 mixed GGNs (mGGNs)] in 106 patients were included in this study, including 27 men (25.5%) and 79 women (74.5%) whose ages ranged from 28 to 76 years, with an average age of 55.61 ± 11.50 years. The surgical and pathological analysis confirmed 61 IAC lesions in 56 patients and 57 MIA lesions in 53 patients; 42 lesions were located in the left lung (18 in the upper left lobe and 24 in the lower left lobe), and 76 in the right lung (10 in the upper right lobe, 42 in the middle right lobe, and 24 in the lower right lobe) (**Figure 2**). Thirteen patients had a history of smoking, and eight had a history of a malignant tumor, all of which were lung cancer, two of them also had a history of thyroid cancer. Ninety-six patients had single lung GGNs, whereas 10 had multiple,

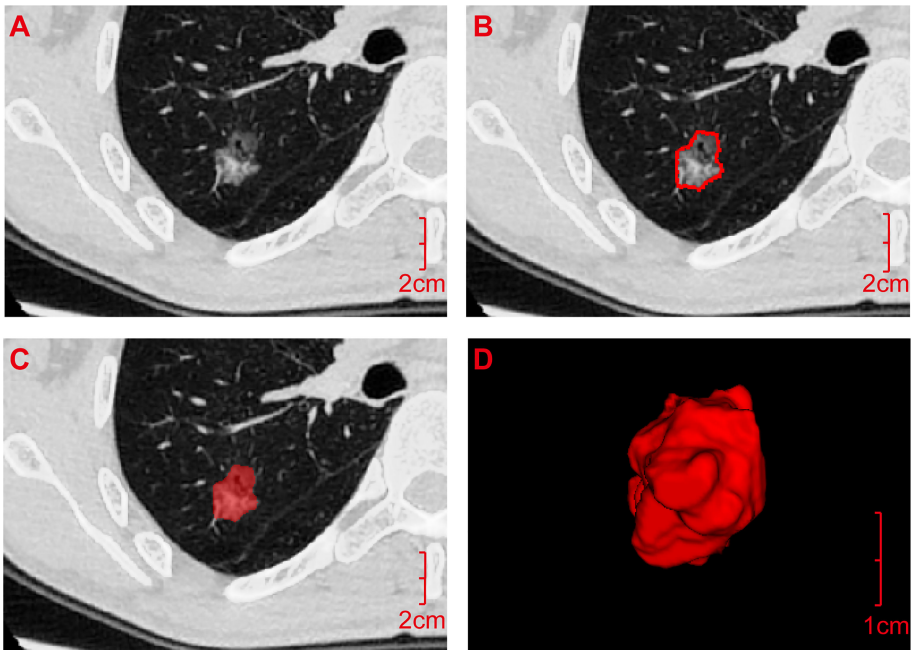


FIGURE 1 | Acquisition process for the radiomics features. **(A)** Digital Imaging and Communications in Medicine image of the transverse section of the ground-glass nodule was pre-processed by A.K. software and then imported into ITK-SNAP software. **(B)** ROI was manually outlined layer by layer along the inner edge of the lesion based on pixel points. **(C)** ROI was outlined. **(D)** It was fused and saved as an ROI three-dimensional image and imported into A.K. software together with the pre-processed original image **(A)** in batches to quantitatively calculate the radiomics features. ROI, Region of interest.

among which two patients had three lung GGNs resected at the same time. Postoperative pathology results showed three IAC lesions in one patient and two IAC lesions and one MIA lesion in the other patient. All patients underwent video-assisted thoracoscopic surgery, 56 patients underwent lobectomy, 40 patients underwent sublobar resection (segmental resection or wedge resection), and 10 patients underwent lobectomy and sublobar resection. No lymph node or distant metastasis was found in any patient after the operation. The random function of R language software (Version 3.5.1) was used to divide the 118 lung GGNs into the training group and verification group at a ratio of 7:3. There were 83 lesions (including 43 IAC lesions and 40 MIA lesions) in the training group and 35 lesions (including 18 IAC lesions and 17 MIA lesions) in the verification group. There was no significant difference between IAC and MIA in terms of age, gender, smoking history, tumor history, and lung GGN location in both groups (P>0.05) (**Table 1**).

3.2 Acquisition and Screening of Radiomics Features

A total of 396 valid radiomics features in six categories were collected by A.K. software for each lung GGN (**Table 2**): 42 histogram, 9 form factor, 144 texture, 11 GLCM, 180 RLM, and 10 GLSZM features. The ICC was used for consistency analysis, and the characteristic features with an ICC <0.8 were eliminated. Based on the data of the training group, random forest combined with hyperparameter tuning was used to tune parameters and evaluate the importance of the radiomics features (19). Then we extract the top-n features for training and evaluation, and the experiment reveals that the top-9 features yields the best outcome. Finally, nine optimal radiomics features were screened out. The names, categories, importance, and ICC of each radiomics features were shown in **Table 3** and **Figure 3**, and the difference between IAC and MIA in each feature was statistically significant ($P<0.05$) (**Table 4**). A total of nine features were classified into three main categories: RLM, gray level co-occurrence matrix, and histogram features. Among them, the seven RLM features included three short run low grey level emphasis features, one long run low grey level emphasis feature, one run length nonuniformity feature, one grey level nonuniformity feature, and one low grey level run emphasis feature. There was one gray level co-occurrence matrix feature and one histogram feature, which were GLCM energy and frequency size, respectively. **Figure 4** shows the distribution of the values of the radiomics features of all the IAC and MIA patients mentioned above in the training and test datasets.

3.3 Prediction Efficacy of the Radiomics Random Forest Model

Based on the data of the training group, random forest combined with hyperparameter tuning results were used for prediction modeling, and prediction analysis of the training group and verification group based on the random forest algorithm was carried out again with this set of hyperparameter settings; the

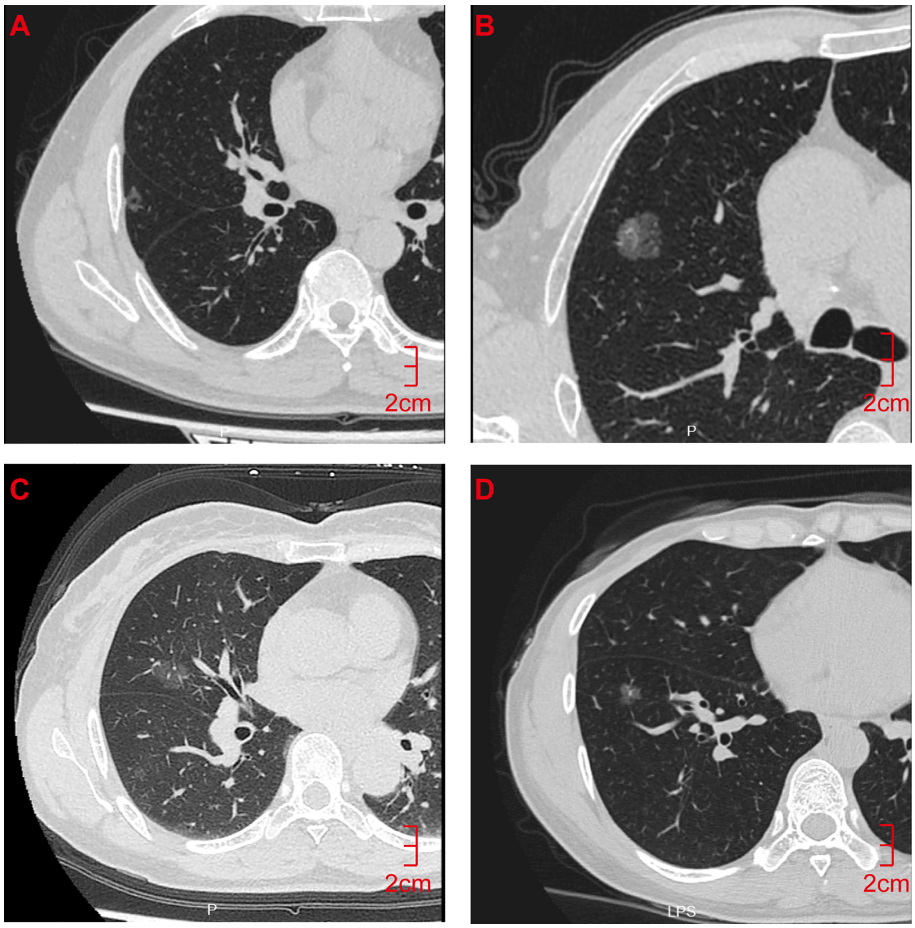


FIGURE 2 | CT Findings of representative IAC and MIA nodules. **(A)** M/55y, pGGN of the right upper lobe with a long diameter of 1.2 cm, pathological diagnosis showed that the lesion was IAC. **(B)** M/32y, mGGN of the right upper lobe with a long diameter of 1.5 cm, pathological diagnosis showed that the lesion was IAC. **(C)** F/62y, pGGN with a long diameter of 1.2 cm in the right middle lobe, pathological diagnosis showed that the lesion was MIA. **(D)** F/55y, pGGN with a long diameter of 1.0 cm in the right lower lobe, pathological diagnosis showed that the lesion was MIA. CT, computed tomography; pGGN, pure ground-glass nodule; mGGN, mixed ground-glass nodules; IAC, invasive adenocarcinoma; MIA, minimally invasive adenocarcinoma.

ROC curve was used to evaluate the prediction efficacy of the model (**Figure 5**). The accuracy for the training group was 0.89 (95%CI, 0.73 - 0.99), sensitivity was 0.98 (95%CI 0.78 - 1.00), specificity was 0.81 (95%CI, 0.59 - 1.00), and area under the curve (AUC) was 0.97(95%CI, 0.92-1.00); the accuracy for the verification group was 0.80 (95%CI, 0.58 - 0.93), sensitivity was 0.82 (95%CI, 0.55 - 1.00), specificity was 0.78 (95%CI, 0.57 - 1.00), and AUC was 0.92 (95%CI, 0.83 - 1.00). The calibration

TABLE 1 | Comparison of the general clinical data of patients in the training and verification groups (N = 118).

Clinical data	Training group			Verification group		
	IAC	MIA	P^1	IAC	MIA	P
Number of cases (cases)	42	38		16	15	
Number of lesions (number)	43	40		18	17	
Age (years)	56.21 ± 11.15	54.23 ± 9.98	0.3978	56.94 ± 12.80	54.71 ± 11.62	0.5938
Gender (male/female)	(12/30)	(9/29)	0.6198	(5/11)	(4/11)	0.2504
Smoking history (cases)	4	5	0.1109	2	2	0.6406
Tumor history (cases)	3	2	0.2476	1	2	0.0629
GGN location (left/right)	(16/27)	(12/28)	0.4876	(9/9)	(5/12)	0.2140

P<0.05 indicates statistically significant difference.
GGN, ground-glass nodule; IAC, invasive adenocarcinoma; MIA, minimally invasive adenocarcinoma.

TABLE 2 | Types and numbers of valid radiomics features.

Type of valid radiomics features	n
Histogram features	42
Form factor features	9
Texture features	144
Gray level co-occurrence matrix features	11
Run-length matrix features	180
Gray level zone size matrix features	10

curve of the prediction model (**Figure 6**) showed good agreement between the predicted probability and actual probability.

4 DISCUSSION

4.1 Epidemiology

Lung cancer is the leading cause of cancer-related death worldwide (23) and at present, it has the fastest growing prevalence and mortality among human malignancies (24) and the trend is rising. Adenocarcinoma is the most common pathological type of lung cancer, accounting for approximately 50% of cases of lung cancer. The proportion of males is higher than that of females, and smoking is the most important risk factor (25). In this study, the proportion of female patients was much higher than that of male patients (79/27), and the proportion of smokers (13/106) was relatively low, which is inconsistent with previous reports. It might be related to the case selection in this study, and at the same time, it is necessary to consider that the incidence in females has been increasing year after year due to indoor air pollution and second-hand smoke exposure, among other factors. There were no significant differences in age, gender, smoking history, tumor history, or GGN location between IAC and MIA patients, which is consistent with the literature (26). The National Lung Screening Trial in the United States showed that CT low-dose screening was helpful in reducing the mortality of lung cancer (27). In recent years, the detection rate of GGN lung adenocarcinoma has been increasing, and early diagnosis and accurate pathological classification have become the keys to treatment. However, due to the overlapping of traditional imaging manifestations among pathological subtypes, it is still difficult to accurately classify them, especially distinguishing between IAC and MIA, which are both invasive lesions. Therefore, a systematic and objective differential diagnosis method must be urgently developed.

TABLE 3 | Names, categories, importance, and ICC of the selected radiomics features.

Radiomics features	Importance	ICC
Run-length matrix features		
Short run low grey level emphasis_angle135_offset1	1.06635	0.988
Run length nonuniformity_AllDirection_offset1	1.06588	0.997
Long run low grey level emphasis_AllDirection_offset4	1.06569	0.993
Grey level nonuniformity_AllDirection_offset1	1.06564	0.984
Short run low grey level emphasis_AllDirection_offset4	1.06563	0.995
Low grey level run emphasis_angle90_offset1	1.06560	0.982
Short run low grey level emphasis_angle135_offset7	1.06554	0.965
Gray level co-occurrence matrix features		
GLCM energy_AllDirection_offset1_SD	1.06594	0.999
Histogram features		
Frequency size	1.06570	0.996

GLCM, gray level co-occurrence matrix; SD, square deviation; ICC, intraclass correlation coefficient.

4.2 Correlation Between Radiomics Features and Pathological Subtypes of GGN Lung Adenocarcinoma

Through univariate analysis, in the verification group, there was no significant difference in tumor history between IAC and MIA (P=0.0629). However, due to the small sample size, the results presented here should be carefully considered. Notably, since a purely radiomics prediction model was employed in this study, biological differences in tumor history have limited potential impact on model efficacy. When using combined clinical-radiomic prediction models, propensity score matching should be used to control for confounding variables in the presence of statistically significant biological differences in tumor history to improve data comparability. Using radiomics, numerous imaging features can be extracted *via* software analysis of lesion heterogeneity, which is objective, does not cost much and facilitates the prediction of clinical outcomes (13). In this study, nine optimal radiomics features were screened out by combining random forest with hyperparameter tuning, which were classified into three categories: RLM, gray level co-occurrence matrix, and histogram features. Except for one histogram feature that belongs to the low-order texture, the other eight features belong to the high-order texture which shows the distribution of pixel points. This also reveals that high-order texture can better reflect the spatial heterogeneity changes of lung GGNs.

The histogram is a function of the image gray level (28) which describes and compares the distributions of pathological or biological indicators quantitatively. Histogram features are related to the attributes of a single pixel, and the distribution of voxel intensity in CT images is described by common and basic indices; thus, the calculation results of voxel values are more accurate. It has been reported that the histogram pattern based on CT pixels can assist in distinguishing AAH from bronchioloalveolar carcinoma (29, 30). Although frequency size was the only histogram feature selected in this study, it ranked fourth in importance (**Figure 3**), which reflects its value in terms of the differential diagnosis of IAC and MIA. The GLCM is a two-dimensional gray histogram that examines a pair of pixels separated by a fixed spatial relationship, which reflects the change speed and amplitude of pixel gray levels at different intervals and in different directions in the image. It is the basis for analyzing the arrangement rules and local patterns of images, including 11 indices such as energy, entropy, inertia, and correlation (31), which reflect the internal characteristics and spatial heterogeneity of tumors. Energy is a set of feature values that indicates the complexity of image texture. A large energy

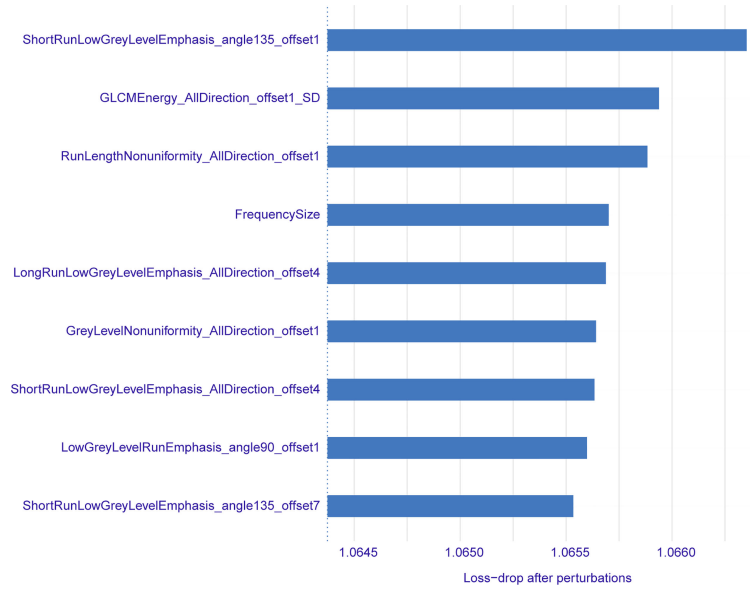


FIGURE 3 | Names, importance, and sorting of radiomics features.

value indicates that the image has very good uniformity or very similar pixels, and vice versa. Heterogeneity of malignant tumors is caused by tissue structure changes resulting from uneven distribution of cell density, hemorrhage, necrosis, and mucoid degeneration, among other factors (32). Pathologically, IAC is more heterogeneous than MIA, and the complexity of image texture is also higher, though it is not easy to detect using the naked eye. The energy value of IAC in this study was significantly less than that of MIA (**Table 4**), which objectively reflected the difference in image texture between them. The RLM features mainly reflect the roughness and directionality of the image. Directional texture will have a longer run length at a certain angle, in which the value of short run emphasis is larger on rougher images, and that of long run emphasis is larger on smoother images. The run length is related to the gray level distribution of the image, and the heterogeneity of the tumor often reflects the gray level changes of the image; thus, the RLM is very sensitive to the texture changes of lung GGNs. Among the nine optimal radiomics features that were finally selected, seven were RLM features, and short run emphasis ranked first in

TABLE 4 | Comparative analysis of selected radiomics features between IAC and MIA.

Radiomics Features	Pathological type	Value	*P*
Short run low grey level emphasis_angle135_offset1	IAC	$4.24\times10^{-4} \pm 4.31\times10^{-4}$	0.0000
	MIA	$1.55\times10^{-3} \pm 1.13\times10^{-3}$	
Run length nonuniformity_AllDirection_offset1	IAC	1705.55 ± 1686.94	0.0013
	MIA	320.71 ± 380.09	
Long run low grey level emphasis_AllDirection_offset4	IAC	$5.81\times10^{-3} \pm 6.97\times10^{-3}$	0.0005
	MIA	$2.39\times10^{-2} \pm 1.80\times10^{-2}$	
Grey level nonuniformity_AllDirection_offset1	IAC	19.56 ± 24.42	0.0000
	MIA	4.72 ± 5.61	
Short run low grey level emphasis_AllDirection_offset4	IAC	$4.20\times10^{-4} \pm 4.27\times10^{-4}$	0.0000
	MIA	$1.54\times10^{-3} \pm 1.12\times10^{-3}$	
Low grey level run emphasis_angle90_offset1	IAC	$1.50\times10^{-3} \pm 1.76\times10^{-3}$	0.0000
	MIA	$6.05\times10^{-3} \pm 4.51\times10^{-3}$	
Short run low grey level emphasis_angle135_offset7	IAC	$4.20\times10^{-4} \pm 4.26\times10^{-4}$	0.0000
	MIA	$1.53\times10^{-3} \pm 1.12\times10^{-3}$	
GLCM energy_AllDirection_offset1_SD	IAC	$1.21\times10^{-8} \pm 4.18\times10^{-8}$	0.0000
	MIA	$5.93\times10^{-7} \pm 2.13\times10^{-6}$	
Frequency size	IAC	1821.46 ± 1838.75	0.0000
	MIA	340.68 ± 417.84	

P<0.05 indicates statistically significant difference.
IAC, invasive adenocarcinoma; MIA, minimally invasive adenocarcinoma; GLCM, gray level co-occurrence matrix; SD, square deviation.

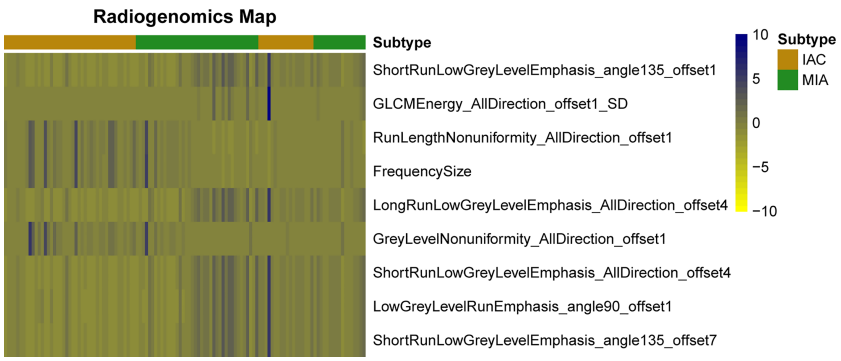


FIGURE 4 | Heat map of radiomics features. Each row represents a feature, and each column represents a lung ground-glass nodule. The figure shows the difference between invasive adenocarcinoma and minimal invasive adenocarcinoma in each feature and indicates the classification ability of the features.

importance (**Figure 3**). This shows that the RLM is most valuable for distinguishing IAC from MIA.

Form factor features are of great value in the differential diagnosis of pathological subtypes of GGN lung adenocarcinoma. Chae et al. (33) showed that mass (volume*density) was significantly different between pre-invasive lesions (AAH/AIS) and invasive lesions (MIA/IAC) presenting as mGGNs. However, none of the form factor features in this study were selected as an optimal feature, which might be related to the similarity and overlapping of morphological features between IAC and MIA, which is why it is more difficult to differentiate them.

4.3 Value and Superiority of the Radiomics Random Forest Model

Quantitative features such as the geometry, wavelet, and texture in the internal space of tumors were collected in a high-throughput way using radiomics software. The characteristic variables were screened out by computer artificial intelligence technology, and a quantitative prediction model was constructed, which made it a brand-new imaging diagnosis decision-making and analysis tool. Its application scope covers the qualitative analysis, clinical staging and grading, treatment outcome evaluation, and prognosis prediction of tumors (34).

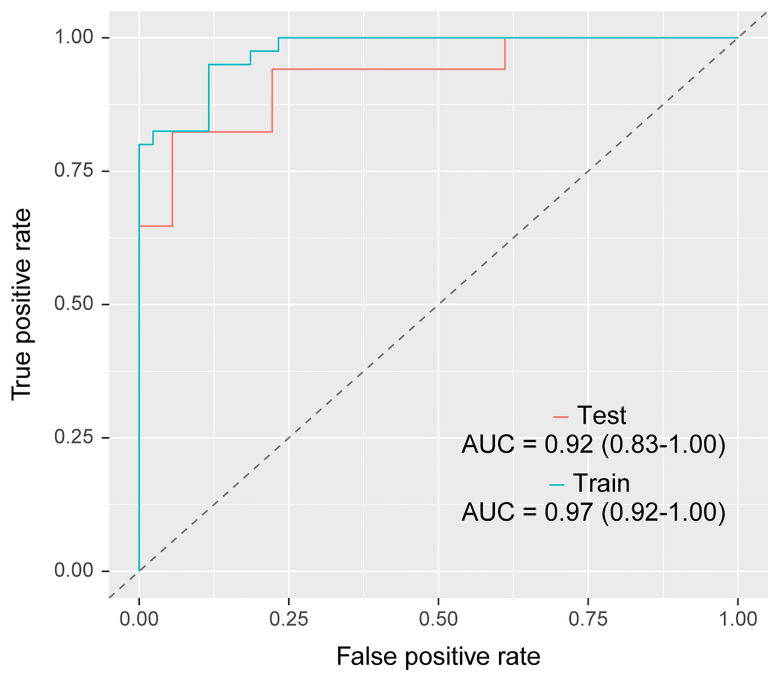


FIGURE 5 | Receiver operating characteristic curve analysis of the radiomics random forest prediction model in differentiating lung ground-glass nodule-type invasive adenocarcinoma and minimal invasive adenocarcinoma in the training group (blue line) and the verification group (red line). The areas under the curve of the training and verification groups were 0.97(95%CI, 0.92-1.00) and 0.92 (95%CI, 0.83 - 1.00), respectively. CI, confidence intervals.

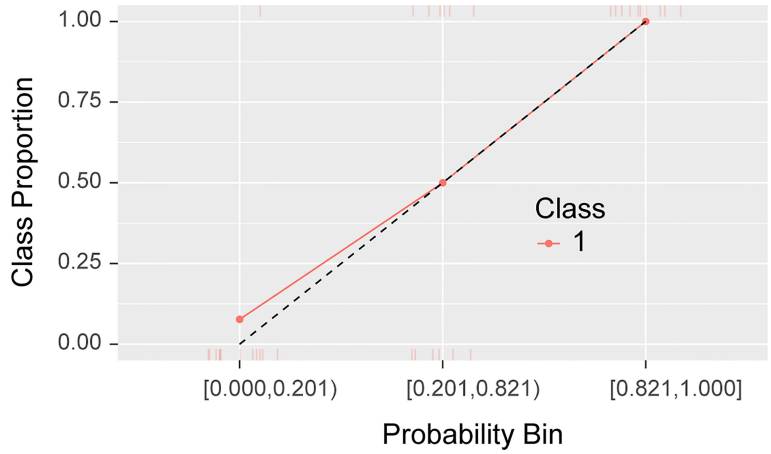


FIGURE 6 | Calibration curve of the radiomics random forest prediction mode.

At present, studies using radiomics prediction models to distinguish pathological subtypes of lung adenocarcinoma mainly focus on MIA/IAC and AAH/AIS or IAC and AIS/MIA, and have achieved promising results. For example, the predictive model between IVA and AIS/MIA based on pGGNs established by Xu et al. (35), the AUC value of the combined model was 0.848 (95% CI, 0.750-0.946); the combined clinical model of Wu et al. (36) - The AUC values of the radiological model were 0.917 and 0.876 in the training and validation groups, respectively. It is worth noting that in 2021, the "WHO Classification of Thoracic Tumors (5th Edition)" has excluded AIS from the category of lung malignancies and classified it as a glandular precursor lesion together with AAH. Although both MIA and IAC require surgical treatment, MIA is suitable for sublobar resection (wedge resection or segmentectomy) with a 5-year disease-free survival rate of approximately 100% after complete resection, and IAC is suitable for standard lobectomy and extensive lobectomy. With lymph node dissection, the 5-year disease-free survival rate was 74.6%. Therefore, accurate preoperative identification of IAC and MIA will help guide the selection of surgical methods and the judgment of prognosis. Currently, studies to differentiate MIA from IAC are very rare. A previous study (18) constructed a combined prediction model integrating lesion shape and radiological features to distinguish MIA from IAC, with an AUC of 0.888. In this study, by using the same equipment, using the same scanning protocol and the same reconstruction scheme to acquire images, and using AK software for image preprocessing to ensure image consistency, we finally obtained AUC superior to the above studies values.

Fan et al. (37) established an individualized prediction model based on the patient's age, spicule sign, pleural indentation sign, and radiomic labels, and established a clinical model based on the patient's age, spicule sign, and pleural indentation sign to distinguish GGN lung adenocarcinoma from invasive lesions (AAH/AIS). The results showed that the AUC increased from 0.743 in the clinical model to 0.934 in the individualized prediction model, indicating the importance of radiomic labeling. She et al. (38) included 402 patients with lung GGNs and extracted 60 radiomics features, among which five features were the most critical diagnostic factors. The results showed that the AUCs of the radiomics prediction model in the training group and verification group were 0.95 and 0.89, respectively, indicating that radiomics had advantages in differentiating IAC from MIA/AIS. Weng et al. (18) included 119 pulmonary mGGN patients to differentiate IAC from MIA and extracted 396 radiomics features, among which four were optimal distinguishing features for establishing a radiomics model. The results showed that the AUCs of the radiomics feature model for the training and verification groups were 0.854 and 0.813, respectively. Then, a CT feature model was established using lesion morphology and the diameter of the solid components; the AUC was 0.755. Finally, the lesion morphology and radiomics features were combined, and the AUC was 0.888. In this study, A.K. software was used to collect the radiomics features; the optimal radiomics features were screened out by random forest combined with hyperparameter tuning and a prediction model was established. The results showed that the AUCs of the training and verification groups were 0.97 (95%CI, 0.92 - 1.00) and 0.92 (95%CI, 0.83 - 1.00), respectively, indicating that the prediction model could differentiate IAC and MIA presenting as lung GGN non-invasively. This study established a pure radiomics labeling model, and the differentiation objects were IAC and MIA with more similar pathological features. The prediction efficacy was similar to that of individualized prediction models, which might be related to the fact that this study collected cases scanned with the same CT machine and the same scanning protocol, and pre-processed all images before outlining ROIs, which reduced the influences of equipment and scanning parameters on the results to some extent.

Chae et al. (33) used an artificial neural network to establish a radiomics prediction model to distinguish pre-invasive lesions (AAH/AIS) from invasive lesions (MIA/IAC) of mGGN lung adenocarcinoma and achieved good results (AUC=0.981). Although an artificial neural network can solve dichotomy problems well, its generalization ability in specific models is limited due to potential over-fitting and a complex structure.

Random forest is an algorithm that integrates multiple decision trees through the idea of ensemble learning, and it essentially belongs to the ensemble learning method in machine learning. It can handle both continuous data and discrete data and can also handle missing data and sort the importance of variables. The results of random forest models have higher accuracy and generalization performance; thus, they are often used to predict the risk of diseases and susceptibility of patients (39).

4.4 Limitations

This study had some limitations. First, owing to the retrospective nature of the study, there is potential bias; thus prospective studies are required to confirm our results. Second, the manual sketching of ROIs made it difficult to eliminate bronchi and blood vessels in the nodules; thus, the accuracy of some features might be affected. Third, the boundary between some nodules and normal lung tissue was unclear, and boundary leakage might occur during image segmentation. Moreover, similar to some previous studies (22, 35, 36, 40), this study divides the lesions are divided into train and test sets – this leads to the possibility of overfitting as two lesions from the same patient may end up in different subsets. Finally, the sample size was too small, and the prediction accuracy of the model might be unstable to some extent. Therefore, future studies with larger sample sizes are warranted.

5 CONCLUSION

The radiomics prediction model established by combining random forest with hyperparameter tuning could effectively differentiate IAC and MIA presenting as lung GGN and could provide a noninvasive, low-cost, rapid, and reproducible preoperative prediction method that is clinically applicable.

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 human participants were reviewed and approved by Dongyang Hospital ethics review board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Z-HH and X-MY designed the study. F-HZ and H-JF collected and analyzed the relevant literature. K-FS, LZ, Z-ZP, C-LF, Z-BY, M-KW, and J-HS analyzed the literature and constructed the figures. All authors participated in writing the manuscript and have approved the final version of the manuscript for submission.

REFERENCES

1. Austin JH, Müller NL, Friedman PJ, Hansell DM, Naidich DP, Remy-Jardin M, et al. Glossary of Terms for CT of the Lungs: Recommendations of the Nomenclature Committee of the Fleischner Society. *Radiology* (1996) 200:327–31. doi: 10.1148/radiology.200.2.8685321
2. Lee HY, Lee KS. Ground-Glass Opacity Nodules: Histopathology, Imaging Evaluation, and Clinical Implications. *J Thorac Imaging* (2011) 26:106–18. doi: 10.1097/RTI.0b013e3181fbaa64
3. Chang B, Hwang JH, Choi YH, Chung MP, Kim H, Kwon OJ, et al. Natural History of Pure Ground-Glass Opacity Lung Nodules Detected by Low-Dose CT Scan. *Chest* (2013) 143:172–8. doi: 10.1378/chest.11-2501
4. Kim HY, Shim YM, Lee KS, Han J, Yi CA, Kim YK. Persistent Pulmonary Nodular Ground-Glass Opacity at Thin-Section CT: Histopathologic Comparisons. *Radiology* (2007) 245:267–75. doi: 10.1148/radiol.2451061682
5. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, et al. International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society International Multidisciplinary Classification of Lung Adenocarcinoma. *J Thorac Oncol* (2011) 6:244–85. doi: 10.1097/JTO.0b013e318206a221
6. WHO Classification of Tumours Editorial Board. WHO Classification of Tumours. In: *Thoracic Tumours, 5th ed.* Lyon: IARC Press (2021). 1–565.
7. Soda H, Nakamura Y, Nakatomi K, Tomonaga N, Yamaguchi H, Nakano H, et al. Stepwise Progression From Ground-Glass Opacity Towards Invasive Adenocarcinoma: Long-Term Follow-Up of Radiological Findings. *Lung Cancer* (2008) 60:298–301. doi: 10.1016/j.lungcan.2007.09.001
8. Naidich DP, Bankier AA, MacMahon H, Schaefer-Prokop CM, Pistolesi M, Goo JM, et al. Recommendations for the Management of Subsolid Pulmonary Nodules Detected at CT: A Statement From the Fleischner Society. *Radiology* (2013) 266:304–17. doi: 10.1148/radiol.12120628
9. Boland JM, Froemming AT, Wampfler JA, Maldonado F, Peikert T, Hyland C, et al. Adenocarcinoma *in Situ*, Minimally Invasive Adenocarcinoma, and Invasive Pulmonary Adenocarcinoma- Analysis of Interobserver Agreement, Survival, Radiographic Characteristics, and Gross Pathology in 296 Nodules. *Hum Pathol* (2016) 5:41–50. doi: 10.1016/j.humpath.2015.12.010
10. Dembitzer FR, Flores RM, Parides MK, Beasley MB. Impact of Histologic Subtyping on Outcome in Lobar vs Sublobar Resections for Lung Cancer: A Pilot Study. *Chest* (2014) 146:175–81. doi: 10.1378/chest.13-2506
11. Van Schil PE, Asamura H, Rusch VW, Mitsudomi T, Tsuboi M, Brambilla E, et al. Surgical Implications of the New IASLC/ATS/ERS Adenocarcinoma Classification. *Eur Respir J* (2012) 39:478–86. doi: 10.1183/09031936.00027511
12. Macke RA, Schuchert MJ, Odell DD, Wilson DO, Luketich JD, Landreneau RJ. Parenchymal Preserving Anatomic Resections Result in Less Pulmonary Function Loss in Patients With Stage I non- Small Cell Lung Cancer. *J Cardiothorac Surg* (2015) 10:49. doi: 10.1186/s13019-015-0253-6
13. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: Extracting More Information From Medical Images Using Advanced Feature Analysis. *Eur J Cancer* (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036
14. Nelson DA, Tan TT, Rabson AB, Anderson D, Degenhardt K, White E. Hypoxia and Defective Apoptosis Drive Genomic Instability and Tumorigenesis. *Genes Dev* (2004) 18:2095–107. doi: 10.1101/gad.1204904
15. Ma J, Wang Q, Ren Y, Hu H, Zhao J. Automatic Lung Nodule Classification With Radiomics Approach. *Int Soc Optics Photonics* (2016) 20:789–96. doi: 10.1117/12.2220768
16. 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. *Radiother Oncol* (2016) 119:480–6. doi: 10.1016/j.radonc.2016.04.004

17. Yu W, Tang C, Hobbs BP, Li X, Koay EJ, Wistuba II, et al. Development and Validation of a Predictive Radiomics Model for Clinical Outcomes in Stage I Non-Small Cell Lung Cancer. *Int J Radiat Oncol Biol Phys* (2018) 102:1090–7. doi: 10.1016/j.ijrobp.2017.10.046
18. Weng Q, Zhou L, Wang H, Hui J, Chen M, Pang P, et al. A Radiomics Model for Determining the Invasiveness of Solitary Pulmonary Nodules That Manifest as Part-Solid Nodules. *Clin Radiol* (2019) 74:933–43. doi: 10.1016/j.crad.2019.07.026
19. Breiman L. Random Forests. *Mach Learn* (2001) 45:5–32. doi: 10.1023/A1010933404324
20. Wu G, Woodruff HC, Shen J, Refaee T, Sanduleanu S, Ibrahim A, et al. Diagnosis of Invasive Lung Adenocarcinoma Based on Chest CT Radiomic Features of Part-Solid Pulmonary Nodules: A Multicenter Study. *Radiology* (2020) 297:451–8. doi: 10.1148/radiol.2020192431
21. Wang X, Li Q, Cai J, Wang W, Xu P, Zhang Y, et al. Predicting the Invasiveness of Lung Adenocarcinomas Appearing as Ground-Glass Nodule on CT Scan Using Multi-Task Learning and Deep Radiomics. *Trans Lung Cancer Res* (2020) 9:1397. doi: 10.21037/tlcr-20-370
22. Sun Y, Li C, Jin L, Gao P, Zhao W, Ma W, et al. Radiomics for Lung Adenocarcinoma Manifesting as Pure Ground-Glass Nodules: Invasive Prediction. *Eur Radiol* (2020) 30:3650–9. doi: 10.1007/s00330-020-06776-y
23. Global Burden of Disease Cancer Collaboration, Fitzmaurice C, Allen C, Barber RM, Barregard L, Bhutta ZA, et al. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability- Adjusted Life-Years for 32 Cancer Groups, 1990 to 2015: A Systematic Analysis for the Global Burden of Disease Study. *JAMA Oncol* (2017) 3:524–48. doi: 10.1001/jamaoncol.2016.5688
24. Didkowska J, Wojciechowska U, Mańczuk M, Łobaszewski J. Lung Cancer Epidemiology: Contemporary and Future Challenges Worldwide. *Ann Transl Med* (2016) 4:150. doi: 10.21037/atm.2016.03.11
25. McGuire S. World Cancer Report 2014. Geneva, Switzerland: World Health Organization, International Agency for Research on Cancer, WHO Press. *Adv Nutr* (2016) 7:418–9. doi: 10.3945/an.116.012211
26. Song YS, Park CM, Park SJ, Lee SM, Jeon YK, Goo JM. Volume and Mass Doubling Times of Persistent Pulmonary Subsolid Nodules Detected in Patients Without Known Malignancy. *Radiology* (2014) 273:276–84. doi: 10.1148/radiol.14132324
27. National Lung Screening Trial Research Team. Reduced Lung-Cancer Mortality With Low-Dose Computed Tomographic Screening. *N Engl J Med* (2011) 365:395–409. doi: 10.1056/NEJMoa1102873
28. Castellano G, Bonilha L, Li LM, Cendes F. Texture Analysis of Medical Images. *Clin Radiol* (2004) 59:1061–9. doi: 10.1016/j.crad.2004.07.008
29. Nomori H, Ohtsuka T, Naruke T, Suemasu K. Differentiating Between Atypical Adenomatous Hyperplasia and Bronchioloalveolar Carcinoma Using the Computed Tomography Number Histogram. *Ann Thorac Surg* (2003) 76:867–71. doi: 10.1016/s0003-4975(03)00729-x
30. Ikeda K, Awai K, Mori T, Kawanaka K, Yamashita Y, Nomori H. Differential Diagnosis of Ground-Glass Opacity Nodules: CT Number Analysis by Three-Dimensional Computerized Quantification. *Chest* (2007) 132:984–90. doi: 10.1378/chest.07-0793
31. Haralick RM, Sternberg SR, Zhuang X. Image Analysis Using Mathematical Morphology. *IEEE Trans Pattern Anal Mach Intell* (1987) 9:532–50. doi: 10.1109/TPAMI.1987.4767941
32. Ganeshan B, Goh V, Mandeville HC, Ng QS, Hoskin PJ, Miles KA. Non–small Cell Lung Cancer: Histopathologic Correlates for Texture Parameters at CT. *Radiology* (2013) 266:326–36. doi: 10.1148/radiol.12112428
33. Chae HD, Park CM, Park SJ, Lee SM, Kim KG, Goo JM. Computerized Texture Analysis of Persistent Part-Solid Ground-Glass Nodules: Differentiation of Preinvasive Lesions From Invasive Pulmonary Adenocarcinomas. *Radiology* (2014) 273:285–93. doi: 10.1148/radiol.14132187
34. Yang B, Guo L, Lu G, Shan W, Duan L, Duan S. Radiomic Signature: A non-Invasive Biomarker for Discriminating Invasive and non-Invasive Cases of Lung Adenocarcinoma. *Cancer Manag Res* (2019) 11:7825–34. doi: 10.2147/CMAR.S217887
35. Xu F, Zhu W, Shen Y, Wang J, Xu R, Qutesh C, et al. Radiomic-Based Quantitative CT Analysis of Pure Ground-Glass Nodules to Predict the Invasiveness of Lung Adenocarcinoma. *Front Oncol* (2020) 10:872. doi: 10.3389/fonc.2020.00872
36. Wu L, Gao C, Xiang P, Zheng S, Pang P, Xu M. CT-Imaging Based Analysis of Invasive Lung Adenocarcinoma Presenting as Ground Glass Nodules Using Peri- and Intra-Nodular Radiomic Features. *Front Oncol* (2020) 10:838. doi: 10.3389/fonc.2020.00838
37. Fan L, Fang M, Li Z, Tu W, Wang S, Chen W, et al. Radiomics Signature: A Biomarker for the Preoperative Discrimination of Lung Invasive Adenocarcinoma Manifesting as a Ground-Glass Nodule. *Eur Radiol* (2019) 29:889–97. doi: 10.1007/s00330-018-5530-z
38. She Y, Zhang L, Zhu H, Dai C, Xie D, Xie H, et al. The Predictive Value of CT-Based Radiomics in Differentiating Indolent From Invasive Lung Adenocarcinoma in Patients With Pulmonary Nodules. *Eur Radiol* (2018) 28:5121–8. doi: 10.1007/s00330-018-5509-9
39. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine Learning for Medical Imaging. *Radiographics* (2017) 37:505–15. doi: 10.1148/rg.2017160130
40. Autrusseau PA, Labani A, De Marini P, Leyendecker P, Hintzpeter C, Ortlieb AC, et al. Radiomics in the Evaluation of Lung Nodules: Intrapatient Concordance Between Full-Dose and Ultra-Low-Dose Chest Computed Tomography. *Diagn Interventional Imaging* (2021) 102:233–9. doi: 10.1016/j.diii.2021.01.010

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.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 24 May 2022
doi: 10.3389/fonc.2022.906888

Prediction of Tumor Mutation Load in Colorectal Cancer Histopathological Images Based on Deep Learning

*Yongguang Liu[1], Kaimei Huang[2,3], Yachao Yang[4], Yan Wu[2,3] and Wei Gao[5]**

[1] Department of Anorectal Surgery, Weifang People's Hospital, Weifang, China, [2] Genies (Beijing) Co., Ltd., Beijing, China, [3] Qingdao Geneis Institute of Big Data Mining and Precision Medicine, Qingdao, China, [4] School of Mathematics and Statistics, Hainan Normal University, Haikou, China, [5] Department of Internal Medicine-Oncology, Fujian Cancer Hospital and Fujian Medical University Cancer Hospital, Fuzhou, China

Colorectal cancer (CRC) is one of the most prevalent malignancies, and immunotherapy can be applied to CRC patients of all ages, while its efficacy is uncertain. Tumor mutational burden (TMB) is important for predicting the effect of immunotherapy. Currently, whole-exome sequencing (WES) is a standard method to measure TMB, but it is costly and inefficient. Therefore, it is urgent to explore a method to assess TMB without WES to improve immunotherapy outcomes. In this study, we propose a deep learning method, DeepHE, based on the Residual Network (ResNet) model. On images of tissue, DeepHE can efficiently identify and analyze characteristics of tumor cells in CRC to predict the TMB. In our study, we used ×40 magnification images and grouped them by patients followed by thresholding at the 10th and 20th quantiles, which significantly improves the performance. Also, our model is superior compared with multiple models. In summary, deep learning methods can explore the association between histopathological images and genetic mutations, which will contribute to the precise treatment of CRC patients.

Keywords: immunotherapy, colorectal cancer, deep learning, tumor mutational burden, ResNet

OPEN ACCESS

Edited by:
Min Tang,
Jiangsu University, China

Reviewed by:
Xiangzheng Fu,
Hunan University, China
Youwen Zhang,
University of South Carolina, United States

****Correspondence:***
Wei Gao
13960986882@fjzlhospital.com

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *29 March 2022*
Accepted: *18 April 2022*
Published: *24 May 2022*

Citation:
Liu Y, Huang K, Yang Y, Wu Y and Gao W (2022) Prediction of Tumor Mutation Load in Colorectal Cancer Histopathological Images Based on Deep Learning. Front. Oncol. 12:906888. doi: 10.3389/fonc.2022.906888

INTRODUCTION

Colorectal cancer (CRC), including colon cancer and rectal cancer, is one of the top 3 malignant tumors in the world for morbidity and mortality (1–3). According to statistics from the American Cancer Society, the estimated death toll in 2021 even reached 149,500 (4). In China, 25% of patients experience metastasis during diagnosis or treatment, and the 5-year survival rate of patients is less than 5% (5). The treatment of CRC is mostly based on surgery and chemotherapy. However, because tumor cells grow rapidly and are prone to metastasis, surgery and chemotherapy can only temporarily relieve the disease but cannot completely cure it. Immunotherapy kills tumors by activating the host immune system that is anti-deteriorating and durable; this has become the focus of the cancer treatment field in the new era.

In tumor immunotherapy, programmed death receptor programmed cell death-1/programmed cell death-ligand 1 (PD-1/PD-L1) inhibitors and cytotoxic T lymphocyte-associated antigen 4 (CTLA-4) inhibitors are the main immune checkpoint inhibitors (ICIs) (6, 7). Several clinical studies have proven that compared with platinum-based chemotherapy and surgery, immunotherapy can improve the overall survival rate of patients in most cancers (8–14).

However, not all patients respond well in the clinical treatment with ICIs. Several studies have found that the efficacy of ICIs is closely related to the level of PD-L1 (15). So mastering the immune microenvironmental response of patients is a critical requirement. Previously, PD-L1 expression was the main marker for predicting the effect of immunotherapy. Several solid tumor studies have shown the effectiveness of PD-L1 expression detection, such as melanoma, CRC, and non-small cell lung cancer (NSCLC) (16, 17). However, as immunotherapy research continues to progress, the insufficient detection of PD-L1 expression has shown that it is no longer the only criterion to predict the efficacy of ICIs (18). In this regard, tumor mutational burden (TMB) appears as a marker of ICI efficacy identification and plays an irreplaceable role in optimizing targeted regimens and developing well-tolerated drugs for physicians.

The definition of TMB is the number of mutations per megabase in the coding region of the tumor exome. With an increase of the TMB, the degree of acquired somatic mutations will increase and more tumor-specific neoantigens will be released. A portion of the antigen is presented on the human leukocyte antigen (HLA) molecules on the surface of the cancer cells, thus triggering the recognition and processing of the tumor cells by the immune system. The reflection of TMB on PD-L1 levels affects the formulation of ICI regimens for patients in the clinic, which has aroused great interest in the determination of the TMB in tumors by researchers in various fields. According to current research, there is almost no correlation between TMB and PD-L1 expression in many cancers and their subtypes, including NSCLC, CRC, melanoma, and pancreatic cancer, which indicates that the TMB can serve as an independent prognostic marker (19, 20). Cao et al. (21) compiled the survival indicators of 103,078 patients with different cancers and included 45 immune-related studies; they finally found that TMB-H (high tumor mutational burden) patients achieved better survival after receiving immunotherapy. In 2020, the Food and Drug Administration (FDA) for first time approved TMB to be used as a diagnostic marker for pan-cancer immunotherapy when unresectability or metastasis occurs (22). In general, TMB-H has predictive and prognostic potential for the immunotherapy of solid tumors.

Although many studies have proven that the TMB performs well as a marker in ICI treatment, it is still hard to accurately measure and define the threshold of TMB-H (23). At present, whole-exome sequencing (WES) is the main method of TMB quantification that quantifies the TMB directly and comprehensively. WES data sets are often used in tumors to show the correlation between ICI reaction and TMB status (24, 25). Although this method can measure the TMB with high standards, it has some stringent requirements. For example, it not only requires fresh and high-quality samples but also is expensive and has a long working time (26). Therefore, low-cost targeted sequencing panel detection is often used as an alternative measurement to WES; it infers full mutation burden from a narrower sequencing space, leading to the development of an integrated MSK-IMPACT assay by Zehir et al. The assay can evenly cover clinically relevant genes and fusions of target genes, so the TMB content can be accurately estimated (27). However, targeted sequencing has some fatal drawbacks. Buchhalter et al. (28) evaluated the Illumina TSO500 panel and found that the 1.5–3-Mbp panel is more suitable for TMB estimation, and lower ranges will bring errors in TMB estimation, while targeted sequencing cannot detect small sequencing ranges and only targets tumor cells with repeated mutations. Deep learning has shown an ultrahigh level in processing complex and large amounts of information in histopathological images. Deep convolutional neural networks (CNN) have yielded many shocking research results in image feature recognition of cancer histopathology (29, 30). Moreover, Mika et al. developed the Image2TMB method with deep learning to measure the TMB in lung adenocarcinoma pathological tissue images at three scales (×5, ×10, and ×20 magnification) (31). Finally, the performance of the ×20 scale is the best with an area under the curve (AUC) of 0.81, showing that high-resolution scale images are more conducive to the prediction of markers. Therefore, there is indeed a correlation between the tumor somatic mutations and gene mutations, and deep learning can evaluate this correlation well. To further understand the capability of deep learning for tumor cell somatic recognition in histopathological images and to explore efficient strategies for TMB measurement, this paper builds a deep learning model, DeepHE. This can automatically analyze the TMB in pathological images and predict the probability of their potential TMB from CRC whole slices [whole-section images (WSIs)] in The Cancer Genome Atlas (TCGA). We downloaded the entire formalin-fixed paraffin-embedded (FFPE) tissue data of CRC at ×40 resolution and grouped them by patients. To train a more efficient model, a higher resolution compared to that of previous studies is chosen. Also, the classification by patients avoids errors caused by the same pathological tissue images from one patient. The images of 509 patients remained after DeepHE sorted and filtered data by patients, then we segmented the WSI slice and performed color normalization. To improve its feature recognition ability and speed up the convergence, this research introduced a residual network model derived from CNN technology and then trained the model of ResNet50 by 2-fold cross-validation. The performance of the five models including ResNet18, ResNet34, VGG16, AlexNet, and ResNet50 showed superiority. This study provides an important way for patients to benefit from ICI treatment and explores the relationship between the TMB and the tumor immune microenvironment.

RESULTS

The Workflow of DeepHE

Figure 1 shows the workflow of DeepHE. The ×40 scale images contain more details, and our model still performs well in predicting the TMB on CRC in a short time. From 611 patients, a total of 509 patients were left by professional pathologists, of which 12 patients were removed because they lacked TMB information or the tumor area could not be

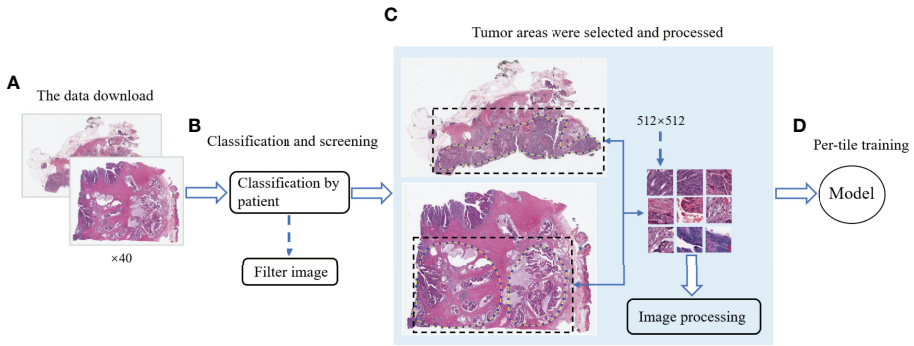


FIGURE 1 | The workflow of this study. **(A)** Download the CRC image data of ×40 resolution from TCGA. **(B)** Categorize the data by patients and remove the unqualified images. **(C)** Mark the tumor area and segment it, and then perform noise removal and color normalization processing. **(D)** Model training and testing.

annotated. The clinical information of the remaining 509 CRC patients was also obtained and collated from TCGA (**Table 1**). Then 1,586,826 qualified slices were derived from the images of these patients, which are approximately 3,117 slices in each group. After that, each group was randomly divided into two groups as the training set and validation set, resulting in two sets of 793,413 H&E slices for model training and testing.

DeepHE Achieves Relatively Good Performance in the Tumor Mutational Burden Prediction

According to the standard of dividing the TMB level in a previous related work, this study uses the thresholds of 10 and 20 (32). Then, the research divided the data into TMB-H (TMB high) and TMB-L (TMB low), satisfying H:L (10) = 83:426 (with a threshold of 10) and H:L (20) = 73:426 (with a threshold of 20), after which the model was trained and tested on the FFPE dataset. When the data set used a TMB threshold of 10, the ratio of the number of patients between TMB-H and TMB-L is 83:426.

The performances of DeepHE in predicting TMB at different thresholds were shown in **Figure 2**. ResNet50 with more hidden layers in the residual network is selected to capture a more detailed feature information. The results also showed that the AUC of ResNet50 reached 0.729 under 2-fold cross-validation. Then, when the threshold is 20, H:L (20) = 73:436, the AUC of ResNet50 is 0.774. During the whole trial, 30 epochs were maintained, which was the best value obtained after many attempts. At the beginning of the experiment, the epoch was set to 50. However, as the epoch was set greater than 30, the training ACC result has stabilized at around 0.9679, which means that a larger value will only take more time.

DeepHE Achieves Higher Areas Under the Curve Than Those of Existing Methods in the Tumor Mutational Burden Prediction

To verify the superiority of the model performance, we tested four models that have contributed greatly to the current image recognition field based on the same process, namely, ResNet18, ResNet34, VGG16, and AlexNet. The comparisons between their results and our model are shown in **Figure 3**. We used a sliding window to visualize the probability value on each small slice, classified and counted the TMB level on the slice contained in each complete WSI. When the ratio of the number of slices containing TMB-H in a patient to the number of all his eligible slices is greater than 50%, the patient was identified as TMB-H and *vice versa* for TMB-L. It is worth noting that the eligibility here refers to slices owned by the patient that were not screened out and participated in the training. **Figure 3A** shows the ROC curves of the TMB divided by 10 into all methods, and **Figure 3B** shows the case where 20 is the threshold. After training of ResNet18, the AUC is 0.720; at H:L (20), the AUC is 0.736. It can be found that the results of ResNet18 are very close to that of ResNet50, but there are still some gaps. We suspect that as the number of layers increased, the results would get closer to our model or even surpass it. As a result, we experimented with ResNet34. When the threshold is 10, the AUC value of the ROC curve is 0.716, and when H:L (20) = 73:426, the AUC is 0.715. Next, we tested VGG16 and AlexNet. When AlexNet was split at 20, the AUC only reached 0.685, so it was not tested with a threshold of 10. In the study, VGG16 made the TMB at H:L (10) = 83:426, the AUC value of the ROC curve was 0.677, H:L (20) = 73:426, and the AUC value was 0.701. The reason is that compared with AlexNet, VGG16 has more convolutional layers and smaller pooling kernels, which can extract more detailed information, but the model requires more parameters to participate, which will occupy more or a large memory space (33). Taken together, the results are that ResNet50 performs well, while AlexNet performs the least well. For ResNet, VGG, and AlexNet models, AlexNet has the least number of convolutional layers, which may be one of the reasons for its worst effect.

At the same threshold, the performances of all models are comparable. We perform statistics on performance metrics of all models used in the study in **Tables 2**, **3**, namely, accuracy (ACC), precision, recall, and F1 score. After statistics highlighting the optimal value in red, it was found that DeepHE based on ResNet50 has always maintained a relatively high performance.

TABLE 1 | Clinical Information for TCGA Colorectal Cancer Patients.

Clinical variable	Category	Number of patients
Tumor stage	Stage I	88
	Stage II	178
	Stage III	147
	Stage IV	75
	Unknown	21
Prior malignancy	Yes	55
	No	451
	Unknown	2
AJCC pathologic T	T1	15
	T2	91
	T3	341
	T4	58
	Unknown	4
AJCC pathologic N	N0	282
	N1	130
	N2	91
	Nx	2
	Unknown	4
AJCC pathologic M	M0	366
	M1	74
	Mx	58
	Unknown	10
Gender	Women1	246
	Men0	260
	Unknown	3
Vital status	Alive	395
	Dead	111
	Unknown	3
Age at index	≥66	272
	<66	237
New tumor event after initial treatment	Yes	91
	No	332
	Unknown	86

DISCUSSION

TMB has emerged as a biomarker responsive to the efficacy of immunotherapy and has been approved by the FDA. In 2018, Gandara et al. (34) for first time demonstrated that the TMB can stably predict the effect of immunotherapy. In this study, the content of the TMB was determined on pathological images of CRC. For the selection of the threshold, the researchers found that in patients with NSCLC, when the TMB ≥10 mut/Mb was used as the cutoff point, it was found that patients with high TMB content were more responsive to immunotherapy. ICI treatment prolonged the progression-free survival rate of these patients and far exceeded the effect of platinum-based doublet chemotherapy (35). Therefore, the ResNet50-based DeepHE had a threshold of 10 and 20, respectively. When the threshold was 10, the area under the ROC curve reached 0.729. At the same time, when the

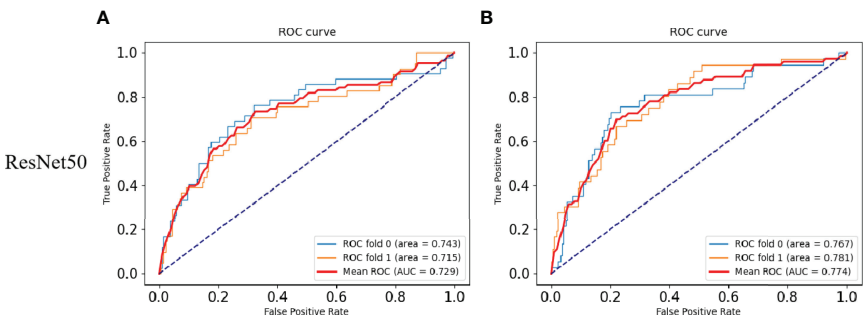


FIGURE 2 | Results of the TMB prediction model. **(A)** ROC plot of the ResNet50 model with a TMB cutoff of 10 under 2-fold cross-validation. **(B)** ROC plot of the ResNet50 model with a TMB cutoff of 20 under 2-fold cross-validation.

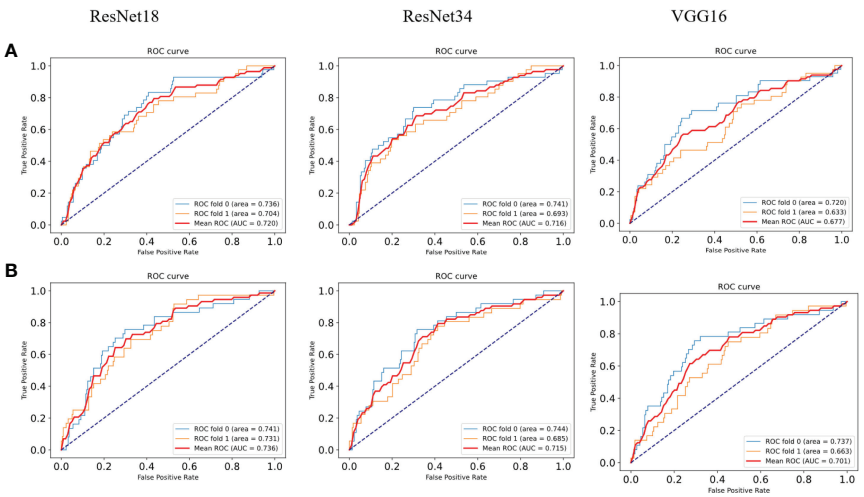


FIGURE 3 | ROC curves of the comparison models. **(A)** ROC plots for different models with a TMB cutoff of 10 under 2-fold cross-validation. **(B)** ROC plots for different models with a TMB cutoff of 20 under 2-fold cross-validation.

threshold was 20, the AUC of ResNet50 was 0.774. Ciardiello et al. (36) studied colon cancer with mismatch repair deficiency (dMMR), and high microsatellite instability (MSI-H) concluded that immunotherapy has a certain clinical therapeutic effect. The TMB has been reported as an important marker for concomitant CRC immunotherapy, which fully justifies our trial design. Furthermore, the addition of residual blocks further improves the performance of the model. Therefore, our research is bound to have an important impact on improving the survival rate of cancer patients. The detection of TMB mainly relies on WES technology, which is expensive, requires a lot of time (about 60 days), will delay the treatment time of cancer patients, and is likely to make the best treatment time missed. In contrast, the DeepHE method does not require a large amount of biopsy tissue sample, a lot of manpower, material resources, and time. It only needs to run the trained model on the pathological images of patients.

In recent years, machine learning methods have been widely used in biomedical research like drug repositioning (37, 38) and single-cell analysis (39). Among all of these fields, deep learning showed advantages over many previous related technologies (40, 41). For example, in lung cancer research, deep learning methods can be used to identify biomarker genes on pathological images (42, 43). The success of Residual Networks in the ImageNet Large Scale Visual Recognition Competition in 2015 brought the ResNet model into the limelight. When ResNet18 predicted MSI on H&E tissue section images of gastric adenocarcinoma (STAD) and CRC, it not only had a shorter training time but also achieved an AUC of 0.84 (44). Moreover, it has been reported that ResNet50 has demonstrated exciting performance results in breast cancer and skin cancer classification (45, 46). ResNet18 includes convolution layers and fully connected layers. Compared with ResNet50, ResNet18 lacks the reduction of the corresponding dimension and the function given by the batch norm (BN) layer, which may be the reason why the performance of ResNet18 in predicting the TMB is slightly lower than that of ResNet50 (47). ResNet34, VGG16, and AlexNet can be regarded as the classic models in the deep CNN. After AlexNet was proposed in 2012, it has triggered a boom in its application and plays an important role in the research of medical images. Notably, the structure of VGG16 is very simple. However, the number of VGG network channels is too large, and its structure determines that it requires more parameters and brings more memory usage (48). In addition, the VGG16 network structure is too densely connected, resulting in a long training time, these factors may lead to the relatively poor effect of VGG16 in this study. Currently, the research of deep learning on medical images is quite mature, and most of its achievements have also been recognized and practically applied in clinical practice.

In our research, we used deep learning to identify and analyze CRC histopathological images and achieved the purpose of predicting the TMB. However, the content of the TMB in

TABLE 2 | Comparison of the performance of different models (TMB cutoff = 10).

Model	ACC	Precision	Recall	F1-score
ResNet18	0.820	0.640	0.562	0.575
ResNet34	0.815	0.607	0.548	0.552
ResNet50	0.830	0.681	0.587	0.605
VGG16	0.830	0.681	0.587	0.605

TABLE 3 | Comparison of performance of different models (TMB cutoff = 20).

Model	ACC	Precision	Recall	F1-score
ResNet18	0.835	0.622	0.564	0.575
ResNet34	0.845	0.649	0.548	0.555
ResNet50	0.850	0.677	0.567	0.582
VGG16	0.845	0.620	0.530	0.530
AlexNet	0.840	0.643	0.563	0.575

tumors was seldom estimated using deep learning previously. The reason may be the difference between gene level and phenotype level. DeepHE divides the complete WSI into 512 × 512 H&E slices and predicts the TMB probability. Our results, shown in **Figure 2**, illustrate the promise of exploring associations between cancer genotypes and phenotypes.

To save the detection time of the TMB, researchers have also explored single-panel sequencing methods and more clinically practical panel-based methods. However, many factors like the protein-coding regions of the panel, the association of selected genes with tumors, and the number of genes will affect the ACC of the results (49). DeepHE extracts the potential features of TMB on the pathological images of tumors through deep learning, which does not depend on the selection of genes. Therefore, these shortcomings of the gene panel do not exist for the DeepHE method.

Although DeepHE showed good performance, it was still controversial in many aspects. Firstly, our samples have been delineated and annotated by professional pathologists, and the pathological images have been segmented and screened. The addition of professional pathologists makes this study more professional in medicine and more promising for clinical use. These steps are not required by some TMB detection methods, but those methods still achieved good results. Second, our data all came from FFPE images in TCGA, and no independent validation set was established. Moreover, much complex information and noise in these images cannot be completely analyzed and removed that could affect the ACC and persuasiveness of DeepHE.

MATERIALS AND METHODS

Data Sources

TCGA (https://www.cancer.gov/tcga) is a joint cancer multi-omics analysis database co-founded by the National Cancer Institute and the National Human Genome Research Institute in 2006. From TCGA database, we have collected all available FFPE images of CRC patients; this type of images is often used for clinical diagnostic analysis and is stored simply. This manner does not affect the pathology contained in FFPE tissue, thus guaranteeing the ACC of our model (50). The data resolution was chosen as ×40 and submitted to professional pathologists in SVS format. The TMB distribution of patients is known and published in TCGA, and 20 or 10 were used as the cut points to categorize patients. Patients whose TMB content is higher than the threshold were marked as 1 and recorded as TMB-H, while those lower than the threshold are marked as 0 and recorded as TMB-L.

Data Processing

In this study, CRC images were downloaded from TCGA and were classified by patients. There are 611 sets of image data in total. According to the diagnosis of pathologists, 6 groups of data have unclear tumor areas in the images, and 96 patients were excluded because the TMB information is missing. WSIs of 509 CRC patients were used for model training in this study. There are thousands of pixels in a WSI, which often contains too much complex information and is not conducive to the analysis of the features on these images by DeepHE. The DeepHE model divided the patient's WSI scan into H&E slices of 512 × 512 pixels. These slices were used for subsequent training, as shown in **Figure 1C**.

Noise information, blurred areas, and blank areas in H&E slices have a non-negligible impact on model training, such as false-positive results and capture feature deviations; then, that information must be paid attention to. As an open-source computer vision library for image processing, OpenCV has powerful and reliable image processing capabilities to reduce the research cost and time of researchers (51). Based on OpenCV, this research regarded H&E slices as pixel matrices and performs segmentation, signature detection, and noise removal for specified targets. After reading the slice data, OpenCV was used to calculate the ratio of the number of blank area pixels in the slice to the slice area, and 70% was used as the threshold to screen samples suitable for predicting target genes. Denote K_0 as the real value of the pixel in the image, and the noise pixel as L, then $K = K_0 + L$. OpenCV collected many pixels in the image and calculated the average value to make the value of L tend to 0, then used the average values to represent the new pixel values that achieve smooth filtering and noise elimination. In addition, OpenCV also shows the functions of image edge expansion (filling), highlighting important parts of the image and adjusting brightness to improve image quality, as shown in **Figure 4**.

In H&E slices of CRC, some images inevitably contain more microvessels, inflammatory cells, microfibrils in the background, and are wrinkled and blurry. Once these images were identified, they were abandoned. After a series of program operations and careful screening, a total of 1,586,826 CRC H&E slices remained. The high quality of these slices ensures the ACC and reliability of the results of the TMB prediction model.

Hematoxylin and eosin staining are commonly used staining methods in FFPE sectioning technology. Hematoxylin can differentiate cell structures into various colors, while eosin

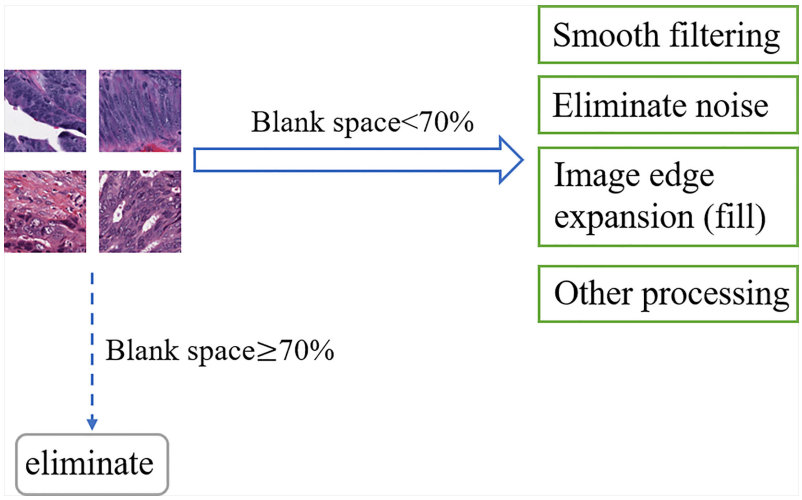


FIGURE 4 | OpenCV process.

stains the cytoplasm and intercellular substance. There are often differences in the color of each structure. And there are many reasons for color differences, including temperature, solution dose, tissue or cell type, changes in cell cycle, and histopathology (52). Therefore, in the study, H&E slices often show different colors on the same structure, which increases the difficulty of DeepHE in capturing the target information during the training process. To address this issue, we incorporated a color normalization method into the development of the DeepHE model. The color normalization employs an unsupervised deep convolutional Gaussian mixture model (DCGMM) to identify color information in H&E tissue slice images and converts them into a reference image, as shown in **Figure 5**. The color normalization method only transforms the chromaticity of H&E images; the spatial structure and pathological information on it do not change. This method does not require labels and premise assumptions; it also has the capability of automatic learning (53).

The Gaussian mixture model (GMM) can be regarded as a linear group sum of multiple Gaussian functions. Using GMM, the number of clusters n can be specified, and the random variable data are f_n. The training of GMM is to cluster points according to the distance between two different pixels to maximize the expectation:

$$\beta(n) = \sqrt{(f_n - \theta_n)^T \sum_n^{-1} (f_n - \theta_n)} \tag{4.1}$$

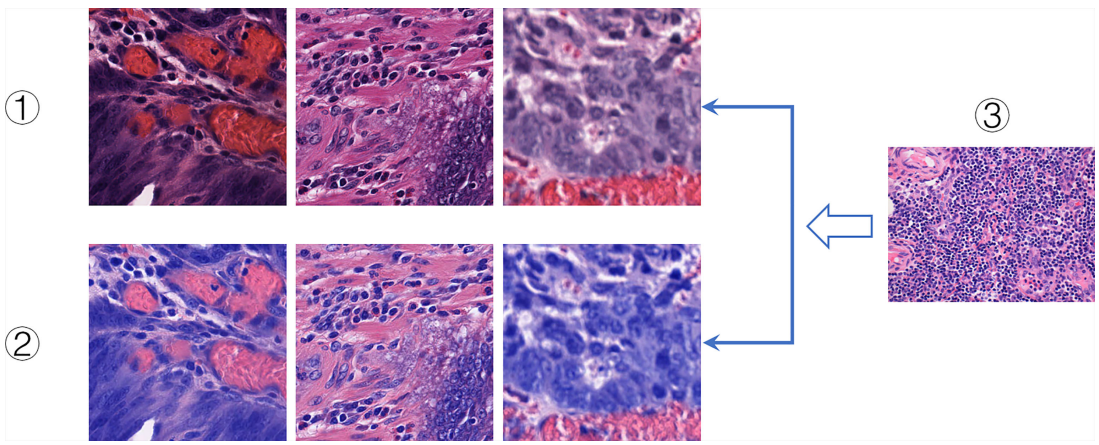


FIGURE 5 | Color normalization. I. Original slice images. II. Color-normalized slice images. III. Reference images.

where θ_n represents the mean, and the process is very similar to the k-means algorithm.
When the normal distribution is ∂, the n-nt data are $\partial(f_n|\theta_n, \gamma_n)$, and GMM satisfies:

$$\alpha(n) = \sum_{n=1}^{N} \lambda_n \partial(f_n|\theta_n, \gamma_n) \tag{4.2}$$

where γ_n is the covariance matrix $\sum_{n=1}^{N} \lambda_n$ is the weight of $\partial(f_n|\theta_n, \gamma_n)$ which satisfies the condition $\sum_{n=1}^{N} W_n = 1$ (54).

The DCGMM model is a probability distribution model based on GMM. The model is generated by linear superposition of an N-dimensional GMM. When λ_n is the prior condition of f_n and the data are n, it satisfies:

$$P(n) = \frac{\lambda_n \partial(f_n|\theta_n, \gamma_n)}{\sum_{i=1}^{N} \lambda_j \partial(f_n|\theta_n, \gamma_n)} \tag{4.3}$$

When f_n is the submodel data, all submodels can finally form a DCGMM model whose (natural) log-likelihood function is:

$$\ln P(x_k|(\lambda_n, \theta_n, \gamma_n)) = \sum_{k=1}^{K} \ln P(x) \tag{4.4}$$

At this time, the selection and change of the parameter (λ_n, θ_n, γ_n) have a decisive effect on the effect of the DCGMM model.

Deep Learning Algorithms

ResNet50 is one of the methods in ResNet. It contains two basic blocks, Conv Block and Identity Block. Usually, Bottlenecks is included in the four blocks, and the number of channels is reduced by a 1 × 1 convolutional layer to half, followed by 3 × 3 and a 1 × 1 convolution to achieve dimensionality reduction and pooling of these images, reducing the amount of subsequent computation and outputs to the next block. Identity Block does not change the dimension of the data itself; it performs the mapping of the data itself. As a result, the network can be extended to a deeper level, which will make the model feature extraction better and improve the model's classification ACC of image features. The nature of the residual block skip link reduces the training time, which is shown in **Figure 6**. After the data are output, they can linearly reach the input layer of the following block through the skip link, so that the network only needs to learn the differential information between the input layer and the output layer. Compared with traditional CNN, the introduction of ResNet reduces the loss of information and optimizes the model generalization ability and training speed (55).

In this study, ResNet50 is mainly used to form the neural network part of the DeepHE model, and Conv is used as the convolution layer. After the image data x is input, it will first enter the convolution layer with 32n×n kernels and perform feature extraction by weight. Conv Block will change the network dimensions, so they cannot be directly connected in series. Then, the BN layer is added to normalize the Conv results, smooth the landscape of the entire loss function, and improve the feature extraction accuracy and generalization ability of the network (56).

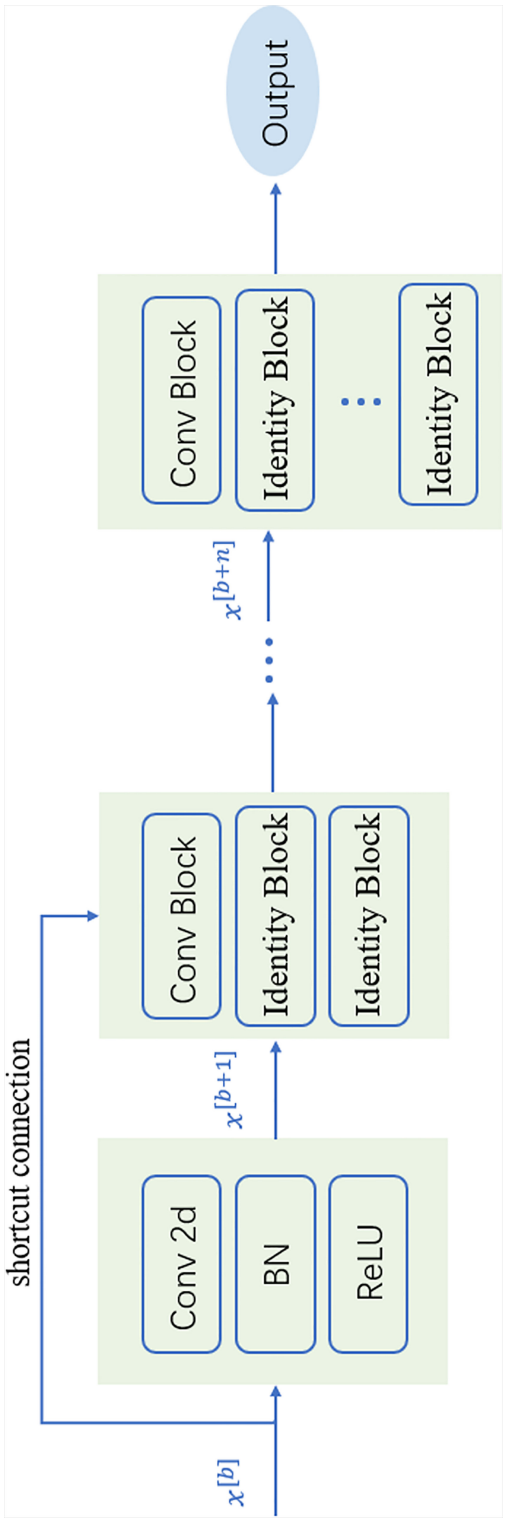


FIGURE 6 | Residual network.

We choose ReLU as the activation function, which can be regarded as an identity mapping model for forwarding calculation. This makes the network sparse and at the same time acts as a regularization to realize the repeated comparison of extracted information and feature confirmation.

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

WG provided the idea of the thesis. YL and KH performed the study and experiments. YW supervised the experiments. YY and WG wrote the article. YL modified and reviewed the article. All authors contributed to the interpretation of data and to the writing and revision of the article. All authors contributed to the article and approved the submitted version.

FUNDING

The study was funded by the Joint Funds for the Innovation of Science and Technology, Fujian province (Grant No. 2019Y9038).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.906888/full#supplementary-material

REFERENCES

1. Snf~ a, Miljua D, Jovanovi V. Registration of Malignant Diseases in Estimating Global Cancer Burden. *Glasnik Javnog Zdravlja* (2021) 95 (1):73–84. doi: 10.5937/gjz2101073q
2. Hong J, Lin X, Hu X, Wu X, Fang W. A Five-Gene Signature for Predicting the Prognosis of Colorectal Cancer. *Curr Gene Ther* (2021) 21(4):280–9. doi: 10.2174/1566523220666201012151803
3. Liu H, Qiu C, Wang B, Bing P, Tian G, Zhang X, et al. Evaluating DNA Methylation, Gene Expression, Somatic Mutation, and Their Combinations in Inferring Tumor Tissue-of-Origin. *Front Cell Dev Biol* (2021) 9:619330. doi: 10.3389/fcell.2021.619330
4. Viale PH. The American Cancer Society' s Facts & Figures: 2020 Edition. *J Adv Practit Oncol* (2020) 11:135–6. doi: 10.6004/jadpro.2020.11.2.1
5. Tepfa B, Mlakar DN, Stefanovj M, et al. The Impact of V Years of the National Colorectal Cancer Screening Program on Colorectal Cancer Incidence and 5-Year Survival. *Eur J Cancer Prev* (2020) 30(4):304–10. doi: 10.1097/CEJ.0000000000000628
6. Hanahan D, Weinberg RA. Hallmarks of Cancer: The Next Generation. *Cell* (2011) 144:646–74. doi: 10.1016/j.cell.2011.02.013
7. McQuade JL, Daniel CR, Hess KR, Mak C, Wang DY, Rai RR, et al. Association of Body-Mass Index and Outcomes in Patients With Metastatic Melanoma Treated With Targeted Therapy, Immunotherapy, or Chemotherapy: A Retrospective, Multicohort Analysis. *Lancet Oncol* (2018) 19(3):310–22. doi: 10.1016/S1470-2045(18)30078-0
8. Reck M, Rodríguez-Abreu D, Robinson AG, et al. Pembrolizumab Versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer. *New Engl J Med* (2016) 375 19:1823–33. doi: 10.1056/NEJMoa1606774
9. Taube JM, Klein AP, Brahmer JR, et al. Association of PD-1, PD-1 Ligands, and Other Features of the Tumor Immune Microenvironment With Response to Ant ‰ PD-1 Therapy. *Clin Cancer Res* (2014) 20:5064–74. doi: 10.1158/1078-0432.CCR-13-3271
10. He W, Li Q, Lu Y, Ju D, Gu Y, Zhao K, et al. Cancer Treatment Evolution From Traditional Methods to Stem Cells and Gene Therapy. *Curr Gene Ther* (2021). doi: 10.2174/1566523221666211119110755
11. Sindhu RK, Madaan P, Chandel P, Akter R, Adilakshmi G, Rahman MH. Therapeutic Approaches for the Management of Autoimmune Disorders *via* Gene Therapy: Prospects, Challenges, and Opportunities. *Curr Gene Ther* (2021). doi: 10.2174/1566523221666210916113609
12. Zhao Y, Zhao Z, Wang Y, Zhou Y, Ma Y, Zuo W. Single-Cell RNA Expression Profiling of ACE2,the Receptor of SARS-COV-2. *Am J Respir Crit Care Med* (2020) 202(5):756–9. doi: 10.1101/2020.01.26.919985
13. Yang J, Hui Y, Zhang Y, Zhang M, Ji B, Tian G, et al. Application of Circulating Tumor DNA as a Biomarker for Non-Small Cell Lung Cancer. *Front Oncol* (2021) 11:725938. doi: 10.3389/fonc.2021.725938
14. Song Z, Chen X, Shi Y, Huang R, Wang W, Zhu K, et al. Evaluating the Potential of T Cell Receptor Repertoires in Predicting the Prognosis of Resectable Non-Small Cell Lung Cancers. *Mol Ther Methods Clin Dev* (2020) 18:73–83. doi: 10.1016/j.omtm.2020.05.020
15. Lin H, Wei S, Hurt EM, Green MD, Zhao L, Vatan L, et al. Amendment History: Erratum (April 2018) Host Expression of PD-L1 Determines Efficacy of PD-L 1 Pathway Blockade-Mediated Tumor Regression. *J Clin Invest* (2018) 128(2):80515. doi: 10.1172/JCI96113
16. Brahmer JR, Drake CG, Wollner IS, Powderly JD, Picus J, Sharfman WH, et al. Phase I Study of Single-Agent Anti-Programmed Death-1 (MDX-1106) in Refractory Solid Tumors: Safety, Clinical Activity, Pharmacodynamics, and Immunologic Correlates. *J Clin Oncol* (2010) 28 19:3167–75. doi: 10.1200/JCO.2009.26.7609
17. Sun J, Zheng Y, Mamun M, Xiaojing L, Xiaoping C, Yongshun G, et al. Research Progress of PD-1/PD-L1 Immunotherapy in Gastrointestinal Tumors. *Biomed Pharmacother* (2020) 129(2020):1–7. doi: 10.1016/j.biopha.2020.110504
18. Rizvi HA, Sánchez-Vega F, La KC, Chatila W, Jonsson P, Halpenny D, et al. Molecular Determinants of Response to Anti-Programmed Cell Death (PD)-1 and Anti-Programmed Death-Ligand 1 (PD-L1) Blockade in Patients With Non-Small-Cell Lung Cancer Profiled With Targeted Next-Generation Sequencing. *J Clin Oncol* (2018) 36(7):633–41. doi: 10.1200/JCO.2017.75.3384
19. Yarchoan M, Albacker LA, Hopkins AC, Montesion M, Murugesan K, Vithayathil TT, et al. PD-L1 Expression and Tumor Mutational Burden Are Independent Biomarkers in Most Cancers. *JCI Insight* (2019) 4(6):e126908. doi: 10.1172/jci.insight.126908
20. Yarchoan M, Hopkins AC, Jaffee EM. Tumor Mutational Burden and Response Rate to PD-1 Inhibition. *N Engl J Med* (2017) 377(25):2500–1. doi: 10.1056/NEJMc1713444
21. Cao D, Xu H, Xu X, Guo T, Ge W. High Tumor Mutation Burden Predicts Better Efficacy of Immunotherapy: A Pooled Analysis of 103078 Cancer Patients. *OncoImmunology* (2019) 8(9):e1629258. doi: 10.1080/2162402X.2019.1629258
22. Fridland S, Chae YK. 71 Tumors With Higher Heterogeneity Were Associated With Superior Survival Outcome Amongst Stage I Lung Cancer Patients With Low Tumor Mutational Burden (TMB). *J ImmunoTher Cancer* (2021). doi: 10.1136/jitc-2021-SITC2021.071
23. McGrail DJ, Pilié PG, Rashid NU, Voorwerk L, Slagter M, Kok M, et al. High Tumor Mutation Burden Fails to Predict Immune Checkpoint Blockade Response Across All Cancer Types. *Ann Oncol* (2021) 32(5):661–72. doi: 10.1016/j.annonc.2021.02.006

24. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Mutational Landscape Determines Sensitivity to PD-1 Blockade in Non-Small Cell Lung Cancer. *Science* (2015) 348(6230):124–8. doi: 10.1126/science.aaa1348
25. Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, et al. Genetic Basis for Clinical Response to CTLA-4 Blockade in Melanoma. *N Engl J Med* (2014) 371(23):2189–99. doi: 10.1056/NEJMoa1406498
26. Li R, Han D, Shi J, Han Y, Tan P, Zhang R, et al. Choosing Tumor Mutational Burden Wisely for Immunotherapy: A Hard Road to Explore. *Biochim Biophys Acta Rev Cancer* (2020) 1874(2):188420. doi: 10.1016/j.bbcan.2020.188420
27. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, et al. Mutational Landscape of Metastatic Cancer Revealed From Prospective Clinical Sequencing of 10,000 Patients. *Nat Med* (2017) 23:703–13. doi: 10.1038/nm.4333
28. Buchhalter I, Rempel E, Endris V, Allgäuer M, Neumann O, Volckmar AL, et al. Size Matters: Dissecting Key Parameters for paneŒ Based Tumor Mutational Burden Analysis. *Int J Cancer* (2019) 144(4):848–58. doi: 10.1002/ijc.31878
29. 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
30. 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
31. Jain MS, Massoud TF. Predicting Tumour Mutational Burden From Histopathological Images Using Multiscale Deep Learning. *Nat Mach Intell* (2020) (2):356–62. doi: 10.1101/2020.06.15.153379
32. Hsiehchen D, Espinoza M, Valero C, Ahn C, Morris LGT. Impact of Tumor Mutational Burden on Checkpoint Inhibitor Drug Eligibility and Outcomes Across Racial Groups. *J Immunother Cancer* (2021) 9(11):e003683. doi: 10.1136/jitc-2021-003683
33. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR* (2015) arXiv: 1409, 1556v6. doi: 10.48550/arXiv.1409.1556.
34. Gandara DR, Paul SM, Kowanetz M, Schleifman E, Zou W, Li Y, et al. Blood-Based Tumor Mutational Burden as a Predictor of Clinical Benefit in Non-Small-Cell Lung Cancer Patients Treated With Atezolizumab. *Nat Med* (2018) 24:1441–8. doi: 10.1038/s41591-018-0134-3
35. Hellmann MD, Ciuleanu TE, Pluzanski A, Lee JS, Otterson GA, Audigier-Valette C, et al. Nivolumab Plus Ipilimumab in Lung Cancer With a High Tumor Mutational Burden. *N Engl J Med* (2018) 378:2093–104. doi: 10.1056/NEJMoa1801946
36. Ciardiello D, Vitiello PP, Cardone C, Martini G, Troiani T, Martinelli E, et al. Immunotherapy of Colorectal Cancer: Challenges for Therapeutic Efficacy. *Cancer Treat Rev* (2019) 76:22–32. doi: 10.1016/j.ctrv.2019.04.003
37. Yang J, Peng S, Zhang B, Houten S, Schadt E, Zhu J, et al. Human Geroprotector Discovery by Targeting the Converging Subnetworks of Aging and Age-Related Diseases. *Geroscience* (2020) 42(1):353–72. doi: 10.1007/s11357-019-00106-x
38. Tang X, Cai L, Meng Y, Xu J, Lu C, Yang J. Indicator Regularized Non-Negative Matrix Factorization Method-Based Drug Repurposing for COVID-19. *Front Immunol* (2020) 11:603615. doi: 10.3389/fimmu.2020.603615
39. Xu J, Cai L, Liao B, Zhu W, Yang J. CMF-Impute: An Accurate Imputation Tool for Single-Cell RNA-Seq Data. *Bioinformatics* (2020) 36(10):3139–47. doi: 10.1093/bioinformatics/btaa109
40. Zhao T, Hu Y, Peng J, Cheng L. DeepLGP: A Novel Deep Learning Method for Prioritizing lncRNA Target Genes. *Bioinformatics* (2020) 36(16):4466–72. doi: 10.1093/bioinformatics/btaa428
41. Meng Y, Lu C, Jin M, Xu J, Zeng X, Yang J. A Weighted Bilinear Neural Collaborative Filtering Approach for Drug Repositioning. *Brief Bioinform* (2022) 23(2):bbab581. doi: 10.1093/bib/bbab581
42. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-Level Classification of Skin Cancer With Deep Neural Networks. *Nature* (2017) 542:115–8. doi: 10.1038/nature21056
43. 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(2):256–64. doi: 10.1016/j.eururo.2020.04.023
44. 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) 1–3. doi: 10.1038/s41591-019-0462-y
45. Abd ElGhany S, Ramadan Ibraheem M, Alruwaili M, Elmogy M. Diagnosis of Various Skin Cancer Lesions Based on Fine-Tuned ResNet50 Deep Network. *Cmc-computers Mat Continua* (2021) 68(1):117–35. doi: 10.32604/cmc.2021.016102
46. Raihan M, Suryanegara M. Classification of COVID-19 Patients Using Deep Learning Architecture of InceptionV3 and 2021 4th International Conference of Computer and Informatics Engineering (IC2IE). (2021). pp: 46–50. doi: 10.1109/IC2IE53219.2021.9649255
47. Miao H, Zeng Q, Xu S, Chen Z. miR-1-3p/CELSR3 Participates in Regulating Malignant Phenotypes of Lung Adenocarcinoma Cells. *Curr Gene Ther* (2021) 21(4):304–12. doi: 10.2174/1566523221666210617160611
48. Xu P, Zhao J, Zhang J. Identification of Intrinsically Disordered Protein Regions Based on Deep Neural Network-Vgg16. *Algorithms* (2021) 14:107. doi: 10.3390/a14040107
49. Endris V, Buchhalter I, Allgäuer M, Rempel E, Lier A, Volckmar AL, et al. Measurement of Tumor Mutational Burden (TMB) in Routine Molecular Diagnostics: *In Silico* and reaŒ Life Analysis of Three Larger Gene Panels. *Int J Cancer* (2019) 144(9):2303–12. doi: 10.1002/ijc.32002
50. Grillo F, Bruzzone M, Pigozzi S, Prosapio S, Migliora P, Fiocca R, et al. Immunohistochemistry on Old Archival Paraffin Blocks: Is There an Expiry Date? *J Clin Pathol* (2017) 70(11):988–93. doi: 10.1136/jclinpath-2017-204387
51. Zhi-kun C. Realization of Image Process Technology for Computer Vision Based on OpenCV. *Mach Electron* (2010) 512(6):54–7.
52. Bejnordi BE, Litjens G, Timofeeva N, Otte-Höller I, Homeyer A, Karssemeijer N, et al. Stain Specific Standardization of Whole-Slide Histopathological Images. *IEEE Trans Med Imaging* (2016) 35:404–15. doi: 10.1109/TMI.2015.2476509
53. Zanjani FG, Zinger S, Bejnordi BE, van der Laak JAWM, de With PHN. Histopathology Stain-Color Normalization Using Deep Generative Models. In *1st Conference on Medical Imaging with Deep Learning (MIDL)* (2018). pp: 1–11.
54. Neal RM. Pattern Recognition and Machine Learning. *Technometrics* (2007) 49:366. doi: 10.1198/tech.2007.s518
55. Chan H-P, Samala RK, Hadjiiski LM, Zhou C.. Deep Learning in Medical Image Analysis. *Adv Exp Med Biol* (2020) 1213:3–21. doi: 10.1007/978-3-030-33128-3_1
56. He K, Zhang X, Ren S, Sun J.. Deep Residual Learning for Image Recognition 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016). pp. 770–8. doi: 10.1109/CVPR.2016.90.

Conflict of Interest: KH and YW are employed by Genies (Beijing) Co., Ltd.

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

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

frontiers | Frontiers in Oncology

METHODS
published: 15 July 2022
doi: 10.3389/fonc.2022.905955

Ultrasound Image Classification of Thyroid Nodules Based on Deep Learning

Jingya Yang[1,2†], Xiaoli Shi[2,3†], Bing Wang[1], Wenjing Qiu[1,2], Geng Tian[2,3], Xudong Wang[1], Peizhen Wang[1] and Jiasheng Yang[1*]*

[1] School of Electrical & Information Engineering, Anhui University of Technology, Ma'anshan, China, [2] Scientific System, Geneis Beijing Co., Ltd., Beijing, China, [3] Qingdao Genesis Institute of Big Data Mining and Precision Medicine, Qingdao, China

A thyroid nodule, which is defined as abnormal growth of thyroid cells, indicates excessive iodine intake, thyroid degeneration, inflammation, and other diseases. Although thyroid nodules are always non-malignant, the malignancy likelihood of a thyroid nodule grows steadily every year. In order to reduce the burden on doctors and avoid unnecessary fine needle aspiration (FNA) and surgical resection, various studies have been done to diagnose thyroid nodules through deep-learning-based image recognition analysis. In this study, to predict the benign and malignant thyroid nodules accurately, a novel deep learning framework is proposed. Five hundred eight ultrasound images were collected from the Third Hospital of Hebei Medical University in China for model training and validation. First, a ResNet18 model, pretrained on ImageNet, was trained by an ultrasound image dataset, and a random sampling of training dataset was applied 10 times to avoid accidental errors. The results show that our model has a good performance, the average area under curve (AUC) of 10 times is 0.997, the average accuracy is 0.984, the average recall is 0.978, the average precision is 0.939, and the average F1 score is 0.957. Second, Gradient-weighted Class Activation Mapping (Grad-CAM) was proposed to highlight sensitive regions in an ultrasound image during the learning process. Grad-CAM is able to extract the sensitive regions and analyze their shape features. Based on the results, there are obvious differences between benign and malignant thyroid nodules; therefore, shape features of the sensitive regions are helpful in diagnosis to a great extent. Overall, the proposed model demonstrated the feasibility of employing deep learning and ultrasound images to estimate benign and malignant thyroid nodules.

Keywords: thyroid nodule, ultrasound images, deep learning, convolutional neural network, Grad-CAM, feature extraction

OPEN ACCESS

Edited by:
Min Tang,
Jiangsu University, China

Reviewed by:
Weinan Zhou,
University of Illinois at Urbana-Champaign, United States
Wang Xuefeng,
Chinese Academy of Forestry, China

***Correspondence:**
Peizhen Wang
pzhwang@ahut.edu.cn

Jiasheng Yang
jsyang.mcc@gmail.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *28 March 2022*
Accepted: *22 June 2022*
Published: *15 July 2022*

Citation:
Yang J, Shi X, Wang B, Qiu W, Tian G, Wang X, Wang P and Yang J (2022) Ultrasound Image Classification of Thyroid Nodules Based on Deep Learning. Front. Oncol. 12:905955. doi: 10.3389/fonc.2022.905955

INTRODUCTION

Thyroid nodules can be divided into benign and malignant nodules. According to research, the incidence of thyroid cancer has increased by 2.4 times in the past 30 years, which is one of the fastest-growing malignant tumors (1–4).

There are two ways to diagnose thyroid nodules, fine needle aspiration (FNA) and ultrasound. FNA is the gold standard for thyroid nodule detection, but it is traumatic to the human body (5). Acharya et al. indicated that 70% of the diagnoses with FNA were benign; about 4% (1% to 10%) were malignant or suspected malignant (6). Therefore, FNA is not suitable for all thyroid nodules. Ultrasound is a non-invasive and radiation-free method, which has low cost and short time and can even show a few millimeters of lesions, so it has been widely used (7, 8). However, the diagnosis of ultrasound depends on the experience and judgment of doctors to a great extent, which may lead to misdiagnosis (9). The computer-aided diagnosis (CAD) system could help doctors diagnose thyroid nodules objectively (10–14). Park et al. confirmed that CAD and radiology were generally comparable; CAD is feasible to assist doctors in diagnosis (15). Chang et al. extracted six features and calculated the F-score of these feature sets and screened out the main texture features through support vector machines (SVMs) for subsequent classification (16). Lyra et al. adopted the gray-level co-occurrence matrix (GLCM) to characterize texture features of thyroid nodules (17). Keramidas et al. tried to classify thyroid nodules by local binary patterns (LBPs) (18). Acharya et al. segmented thyroid nodule images manually, extracted four texture features from images, including fractal dimension (FD), LBP, Fourier spectrum descriptor (FS), and Laws' texture energy (LTE), and then used these feature vectors to predict thyroid nodules using SVM, decision tree (DT), Sugeno fuzzy, gaussian mixture model (GMM), K-nearest neighbor (KNN), radial basis probabilistic neural network (RBPNN), and naive bayes classifier (NBC) (19). Ma et al. tried to classify thyroid nodules by five different machine learning methods, namely, deep neural network (DNN), SVM, central clustering methods, KNN, and logistic regression; the accuracy was 0.87 (20).

These methods need to extract and even fuse many features manually to achieve high results. While deep learning can deal with massive data and learn deeper and more abstract features automatically, it also avoids complex manual feature extraction (21–23). Guan et al. employed the Inception v3 model to classify thyroid nodules (24). Chi et al. proposed a deep learning framework to extract features from ultrasound images. The proposed model achieved 96.34% accuracy, 86% sensitivity, and 99% specificity on their own database (25). Peng et al. developed a ThyNet model to classify thyroid nodules from multiple hospitals; the area under the receiver operating characteristic curve (AUROC) reached 0.922, and the AUROC of the radiologist's diagnosis improved from 0.837 to 0.875 with the aid of ThyNet, and from 0.862 to 0.873 in the clinical test. The frequency of FNA is reduced in the simulated scenario, while the missed diagnosis of malignant tumors was reduced from 18.9% to 17% as well (26). Avola et al. proposed a knowledge-driven classification framework; the proposed framework finally achieved an AUC value of 98.79% (27). Ye et al. applied a residual network pretrained by transfer learning to classify thyroid nodules; the highest accuracy reached 93.75% (28). Ma et al. tried to classify thyroid nodules by two pretrained convolutional neural networks (CNNs); the feature maps obtained by the two CNNs were fused (29). Sun et al. transferred the CNN model learned from ImageNet as a pretrained feature extractor to a new dataset of ultrasound images. The proposed method combined traditional low-level features extracted from the histogram of oriented gradient (HOG) and LBP with high-level deep features extracted from CNN models to form a hybrid feature space. The experimental accuracy was 93.10% (30). Furthermore, Chen et al. used a deep learning ultrasound text classifier for predicting thyroid nodules. The method achieved 93% and 95% accuracy on real medical datasets and standard datasets, respectively (31).

In this paper, 508 ultrasound images were collected from the Third Hospital of Hebei Medical University in China, which are the same as that by Ma et al. (20). The main work is as follows. First, the ResNet18 model, combined with transfer learning, was employed to classify benign and malignant thyroid nodules (32). Second, a heatmap was used to visualize the model's attention on thyroid nodule images. Next, the highlighted regions expressed by heatmaps were extracted and analyzed from original images. Finally, we found that the characteristics of thyroid nodules with different properties were quite different ($P < 0.05$ means the difference is statistically significant.). In summary, our method is effective in the classification of thyroid nodule images.

MATERIAL AND METHOD

Ultrasound Image Data

In this paper, we used the ultrasound images of thyroid nodules provided by the Third Hospital of Hebei Medical University in China. Images of thyroid nodules had been checked by an experienced doctor; artifacts and blurred images were excluded. The pathologist determined the benign and malignant nodules according to the pathological diagnosis. Finally, we collected 508 ultrasound images, of which 415 were benign nodules and 93 were malignant nodules. Each ultrasound image corresponded to a patient. Among them, 70% of the dataset served as the training set and 30% as the test set. There was no overlap between the training and testing sets. **Table 1** shows the distribution of benign and malignant nodules in the training and testing groups.

The complete process of predicting benign and malignant thyroid nodules is shown in **Figure 1**, which can be divided into three components. The Resnet18 model was employed to diagnose whether the nodule is benign or malignant; the heatmap shows the highlighted regions of the model, which were then extracted and analyzed.

TABLE 1 | The distribution of thyroid nodules in the training and testing groups.

Dataset	Benign	Malignant	Total
Train	291	66	357
Test	124	27	151
Total	415	93	508

A Model for Predicting Thyroid Nodules With Ultrasound Images

The convolution neural network performs well in image classification. The CNN has an input layer, hidden layers, and an output layer. With the image always as input, the hidden layers are used to extract features in the image. ResNet is a classical convolutional neural network, which was proposed by He et al. (32); it was the champion at the 2015 ImageNet Large Scale Visual Recognition Challenge (ILSVRC). The structure of the ResNet18 network is shown in **Figure 2**. A 7*7 convolution and a 3*3 max pooling operation were employed successively. Then it was followed by four layers, with each layer containing two basic blocks, and each basic block containing two convolutions, a batch normalization, an activation function, and a shortcut connection. Then it was modified to two classes. Among them, the role of batch normalization is to speed up the training and convergence of the network, preventing the gradient from vanishing, exploding, and overfitting at the same time. The activation function adds non-linear factors to improve the expressive ability of the neural network. The shortcut connection directly bypasses the input information to the output; this ensures the integrity of information transmission. The constructure of layers is depicted in **Figure 2**; the dotted line indicates that the number of channels has changed, and the solid line indicates that the number of channels has not changed.

Although ResNet18 predicts well in natural images, the results cannot be guaranteed to be optimal due to the huge difference between ultrasonic images and natural images. In addition, it is very complicated to train the neural network from scratch, which needs a lot of computing and memory resources. Transfer learning extracts basic information from the source dataset and applies it to the target domain by fine-tuning parameters. It can use fewer computing resources and shorter time to train the model to obtain better results.

Cross entropy (CE) is a commonly used loss function. We can obtain results that are consistent with expectations by using cross entropy as the loss function and minimizing the target loss function as the goal. Unfortunately, our dataset is unbalanced. If CE is used as the objective function, different categories will be given equal weights when calculating, which may greatly interfere with the learning process of model parameters. Many negative samples constitute a large part of the loss, thus controlling the direction of gradient updates, making the final trained model more inclined to classify samples into this type. Therefore, in order to avoid the hidden dangers that may be caused by dataset imbalance, focal loss was employed (33). Focal loss was proposed by Lin et al. in 2017; it was applied to solve the problem of data imbalance and difficult samples. Focal loss is derived from the CE loss function; CE is defined as:

$$CE(p,y) = \begin{cases} -log(p), & if \quad y = 1 \\ -log(1-p), & otherwise \end{cases} \tag{1}$$

where y represents the labels of negative and positive samples, corresponding to 0 or 1 in binary classification, and p ∈ [0,1] represents the estimated probability of the class labeled y = 1.

$$p_t = \begin{cases} p, & if \quad y = 1 \\ 1-p, & otherwise \end{cases} \tag{2}$$

At this point, CE becomes:

$$CE(p,y) = -log(p_t) \tag{3}$$

Lin et al. added a modulating factor on the basis of CE: $(1-p_t)^{\gamma}$; α is the adjustment factor. Finally, focal loss can be written

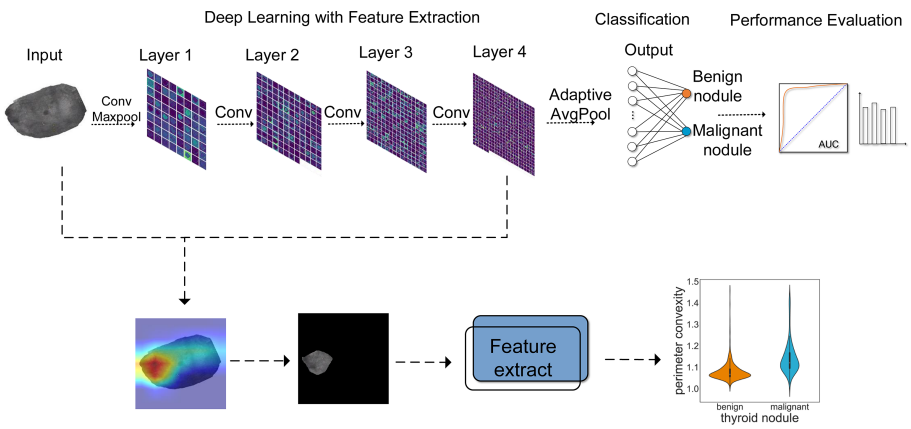


FIGURE 1 | The workflow for thyroid module classification with ResNet18. Layer 1~layer 4 show the process of image analysis by ResNet18. With the increase of layers, the features extracted by the model became more abstract. AUC and other evaluation indicators were used to evaluate the effect of model classification. In addition, a heatmap was employed to visualize the prediction results, by which we extracted and analyzed the highlighted areas.

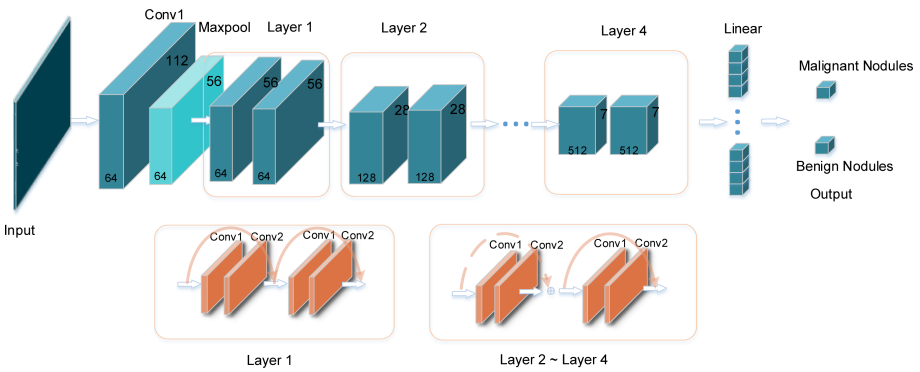


FIGURE 2 | The specific structure of the ResNet18 model. The input is the image of thyroid nodules with the same size. After the convolution layers and pooling layers, the image features are extracted automatically. The output layer is the result of classification: benign or malignant nodule.

as:

$$FL(p_t) = -\alpha_t(1-p_t)^{\gamma}log(p_t) \tag{4}$$

In summary, our framework consists of the following steps.

1. Assign a label for each thyroid nodule image.
2. Resize input images to 224*224, then transform them by flip, rotate, and so on to generate a more complex and diverse dataset, then normalize images to improve the accuracy and generality.
3. Pretrain the ResNet18 model on ImageNet to learn general image feature parameters from natural images.
4. Fine-tune the pretrained model, and adjust the 1000 classification on ImageNet to 2 in our dataset.
5. Adjust the parameters about the network, set the learning rate to 0.0001, and employ the Adam optimizer and focal loss function to optimize the network.
6. Perform pooling and fully connect the layers, then output the prediction results.

Visualizing the Highlighted Regions by Heatmap

CNN's performance is excellent, but it sacrifices intuition and interpretability, and it is hard to interpret the prediction results of the model. Class activation mapping (CAM) is a modern technique for model interpretation (34). However, the disadvantage is that CAM depends on the global average pooling layer; if not, we need to change and retrain the model. Therefore, the use of CAM is not convenient. Grad-CAM alleviates this problem (35). It does not need to modify the structure of the existing model, which makes it applicable for any CNN-based architecture. Grad-CAM shows the contribution distribution of the model output by heatmap, and the contribution is shown by colors; the red color in the heatmap represents a large contribution to the output, which is the main basis for judgment, while blue represents a small contribution. As shown in formula (5), let y be the probability of class and ∂_k be the feature map of the last convolutional layer of the network. Compute the gradient of y with respect to ∂_k and take a global average pooling of all pixels (i and j are width and height, respectively) to obtain a weight ∂_k, which represents the importance of feature map k to discriminate the thyroid nodule category.

$$\partial_k = \frac{1}{Z}\sum_{ij}\frac{\partial y}{\partial A_{ij}^k} \tag{5}$$

Next, weighted combinations were performed to sum the feature maps. Then, a ReLU function was followed, because we only focus on the areas that have a positive impact on the class judgment, as described in formula (6). Grad-CAM shows the areas of positive impact clearly by the heatmap.

$$L_{Grad-CAM} = ReLU\left(\Sigma_k \partial_k A^k\right) \tag{6}$$

Extracting and Analyzing the Highlighted Regions in Heatmaps

After visualizing the area of interest of the model by Grad-CAM, the highlighted regions based on the heatmap were extracted. First of all, according to the corresponding relationship between HSV and RGB colors, the image was converted from RGB to HSV color space. Then, the red areas in the heatmap were captured; we extracted the boundaries of the red area and overlapped them with the original image. Then the extracted areas were analyzed.

In general, image feature extraction includes shape, texture, and color feature extraction. These features depict images from different aspects. Shape features are mostly used to describe shapes of objects in images (36–38). Combined with the extracted areas of thyroid nodules, the form parameter, area convexity, and perimeter convexity were adopted to describe the shape. These features might not be the most perfect representation of the properties of the target regions but may be the most appropriate description of the shape of the regions.

The form parameter refers to the ratio of the area to the square of perimeter in one region, which indicates the complexity of the edge. The formula of the form parameter is

as follows:

$$F = \frac{4 * pi * S}{P^2} \quad (7)$$

where S represents the area of the region and P represents the perimeter of the region. A smaller value for the form parameter indicates a more complex edge of the region.

The area convexity is the ratio of the area to the convex hull area. The convex hull refers to the smallest convex polygon containing the specified area; that is to say, all points in the target region are on or inside the surrounding convex hull area. The formula for area convexity is as follows.

$$S = \frac{S}{S_{ch}} \quad (8)$$

S_{ch} represents the convex hull area. The area convexity is smaller than or equal to 1; the smaller the value is, the more complex the edge of the region is.

Similarly, the perimeter convexity refers to the ratio of the perimeter to the convex hull perimeter of the region.

$$P = \frac{P}{P_{ch}} \quad (9)$$

where P_{ch} represents the perimeter of the convex hull, and the perimeter convexity is equal to or larger than 1. The larger the value, the more complex the edge of the region.

Evaluation Criteria

Our model was evaluated by ROC curve, accuracy, recall, precision, and F1 score (39, 40), which are defined as follows:

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP} \quad (10)$$

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

$$Precision = \frac{TP}{TP + FP} \quad (12)$$

$$F1\ \ score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (13)$$

where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative, respectively.

RESULTS

Our Model Performs Better Than Existing Methods

Our dataset was randomly divided into 10 times, and we calculated the average of the 10 times to avoid the chance of the results caused by one division. Compared with AlexNet, Inception_v3, and VGG16, the performance of our model was best, the AUC was 0.997, and as shown in **Figure 3A**, the average accuracy, recall, precision, and F1 score were 0.984, 0.978, 0.997, and 0.957, respectively; our model performed well in the above indicators. The results are shown in **Figure 3B**.

Visualizing With Grad-CAM

As we mentioned in the "Visualizing the highlighted regions by heatmap" section, the model achieved unprecedented accuracy in image classification, but the interpretability was poor. Visualization is helpful to understand and debug the model. Grad-CAM was used to validate model predictions on images; it adopted the final convolutional layer gradients to generate the positioning heatmaps when predicting. **Figure 4A** shows original images, where (a1)~(a3) are benign nodules and (a4)~(a6) are malignant nodule images. **Figure 4B**) shows the corresponding heatmaps by Grad-CAM. The highlighted regions in the heatmaps are shown in red, and the weak regions are shown in blue. The red and blue marks represent the regions of strong and weak emphases, respectively.

Region Extraction and Analysis

OpenCV was used to perform region extraction on the red highlight in the heatmap. The extracted results are shown in **Figure 5**; samples in (a) are the extraction result of benign thyroid nodules, and those in (b) are malignant. We found that

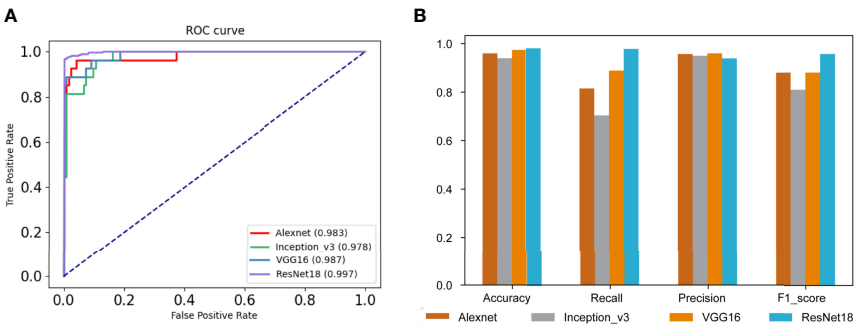


FIGURE 3 | Evaluations of model results. **(A)** The receiver operating characteristic (ROC) curves and the area under the curve (AUC) of our model and other comparative models. **(B)** The performance of our model and other comparative models on accuracy, recall, precision, and F1 score.

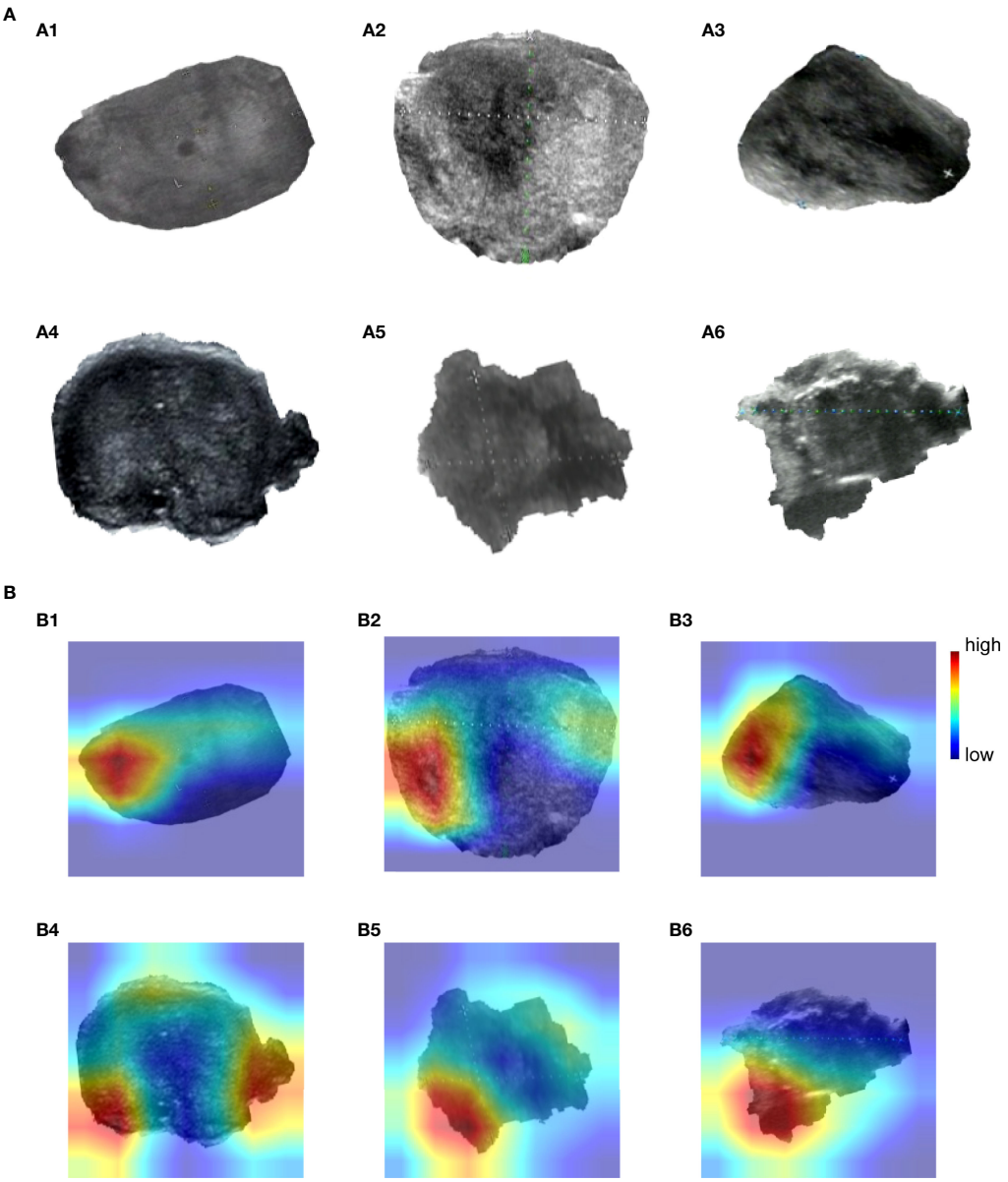


FIGURE 4 | Grad-CAM visualizes highlighted regions. **(A)** (a1)~(a6) are original images, (a1)~(a3) are benign nodules, (a4)~(a6) are malignant nodules. **(B)** (b1)~(b6) are heatmaps drawn by Grad-CAM, corresponding to **Figure 4** (a1)~(a6).

the outline boundary of benign nodules was relatively regular, while the outline boundary of malignant nodules was relatively irregular. To verify whether almost all thyroid nodules fit this phenomenon, the shape features of thyroid nodules were calculated according to the above formulas. As a result, the form parameter, area convexity, and perimeter convexity were statistically different for benign and malignant nodules (their p-values were 1.68e-27, 7.01e-32, and 8.1e-33, respectively. $P < 0.05$ means the difference is statistically significant). As shown in **Figure 6**, violin plots were employed to describe the distribution of values of benign and malignant thyroid nodules with the abovementioned features.

DISCUSSION

The prevalence of thyroid nodules is increasing year by year, and people's awareness of health management is also gradually improving. As a result, the burden of ultrasound doctors in hospitals and physical examination institutions is increasing. If an AI-assisted diagnosis system can be used to assist doctors to distinguish ultrasound image data, the pressure of doctors will be relieved and work efficiency will be improved. Moreover, the interpretation of ultrasound images largely depends on the clinical experience of radiologists. It was reported that the sensitivity of radiologists varied from 40.3% to 100%, and

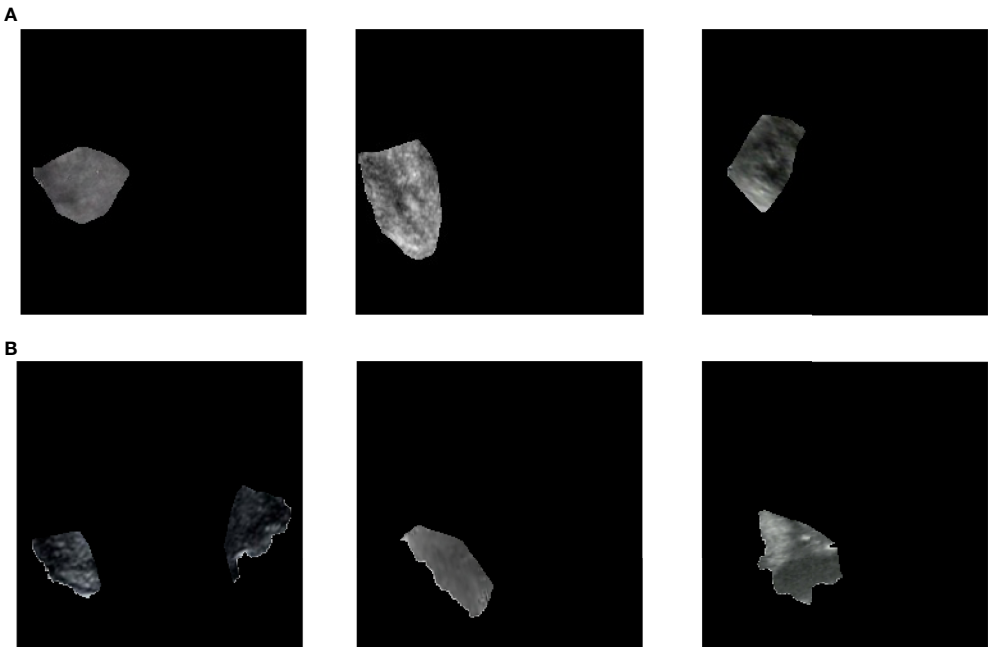


FIGURE 5 | Extract highlighted regions in heatmaps. These heatmaps correspond to **Figure 4A**) or **(B)**. Samples in **(A)** are the extraction result of benign thyroid nodules, which correspond to **Figure 4** (A1–A3) or (B1–B3). Samples in **(B)** are the extraction result of malignant thyroid nodules, which correspond to **Figure 4** (a4) ~(a6) or (b4)~(b6).

the specificity from 50% to 100% (41–44). Computer-aided diagnosis can provide more objective and more accurate results, which is very helpful to doctors with less experience. Grani et al. showed that many thyroid nodules removed by surgery were not malignant, which increased the economic burden and physical pain for patients (45), whereas the AI-assisted diagnosis system based on deep learning algorithms could lower the false positive rate and then help to reduce unnecessary FNAB and surgery.

In this paper, the ResNet18 framework was applied to train the model and Grad-CAM was proposed to highlight sensitive regions in the ultrasound images. Finally, the 10-time average AUC of our proposed method was 0.997, and the average accuracy was 0.984, which is higher than the accuracy of 0.89 designed by Ma et al. (20). Moreover, the shape features of the sensitive regions rather than other features are more helpful in the discrimination of benign and malignant tumors. From the perspective of methodology, the performance of neural network-based methods is generally higher than the traditional feature-based methods. CNNs can learn efficient and useful features automatically, avoiding the time-consuming and laborious task of obtaining features manually.

Although the proposed method achieved a supportive result, it still had some limitations. Firstly, the number of images for training and testing is insufficient and multicenter data are not available. In the future work, we will collect more data to validate

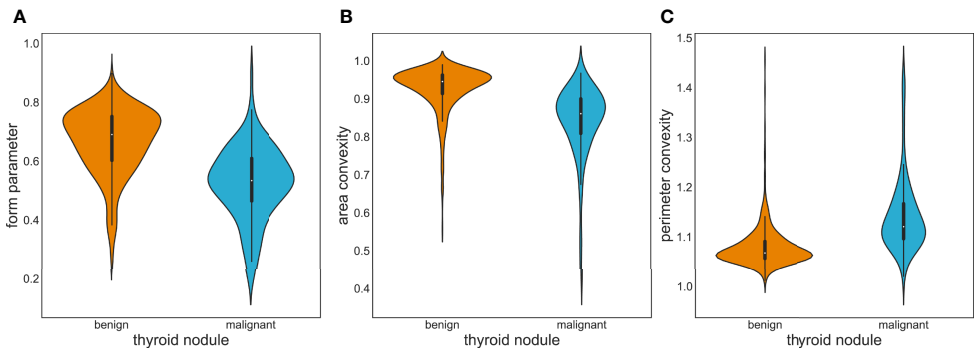


FIGURE 6 | Violin plots of image feature distribution with benign and malignant nodules. **(A)** Form parameter. **(B)** Area convexity. **(C)** Perimeter convexity.

the model performance. Secondly, a more detailed classification of thyroid nodules may be tried using a variety of algorithms as benign thyroid nodules can be divided into benign follicular nodules and follicular adenomas, etc., clinically, and malignant thyroid nodules can be divided into papillary, follicular, etc., medullary carcinomas. Finally, yet importantly, the results were obtained based on static ultrasound images and we should consider how to better assist doctors in making decisions in a real clinical environment.

It is believed that deep learning in diagnosing thyroid nodules has a bright future. Deep learning algorithms have been widely concerned and applied in various fields. They can map unstructured information to structured forms and learn relevant information automatically. The automatic and intelligent method not only improves the efficiency of diagnosis but also ensures the reliability, which may significantly affect early diagnosis and subsequent treatment.

CONCLUSION

In this paper, we have explored the problem of thyroid nodule classification. ResNet18 was deployed with 508 thyroid nodule ultrasound images. Due to insufficient datasets, transfer learning was adopted. At the same time, considering the imbalance of the dataset, focal loss was employed to adjust the weight of data. Finally, the AUC was 0.997, which means that we can predict almost all thyroid nodules correctly. Moreover, in order to visualize the model's attention in thyroid nodule images and help understand the model's predictions more easily, Grad-CAM was used to identify sensitive regions in the learning process of ultrasound images, which were an important reference of image prediction. The regions concerned by the model were segmented and analyzed. Finally, in the region of interest, there are differences between benign and malignant nodules. The results of this study show that our model could diagnose well benign and malignant thyroid nodules in ultrasound images. Besides, our localization information could be regarded as a second opinion for clinical decision-making. The proposed method could assist doctors in making better decisions, reducing the time for human participation, and improving the efficiency of 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 authors.

AUTHOR CONTRIBUTIONS

PW and JSY designed the study; JYY, XS, BW, WQ, and GT performed the study, analyzed the data, and interpreted the data; JYY and JSY wrote the manuscript; BW, JYY, GT, and PW reviewed the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (Nos. 51574004, 62172004) and Natural Science Foundation of the Higher Education Institutions of Anhui Province, China (Nos. KJ2019A0085, KJ2019ZD05).

REFERENCES

1. Camargo RY, Tomimori EK. [Usefulness of Ultrasound in the Diagnosis and Management of Well-Differentiated Thyroid Carcinoma]. *Arq Bras Endocrinol Metabol* (2007) 51:783–92. doi: 10.1590/S0004-27302007000500016
2. Enewold L, Zhu K, Ron E, Marrogi AJ, Stojadinovic A, Peoples GE, et al. Rising Thyroid Cancer Incidence in the United States by Demographic and Tumor Characteristics, 1980-2005. *Cancer Epidemiol Biomarkers Prev* (2009) 18:784–91. doi: 10.1158/1055-9965.EPI-08-0960
3. Russ G, Leboulleux S, Leenhardt L, Hegedüs L. Thyroid Incidentalomas: Epidemiology, Risk Stratification With Ultrasound and Workup. *Eur Thyroid J* (2014) 3:154–63. doi: 10.1159/000365289
4. Liu H, Qiu C, Wang B, Bing P, Tian G, Zhang X, et al. Evaluating DNA Methylation, Gene Expression, Somatic Mutation, and Their Combinations in Inferring Tumor Tissue-Of-Origin. *Front Cell Dev Biol* (2021) 9:619330. doi: 10.3389/fcell.2021.619330
5. Baloch ZW, Fleisher S, Livolsi VA, Gupta PK. Diagnosis of "Follicular Neoplasm": A Gray Zone in Thyroid Fine-Needle Aspiration Cytology. *Diagn Cytopathol* (2002) 26:41–4. doi: 10.1002/dc.10043
6. Acharya UR, Swapna G, Sree SV, Molinari F, Gupta S, Bardales RH, et al. A Review on Ultrasound-Based Thyroid Cancer Tissue Characterization and Automated Classification. *Technol Cancer Res Treat* (2014) 13:289–301. doi: 10.7785/tcrt.2012.500381
7. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 2015 American Thyroid Association Management Guidelines for Adult Patients With Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid* (2016) 26:1–133. doi: 10.1089/thy.2015.0020
8. Wang L, Zhang L, Zhu M, Qi X, Yi Z. Automatic Diagnosis for Thyroid Nodules in Ultrasound Images by Deep Neural Networks. *Med Imag Anal* (2020) 61:101665. doi: 10.1016/j.media.2020.101665
9. Li X, Wang S, Xi W, Zhu J, Liu S. Fully Convolutional Networks for Ultrasound Image Segmentation of Thyroid Nodules". *In: 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*. Exeter, UK: IEEE (2018).
10. Huang L, Li X, Guo P, Yao Y, Liao B, Zhang W, et al. Matrix Completion With Side Information and Its Applications in Predicting the Antigenicity of Influenza Viruses. *Bioinformatics* (2017) 33:3195–201. doi: 10.1093/bioinformatics/btx390
11. Cheng L, Hu Y, Sun J, Zhou M, Jiang Q. DincRNA: A Comprehensive Web-Based Bioinformatics Toolkit for Exploring Disease Associations and ncRNA Function. *Bioinformatics* (2018) 34:1953–6. doi: 10.1093/bioinformatics/bty002
12. Yang J, Peng S, Zhang B, Houten S, Schadt E, Zhu J, et al. Human Geroprotector Discovery by Targeting the Converging Subnetworks of

Aging and Age-Related Diseases. *Geroscience* (2020) 42:353–72. doi: 10.1007/s11357-019-00106-x
13. Hunt C, Montgomery S, Berkenpas JW, Sigafoos N, Cao R. Recent Progress of Machine Learning in Gene Therapy. *Curr Gene Ther* (2021) 22:132–43. doi: 10.2174/1566523221666210622164133
14. Meng Y, Lu C, Jin M, Xu J, Zeng X, Yang J. A Weighted Bilinear Neural Collaborative Filtering Approach for Drug Repositioning. *Brief Bioinform* (2022) 23(2):bbab581. doi: 10.1093/bib/bbab581
15. Park VY, Han K, Seong YK, Park MH, Kim EK, Moon HJ, et al. Diagnosis of Thyroid Nodules: Performance of a Deep Learning Convolutional Neural Network Model vs. Radiologists. *Sci Rep* (2019) 9:17843. doi: 10.1038/s41598-019-54434-1
16. Chang CY, Tsai MF, Chen SJ. Classification of the Thyroid Nodules using Support Vector Machines. In: IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence). Hong Kong, China: IEEE (2008). p. 3093–8.
17. Lyra ME, Lagopati N, Charalabatou P, Vasoura E, Skouroliakou K. *Texture Characterization in Ultasonograms of the Thyroid Gland*. Corfu, Greece: IEEE (2010).
18. Keramidas EG. Efficient and Effective Ultrasound Image Analysis Scheme for Thyroid Nodule Detection. *Springer-Verlag* (2007) 4633:1052–60. doi: 10.1007/978-3-540-74260-9_93
19. Acharya UR, Vinitha SSree, Krishnan MM, Molinari F, Garberoglio R, Suri JS. Non-Invasive Automated 3D Thyroid Lesion Classification in Ultrasound: A Class of ThyroScan™ Systems. *Ultrasonics* (2012) 52:508–20. doi: 10.1016/j.ultras.2011.11.003
20. Ma X, Xi B, Zhang Y, Zhu L, Yang J. A Machine Learning-Based Diagnosis of Thyroid Cancer Using Thyroid Nodules Ultrasound Images. *Curr Bioinf* (2020) 15(4):349–58. doi: 10.2174/1574893614666191017091959
21. Zhao T, Hu Y, Peng J, Cheng L. DeepLGP: A Novel Deep Learning Method for Prioritizing lncRNA Target Genes. *Bioinformatics* (2020) 36:4466–72. doi: 10.1093/bioinformatics/btaa428
22. Du B, Tang L, Liu L, Zhou W. Predicting LncRNA-Disease Association Based on Generative Adversarial Network. *Curr Gene Ther* (2022) 22:144–51. doi: 10.2174/1566523221666210506131055
23. 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
24. Guan Q, Wang Y, Du J, Qin Y, Lu H, Xiang J, et al. Deep Learning Based Classification of Ultrasound Images for Thyroid Nodules: A Large Scale of Pilot Study. *Ann Transl Med* (2019) 7:137. doi: 10.21037/atm.2019.04.34
25. Chi J, Walia E, Babyn P, Wang J, Groot G, Eramian M. Thyroid Nodule Classification in Ultrasound Images by Fine-Tuning Deep Convolutional Neural Network. *J Dig Imaging* (2017) 30:477–86. doi: 10.1007/s10278-017-9997-y
26. 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 Dig Health* (2021) 3:e250–9. doi: 10.1016/S2589-7500(21)00041-8
27. Avola D, Cinque L, Fagioli A, Filetti S, Grani G, Rodolà E. Knowledge-Driven Learning *via* Experts Consult for Thyroid Nodule Classification. *ArXiv E-prints*. (2020). arXiv:2005.14117. 27. doi: 10.48550/arXiv.2005.14117
28. Ye Z, Zhuang F, Jian F. An Image Augmentation Method using Convolutional Network for Thyroid Nodule Classification by Transfer Learning, In: 2017 3rd IEEE International Conference on Computer and Communications (ICCC). Chengdu, China: IEEE (2017). 1819–23.
29. Ma J, Wu F, Zhu J, Xu D, Kong D. A Pre-Trained Convolutional Neural Network Based Method for Thyroid Nodule Diagnosis. *Ultrasonics* (2017) 73:221–30. doi: 10.1016/j.ultras.2016.09.011
30. Sun W, Liu T, Xie S, Yu J, Niu L, Sun W. Classification of Thyroid Nodules in Ultrasound Images using Deep Model Based Transfer Learning and Hybrid Features In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). New Orleans, LA, USA: IEEE (2017). p. 919–23.
31. Chen D, Niu J, Qiao P, Yue L, Mei W. A Deep-Learning Based Ultrasound Text Classifier for Predicting Benign and Malignant Thyroid Nodules. *In: International Conference on Green Informatics*. Fuzhou, China: IEEE (2017). p. 199-204
32. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet Large Scale Visual Recognition Challenge. *Int J Comput Vision* (2015) 115:211–52. doi: 10.1007/s11263-015-0816-y
33. Lin TY, Goyal P, Girshick R, He K, Dollar P. Focal Loss for Dense Object Detection. *IEEE Trans Pattern Anal Mach Intell* (2020) 42:318–27. doi: 10.1109/TPAMI.2018.2858826
34. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning Deep Features for Discriminative Localization. *CVPR* (2016) 2921–9. doi: 10.1109/CVPR.2016.319
35. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations From Deep Networks *via* Gradient-Based Localization. *Int J Comput Vision* (2020) 128:336–59. doi: 10.1007/s11263-019-01228-7
36. Wu SG, Bao FS, Xu EY, Wang YX, Xiang QL. *A Leaf Recognition Algorithm for Plant Classification Using Probabilistic Neural Network*. IEEE (2007).
37. Yang M, Kpalma K, Ronsin J. A Survey of Shape Feature Extraction Techniques. *InTech* (2007) 15(7):43–90. doi: 10.5772/6237
38. Lande MV, Bhanodiya P, Jain P. *An Effective Content-Based Image Retrieval Using Color, Texture and Shape Feature*. Intelligent Computing, Networking, and Informatics (2014).
39. Cheng L, Jiang Y, Ju H, Sun J, Peng J, Zhou M, et al. InfAcrOnt: Calculating Cross-Ontology Term Similarities Using Information Flow by a Random Walk. *BMC Genomics* (2018) 19:919. doi: 10.1186/s12864-017-4338-6
40. Puranik N, Yadav D, Chauhan PS, Kwak M, Jin JO. Exploring the Role of Gene Therapy for Neurological Disorders. *Curr Gene Ther* (2021) 21(1):11–22. doi: 10.2174/1566523220999200917114101
41. Park SH, Kim SJ, Kim EK, Kim MJ, Son EJ, Kwak JY. Interobserver Agreement in Assessing the Sonographic and Elastographic Features of Malignant Thyroid Nodules. *AJR Am J Roentgenol* (2009) 193:W416–423. doi: 10.2214/AJR.09.2541
42. Kim SH, Park CS, Jung SL, Kang BJ, Kim JY, Choi JJ, et al. Observer Variability and the Performance Between Faculties and Residents: US Criteria for Benign and Malignant Thyroid Nodules. *Kor J Radiol* (2010) 11:149–55. doi: 10.3348/kjr.2010.11.2.149
43. Park CS, Kim SH, Jung SL, Kang BJ, Kim JY, Choi JJ, et al. Observer Variability in the Sonographic Evaluation of Thyroid Nodules. *J Clin Ultrasound* (2010) 38:287–93. doi: 10.1002/jcu.20689
44. Kim HG, Kwak JY, Kim EK, Choi SH, Moon HJ. Man to Man Training: Can it Help Improve the Diagnostic Performances and Interobserver Variabilities of Thyroid Ultrasonography in Residents? *Eur J Radiol* (2012) 81:e352–356. doi: 10.1016/j.ejrad.2011.11.011
45. Grani G, Lamartina L, Ascoli V, Bosco D, Biffoni M, Giacomelli L, et al. Reducing the Number of Unnecessary Thyroid Biopsies While Improving Diagnostic Accuracy: Toward the "Right" TIRADS. *J Clin Endocrinol Metab* (2019) 104:95–102. doi: 10.1210/jc.2018-01674

Conflict of Interest: JYY, XS, WQ, and GT are currently employed in Geneis Beijing Co., Ltd.

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

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

Frontiers 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 imrpove diagnosis, therapeutics and management strategies.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact