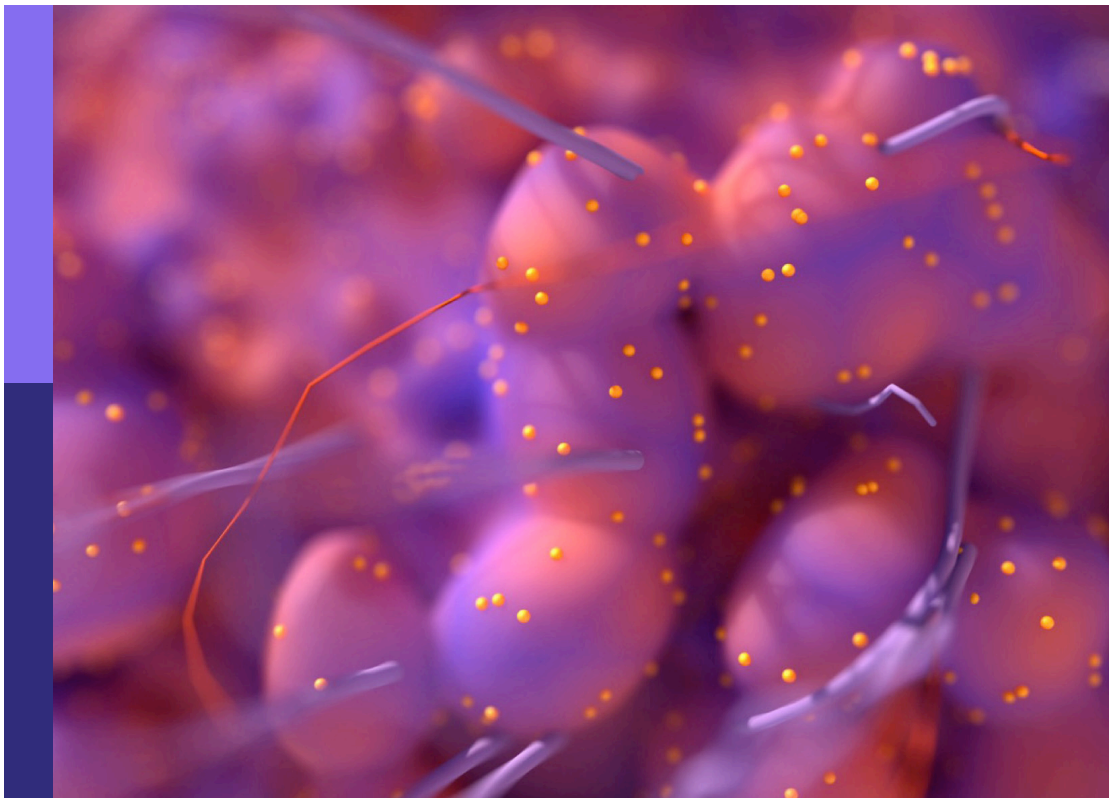
frontiers | Research Topics

Advances and new insights into cancer characterization: When novel imaging meets quantitative imaging biomarkers

vol. II

Edited by
Guang Yang, Tingying Peng and Kyung Hyun Sung

Published in
Frontiers in Oncology



ISSN 1664-8714
ISBN 978-2-83252-163-2
DOI 10.3389/978-2-83252-163-2

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

Advances and new insights into cancer characterization: When novel imaging meets quantitative imaging biomarkers vol. II

Topic editors

Guang Yang — Imperial College London, United Kingdom

Tingying Peng — Helmholtz AI, Helmholtz zentrum München, Germany

Kyung Hyun Sung — Department of Radiology, UCLA Health System, United States

Citation

Yang, G., Peng, T., Sung, K. H., eds. (2023). *Advances and new insights into cancer characterization: When novel imaging meets quantitative imaging biomarkers vol. II*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83252-163-2

Table of contents

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 14 April 2022
doi: 10.3389/fonc.2022.876120

Evaluation of Amide Proton Transfer-Weighted Imaging for Risk Factors in Stage I Endometrial Cancer: A Comparison With Diffusion-Weighted Imaging and Diffusion Kurtosis Imaging

Xingxing Jin[1†], Ruifang Yan[1†], Zhong Li[1†], Gaiyun Zhang[1], Wenling Liu[1], Hongxia Wang[1], Meng Zhang[1], Jinxia Guo[2], Kaiyu Wang[2] and Dongming Han[1]*

[1] Department of Magnetic Resonance Imaging (MRI), The First Affiliated Hospital, Xinxiang Medical University, Weihui, China, [2] Magnetic Resonance Imaging (MRI) Research China, General Electric (GE) Healthcare, Beijing, China

OPEN ACCESS

Edited by:
Guang Yang,
Imperial College London,
United Kingdom

Reviewed by:
Xiaotang Yang,
Shanxi Provincial Cancer Hospital,
China
Yoshitaka Masutani,
Hiroshima City University, Japan

***Correspondence:**
Dongming Han
625492590@qq.com

[†]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: 15 February 2022
Accepted: 11 March 2022
Published: 14 April 2022

Citation:
Jin X, Yan R, Li Z, Zhang G, Liu W, Wang H, Zhang M, Guo J, Wang K and Han D (2022) Evaluation of Amide Proton Transfer-Weighted Imaging for Risk Factors in Stage I Endometrial Cancer: A Comparison With Diffusion-Weighted Imaging and Diffusion Kurtosis Imaging. Front. Oncol. 12:876120. doi: 10.3389/fonc.2022.876120

Background: Endometrial cancer (EC) is one of the most common gynecologic malignancies in clinical practice. This study aimed to compare the value of diffusion-weighted imaging (DWI), diffusion kurtosis imaging (DKI), and amide proton transfer-weighted imaging (APTWI) in the assessment of risk stratification factors for stage I EC including histological subtype, grade, stage, and lymphovascular space invasion (LVSI).

Methods: A total of 72 patients with stage I EC underwent pelvic MRI. The apparent diffusion coefficient (ADC), mean diffusivity (MD), mean kurtosis (MK), and magnetization transfer ratio asymmetry (MTRasym at 3.5 ppm) were calculated and compared in risk groups with the Mann–Whitney *U* test or independent samples *t*-test. Spearman's rank correlation was applied to depict the correlation of each parameter with risk stratification. The diagnostic efficacy was evaluated with receiver operating characteristic (ROC) curve analysis and compared using the DeLong test. A multivariate logistic regression was conducted to explore the optimal model for risk prediction.

Results: There were significantly greater MTRasym (3.5 ppm) and MK and significantly lower ADC and MD in the non-adenocarcinoma, stage IB, LVSI-positive, high-grade, and non-low-risk groups (all $p < 0.05$). The MK and MTRasym (3.5 ppm) were moderately positively correlated with risk stratification as assessed by the European Society for Medical Oncology (EMSO) clinical practice guidelines ($r = 0.640$ and 0.502, respectively), while ADC and MD were mildly negatively correlated with risk stratification ($r = -0.358$ and -0.438, respectively). MTRasym (3.5 ppm), MD, and MK were identified as independent risk predictors in stage I EC, and optimal predictive performance was obtained with their combinations (AUC = 0.906, sensitivity = 70.97%, specificity = 92.68%). The results of the validation model were consistent with the above results, and the calibration curve showed good accuracy and consistency.

Conclusions: Although similar performance was obtained with each individual parameter of APTWI, DWI, and DKI for the noninvasive assessment of aggressive behavior in stage I EC, the combination of MD, MK, and MTRasym (3.5 ppm) provided improved predictive power for non-low-risk stage I EC and may serve as a superior imaging marker.

Keywords: endometrial cancer, amide proton transfer-weighted imaging, diffusion kurtosis imaging, diffusion-weighted imaging, risk factors

INTRODUCTION

Endometrial cancer (EC) is one of the most common gynecologic malignancies in clinical practice, and 80% of newly diagnosed patients are in stage I (1, 2). According to the histological subtype, grade, International Federation of Gynecology and Obstetrics (FIGO) stage, and lymphovascular space invasion (LVSI), the European Society for Medical Oncology (ESMO) clinical practice guidelines classify stage I EC into low risk, intermediate risk, intermediate high risk, and high risk (3). For low-risk patients, lymphadenectomy is likely to lead to complications and increased care costs, thereby reducing their survival benefit, but in non-low-risk (intermediate-, intermediate-high-, and high-risk) patients, lymphadenectomy is necessary and effective (4). The histological subtype, grade, FIGO stage, and LVSI, which are obtained mainly by preoperative magnetic resonance imaging (MRI) and biopsy at present, are important factors for the risk stratification of stage I EC (3, 5). However, the accuracy of FIGO stage evaluation using conventional T1-weighted (T1W) and T2-weighted (T2W) MRI may be influenced by factors such as adenomyosis, leiomyomas, myometrial compression, and loss of the junctional zone (6, 7). In addition, biopsy has the disadvantages of invasiveness, inadequate sampling, and susceptibility to operator experience (8, 9). Therefore, it is of great interest to discover a noninvasive and effective means for assessing stage I EC risk factors for stratification, thus complementing existing methods.

Diffusion and molecular MR imaging techniques have been explored for the diagnosis and differentiation of EC. Diffusion-weighted (DW) MRI detects the diffusion movement of water molecules in tissues (10). Jiang et al. showed that diffusion-weighted imaging (DWI) can help differentiate EC from normal endometrial parenchyma (11). Diffusion kurtosis imaging (DKI), as an evolutionary technique of DWI, takes into account the non-Gaussian distribution of the diffusion movement of water molecules in the tissue and is considered as a more accurate imaging technique to characterize the microstructure of the lesion (12, 13). Amide proton transfer-weighted imaging (APTWI) is a molecular imaging method that utilizes the chemical exchange between amide protons and water molecules to quantify the mobile proteins and peptides in tissues (14). Yue et al. and Takayama et al. indicated that DKI and APTWI can play active roles in the histological grading assessment of EC (15, 16). Meng et al. found that both DKI and APTWI can be used in the diagnosis of EC with different clinical and histological types (17). Based on these results, we hypothesized that DWI-, DKI-, and APTWI-related parameters may be useful predictors of the risk stratification factors for stage I EC.

A few studies have briefly reported the role of these techniques in stage I EC risk stratification (18, 19). However, these studies either explored the application of only a single imaging technique or assessed only risk stratification without evaluating the risk stratification factors. The aims of this study were to compare the value of DWI, DKI, and APTWI in assessing the risk stratification factors for stage I EC, including histological subtype, grade, FIGO stage, and LVSI, and to explore the advantage of including multiple parameters from MRI in differentiating low-risk and non-low-risk stage I EC patients.

MATERIALS AND METHODS

Study Population

The local institutional review board approved the present study, and all participants provided written informed consent. A series of 132 consecutive female patients with suspected EC on computed tomography (CT) or ultrasound (US) underwent pelvic MRI between July 2018 and June 2021. Sixty participants were excluded for the following reasons: 1) having FIGO stage II, III, or IV (n = 32); 2) having claustrophobia or other diseases or conditions that prevent them from completing all the sequences (n = 4); 3) inadequate imaging quality in DWI, DKI, or APTWI for analysis due to severe artifacts (n = 5); 4) received relevant treatment prior to scanning (n = 6); 5) having the largest area of the lesion <50 pixels (392 mm^2) in the axial plane of DWI, DKI, or APTWI (n = 7); 6) histological findings of non-EC (n = 4); and 7) uncertain histological findings (n = 2). Ultimately, 72 patients were enrolled in the present study (**Figure 1**).

MRI Protocols

All pelvic MRI examinations were acquired with a 3.0-T MRI system (Discovery MR750, GE Healthcare, Waukesha, WI, USA) using a 16-channel phased-array body coil. Participants were given 40 mg of hyoscine butylbromide (Buscopan; Boehringer, Ingelheim, Germany) intramuscularly or intravenously prior to the examination to minimize bowel motion. All participants were placed in the supine position with their feet first and a partially full bladder. Two-dimensional oblique axial (perpendicular to the long axis of the cervix) T1W imaging (T1WI), T2W imaging (T2WI), and DWI were performed first. Subsequently, all slices containing lesions were selected from the images from DWI, and their position, layer thickness, and layer

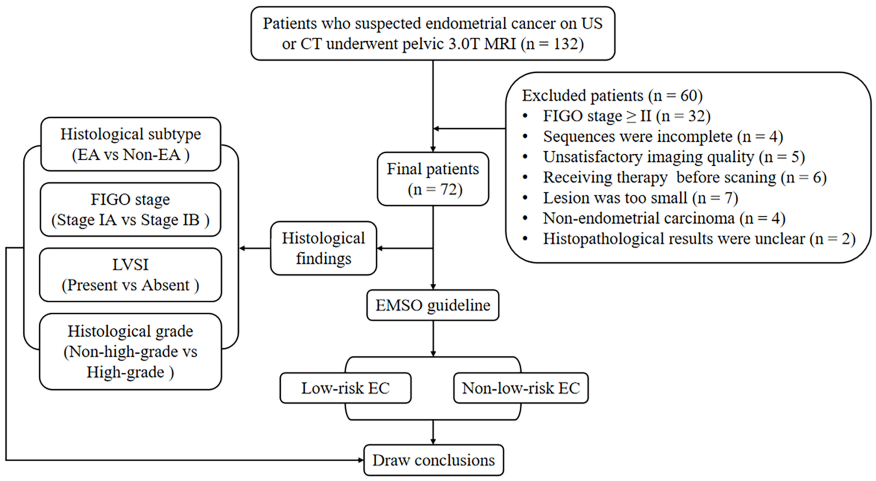


FIGURE 1 | Flowchart of the present study.

spacing were copied to DKI and APTWI for the corresponding scans. Finally, a three-dimensional axial contrast-enhanced sequence was performed *via* intravenous injection (0.1 ml/kg, 2.0 ml/s) of gadopentetate dimeglumine (Gd-DTPA; Bayer Pharmaceutical, Berlin, Germany) using an automatic injector. Details of the protocol are provided in **Table 1**.

Image Post-Processing

All images were transferred to the Advantage Workstation (version 4.6; GE Healthcare) and post-processed with the apparent diffusion coefficient (ADC), DKI, and amide proton transfer (APT) processing tools independently by two genitourinary radiologists (XJ and RY, with 7 and 15 years of experience, respectively.) who were unaware of each other's outcomes and of the clinical and histological information.

The DWI parameter was drawn from the following formula:

$$S_b/S_0 = \exp(-b \times \mathrm{ADC}) \quad (1)$$

where b is the diffusion sensitizing factor, S_0 and S_b are the signal intensities (SIs) under zero and nonzero b values, respectively, and ADC is the apparent diffusion coefficient (10). The DKI parameter was derived from the following function:

$$S_b = S_0 \times \exp(-b \times D_{app} + b^2 \times D_{app}2 \times K_{app}/6) \quad (2)$$

where D_{app} denotes the diffusion coefficient corrected for non-Gaussian bias and K_{app} denotes the degree of deviation from the Gaussian distribution. MD and MK reflect the average D_{app} and K_{app} values for all directions, respectively (12). The APTWI parameter was calculated using the following equation:

TABLE 1 | Imaging protocol parameters.

Parameters	T1WI	T2WI	DWI	DKI	APTWI	Contrast-enhanced imaging
Sequence	2D FSE	2D FSE	2D SS-EPI	2D SS-EPI	2D EPI	3D LAVA
Orientation	Axial	Axial	Axial	Axial	Axial	Axial
Repetition time/echo time (ms)	605/8	5,455/109	6,000/60.5	2,500/58.9	3,000/12	4.2/2.1
Field of view (cm^2)	36 × 36	36 × 36	36 × 36	36 × 36	36 × 36	36 × 36
Matrix	320 × 224	320 × 224	128 × 128	128 × 128	128 × 128	320 × 320
Bandwidth (Hz/pixel)	62.50	83.33	250	250	250	83.33
Slice thickness (mm)	5	5	5	5	5	1
No. of sections	20	20	20	Based on lesion size	Based on lesion size	80
No. of excitation	1	1	1 (b = 0) 4 (b = 1,000)	2	1	0.7
Diffusion encoding directions	–	–	1	30	–	–
Fat suppression	–	STIR	STIR	SPECIAL	STIR	FLEX
b-values (s/mm^2)	–	–	0, 1000	0, 500, 1,000,1,500, 2,000	–	–
Respiratory compensation	Free	Free	Free	Free	Free	Breath holding
Scan time	1 min, 57 s	1 min, 33 s	1 min, 24 s	5 min, 28 s	2 min, 36 s (single slice)	9 s (each phase)

Saturation pulses (T_{sat}) of 0.5 s and a saturation level of 2.0 μT were used to perform APTWI. A total of 52 frequencies, including a frequency 5000 Hz (3 times) away from the resonant frequency and 49 offsets ranging from -600 to +600 Hz with an interval of 25 Hz, were used for signal normalization of APTWI and z-spectrum scans. The water saturation shift reference (WASSR) was applied for B_0 correction. The number of DKI diffusion gradient directions is 30.

T1WI, T1-weighted imaging; T2WI, T2-weighted imaging; DWI, diffusion-weighted imaging; DKI, diffusion kurtosis imaging; APTWI, amide proton transfer-weighted imaging; FSE, fast spin echo; SS-EPI, single-shot echo planar imaging; LAVA, liver acquisition with volume assessment; FLEX, flexible; STIR, short-inversion time (TI) recovery; SPECIAL, spectral inversion at lipids.

$$\text{MTRasym}\ (3.5\ \text{ppm}) = [S_{\text{sat}}(-3.5\ \text{ppm}) - S_{\text{sat}}(+3.5\ \text{ppm})]/S_0 \quad (3)$$

where S_{sat} and S_0 denote the SIs obtained with and without selective saturation, respectively, and MTRasym (3.5 ppm) is the asymmetric magnetization transfer ratio at 3.5 ppm (14). With the DWI and contrast-enhanced images as references, the region of interest (ROI) of the lesion was manually delineated layer by layer along the inside of the tumor margin on axial T2WI, where areas with necrosis, apparent signs and hemorrhage artifacts, cystic degeneration, and blood vessels were avoided (17). All ROIs were automatically copied to each parameter map by the software to calculate the mean values.

Histopathological Analysis

All lesion specimens were harvested through surgery with a median interval of 10 days (range, 1–24 days) between pelvic MRI examination and surgery. An experienced pathologist analyzed these specimens without knowledge of the clinical and imaging findings. The histological grade, subtype, and LVSI were confirmed by hematoxylin/eosin (HE) staining. The depth of myometrial invasion was assessed using the FIGO staging system (20). The new ESMO clinical practice guidelines were used to assess the risk stratification, and eventually, all included participants were classified into four groups: low, intermediate, high-intermediate, and high risk (3). Then, the low-risk stage I EC patients were categorized into the low-risk group and the intermediate, high-intermediate, and high-risk stage I EC patients, who are usually considered to have undergone lymphadenectomy, were categorized into the non-low-risk group (18).

Statistical Analysis

The intraclass correlation coefficient (ICC) was calculated to assess inter-observer agreement ($r < 0.40$, poor; $0.40 \le r < 0.60$, fair; $0.60 \le r < 0.75$, good; and $r \ge 0.75$, excellent) (21). After checking the normality of the data with the Shapiro–Wilk test, the Mann–Whitney *U* test was used for the comparison of non-normally distributed data (median and interquartile range) and the independent samples *t*-test used for the comparison of normally distributed data (mean ± standard deviation). The diagnostic efficacy of the different parameters was described by the area under the receiver operating characteristic (ROC) curve (AUC) and compared with DeLong analysis. A multivariate logistic regression was performed to explore the optimal differentiation performance with multiple parameters. The regression model was also verified using calibration curves with bootstrapping (1,000 samples) (22). Spearman's rank correlation was applied to evaluate the correlation of each parameter with risk stratification ($r \ge 0.75$, good; $0.50 \le r < 0.75$, moderate; $0.25 \le r < 0.50$, mild; and $r < 0.25$, little or none) (23). All analyses were performed by Stata (version 16.0; StataCorp, College Station, TX, USA) and MedCalc (version 15.0; MedCalc Software, Oostende, Belgium) software. A $p < 0.05$ was considered statistically significant.

RESULTS

Basic Information

Table 2 and **Figure 2** present the clinicopathological and imaging information of the patients, respectively.

Consistency Between Two Radiologists for Quantification

The quantification parameters measured by the two radiologists showed excellent consistency, and the ICCs of ADC, MTRasym (3.5 ppm), MD, and MK were 0.896 (95% CI = 0.834–0.935), 0.844 (95% CI = 0.752–0.903), 0.881 (95% CI = 0.809–0.925), and 0.861 (95% CI = 0.778–0.913), respectively. The average results of the two radiologists were used for the final analysis.

Assessment of Risk Factors

The differentiation of adenocarcinoma from non-adenocarcinoma in stage I EC showed significantly greater MTRasym (3.5 ppm) and MK and significantly lower ADC and MD in the non-adenocarcinoma group than those in the adenocarcinoma group (all $p < 0.05$). The AUCs of MD, MTRasym (3.5 ppm), MK, and ADC were 0.839, 0.793, 0.830, and 0.797, respectively. No statistically significant differences among these AUCs were found (**Tables 3**, **4** and **Figure 3A**).

The results for the differentiation of stages IA and IB in stage I EC are shown in **Tables 3**, **4** and **Figure 3B**. Significantly greater MTRasym (3.5 ppm) and MK but significantly lower ADC and MD were found in the stage IA group compared with those in the stage IB group (all $p < 0.05$). The AUCs of ADC, MD, MTRasym (3.5 ppm), and MK were 0.665, 0.723, 0.748, and 0.864, respectively, and there were significant differences between the AUCs of MK and ADC and between the AUCs MK and MD (Z = 2.779 and 2.074, p = 0.006 and 0.04, respectively).

TABLE 2 | Clinicopathological features of the patients.

Variable	Data
Age (years), mean ± SD	58.89 ± 7.53
Maximum diameter (mm), mean ± SD	52.07 ± 15.31
FIGO stage, *n* (%)	
IA	44 (61.11)
IIB	28 (38.89)
Histologic subtype, *n* (%)	
Adenocarcinoma	67 (93.06)
Non-adenocarcinoma	5 (6.94)
Clear cell	3 (4.17)
Serous	2 (2.77)
Lymphovascular space invasion, *n* (%)	
Present	18 (25.00)
Absent	54 (75.00)
Histological grade, *n* (%)	
Grade 1	31 (43.06)
Grade 2	24 (33.33)
Grade 3	17 (23.61)
Risk stratification, *n* (%)	
Low	41 (56.94)
Intermediate	7 (9.72)
High-intermediate	12 (16.67)
High-risk group	12 (16.67)

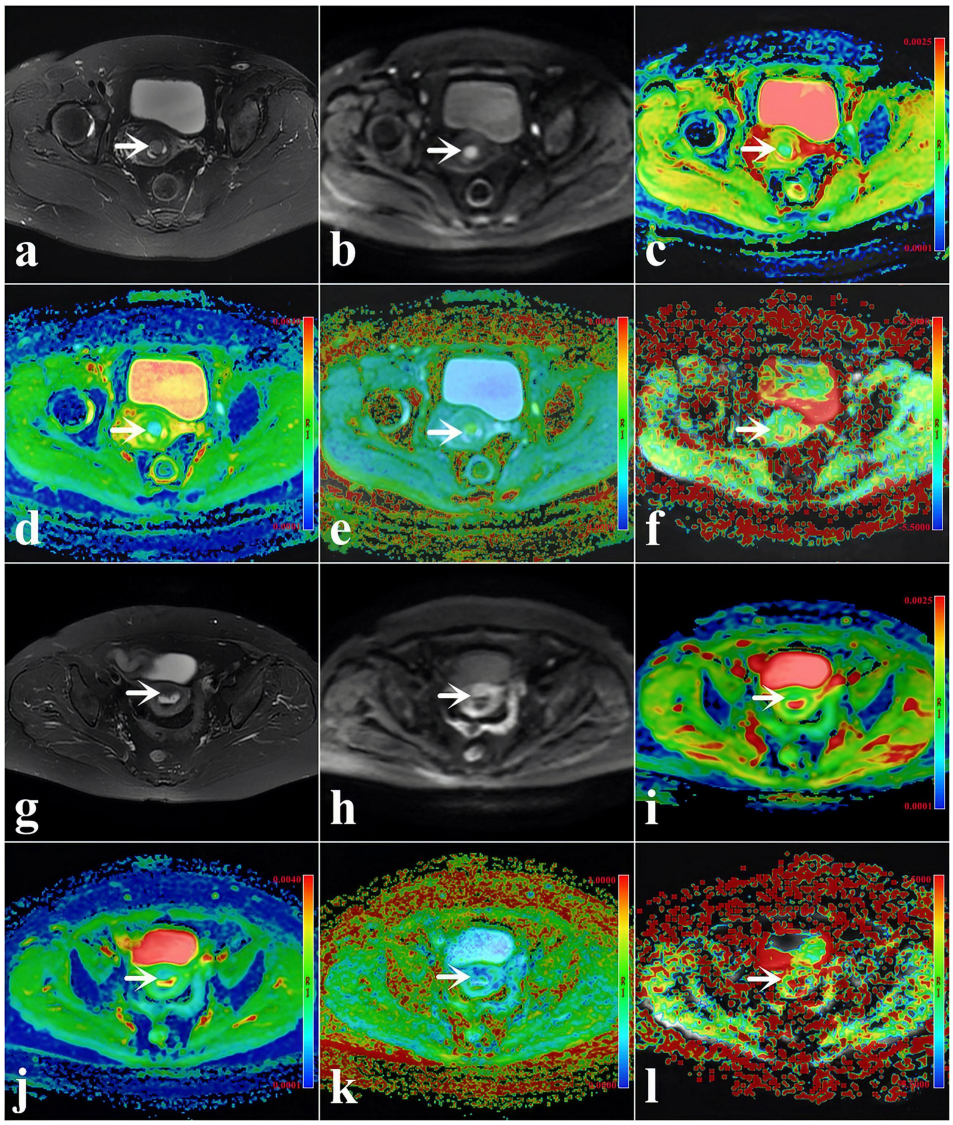


FIGURE 2 | **(A–F)** A 48-year-old woman with low-risk endometrial cancer (EC) [*arrowheads*, endometrioid type; grade 1, stage IA, lymphovascular space invasion (LVSI)-negative]. **(G–L)** A 61-year-old woman with high-intermediate-risk EC (*arrowheads*, endometrioid type; grade 3, stage IA, LVSI-positive). **(A, G)** T2-weighted imaging (T2WI) maps (fat suppression). **(B, H)** Diffusion-weighted imaging (DWI) original maps (b = 1,000 s/mm^2). **(C, I)** Pseudo-colored maps of the apparent diffusion coefficient (ADC). **(D, J)** Pseudo-colored maps of the mean diffusivity (MD). **(E, K)** Pseudo-colored maps of the mean kurtosis (MK). **(F, L)** Pseudo-colored maps of the magnetization transfer ratio asymmetry (MTRasym, at 3.5 ppm).

Compared with the quantification of the LVSI-negative group in stage I EC, the MTRasym (3.5 ppm) and MK in the LVSI-positive groups were significantly greater and the ADC and MD were significantly lower (all $p < 0.05$). The AUCs of ADC, MD, MK, and MTRasym (3.5 ppm) were 0.693, 0.698, 0.767, and 0.775, respectively, with no significant difference for each between the LVSI-negative and LVSI-positive groups (**Tables 3**, **4** and **Figure 3C**).

Quantification of the high-grade group in stage I EC was for grade III, while that of the non-high-grade group was for grades I and II. MTRasym (3.5 ppm) and MK were found significantly greater while ADC and MD were significantly lower in the high-grade group than those in the non-high-grade group (all $p < 0.05$). The AUCs of ADC, MD, MTRasym (3.5 ppm), and MK were 0.690, 0.693, 0.828, and 0.903, respectively, and significant differences were found between the AUCs of MK and ADC and between the AUCs of MK and MD (Z = 2.625 and 2.974, p = 0.008 and 0.003, respectively) (**Tables 3**, **4** and **Figure 3D**).

Assessment of Risk Stratification

MK and MTRasym (3.5 ppm) were moderately positively correlated with risk stratification, with r values of 0.640 (95% CI = 0.479–0.759, $p < 0.001$) and 0.502 (95% CI = 0.306–0.657, $p < 0.001$), respectively. ADC and MD were mildly negatively correlated with risk

TABLE 3 | Comparison of the different parameters among different groups.

Parameters	MTRasym (3.5 ppm) (%)	ADC (×10^{-3} mm^2/s)	MK	MD (×10^{-3} mm^2/s)
Histological subtype				
Adenocarcinoma	3.54 ± 0.41	0.98 (0.92–1.04)	0.79 (0.76–0.83)	1.22 (1.10–1.27)
Non-adenocarcinoma	3.96 ± 0.30	0.90 (0.86–0.92)	0.85 (0.83–0.87)	1.13 (0.99–1.15)
t/z value	2.873	−2.515	−2.448	−2.171
p-value	**0.005**[b]	**0.009**[b]	**0.011**	**0.027**[b]
FIGO stage				
IA	3.48 (3.22–3.66)	1.00 ± 0.10	0.77 (0.75–0.81)	1.22 ± 0.10
IIB	3.78 (3.52–4.14)	0.95 ± 0.06	0.84 (0.81–0.87)	1.12 ± 0.11
t/z value	−3.535	2.873	−5.181	3.524
p-value	**<0.001**[a]	**0.005**[b]	**<0.001**[a]	**0.001**[b]
Lymphovascular space invasion				
Positive	3.85 ± 0.34	0.94 ± 0.06	0.83 (0.80–0.86)	1.13 ± 0.11
Negative	3.48 ± 0.40	0.99 ± 0.09	0.78 (0.75–0.82)	1.20 ± 0.10
t/z value	3.889	−2.521	−3.375	2.604
p-value	**<0.001**[b]	**0.014**[b]	**0.001**[a]	**0.015**[b]
Histological grade				
Non-high grade (grades I and II)	3.46 ± 0.38	0.98 (0.92–1.04)	0.78 (0.75–0.81)	1.20 ± 0.10
High grade (grade III)	3.92 ± 0.32	0.93 (0.90–0.99)	0.86 (0.83–0.87)	1.12 ± 0.11
t/z value	−5.030	−2.361	−4.999	2.582
p-value	**<0.001**[b]	**0.018**[a]	**<0.001**[a]	**0.016**[b]
Risk stratification				
Low risk	3.41 (3.14–3.64)	1.01 ± 0.09	0.76 (0.74–0.80)	1.23 ± 0.09
Non-low risk (intermediate, high-intermediate, and high)	3.77 (3.51–4.13)	0.94 ± 0.06	0.83 (0.80–0.86)	1.12 ± 0.11
t/z value	−4.055	3.261	−5.101	4.328
p-value	**<0.001**[a]	**0.002**[b]	**<0.001**[a]	**<0.001**[b]

Values shown in bold denote statistical significance in the comparison.

MTRasym, magnetization transfer ratio asymmetry; ADC, apparent diffusion coefficient; MK, mean kurtosis; MD, mean diffusivity.

[a]*Comparisons performed using Mann–Whitney U test.*

[b]*Comparisons performed using independent t-test.*

TABLE 4 | Predictive performance of the different parameters.

Parameters	AUC (95% CI)	*p*-value	Cutoff	Sensitivity (%)	Specificity (%)	Youden index (%)
Histological subtype						
MTRasym (3.5 ppm) (%)	0.797 (0.686–0.883)	<0.001	3.780	77.61	80.00	57.61
ADC (×10^{-3} mm^2/s)	0.839 (0.733–0.915)	<0.001	0.928	71.64	100.00	71.64
MK	0.830 (0.723–0.908)	<0.001	0.831	76.12	80.00	56.12
MD (×10^{-3} mm^2/s)	0.793 (0.681–0.879)	<0.001	1.159	64.18	100.00	64.18
FIGO stage						
MTRasym (3.5 ppm) (%)	0.748 (0.632–0.843)	<0.001	3.750	86.36	53.57	39.94
ADC (×10^{-3} mm^2/s)	0.665 (0.544–0.772)	0.011	1.043	29.55	96.43	25.97
MK	0.864 (0.763–0.933)	<0.001	0.784	68.18	96.43	64.61
MD (×10^{-3} mm^2/s)	0.723 (0.605–0.822)	<0.001	1.233	50.00	85.71	35.71
Lymphovascular space invasion						
MTRasym (3.5 ppm) (%)	0.775 (0.662–0.865)	<0.001	3.520	57.41	94.44	51.85
ADC (×10^{-3} mm^2/s)	0.693 (0.574–0.797)	0.003	0.993	48.15	88.89	37.04
MK	0.767 (0.652–0.859)	<0.001	0.784	55.56	94.44	50.00
MD (×10^{-3} mm^2/s)	0.698 (0.578–0.801)	0.006	1.213	55.56	77.78	33.33
Histologic grade						
MTRasym (3.5ppm) (%)	0.828 (0.721–0.907)	<0.001	3.750	83.64	70.59	54.22
ADC (×10^{-3} mm^2/s)	0.690 (0.570–0.794)	0.007	1.036	29.09	94.12	23.21
MK	0.903 (0.810–0.960)	<0.001	0.815	78.18	94.12	72.30
MD (×10^{-3} mm^2/s)	0.693 (0.573–0.796)	0.006	1.217	50.91	76.47	27.38
Risk stratification						
MTRasym (3.5 ppm) (%)	0.780 (0.667–0.870)	<0.001	3.490	83.87	56.10	39.97
ADC (×10^{-3} mm^2/s)	0.709 (0.590–0.810)	<0.001	0.915	41.94	87.80	29.74
MK	0.853 (0.750–0.925)	<0.001	0.784	93.55	70.73	64.28
MD (×10^{-3} mm^2/s)	0.766 (0.652–0.858)	<0.001	1.233	87.10	53.66	40.76
Combined diagnosis	0.906 (0.814–0.962)	<0.001	–	70.97	92.68	63.65

The combined diagnosis represents MTRasym (3.5 ppm) + D + MK.

AUC, area under the receiver operating characteristic curve; MTRasym, magnetization transfer ratio asymmetry; ADC, apparent diffusion coefficient; MK, mean kurtosis; MD, mean diffusivity.

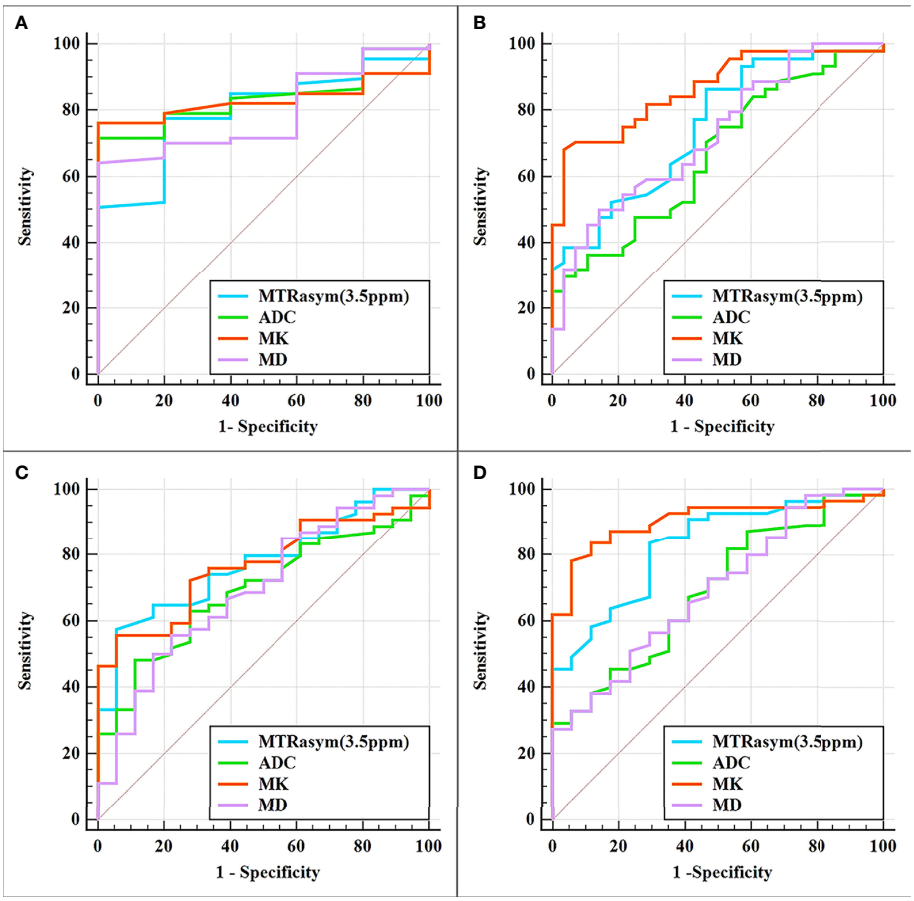


FIGURE 3 | Curves showing each parameter using receiver operating characteristic (ROC) analysis for the differentiation of adenocarcinoma and non-adenocarcinoma **(A)**, stage IA and stage IB **(B)**, LVSI-positive and LVSI-negative **(C)**, and high-grade and non-high-grade **(D)** stage I endometrial cancer (EC). Details of the area under the curves and the 95% CIs of each index are shown in **Table 5**.

stratification, with r values of −0.358 (95% CI = −0.544 to −0.138, p = 0.002) and −0.438 (95% CI = −0.608 to −0.229, $p < 0.001$), respectively (**Figure 4**). Only the difference in the r values between MK and ADC was statistically significant ($Z = 2.253$, $p = 0.024$) (**Figure 4**).

MTRasym (3.5 ppm) and MK were significantly greater while ADC and MD were significantly lower in the non-low-risk group than those in the low-risk group (all $p < 0.05$) (**Table 3**). The AUCs of ADC, MD, MTRasym (3.5 ppm), and MK increased successively, which were 0.709, 0.766, 0.780, and 0.853, respectively, but only the differences between the AUCs of MK and ADC were significant ($Z = 1.981$, $p = 0.047$) (**Tables 3**, **4** and **Figure 5A**).

The potential risk-related factors of age, tumor size, ADC, MD, MK, and MTRasym (3.5 ppm) were investigated in the logistic regression analysis to explore their value for the stratification of low- and non-low-risk stage I EC patients.

Univariate analysis showed statistical significance for ADC, MTRasym (3.5 ppm), MD, and MK as risk predictors (p-values of 0.004, <0.001, <0.001, and <0.001, respectively), while multivariate analysis revealed that MTRasym (3.5 ppm), MK, and MD were independent predictors (p-values of 0.005, 0.034, and 0.015, respectively).

The combination of the independent predictors [MD, MK, and MTRasym (3.5 ppm)] showed optimal predictive performance (AUC = 0.906, sensitivity = 70.97%, specificity = 92.68%, $p < 0.001$), which was significantly better than those of ADC (AUC = 0.709, $Z = 3.013$, $p = 0.003$), MTRasym (3.5 ppm) (AUC = 0.780, $Z = 2.852$, $p = 0.004$), and MD (AUC = 0.766, $Z = 2.787$, $p = 0.005$) individually, but not MK (AUC = 0.853, $Z = 1.414$, $p = 0.157$) (**Table 5** and **Figure 5A**).

The calibration curves generated by the analysis of bootstrapped samples are shown in **Figures 5B, C** and were used to validate the multi-parameter regression model that included MD, MK, and MTRasym (3.5 ppm). There was high consistency between the predicted and the observed risk stratification for stage I EC.

DISCUSSION

Both ADC and MD can be used to reflect the degree of the restricted diffusion movement of water molecules in tissues. Generally, the higher the density of tissue cells, the more

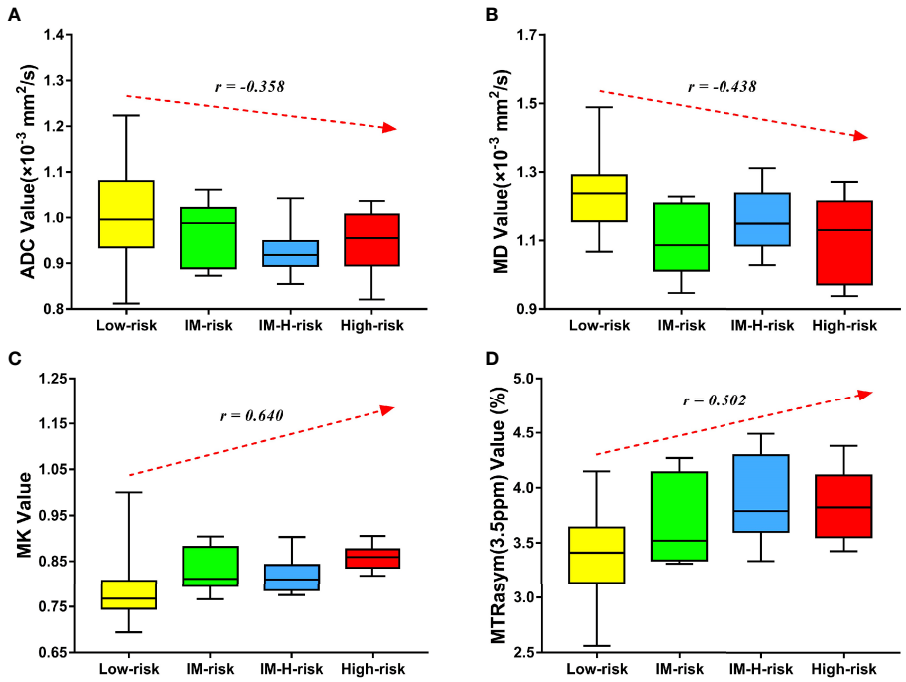


FIGURE 4 | Correlation of various parameters with risk stratification (*LM*, intermediate; *LM-H*, intermediate-high). The apparent diffusion coefficient (ADC) **(A)** and mean diffusivity (MD) **(B)** were mildly negatively correlated with risk stratification (r = –0.358 and –0.438, respectively). The mean kurtosis (MK) **(C)** and magnetization transfer ratio asymmetry (MTRasym, at 3.5 ppm) **(D)** were moderately positively correlated with risk stratification (r = 0.640 and 0.502, respectively).

significant the limitation of the diffusion movement of water molecules and, thus, the smaller the ADC and MD (11, 12). ADC and MD have been used to evaluate stage I EC in several studies. The study of An et al. showed that the ADC histogram was conducive to the evaluation of stage I EC histological subtype, grade, FIGO stage, and even risk stratification (18). Meng et al. used the average ADC and MD based on the total tumor volume for stage I EC risk stratification assessment, and the results showed that these values decreased with the increase in risk stratification; significant differences in the ADC and MD between the low-risk and non-low-risk groups were observed (19). In the present study, the ADC and MD were lower in the non-adenocarcinoma, stage IB, high-grade, and non-low-risk groups than those in the adenocarcinoma, stage IA, non-high-grade, and low-risk groups (all $p < 0.05$), which was generally consistent with the above findings. The results of both ADC and MD being lower in the LVSI-positive group than those in the LVSI-negative group were similar to the findings of Ma et al. (24) and Yamada et al. (13), suggesting that both parameters can be helpful for LVSI assessment in stage I EC. Due to the tighter tissue structure in patients of the LVSI-positive group, a more significant restriction of water molecule diffusion within it may be the main reason for the above results. MD rather than ADC in the multivariate regression analysis was found as an independent predictor, which might be related to the fact that MD was calculated by taking into account the restricted diffusion of water molecules in all directions and therefore could assess the diffusion of water molecules more accurately than ADC (12, 15).

MK is a representative parameter of DKI that is mainly used to reflect the degree of deviation from the Gaussian distribution of water molecule diffusion movement in tissues (12). Usually, malignant lesions with complex tissue structures are assumed to have a higher degree of deviation from the Gaussian distribution of water molecule diffusion movement, which means larger MK values (13, 19). Previous studies have shown that MK can provide a valid assessment of the histological type, grade, stage, and LVSI status of EC patients due to differences in the cell density, nuclear heterogeneity, and other factors (13, 24–26). However, these studies included patients with different FIGO stages of EC, so it may be difficult to provide a more definitive reference for the management of patients with stage I EC. Our results for MK in stage I EC patients for the groups in histological subtypes, grades, FIGO stages, and even LVSI status were similar to those described above. The results also showed that MK was not only effective in assessing the above risk factors but also one of the independent predictors for discriminating between non-low-risk and low-risk stage I EC patients, which was consistent with the results of the study conducted by Meng et al. (19) using the old ESMO clinical practice guidelines, indicating that MK can play a reliable role in the risk assessment of stage I EC patients.

APTWI is a MRI molecular imaging technique, and MTRasym (3.5 ppm) characterizes the heterogeneous metabolism of mobile proteins and peptides due to changes in the histopathology and genetic expression of tumors (14, 27, 28). Previous investigations have revealed that a higher MTRasym (3.5 ppm) indicated a high level of mobile protein and peptide metabolism, which was

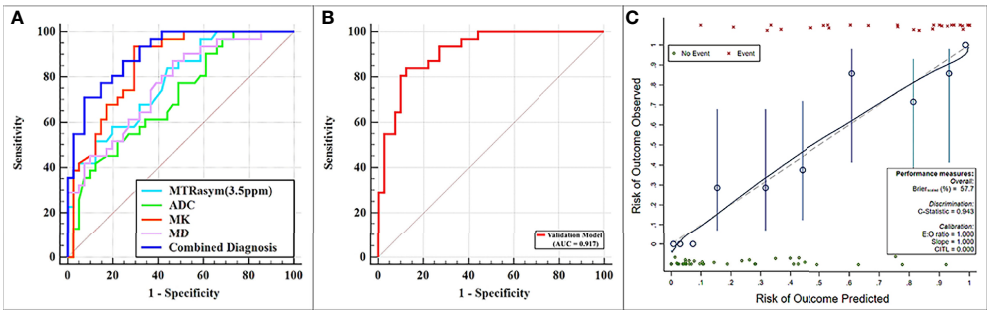


FIGURE 5 | **(A)** Receiver operating characteristic (ROC) curves showing each parameter and the combination of mean diffusivity (MD), mean kurtosis (MK), and magnetization transfer ratio asymmetry (MTRasym, at 3.5 ppm) for the differentiation of low- and non-low-risk stage I endometrial cancer (EC) patients. **(B, C)** ROC and calibration curves for predicting risk stratification of stage I EC patients in the validation model.

TABLE 5 | Univariate and multivariate analyses for the identification of low- and non-low-risk EC patients.

Parameters	Univariate analyses		Multivariate analyses	
	OR (95% CI)	*p*-value	OR (95% CI)	*p*-value
Age (years)	1.386[a] (0.855–2.247)	0.186	–	–
Tumor size (mm)	1.445[a] (0.885–2.393)	0.139	–	–
MTRasym (3.5 ppm) (%)	4.334[a] (2.009–9.349)	<0.001	3.897[a] (1.501–10.119)	**0.005**
ADC ($\times10^{-3}$ mm^2/s)	0.402[a] (0.215–0.750)	0.004	0.524[a] (0.220–1.249)	0.145
MK	4.528[a] (2.113–9.704)	<0.001	2.781[a] (1.083–7.142)	**0.034**
MD	0.304[a] (0.158–0.585)	<0.001	0.312[a] (0.121–0.799)	**0.015**

Values in bold are statistically significant.

OR, odds ratio; CI, confidence interval; MTRasym, magnetization transfer ratio asymmetry; ADC, apparent diffusion coefficient; MK, mean kurtosis; MD, mean diffusivity.

[a]*OR per 1 standard deviation.*

associated with more active cell proliferation, more microscopic necrosis (29), greater microvascular density (30), and an appropriate pH level (31). Only a few studies have explored the value of APTWI for the assessment of EC. The study by Takayama et al. showed that MTRasym (3.5 ppm) was positively correlated with the histological grade of endometrial adenocarcinoma (16), and the work by Meng et al. revealed that MTRasym (3.5 ppm) can be used to differentiate EC of different clinical types, histological grades, subtypes, and risk stratification (17, 19). In the present study, MTRasym (3.5 ppm) showed similar performance to that in the aforementioned studies in the identification of stage I EC patients with different histological subtypes, grades, and risk stratification. To our knowledge, our study is the first to conduct the evaluation of APTWI for identifying stage I EC patients with different FIGO stages and LVSI status. The higher MTRasym (3.5 ppm) in the stage IB and LVSI-positive groups was speculated to be related to the fact that the EC in these groups has more active cell proliferation, which leads to an increased content of mobile protein peptides in the tissues (13, 25).

Several limitations of this research should be taken into account. Firstly, our study was designed at a single institution with a relatively small number of patients, which may have contributed to selection bias. Secondly, both APTWI and DKI based on echo planar imaging acquisition had poor signal-to-noise ratios and low spatial resolution, making the assessment of small EC lesions difficult (largest area, <50 pixels). Thirdly, the APTWI sequence used in the present study was two-dimensional, and although we replicated the position, layer spacing, and layer thickness of the previous sequence layer by layer throughout the scanning procedure, this not only led to an increase in the scanning time but also may have introduced errors. In the future, we will include a larger population, attempt to conduct multi-institutional studies, and refine the relevant scanning techniques to make the findings more complete and reliable.

CONCLUSION

Although a similar performance was obtained with each single parameter of APTWI, DWI, and DKI for the noninvasive assessment of aggressive behavior in stage I EC, the combination of MD, MK, and MTRasym (3.5 ppm) provided improved predictive power for non-low-risk stage I EC and may serve as a superior imaging marker.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the First Affiliated Hospital of Xinxiang Medical

University. The patients/participants provided written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

DH: study concepts, and design. XJ, ZL, and GZ: literature research. WL and HW: clinical studies. MZ and RY: data analysis. XJ and ZL: manuscript preparation. JG and KW: manuscript editing. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Roentgen Imaging Research Project (HN-20201017-002) and the Key Project of Henan Province Medical Science and Technology Project (LHGJ20200519).

ACKNOWLEDGMENTS

We acknowledge the financial support from the National Natural Science Foundation of China.

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2020. *CA Cancer J Clin* (2020) 70:7–30. doi: 10.3322/caac.21590
2. Stelloo E, Nout RA, Osse EM, Jürgenliemk-Schulz IJ, Jobsen JJ, Lutgens LC, et al. Improved Risk Assessment by Integrating Molecular and Clinicopathological Factors in Early-Stage Endometrial Cancer-Combined Analysis of the PORTEC Cohorts. *Clin Cancer Res* (2016) 22:4215–24. doi: 10.1158/1078-0432.CCR-15-2878
3. Colombo N, Creutzberg C, Amant F, Bosse T, González-Martín A, Ledermann J, et al. ESMO-ESGO-ESTRO Consensus Conference on Endometrial Cancer: Diagnosis, Treatment and Follow-Up. *Ann Oncol* (2016) 27:16–41. doi: 10.1093/annonc/mdv484
4. ASTEC study group, Kitchener H, Swart AM, Qian Q, Amos C, Parmar MK. Efficacy of Systematic Pelvic Lymphadenectomy in Endometrial Cancer (MRC ASTEC Trial): A Randomised Study. *Lancet* (2009) 373:125–36. doi: 10.1016/S0140-6736(08)61766-3
5. Nougaret S, Horta M, Sala E, Lakhman Y, Thomassin-Naggara I, Kido A, et al. Endometrial Cancer MRI Staging: Updated Guidelines of the European Society of Urogenital Radiology. *Eur Radiol* (2019) 29:792–805. doi: 10.1007/s00330-018-5515-y
6. Luomaranta A, Leminen A, Loukovaara M. Magnetic Resonance Imaging in the Assessment of High-Risk Features of Endometrial Carcinoma: A Meta-Analysis. *Int J Gynecol Cancer* (2015) 25:837–42. doi: 10.1097/IGC.0000000000000194
7. Sala E, Rockall AG, Freeman SJ, Mitchell DG, Reinhold C. The Added Role of MR Imaging in Treatment Stratification of Patients With Gynecologic Malignancies: What the Radiologist Needs to Know. *Radiology* (2013) 266:717–40. doi: 10.1148/radiol.12120315
8. Werner HM, Trovik J, Marcickiewicz J, Tingulstad S, Staff AC, Engh ME, et al. A Discordant Histological Risk Classification in Preoperative and Operative Biopsy in Endometrial Cancer Is Reflected in Metastatic Risk and Prognosis. *Eur J Cancer* (2013) 49:625–32. doi: 10.1016/j.ejca.2012.09.006
9. Garcia TS, Appel M, Rivero R, Kliemann L, Wender MC. Agreement Between Preoperative Endometrial Sampling and Surgical Specimen Findings in Endometrial Carcinoma. *Int J Gynecol Cancer* (2017) 27:473–8. doi: 10.1097/IGC.0000000000000922
10. Nougaret S, Reinhold C, Alsharif SS, Addley H, Arceneau J, Molinari N, et al. Endometrial Cancer: Combined MR Volumetry and Diffusion-Weighted Imaging for Assessment of Myometrial and Lymphovascular Invasion and Tumor Grade. *Radiology* (2015) 276:797–808. doi: 10.1148/radiol.15141212
11. Jiang JX, Zhao JL, Zhang Q, Qing JF, Zhang SQ, Zhang YM, et al. Endometrial Carcinoma: Diffusion-Weighted Imaging Diagnostic Accuracy and Correlation With Ki-67 Expression. *Clin Radiol* (2018) 73:413.e1–6. doi: 10.1016/j.crad.2017.11.011
12. Jensen JH, Helpern JA, Ramani A, Lu H, Kaczynski K. Diffusional Kurtosis Imaging: The Quantification of Non-Gaussian Water Diffusion by Means of Magnetic Resonance Imaging. *Magn Reson Med* (2005) 53:1432–40. doi: 10.1002/mrm.20508
13. Yamada I, Sakamoto J, Kobayashi D, Miyasaka N, Wakana K, Oshima N, et al. Diffusion Kurtosis Imaging of Endometrial Carcinoma: Correlation With Histopathological Findings. *Magn Reson Imaging* (2019) 57:337–46. doi: 10.1016/j.mri.2018.12.009
14. Zhou J, Payen JF, Wilson DA, Traystman RJ, van Zijl PC. Using the Amide Proton Signals of Intracellular Proteins and Peptides to Detect pH Effects in MRI. *Nat Med* (2003) 9:1085–90. doi: 10.1038/nm907
15. Yue W, Meng N, Wang J, Liu W, Wang X, Yan M, et al. Comparative Analysis of the Value of Diffusion Kurtosis Imaging and Diffusion-Weighted Imaging in Evaluating the Histological Features of Endometrial Cancer. *Cancer Imaging* (2019) 19:9. doi: 10.1186/s40644-019-0196-6
16. Takayama Y, Nishie A, Togao O, Asayama Y, Ishigami K, Ushijima Y, et al. Amide Proton Transfer MR Imaging of Endometrioid Endometrial Adenocarcinoma: Association With Histologic Grade. *Radiology* (2018) 286:909–17. doi: 10.1148/radiol.2017170349
17. Meng N, Wang X, Sun J, Huang Z, Yang Z, Shang J, et al. Evaluation of Amide Proton Transfer-Weighted Imaging for Endometrial Carcinoma Histological Features: A Comparative Study With Diffusion Kurtosis Imaging. *Eur Radiol* (2021) 31:8388–98. doi: 10.1007/s00330-021-07966-y
18. An T, Kim CK. Pathological Characteristics and Risk Stratification in Patients With Stage I Endometrial Cancer: Utility of Apparent Diffusion Coefficient Histogram Analysis. *Br J Radiol* (2021) 94:20210151. doi: 10.1259/bjr.20210151
19. Meng N, Fang T, Feng P, Huang Z, Sun J, Wang X, et al. Amide Proton Transfer-Weighted Imaging and Multiple Models Diffusion-Weighted Imaging Facilitates Preoperative Risk Stratification of Early-Stage Endometrial Carcinoma. *J Magn Reson Imaging* (2021) 54:1200–11. doi: 10.1002/jmri.27684
20. Pecorelli S. Revised FIGO Staging for Carcinoma of the Vulva, Cervix, and Endometrium. *Int J Gynaecol Obstet* (2009) 105:103–4. doi: 10.1016/j.ijgo.2009.02.012
21. Shieh G. Choosing the Best Index for the Average Score Intraclass Correlation Coefficient. *Behav Res Methods* (2016) 48:994–1003. doi: 10.3758/s13428-015-0623-y
22. Xu C, Yu Y, Li X, Sun H. Value of Integrated PET-IVIM MRI in Predicting Lymphovascular Space Invasion in Cervical Cancer Without Lymphatic Metastasis. *Eur J Nucl Med Mol Imaging* (2021) 48:2990–3000. doi: 10.1007/s00259-021-05208-3
23. Chu CJ, Chan A, Song D, Staley KJ, Stufflebeam SM, Kramer MA. A Semi-Automated Method for Rapid Detection of Ripple Events on Interictal Voltage Discharges in the Scalp Electroencephalogram. *J Neurosci Methods* (2017) 277:46–55. doi: 10.1016/j.jneumeth.2016.12.009
24. Ma X, Ren X, Shen M, Ma F, Chen X, Zhang G, et al. Volumetric ADC Histogram Analysis for Preoperative Evaluation of LVSI Status in Stage I Endometrioid Adenocarcinoma. *Eur Radiol* (2022) 32:460–9. doi: 10.1007/s00330-021-07996-6
25. Song JC, Lu SS, Zhang J, Liu XS, Luo CY, Chen T. Quantitative Assessment of Diffusion Kurtosis Imaging Depicting Deep Myometrial Invasion: A Comparative Analysis With Diffusion-Weighted Imaging. *Diagn Interv Radiol* (2020) 26:74–81. doi: 10.5152/dir.2019.18366
26. Chen T, Li Y, Lu SS, Zhang YD, Wang XN, Luo CY, et al. Quantitative Evaluation of Diffusion-Kurtosis Imaging for Grading Endometrial Carcinoma: A Comparative Study With Diffusion-Weighted Imaging. *Clin Radiol* (2017) 72:995.e11–995.e20. doi: 10.1016/j.crad.2017.07.004
27. Zhuo Z, Qu L, Zhang P, Duan Y, Cheng D, Xu X, et al. Prediction of H3K27M-Mutant Brainstem Glioma by Amide Proton Transfer-Weighted Imaging and Its Derived Radiomics. *Eur J Nucl Med Mol Imaging* (2021) 48:4426–36. doi: 10.1007/s00259-021-05455-4

28. Ohno Y, Yui M, Koyama H, Yoshikawa T, Seki S, Ueno Y, et al. Chemical Exchange Saturation Transfer MR Imaging: Preliminary Results for Differentiation of Malignant and Benign Thoracic Lesions. *Radiology* (2016) 279(2):578–89. doi: 10.1148/radiol.2015151161
29. Togao O, Yoshiura T, Keupp J, Hiwatashi A, Yamashita K, Kikuchi K, et al. Amide Proton Transfer Imaging of Adult Diffuse Gliomas: Correlation With Histopathological Grades. *Neuro Oncol* (2014) 16:441–8. doi: 10.1093/neuonc/not158
30. Zheng S, van der Bom IM, Zu Z, Lin G, Zhao Y, Gounis MJ. Chemical Exchange Saturation Transfer Effect in Blood. *Magn Reson Med* (2014) 71:1082–92. doi: 10.1002/mrm.24770
31. Ray KJ, Simard MA, Larkin JR, Coates J, Kinchesh P, Smart SC, et al. Tumor pH and Protein Concentration Contribute to the Signal of Amide Proton Transfer Magnetic Resonance Imaging. *Cancer Res* (2019) 79:1343–52. doi: 10.1158/0008-5472.CAN-18-2168

Conflict of Interest: JG and KW were employed by GE Healthcare.

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

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

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 19 May 2022
doi: 10.3389/fonc.2022.856306

MRI Feature-Based Nomogram Model for Discrimination Between Non-Hypervascular Pancreatic Neuroendocrine Tumors and Pancreatic Ductal Adenocarcinomas

Jiake Xu[1†], Jie Yang[2†], Ye Feng[1], Jie Zhang[1], Yuqiao Zhang[2], Sha Chang[2], Jingqiang Jin[3] and Xia Du[2*]*

[1] Department of Gastroenterology, Kunshan Second People's Hospital, Kunshan, China, [2] Department of Radiology, The Affiliated Hospital of Guizhou Medical University, Guiyang, China, [3] Department of Radiology, Kunshan Second People's Hospital, Kunshan, China

OPEN ACCESS

Edited by:
Guang Yang,
Imperial College London,
United Kingdom

Reviewed by:
Stefano Francesco Crinò,
University of Verona, Italy
Shaofeng Duan,
GE Healthcare, China

***Correspondence:**
Xia Du
12132950@qq.com
Jingqiang Jin
568374091@qq.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: 26 January 2022
Accepted: 19 April 2022
Published: 19 May 2022

Citation:
Xu J, Yang J, Feng Y, Zhang J, Zhang Y, Chang S, Jin J and Du X (2022) MRI Feature-Based Nomogram Model for Discrimination Between Non-Hypervascular Pancreatic Neuroendocrine Tumors and Pancreatic Ductal Adenocarcinomas. Front. Oncol. 12:856306. doi: 10.3389/fonc.2022.856306

This study aimed to investigate whether magnetic resonance imaging (MRI) features could differentiate non-hypervascular pancreatic neuroendocrine tumors (PNETs) from pancreatic ductal adenocarcinomas (PDACs). In this study, 131 patients with surgically and pathologically proven non-hypervascular PNETs (n = 44) or PDACs (n = 87) were enrolled. Two radiologists independently analyzed MRI imaging findings and clinical features. Relevant features in differentiating non-hypervascular PNETs from PDACs were identified *via* univariate and multivariate logistic regression models. The MRI feature-based nomogram was constructed based on multivariable logistic analysis and the reliability of the constructed nomogram was further validated. The results showed that tumor margin (P = 0.012; OR: 6.622; 95% CI: 1.510, 29.028), MPD dilation (P = 0.047; OR: 4.309; 95% CI: 1.019, 18.227), and signal in the portal phase (P < 0.001; OR: 53.486; 95% CI: 10.690, 267.618) were independent discriminative MRI features between non-hypervascular PNETs and PDACs. The discriminative performance of the developed nomogram was optimized compared with single imaging features. The calibration curve, C-index, and DCA validated the superior practicality and usefulness of the MRI-based nomogram. In conclusion, the radiologically discriminative model integrating various MRI features could be preoperatively and easily utilized to differentiate non-hypervascular PNETs from PDACs.

Keywords: pancreatic neuroendocrine tumors, pancreatic ductal adenocarcinoma, magnetic resonance imaging, nomogram, non-hypervascular

INTRODUCTION

Pancreatic neuroendocrine tumors (PNETs) are rare and heterogeneous pancreatic tumors, which arise from pancreatic neuroendocrine cells (1, 2). PNETs have various clinical behaviors and account for about 2%–10% of all pancreatic neoplasms (3). With the development of imaging technology, the detection and diagnosis rates of PNETs have been increasing in recent years (4).

The common imaging features of PNETs have been summed up as a hypervascular and well-defined solid pancreatic mass, which are exhibited as relatively hyperenhancement in the arterial phase of computed tomography (CT) or magnetic resonance imaging (MRI) (4, 5). Nevertheless, several studies showed that up to 42% of PNETs exhibit non-enhancement in the arterial phase (6). Therefore, such overlapping imaging features between hypovascular PNETs and low-enhancement pancreatic ductal adenocarcinomas (PDACs) make it difficult to preoperatively distinguish the two tumors on imaging. Remarkably, PDACs have a relatively worse prognosis and lower survival rates and, more importantly, lower resectability rate with a wider range of excision compared with PNETs (7). Therefore, it is of great clinical significance to preoperatively discriminate between PNETs and PDACs.

Recently, several studies have focused on the difference of CT imaging features between non-hypervascular PNETs and PDACs (8, 9). For instance, Karmazanovsky et al. have reported that a series of CT imaging features, such as a well-defined margin, morphologic characteristic, and enhancement pattern, contribute to differentiate non-hypervascular PNETs from PDACs (8). Moreover, Xue's group has constructed a combined model, integrating CT-based radiomics signature and clinical–radiological features and exhibiting a better performance on the discrimination between atypical non-functional neuroendocrine tumors and PDACs (10). Unfortunately, few studies have reported the value of MRI features in discriminating non-hypervascular PNETs from PDACs. Due to its superior assessment performance for pancreatic parenchyma, pancreatic ducts, and peripancreatic soft tissues or vessels (11), MRI may be helpful in improving the differential diagnosis of non-hypervascular PNETs from PDACs. Moreover, to our knowledge, the radiological identification model integrating various MRI features for differentiating non-hypervascular PNETs from PDACs has not been reported.

Therefore, the purpose of our study was to evaluate whether MRI features are helpful to differentiate non-hypervascular PNETs from PDACs. Furthermore, we tried to develop a radiological identification model integrating significant MRI features for the precise differentiation of non-hypervascular PNETs from PDACs.

MATERIALS AND METHODS

Patient Selection

This retrospective study was approved by the Institutional Review Board of the Affiliated Hospital of Guizhou Medical University and the requirement for informed consent was waived. Between April 2012 and May 2019, all patients with surgically and pathologically proven PNETs in our hospital were enrolled. The inclusion criteria were as follows: a) patients who underwent MRI examination with a standardized MRI protocol within 2 weeks before surgery and b) patients who did not receive local or systemic treatment prior to surgery. The exclusion criteria, on the other hand, were as follows: a) missing MR imaging data or poor-quality MR images, b) patients underwent treatment before MRI examination, and c) patients diagnosed with hypervascular PNETs. The detailed data are presented in **Figure 1**. Hypervascular PNETs were defined as tumors that exhibited hyperintensity in the MRI arterial phase compared with the adjacent normal pancreas parenchyma. On the other hand, non-hypervascular PNETs are considered atypical tumors. All PNETs in this study were defined by three experienced reviewers. Finally, 44 patients with non-hypervascular PNETs were enrolled in this cohort.

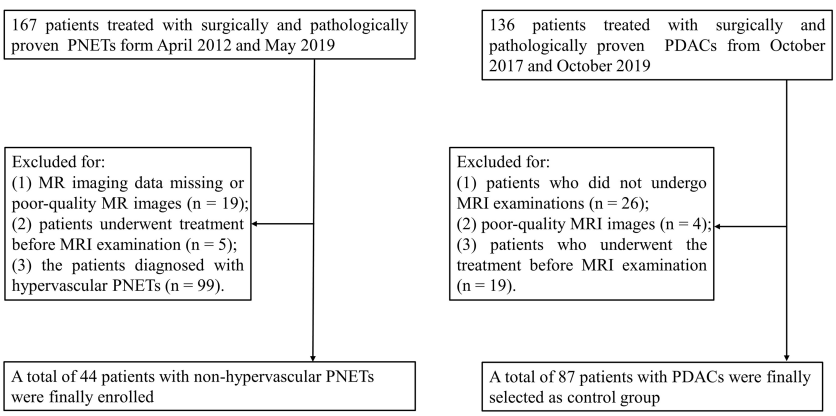


FIGURE 1 | Flowchart of the selection of patients with non-hypervascular pancreatic neuroendocrine tumors (PNETs) and pancreatic ductal adenocarcinomas (PDACs).

In addition, 136 consecutive patients with surgically and pathologically proven PDACs between May 2018 and October 2019 were enrolled as a comparison group. Among these cases, 49 patients were excluded for the following reasons: a) patients did not undergo MRI examinations (n = 26), b) patients' MRI images were of poor quality (n = 4), and c) patients underwent treatment before the MRI examination (n = 19).

MRI Protocol

All cases in this study were examined by contrast-enhanced MRI in a 1.5-T MRI scanner (Magnetom Aera; Siemens Medical Solutions, Germany) with a standardized scan protocol. The examination protocol is as follows: fat-suppressed T2-weighted two-dimensional turbo spin-echo (TSE) and diffusion-weighted imaging (DWI) with b-values = 0 and 500 s/mm^2 utilizing respiratory triggering. Three-dimensional T1-weighted volumetric interpolated breath-hold examination (VIBE) was conducted once before and three times after intravenous injection. Acquisitions were obtained at 20, 90, and 180 s after injection of gadopentetate dimeglumine at a rate of 3 ml/s and a dose of 0.1 mmol/kg during the hepatic arterial, portal, and delayed phases, respectively. All detailed MRI sequences are specifically shown in **Table 1**.

Imaging Analysis

The MRI images were acquired, evaluated, and processed *via* a picture archiving communication system (PACS) workstation. All MRI images were reviewed by two pancreatic radiologists (with 6 and 18 years of experience, respectively), who were blinded to the pathological and clinical data. Furthermore, all images were evaluated independently by two radiologists. In the event that there is inconsistency, another more experienced observer was invited for an opinion, and a majority decision was finally reached.

Qualitative data included the following: a) tumor location (head/neck, body, or tail), b) tumor consistency (solid, cystic, or solid and cystic), c) tumor margins (well-defined/ill-defined), d) main pancreatic duct (MPD) dilation (presence/absence), e) pancreatic atrophy (presence/absence), f) bile duct (BD) dilation (presence/absence), g) infiltration of peripancreatic fat (presence/absence), h) invasion of peripancreatic vessels (presence/absence), i) lymph node invasion (presence/absence), and j) signal on T2-weighted portal venous and delayed phase MRI images (defined as hypointense or iso-/hyperintense in comparison with the surrounding normal pancreatic parenchyma). MPD dilation was defined as MPD with a diameter larger than 3 mm. Invasion of peripancreatic vessels was considered based on the following: a) a mass adjoined >90° of the vascular circumference, b) occlusion of major peripancreatic arteries, and c) a mass adjoined >180° of the circumference of the portal vein (PV) or superior mesenteric vein (SMV). Lymph node invasion was considered as having at least one or more peripancreatic lymph nodes that are larger than 1 cm in diameter. Tumor consistency was divided according to the following three types: a) solid exhibiting an enhancing solid part of >90% of the mass, b) cystic exhibiting an enhancing solid part of <50% of the mass, and c) solid and cystic exhibiting an enhancing solid portion of 50%–90% of the mass.

Quantitative data analysis included a) tumor size (the maximal diameter of the tumors) and b) apparent diffusion coefficient (ADC) values (b = 500 s/mm^2). The ADC values were assessed *via* ROIs on the ADC images. All ROIs were manually drawn to include the largest part of the mass, avoiding the adjacent pancreas parenchyma, large vessels, and areas of hemorrhage or necrosis. All quantitative data were measured thrice by one experienced pancreatic radiologist, and the average values were finally used for further research.

Pathological Analysis

Histopathologic analysis of all excised lesions was performed by two experienced pancreatic pathologists. The pathological grade of all PNETs was classified according to the 2017 World Health Organization classification as follows: G1 (low grade), G2 (intermediate grade), or G3 (high grade) (12).

Construction and Validation of the MRI-Based Nomogram

In order to build a combined nomogram integrating various MRI features, we constructed a multivariable logistic regression model to identify the preoperatively discriminative radiological findings between non-hypervascular PNETs and PDACs and then integrated all significant features to construct a valuable discriminative radiological model. Furthermore, validation of the performance of the developed MRI-based nomogram was evaluated *via* the calibration curve and concordance index (C-index). The calibration curve was performed to graphically describe discriminative outcomes versus real outcomes, and the C-index was conducted to assess the discriminative performance of the developed MRI-based nomogram.

Statistical Analysis

All quantitative parameters were represented as mean ± standard deviation (SD) or median [interquartile range (IQR) = 25–75], and categorical parameters were represented as number

TABLE 1 | MRI sequences and parameters.

Parameters	Repetition time (ms)	Echo time (ms)	Section thickness (mm)	*N* excitations	Matrix	Bandwidth (Hz/pixel)	Flip angle (°)
T1-weighted imaging	6.87	2.38/4.76	4	1	320 * 240	430/490	10
T2-weighted imaging	2,400	94	5.5	2	384 * 218	194	160
Diffusion-weighted imaging	5,100	55	6	2	84 * 128	1,562	/
Contrast-enhanced imaging	4.36	2	3.5	2	320 * 195	64	10

The 1.5-T MRI imager was a Magnetom Aera (Siemens Medical Solutions) unit.

(percentage). Depending on the distribution of variables, continuous variables between two groups were analyzed *via* Student's *t*-test or the Mann–Whitney *U* test, and categorical parameters between two groups were analyzed *via* the χ^2 test or Fisher's exact test. The dependent discriminative parameters for differentiating non-hypervascular PNETs from PDACs were analyzed using receiver-operating characteristic (ROC) curve analysis, and then the corresponding sensitivity, specificity, and area under the ROC curve (AUC) were calculated. κ statistics was performed to assess the interobserver variability for categorical parameters. The grade of agreement was classified as follows: slight ($\kappa < 0.20$), fair ($\kappa = 0.21$–0.40), moderate ($\kappa = 0.41$–0.60), substantial ($\kappa = 0.61$–0.80), and outstanding ($\kappa > 0.80$).

Univariate and multivariate logistic regression analyses were used to identify the independent risk factors of the two groups. Then, a discriminative nomogram based on the significant MRI features was formulated *via* the rms package in R project. The C-index and the calibration curve were performed to assess the performance of the preliminary MRI-based nomogram. All statistical analyses were performed using SPSS software (version 22.0; IBM, Armonk, NY, USA) and R project (version 3.5.0). The tests were two-sided and *P*-value <0.05 was defined as statistically significant.

RESULTS

Patient and Tumor Characteristics

A total of 131 patients, consisting of 44 with non-hypervascular PNETs and 87 with PDACs, were finally included in this retrospective study. There were no multifocal masses in both cohorts. In the non-hypervascular PNET group, the age of the patients (median age, 55.6 ± 14.6 years) ranged from 19 to 79, and in the PDAC group, the patients' age (median age, 57.7 ± 12.6 years) ranged from 32 to 79. Furthermore, in this cohort, the ratio of female patients in the non-hypervascular PNET group was 65.9% (29/44), which was slightly larger than that in the PDAC group (52.3%, 46/87). Based on the 2017 WHO classification, among the non-hypervascular PNET cohort, 15 masses were classified as G1 (34.1%) and 65.9% masses were defined as G2/G3.

Interobserver Agreement for Qualitative MRI Features

An outstanding interobserver agreement was achieved for tumor consistency ($\kappa = 0.812$), tumor margin ($\kappa = 0.878$), MPD dilation ($\kappa = 0.924$), pancreatic atrophy ($\kappa = 0.873$), BD dilation ($\kappa = 0.819$), and signal in T2-weighted images, portal phase, and delayed phase ($\kappa = 0.846$, 0.896, and 0.876, respectively), and a substantial agreement was obtained for invasion of peripancreatic vessels ($\kappa = 0.798$) and infiltration of peripancreatic fat ($\kappa = 0.786$).

Analysis of the Predictive Factors for MR Imaging Features in Differentiating Non-Hypervascular PNETs From PDACs

To investigate the predictive value of MR imaging features in differentiating non-hypervascular PNETs from PDACs, univariate and multivariate analyses were performed, and these are shown in **Tables 2**, **3**. The univariate analysis data exhibited that tumor margin ($P < 0.001$), MPD dilation ($P < 0.001$), pancreatic atrophy ($P = 0.03$), infiltration of peripancreatic fat ($P = 0.001$), invasion of peripancreatic vessels ($P = 0.004$), and signal in the portal phase ($P < 0.001$) were considered as significantly different MR imaging features between the non-hypervascular PNETs and PDACs. Next, the above significant MRI parameters were further subjected to multivariate analysis. The multivariate analysis data showed that the well-defined tumor margin ($P = 0.012$; OR: 6.622; 95% CI: 1.510, 29.028), the absence of MPD dilation ($P = 0.047$; OR: 4.309; 95% CI: 1.019, 18.227), and hyperintensity in the portal phase ($P < 0.001$; OR: 53.486; 95% CI: 10.690, 267.618) were independent risk factors for discriminating non-hypervascular PNETs from PDACs (**Figure 2**). The ROC analysis exhibited that the sensitivity, specificity, and AUC of the tumor margin, MPD dilation, intensity in the portal phase, and the combined MR imaging features were 84.62%, 57.47%, and 0.710; 92.31%, 67.82%, and 0.801; 82.05%, 74.71%, and 0.814; and 82.05%, 86.21%, and 0.900, respectively (**Figure 3**).

Construction and Validation of the MRI-Based Nomogram for Discrimination Between Non-Hypervascular PNETs and PDACs

To develop a visual and individualized differential model, we have combined various significant MR imaging features in multivariate logistic regression to construct a novel nomogram for discriminating non-hypervascular PNETs from PDACs (**Figure 4A**). In this cohort, the C-index was calculated to assess the discriminative performance of various MRI features. The results represented that the C-index for differential diagnosis with the tri-combined nomogram was 0.914 (95% CI: 0.036, 0.134), which was larger than other C-indices for the other single or bi-combined variables (**Table 4**).

To validate the discriminative effect of the developed nomogram, the calibration curve and decision curve analysis (DCA) were performed. The calibration curve results implied better consistency between estimation and observation for the discrimination performance of the two neoplasms (**Figure 4B**). Furthermore, as shown in **Figure 5**, the DCA exhibited that the developed nomogram represented better discriminative net benefits with a broader scope of threshold probabilities compared with the single MR imaging features, implying that the MRI feature-based nomogram can serve as a more effective method for differentiating non-hypervascular PNETs from PDACs.

TABLE 2 | Clinical and radiological characteristics.

	Non-hypervascular PNETs (*n* = 44)	PDACs (*n* = 87)	*P*-value
Age (years)[a]	55.6 ± 14.6	57.7 ± 12.6	0.401
Sex			0.154
Female	29 (65.91%)	46 (52.87%)	
Male	15 (34.09%)	41 (47.13%)	
Location			0.198
Head/neck	23 (52.27%)	52 (59.77%)	
Body	4 (9.09%)	2 (2.3%)	
Tail	17 (38.64%)	33 (37.93%)	
Tumor size[a]	3.3 ± 1.4	3.7 ± 1.2	0.147
Tumor consistency			0.369
Cystic	4 (9.09%)	3 (3.45%)	
Solid	34 (77.27%)	69 (79.31%)	
Solid and cystic	6 (13.64%)	15 (17.24%)	
Tumor margin			<0.001
Well-defined	33 (75%)	27 (31.03%)	
Ill-defined	11 (25%)	60 (68.97%)	
MPD dilation			<0.001
Absence	36 (81.82%)	31 (35.63%)	
Presence	8 (18.18%)	56 (64.37%)	
Pancreatic atrophy			0.030
Absence	38 (86.36%)	60 (68.97%)	
Presence	6 (13.64%)	27 (31.03%)	
BD dilation			0.082
Absence	37 (84.09%)	61 (70.11%)	
Presence	7 (15.91%)	26 (29.89%)	
Infiltration of peripancreatic fat			0.001
Absence	33 (75%)	37 (42.53%)	
Presence	11 (25%)	50 (57.47%)	
Lymph node invasion			0.351
Absence	29 (65.91%)	50 (57.47%)	
Presence	15 (34.09%)	37 (42.53%)	
Invasion of peripancreatic vessels			0.004
Absence	35 (79.55%)	47 (54.02%)	
Presence	9 (20.45%)	40 (45.98%)	
Signal in T2-weighted images			0.292
Hypointense	1 (2.27%)	1 (1.15%)	
Isointense	13 (14.94%)	16 (18.39%)	
Hyperintense	30 (68.18%)	70 (80.46%)	
Signal in the portal phase			<0.001
Hypointense	8 (18.18%)	69 (79.31%)	
Isointense	15 (34.09%)	10 (11.49%)	
Hyperintense	21 (47.73%)	8 (9.2%)	
Signal in the delayed phase			0.061
Hypointense	66 (150%)	26 (29.89%)	
Isointense	20 (45.45%)	15 (17.24%)	
Hyperintense	1 (2.27%)	3 (3.45%)	
ADC ($\times10^{-3}$ mm^2/s)[b]	1.25 (0.81–1.49)	1.16 (0.80–1.48)	0.153

Unless otherwise indicated, data are the number of lesions, with percentage in parentheses.

[a]*Data are expressed as mean ± standard deviation.*

[b]*Data are expressed as [interquartile range (IQR) = 25–75].*

ADC, apparent diffusion coefficient; MPD, main pancreatic duct; BD, bile duct.

DISCUSSION

In the present study, we not only investigated the performance of MRI features for discrimination of non-hypervascular PNETs and PDACs but also constructed and validated a more practical differential diagnosis model merging diverse MRI findings, which showed better diagnostic efficiency than separated MRI features. In this cohort, tumor margin, MPD dilation, and intensity in the portal phase were significantly discriminative MRI features between non-hypervascular PNETs and PDACs.

Several studies have reported that about 20%~43% of PNETs exhibited iso- or hypointensity in the arterial phase of contrast-enhanced MDCT or MRI, which is similar to the imaging findings in the arterial phase of pancreatic cancer, leading to misdiagnosis (13, 14). In our study, 26.3% (44/167) of PNET patients showed non-hyperintensity in the arterial phase of

TABLE 3 | Univariate and multivariate analyses for relevant MRI features for differentiating non-hypervascular PNETs and PDACs.

Risk factors	Univariate analysis			Multivariate analysis		
	OR	95% CI	*P*-value	OR	95% CI	*P*-value
Age (years)	0.988	0.962, 1.016	0.398			
Sex						
Female	1.723	0.812, 3.656	0.156			
Male[a]						
Location						
Head/neck[a]						
Body	4.522	0.773, 26.465	0.094			
Tail	1.165	0.543, 2.500	0.696			
Tumor size	0.776	0.572, 1.051	0.102			
Tumor consistency						
Cystic[a]						
Solid	1.232	0.439, 3.457	0.692			
Solid and cystic	3.333	0.567, 19.593	0.183			
Tumor margin (well-defined)	6.667	2.937, 15.132	**<0.001**	6.622	1.510, 29.028	**0.012**
MPD dilation (absence)	9.482	3.900, 23.053	**<0.001**	4.309	1.019, 18.227	**0.047**
BD dilation (absence)	2.253	0.890, 5.705	0.087			
Pancreatic atrophy (absence)	2.850	1.007, 7.544	**0.035**	1.947	0.423, 8.961	0.392
Infiltration of peripancreatic fat (absence)	4.054	1.814, 9.058	**0.001**	2.892	0.586, 29.028	0.192
Lymph node invasion (absence)	1.431	0.673, 3.042	0.352			
Invasion of peripancreatic vessels (absence)	3.310	1.421, 7.706	**0.006**	1.025	0.179, 5.875	0.978
Signal in T2-weighted images			0.292			
Hypointense[a]						
Isointense	1.896	0.812, 4.425	0.139			
Hyperintense	2.333	0.141, 38.548	0.554			
Signal in the portal phase						
Hypointense[a]						
Isointense	12.937	4.374, 38.268	**<0.001**	28.298	5.956, 134.463	**<0.001**
Hyperintense	22.641	7.547, 67.675	**<0.001**	53.486	10.690, 267.618	**<0.001**
Signal in the delayed phase						
Hypointense[a]						
Isointense	1.904	0.848, 4.274	0.119			
Hyperintense	7.615	0.757, 76.584	0.085			
ADC ($\times10^{-3}$ mm^2/s)	1.993	0.771, 5.155	0.155			

[a]*Data were utilized as the reference variable.*

OR, odds ratio; CI, confidence interval; ADC, apparent diffusion coefficient; MPD, main pancreatic duct; BD, bile duct.

The values provided in bold type mean <0.05 and statistically significant.

dynamic enhanced MRI, which coincided with previous studies (8). Among these cases, 68.8% of non-hypervascular PNETs were classified as G2/G3, which coincided with previous studies (8).

In this study, a well-defined tumor margin has been more often shown in the non-hypervascular PNETs than in PDACs. In a recent research by Jeon et al., the authors reported that tumor margin can serve as a discriminatively morphologic characteristic between non-hypervascular PNETs and PDACs (15). Likewise, Karmazanovsky et al. have also suggested that smooth and regular margins are more commonly observed in non-hypervascular PNETs in comparison with PDACs (58% vs. 25%) (8). This may be partly explained by the fact that non-hypervascular PNETs show less infiltration into surrounding tissues than PDACs.

In addition, the absence of MPD dilation, in our cohort, is another significant MRI feature to discriminate non-hypervascular PNETs from PDACs, which may be associated with the location of origin of the two tumors. Notably, PNETs originated from progenitor islet cells of pancreatic parenchyma. However, PDACs originated from the ductal epithelium of the pancreas, which is more likely to infiltrate into the pancreatic duct and cause dilation or obstruction of the pancreatic duct. Several recent studies have shown that the absence of MPD dilation in both CT and MRI features is more likely to occur in PNETs, which is in agreement with our research results (16, 17).

Furthermore, our study also investigated that, compared with PDACs, iso- or hyperintensity in the portal phase in the dynamic enhanced MRI more commonly appeared in non-hypervascular PNETs. A recent study has also reported that there is a significant difference in the enhancement patterns of MRI between non-hypervascular PNETs and PDACs (15). Moreover, in that study, the enhancement degree of the portal phase in the non-hypervascular PNETs was obviously higher than that in the PDACs, which is consistent with our study's results. Similarly, in another study, hyperenhancement in the portal venous phase and persistent iso-enhancement were the significant independent CT features of non-hypervascular PNETs (8). This can be attributed to the pathological nature of PDACs (having lower vascularity and a higher rate of fibrosis compared with PNETs). Furthermore, in our cohort, other MRI features, including pancreatic atrophy, BD dilation, infiltration of peripancreatic fat, and invasion of peripancreatic vessels, are not independent

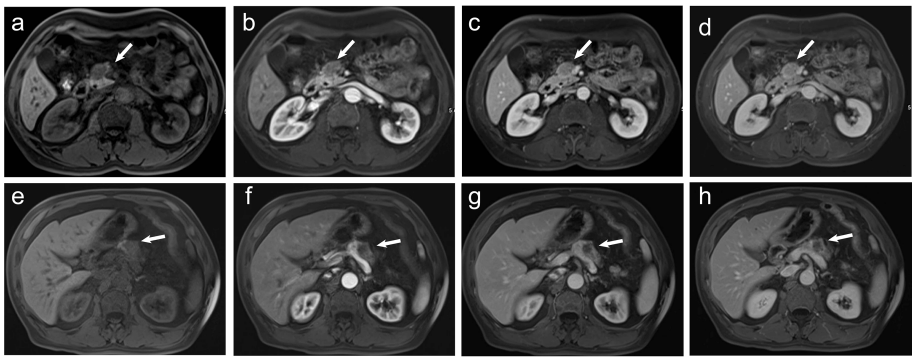


FIGURE 2 | **(A–D)** A 60-year-old man pathologically diagnosed with G2 non-hypervascular PNETs. **(A)** T1-weighted imaging shows a well-defined hypointense mass in the head of the pancreas (white arrow). **(B)** The mass in the arterial phase shows hypointensity (white arrow). **(C, D)** The mass in both portal **(C)** and delayed phases **(D)** shows relative isointensity (white arrow). **(E–H)** A 60-year-old man pathologically diagnosed with PDACs. **(E)** T1-weighted imaging shows an ill-defined hypointense mass in the body of the pancreas (white arrow). **(F)** The mass in the arterial phase shows hypointensity (white arrow). **(G)** The mass in the portal phase shows obvious hypointensity with mild pancreatic duct dilation (white arrow). **(H)** The mass in the delayed phase shows obvious hypointensity (white arrow).

predictors for the discrimination of non-hypervascular PNETs and PDACs. However, in two other recent studies, peripancreatic infiltration and pancreatic parenchymal atrophy are discriminative CT features between PNETs and PDACs, which may be caused by the different sample sizes and inclusion criteria (8, 18).

Although there are overlapping imaging findings between non-hypervascular PNETs and PDACs, the treatment strategies and prognosis are totally different. Specifically, compared with PNETs, thorough surgical approaches, such as the Whipple procedure or pylorus-preserving pancreatoduodenectomy, are more beneficial to patients. Undoubtedly, preoperative imaging discrimination of non-hypervascular PNETs from PDACs is of great importance in developing treatment strategies, improving patient outcomes (4, 15, 19). For example, previous studies reported that endoscopic ultrasound (EUS) features are correlated with malignancy of non-hypovascular solid pancreatic tumors, and EUS tissue acquisition (EUS-TA) is helpful for obtaining pathological results before surgery. However, such imaging strategies are still invasive and complex (20, 21). Thus, in this study, a preliminary MRI-based nomogram merging the various significant MRI parameters

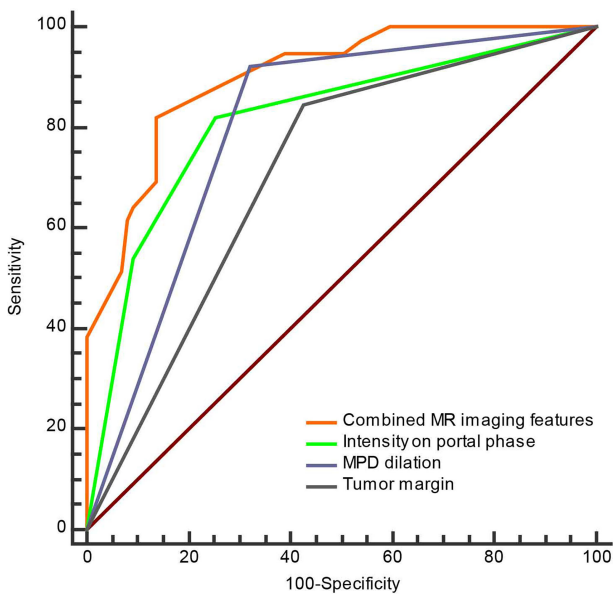


FIGURE 3 | The ROC curve of separate three MRI features and combined MRI features for discrimination of the non-hypervascular PNETs from PDACs.

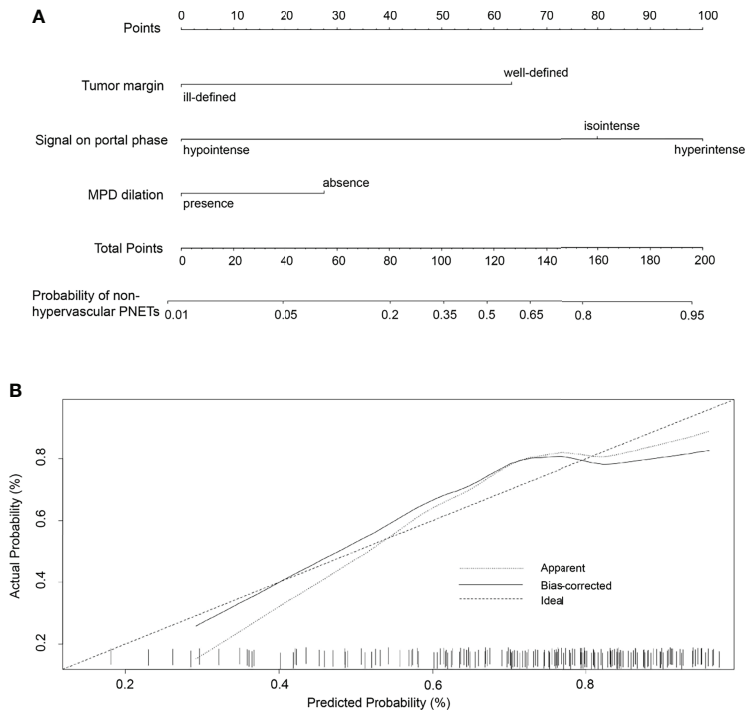


FIGURE 4 | **(A)** The developed nomogram integrating three statistically significant MRI features to discriminate non-hypervascular PNETs from PDACs. **(B)** The calibration curve of the MRI-based nomogram to discriminate non-hypervascular PNETs from PDACs.

derived from multivariate regression analysis was developed for the individualized discrimination of non-hypervascular PNETs from PDACs. The developed nomogram is of great clinical significance because it eliminates the complicated equation and the calculation of the regression analysis model and enables clinicians to intuitively and graphically calculate the probability of disease (22). As far as we know, this is the first time that a radiologically user-friendly model was constructed combining diverse MR imaging findings to improve differential diagnostic performance. Furthermore, the preliminarily developed model was further assessed by the calibration curve, C-index, and DCA, to determine its practicality and accuracy. As shown by the results, the calibration curve represented a favorable coherence between the nomogram-estimated and the actually observed probability, and the C-index validated that this developed nomogram optimized the accuracy of discrimination. Meanwhile, DCA showed the best clinical benefit with a wider range of threshold probability, guaranteeing the dependability of the developed nomogram.

This study also has several limitations. First, there was inevitably inherent selection bias because only cases with surgically resected PNETs and PDACs were enrolled in this retrospective study. Second, due to some technical difficulties, this study only enrolled a relatively large sample of non-hypervascular PNETs from a single center; thus, our results may not represent the true spectrum of PNETs. Third, owing to the limitation of the number of cases, internal and external validation were not performed. Fourth, since non-hypervascular PNETs are relatively rare, we did not create a test group and a validation group nor included an external validation group. Although a very clinically valuable discrimination model was constructed in this study, it still needs to be verified by further

TABLE 4 | Discriminatory capabilities of the nomogram and independent MRI features.

Factors	C-index	95% CI
Tumor margin	0.719	0.199, 0.361
Signal in the portal phase	0.817	0.109, 0.256
MPD dilation	0.731	0.193, 0.345
Nomogram incorporating (tumor margin + signal in the portal phase)	0.896	0.044, 0.162
Nomogram incorporating (tumor margin + MPD dilation)	0.779	0.147, 0.294
Nomogram incorporating (signal in the portal phase + MPD dilation)	0.874	0.066, 0.185
Nomogram incorporating (tumor margin + signal in the portal phase + MPD dilation)	0.914	0.036, 0.134

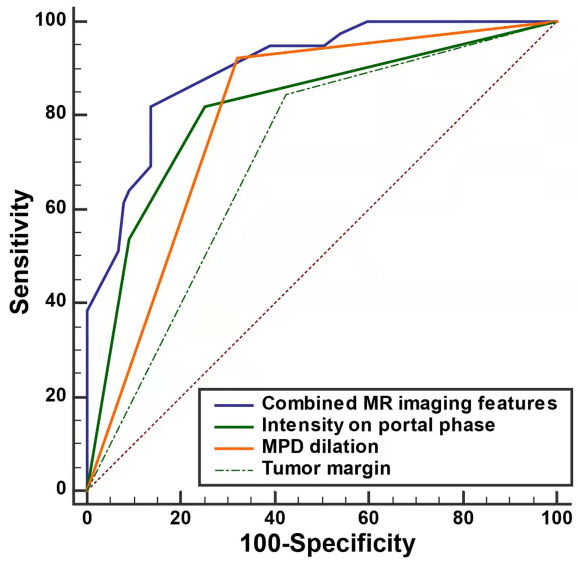


FIGURE 5 | The decision curve analysis for the MRI-based nomogram. The red line displays the integrated MRI features, the blue line the signal in the portal phase, the purple line the MPD dilation, the green line the tumor margin, the gray line the hypothesis that all patients had non-hypervascular PNETs, and the black line the hypothesis that all patients had PDACs. The *x*-axis displays the high-risk threshold and the *y*-axis represents the net benefit.

experimental studies. Thus, to validate the results of this study in the future, a much larger sample and multicenter data should be utilized. Finally, although the practicability and accuracy of the developed model were assessed, this tool did not integrate another clinical variable. Therefore, this tool should be further improved and validated in another cohort.

In conclusion, we thoroughly investigated the significantly useful MRI features, including tumor margin, MPD dilation, and intensity in the portal phase, to discriminate non-hypervascular PNETs from PDACs. Notably, a radiologically discriminative nomogram incorporating diverse MRI parameters was constructed and validated, which may improve the efficiency and accuracy of diagnosis and provide more efficient communication among radiologists, clinicians, and patients.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board of the Affiliated Hospital of Guizhou Medical University. Written informed consent for participation was not required for this study in accordance with national legislation and institutional requirements. 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

JX and XD participated in the design of the study. JY, YF, JZ, YZ, and SC were responsible for collecting the patients' data. JX and JY performed the statistical analysis. JX and JY wrote the manuscript. JJ and XD revised the manuscript. All authors contributed to the article and agreed to the submitted version.

REFERENCES

1. Falconi M, Eriksson B, Kaltsas G, Bartsch DK, Capdevila J, Caplin M, et al. Enets Consensus Guidelines Update for the Management of Patients With Functional Pancreatic Neuroendocrine Tumors and Non-Functional Pancreatic Neuroendocrine Tumors. *Neuroendocrinology* (2016) 103 (2):153–71. doi: 10.1159/000443171
2. Ma ZY, Gong YF, Zhuang HK, Zhou ZX, Huang SZ, Zou YP, et al. Pancreatic Neuroendocrine Tumors: A Review of Serum Biomarkers, Staging, and Management. *World J Gastroenterol* (2020) 26(19):2305–22. doi: 10.3748/wjg.v26.i19.2305
3. Ito T, Igarashi H, Jensen RT. Pancreatic Neuroendocrine Tumors: Clinical Features, Diagnosis and Medical Treatment: Advances. *Best Pract Res Clin Gastroenterol* (2012) 26(6):737–53. doi: 10.1016/j.bpg.2012.12.003
4. Sun H, Zhou J, Liu K, Shen T, Wang X, Wang X. Pancreatic Neuroendocrine Tumors: Mr Imaging Features Preoperatively Predict Lymph Node Metastasis. *Abdom Radiol (NY)* (2019) 44(3):1000–9. doi: 10.1007/s00261-018-1863-y
5. Gandhi NS, Feldman MK, Le O, Morris-Stiff G. Imaging Mimics of Pancreatic Ductal Adenocarcinoma. *Abdom Radiol (NY)* (2018) 43(2):273–84. doi: 10.1007/s00261-017-1330-1
6. Khanna L, Prasad SR, Sunnapwar A, Kondapaneni S, Dasyam A, Tammisetti VS, et al. Pancreatic Neuroendocrine Neoplasms: 2020 Update on Pathologic and Imaging Findings and Classification. *Radiographics* (2020) 40(5):1240–62. doi: 10.1148/rg.2020200025
7. Iovanna J, Dusetti N. Speeding Towards Individualized Treatment for Pancreatic Cancer by Taking an Alternative Road. *Cancer Lett* (2017) 410:63–7. doi: 10.1016/j.canlet.2017.09.016
8. Karmazanovsky G, Belousova E, Schima W, Glotov A, Kalinin D, Kriger A. Nonhypervascular Pancreatic Neuroendocrine Tumors: Spectrum of Mdct Imaging Findings and Differentiation From Pancreatic Ductal Adenocarcinoma. *Eur J Radiol* (2019) 110:66–73. doi: 10.1016/j.ejrad.2018.04.006
9. Ohki K, Igarashi T, Ashida H, Shiraishi M, Nozawa Y, Ojiri H. Differentiation Between Non-Hypervascular Pancreatic Neuroendocrine Tumour and Pancreatic Ductal Adenocarcinoma on Dynamic Computed Tomography and Non-Enhanced Magnetic Resonance Imaging. *Pol J Radiol* (2019) 84: e153–61. doi: 10.5114/pjr.2019.84193
10. He M, Liu Z, Lin Y, Wan J, Li J, Xu K, et al. Differentiation of Atypical Non-Functional Pancreatic Neuroendocrine Tumor and Pancreatic Ductal Adenocarcinoma Using Ct Based Radiomics. *Eur J Radiol* (2019) 117:102–11. doi: 10.1016/j.ejrad.2019.05.024

11. Petitclerc L, Sebastiani G, Gilbert G, Cloutier G, Tang A. Liver Fibrosis: Review of Current Imaging and Mri Quantification Techniques. *J Magn Reson Imaging* (2017) 45(5):1276–95. doi: 10.1002/jmri.25550
12. Inzani F, Petrone G, Rindi G. The New World Health Organization Classification for Pancreatic Neuroendocrine Neoplasia. *Endocrinol Metab Clin North Am* (2018) 47(3):463–70. doi: 10.1016/j.ecl.2018.04.008
13. Dromain C, Deandreis D, Scoazec JY, Goere D, Ducreux M, Baudin E, et al. Imaging of Neuroendocrine Tumors of the Pancreas. *Diagn Interv Imaging* (2016) 97(12):1241–57. doi: 10.1016/j.diii.2016.07.012
14. Wang Y, Chen X, Wang J, Cui W, Wang C, Chen X, et al. Differentiation Between Non-Hypervascular Pancreatic Neuroendocrine Tumors and Mass-Forming Pancreatitis Using Contrast-Enhanced Computed Tomography. *Acta Radiol* (2021) 62(2):190–7. doi: 10.1177/0284185120921503
15. Jeon SK, Lee JM, Joo I, Lee ES, Park HJ, Jang JY, et al. Nonhypervascular Pancreatic Neuroendocrine Tumors: Differential Diagnosis From Pancreatic Ductal Adenocarcinomas at Mr Imaging-Retrospective Cross-Sectional Study. *Radiology* (2017) 284(1):77–87. doi: 10.1148/radiol.2016160586
16. Ren S, Chen X, Wang Z, Zhao R, Wang J, Cui W, et al. Differentiation of Hypovascular Pancreatic Neuroendocrine Tumors From Pancreatic Ductal Adenocarcinoma Using Contrast-Enhanced Computed Tomography. *PloS One* (2019) 14(2):e0211566. doi: 10.1371/journal.pone.0211566
17. Xiao B, Jiang ZQ, Hu JX, Zhang XM, Xu HB. Differentiating Pancreatic Neuroendocrine Tumors From Pancreatic Ductal Adenocarcinomas by the "Duct-Road Sign": A Preliminary Magnetic Resonance Imaging Study. *Med (Baltimore)* (2019) 98(35):e16960. doi: 10.1097/MD.0000000000016960
18. Guo C, Zhuge X, Wang Q, Xiao W, Wang Z, Wang Z, et al. The Differentiation of Pancreatic Neuroendocrine Carcinoma From Pancreatic Ductal Adenocarcinoma: The Values of Ct Imaging Features and Texture Analysis. *Cancer Imaging* (2018) 18(1):37. doi: 10.1186/s40644-018-0170-8
19. Hain E, Sindayigaya R, Fawaz J, Gharios J, Bouteloup G, Soyer P, et al. Surgical Management of Pancreatic Neuroendocrine Tumors: An Introduction. *Expert Rev Anticancer Ther* (2019) 19(12):1089–100. doi: 10.1080/14737140.2019.1703677
20. Paiella S, Landoni L, Rota R, Valenti M, Elio G, Crino SF, et al. Endoscopic Ultrasound-Guided Fine-Needle Aspiration for the Diagnosis and Grading of Pancreatic Neuroendocrine Tumors: A Retrospective Analysis of 110 Cases. *Endoscopy* (2020) 52(11):988–94. doi: 10.1055/a-1180-8614
21. Crino SF, Brandolese A, Vieceli F, Paiella S, Conti Bellocchi MC, Manfrin E, et al. Endoscopic Ultrasound Features Associated With Malignancy and Aggressiveness of Nonhypovascular Solid Pancreatic Lesions: Results From a Prospective Observational Study. *Ultraschall Med* (2021) 42(2):167–77. doi: 10.1055/a-1014-2766
22. Lai J, Lin J, Wang H, Sun Y, Li Y, Tian H, et al. Establishment and Validation of a Preoperative Mri-Based Nomogram for Predicting the Risk of Malignancy in Patients With Breast Tumor. *J Cancer* (2021) 12(3):799–806. doi: 10.7150/jca.49441

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: 30 May 2022
doi: 10.3389/fonc.2022.868823

The Role of Preoperative ^{18}F-fluorodeoxyglucose Positron Emission Tomography/ Computed Tomography in Retroperitoneal Sarcoma

OPEN ACCESS

Edited by:
Georgios S. Limouris, National and Kapodistrian University of Athens, Greece

Reviewed by:
Satvinder Singh Mudan, Imperial College London, United Kingdom
Robert J. Canter, University of California, Davis, United States

***Correspondence:**
Kyo Won Lee
Leekw1980.lee@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 February 2022*
Accepted: *03 May 2022*
Published: *30 May 2022*

Citation:
Jo SJ, Kim KD, Lim SH, Kim J, Hyun SH, Park JB and Lee KW (2022) The Role of Preoperative ^{18}F-fluorodeoxyglucose Positron Emission Tomography/Computed Tomography in Retroperitoneal Sarcoma. Front. Oncol. 12:868823. doi: 10.3389/fonc.2022.868823

***Sung Jun Jo*[1], *Kyeong Deok Kim*[1], *So Hee Lim*[1], *Jinseob Kim*[2], *Seung Hyup Hyun*[3], *Jae Berm Park*[1] *and Kyo Won Lee*[1*]**

[1] Department of Surgery, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea, [2] Department of Epidemiology, School of Public Health, Seoul National University, Seoul, South Korea, [3] Department of Nuclear Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea

^{18}F-fluorodeoxyglucose positron emission tomography/computed tomography (^{18}F-FDG PET/CT) was used to predict pathologic grades based on the maximum standardized uptake value (SUVmax) in soft tissue sarcoma and bone sarcoma. In retroperitoneal sarcoma (RPS), the effectiveness of PET was not well known. This study was designed to investigate the association of SUVmax with histopathologic grade and evaluate the usefulness of ^{18}F-FDG PET/CT before operation. Patients at Samsung Medical Center undergoing primary surgery for retroperitoneal sarcoma with preoperative ^{18}F-FDG PET/CT imaging between January 2001 and February 2020 were investigated. The relationship between SUVmax and histologic features was assessed. The association of SUVmax with overall survival (OS), local recurrence (LR), and distant metastasis (DM) were studied. Of the total 129 patients, the most common histologic subtypes were liposarcoma (LPS; 68.2%) and leiomyosarcoma (LMS; 15.5%). The median SUVmax was 4.5 (range, 1- 29). Moreover, SUVmax was correlated with tumor grade ($p < 0.001$, Spearman coefficient; 0.627) and mitosis ($p < 0.001$, Spearman coefficient; 0.564) and showed a higher value in LMS (12.04 ± 6.73) than in dedifferentiated liposarcoma (DDLPS; 6.32 ± 4.97, $p = 0.0054$). SUVmax was correlated with pathologic parameters (tumor grade and mitosis) in RPS and was higher in the LMS group than the DDLPS group. The optimal SUVmax threshold to distinguish high tumor grade was 4.8. Those with a SUVmax greater than the threshold showed poor prognosis regarding OS, LR, and DM ($p < 0.001$).

Keywords: PET, RPS, SUVmax, LPS, DDLPS

INTRODUCTION

Retroperitoneal sarcoma (RPS) is a rare neoplasm of mesenchymal origin derived from connective tissue. The most common histologic types are liposarcoma (LPS) and leiomyosarcoma (LMS), which account for 70% of all RPSs (1, 2).

Researches on the optimal treatment of RPS are in progress. Hospitals in many countries around the world are conducting continuous research together (2–4). For example, Almond, L.M., et al. was reported that neoadjuvant chemotherapy can improve the likelihood of negative resection margins in patients with locally advanced and high-risk primary sarcomas (5). Bonvalot, S., et al. was reported that preoperative radiotherapy had no clinical benefit on RPS (6). However, peri-operative treatments on RPS are still controversial and surgical resection, including that of adjacent organs, is accepted as the standard treatment (7–9).

Preoperative diagnosis and identification of tumor extent are important to determine extensive surgical resection including adjacent organs. Percutaneous biopsy and computed tomography (CT) are robust preoperative diagnosis methods and can safely determine histologic subtype and presence of metastasis (10, 11). However, percutaneous biopsy has limitations in that the accuracy is low (67.2%) and it is difficult to distinguish the tumor grade (12). In addition, CT scan has the disadvantage of being inaccurate in discriminating histologic subtypes of heterogenous tumors (13). Due to these limitations, diagnostic tool that can increase the accuracy of diagnosis is needed. The ^{18}F-fluorodeoxyglucose positron emission tomography/computed tomography (^{18}F-FDG PET/CT) can play a complementary role as it differentiate high-grade portion of heterogenous tumors and perform targeted biopsy (14).

There have been several studies on the use of ^{18}F-FDG PET/CT in sarcomas, but most included both bone sarcoma and soft tissue sarcoma (15, 16). Alternatively, whole soft tissue sarcomas not specific to RPS have also been targeted (17, 18). Previously, our research team conducted a study on the association between maximum standardized uptake value (SUVmax) and retroperitoneal LPS (19). However, there was a limitation that only LPS was included.

In this study, we aimed to investigate the prognostic significance of SUVmax in RPS and to find out whether SUVmax shows different values depending on the histologic subtypes.

METHODS

Patients

We retrospectively investigated patients undergoing primary surgery for RPS with preoperative ^{18}F-FDG PET/CT imaging at Samsung Medical Center between January 2001 and February 2020. The diagnoses were determined according to the World Health Organization 2013 classification of specimens collected during surgery by pathologists specialized in sarcoma. The following patients were excluded: pediatric patients (those under 19 years of age); patients diagnosed with another malignant disease; patients who received pre-operative treatment, such as chemo-radiation therapy, before obtaining PET imaging; patients diagnosed with distant metastasis; and patients diagnosed with visceral sarcoma (tumors that clearly originated from a visceral organ, such as uterine sarcoma and sarcoma of the prostate, bladder, or vesicles), a benign tumor, carcinosarcoma, or a gastrointestinal tumor.

Data on underlying diseases, gender, BMI, and surveillance, such as [overall survival (OS), local recurrence (LR), and distant metastasis (DM)] were collected from patient medical records.

Pathologic Characteristics

All pathologic records, based on surgical specimens, were reviewed by specialized sarcoma pathologists. Tumor histologic subtype, size, mitosis, necrosis, and multifocality were analyzed. Tumor grade was determined using the French Federation of Cancer Centers Sarcoma Group Grading System (FNCLCC).

^{18}F-FDG PET/CT Imaging

All ^{18}F-FDG PET/CT images were taken to confirm metastasis to other organs before surgery, and interpretations were made by nuclear medicine specialists. All patients fasted for at least 6 hours before PET/CT imaging, and their blood glucose level was required to be less than 200 mg/dL. Whole-body PET and unenhanced CT images were acquired using a PET/CT scanner (Discovery STE, GE Healthcare, Waukesha, WI, USA). Whole-body CT was performed using a 16-slice helical CT with 30 to 170 mAs adjusted to the patient's body weight at 140-kVp and 3.75-mm section width. After the CT scan, an emission scan was performed from the thigh to the basal skull for 2.5 min per frame in three-dimensional mode 60 minutes after intravenous ^{18}F-FDG injection (5.0 MBq/kg). The ordered subsets expectation-maximization algorithm (20 subsets and 2 iterations) with a 128 $\times$128 matrix and voxel size of 3.9 $\times$ 3.9 $\times$ 3.3 mm was used to reconstruct PET images utilizing CT data to correct attenuation. Regarding SUVmax measurement, we placed a spherical volume of interest with a diameter of 3 cm at a location where the tumor tissue had the highest metabolic activity using Volume Viewer (Advantage Workstation 4.4, GE Healthcare). SUVmax was normalized to patient body weight.

Statistical Analysis

Factors affecting the prognosis of RPS were analyzed through univariate and multivariate Cox regression models. The Cox proportional hazards model was used to evaluate prognostic variables, and an estimated hazard ratio (HR) with its 95% confidence interval (95% CI) was presented. $P < 0.05$ was considered to represent a statistically significant comparison.

The Analysis of Variance (ANOVA) test was used to analyze the correlation between SUVmax and histologic subtypes. The receiver-operating characteristic (ROC) methodology was used to calculate the ideal threshold to distinguish high-grade sarcoma. The area under the curve (AUC) was calculated for each parameter using the non-parametric method to represent the overall predictive or prognostic performance.

Regarding survival analysis, Kaplan-Meier estimates and the log-rank test were used, and OS, LR, and DM were analyzed using time-to-event regression. Specifically, OS was calculated from the date of surgery to the date of death, LR was identified in CT scans, and the duration was calculated based on the CT scan date. DM was defined as a tumor found in organs such as liver,

lung, brain, and bone, and the date of its diagnosis corresponded to when the tumor was detected by clinical symptoms or imaging tests. All analyses were performed using R version 4.0.4 (The R Core Team, Vienna, Austria).

Ethical Approval

The study protocol conformed to the ethical guidelines of the Declaration of Helsinki and was approved by the Institutional Review Board of Samsung Medical Center (IRB No. 2021-09-062-001)

Informed Consent

The need for informed consent was waived by the institutional review board of Samsung Medical Center due to the retrospective nature of the study.

RESULTS

Clinicopathologic Data

In total, 136 patients who underwent primary surgery for RPS between 2001 and 2020 and underwent preoperative ^{18}F-FDG PET/CT to determine the presence of metastasis were identified. Three patients with Ewing's sarcoma were excluded. Four patients were excluded due to insufficient pathological data such as mitosis and necrosis. After excluding these patients, data from a total of 129 patients were investigated. The histologic subtypes were dominantly LPS (68.2%) and LMS (15.5%). DDLPS accounted for 68% of the LPS patients, followed by well-differentiated liposarcoma (WDLPS) and pleomorphic liposarcoma (PLS). There was no significant difference in the distribution of tumor grades. Demographic and clinicopathological details are shown in **Table 1**.

Correlation Between SUVMmax and Pathologic Characteristics

The median SUVmax was 4.5 (range, 0.4-29). Tumor SUVmax was correlated with a higher tumor grade ($p < 0.001$, Spearman coefficient; 0.627) and mitosis ($p < 0.001$ Spearman coefficient; 0.564). In addition, SUVmax was different depending on the histologic subtype. The LPS group showed a lower SUVmax than the LMS group. When comparing SUVmax among the three groups, values were obtained in this order: WDLPS (2.32 ± 0.89), DDLPS (6.32 ± 4.97), and LMS (12.04 ± 6.73). The differences were statistically significant (**Figure 1**).

TABLE 1 | Characteristics of patients.

Variable		Value
Age, years (mean)		56.4 ± 12.2
Gender (%)	F	67 (51.9)
	M	62 (48.1)
BMI, kg/m^2 (mean)		23.5 ± 3.0
Underlying disease		
DM	Yes	11
	No	118
HTN	Yes	39
	No	90
Chronic renal disease	Yes	1
	No	128
Histologic subtype (%)	Well-differentiated liposarcoma	24 (18.6)
	Dedifferentiated liposarcoma	60 (46.5)
	Pleomorphic liposarcoma	4 (3.1)
	Leiomyosarcoma	20 (15.5)
	Malignant peripheral nerve sheath tumor	4 (3.1)
	Perivascular epithelioid cell tumor	1 (0.8)
	Other	16 (12.4)
FNCLCC grade (%)	1	29 (22.5)
	2	36 (27.9)
	3	64 (49.6)
SUVmax (median [range])		4.5 [0.4, 29.0]
Tumor size, mm (mean)		166.4 ± 101.3
Multifocality (%)	Yes	23 (17.8)
	No	106 (82.2)
Necrosis (%)	Absent	60 (46.5)
	<50%	60 (46.5)
	≥50%	9 (7.0)
Mitosis (%)	<9/10 HPF	95 (73.6)
	10-19/10 HPF	24 (18.6)
	≥20/10 HPF	10 (7.8)
Local recurrence (%)	Yes	54 (41.9)
	No	75 (58.1)
Distant metastasis (%)	Yes	17 (13.2)
	No	112 (86.8)
Follow up months after primary surgery, month (median[range])		37.8 [20.3, 71.9]

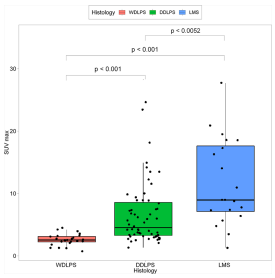


FIGURE 1 | Comparison of median SUVmax with histologic subtypes.

Prognostic Factors for RPS and SUVmax

Univariate analysis of the prognostic factors associated with OS was performed considering all patients with RPS. The factors significantly associated with OS were high-tumor grade (grade III, p = 0.003), SUVmax (p < 0.001), mitosis (≥ 20/10 high power fields (HPF), p < 0.001), and necrosis (≥50%, p < 0.001). In the multivariate analysis, SUVmax (p = 0.004) was the only factor significantly associated with OS. When analyzing the OS factors by histologic subtype, tumor grade (grade III, p = 0.011) and SUVmax (p < 0.001) were significant prognostic factors in the LPS group, consistent with RPS. However, there were no statistically significant risk factors in the LMS group. The details of the analyses are shown in **Table 2**.

Univariate analysis of prognostic factors for LR was performed considering all RPS patients. The SUVmax (p < 0.001), high tumor grade (p < 0.001), mitosis (≥20/10 HPF, p = 0.024), WDLPS (p = 0.004), LMS (p = 0.011) and necrosis (≥ 50%, p < 0.001) were significantly associated with LR. Within the multivariate analysis, the only factors independently associated with LR were high tumor grade (p = 0.014), WDLPS (p = 0.035) and necrosis (≥ 50%, p = 0.005). However, in the analysis conducted within histologic subtypes, SUVmax (p < 0.001) and high tumor grade (p = 0.002) were the main factors for LPS LR (**Table 3**).

Optimal Threshold to Distinguish High Grade Sarcoma

Receiver Operating Characteristic (ROC) curve analysis demonstrated that the Area Under the ROC curve (AUC) for high tumor grade (Grade III) was maximal when the threshold SUVmax was 4.8. The AUC for high tumor grade at the cut-off SUVmax was 0.820 (p < 0.001). At this threshold, the values of sensitivity and specificity were 0.77 and 0.80, respectively (**Figure 2**).

Outcome prediction Using an Optimal SUVmax threshold

The SUVmax threshold was used to distinguish a high SUVmax group and a low SUVmax group survival analysis was performed with respect to OS, LR, and DM. Considering the entire RPS group, the high SUVmax group showed a poor prognosis regarding OS, LR, and DM (p < 0.001). When analyzed by histologic subtype, the LPS patients with high SUVmax showed poor prognosis regarding OS (p < 0.001) and LR (p = 0.004). However, there were no such differences in the LMS group (**Figure 3**).

DISCUSSION

This study analyzed the relationship between SUVmax and the pathologic characteristics and prognosis of RPS. We showed that SUVmax is associated with high-grade RPS. In addition, we demonstrated that the range of SUVmax varies according to histologic subtype.

TABLE 2 | Univariate and multivariate analyses of risk factors associated with overall survival.

Variables	Univariate		Multivariate	
	HR (95% CI)	p value	HR (95% CI)	p value
Male	1.9 (0.98,3.66)	0.057		
Age	1.03 (1,1.06)	0.033		
SUVmax	1.11 (1.07,1.16)	< 0.001	1.09 (1.03,1.15)	0.004
Tumor size	1 (1,1)	0.815		
FNCLCC grade: ref. = 1				
2	0.93 (0.19,4.61)	0.926	0.76 (0.15,4.01)	0.749
3	6.06 (1.84,19.98)	0.003	4.4 (0.83,23.45)	0.083
Histology: ref.=DDLPS				
WDLPS	0.35 (0.1,1.21)	0.097		
LMS	0.89 (0.33,2.41)	0.815		
MPNST	1.48 (0.34,6.38)	0.597		
Other	1.18 (0.43,3.22)	0.746		
Necrosis: ref.= Absent				
<50%	3.26 (1.46,7.28)	0.004	0.81 (0.24,2.74)	0.74
≥50%	6.49 (2.1,20.02)	0.001	1.37 (0.33,5.73)	0.666
Mitosis: ref.= <9/10 HPF				
10-19/10 HPF	2.26 (1.06,4.81)	0.035	0.7 (0.26,1.9)	0.484
≥20/10 HPF	4.63 (1.83,11.7)	0.001	0.77 (0.21,2.81)	0.69

TABLE 3 | Univariate and multivariate analyses of risk factors associated with local recurrence.

Variables	Univariate		Multivariate	
	HR (95% CI)	p value	HR (95% CI)	p value
Male	1.14 (0.67,1.95)	0.632		
Age	1.01 (0.99,1.03)	0.503		
SUVmax	1.08 (1.04,1.12)	< 0.001	1.01 (0.94,1.08)	0.855
Tumor size	1 (1,1)	0.582		
FNCLCC grade: ref. = 1				
2	8.43 (1.9,37.39)	0.005	7.27 (1.57,33.75)	0.011
3	15.38 (3.69,64.04)	< 0.001	8.13 (1.54,42.92)	0.014
Histology: ref.=DDLPS				
WDLPS	0.12 (0.03,0.51)	0.004	0.19 (0.04,0.89)	0.035
LMS	2.32 (1.22,4.43)	0.011	1.93 (0.81,4.6)	0.137
MPNST	3.22 (0.96,10.78)	0.058	2.77 (0.75,10.19)	0.126
Other	0.46 (0.14,1.52)	0.206		
Necrosis: ref.= Absent				
<50%	3.35 (1.77,6.33)	< 0.001	1.56 (0.76,3.21)	0.23
≥50%	13.9 (5,38.6)	< 0.001	6 (1.74,20.7)	0.005
Mitosis: ref.= <9/10 HPF				
10-19/10 HPF	3.38 (1.79,6.39)	< 0.001	1.37 (0.61,3.09)	0.449
≥20/10 HPF	2.99 (1.15,7.75)	0.024	1.79 (0.4,7.91)	0.444

Distinction Between DDLPS and LMS

Our key finding was that higher SUVmax was found in LMS (12.04 ± 6.73) than DDLPS (6.32 ± 4.97). DDLPS and LMS are potential candidates for neoadjuvant chemotherapy, as the micrometastasis potential is lowered, and unresectable tumors can be reduced in size before surgery (5). Anthracycline-based adjuvant chemotherapy is the cornerstone of first-line treatment for localized soft tissue sarcoma (20). However, based on many retrospective studies, different histology-driven -chemotherapy options can be applied to DDLPS and LMS. In addition, multicenter prospective research (STRASS-2) is ongoing to determine whether these treatments affect prognosis (21). The distinction between high-grade LPS and LMS is becoming increasingly important to clinical decision-making considering these studies. Our findings suggest that ^{18}F-FDG PET/CT can be useful in distinguishing these two histologic subtypes preoperatively.

Detecting High-Grade RPS Through ^{18}F-FDG PET/CT Imaging

Due to its multifocal nature and large size, RPS can be difficult to target accurately during biopsy at the time of detection. In addition, preoperative biopsies tend to underestimate the final grade, most likely due to sampling error (22). For example, in LPS, when a solid portion and a fatty portion exist together, the high-grade portion is likely to be the solid portion. However, when there are several solid portions, it is difficult to predict the high-grade portion with CT. Because of these difficulties, the TARPSWG guidelines suggest that ^{18}F-FDG PET/CT be available for defining biopsy target areas (2). The current study demonstrated that tumor -SUVmax was correlated with higher tumor grade ($p < 0.001$, Spearman coefficient; 0.627) and mitosis ($p < 0.001$ Spearman coefficient; 0.564). This result is similar to other studies showing the association between pathologic characteristics and SUVmax (22, 23). These results support the TARPSWG guidelines recommendation to set SUVmax as the biopsy target area.

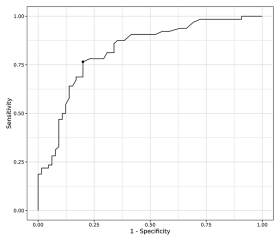


FIGURE 2 | Receiver Operation Characteristic (ROC) curve for SUVmax.

Prognosis Prediction Using SUVmax

A previous study conducted by our research team demonstrated that a SUVmax cut-off of 4.5 stratified RPS tumor grades and prognosis. In this study, only LPS was used, and there was a limitation in that SUVmax was determined in a heterogeneous population including metastatic and recurrent tumors (19). Subramaniam et al. also reported that when the SUVmax was higher than 5.0, the prognosis was poor, and high SUVmax and tumor grade were related. This study investigated a homogenous population; only the DDLPS and LMS groups were studied. However, the small number of patients has been mentioned as a limitation (22). In both studies referenced above, OS and relapse-free survival (RFS) were mentioned in the analysis of SUVmax and prognosis.

The current study investigated a relatively large number of patients given the low prevalence of RPS, excluding those with metastatic or recurrent tumors. In addition, the present study showed a correlation between SUVmax and DM, which has not been shown in other studies to our knowledge. The cut-off SUVmax (4.8) was a good measure for predicting prognosis

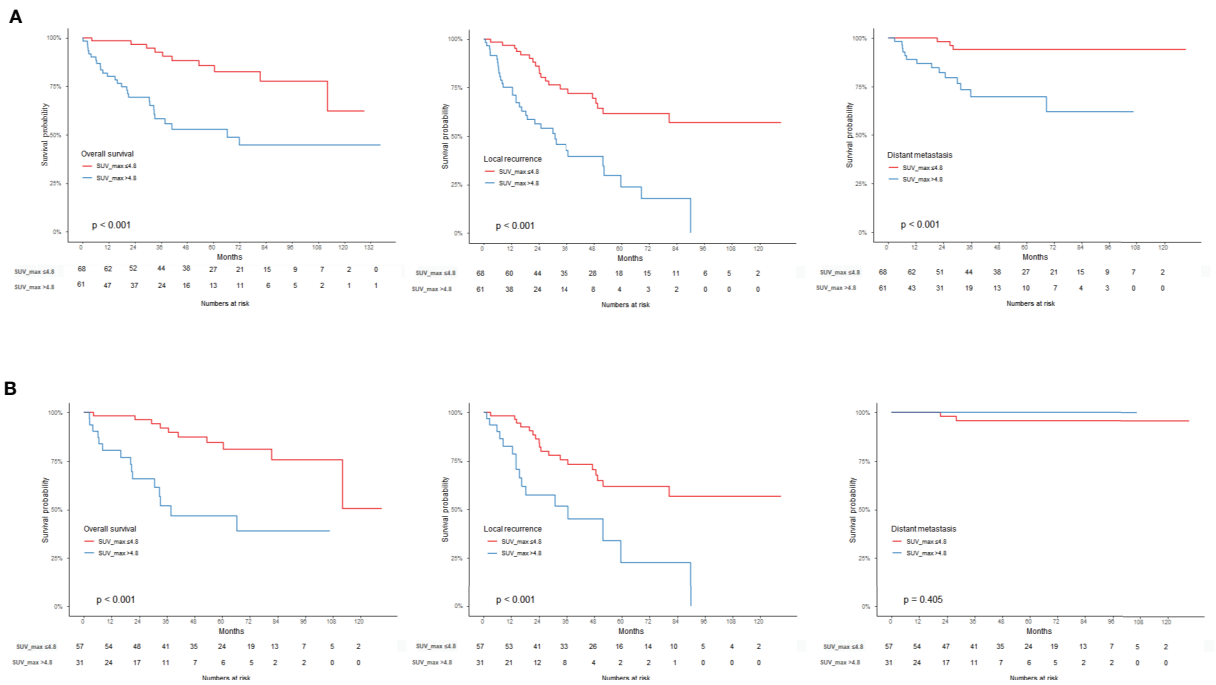


FIGURE 3 | Kaplan-Meier survival graph compared to the SUVmax threshold of 4.8. **(A)** OS, LR, DM in RPS group, **(B)** OS, LR, DM in LPS group.

but showed relatively low sensitivity (0.77) for predicting tumor grade and was not particularly useful in the LMS group. Therefore, our results indicate that ^{18}F-FDG PET/CT may be a useful measure of prognosis or high tumor grade for LPS considering its relatively high specificity (0.8).

Limitations

The current study is limited by its retrospective nature and the small number of LMS patients. A large-volume study is needed to find the SUVmax that can differentiate between DDLPS and LMS and to further evaluate the role of ^{18}F-FDG PET/CT in recurrent and metastatic tumors.

CONCLUSION

Tumor SUVmax was correlated with RPS pathologic parameters (tumor grade and, mitosis) and was higher in LMS than DDLPS. In addition, prognosis with respect to (OS, LR, and DM) was poor for patients with high SUVmax ($p < 0.001$). A SUVmax of 4.8 is the optimal threshold to rule out high-grade tumors, and prognosis can be predicted using this value.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board of Samsung Medical Center. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

SJ: Investigation, methodology, writing -original draft, writing – review & editing. KK: Investigation, resources. SL: Data curation, investigation. JK: Data curation, validation, visualization. SH: Writing – review & editing, investigation. JP: Project administration, resources. KL: Investigation, methodology, writing – review & editing. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the SungKyunKwan University and the BK21 FOUR (Graduate School Innovation) funded by the Ministry of Education (MOE, Korea) and National Research Foundation of Korea (NRF).

REFERENCES

1. Brennan MF, Antonescu CR, Moraco N, Singer S. Lessons Learned From the Study of 10,000 Patients With Soft Tissue Sarcoma. *Ann Surg* (2014) 260 (3):416–21; discussion 21-2. doi: 10.1097/SLA.0000000000000869
2. Trans-Atlantic, RPS Working Group. Management of Primary Retroperitoneal Sarcoma (RPS) in the Adult: A Consensus Approach From the Trans-Atlantic RPS Working Group. *Ann Surg Oncol* (2015) 22(1):256–63. doi: 10.1245/s10434-014-3965-2
3. Gronchi A, Strauss DC, Miceli R, Bonvalot S, Swallow CJ, Hohenberger P, et al. Variability in Patterns of Recurrence After Resection of Primary Retroperitoneal Sarcoma (RPS): A Report on 1007 Patients From the Multi-Institutional Collaborative RPS Working Group. *Ann Surg* (2016) 263(5):1002–9. doi: 10.1097/sla.0000000000001447
4. MacNeill AJ, Gronchi A, Miceli R, Bonvalot S, Swallow CJ, Hohenberger P, et al. Postoperative Morbidity After Radical Resection of Primary Retroperitoneal Sarcoma: A Report From the Transatlantic RPS Working Group. *Ann Surg* (2018) 267(5):959–64. doi: 10.1097/sla.0000000000002250
5. Almond LM, Gronchi A, Strauss D, Jafri M, Ford S, Desai A. Neoadjuvant and Adjuvant Strategies in Retroperitoneal Sarcoma. *Eur J Surg Oncol* (2018) 44 (5):571–9. doi: 10.1016/j.ejso.2018.02.001
6. Bonvalot S, Gronchi A, Le Péchoux C, Swallow CJ, Strauss D, Meeus P, et al. Preoperative Radiotherapy Plus Surgery Versus Surgery Alone for Patients With Primary Retroperitoneal Sarcoma (EORTC-62092: STRASS): A Multicentre, Open-Label, Randomised, Phase 3 Trial. *Lancet Oncol* (2020) 21(10):1366–77. doi: 10.1016/s1470-2045(20)30446-0
7. Gronchi A, Lo Vullo S, Fiore M, Mussi C, Stacchiotti S, Collini P, et al. Aggressive Surgical Policies in a Retrospectively Reviewed Single-Institution Case Series of Retroperitoneal Soft Tissue Sarcoma Patients. *J Clin Oncol* (2009) 27(1):24–30. doi: 10.1200/jco.2008.17.8871
8. Gronchi A, Bonvalot S, Le Cesne A, Casali PG. Resection of Uninvolved Adjacent Organs can be Part of Surgery for Retroperitoneal Soft Tissue Sarcoma. *J Clin Oncol* (2009) 27(12):2106–7. doi: 10.1200/JCO.2008.21.5467
9. Pasquali S, Vohra R, Tsimopoulou I, Vijayan D, Gourevitch D, Desai A. Outcomes Following Extended Surgery for Retroperitoneal Sarcomas: Results From a UK Referral Centre. *Ann Surg Oncol* (2015) 22(11):3550–6. doi: 10.1245/s10434-015-4380-z
10. Hwang SY, Warrier S, Thompson S, Davidson T, Yang JL, Crowe P. Safety and Accuracy of Core Biopsy in Retroperitoneal Sarcomas. *Asia Pac J Clin Oncol* (2016) 12(1):e174–8. doi: 10.1111/ajco.12125
11. Wilkinson MJ, Martin JL, Khan AA, Hayes AJ, Thomas JM, Strauss DC. Percutaneous Core Needle Biopsy in Retroperitoneal Sarcomas Does Not Influence Local Recurrence or Overall Survival. *Ann Surg Oncol* (2015) 22 (3):853–8. doi: 10.1245/s10434-014-4059-x
12. Almond LM, Tirotta F, Tattersall H, Hodson J, Cascella T, Barisella M, et al. Diagnostic Accuracy of Percutaneous Biopsy in Retroperitoneal Sarcoma. *Br J Surg* (2019) 106(4):395–403. doi: 10.1002/bjs.11064
13. Morosi C, Stacchiotti S, Marchianò A, Bianchi A, Radaelli S, Sanfilippo R, et al. Correlation Between Radiological Assessment and Histopathological Diagnosis in Retroperitoneal Tumors: Analysis of 291 Consecutive Patients at a Tertiary Reference Sarcoma Center. *Eur J Surg Oncol* (2014) 40(12):1662–70. doi: 10.1016/j.ejso.2014.10.005
14. Messiou C, Morosi C. Imaging in Retroperitoneal Soft Tissue Sarcoma. *J Surg Oncol* (2018) 117(1):25–32. doi: 10.1002/jso.24891
15. Bastiaannet E, Groen H, Jager PL, Cobben DC, van der Graaf WT, Vaalburg W, et al. The Value of FDG-PET in the Detection, Grading and Response to Therapy of Soft Tissue and Bone Sarcomas; a Systematic Review and Meta-Analysis. *Cancer Treat Rev* (2004) 30(1):83–101. doi: 10.1016/j.ctrv.2003.07.004
16. Folpe AL, Lyles RH, Sprouse JT, Conrad EU3rd, Eary JF. (F-18) Fluorodeoxyglucose Positron Emission Tomography as a Predictor of Pathologic Grade and Other Prognostic Variables in Bone and Soft Tissue Sarcoma. *Clin Cancer Res* (2000) 6(4):1279–87.
17. Roberge D, Vakilian S, Alabed YZ, Turcotte RE, Freeman CR, Hickeson M. FDG PET/CT in Initial Staging of Adult Soft-Tissue Sarcoma. *Sarcoma* (2012) 2012:960194. doi: 10.1155/2012/960194
18. Schwarzbach MH, Dimitrakopoulou-Strauss A, Willeke F, Hinz U, Strauss LG, Zhang YM, et al. Clinical Value of [18-F]] Fluorodeoxyglucose Positron Emission Tomography Imaging in Soft Tissue Sarcomas. *Ann Surg* (2000) 231 (3):380–6. doi: 10.1097/00000658-200003000-00011
19. Rhu J, Hyun SH, Lee KH, Jo SJ, Lee KW, Park JB, et al. Maximum Standardized Uptake Value on (18)F-Fluorodeoxyglucose Positron Emission Tomography/Computed Tomography Improves Outcome Prediction in Retroperitoneal Liposarcoma. *Sci Rep* (2019) 9(1):6605. doi: 10.1038/s41598-019-43215-5
20. Frustaci S, Gherlinzoni F, De Paoli A, Bonetti M, Azzarelli A, Comandone A, et al. Adjuvant Chemotherapy for Adult Soft Tissue Sarcomas of the Extremities and Girdles: Results of the Italian Randomized Cooperative Trial. *J Clin Oncol* (2001) 19(5):1238–47. doi: 10.1200/JCO.2001.19.5.1238
21. van Houdt WJ, Raut CP, Bonvalot S, Swallow CJ, Haas R, Gronchi A. New Research Strategies in Retroperitoneal Sarcoma. The Case of TARPSWG, STRASS and RESAR: Making Progress Through Collaboration. *Curr Opin Oncol* (2019) 31(4):310–6. doi: 10.1097/cco.0000000000000535
22. Subramaniam S, Callahan J, Bressel M, Hofman MS, Mitchell C, Hendry S, et al. The Role of (18) F-FDG PET/CT in Retroperitoneal Sarcomas-A Multicenter Retrospective Study. *J Surg Oncol* (2021) 123(4):1081–7. doi: 10.1002/jso.26379
23. Liu DN, Li ZW, Wang HY, Zhao M, Zhao W, Hao CY. Use of 18F-FDG-PET/CT for Retroperitoneal/Intra-Abdominal Soft Tissue Sarcomas. *Contrast Media Mol Imaging* (2018) 2018:2601281. doi: 10.1155/2018/2601281

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

MINI REVIEW
published: 09 June 2022
doi: 10.3389/fonc.2022.872656

Radiomics-Guided Precision Medicine Approaches for Colorectal Cancer

Mohammed I. Quraishi *

Department of Radiology, University of Tennessee Medical Center, Knoxville, TN, United States

The concept of precision oncology entails molecular profiling of tumors to guide therapeutic interventions. Genomic testing through next-generation sequencing (NGS) molecular analysis provides the basis of such highly targeted therapeutics in oncology. As radiomic analysis delivers an array of structural and functional imaging-based biomarkers that depict these molecular mechanisms and correlate with key genetic alterations related to cancers. There is an opportunity to synergize these two big-data approaches to determine the molecular guidance for precision therapeutics. Colorectal cancer is one such disease whose therapeutic management is being guided by genetic and genomic analyses. We review the rationale and utility of radiomics as a combinative strategy for these approaches in the management of colorectal cancer.

Keywords: radiomics, colorectal, cancer, precision medicine, personalized medicine, precision oncology, genomics

OPEN ACCESS

Edited by:
Tingying Peng,
Helmholtz zentrum München,
Germany

Reviewed by:
Shuji Ogino,
Brigham and Women's Hospital and
Harvard Medical School, United States

***Correspondence:**
Mohammed I. Quraishi
mquraishi@utmck.edu

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: *06 May 2022*
Published: *09 June 2022*

Citation:
Quraishi MI (2022) Radiomics-Guided Precision Medicine Approaches for Colorectal Cancer. Front. Oncol. 12:872656. doi: 10.3389/fonc.2022.872656

REVIEW

What is Precision Medicine/Oncology?

In the early part of this century, the term *personalized medicine* was heavily used to promote a new paradigm of treatment tailored to an individual. However, as this field developed, this term was seen as a misnomer as it highlights the individual and not the disease process. The term *precision* medicine more aptly describes how treatment is tailored to tumor-specific features that can be shared between individuals (1). Features of a disease state can be found in genetic, imaging, and histological information, and have been explored using deep analytics with genomics, radiomics, and pathomics, respectively.

Precision oncology, the application of this paradigm to oncology, has potential to revolutionize cancer management. This is a developing field with robust ongoing research. The greatest strides have been in genomics with radiomics not far behind.

The Genetics/Genomics Landscape of Colorectal Cancer

The theory of genomics in precision oncology is that treatment can be tailored based on the genetic make-up of the cancer (2). The most suitable therapy can be selected either by correlating between the unique genetic fingerprint of the malignancy and therapy options, or by impeding the driver mutation of an identified specific oncogene.

There is robust research in this field mainly due to recent availability and affordability of high-quality NGS molecular analysis (3). An example of that is the HER2 positivity, a clinically relevant genomic marker for breast cancer that predicts response to trastuzumab-based therapies (3).

Trastuzumab binds to the HER2 receptor which results in inactivation of the intracellular tyrosine kinase and, therefore, handicaps cell proliferation (4). Lung cancer management has also been more recently transformed by precision oncology approaches. Currently, instead of treating all patients with platinum-based doublets, treatment takes into consideration genetically defined subsets. For example, PD-L1 expression denotes a stronger response to immunotherapy (3). More recently, genomic subtyping has been shown to be prognostic in pulmonary large-cell neuroendocrine carcinoma (5).

The advent of NGS has made finding mutations in colorectal cancer more applicable to clinical practice. Genetic testing in colorectal cancer for prognostication and therapy selection is now standard of practice. In EGFR-expressing colorectal cancer (CRC), the mainstay of treatment is EGFR-targeting antibodies. However, only a limited percentage of CRC with EGFR expression respond. Currently, prognostication of treatment response with anti-EGFR correlates with specific KRAS, NRAS, and to a lesser extent BRAF V600E mutations. These mutations predict poor treatment response to anti-EGFR treatment. The presence of the KRAS or NRAS mutation confers not only poor response but also shorter survival if anti-EGFR treatment is used (6).

Beyond anti-EGFR therapy, newer therapeutic approaches are increasingly being guided by genetic/genomic advancements. Clinical trials studying the targeting of the BRAF V600E mutation with triple therapy (anti-RAF, anti-MEK, and cetuximab) are showing promise (6). Amplification of HER2 (ERBB2) also predicts poor anti-EGFR response as seen in the HERACLES trial (6, 7). Fusion involving the NTRK family of genes are seen in some CRC. Anti-NTRK treatment has shown promise in this subtype of CRC. Further, fusion in ALK, ROS, and RET all are present in different subtypes of CRC and show promise with targeted therapies (6).

Tissue heterogeneity is the main culprit in poor clinical adoption of new genomic insights for colorectal cancer. Within the same tumor there can be heterogeneous genetics that can be missed due to biopsy sampling error. Tissue heterogeneity can also be seen in colorectal cancer between the primary site and metastatic sites. Even more challenging are the changes in the genetic make-up of the disease seen after therapy. Furthermore, it is practically difficult to subject the patient to repeated biopsies for tissue/molecular/genetic analyses (6). It can be postulated that these issues can be allayed by radiomics. The phenotypic heterogeneity of tissue and its metastatic foci can be deciphered from imaging data. Moreover, imaging is easily repeatable, making it feasible to continually assess radiomic endpoints that represent intratumoral patterns of therapy response. In a synergistic fashion, radiogenomics, we believe, will enable precision oncology to deliver at its highest potential.

Radiomics as a Non-Invasive Imaging Correlate to Pathophysiologic and Genetic Basis of CRC

As a quantitative image analysis technique, radiomic analysis yields signatures (set of features) that may be shown to correlate with molecular processes or structural changes implicated in tumor behavior and its responsiveness to therapy. A wide range of radiomic signatures have been extracted in a number of studies focusing on CRC and were found to be useful in improving the way CRC is classified (8), stratified (9), and assessed for their natural history and therapy response (10). Radiomics is also improving our ability to develop more effective therapies to fight cancer by way of providing endpoints for pharmacokinetic/pharmacodynamic assessment as well as early therapy assessment response (or even predictive) biomarkers.

The emerging landscape of radiomics in CRC includes applications of this technique for the improved classification of CRC. Studies have shown that CT-based radiomic signatures were able to distinguish between high-grade and low-grade (AUC 0.7; 0.9) (11), and between the stage I-II and stage III-IV of CRC (AUC 0.8) (12), and for the prediction of microsatellite instability (MSI) in CRC which has implications for therapy responsiveness. Radiomics should be seen as a complimentary approach (and not as an alternative) to standard clinicopathologic methods used to guide clinical management. In a recent randomized controlled trial, radiomics was performed on preoperative CT in stage II and III CRC and it was observed that the combined clinical-radiomics model predicted the preoperative MSI status more accurately than the clinical or radiomic models alone (AUC 0.8) (13).

Radiomics has the potential to noninvasively predict genetic mutations/alterations which have implications in the personalized management of CRC. In a study by Tien et al., it was observed that CT-based radiomics predicted the KRAS/NRAS/BRAF mutations while clinical background, tumor stage and histological grade had no significant association (AUC 0.8) (14). FDG-PET-based radiomics analysis has also been used to determine various genetic alterations in CRC. Kao et al. observed that CRC with a mutated KRAS showed a 25th-percentile increase in the standardized uptake value (SUV) in their metabolic tumor volume (MTV) (15). It was also shown that the mutated TP53 was associated with an increased value of short-run low gray-level emphasis derived from the gray-level run length matrix, while the APC mutants exhibited lower low gray-level zone emphasis derived from the gray-level zone length matrix (15). In another study, CT-based radiomic analysis that quantified the temporal decrease in the tumor spatial heterogeneity and boundary infiltration were found to better predict the sensitivity EGFR-targeting therapy in metastatic CRC as compared to the standard endpoints, such as KRAS mutations and tumor shrinkage as per RECIST 1.1 (16).

More sophisticated radiomic analyses are possible with MRI, a modality used regularly in the management of CRC. A 30-feature radiomic signature extracted from pre- and post-therapy MRI (along with tumor length) was shown to predict the pathologic complete response to locally advanced CRC (AUC 0.98) (17). Another study showed that T2W-MRI based radiomics was superior to T2W/DWI at predicting pathologic complete response to neoadjuvant therapy for CRC (AUC 0.93) (18). FDG-PET based radiomics has also been found to be useful for therapy response prediction. In a recent study, it was shown that the radiomics-based tumor heterogeneity along with low

tumor volume, both measured on FDG PET/CT were associated with improved clinical outcome (AUC 0.8) (19). A combined multiparametric approach, using FDG-PET and MR-based radiomics assessment was found highly predictive of neoadjuvant therapy response (20).

Combining Radiomics With Genomics for Precision Therapy Guidance for CRC

As shown by several studies mentioned above, radiomic biomarkers can be correlated with and potentially predict genetic mutations on one hand and clinical outcomes on the other. Furthermore, it can be proven to be a highly useful methodology when combined with genetic/genomic studies to understand the complete genotype-phenotype relationship that is implicated in the natural history and therapy responsiveness of CRC. Such radiogenomic methodologies have been employed in diseases, including Non-small lung and breast cancers (21) and utilizing it may be of great benefit in the precision management of CRC as well.

This combinative multi-omics approach has been used to evaluate the pathophysiologic underpinnings and the natural history they impact (22, 23). Radiogenomics involves extracting radiomic features from a CT, MR or a PET scan and combining it with the mRNA expression data to create a radiogenomic signature and correlating it with the clinicopathological event under study (see **Figure 1**). This was performed in a study performed by Duo et al, in which they employed a radiogenomic approach using FDG-PET/CT and messenger RNA data to develop a signature that depicted the epithelial–mesenchymal transition in non–small cell lung cancer (25). Such methodologies have been attempted in the realm of CRC as well. Bogdan et al. studied the combined signature from CT and ABC22, CD166, CDKNV1 and IHBBB gene expression and histologic grading to be an statistically significant prognosticator for CRC screening and management (26).

Studies have been performed to study the genetic changes and the radiomics features and if they can be used to guide precision therapy. Lambin et al. have demonstrated that radiomics can identify the gene expressions related to tumor cells in response to doxycycline and radiation treatment (27). Radiomic signatures that can predict the V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (KRAS) mutation in CRC have been studied, where it was interesting to note that radiomic signal based on CT was superior to FDG PET for that purpose (28). In a retrospective study performed by Petskova et al., a radiogenomic analysis using the genomic sequences and the pretreatment MRI was performed, which showed that quantitative assessment (radiomics as opposed to qualitative assessment) showed an association with genetic mutation related to CRC (29).

These approaches, while potentially highly beneficial are not without their challenges. The robustness of radiomics/radiogenomics in terms of inter- and/or intra-reader agreement continues to be scrutinized. Having said that, many studies have concluded that these techniques are reliable in that regard, especially when the models have large datasets to build on and the segmentation capabilities improved (which can be challenging in CRC lesions) (30).

Other Considerations

Here we have focused on the emerging role of radiogenomics for colorectal cancer as we believe it will be a pillar of precision oncology. However, it should be noted that beyond imaging biomarkers and genetic make-up, environmental and epidemiological factors also affect cancer behavior. Varied

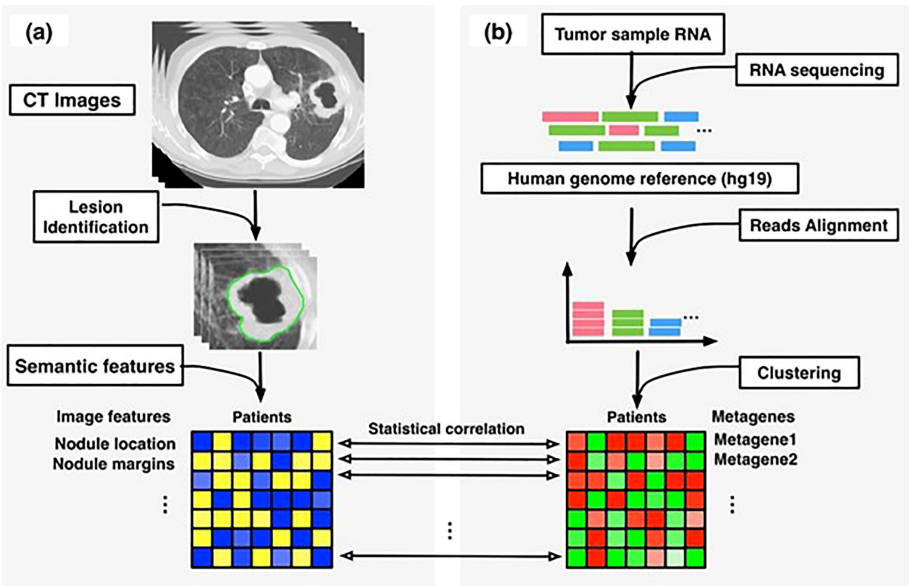


FIGURE 1 | Overview of radiogenomic analysis to identify associations between, **(A)** semantic features at CT and, **(B)** RNA sequencing data. [reused from Zhou et al. (24) under the CC-BY license].

factors such as alcohol intake, processed meat consumption, gut flora makeup, cigarette smoking, infanthood bottle feeding, sedentary lifestyle, and obesity amongst others are implicated in cellular epigenetic and genetic alterations which can predispose an individual to CRC (31). This is termed the exposome, which describes the sum of exposures and their interactions (31). The field of molecular pathological epidemiology (MPE) looks at how epidemiologic factors affect molecular pathology, and as MPE develops there is potential for finding exposome biomarkers (32). Studying the interplay between imaging, genetics, pathology, and the exposome to develop a more precise tumor signature is the terminus of precision oncology.

CONCLUSION

Precision management of CRC is poised to benefit from the noninvasive radiomic correlates to the genetic information implicated in its driving decision making process as well as the combined radiogenomics approaches that study the genotype-phenotype relationship at its core.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

1. Desmond-Hellmann S, Sawyers CL, Cox DR, Fraser-Liggett C, Galli SJ, Goldstein DB, et al. National Research Council. Toward Precision Medicine: Building a Knowledge Network for Biomedical Research and a New Taxonomy of Disease. Washington, DC: The National Academies Press (2011). doi: 10.17226/13284
2. Bluemke DA. Issue Summary 1 [Audio Podcast Episode]. In: *Radiology Podcasts*. Oak Brook, IL: Radiological Society of North America (2020). Available at: https://rsnaradiology.libsyn.com/issue-summary-1-august-2020.
3. Schwartzberg L, Kim ES, Liu D, Schrag D. Precision Oncology: Who, How, What, When, and When Not? *Am Soc Clin Oncol Educ Book* (2017) 37:160–9. doi: 10.1200/EDBK_174176
4. Hudis CA. Trastuzumab - Mechanism of Action and Use in Clinical Practice. *N Engl J Med* (2007) 357:399–51. doi: 10.1056/NEJMra043186
5. Zhuo M, Guan Y, Yang X, Hong L, Wang Y, Li Z, et al. The Prognostic and Therapeutic Role of Genomic Subtyping by Sequencing Tumor or Cell-Free DNA in Pulmonary Large-Cell Neuroendocrine Carcinoma. *Clin Cancer Res* (2020) 26(4):892–901. doi: 10.1158/1078-0432.CCR-19-0556
6. Mondaca S, Yaeger R. Colorectal Cancer Genomics and Designing Rational Trials. *Ann Transl Med* (2018) 6(9):159. doi: 10.21037/atm.2018.03.27
7. Sartore-Bianchi A, Trusolino L, Martino C, Bencardino K, Lonardi S, Bergamo F, et al. Dual-Targeted Therapy With Trastuzumab and Lapatinib in Treatment-Refractory, KRAS Codon 12/13 Wild-Type, HER2-Positive Metastatic Colorectal Cancer (HERACLES): A Proof-of-Concept, Multicentre, Open-Label, Phase 2 Trial. *Lancet Oncol* (2016) 17(6):738–46. doi: 10.1016/S1470-2045(16)00150-9
8. Li Wen Y, Leech M. Review of the Role of Radiomics in Tumour Risk Classification and Prognosis of Cancer. *Anticancer Res* (2020) 40(7):3605–18. doi: 10.21873/anticanres.14350
9. Osman SOS, Leijenaar RTH, Cole AJ, Lyons CA, Hounsell AR, Prise KM, et al. Computed Tomography-Based Radiomics for Risk Stratification in Prostate Cancer. *Int J Radiat Oncol Biol Phys* (2019) 105(2):448–56. doi: 10.1016/j.ijrobp.2019.06.2504
10. Hsu CY, Doubrovin M, Hua CH, Mohammed O, Shulkin BL, Kaste S, et al. Radiomics Features Differentiate Between Normal and Tumoral High-Fdg Uptake. *Sci Rep* (2018) 8(1):3913. doi: 10.1038/s41598-018-22319-4
11. 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
12. Liang C, Huang Y, He L, Chen X, Ma Z, Dong D, et al. The Development and Validation of a CT-Based Radiomics Signature for the Preoperative Discrimination of Stage I-II and Stage III-IV Colorectal Cancer. *Oncotarget* (2016) 7(21):31401–12. doi: 10.18632/oncotarget.8919
13. Golia Pernicka JS, Gagniere J, Chakraborty J, Yamashita R, Nardo L, Creasy JM, et al. Radiomics-Based Prediction of Microsatellite Instability in Colorectal Cancer at Initial Computed Tomography Evaluation. *Abdom Radiol (NY)* (2019) 44(11):3755–63. doi: 10.1007/s00261-019-02117-w
14. Yang L, Dong D, Fang M, Zhu Y, Zang Y, Liu Z, et al. Can CT-Based Radiomics Signature Predict KRAS/NRAS/BRAF Mutations in Colorectal Cancer? *Eur Radiol* (2018) 28(5):2058–67. doi: 10.1007/s00330-017-5146-8
15. Chen SW, Shen WC, Chen WT, Hsieh TC, Yen KY, Chang JG, et al. Metabolic Imaging Phenotype Using Radiomics of [18F]FDG PET/CT Associated With Genetic Alterations of Colorectal Cancer. *Mol Imaging Biol* (2019) 21(1):183–90. doi: 10.1007/s11307-018-1225-8
16. Dercle L, Lu L, Schwartz LH, Qian M, Tejpar S, Eggleton P, et al. Radiomics Response Signature for Identification of Metastatic Colorectal Cancer Sensitive to Therapies Targeting EGFR Pathway. *J Natl Cancer Inst* (2020) 112(9):902–12. doi: 10.1093/jnci/djaa017
17. Jiao S, Xia W, Yamaguchi H, Wei Y, Chen MK, Hsu JM, et al. PARP Inhibitor Upregulates PD-L1 Expression and Enhances Cancer-Associated Immunosuppression. *Clin Cancer Res* (2017) 23(14):3711–20. doi: 10.1158/1078-0432.CCR-16-3215
18. Horvat N, Veeraraghavan H, Khan M, Blazic I, Zheng J, Capanu M, et al. MR Imaging of Rectal Cancer: Radiomics Analysis to Assess Treatment Response After Neoadjuvant Therapy. *Radiology* (2018) 287(3):833–43. doi: 10.1148/radiol.2018172300
19. van Helden EJ, Vacher YJL, van Wieringen WN, van Velden FHP, Verheul HMW, Hoekstra OS, et al. Radiomics Analysis of Pre-Treatment [18F]FDG PET/CT for Patients With Metastatic Colorectal Cancer Undergoing Palliative Systemic Treatment. *Eur J Nucl Med Mol Imaging* (2018) 45(13):2307–17. doi: 10.1007/s00259-018-4100-6
20. Giannini V, Mazzetti S, Bertotto I, Chiarenza C, Cauda S, Delmastro E, et al. Predicting Locally Advanced Rectal Cancer Response to Neoadjuvant Therapy With 18F-FDG PET and MRI Radiomics Features. *Eur J Nucl Med Mol Imaging* (2019) 46(4):878–88. doi: 10.1007/s00259-018-4250-6
21. Ma DN, Gao XY, Dan YB, Zhang AN, Wang WJ, Yang G, et al. Evaluating Solid Lung Adenocarcinoma Anaplastic Lymphoma Kinase Gene Rearrangement Using Noninvasive Radiomics Biomarkers. *Onco Targets Ther* (2020) 13:6927–35. doi: 10.2147/OTT.S257798
22. Wu J, Tha KK, Xing L, Li R. Radiomics and Radiogenomics for Precision Radiotherapy. *J Radiat Res* (2018) 59(suppl_1):i25–31. doi: 10.1093/jrr/rrx102
23. 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
24. Zhou M, Leung A, Echegaray S, Gentles A, Shrager JB, Jensen KC, et al. Non–small Cell Lung Cancer Radiogenomics Map Identifies Relationships Between Molecular and Imaging Phenotypes With Prognostic Implications. *Radiology* (2018) 286(1):307–15.
25. Yamamoto S, Huang D, Du L, Korn RL, Jamshidi N, Burnette BL, et al. Radiogenomic Analysis Demonstrates Associations Between (18)F-Fluoro-2-Deoxyglucose PET, Prognosis, and Epithelial-Mesenchymal Transition in Non-Small Cell Lung Cancer. *Radiology* (2016) 280(1):261–70. doi: 10.1148/radiol.2016160259
26. Badic B, Hatt M, Durand S, Jossic-Corcos CL, Simon B, Visvikis D, et al. Radiogenomics-Based Cancer Prognosis in Colorectal Cancer. *Sci Rep* (2019) 9(1):9743. doi: 10.1038/s41598-019-46286-6

27. Panth KM, Leijenaar RT, Carvalho S, Lieuwes NG, Yaromina A, Dubois L, et al. Is There a Causal Relationship Between Genetic Changes and Radiomics-Based Image Features? An *In Vivo* Preclinical Experiment With Doxycycline Inducible GADD34 Tumor Cells. *Radiother Oncol* (2015) 116 (3):462–6. doi: 10.1016/j.radonc.2015.06.013
28. Taguchi N, Oda S, Yokota Y, Yamamura S, Imuta M, Tsuchigame T, et al. CT Texture Analysis for the Prediction of KRAS Mutation Status in Colorectal Cancer *via* a Machine Learning Approach. *Eur J Radiol* (2019) 118:38–43. doi: 10.1016/j.ejrad.2019.06.028
29. Horvat N, Veeraraghavan H, Pelossof RA, Fernandes MC, Arora A, Khan M, et al. Radiogenomics of Rectal Adenocarcinoma in the Era of Precision Medicine: A Pilot Study of Associations Between Qualitative and Quantitative MRI Imaging Features and Genetic Mutations. *Eur J Radiol* (2019) 113:174–81. doi: 10.1016/j.ejrad.2019.02.022
30. Horvat N, Bates DDB, Petkovska I. Novel Imaging Techniques of Rectal Cancer: What do Radiomics and Radiogenomics Have to Offer? A Literature Review. *Abdom Radiol* (2019) 44(11):3764–74. doi: 10.1007/s00261-019-02042-y
31. Akimoto N, Ugai T, Zhong R, Hamada T, Fujiyoshi K, Giannakis M, et al. Rising Incidence of Early-Onset Colorectal Cancer — a Call to Action. *Nat Rev Clin Oncol* (2020) 18:230–43.
32. Ogino S, Nowak JA, Hamada T, Milner DAJr, Nishihara R. Insights Into Pathogenic Interactions Among Environment, Host, and Tumor at the Crossroads of Molecular Pathology and Epidemiology. *Annu Rev Pathol* (2019) 14:83–103. doi: 10.1146/annurev-pathmechdis-012418-012818

Conflict of Interest: The 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.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 01 July 2022
doi: 10.3389/fonc.2022.865548

A Radiomics Nomogram Integrated With Clinic-Radiological Features for Preoperative Prediction of DNA Mismatch Repair Deficiency in Gastric Adenocarcinoma

Yahan Tong[1,2,3†], Jiaying Li[4,5†], Jieyu Chen[1,2†], Can Hu[2,3,6], Zhiyuan Xu[2,3,6], Shaofeng Duan[7], Xiaojie Wang[8], Risheng Yu[8] and Xiangdong Cheng[2,3,6*]*

[1] Department of Radiology, The Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital), Hangzhou, China, [2] Institute of Basic Medicine and Cancer (IBMC), Chinese Academy of Sciences, Hangzhou, China, [3] Key Laboratory of Prevention, Diagnosis and Therapy of Upper Gastrointestinal Cancer of Zhejiang Province, Hangzhou, China, [4] Department of Radiology, The First Clinical Medical College of Zhejiang Chinese Medical University, Hangzhou, China, [5] Department of Radiology, The First Affiliated Hospital of Zhejiang Chinese Medical University, Hangzhou, China, [6] Department of Gastric Surgery, The Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital), Hangzhou, China, [7] Precision Health Institution, GE Healthcare, Shanghai, China, [8] Department of Radiology, Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

OPEN ACCESS

Edited by:
Rathan Subramaniam,
University of Otago, New Zealand

Reviewed by:
Ning Mao,
Yantai Yuhuangding Hospital, China
Israel Tojal da Silva,
AC Camargo Cancer Center, Brazil

***Correspondence:**
Xiangdong Cheng
chengxd@zjcc.org.cn
Risheng Yu
risheng-yu@zju.edu.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: 30 January 2022
Accepted: 26 May 2022
Published: 01 July 2022

Citation:
Tong Y, Li J, Chen J, Hu C, Xu Z, Duan S, Wang X, Yu R and Cheng X (2022) A Radiomics Nomogram Integrated With Clinic-Radiological Features for Preoperative Prediction of DNA Mismatch Repair Deficiency in Gastric Adenocarcinoma. Front. Oncol. 12:865548. doi: 10.3389/fonc.2022.865548

Purpose: To develop and validate a radiomics nomogram integrated with clinic-radiological features for preoperative prediction of DNA mismatch repair deficiency (dMMR) in gastric adenocarcinoma.

Materials and Methods: From March 2014 to August 2020, 161 patients with pathologically confirmed gastric adenocarcinoma were included from two centers (center 1 as the training and internal testing sets, n = 101; center 2 as the external testing sets, n = 60). All patients underwent preoperative contrast-enhanced computerized tomography (CT) examination. Radiomics features were extracted from portal-venous phase CT images. Max-relevance and min-redundancy (mRMR) and least absolute shrinkage and selection operator (LASSO) methods were used to select features, and then radiomics signature was constructed using logistic regression analysis. A radiomics nomogram was built incorporating the radiomics signature and independent clinical predictors. The model performance was assessed using receiver operating characteristic (ROC) curve analysis, calibration curve, and decision curve analysis (DCA).

Results: The radiomics signature, which was constructed using two selected features, was significantly associated with dMMR gastric adenocarcinoma in the training and internal testing sets ($P < 0.05$). The radiomics signature model showed a moderate discrimination ability with an area under the ROC curve (AUC) of 0.81 in the training set, which was confirmed with an AUC of 0.78 in the internal testing set. The radiomics nomogram consisting of the radiomics signature and clinical factors (age, sex, and

location) showed excellent discrimination in the training, internal testing, and external testing sets with AUCs of 0.93, 0.82, and 0.83, respectively. Further, calibration curves and DCA analysis demonstrated good fit and clinical utility of the radiomics nomogram.

Conclusions: The radiomics nomogram combining radiomics signature and clinical characteristics (age, sex, and location) may be used to individually predict dMMR of gastric adenocarcinoma.

Keywords: gastric cancer/adenocarcinoma, radiomics, tomography, X-ray computed, nomogram, DNA mismatch repair deficiency

INTRODUCTION

Globally, gastric cancer (GC) is one of the most common malignant tumors and is a common cause of cancer-related death (1, 2). The symptoms of early GC are occult and often neglected, so many patients in China have locally advanced disease at the time of diagnosis (3). Since microsatellite instability (MSI) was found in hereditary non-polyposis colorectal cancer in 1993, it has been detected in many forms of malignant tumor, such as lung and bladder cancers (4–6). Increasingly, clinical trials have confirmed that MSI/DNA mismatch repair deficiency (dMMR) plays an important role in the occurrence and prognosis of GC (7–9). The Cancer Genome Atlas has identified MSI or dMMR as a hallmark of the second molecular subtype of GC (10, 11). MSI or dMMR status in GC is crucial for clinical decision making, as it identifies patients with different treatment responses and prognoses of GC (12–14). According to the 2021 guidelines of the National Comprehensive Cancer Network (NCCN) for GC (15), all newly diagnosed GC patients should be tested for MSI by polymerase chain reaction (PCR)-based molecular testing or DNA mismatch repair (MMR) protein using immunohistochemistry (IHC). Conventional MSI/MMR testing is recommended, but many patients remain untested. Testing for MSI/MMR is expensive, and interobserver variability in interpretation has been found among the different primary modalities (16, 17). Presurgery prediction of mismatch repair gene expression in GC would be of great significance for the selection of the treatment plan and treatment method and the evaluation of prognosis. There is a critical need for development of an objective, broadly accessible, and cost-efficient testing method for patients with GC.

Radiomics can provide more information than conventional CT images. The rise of radiomics makes it possible to convert imaging data into high-dimensional feature data, and the multiple quantitative features extracted from original images by bioinformatics can predict the underlying biological behavior of tumors (18–20). In recent years, many studies have found that certain radiomics features have diagnostic and prognostic value (21–23). Zhang et al. reported that the magnetic resonance imaging (MRI) texture signature may serve as a potential predictive biomarker for immunophenotyping and overall survival of intrahepatic cholangiocarcinoma patients (23). In the field of radio-genomics, imaging features are allied to genotype. Tumors with poor prognosis also tend to have greater genomic heterogeneity of tumor tissues (24). Radio-genomics is an evolution on the foundation of radiomics, which assumes that genomic heterogeneity at the microscopic level can present as tumor heterogeneity, and variation in the microenvironment of the lesion may be manifested as morphological characteristics and macroscopic images (25). Hence, the application of radiomics offers a new path to remove the limitations of traditional biopsy methods. Kim et al. found that the texture features based on multiparametric MRI were particularly connected with the isocitrate dehydrogenase mutation and tumor aggressiveness in diffuse lower-grade glioma (26). In recent years, radiomics nomograms, which are based on multiple variables, have been widely accepted as a user-friendly tool for predicting prognosis and have been used successfully to forecast the genotype of malignant tumors preoperatively (27–29). Wang et al. reported that the radiomics nomogram integrated with clinic-radiological features holds promise for clinical use as a non-invasive tool in the individual prediction of lymph node metastasis in GC (30). Wang et al. found that the nomogram-integrated CT-radiomics signature and CT-reported T stage can enhance prediction of the human epidermal growth factor receptor 2 status of esophagogastric junction adenocarcinoma before surgery (31).

Therefore, in this research, we aimed to develop and validate a radio-clinical nomogram based on a combination of radiomics signature and clinical risk factors for the preoperative prediction of DNA mismatch repair deficiency in patients with gastric adenocarcinoma.

MATERIALS AND METHODS

Patients

This retrospective study was approved by the review board of our institution (The Cancer Hospital of the University of Chinese Academy of Science). The requirement for informed consent was waived. This study retrospectively collected data from 1,456 patients with pathologically confirmed GC who underwent radical gastrectomy between March 2014 and August 2020 at two centers. In total, 161 patients were enrolled according to the inclusion and exclusion criteria (detailed below). Among these, 101 cases from center 1 (The Cancer Hospital of the University of Chinese Academy of Science) were used as the training and internal testing sets, and 60 cases from center 2 (The Second Affiliated Hospital Zhejiang University School of Medicine) were used as the external testing set. A flowchart of the patient record

selection process is shown in **Figure 1**. All patients underwent preoperative contrast-enhanced CT examination of the abdomen.

The inclusion criteria were as follows: (1) postoperative pathologically confirmed gastric adenocarcinoma at stage 3 or 4 according to the diagnostic criteria of the 8th edition of the American Joint Committee on Cancer Staging Manual; (2) patients who underwent contrast-enhanced CT of the upper abdomen or the whole abdomen within 1 month before surgery; (3) IHC detection was performed on pathological tissue to evaluate the MMR status postoperatively. The exclusion criteria were the following: (1) incomplete clinical or pathological information (2); pathological type signet ring cell carcinoma or mucinous adenocarcinoma; (3) treatment was given before surgery; (4) poor CT image quality with longest diameter of less than 5mm. Patients' clinical and imaging data including sex, age, tumor location, and MMR status were recorded. The location of GC was based on pathology, including cardia, gastric body, and gastric antrum.

CT Image Acquisition

All patients underwent contrast-enhanced abdominal CT using the following multidetector row CT systems: BrightSpeed, Optima CT680 Series (GE Medical Systems), and Siemens Somatom Definition AS 64, Perspective (Siemens Medical Systems). The acquisition parameters were as follows: tube voltage, 120–130 kV; tube current, 150–300 mAs; reconstructed axial-section thickness 5 mm, slice interval 5 mm, pitch 0.6. The contrast agents were Ultravist (Bayer Schering Pharma, Berlin, Germany), Optiray (Liebel-Flarsheim Canada Inc., Kirkland, Quebec, Canada), and Iohexol (Beijing North Road Pharmaceutical Co. Ltd., Beijing, China). A total of 70–100 ml of contrast agent was administered using a pump injector into an antecubital vein. Arterial phase and

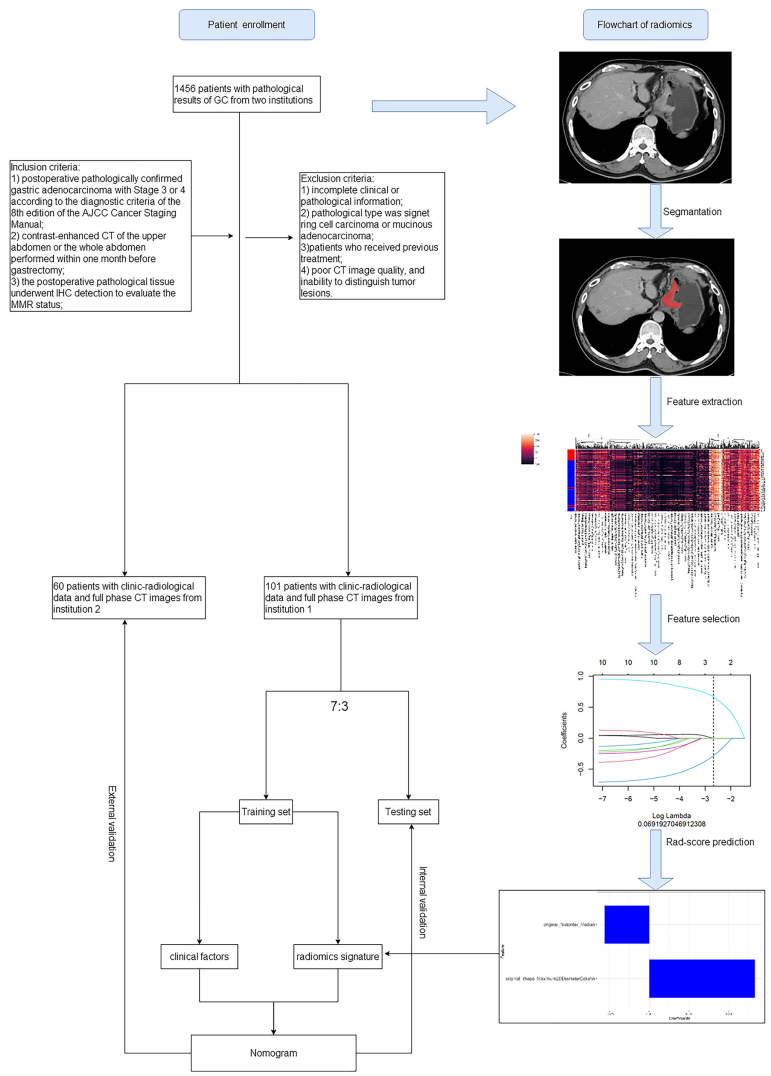


FIGURE 1 | Flowchart of the recruiting study population and model construction.

portal venous phase contrast-enhanced CT scans were performed after delays of 30–35 s and 50–60 s after injection of the contrast medium, respectively.

Mismatch Repair Protein Status

IHC was used to evaluate the results of MMR protein status according to the 2021 Gastric Cancer NCCN guidelines as follows: FmutL homologue 1 (MLH1), mutS homologue 2 (MSH2), mutS homologue 6 (MSH6), and PMS1 homologue 2 (PMS2) proteins were detected, which were positively located in the nucleus. Any protein expression loss was evaluated as dMMR (mismatch repair function defect), and all four protein expressions were positive as pMMR (mismatch repair function complete).

Tumor Segmentation

The portal-venous phase CT images of GC patients were acquired from the picture archiving and communication systems. The patient's abdominal portal venous phase CT digital image was exported in digital imaging and communications in medicine (DICOM) format. Radiologists with over 5 years of experience in interpreting abdominal diseases examined each layer of the patients' CT images. Two radiologists outlined the regions in each patient's CT images. Lesions were delineated using ITK-SNAP (version 3.8.0, http://www.itksnap.org) as shown in **Figure 2**.

For the tumor regions of interest (ROIs), radiologists reviewed all of each patient's CT image slices and selected the largest tumor area slice to segment. The ROI was selected to cover the whole area of the tumor. Observer 1 delineated the lesions of all patients with GC. Observer 2 confirmed the tumor segmentation (32). If the segmented lesions were inconsistent between the two observers, consensus was reached by discussion. During the delineation process, ROI selection avoided the areas of gastric air, necrosis, and adipose tissue.

Radiomics Feature Extraction and Selection

The radiomics feature extraction process for this study was performed using YITU AI Enabler, which is an integrated machine learning platform for medical data analysis using well-established python pyradiomics (version 3.0.1) and the scikit-learn (version 0.22) package. Resampling through the radiomics features was first extracted based on the original image data set. Then a feature stability check was performed on the features extracted within the lesion ROI and the extended lesion ROI to filter out unstable features with minor change of ROI using an intra-class correlation algorithm. The extended lesion ROI was made by extending the boundary of lesion ROI by one image pixel. We used max-relevance and min-redundancy (mRMR) and least absolute shrinkage and selection operator (LASSO) methods to select features, and then the rad-score of each GC patient was calculated by their coefficients.

Construction of a Predictive Model

Multivariable logistic regression analysis was used to develop a prediction model by combining significant rad-score, sex, age, and tumor location (with P values less than 0.05 in the univariable analysis). In the training set, for clinicians' convenience we constituted the model as a radio-clinical nomogram based on multivariable logistic regression analysis. Finally, the generalization ability of the nomogram was evaluated in the internal and external testing sets.

Performance of the Radiomics Nomogram

The predictive performance of the radiomics nomogram was evaluated using the receiver operator characteristic (ROC) curve, calibration curve, and decision curve analysis (DCA). Model evaluation 10-fold cross-validation was used in model training, and the diagnostic performance of radiomics, clinical, and radio-clinical models were validated in the internal testing set. The area

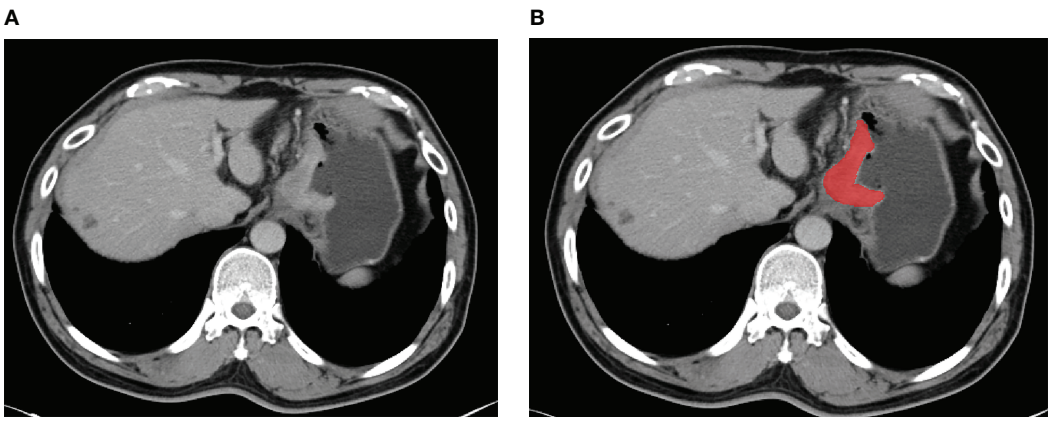


FIGURE 2 | An example of manual segmentation in gastric cancer. **(A)** Localized thick wall of gastric cancer with enhancement is observed on the portal venous phase computed tomography (CT) image. **(B)** Manual segmentation on the same axial slice is depicted with red label.

under the ROC curve (AUC), sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV) of the nomogram were calculated. DCA analysis was performed to assess the model's clinical utility by calculating the net benefits at different threshold probabilities. Finally, generalization of the radiomics nomogram was evaluated in the independent external testing set.

Statistical Analysis

All statistical analyses were performed using R software (version 3.4.1; http://www.Rproject.org) and IBM SPSS Statistics (Version 26.0; IBM Corp., New York, USA). Quantitative data were described by mean ± standard deviation, and qualitative data by frequency (percent). Normally distributed continuous data were compared using the Student's t-test. The chi-square test was used to compare the distribution of categorical data between groups. A multivariate logistic regression analysis was applied to determine the independent predictors among all the clinical variables. $P < 0.05$ was considered statistically significant. The "glmnet" package was used for LASSO logistic regression analysis. The multivariable logistic regression analysis and calibration plots were conducted using the "rms" package. The ROC plots of radiomics signature were performed with the "pROC" package. The "rmda" package was applied for decision curve analysis (DCA).

RESULTS

Clinical Characteristics

Among 101 patients with GC from center 1, there were dMMR (n = 35) and pMMR (n = 66) cases. The patients were randomly divided into a training set of 71 cases and an internal testing set of 30 cases. In the training set, statistically significant differences in sex, age, and tumor location were found between dMMR and pMMR GC patients ($P < 0.05$). In the training and internal testing sets, a significantly higher rad-score was found in dMMR than in pMMR in both cohorts ($P < 0.05$). Among 60 patients with GC from center 2 as an external testing set, there were dMMR (n = 21) and pMMR (n = 39) cases. Additional details are provided in **Table 1**.

Radiomics Feature Selection and Radiomics Signature Construction

A total of 1,648 radiomics features were extracted from CT images of each GC patient, among which 989 features with good stability were selected for radiomics model establishment. Initially, mRMR was performed to eliminate the redundant and irrelevant features, and 30 features were retained. Then, LASSO was conducted to choose the optimized subset of features to construct the final model. The optimal λ in the LASSO logistic regression analysis with 10-fold cross-validation was used to select the best radiomics feature with a non-zero coefficient, as shown in **Figure 3**. Finally, two radiomics features were selected to construct the radiomics signature, and the rad-score was calculated by summing the selected features weighted by their coefficients. The final formula for rad-score is as follows:

$$Radscore = 0.666 * original\_shape\_Maximum2DDiameterColumn + -0.283 * original\_firstorder\_Median + -0.747$$

Development of an Individualized Radiomics Nomogram

Univariate analysis showed that sex, age and tumor location with P values less than 0.05 were independent clinical risk factors for MMR status in GC patients. Multivariable analysis was performed to develop a prediction model by combining the rad-score, sex, age, and tumor location **(Table 2)**. Further, the radiomics nomogram is visualized in **Figure 4**. The formula for the nomoscore is as follows:

$$\begin{aligned} Nomoscore = {} & (Intercept) * -7.56566042486333 \\ & + Age * 0.127948643930096 \\ & + Location * -1.49110528477808 \\ & + Sex * 1.64766133092359 \\ & + Radscore * 2.22277808425775 \end{aligned}$$

TABLE 1 | Clinic-radiological characteristics of patients in the training and testing sets.

Characteristic	Training set		Internal testing set		External testing set	
	dMMR	pMMR	dMMR	pMMR	dMMR	pMMR
Age (Y)						
mean (sd)	72.8 (9.1)	65.6 (10.4)	69.3 (7)	68.1 (8.1)	70.0 (8.5)	64.8 (9.3)
Sex						
Male	13 (54.2)	38 (80.9)	6 (54.5)	15 (78.9)	9 (42.9)	33 (84.6)
Female	11 (45.8)	9 (19.1)	5 (45.5)	4 (21.1)	12 (57.1)	6 (15.4)
Location						
Cardia	1 (4.2)	14 (29.8)	1 (9.1)	7 (36.8)	17 (81.0)	12 (30.8)
Gastric body	10 (41.7)	19 (40.4)	5 (45.5)	7 (36.8)	3 (14.3)	21 (53.8)
Antrum	13 (54.2)	14 (29.8)	5 (45.5)	5 (26.3)	1 (4.7)	6 (15.4)
Rad-sore						
Median [iqr]	-0.2 [-0.7, 0.6]	-1.1 [-1.5,-0.8]	-0.6 [-0.8, 0.0]	-1.1 [-1.3,-0.7]	-1.1 [-1.4,-0.8]	-1.5 [-1.9,-1.0]

Performance of the Radiomics Nomogram

Table 3 lists the performance of the radiomics nomogram in the training, internal, and external testing sets. The prediction model based on the radiomics features provided only moderate predictive power, as shown in **Figure 5**. The AUC value of radiomics signature in the training set and internal testing sets was 0.81 and 0.78, respectively. The predictive model based on clinical features alone showed that the AUC values in the training and internal testing sets were 0.82 and 0.69, respectively. The radiomics nomogram model combining clinical factors and radiomics features shows superior ability to differentiate MMR status compared with the other two models generated with clinical features and radiomics features alone. The AUC values of the radiomics nomogram in the training set and internal testing set were 0.93 and 0.82 **(Figure 5).** The external testing set radiomics nomogram showed an AUC value of 0.83 **(Figure 6)**. The calibration curve of the radiomics nomogram showed good predictions in both the training and validation cohorts **(Figure 7)**. The DCA of the radiomics nomogram demonstrated the higher overall net benefit compared to the clinics model, showing an excellent clinical utility in distinguishing MMR status (**Figure 8**).

DISCUSSION

In the present study, we developed and validated a radio-clinical nomogram for the prediction of the MMR status of GC perioperatively. The user-friendly nomogram, which consisted of the radiomics signature, sex, age, and tumor location, showed good performance in both cohorts and may effectively stratify patients according to MMR status. The combined analysis of multiple radiomics and clinical markers as a signature is the approach that demonstrates the most promise to change clinical practice (21, 33).

Since MSI was detected in many different types of tumors, the MMR status of tumors has become an important determinant in the choice of therapeutic method. In recent years, immunotherapy has gradually attracted attention and has developed rapidly. Immune checkpoint inhibitors, including anti-programmed death-1 and anti-cytotoxic T-lymphocyte-associated protein-4 antibodies, were effective for MSI-high or dMMR solid tumors in many trials (34). In 2017, the Food and Drug Administration of the United States approved pembrolizumab to treat patients with dMMR/MSI-H non-resectable or solid metastatic tumors. The MSI status is currently used as a biomarker for cancer immunotherapy (35). In addition, MMR status plays an important role in predicting the efficacy of neoadjuvant chemotherapy (13, 14). Accurate prediction of the DNA mismatch repair deficiency status is consequential for the selection of individualized treatment plans in patients with GC. A recent study found that deep learning can differentiate routine hematoxylin and eosin (H&E)-stained, formalin-fixed, paraffin-embedded digital whole-slide images (WSIs) of colorectal cancer into those with microsatellite stability and microsatellite

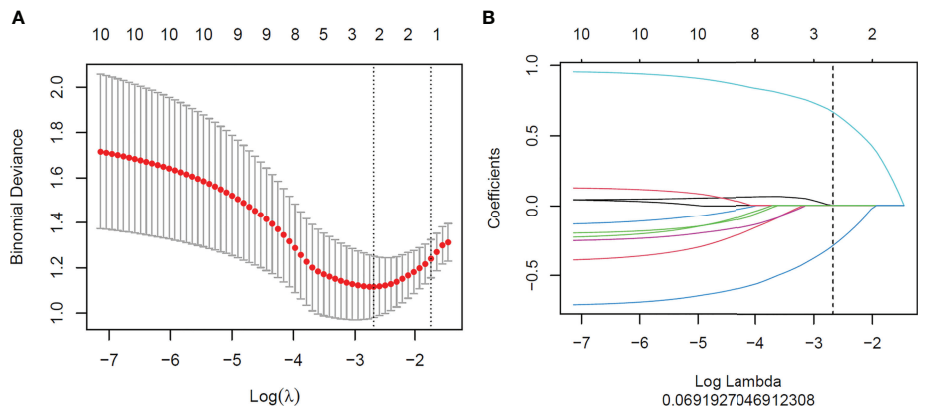


FIGURE 3 | Feature selection with the least absolute shrinkage and selection operator (LASSO) binary logistic regression model. **(A)** Tuning parameter (λ) selection of the LASSO model. Binomial deviance was drawn versus log(λ). Vertical dotted lines were plotted at the best value using 10-fold cross-validation to tune parameter (λ) selection in the LASSO model. **(B)** LASSO coefficient profiles of the features. Each colored line represents the corresponding coefficient of each feature. A vertical dotted line was drawn at the selected λ, where non-zero coefficients were obtained with two features.

TABLE 2 | Univariate and multivariate logistic regression analysis of the clinic-radiological features.

Characteristics	Univariate analysis			Multivariate analysis		
	OR	95% CI	P value	OR	95% CI	P value
Age	1.08	[1.02;1.14]	<0.01	1.14	[1.03;1.25]	<0.01
Sex	3.57	[1.21;10.55]	0.021	5.19	[0.88;30.54]	0.068
Location	0.37	[0.17;0.79]	0.012	0.23	[0.07;0.73]	0.013
Rad score	5.51	[2.30;13.18]	<0.01	9.23	[2.95;28.92]	<0.01

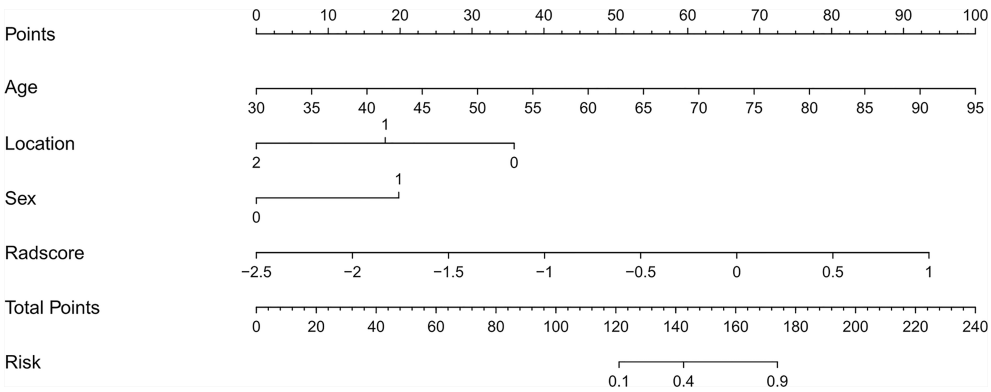


FIGURE 4 | The CT-based radiomics nomogram. The radiomics nomogram was built in the training cohort, with the radiomics signature, sex (0 is male, 1 is female), age, and tumor location (0 is antrum, 1 is gastric body, 2 is cardia).

TABLE 3 | Predictive performance of the radiomics nomogram.

Radiomics nomogram	AUC (95%CI)	Accuracy	Sensitivity	Specificity	PPV	NPV
Training set	0.93 (0.85–1.00)	0.873	0.917	0.851	0.759	0.952
Internal testing set	0.82 (0.66–0.98)	0.733	0.616	0.824	0.727	0.737
External testing set	0.83 (0.73–0.94)	0.767	0.821	0.667	0.821	0.667

instability, with an AUC of up to 0.84 (36). Rikiya et al. developed a deep learning model using 100 H&E-stained WSIs and found that they performed better than human experts (gastrointestinal pathologists) at detecting MSI in routine H&E-stained WSI (37). Some researchers have begun to use artificial intelligence to predict gene expression status non-invasively. The present study used more easily available imaging data and achieved good predictive performance. Radiomics enables non-invasive detection of the revealing relationship between invisible high-dimensional image features and pathophysiological characteristics. Radiomics has developed rapidly in recent years, and now more than 1,000 radiomics features are available for various aspects of tumor heterogeneity (38). The advantage of this study was presumably that it took radiomics scores, incorporating numerous quantitative features, into consideration. Radio-genomics builds on radiomics, which hypothesizes that genomic heterogeneity at the microscopic level may manifest in the tumor, and changes in the microenvironment within the tumor can be expressed on macroscopic images (18). Yang et al. reported that the proposed CT-based radiomics signature is associated with KRAS/NRAS/BRAF mutations; their study indicated that CT may be useful for the analysis of tumor genotype in colorectal cancer and thus helpful to determine therapeutic strategies (39). Combining analysis of clinical features and CT-based radiomics signature may improve predictive efficacy and allow patients to non-invasively choose individualized treatment plans (40).

In this study, dMMR accounted for only 8 percent of GC, a cohort of 80 out of 1,000 patients. Therefore, the sample size of this experiment is small. Pathophysiological characteristics are the foundation of the radiomics features. Since histopathological types and grades have more influence on image performance than genotypes in GC, this study limited the pathological type of GC to confirmed gastric adenocarcinoma at stage 3 or 4, excluding signet ring cell carcinoma or mucinous adenocarcinoma. The above criteria aimed to minimize the influence of factors other than DNA mismatch repair status on image performance.

In our study, radiomics signature comprised two robust radiomics features and manifest moderate predictive efficacy. Texture features consider the interaction between neighboring pixels and are therefore more propitious to quantifying tumor heterogeneity (41). The LASSO algorithm was used for feature redundancy elimination. This method has two primary preponderances. First, it allows features to be selected on the foundation of their univariable association with the outcome without overfitting. Next, it enables a signature to be constructed by a group of selected features (42). In this study, two texture features related to dMMR were selected to build the radiomics signature, which were intended to reveal tumor characteristics that are not apparent in the visual image (43). The two features were Original first-order Median and Original shape-Maximum 2D diameter Column. Original first-order Median is a first-order feature, while first-order statistics describes the distribution of voxel intensities within the image region defined by the mask through commonly used and basic metrics. The meaning of "median" in this context is the median gray-level intensity within the ROI. Original shape-Maximum 2D diameter Column is a shape feature. Maximum 2D diameter Column is defined as the largest pair-wise Euclidean distance between tumor surface mesh vertices in the row-slice plane. Using this approach, we attempted to develop a radiomics signature for the prediction of DNA mismatch repair

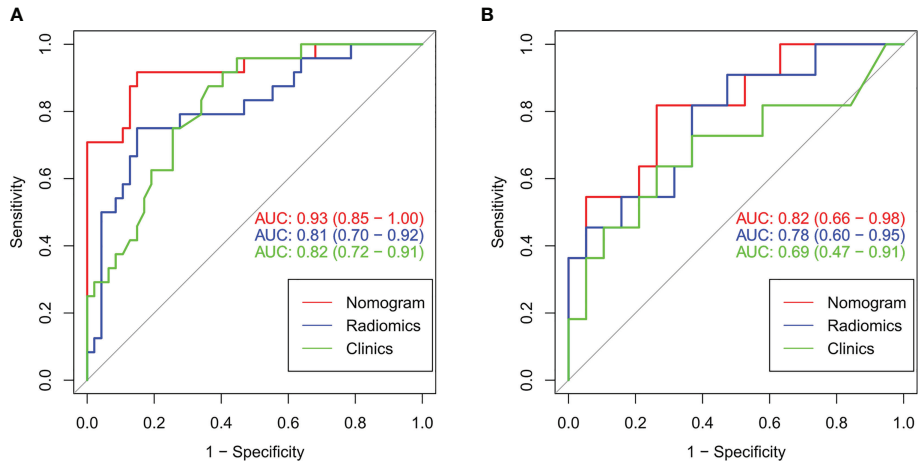


FIGURE 5 | The ROC curves (AUC) of the three models in the training set **(A)** and internal testing set **(B)**.

deficiency in patients with GC. Our radiomics signature exhibited moderate discrimination, with an AUC of 0.81 in the training set and 0.78 in the internal testing set.

In this study, we extracted 2D CT annotations radiomics features based on single CT image slices. Meng et al. conducted a multicenter study comprehensively comparing the representation and discrimination capacity of 2D and 3D radiomics features regarding GC. The results based on three tasks showed that 2D and 3D models showed comparable ability to characterize GC. Their study indicated that 2D CT annotations might be a better choice than 3D in GC radiomics studies, because the latter may add noise (44).

Furthermore, the present study was not limited to the use of a single CT image slice. The importance of clinical characteristics should not be neglected, and the radiomics-derived data cannot predict all clinical decision problems. The univariate analysis showed that three clinical features (gender, age, and tumor location) were independent predictors. We then constructed the nomogram, a user-friendly, graphical analog computation device. The nomogram has clinical significance in the support of

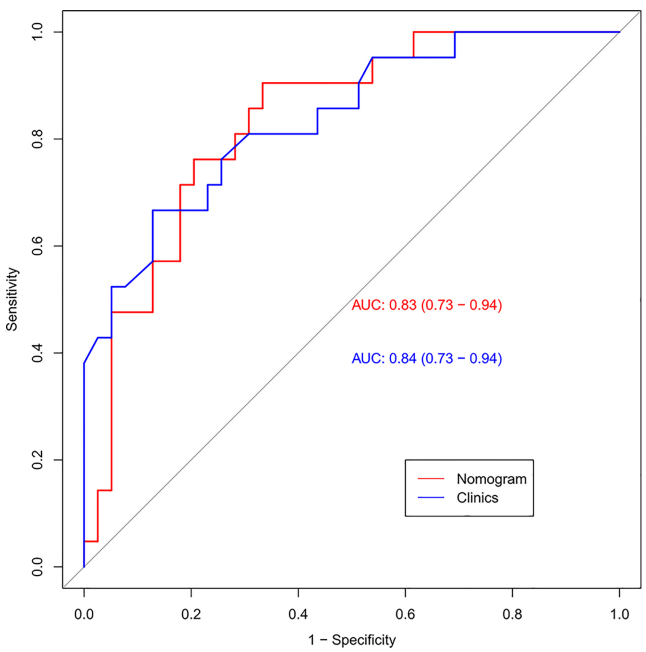


FIGURE 6 | The ROC curves (AUC) of the external testing set.

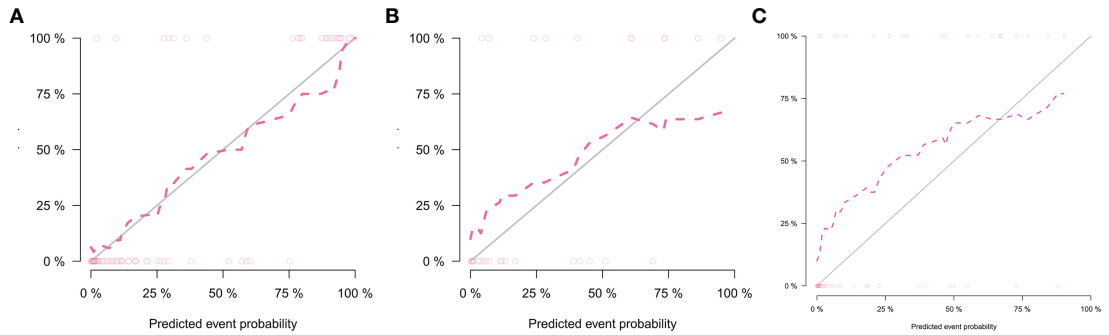


FIGURE 7 | Calibration curves of the nomogram in the training set **(A)**, internal testing set **(B)**, and external testing set **(C)**.

clinicians selecting individualized treatment for patients with GC. The AUC of the nomogram was 0.93, suggesting that the radiomics nomogram achieved greater predictive efficacy than either the radiomics signature or the clinical predictive model alone. The calibration and discrimination in the internal and external validation sets were also good. As a previous study revealed, dMMR GC typically has an antral location (45). Consistent with former research, the tumor locations in the present study were significantly different between the dMMR group and the pMMR group, with dMMR GC more likely to occur in the gastric antrum. The results of the present study also showed that pMMR (without DNA mismatch repair gene deficiency) was more likely to occur in men and at a younger age than the defective form. In contrast, Wang et al. reported that dMMR GC was more common in men (65% vs. 35%) (46) and that most of the cases were stage 2. In the present study, dMMR GC was more common in women, and this difference may be due to Wang's study including mostly dMMR cases at stage 2, while the present study only included stage 3 or 4 GC patients.

The strength of our study is that the radiomics nomogram consists of only three clinical factors that are easily accessible preoperatively. Thus, the nomogram developed here may be used as a credible and non-invasive modality to preoperatively predict DNA mismatch repair deficiency in GC.

Our study was subject to some limitations. Firstly, the sample size of this study is small, including few patients with dMMR GC. Secondly, the tumor segmentation was manually sketched, which is time-consuming and laborious. In future work, computer algorithm-assisted automatic segmentation should be used. Thirdly, due to the retrospective nature of our study, selection bias was difficult to avoid, and patients not eligible for surgery were excluded. Fourthly, the slice thickness of most segmented CT images is 5 mm, and the volume effect of segmented CT images with a diameter of less than 5 mm is clear.

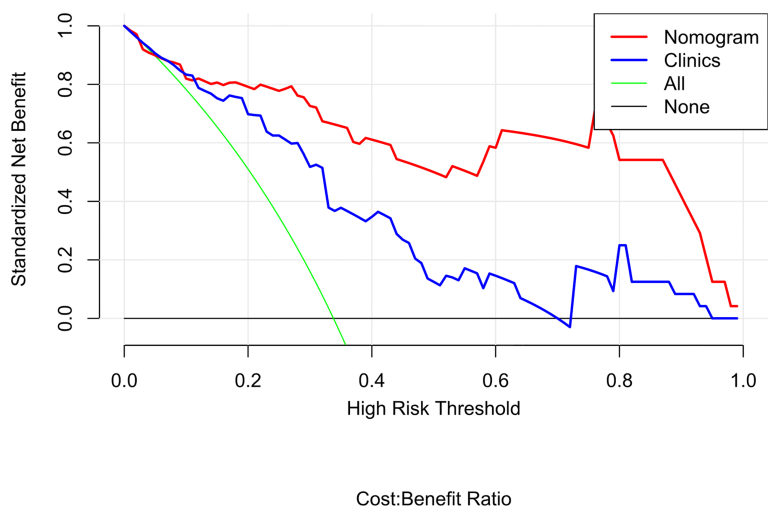


FIGURE 8 | Decision curve analysis (DCA) for the radiomics nomogram and clinics model. The DCA indicated that more net benefits within the most of thresholds probabilities were achieved using the radiomics nomogram.

CONCLUSIONS

In conclusion, our study demonstrated that the radiomics nomogram based on radiomics signature and clinical characteristics (age, sex, and tumor location) may be used for personalized preoperative prediction of DNA mismatch repair deficiency of GC and thereby assist in clinical decision-making.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Cancer Hospital of the University of Chinese Academy of Science. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

YT completed the initial manuscript and designed the whole study. JL and JC collected the clinical and imaging data and participated in revising the manuscript. CH collected the patients and recorded the needed information. ZX participated in its design. SD participated in the statistics and provided the result interpretation. XW helped collect the cases. RY helped collect the cases and reviewed the manuscript. XC revised the manuscript and guaranteed the entire study. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Key Laboratory of Prevention, Diagnosis and Therapy of Upper Gastrointestinal Cancer of Zhejiang Province (2022E10021) and the Medical Health Science and Technology Project of Zhejiang Province (2022KY655).

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2018. *CA Cancer J Clin* (2018) 68:7–30. doi: 10.3322/caac.21442
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence Andmortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2018) 68:394–424. doi: 10.3322/caac.21492
3. Yang L, Zheng R, Wang N, Yuan Y, Liu S, Li H, et al. Incidence and Mortality of Stomach Cancer in China, 2014. *Chin J Cancer Res* (2018) 0(3):291–8. doi: 10.21147/j.issn.1000-9604.2018.03.01
4. Peltomäki P, Lothe RA, Aaltonen LA, Pylkkänen L, Nyström-Lahti M, Seruca R, et al. Microsatellite Instability Is Associated With Tumors That Characterize the Hereditary Non-Polyposis Colorectal Carcinoma Syndrome. *Cancer Res* (1993) 53(24):5853–5. doi: 10.1002/1097-0142(19931215)72:12
5. Gonzalez-Zulueta M, Ruppert JM, Tokino K, Tsai YC, Spruck CH3rd, Miyao N, et al. Microsatellite Instability in Bladder Cancer. *Cancer Res* (1993) 53 (23):5620–3. doi: 10.1016/0165-4608(93)90031-G
6. Shridhar V, Siegfried J, Hunt J, del Mar Alonso M, Smith DI. Genetic Instability of Microsatellite Sequences in Many non-Small Cell Lung Carcinomas. *Cancer Res* (1994) 54(8):2084–87. doi: 10.1002/1097-0142 (19940415)73:8
7. Machado AM, Figueiredo C, Touati E, Máximo V, Sousa S, Michel V, et al. Helicobacter Pylori Infection Induces Genetic Instability of Nuclear and Mitochondrial DNA in Gastric Cells. *Clin Cancer Res* (2009) 15(9):2995–3002. doi: 10.1158/1078-0432.CCR-08-2686
8. Ling ZQ, Tanaka A, Li P, Nakayama T, Fujiyama Y, Hattori T, et al. Microsatellite Instability With Promoter Methylation and Silencing of Hmlh1 can Regionally Occur During Progression of Gastric Carcinoma. *Cancer Lett* (2010) 297:244–51. doi: 10.1016/j.canlet.2010.05.017
9. Usui G, Matsusaka K, Mano Y, Urabe M, Funata S, Fukayama M, et al. DNA Methylation and Genetic Aberrations in Gastric Cancer. *Digestion* (2021) 102:25–32. doi: 10.1159/000511243
10. Chia NY, Tan P. Molecular Classification of Gastric Cancer. *Ann Oncol* (2016) 27:763 –69. doi: 10.1093/annonc/mdw040
11. Shitara K, Özgüroğlu M, Bang YJ, Di Bartolomeo M, Mandalà M, Ryu MH, et al. Molecular Determinants of Clinical Outcomes With Pembrolizumab Versus Paclitaxel in a Randomized, Open-Label, Phase III Trial in Patients With Gastroesophageal Adenocarcinoma. *Ann Oncol* (2021) 32:1127–36. doi: 10.1016/j.annonc.2021.05.803
12. Cancer Genome Atlas Research Network. Comprehensive Molecular Characterization of Gastric Adenocarcinoma. *Nature* (2014) 513 (7517):202–9. doi: 10.1038/nature13480
13. Li Z, Gao X, Peng X, May Chen MJ, Li Z, Wei B, et al. Multi-Omics Characterization of Molecular Features of Gastric Cancer Correlated With Response to Neoadjuvant Chemotherapy. *Sci Adv* (2020) 6:eaay4211. doi: 10.1126/sciadv.aay4211
14. Choi YY, Kim H, Shin SJ, Kim HY, Lee J, Yang HK, et al. Microsatellite Instability and Programmed Cell Death-Ligand 1 Expression in Stage II/III Gastric Cancer: *Post Hoc* Analysis of the CLASSIC Randomized Controlled Study. *Ann Surg* (2019) 270:309–16. doi: 10.1097/SLA.0000000000002803
15. National Comprehensive Cancer Network. *NCCN Guidelines Version 2 - Gastric Cancer* (2022). Available at: https://www.nccn.org/professionals/physician_gls/pdf/gastric.pdf (Accessed January 11, 2022).
16. Boland CR, Goel A. Microsatellite Instability in Colorectal Cancer. *Gastroenterology* (2010) 138:2073–87. doi: 10.1053/j.gastro.2009.12.064
17. Kawakami H, Zaanan A, Sinicrope FA. Microsatellite Instability Testing and its Role in the Management of Colorectal Cancer. *Curr Treat Opt Oncol* (2015) 16:30. doi: 10.1007/s11864-015-0348-2
18. 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) 48:441–46. doi: 10.1016/j.ejca.2011.11.036
19. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding Tumour Phenotype by Noninvasive Imaging Using Aquantitative Radiomics Approach. *Nat Commun* (2014) 5:4006. doi: 10.1038/ncomms5644
20. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are More Than Pictures, They are Data. *Radiology* (2016) 278:563–77. doi: 10.1148/radiol.2015151169
21. Huang Y, Liu Z, He L, Chen X, Pan D, Ma Z, et al. Radiomics Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early Stage (I or II) non-Small Cell Lung Cancer. *Radiology* (2016) 281:947–57. doi: 10.1148/radiol.2016152234

22. Wilson R, Devaraj A. Radiomics of Pulmonary Nodules and Lung Cancer. *Transl Lung Cancer Res* (2017) 6:86–91. doi: 10.21037/tlcr.2017.01.04
23. Zhang J, Wu Z, Zhao J, Liu S, Zhang X, Yuan F, et al. Intrahepatic Cholangiocarcinoma: MRI Texture Signature as Predictive Biomarkers of Immunophenotyping and Survival. *Eur Radiol* (2021) 31:3661–72. doi: 10.1007/s00330-020-07524-y
24. Bodalal Z, Trebeschi S, Nguyen-Kim TDL, Schats W, Beets-Tan R. Radiogenomics: Bridging Imaging and Genomics. *Abdom Radiol* (2019) 44:1960–84. doi: 10.1007/s00261-019-02028-w
25. 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) 14:749–62. doi: 10.1038/nrclinonc
26. Kim M, Jung SY, Park JE, Jo Y, Park SY, Nam SJ, et al. Diffusion- and Perfusion-Weighted MRI Radiomics Model may Predict Isocitrate Dehydrogenase (IDH) Mutation and Tumor Aggressiveness in Diffuse Lower Grade Glioma. *Eur Radiol* (2020) 30(4):2142–51. doi: 10.1007/s00330-019-06548-3
27. Zhang W, Fang M, Dong D, Wang X, Ke X, Zhang L, et al. Development and Validation of a CT-Based Radiomic Nomogram for Preoperative Prediction of Early Recurrence in Advanced Gastric Cancer. *Radioth Oncol* (2020) 145:13–20. doi: 10.1016/j.radonc.2019.11.023
28. Tan Y, Zhang ST, Wei JW, Dong D, Wang XC, Yang GQ, et al. A Radiomics Nomogram may Improve the Prediction of IDH Genotype for Astrocytoma Before Surgery. *Eur Radiol* (2019) 29:3325–37. doi: 10.1007/s00330-019-06056-4
29. Wu S, Zheng J, Li Y, Yu H, Shi S, Xie W, et al. A Radiomics Nomogram for the Preoperative Prediction of Lymph Node Metastasis in Bladder Cancer. *Clin Cancer Res* (2017) 23:6904–11. doi: 10.1158/1078-0432.CCR-17-1510
30. Wang Y, Liu W, Yu Y, Liu JJ, Xue HD, Qi YF, et al. CT Radiomics Nomogram for the Preoperative Prediction of Lymph Node Metastasis in Gastric Cancer. *Eur Radiol* (2020) 30:976–86. doi: 10.1007/s00330-019-06398-z
31. Wang S, Chen Y, Zhang H, Liang Z, Bu J. The Value of Predicting Human Epidermal Growth Factor Receptor 2 Status in Adenocarcinoma of the Esophagogastric Junction on CT-Based Radiomics Nomogram. *Front Oncol* (2021) 11:707686. doi: 10.3389/fonc.2021.707686
32. Du F, Tang N, Cui Y, Wang W, Zhang Y, Li Z, et al. A Novel Nomogram Model Based on Cone-Beam CT Radiomics Analysis Technology for Predicting Radiation Pneumonitis in Esophageal Cancer Patients Undergoing Radiotherapy. *Front Oncol* (2020) 10:596013. doi: 10.3389/fonc.2020.596013
33. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. *J Clin Oncol* (2016) 34:2157–64. doi: 10.1200/jco.2015.65.9128
34. Nan HM, Song YJ, Yun HY, Park JS, Kim H. Effects of Dietary Intake and Genetic Factors on Hypermethylation of the Hmlh1 Gene Promoter in Gastric Cancer. *World J Gastroenterol* (2005) 11:3834–41. doi: 10.3748/wjg.v11.i25.3834
35. 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) 372:2509–20. doi: 10.1056/NEJMoa1500596
36. Kather JN, Pearson AT, Halama N, Jäger D, Krause J, Loosen SH, et al. Deep Learning can Predict Microsatellite Instability Directly From Histology in Gastrointestinal Cancer. *Nat Med* (2019) 25:1054–6. doi: 10.1038/s41591-019-0462-y
37. Yamashita R, Long J, Longacre T, Peng L, Berry G, Martin B, et al. Deep Learning Model for the Prediction of Microsatellite Instability in Colorectal Cancer: A Diagnostic Study. *Lancet Oncol* (2021) 22:132–41. doi: 10.1016/S1470-2045(20)30535-0
38. Nie P, Yang G, Wang Z, Yan L, Miao W, Hao D, et al. A CT-Based Radiomics Nomogram for Differentiation of Renal Angiomyolipoma Without Visible Fat From Homogeneous Clear Cell Renal Cell Carcinoma. *Eur Radiol* (2020) 30:1274–84. doi: 10.1007/s00330-019-06427-x
39. Yang L, Dong D, Fang M, Zhu Y, Zang Y, Liu Z, et al. Can CT-Based Radiomics Signature Predict KRAS/NRAS/BRAF Mutations in Colorectal Cancer? *Eur Radiol* (2018) 28:2058–67. doi: 10.1007/s00330-017-5146-8
40. Fan S, Li X, Cui X, Zheng L, Ren X, Ma W, et al. Computed Tomography-Based Radiomic Features Could Potentially Predict Microsatellite Instability Status in Stage II Colorectal Cancer: A Preliminary Study. *Acad Radiol* (2019) 26:1633–40. doi: 10.1016/j.acra.2019.02.009
41. Park H, Lim Y, Ko ES, Cho HH, Lee JE, Han BK, et al. Radiomics Signature on Magnetic Resonance Imaging: Association With Disease-Free Survival in Patients With Invasive Breast Cancer. *Clin Cancer Res* (2018) 24:4705–14. doi: 10.1158/1078-0432.CCR-17-3783
42. Hepp T, Schmid M, Gefeller O, Waldmann E, Mayr A. Approaches to Regularized Regression - A Comparison Between Gradient Boosting and the Lasso. *Methods Inf Med* (2016) 55:422–30. doi: 10.3414/ME16-01-0033
43. Lee SE, Han K, Kwak JY, Lee E, Kim EK. Radiomics of US Texture Features in Differential Diagnosis Between Triple-Negative Breast Cancer and Fibroadenoma. *Sci Rep* (2018) 8:13546. doi: 10.1038/s41598-018-31906-4
44. Meng L, Dong D, Chen X, Fang M, Wang R, Li J, et al. 2D and 3D CT Radiomic Features Performance Comparison in Characterization of Gastric Cancer: A Multi-Center Study. *IEEE J BioMed Health Inform* (2021) 25:755–63. doi: 10.1109/JBHI.2020.3002805
45. Cristescu R, Lee J, Nebozhyn M, Kim KM, Ting JC, Wong SS, et al. Molecular Analysis of Gastric Cancer Identifies Subtypes Associated With Distinct Clinical Outcomes. *Nat Med* (2015) 21:449–56. doi: 10.1038/nm.3850
46. Wang J, Xi Y, Zhao J, Rong X, Lu W, Wang Y. The Clinicopathological Characteristics and Prognoses of dMMR Gastric Adenocarcinoma Patients. *Gastroenterol Res Pract* (2021) 2021:4269781. doi: 10.1155/2021/4269781

Conflict of Interest: Author SD was employed by GE Healthcare.

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

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

frontiers | Frontiers in Oncology

REVIEW
published: 13 July 2022
doi: 10.3389/fonc.2022.893972

Skin Cancer Classification With Deep Learning: A Systematic Review

Yinhao Wu[1], *Bin Chen*[2], *An Zeng*[3], *Dan Pan*[4], *Ruixuan Wang*[5] *and Shen Zhao*[1]*

[1] School of Intelligent Systems Engineering, Sun Yat-Sen University, Guangzhou, China, [2] Affiliated Hangzhou First People's Hospital, Zhejiang University School of Medicine, Zhejiang, China, [3] School of Computer Science and Technology, Guangdong University of Technology, Guangzhou, China, [4] School of Electronics and Information, Guangdong Polytechnic Normal University, Guangzhou, China, [5] School of Computer Science and Engineering, Sun Yat-Sen University, Guangzhou, China

OPEN ACCESS

Edited by:
Guang Yang,
Imperial College London,
United Kingdom

Reviewed by:
Anju Yadav,
Manipal University Jaipur, India
Dong Zhang,
University of British Columbia, Canada
Zhiyuan Zhu,
Beijing Normal University, China
Sheng Lian,
Fuzhou University, China
Tarun Jain,
Manipal University Jaipur, India

***Correspondence:**
Shen Zhao
z-s-06@163.com

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

Received: 11 March 2022
Accepted: 16 May 2022
Published: 13 July 2022

Citation:
Wu Y, Chen B, Zeng A, Pan D, Wang R and Zhao S (2022) Skin Cancer Classification With Deep Learning: A Systematic Review. Front. Oncol. 12:893972. doi: 10.3389/fonc.2022.893972

Skin cancer is one of the most dangerous diseases in the world. Correctly classifying skin lesions at an early stage could aid clinical decision-making by providing an accurate disease diagnosis, potentially increasing the chances of cure before cancer spreads. However, achieving automatic skin cancer classification is difficult because the majority of skin disease images used for training are imbalanced and in short supply; meanwhile, the model's cross-domain adaptability and robustness are also critical challenges. Recently, many deep learning-based methods have been widely used in skin cancer classification to solve the above issues and achieve satisfactory results. Nonetheless, reviews that include the abovementioned frontier problems in skin cancer classification are still scarce. Therefore, in this article, we provide a comprehensive overview of the latest deep learning-based algorithms for skin cancer classification. We begin with an overview of three types of dermatological images, followed by a list of publicly available datasets relating to skin cancers. After that, we review the successful applications of typical convolutional neural networks for skin cancer classification. As a highlight of this paper, we next summarize several frontier problems, including data imbalance, data limitation, domain adaptation, model robustness, and model efficiency, followed by corresponding solutions in the skin cancer classification task. Finally, by summarizing different deep learning-based methods to solve the frontier challenges in skin cancer classification, we can conclude that the general development direction of these approaches is structured, lightweight, and multimodal. Besides, for readers' convenience, we have summarized our findings in figures and tables. Considering the growing popularity of deep learning, there are still many issues to overcome as well as chances to pursue in the future.

Keywords: generative adversarial networks, convolutional neural network, deep learning, skin cancer, image classification

1 INTRODUCTION

Given the rising prevalence of skin cancer and the significance for early detection, it is crucial to develop an effective method to automatically classify skin cancer. As the largest organ of the human body (1), the skin shoulders the responsibility of protecting other human systems, which increases its vulnerability to disease (2). Melanoma was the most common cancer in both men and women with approximately 300,000 new cases (3) diagnosed globally in 2018. In addition to melanoma, two other major skin cancer diseases, basal cell carcinoma (BCC) and squamous cell carcinoma (SCC), also had a relatively high incidence, with over 1 million cases in 2018 (4). As (5) reported, more skin cancers are diagnosed each year than all other cancers combined in the United States. Fortunately, if detected early, the chances of cure will be greatly improved. According to (4), melanoma has a 5-year survival rate of 99% when it does not metastasize. If it metastasizes to other organs in the body, its survival rate reduces to 20%. However, because early indications of skin cancer are not always visible, diagnostic results are often dependent on the dermatologist's expertise (6). For inexperienced practitioners, an automatic diagnosis system is an essential tool for more accurate diagnoses. Beyond that, diagnosing skin cancer with naked eyes is highly subjective and rarely generalizable (7). Therefore, it is necessary to develop an automatic classification method for skin cancer that is more accurate, less expensive, and quicker to diagnose (8). Besides, implementing such automated diagnostic systems can effectively minimize mortality from skin cancers, benefiting both patients and the healthcare systems (9).

However, owing to the complexity and diversity of skin disease images, achieving automatic classification of skin cancer is challenging. First of all, different skin lesions have lots of interclass similarities, which could result in misdiagnosis (10). For example, there exist various mimics of BCC in histopathological images, such as SCC and other skin diseases (11). As a result, it is difficult for the diagnosis systems to effectively discriminate skin malignancies from their known imitators. Secondly, several skin lesions differ within their same class in terms of color, feature, structure, size, and location (12). For example, the appearance of BCC and its subcategories is almost different. This makes it difficult to classify different subcategories of the same category. Furthermore, the classification algorithms are highly sensitive to the types of camera devices used to capture images. When the test images come from a different domain, their performance suffers (13).

Although traditional machine learning approaches are capable of performing well in particular skin cancer classification tasks, these algorithms are ineffective for complicated diagnostic demands in clinical practice. Traditional machine learning methods for skin cancer diagnosis typically involve extracting features from skin-disease images and then classifying the extracted features (14). For example, ABCD Rule (15), Menzies Method (16), and 7-Point Checklist (17) are effective methods for extracting various features from skin disease images. The handcrafted features are then classified using several classification methods such as SVM (18), XGBoost (19), and decision tree (20). Due to the restricted number of selected features, machine learning algorithms can often only classify a subset of skin cancer diseases and cannot generalize to a broader range of disease types (21). Besides, given the wide variety of skin cancers, it is not effective to identify each form of cancer solely based on handcrafted features (22).

Without the need for domain expertise and feature extraction, deep learning algorithms have been widely used for skin cancer classification in recent years; however, there are still several difficulties and challenges ahead. Compared with traditional machine learning methods, deep learning algorithms can analyze data from a large-scale dataset faster and more accurately, which allows them to effectively extract relevant characteristics (23). At the same time, deep learning algorithms can also aid clinicians in more thorough data analysis and examination of test results (24). A number of studies, such as (25–27) demonstrated that deep learning algorithms can diagnose at a level comparable to that of a dermatologist. However, these algorithms still have many obstacles to becoming a complete diagnostic system. Firstly, data imbalance and the lack of a large volume of labeled images have hindered the widespread use of deep learning methods in skin cancer classification (12). When these algorithms are used to classify skin cancers that are rare in the training dataset, they frequently result in a misdiagnosis (28). Furthermore, when working with high-resolution images (such as pathological images) with millions of pixels, the deep learning models often result in significant computing costs and additional training time (29). Besides, different noises will be generated as a result of the various conditions (such as different imaging devices, backgrounds). Therefore, the robustness and generalization ability of these algorithms should also be taken into account (30).

These years, a number of reviews that detail the diagnostic breakthroughs in skin cancer classification have been published; however, no review has provided a specific analysis of frontier challenges in skin cancer classification tasks, such as data imbalance and limitation, domain adaptability, model robustness, and model efficiency (31). reviewed the recent developments in skin lesion classification using dermoscopic images (32). presented a detailed overview of studies on using CNNs to classify skin lesions (33). showed how the use of CNNs in correctly identifying skin cancer has developed (34). presented a review of different machine learning algorithms in dermatology diagnosis, as well as some of the obstacles and limitations (12). and (28) summarized a number of deep learning-based approaches for skin cancer classification, as well as various challenges and difficulties (35). provided an in-depth review of the current articles about melanoma classification and compared their results with human experts (36). summarized the latest CNN-based methods in skin lesion classification by utilizing image data and patient data (37). provided a review of deep learning-based methods for early diagnosis of skin cancer. We present these relevant surveys with details and highlights in **Table 1**. By summarizing the previous reviews, we find that all of the preceding publications methodically studied a specific topic

TABLE 1 | A summary of the current review related to skin cancer classification.

Ref.	Title	Venue	Remarks
(32)	Skin Cancer Classification Using Convolutional Neural Networks: Systematic Review	Journal of Medical Internet Research	This study presents a detailed overview of studies on using CNNs to classify skin lesions.
(31)	Techniques and algorithms for computer aided diagnosis of pigmented skin lesions—A review	Biomedical Signal Processing and Control	This paper gives a review of the recent developments in skin lesion classification using dermoscopic images.
(33)	Classification of Skin cancer using deep learning, Convolutional Neural Networks -Opportunities and vulnerabilities-A systematic Review	International Journal for Modern Trends in Science and Technology	This article reviews the development of deep learning for skin cancer classification tasks.
(34)	Machine Learning in Dermatology: Current Applications, Opportunities, and Limitations	Dermatology and Therapy volume	This paper reviews the fundamentals of machine learning and its wide range of applications in dermatology.
(12)	Artificial intelligence-based image classification methods for diagnosis of skin cancer: Challenges and opportunities	Computers in Biology and Medicine	This review discusses the developments in AI-based methods for skin cancer diagnosis, as well as challenges and future directions to enhance them.
(35)	Skin cancer classification *via* convolutional neural networks: systematic review of studies involving human experts	European Journal of Cancer	This paper analyses studies comparing AI–based skin cancer classifiers with dermatologists.
(37)	Skin Cancer Detection: A Review Using Deep Learning Techniques	International Journal of Environmental Research and Public Health	This paper provides a review of deep learning-based methods for early diagnosis of skin cancer.
(36)	Integrating Patient Data Into Skin Cancer Classification Using Convolutional Neural Networks: Systematic Review	Journal of Medical Internet Research	This review summarizes the latest CNN-based methods in skin lesion classification by utilizing image data and patient data.
(28)	Skin disease diagnosis with deep learning: A review	Neurocomputing	This paper analyses several deep learning algorithms for diagnosing skin diseases from a variety of perspectives based on the challenges at hand.

in skin cancer classification. However, most of them treated skin cancer classification as a classical classification problem, without addressing the model's significant practical constraints in clinical work, such as data imbalance and limitation, cross-domain adaptability, model robustness, and model efficiency. Although several earlier reviews summarized some of the methods to solve the abovementioned frontier problems, their summaries were incomplete. Some novel techniques were not covered, such as pruning, knowledge distillation, and transformer. Therefore, in this review, we comprehensively summarize the frontier challenges in skin cancer classification and provide corresponding solutions by analyzing articles published until the year 2022. It gives readers in-depth information on the advances and limitations of deep learning in skin cancer classification and also provides different ideas for researchers to improve these algorithms.

The rest of this paper is organized as follows: first of all, Section 2 introduces three types of dermatological images and several popular public datasets. In Section 3, we review several typical CNN frameworks and frontier problems with their corresponding solutions in skin cancer classification tasks. A brief conclusion is given in Section 4.

2 DERMATOLOGICAL IMAGES AND DATASETS

High-quality images of skin diseases are important for both dermatologists and automated diagnostic systems. On the one hand, dermatologists rely on high-resolution (HR) images to make diagnoses when direct observation is impossible (38). This is especially common in telemedicine, medical consultations, and regular clinics (39). On the other hand, training reliable algorithms has always necessitated the use of high-quality data. In particular, deep learning algorithms always need a vast volume of labeled data for a better accuracy (28). As a result, high-quality dermatological images are critical for both clinical diagnosis and the design of new algorithms. In this section, we go over three different types of images commonly used in skin cancer diagnosis, as well as some public datasets.

2.1 Dermatological Images

The three main types of image modalities used to diagnose skin diseases are clinical images, dermoscopy images, and histopathological images (see **Figure 1**). Clinical images are frequently captured by mobile devices for remote diagnosis or as medical records. Dermoscopy images and histopathological images are commonly utilized in clinical diagnosis to assess the severity of the illness. In the next part, we introduce them separately.

2.1.1 Clinical Images

Clinical images are obtained by photographing the skin disease site directly with a camera. They can be used as a medical record for patients and provide different insights for dermoscopy images (12). The biggest issue of utilizing clinical images for skin cancer classification is that they include limited morphological information while also introducing considerable inaccuracies into the diagnostic results, owing to the effect of diverse imaging settings (such as lighting, angle, and so on) (40).

2.1.2 Dermoscopy Images

Dermoscopy images are captured with dermoscopy, a type of optical observation tool used to assess the fine details of skin diseases (41). Clinicians frequently utilize dermoscopy to diagnose benign nevi and malignant melanoma (42). It serves as a bridge between clinical and pathological aspects, and thus dermoscopy is often referred to as a dermatologist's stethoscope.

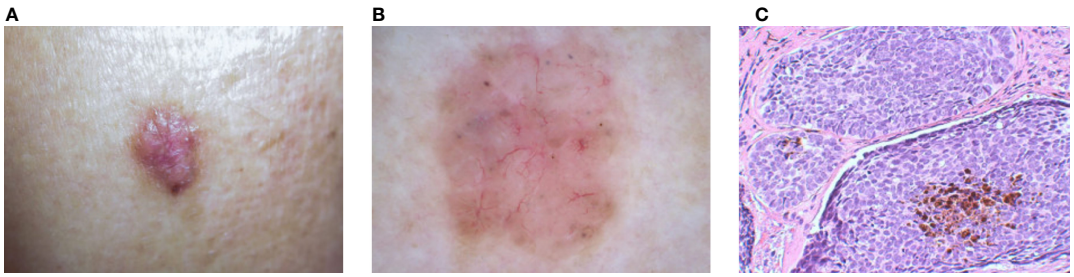


FIGURE 1 | Examples of three types of dermatological images of BCC to show their differences and relationships: **(A)** Clinical image. **(B)** Dermoscopy image. **(C)** Histopathological image.[1]

Dermoscopy images provide a clear visualization of the skin's surface and are used to analyze the color and microstructure of the epidermis (43). For some skin diseases, there are already numerous diagnostic guidelines based on dermoscopy images (44), for example, the ABCD Rule Law (15), the CASH Rule Law (45), and the Menzies Method (16). When using dermoscopy images for skin cancer diagnosis, the range of structures that can be observed is limited, and its diagnostic accuracy is occasionally affected by the experience of dermatologists (46).

2.1.3 Histopathological Images

Histopathological images were obtained using microscopes to scan tissue slides and then digitalize as images (28). They are utilized to show the vertical structure and complete internal characteristics of the diseased tissue. In the clinic, pathological examinations serve as the "gold standard" for diagnosing almost all types of cancers, as they are often used to distinguish between types of cancers and guide appropriate treatment plans based on pathological changes. However, different histopathological images of skin cancer exhibit different morphologies, scales, textures, and color distributions, which makes it difficult to find a common pattern for diagnosis (12).

2.2 Datasets

To create a trustworthy and robust skin cancer classification system, a variety of datasets with all kinds of dermatological images are required. As the need for medical imaging resources in academia grows, more and more datasets are becoming publicly available. To provide readers with a reference, we introduce several commonly used skin-disease datasets in the next part, along with the works based on these datasets.

2.2.1 PH^2 Dataset

The PH^2 dataset is constructed by (47) to support the research of classification and segmentation methods. It contains 200 color dermoscopy images (768 × 560) of three types of skin diseases, including common nevi, atypical nevi, and melanomas. Besides, it contains complete medical annotations, such as lesion segmentation results and pathological diagnosis.

PH^2 is frequently used as a dataset for testing the diagnostic algorithms of skin disease. For example (48), used the SegNet framework to automatically diagnose and segment the dermoscopic images in PH^2 and finally obtained the classification accuracy of 94% (49). proposed a novel deep convolutional network for feature extraction and classification of skin lesions. The model was mainly divided into three stages. The first stage was for data augmentation and image contrast enhancement. The second stage used CNN to extract information from the boundary of the lesion area. The third stage used Hamming distance to fuse and select features obtained with pretrained Inception v3. Finally, the model obtained a classification accuracy of 98.4%, 95.1%, and 94.8% on the PH^2, ISIC-2016, and ISIC-2017 datasets, respectively (50). proposed a Multi-Focus Segmentation Network for skin cancer disease segmentation tasks based on the PH^2 dataset by utilizing feature maps of different scales. Two boundary attention modules and two reverse attention modules were utilized to generate skin lesion masks. Finally, the experimental results revealed that the proposed method achieved a dice similarity coefficient of 0.954 and an IoU index of 0.914 on the PH^2 dataset. In addition to the above works, the PH^2 dataset is being utilized by an increasing number of algorithms to validate their effectiveness and accuracy.

2.2.2 MED-NODE Dataset

The MED-NODE Dataset[3] is collected by the Department of Dermatology of the University Medical Center Groningen (UMCG), which contains 170 digital images of melanoma (51) and nevi case (52). It is used to build and evaluate the MED-NODE system for detecting skin cancer with macroscopic images (53).

On the MED-NODE dataset, a variety of approaches provided significant classification results. For example, in order to improve the generalization ability of the model and alleviate the problem of data imbalance (54), proposed a model for melanoma classification based on transfer learning and ensemble learning. Finally, the model achieved 93% classification accuracy on the MED-NODE dataset, surpassing other state-of-the-art methods (55). applied AlexNet for the skin

[1] https://dermnetnz.org/topics/basal-cell-carcinoma.

[3] https://www.cs.rug.nl/~imaging/databases/melanoma_naevi/.

cancer classification task by using three different transfer learning methods, including fine-tuning the weight parameters of the model, replacing the classification layer function, and performing data augmentation on the original dataset. In the end, they achieved an accuracy of 96.86% on the MED-NODE dataset. Then (56), used two additional networks for the skin cancer classification task, including ResNet-101 and GoogleNet. Finally, experiment results revealed that GoogleNet achieved the best classification accuracy of 99.29% on the MED-NODE dataset. It can be seen that various convolutional neural networks have obtained decent classification results on this dataset; however, the number of skin disease images included is relatively restricted.

2.2.3 HAM10000 Dataset

The HAM10000[4] dataset was collected by the International Skin Imaging Collaboration (ISIC) to solve the problem of data imbalance and data limitation in skin-disease datasets. It contains 10,015 dermoscopic images with seven representative diseases in pigmented skin lesions: nematode disease and intraepithelial carcinoma, basal cell carcinoma, benign keratoid lesions, cutaneous fibroma, melanoma, melanocyte nevi, and vascular lesions (including hemangiomas, purulent granulomas, and subcutaneous hemorrhage) (57, 58).

The HAM10000 dataset is widely used by many scholars due to its diversity of skin lesions. For example (25), used four novel deep CNN models, DenseNet-201, ResNet-152, Inception-v3, and InceptionResNet-v2 to classify eight different types of skin cancers on the HAM10000 and PH[2] datasets. Finally, experimental results indicated that the diagnostic level of these CNN models exceeds the dermatologists in terms of ROC AUC score (59). trained 30 different models on the HAM10000 dataset to explore the classification performance of different models. At the same time, they also used two locally interpretable methods GradCAM and Kernel SHAP techniques to observe the mechanism of the classification model. Finally, the model achieved an average AUC of 0.85 (60). designed a method for classifying seven skin diseases that used ensemble learning and the one-versus-all (OVA) strategy. Finally, they achieved a classification accuracy of 0.9209 on the HAM10000 dataset (61). obtained the best classification result by combining Inception ResNet-v2 with Soft-Attention mechanism on the HAM10000 dataset, with an accuracy of 0.934, an AUC of 0.984, and an average precision of 0.937. With the in-depth study of skin cancer classification tasks by scholars, more and more novel classification methods are being tested on the HAM10000 dataset for a better comparison, where the adoption of the Soft-Attention module yields the best classification results.

2.2.4 Derm7pt Dataset

The Derm7pt dataset contains approximately 2,000 clinical and dermoscopy color images of skin disease, as well as structured information for training and assessing CAD systems. It serves as a database for analyzing the prediction results of the seven-point malignancy checklist of skin lesion (62).

Due to the multimodal information contained in the Derm7pt dataset, it has gradually been widely used to test various multitask networks. When releasing the dataset (62), also proposed a multitask network for predicting melanoma with seven-point checklist criteria and diagnostic results. The model used different loss functions to handle different input modalities, while being able to make predictions on missing data at the output. Finally, the model achieved a classification accuracy of 73.7% on the Derm7pt dataset, also benchmarking the approach. To increase its interpretability (63), created a multitask model based on the Derm7pt dataset to illustrate the mechanism between different tasks. Learnable gates were used in the model to show how the method used or combined features from various tasks. This strategy may be used to investigate how CNN models behave, potentially enhancing their clinical utility (64). proposed a deep convolutional network for skin lesion classification on the Derm7pt dataset. Meanwhile, they implemented regularized DropOut and DropBlock to increase the model's generalization capabilities and reduce overfitting. In addition, to address the dataset's imbalance and limitation, they devised a novel loss function that assigns different weights to various samples, as well as an end-to-end cumulative learning technique. Finally, the method achieved excellent classification performance on the Derm7pt dataset and ISIC dataset while with low computational resources. The release of the Derm7pt dataset has a great boost in promoting the use of multimodal data in skin cancer classification tasks, as well as new ideas and solutions.

2.2.5 BCN20000 Dataset

The BCN20000[5] dataset comprises 5,583 skin lesions and 19,424 dermoscopic images taken using high-resolution dermoscopy. They were all gathered between 2010 and 2016. At the same time, the collector employed a variety of computer vision techniques to remove noise, background, and other interference from the images. Finally, they were carefully reviewed by numerous experts to ensure the diagnosis' validity (65).

BCN20000 is commonly utilized in skin cancer classification and segmentation tasks as part of the dataset for the ISIC-2019 competition. For example, in order to protect the data privacy and avoid data abuse (66), proposed a Distributed Analytics method for distributed training of skin disease images, which ensures that the training data remains in the original institution. Finally, after training on the BCN20000 dataset, the model achieves classification accuracy comparable to the centralized distribution. To evaluate the robustness of different CNN models (67), generated a series of out-of-distribution (OOD) images by using different data augmentation methods based on BCN20000, HAM10000, and other skin-disease datasets. This method establishes a benchmark for OOD testing and considerably facilitates the clinical use of skin cancer classification methods. Specially, by using different data augmentation methods with an ensemble learning strategy (including EfficientNets, SENet, and ResNeXt101_wsl) (68), achieved the first-place classification

[4] https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DBW86T.

[5] https://www.isic-archive.com/, 2019.

TABLE 2 | Characteristics of different skin-disease datasets.

Dataset	No. of images	Modality of images	No. of lesion types	Image format	Published year	Goal of publication
PH[2]	200	Dermoscopic	3	.bmp	2013	To facilitate the development of computer-aided diagnosis systems in the segmentation and classification of melanoma.
MED-NODE	170	Macroscopic	2	.jpg	2015	To build and evaluate the MED-NODE system for detecting skin cancer with dermoscopic images.
HAM10000	10,015	Dermoscopic	8	.jpg	2018	To address the small size and insufficient diversity of images in the skin-disease dataset.
Derm7pt	2,000	Dermoscopic Structured data	15	.jpg	2018	As a database for the analysis of a seven-point malignant checklist for skin lesions.
BCN20000	19,424	Dermoscopic	9	.jpg	2019	Used to analyze skin cancer lesions in hard-to-diagnose locations such as nails and mucous membranes.
ISIC Archive	>13,000	Dermoscopic	9	.jpg, DICOM	2016–2020	To reduce skin cancer mortality while promoting the development and use of digital skin imaging.

result with a balanced accuracy of 74.2% on the BCN20000 dataset.

2.2.6 ISIC Dataset

To reduce skin cancer mortality while promoting the development and use of digital skin imaging (69), the International Skin Imaging Collaboration (ISIC) has established a publicly available skin disease dataset[6] for the computer science community around the world. Currently, ISIC Archive comprises over 13,000 representative dermoscopic images from clinical facilities throughout the world, all of which have been inspected and annotated by experts to ensure image quality (70).

The majority of studies that utilized the ISIC dataset focused on skin cancer classification and segmentation tasks, with the binary classification task being the most popular. For example (71), designed different modules based on VGGNet for skin disease classification (melanoma or benign) and benchmarked for the ISIC-2016 dataset. In the end, results showed that this method obtained excellent performance with an accuracy of 0.8133 and a sensitivity of 0.7866 (51). achieved the best classification results with an AUC of 0.911 and balanced multiclass accuracy of 0.831 on three skin cancer classification tasks of ISIC-2017 by using an ensemble of ResNet-50 networks on normalized images (72). used ensemble learning with a stacking scheme and obtained the classification results with an accuracy of 0.885 and an AUC of 0.983 in the ISIC-2018 competition (73). employed two bias removal techniques, "Learning Not to Learn" (LNTL) and "Turning a Blind Eye" (TABE), to alleviate irregularities in model predictions and spurious changes in melanoma images. Among them, the LNTL method combined a new regularization loss with a gradient inversion layer to enable the model to debias the CNN's features in backpropagation. The TABE method reduced biases by using different auxiliary classifiers to identify biases in features. Finally, the experimental results revealed that TABE had a more effective denoising effect, with an improvement of 11.6% in the AUC score benchmark on the ISIC dataset. Since the ISIC dataset is widely used in competitions and research, readers can find more methods for comparison on the competition leaderboard or on the Internet.

Table 2 summarizes the above datasets to show the different characteristics between them. What we summarized are the most common datasets in the skin cancer classification task and may not be the most exhaustive summary. Readers can find more datasets from various sources online. At the same time, it can be seen from the above summary that most of the images in the skin-disease dataset are dermoscopic images, while clinical images and histopathological images are still relatively rare. Furthermore, most skin-disease datasets have a relatively small number of images compared with datasets of natural images, which poses certain challenges for skin cancer classification tasks.

3 METHODS FOR TYPICAL AND FRONTIER PROBLEMS IN SKIN CANCER CLASSIFICATION

In the past few years, many scholars have been working on developing computer-aided diagnostic (CAD) systems for skin cancer classification. Before the emergence of deep learning, the CAD systems were primarily designed by machine learning (ML) algorithms (74). However, due to the complexity of feature engineering and limitations of handcrafted features, these ML-based methods can only diagnose a subset of skin diseases. Deep learning algorithms, on the other hand, can automatically learn semantic features from large-scale datasets with higher accuracy and efficiency. As a result, deep learning-based methods such as Convolutional Neural Network (CNN) have been used to solve the great majority of skin cancer classification problems in recent years and obtained satisfactory results.

However, as we dig deeper into the challenges of skin cancer classification, it appears that they are not as straightforward as the challenges in the non-medical domain (e.g., ImageNet, PASCAL-VOC, MS-COCO) (75) (12). Firstly, many datasets of skin images are imbalanced due to the disproportions among different skin cancer classes, which increases the risk of misdiagnosis by the diagnostic system. Also, since correct annotation needs a great amount of expertise knowledge and is time-consuming and labor-intensive, many datasets only provide

[6] https://www.isic-archive.com/.

a limited number of images (e.g., the ISIC dataset is the largest publicly available skin disease dataset until now, which contains about 13,000 skin images). As a result, more labeled data is required to design a more accurate system. Besides, when the amount of training data is insufficient, the model's generalization performance degrades. In addition, different noises generated by different devices or different shooting conditions also bring biases to the model, resulting in a reduction in diagnosis. Furthermore, the operational efficiency and resource consumption of the model also limit its clinical implementation on various medical devices.

As a result, in the following part, we present a complete overview of the use of deep learning methods in skin cancer classification. We begin by introducing the use of typical CNN frameworks in skin cancer classification, then review the frontier challenges in skin cancer classification and provide related solutions. We summarize these methods in **Tables 3–6**. Among them, **Table 3** summarizes the use of typical frameworks in skin cancer classification, as well as their highlights and limitations. **Tables 4–6** summarize the approaches to address the frontier issues of data imbalance and limitation, model generalization ability and robustness, and model computational efficiency in skin cancer classification. At the same time, we list publications based on the same or similar dataset together to make it easier for readers to compare different approaches.

3.1 Typical CNN Frameworks for Skin Cancer Classification

During the early stages of the development of CNN, people usually used self-building networks for a specific task. For example (76), presented a self-supervised model for melanoma detection. Firstly, a deep belief network and self-advised SVM were used to train the labeled and unlabeled images. After that, a

TABLE 3 | References of skin cancer classification with typical CNN frameworks.

Ref.	Dataset	CNN Architecture	Highlights	Limitations	Performance
(76)	Self-collected dataset	Deep Belief Network, SVM	By combining deep belief networks and SVM classifiers to handle skin cancer diagnosis tasks with limited datasets, as well as outliers and erroneous data.	The generalization ability of the model is limited.	Accuracy: 0.89
(77)	Self-collected dataset	Resnet-34, ResNet-50 ResNet-101 and ResNet-152	Proposed how to improve deep learning-based dermoscopy classification and dataset creation.	Data from more modalities, such as the patient's medical history, information on other symptoms, are not considered.	Accuracy: 0.85
(78)	Online repositories and the Stanford University Medical Center	Inception-v3	Used a CNN framework to train a large-scale skin disease dataset and achieve superior results on par with dermatologists. The method was also developed for mobile devices.	More research is required to assess its performance in clinical practice. At the same time, this method is limited to some extent by the amount of data.	Accuracy: 0.6375 (avg.)
(79)	MED-NODE	Deep CNN	Compared with previous methods, it directly used CNN to automatically extract features for skin disease images, also had a higher classification accuracy.	Due to the large noise interference of clinical images, there are still some misclassifications.	Accuracy: 0.81 PPV: 0.75, NPV: 0.86
(71)	ISIC-2016	VGG-16	It reduces the training time of the model by using the transfer learning strategy while obtaining higher sensitivity and precision.	It is prone to overfitting due to the limited amount of training images.	Accuracy: 0.813 Sensitivity: 0.787
(80)	ISIC-2017, IAD	Inception-v2	Introducing sonification into the diagnosis of skin cancer lesions to improve the sensitivity of the model.	Differences in the diagnosis of pathologists can affect the prediction results of the model.	AUC: 0.976 Sensitivity: 0.86 Specificity: 0.91
(27)	ISIC-2017	DenseNet, Dual Path Nets Inception-v4, Inception-ResNet-v2MobileNetV2, PNASNet, ResNet SENet, Xception	By analyzing 13 factors from 9 different models, they systematically evaluated the factors influencing the choice of CNN structure.	The dataset used in this article is too limited, and it only focuses on the melanoma classification task.	Top accuracy: 0.827
(81)	IAD	VGG-19	Adopted VGG-19 network to evaluate the thickness of melanoma for the first time.	There are no more pre-training methods utilized for comparison, and precisely predicting melanoma thickness would be more clinically significant.	Accuracy: 0.872 Specificity: 0.840
(82)	Derm7pt	Inception-v3	A multi-task network was designed to classify the seven-point checklist and skin disease diagnosis. Different loss functions were also designed to handle different input modalities, such as clinical and dermoscopic images, and patient diagnostic results.	Some criteria of the 7-point checklist are unable to be distinguished.	Accuracy: 0.737
(60)	HAM10000	Deep CNN models	Proposed a method combining CNN with one-versus-all (OVA) for skin disease classification.	The model has not been tested on datasets from various domains and may have a large variance.	Accuracy: 0.929
(83)	HAM10000 ISIC-2019	ResNeXt, SeResNeXt, DenseNet Xception, and ResNet	Adopted a grid search strategy to find the best ensemble learning methods for skin cancer classification.	The amount of training data is still insufficient, and most of models employed in ensemble learning are from the same network architecture.	Accuracy: 0.88 F1 score: 0.89

TABLE 4 | Different methods for solving data imbalance and data limitation.

Ref.	Dataset	Highlights	Limitations	Performance
(84)	ISIC-2017	By coupling seven GANs to generate seven skin-disease images. At the same time, they improved the efficiency of the model by making the initial layers of GANs share the same parameters.	The model was unable to distinguish the lesion area well when it was mixed with the skin surface, and artifacts such as human hair can also affect the generation of new images.	Accuracy: 0.816 AUC: 0.88
(85)	ISIC-2018	Proposed a GAN architecture that was customized to the style of skin lesions. At the same time, it can generate higher resolution and more diverse skin disease images by adjusting the progressive growth structure of the generator and discriminator in the GAN network.	The content of the GAN-generated synthetic dataset was not complicated enough when compared with the original dataset, and it was also not diverse enough.	Accuracy: 0.952 Sensitivity: 0.832 Specificity: 0.743
(86)	ISIC-2018	Utilized conditional generative adversarial networks (CGAN) to extract key information from all layers to generate skin lesion images with different textures and shapes while ensuring the stability of training.	The amount of data used for training was relatively limited.	Accuracy: 0.941 Precision: 0.915 Recall: 0.799
(87)	ISIC Archive	Explored four types of data augmentation methods and a multiple layers augmentation method in melanoma classification.	The data augmentation methods evaluated in this paper were limited and not validated on a large amount of datasets.	Accuracy: 0.829
(88)	HAM10000	They adopted a variational autoencoder network to get domain-dependent noise vectors. Also, a student-like distribution was employed to increase image diversity, and an auxiliary classifier was used to create images of certain classes.	Due to the specificity of medical images, different image generation models may generate skin disease images that did not belong to the same class.	Accuracy: 0.925
(89)	HAM10000	It combined the attention mechanism with PGGAN to obtain global features of skin lesions images, also introduced the Two-Timescale Update Rule to generate features with high fine-grainedness, while increasing the stability of GAN.	Due to the limitation of hardware conditions, this data augmentation method was only evaluated on the resolution of 256 × 256, rather than the original resolution of 600 × 450 in HAM10000 dataset.	AUC: 0.793
(90)	HAM10000	Proposed a class-weighted loss function and a focal loss to overcome the problem of data imbalance.	There is no artifact removal for the images in the training dataset, which leads the model to be biased. Also, it has a relatively high computational complexity.	Accuracy: 0.93 Recall: 0.86
(91)	HAM10000 ISIC-2019	A novel loss function was combined with the balanced mini-batch logic of the data level to alleviate the imbalance problem of the dermatology dataset.	The classification accuracy for rare skin diseases with limited data needs to be improved further.	Accuracy: 0.8997 Recall: 0.8613
(92)	HAM10000	Proposed a two-stage technique for determining the appropriate augmentation procedure for mobile devices.	Given the particularity of lightweight CNN, more data augmentation methods and data need to be considered to alleviate the problem of overfitting.	Accuracy: 0.853
(93)	PAD-UFES	Designed two algorithms based on evolutionary algorithm and also applied weighted loss function and oversampling to alleviate the problem of data imbalance.	A larger dataset was necessary to improve the performance further.	Accuracy: 0.92 Recall: 0.94
(94)	PH^2	Proposed novel a data augmentation method based on a oversampling technique (SMOTE).	The proposed data augmentation method was not validated in the deep learning architectures, and experiments on larger datasets were also required.	Accuracy: 0.922 Sensitivity: 0.808 Specificity: 0.951

bootstrap approach was used to randomly choose the training images for the network to improve the generalization ability and decrease the redundancy of the model. Experiments showed that the proposed method outperformed other methods like KNN and SVM. Then (79), designed a simple CNN network work for detecting melanoma. Firstly, all input images were preprocessed to eliminate the effects of noise and artifacts. The processed images were then fed into a pretrained CNN to detect if they were melanoma or benign. Finally, experiment results showed that CNN outperformed other classification methods.

With the development of deep learning, various well-known networks, such as VGGNet (117), GoogleNet (118), and ResNet (119), have been applied to skin cancer classification with favorable results. The most landmark work was (78). It was the first time that a CNN has been utilized to train large-scale clinical images for skin cancer classification. They designed an end-to-end network for automated skin cancer classification using Inception v3. A total of 129,450 clinical images with 2,032 distinct skin diseases were utilized for training the model. Meanwhile, to make use of the fine-grained information in taxonomy structure, they proposed a disease partitioning algorithm to divide skin cancers into fine-grained classes (e.g., melanoma was subdivided into amelanotic melanoma and acrolentiginous melanoma). In the end, the results of the experiments indicated that the skin cancer classification system could attain diagnostic levels equivalent to dermatologists. In the same year (71), successfully implemented VGGNet for skin lesion classification (melanoma or benign) and benchmarked

TABLE 5 | Different methods for improving model generalization ability and robustness.

Ref.	Dataset	Highlights	Limitations	Performance
(95)	DermIS DermQuest	Investigated the advantages of large-scale supervised pre-training for medical imaging applications.	In addition to the analysis of the weights and features of the model, it is necessary to conduct a comprehensive analysis of other features such as network structure to explore the importance of pre-training.	Accuracy: 0.871 (DermIS) Accuracy: 0.974 (DermQuest)
(96)	HAM10000 MoleMap	Proposed transfer learning and adversarial learning in skin disease classification to improve the generalization ability of models to new samples and reduce cross-domain shift.	When the data domain and target domain are significantly different, the method's overall accuracy suffers.	Accuracy: 0.909 AUC: 0.967
(97)	HAM10000	Performed adversarial training on MobileNet and VGG-16 using the innovative attacking models FGSM and PGD for skin cancer classification.	The number of datasets tested for this experiment is very limited, and there may be local optimizations.	Accuracy: 0.7614
(98)	ISIC-2016	Proposed a comprehensive deep learning framework combining adversarial training and transfer learning for melanoma classification. At the same time, focal loss was introduced to iteratively optimize the network to better learn hard samples.	This method does not consider more types of skin diseases, and it had a high computational cost.	Accuracy: 0.812 Sensitivity: 0.918
(99)	ISIC2017 HAM10000	Presented a Multi-view Filtered Transfer Learning approach to extract useful information from the original samples for domain adaption, thereby improving representation ability for skin disease image.	The effectiveness of this domain adaptation method should be validated on more dermatology datasets.	Accuracy: 0.918 AUC: 0.879
(100)	ISBI-2017, PH^2	Proposed an adversarial training method combined with attention module to enhance the robustness of the model in skin-disease classification and segmentation.	Due to the limited amount of training data and the unclear boundaries of skin disease images, the model still suffers from under-segmentation and over-segmentation.	Accuracy: 0.968 Sensitivity: 0.962 Specificity: 0.941
(101)	ISIC-2018	Using seven universal adversarial perturbations to investigate the vulnerability of the classification model.	This method does not perform adversarial training on more skin disease datasets, so the robustness of its model needs to be further improved.	Accuracy: 0.873
(102)	ISIC-2019	Proposed Monte Carlo dropout, Ensemble MC dropout, and Deep Ensemble for uncertainty quantification.	Further optimization of the robustness of the model is required, and the model should also be tested for noise detection to provide a confidence score.	Accuracy: 0.90 AUC: 0.945
(103)	ISIC Archive MED-NODE Dermofit	Proposed a transfer learning method to address the shortage of data in skin lesion images. Also, they utilized a hybrid deep CNN model to accurately extract features and ensure training stability while avoiding overfitting.	The model requires a considerable amount of computational resources while also lacking the diversity of domains.	Accuracy: 0.853 F1 score: 0.891
(104)	HAM10000, Dermofit, Derm7pt, MSK PH^2, SONIC, UDA	Proposed to improve the generalization performance of the model by combining data augmentation and domain alignment. Designed a Bayesian generative model for continual learning based on a fixed pretrained feature extractor.	Due to the privacy of medical images, this trained model may underperform on ethnic groups with a small proportion of the population.	Accuracy: 0.670
(105)	Skin7, Skin40		To increase the method's overall performance, better pre-training of the extractor can be investigated.	Mean class recall: 0.65

for ISIC datasets by using dermoscopic images. In this study, they designed three different modules based on VGG-16 as comparison. The first module trained the network from initial weights. The second module used pretrained VGG-16 for training and then used the current dataset to train the fully connected classifier. The third module also used transfer learning to train the network, but weights in the high-level part of the convolutional layers were initialized from the first module. In the end, results showed that the third module obtained excellent performance in skin cancer classification. Different from previous classification tasks (81), utilized a CNN framework (VGG-19) for the first time to evaluate the thickness of melanoma. They began by locating the lesion and cropping the region of interest (ROI). To solve the problem of data limitation and data imbalance; they then employed the Synthetic Minority Over-sampling technique to generate synthetic samples. After that, the pretrained VGG-19 was used for the thickness prediction. Finally, the results demonstrated that the algorithm can estimate the thickness of melanoma with an accuracy of 87.5%. For the first time, a multitask network was proposed by (82) based on Inception v3 by utilizing three different modalities of data to predict seven-point criteria. In addition, they designed a multimodal–multitask loss function to tackle the combinations of input modalities, which was also able to make predictions with incomplete information. Finally, results showed the superior performance in classifying skin lesions and the seven-point criteria. Besides, the proposed method had the ability to identify discriminating information and generate feature vectors for image retrieval (80). built two systems for skin disease classification based on the novel deep learning algorithms. Additionally, they added a sonification-derived layer to increase the sensitivity of the model. In the first

TABLE 6 | Different methods for improving model efficiency.

Ref.	Dataset	Highlights	Limitations	Performance
(106)	Self-collected	Proposed a knowledge distillation method to transfer knowledge between various models simultaneously.	The proposed method sacrifices local accuracy for higher global accuracy, with some additional classification errors on local objects.	Accuracy: 0.75
(107)	Public repositories	Proposed a MobileNet-based classification method and successfully deployed it on an Android application.	To improve the model's classification accuracy, more sophisticated sampling strategies and data preprocessing can be adopted.	Accuracy: 0.944
(108)	HAM10000	Presented an assessment of the effectiveness for the attention module and self-attention module in skin cancer classification based on ResNet architecture.	Only limited number of attention mechanisms are used for comparison.	Accuracy: 0.622 (attention) Accuracy: 0.737 (self-attention)
(109)	HAM10000 SH-11	Proposed a weight pruning strategy for lightweight neural networks to make up for the accuracy loss and improve model performance and reliability in the skin cancer classification.	The proposed pruning method is only validated on the skin disease dataset, and more kinds of medical images are needed to validate its effectiveness.	Accuracy: 0.975 AUC: 0.931
(110)	HAM10000, PH^2, Dermofit Derm7pt, MSK, UDA	Designed a new pruning method "MergePrune" to reduce the computational cost of retraining the network by combining pruning and training into a single stage.	To assess this strategy, more domain data is needed, such as clinical images, patient meta-data.	Accuracy: 0.776 (avg.)
(111)	ISIC-2017	Proposed a classification method that incorporated the attention residual learning (ARL) mechanism to EfficientNet for skin cancer diagnoses.	The interpretability of the model needs to be further strengthened.	Accuracy: 0.873 AUC: 0.867
(112)	ISIC-2017	Three different lightweight networks MobileNet, MobileNetV2, and NASNetMobile were were evaluated for skin cancer classification.	The number of lightweight networks and hyperparameters used for testing are relatively restricted.	Accuracy: 0.82 Precision: 0.812
(113)	ISIC-2017, PH^2	Proposed an MT-TransUNet network to segment and classify skin lesions simultaneously.	The model finds it difficult with low-contrast skin disease images, and its segmentation performance is vulnerable to occlusions in the skin image.	Accuracy: 0.912
(114)	PH^2, DermQuest	Built a pruning framework to simplify the complicated architectures by choosing the most informative color channels in skin lesion detection. Also, it carried out a hardware-level analysis of the complexity of different skin cancer classification networks.	The proposed method works well for simple networks, but it may not perform as well for more complicated networks.	Accuracy: 0.9811 (PH^2) Accuracy: 0.9892 (DermQuest)
(115)	SD-198, SD-260	Proposed a knowledge distillation method based on curriculum training to distinguish herpes zoster from other skin diseases.	It requires manual tuning of hyperparameters according to different models and datasets.	Accuracy: 0.935
(116)	DermIS, DermQuest, DermNZ, "11K Hands"	Proposed an expert system "i-Rash" based on SqueezeNet to classify four skin diseases.	More clinical data and skin-disease images are needed to further improve the generalization of the model.	Accuracy: 0.972 Sensitivity: 0.944 Specificity: 0.981

system, a CNN architecture was proposed based on Inception v2 to identify skin diseases (benign or malignant) with dermoscopic images. The second system transformed the feature representation generated in the preceding system into sound data. Then, this sound information was then put into a machine learning classifier or translated to spectrograms for further analysis. In the end, both systems performed exceptionally well in terms of classification and sonification. After the deep learning methods achieved excellent results in the skin cancer classification task (77), proposed how to improve deep learning-based dermoscopy classification and dataset creation. They analyzed four ResNet architectures in dermoscopic classification, namely, ResNet-34, ResNet-50, ResNet-101, and ResNet-152, to apprehend the mechanisms and certain error causes. First, four ResNet networks were trained at their best fits to see if the structural differences between the models would result in different classification results. After testing with several epochs, they found that the accuracy of different models tended to be consistent and varied with different hyperparameter settings. Meanwhile, they had a high level of stability during training. Therefore, the training errors of the classification models were attributed to incorrect annotations and the complexity of medical images.

Gradually, people discovered that applying a single CNN to a CAD system typically did not produce the desired results due to the large variances in deep neural networks. After that, ensemble learning was proposed as a way to limit the error generated by a single model by training multiple models and then combining

their results to get the final classification results (27). compared the performance between ensemble models and a single model by utilizing nine different CNN architectures in skin cancer classification. After different comparative experiments, they found the significance of ensemble learning for obtaining optimal classification models. In addition, they investigated the effectiveness between two different selection strategies in ensemble learning: random selection and utilizing a validation set. For the smaller ensemble models, they found that the second method had more advantages, but the first was also effective. For the larger ensemble models, it was possible to get away with merely picking models arbitrarily. Based on the same method (60), proposed two different methods for skin cancer classification while reducing the complexity of the model by using an OVA strategy: i) alone CNN model and ii) the incorporation of seven CNN models. In the first method, images from the dataset were directly put into the single CNN model for the final prediction. In the second method, a one-versus-all (OVA) strategy was used to combine seven separate models with two classes to obtain the final prediction. Each class in this method was classified according to true and false labels, thus increasing the efficiency of the model. The results revealed that the second method outperformed the first in terms of classification accuracy (83). adopted a grid search strategy to find the best ensemble learning methods for the classification of seven skin lesions. During the training, five CNN networks, ResNeXt, SeResNeXt, ResNet, Xception, and DenseNet, were used as baseline. After that, two ensemble learning strategies, namely, average ensemble and weighted ensemble, were conducted to find the optimal model. In the end, results showed that the weighted ensemble model had more advantages than the average ensemble model.

3.2 Data Imbalance and Data Limitation in Skin Disease Datasets

Data imbalance and data limitation in skin disease datasets are common problems in the skin cancer classification tasks. In fact, benign lesions account for the majority of data in many skin disease datasets. Meanwhile, many skin disease datasets have large inequities in the number of samples among different skin disease classes. Only the common skin diseases, such as BCC, SCC, and melanoma, are included in the majority of skin disease datasets. Other skin cancer diseases (such as appendiceal carcinomas and cutaneous lymphoma) are relatively rare in these datasets, making it difficult for algorithms to classify them correctly (28). Besides, the skin lesions in most of the current datasets are from fair-skinned people, with only a few from dark-skinned people (12). It has been demonstrated that deep learning frameworks that have been validated for skin cancer diagnosis in fair-skinned people are more likely to misdiagnose those with different races or ethnicity (120). At the same time, the quantity of skin disease images is also relatively restricted. For example, ISIC-2020 (121) is the dataset with the largest number images so far, with about 30,000 skin disease images. Although large amounts of skin disease images can be obtained from websites or medical institutions without any diagnosis information, labeling them takes professional knowledge and can be extremely challenging and time-consuming. What is more, sufficient labeled data are a requirement for training a reliable model. When only a limited number of images are provided, overfitting is more likely to occur. As a result, for the skin cancer classification task, a considerable amount of labeled data is required.

Generative adversarial networks (GAN) are widely thought to be a preferable alternative, as they can generate artificial data to compensate for data imbalance in terms of positive and negative proportions, rare cases, and different people (84). designed a data augmentation method based on generative adversarial networks to address the shortcomings of skin lesion images in melanoma detection. Firstly, they utilized several data processing methods to locate and eliminate hairs and other artifacts of the input images. Then they used two convolutional GANs, namely, DCGANs, to generate 350 images of melanoma and 750 images of seborrheic keratosis. Finally, the results demonstrated that combining the processing module and generative adversarial networks resulted in superior performance when compared with other baselines. Although GAN is extensively employed for data augmentation, the images it generated are typically low-resolution. To overcome this issue (85), proposed a style-based GAN to generate more high-quality images in skin lesion classification. Then these synthetic images were added to the training set to the pretrained ResNet-50 model. The experiment showed that the proposed style-based GAN method outperformed other GAN-based methods in terms of Inception Score (IS), Fréchet Inception Distance (FID), precision, and recall. What is more, the accuracy, sensitivity, specificity, and other indicators of the classification model also improved. In (88), the author proposed a GAN-based framework "TED-GAN" to generate skin lesion images artificially in skin cancer classification. Instead of using random Gaussian distribution to sample the noise vector in GAN, they used informative noise that was obtained from a separate network for the first time to generate the medical images. TED-GAN had four parts: one variational auto-encoder, two GANs, and one auxiliary classifier. Firstly, an auto-encoder network was trained to get the vector containing the image manifold's information. Then one of the GANs sampled output of the auto-encoder to ensure the stability of training and make it more convenient to use the domain information. After that, the other GAN obtained more training data from the prior GAN. In addition, an auxiliary classifier was added to this GAN network, then the two were trained together to generate images of various skin diseases. In the end, experiment results showed that TED-GAN had a positive effect on skin cancer classification as it provided more images for training. Although data augmentation methods such as GAN may successfully increase the number of skin cancer images and alleviate the problem of data imbalance, the generated data usually have identical distributions, limiting the improvement in model performance. To solve this issue (89), proposed a data augmentation method based on PGGAN, namely, SPGGAN, to generate skin lesion images with different types and data distributions. Firstly, an attention module was

added into SPGGANs to obtain the global and local information from skin lesion images, also enabling PGGAN to generate more diverse high-quality samples. Then, the Two-Timescale Update Rule (TTUR) was added to SPGGANs to reduce the signal magnitude increase and hence enhance the stability of the model. Finally, experiments showed that the GAN-based data augmentation approach can lead to an improvement in the classification in terms of accuracy, sensitivity, F1 score, and other metrics. Since skin lesions often contain irregular boundaries, varied textures, and shapes, it makes the training of the GAN framework sometimes unstable. To address this issue (86), utilized conditional generative adversarial networks (CGANs) to extract key information from all layers and generate skin lesion images. The proposed CGAN has two modules: a generator module and a discriminator module. The generator module was to extract useful features from high-level and low-level layers and generate synthetic images. The discriminator module was to accurately map latent feature components by combining auxiliary information with training images. After that, augmented images with original datasets were put into the pretrained ResNet-18 network for the classification task. Experiments showed that this model achieved superior results compared with other datasets.

Another popular method for resolving data imbalance is to apply weights to various samples in the loss function. The goal is to calculate the losses differently depending on whether the samples are in the majority or minority. For example (90), proposed an end-to-end framework for classifying seven skin lesions in the HAM10000 dataset. Especially, a class-weighted learning strategy was utilized to overcome the problem of data imbalance in the dataset by assigning different weights to different lesion classes in computing the loss function. Meanwhile, focus loss was used to further increase the model's classification performance. It concentrated training on tough examples, preventing the classifier from being overwhelmed by easy samples. Experiment results revealed that the model obtained an average accuracy of 93%, outperforming dermatologists' 84% accuracy. Although the problem of data imbalance can be alleviated through the design of the loss function, there exists a problem of slow learning of the minority classes. To solve the issue (91), proposed a hybrid strategy for skin cancer classification. It combined a loss function method at the algorithm level with a balanced mini-batch logic method for real-time image augmentation at the data level. By applying the balanced mini-batch and real-time image augmentation method, the new loss function can improve its learning ability in minority samples, thereby improving training efficiency. When compared with the previous strategy, this method improved the learning effectiveness of minority classes on an imbalanced dataset by increasing m-Recall by 4.65% and decreasing the standard deviation of recalls by 4.24%. In addition to designing a new loss function (93), also designed two new algorithms based on evolutionary algorithms, the Mixup Extrapolation Balancing (MUPEB) and the Differential Evolution (DE), to solve the problem of data imbalance in melanoma classification. The MUPEB method included a set of operations to mix and interpolate the dataset until it was balanced. The DE method mixed and combined three random images with varied clinical information to achieve data balance. Apart from that, weighted loss function and oversampling were also used to alleviate data imbalance. In the end, this algorithm increased the model's classification precision and recall by 1% and 11%, respectively.

Data augmentation is an ideal solution to artificially increase the amount of data by generating new data points from existing data. It scales the number of images by random rotating, padding, rescaling, flipping, translation, etc. At the same, with the development of technology, various novel approaches for data augmentation have been presented in skin cancer classification (58, 122). released the HAM10000 dataset by natural data augmentation; the images of skin lesions were captured at various magnifications or angles, or with multiple cameras. To evaluate the effectiveness of data augmentation methods while determining the most effective method (87), explored four types of data augmentation methods (geometric transformation, adding noise, color transformation, and image mix) and a multiple-layer augmentation method (augmented images by more than one operation) in melanoma classification. The first step was to preprocess the images to remove artifacts such as body hair on the images. Then each augmentation method was assessed to decide the optimal augmentation method. In the end, they found that single-layer augmentation outperformed multiple-layer augmentation methods. Besides, the region of interest (ROI)-mix method achieved the best performance compared with other approaches (92). proposed a two-stage strategy data augmentation method on mobile devices successfully with limited computing resources. The first stage was to search the optimal augmentation method in the Low-Cost-Augment (LCA) space. The second stage was to fine-tune the deep CNNs with augmented images and choose the model with the highest accuracy. Finally, the augmented images were trained with EfficientNets, which resulted in better accuracy and computational efficiency. Different from previous data augmentation methods (94), proposed a novel Synthetic Minority Oversampling Technique (SMOTE) to solve the problem of image scarcity and imbalance in the skin lesion dataset. Firstly, all images in the PH^2 dataset were preprocessed for ensuring cleaning. Then in the data augmentation stage, the covariance matrix (CM) was exploited by SMOTE to find dependent connections between attributes. Then they built surrogate instances based on the estimated CM to balance the number of minority class and majority class. Finally, all augmented images were utilized to train the SqueezeNet and it resulted in a significant improvement in terms of accuracy, sensitivity, specificity, and F1 score.

3.3 Poor Generalization Ability Across Different Domains

In the skin cancer classification task, the generalization ability of the model is often inferior to that of an experienced dermatologist. Firstly, owing to the small scale of skin image datasets, even if a large amount of similar data is artificially

generated, the overfitting problem still exists. Secondly, the majority of research exclusively focuses on dermatological images taken using standardized medical equipment, such as dermoscopic and histological images (78). Little research has been conducted on dermatological images captured by other devices. When a trained model is applied to a new dataset with a different domain, its performance suffers significantly.

Transfer learning (TL) is commonly utilized for improving the generalization ability of computer-aided diagnostic systems in test data. The fundamental idea of TL is to preserve information gained while addressing a problem and implement it to a new but relevant problem (52). It can not only drastically reduce the time overhead and labor cost associated with partial repetitive labor but also compensate for the flaw in the skin disease datasets (96). presented two methods to improve the generalization ability of models to new samples and reduce cross-domain shift. The first method used a transfer learning strategy with two steps to acquire new knowledge from diverse domains. It began with pretraining on ImageNet and fine-tuned the model with a single skin dataset. In the end, they used the target set to fine-tune the model to get the prior information. The second method used a pixel-wise image synthesizing adaptation method to transfer the features between the source domain and target domain. In comparison to the previous transfer learning approach, this method was semi-supervised and did not need any labels for domain adaptation. Finally, cross-domain experiments showed that in order to improve classification performance, the proposed methods had the ability to transform images between different modalities. In order to solve the problem of class imbalance in skin lesion datasets, To address the problem of poor generalization performance due to low interclass variance and class imbalance in skin disease images (98), proposed a two-stage framework with adversarial training and transfer learning in melanoma detection. The first stage was to solve the data scarcity and class imbalance problem by generating underrepresented class samples. The second stage was to train deep neural networks for melanoma classification, by using newly synthesized images and original datasets. A focal loss was proposed to assist the model in learning from hard examples. In the end, results showed the significant improvement of the classification performance and superiority of the proposed method. With the application of transfer learning in skin cancer diagnosis, it has been discovered that most existing transfer learning methods only extract knowledge from the source data to learn, but many inaccurate samples that are very different from the target data are incorporated into the process. Meanwhile, most skin cancer classification methods simply learn from raw skin disease images, which makes information from different aspects (such as texture, shape, etc.) interfered by noise during the learning process. Therefore (99), proposed a multi-view-filtered transfer learning (MFTL) method to solve the poor scalability problem of skin cancer classification models. MFTL consisted primarily of two modules: a multi-view-weighing representation module and a filtered domain adaption module. The first module put the view weights obtained from the feature representation procedure to the final prediction. The second module selected key source information to transfer the knowledge between the source domain and target domain. Finally, the result showed a significantly improved performance in classifying melanoma and seborrheic keratosis. In (103), the author proposed a transfer learning approach to address the issue of insufficient data in the medical image datasets, as well as to improve the performance of other related medical image classification tasks. The proposed approach first trained deep learning models on a great amount of unlabeled images for a specific task, as the volume of unlabeled medical images has increased significantly. Then the models were fine-tuned on a relatively small-labeled dataset to perform the same task. Besides, they utilized a hybrid deep CNN model to accurately extract features and ensure training stability while avoiding overfitting. Experiments showed the effectiveness in the skin cancer and breast cancer classification in terms of classification accuracy, recall, precision, and F1 score. With the growing use of transfer learning in the field of computer vision, an increasing number of studies have proved that large-scale pretraining on natural images can be beneficial in a variety of tasks. However, research on medical images is still limited. With this purpose (95), investigated the advantages of large-scale supervised pretraining with three medical images: chest radiography, mammography, and dermatological images. Five tasks including in-domain performance, generalization under distribution shift, data efficiency, subgroup fairness, and uncertainty estimation were conducted to test if large-scale pretraining aided in the modeling of medical images. Finally, experiment results indicated that, despite significant differences from the pretraining data, employing larger pretraining datasets can achieve significant improvements across a wide range of medical disciplines. Besides, they discovered that pretraining at scale may allow downstream tasks to more effectively reuse deeper features.

In addition to TL, many novel methods such as adding innovative regularization terms, estimating model uncertainty, and lifelong learning models are beginning to be introduced into the skin cancer classification task to improve the generalization ability of the model across different domains (104). proposed a method that can improve the generalization ability of a model under limited samples by combining data augmentation and domain alignment. They observed in medical images that domain changes were compact and related to a certain extent. To be able to model such dependencies, the author introduced a dependency regularization term to learn a representative feature space that captured sharable information across different medical image domains. At the same time, a variational encoder was used to ensure that the latent features followed a predetermined distribution. Finally, through theoretical derivation, the author obtained the upper bound of empirical risk for any relevant target domain under this method, which alleviated the problem of overfitting. Finally, the generalization ability of the model was well confirmed on seven skin-disease datasets. In order to obtain the uncertainty quantification (UQ) of the deep learning model

to prevent overfitting (102), proposed three indicators Monte Carlo (MC) dropout, Ensemble MC (EMC) dropout, and Deep Ensemble (DE) to solve this problem. They next presented a novel hybrid Bayesian deep learning model based on the three-way decision (TWD) theory to obtain the residual uncertainty after using the three methods of MC, EMC, and DE. It also enabled different UQ methods to be used in different neural networks or different classification stages. Finally, the experimental findings demonstrated that the proposed model can be employed efficiently in analyzing different stages of medical images, and the model's uncertainty was accurately quantified. Since the deep learning model might forget much of the previous information while learning new data, updating the system with more new data would reduce the performance of the previous learning, which poses a greater challenge to the medical autonomous diagnosis system. To this end (105), designed a Bayesian generative model for continual learning based on a fixed pretrained feature extractor. Different from the previous continual learning method, which stored a small number of images for each old class, the proposed method stored the statistical information of each class based on the previous feature extractor, which can make the model naturally keep the knowledge of each old class from being used. Therefore, there was no need to store or regenerate old images. Finally, the model performed well on both the Skin7 and Skin40 datasets, and it was able to retain some images from previous classes during continual learning. The model's scalability and generalization have been greatly enhanced.

3.4 Noises From Heterogeneous Devices and Images

Various noises obtained from heterogeneous sources and skin disease images pose challenges to the robustness of models in the task of skin cancer classification. When trained on high-quality skin lesion datasets, the deep learning model can reach the same diagnostic level as dermatologists, even surpassing them. However, since the skin cancer classification model is sensitive to images captured with different devices, lighting settings, and backgrounds, it frequently fails to obtain satisfactory classification results when tested with different images. Furthermore, photographic images (such as smartphone images) vary greatly in terms of zoom, perspective, and lighting, making classification much more difficult.

Therefore, many scholars have worked to integrate adversarial training into the field of skin cancer classification to enhance the robustness of the classification models. In (100), the author introduced a novel Attention-based DenseUnet (Att-DenseUnet) network combined with adversarial training for skin lesion segmentation and classification. With the addition of the attention module, the model can pay more attention to discriminative features while also successfully suppressing irrelevant features in the DenseBlocks output. In this way, the interference of artifacts on skin disease images is reduced. Att-DenseUnet had two main modules: Segmentor module and Discriminator module. The segmentor module was a U-Net shape structure, which contained a down-sampling path, up-sampling path, and related attention module to ensure the information transfer between different layers. Additionally, it adopted an attention module to focus on the essential features and speed up the training process. The discriminator module employed the adversarial training to impose the segmentor module to obtain diverse features with different sizes and shapes and direct the attention module to concentrate on the multiscale lesions. Besides, they used the adversarial loss to prevent overfitting by providing the regularization term for the networks. Finally, the results showed that this network achieved excellent performance and was robust enough for different skin image datasets. In clinical applications, it has been discovered that noises that are difficult for humans to detect frequently cause significant interference to the diagnostic model, limiting the utility of deep learning in the actual world. To improve the model's robustness (97), performed adversarial training on MobileNet and VGG-16 using the innovative attacking models FGSM and PGD for skin cancer classification. Firstly, two white-box attacks based on Projected Gradient Descent (PGD) and Fast Gradient Sign Method (FGSM) were used to test the robustness of these models. Then, to increase the robustness of these models, the author did the adversarial training based on PGD against white-box attacks. In the end, the results showed that the robustness of these models significantly improved. To further increase the difficulty of adversarial attacks instead of simple adversarial attacks (101), used the more realistic and riskier Universal Adversarial Perturbation (UAP) to adversarially train seven classification models (VGG-16, VGG-19, ResNet-50, Inception ResNet-V2, DenseNet-21, and DenseNet-169). During the adversarial attack phase, the author used an iterative algorithm to generate perturbations for non-targeted and targeted attacks and the Fast Gradient Sign Method (FGSM) was used to generate perturbations for input images. After that, they conducted adversarial retraining to improve the robustness of these seven models. The results showed that these models were easily deceived when applied to adversarial attacks. In addition, they found the limited effect of adversarial retraining on non-targeted perturbations. Although adversarial retraining considerably lowered the vulnerability to adversarial perturbations in targeted attacks, it did not totally avoid it.

3.5 Toward Faster and More Efficient Classification Models

Although an increasing number of deep learning algorithms have been successfully applied to skin cancer classification with excellent classification results, the computational complexity of the model still needs to be considered. Firstly, due to improvements in imaging technology, many skin disease images with high resolution have large pixels. For example, histological scans are made up of millions of pixels, and their resolution is often larger than 50,000 × 50,000 (123). As a result, training them takes longer time and additional computing resources. Secondly, the computational complexity in the deep learning model is increasing as its accuracy improves, which

demands their implementation to various medical equipment or mobile devices at a higher cost. Here we introduce three latest methods when designing an effective network for skin cancer classification.

Over the past few years, many Lightweight Convolutional Neural Networks have been designed and successfully applied in skin cancer classification to meet the demands of practical applications. Subsequently, many scholars used lightweight CNN for the task of skin cancer classification and successfully employed it to various mobile devices. For example (107), proposed an automated classification method based on MobileNet and successfully deployed it on an Android application or a website for public use. With the vigorous development of mobile health (mHealth), more and more mobile applications are designed for cancer classification and prediction. However, the application of automatic classification of skin cancer is still limited. To solve this problem (116), proposed an innovative expert system based on SqueezeNet, namely, "i-Rash," to classify four classes of skin diseases in real time. Due to the limited size of "i-Rash" (i.e., 3 MB), identifying an unknown image for the system only took 0.09 s. Inspired by predecessors (111), proposed a novel method that incorporated attention residual learning (ARL) mechanism to EfficientNet with fewer parameters. Besides, they also investigated how the mechanism related to the existing attention mechanisms: Squeeze and Excitation (SE). Through the comparison of experimental results between models with and without SE, they speculated that the attention module accounts for a large portion of EfficientNet's outstanding performance. What is more, the addition of ARL increased the accuracy of the EfficientNet and its variance. In (112), three different lightweight models (including MobileNet, MobileNetV2, NASNetMobile) were adopted for skin cancer classification. To find the model with the best performance, they tested a total of nine models with three different batch sizes. In the end, they found that the NASNetMobile model showed the best performance with a batch size of 16. Meanwhile, they benchmarked the lightweight models with fewer parameters and less computational time.

Pruning is an effective way to remove parameters from an existing network to maintain the accuracy of the network while increasing its efficiency (124). To enable CNN to be used in medical devices with limited power and resources (114), built a pruning framework to simplify the complicated architectures by choosing the most informative color channels in skin lesion detection. The proposed method is to achieve two purposes: removing redundant color channels and simplifying the whole network. Firstly, all color channels were put into the network. Then the weights that associated with the non-essential color channels were deleted to select the most informative color channel. After that, to generate a simplified network, they utilized CNN models as the target network and trained them on the chosen color channels. Besides, the requirements of these models were calculated from hardware perspectives to analyze the complexity of various networks. Finally, results showed that this color channel pruning strategy improved segmentation accuracy while also simplifying the network. Designing an efficient and generalizable deployment strategy is an extremely challenging problem for lightweight networks. To this end (109), proposed a weight pruning strategy for lightweight neural networks to make up for the accuracy loss and improve model performance and reliability in the skin cancer classification. Five lightweight CNNs, namely, SqueezeNet, MnasNet, MobileNetV2, ShuffleNetV2, and Xception, were investigated in this task. Firstly, a dense–sparse–dense (DSD) training strategy was used to avoid the underfitting and high bias of the networks. Then, a detailed analysis was used for building a pruning method including not just pruning connections with various relations but also reviewing a novel pruning mechanism that can remove the weights according to the distribution in each layer adaptively. In the end, the pruning strategy achieved higher accuracy and less computation compared with unpruned networks (110). designed a new pruning method "MergePrune" to reduce the computational cost of retraining the network by combining pruning and training into a single stage. Firstly, different units were assigned to learn each domain independently as they contribute differently to the classification result. Then, for one domain, determined culprit network units with high "culpability" scores were pruned and then reset and assigned to learn new domains. At the same time, non-culprit units were preserved. MergePrune was implemented to reduce the amount of computation and improve the efficiency of the classification model. Finally, the results showed that the network can perform accurately and effectively on real-world clinical imaging data with various domains, even with high pruning ratios.

Knowledge distillation is the process of distilling information from a huge model or group of models to a smaller model that could be successfully implemented with real-world restrictions (106, 125). proposed a knowledge distillation-based method that enabled to transfer knowledge between models simultaneously in skin cancer classification and brain tumor detection. Firstly, a pretrained ResNet-50 was chosen as a base model as its excellent performance out of the box. Then, with the significant degree of resemblance across the images in the medical image dataset, they let the knowledge transfer only between the two bottom-most layers. As a result, high-level visual comprehension was preserved, and information was added to the granular distinction in this way. The findings of the experiments were revealed in order to gather remote knowledge and enhance global accuracy; some local accuracy was lost. To improve the robustness and reduce the computational cost of the model (115), proposed a knowledge distillation method based on curriculum training in distinguishing herpes zoster from other skin diseases. Firstly, three kinds of model, namely, basic models, mobile models, and ensemble models, were chosen for benchmark. Then, to improve the performance of a single network, an ensemble knowledge distillation was utilized. This allowed the student network to learn more robust and representative features from the network while keeping a low

computational cost. After that, they proposed curriculum training for ensemble knowledge distillation in order to distill ensemble teachers more efficiently with an adaptive learning technique. In the end, the results showed that the proposed method achieved improved performance while obtaining higher efficiency.

Transformer (126) is a deep learning model designed by the Google team in 2017 that was originally utilized in Natural Language Processing (NLP) and is now frequently employed in medical image processing, such as skin lesion images. It uses the self-attention mechanism to weigh the relevance of different parts of the input data series, resulting in shorter training periods and improved accuracy (126, 127). The introduction of the attention mechanism has generated great interest in the research community, but there is still a lack of systematic ways to select hyperparameters that guarantee model improvement. To this end (108), presented an assessment of the effectiveness for the attention module and self-attention module in skin cancer classifications based on ResNet architecture. Among the two modules, the attention module was used to recompute the features of the input tensor in each layer. The self-attention module was used to connect multiple positions of input images to obtain different representations of the input. In the experiment stage, the author investigated and compared a variety of alternative attention mechanisms with images from the HAM10000 dataset. In the end, the results showed that many of the self-attention structures outperformed the ResNet-based architectures, while containing fewer parameters. At the same time, applying the attention mechanism reduced the image noise; however, it did not behave consistently across different structural parameters. In solving the skin cancer classification problem, people often treat it as a simple classification task, ignoring the potential benefits of lesion segmentation. To this end (113), proposed an approach that combined the attention module with the CNN module for skin cancer classification. The CNN module was in charge of getting lesion texture information, while the attention module was responsible for obtaining context information such as the shape and size of the lesion. In addition, dual-task and attended region consistency losses were adopted to mediate the classification and segmentation heads without pixel-level annotation, which increased the robustness of the model when it trained with various augmented images. Finally, MT-TransUNet achieved excellent performance in the skin lesion segmentation and classification. At the same time, it preserved compelling computational efficiency and speed.

4 CONCLUSION

With the development of science and technology, the diagnosis accuracy and efficiency for skin cancer classification are constantly improving. In the previous clinical diagnosis scenarios of skin cancer, the final diagnosis often depends on the imaging quality and the experience of dermatological experts, which is highly subjective and has a high rate of misdiagnosis. With the advent of machine learning, various CAD systems have been designed to aid the dermatologists to diagnose skin cancer diseases. In some skin cancer classification tasks, these CAD systems achieved excellent performance by utilizing handcrafted features. Recently, with the success of deep learning in medical image analysis, several researchers have applied deep learning methods for skin cancer classification in an end-to-end manner and achieved satisfactory results. It is expected that in the future, artificial intelligence and the diagnosis of skin cancer diseases would become closely associated.

In this study, we present a comprehensive overview of the most recent breakthroughs in deep learning algorithms for skin cancer classification. Firstly, we introduced three different types of dermatological images used in diagnosis and some commonly used datasets. Next, we present the applications of typical CNN-based methods in skin cancer classification. After that, we introduce several frontier problems in the skin cancer classification task, such as data imbalance and limitation, cross-domain adaptability, model robustness, and model efficiency, along with relevant deep learning-based approaches. Finally, we provide a summary of the entire review. We draw the key information as follows:

- Skin cancer develops as a result of uncontrolled cell proliferation in the skin. It frequently appears on sun-exposed skin. The three major types of skin cancers are basal cell carcinoma (BCC), squamous cell carcinoma (SCC), and melanoma. Early skin cancer classification increases the chances of a successful treatment (refer to Section 1 for more information).
- Clinical images, dermoscopic images, and histopathological images are three common types of images used for skin disease diagnosis. Among them, the most common forms of images are dermoscopy images. With the growing need for medical imaging resources in academia, more and more datasets are becoming publicly available. We list several popular datasets for skin-disease images along with works based on these datasets. However, compared with natural image datasets, the diversity and quantity of skin-disease datasets are still very limited, which also brings great challenges to the automatic diagnosis of skin cancer (refer to Section 2 for more information).
- When using CNN-based methods for skin cancer classification, VGGNet, GoogleNet, ResNet, and their variants are the most often used deep learning models. Also, ensemble learning was proposed to limit the error generated by only a single model and achieved satisfactory results. Although various deep learning models have performed admirably on skin cancer classification tasks, several challenges still exist and need to be resolved, such as imbalanced datasets, a lack of labeled data, cross-domain generalization ability, noisy data from heterogeneous devices and images, and how to design effective models for complicated classification tasks. To address the challenges, methods include generative adversarial networks, data augmentation, designing new loss functions, transfer

learning, continual learning, adversarial training, lightweight CNN, pruning strategy, knowledge distillation, and transformer. It can be expected that AI has the potential to play an active role in a paradigm shift in skin cancer diagnosis in the near future (refer to Section 3 for more information).

In comparison to other comparable reviews, this paper presents a comprehensive review in the topic of skin cancer classification with a focus on contemporary deep learning applications. It can be seen that the general evolutionary trend of these frameworks is structured, lightweight, and multimodal. With the help of this essay, one can gain an intuitive understanding of the core principles and issues in this field. Furthermore, anyone eager to engage in this field in the future should explore a number of different approaches to dealing with these issues. It is believed that the problems described above will become the research hotspots of scholars for a long time to come.

AUTHOR CONTRIBUTIONS

YW was responsible for writing the paper and supplementing materials. BC and RW were responsible for proposing amendments. AZ was responsible for revising and reviewing the article. DP was responsible for making the article more complete. SZ was the overall guide and was responsible for the whole project. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (Youth Project) under Grant 62101607 and National Natural Science Foundation of China under Grant 62071502.

REFERENCES

1. G Swann ed. *Journal of Visual Communication in Medicine* (2010), pp. 148–9. doi: 10.3109/17453054.2010.525439. New York, NY, United States: Foundation of Computer Science (FCS)
2. Montagna W. *The Structure and Function of Skin*. Amsterdam, Netherlands: Elsevier (2012).
3. Samuel E, Moore M, Voskoboynik M, Shackleton M, Haydon A. An Update on Adjuvant Systemic Therapies in Melanoma. *Melanoma Manage* (2019) 6: MMT28. doi: 10.2217/mmt-2019-0009
4. ACS. Cancer Facts & Figures 2018. In: *Cancer Facts Fig.* Atlanta, GA,U.S.: American Cancer Society (ACS) (2018). p. 1–71.
5. Rogers HW, Weinstock MA, Feldman SR, Coldiron BM. Incidence Estimate of Nonmelanoma Skin Cancer (Keratinocyte Carcinomas) in the Us Population, 2012. *JAMA Dermatol* (2015) 151: 1081–6. doi: 10.1001/jamadermatol.2015.1187
6. Sheha MA, Mabrouk MS, Sharawy A. Automatic Detection of Melanoma Skin Cancer Using Texture Analysis. *Int J Comput Appl* (2012) 42: 22–6. doi: 10.5120/5817-8129
7. Massone C, Di Stefani A, Soyer HP. Dermoscopy for Skin Cancer Detection. *Curr Opin Oncol* (2005) 17:147–53. doi: 10.1097/01.cco.0000152627.36243.26
8. Hoang L, Lee SH, Lee EJ, Kwon KR. Multiclass Skin Lesion Classification Using a Novel Lightweight Deep Learning Framework for Smart Healthcare. *Appl Sci* (2022) 12:2677. doi: 10.3390/app12052677
9. Anas M, Gupta K, Ahmad S. Skin Cancer Classification Using K-Means Clustering. *Int J Tech Res Appl* (2017) 5:62–5.
10. Ali AR, Li J, O'Shea SJ, Yang G, Trappenberg T, Ye X. A Deep Learning Based Approach to Skin Lesion Border Extraction With a Novel Edge Detector in Dermoscopy Images, in: *International Joint Conference on Neural Networks (IJCNN)* Manhattan, New York, U.S.:Institute of Electrical and Electronics Engineers (IEEE) (2019) 1–7. pp
11. Stanoszek LM, Wang GY, Harms PW. Histologic Mimics of Basal Cell Carcinoma. *Arch Pathol Lab Med* (2017) 141:1490–502. doi: 10.5858/arpa.2017-0222-RA
12. Goyal M, Knackstedt T, Yan S, Hassanpour S. Artificial Intelligence-Based Image Classification for Diagnosis of Skin Cancer: Challenges and Opportunities. *Comput Biol Med* (2020) 127:104065. doi: 10.1016/j.compbiomed.2020.104065
13. Weingast J, Scheibböck C, Wurm EM, Ranharter E, Porkert S, Dreiseitl S, et al. A Prospective Study of Mobile Phones for Dermatology in a Clinical Setting. *J Telemed Telecare* (2013) 19:213–8. doi: 10.1177/1357633x13490890
14. Arroyo JLG, Zapirain BG. Automated Detection of Melanoma in Dermoscopic Images. In: *Computer Vision Techniques for the Diagnosis of Skin Cancer*. New York, NY, United States:Springer (2014). p. 139–92.
15. Nachbar F, Stolz W, Merkle T, Cognetta AB, Vogt T, Landthaler M, et al. The Abcd Rule of Dermatoscopy: High Prospective Value in the Diagnosis of Doubtful Melanocytic Skin Lesions. *J Am Acad Dermatol* (1994) 30:551–9. doi: 10.1016/S0190-9622(94)70061-3
16. Menzies SW, Ingvar C, Crotty KA, McCarthy WH. Frequency and Morphologic Characteristics of Invasive Melanomas Lacking Specific Surface Microscopic Features. *Arch Dermatol* (1996) 132:1178–82. doi: 10.1001/archderm.1996.03890340038007
17. Argenziano G, Fabbrocini G, Carli P, De Giorgi V, Sammarco E, Delfino M. Epiluminescence Microscopy for the Diagnosis of Doubtful Melanocytic Skin Lesions: Comparison of the Abcd Rule of Dermatoscopy and a New 7-Point Checklist Based on Pattern Analysis. *Arch Dermatol* (1998) 134:1563–70. doi: 10.1001/archderm.134.12.1563
18. Noble WS. What Is a Support Vector Machine? *Nat Biotechnol* (2006) 24:1565–7. doi: 10.1038/nbt1206-1565
19. Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, et al. Xgboost: Extreme Gradient Boosting. In: *R Package Version 0.4-2*. Brookline, Massachusetts, U.S: Microtome Publishing (2015). 24 p. 1–4.
20. Safavian SR, Landgrebe D. A Survey of Decision Tree Classifier Methodology. *IEEE Trans Syst Man Cybernet* (1991) 21:660–74. doi: 10.1109/21.97458
21. Pomponiu V, Nejati H, Cheung NM. Deepmole: Deep Neural Networks for Skin Mole Lesion Classification, in: *IEEE International Conference on Image Processing (ICIP)*, Manhattan, New York, U.S:Institute of Electrical and Electronics Engineers (IEEE) (2016) 2623–7. pp
22. Habif TP, Chapman MS, Dinulos JG, Zug KA. *Skin Disease E-Book: Diagnosis and Treatment*. Amsterdam, Netherlands:Elsevier Health Sciences (2017).
23. Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, Lew MS. Deep Learning for Visual Understanding: A Review. *Neurocomputing* (2016) 187:27–48. doi: 10.1016/j.neucom.2015.09.116
24. Castiglioni I, Rundo L, Codari M, Di Leo G, Salvatore C, Interlenghi M, et al. Ai Applications to Medical Images: From Machine Learning to Deep Learning. *Physica Med* (2021) 83:9–24. doi: 10.1016/j.ejmp.2021.02.006
25. Rezvantalab A, Safigholi H, Karimijeshni S. Dermatologist Level Dermoscopy Skin Cancer Classification Using Different Deep Learning Convolutional Neural Networks Algorithms. *ArXiv* (2018) abs/1810.10348:arXiv:1810.10348. doi: 10.48550/arXiv.1810.10348
26. Li KM, Li EC. Skin Lesion Analysis Towards Melanoma Detection *via* End-to-End Deep Learning of Convolutional Neural Networks. *ArXiv* (2018) abs/1807.08332:arXiv:1807.08332. doi: 10.48550/arXiv.1807.08332

27. Perez F, Avila S, Valle E. Solo or Ensemble? Choosing a Cnn Architecture for Melanoma Classification. *Proc IEEE/CVF Conf Comput Vision Pattern Recognit Workshops* (2019), 0–0. doi: 10.1109/CVPRW.2019.00336
28. Li H, Pan Y, Zhao J, Zhang L. Skin Disease Diagnosis With Deep Learning: A Review. *Neurocomputing* (2021) 464:364–93. doi: 10.1016/j.neucom.2021.08.096
29. Hameed N, Ruskin A, Hassan KA, Hossain MA. A Comprehensive Survey on Image-Based Computer Aided Diagnosis Systems for Skin Cancer. In: *International Conference on Software, Knowledge, Information Management & Applications (SKIMA) (IEEE), 10th*. Manhattan, New York, U.S: Institute of Electrical and Electronics Engineers (IEEE) (2016). p. 205–14.
30. Kuntz S, Krieghoff-Henning E, Kather JN, Jutzi T, Höhn J, Kiehl L, et al. Gastrointestinal Cancer Classification and Prognostication From Histology Using Deep Learning: Systematic Review. *Eur J Cancer* (2021) 155: 200–15. doi: 10.1016/j.ejca.2021.07.012
31. Pathan S, Prabhu KG, Siddalingaswamy P. Techniques and Algorithms for Computer Aided Diagnosis of Pigmented Skin Lesions—A Review. *Biomed Signal Process Control* (2018) 39:237–62. doi: 10.1016/j.bspc.2017.07.010
32. Brinker TJ, Hekler A, Utikal JS, Grabe N, Schadendorf D, Klode J, et al. Skin Cancer Classification Using Convolutional Neural Networks: Systematic Review. *J Med Internet Res* (2018) 20:e11936. doi: 10.2196/11936
33. Manne R, Kantheti S, Kantheti S. Classification of Skin Cancer Using Deep Learning, Convolutionalneural Networks-Opportunities and Vulnerabilities-a Systematic Review. *Int J Modern Trends Sci Technol* (2020) 6:2455–3778. doi: 10.46501/ijmtst061118
34. Chan S, Reddy V, Myers B, Thibodeaux Q, Brownstone N, Liao W. Machine Learning in Dermatology: Current Applications, Opportunities, and Limitations. *Dermatol Ther* (2020) 10:365–86. doi: 10.1007/s13555-020-00372-0
35. Haggenmüller S, Maron RC, Hekler A, Utikal JS, Barata C, Barnhill RL, et al. Skin Cancer Classification *via* Convolutional Neural Networks: Systematic Review of Studies Involving Human Experts. *Eur J Cancer* (2021) 156:202–16. doi: 10.1016/j.ejca.2021.06.049
36. Höhn J, Hekler A, Krieghoff-Henning E, Kather JN, Utikal JS, Meier F, et al. Integrating Patient Data Into Skin Cancer Classification Using Convolutional Neural Networks: Systematic Review. *J Med Internet Res* (2021) 23:e20708. doi: 10.2196/20708
37. Dildar M, Akram S, Irfan M, Khan HU, Ramzan M, Mahmood AR, et al. Skin Cancer Detection: A Review Using Deep Learning Techniques. *Int J Environ Res Public Health* (2021) 18:5479. doi: 10.3390/ijerph18105479
38. Sheng M, Tang M, Lin W, Guo L, He W, Chen W, et al. The Value of Preoperative High-Resolution Mri With Microscopy Coil for Facial Nonmelanoma Skin Cancers. *Skin Res Technol* (2021) 27:62–9. doi: 10.1111/srt.12909
39. Moreno-Ramirez D, Ferrandiz L. Skin Cancer Telemedicine. In: *Telemedicine in Dermatology*. New York, NY, United States:Springer (2012). p. 113–21.
40. Fujisawa Y, Otomo Y, Ogata Y, Nakamura Y, Fujita R, Ishitsuka Y, et al. Deep-Learning-Based, Computer-Aided Classifier Developed With a Small Dataset of Clinical Images Surpasses Board-Certified Dermatologists in Skin Tumour Diagnosis. *Br J Dermatol* (2019) 180:373–81. doi: 10.1111/bjd.16924
41. Zhang L, Yang G, Ye X. Automatic Skin Lesion Segmentation by Coupling Deep Fully Convolutional Networks and Shallow Network With Textons. *J Med Imaging* (2019) 6:024001. doi: 10.1117/1.JMI.6.2.024001
42. Kasuya A, Aoshima M, Fukuchi K, Shimauchi T, Fujiyama T, Tokura Y. An Intuitive Explanation of Dermoscopic Structures by Digitally Reconstructed Pathological Horizontal Top-Down View Images. *Sci Rep* (2019) 9:1–7. doi: 10.1038/s41598-019-56522-8
43. Pellacani G, Seidenari S. Comparison Between Morphological Parameters in Pigmented Skin Lesion Images Acquired by Means of Epiluminescence Surface Microscopy and Polarized-Light Videomicroscopy. *Clinics Dermatol* (2002) 20:222–7. doi: 10.1016/S0738-081X(02)00231-6
44. Ali ARH, Li J, Yang G. Automating the Abcd Rule for Melanoma Detection: A Survey. *IEEE* (2020) 8:83333–46. doi: 10.1109/ACCESS.2020.2991034
45. Henning JS, Dusza SW, Wang SQ, Marghoob AA, Rabinovitz HS, Polsky D, et al. The Cash (Color, Architecture, Symmetry, and Homogeneity) Algorithm for Dermoscopy. *J Am Acad Dermatol* (2007) 56:45–52. doi: 10.1016/j.jaad.2006.09.003
46. Lorentzen H, Weismann K, Petersen CS, Grønhøj Larsen F, Secher L, Skødt V. Clinical and Dermatoscopic Diagnosis of Malignant Melanoma: Assessed by Expert and Non-Expert Groups. *Acta Dermato-venereol* (1999) 79:301–4. doi: 10.1080/000155599750010715
47. Mendonça T, Ferreira PM, Marques JS, Marcal AR, Rozeira J. Ph 2-a Dermoscopic Image Database for Research and Benchmarking, in: *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Manhattan, New York, U.S: Institute of Electrical and Electronics Engineers (IEEE) (2013). pp. 5437–40.
48. Brahmbhatt P, Rajan SN. Skin Lesion Segmentation Using Segnet With Binary Cross-Entropy. In: *International Conference on Artificial Intelligence and Speech Technology (Aist2019), 15th*, vol. 14. Delhi, India: Excel India Publishers (2019).
49. Saba T, Khan MA, Rehman A, Marie-Sainte SL. Region Extraction and Classification of Skin Cancer: A Heterogeneous Framework of Deep Cnn Features Fusion and Reduction. *J Med Syst* (2019) 43:1–19. doi: 10.1007/s10916-019-1413-3
50. Basak H, Kundu R, Sarkar R. Mfsnet: A Multi Focus Segmentation Network for Skin Lesion Segmentation. *Pattern Recognit* (2022) 128:108673. doi: 10.1016/j.patcog.2022.108673
51. Matsunaga K, Hamada A, Minagawa A, Koga H. Image Classification of Melanoma, Nevus and Seborrheic Keratosis by Deep Neural Network Ensemble. *ArXiv* (2017) abs/1703.03108:arXiv:1703.03108. doi: 10.48550/arXiv.1703.03108
52. West J, Ventura D, Warnick S. *Spring Research Presentation: A Theoretical Foundation for Inductive Transfer*. Brigham Young University: College of Physical and Mathematical Sciences (2007).
53. Giotis I, Molders N, Land S, Biehl M, Jonkman MF, Petkov N. Med-Node: A Computer-Assisted Melanoma Diagnosis System Using non-Dermoscopic Images. *Expert Syst Appl* (2015) 42:6578–85. doi: 10.1016/j.eswa.2015.04.034
54. Manzo M, Pellino S. Bucket of Deep Transfer Learning Features and Classification Models for Melanoma Detection. *J Imaging* (2020) 6:129. doi: 10.3390/jimaging6120129
55. Hosny KM, Kassem MA, Foaud MM. Classification of Skin Lesions Using Transfer Learning and Augmentation With Alex-Net. *PloS One* (2019) 14: e0217293. doi: 10.1371/journal.pone.0217293
56. Hosny KM, Kassem MA, Foaud MM. Skin Melanoma Classification Using Roi and Data Augmentation With Deep Convolutional Neural Networks. *Multimedia Tools Appl* (2020) 79:24029–55. doi: 10.1007/s11042-020-09067-2
57. Codella N, Rotemberg V, Tschandl P, Celebi ME, Dusza S, Gutman D, et al. Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (Isic). *ArXiv* (2019) abs/1902.03368:arXiv:1902.03368. doi: 10.48550/arXiv.1902.03368
58. Tschandl P, Rosendahl C, Kittler H. The Ham10000 Dataset, a Large Collection of Multi-Source Dermatoscopic Images of Common Pigmented Skin Lesions. *Sci Data* (2018) 5:1–9. doi: 10.1038/sdata.2018.161
59. Young K, Booth G, Simpson B, Dutton R, Shrapnel S. Deep Neural Network or Dermatologist? In: *Interpretability of Machine Intelligence in Medical Image Computing and Multimodal Learning for Clinical Decision Support*. New York, NY, United States:Springer (2019). p. 48–55.
60. Polat K, Koc KO. Detection of Skin Diseases From Dermoscopy Image Using the Combination of Convolutional Neural Network and One-Versus-All. *J Artif Intell Syst* (2020) 2:80–97. doi: 10.33969/AIS.2020.21006
61. Datta SK, Shaikh MA, Srihari SN, Gao M. Soft Attention Improves Skin Cancer Classification Performance. In: *Interpretability of Machine Intelligence in Medical Image Computing, and Topological Data Analysis and Its Applications for Medical Data*. New York, NY, United States:Springer (2021). p. 13–23.
62. Kawahara J, Daneshvar S, Argenziano G, Hamarneh G. Seven-Point Checklist and Skin Lesion Classification Using Multitask Multimodal Neural Nets. *IEEE J Biomed Health Inf* (2018) 23:538–46. doi: 10.1109/JBHI.2018.2824327
63. Coppola D, Lee HK, Guan C. Interpreting Mechanisms of Prediction for Skin Cancer Diagnosis Using Multi-Task Learning. *Proc IEEE/CVF Conf Comput Vision Pattern Recognit Workshops* (2020), 734–5. doi: 10.1109/CVPRW50498.2020.00375

64. Yao P, Shen S, Xu M, Liu P, Zhang F, Xing J, et al. Single Model Deep Learning on Imbalanced Small Datasets for Skin Lesion Classification. *IEEE Trans Med Imaging* (2021) 41:1242–54. doi: 10.1109/TMI.2021.3136682
65. Combalia M, Codella NC, Rotemberg V, Helba B, Vilaplana V, Reiter O, et al. Bcn20000: Dermoscopic Lesions in the Wild. *ArXiv* (2019) abs/1908.02288:arXiv:1908.02288. doi: 10.48550/arXiv.1908.02288
66. Mou Y, Welten S, Yediel YU, Kirsten T, Beyan OD. Distributed Learning for Melanoma Classification Using Personal Health Train. *ArXiv* (2021) abs/2103.13226:arXiv:2103.13226. doi: 10.48550/arXiv.2103.13226
67. Maron RC, Schlager JG, Haggenmüller S, von Kalle C, Utikal JS, Meier F, et al. A Benchmark for Neural Network Robustness in Skin Cancer Classification. *Eur J Cancer* (2021) 155:191–9. doi: 10.1016/j.ejca.2021.06.047
68. Gessert N, Nielsen M, Shaikh M, Werner R, Schlaefer A. Skin Lesion Classification Using Ensembles of Multi-Resolution Efficientnets With Meta Data. *MethodsX* (2020) 7:100864. doi: 10.1016/j.mex.2020.100864
69. Gutman D, Codella NC, Celebi E, Helba B, Marchetti M, Mishra N, et al. Skin Lesion Analysis Toward Melanoma Detection: A Challenge at the International Symposium on Biomedical Imaging (Isbi) 2016, Hosted by the International Skin Imaging Collaboration (Isic). *IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)* (2016) abs/1605.01397:arXiv:1605.01397. doi: 10.1109/ISBI.2018.8363547
70. Cassidy B, Kendrick C, Brodzicki A, Jaworek-Korjakowska J, Yap MH. Analysis of the Isic Image Datasets: Usage, Benchmarks and Recommendations. *Med Image Anal* (2022) *75*:102305. doi: 10.1016/j.media.2021.102305
71. Lopez AR, Giro-i Nieto X, Burdick J, Marques O. Skin Lesion Classification From Dermoscopic Images Using Deep Learning Techniques, in: *13th IASTED International Conference on Biomedical Engineering (BioMed)*, Manhattan, New York, U.S: Institute of Electrical and Electronics Engineers (IEEE) (2017). pp. 49–54.
72. Nozdryn-Plotnicki A, Yap J, Yolland W. (2018). Ensembling Convolutional Neural Networks for Skin Cancer Classification. International Skin Imaging Collaboration (ISIC) Challenge on Skin Image Analysis for Melanoma Detection. MICCAI.
73. Bevan PJ, Atapour-Abarghouei A. Skin Deep Unlearning: Artefact and Instrument Debiasing in the Context of Melanoma Classification. *ArXiv* (2021) abs/2109.09818:arXiv:2109.09818. doi: 10.48550/arXiv.2109.09818
74. Ali AR, Li J, Yang G, O'Shea SJ. A Machine Learning Approach to Automatic Detection of Irregularity in Skin Lesion Border Using Dermoscopic Images. *PeerJ Comput Sci* (2020) 6:e268. doi: 10.7717/peerj-cs.268
75. Krizhevsky A, Sutskever I, Hinton GE. Imagenet Classification With Deep Convolutional Neural Networks. *Adv Neural Inf Process Syst* (2012) 60:84–90. doi: 10.1145/3065386
76. Masood A, Al-Jumaily A, Anam K. Self-Supervised Learning Model for Skin Cancer Diagnosis, in: *7th International IEEE/EMBS Conference on Neural Engineering (NER)*, Manhattan, New York, U.S.:Institute of Electrical and Electronics Engineers (IEEE) (2015). 1012–5 pp. doi: 10.1109/NER.2015.7146798
77. Mishra S, Imaizumi H, Yamasaki T. Interpreting Fine-Grained Dermatological Classification by Deep Learning. *Proc IEEE/CVF Conf Comput Vision Pattern Recognit Workshops* (2019), 0–0. doi: 10.1109/CVPRW.2019.00331
78. 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
79. Nasr-Esfahani E, Samavi S, Karimi N, Soroushmehr S, Jafari M, Ward K, et al. Melanoma Detection by Analysis of Clinical Images Using Convolutional Neural Network, in: *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Manhattan, New York, U.S: Institute of Electrical and Electronics Engineers (IEEE) (2016). 1373–6 pp. doi: 10.1109/EMBC.2016.7590963
80. Walker B, Rehg J, Kalra A, Winters R, Drews P, Dascalu J, et al. Dermoscopy Diagnosis of Cancerous Lesions Utilizing Dual Deep Learning Algorithms via Visual and Audio (Sonification) Outputs: Laboratory and Prospective Observational Studies. *EBioMedicine* (2019) 40:176–83. doi: 10.1016/j.ebiom.2019.01.028
81. Jaworek-Korjakowska J, Kleczek P, Gorgon M. Melanoma Thickness Prediction Based on Convolutional Neural Network With Vgg-19 Model Transfer Learning. *Proc IEEE/CVF Conf Comput Vision Pattern Recognit Workshops* (2019), 0–0. doi: 10.1109/CVPRW.2019.00333
82. Kawahara J, Daneshvar S, Argenziano G, Hamarneh G. Seven-Point Checklist and Skin Lesion Classification Using Multitask Multimodal Neural Nets. *IEEE J Biomed Health Inf* (2019) 23:538–46. doi: 10.1109/JBHI.2018.2824327
83. Rahman Z, Hossain MS, Islam MR, Hasan MM, Hridhee RA. An Approach for Multiclass Skin Lesion Classification Based on Ensemble Learning. *Inf Med Unlocked* (2021) 25:100659. doi: 10.1016/j.imu.2021.100659
84. Bisla D, Choromanska A, Berman RS, Stein JA, Polsky D. Towards Automated Melanoma Detection With Deep Learning: Data Purification and Augmentation. *Proc IEEE/CVF Conf Comput Vision Pattern Recognit Workshops* (2019), 0–0. doi: 10.1109/CVPRW.2019.00330
85. Qin Z, Liu Z, Zhu P, Xue Y. A Gan-Based Image Synthesis Method for Skin Lesion Classification. *Comput Methods Programs Biomed* (2020) 95:105568. doi: 10.1016/j.cmpb.2020.105568
86. Kaur R, GholamHosseini H, Sinha R. Synthetic Images Generation Using Conditional Generative Adversarial Network for Skin Cancer Classification, in: *TENCON 2021-2021 IEEE Region 10 Conference (TENCON)*, Manhattan, New York, U.S: Institute of Electrical and Electronics Engineers (IEEE) (2021). 381–6 pp.
87. Lee KW, Chin RKY. The Effectiveness of Data Augmentation for Melanoma Skin Cancer Prediction Using Convolutional Neural Networks, in: *IEEE 2nd International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, Manhattan, New York, U.S:Institute of Electrical and Electronics Engineers (IEEE) (2020). 1–6 pp.
88. Ahmad B, Jun S, Palade V, You Q, Mao L, Zhongjie M. Improving Skin Cancer Classification Using Heavy-Tailed Student T-Distribution in Generative Adversarial Networks (Ted-Gan). *Diagnostics* (2021) 11:2147. doi: 10.3390/diagnostics11112147
89. Abdelhalim ISA, Mohamed MF, Mahdy YB. Data Augmentation for Skin Lesion Using Self-Attention Based Progressive Generative Adversarial Network. *Expert Syst With Appl* (2021) 165:113922. doi: 10.1016/j.eswa.2020.113922
90. Le DN, Le HX, Ngo LT, Ngo HT. Transfer Learning With Class-Weighted and Focal Loss Function for Automatic Skin Cancer Classification. *ArXiv* (2020) abs/2009.05977:arXiv:2009.05977. doi: 10.48550/arXiv.2009.05977
91. Pham TC, Doucet A, Luong CM, Tran CT, Hoang VD. Improving Skin-Disease Classification Based on Customized Loss Function Combined With Balanced Mini-Batch Logic and Real-Time Image Augmentation. *IEEE* (2020) 8:150725–37. doi: 10.1109/ACCESS.2020.3016653
92. Shen S, Xu M, Zhang F, Shao P, Liu H, Xu L, et al. Low-Cost and High-Performance Data Augmentation for Deep-Learning-Based Skin Lesion Classification. *ArXiv* (2021) abs/2101.02353:arXiv:2101.02353. doi: 10.34133/2022/9765307
93. Castro PB, Krohling B, Pacheco AG, Krohling RA. An App to Detect Melanoma Using Deep Learning: An Approach to Handle Imbalanced Data Based on Evolutionary Algorithms, in: *International Joint Conference on Neural Networks (IJCNN)*, Manhattan, New York, U.S.:Institute of Electrical and Electronics Engineers (IEEE) (2020). 1–6 pp.
94. Abayomi-Alli OO, Damasevicius R, Misra S, Maskeliunas R, Abayomi-Alli A. Malignant Skin Melanoma Detection Using Image Augmentation by Oversampling in Nonlinear Lower-Dimensional Embedding Manifold. *Turkish J Electric Eng Comput Sci* (2021) 29:2600–14. doi: 10.3906/elk-2101-133
95. Mustafa B, Loh A, Freyberg J, MacWilliams P, Wilson M, McKinney SM, et al. Supervised Transfer Learning at Scale for Medical Imaging.*ArXiv* (2021) abs/2101.05913:arXiv:2101.05913. doi: 10.48550/arXiv.2101.05913
96. Gu Y, Ge Z, Bonnington CP, Zhou J. Progressive Transfer Learning and Adversarial Domain Adaptation for Cross-Domain Skin Disease Classification. *IEEE J Biomed Health Inf* (2020) 24:1379–93. doi: 10.1109/JBHI.2019.2942429
97. Huq A, Pervin MT. Analysis of Adversarial Attacks on Skin Cancer Recognition, in: *International Conference on Data Science and Its*

Applications (ICoDSA), Manhattan, New York, U.S.:Institute of Electrical and Electronics Engineers (IEEE) (2020). 1–4 pp.
98. Zunair H, Hamza AB. Melanoma Detection Using Adversarial Training and Deep Transfer Learning. *Phys Med Biol* (2020) 65:135005. doi: 10.1088/1361-6560/ab86d3
99. Bian J, Zhang S, Wang S, Zhang J, Guo J. Skin Lesion Classification by Multi-View Filtered Transfer Learning. *IEEE* (2021) 9:66052–61. doi: 10.1109/ACCESS.2021.3076533
100. Wei Z, Song H, Chen L, Li Q, Han G. Attention-Based Denseunet Network With Adversarial Training for Skin Lesion Segmentation. *IEEE Access* (2019) 7:136616–29. doi: 10.1109/ACCESS.2019.2940794
101. Hirano H, Minagi A, Takemoto K. Universal Adversarial Attacks on Deep Neural Networks for Medical Image Classification. *BMC Med Imaging* (2021) 21:1–13. doi: 10.1186/s12880-020-00530-y
102. Abdar M, Samami M, Mahmoodabad SD, Doan T, Mazoure B, Hashemifesharaki R, et al. Uncertainty Quantification in Skin Cancer Classification Using Three-Way Decision-Based Bayesian Deep Learning. *Comput Biol Med* (2021) 13:104418. doi: 10.1016/j.compbiomed.2021.104418
103. Alzubaidi L, Al-Amidie M, Al-Asadi A, Humaidi AJ, Al-Shamma O, Fadhel MA, et al. Novel Transfer Learning Approach for Medical Imaging With Limited Labeled Data. *Cancers* (2021) 13:1590. doi: 10.3390/cancers13071590
104. Li H, Wang Y, Wan R, Wang S, Li TQ, Kot A. Domain Generalization for Medical Imaging Classification With Linear-Dependency Regularization. *Adv Neural Inf Process Syst* (2020) 33: 3118–29.
105. Yang Y, Cui Z, Xu J, Zhong C, Wang R, Zheng WS. Continual Learning With Bayesian Model Based on a Fixed Pre-Trained Feature Extractor. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (Springer)*. New York, NY, United States: Springer (2021). p. 397–406.
106. Goldstein O, Kachuee M, Sarrafzadeh M. Decentralized Knowledge Transfer on Edge Networks for Detecting Cancer in Images, in: *IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, Manhattan, New York, U.S:Institute of Electrical and Electronics Engineers (IEEE) (2021). 1–5 pp.
107. Velasco J, Pascion C, Alberio JW, Apuang J, Cruz JS, Gomez MA, et al. A Smartphone-Based Skin Disease Classification Using Mobilenet Cnn. *ArXiv* (2019) abs/1911.07929:arXiv:1911.07929. doi: 10.30534/ijatcse/2019/116852019
108. Pedro R, Oliveira AL. Assessing the Impact of Attention and Self-Attention Mechanisms on the Classification of Skin Lesions. *ArXiv* (2021) abs/2112.12748:arXiv:2112.12748. doi: 10.48550/arXiv.2112.12748
109. Xiang K, Peng L, Yang H, Li M, Cao Z, Jiang S, et al. A Novel Weight Pruning Strategy for Light Weight Neural Networks With Application to the Diagnosis of Skin Disease. *Appl Soft Comput* (2021) 111:107707. doi: 10.1016/j.asoc.2021.107707
110. Bayasi N, Hamarneh G, Garbi R. Culprit-Prune-Net: Efficient Continual Sequential Multi-Domain Learning With Application to Skin Lesion Classification. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. New York, NY, United States:Springer (2021). p. 165–75.
111. Alche MN, Acevedo D, Mejail M. Efficientarl: Improving Skin Cancer Diagnoses by Combining Lightweight Attention on Efficientnet. *Proc IEEE/CVF Int Conf Comput Vision* (2021), 3354–60. doi: 10.1109/ICCVW54120.2021.00374
112. Yilmaz A, Kalebasi M, Samoylenko Y, Guvenilir ME, Uvet H. Benchmarking of Lightweight Deep Learning Architectures for Skin Cancer Classification Using Isic 2017 Dataset. *ArXiv}* (2021) abs/2110.12270:arXiv:2110.12270. doi: 10.48550/arXiv.2110.12270
113. Chen J, Chen J, Zhou Z, Li B, Yuille A, Lu Y. Mt-Transunet: Mediating Multi-Task Tokens in Transformers for Skin Lesion Segmentation and Classification. *ArXiv* (2021) abs/2112.01767:arXiv:2112.01767. doi: 10.48550/arXiv.2112.01767
114. Hajabdollahi M, Esfandiarpoor R, Khadivi P, Soroushmehr SMR, Karimi N, Samavi S. Simplification of Neural Networks for Skin Lesion Image Segmentation Using Color Channel Pruning. *Comput Med Imaging Graphics* (2020) 82:101729. doi: 10.1016/j.compmedimag.2020.101729
115. Back S, Lee S, Shin S, Yu Y, Yuk T, Jong S, et al. Robust Skin Disease Classification by Distilling Deep Neural Network Ensemble for the Mobile Diagnosis of Herpes Zoster. *IEEE* (2021) 9:20156–69. doi: 10.1109/ACCESS.2021.3054403
116. Hameed N, Shabut A, Hameed F, Cirstea S, Harriet S, Hossain A. Mobile Based Skin Lesions Classification Using Convolution Neural Network. *Ann Emerging Technol Comput (AETiC)* (2020) 4:26–37. doi: 10.33166/AETiC.2020.02.003
117. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR* (2014) abs/1409.1556:arXiv:1409.1556. doi: 10.48550/arXiv.1409.1556
118. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going Deeper With Convolutions. *Proc IEEE Conf Comput Vision Pattern Recognit* (2015), 1–9. doi: 10.1109/CVPR.2015.7298594
119. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. *Proc IEEE Conf Comput Vision Pattern Recognit* (2016), 770–8. doi: 10.1109/CVPR.2016.90
120. Marcus G, Davis E. Rebooting Ai. In: *Building Artificial Intelligence We can Trust (Vintage)*. New York, NY, United States: Vintage (2019).
121. Rotemberg V, Kurtansky N, Betz-Stablein B, Caffery L, Chousakos E, Codella N, et al. A Patient-Centric Dataset of Images and Metadata for Identifying Melanomas Using Clinical Context. *Sci Data* (2021) 8:81–8. doi: 10.1038/s41597-021-00815-z
122. Mikołajczyk A, Grochowski M. Data Augmentation for Improving Deep Learning in Image Classification Problem, in: *international interdisciplinary PhD workshop (IIPhDW)*, Manhattan, New York, U.S.:Institute of Electrical and Electronics Engineers (IEEE) (2018) 117–22 pp.
123. Tizhoosh HR, Pantanowitz L. Artificial Intelligence and Digital Pathology: Challenges and Opportunities. *J Pathol Inf* (2018) 9:1–6 doi: 10.4103/jpi.jpi_53_18
124. Blalock D, Ortiz JJG, Frankle J, Guttag J. What is the State of Neural Network Pruning? *ArXiv* (2020) abs/2003.03033:arXiv:2003.03033.
125. Hinton GE, Vinyals O, Dean J. Distilling the Knowledge in a Neural Network. *ArXiv* (2015) abs/1503.02531:ArXiv abs/1503.02531. doi: 10.48550/arXiv.1503.02531
126. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is All You Need. *Adv Neural Inf Process Syst* (2017) 30:5998–6008. doi: 10.48550/arXiv.1706.03762
127. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *ArXiv*(2020) abs/2010.11929:arXiv:2010.11929. doi: 10.48550/arXiv.2010.11929

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

TYPE Original Research
PUBLISHED 11 October 2022
DOI 10.3389/fonc.2022.879553

OPEN ACCESS

EDITED BY
Tingying Peng,
Helmholtz zentrum München,
Germany

REVIEWED BY
Chandra Bortolotto,
San Matteo Hospital Foundation
(IRCCS), Italy
Guang Yang,
Imperial College London,
United Kingdom

*CORRESPONDENCE
Chiara Giraudo
chiara.giraudo@unipd.it

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

RECEIVED 19 February 2022
ACCEPTED 20 September 2022
PUBLISHED 11 October 2022

CITATION
Giraudo C, Fichera G, Del Fiore P,
Mocellin S, Brunello A, Rastrelli M and
Stramare R (2022) Tumor cellularity
beyond the visible in soft tissue
sarcomas: Results of an ADC-based,
single center, and preliminary
radiomics study.
Front. Oncol. 12:879553.
doi: 10.3389/fonc.2022.879553

Tumor cellularity beyond the visible in soft tissue sarcomas: Results of an ADC-based, single center, and preliminary radiomics study

Chiara Giraudo[1]*, Giulia Fichera[1], Paolo Del Fiore[2], Simone Mocellin[2,3], Antonella Brunello[4], Marco Rastrelli[2,3] and Roberto Stramare[1]

[1]Department of Medicine – DIMED, University of Padova, Padova, Italy, [2]Soft-Tissue, Peritoneum and Melanoma Surgical Oncology Unit, Veneto Institute of Oncology - IOV Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS), Padova, Italy, [3]Department of Surgery, Oncology and Gastroenterology (DISCOG), University of Padua, Padua, Italy, [4]Department of Oncology, Medical Oncology 1 Unit, Veneto Institute of Oncology - IOV Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS), Padua, Italy

Purpose: Soft tissue sarcomas represent approximately 1% of all malignancies, and diagnostic radiology plays a significant role in the overall management of this rare group of tumors. Recently, quantitative imaging and, in particular, radiomics demonstrated to provide significant novel information, for instance, in terms of prognosis and grading. The aim of this study was to evaluate the prognostic role of radiomic variables extracted from apparent diffusion coefficient (ADC) maps collected at diagnosis in patients with soft tissue sarcomas in terms of overall survival and metastatic spread as well as to assess the relationship between radiomics and the tumor grade.

Methods: Patients with histologically proven soft tissue sarcomas treated in our tertiary center from 2016 to 2019 who underwent an Magnetic Resonance (MR) scan at diagnosis including diffusion-weighted imaging were included in this retrospective institution review board–approved study. Each primary lesion was segmented using the b50 images; the volumetric region of interest was then applied on the ADC map. A total of 33 radiomic features were extracted, and highly correlating features were selected by factor analysis. In the case of feature/s showing statistically significant results, the diagnostic accuracy was computed. The Spearman correlation coefficient was used to evaluate the relationship between the tumor grade and radiomic features selected by factor analysis. All analyses were performed applying $p<0.05$ as a significant level.

Results: A total of 36 patients matched the inclusion criteria (15 women; mean age 58.9 ± 15 years old). The most frequent histotype was myxofibrosarcoma (16.6%), and most of the patients were affected by high-grade lesions (77.7%). Seven patients had pulmonary metastases, and, altogether, eight were deceased. Only the feature Imc1 turned out to be a predictor of metastatic spread (p=0.045 after Bonferroni correction) with 76.7% accuracy. The value -0.16 showed 73.3% sensitivity and 71.4% specificity, and patients with metastases showed lower values (mean Imc1 of metastatic patients -0.31). None of the examined variables was a predictor of the overall outcome (p>0.05, each). A moderate statistically significant correlation emerged only between Imc1 and the tumor grade (r=0.457, p=0.005).

Conclusions: In conclusion, the radiomic feature Imc1 acts as a predictor of metastatic spread in patients with soft tissue sarcomas and correlates with the tumor grade.

KEYWORDS

radiomics, soft tissue sarcoma, magnetic resonance, tumor grade, metastasis

Introduction

Soft tissue sarcomas are rare tumors representing approximately 1% of all malignancies, and diagnostic radiology plays a significant role not only at diagnosis and staging but also during follow-up (1–3). In particular, MR imaging, because of its intrinsic soft tissue contrast, is considered the main tool for investigating the primary site. Furthermore, this technique carries additional advantages due to the possibility to perform quantitative and functional analyses. In fact, the information about tumor perfusion, chemical composition, and cellularity can be easily assessed by dynamic contrast-enhanced (DCE) techniques, spectroscopy, and diffusion-weighted imaging (DWI), respectively (3–9). Nowadays, DWI is considered crucial in oncological imaging in general (10, 11). For soft tissue sarcomas, it has been demonstrated that it also contributes in distinguishing recurrences from postsurgical scars, improves specificity in defining tumor margin infiltration, and predicts the response to treatment (8–10). Despite these encouraging results, potential pitfalls and controversies must be addressed, especially for myxoid tumors due to their high mucin content. In fact, in this case, the distinction between benign and malignant lesions by DWI could be hampered (10).

The recent technical developments allowed radiologists to move further in their contribution to the overall diagnostic management of cancer patients as demonstrated by the increasing use of complex analyses such as histogram and radiomics. Indeed, these types of computations were demonstrated to be very useful for different types of tumors including soft tissue sarcomas, applying various sequences (12–15). For instance, Corino et al. already showed that the variables of first order (FOS) extracted from apparent diffusion coefficient (ADC) maps allow the distinction of the lesions of different grades. Similar conclusions were drawn by Xu and colleagues using T1w and T2w fat-sat imaging (15, 16). Using DCE-MRI at the baseline in patients with high-grade non-metastatic soft tissue sarcomas, Crombe et al. showed that radiomic variables have an important prognostic role (17). Lastly, Gao and colleagues did not predict the response to radiotherapy by DWI but better predicted the treatment effect score applying delta radiomics (18).

Despite this growing evidence, up to now, to the best of our knowledge, the prognostic value of radiomics regarding metastatic spread has not been fully investigated yet.

Thus, the aim of this study was to evaluate the prognostic role of radiomic variables extracted from ADC maps collected at diagnosis in patients with soft tissue sarcomas in terms of overall survival and metastatic spread as well as to assess the relationship between radiomics and the tumor grade.

Methods

Patients with histologically proven soft tissue sarcomas treated in our tertiary center from 2016 to 2019 who

underwent an MR scan at diagnosis by a 1.5 T scanner (Siemens Avanto 1.5T, Siemens Healthcare, Siemens, Erlangen, Germany) including an axial short tau inversion recovery–DWI sequence with 6 mm slice thickness and two b-values (i.e., b50 and b 800) were examined for this preliminary retrospective single-center institution review board–approved study. One radiologist with 12 years of experience in musculoskeletal imaging segmented each lesion along tumor margins using the b50 images because of the higher spatial resolution than the ADC map. The volumetric region of interest was then applied on the map (Figure 1). From each segmented volume, the 33 radiomic features of two classes were extracted: intensity-based features (FOS) and texture features [gray-level co-occurrence matrix (GLCM as well as gray-level run length matrix (GLRLM)]. The segmentation and extraction of radiomics features have been performed by an open-source software (3D Slicer, www.slicer.org).

Statistical analysis

Descriptive statistics were applied for demographics, tumor histotype, site, grade, and metastatic spread. Factor analysis was applied to select highly correlating radiomic features. Then,

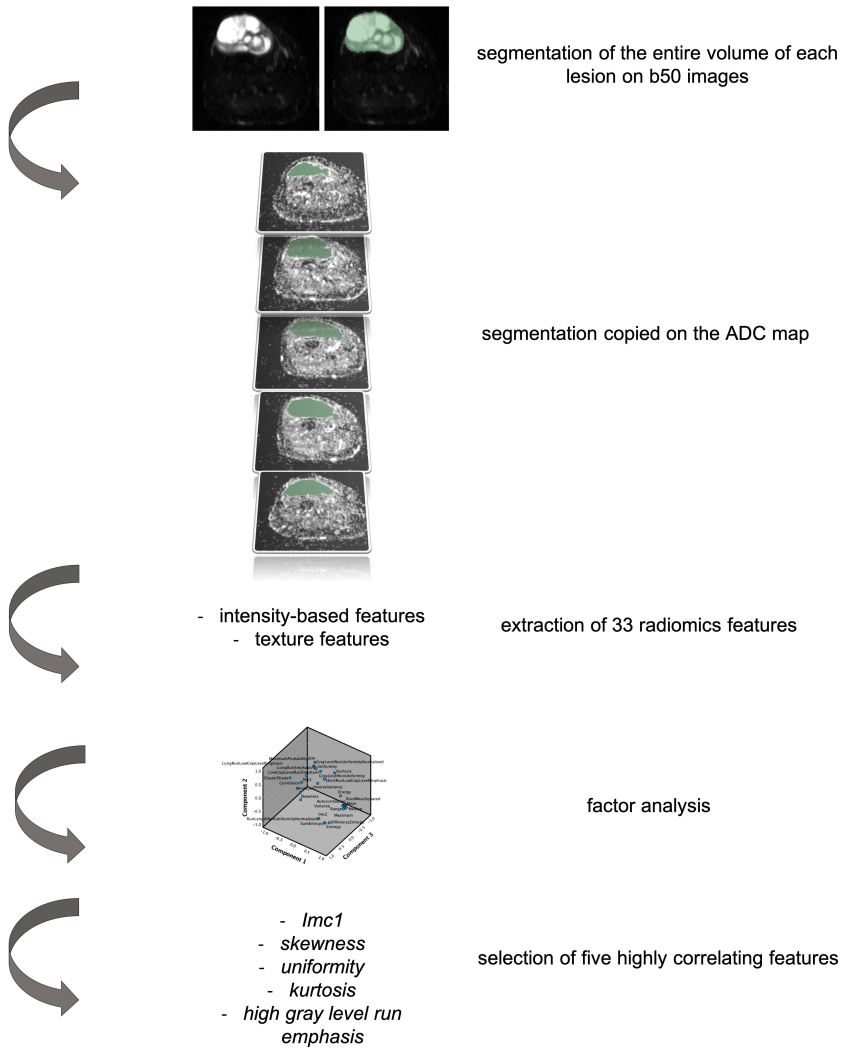


FIGURE 1
Representation of the segmentation process and data extraction from the primary lesion of patients with soft tissue sarcoma included in the study. In particular, it is hereby represented by a 76-year-old woman affected by a myxofibrosarcoma of the right lower leg. The axial b50 DWI images have been used for segmentation of the entire lesion. The volumetric segmentation has been then applied on the apparent diffusion coefficient (ADC) map and the extraction of 33 radiomic features performed. Factor analysis allowed the selection of five highly correlating features.

stepwise regression analysis was used to evaluate if any of the selected radiomic variables had a predicting role on the overall outcome (dead/alive) and/or the metastatic spread. Moreover, the Bonferroni correction was applied to correct the statistical significance level for multiple tests. In the case of feature/s showing statistically significant results, the diagnostic accuracy was computed using ROC curves and the value/s with the highest Youden index were selected as a cut-off. The Spearman correlation coefficient was used to evaluate the relationship between the tumor grade and all radiomic features selected by factor analysis.

To evaluate the robustness of the proposed method, all segmentations and data extraction were repeated by a second reader with 4 years of experience in oncological imaging and the intraclass correlation coefficient (ICC) of the variables highly correlating at factor analysis were computed. ICC values >.750 were considered excellent (19).

All statistical analyses were performed with SPSS (IBM SPSS Statistics version 27, IBM Armonk, NY, USA), applying $p<0.05$ as a significant level.

Results

From an overall amount of 80 cases treated in our center, 36 patients matched the inclusion criteria (15 women; mean age 58.9 ± 15 years old) and were examined. The characteristics of the examined population are summarized in Table 1. The most frequent histotype was myxofibrosarcoma (six patients, 16.6%), and most of the patients were affected by high-grade lesions (i.e., 28 had grade III lesions, 77.7%). Seven patients had pulmonary metastases, and, altogether, eight were deceased. On average, the survival was of 56.9 ± 22 months. Most of the lesions affected the lower limbs (29, 80.5%).

Factor analysis allowed the extraction of five highly correlating variables: three of first order (kurtosis, skewness, and uniformity) and two of second order (informational measure of correlation (Imc1) and high gray-level run emphasis).

Only the feature Imc1 turned out to be a predictor of metastatic spread ($p=0.045$ after Bonferroni correction) with 76.7% accuracy (Figure 2). The value -0.16 showed 73.3% sensitivity and 71.4% specificity, and patients with metastases showed lower values (mean Imc1 of metastatic patients -0.31). None of the examined variables was a predictor of the overall outcome ($p>0.05$ each).

A moderate statistically significant correlation emerged only between Imc1 and the tumor grade ($r=0.457$, $p=0.005$) (Figure 3).

The proposed method of the segmentation and extraction of radiomic features showed high reproducibility: kurtosis ICC = .870 [95% CI,.748 –.933], skewness ICC = .765 [95% CI,.544 –.879], uniformity ICC = .814 [95% CI,.640 –.904], Imc1 ICC = .786 [95%CI,.584–.890], and high gray-level run emphasis ICC=.842 [95% CI,.692 –.918].

TABLE 1 Characteristics of the examined population.

Gender(female/male)	14/22
Age(years)	58.9 ± 15 (mean ± SD)(range 18–82)
Histotype	6 myxofibrosarcoma
	6 undifferentiated pleomorphic sarcoma
	5 leiomyosarcoma
	4 myxoid liposarcoma
	3 aggressive fibromatosis
	2 synovial sarcoma
	2 solitary fibrous tumor
	2 liposarcoma
	2 round cell liposarcoma
	1 angiosarcoma
	1 hemangioendothelioma
	1 malignant peripheral nerve sheath tumor
	1 spindle cell sarcoma
Site	29 lower limbs
	3 upper limbs
	2 pelvis
	2 thorax
	1 neck
Grade	3 grade I
	5 grade II
	28 grade III
Metastatic(yes/no)	7/29
Deceased(yes/no)	8/28
Survival in months(median)	55

Discussion

To the best of our knowledge, this is one of the largest studies investigating the role of radiomic features extracted from the ADC maps of patients with soft tissue sarcomas at diagnosis demonstrating that the Imc1 is a predictor of metastatic spread and correlates with the tumor grade. Similar results were obtained using T1w and T2w fat-sat sequences by other groups. For instance, Tian and colleagues, applying a machine learning model, and Vallieres et al., associating the above-mentioned MR sequences with positron emission tomography (PET) information, obtained an early detection of pulmonary metastases (13, 20). This last evidence, together with other previous promising results provided by PET/MR-based histogram analyses and texture features, suggests that the role of hybrid imaging should be further assessed, especially about

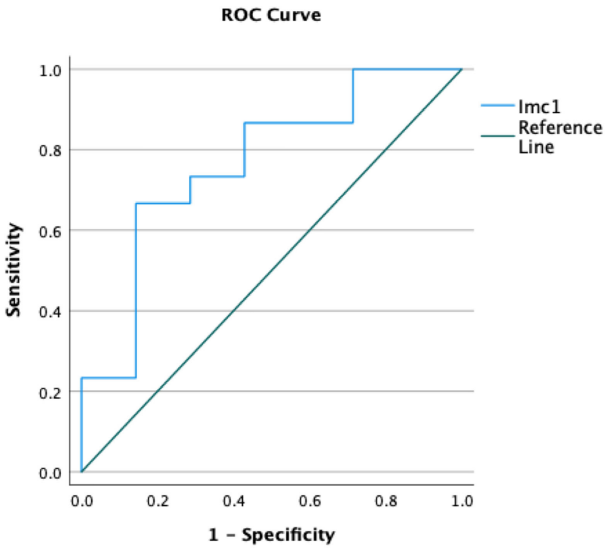


FIGURE 2
Receiver operating curves (ROCs) demonstrating the accuracy of the radiomic feature Imc1 in predicting the occurrence of metastases in patients with soft tissue sarcomas.

the possibility to simultaneously collect data regarding metabolic activity and functional information (12, 20).

Regarding, in particular, the Imc1 feature, it quantifies the complexity of the texture and its importance has been previously demonstrated in several studies on different types of cancer (21–24). For instance, our results are in line with those of Liao and colleagues who identified more negative values in patients with brain metastasis due to non-small lung cancer with poor local tumor control, thus suggesting that tumors with high intralesion heterogeneity might be associated with a worse clinical course, in our study represented by the occurrence of metastasis. Moreover, together with other features, it was part of a "radiomic signature" indicating patients with ground glass nodules at risk for invasive adenocarcinoma (25). The feature Imc1 may even be considered a novel biomarker of metastatic spread in patients with soft tissue sarcomas considering that it showed an excellent repeatability in our dataset and strong robustness in a previously published computed tomography–based computational model (26). Nevertheless, it should be considered that further studies on a larger population are needed to fully assess the hereby presented evidence; since, for instance, Corino et al., in a group of 19 patients with soft tissue sarcomas, showed that only the features of the first order were the best classifier of the tumor grade (15).

For the overall role of radiomics for this heterogeneous group of tumors, as previously underlined, different MR sequences have been used with very promising results (17, 18) and probably not only more comparisons among sequences are necessary (27), but it should also be assessed if a model combining information deriving from multiple sequences could provide additional and more robust results.

In contrast to part of the literature, in our study, none of the selected radiomic features turned out to be a predictor of the overall outcome. This discrepancy might be due to the heterogeneity of our sample in terms of the histotype and grade. In fact, radiomics acted as prognostic factor in selected groups like patients with myxoid and liposarcomas only and patients with high-grade lesions (28–30). Moreover, in the above-mentioned studies, other types of sequences (i.e., T1w, T2 fat-sat, or DCE-MRI) than DWI have also been used. We strongly encourage further investigation in this direction to also gain new knowledge about the potential prognostic role of this sequence in the specific subgroups of patients.

This study is affected by several limits. First, as noted above, the sample size and the large variety of histotypes did not allow to investigate the predictive role of this technique on the type of tumor and the relationship between different histotypes and radiomic features. Moreover, the small population also hampered the subdivision of our sample in test and validation sets. Certainly, the rarity of the disease contributed to these flaws and multicenter studies are necessary to fill these voids and provide further insights into the role of radiomics for these tumors.

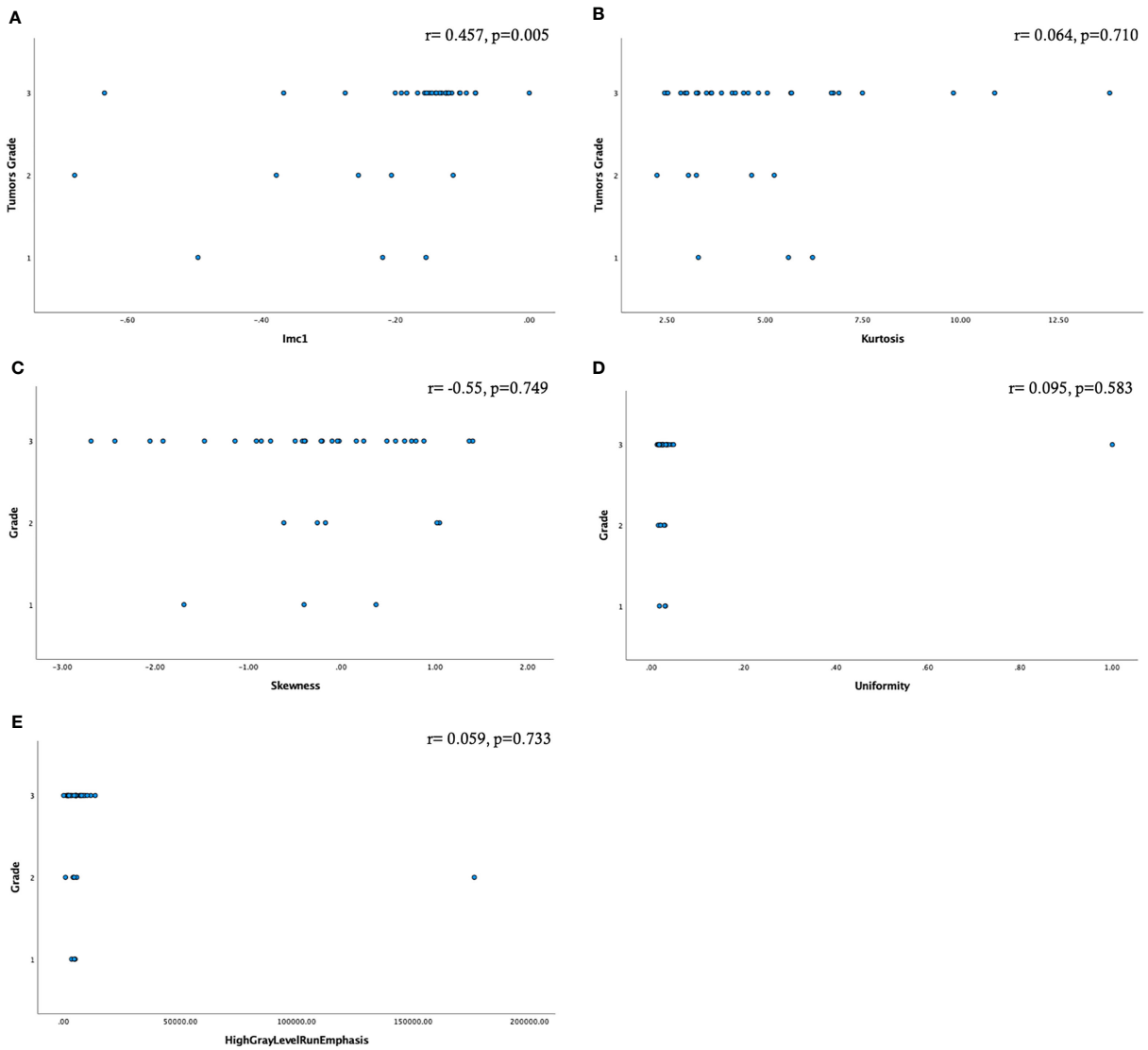


FIGURE 3
Scatter plots showing the relationship among the radiomic features selected by factor analysis and the tumor grade **(A–E)**. Only the feature Imc1 showed a moderate statistically significant correlation **(A)**.

Then, the lower spatial resolution of DWI in contrast to conventional imaging should be considered. Nevertheless, as previously mentioned, analyses showed a high repeatability and several *ex vivo* and *in vivo* studies demonstrated a high reliability of radiomics using this type of sequence (31, 32).

In conclusion, the radiomic feature Imc1 acts as a predictor of metastatic spread in patients with soft tissue sarcomas and correlates with the tumor grade. Further studies on a larger sample and including delta-radiomics analyses at follow-up are expected to provide new insights on the potential impact of this evidence on the therapeutic and overall management of this rare group of tumors.

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 reviewed and approved by the Ethics Committee of the Veneto Institute of Oncology (Approval No. 16336/21 CESC-IOV) on 1 September 2021.

Author contributions

All authors significantly contributed to this study and approved the final version of the manuscript. GC: study design, data analysis, and manuscript drafting. FG: data collection and review and editing of the manuscript. DP: data collection and review and editing of the manuscript. MS: data interpretation and review and editing of the manuscript. BA: assisted in data collection and review and editing of the manuscript. RM: study design and review and editing of the manuscript. SR: review and editing of the manuscript and study coordination. All authors contributed to the article and approved the submitted version.

Conflict of interest

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

Publisher's note

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

References

1. Von Mehren M, Randall LR, Benjamin RS, Boles S, Bui MM, Ganjoo KN, et al. Soft tissue sarcoma, version 2. 2018. NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw* (2018) 16:5. doi: 10.6004/jnccn.2018.0025

2. Robinson E, Bleakney RR, Ferguson PC, O'Sullivan B. Oncodiagnosis panel: 2007. multidisciplinary management of soft-tissue sarcoma. *RadioGraphics* (2008) 28:7. doi: 10.1148/rg.287085167

3. Varma DGK. Imaging of soft-tissue sarcomas. *Curr Oncol Rep* (2000) 2:6. doi: 10.1007/s11912-000-0100-2

4. Subhawong TK, Jacobs MA, Fayad LM. Diffusion-weighted MR imaging for characterizing musculoskeletal lesions. *RadioGraphics* (2014) 34:5. doi: 10.1148/rg.345140190

5. Li X, Wang Q, Dou Y, Zhang Y, Tao J, Yang L, et al. Soft tissue sarcoma: can dynamic contrast-enhanced (DCE) MRI be used to predict the histological grade? *Skeletal Radiol* (2020) 49:11. doi: 10.1007/s00256-020-03491-z

6. Lee JH, Yoon YC, Seo SW, Choi YL, Kim HS. Soft tissue sarcoma: DWI and DCE-MRI parameters correlate with ki-67 labeling index. *Eur Radiol* (2020) 30:2. doi: 10.1007/s00330-019-06445-9

7. Subhawong TK, Wang X, Durand DJ, Jacobs MA, Carrino JA, Machado AJ, et al. Proton MR spectroscopy in metabolic assessment of musculoskeletal lesions. *AJR* (2012) 198:1. doi: 10.2214/AJR.11.6505

8. Del Grande F, Subhawong TK, Weber K, Aro M, Mugera C, Fayad LM. Detection of soft-tissue sarcoma recurrence: Added value of functional MR imaging techniques at 3.0 T. *Radiology* (2014) 271:2. doi: 10.1148/radiol.13130844

9. Soldatos T, Ahlawat S, Montgomery E, Chalian M, Jacobs MA, Fayad LM. Multiparametric assessment of treatment response in high-grade soft-tissue sarcomas with anatomic and functional MR imaging sequences. *Radiology* (2016) 278:3. doi: 10.1148/radiol.2015142463

10. Messina C, Bignone R, Bruno A, Bruno A, Bruno F, Calandri M, et al. Diffusion-weighted imaging in oncology: An update. *Cancers* (2020) 12:6. doi: 10.3390/cancers12061493

11. Giraudo C, Karanikas G, Weber M, Raderer M, Jager U, Simonitsch-Klupp I, et al. Correlation between glycolytic activity on [18F]-FDG-PET and cell density on diffusion-weighted MRI in lymphoma at staging. *J Magn Reson Imaging* (2018) 47:5. doi: 10.1002/jmri.25884

12. Orsatti G, Zucchetta P, Varotto A, Crimì F, Weber M, Cecchin D, et al. Volumetric histograms-based analysis of apparent diffusion coefficients and standard uptake values for the assessment of pediatric sarcoma at staging: preliminary results of a PET/MRI study. *Radiol Med* (2021) 126:6. doi: 10.1007/s11547-021-01340-0

13. Tian L, Zhang D, Bao S, Nie P, Hao D, Liu Y, et al. Radiomics-based machine-learning method for prediction of distant metastasis from soft-tissue sarcomas. *Clin Radiol* (2021) 76:2. doi: 10.1016/j.crad.2020.08.038

14. Tagliafico AS, Bignotti B, Rossi F, Valdora F, Martinoli C. Local recurrence of soft tissue sarcoma: a radiomic analysis. *Radiol Oncol* (2019) 53:3. doi: 10.2478/raon-2019-0041

15. Corino VDA, Montin E, Messina A, Casali PG, Gronchi A, Marchianò A, et al. Radiomic analysis of soft tissues sarcomas can distinguish intermediate from high-grade lesions. *J Magn Reson Imaging* (2018) 47:3. doi: 10.1002/jmri.25791

16. Xu W, Hao D, Hou F, Zhang D, Wang H. Soft tissue sarcoma: Preoperative MRI-based radiomics and machine learning may be accurate predictors of histopathologic grade. *AJR* (2020) 215:4. doi: 10.2214/AJR.19.22147

17. Crombe A, Saut O, Guigui J, Italiano A, Buy X, Kind M. Influence of temporal parameters of DCE-MRI on the quantification of heterogeneity in tumor vascularization. *J Magn Reson Imaging* (2019) 50:6. doi: 10.1002/jmri.26753

18. Gao Y, Kalbasi A, Hsu W, Ruan D, Fu J, Shao J, et al. Treatment effect prediction for sarcoma patients treated with preoperative radiotherapy using radiomics features from longitudinal diffusion- weighted MRI. *Phys Med Biol* (2020) 65:17. doi: 10.1088/1361-6560/ab9e58

19. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess* (1994) 6:4. doi: 10.1037/1040-3590.6.4.284

20. Vallieres M, Freeman CR, Skamene SR, Naqa IE. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Phys Med Biol* (2015) 60:14. doi: 10.1088/0031-9155/60/14/5471

21. Tietz E, Truhn D, Mueller-Franzes G, Berres ML, Hamesch K, Lang SA, et al. A radiomics approach to predict the emergence of new hepatocellular carcinoma in computed tomography for high-risk patients with liver cirrhosis. *Diagnostics* (2021) 11:9. doi: 10.3390/diagnostics11091650

22. Liao CY, Lee CC, Yang HC, Chen CJ, Chung WY, Wu HS, et al. Enhancement of radiosurgical treatment outcome prediction using MRI radiomics in patients with non-small cell lung cancer brain metastases. *Cancers* (2021) 13:16. doi: 10.3390/cancers13164030

23. Wang X, Wan Q, Chen H, Li Y, Li X. Classification of pulmonary lesion based on multiparametric MRI: utility of radiomics and comparison of machine learning methods. *Eur Radiol* (2020) 30:8. doi: 10.1007/s00330-020-06768-y

24. Shin J, Lin JS, Huh YM, Kim JH, Hyung WJ, Chung JJ, et al. A radiomics-based model for predicting prognosis of locally advanced gastric cancer in the preoperative setting. *Sci Rep* (2021) 11:1. doi: 10.1038/s41598-021-81408-z

25. Jiang Y, Che S, Ma S, Liu X, Guo Y, Liu A, et al. Radiomic signature based on CT imaging to distinguish invasive adenocarcinoma from minimally invasive adenocarcinoma in pure ground-glass nodules with pleural contact. *Cancer Imaging* (2021) 21:1. doi: 10.1186/s40644-020-00376-1

26. Bernatowicz K, Grussu F, Ligero M, Garcia A, Delgado E, Perez-Lopez R. Robust imaging habitat computation using voxel-wise radiomics features. *Sci Rep* (2021) 11:1. doi: 10.1038/s41598-021-99701-2

27. Hu P, Chen L, Zhou Z. Machine learning in the differentiation of soft tissue neoplasms: Comparison of fat-suppressed T2WI and apparent diffusion coefficient (ADC) features-based models. *J Digital Imaging* (2021) 34:5. doi: 10.1007/s10278-021-00513-7

28. Crombe A, Le Loarer F, Sitbon M, Italiano A, Stoeckle E, Buy X, et al. Can radiomics improve the prediction of metastatic relapse of myxoid/round cell liposarcomas? *Eur Radiol* (2020) 30:5. doi: 10.1007/s00330-019-06562-5

29. Crombe A, Fadli D, Buy X, Italiano A, Saut O, Kind M. High-grade soft-tissue sarcomas: Can optimizing dynamic contrast-enhanced MRI postprocessing improve prognostic radiomics models? *J Magn Reson Imaging* (2020) 52:1. doi: 10.1002/jmri.27040

30. Crombe A, Perier C, Kind M, De Senneville BD, Le Loarer F, Italiano A, et al. T2-based MRI delta-radiomics improve response prediction in soft-tissue sarcomas treated by neoadjuvant chemotherapy. *J Magn Reson Imaging* (2019) 50:2. doi: 10.1002/jmri.26589

31. Dreher C, Kuder TA, Koenig F, Mlynarska-Bujny A, Tenconi C, Paech D, et al. Radiomics in diffusion data: a test-retest, inter- and intra-reader DWI phantom study. *Clin Radiol* (2020) 75:10. doi: 10.1016/j.crad.2020.06.024

32. Peerlings J, Woodruff HC, Winfield JM, Ibrahim A, Van Beers BB, Heerschap A, et al. Stability of radiomics features in apparent diffusion coefficient maps from a multi-centre test-retest trial. *Sci Rep* (2019) 9:1. doi: 10.1038/s41598-019-41344-5

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 24 October 2022
DOI 10.3389/fonc.2022.896593

OPEN ACCESS

EDITED BY
Kyung Hyun Sung,
UCLA Health System, United States

REVIEWED BY
Guoqiang Shao,
Nanjing Medical University, China
Jianhua Gong,
Chinese Academy of Medical Sciences and Peking Union Medical College, China

*CORRESPONDENCE
Hui Wang
wanghui@xinhuamed.com.cn
Suyun Chen
nuyus@outlook.com

†These authors share last authorship

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

RECEIVED 15 March 2022
ACCEPTED 20 September 2022
PUBLISHED 24 October 2022

CITATION
Li C, Wang S, Li C, Yin Y, Feng F, Fu H, Wang H and Chen S (2022) Improved risk stratification by PET-based intratumor heterogeneity in children with high-risk neuroblastoma. *Front. Oncol.* 12:896593. doi: 10.3389/fonc.2022.896593

Improved risk stratification by PET-based intratumor heterogeneity in children with high-risk neuroblastoma

Chao Li[1], Shaoyan Wang[1], Can Li[2], Yafu Yin[1], Fang Feng[1], Hongliang Fu[1], Hui Wang[1*†] and Suyun Chen[1*†]

[1]Department of Nuclear Medicine, Xinhua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China, [2]Department of Pathology, Xinhua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, China

Purpose: The substratification of high-risk neuroblastoma is challenging, and new predictive imaging biomarkers are warranted for better patient selection. The aim of the study was to evaluate the prognostic role of PET-based intratumor heterogeneity and its potential ability to improve risk stratification in neuroblastoma.

Methods: Pretreatment ^{18}F-FDG PET/CT scans from 112 consecutive children with newly diagnosed neuroblastoma were retrospectively analyzed. The primary tumor was segmented in the PET images. SUVs, volumetric parameters including metabolic tumor volume (MTV) and total lesion glycolysis (TLG), and texture features were extracted. After the exclusion of imaging features with poor and moderate reproducibility, the relationships between the imaging indices and clinicopathological factors, as well as event-free survival (EFS), were assessed.

Results: The median follow-up duration was 33 months. Multivariate analysis showed that PET-based intratumor heterogeneity outperformed clinicopathological features, including age, stage, and MYCN, and remained the most robust independent predictor for EFS [training set, hazard ratio (HR): 6.4, 95% CI: 3.1–13.2, $p < 0.001$; test set, HR: 5.0, 95% CI: 1.8–13.6, $p = 0.002$]. Within the clinical high-risk group, patients with a high metabolic heterogeneity showed significantly poorer outcomes (HR: 3.3, 95% CI: 1.6–6.8, $p = 0.002$ in the training set; HR: 4.4, 95% CI: 1.5–12.9, $p = 0.008$ in the test set) compared to those with relatively homogeneous tumors. Furthermore, intratumor heterogeneity outran the volumetric indices (MTVs and TLGs) and yielded the best performance of distinguishing high-risk patients with different outcomes with a 3-year EFS of 6% vs. 47% ($p = 0.001$) in the training set and 9% vs. 51% ($p = 0.004$) in the test set.

Conclusion: PET-based intratumor heterogeneity was a strong independent prognostic factor in neuroblastoma. In the clinical high-risk group, intratumor heterogeneity further stratified patients with distinct outcomes.

KEYWORDS
pediatric, neuroblastoma, ^{18}F-FDG, PET/CT, intratumor heterogeneity, radiomics

Introduction

Neuroblastoma is the most common extracranial solid tumor in children and is remarkable for its heterogeneity (1). Risk stratification using a combination of clinical and biological factors, such as age at diagnosis, stage, histology, and MYCN status, is of paramount importance to effectively inform therapeutic approaches. At the time of presentation, about 60% of children are classified as high risk (2). The incorporation of intensive multimodality therapy has increased the 5-year survival for high-risk neuroblastoma from less than 20% to ~50% (3). However, a notable subset of patients do not respond to induction therapy and have a dismal outcome, with a long-term survival of less than 15% (4). The improved outcome for the survivors has come at a cost of significant early or long-term toxicity. The early identification of these different subsets of patients may facilitate a more precisely tailored treatment, which remains an important unmet need.

Intratumor heterogeneity, resulting from subclonal genetic diversity within a tumor, manifests in spatial variation in stromal architecture and consumption of oxygen and glucose (5). It has been associated with poor prognosis and predisposes patients to inferior response to anticancer therapies (6). Medical images can depict the spatial heterogeneity in individual tumors andquantify the overall functional characteristics. Various approaches for the assessment of intratumor heterogeneity in PET images have been investigated, including simple visual analysis, histogram quantifying voxel distributions, and texture features quantifying spatial complexity (7, 8). A growing body of evidence suggests that PET-based intratumor heterogeneity might have predictive or prognostic value in various malignancies (9, 10).

^{123}I-meta-iodobenzylguanidine (mIBG) scan has been the main imaging modality for neuroblastoma. For high-risk diseases, however, the limited prognostic value of pretreatment mIBG score was reported (11, 12). On the other hand, ^{18}F-FDG PET/CT is increasingly used in neuroblastoma, particularly in tumors not taking up mIBG. SUVmax has been reported to correlate with MYCN amplification (13) and may serve as a prognostic biomarker in neuroblastoma (14, 15). Volumetric parameters derived from ^{18}F-FDG PET, including metabolic tumor volume (MTV) and total lesion glycolysis (TLG), were previously reported as significant prognostic factors in neuroblastoma (16). To date, there is limited evidence regarding the role of intratumor metabolic heterogeneity in neuroblastoma.

Our key objectives were to investigate the prognostic role of PET-based intratumor heterogeneity and whether it could be used to further risk-stratify neuroblastoma.

Materials and methods

Patients

This study included 129 consecutive pediatric patients with histologically proven neuroblastoma between October 2011 and September 2020. The inclusion criteria were as follows: 1) newly diagnosed neuroblastoma with no previous anticancer treatment, 2) underwent baseline ^{18}F-FDG PET/CT scan, 3) not accompanied by other malignancies, and 4) at least 6 months of follow-up. Patients were excluded if they had primary intracranial neuroblastoma, ganglioneuroma, no predominant primary tumor site, refused treatment, or had received chemotherapy before the PET scan (Figure 1). Clinicopathological prognostic indices, such as age, stage, risk stratification, MYCN, lactate dehydrogenase (LDH), and ferritin, were collected. This retrospective study was approved by the institutional review board, and the requirement for informed consent was waived.

PET/CT imaging

^{18}F-FDG was administered at a dose of 5.18 MBq/kg after at least 4–6 h of fasting. PET/CT scans from the skull to the proximal thigh were acquired about 60 min after injection using a Biograph mCT-64 scanner (Siemens). When metastasis was suspected to involve the extremities, imaging from the vertex to the toes including the arms was performed. Chloral hydrate sedation (50 mg/kg) was used 30 min before scanning for children unable to follow instructions. PET images were reconstructed using 3D ordered subset expectation maximization (3 iterations, 24 subsets). CT scans were acquired with 100-kV tube voltage, automated tube current modulation, 3-mm slice thickness, and a pitch of 1.5.

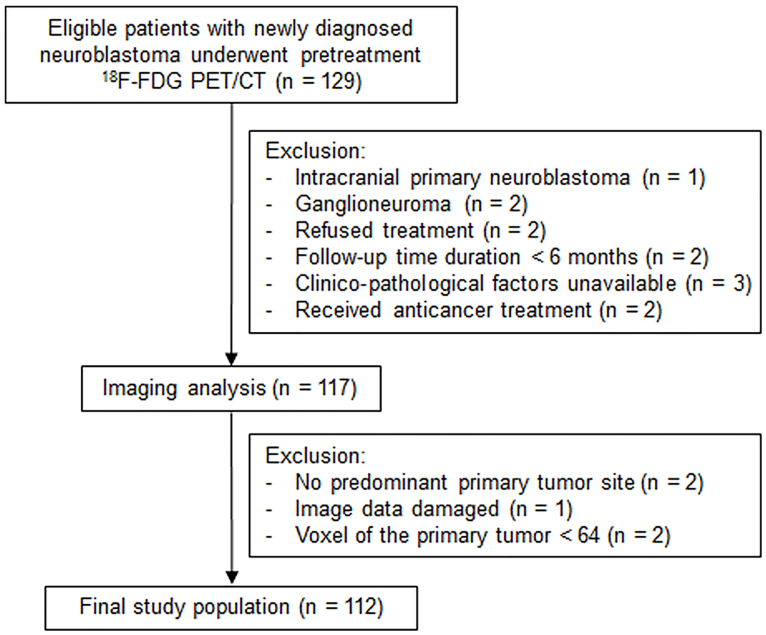


FIGURE 1
Flowchart shows study population selection, with exclusion criteria.

Imaging segmentation and feature extraction

Segmentation and feature extraction were performed using the LIFEx software (Version 6.31, http://www.lifexsoft.org). To investigate the voxel relationships inside the entire tumor, volumes of interest (VOIs) covering the whole primary tumor were delineated manually in PET images by a nuclear medicine physician with more than 11 years of PET/CT experience without knowledge of clinical information. In some cases, the primary tumor fused with the metastatic lesions and was delineated with reference to recent contrast-enhanced CT or MRI images. As PET has relatively large voxels compared with CT and MRI, each VOI must contain at least 64 contiguous voxels according to the LIFEx user guide. Two patients were excluded due to small voxels. Imaging indices were computed after a resampling step using 64 bins (size bin of 0.3) without spatial resampling. MTV and TLG with a threshold of 41% of SUVmax (MTV41%, TLG41%), which has been reported to correspond best with the actual dimensions of the tumor for tumor boundary delineation (17), were extracted from the same VOIs.

Clinical endpoints and risk stratification

Event-free survival (EFS) was calculated as the time from the start date of cancer treatment to the date of relapse, progression, or death from any cause. All the patients received risk-adapted treatment according to the Chinese Children Cancer Group-NB-2009/2014. The risk categorization schema was consistent with the Children's Oncology Group protocol (2). Briefly, patients were classified into low-, intermediate-, and high-risk categories based on age, stage, and other histopathological factors. High-risk disease was defined as ≥18 months of age and either disseminated disease or localized disease with unfavorable markers, such as MYCN amplification.

Statistical analysis

To determine robust features, half of the patients were selected randomly and segmented independently by another nuclear medicine physician with 6 years of PET/CT experience. We evaluated the reproducibility of features using a two-way random, absolute agreement intraclass correlation coefficient (ICC). Using the lower bounds of the 95% confidence interval (CI) of the ICC value ($ICC_{lb95\%}$) (18), the reproducibility of each feature was categorized as follows: poor, $ICC_{lb95\%}$ <0.50; moderate, $ICC_{lb95\%}$ of 0.50–0.75; good, $ICC_{lb95\%}$ of 0.75–0.90; and excellent, $ICC_{lb95\%}$ ≥0.90. Robust features with good or excellent reproducibility were qualified for further analysis.

The Mann–Whitney *U* test and chi-squared test were used for comparing variables between groups. The Benjamini–Hochberg stepwise method was performed to control the false discovery rate and adjusted *p*-values were calculated. Correlations among the parameters were determined by the Pearson and Spearman rank

correlation. To avoid redundancy, factors with poorer predictive validity in the pairs of indices that showed correlation coefficient (r) ≥0.8 were omitted (19, 20). Logistic regression analyses with forward selection were performed to evaluate the relationship between imaging indices and MYCN amplification. Then, the entire cohort was randomly split into a training set (n = 77) and a test set (n = 35). Prognostic factors were identified by univariate and multivariable Cox regression analyses in the training set and then validated in the test set. Receiver-operating characteristic curve (ROC) analyses and the Youden index were used to determine the optimal cutoff values. Survival estimates were evaluated by the Kaplan–Meier analysis and log-rank test. All statistical analyses were performed using SPSS 25.0 (IBM, Chicago, IL, USA), except that the adjusted p-values were obtained on R software (Version 4.0.3, http://www.r-project.org/). A two-sided p-value <0.05 was considered statistically significant.

Results

Patient characteristics

As a result, a total of 112 children were identified. The patient characteristics are summarized in Table 1. There were 39 girls (median age 34 months, range 1–153 months) and 73 boys (median age 36 months, range 2–150 months). Ninety patients had neuroblastoma and 22 had ganglioneuroblastoma (GNB). Most of them presented disseminated disease (2 with stage 4S, 79 with stage 4). With a median follow-up of 33 months, 51 disease relapse/progression and 34 deaths occurred. The 3-year EFS rate was 47%.

All the patients had an FDG-avid primary tumor with a median SUVmax of 5.8 (range 1.6–26.5). Seven tumors had SUVmax lower than 2.5 (1.6–2.4), all of which were higher than the liver

TABLE 1 Patient characteristics.

Characteristics	Total (n = 112) No. (%)	Training set (n = 77) No. (%)	Test set (n = 35) No. (%)	p-value
Median age (months)	37 ± 26	37 ± 28	37 ± 22	0.615
≥18 months	84 (75%)	58 (75%)	26 (74%)	0.906
Sex				0.438
Female	39 (35%)	25 (32%)	14 (40%)	
Male	73 (65%)	52 (68%)	21 (60%)	
Pathology				0.081
GNB intermixed/well-differentiated	16 (14%)	8 (10%)	8 (23%)	
GNB nodular, neuroblastoma	96 (86%)	69 (90%)	27 (77%)	
MYCN (n = 90)				0.320
Non-amplified	70 (78%)	48/64 (75%)	22/26 (85%)	
Amplified	20 (22%)	16/64 (25%)	4/26 (15%)	
Location				0.352
Abdominal and pelvic	92 (82%)	65 (84%)	27 (77%)	
Others	20 (18%)	12 (16%)	8 (23%)	
Stage				0.889
1, 2, 3, 4S	33 (29%)	23 (30%)	10 (29%)	
4	79 (71%)	54 (70%)	25 (71%)	
Risk stratification				0.910
Low	7 (6%)	5 (6%)	2 (6%)	
Intermediate	26 (23%)	17 (22%)	9 (26%)	
High	79 (71%)	55 (71%)	24 (69%)	
Laboratory tests				
Ferritin ≥92 ng/ml[a]	63/96 (66%)	45/64 (70%)	18/32 (56%)	0.171
LDH ≥587 U/L[a]	51/100 (51%)	36/66 (55%)	15/34 (44%)	0.323
Metabolic parameters				
SUVmax	6.3 ± 3.5	6.1 ± 3.2	6.7 ± 4.3	0.488
SUVpeak	4.6 ± 2.4	4.5 ± 2.1	4.8 ± 3.1	0.713
Endpoints				
Progression	51 (46%)	33 (43%)	18 (51%)	0.399
Death	34 (30%)	21 (27%)	13 (37%)	0.292

LDH, lactate dehydrogenase; GNB, ganglioneuroblastoma.
[a]Cutoff values for ferritin and LDH were set according to INRG (2).

background. High-risk neuroblastoma showed significantly higher FDG uptake (SUVmax and SUVpeak, $p < 0.001$) and volumetric values (MTV, TLG, and TLG41%, all $p < 0.001$; MTV41%, $p = 0.040$) than those with non-high-risk disease (Figure S1).

Imaging feature selection

Sixty-nine imaging features were obtained per VOI. The steps used to reduce feature dimension are summarized in Table S1. The ICC revealed that most of the imaging features could be reproduced well (Table S2). Fifty-one out of the 59 feature pairs had $ICC_{lb95\%}$ ≥0.75 (excellent reproducibility in 39 and good reproducibility in 12) and were qualified for subsequent analyses.

Imaging model for predicting MYCN amplification

MYCN status was available in 90 patients and was amplified in 20 patients. The majority of imaging features (44/51) were significantly different between the MYCN-amplified and the non-amplified groups (Table S3). After false discovery correction, 34 remained statistically significant. For example, FDG uptake was significantly higher in the MYCN-amplified tumor (SUVmax: 7.9, 95% CI: 6.7–9.9 vs. 5.1, 95% CI: 4.9–6.6, $p < 0.001$, adjusted $p = 0.005$).

The ROC analysis showed that all of the above 34 features had AUCs higher than 0.7 to predict MYCN amplification. Histogram_Kurtosis, which reflects the shape of the histogram distribution relative to a normal distribution, yielded the highest AUC of 0.853 ($p < 0.001$). After multicollinearity reduction, nine features were entered into multivariate logistic regression analysis. A radiomic model composed of two features [Histogram_Kurtosis and gray-level non-uniformity from gray-level zone length matrix (GLZLM_GLNU), which reflects the non-uniformity of the gray levels of the homogeneous zones in 3D] was built subsequently, resulting in an AUC of 0.871 (Figure 2, $p < 0.001$) with the following equation:

Predicted probability = EXP (−0.287 − 0.228 × Histogram_Kurtosis + 0.021 × GLZLM_GLNU)/(1 + EXP (−0.287 − 0.228 × Histogram_Kurtosis + 0.021 × GLZLM_GLNU).

Development of rad-risk to predict EFS

The distribution of key variables including age, stage, MYCN, and conventional metabolic parameters was similar between the training and test sets (Table 1). In the training set,

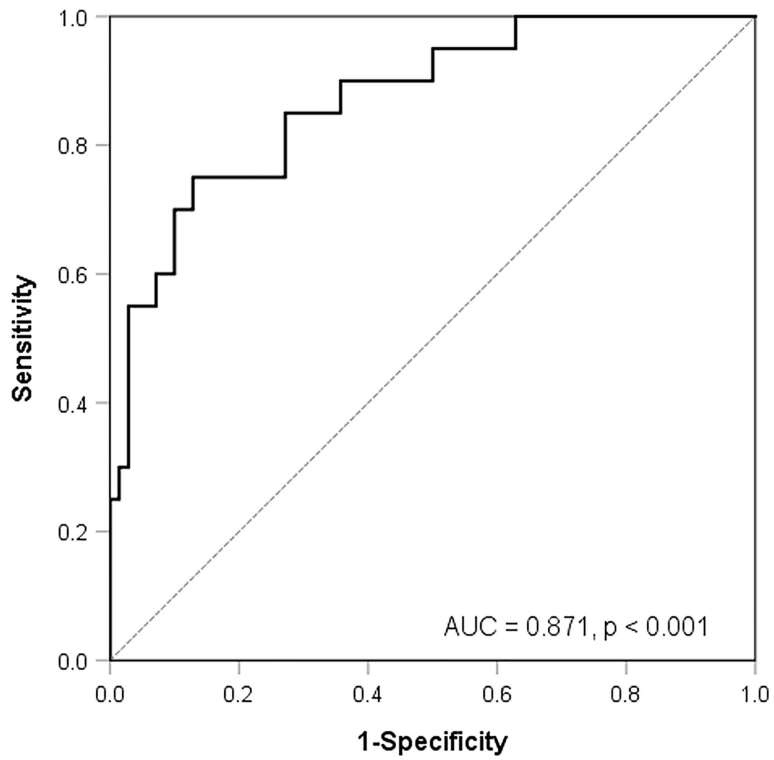


FIGURE 2
Receiver-operating characteristic curve analysis for the prediction of MYCN amplification according to a model composed of two texture features.

univariate Cox regression analysis revealed that 14 first-order and 8 second-order indices correlated with EFS ($p < 0.05$). After feature dimension reduction, two first-order indices, namely, SUVmax and Histogram_Entropy reflecting the randomness of the voxel distribution, were retained (Table 2). Three second-order indices retained were as follows: one from the gray-level co-occurrence matrix (GLCM): GLCM_energy, which reflects the uniformity of gray-level voxel pairs; two from the gray-level run-length matrix (GLRLM), namely, the gray-level non-uniformity (GLRLM_GLNU), which measures the non-uniformity of the gray levels, and run-length non-uniformity (GLRLM_RLNU), which quantifies the non-uniformity of the length of the homogeneous runs. The AUCs for SUVmax, Histogram_Entropy, GLCM_energy, GLRLM_GLNU, and GLRLM_RLNU to predict progression were 0.611, 0.666, 0.645, 0.689, and 0.733, respectively. Multivariate Cox regression analyses revealed that GLRLM_RLNU with a cutoff value of 1,828 and Histogram_Entropy with a cutoff value of 3.3 outperformed other imaging indices and were significant to predict events. In addition, imaging features extraction was performed in the high-risk group separately, and the results are presented in the Supplementary Materials (Supplementary Data and Figure S2).

Then, patients in the training set and the test set were divided into three groups according to whether GLRLM_RLNU ≥1,828 and Histogram_Entropy ≥3.3: patients with neither of these two risk factors, those with either one of the factors, and those with both. Patients with neither or either one of these factors demonstrated similar survival curves both in the training set and test set (Figure 3A, $p = 0.697$; Figure 3B, $p = 0.383$) and, thus, were combined and categorized as low rad-risk. Patients with both factors had a significantly worse prognosis (training set, HR: 6.4, 95% CI: 3.1–13.2, $p < 0.001$; test set, HR: 5.0, 95% CI: 1.8–13.6, $p = 0.002$) and were categorized as high rad-risk. The 3-year EFS of low vs. high rad-risk was 71% vs. 6% in the training set and 69% vs. 17% in the test set, respectively (both $p < 0.001$).

Multivariate analysis

Clinicopathological factors including age, stage, MYCN, LDH, and ferritin significantly correlated with EFS in the training set (Table 2). As MYCN status, LDH, and ferritin were unavailable in several patients, we firstly integrated rad-risk with age and stage into the multivariate analysis. After adjustment for clinical covariates (Table 3), rad-risk obtained independent significance with HR of 4.3 (95% CI: 2.0–9.1, $p < 0.001$), while age showed marginal significance (HR: 6.8, 95% CI: 0.9–52.5, $p = 0.066$). After incorporating MYCN into the model, only rad-risk remained significant (HR: 8.8, 95% CI: 3.7–21.0, $p < 0.001$). Furthermore, we integrated LDH and ferritin into the multivariate analyses separately or together, and rad-risk was the only factor that retained significance.

Similarly, after adjusting for clinicopathological variables separately or together in the multivariate analysis, high rad-risk was confirmed to be the most significant factor to predict EFS in the test set (Table 3).

TABLE 2 Univariate Cox regression analyses for event-free survival.

Variables	Training set ($n = 77$)			Test set ($n = 35$)		
	HR	95% CI	*p*-value	HR	95% CI	*p*-value
Clinicopathological factors						
Age ≥18 months	13.2	1.8–96.8	0.011	2.1	0.6–7.3	0.246
Stage 4 vs. 1, 2, 3, 4S	4.5	1.6–12.9	0.005	2.5	0.6–10.9	0.224
MYCN amplification[a]	2.8	1.2–6.4	0.014	0.5	0.1–4.1	0.527
LDH ≥587[b]	3.0	1.3–6.9	0.008	3.3	1.3–8.6	0.015
Ferritin ≥92[c]	3.0	1.1–7.8	0.026	1.9	0.7–5.3	0.246
First-order imaging indices						
SUVmax ≥5.5	3.2	1.5–6.7	0.003	2.9	1.0–8.1	0.049
Histogram_Entropy ≥3.3	3.8	1.8–7.9	<0.001	3.5	1.1–10.6	0.029
Second-order imaging indices						
GLCM_Energy ≤0.02	3.5	1.7–7.1	0.001	2.1	0.8–5.6	0.146
GLRLM_RLNU ≥1,828	5.1	2.3–11.3	<0.001	5.7	2.0–16.4	0.001
GLRLM_GLNU ≥575	2.9	1.5–5.8	0.002	2.1	0.6–6.6	0.224

CI, confidence interval; HR, hazard ratio; GLCM, gray level co-occurrence matrix; GLNU, gray-level non-uniformity; GLRLM, gray-level run-length matrix; rad-risk, radiomic risk; RLNU, run-length non-uniformity.
[a]MYCN amplification status was available in 64 patients in the training set and 26 in the test set.
[b]LDH was available in 66 patients in the training set and 34 in the test set.
[c]Ferritin was available in 64 patients in the training set and 32 in the test set.

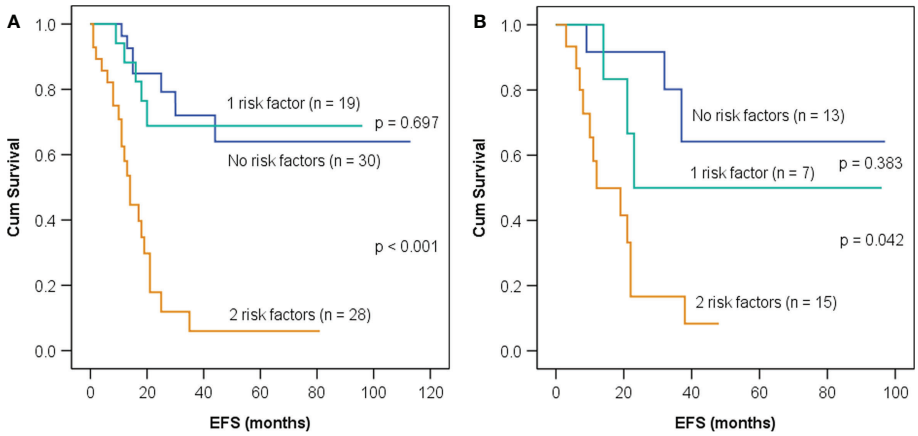


FIGURE 3
Kaplan–Meier event-free survival (EFS) curves in children with neuroblastoma having neither, one, or both imaging risk factors—GLRLM_RLNU ≥1,828 and Histogram_Entropy ≥3.3—in the training set **(A)** and the test set **(B)**.

Refinement of risk stratification in neuroblastoma

None of the seven patients with clinical low-risk diseases had a high rad-risk, and only 2 of the 26 patients with clinical intermediate-risk diseases had a high rad-risk, indicating that the majority of patients with clinical non-high-risk had a relatively homogeneous tumor. Due to limited cases with a high rad-risk in the clinical non-high-risk group, the significance of rad-risk in the risk substratification in this group could not be statistically analyzed.

Seventy-nine patients had high-risk neuroblastoma: 55 patients in the training set and 24 in the test set. We further evaluated whether adding rad-risk could refine risk stratification and compared it with volumetric indices, including MTV, MTV41%, TLG, and TLG41%. In the training set, ROC analyses were performed (Figure S3) and optimal cutoff values were determined to be 120 ml for MTV, 65 ml for MTV41%, 426 g for TLG, and 141 g for TLG41%, respectively.

TABLE 3 Multivariate Cox regression analyses for event-free survival.

Models	Training set			Test set		
	HR	95% CI	*p*-value	HR	95% CI	*p*-value
Multivariate model 1[a]	*n* = 77			*n* = 35		
Age ≥18 months	6.8	0.9–52.5	0.066	/	/	/
High rad-risk	4.3	2.0–9.1	<0.001	5.0	1.8–13.6	0.002
Multivariate model 2[b]	*n* = 64			*n* = 26		
High rad-risk	8.8	3.7–21.0	<0.001	6.7	1.7–26.0	0.007
Multivariate model 3[c]	*n* = 66			*n* = 34		
Age ≥18 months	6.5	0.8–49.9	0.074	/	/	/
LDH ≥587 U/L	/	/	/	3.2	1.1–8.7	0.026
High rad-risk	4.4	2.0–9.6	<0.001	4.8	1.7–13.6	0.003
Multivariate model 4[d]	*n* = 64			*n* = 32		
Age ≥18 months	7.0	0.9–54.1	0.062	/	/	/
High rad-risk	3.9	1.8–8.5	<0.001	4.4	1.6–12.2	0.004
Multivariate model 5[e]	*n* = 51			*n* = 25		
High rad-risk	8.3	3.2–21.4	<0.001	12.9	2.6–63.6	0.002

[a]Multivariate model 1 includes age, stage, and rad-risk (n = 112).
[b]Multivariate model 2 includes age, stage, MYCN, and rad-risk (n = 90).
[c]Multivariate model 3 includes age, stage, LDH, and rad-risk (n = 100).
[d]Multivariate model 4 includes age, stage, ferritin, and rad-risk (n = 96).
[e]Multivariate model 5 includes age, stage, MYCH, LDH, ferritin, and rad-risk (n = 76).

As shown in Figure 4, all of the five imaging indices significantly correlated with EFS in the training set. The 3-year EFS for patients with high vs. low MTV, MTV41%, TLG, and TLG41% were 17% vs. 61% ($p = 0.013$), 14% vs. 53% ($p = 0.015$), 18% vs. 40% ($p = 0.014$), and 16% vs. 61% ($p = 0.025$), respectively. However, the volumetric features failed to retain significance in the test set, except TLG (Figure 5). Rad-risk yielded the best performance to distinguish high-risk patients with different outcomes, with a 3-year EFS of 6% vs. 47% ($p = 0.001$, Figure 4E) in the training set and 9% vs. 51% ($p = 0.004$, Figure 5E) in the test set. High rad-risk was associated with a 2.3–3.4 times higher risk of progression (HR: 3.3, 95% CI: 1.6–6.8, $p = 0.002$ in the training set; HR: 4.4, 95% CI: 1.5–12.9, $p = 0.008$ in the test set). Two patients with high-risk neuroblastoma and a high or low rad-risk are presented in Figure 6.

Discussion

The substratification of high-risk neuroblastoma is challenging, and new predictive biomarkers are warranted for better patient selection. In this study, we confirmed that PET-based intratumor heterogeneity independently correlated with EFS in neuroblastoma both in the training set and the test set. It further improved the risk stratification in high-risk neuroblastoma, with a 3-year EFS of 6%–9% for the highly heterogeneous tumors compared to 47%–51% for the relatively homogeneous ones.

Radiomics, extracting quantitative features from medical images, has rapidly evolved throughout these years. Compared to histological biopsy only capturing a small proportion of tumor tissue that could underestimate the mutational burden (21), a great advantage of radiomics is its ability to visualize the characteristics of the whole tumor non-invasively. It fully depicts spatial intratumor heterogeneity, which has been associated with poor prognosis. Studies showed that radiomic features in PET images correlated with heterogeneity at the cellular and genomic levels and had significant prognostic value in various malignancies (22–24). On the other hand, tumor necrosis results from increased tumor size, intratumor hypoxia, and nutrient deprivation. Both the presence and the extent of necrosis correlated with poor prognosis (25, 26). A necrotic core appears as non-FDG-avid area within the tumor. To investigate the spatial voxel relationships inside the entire tumor, the current study examined the VOI covering the whole mass (including the necrotic region) instead of putting a threshold of SUVmax on VOI segmentation. Our results partly confirmed previous studies that texture features significantly correlated with tumor size or volume (27, 28). We found four second-order indices reflecting tumor heterogeneity, including GLRLM_RLNU, highly correlated with volumetric indices. The latter is usually considered a reflection of tumor burden, while texture features correlate with tumor heterogeneity. A larger tumor results in a higher level of intratumor hypoxia and necrosis and leads to higher spatial complexity and heterogeneity (27). Hatt et al. found that radiomic heterogeneity quantification provided valuable complementary information for large tumors (>10 cm^3) (27). In our study, only one patient had a tumor volume less than 10 ml. The median volume for our whole cohort was 160 ml.

MYCN amplification is the most common genomic alteration in neuroblastoma, occurring in approximately 20% of the patients (29). It is highly associated with advanced stage and poor prognosis; thus, it has been incorporated into the mostly used neuroblastoma protocol. Radiomic models derived from contrast-enhanced CT have been shown to accurately predict MYCN amplification (30–32). Wu et al. (31) suggested that three-phase CT had a higher value than non-contrast CT scan, which could be explained by tumor angiogenesis promoted by MYCN amplification. Different from the density heterogeneous and vascular structure complexity depicted by CT scans, PET imaging semi-quantifies the glucose consumption of tumor parenchyma and reflects the uneven spatial distribution of cellular metabolism, hypoxia, necrosis, and proliferation. In our study, MYCN amplification occurred in 22% of the patients. Two patients had divergent MYCN results, potentially resulting from the heterogeneity of the tumor or underestimation of the mutational burden by biopsy bias. In line with a prior study by Sung et al. (33), we found that SUVmax and TLG had the potential to predict MYCN with AUCs of 0.771 and 0.776, respectively. However, histogram metrics and several second-order indices showed superior performance. Consequently, a radiomic model containing two PET features, namely, Histogram_Kurtosis and GLZLM_GLNU, was built and showed the strongest predictive power with an AUC of 0.871. Histogram_Kurtosis reflects the shape of the histogram distribution relative to a normal distribution. GLZLM_GLNU reflects the non-uniformity of the gray levels of the homogeneous zones in 3D. These two features have been proven to be promising parameters as biomarkers of tumor heterogeneity in various malignancies (34–36). A higher GLZLM_GLNU and a lower Histogram_Kurtosis, which indicate higher spatial heterogeneity, correlated with a higher possibility of MYCN amplification. Recently, Qian et al. reported that the radiomic signature containing both PET and CT features had a good ability to predict MYCN amplification (37). However, the majority of the features were obtained from wavelet transformed images, which decompose an image by using spatially oriented frequency filters but require intensive computation and may suffer from low reproducibility (38). Despite the methodology differences, we both showed that a high intratumor heterogeneity was associated with MYCN amplification. Since neuroblastoma is remarkably heterogeneous, which might require at least two solid tumor areas to provide a more accurate genomic diagnosis (39), texture features fully portraying the entire tumor might provide important complementary information about molecular profiling.

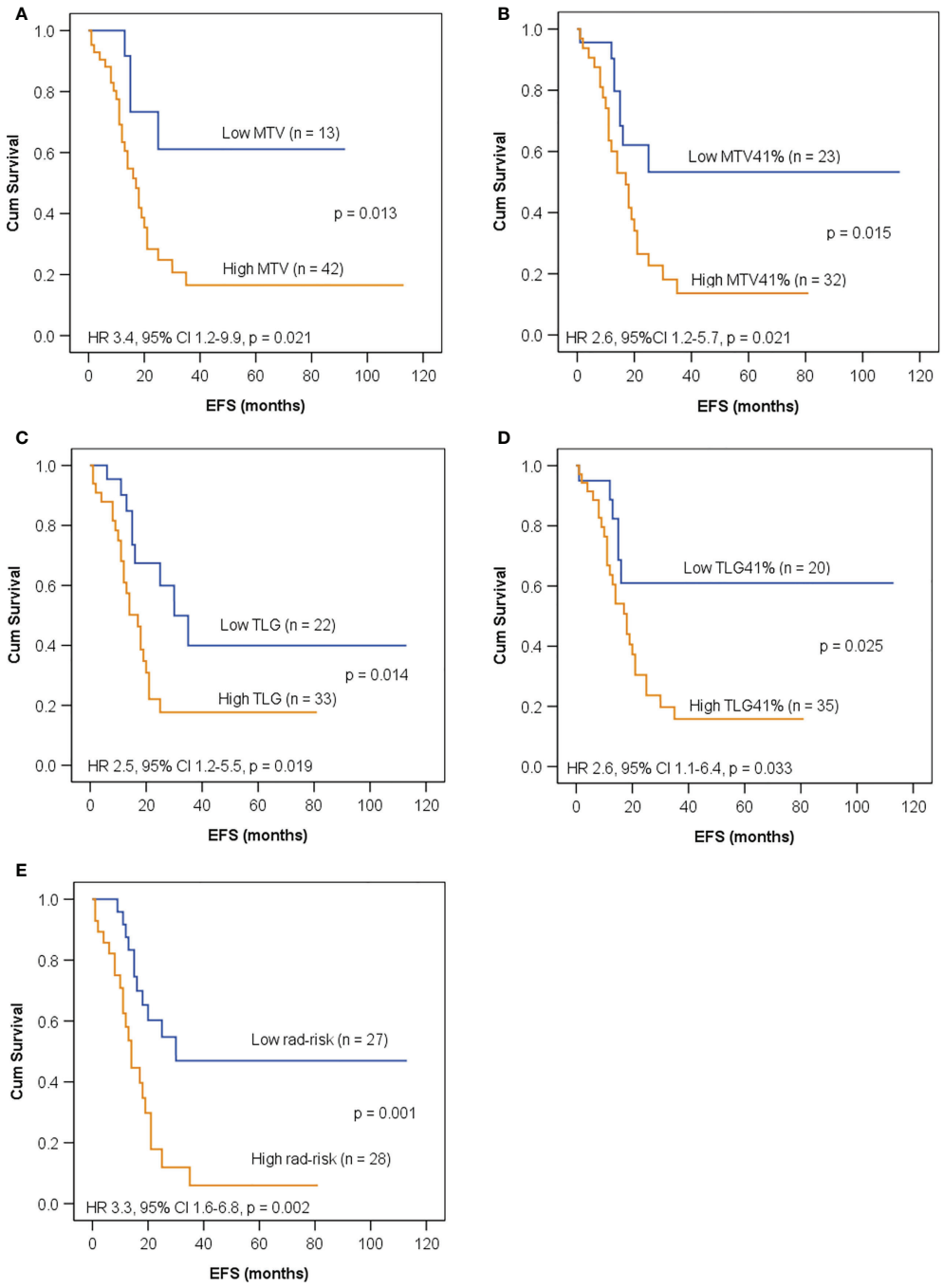


FIGURE 4
Kaplan–Meier curves for EFS in children with high-risk neuroblastoma in the training set according to **(A)** metabolic tumor volume (MTV) with a cutoff value of 120 ml; **(B)** MTV41% with a cutoff value of 65 ml; **(C)** total lesion glycolysis (TLG) with a cutoff value of 426 g; **(D)** TLG41% with a cutoff value of 141 g; **(E)** rad-risk.

Our second step was to evaluate whether intratumor heterogeneity could provide prognostic information in pretreatment neuroblastoma. A recent study reported that high intratumor metabolic heterogeneity on ^{18}F-FDG PET/CT was a strong prognostic factor in 38 children with newly diagnosed neuroblastoma (40), and it was the first report identifying metabolic heterogeneity as a prognostic biomarker of neuroblastoma. The authors used the area under the curve of the cumulative SUV-volume histograms (AUC-CSHs), which is a histogram-based first-order feature that describes the percentage

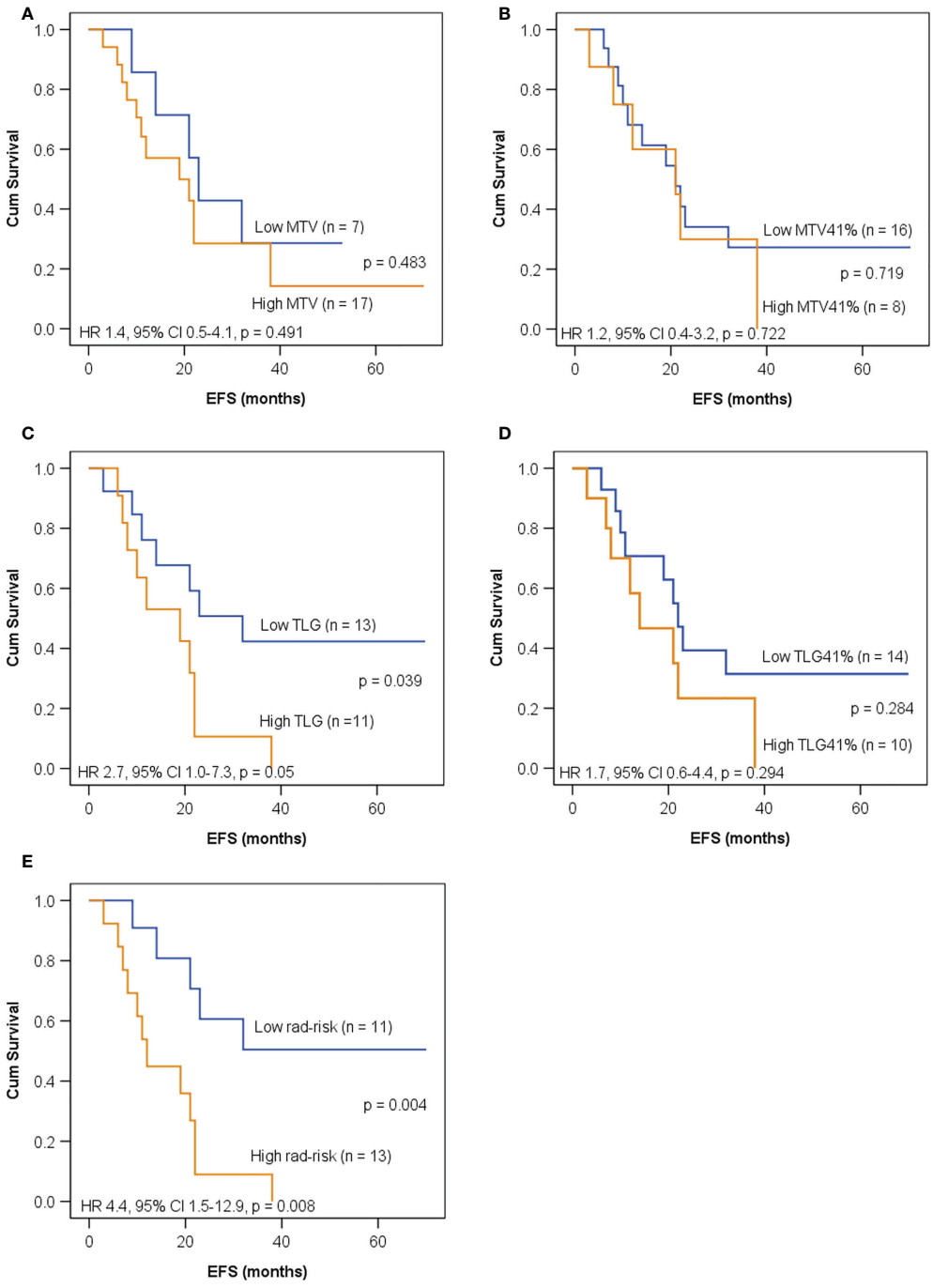


FIGURE 5
Kaplan–Meier curves for EFS in children with high-risk neuroblastoma in the test set according to **(A)** MTV with a cutoff value of 120 ml; **(B)** MTV41% with a cutoff value of 65 ml; **(C)** TLG with a cutoff value of 426 g; **(D)** TLG41% with a cutoff value of 141 g; **(E)** rad-risk.

of total tumor volume above the percent threshold of SUVmax, as an intratumor heterogeneity index. Lower AUC-CSH indicated higher heterogeneity of the tumor and poorer outcomes. Although the histogram analysis appears promising and simple, the major pitfalls of the histogram analysis are the lack of information on the spatial organization of tumors and that it is not straightforward which might lead to errors (34, 41). In another recently published study of 18 children with high-risk neuroblastoma, Fiz et al. demonstrated that intratumor heterogeneity on 18fluorine-dihydroxyphenylalanine (18F-DOPA) PET/CT was closely

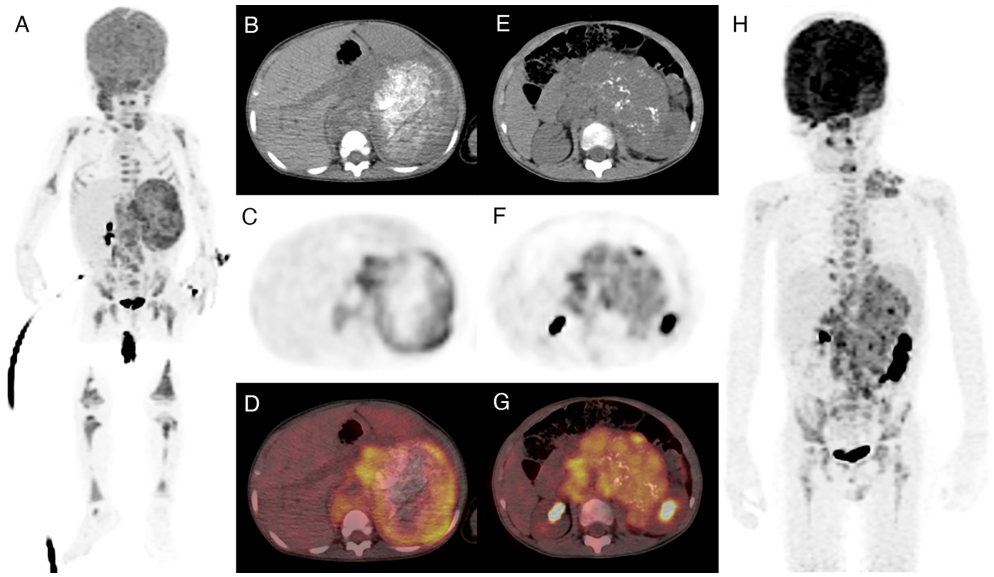


FIGURE 6
Two patients with a high-risk neuroblastoma and high or low intratumor heterogeneity. Both patients had amplified MYCN and stage 4 diseases. **(A–D)** A 20-month-old girl with a highly heterogeneous FDG uptake in the primary tumor (high rad-risk). She progressed 21 months after diagnosis. **(E–H)** A 5-year-old boy with a relatively homogeneous FDG uptake (low rad-risk). The patient remained recurrence free within 5 years of follow-up.

associated with metastatic burden and had certain prognostic value (42). In the current study, we further expanded that intratumor heterogeneity was a prognostic biomarker in neuroblastoma, with a much larger cohort and a higher order of texture analysis, which further improves quantitative histogram approaches by introducing the spatial dimension. Multivariate analysis identified GLRLM_RLNU and Histogram_Entropy as the independently significant predictors for EFS. GLRLM_RLNU gives the size of homogeneous runs for each gray level. A similar run length results in low values of GLRLM_RLNU. On the contrary, a high value is indicative of heterogeneity. Studies have reported that GLRLM_RLNU extracted from PET had the potential for predicting treatment response and prognosis (43, 44). On the other hand, Histogram_Entropy measures the randomness of voxel distribution and has been established as an important biomarker reflecting heterogeneity in various MRI and PET studies (43, 45). In accordance with previous studies (40, 42), we found that high rad-risk, defined as patients with both a high GLRLM_RLNU and a high Histogram_Entropy, indicating a high intratumor heterogeneity, was the most significant independent factor for EFS after adjusting for clinicopathological factors.

To further evaluate the ability of rad-risk in the refinement of risk stratification, we incorporated rad-risk into the existing risk stratification schema and compared it to volumetric indices. The results showed that the majority of patients with clinical non-high risk had a low rad-risk, indicating a relatively homogeneous tumor. Among high-risk neuroblastoma, rad-risk effectively distinguished patients with distinct outcomes both in the training and test sets. In addition, despite that intratumor heterogeneity highly correlated with MTV and TLG, rad-risk outperformed the volumetric indices and showed the highest ability to predict the outcome. These findings indicate that PET-based intratumor heterogeneity might have independent prognostic information, which may help substratify neuroblastoma patients for more refined risk-adapted treatment approaches in the future.

The limitations of this study are as follows: first, this is a retrospective study with a relatively small sample size in a single center. Second, ^{123}I-mIBG scans were not performed in our cohort, since ^{123}I-MIBG is not yet available in our country. The disadvantages of the ^{123}I-mIBG scan, including limited spatial resolution and lower sensitivity in soft tissue lesions or small lesions, limit its value in radiomic analysis in neuroblastoma. Future efforts in PET-based texture features using novel radiopharmaceuticals such as ^{18}F-fluorometaguanidine and ^{124}I-mIBG might yield important predictive or prognostic information. Third, this study evaluated the features of primary tumor and captured less information outside the primary site, such as metastatic lesions or metastatic burden, which could be of important prognostic value. An additional limitation is that no separate cohort was used for validation regarding the prediction of MYCN amplification due to the limited number of patients with amplified MYCN. A large cohort with external validation should be warranted in the future.

Conclusions

In summary, PET-based intratumor heterogeneity could serve as a powerful and non-invasive approach to predict MYCN amplification and survival outcome in newly diagnosed neuroblastoma, providing a potential approach to refine the risk stratification in children with high-risk diseases. Further validation with a larger cohort is required.

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

This retrospective study was approved by the Ethics Committee of Xin Hua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine and the requirement for informed consent was waived.

Author contributions

ChL, HW, and SC contributed to the conception and design of the study. ChL and SC organized the database. ChL and SC performed the statistical analysis. CaL, SW, YY, FF, and HF performed the data analysis and interpretation. ChL and SC wrote the first draft of the manuscript. CaL, SW, YY, FF, and HF wrote sections of the manuscript. ChL, HW, and SC edited the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This study has been supported by the National Natural Science Funds (81801731 and 81901775).

Conflict of interest

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

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.896593/full#supplementary-material

References

1. van Groningen T, Koster J, Valentijn LJ, Zwijnenburg DA, Akogul N, Hasselt NE, et al. Neuroblastoma is composed of two super-enhancer-associated differentiation states. *Nat Genet* (2017) 49(8):1261–6. doi: 10.1038/ng.3899

2. Cohn SL, Pearson AD, London WB, Monclair T, Ambros PF, Brodeur GM, et al. The international neuroblastoma risk group (INRG) classification system: An INRG task force report. *J Clin Oncol* (2009) 27(2):289–97. doi: 10.1200/JCO.2008.16.6785

3. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2021. *CA Cancer J Clin* (2021) 71(1):7–33. doi: 10.3322/caac.21654

4. Tadeo I, Berbegall AP, Castel V, Garcia-Miguel P, Callaghan R, Pahlman S, et al. Extracellular matrix composition defines an ultra-high-risk group of neuroblastoma within the high-risk patient cohort. *Br J Cancer* (2016) 115 (4):480–9. doi: 10.1038/bjc.2016.210

5. Junttila MR, de Sauvage FJ. Influence of tumour micro-environment heterogeneity on therapeutic response. *Nature* (2013) 501(7467):346–54. doi: 10.1038/nature12626

6. Meacham CE, Morrison SJ. Tumour heterogeneity and cancer cell plasticity. *Nature* (2013) 501(7467):328–37. doi: 10.1038/nature12624

7. Chicklore S, Goh V, Siddique M, Roy A, Marsden PK, Cook GJ. Quantifying tumour heterogeneity in 18F-FDG PET/CT imaging by texture analysis. *Eur J Nucl Med Mol Imaging* (2013) 40(1):133–40. doi: 10.1007/s00259-012-2247-0

8. van Velden FH, Cheebsumon P, Yaqub M, Smit EF, Hoekstra OS, Lammertsma AA, et al. Evaluation of a cumulative SUV-volume histogram method for parameterizing heterogeneous intratumoural FDG uptake in non-small cell lung cancer PET studies. *Eur J Nucl Med Mol Imaging* (2011) 38(9):1636–47. doi: 10.1007/s00259-011-1845-6

9. Yoo SH, Kang SY, Cheon GJ, Oh DY, Bang YJ. Predictive role of temporal changes in intratumoral metabolic heterogeneity during palliative chemotherapy in patients with advanced pancreatic cancer: A prospective cohort study. *J Nucl Med* (2020) 61(1):33–9. doi: 10.2967/jnumed.119.226407

10. Pinho DF, King B, Xi Y, Albuquerque K, Lea J, Subramaniam RM. Value of intratumoral metabolic heterogeneity and quantitative (18)F-FDG PET/CT parameters in predicting prognosis for patients with cervical cancer. *AJR Am J roentgenol* (2020) 214(4):908–16. doi: 10.2214/AJR.19.21604

11. Yanik GA, Parisi MT, Shulkin BL, Naranjo A, Kreissman SG, London WB, et al. Semiquantitative mIBG scoring as a prognostic indicator in patients with stage 4 neuroblastoma: A report from the children's oncology group. *J Nucl Med* (2013) 54(4):541–8. doi: 10.2967/jnumed.112.112334

12. Katzenstein HM, Cohn SL, Shore RM, Bardo DM, Haut PR, Olszewski M, et al. Scintigraphic response by 123I-metaiodobenzylguanidine scan correlates with event-free survival in high-risk neuroblastoma. *J Clin Oncol* (2004) 22(19):3909–15. doi: 10.1200/JCO.2004.07.144

13. Liu CJ, Lu MY, Liu YL, Ko CL, Ko KY, Tzen KY, et al. Risk stratification of pediatric patients with neuroblastoma using volumetric parameters of 18F-FDG and 18F-DOPA PET/CT. *Clin Nucl Med* (2017) 42(3):e142–e8. doi: 10.1097/RLU.0000000000001529

14. Lee JW, Cho A, Yun M, Lee JD, Lyu CJ, Kang WJ. Prognostic value of pretreatment FDG PET in pediatric neuroblastoma. *Eur J Radiol* (2015) 84 (12):2633–9. doi: 10.1016/j.ejrad.2015.09.027

15. Papathanasiou ND, Gaze MN, Sullivan K, Aldridge M, Waddington W, Almuhaideb A, et al. 18F-FDG PET/CT and 123I-metaiodobenzylguanidine imaging in high-risk neuroblastoma: Diagnostic comparison and survival analysis. *J Nucl Med* (2011) 52(4):519–25. doi: 10.2967/jnumed.110.083303

16. Li C, Zhang J, Chen S, Huang S, Wu S, Zhang L, et al. Prognostic value of metabolic indices and bone marrow uptake pattern on preoperative 18F-FDG PET/CT in pediatric patients with neuroblastoma. *Eur J Nucl Med Mol Imaging* (2018) 45(2):306–15. doi: 10.1007/s00259-017-3851-9

17. Boellaard R, Delgado-Bolton R, Oyen WJ, Giammarile F, Tatsch K, Eschner W, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging* (2015) 42(2):328–54. doi: 10.1007/s00259-014-2961-x

18. Peerawaranun P, Landier J, Nosten FH, Nguyen TN, Hien TT, Tripura R, et al. Intracluster correlation coefficients in the greater Mekong subregion for sample size calculations of cluster randomized malaria trials. *Malar J* (2019) 18 (1):428. doi: 10.1186/s12936-019-3062-x

19. Dissaux G, Visvikis D, Da-Ano R, Pradier O, Chajon E, Barillot I, et al. Pretreatment (18)F-FDG PET/CT radiomics predict local recurrence in patients treated with stereotactic body radiotherapy for early-stage non-small cell lung cancer: A multicentric study. *J Nucl Med* (2020) 61(6):814–20. doi: 10.2967/jnumed.119.228106

20. Mukaka MM. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Med J* (2012) 24(3):69–71.

21. Gerlinger M, Rowan AJ, Horswell S, Math M, Larkin J, Endesfelder D, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* (2012) 366(10):883–92. doi: 10.1056/NEJMoa1113205

22. Basler L, Gabrys HS, Hogan SA, Pavic M, Bogowicz M, Vuong D, et al. Radiomics, tumor volume, and blood biomarkers for early prediction of pseudoprogression in patients with metastatic melanoma treated with immune checkpoint inhibition. *Clin Cancer Res* (2020) 26(16):4414–25. doi: 10.1158/1078-0432.CCR-20-0020

23. 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

24. Moon SH, Kim J, Joung JG, Cha H, Park WY, Ahn JS, et al. Correlations between metabolic texture features, genetic heterogeneity, and mutation burden in patients with lung cancer. *Eur J Nucl Med Mol Imaging* (2019) 46(2):446–54. doi: 10.1007/s00259-018-4138-5

25. Minervini A, Di Cristofano C, Gacci M, Serni S, Menicagli M, Lanciotti M, et al. Prognostic role of histological necrosis for nonmetastatic clear cell renal cell carcinoma: Correlation with pathological features and molecular markers. *J Urol* (2008) 180(4):1284–9. doi: 10.1016/j.juro.2008.06.036

26. Rakheja R, Makis W, Tulbah R, Skamene S, Holcroft C, Nahal A, et al. Necrosis on FDG PET/CT correlates with prognosis and mortality in sarcomas. *AJR Am J Roentgenol* (2013) 201(1):170–7. doi: 10.2214/AJR.12.9795

27. Hatt M, Majdoub M, Vallières M, Tixier F, Le Rest CC, Groheux D, et al. 18F-FDG PET uptake characterization through texture analysis: investigating the complementary nature of heterogeneity and functional tumor volume in a multi-cancer site patient cohort. *J Nucl Med* (2015) 56(1):38–44. doi: 10.2967/jnumed.114.144055

28. Pfaehler E, Mesotten L, Zhovannik I, Pieplenbosch S, Thomeer M, Vanhove K, et al. Plausibility and redundancy analysis to select FDG-PET textural features in non-small cell lung cancer. *Med Phys* (2021) 48(3):1226–38. doi: 10.1002/mp.14684

29. Matthay KK, Maris JM, Schleiermacher G, Nakagawara A, Mackall CL, Diller L, et al. Neuroblastoma. *Nat Rev Dis Primers* (2016) 2:16078. doi: 10.1038/nrdp.2016.78

30. Chen X, Wang H, Huang K, Liu H, Ding H, Zhang L, et al. CT-based radiomics signature with machine learning predicts MYCN amplification in pediatric abdominal neuroblastoma. *Front Oncol* (2021) 11:687884. doi: 10.3389/fonc.2021.687884

31. Wu H, Wu C, Zheng H, Wang L, Guan W, Duan S, et al. Radiogenomics of neuroblastoma in pediatric patients: CT-based radiomics signature in predicting MYCN amplification. *Eur Radiol* (2021) 31(5):3080–9. doi: 10.1007/s00330-020-07246-1

32. Di Giannatale A, Di Paolo PL, Curione D, Lenkowicz J, Napolitano A, Secinaro A, et al. Radiogenomics prediction for MYCN amplification in neuroblastoma: A hypothesis generating study. *Pediatr Blood Cancer* (2021) 68 (9):e29110. doi: 10.22541/au.161144128.80677267/v1

33. Sung AJ, Weiss BD, Sharp SE, Zhang B, Trout AT. Prognostic significance of pretreatment (18)F-FDG positron emission tomography/computed tomography in pediatric neuroblastoma. *Pediatr Radiol* (2021) 51(8):1400–5. doi: 10.1007/s00247-021-05005-y

34. Just N. Improving tumour heterogeneity MRI assessment with histograms. *Br J Cancer* (2014) 111(12):2205–13. doi: 10.1038/bjc.2014.512

35. Acar E, Turgut B, Yigit S, Kaya G. Comparison of the volumetric and radiomics findings of 18F-FDG PET/CT images with immunohistochemical prognostic factors in local/locally advanced breast cancer. *Nucl Med Commun* (2019) 40(7):764–72. doi: 10.1097/MNM.0000000000001019

36. Toyama Y, Hotta M, Motoi F, Takanami K, Minamimoto R, Takase K. Prognostic value of FDG-PET radiomics with machine learning in pancreatic cancer. *Sci Rep* (2020) 10(1):17024. doi: 10.1038/s41598-020-73237-3

37. Qian L, Yang S, Zhang S, Qin H, Wang W, Kan Y, et al. Prediction of MYCN amplification, 1p and 11q aberrations in pediatric neuroblastoma *via* pre-therapy 18F-FDG PET/CT radiomics. *Front Med (Lausanne)* (2022) 9:840777. doi: 10.3389/fmed.2022.840777

38. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res* (2017) 77(21):e104–e7.

39. López-Carrasco A, Berbegall AP, Martín-Vañó S, Blanquer-Maceiras M, Castel V, Navarro S, et al. Intra-tumour genetic heterogeneity and prognosis in high-risk neuroblastoma. *Cancers (Basel)* (2021) 13(20):5173. doi: 10.3390/cancers13205173

40. Liu J, Si Y, Zhou Z, Yang X, Li C, Qian L, et al. The prognostic value of (18) F-FDG PET/CT intra-tumoural metabolic heterogeneity in pretreatment neuroblastoma patients. *Cancer Imaging* (2022) 22(1):32. doi: 10.1186/s40644-022-00472-4

41. Rose CJ, O'Connor JP, Cootes TF, Taylor CJ, Jayson GC, Parker GJ, et al. Indexed distribution analysis for improved significance testing of spatially heterogeneous parameter maps: Application to dynamic contrast-enhanced MRI biomarkers. *Magn Reson Med* (2014) 71(3):1299–311. doi: 10.1002/mrm.24755

42. Fiz F, Bottoni G, Bini F, Cerroni F, Marinozzi F, Conte M, et al. Prognostic value of texture analysis of the primary tumour in high-risk neuroblastoma: An (18) f-DOPA PET study. *Pediatr Blood Cancer* (2022): e29910. doi: 10.1002/pbc.29910

43. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* (2014) 5:4006. doi: 10.1038/ncomms5006

44. Chen SW, Shen WC, Lin YC, Chen RY, Hsieh TC, Yen KY, et al. Correlation of pretreatment (18)F-FDG PET tumor textural features with gene expression in pharyngeal cancer and implications for radiotherapy-based treatment outcomes. *Eur J Nucl Med Mol Imaging* (2017) 44(4):567–80. doi: 10.1007/s00259-016-3580-5

45. Atkinson C, Ganeshan B, Endozo R, Wan S, Aldridge MD, Groves AM, et al. Radiomics-based texture analysis of (68)Ga-DOTATATE positron emission tomography and computed tomography images as a prognostic biomarker in adults with neuroendocrine cancers treated with (177)Lu-DOTATATE. *Front Oncol* (2021) 11:686235. doi: 10.3389/fonc.2021.686235

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 24 October 2022
DOI 10.3389/fonc.2022.952457

OPEN ACCESS

EDITED BY
Guang Yang,
Imperial College London,
United Kingdom

REVIEWED BY
Ping Li,
Hong Kong Polytechnic University,
Hong Kong SAR, China
Qiang Zhang,
Dalian University of Technology, China
Qiong Wang,
Shenzhen Institutes of Advanced Technology (CAS), China

*CORRESPONDENCE
Moon-Tong Chan
tmtchan@hkmu.edu.hk

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

RECEIVED 25 May 2022
ACCEPTED 18 July 2022
PUBLISHED 24 October 2022

CITATION
Huang X, Lin Z, Huang S, Wang FL, Chan M-T and Wang L (2022) Contrastive learning–guided multi-meta attention network for breast ultrasound video diagnosis. *Front. Oncol.* 12:952457. doi: 10.3389/fonc.2022.952457

Contrastive learning–guided multi-meta attention network for breast ultrasound video diagnosis

Xiaoyang Huang[1], Zhi Lin[1], Shaohui Huang[1], Fu Lee Wang[2], Moon-Tong Chan[2]* and Liansheng Wang[1]

[1]Department of Computer Science, School of Informatics, Xiamen University, Xiamen, China, [2]School of Science and Technology, Hong Kong Metropolitan University, Hong Kong, Hong Kong SAR, China

Breast cancer is the most common cause of cancer death in women. Early screening and treatment can effectively improve the success rate of treatment. Ultrasound imaging technology, as the preferred modality for breast cancer screening, provides an essential reference for early diagnosis. Existing computer-aided ultrasound imaging diagnostic techniques mainly rely on the selected key frames for breast cancer lesion diagnosis. In this paper, we first collected and annotated a dataset of ultrasound video sequences of 268 cases of breast lesions. Moreover, we propose a contrastive learning–guided multi-meta attention network (CLMAN) by combining a deformed feature extraction module and a multi-meta attention module to address breast lesion diagnosis in ultrasound sequence. The proposed feature extraction module can autonomously acquire key information of the feature map in the spatial dimension, whereas the designed multi-meta attention module is dedicated to effective information aggregation in the temporal dimension. In addition, we utilize a contrast learning strategy to alleviate the problem of high imaging variability within ultrasound lesion videos. The experimental results on our collected dataset show that our CLMAN significantly outperforms existing advanced methods for video classification.

KEYWORDS

ultrasound sequence, video classification, breast lesion, contrastive learning, multi-meta attention network

1 Introduction

According to the World Cancer Report (1), the number of new cases of breast cancer has reached 2.26 million worldwide in 2020, and breast cancer accounts for about 6.9% of all deaths from cancer worldwide, ranking fifth. Early detection and timely treatment can effectively improve the survival prognosis of breast cancer patients, prolong their survival years, and improve the people's living standards. Because it is non-invasive, inexpensive, safe, and free of ionizing radiation, ultrasound imaging is currently the most commonly used technique for the early detection of breast lesions. However, ultrasound imaging provides low-quality imaging, mainly because interference from the ultrasound reflective wavefront causes speckle noise on imaging. During the acquisition or examination of a breast sequence, the operator usually needs to apply processing such as filtering, adjusting brightness levels, and scaling the image to improve the quality of ultrasound imaging, whereas interpreting ultrasound imaging usually requires an experienced and well-trained radiologist. However, in some cases, the breast lesion in the ultrasound imaging is ambiguous, and even experienced radiologists are unable to accurately determine its benignity or malignancy [in medical practice, BI-RADS 1–3 are usually considered benign, BI-RADS 4 for suspicious malignant, and BI-RADS 5–6 for malignant (2)].

Computer-assisted technology has provided new ideas for the diagnosis of breast lesions by ultrasound imaging. With the help of computer-aided diagnosis, the operation-dependent impact of ultrasound imaging can be minimized. At the same time, computer-aided diagnosis can also reduce the workload of radiologists. Most of the existing computer-aided diagnostic techniques analyze a single frame (key frames) in the video sequence of pathology acquisition. Although it helps to reduce the computer diagnostic time, it also reveals two significant problems: first, it is challenging to select typical key frames representing pathology samples; second, too much pathology diagnostic information is lost in the video sequence. The field of benign and malignant classification for breast lesions by ultrasound video sequences is in urgent need of research.

Therefore, we propose an automatic diagnosis model for ultrasound sequences, which uses deep learning methods to achieve high accuracy in classification recognition to assist medical diagnosis tasks. The designed diagnostic model weighs spatial dimensional information through the non-local module, on the one hand, and adaptive and fine-grained attention weight scoring for each feature dimension of each frame through the multi-meta attention module, on the other hand, focusing on the key information in the samples in a self-learning manner. This approach can accept samples of different sequence lengths and make full use of the potential connections between frames in the sample by weighting and aggregating the features of each frame through the aggregation module to improve the accuracy of diagnosis.

The contributions of this work can be summarized as follows: a) We develop a new network for learning video-level classification of breast lesions. b) We collected an ultrasound video dataset (268 sequences) for breast lesion classification. c) A deformed feature extraction module is proposed to facilitate high-quality deep feature representation, whereas a multi-meta attention module is developed to acquire key feature information at the video level adaptively. d) The experimental results show that our network achieves a new state-of-the-art performance in the breast ultrasound lesion classification task on our collected dataset.

2 Related work

2.1 Breast ultrasound classification

Classification of breast lesion pathology is a primary task in computer-aided diagnosis projects. Researchers working on breast ultrasound-related topics have proposed a number of effective deep learning schemes. Han et al. (3) used deep convolutional networks pre-trained on grayscale nature images to discriminate between benign and malignant. Although the lesion regions of interest used in this scheme were all provided by radiologists, this study demonstrated that breast lesion features extracted by deep learning–based networks can achieve comparable classification performance to hand-designed feature methods. To further avoid the potential missing effects that result from manual intervention in the region of interest selection, Cheng et al. (4) proposed the utilization of an unsupervised stacked denoising auto-encoder to extract high-level feature representations for breast lesion imaging with supervised fine-tuning training. Diagnosis models constructed in a deep learning manner usually require a large amount of training data to achieve significant classification results. However, because most cases are benign, the imbalance of medical data makes it particularly difficult to collect sufficient training samples. To alleviate the problem of model underfitting due to data scarcity, Fujioka et al. (5) and Pan et al. (6) started to use generative adversarial networks to simulate and enhance breast ultrasound sample data. The synthesized images will be further used for the training of convolutional neural networks. The semi-automatic classification model proposed by Bocchi et al. (7) is an outstanding early work to study breast lesion classification based on ultrasound video sequence data. In their proposed method, each imaging frame of the video is independently classified as benign or malignant after semi-automatic segmentation and morphological feature extraction. Subsequently, the classification results of all frames of the video are integrated to obtain reliable video-level results. This scheme results in a substantial improvement in the correct classification rate compared with the results of a single-image

frame. At the same time, the uncertainty of classification judgments for certain frames reflects the clinical situation that lesions may present different characteristic manifestations when viewed from different viewpoints.

2.2 Contrastive learning

Traditional supervised learning methods rely heavily on a large amount of labeled training data available. In addition to the expensive labeling cost, this approach is also vulnerable to generalization error, spurious correlations, adversarial attaches, etc. (8). More and more studies start to find new ways out and start to learn feature representation by self-supervised learning. Contrastive learning is a discriminative approach, which aims to group similar samples closer together and dissimilar samples as far away from each other as possible. For computer vision tasks, methods such as MoCo (9), SwAV (10), and SimCLR (11) have produced comparable results to the state-of-the-art supervised methods in ImageNet (12) dataset. He et al. (9) proposed the momentum contrast method for unsupervised visual representation learning, which trains visual representation by constructing dynamic dictionaries with queueing and moving average encoders to match with encoded queries encoder. Compared with the direct comparison of features in general contrast learning, Caron et al. (10) save computational overhead by clustering data and computing online for different enhancements of the same image. Chen et al. (11) save computational overhead by incremental image augmentation and by feature representation and introducing a learnable linear transformation between the feature representation and contrast loss, further substantially improving the quality of the learned feature.

2.3 Attention mechanism

In the field of image classification, the attention mechanism is used to extract key regions and recognize images by spatial invariance. The STN (Spatial Transformer Networks) proposed by Jaderberg et al. (13) effectively addresses the insensitivity of convolutional networks to different viewpoints of the same thing through the attention mechanism. Wang et al. (14) proposed the non-local model to apply the self-attention mechanism to the computer vision tasks. For an input feature image, each pixel value is derived from the weighted average of other pixel features. SENet (15) proposed the squeeze-and-excitation module, which enhances important channels and suppresses invalid channels by automatically learning the importance of different channel features, thus improving model accuracy and reducing computational effort and complexity. Woo et al. (16) propose CBAM (Convolutional Block Attention Module) based on SENet. It extends the attentional dimension from focusing on the channel dimension to the spatial dimension.

3 Method

Figure 1 shows the schematic illustration of the designed contrastive learning–guided multi-meta attention network (CLMAN). The network determines the input breast ultrasound sequence as benign or malignant, as well as the predicted score given to that. CLMAN consists of two main modules: a feature extraction module and a multi-meta attention module. The feature extraction module performs self-supervision training on the breast ultrasound video dataset by the contrast learning method before the formal training to the learn high-quality feature extraction patterns. For a given breast ultrasound sequence containing T frames, CLMAN first performs feature extraction on each frame by a pretrained feature extraction module to obtain independently encoded high-level feature vectors. Subsequently, the high-level feature vectors are aggregated for each frame in the multi-meta attention module. The module performs adaptive and fine-grained weight scoring along each feature dimension of each frame to form a compact and differentiated representation of breast lesions. Finally, the aggregated video-level feature vectors are used to determine the pathology of breast lesions by a linear classifier.

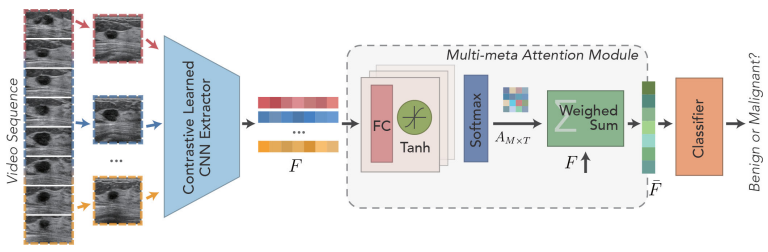


FIGURE 1
Schematic illustration of the developed Contrastive Learning guided Multi-meta Attention Network (CLMAN) for breast lesion classification in ultrasound sequence.

3.1 Deformed feature extraction module

As shown in Figure 2, the feature extraction module is designed to extract features in a sequence and obtain a high-quality feature encoding vector for each frame, which is used for downstream tasks. The module is based on ResNet-18 (17) because the residual structure adopted effectively solves the problem of model degradation due to its depth, and the constant mapping also enhances the information transfer between the upper and lower layers. Because of the inherent multi-frame nature of a sequence, video classification tasks often take smaller batch sizes. Although the amount of training data per batch is sufficient in terms of the number of images, the general batch normalization approach may not be applicable when the model goes normalization because of the high similarity of pixel feature distribution across frames within the same video. In view of this, the group normalization (18) is used in each bottleneck structure in the basic feature extraction module to guarantee the stability of the distribution of the input features. For the problems of low quality and poor contrast of ultrasound imaging, it is especially important to focus on critical regions and suppress invalid regions effectively. The non-local (14) module is introduced and placed in the third and fourth stages of the feature extraction module for capturing spatially distant relationships. It focuses on the correlation between larger objects when the model level is shallow and pays more attention to the correlation between smaller objects when the model level is deep.

Suppose a breast ultrasound sequence $V = \{v_t | t \in [0,T]\}$, where v_t denotes the tth frame and T denotes the index of frames in the sequence. The feature extraction module $\Theta(\cdot)$ extracts features from each frame to obtain the high-quality feature coding vector $F = \{f_t | t \in [0,T]\}$ for the whole sequence, which is given by

$$f_t = \Theta(v_t), \quad t \in [0, T] \tag{1}$$

3.2 Contrast learning strategy

Breast ultrasound tumors tend to be characterized by large intraclass disparities and small interclass disparities in visual presentation. Moreover, the cross-sectional visualization of lesions presented at different stages within the same sequence often varies greatly. How to identify the diversity of different cross-sections of the same lesion is the basis for the correct classification of multi-frame sequences. Inspired by SimCLR (11), we borrowed this method of learning different data augmentation of the same image as positive samples together with negative samples composed of other images to train to determine the proximity of two features and applied it to video data, as shown in Figure 3.

For any N sequence clips, M frames are selected randomly as training samples, and then, the augmented training samples are extracted by the feature extraction network to obtain the high-level feature vectors. The extracted features are cascaded through MLP layers to obtain a tighter feature representation for the model to learn a better similarity representation. Assuming that the training samples provided for learning are $\{v_m^n | \forall_n \in [0,N], \ \forall_m \in [0,M]\}$, the final feature representation can be obtained by the following:

$$\bar{f}_m^n = \mathrm{MLP}(\Theta(\tau(v_m^n))) \tag{2}$$

where τ denotes data augmentation. In the data augmentation stage, we mainly adopt the random combination of flip, crop, scale, modulation of brightness, contrast, and elastic transformation to increase the diversity of sample data.

3.3 Multi-meta attention module

The multi-meta attention module is applied to aggregate high-level feature vectors across frames of video to provide a compact and differentiated representation of mammary nodules. The module adaptively weighs all frames at a fine-grained level along each feature dimension, leveraging the valuable or discriminatory parts of each frame to facilitate commonality recognition without easily discarding or trivializing low-quality frames as the previous approaches have done. The feature extraction module trained by the contrast learning strategy is

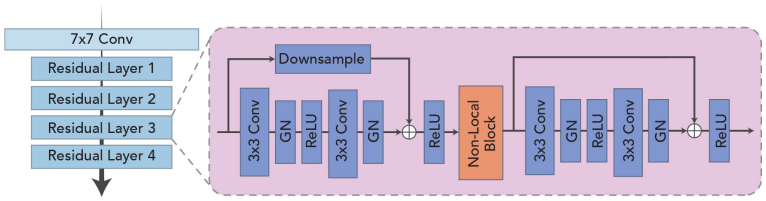


FIGURE 2
Schematic illustration of the deformed feature extraction module.

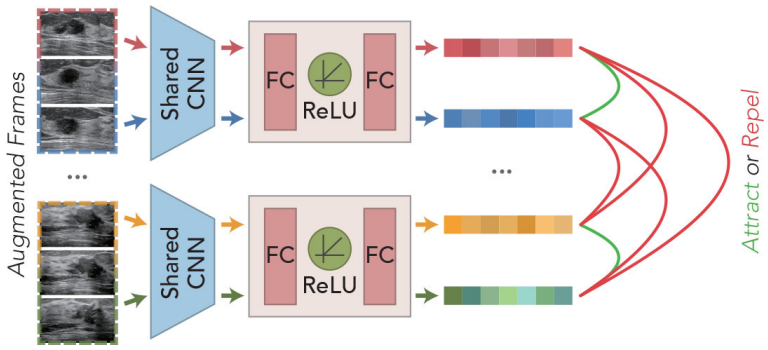


FIGURE 3
Schematic illustration of the predecessor task for the contrast learning strategy.

used to extract feature representations for each frame of the original sequence, denoted as follows:

$$F^t = \left[f_1^t\ f_2^t\ \cdots\ f_m^t\right]_{m\times 1}^T \tag{3}$$

where F^t denotes the tth frame feature vector with m dimensions.

As shown in Figure 1, a cascading attention module is applied to each frame feature to capture the attention representation better. Each attention module consists of a filter and an activation layer, which are cascaded to perform nonlinear feature learning:

$$E_l^t = \sigma\left(W_l E_{l-1}^t + b_l\right) \tag{4}$$

where the fully connected layer is used as the filter and the Tanh function is used as the activation layer $\sigma(\cdot)$ for nonlinearly transformation. When l = 1, E_{l-1}^t is defined as F^t. For the obtained attention vectors of each frame, the attention linear weights corresponding to each of the F^t channels are obtained by Softmax operation:

$$A^t = \begin{bmatrix} \frac{exp\left(e_1^t\right)}{\sum_{j=1}^T exp\left(e_1^j\right)} \\ \frac{exp\left(e_2^t\right)}{\sum_{j=1}^T exp\left(e_2^j\right)} \\ \vdots \\ \frac{exp\left(e_m^t\right)}{\sum_{j=1}^T exp\left(e_m^j\right)} \end{bmatrix}_{m\times 1} \tag{5}$$

The final aggregated feature is computed by multiplying the attention weights by the cumulative sum of the feature vectors, as shown in Figure 4. The specific aggregation operation can be expressed as follows:

$$\bar{F} = \sum_{t=1}^T A^t \dot{o} F^t \tag{6}$$

where ȯ denotes the matrix bitwise product. This aggregation module can weigh the importance of features at the dimensional level. Theoretically, it can achieve the best aggregation with good training. CLMAN uses a fair treatment of each frame of information to maximize the use of any of its valuable local features to facilitate the recognition of lesion sequence. Meanwhile, it is worth noting that the formula $\bar{F}$

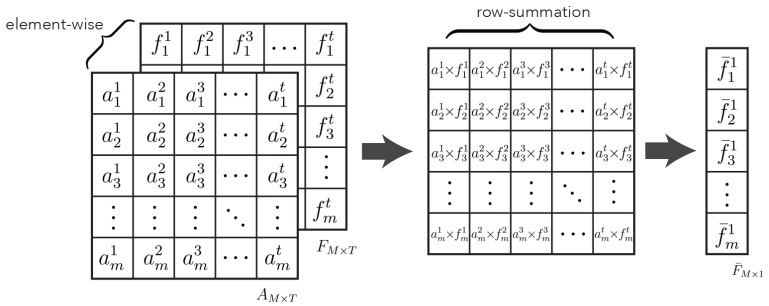


FIGURE 4
Schematic illustration of multi-meta attention operation .

degrades to average pooling when each item in the attention matrix A is equal, and of course, the same formula also applies to maximum pooling in extreme cases.

In addition, using this module of mid-term aggregation of features allows the linear classifier to process sequence clips of arbitrary length, whereas the aggregation results $\bar{F}$ have the same vector dimension as the individual features F^t and the order remain constant, i.e., this aggregation module is insensitive to sequence order and temporal information and is generalizable to ultrasound sequence clips of arbitrary start and end points. The module's parameters can be obtained by the standard backpropagation and gradient descent for supervised learning.

4 Experiments

4.1 Dataset

To evaluate the effectiveness of the developed network, we collected a dynamic breast ultrasound video sequences dataset with 268 videos, of which 152 sequences are malignant and 116 sequences are benign. All sequences are acquired by GE Healthcare equipment (Chicago, IL, USA), with L12-5 ultrasound probe and sampling frequency of 12 MHz, supported by the Xiamen University Xiang'an Hospital. A total of 107 of these sequences are randomly selected as the test set (about 40% of the total data volume), and the rest of the sequences are used as the training set. Data for both the training and test sets are obtained from cases of patients aged 20 years and older with definite benign or malignant pathological findings (BI-RADS categories 3 to 5) of breast lesions as determined by ultrasound.

4.2 Evaluation metrics

The six widely used metrics are utilized for quantitatively comparing different breast lesion ultrasound sequence classification methods. These are accuracy, average precision (AP), sensitivity, specificity, F_1 score, and area under curve (AUC).

4.3 Implementation details

Our network is implemented on PyTorch (19) and trained using a SGD (Stochastic Gradient Descent) (20) with 320 epochs, an initial learning rate of 1×10^{-4}, a momentum of 0.9, and a weight decay of 5×10^{-4}. The sample length T is set to 16, whereas cross-entropy loss is set as the loss function. The whole architecture is trained on one GeForce RTX 2080 Ti GPU, and each GPU has a batch size of 8. In the contrast learning phase, NT-Xent (11) is used as the loss function, and the LARS (21) optimizer is used to train the model in the pre-task with 8,192 epochs, an initial learning rate of 9×10^{-3}, and a weight decay of 1×10^{-6}. The batch size here is set to 64. The learning rate is adjusted using the Cosine Annealing (22).

4.4 Ablation study

4.4.1 Effectiveness of deformed extraction modules

We establish separate control groups based on ResNet-18 and compare the use of different components on the classification performance. As shown in Table 1, "ResNet18 (vanilla)" indicates the most primitive ResNet-18 architecture, "GN" denotes Group Norm, and "NL" denotes non-local module. To avoid the effect of the contrast learning strategy, none of the four settings in Table 1 use that strategy. Compared with the plain ResNet-18 architecture, the feature extraction module with group norm has 4.68%, 6.33%, 1.92% and 3.18% improvement in accuracy, specificity, F1, and AUC, respectively. Meanwhile, the feature extraction module with the non-local module shows a steady increase in all six metrics, with 10.28% increase in accuracy, 6.4% increase in AP, 3.05% increase in sensitivity, 13.15% increase in specificity, 5.64% increase in F_1, and 11.26% increase in AUC. The feature extraction module with the group norm and non-local module achieves the average

TABLE 1 Quantitative comparisons for the effectiveness of deformed extraction modules.

Methods	Acc	AP	Sens	Spec	F_1	AUC
ResNet18 (vanilla)	70.09	77.68	75.00	68.67	78.08	71.75
ResNet18 + GN	74.77	77.01	74.29	75.00	80.00	74.93
ResNet18 + NL	80.37	**84.08**	78.05	81.82	83.72	83.01
ResNet18 + GN + NL (ours)	**82.24**	81.16	**82.05**	**82.35**	**85.50**	**84.85**

"GN" denotes Group Norm, and "NL" denotes Non-local module.
The bold values/numbers means that it is the largest among all the values at the column.

best performance, with accuracy of 82.24%, AP of 81.16%, sensitivity of 82.08%, specificity of 82.35%, F_1 of 85.50%, and AUC of 84.85%. It indicates that using the group norm and non-local module for the feature extraction module to obtain high-quality deep features has a certain facilitation effect.

4.4.2 Effectiveness of contrast learning strategy

The feature extraction module used by our network is pre-trained by a contrast learning strategy to effectively identify different geometric patterns of the same lesion under the same sequence imaging before formally training. Table 2 verifies the impact of the contrast learning strategy, which is denoted as "CL", on the model performance. The experiments show that the performance of the CLMAN decreases when the contrast learning strategy is removed. Specifically, accuracy, AP, sensitivity, specificity, F_1, and AUC decreased by 6.55%, 11.62%, 12.39%, 3.57%, 5.54%, and 7.79%, respectively. It suggests that the contrast learning strategy can effectively alleviate the problem of large intraclass differences in the visual presentation of ultrasound lesions.

4.4.3 Effectiveness of multi-meta attention module

We conduct ablation experiments of multi-meta attention modules on the CLMAN model. First, the experiment considers the degenerate version of our multi-attention module, i.e., average pooling, as well as the extreme case of the maximum pooling and then compares them. Second, the LSTM (Long Short Term Memory) methods for long sequence feature capture are also compared in this experiment. In addition, we also compared attention modules proposed by other studies to demonstrate the advantage of the multi-meta attention module in video tasks. As shown in Table 3, "Multi-meta Att" denotes the multi-meta attention module, and "Average" and "Max-pooling" represent the degenerate average pooling and the extreme maximum pooling, respectively. According to Table 3 the long sequence feature capture capability of LSTM is not fully applicable to ultrasound video imaging aggregation. The proposed classic attention modules that have often been effective in the past do not seem to be up to our video task. Meanwhile, the simple average pooling and maximum pooling methods achieved the best in terms of sensitivity or specificity, but the other metrics were not satisfactory. The proposed multi-meta attention scheme shows a 1.87% improvement in accuracy, 4.66% improvement in AP, 3.94% improvement in F_1, and 5.12% improvement in AUC, with a stronger comprehensive capability. It indicates that the model has different fine-grained trade-offs for each part of the features, whereas such weights are learnable, and the simple and crude average pooling and maximum pooling approaches limit this adaptive capability.

4.5 Comparisons with state of the arts

To demonstrate the effectiveness and feasibility of the designed CLMAN model, Table 4 selects from five papers nine existing methods commonly used to handle video

TABLE 2 Quantitative comparisons for the effectiveness of contrast learning strategy.

Methods	Acc	AP	Sens	Spec	F_1	AUC
Without CL Guided	82.24	81.16	82.05	82.35	85.50	84.85
With CL (ours)	**88.79**	**92.78**	**94.44**	**85.92**	**91.04**	**92.64**

"CL" denotes the contrast learning strategy.
The bold values/numbers means that it is the largest among all the values at the column.

TABLE 3 Quantitative comparisons for the effectiveness of multi-meta attention module.

Methods	Acc	AP	Sens	Spec	F_1	AUC
With External Attention (23)	79.44	81.39	76.19	81.54	82.81	79.98
With Self-attention (24)	80.37	77.85	79.49	80.88	83.97	78.50
With Efficient Multi-head Self-attention (25)	80.37	78.51	81.08	80.00	84.21	77.49
With LSTM	83.18	77.91	96.43	78.48	87.32	79.47
With Average	86.92	85.26	**96.88**	82.67	89.86	87.48
With Max-pooling	85.05	88.12	80.43	**88.52**	87.10	87.52
With Multi-meta Attn (ours)	**88.79**	**92.78**	94.44	85.92	**91.04**	**92.64**

"Multi-meta Att" denotes the multi-meta attention module, and "Average" and "Max-pooling" represent the average pooling and the maximum pooling, respectively.
The bold values/numbers means that it is the largest among all the values at the column.

TABLE 4 Quantitative comparisons of our network and compared methods on the collected ultrasound sequence dataset.

Methods	Acc	AP	Sens	Spec	F_1	AUC
R3D (26)	75.70	80.03	82.14	73.42	81.69	77.56
Times Former (27)	77.57	71.34	77.78	77.46	82.09	72.08
MC3 (26)	77.57	80.83	81.25	76.00	82.61	78.86
P3D (28)	80.37	81.33	89.66	76.92	85.11	81.28
R(2+1)D (26)	82.24	87.36	90.32	78.95	86.33	84.85
TIN Res34 (29)	84.48	86.85	82.33	86.35	86.87	86.79
TIN Res50 (29)	85.05	85.53	88.94	83.10	88.06	86.66
TIN Res18 (29)	85.24	89.52	82.50	87.11	87.52	87.30
TSM (30)	86.92	89.72	91.67	84.51	89.55	90.22
CLMA-Net (ours)	**88.79**	**92.78**	**94.44**	**85.92**	**91.04**	**92.64**

The bold values/numbers means that it is the largest among all the values at the column.

classification task for comparison, including R3D (23), Times Former (27), MC3 (23), P3D (24), R(2 + 1)D (23), TIN(Res18, Res34, Res50)(29), and TSM (25). For providing a fair comparison, we obtain the classification results of all compared methods by exploiting their public implementations or by implementing them. We train these networks on our dataset and only set the batch size and epoch number to the same as ours.

CLMAN performs on par with the best of the methods compared and even better suited for video-level classification tasks of breast ultrasound sequence, with accuracy improved by 1.87%, AP improved by 3.06%, sensitivity improved by 2.77%, specificity by 1.44%, F_1 improved by 1.49%, and AUC improved by 2.42%.

More visually, Figure 5 shows the ROC curves of CLMAN with the above five of the nine methods. The performance of R3D, MC3, P3D, and R(2 + 1)D is similar, and the AUC remains around 80%, whereas the area of TSM and CLMAN is comparable, both exceeding 90%.

5 Conclusion

In this paper, we first collected 268 video sequences constituting a video dataset for breast ultrasound classification. Moreover, we propose a CLMAN for lesion diagnosis of ultrasound breast sequences in arbitrary length. Our approach

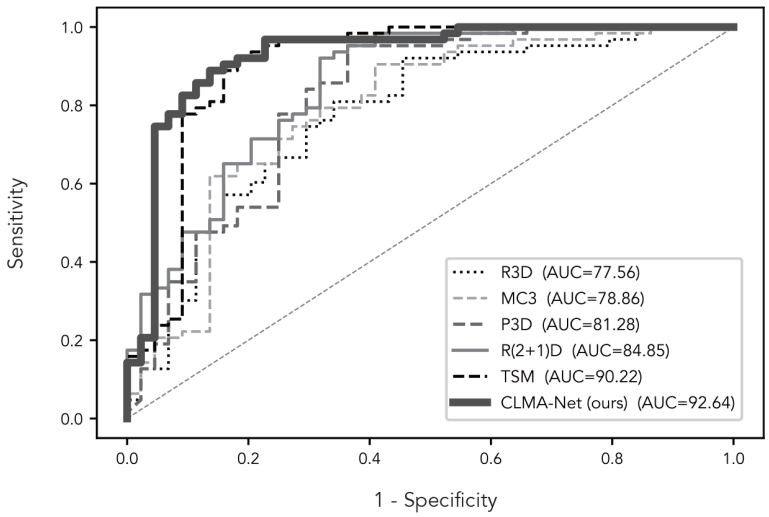


FIGURE 5
ROC curves of our network and compared methods.

is able to learn the attention weights of each feature dimension adaptively and autonomously in both spatial and temporal dimensions while using a contrast learning predecessor task to effectively address several challenges of the ultrasound video sequence classification problem. Experimental results on the collected dataset show that our network achieves superior diagnostic performance for breast lesions than the state-of-the-art video classification methods.

Data availability statement

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

Author contributions

XH and SH participated in the design of the study. XH and SH collected the data. ZL and XH performed the statistical analysis. ZL and XH wrote the manuscript. ZL, M-TC, and LW revised the manuscript. FW makes its contributions to prepare and revise the response letter for revised paper,and provide a research grant to cover the publication fee. All authors contributed to the article and agreed to the submitted version.

Acknowledgments

This work was supported by the Research Grants Council of Hong Kong (Project: UGC/FDS16/M08/18).

Conflict of interest

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

Publisher's note

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

References

1. Wild C, Weiderpass E, Stewart BW. *World cancer report: cancer research for cancer prevention*. France: IARC Press (2020).

2. Asbeutah AM, Karmani N, Asbeutah AA, Echreshzadeh YA, AlMajran AA, Al-Khalifah KH. Comparison of digital breast tomosynthesis and digital mammography for detection of breast cancer in kuwaiti women. *Med Principles Pract* (2019) 28:10–5. doi: 10.1159/000495753

3. Han S, Kang HK, Jeong JY, Park MH, Kim W, Bang WC, et al. A deep learning framework for supporting the classification of breast lesions in ultrasound images. *Phys Med Biol* (2017) 62:7714–28. doi: 10.1088/1361-6560/aa82ec

4. Cheng JZ, Ni D, Chou YH, Qin J, Tiu CM, Chang YC, et al. Computer-aided diagnosis with deep learning architecture: applications to breast lesions in us images and pulmonary nodules in ct scans. *Sci Rep* (2016) 6:1–13. doi: 10.1038/srep24454

5. Fujioka T, Kubota K, Mori M, Kikuchi Y, Katsuta L, Kimura M, et al. Efficient anomaly detection with generative adversarial network for breast ultrasound imaging. *Diagnostics* (2020) 10:456. doi: 10.3390/diagnostics10070456

6. Pang T, Wong JHD, Ng WL, Chan CS. Semi-supervised gan-based radiomics model for data augmentation in breast ultrasound mass classification. *Comput Methods Programs Biomed* (2021) 203:106018. doi: 10.1016/j.cmpb.2021.106018

7. Bocchi L, Gritti F, Manfredi C, Giannotti E, Nori J. Semiautomated breast cancer classification from ultrasound video, in: *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*. Piscataway, NJ: IEEE (2012). p. 1112–5.

8. Liu X, Zhang F, Hou Z, Mian L, Wang Z, Zhang J, et al. Self-supervised learning: Generative or contrastive, in: *IEEE Transactions on Knowledge and Data Engineering*. Piscataway, NJ: IEEE (2021).

9. He K, Fan H, Wu Y, Xie S, Girshick R. Momentum contrast for unsupervised visual representation learning, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ: IEEE (2020). p. 9729–38.

10. Caron M, Misra I, Mairal J, Goyal P, Bojanowski P, Joulin A. Unsupervised learning of visual features by contrasting cluster assignments. (2020) arXiv preprint (La Jolla: Neural information processing systems foundation). arXiv:2006.09882.

11. Chen T, Kornblith S, Norouzi M, Hinton G. (2020). A simple framework for contrastive learning of visual representations, in: *International conference on machine learning (PMLR)*, New York: PMLR. pp. 1597–607.

12. Deng J, Dong W, Socher R, LJ L, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database, in: *2009 IEEE conference on computer vision and pattern recognition (Ieee)*. Piscataway, NJ: IEEE (2009). p. 248–55.

13. Jaderberg M, Simonyan K, Zisserman A. Spatial transformer networks. *Adv Neural Inf Process Syst* (2015) 28:2017–25.

14. Wang X, Girshick R, Gupta A, He K. Non-local neural networks, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*. (Piscataway, NJ: IEEE) (2018). p. 7794–803.

15. Hu J, Shen L, Sun G. Squeeze-and-excitation networks, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*. (Piscataway, NJ: IEEE) (2018). p. 7132–41.

16. Woo S, Park J, Lee JY, Kweon IS. in: *Cbam: Convolutional block attention module. Proceedings of the European conference on computer vision (ECCV)*. New York: Springer (2018). p. 3–19.

17. 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*. New York: Springer (2016). p. 770–8.

18. Wu Y, He K. Group normalization, in: *Proceedings of the European conference on computer vision (ECCV)*. New York: Springer (2018). p. 3–19.

19. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. Pytorch: An imperative style, high-performance deep learning library. *Adv Neural Inf Process Syst* (2019) 32:8026–37.

20. Robbins H, Monro S. A stochastic approximation method. *Ann Math Stat* (1951) 22:400–7. doi: 10.1214/aoms/1177729586

21. You Y, Gitman I, Ginsburg B. Large Batch training of convolutional networks. (2017). arXiv preprint arXiv:1708.03888.

22. Loshchilov I, Hutter F. Sgdr: Stochastic gradient descent with warm restarts. (2016). arXiv preprint (San Diego: International Conference on Learning Representations). arXiv:1608.03983.

23. Tran D, Wang H, Torresani L, Ray J, LeCun Y, Paluri M. A closer look at spatiotemporal convolutions for action recognition, in: *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. Piscataway, NJ: IEEE (2018). p. 6450–9.

24. Qiu Z, Yao T, Mei T. Learning spatio-temporal representation with pseudo-3d residual networks, in: *Proceedings of the IEEE International Conference on Computer Vision*. Piscataway, NJ: IEEE (2017). p. 5533–41.

25. Lin J, Gan C, Han S. Tsm: Temporal shift module for efficient video understanding, in: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. Piscataway, NJ: IEEE (2019). p. 7083–93.

26. Tran D, Wang H, Torresani L, Ray J, LeCun Y, Paluri M. A closer look at spatiotemporal convolutions for action recognition. *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition* (2018), 6450–9.

27. Bertasius G, Wang H, Torresani L. Is space-time attention all you need for video understanding? *325 Proceedings of the International Conference on Machine Learning (ICML)* (2021), 2(3):4.

28. Qiu Z, Yao T, Mei T. Learning spatio-temporal representation with pseudo-3d residual networks. *Proceedings of the IEEE International Conference on Computer Vision* (2017), 5533–41.

29. Shao H, Qian S, Liu Y. Temporal interlacing network. *Proceedings of the AAAI Conference on Artificial Intelligence* (2020), Vol. 34, No. 07, 11966–73.

30. Lin J, Gan C, Han S. Tsm: Temporal shift module for efficient video understanding. *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2019), 7083–93.

frontiers | Frontiers in Oncology

TYPE Case Report
PUBLISHED 04 January 2023
DOI 10.3389/fonc.2022.927086

OPEN ACCESS

EDITED BY
Kyung Hyun Sung,
UCLA Health System, United States

REVIEWED BY
Nguyen Minh Duc,
Pham Ngoc Thach University of Medicine, Vietnam
Baixin Shen,
Nanjing Medical University, China

*CORRESPONDENCE
Yanwei Miao
ywmiao716@163.com

†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 23 April 2022
ACCEPTED 05 December 2022
PUBLISHED 04 January 2023

CITATION
Tian S, Pan T, Gao B, Li W, Liu J, Zou K and Miao Y (2023) Case report: Primary intracranial mucosa-associated lymphoid tissue lymphoma presenting as two primary tumors involving the cavernous sinus and extra-axial dura, respectively. *Front. Oncol.* 12:927086. doi: 10.3389/fonc.2022.927086

Case report: Primary intracranial mucosa-associated lymphoid tissue lymphoma presenting as two primary tumors involving the cavernous sinus and extra-axial dura, respectively

Shiyun Tian[1†], Tao Pan[2†], Bingbing Gao[1], Wanyao Li[1], Jiashen Liu[3], Kun Zou[4] and Yanwei Miao[1*]

[1]Department of Radiology, First Affiliated Hospital of Dalian Medical University, Dalian, China, [2]The Interventional Therapy Department, First Affiliated Hospital, Dalian Medical University, Dalian, China, [3]Department of Pathology, First Affiliated Hospital, Dalian Medical University, Dalian, China, [4]Department of Radiation Therapy, First Affiliated Hospital, Dalian Medical University, Dalian, China

Primary intracranial mucosa-associated lymphoid tissue (MALT) lymphoma is a rare type of brain tumor, with only a few reported cases worldwide that mostly have only one lesion with conventional magnetic resonance imaging (MRI) findings. Here, we present a special case of intracranial MALT lymphoma with two mass lesions radiographically consistent with meningiomas on MRI before the operation. A 66-year-old woman was admitted to the hospital with intermittent right facial pain for 1 year, aggravated for the last month. Brain MRI showed two extracerebral solid masses with similar MR signal intensity. One mass was crescent-shaped beneath the skull, and the other was in the cavernous sinus area. Lesions showed isointensity on T1WI and T2WI and an intense homogeneous enhancement after contrast agent injection. Both lesions showed hyperintensity in amide proton transfer–weighted images. The two masses were all surgically resected. The postoperative pathology indicated extranodal marginal zone B-cell lymphoma of MALT. To improve awareness of intracranial MALT lymphoma in the differential diagnosis of extra-axial lesions among clinicians, we present this report and briefly summarize previously reported cases to describe the clinical, pathological, radiological, and treatment features.

KEYWORDS

extranodal marginal zone lymphoma, mucosa-associated lymphoid tissue, primary intracranial tumor, central nervous system, magnetic resonance imaging, treatment

Introduction

Extranodal marginal zone lymphoma of the mucosa-associated lymphoid tissue (MALT lymphoma), as defined in the 2016 World Health Organization classification of lymphoid neoplasms (1), is a type of mature B-cell lymphoid neoplasm. At the very beginning, MALT lymphoma is described as a subtype of gastric lymphoma. Then, studies found that the MALT lymphoma may arise in almost all organs of the human body, including unusual sites, such as the dura (2). MALT lymphoma derived from dura is extremely rare; only a few cases were diagnosed and reported (3–10). The diagnoses are mostly based on conventional magnetic resonance imaging (MRI) image, and some combine with diffusion–weighted imaging (DWI) (11). This case report presented a MALT lymphoma with two separated lesions in the cranial: along the right fronto–temporo–parietal extra–axial dura and in the right cavernous sinus (CS). On the other hand, the patient also underwent amide proton transfer–weighted (APTw) and contrast–enhanced fluid–attenuated inversion recovery (FLAIR) examinations in addition to the conventional MR sequences.

Case report

A 66–year–old woman was admitted to the hospital with intermittent right facial pain for 1 year, aggravated for the last month, the pain was knife–like accompanied with numbness of the upper and lower lips and had no obvious inducement before onset. The patient also had nausea without vomiting. No other objective neurologic findings were detected. The patient had a history of type 2 diabetes mellitus for more than 10 years.

Brain MRI showed two extracerebral solid masses with similar MR signal intensity. One mass was fusiform in the cavernous sinus (CS) area (Figures1 A–C), the other was crescent-shaped beneath the fronto-temporo-parietal region skull (Figures1 D–F). The lesions showed a slight hypointensity to isointensity on T1–weighted image (T1WI), an isointensity on T2–weighted image (T2WI), and an intense homogeneous enhancement after contrast agent injection. (Figures 1G–I). For more details, for the tumor beneath cranium, short striped flow voids signals inside and moderate edema zone around (Figure 1E) are were observed. As the CS lesion, the ipsilateral cavernous segment of internal carotid artery had no stenosis (Figures 1J, K). The right CS lesion spread to the posterior cranial fossa, involving the internal auditory canal and encircled the cisternal segment of the right trigeminal nerve, auditory nerve and facial nerve (Figures 1J, K). The serrated enhancement was due to leptomeningeal involvement along the inner side of the fronto-temporo-parietal region lesion on the post-contrast fluid-attenuated inversion recovery (FLAIR) image (Figure 1L). APTw image (Figure 1M) showed relatively hyperintense compared to the normal brain tissue, the mean APTw value of the tumor was 3.1% (2.2%–4.3%). Overall, the MR features of these two lesions were misdiagnosed asconsistent with meningiomas.

The surgery was performed to resect the lesions on the right forehead and parasellar *via* the right frontotemporal parietal approach. The tumor was 9.0 × 7.5 cm in size at the top of the right forehead region, reaching from the base of the anterior cranial fossa forward to the central anterior gyrus backward. The solid tumor showed inconsistent hardness supplied by dural and lateral fissure vessels and adheres closely to the brain parenchyma. The dura near the tumor was enlarged up to 2 cm. The lesion was completely removed. The right parasellar tumor was relatively soft grayish white and dark brown mass with a size of 2.5 × 2.0 × 1.8 cm, wrapping the internal carotid artery and III–V cranial nerves. The parasellar tumor is only partially removed along the internal carotid artery and the lateral cranial nerve space due to the involvement of the cranial nerve.

Microscopic examination of the biopsy revealed fragments of dense connective tissue infiltrated by closely packed, medium–sized, and monomorphic lymphocytes and scattered plasma cells (Figure 2A). The lymphocytic infiltrates were positive for CD20 (Figure 2B), CD138, CD38, CD21 (Figure 2C), CD79a (Figure 2D), and MUM1. These lymphocytes and plasma cells are monotypic for kappa light chain (Figure 2E) expression, and they are essentially negative for lambda light chain (Figure 2F). The pathological diagnosis was MALT lymphoma with plasma cell differentiation and amyloidosis of the stroma and vascular wall.

A month and a half after the surgery, the patient received 4 weeks of whole–brain radiotherapy of 40 Gy in total. Follow–up of the initial MRI 8 months after the operation illustrated no evidence of tumor recurrence (Figures 3A–L). During the last 27 months of follow–up, the patient had no complaints in her daily life.

Discussion

Extranodal marginal zone MALT lymphoma accounts for 7% to 8% of newly diagnosed lymphomas (12). Although the World Health Organization classification published in 2008 still defines the stomach as the most common organ of origin (accounting for roughly 50% of MALT lymphomas), recent data have suggested a decline in the percentage of gastric MALT lymphomas (12). It can occur in a variety of extranodal locations, but it is most common in organs where lymphocytes are generally absent, such as the stomach, salivary glands, and thyroid. To our knowledge, only a few cases have been reported to occur in the dura mater (3, 4). Louveau et al. (12) discovered functional lymphatic vessels in the dural sinus, which could be the cause of the issue. The published cases were all single lesions. To the best of our knowledge, we are the first to present a rare case of two simultaneous lesions occurring in different cranial regions.

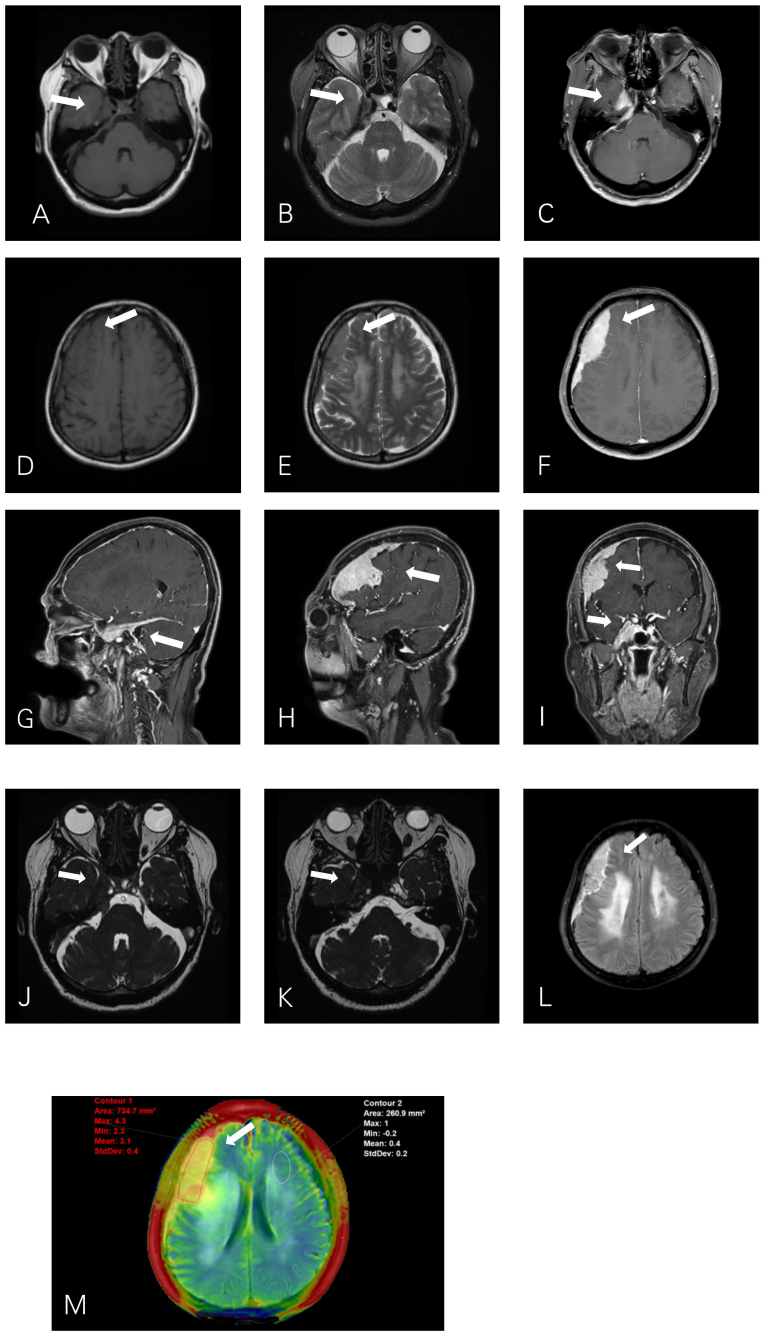


FIGURE 1
Axial T1WI **(A)** and T2WI **(B)** show a fusiform lesion isointensity to gray matter in the right cavernous sinus (CS). T1WI **(D)** and T2WI **(E)** show a crescent–shaped extra–axial mass in the right fronto–temporo–parietal region dura. The right CS lesion **(J)** extends into the posterior cranial fossa into the internal auditory canal and encircles the auditory nerve and facial nerve **(K)**. Post–contrast T1–weighted **(C, F, G, H, I)** images show intense homogeneous enhancement of the mass lesions with both dural tails visible. The leptomeningeal involvement of the right extra–axial dura lesion shows serrated enhancement on contrast–enhanced T2–FLAIR **(L)**. APTw image **(M)** shows a relative homogeneous tumor mass, with APTw signal intensity rates of 2.2%–4.3% (mean: 3.1%).

MALT lymphoma has a wide range of clinical presentations, owing to variances in signs and symptoms associated with different extranodal organs. According to the literature, there is a slight female preponderance among patients with MALT lymphoma, and the median age at diagnosis is about 65 years. Patients with cranial MALT lymphoma present non–specific neurologic symptoms, including headache, meningeal signs, and cranial nerve involvement (4). The clinical features are

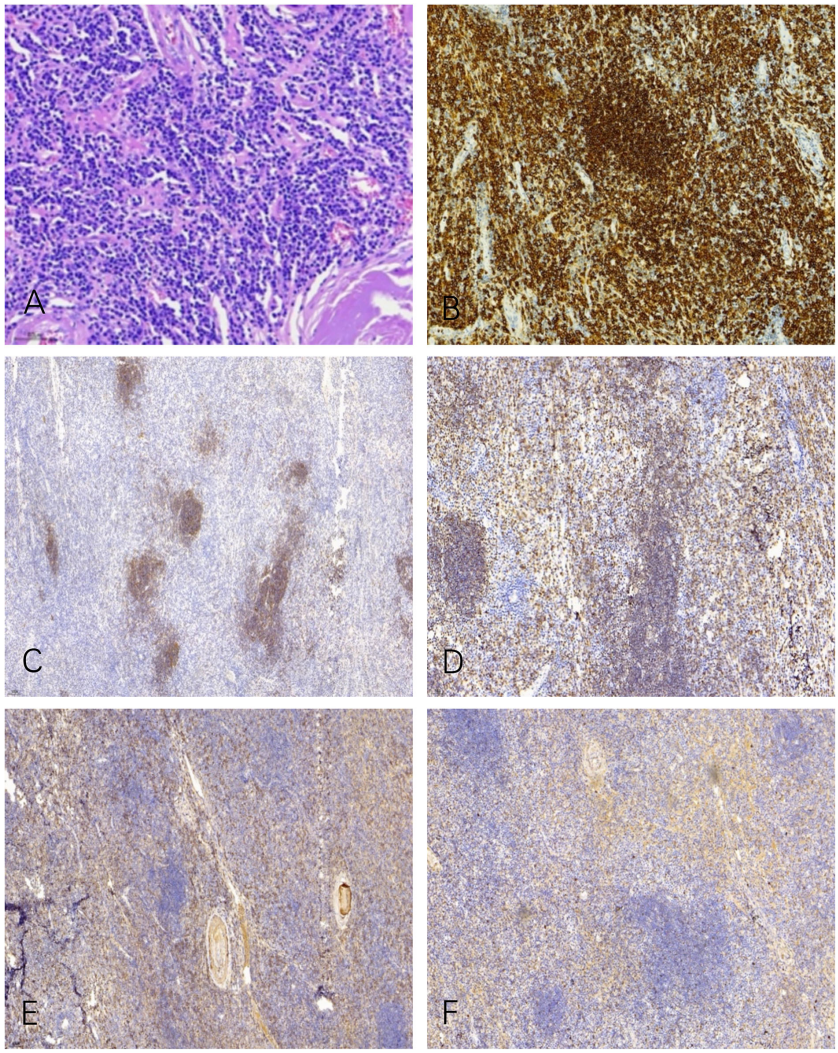


FIGURE 2
Microscopic examination shows a diffuse infiltrate of lymphocytes, plasma cells, and some cells with intermediate lymphoplasmacytic morphology **(A)**. The lymphocytic infiltrates were positive for CD20 **(B)**, CD21 **(C)**, and CD79a **(D)**. These lymphocytes and plasma cells are monotypic for kappa light chain **(E)** expression, and they are essentially negative for lambda light chain **(F)**.

summarized in Table 1. In this case, the patient complained about a severe knife–like pain on the right side of her face. The MRI showed that the trigeminal nerve was involved by the right CS lesion, and it was further confirmed that the lesion wrapped the III–V nerves during the operation.

Radiologically, MALT lymphoma is mainly a single hyperattenuated lesion on CT, reflecting the highly dense tumor cells. MRI reveals a homogeneous isointensity–to–hypointensity lesion on T1WI and isointensity on T2WI, consistent with the dense cells and increased fibrous tissues in the tumor. Vasogenic edema is typically noted in the adjacent brain parenchyma. MALT lymphomas are uniformly enhanced on post–contrast T1WIs, and the interface between the tumor and brain parenchyma is blurred (13), with a long and wide meningeal tail sign. In this case, the contrasted T2–FLAIR sequence shows a sawtooth–like enhancement in the brain–tumor interface, which is considered as a sign of the involvement of the pia mater. The permeability of meningeal vessels increases with the breakdown of the blood–brain, blood– cerebrospinal fluid (CSF), or blood–nerve barrier (14). As T2–FLAIR sequences are thought not to show signals in the leptomeningeal vasculatures at a normal flow, a leptomeningeal enhancement on the contrast–enhanced T2–FLAIR images indicates a true leptomeningeal involvement (8, 15, 16).

Advanced MRI technology can add more diagnostic information and improve the accuracy of diagnosis. However, to our knowledge, only one MALT lymphoma case with DWI has been published. They observed diffusion restriction on DWI

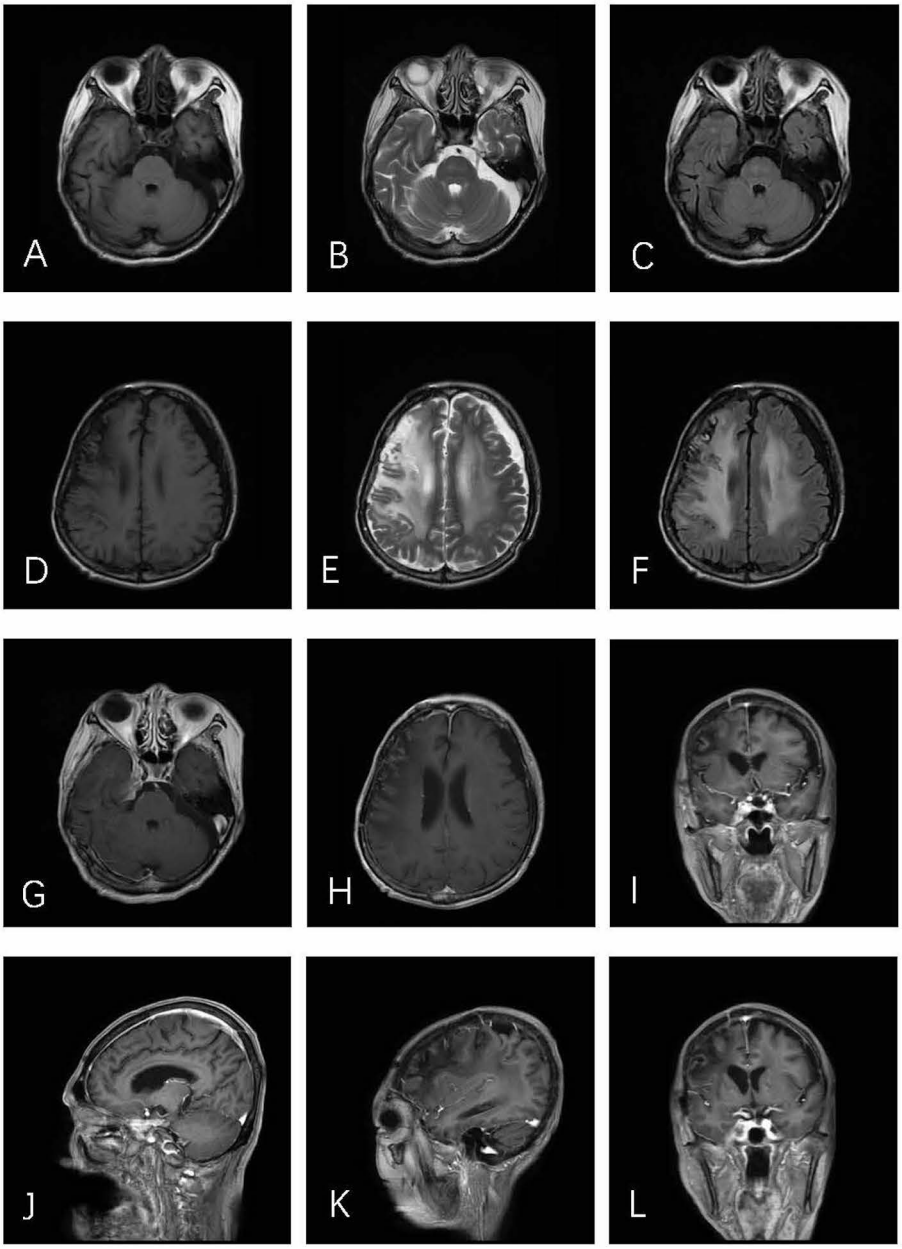


FIGURE 3
Axial T1WI **(A)**, T2WI **(B)**, and T2–FLAIR **(C)** demonstrate postoperative alterations in CS. In axial T1WI **(D)**, T2WI **(E)**, and T2–FLAIR **(F)** images, an obvious edema zone can be seen in the right parietal frontal brain, which may be related to brain changes after radiotherapy. No tumor recurrence was found in post–contrast T1–weighted **(G–L)** images.

(b–value =1,000 s/mm^2) with a decreased apparent diffusion coefficient (ADC) value of 0.64×10^{-3} mm^2/s (11). The patient in this case underwent APTw sequence, and a relative hyperintense mass with a mean APTw value of 3.1% in the APTw. The elevated APTw may be result from a high protein and amino acid concentration in the increased lymphoid cells. The APTw changes are also helpful to the differentiation and grading of tumors. Jiang et al. (16) found that the $APTW_{max}$ value of lymphomas was lower than that of high–grade gliomas (3.38% ± 1.06% and 4.36% ± 1.30%, respectively). Recent studies on meningiomas have demonstrated that the normalized magnetization transfer ratio asymmetry ($nMTR_{asym}$) of atypical meningiomas was significantly greater than that of benign meningiomas (2.46% *vs.* 1.67%, $P < 0.001$) (17).

TABLE 1 Summary of Intracranial MALT Lymphomas up to 4 November 2022.

References	Age/ sex	Immune status	Clinical features	Involvement site	CT/ MRI	Treatment	Follow–up
Shaia et al. (2)	61/F	N/A	Nausea and vomiting	Right posterior fossa	–/+	Oral steroid therapy, surgery, R/T	3/6 months: relapse–free
Ferguson et al. (4)	29/F	Immunocompetent	Exophthalmos and visual loss	Right optic foramen and cavernous sinus	–/+	Surgery, R/T	3 years: relapse–free
Choi et al. (5)	69/M	N/A	Headache	Anterosuperior	–/+	Surgery, C/T	30 months: relapse–free
Sanjeevi et al. (6)	46/F	N/A	Headache and ophthalmalgia	Left cavernous sinus	–/+	surgery, R/T	N/A
Neidert et al. (8)	44/M	N/A	facial numbness	right fronto–parietal	–/+	surgery, R/T	2 years: relapse–free
Yang et al. (11)	59/M	Immuno competent	Ptosis and blurred vision	Right cavernous sinus	+/+	Surgery, R/T, C/T	2 years: relapse–free

N/A, not available, C/T, chemotherapy, R/T, radiotherapy.
-/+ represents the absence or availability of corresponding CT and MRI images.

A wide variety of lesions can involve the dura, ranging from benign to malignant neoplastic, infectious, and granulomatous etiologies. Frequently, these lesions have imaging characteristics similar to those of a meningioma and are often mistaken for one. MALT lymphoma is often misdiagnosed as meningioma because of its origin in the dura mater. As shown in this case, compared with meningiomas, the meningeal tail of MALT lymphoma is longer and not smooth, and the interface presents a serrated change. Solitary fibrous tumor/hemangiopericytomas also originate from dura and is characterized by a narrow base connected to the dura, and "mushroom" grows to the adjacent brain in a lobulated shape. Intracranial Rosai–Dorfman disease is a rare benign histiocytosis that primarily affects men. The distinguishing point from MALT lymphoma is that the edema of the adjacent brain parenchyma is obvious.

The treatment of MALT lymphoma includes local surgical resection, radiotherapy, or chemotherapy. This case was misdiagnosed as meningiomas and underwent surgical resection. During the operation, the tumor adhered to the cranial nerve, resulting in a small part of the residual. Radiotherapy was performed after the operation. The patient's life was normal after postoperative radiotherapy. Although surgical resection of tumors may cure many patients with MALT lymphoma, the use of this strategy is gradually decreasing (12). This is because postoperative sequelae and organ dysfunction are more harmful than lymphoma itself (12). In one case (3), six cycles of rituximab + bendamustine allowed intracranial MALT lymphoma complete remission for more than 2 years without the need for invasive surgery. Therefore, surgery is mainly limited to histopathological diagnosis, management of treatment complications, or treatment of recurrent diseases in patients who are not suitable for other treatments. As an inert malignant tumor, MALT lymphoma usually has a good curative effect.

In conclusion, we report a case in which two MALT lymphomas occur simultaneously in different skull regions. These specific MRI features combined with APTw, post–contrast T1WI, and contrast–enhanced T2 FLAIR images could help in making a directive diagnosis before the operation, which can further help neurosurgeons make appropriate preoperative treatment plans for patients.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

Ethics statement

The studies involving human participants were reviewed and approved by the Local Ethics Committee of the First Affiliated Hospital of Dalian Medical University. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

ST and TP reviewed the literature, designed the article and wrote the report. BG and WL collected and analyzed the data. JL prepared histology figures and provided immunohistochemical analysis. KZ provided information on radiotherapy. YM revised the report critically for important intellectual content and gave

final approval of the version to be published. All authors contributed to the article and approved the submitted version.

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

References

1. Swerdlow SH, Campo E, Pileri SA, Harris NL, Stein H, Siebert R, et al. The 2016 revision of the world health organization classification of lymphoid neoplasms. *Blood* (2016) 127(20):2375–90. doi: 10.1182/blood–2016–01–643569

2. Shaia J, Kerr PB, Saini A, Roberti F, Kapil J, Jones R, et al. Mucosa–associated lymphoma tissue of the dura presenting as meningioma. *South Med J* (2010) 103 (9):950–2. doi: 10.1097/SMJ.0b013e3181eb3477

3. Villeneuve A, Rubin F, Bonfils P. Meningeal marginal zone b–cell lymphoma: The meningioma trap. *Eur Ann Otorhinolaryngol Head Neck Dis* (2018) 135 (2):131–2. doi: 10.1016/j.anorl.2017.12.002

4. Ferguson SD, Musleh W, Gurbuxani S, Shafizadeh SF, Lesniak MS. Intracranial mucosa–associated lymphoid tissue (MALT) lymphoma. *J Clin Neurosci* (2010) 17(5):666–9. doi: 10.1016/j.jocn.2009.10.001

5. Choi JY, Chung JH, Park YJ, Jung GY, Yoon TW, Kim YJ, et al. Extranodal marginal zone b–cell lymphoma of mucosa–associated tissue type involving the dura. *Cancer Res Treat* (2016) 48(2):859–63. doi: 10.4143/crt.2014.334

6. Sanjeevi A, Krishnan J, Bailey PR, Catlett J. Extranodal marginal zone b–cell lymphoma of malt type involving the cavernous sinus. *Leuk Lymphoma* (2001) 42 (5):1133–7. doi: 10.3109/10428190109097736

7. Matmati KS, Matmati N, Hannun YA, Rumboldt Z, Patel S, Lazarchick J, et al. Dural MALT lymphoma with disseminated disease. *Hematol Rep* (2010) 2(1): e10. doi: 10.4081/hr.2010.e10

8. Neidert MC, Leske H, Burkhardt J–K, Rushing EJ, Bozinov O. A 44–year old Male with right–sided facial numbness: Correspondence. *Brain Pathol* (2015) 25 (1):113–4. doi: 10.1111/bpa.12234

9. Yang C–C, Chen T–Y, Tsui Y–K, Ko C–C. Primary marginal zone b–cell lymphoma of the cavernous sinus: a case report and review of the literature. *BMC Med Imaging* (2021) 21(1):25. doi: 10.1186/s12880–021–00556–w

10. Yeh CH, Tsui YK, Liu H, Chuang SS. Primary IgG4–producing extranodal marginal zone lymphoma of mucosa–associated lymphoid tissue (MALT lymphoma) in the cavernous sinus: A mimicker of IgG4–related disease/ hypertrophic pachymeningitis. *Pathol Int* (2021) 71(4):278–80. doi: 10.1111/ pin.13070

11. Raderer M, Kiesewetter B, Ferreri AJM. Clinicopathologic characteristics and treatment of marginal zone lymphoma of mucosa–associated lymphoid tissue (MALT lymphoma): Characteristics and management of MALT lymphoma. *CA Cancer J Clin* (2016) 66(2):152–71. doi: 10.3322/caac.21330

12. Louveau A, Smirnov I, Keyes TJ, Eccles JD, Rouhani SJ, Peske JD, et al. Structural and functional features of central nervous system lymphatic vessels. *Nature* (2015) 523(7560):337–41. doi: 10.1038/nature14432

13. Ercan N, Gultekin S, Celik H, Tali TE, Oner YA, Erbas G. Diagnostic value of contrast–enhanced fluid– attenuated inversion recovery MR imaging of intracranial metastases. *AJNR Am J Neuroradiol* (2004) 25(5):761–5.

14. Iwamoto FM, DeAngelis LM, Abrey LE. Primary dural lymphomas: A clinicopathologic study of treatment and outcome in eight patients. *Neurology* (2006) 66(11):1763–5. doi: 10.1212/01.wnl.0000218284.23872.eb

15. Griffiths PD, Coley SC, Romanowski CAJ, Hodgson T, Wilkinson ID. Contrast–enhanced fluid–attenuated inversion recovery imaging for leptomeningeal disease in children. *AJNR Am J Neuroradiol* (2003) 24(4):719–23.

16. Jiang S, Yu H, Wang X, Lu S, Li Y, Feng L, et al. Molecular MRI differentiation between primary central nervous system lymphomas and high–grade gliomas using endogenous protein–based amide proton transfer MR imaging at 3 Tesla. *Eur Radiol* (2016) 26(1):64–71. doi: 10.1007/s00330–015–3805–1

17. Joo B, Han K, Choi YS, Lee S–K, Ahn SS, Chang JH, et al. Amide proton transfer imaging for differentiation of benign and atypical meningiomas. *Eur Radiol* (2018) 28(1):331–9. doi: 10.1007/s00330–017–4962–1

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