HEAD AND NECK

A deep learning model combining multimodal radiomics, clinical and imaging features for differentiating ocular adnexal lymphoma from idiopathic orbital inflammation

Xiaoyang Xie[1] · Lijuan Yang[2] · Fengjun Zhao[1] · Dong Wang[2] · Hui Zhang[2] · Xuelei He[1] · Xin Cao[1] · Huangjian Yi[1] · Xiaowei He[1] · Yuqing Hou[1]

Abstract

Objectives To evaluate the value of deep learning (DL) combining multimodal radiomics and clinical and imaging features for differentiating ocular adnexal lymphoma (OAL) from idiopathic orbital inflammation (IOI).

Methods Eighty-nine patients with histopathologically confirmed OAL ($n = 39$) and IOI ($n = 50$) were divided into training and validation groups. Convolutional neural networks and multimodal fusion layers were used to extract multimodal radiomics features from the T1-weighted image (T1WI), T2-weighted image, and contrast-enhanced T1WI. These multimodal radiomics features were then combined with clinical and imaging features and used together to differentiate between OAL and IOI. The area under the curve (AUC) was used to evaluate DL models with different features under five-fold cross-validation. The Student t-test, chi-squared, or Fisher exact test was used for comparison of different groups.

Results In the validation group, the diagnostic AUC of the DL model using combined features was 0.953 (95% CI, 0.895–1.000), higher than that of the DL model using multimodal radiomics features (0.843, 95% CI, 0.786–0.898, $p < 0.01$) or clinical and imaging features only (0.882, 95% CI, 0.782–0.982, $p = 0.13$). The DL model built on multimodal radiomics features outperformed those built on most bimodalities and unimodalities ($p < 0.05$). In addition, the DL-based analysis with the orbital cone area (covering both the orbital mass and surrounding tissues) was superior to that with the region of interest (ROI) covering only the mass area, although the difference was not significant ($p = 0.33$).

Conclusions DL-based analysis that combines multimodal radiomics features with clinical and imaging features may help to differentiate between OAL and IOI.

Key Points

• *It is difficult to differentiate OAL from IOI due to the overlap in clinical and imaging manifestations.*
• *Radiomics has shown potential for noninvasive diagnosis of different orbital lymphoproliferative disorders.*
• *DL-based analysis combining radiomics and imaging and clinical features may help the differentiation between OAL and IOI.*

Keywords Deep learning · Magnetic resonance imaging · Lymphoma · Orbital pseudotumor · Differential diagnosis

Abbreviations

AUC	Area under the curve
DL	Deep learning
IOI	Idiopathic orbital inflammatory
OAL	Ocular adnexal lymphoma
OLPD	Orbital lymphoproliferative disorder
T1WI	T_1-weighted image
T1WI+C	Contrast-enhanced T1WI
T2WI	T_2-weighted image

✉ Lijuan Yang
ylijuan@126.com

✉ Fengjun Zhao
fjzhao@nwu.edu.cn

✉ Yuqing Hou
houyuqin@nwu.edu.cn

1 Xi'an Key Lab of Radiomics and Intelligent Perception, School of Information Science and Technology, Northwest University, Xi'an 710069, Shaanxi, China

2 Department of Radiology, Xi'an Fourth Hospital, Xi'an 710004, Shaanxi, China

🐾 Springer

Introduction

Ocular adnexal lymphoma (OAL) and idiopathic orbital inflammation (IOI) are two typical orbital lymphoproliferative disorders (OLPDs) [1, 2]. OAL is the most common primary orbital malignant tumor [3–6], accounting for about 1–2% of all non-Hodgkin's lymphomas [7, 8] and 8% of extranodal lymphomas [9–11]. IOI, also termed as orbital inflammatory pseudotumor, is the third most common disease of orbital inflammation [12–14], accounting for about 10% of all orbital masses [15–17]. OAL and IOI have considerable overlap in clinical and imaging manifestations [2, 14, 18], but the treatment options and prognosis are very different. The first-line treatment for OAL is low-dose radiation therapy [19, 20], while IOI is sensitive to corticosteroid therapy [21–23]. Pathological diagnosis with biopsy is the gold standard for differentiating between OAL and IOI, but it is invasive and may cause many complications, such as pain, bleeding, infection, and tumor spread [23–25]. In addition, biopsy relies on accurate localization of lesions, which is often troublesome for those lesions in the orbital apex and around the optic nerve [25, 26].

Magnetic resonance imaging (MRI) provides an option for non-invasive diagnosis of OLPDs [27, 28]. Some studies reported imaging findings including homogeneity [2, 27], signal intensity of T_1-weighted image (T1WI) and T_2-weighted image (T2WI) [2, 27, 29], tumor boundary [27, 30, 31], the "flow void sign" [30, 32], and pattern of enhancement [27], may be useful for assessing the manifestation of orbital diseases. The diffusion and perfusion parameters derived from intravoxel incoherent motion diffusion-weighted imaging (IVIM-DWI) further improved the diagnostic performance between different OLPDs [33]. Moreover, some clinical characteristics, such as age [34, 35], laterality [31], tumor shape [27, 29], and sinusitis [30], were also considered helpful to describe the differences between OAL and IOI. However, conclusions obtained from these findings are not completely consistent and are sometimes even contradictory to each other. For example, some studies demonstrated that IOI has a higher signal intensity in T1WI and T2WI [29, 31], more regular in shape [29], and more bilateral compared with OAL [31], while other studies reported that there are no significant differences in signal intensity [32], tumor shape [27, 32, 36], and laterality [30, 32] between OAL and IOI.

Radiomics, an image-based computer-aided diagnostic technique, shifts the visual assessment of radiologists to the high-throughput mining of quantitative features from medical images by machines and has demonstrated great potential in diagnosis, prognosis, and prediction of a wide range of diseases [24, 37, 38]. To date, radiomics features including first-order statistics (or intensity histograms), shape features, texture features [31, 34, 39], and bag-of-features (BOF)-based features [35] have been used to distinguish benign from malignant OLPDs, with results superior to or comparable with experienced radiologists' visual assessment [34, 35]. In addition, most radiomics features were extracted from multiple MR sequences, including T1WI, T2WI, and contrast-enhanced T1WI (T1WI + C) with [39] or without diffusion-weighted imaging (DWI) [31, 34, 35], which preliminary proved the prospect of multimodality in improving the accuracy of OLPD diagnosis. However, these radiomics studies required complex analysis steps, including handcrafted feature extraction, feature selection, and predictive model building, increasing the inconsistency of results across studies due to the stochastic nature of each step. Moreover, they only treated features from different MR sequences as added features and performed features selection by linear or simple algorithms, without investigating the complex nonlinear relationships between features from different modalities.

Deep learning (DL) has the capacity to learn effective representation directly from medical images and to couple feature extraction, feature selection, and predictive model building into one neural network model through end-to-end learning, thus greatly simplifying the process of radiomics analysis [40]. Currently, DL models have shown expert-level performances in various medical image-based diagnostic tasks, such as the differentiation of benign and malignant renal tumors [41], grading of non-small cell lung cancer [42], and prediction of lymph node metastasis in breast cancer [43]. However, to the best of our knowledge, the performance of DL models has not been evaluated in the differential diagnosis of OLPDs. In addition, current DL-based analysis mainly focuses on unimodal images, and further study is needed to take full advantage of the diverse features from different modalities.

In this study, we developed a DL model combining multimodal radiomics and clinical and imaging features for the differentiation of OAL and IOI. First, multimodal radiomics features were extracted from three MR sequences. Then, the multimodal radiomics features were combined with clinical and imaging features to differentiate OAL from IOI. The purpose of our study was to evaluate whether the DL model that employed the combined features could accurately differentiate benign from malignant OLPDs.

Materials and methods

Patients

This retrospective study was approved by our institutional review board and written informed consent from patients was waived. Personal information of all patient data was de-identified prior to analysis. From July 2014 and October 2020, a total of 105 consecutive patients with OLPDs were collected from the Radiology Department of Xi'an Fourth Hospital. The inclusion criteria for the study were (1) histopathologically

confirmed primary OAL or IOI by biopsy or surgery; (2) performing MRI examination less than 14 days before the surgical biopsy; (3) complete multimodal MRI data (i.e., T1WI, T2WI, and T1WI+C); (4) adequate quality of the images for analysis (without motion or artifacts). Finally, a total of 89 patients (39 OALs and 50 IOIs) were included in this study. Other details of patient inclusion and exclusion are shown in Fig. S1.

Imaging data acquisition

All patients underwent MRI examination with a 3.0-T MR scanner (Signa HDxt, GE Healthcare) equipped with an eight-channel high-resolution head coil. Fast spin-echo (FSE) T1WI, T2WI, and T1WI+C with fat saturation in axial, parasagittal (parallel to the optic nerve), and coronal planes were acquired. T1WI+C was performed after an intravenous bolus injection of 0.2 ml/kg gadolinium-DTPA (Magnevist). The MRI parameters are detailed in Table S1.

Analysis workflow

The multimodal MR images of 89 patients with OLPDs were randomly divided into two independent groups in a ratio of approximately 4:1 to obtain a training group (71 patients, 31 OALs and 40 IOIs) and a validation group (18 patients, 8 OALs and 10 IOIs) for the development and validation of the DL model, respectively. The workflow of the study consisted of region of interest (ROI) segmentation, multimodal radiomics feature extraction, clinical and imaging feature extraction, and the differentiation of OLPDs with the combined features (Fig. 1).

Region of interest segmentation

ROIs used for radiomics analysis were manually segmented from the axial MRI by two radiologists who were blinded to the histopathological results. A radiologist with 8 years of head and neck radiology experience checked the T1WI+C data on the Medical Imaging Interaction Toolkit (MITK) software (v.2015.5.0), and delineated both the OLPD mass and the entire orbital cone area (covering both the OLPD mass and surrounding tissues) (Fig. 2). Corresponding ROIs on T1WI and T2WI were mapped from the segmentation results on T1WI+C. A senior radiologist with 14 years of radiology experience checked and revised the segmented results. In patients with bilateral involvement, only the largest mass was segmented and used for subsequent analysis.

Multimodal radiomics feature extraction

The convolutional neural network (CNN) model was used to extract radiomics features from the ROI of each MR sequence with the following architecture: two convolution layers, two pooling layers, and one fully connected layer. Using three CNN models, radiomics features were extracted from three different MR sequences (T1WI, T2WI, and T1WI+C) and then fused by the MFL [44] module (Figs. 1 and S2) to generate the multimodal radiomics features. Theoretically, the MFL integrates the radiomics features of multiple MRIs in a probabilistic manner based on multinomial sampling; therefore, it can prevent over-learning of the features of a particular modality. In this way, we first randomly selected the radiomics features of each MRI, and then obtained multimodal radiomics features by summing the radiomics
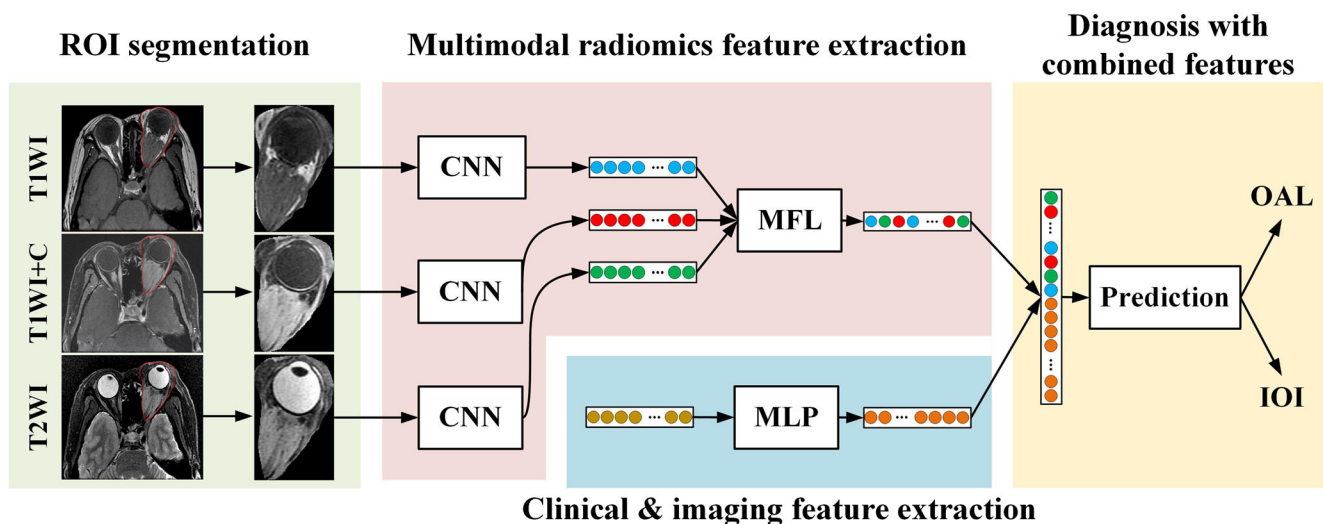
Fig. 1 Workflow of the DL model for differentiating IOI and OAL, where CNN, MFL, and MLP represent convolutional neural network, multimodality fusion layer, and multilayer perceptron, respectively
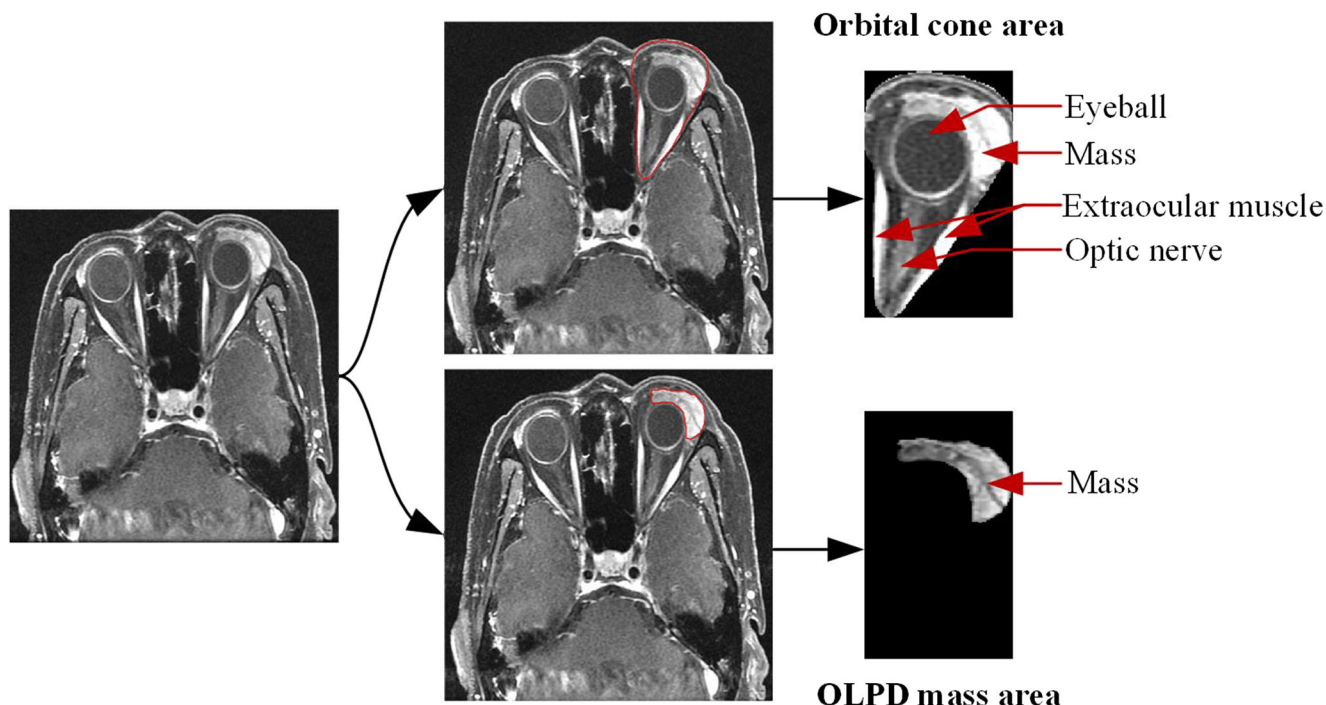
Fig. 2 Two different regions of interest (ROIs), including the orbital cone area (covering both the OLPD mass and surrounding tissues) and the OLPD mass area

features of multiple MR sequences. Other details of the CNN model and fusion process were summarized in Supplementary Methods.

To validate the performance of the MFL in the radiomics model, other multimodality fusion strategies including feature concatenation and decision fusion were also conducted. Feature concatenation stringed together the radiomics features extracted from each MR sequence to obtain the fused multimodality features, while decision fusion was performed by voting the predictive results acquired by the radiomics features on each MR sequence.

Clinical and imaging feature extraction

Clinical characteristics of each patient were first collected including age, gender, and laterality that might contribute to the differentiation of OLPDs. Then, imaging characteristics for routine diagnosis were also recorded. Specifically, two radiologists (same as the two in the ROI segmentation) independently evaluated the multimodal MR images of each patient from the following aspects: involved quadrants, involvement of the orbital area, shape, and borders of the lesions, the signal intensity on T1WI and T2WI, enhancement pattern on T1WI+C and homogeneity (Table 1). Considering that the clinical and imaging characteristics may not be directly relevant

for differentiating OLPDs, a multilayer perceptron (MLP) was used to generate the high-level representation of these characteristics, which consists of three fully connected layers with 64, 32, and 16 hidden units, respectively.

Differentiation of OLPDs with different features

In the diagnosis part of the DL model (Fig. 1), the extracted multimodal radiomics features were firstly concatenated with the clinical and imaging features, and then the combined features were passed through a fully connected layer and an output layer with the softmax activation function to finalize the distinction between OAL and IOI. To validate the performance of the DL model with the combined features (DL-both), DL models with only multimodal radiomics features (DL-rad) or clinical and imaging features (DL-clin) were also used to distinguish between OAL and IOI.

It is worth noting that the DL-both and DL-rad models were trained on MR slices of all patients in the training group, which actually increased the sample size of DL models (from 71 to 541) and alleviated the issue of overfitting. In the inference phase, the prediction results were firstly acquired by applying the DL model on all slices in the validation group, and then the diagnostic result of each patient was obtained by majority voting the prediction results on all the slices of this patient.

Table 1 Clinical and imaging characteristics of OAL and IOI patients

Parameters	Characters	IOI (n = 50)	OAL (n = 39)	p value
Age, mean (SD)		51.3 (16.9)	63.4 (13.4)	< 0.001
Gender	Female	27	15	0.14
	Male	23	24	
Laterality	Unilateral	41	38	0.022
	Bilateral	9	1	
Involved quadrants	Upper inner eye	26	18	0.49
	Lower inner eye	28	13	
	Upper outer eye	24	21	
	Lower outer eye	21	11	
Involvement of the orbital area	Anterior orbit preseptal space	27	25	0.81
	Intramuscular cone	17	12	
	Extraconal space	18	20	
	Lacrimal area	8	6	
Shape	Irregular	46	33	0.27
	Regular	4	6	
Border	Well defined	5	8	0.16
	Ill defined	45	31	
Signal intensity on T1WI	Low	2	2	0.91
	Iso	45	34	
	High	3	3	
Signal intensity on T2WI	Low	3	1	0.21
	Iso	10	14	
	High	37	24	
Enhancement pattern on T1WI+C	Mild	12	14	< 0.001
	Moderate	3	17	
	Significant	35	8	
Homogeneity	Homogeneous	17	38	< 0.001
	Heterogeneous	33	1	

The signal intensity on T1WI and T2WI was compared with that of extraocular muscle

Performance evaluation and statistical analysis

Using fivefold cross-validation, the performance of each DL model was evaluated by averaging the results over five times. The receiver operating characteristic (ROC) curve and the area under the curve (AUC) were used to compare the DL models with different features, ROIs, and multimodality fusion strategies. Other metrics including the accuracy (ACC), sensitivity (SEN), specificity (SPE), positive predictive value (PPV), and negative predictive value (NPV) were also calculated with the cut-off value determined by the maximum Youden index.

Statistical analysis was performed using SPSS software (version 25.0). Student t-test was applied for comparison of continuous variables and different models, and Pearson chi-squared test or Fisher exact test was used for comparison of categorical variables. In all the analyses, p values less than 0.05 were considered statistically significant.

Results

Clinical and imaging characteristics

The clinical and imaging characteristics of the 89 patients enrolled in this study were summarized in Table 1. Compared with IOI, OAL was more likely to occur in elderly patients (p < 0.001) and tended to develop unilaterally (p = 0.022). Most OAL patients have moderate or mild degree of enhancement in T1WI+C, while most IOI patients have significant or mild degree of enhancement (p < 0.001). In addition, the MR images of OAL were more homogeneous in texture compared with those of IOI (p < 0.001).

Evaluation of different ROIs in MRI

The results in Tables 2 and S2 showed that the performance of the DL-rad model with the two ROIs was similar in the

Table 2 Diagnostic performance of different ROIs in the training and validation groups

ROI	Group	AUC (95% CI)	ACC	SEN	SPE	NPV	PPV
Mass	Training	0.978 (0.924, 1.000)	96.0%	98.0%	94.5%	98.2%	93.8%
	Validation	0.801 (0.714, 0.889)	78.8%	70.0%	86.0%	79.3%	82.8%
Cone	Training	0.966 (0.935, 0.997)	94.1%	92.9%	95.0%	94.7%	94.0%
	Validation	0.843 (0.786, 0.898)	84.4%	85.0%	84.0%	89.4%	83.4%

Mass and cone represent the OLPD mass area and the entire orbital cone area, respectively

training group, but the diagnostic AUC of the cone area was better than that of the mass area (0.843 vs. 0.801 for patients, and 0.696 vs. 0.679 for slices) in the validation group. Other metrics including ACC, SEN, NPV, and PPV of the orbital cone area were also higher than those of OLPD mass area. Therefore, the orbital cone area was selected as the preferred ROI for multimodal radiomics analysis even though there was no significant difference in AUC between the two ROIs ($p = 0.33$ for patients, and $p = 0.63$ for slices).

Evaluation of different multimodality fusion strategies

In the validation group, the MFL strategy was able to achieve better performance on all metrics compared to the other two fusion methods (Table 3). At the slice level, all metrics of the MFL except for SPE were higher than those of feature concatenation and decision fusion (Table S3). In addition, relatively large differences caused by overfitting between the training and validation groups in the other fusion strategies were mitigated by the MFL strategy. Therefore, we used MFL to fuse the radiomics features of three MR sequences, even though the AUC was not significantly different.

Evaluation of multimodal radiomics features

The optimal result of the DL-rad model was achieved with the multimodal radiomics features incorporating T1WI, T1WI+C, and T2WI (Table 4 and Fig. 3), with significantly better AUC values than those of bimodalities (T1C-T2, T1-T2) and

unimodalities (T1WI and T2WI) ($p < 0.05$). The slice-level AUC of multimodality was also significantly better than T1-T2, T1WI, T1WI+C, and T2WI ($p < 0.05$) (Table S4 and Fig. S3). Meanwhile, the diagnostic AUC of bimodal radiomics features was also better than that of corresponding unimodal radiomics features, except for the T1C-T2 bimodality, which had a slightly lower performance than the T1WI+C modality. Most remaining metrics for multimodality were superior to those for bimodalities and unimodalities in both patient and slice levels.

Differentiation of OAL and IOI with different features

The diagnostic results of DL models with the combined features (DL-both), multimodal radionics features (DL-rad), and clinical and imaging features (DL-clin) are given in Tables 5 and S5 and Figs. 4 and S4. In the validation group, the patient- and slice-level AUCs of the DL-both/DL-rad model were 0.953 (95% CI, 0.895–1.000)/0.843 (95% CI, 0.786–0.898), and 0.909 (95% CI, 0.805–1.000)/0.696 (95% CI, 0.662–0.730), respectively. This indicated that the combined features significantly improved the differentiation performance ($p < 0.01$ for patients, and $p < 0.001$ for slices). Superiority of the DL-both over DL-rad model was also found in other metrics, such as the slice-level ACC, NPV, and PPV ($p < 0.05$). In addition, all the metrics of the DL-both model were also higher than those of the DL-clin model in the patient level (although there was no significant difference), which validated the effectiveness of the DL-both model with the combined features.

Table 3 Diagnostic performance of different fusion strategies in the training and validation groups

Strategy	Group	AUC (95% CI)	ACC	SEN	SPE	NPV	PPV
MFL (ours)	Training	0.966 (0.935, 0.997)	94.1%	92.9%	95.0%	94.7%	94.0%
	Validation	0.843 (0.786, 0.898)	84.4%	85.0%	84.0%	89.4%	83.4%
Concat	Training	0.987 (0.973, 1.000)	96.1%	94.8%	97.0%	96.1%	96.0%
	Validation	0.810 (0.722, 0.898)	80.2%	80.0%	78.0%	84.9%	77.3%
Decision	Training	1.000 (1.000,1.000)	100%	100%	100%	100%	100%
	Validation	0.809 (0.714, 0.905)	80.6%	81.2%	80.0%	85.1%	77.7%

MFL, concat and decision represent the multimodal fusion layer, feature concatenation, and decision fusion, respectively

Table 4 Diagnostic performance of different modalities in the validation group

Modality	AUC (95% CI)	ACC	SEN	SPE	NPV	PPV
T1	0.722 (0.618, 0.826)	74.4%	70.0%	78.0%	77.9%	72.5%
T2	0.710 (0.563, 0.857)	72.2%	80.0%	66.0%	81.1%	65.9%
T1C	0.787 (0.685, 0.889)	78.9%	81.2%	82.0%	84.8%	82.6%
T1-T2	0.732 (0.628, 0.836)	76.7%	75.0%	78.0%	81.9%	77.3%
T1-T1C	0.795 (0.702, 0.887)	78.9%	82.5%	76.0%	86.1%	73.7%
T1C-T2	0.773 (0.717, 0.827)	78.8%	67.5%	88.0%	77.7%	83.8%
T1-T1C-T2	0.843 (0.786, 0.898)	84.4%	85.0%	84.0%	89.4%	83.4%

T1, T2, and T1C represent T1WI, T2WI, and T1WI+C, respectively

Discussion

We built a DL model trained by both the multimodal radiomics features and clinical and imaging features for the differentiation of OAL and IOI, which performed well in both the training and testing groups, outperforming models trained by multimodal radiomics features or clinical and imaging features. This suggested that both the radiomics features and clinical and imaging information were associated with the types of OLPDs. In addition, the relatively high sensitivity (SEN = 92.5%) and negative predictive value (NPV = 94.1%) in the validation group indicated that the false-negative rate of the differentiation has been greatly depressed. This is particularly beneficial for OAL patients, as most of them can be precisely identified by MRI and will receive early follow-up treatment.

To investigate the diagnostic performance of different DL models, we counted the number of correctly and incorrectly predicted patients in the fold with median AUC (Fig. 5). There are three, three and one incorrectly predicted patients for the
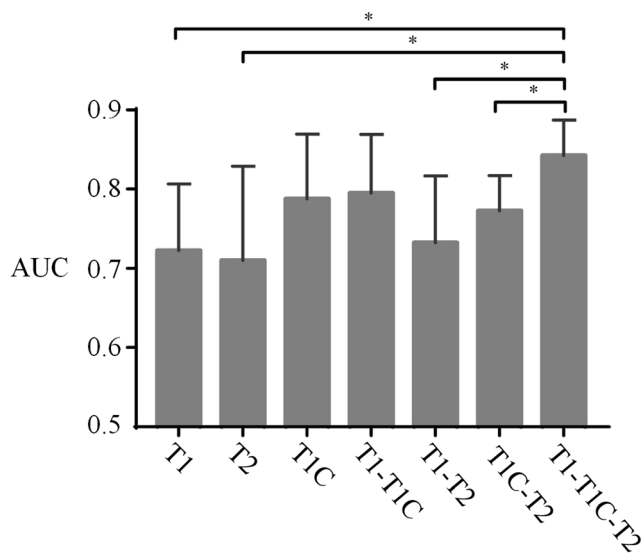
Fig. 3 Diagnostic AUC of the DL-rad model in the validation group with different modalities, where T1, T2, and T1C represent T1WI, T2WI, and T1WI+C, respectively

DL-clin, DL-rad, and DL-both models, respectively. The case in Fig. 6a was small in size and irregular in shape, leading to the failure of the DL-clin model. The case in Fig. 6b belonged to a diffuse IOI lesion, which was less common and the changes in the surrounding orbital structures were very similar to OAL, resulting in the mistake of the DL-rad model. The case in Fig. 6c had homogenous intensity in T1WI+C and T2WI and clear boundary in T1WI+C, so the predicted results of the DL-clin and DL-rad models were more biased toward OAL, but the DL-both model was able to correctly identify it as IOI. The case in Fig. 6d was an atypical lymphoid tissue hyperplasia with the potential to transform into OAL at a later stage, which may be the reason for the misclassification of DL-both model.

This is the first study to deploy DL-based analysis to distinguish between different OLPDs, which can simplify radiomics analysis by extracting features directly from MR images through end-to-end learning. To roughly assess the DL-based analysis, we compared the results of the DL model with those of radiomics analyses. Specifically, Guo et al used T2WI and T1WI+C to train a radiomics model for predicting OLPDs and obtained an AUC of 0.73 (95% CI, 0.65–0.88) [34], while our DL-rad model, also built on T2WI and T1WI+C (Table 4), elevated the AUC value to 0.773 (95% CI, 0.717–0.827). The results of our DL-rad model on T1WI+C sequence were comparable to those of the BOF-based radiomics analysis [35], with the AUC of 0.787 (95% CI, 0.685–0.889) and 0.803 (95% CI, 0.725–0.880), respectively. Furthermore, DL-based analysis showed that the T1WI+C was the most valuable modality to distinguish OAL from IOI, which was consistent with previous radiomics analyses [34, 35].

Radiomics analyses have shown the advantages of multimodality over unimodality in the differentiation of OLPDs, and the performance became even better with increasing MR sequences. For example, the AUCs obtained by radiomics analysis on T2-T1C [34], T1-T2-T1C [31], and T1-T2-T1C-DWI [39] were 0.73, 0.78, and 0.87, respectively. In our study, the DL model built on multimodalities also performed better than the models built on bimodalities and unimodalities (Table 4 and Fig. 3). In addition, the MFL-based multimodal

Table 5 Diagnostic performance of different models in the training and validation groups

Model	Group	AUC (95% CI)	ACC	SEN	SPE	NPV	PPV
DL-both	Training	0.998 (0.993, 1.000)	99.2%	99.4%	99.0%	99.5%	98.7%
	Validation	0.953 (0.895, 1.000)	91.1%	92.5%	90.0%	94.1%	88.3%
DL-rad	Training	0.966 (0.935, 0.997)	94.1%	92.9%	95.0%	94.7%	94.0%
	Validation	0.843 (0.786, 0.898)	84.4%	85.0%	84.0%	89.4%	83.4%
DL-clin	Training	0.966 (0.938, 0.995)	91.8%	89.0%	94.0%	92.0%	92.0%
	Validation	0.882 (0.782, 0.982)	87.7%	90.0%	86.0%	92.1%	84.7%

Fig. 4 Diagnostic ROC curves of different models in the training group (**a**) and the validation group (**b**), where the DL-both, DL-rad, and DL-clin represent the DL models built with the combined features, multimodal radiomics features and clinical and imaging features, respectively
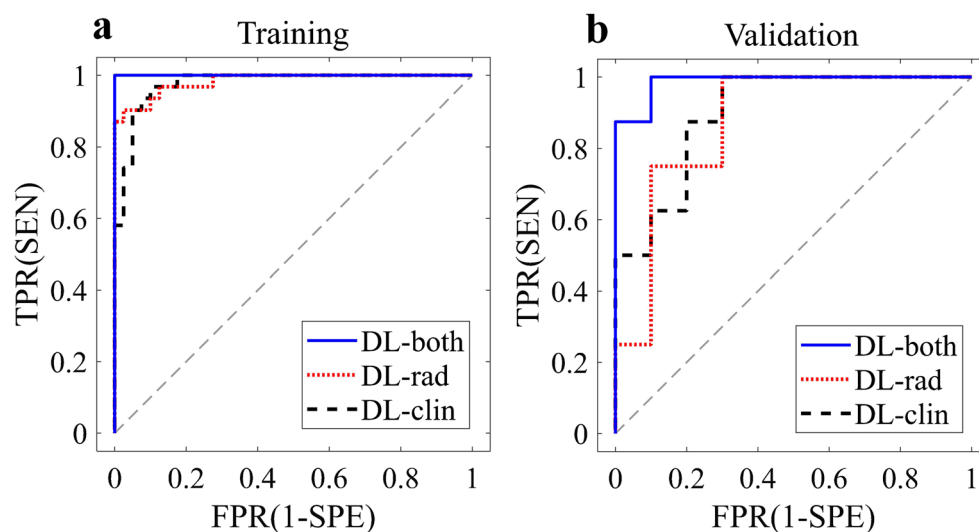
fusion effectively considered the correlation between different modalities and discarded redundant features during the training process, similar to the dropout operation used to prevent overfitting [44]. Therefore, the MFL strategy achieved better performance than the feature concatenation and decision fusion, even though there were no significant differences.

DL-based analysis with the orbital cone ROI covering both the OLPD mass and surrounding tissues performed better than the ROI covering only the OLPD mass (Table 2). This may be reasonable because different types of OLPDs have different patterns of invasion into surrounding orbital tissues. Specifically, OAL usually surrounds orbital structures, such as eyeball walls, extraocular muscles, and optic nerves, but does not cause the deformation of these structures [2, 36, 45]. However, IOI often invades lacrimal glands, causing thickening of eye rings and extraocular muscles [14, 15, 23]. Therefore, the orbital cone area provides more diagnostically relevant information than the ROI covering only the mass area. Similar conclusions can also be found in the radiomics analysis of other tumors [46, 47].

There are some limitations to this study. First, the sample size was relatively limited because of the small number of patients with pathologically confirmed OAL or IOI and the absence of some MR sequences. How to use incomplete

Fig. 5 Diagnostic performance of different DL models in the validation group with 18 patients. Symbols × and √ indicate that the predictive results were wrong and correct, respectively

DL-Clin	DL-Rad	DL-both	Count
√	√	√	12
×	√	√	2
√	×	√	2
×	×	√	1
√	√	×	1
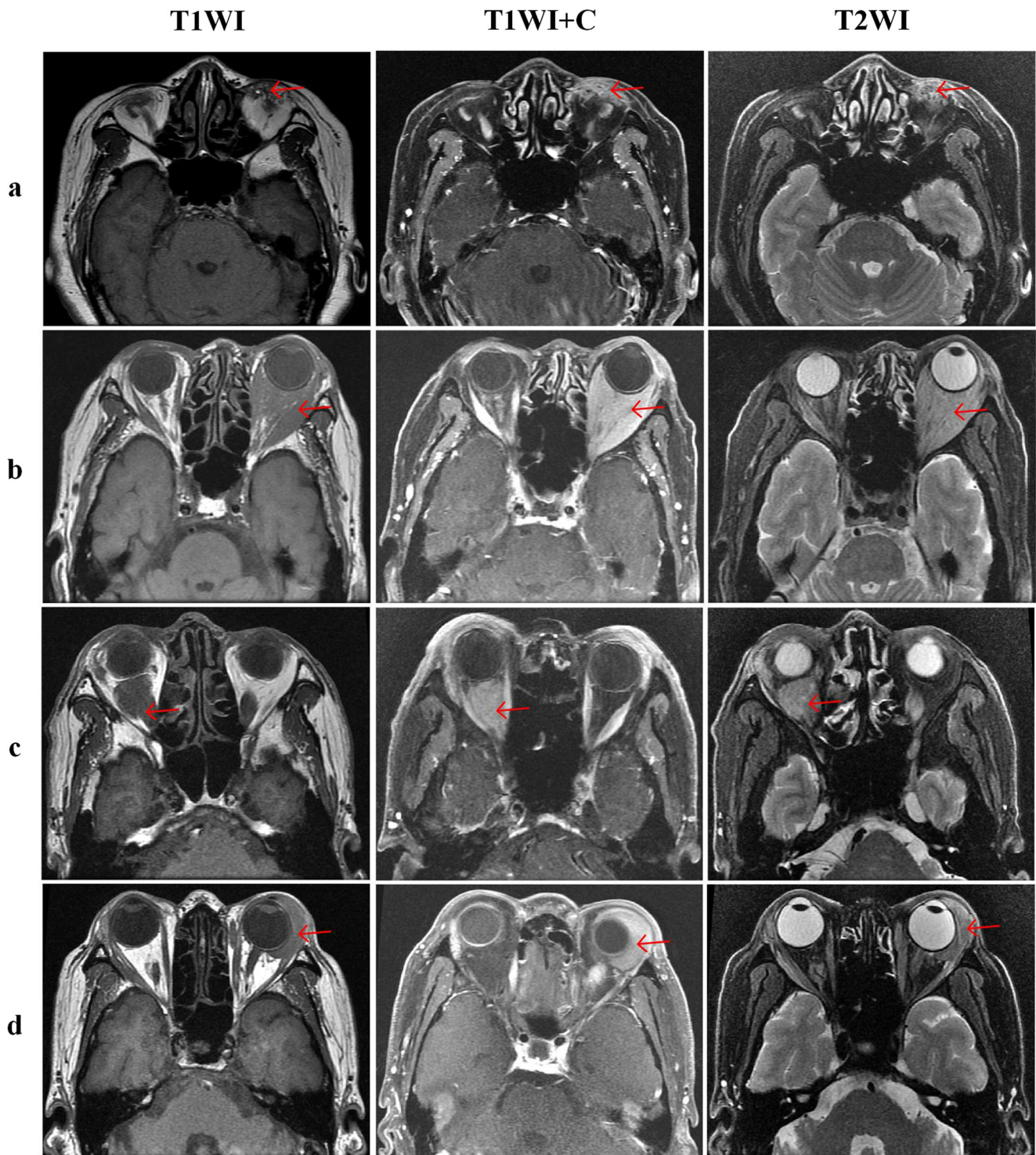
Fig. 6 Four typical cases of misclassification. **a** A 60-year-old woman with IOI was correctly predicted by DL-both and DL-rad models, but the DL-clin model failed. **b** An 82-year-old woman with IOI was correctly predicted by DL-both and DL-clin models, but the DL-rad model failed. **c** A 61-year-old man with IOI was correctly predicted by the DL-both model, but the DL-clin and DL-rad models failed. **d** A 54-year-old man with IOI was correctly predicted by DL-clin and DL-rad models, but the DL-both model failed

multimodal data for radiomics analysis of OLPDs is the focus of our future research. Second, due to the limited amount of samples, we only divided the data into independent training and validation groups to build and validate the DL-based model, respectively. Nevertheless, external validation is helpful to further validate the developed model and will be performed in the future as more samples are collected. Third, similar to most radiomics studies [31, 34, 35], the DL-based

analysis was only performed on routinely used MR sequences, such as the T1WI, T2WI, and T1WI+C. We recently noticed the potential of DWI with apparent diffusion coefficient (ADC) and dynamic contrast-enhanced MRI (DCE-MRI) in providing additional functional information on orbital lesions [48, 49]. Therefore, the addition of DWI and DCE-MRI may further improve diagnostic accuracy if these functional imagings are routinely performed in the future.

In conclusion, we proposed a DL-based model for OLPD diagnosis, which extracted radiomics features from multimodal MRIs by CNNs and fused these features with the MFL module. The obtained multimodal radiomics features were then combined with the clinical and imaging features extracted by the MLP and used together to differentiate between different OLPDs. The promising results suggest that the developed DL-based analysis, combining multimodal radiomics and clinical and imaging features, may be used as a differential diagnostic tool for OAL and IOI.

Declarations

Guarantor The scientific guarantor of this publication is Fengjun Zhao.

Conflict of interest The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Statistics and biometry One of the authors has significant statistical expertise.

Informed consent Written informed consent was waived by the Institutional Review Board.

Ethical approval Institutional Review Board approval was obtained.

Methodology
• retrospective
• diagnostic or prognostic study
• performed at one institution

References

1. Li EY, Yuen HK, Cheuk W (2015) Lymphoproliferative Disease of the Orbit. Asia Pac J Ophthalmol (Phila) 4:106–111
2. Cytryn AS, Putterman AM, Schneck GL, Beckman E, Valvassori GE (1997) Predictability of magnetic resonance imaging in differentiation of orbital lymphoma from orbital inflammatory syndrome. Ophthal Plast Reconstr Surg 13:129–134
3. Bairey O, Kremer I, Rakowsky E, Hadar H, Shaklai M (1994) Orbital and adnexal involvement in systemic non-Hodgkin's lymphoma. Cancer 73:2395–2399
4. Ferry JA, Fung CY, Zukerberg L et al (2007) Lymphoma of the Ocular Adnexa: A Study of 353 Cases. Am J Surg Pathol 31:170–184
5. Watkins LM, Carter KD, Nerad JA (2011) Ocular adnexal lymphoma of the extraocular muscles: case series from the University of Iowa and review of the literature. Ophthal Plast Reconstr Surg 27:471–476
6. Ponzoni M, Govi S, Licata G et al (2013) A reappraisal of the diagnostic and therapeutic management of uncommon histologies of primary ocular adnexal lymphoma. Oncologist 18:876–884
7. Eckardt AM, Lemound J, Rana M, Gellrich NC (2013) Orbital lymphoma: diagnostic approach and treatment outcome. World J Surg Oncol 11:73
8. Sniegowski MC, Roberts D, Bakhoum M et al (2014) Ocular adnexal lymphoma: validation of American Joint Committee on Cancer seventh edition staging guidelines. Br J Ophthalmol 98:1255–1260
9. Coupland SE, Krause L, Delecluse H-J et al (1998) Lymphoproliferative lesions of the ocular adnexa. Ophthalmology 105:1430–1441
10. Fung CY, Tarbell NJ, Lucarelli MJ et al (2003) Ocular adnexal lymphoma: Clinical behavior of distinct World Health Organization classification subtypes. Int J Radiat Oncol Biol Phys 57:1382–1391
11. Sullivan TJ, Whitehead K, Williamson R et al (2005) Lymphoproliferative disease of the ocular adnexa: a clinical and pathologic study with statistical analysis of 69 patients. Ophthal Plast Reconstr Surg 21:177–188
12. William M, Lessner AM (2010) Orbital pseudotumor. Am J Clin Oncol 33:304–306
13. Rubin PA, Foster CS (2004) Etiology and management of idiopathic orbital inflammation. Am J Ophthalmol 138:1041–1043
14. Rosenbaum JT, Phal PM, Lutt JR, Lim LL (2008) Orbital inflammatory disease. Semin Arthritis Rheum 37:207–222
15. Lee EJ, Jung SL, Kim BS et al (2005) MR Imaging of orbital inflammatory pseudotumors with extraorbital extension. Korean J Radiol 6:82–88
16. Gordon LK (2006) Orbital inflammatory disease: a diagnostic and therapeutic challenge. Eye (Lond) 20:1196–1206
17. Swamy BN, McCluskey P, Nemet A et al (2007) Idiopathic orbital inflammatory syndrome: clinical features and treatment outcomes. Br J Ophthalmol 91:1667–1670
18. Akansel G, Hendrix L, Erickson BA et al (2005) MRI patterns in orbital malignant lymphoma and atypical lymphocytic infiltrates. Eur J Radiol 53:175–181
19. Roshdy N, Shahin M, Kishk H et al (2010) MRI in diagnosis of orbital masses. Curr Eye Res 35:986–991
20. Kalogeropoulos D, Papoudou-Bai A, Kanavaros P, Kalogeropoulos C (2018) Ocular adnexal marginal zone lymphoma of mucosa-associated lymphoid tissue. Clin Exp Med 18:151–163
21. Uehara F, Ohba N (2002) Diagnostic imaging in patients with orbital cellulitis and inflammatory pseudotumor. Int Ophthalmol Clin 42:133–142
22. Gordon LK (2003) Diagnostic dilemmas in orbital inflammatory disease. Ocul Immunol Inflamm 11:3–15
23. Brannan P (2007) A review of sclerosing idiopathic orbital inflammation. Curr Opin Ophthalmol 5:402–404
24. Lambin P, Rios-Velazquez E, Leijenaar R et al (2012) Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer 48:441–446

25. Woolf DK, Ahmed M, Plowman PN (2012) Primary lymphoma of the ocular adnexa (orbital lymphoma) and primary intraocular lymphoma. Clin Oncol (R Coll Radiol) 24:339–344

26. Decaudin D, de Cremoux P, Vincent-Salomon A, Dendale R, Rouic LL (2006) Ocular adnexal lymphoma: a review of clinicopathologic features and treatment options. Blood 108:1451–1460

27. Xian J, Zhang Z, Wang Z et al (2010) Value of MR imaging in the differentiation of benign and malignant orbital tumors in adults. Eur Radiol 20:1692–1702

28. Ferreira TA, Saraiva P, Genders SW, Buchem MV, Luyten GPM, Beenakker JW (2018) CT and MR imaging of orbital inflammation. Neuroradiology 60:1253–1266

29. Ben Simon GJ, Annunziata CC, Fink J, Villablanca P, McCann JD, Goldberg RA (2005) Rethinking orbital imaging establishing guidelines for interpreting orbital imaging studies and evaluating their predictive value in patients with orbital tumors. Ophthalmology 112:2196–2207

30. Haradome K, Haradome H, Usui Y et al (2014) Orbital lymphoproliferative disorders (OLPDs): value of MR imaging for differentiating orbital lymphoma from benign OPLDs. AJNR Am J Neuroradiol 35:1976–1982

31. Yuan Y, Chu G, Gong T et al (2021) To explore MR imaging radiomics for the differentiation of orbital lymphoma and IgG4-related ophthalmic disease. Biomed Res Int 2021:6668510

32. Xu XQ, Hu H, Liu H et al (2017) Benign and malignant orbital lymphoproliferative disorders: Differentiating using multiparametric MRI at 3.0T. J Magn Reson Imaging 45:167–176

33. Jiang H, Wang S, Li Z et al (2020) Improving diagnostic performance of differentiating ocular adnexal lymphoma and idiopathic orbital inflammation using intravoxel incoherent motion diffusion-weighted MRI. Eur J Radiol 130:109191

34. Guo J, Liu Z, Shen C et al (2018) MR-based radiomics signature in differentiating ocular adnexal lymphoma from idiopathic orbital inflammation. Eur Radiol 28:3872–3881

35. Hou Y, Xie X, Chen J et al (2020) Bag-of-features-based radiomics for differentiation of ocular adnexal lymphoma and idiopathic orbital inflammation from contrast-enhanced MRI. Eur Radiol 31:24–33

36. Klingenstein A, Garip-Kuebler A, Priglinger S, Hintschich C, Mueller-Lisse UG (2021) Morphologic cross-sectional imaging features of IgG4-related orbitopathy in comparison to ocular adnexal lymphoma. Clin Ophthalmol 15:1119–1127

37. Kumar V, Gu Y, Basu S et al (2012) Radiomics: the process and the challenges. Magn Reson Imaging 30:1234–1248

38. Visvikis D, Cheze Le Rest C, Jaouen V, Hatt M (2019) Artificial intelligence, machine (deep) learning and radio(geno)mics: definitions and nuclear medicine imaging applications. Eur J Nucl Med Mol Imaging 46:2630–2637

39. Duron L, Heraud A, Charbonneau F et al (2021) A magnetic resonance imaging radiomics signature to distinguish benign from malignant orbital lesions. Invest Radiol 56:173–180

40. Afshar P, Mohammadi A, Plataniotis KN, Oikonomou A, Benali H (2019) From handcrafted to deep-learning-based cancer radiomics: challenges and opportunities. IEEE Signal Process Mag 36:132–160

41. Zhou L, Zhang Z, Chen Y-C, Zhao Z-Y, Yin X-D, Jiang H-B (2019) A deep learning-based radiomics model for differentiating benign and malignant renal tumors. Transl Oncol 12:292–300

42. Hosny A, Parmar C, Coroller TP et al (2018) Deep learning for lung cancer prognostication: a retrospective multi-cohort radiomics study. PLoS Med 15:e1002711

43. Sun H, Zeng X, Xu T, Peng G, Ma Y (2020) Computer-aided diagnosis in histopathological images of the endometrium using a convolutional neural network and attention mechanisms. IEEE J Biomed Health Inform 24:1664–1676

44. Choi J-H, Lee J-S (2019) EmbraceNet: a robust deep learning architecture for multimodal classification. Inf Fusion 51:259–270

45. Sullivan TJ, Valenzuela AA (2006) Imaging features of ocular adnexal lymphoproliferative disease. Eye (Lond) 20:1189–1195

46. Xu L, Wan Y, Luo C et al (2021) Integrating intratumoral and peritumoral features to predict tumor recurrence in intrahepatic cholangiocarcinoma. Phys Med Biol 66:125001

47. Hu Y, Xie C, Yang H et al (2020) Assessment of intratumoral and peritumoral computed tomography radiomics for predicting pathological complete response to neoadjuvant chemoradiation in patients with esophageal squamous cell carcinoma. JAMA Netw Open 3:e2015927

48. Politi LS, Forghani R, Godi C et al (2010) Ocular Adnexal Lymphoma: Diffusion-weighted MR Imaging for Differential Diagnosis and Therapeutic Monitoring. Radiology 256:565–574

49. Hu H, Xu X-Q, Liu H, Hong X-N, Shi H-B, Wu F-Y (2017) Orbital benign and malignant lymphoproliferative disorders: differentiation using semi-quantitative and quantitative analysis of dynamic contrast-enhanced magnetic resonance imaging. Eur J Radiol 88:88–94