European Journal of Radiology 170 (2024) 111250

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Meningioma consistency assessment based on the fusion of deep learning features and radiomics features

Jiatian Zhang [a,c,1], Yajing Zhao [b,1], Yiping Lu [b], Peng Li [c], Shijie Dang [c], Xuanxuan Li [b], Bo Yin [b,*], Lingxiao Zhao [c,*]

[a] School of Biomedical Engineering (Suzhou), Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei 230022, China
[b] Department of Radiology, Huashan Hospital, Fudan University, 12 Wulumuqi Rd. Middle, Shanghai 200040, China
[c] Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou 215163, China

ARTICLE INFO

Keywords:
Meningioma
Tumor consistency
Classification
Radiomics
Deep learning

ABSTRACT

Purpose: This study aims to combine deep learning features with radiomics features for the computer-assisted preoperative assessment of meningioma consistency.

Methods: 202 patients with surgery and pathological diagnosis of meningiomas at our institution between December 2016 and December 2018 were retrospectively included in the study. The T2-fluid attenuated inversion recovery (T2-Flair) images were evaluated to classify meningioma as soft or hard by professional neurosurgeons based on Zada's consistency grading system. All the patients were split randomly into a training cohort (n = 162) and a testing cohort (n = 40). A convolutional neural network (CNN) model was proposed to extract deep learning features. These deep learning features were combined with radiomics features. After multiple feature selections, selected features were used to construct classification models using four classifiers. AUC was used to evaluate the performance of each classifier. A signature was further constructed by using the least absolute shrinkage and selection operator (LASSO). A nomogram based on the signature was created for predicting meningioma consistency.

Results: The logistic regression classifier constructed using 17 radiomics features and 9 deep learning features provided the best performance with a precision of 0.855, a recall of 0.854, an F1-score of 0.852 and an AUC of 0.943 (95 % CI, 0.873–1.000) in the testing cohort. The C-index of the nomogram was 0.822 (95 % CI, 0.758–0.885) in the training cohort and 0.943 (95 % CI, 0.873–1.000) in the testing cohort with good calibration. Decision curve analysis further confirmed the clinical usefulness of the nomogram for predicting meningioma consistency.

Conclusions: The proposed method for assessing meningioma consistency based on the fusion of deep learning features and radiomics features is potentially clinically valuable. It can be used to assist physicians in the preoperative determination of tumor consistency.

1. Introduction

Worldwide, meningioma is one of the most common intracranial brain tumors [1]. Current therapeutic options for treating meningiomas mainly include surgery, radiotherapy and chemotherapy. For asymptomatic meningiomas, routine follow-up is preferable for small tumors with a suggested cutoff of 3 cm [2]. For meningiomas with large size or invasive manifestations, surgery is chosen as the first-line treatment. Before surgery, a comprehensive plan must be made by a group of neurosurgeons and neuroradiologists, including the patient's clinical information, tumor features, surgery-related factors, etc. [2].

Meningioma consistency, defined as the mechanical firmness of the tumor tissue, is extremely important for selecting the surgical approach and evaluating the prognosis [3–5]. Soft tumors can be removed by cutting and aspiration. It is more difficult to remove hard tumors, especially skull base meningiomas, which are associated with complex and critical neural and vascular structures. Surgery of hard tumors requires additional instruments, such as ultrasonic aspiration and

* Corresponding authors.
E-mail addresses: yinbo@fudan.edu.cn (B. Yin), hitic@sibet.ac.cn (L. Zhao).
[1] Jiatian Zhang and Yajing Zhao share the first authorship.

https://doi.org/10.1016/j.ejrad.2023.111250
Received 5 December 2022; Received in revised form 27 September 2023; Accepted 30 November 2023
Available online 5 December 2023

electrophysiological monitoring [6]. Consequently, a comprehensive and accurate preoperative prediction of meningioma consistency can greatly facilitate appropriate surgical planning. Currently, the main method for predicting the consistency of an intracranial tumor is to use magnetic resonance sequences, including conventional magnetic resonance imaging (MRI), magnetic resonance spectroscopy (MRS) and magnetic resonance elastography (MRE). The conventional MRI scanning sequences for meningioma include T1-weighted imaging (T1WI), T2-weighted imaging (T2WI), T1-weighted contrast-enhanced imaging (T1C), T2-Flair and diffusion-weighted imaging (DWI). The T2-Flair sequence is indispensable for central nervous system scanning, which effectively suppresses high signal in the cerebrospinal fluid to obtain better tissue contrast. It is beneficial for the detection of brain lesions. Fig. 1 shows T2-Flair slices containing hard meningiomas and soft meningiomas. It is difficult to distinguish the consistency of the tumor with the human eyes directly from the original MRI slices. A few papers have described the correlation between the meningioma consistency and the signal intensity of T2WI or T2-Flair. For example, the ratio between the signal intensity of the tumor and that of the healthy grey matter of the brain can be used to predict meningioma consistency. However, the validity of such a prediction is still controversial [1,4,5,7–10]. The apparent diffusion coefficient (ADC) value of DWI is also correlated with tumor consistency. Our previous study suggested that tumor consistency correlates with the ratio of ADC values between pituitary adenoma and normal brain tissue, and the ADC ratio decreases with increasing collagen content and predicts hard consistency of tumors for ADC $<$ 1.077 [11].

Radiomics performs an automated high-pass extraction of large amounts of quantitative features from medical images [12]. Applying machine learning techniques, researchers can use these radiomics features to build predictive models and analyze massive image features. In the past few years, many radiomics models have been proposed for classification [13], staging [14] and survival prediction of tumors [15]. More recently, some studies have employed radiomics techniques to solve the meningioma consistency prediction problem with good results [16–18]. Most radiomics features are manually defined [19], including intensity, shape, texture and wavelets. The number of radiomics features can reach tens of thousands with good interpretability. However, these features are usually shallow and low-order image features, which limit the effectiveness of radiomics [20].

In recent years, deep learning techniques have been widely used in the field of medical image analysis. Unlike radiomics techniques, CNN models used in deep learning methods extract convolutional feature maps [21]. Convolutional kernels of CNN models can capture the deep features of tumors by acting on different regions.

Deep learning features have proven to be an effective complement to radiomics features, yielding good results on a variety of clinical issues for diagnosing soft-tissue tumors [22], malignant nodules [23] and glioblastoma multiforme [20]. However, the prediction of meningioma consistency based on the fusion of radiomics features and deep learning features has not been reported.

Deep networks may not necessarily be omnipotent solutions. They usually rely on massive data for training network parameters. The insufficient amount of data may limit the deep network performance in the medical image analysis domain. We designed a shallow CNN architecture and trained it using a collection of T2-Flair images. Features were extracted from the last fully connected hidden layer of the proposed CNN architecture and then fused with the radiomics features to construct the classifiers.

This study seeks to integrate deep learning features with radiomics features to facilitate the computer-assisted preoperative evaluation of meningioma consistency. In our method, deep learning features were extracted using a self-defined CNN model. To validate the effectiveness of our feature fusion strategy, we compared the approach with several different feature combination methods.

2. Material and methods

2.1. Study participants

Our institutional review board approved the research protocol in this retrospectively study and waived formal informed consent. All the cases underwent data masking and were numbered before the experiment. Fig. 2 depicts the overall experimental design of this study. Between December 2016 and December 2018, 295 patients suspected of having meningiomas with clinical manifestations underwent MRI before surgery at our institution. After surgery, the same neurosurgeon with the experience of 15 years promptly formed a surgical record. The inclusion criteria were as follows: (1) pathologically confirmed meningioma; (2) complete T2-Flair images and surgical records (must include a clear description of the consistency); (3) no history of meningioma treatment before surgery. The exclusion criteria were as follows: (1) poor image quality, incomplete surgical records; (2) history of meningioma-related treatment; (3) other intracranial tumors or other diseases. 61 cases without complete surgical records, 22 cases receiving meningioma-related treatments, and 10 with poor-quality images were excluded. In total, 202 patients were screened for this study.

All the cases were split randomly into a training cohort (n = 162) and a testing cohort (n = 40) on the patient level. The training set was used for feature selection, hyperparameter optimization, model training, model validation. The testing set was only used to evaluate the models' performance.

2.2. Meningioma consistency assessment

After surgery, the same neurosurgeon assessed the meningioma consistency according to Zada's consistency grading system, according to the following grading criteria: (1) extremely soft tumor, internal debulking by suction only; (2) soft tumor, internal debulking mostly by suction, and remaining fibrous strands resected with easily folded capsule; (3) average consistency, the tumor cannot be freely suctioned and requires mechanical debulking, and the capsule then fold with relative ease; (4) firm tumor, high degree of mechanical debulking required, and capsule remains difficult to fold; (5) extremely firm,

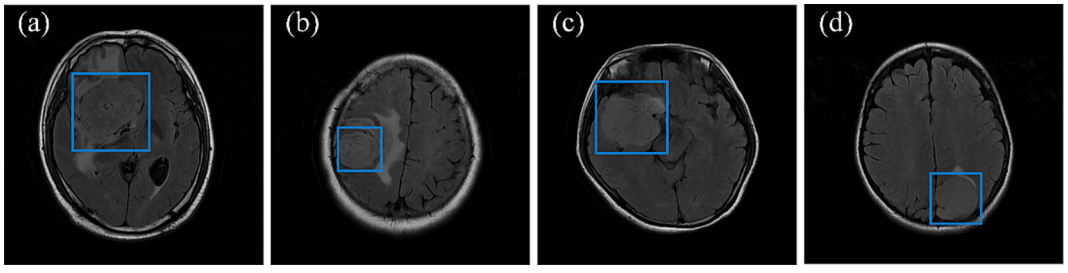

Fig. 1. Example images containing meningiomas of different consistency in T2-Flair images: (a) and (b) show slices containing hard meningiomas, (c) and (d) show slices containing soft meningiomas.

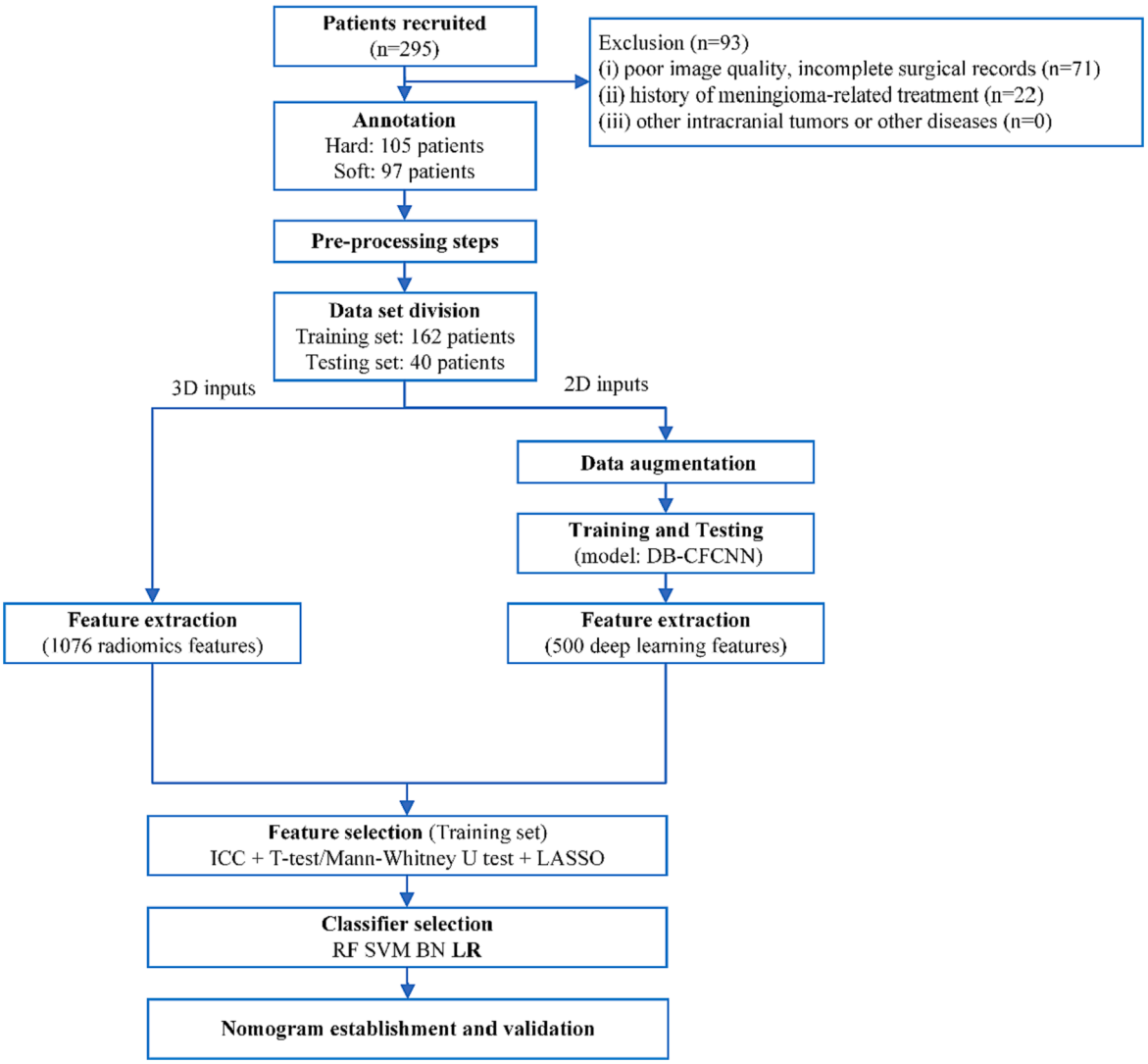

Fig. 2. The flowchart of our study. DB-CFCNN: the dual-branch architecture with central focused pooling layer; RF: random forest; SVM: support vector machine; BN: Bayesian network; LR: logistic regression.

calcified tumor, approaches the density of bone, and capsule does not fold [24]. We reviewed surgical records and classified these cases as soft (1–2 grade, aspiration removal) and hard (3–5 grade, non-suckable, requiring scissors or ultrasonic tumor devices).

2.3. Data acquisition and tumor annotation

All MRI images were scanned using two 3.0-Telsa MRI scanners with 32-channel head coils: (1) DISCOVERY MR750W; G.E., Milwaukee, MI, USA, and (2) MAGNETOM Verio; Siemens Healthcare, Germany. The T2-Flair sequence was acquired with the following parameters: repetition time (TR) = 8000–9000 ms; echo time (TE) = 94–150 ms; matrix = 358 × 512; excitation = 2; field of view (FOV) = 240 × 240 mm; bandwidth = 122 Hz/pixel; slice thickness = 3 mm; slice number = 16.

Regions of interest (ROIs) were manually outlined slice by slice on T2-Flair sequences by two radiologists with six-year experience using ITK-SNAP software (version 3.8.0, https://www.itksnap.org) respectively and the intersection was taken as the final result. The results were exported as a zip archive in NII format. The physicians were unaware of the clinical findings.

2.4. Radiological evaluation

Radiological semantic features (tumor visual characteristics) were assessed after the negotiation by 2 radiologists with six-year neuroradiological experience who were blinded to clinical information and histopathological diagnosis. These semantic features included maximum diameter, irregular shape, cystic composition, necrosis or hemorrhage, bone invasion and *peri*-tumoural edema.

2.5. Radiomics features extraction

All radiomics features were obtained by the Pyradiomics 3.0.1 [25], which is an open-source python package for extracting radiomics features from medical images. As our data were acquired using two different MRI scanners, some pre-processing steps were conducted on these images before feature extraction. Image intensities were first normalized to 0–100. Since radiomics features extracted from images were voxel-based, the images were then resampled to the same resolution (3 × 3 × 3 mm) using cubic b-spline to ensure isotropy. In addition, a 5 bin width and 300 voxel array shift were set to these images.

Six types of features were extracted from both original and derived images (wavelet and Laplacian of Gaussian (LoG)), including 3D-based shape features, first-order statistics, grayscale co-occurrence matrix

(glcm), grayscale run length matrix (glrlm), grayscale size zone matrix (glszm) and grayscale dependence matrix (gldm). The wavelet filters consisted of various combinations of high-pass and low-pass filters in three dimensions, namely LLH, LHL, LHH, HLL, HLH, HHL, HHH. Furthermore, the sigma values for LoG were set at 3.0 and 5.0. A total of 1076 features were involved in our study. Details about the extraction and description of involved radiomics features can be found in the Pyradiomics documentation (https://pyradiomics.readthedocs.io).

2.6. Deep learning feature extraction

T2-Flair images used in our study were divided into a training cohort (n = 162) and a testing cohort (n = 40) on the patient level. Such a training set was rather small for training a CNN. For each case in the training cohort, we selected all slices labeled as tumors and performed a data augmentation operation. After data augmentation, the number of T2-Flair images used for training and validating was 3032. For each case in the testing cohort, we selected all slices labeled as tumors for inclusion. We adopted stratified sampling, where 80 % of the training cohort was selected for training and the remaining 20 % was used for validation. It was guaranteed that slices of each case would not appear in both the training and validation sets. A rectangular box covering the tumor area was taken per slice. The length and width of the rectangular box were equal to ensure that the tumor shape did not change when the image was resized. Meningioma images used in this study were grayscale images, and we first normalized the grayscale values of these images to 0–255. Since the grayscale values of the tumor and surrounding tissues did not differ much, we used histogram equalization to enhance the contrast of the tumor and better show the inhomogeneous texture inside the tumor.

We designed the CNN architecture employed in our study using PyTorch 1.2. Fig. 3 shows the flowchart of our CNN Architecture. It has a dual-branch architecture, where adaptive max pooling is used in one branch to process the input. The texture information of the tumor is retained, which is closely related to the consistency of the tumor. The other branch consists of a normal convolutional layer (using convolution + batch normalization (BN) + LeakyReLU structure, called conv block) and a central focused pooling layer [26]. This CNN architecture is named DB-CFCNN. Its conv block uses the small receptive field of the 3 × 3 convolution kernel to maintain the sensitivity of local features. The center-focused pooling layer is helpful to largely eliminate irrelevant edge features of the patch while preserving the central features at the same time. The splicing operation occurs after obtaining the same size output (20 × 20) from both branches. The information from both branches is integrated using the conv block and the center-focused pooling.

Two fully connected layers are used in the classification layer, with output dimensions of 500 and 2. Outputs of the DB-CFCNN model are meningioma consistency predictions (hard or soft).

2.7. Parameter settings and implementation details

DB-CFCNN training details are elaborated in this section. For the input image size, we employed bicubic interpolation to resize the tumor-containing slices, which varied in size from 100 to 21,316 pixels, to 80 × 80. Multiple experiments were conducted using different input image sizes, and it was determined that the most optimal results were achieved when utilizing the dimension of 80 × 80. The total number of epochs for training was set to 300. The initial value of the learning rate was set to 0.01, which was adaptively adjusted using the ReduceLROnPlateau method. Stochastic gradient descent (SGD) was used as the optimizer. We experimented with various batch sizes (8/16/32). A batch size of 32 gave the best result. Our experiment was performed on a PC workstation with a Linux operating system. The deep learning networks were implemented using PyTorch 1.2.

2.8. Feature selection and establishment of classification model

After feature extraction, 1076-dimensional radiomics features were fused with 500-dimensional deep learning features. We standardized 1576-dimensional features for each patient on the training set, and then applied standardized parameters to the testing set. To remove redundant features, we performed the following feature selection steps on the training set.

First, as feature extraction relies on tumor regions manually outlined by physicians, the dataset underwent retesting analysis. Each tumor was independently segmented by two radiologists with six-year experience, so the consistency of the extracted features was assessed by the intra-class correlation coefficient (ICC) [27]. Features with ICC ≥ 0.85 were robust.

Second, selected features were interclass compared using *T*-test (normally distributed) or Mann-Whitney *U* test (skewed distribution). Features with p-values less than 0.05 were selected for further analysis.

Finally, LASSO was used to search for the optimal combination of features. Tenfold cross-validation was performed to search the model's optimal hyperparameter alpha (α).

Features with non-zero coefficients were screened to construct our machine learning classifiers RF, SVM, BN and LR. The training methods for the classifiers were grid search and ten-fold cross-validation to make the model more credible. Precision, recall, F1-score and AUC were used as evaluation metrics in our assessment for internal testing. For assessing the necessity of different types of features on the model, we also built

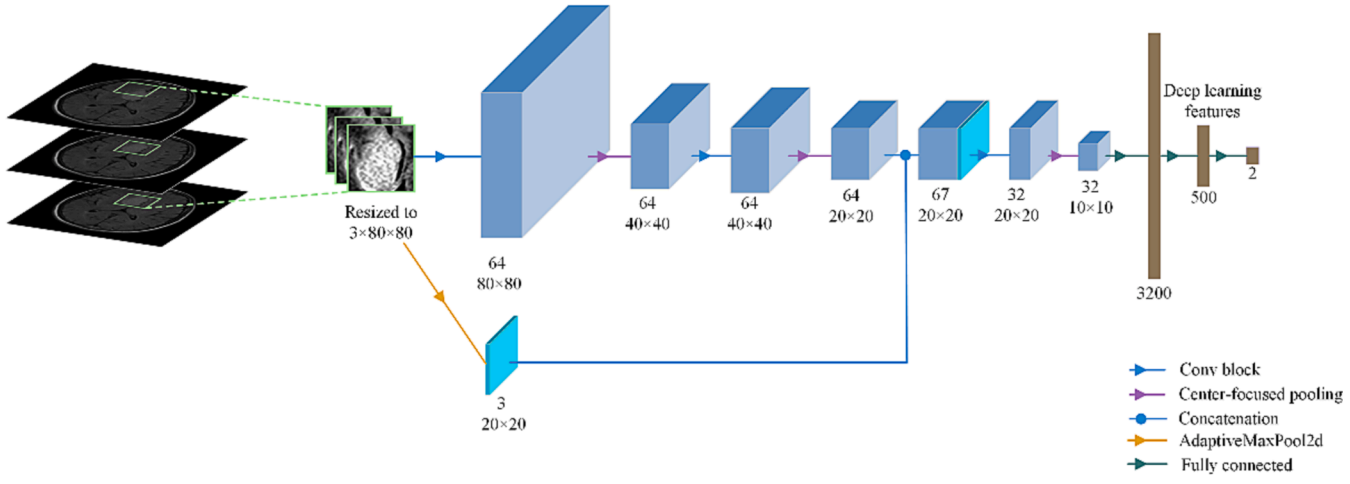

Fig. 3. The architecture of DB-CFCNN. After training the model, 500 features from the fully connected layer were extracted as deep learning features.

clinical features-only model, radiomics features-only model, deep learning features-only model and a model that integrated all three types of features.

2.9. Signature calculation and nomogram construction

After feature selection, features with non-zero coefficients were constructed as the signature. The signature for each patient can be calculated by a linear combination of selected features multiplied by their respective non-zero coefficients.

A nomogram was built utilizing the signature. The performance of the nomogram was validated using the C-index, calibration curve and decision curve analysis (DCA). The C-index served to measure the likelihood of concordance between predicted and actual outcomes. The calibration curve was used to estimate the goodness of fit. The DCA evaluated the clinical utility of the nomogram by quantifying the net benefit at different threshold probabilities.

2.10. Statistical analysis

Statistical analysis was performed using statistical software, SPSS version 26.0, IBM, China. When analyzing the baseline clinical information and radiological semantic features of the patients, we performed unpaired *t*-test for continuous variables including age, maximum diameter and Ki-67 value to identify their correlations with meningioma consistency. A p-values < 0.05 was considered statistically significant. We applied the Chi-Square test or Fisher's exact probability test to search for correlations of discrete variables and categorical variables, including sex, tumor location, WHO grade, irregular shape, cystic composition, necrosis or hemorrhage, bone invasion, *peri*-tumoural edema and pathological classification. In the feature selection phase, radiomics features and deep learning features were compared using *T*-test (normally distributed) or Mann-Whitney *U* test (skewed distribution).

3. Results

3.1. Patient characteristics

The mean (±SD) age of the 202 patients was 55.03 ± 11.41 years and females accounted for 74.8 %. According to Zada's consistency grading system, 97 cases (26 men, 71 women; mean age, 53.53 years ± 12.36 [SD]) were labeled as "Soft" and 105 cases (25 men, 80 women; mean age, 55.90 years ± 10.80) as "Hard". After data division, the training set consisted of 162 cases, comprising 80 soft cases and 82 hard cases. The testing set included 40 cases, consisting of 17 soft cases and 23 hard cases. The baseline clinical information and radiological semantic features of patients were summarized in Table 1. Six clinical factors were outlined including age, sex, tumor location, the WHO grade, Ki-67 value and pathological classification. Meningioma consistency was associated with the WHO grade ($p = 0.003$) and Ki-67 value ($p = 0.007$). For radiological semantic features, there was a significant difference in the maximum diameter of meningioma between soft tumors (3.83 ± 1.52) and hard tumors (3.41 ± 1.40) ($p = 0.04$). The distribution of these features in both the training and testing sets can be found in Supplemental Table 1.

3.2. Feature selection and signature construction

The largest tumor slice of T2-Flair images was selected and fed into the well-trained DB-CFCNN to extract a total of 500 deep learning features. These deep learning features were fused with 1076 radiomics features. After a three-step feature selection, 26 features with non-zero coefficients were filtered using LASSO to construct the signature. The selected features and corresponding coefficients are displayed in Fig. 4. The coefficients of features are provided in Supplemental Table 2.

Table 1
Clinical and radiological semantic features of patients.

	Soft	Hard	p-value
Patient count	97	105	
Age (mean ± SD)	53.53 ± 12.36	55.90 ± 10.83	0.148
Sex			
Male	26	25	0.625
Female	71	80	
Location			
Cerebral convexity	39	43	0.355
Skull base	11	11	
Sphenoid	4	14	
Cerebellopontine angle	5	9	
Parasagittal	6	4	
Petroclival	5	3	
Tentorium	5	3	
Sellar	4	3	
Falx	3	3	
Foramen magnum	2	3	
Parasellar	3	0	
Cerebellar convexity	5	1	
Lateral ventricle	2	4	
Anterior clinoid process	1	2	
Tuberculum sellae	2	1	
Orbital	0	1	
WHO grade			
1	84	103	0.003
2	12	2	
3	1	0	
Ki-67 value (%)	3.29 ± 2.46	2.50 ± 1.48	0.007
Pathological classification			
Atypical	9	2	0.126
Secretory	1	1	
Transitional	2	4	
Chordoid	3	0	
Anaplastic	1	0	
Psammomatous	0	1	
Epithelial	44	52	
Microcystic	0	1	
Fibrous	33	37	
Angiomatous	4	7	
Maximum diameter (cm)	3.83 ± 1.52	3.41 ± 1.40	0.04
Irregular shape			
Regular	53	70	0.08
Irregular	44	35	
Cystic composition			
Yes	14	15	0.976
No	83	90	
Necrosis or hemorrhage			
Yes	4	8	0.294
No	93	97	
Bone invasion			
Yes	22	32	0.211
No	75	73	
Peri-tumoural edema			
Yes	45	53	0.562
No	52	52	

3.3. Selection of machine learning classifier

We trained four machine learning classifiers using the 26 features in Fig. 4. Table 2 evaluates the diagnostic value of various machine

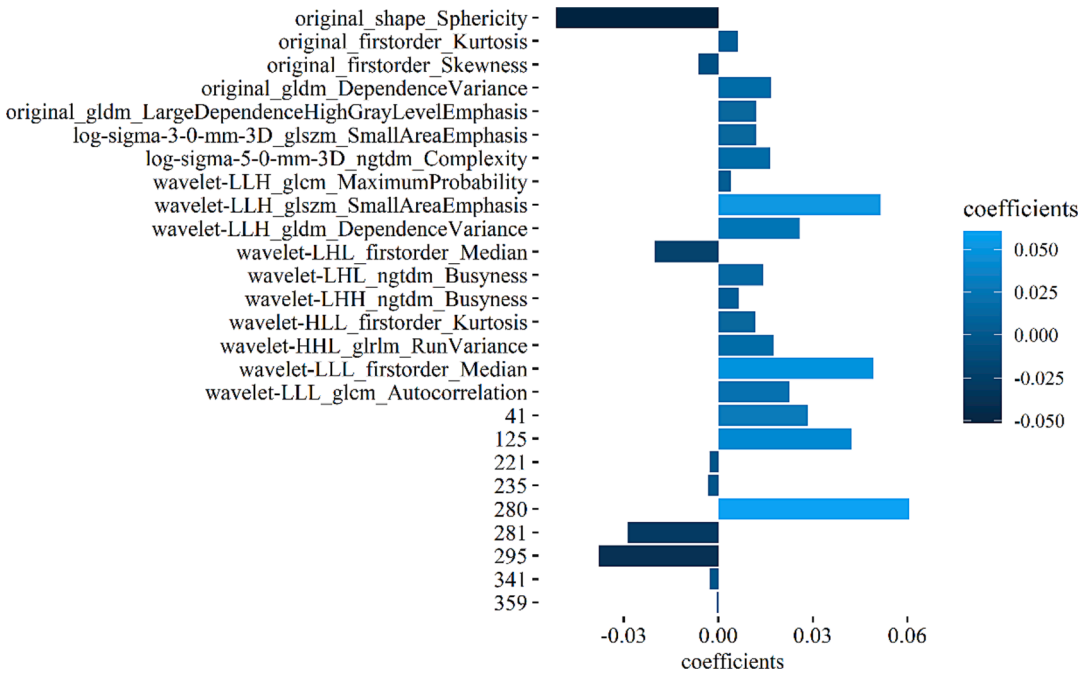

Fig. 4. The selected features and corresponding coefficients. Rows 1 to 17 are radiomics features and 18–26 are deep learning features (Deep learning features are indexed from 1 to 500).

Table 2
Performance comparison of different classifiers (testing set). The features used are 9 deep learning features extracted from DB-CFCNN and 17 radiomics features.

Classifiers	Precision	Recall	F1 score	AUC (95 % CI)
RF	0.817	0.780	0.787	0.798 (0.732–0.866)
SVM	0.776	0.756	0.759	0.766 (0.700–0.830)
BN	0.843	0.756	0.749	0.783 (0.723–0.854)
LR	**0.855**	**0.854**	**0.852**	**0.943 (0.873–1.000)**

Notes: The value of AUC is reported by 95 % confidence interval (CI) in the bracket.

Table 3
Performance comparison of different feature fusion approaches (testing set). All these approaches use LR classifier.

Feature types			Performance			
CF	RADF	DLF	Precision	Recall	F1-score	AUC
√			0.600	0.529	0.563	0.701 (0.630–0.752)
	√		0.729	0.683	0.726	0.768 (0.718–0.840)
		√	0.705	0.705	0.705	0.766 (0.716–0.840)
	√	√	0.855	0.854	0.852	**0.943 (0.873–1.000)**
√	√	√	0.743	0.732	0.734	0.737 (0.690–0.822)

Notes: CF, clinical feature; RADF, radiomics features; DLF, DB-CFCNN's deep learning features.

learning classifiers (RF, SVM, BN and LR) for tumor consistency. None of them showed overfitting during training. For logistic regression on the testing set, the AUC was 0.943 (95 % CI, 0.873–1.000), the precision was 0.855 and the recall was 0.854. We chose LR classifier for the subsequent analysis.

3.4. Comparison of models using different types of features

To validate the best combination of features, we also built clinical factor-only model (using the maximum diameter), radiomics features-only model, deep learning features-only model and a model that integrated all three types of features.

Our testing set was used to evaluate the models, and the results are outlined in Table 3. In all approaches, 9 DB-CFCNN's deep learning features and 17 radiomics features demonstrated the best performance, which provided the best precision of 85.5 % and the best AUC of 0.943 (95 % CI, 0.873–1.000). The comprehensive classification model incorporating maximum diameter, 9 deep learning features, and 17 radiomics features offered a precision of 74.3 % (AUC 0.737).

3.5. Nomogram establishment and validation

Comparing the feature fusion approaches in Table 3, we chose the model with the highest AUC to construct the nomogram. The nomogram can be used to assist the surgeon to predict the meningioma consistency.

We calculated the radiomics scores of Fig. 1(a) and (c). As shown in Fig. 5(a), Fig. 1(a) had a radiomics score of −0.203, which corresponded to a probability of less than 20 % of soft consistency. Thus, the meningioma consistency was predicted to be hard. Fig. 1(c) had a radiomics score of 0.329, the total points were about 71, corresponding to a probability of more than 90 % of a soft meningioma. Fig. 5(b) and Fig. 5 (c) showed the calibration curves on training and testing sets. The calibration curve showed good agreement between the proposed nomogram and meningioma consistency prediction. The C-index was 0.822 (95 % CI, 0.758–0.885) in the training cohort and 0.943 (95 % CI, 0.873–1.000) in the testing cohort, indicating favorable discrimination. Furthermore, the comparison of models employing different types of features was conducted using DCA, as depicted in Fig. 6. The nomogram had a high overall net benefit over most reasonable threshold probability ranges. This can greatly facilitate the application of the model in clinical settings.

4. Discussions

Meningioma is one of the most common intracranial brain tumors [1]. Accurate preoperative assessment of meningioma consistency is important for planning surgical access, resection modality and prognostic evaluation. Soft tumors can be removed by cutting and aspiration. It is more difficult to remove hard tumors, especially skull base

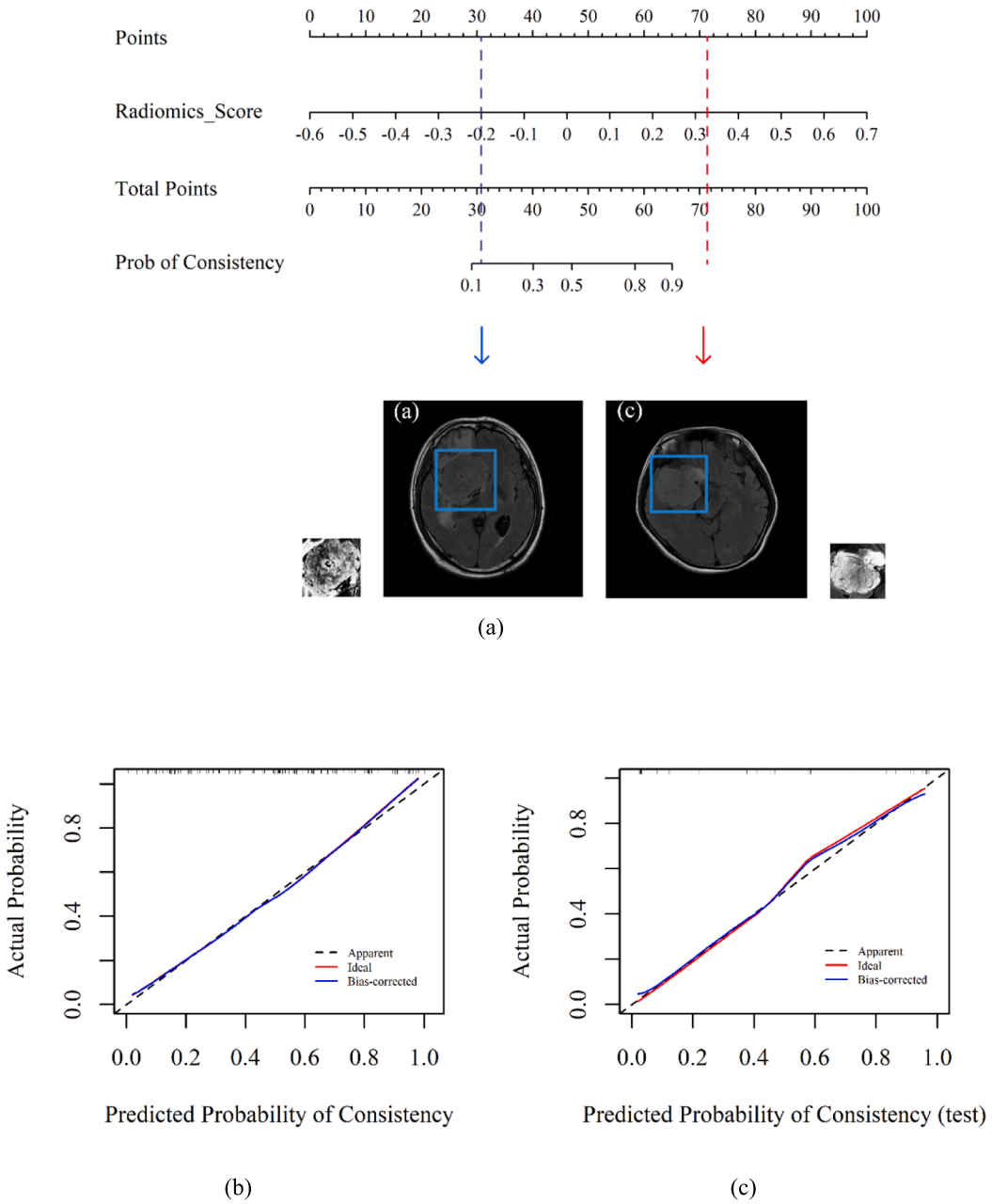

Fig. 5. (a) The nomogram for predicting the meningioma consistency. The blue line and the red line provide illustrative examples. The blue line represents a hard meningioma (Fig. 1(a)), and the red line represents a soft meningioma (Fig. 1(c)). The two small images next to the two MRI slices show the tumor area after contrast enhancement. (b)(c) Calibration curves of the nomogram for the training set and testing set. In a well-calibrated model, its predictions should fall on a 45-degree diagonal line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

meningiomas, which require additional instruments. The main method for predicting meningioma consistency is to use magnetic resonance sequences. Some papers have described the correlation between the meningioma consistency and the signal intensity of MRI [1,4,5,7–9]. A few studies have used radiomics to predict meningioma consistency [17]. To the best of our knowledge, this study presents the first attempt to introduce deep learning features to assess meningioma consistency.

In our study, the patient's baseline clinical information and radiological semantic features were used to analyze the potential relationship with consistency. There were significant correlations between meningioma consistency and the WHO grade ($p = 0.003$), Ki-67 value ($p = 0.007$) and maximum diameter ($p = 0.04$). In our dataset, high-grade meningiomas (2, 3) made up a larger proportion of soft tumors (13.4 %) than hard meningiomas (1.9 %). The mean (±SD) Ki-67 value for soft meningioma was 3.29 ± 2.46 and for hard meningioma was 2.50 ± 1.48. And there was a significant difference in the maximum diameter of meningioma between soft tumors (3.83 ± 1.52) and hard tumors (3.41 ± 1.40). Epithelial meningiomas with a cellular component are softer, while fibrous and calcified meningiomas are harder, such as fibrous meningiomas and psammomatous types. Our study showed that soft meningiomas had a higher WHO grade and a larger Ki-67 value. High Ki-67 means that the tumor cells are more proliferative, faster growing, more aggressive and more likely to develop distant metastases [28]. Therefore, we speculate that soft meningiomas have a faster growth rate and larger maximum diameter. In addition, the consistency of meningioma may be potentially related to the likelihood of necrosis, which will be investigated in follow-up work. The WHO grade and Ki-67 value are pathological diagnostic findings, so we did not include these two

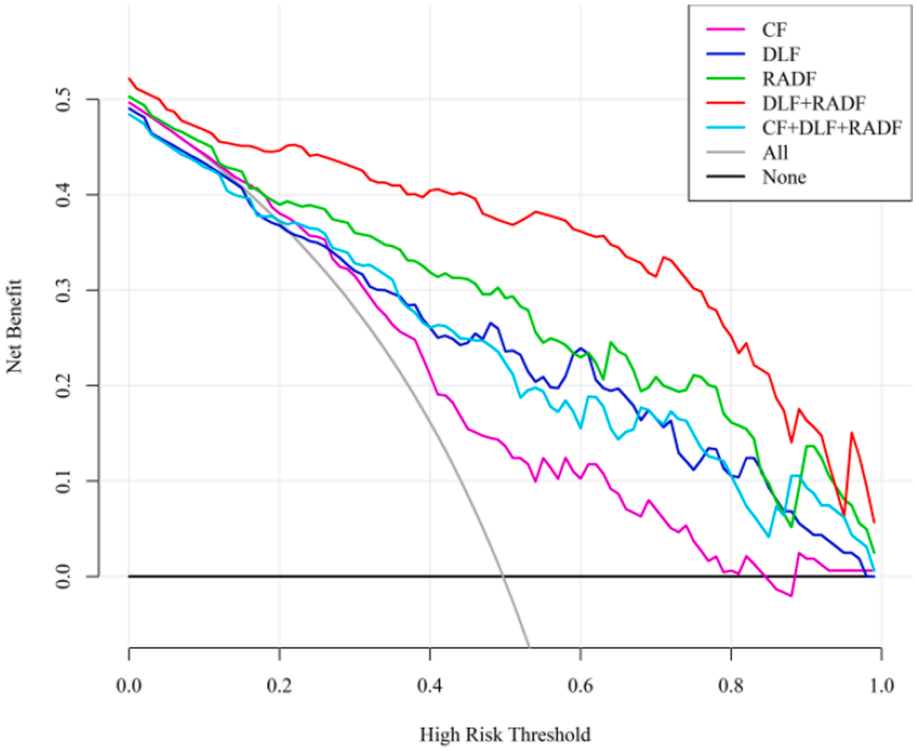

Fig. 6. DCA of models constructed with different feature combinations. The horizontal coordinate is the threshold probability, and the vertical coordinate is the net benefit. The All curve is an intervention for all. The None curve is an intervention for none.

factors in the classification model construction. A precision of 60.0 % (AUC 0.701) was obtained using the maximum diameter only to construct the classification model. After adding this feature to deep learning features and radiomics features, there was no better performance in AUC, resulting in 0.737 (95 % CI, 0.690–0.822) in the testing set.

DB-CFCNN was designed as a feature extractor to extract deep learning features of tumors. As shown in Table 3, the combination of 9 DB-CFCNN's deep learning features and 17 radiomics features performed best than other feature combination approaches, (precision 85.5 %, AUC 0.943), indicating the potential advantages of the combination of deep learning features and radiomics features. As shown in Fig. 5, the nomogram established by the 26 features showed satisfactory predictive performance across the meningioma consistency in both the training set (C-index: 0.822) and testing set (C-index: 0.943) with good calibration. Deep learning features reflect higher-order imaging modalities and capture more information about the tumor. They are usefully complementary to radiomics features in the consistency prediction of meningiomas. The scheme combining deep learning features and radiomics features to predict meningioma consistency has significant advantages.

Among the 26 features selected, the radiomics feature "wavelet-LLH_glszm_SmallAreaEmphasis" contributed more to the signature construction. Small Area Emphasis (SAE) is a measure of the distribution of small-size zones, with a greater value indicative of smaller-size zones and more fine textures. The deep learning feature "280″ also provided a large coefficient. But it is hardly explainable and corresponds to differences between various disease phenotypes that are not visible to the human eye.

Our approach makes use of features extracted from T2-Flair sequences. T2-Flair suppresses the signal of free water, and its signal intensity can reflect the amount of bound water. It has been suggested that soft tumors contain more bound water compared to normal brain tissue or hard tumors [1]. In contrast, harder tumors have more collagen and less bound water, so harder tumor regions show a lower signal on T2-Flair. Therefore, T2-Flair sequences can better highlight the features of tumors with different consistency.

Despite our promising experimental results, this study still has several limitations. First, our method can be applied to T2-Flair sequences without requiring additional specific imaging parameters. It may be more practical to implement. However, when dealing with images of other MRI modalities, images with different imaging parameters may not be accurately identified by our DB-CFCNN model pre-trained using T2-Flair sequences. This problem could be solved by fine-tuning the network using new datasets in the future. Second, the clinical patient dataset used in this study was obtained from the same institution and was small. Multicenter studies are needed for further evaluation and more patients and hospitals shall be involved. The model's generalization ability will also be comprehensively evaluated using a large dataset in our future work. Third, since no 3D input was used when training the DB-CFCNN model, it cannot perceive 3D features. Fourth, the interpretability of deep learning features was not yet studied. The potential correlation between deep learning features and meningioma consistency needs to be explored more deeply. Fifth, this study is based on single-modality MRI. In the future investigations, we plan to expand our experiments by incorporating multimodal MRI along with MRE. Finally, our study requires tumor images manually annotated at the pixel level by physicians. The manual annotation requires a significant upfront human investment to produce the dataset for the experiment. In such a way, large datasets used for training and testing are very difficult to build because of the non-uniform standards for tumor image data annotation, the small amount of annotated data, and the excessive difficulty of manual annotation. End-to-end weakly supervised learning is also a direction for future research. An additional consideration lies in the heterogeneity of meningiomas. While surgical records lack detailed descriptions regarding the consistency of different regions within the tumor, acquiring a larger collection of 3D images and more comprehensive descriptions of tumor consistency is imperative for future investigations.

The proposed method for assessing meningioma consistency based on the fusion of deep learning features and radiomics features is

potentially clinically valuable. We will further evaluate and validate our approach and integrate it into clinical applications to provide preoperative guidance to physicians.

Funding

This project was supported by Shanghai Sailing Program (No. 21YF1404800), Shanghai Municipal Science and Technology Major Project (No. 2018SHZDZX01) and ZJ Lab, Shanghai Center for Brain-Inspired Technology, Medical Engineering Fund of Fudan University (No. yg2021-029).

CRediT authorship contribution statement

Jiatian Zhang: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis. **Yajing Zhao:** Writing – original draft, Resources, Methodology, Investigation, Data curation. **Yiping Lu:** Writing – review & editing, Resources, Investigation, Data curation. **Peng Li:** Investigation, Formal analysis. **Shijie Dang:** Formal analysis. **Xuanxuan Li:** Resources, Funding acquisition. **Bo Yin:** Writing – reviewing & editing, Supervision, Resources, Project administration, Funding acquisition. **Lingxiao Zhao:** Writing – review & editing, Supervision, Methodology.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to extend their gratitude and acknowledgements to all study participants.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejrad.2023.111250.

References

[1] J.M. Hoover, J.M. Morris, F.B. Meyer, Use of preoperative magnetic resonance imaging t1 and t2 sequences to determine intraoperative meningioma consistency, Surg. Neurol. Int. 2 (2011), https://doi.org/10.4103/2152-7806.85983.
[2] L.B. Nabors, J. Portnow, M. Ahluwalia, J. Baehring, H. Brem, S. Brem, N. Butowski, J.L. Campian, S.W. Clark, A.J. Fabiano, et al., Central nervous system cancers, version 3.20 20, nccn clinical practice guidelines in oncology, J. Natl. Compr. Canc. Netw. 18 (11) (2020) 1537–1570, https://doi.org/10.6004/jnccn.2020.0052.
[3] B. Kendall, P. Pullicino, Comparison of consistency of meningiomas and ct appearances, Neuroradiology 18 (4) (1979) 173–176, https://doi.org/10.1007/bf00345721.
[4] N. Yamaguchi, T. Kawase, M. Sagoh, T. Ohira, H. Shiga, S. Toya, Prediction of consistency of meningiomas with preoperative magnetic resonance imaging, Surg. Neurol. 48 (6) (1997) 579–583, https://doi.org/10.1016/s0090-3019(96)00439-9.
[5] A. Yao, M. Pain, P. Balchandani, R.K. Shrivastava, Can mri predict meningioma consistency?: a correlation with tumor pathology and systematic review, Neurosurg. Rev. 41 (3) (2018) 745–753, https://doi.org/10.1007/s10143-016-0801-0.
[6] K. Erkmen, S. Pravdenkova, O. Al-Mefty, Surgical management of petroclival meningiomas: factors determining the choice of approach, Neurosurg. Focus 19 (2) (2005) 1–12, https://doi.org/10.3171/foc.2005.19.2.8.
[7] K. Watanabe, S. Kakeda, J. Yamamoto, S. Ide, N. Ohnari, S. Nishizawa, Y. Korogi, Prediction of hard meningiomas: quantitative evaluation based on the magnetic resonance signal intensity, Acta Radiol. 57 (3) (2016) 333–340, https://doi.org/10.1177/0284185115578323.
[8] L.A. Ortega-Porcayo, P. Ballesteros-Zebadúa, O.R. Marrufo-Meĺendez, J.J. Rańırez-Andrade, J. Barges-Coll, A. Tecante, M. RańırezGilly, J.L. Gómez-Amador, Prediction of mechanical properties and subjective consistency of meningiomas using t1–t2 assessment versus fractional anisotropy, World Neurosurg. 84 (6) (2015) 1691–1698, https://doi.org/10.1016/j.wneu.2015.07.018.
[9] M. Alyamany, M. Alshardan, A. Jamea, N. ElBakry, L. Soualmi, Y. Orz, Meningioma consistency: Correlation between magnetic resonance imaging characteristics, operative findings, and histopathological features, Asian J. Neurosurg. 13 (02) (2018) 324–328, https://doi.org/10.4103/1793-5482.228515.
[10] A. Thotakura, M. Patibandla, M. Panigrahi, A. Mahadevan, Is it really possible to predict the consistency of a pituitary adenoma preoperatively? Neurochirurgie 63 (6) (2017) 453–457, https://doi.org/10.1016/j.neuchi.2017.06.003.
[11] L. Yiping, X. Ji, G. Daoying, Y. Bo, Prediction of the consistency of pituitary adenoma: a comparative study on diffusion-weighted imaging and pathological results, J. Neuroradiol. 43 (3) (2016) 186–194, https://doi.org/10.1016/j.neurad.2015.09.003.
[12] P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R.G. Van Stiphout, P. Granton, C.M. Zegers, R. Gillies, R. Boellard, A. Dekker, et al., Radiomics: extracting more information from medical images using advanced feature analysis, Eur. J. Cancer 48 (4) (2012) 441–446, https://doi.org/10.1016/j.ejca.2011.11.036.
[13] C.Y. Wang, D.T. Ginat, Preliminary computed tomography radiomics model for predicting pretreatment cd8+ t-cell infiltration status for primary head and neck squamous cell carcinoma, J. Comput. Assist. Tomogr. 45 (4) (2021) 629–636, https://doi.org/10.1097/RCT.0000000000001149.
[14] G. Hamerla, H.-J. Meyer, S. Schob, D.T. Ginat, A. Altman, T. Lim, G.A. Gihr, D. Horvath-Rizea, K.-T. Hoffmann, A. Surov, Comparison of machine learning classifiers for differentiation of grade 1 from higher gradings in meningioma: A multicenter radiomics study, Magn. Reson. Imaging 63 (2019) 244–249, https://doi.org/10.1016/j.mri.2019.08.011.
[15] F. Isensee, P. Kickingereder, W. Wick, M. Bendszus, K.H. Maier-Hein, Brain tumor segmentation and radiomics survival prediction: Contribution to the brats 2017 challenge, in: International MICCAI Brainlesion Workshop, Springer, 2017, pp. 287–297. Doi: 10.1007/978-3-319-75238-925.
[16] A. AlKubeyyer, M.M. Ben Ismail, O. Bchir, M. Alkubeyyer, Automatic detection of the meningioma tumor firmness in MRI images, J. Xray Sci. Technol. 28 (4) (2020) 659–682, https://doi.org/10.3233/XST-200644.
[17] Y. Zhai, D. Song, F. Yang, Y. Wang, X. Jia, S. Wei, W. Mao, Y. Xue, X. Wei, Preoperative prediction of meningioma consistency via machine learning-based radiomics, Front. Oncol. (2021) 1519, https://doi.org/10.3389/fonc.2021.657288.
[18] S. Cepeda, I. Arrese, S. Garcıa-Garcıa, M. Velasco-Casares, T. Escudero-Caro, T. Zamora, R. Sarabia, Meningioma consistency can be defined by combining the radiomic features of magnetic resonance imaging and ultrasound elastography. a pilot study using machine learning classifiers, World Neurosurg. 146 (2021) e1147–e1159, https://doi.org/10.1016/j.wneu.2020.11.113.
[19] H.J. Aerts, E.R. Velazquez, R.T. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, et al., Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach, Nat. Commun. 5 (1) (2014) 1–9, https://doi.org/10.1038/ncomms5006.
[20] J. Lao, Y. Chen, Z.-C. Li, Q. Li, J. Zhang, J. Liu, G. Zhai, A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme, Sci. Rep. 7 (1) (2017) 1–8, https://doi.org/10.1038/s41598-017-10649-8.
[21] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324, https://doi.org/10.1109/5.726791.
[22] S. Shang, J. Sun, Z. Yue, Y. Wang, X. Wang, Y. Luo, D. Zhao, T. Yu, X. Jiang, Multi-parametric mri based radiomics with tumor subregion partitioning for differentiating benign and malignant soft-tissue tumors, Biomed. Signal Process. Control 67 (2021), 102522, https://doi.org/10.1016/j.bspc.2021.102522.
[23] R. Paul, S. Hawkins, M.B. Schabath, R.J. Gillies, L.O. Hall, D.B. Goldgof, Predicting malignant nodules by fusing deep features with classical radiomics features, J. Med. Imaging 5 (1) (2018), 011021, https://doi.org/10.1117/1.JMI.5.1.011021.
[24] G. Zada, P. Yashar, A. Robison, J. Winer, A. Khalessi, W.J. Mack, S.L. Giannotta, A proposed grading system for standardizing tumor consistency of intracranial meningiomas, Neurosurg. Focus 35 (6) (2013) E1, https://doi.org/10.3171/2013.8.FOCUS13274.
[25] J.J. Van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R. G. Beets-Tan, J.-C. Fillion-Robin, S. Pieper, H.J. Aerts, Computational radiomics system to decode the radiographic phenotype, Cancer Res. 77 (21) (2017) e104–e107, https://doi.org/10.1158/0008-5472.CAN-17-0339.
[26] S. Wang, M. Zhou, Z. Liu, Z. Liu, D. Gu, Y. Zang, D. Dong, O. Gevaert, J. Tian, Central focused convolutional neural networks: Developing a data-driven model for lung nodule segmentation, Med. Image Anal. 40 (2017) 172–183, https://doi.org/10.1016/j.media.2017.06.014.
[27] J.L. Fleiss, J. Cohen, The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability, Educ. Psychol. Meas. 33 (3) (1973) 613–619, https://doi.org/10.1177/001316447303300309.
[28] A. Durand, F. Labrousse, A. Jouvet, L. Bauchet, M. Kalamarid'es, P. Menei, R. Deruty, J.J. Moreau, M. F'evre-Montange, J. Guyotat, Who grade ii and iii meningiomas: a study of prognostic factors, J. Neuro-Oncol 95 (3) (2009) 367–375, https://doi.org/10.1007/s11060-009-9934-0.