Appendix 1: Image preprocessing

Before texture feature extraction, image preprocessing is necessary because of different scan parameters and irrelevant information in the images will influence the texture feature extraction process. The involved three preprocessing procedures are as follows:

- Wavelet band-pass filtering is used to prevent the noise in ROIs from interfering with the texture information. This operation is performed by utilizing various weights to band-pass sub-bands (*LHL*, *LHH*, *LLH*, *HLL*, *HHL*, *and HLH*) of tumor region to obtain different information in wavelet domain. Ratios of 1/2, 2/3, 1, 3/2, and 2 are implemented.
- Isotropic resampling is for maintaining rotation invariance ensuring the pixel and slice thickness of th1-e ROI are fixed and same. Different isotropic voxel sizes of 1 mm, 2 mm, 3 mm, 4 mm, 5 mm, and initial in-plane resolution (denoted as 'in-pR') are tested in this work. For example, if the desired parameter is set to 5 mm, the volume with the size 5.5 mm × 5.5 mm × 4.5 mm is isotropically resampled to 5 mm × 5 mm × 5 mm.
- Before computing the texture features, the intensity range is quantized to a specific number of gray levels *N*. Quantization of gray level aims to compress or extend the intensity ranges to the specified intensity ranges because of the fact that all higher-order features involve a distance parameter which normalizes images with different attenuation values. This process maps the voxel values to a finite set $r = \{r_k \in R : k = 1,2,...,N\}$ of reconstruction levels by defining a set $t = \{t_k \in R : k = 1,2,...,N+1\}$ of decision levels. Quantization algorithm and number of gray level are the two important parameters in this process. In this work, we test two quantization algorithms, Equal-probability and Lloyd-Max. Equal-probability quantization attempts to define decision thresholds in the volume, then the reconstructed level r_k has the same quantized number of voxels for all gray levels, nevertheless Lloyd-Max attempts to minimize the mean-squared quantization error of output. The number of gray levels of 8, 16, 32, and 64 are implemented in our experiments.

Appendix 2: Relief algorithm

After feature extraction, a linear normalization operator minimum-maximum method is firstly used to eliminate the magnitude of features and negative effects of large magnitude difference. The formula is as follows and *F* represents features:

$$F_{norm} = \frac{F - F_{\min}}{F_{\max} - F_{\min}} \quad [1]$$

High-dimension features always contain redundant features and result in overfitting for classification task. Consequently, feature selection is necessary for acquiring the most remarkable features. Relief (Relevant feature) is a filtering operator that designs a relevant statistic to measure the importance of each feature. This statistic is a vector, whose every component corresponds to an initial feature, and the importance of each feature subset depends on the sum of relevant statistics. Given a training case x_i, it firstly searches for the nearest neighbor $x_{(i,nb)}$ in the homogeneous samples, which is called "near-hit". Then searches for the nearest neighbor $x_{(i,nm)}$ in the heterogeneous samples, which is called "near-miss". Homogeneous means these samples are from the same category and heterogeneous is the opposite. The importance of the feature *j* is measured by the following relevant statistic:

$$\delta^j = \sum_i -diff(x_i^j, x_{i,nh}^j)^2 + diff(x_i^j, x_{i,nm}^j)^2 \quad [2]$$

where x_a^j denotes the feature *j* of in the case x_a, and $diff(x_a^j, x_b^j)$ denotes as the following:

$$diff(x_a^j, x_b^j) = \left| x_a^j - x_b^j \right| \quad [3]$$

According to formula [3], if the distance between x_i and its near-hit $x_{(i,nb)}$ is more closer than the distance between x_i and its near-miss $x_{(i,nm)}$, it indicates that the feature *j* is positive to classification. On the contrary, if the distance between x_i and its near-hit $x_{(i,nb)}$ is farther than the distance between x_i and its near-miss $x_{(i,nm)}$, it means that the feature *j* is negative to classification. Finally, average value of the estimated results is computed based on different samples and it can get the relevant statistics component of every feature. The larger value indicates the feature is more important. After that, first *k* features subset can be selected for classification.

http://dx.doi.org/10.21037/atm-20-4076

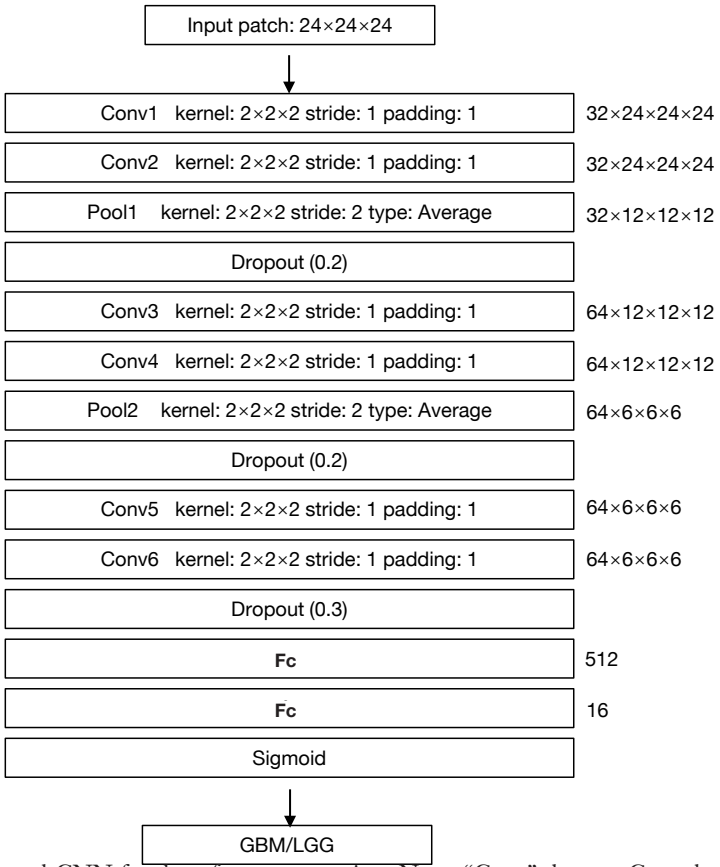

Figure 1 The architecture of the proposed CNN for deep feature extraction. Note: "Conv" denotes Convolutional layer; the numbers follow "Conv" represents the specific layer of convolution; "Pool" denotes pooling layer; the numbers follow "Pool" represents the specific layer of pooling; "Fc" denotes fully connected layer.

http://dx.doi.org/10.21037/atm-20-4076

Table S1 Different kernel functions for kernel fusion

Kernel	Kernel fusion function	Parameters
Linear	$K_{LIN}\left(X_i, X_j\right)=\sum_{v=1}^{4}\omega_v F_{LIN}^{v}\left(X_i, X_j\right)$	$F_{LIN}\left(X_i, X_j\right)=\sum_{k=1}^{N}\left(X_i^k\right)^T \times X_j^k$
Polynomial	$K_{POL}\left(X_i, X_j\right)=\left(1+\sum_{v=1}^{4}\omega_v F_{POL}^{v}\left(X_i, X_j\right)\right)^d$	$F_{POL}\left(X_i, X_j\right)=\sum_{k=1}^{N} X_i^k \times X_j^k$
Gaussian (RBF)	$K_{RBF}\left(X_i, X_j\right)=\exp\left(\frac{-\sum_{v=1}^{4}\omega_v D_{RBF}^{v}\left(X_i, X_j\right)}{2\sigma^2}\right)$	$D_{RBF}\left(X_i, X_j\right)=\sum_{k=1}^{N}\left(X_i^k - X_j^k\right)^2$
Sigmoid	$K_{SIG}\left(X_i, X_j\right)=\tanh\left(\beta \times \sum_{v=1}^{4}\omega_v F_{SIG}^{v}\left(X_i, X_j\right)-\theta\right)$	$F_{SIG}\left(X_i, X_j\right)=\sum_{k=1}^{N} X_i^k \times X_j^k$
Intersection (HIK)	$K_{HIK}\left(X_i - X_j\right)=\sum_{v=1}^{4}\omega_v F_{HIK}^{v}\left(X_i, X_j\right)$	$F_{HIK}\left(X_i, X_j\right)=\sum_{k=1}^{N}\min\left(X_i^k - X_j^k\right)$
Chi-square (χ^2)	$K_{\chi^2}\left(X_i, X_j\right)=\exp\left(\frac{-\sum_{v=1}^{4}\omega_v D_{\chi^2}^{v}\left(X_i, X_j\right)}{2\sigma^2}\right)$	$D_{\chi^2}\left(X_i, X_j\right)=\frac{1}{2}\sum_{k=1}^{N}\frac{\left(X_i^k - X_j^k\right)^2}{\left\|X_i^k + X_j^k\right\|}$

Note: *v =1, 2, 3, 4* responds to the radiomics features on T2 FLAIR and T1ce, deep features on T2 FLAIR and T1ce, respectively; X_i and X_j are the representation of the *i-th* and *j-th* training data; σ, d, β, θ are hyper-parameters determined on experiments; D is the distance function between two objects; F is the basic kernel function between two objects; where k represents the *k-th* feature in the feature vector X_i *and* X_j; N is the dimension of image features.

http://dx.doi.org/10.21037/atm-20-4076

Table S2 Performance of CNN structure with different parameters on internal validation cohort

Kernel size	Stride	Activation	Pooling type	AUC (T2 FLAIR)	AUC (T1ce)
2×2×2	1	Relu	Average	0.81 (0.78, 0.83)	0.82 (0.80, 0.84)
3×3×3	1	Relu	Average	0.76 (0.74, 0.78)	0.80 (0.78, 0.82)
4×4×4	1	Relu	Average	0.80 (0.77, 0.81)	0.76 (0.74, 0.78)
5×5×5	1	Relu	Average	0.76 (0.73, 0.78)	0.75 (0.73, 0.77)
2×2×2	1	Relu	Average	0.81 (0.78, 0.83)	0.82 (0.80, 0.84)
2×2×2	2	Relu	Average	0.78 (0.76, 0.80)	0.80 (0.78, 0.82)
2×2×2	1	Relu	Average	0.81 (0.78, 0.83)	0.82 (0.80, 0.84)
2×2×2	1	LeakyRelu	Average	0.77 (0.75, 0.79)	0.80 (0.78, 0.82)
2×2×2	1	Sigmoid	Average	0.76 (0.74, 0.78)	0.78 (0.76, 0.80)
2×2×2	1	Relu	Max	0.77 (0.74, 0.79)	0.80 (0.78, 0.82)
2×2×2	1	Relu	Average	0.81 (0.78, 0.83)	0.82 (0.80, 0.84)

Table S3 The selected radiomics features on T1ce

Feature	Wavelet band-pass filtering	Isotropic voxel size	Quantization algorithm	Number of gray level
GLCM/Sum Average	2	In-pR	Equal	16
GLCM/Sum Average	2	1 mm	Equal	16
GLCM/Sum Average	2	In-pR	Equal	8
GLSZM/LGRE	2/3	3 mm	Equal	64
GLCM/Variance	2	2 mm	Equal	32
GLCM/Auto Correlation	2	1 mm	Equal	16
GLSZM/LGRE	1/2	4 mm	Equal	64
GLCM/Sum Average	2	1 mm	Equal	8
GLSZM/LGRE	2	3 mm	Equal	32
GLCM/Variance	2	1 mm	Equal	8
GLCM/Auto Correlation	2	1 mm	Equal	8
GLSZM/LGRE	2	In-pR	Equal	64
GLSZM/LGRE	1/2	4 mm	Equal	32
GLCM/Sum Average	2	2 mm	Equal	16
GLSZM/LGRE	1/2	5 mm	Equal	32
GLSZM/LGRE	2	5 mm	Equal	16
GLCM/Sum Average	1/2	1 mm	Equal	16
GLCM/Sum Average	2	2 mm	Equal	8
GLCM/Variance	2	1 mm	Equal	16

Note: “In-pR” denotes initial in-plane resolution.

http://dx.doi.org/10.21037/atm-20-4076

Table S4 The selected radiomics features on T2 Flair

Feature	Wavelet band-pass filtering	Isotropic voxel size	Quantization algorithm	Number of gray level
GLSZM/LGRE	2	4 mm	Equal	64
GLSZM/LGRE	3/2	5 mm	Equal	32
GLSZM/LGRE	2	4 mm	Equal	32
GLSZM/LGRE	3/2	5 mm	Equal	32
GLSZM/SRLGE	3/2	5 mm	Equal	64
GLSZM/LGRE	3/2	5 mm	Equal	64
GLSZM/SRLGE	2	4 mm	Equal	64
GLSZM/LGRE	3/2	3 mm	Equal	64
GLSZM/LGRE	2	3 mm	Equal	64
GLSZM/LGRE	2	2 mm	Equal	64
GLSZM/LGRE	2	3 mm	Equal	32
GLSZM/SRLGE	3/2	5 mm	Equal	64
GLSZM/LGRE	3/2	2 mm	Equal	64
GLSZM/LGRE	2	2 mm	Equal	32
GLSZM/SRLGE	3/2	5 mm	Equal	32
GLSZM/LGRE	3/2	4 mm	Equal	32
GLSZM/SRLGE	2	3 mm	Equal	64
GLSZM/LGRE	2	1 mm	Equal	64
GLSZM/LGRE	3/2	5 mm	Equal	32

Note. “in-pR” denotes initial in-plane resolution.

Table S5 The selected deep learning features on T2 Flair and T1ce

	Selected feature (*i-th* feature of 13824 deep features)
T2 Flair	11439, 2353, 13024, 1146, 3450, 2737, 667, 3834, 273, 6040, 10838, 5265, 1530, 5656, 1432, 4881, 938, 7415, 3428
T1ce	68, 5072, 2391, 452, 2454, 2775, 168, 5060, 2756, 1, 4695, 385, 4676, 80, 4688, 2372, 471, 87, 1325

Table S6 The best performances of different kernel types with mixing coefficients on internal validation cohort

Kernel	Coefficients	AUC	Sensitivity (%)	Specificity (%)
Linear	$\omega_1 = 0.40, \omega_2 = 0.10, \omega_3 = 0.10, \omega_4 = 0.40$	0.90 (0.77-0.97)	81 (17/21)	92 (24/26)
Polynomial	$\omega_1 = 0.40, \omega_2 = 0.10, \omega_3 = 0.10, \omega_4 = 0.40$	0.92 (0.80-0.98)	81 (17/21)	88 (23/26)
Sigmoid	$\omega_1 = 0.35, \omega_2 = 0.10, \omega_3 = 0.15, \omega_4 = 0.40$	0.91 (0.79-0.97)	86 (18/21)	85 (22/26)
Gaussian	$\omega_1 = 0.40, \omega_2 = 0.30, \omega_3 = 0.10, \omega_4 = 0.20$	0.93 (0.82-0.99)	86 (18/21)	88 (23/26)
Intersection	$\omega_1 = 0.35, \omega_2 = 0.10, \omega_3 = 0.20, \omega_4 = 0.35$	0.91 (0.80-0.98)	86 (18/21)	85 (22/26)
Chi-square	$\omega_1 = 0.35, \omega_2 = 0.15, \omega_3 = 0.15, \omega_4 = 0.35$	0.94 (0.85-0.99)	86 (18/21)	92 (24/26)

http://dx.doi.org/10.21037/atm-20-4076