Sum and difference histogram and texture feature

Pixel pairs in image I with relative distances of d1 and d2 are expressed as:

$$y_1 = I(m,n) \tag{2}$$

$$y_2 = I(m+d_1, n+d_2) \tag{3}$$

y_1 and y_2 represent the pixel values at image positions (m, n) and (m + d_1, n + d_2), respectively.

The sum and difference of the pixel pairs y_1 and y_2 are:

$$s_{m,n} = y_1 + y_2 \tag{4}$$

$$d_{m,n} = y_1 - y_2 \tag{5}$$

Histogram of sum and difference of image I is:

$$h_s(i;d_1,d_2) = h_s(i) = Card\{(m,n) \in I \mid s_{m,n} = i\} \tag{6}$$

$$h_d(j;d_1,d_2) = h_d(j) = Card\{(m,n) \in I \mid d_{m,n} = j\} \tag{7}$$

Total number of elements in the sum or difference histogram:

$$N = \sum h_s(i) = \sum h_d(i) \tag{8}$$

Histogram normalization:

$$Ps(i) = \frac{h_s(i)}{N}, i = 2,3 \cdots 2Ng \tag{9}$$

$$Pd(j) = \frac{h_d(j)}{N}, j = -Ng+1, \cdots, Ng+1 \tag{10}$$

Ps is normalized sum histogram and Pd is normalized difference histogram. Mean, variance, energy, correlation, entropy, contrast, homogeneity, cluster shade and cluster prominence are calculated by Ps and Pd:

$$\text{mean} = \frac{1}{2}\sum_i i \cdot Ps(i) = \mu \tag{11}$$

$$\text{variance} = \frac{1}{2}\left\{\sum_i (i-2u)^2 \cdot Ps(i) + \sum_j j^2 \cdot Pd(j)\right\} \tag{12}$$

$$\text{energy} \approx \sum_i Ps(i)^2 \cdot \sum_j Pd(j)^2 \tag{13}$$

$$\text{correlation} = \frac{1}{2}\left\{\sum_i (i-2\mu)^2 \cdot Ps(i) - \sum_j j^2 \cdot Pd(j)\right\} \tag{14}$$

$$\text{entropy} = -\sum_i Ps(i) \cdot \log(Ps(i)) - \sum_j Pd(j) \cdot \log(Pd(j)) \tag{15}$$

$$\text{contrast} = \sum_j j^2 \cdot Pd(j) \tag{16}$$

$$\text{homogeneity} = \sum_j \frac{1}{(i+j)^2} \cdot Pd(j) \tag{17}$$

 https://dx.doi.org/10.21037/tcr-21-2362

$$\text{cluster shade} = \sum_i (i - 2\mu)^3 \cdot Ps(i) \quad [18]$$

$$\text{cluster prominence} = \sum_i (i - 2\mu)^4 \cdot Ps(i) \quad [19]$$

PCA and Relief-F

The main idea of PCA is to carry out orthogonal basis transformation on the data and project the data of n-dimensional space onto the k-dimensional space, meanwhile maximizing the variance of the projected data. The k dimension is a completely new orthogonal feature also known as the principal component. The role of PCA is to find a set of mutually orthogonal coordinate axes sequentially from the original space. The choice of new coordinate axes is closely related to the data itself.

Relief-F algorithm is a feature weighting algorithm, which assigns different weights to features according to the relevance between features and categories. The higher the weight of the feature is, the more powerful the classification ability of the feature is. Thus, features can be selected according to the weight.

LDA, SVM and DT

LDA is a classic algorithm for pattern recognition, which is nonparametric and easy to train. The idea of LDA is to project high-dimensional samples into one-dimensional space, making the distance between samples of the same category as small as possible and the distance between samples of different categories as large as possible.

SVM is a supervised learning algorithm, which uses an optimal hyperplane to separate samples in feature space. In the case of linear inseparability, samples can be mapped to high-dimensional space by kernel function, so that samples can be linearly separable in high-dimensional space. SVM is suitable for small sample classification.

DT is a prediction model expressed in the form of tree structure. DT consists of nodes and directed edges. There are two types of nodes: internal nodes and leaf nodes. Internal nodes represent a feature or attribute, and leaf nodes represent a class. The key of DT is how to choose the optimal partition attribute at each split node. In our study, Iterative Dichotomiser 3 (ID3) algorithm is used to generate DT.

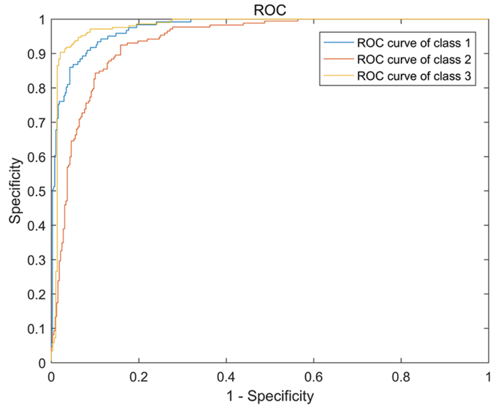

Figure S1 ROC curve of different class. Each class was considered as a positive sample in turn and the rest was treated as a negative sample, thus there are three ROC curves. The AUC of class 1, 2, and 3 were 0.97, 0.93, and 0.98, respectively. ROC, receiver operator characteristic; AUC, area under the curve.

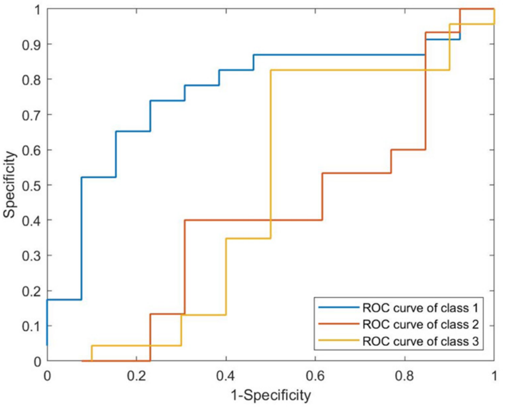

Figure S2 ROC curve of different class in the validation cohort. Each class was considered as a positive sample in turn and the rest was treated as a negative sample, thus there are three ROC curves. The AUC of class 1, 2, and 3 were 0.76, 0.41, and 0.48, respectively. ROC, receiver operator characteristic; AUC, area under the curve.

 https://dx.doi.org/10.21037/tcr-21-2362