

What is claimed is:

1. An evaluation method for multimodal medical data fusion, the method comprising:

acquiring to-be-evaluated medical data in multiple modes of a target object;

5 performing feature extraction on to-be-evaluated medical data in each mode to obtain a plurality of feature vectors;

fusing the a plurality of feature vectors to obtain a fusion feature vector; and

10 inputting the fusion feature vector into a pre-trained multimodal fusion evaluation model to obtain an evaluation result of the to-be-evaluated medical data in multiple modes output by the pre-trained multimodal fusion evaluation model;

wherein the inputting the fusion feature vector into a pre-trained multimodal fusion evaluation model to obtain an evaluation result of the to-be-evaluated medical data in multiple modes output by the pre-trained multimodal fusion evaluation model comprises:

15 splicing each feature vector in the fusion feature vector horizontally to obtain a first matrix $W(In)$ of feature vector, and coding position of the first matrix $W(In)$ of feature vector by a first function to obtain a second matrix $W(P)$ of feature vector, wherein the formula is as follows:

$$p(t) = \begin{cases} \sin(\frac{pos}{10000^{2i/d}}), & i \% 2 = 0 \\ \cos(\frac{pos}{10000^{2i/d}}), & i \% 2 = 1 \end{cases}$$

20 wherein, t represents a sub-vector in the first matrix $W(In)$ of feature vector; $p(t)$ represents a coding result corresponding to the t value; pos indicates which feature vector the vector t belongs to; i represents a sequence number of the vector t in the first matrix $W(In)$ of feature vector; and d represents the number of horizontal dimensions of the first matrix $W(In)$ of feature vector;

25 inputting the second matrix $W(P)$ of feature vector into a second function, and calculating and obtaining a high-dimensional feature representation matrix $W(M)$ on a subspace wherein the formula is as follows:

$$W(M) = \text{Concat}(F(1), F(2), \dots, F(i)) \cdot W^0;$$

$$F(i) = \text{softmax}(\frac{Q(x) K^T(x)}{\sqrt{d(x)}}) \cdot V$$

wherein, a CONCAT function represents the second function, and $F(1), F(2) \dots F(i)$ represents the i th feature sub-vector in the second matrix $W(P)$ of feature vector is conducted the formula F calculation; W^O represents transposition of the first matrix $W(In)$ of feature vector;

5 x in $F(i)$ represents the i th feature sub-vector in the input second matrix $W(P)$ of feature vector; Q, K and V represent the linear sensing layer of parameter n of the hidden layer of the multimodal fusion evaluation model; and $Q(x)$ represents linear regression of x ; and

10 coding the feature vector of each image through an encoder of the multimodal fusion evaluation model, inputting output $W(Out)$ of the encoder to a linear regression layer, converting $W(Out)$ to a low-dimensional feature representation matrix through the linear regression layer, and finally outputting an evaluation result through operation of a softmax function.

15 2. The method according to claim 1, wherein the acquiring the to-be-evaluated medical data in multiple modes of a target object comprises at least three of the following:

acquiring a rectal cancer image data set of the target object as first modal data, wherein the rectal cancer image data set at least comprises a macro-view image, a near-view image and a micro-view image determined according to a tumor region or a regressing tumor region;

20 acquiring a rectal cancer magnetic resonance image data set of the target object as second modal data, wherein the rectal cancer magnetic resonance image data set comprises initial rectal cancer magnetic resonance image data and target rectal cancer magnetic resonance image data; marking a tumor region or a regressing tumor region in the initial rectal cancer magnetic resonance image data and the target rectal cancer
25 magnetic resonance image data respectively to obtain a plurality of slice images containing the tumor region or the regressing tumor region;

acquiring an initial clinical data set and a target clinical data set of the target object as a third modal data, wherein the initial clinical data set and the target clinical data set at least comprise personal information and case information of the target object; and

30 acquiring initial tumor marker information, target tumor marker information, initial blood information and target blood information of the target object as fourth modal data.

3. The method according to claim 2, wherein the performing feature extraction on to-be-evaluated medical data in each mode to obtain a plurality of feature vectors

comprises:

inputting the first modal data and the second modal data to a pre-trained neural network model respectively;

5 performing matrix connection on medical images in the first modal data and the second modal data respectively through a hard-wired layer of the neural network model;

performing convolution calculation and maximum pooling operation on a medical image after the matrix connection through alpha 3D convolution modules of the neural network model to extract high-dimensional feature maps; and

10 converting a high-dimensional feature map extracted from the last 3D convolution kernel into a one-dimensional feature vector through beta up-sampling modules and a full-connection layer of the neural network model to obtain the first feature vector and a second feature vector respectively.

4. The method according to claim 2, wherein the performing feature extraction on to-be-evaluated medical data in each mode to obtain a plurality of feature vectors
15 comprises:

mapping text description features in the third modal data and the fourth modal data into corresponding numerical features; and

mapping the numerical features into a two-dimensional matrix to obtain a third feature vector and a fourth feature vector respectively.

20 5. The method according to claim 3, wherein a training process of the neural network model comprises:

inputting acquired preset to-be-evaluated medical data as a training sample into a corresponding initial neural network model to enable the initial neural network model to output a corresponding initial feature vector;

25 if the initial feature vector meets preset requirements, determining the initial neural network model to be successfully trained to obtain the pre-trained neural network model;

if the initial feature vector does not meet the preset requirements, continuing to train the initial neural network model by adjusting a loss parameter in the initial neural network model until the loss parameter is fitted to reach a preset loss parameter threshold so as
30 obtain the pre-trained neural network model.

6. The method according to claim 1, wherein a cross-entropy loss function is used for parameter back-propagation and update during training of the multimodal fusion evaluation model until the cross-entropy loss function is fitted.

7. An evaluation device for multimodal medical data fusion, wherein the device comprises:

a medical data acquisition module, configured to acquire to-be-evaluated medical data in multiple modes of a target object;

5 an feature vector extraction module, configured to perform feature extraction on to-be-evaluated medical data in each mode to obtain a plurality of feature vectors;

an feature vector fusion module, configured to fuse the a plurality of feature vectors to obtain a fusion feature vector; and

10 a multimodal fusion evaluation module, configured to input the fusion feature vector into a pre-trained multimodal fusion evaluation model to obtain an evaluation result of the to-be-evaluated medical data in multiple modes output by the pre-trained multimodal fusion evaluation model, and specifically configured to:

splice each feature vector in the fusion feature vector horizontally to obtain a first matrix W(In) of feature vector, and code position of the first matrix W(In) of feature vector by a first function to obtain a second matrix W(P) of feature vector, wherein the formula is as follows:

$$p(t) = \begin{cases} \sin(\frac{pos}{10000^{2i/d}}), & i \% 2 = 0 \\ \cos(\frac{pos}{10000^{2i/d}}), & i \% 2 = 1 \end{cases}$$

wherein, t represents a sub-vector in the first matrix W(In) of feature vector; p(t) represents a coding result corresponding to the t value; pos indicates which feature vector the vector t belongs to; i represents a sequence number of the vector t in the first matrix W(In) of feature vector; and d represents the number of horizontal dimensions of the first matrix W(In) of feature vector;

20 input the second matrix W(P) of feature vector into a second function, and calculate and obtain a high-dimensional feature representation matrix W(M) on a subspace, wherein the formula is as follows:

$$W(M) = \text{Concat}(F(1), F(2), \dots, F(i)) \cdot W^0;$$

$$F(i) = \text{softmax}(\frac{Q(x) K^T(x)}{\sqrt{d(x)}}) \cdot V$$

wherein, a CONCAT function represents the second function, and F(1), F(2)... F(i)

represents the i th feature sub-vector in the second matrix $W(P)$ of feature vector is conducted the formula F calculation; W^O represents transposition of the first matrix $W(In)$ of feature vector;

5 x in $F(i)$ represents the i th feature sub-vector in the input second matrix $W(P)$ of feature vector; Q , K and V represent the linear sensing layer of parameter n of the hidden layer of the multimodal fusion evaluation model; and $Q(x)$ represents linear regression of x ; and

10 code the feature vector of each image through an encoder of the multimodal fusion evaluation model, input output $W(Out)$ of the encoder to a linear regression layer, convert $W(Out)$ to a low-dimensional feature representation matrix through the linear regression layer, and finally output an evaluation result through operation of a softmax function.8. An electronic device, wherein the device comprises a processor, a communication interface, a memory and a communication bus, wherein the processor, the communication interface and the memory complete communication with each other through the
15 communication bus;

a memory, configured to store a computer program; and

a processor, configured to implement the method steps according to any one of claims 1 to 6 when executing the program stored on the memory.

20 9. A computer-readable storage medium, where a computer program is stored, wherein method steps according to any one of claims 1 to 6 are implemented when the computer program is executed by a processor.