Study on AO classification of distal radius fractures based on multi-feature fusion

Feng Yang¹, Rikun Cong¹, Mengmeng Xing¹ and Bo Ding¹*

¹Network Information Center, Shandong University of Traditional Chinese Medicine, Jinan, Shandong, 250355, China
*Corresponding author e-mail: dingbo@sducm.edu.cn (ORCID: 0000-0002-7909-8244)

Abstract. Accurate classification of distal radius fracture is of great significance for improving the accuracy and success rate of subsequent bone-setting techniques. Based on the existing clinical distal radius fracture cases of the research group, this paper proposes an AO classification method based on distal radius fracture images based on multi-feature fusion for the problems of poor single feature expression and low accuracy of fracture classification by traditional classifiers and deep learning models. The fusion model uses two traditional features and the depth features extracted by AlexNet. After reducing the dimensions of the above features, a special neural network is designed to effectively fuse the reduced feature vectors. Finally, use the image retrieval classification scheme DML-K proposed in this paper to realize the specific classification of DRF images with the fused feature vectors. The experimental results show that the accuracy of the diagnostic model proposed in this paper on the distal radius fracture data set reaches 83.4%, and the F1 value reaches 0.815. Through horizontal and vertical comparison with other algorithms, the accuracy of the DRF diagnosis model proposed in this paper is increased by 5%, and the F1 value is increased by about 0.2, which fully verifies the effectiveness and feasibility of the method proposed in this study, and is expected to be applied Machine-aided diagnosis of distal radius fracture.

1. Introduction
Distal radius fracture (DRF) is one of the most common fractures in clinical upper limb fractures [1]. It is mostly caused by the hand supporting the ground when falling, or it may be caused by direct violence. It is usually divided into intra-articular fractures and extra-articular fractures. There are many typing methods for DRF. At present, DRF's AO typing has become the most commonly used method recognized in clinical practice with its comprehensive typing methods. The practice has proved that the correct classification of fractures before surgery is of great significance to the selection of subsequent reduction schemes, fracture fixation methods and prognostic evaluation. However, most of the current fracture inspections are X-ray images, which can only obtain images and data superimposed on the entire radius, which makes DRF images lack a clear lesion display. In addition, due to the irregular bone structure of the wrist, many bone fragments, and complex fracture forms such as overlap, compression, rotation, and insertion, it is difficult for orthopedists to judge the specific type of DRF, which can easily cause misdiagnosis and affect subsequent reduction methods. The choice [2]. And there is usually only one physician in the diagnosis process, and the absence of a second auxiliary opinion increases the misdiagnosis rate. Using machine learning and deep learning algorithms to establish a DRF diagnosis model to realize automatic classification of DRF, the
diagnosis result of the model can be referenced by the doctor before the diagnosis is made to provide a reference for the doctor's subsequent diagnosis. Therefore, the realization of DRF classification based on CAD can not only provide a reference for doctors to read pictures, but also reduce the diagnosis burden of doctors and improve diagnosis efficiency, which has very important practical significance.

There are two main feature representation methods currently used in the field of fracture diagnosis, traditional feature representation methods based on machine learning and deep feature representation methods based on deep learning. For example, Yap et al. [3] proposed an automatic method for detecting leg bone fractures based on the texture features of leg bone fracture images. First, locate the boundary of the leg bone fracture image by using the Sobel operator, and then use the gray level co-occurrence matrix(GLCM) algorithm to extract its local texture features, and then use traditional machine learning classifiers such as k-Nearest Neighbor(kNN)[4], Support Vector Machine (SVM) [5] and Extreme Learning Machine(ELM)[6] complete the automatic classification of leg bone fracture images, and achieve a fracture diagnosis accuracy rate of 78.2%. Although the simple use of traditional features has been able to obtain better classification results, the limitation of the high-level semantics of the image itself leads to the lack of traditional feature expression capabilities. Compared with traditional features, the deep learning framework has stronger feature extraction capabilities and can systematically extract features from low-level to high-level from DRF images. For example, Chung et al.[7] used Convolution Neural Network (CNN) to extract and classify the depth features of proximal humeral fracture images. The experimental results show that the classification accuracy of the deep network is significantly higher than that of radiologists and general orthopedics surgeons. Although deep learning has achieved good classification results, there is still room for further improvement in the accuracy of fracture classification based on deep learning. The practice has proved that the comprehensive utilization of multiple features in classification can improve the subsequent classification effect more than using a single feature. For example, Bodla et al. [8] proposed a feature fusion method based on deep heterogeneous features based on the facial features, and experimentally designed a multi-layer CNN model to serially connect multiple types of deep features. Khusnuliarwati et al.[9] realized the serial fusion of features based on multiple traditional local texture features of the image in the form of histograms. The above-mentioned feature fusion method has achieved certain results, but the accuracy rate has not been greatly improved. Analysis of the reason may be that simple serial connection of multiple feature vectors often cannot effectively fuse feature information, and it will cause dimensionality disaster. To solve the above problems, better feature fusion methods should be explored. For example, Bai et al. [10] used the idea of distance metric learning to propose a feature fusion method that assigns weights. The core is to determine the query object belongs to a certain category by calculating the distance from a single query object to each category. Probability value, assign a certain weight to multiple features according to the obtained probability, so that multiple feature vectors can be effectively merged. Shi et al. [11] based on CNN's layered denoising sparse autoencoder (Deep Sparse Autoencoder, DASE) proposed a nonlinear metric learning method to achieve deep fusion of different features. The above studies have achieved varying degrees of progress in related fields. However, because orthopedic image datasets are rarely publicized and are limited by the number and type of fracture image datasets and the special structure of the carpal bone, researchers have designed classifications based on the characteristics of fracture images. Based on this, this paper designs a new DRF image classification and diagnosis model based on multi-feature fusion and distance measurement learning based on the characteristics of DRF images.

The main contributions of this paper are as follows, First, the effective fusion of different types of features is realized by designing new feature fusion methods. Second, based on the algorithm concept of similarity measurement in image retrieval, give specific image retrieval examples, and design a multi-classification algorithm for DRF images. Third, propose a new DRF image diagnosis model, which effectively combines depth features with traditional features and at the same time and give play to the advantages of each of the two features in extracting deep semantic features and underlying features of DRF images, and input them into the new classification algorithm. It is expected to obtain the distinguishing features related to fracture classification while further improving the classification
accuracy. The experimental results prove the effectiveness and feasibility of the proposed algorithm for DRF diagnosis.

2. Materials and method

2.1. Experimental data

The experimental data adopts the DRF data collected by the cooperative hospital Shandong Provincial Hospital of Traditional Chinese Medicine and the Affiliated Hospital of Shandong University of Traditional Chinese Medicine from March 2014 to March 2020, and a large number of DRF clinical diagnosis and treatment cases and clinical teaching collected by the research team in the early stage. The fracture image database established by clinical case information has a total of 1189 samples. In the DRF data set, the specific categories of each DRF image are manually marked by professional orthopedics and traumatologists with more than 10 years of work experience. The classification criteria are as follows:

I Type A, that is, extra-articular fracture. It mainly included comminution and insertion of the outside of the joint, 489 cases in total.

II Type B, namely simple or partial intra-articular fractures. Divided into sagittal fractures of the distal radius, dorsal marginal fractures (back Barton fractures), volar marginal fractures (volar Barton fractures), a total of 344 cases.

III Type C, that is, complex intra-articular fractures. Including simple and comminuted fractures in the joints without comminuted metaphysis, a total of 356 cases. The data is desensitized in batches, including removing the name and identification number fields related to the patient's basic information, and the pinyin of the patient's name in the image. Figure 1 shows three examples of DRF's specific fracture types.

![Figure 1. Examples of three specific types of DRF](image)

2.2. Experimental method

This article proposes an AO classification model suitable for DRF by analyzing the characteristics of DRF images. For the existing labeled DRF X-ray images, the traditional features and depth features are extracted separately. After feature selection, a two-layer convolutional neural network (Convolutional Neural Networks, CNN) is used to correlate the above features to obtain represents the feature vector of the DRF image. Finally, on the basis of image retrieval, distance metric learning is used to design the DRF automatic classification algorithm (DML-K). The specific process of the diagnosis model is shown in Figure 2.
2.3. Image preprocessing

The DRF classification data sets used in the experiment are all X-ray images. X-ray images have the problem of low brightness and poor contrast. Therefore, DRF images must be preprocessed. Contrast Limited Adaptive Histogram Equalization (CLAHE) [12] can improve DRF image brightness, increase image contrast, suppress noise, highlight the internal details of DRF images, and better display DRF images the lesion information.

The implementation process of the CLAHE algorithm is as follows:

I. Image block, divide a single DRF image into M×N individual image blocks.

II. Calculate the gray histogram of a single segmented image block with block as the unit, and then average the pixel value of the image block to each gray level. The number of average pixels in the gray scale can be represented by $N_a$

$$N_a = \frac{N_{xp} - N_{yp}}{N_g}$$  (1)

In formula (1), $N_g$ represents the number of gray levels in the image block; $N_{xp}$ and $N_{yp}$ represent the number of pixels along with the x-axis and y-axis directions of the image block, respectively.

III. Intercept the value greater than $N_L$ in the gray-scale histogram, the calculation formula of $N_L$ is as follows:

$$N_L = SN_a$$  (2)

Among them, S represents the interception coefficient. Calculate the number of intercepted pixels and the total number of pixels averaged to a single gray level.

IV. Histogram equalization. Perform equalization processing on the image block histogram obtained in the above steps, and use the transformation function to obtain the gray value after equalization processing. Figure 3 shows the before and after comparison of the DRF image processed by the CLAHE algorithm.
3. Feature extraction

The current feature representation methods are mainly divided into two types: traditional feature representation methods based on machine learning and deep feature representation methods based on deep learning. The traditional feature representation method is a targeted feature extraction algorithm designed according to the characteristics of the target image, which usually describes a specific type of image information. However, due to the pertinence of its algorithm, it does not apply to all categories of images, so traditional feature extraction has great limitations. In addition, the specific part of the fracture in the DRF image occupies a very small proportion of the entire image, which will result in a weaker ability to locate the required diagnosis of the lesion by a single feature. In recent years, CNN has achieved good results in image feature extraction and can express high-level semantic information of images. CNN can automatically learn the deep features of the image and can extract features that are difficult to achieve by human design without the prior knowledge and complex algorithm design required by traditional features. But CNN also has its limitations. For example, the low-level convolutional layer's receptive field is too small, so that the depth features usually ignore the global features. If we can combine traditional features with depth features in a targeted manner, we can combine the advantages of the two, increase the diversity of feature expression, and achieve a more comprehensive representation of DRF images.

3.1. Traditional feature extraction

Because X-ray images have the characteristics of low gray values, blurred boundaries, and complex texture information, it is usually difficult to perform comprehensive and accurate feature extraction on specific lesions in DRF images. In order to accurately express the shape and texture features of the DRF image, this paper selects two traditional feature extraction methods of Local Binary Patterns (LBP) features [13] and Gabor features [14] to extract the traditional features of DRF images.

LBP can represent the local texture information of the image. The feature extraction steps are as follows: First, divide the image into 16×16 cells, and compare the gray value of each pixel in the cell with the value of the remaining 8 pixels next to each other. When the pixel value at this position is greater than the center position when the value of the pixel is set, the position of the pixel is marked as 1, otherwise it is recorded as 0. After comparing and calculating one by one, an 8-bit binary number is generated in the 3*3 neighbourhood, which is converted into a decimal number according to the conversion rules between the hexadecimals, and the LBP value of the pixel at that position is obtained. After counting the frequency of occurrence of the number, the histogram of each cell is obtained, and the obtained histogram is normalized and connected into a feature vector, totalling 59 dimensions.

The Gabor function can extract relevant features of the image in different scales and directions in the frequency domain. The feature extraction is mainly divided into the following two steps: First, the input image \( I(x, y), x, y \in \Omega \) (\( \Omega \) represents the set of image pixels) and the real and imaginary parts of the Gabor filter \( G(x, y) \) are convolved to obtain the feature image \( R(x, y) \), and then the gray value of
the feature image pixel is obtained. The 48-dimensional feature vector of mean and variance is used as the final Gabor texture feature.

It has been found that combining Gabor features with other local feature descriptors based on image gradients can improve feature expression capabilities. Therefore, this article combines Gabor with LBP local feature descriptors to achieve the purpose of obtaining efficient feature representation. This article uses the feature vector of a DRF image. Among them \( i \) is the total number of images of DRF and \( n \) is the dimension of the feature.

\[
x_i = [f_1, f_2, \cdots, f_{100}, f_{101}, \cdots, f_n]
\]

### 3.2. Deep feature extraction

The depth feature can represent the high-level semantic information of the DRF image, so extracting the depth feature of the DRF image for subsequent classification tasks can improve the accuracy of subsequent fracture type diagnosis. The feature extraction of CNN is done through the convolutional layer, and the local perception information of the image is extracted through the convolution kernel with fixed dimensions. Among them, the shallow convolutional layer can extract low-level features such as edges and textures of the DRF image, and the deeper convolutional layer extracts the deep abstract features of the DRF image. Currently, CNN has been widely used in the field of image classification, but due to its network characteristics, there are certain shortcomings. As mentioned in [15], two 3×3 convolutional layers can be stacked to obtain a 5×5 receptive field. The receptive field of the low-level convolutional layer is very small, and the receptive field can gradually increase with the stacking of the network. Therefore, in order to obtain a larger receptive field, it is often necessary to deepen the number of network layers and increase the size of the convolution kernel, which also leads to a more complex network structure and more network parameters, thereby increasing computing time.

This paper compares the complexity and number of parameters of the current classic CNN frameworks such as LeNet-5, AlexNet, GoogLeNet, and VGG-16 [16] to select the optimal deep feature extraction framework. For example, LeNet-5 defines the image input size as 32×32, which is much smaller than the AlexNet input 227×227. Therefore, AlexNet can retain more DRF image information compared with LeNet-5; also, the number of network layers of VGG-16 is much larger than AlexNet. The network structure of GoogLeNet is more complicated, while the AlexNet network has only 8 layers. Based on the comparison and analysis of the above indicators, this article finally decided to choose AlexNet as the basic framework for DRF image depth feature extraction. The network parameters of each layer of AlexNet used are shown in Table 2. Depth feature extraction of DRF image: The DRF image passes through all the convolutional layers of AlexNet and the last pooling layer and outputs the feature vector, which is the depth feature of the fractured image, with a total of 4096 dimensions.

The depth feature extraction algorithm of DRF image is as follows:

I. The DRF image is used as the input of the AlexNet model, and the convolution kernel with fixed dimensions is used to extract the local perception information of the image. The final fully connected layer is used as the depth feature of the DRF image, which is recorded \( \{ (x^{(i)}, y^{(i)}) | i = 1, 2, \cdots, m \} \).

II. For the output depth feature vector \( x^{(i)} \), the calculation formula \( J(\theta) \) of the cost function is as follows.

\[
J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1 - y^{(i)}) \log (1 - h_\theta(x^{(i)}))
\]
Among them $y^{(i)}$ is the DRF image label, and the output of the softmax layer is represented by $h_{\theta}(x^{(i)})$.

III. Optimize the cost function and use the gradient descent algorithm to obtain the optimal parameter $\theta$.

IV. Use the trained AlexNet deep learning model to extract the depth features $x_j = [g_1, g_2, \cdots g]$ of the DRF image.

3.3. Feature Fusion

Due to the difference in the shooting position of the X-ray image, the target fracture site may be located at the edge of the entire image. If you only focus on the local information, it is easy to cause misjudgment. Since AlexNet requires the input of a fixed image size of $227 \times 227$, and the size of the DRF image in this paper is generally higher than this size, some image information will be discarded before the image is input, and the convolution operation of the local receptive field will also affect the global characteristics of the image. The deep semantic features of images extracted by CNN are often abstract features, which lack the supplement of the underlying features of the image. The traditional features are mostly aimed at the shape and texture of the image, and are usually less affected by the size. Therefore, the deep features of the DRF image extracted by CNN can complement each other, thereby increasing the diversity of features. As mentioned in the introduction, there have been a variety of works using traditional features and depth features for subsequent medical image classification; however, there are not many classification tasks that effectively combine the two methods and apply them to medical images, and feature fusion The choice of method is also not fixed. The issues that need to be considered are how to perform effective feature fusion, as well as the location and method of feature fusion.

This article mainly uses LBP and Gabor algorithms to extract local texture features of DRF images and uses AlexNet to extract high-level semantic features of images. Because each feature focuses on different image scales, there are differences in the role of classification. Although the serial connection of feature vectors can improve the accuracy of subsequent classifications to a certain extent, the dimension of the feature vectors also increases. The sacrifice is the subsequent classification time, and the proportions of different features are also inconsistent. Simple concatenation Fusion is not the best choice [17]. Therefore, it is necessary to design a new feature fusion method to achieve effective fusion between multiple features, which can reduce the fusion feature vector dimension, reduce redundancy, reduce running time, and give full play to the role of different features, and retain Correlation between features.

Based on this, this section builds a two-layer CNN fusion network based on the ideas provided in Reference [18]. The network has a four-layer structure: input layer, hidden layer, softmax layer and output layer. Use gradient descent to minimize the loss function to train the entire feature fusion network to output the optimal fusion result. Figure 4 shows the flow chart of the feature fusion network.
Figure 4. The flow chart of the feature fusion network

The fusion algorithm process is as follows:

The traditional and deep features of the DRF image extracted by the above steps are concatenated as vectors $f$ and used as the input layer of the fusion neural network. The gradient descent method is used to minimize the loss function, and the weight of the hidden layer is learned, and finally, the fused feature vector is obtained. The number of nodes in the hidden layer is the dimension of the fusion feature.

Input: Traditional features $x = \{f_1, f_2, \cdots, f_n\}$, which are obtained $x' = \{g_1, g_2, \cdots, g_n\}$ after PCA dimensionality reduction is performed on the depth features to retain more than 95% of the features.

Algorithm process:

I. The traditional and deep features are serially connected to obtain the serial vector $f$:

$$f = [f_1, f_2, \cdots, f_n, g_1, g_2, \cdots, g_n]$$

II. Forward propagation

The concatenated feature vector $f$ is used as the input of the fusion network and forward propagation is performed. The objective function is as follows:

$$o(\theta, \gamma) = \text{soft} \max(\theta_2 \cdot \tanh(\theta_1 \cdot f + \gamma_1) + \gamma_2) \quad (5)$$

III. Perform back propagation

In the back propagation, formula (4) is used as the cost function of the network, and the gradient descent algorithm is used to minimize the cost function, and the parameter $\theta$ is adjusted according to the direction of the gradient change.

IV. Fusion feature vector of DRF image

The network chooses Tanh as the activation function of the hidden layer, and the output layer is the softmax function. By minimizing the loss function output by the Softmax layer, back-propagating iteratively, and updating the weights of the hidden layer, and adopting the dropout method to avoid over-fitting, the fusion feature vector of the network is finally obtained.

Output Fusion feature $x = \{u_1, u_2, \cdots, u_n\}$ of DRF image.

4. Classification algorithm

In this study, based on the distance metric learning (DML) algorithm in image retrieval (Distance Metric Learning-K, DML) [19-20], based on the $k$NN algorithm, based on the Euclidean distance to measure the similarity between images, we propose The DRF image multi-classification algorithm of DML-K.
First, calculate the Euclidean distance between the DRF image to be queried and the DRF image library, and then take out the first K images with the closest distance to the DRF image to be queried, and weight the feature vectors representing the K images according to the distance between the feature vectors. The weight factor is calculated as follows:

\[ W_k = \frac{1}{|D(q) - D(i)|}, \quad i = 1, 2, \ldots, K \]  

(6)

Among them, \( D(q) \) represents the feature vector of the image to be queried, \( D(i) \) represents the feature vector of the ith reference image, and \(|D(q) - D(i)|\) is the distance between the two image feature vectors. Take its reciprocal as the weight, the closer the distance, the greater the weight, the farther the distance, the smaller the weight. Finally, we add the weights to calculate the probability of querying the specific type of DRF image. The calculation formula is as follows:

\[
S_A = \frac{\sum_{a=1}^{a} W_a}{\sum_{a=1}^{A} W_a + \sum_{b=1}^{B} W_b + \sum_{c=1}^{C} W_c} \]  

(7)

\[
S_B = \frac{\sum_{b=1}^{B} W_b}{\sum_{a=1}^{A} W_a + \sum_{b=1}^{B} W_b + \sum_{c=1}^{C} W_c} \]  

(8)

\[
S_C = \frac{\sum_{c=1}^{C} W_c}{\sum_{a=1}^{A} W_a + \sum_{b=1}^{B} W_b + \sum_{c=1}^{C} W_c} \]  

(9)

\[ A + B + C = K \]  

(10)

Among them, \( S_A, S_B, S_C \) are the probabilities of the three types of query DRF images A, B and C, A is the number of K retrieved images that are most similar to the DRF image to be queried and the fracture type is type A, and B is the retrieved The number of K images with the most similar fracture type to the DRF image to be queried and the fracture type is type B, C is the number of images with fracture type C in the retrieved K DRF images, and the total number of the three is K. That is, \( S_A, S_B, S_C \), the sum of the weights of the fracture image type is compared with the sum of the weights of all the queried images. Generally, we set the \( S_A, S_B, S_C \) threshold (such as \( S_f = 0.5 \)), calculate the value \( S_A, S_B, S_C \) separately, and compare it with the \( S_f \) value respectively. If \( S_A \geq S_f \) the query image is considered to be a type A fracture of the distal radius, if \( S_B \geq S_f \) it is considered to be a type B fracture of the distal radius, \( S_C \geq S_f \) the query image is considered to be a type C fracture of the distal radius.

To avoid the data deviation caused by the imbalance of different types of eigenvalues, the experiment normalized the extracted eigenvectors. In this experiment, to evaluate the classification performance of the proposed DRF diagnosis, we use the classification accuracy as an evaluation index. It refers to the ratio of the number of correctly classified samples to the total number of classified samples. The higher the accuracy, the better the classification performance. As mentioned above, the experimental data set contains a total of 1189 DRF images. We randomly select 70% of the image data for training and the remaining 30% for testing. The experiment is repeated 100 times, and the average classification accuracy and F1 of 100 times are output.

5. Experimental results and analysis

5.1. Parameter setting

All experiments in this article are run on Windows 10, MATLAB R2019a, Inter(R) Core(TM) i9-9820X CPU and 128GB RAM. The training times of the CNN part in the feature fusion model is set to
1,000, the batch size is 100, the learning rate is set to 0.001, the attenuation rate is 0.1, and the optimization algorithm selects Adam. All parameters of the AlexNet network are consistent with the parameters of deep feature extraction in this paper.

In order to verify the effectiveness of the DRF diagnostic classification model proposed in this paper, the DRF classification algorithm proposed in this paper is compared with the experimental results of KNN, SVM, ELM, RF, etc., which are effective classification models in current classic classifiers. In order to achieve the best classification results for the above model, the parameters of the above diagnostic model need to be set. Among them, the main parameters of SVM include penalty factor c, kernel function, and kernel parameter g. The experiment uses a Gaussian radial basis function with a radius of 2.0 as the kernel function, and then performs five-fold cross-validation through the grid search method to obtain the optimal penalty factor c is 8.025 and the nuclear parameter is 0.2176. The optimal k value in the KNN classifier is 7, and the number of nodes in the hidden layer of ELM is 350. RF parameter selection, select the 400th tree for the number of decision trees.

5.2. Evaluation criteria

At present, for medical image data sets, there are mainly two evaluation methods for evaluating the classification performance of diagnostic models, which are divided into image-based and patient-based aspects. Since the clinical DRF data collected by the research group is based on the image level, this paper evaluates the performance of the algorithm proposed in this paper based on the image level. The total number of images is expressed in terms of the number of correct classifications. The classification accuracy based on the image level is:

\[ Acc = \frac{M_i}{M_{all}} \]  

\[ \text{Precision} = \frac{TP}{TP + FP} \]  

\[ \text{Recall} = \frac{TP}{TP + FN} \]  

Among them, TP means that the classification result was originally positive and also positive. TN means that the classification result is negative and negative. FP means that it was originally negative and classified as positive. FN indicates that it was originally positive and classified as negative.

In multi-classification, its accuracy rate is not the only indicator to judge the quality of the diagnostic model. It is to comprehensively weigh precision and recall. Therefore, this paper introduces F-score as another indicator of the model evaluation standard, and F-score as the formula is as follows:

\[ F - score = \frac{(1 + \beta^2) \times P \times R}{\beta^2 \times P + R} \]  

\[ F_1 = \frac{2 \times P \times R}{P + R} \]

5.3. Comparison of classification performance of different feature extraction methods

Feature extraction is a prerequisite for the correct classification and prediction of subsequent images. Practice has proved that accurate and efficient feature representation can not only improve the accuracy of subsequent classification and prediction results, but also reduce running time and improve model diagnosis efficiency. Table 1 lists the experimental results of applying different feature extraction methods to the DRF image data set and using SVM classification. From the comparison of the classification accuracy and F1 value of different feature extraction methods using DRF image data sets, it can be concluded that the feature fusion method proposed in this paper achieved a classification accuracy of 77.2%, and the F1 value was 0.715. Compared with other feature representation methods, the feature fusion method proposed in this paper improves the classification accuracy by 6%.
Table 1. Comparison of classification performance of different feature extraction methods

| Feature representation method       | Acc(%)  | F1     |
|------------------------------------|---------|--------|
| Gabor                              | 61.8±1.3| 0.607  |
| LBP                                | 58.7±2.7| 0.551  |
| CNN                                | 68.2±3.1| 0.671  |
| Features (series connection)       | 71.1±2.5| 0.709  |
| Proposed method                    | 77.2±2.2| 0.715  |

5.4. Classification performance evaluation of DML-K algorithm

In order to evaluate the classification performance of our proposed DML-K algorithm, the DML-K algorithm in this article and other classic classifiers such as SVM, KNN, ELM, RF, etc. have been proven to be applied in other classification fields (such as brain, lung cancer, breast cancer) effective classification algorithms for comparison. From the comparison of classification performance listed in Table 2, it can be seen that for the same kind of feature vector representing DRF image (using the method of this paper), the classification accuracy rate of the DML-K algorithm proposed in this paper has reached 83.4%, and the classification performance should be significant. The classification effect is better than other classic classifiers. And from the comparison of the classification performance of different classifiers, it can be concluded that the classification performance of the traditional machine learning classification method and the same feature representation method show a stable and close classification accuracy rate. The multi-classification performance of the four classic classifiers gaps are all within 4%. But they are all lower than the DML-K classification algorithm proposed in this paper, which proves that the classification algorithm proposed by us is feasible.

Table 2. Classification performance evaluation of different classification algorithms

| Classifier | Acc(%)  | F1     |
|------------|---------|--------|
| SVM        | 77.2±1.8| 0.715  |
| KNN        | 74.9±2.2| 0.769  |
| ELM        | 70.4±3.1| 0.679  |
| RF         | 68.1±3.5| 0.647  |
| DML-K      | 83.4±1.7| 0.815  |

5.5. DML-K parameter selection

Figure 5 shows the impact of the selection of K in the DML-K algorithm on the classification performance of the scheme. Because the DML-K algorithm classifies the three types of fractures in DRF images, the selection of parameters avoids multiples of 3. This article sets the value range of...
parameter K to [4, 6, 8, 10, 12, 14, 16, 18, 20]. It is concluded from Figure 6 that when K=12, the corresponding Acc of the scheme is the highest, reaching 0.83.

5.6. Specific search examples

![Image of specific search examples](image)

Figure 6. Shows a specific search example

6. Discussion and conclusion

This paper proposes and demonstrates a multi-feature DML-K diagnosis scheme for DRF images. The experiment first uses CLAHE to preprocess the DRF image, and then extracts DRF image features from both traditional features and depth features. After the feature vector is reduced in dimension, a two-layer neural network is designed to perform correlation fusion of the extracted features. Multi-feature fusion is used to accurately represent DRF images. In terms of classification, the image retrieval algorithm (DML-K) is used to construct a diagnostic model to predict the specific types of fractures in DRF images. By measuring the distance between the image to be queried and the DRF database images, the first K images with the smallest distance are selected and given a certain weight. Finally, calculate the probability that the first K images are the same label as the image to be queried, so as to realize the specific classification of the DRF image. The diagnosis model proposed in this paper has the following advantages. First, we propose a DML-K image retrieval classification method that can not only help orthopedic surgeons evaluate the specific types of DRF images, but also provide the retrieval set with diagnosis reports for doctors' reference. Second, use a variety of feature fusion to represent DRF images. Traditional features represent the shape and texture features of DRF images. The depth features represent the high-level semantic features of the DRF image, and they complement each other. Third, the algorithm proposed in this paper avoids the problem of dimensionality catastrophe and overfitting by combining different types of features. In addition to the advantages of the algorithm discussed above, this work also has some limitations. First of all, the DRF database only has 1189 images, and a larger DRF image database should be collected later. Secondly, this study only selects two traditional features and depth features for fusion, but many other types of features can be studied to represent DRF images. Therefore, in the follow-up work, finding a more effective feature representation method and performing accurate and effective feature fusion is still the focus of follow-up research. Third, in the experiment, we only use the search algorithm to determine the parameters. Therefore, a better optimization method should be selected in the follow-up to determine the optimal parameters.

Based on the computer-aided diagnosis, this paper proposes a diagnosis model based on multi-feature fusion and DML-K to predict specific types of DRF images. In the experiment, the methods proposed in this paper are analyzed and compared from many aspects, which provides a fast, low-cost, and repeatable diagnostic method for fracture classification. The scheme has certain feasibility and good
robustness. In the future related work of fracture diagnosis research, it is hoped that the characteristics of distal radius images can be further expanded and the accuracy of DRF image prediction can be further improved from the perspective of optimizing the diagnosis model algorithm.

Acknowledgments
The corresponding author of this article is Professor Ding Bo.
This work was funded by the Shandong Province Social Science Planning Research Project "Research on the Path and Method of Chinese Medicine Culture Communication in the New Era" (19CSHJ13). Industry-University Cooperation Collaborative Education Project of the Higher Education Department of the Ministry of Education (201902205004).

References
[1] Song M, Zeng R, Ding Y, et al. (2019) Recent advances in biodegradation controls over Mg alloys for bone fracture management: A review. Journal of Materials Science & Technology, 35: 535–544.
[2] Yap D, Chen Y, Leow W K, et al. (2004) Detecting femur fractures by texture analysis of trabeculae. International conference on pattern recognition, 34: 730–733.
[3] Gan K, Xu D, Lin Y, et al. (2019) Artificial intelligence detection of distal radius fractures: a comparison between the convolutional neural network and professional assessments. Acta Orthopaedica, 90: 394–400.
[4] Froz B, Filho A, Silva A, et al. (2017) Lung nodule classification using artificial crawlers, directional texture and support vector machine. Expert Syst Appl, 69: 176–88.
[5] Huang P, Park S, Yan R, et al. (2017) Added value of computer-aided CT image features for early lung cancer diagnosis with small pulmonary nodules: a matched case-control study. Radiology, 286: 286–95.
[6] Shen W, Zhou M, Yang F, et al. (2017) Multi-crop convolutional neural networks for lung nodule malignancy suspicion classification. Pattern Recognition, 61: 663–673.
[7] Chung S W, Han S S, Lee J W, et al. (2018) Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. Acta Orthopaedica, 89: 468–473.
[8] Bodla N, Hua G, Chellappa R, et al. (2018) Semi-supervised Fused GAN for Conditional Image Generation. Computer Vision and Pattern Recognition, 63: 1278–1286.
[9] Khusnuliaiwati H, Faticah C, Soelaiman R. (2017) Multi-feature fusion using SIFT and LEBP for finger vein recognition. TELKOMNIKA, 15: 478–485.
[10] Bai X, Liu C, Ren P, et al. (2017) Object classification via feature fusion based marginalized kernels. IEEE Geoscience & Remote Sensing Letters, 12: 8–12.
[11] Shi B, Chen Y, Zhang P, et al. (2015) Nonlinear feature transformation and deep fusion for alzheimer’s disease staging analysis//International Workshop on Machine Learning in Medical Imaging, 13: 304–312.
[12] Tomita N, Cheung Y Y, Hassanpour S, et al. (2018) Deep neural networks for automatic detection of osteoporotic vertebral fractures on CT scans. Computers in Biology and Medicine, 12: 8–15.
[13] Pan S J, Yang Q. (2010) A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22: 1345–1359.
[14] Ojala T, Pietikäinen M, Mäenpää T. (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis & Machine Intelligence, 24: 971–987.
[15] Gabor D. (1946) Theory of communication. IEEE Proc London, 93: 56–58.
[16] Zhou FY, Jin LP, Dong J. (2017) Overview of Convolutional Neural Network Research. Chinese Journal of Computers, 40: 1229–1251.
[17] Szegedy C, Liu W, Jia Y, et al. (2015) Going deeper with convolutions// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE. pp. 1–9.
[18] Zhang Y N, Zhao J J, Zhao X, et al. (2019) Long-term benign and malignant prediction method of lung lesions under multimodal fusion. Computer Engineering and Applications, 55: 146–153.

[19] Wei G H, Qiu M, Zhang K X, et al. (2020) A multi-feature image retrieval scheme for pulmonary nodule diagnosis. Medicine, 99: 236–245.

[20] Wei G H, Qi S L, Qian W, et al. (2018) Lung nodule image retrieval algorithm based on similarity measurement. Journal of Northeastern University (Natural Science Edition), 39: 1226–1231.