Computer-Assisted Bone Fractures Detection Based on Depth Feature

Feng Yang*, Guohui Wei*, Hui Cao*, Mengmeng Xing*, Shuxuan Liu* and Jing Liu*

College of Science and Technology Shandong University of Traditional Chinese Medicine, China

*Corresponding author e-mail: caohui63@163.com, ayangfeng163@163.com, bmie530@163.com, 13256688200@163.com, xgglsx@163.com, 13954125302@163.com

Abstract. Nowadays, computer aided diagnosis (CAD) system become popular because it improves the diagnostic efficiency of the disease compared to the early diagnosis of the various diseases for the doctors and the medical expert specialists. Similarly, bone fractures is a common problem due to pressure, accident and osteoporosis. Bone fractures detection using computer vision is getting more and more important in CAD system because it can help to reduce workload of the doctor by screening out the easy case. The purpose of this work was to detect whether fractures occurred in seven different parts of the human body in the MURA database. The process of a bone fractures detection is mainly divided into three steps, which includes preprocessing, feature extraction and classification. In preprocessing, the 40561 sizes of images in the MURA data set were reshaped to same size and the grayscale images were changed to the RGB images. The convolutional neural network (CNN) was then used to extract the depth features of the bone images. Finally, three classic classifiers were selected, such as support vector machine (SVM), extreme learning machine (ELM) and random forest (RF) to detect bone cracks or non-cracks. Then compare it with the Alexnet model and the classification results of the Wu Enda team. The result shows that the classification of SVM based depth features is the best and the accuracy can reach 78.63%. The experimental result indicates that the depth features of the extracted bone images provide a new feature representation method for the prediction of bone fractures, which greatly improves the accuracy of clinical diagnosis of bone fractures.

1. Introduction

The bones form the body's scaffolding structure, providing attachment points for the muscles. With the development of the times, people are more and more likely to have bone fractures. The occurrence of fractures can not only lead to the interruption of force transmission and bone blood supply disorders, but also cause composite damage of bones and surrounding soft tissues [1]. If the diagnosis and treatment of the fractures is not timely, it will also cause the occurrence of fractures complications. The occurrence of bone fractures has caused great damage to the human body, so it is significance for human to accurate and effective diagnosis of bone fractures.
Moreover, the current social environment has extremely high requirements for the correct diagnosis of bone fractures and the current diagnosis of bone fractures is mainly for the doctor's manual reading. Reading a large number of X-ray films not only causes the reader to suffer from reading fatigue, but also determines the deviation. At the same time, due to different doctors' experience and different interpretation standards, it is easy to misdiagnosis in the diagnosis process. By using computer-aided diagnosis technology (CAD) to achieve automatic classification of bone fractures images, before the doctors read the X-ray images, the automatic classification of the fractures, and then the classified images to the physician for diagnosis, not only can reduce the number of doctors to read and overcome people eye inertness and defects that are insensitive to grayscale can also improve the diagnostic efficiency and accuracy of radiologists [2].

The accuracy of bone fractures diagnosis is mainly depending on the quality of the bone images feature extraction and the effectiveness of the classification algorithm. In particular, the quality of bone images feature extraction will ultimately directly affect the diagnosis of the final bone fractures. The researchers designed several special feature extraction schemes for bone sites, such as Martin, D et al [3] had developed a method of automatically detecting bone fractures. Firstly, bone edge were extracted from X-ray images using a non-linear anisotropic diffusion method. Secondly, modified Hough transform with automatic peak detection and magnitude and direction of the gradient were created using the calculate line parameter. This method consistently detected mid-shaft long bone fractures. Mahendran S. K et al [4] concentrated automatic fractures detection using fusion-based classifiers. Contract, Homogeneity, Energy, Entropy, Mean, Variance, Markov Random Field (MRF) and intensity gradient direction (IGD) were extracted from fractures X-ray images as features. Using these features, train the classifier and test the model for detecting fractures in X-ray images. Zheng et al [5] presented fractures classification and feature extraction of X-ray fractures image. They firstly discussed marker-controlled watershed transform based on gradient and homotopy modification to segment X-ray fractures images. Then, marker processing and regionprops function were used to extract region number, region area, region centroid and protuberant polygon of fractures image and Hough transform was be applied to detect and extract lies in the protuberant polygon of X-ray fractures image. Finally compute the angle between fractures line and perpendicular line of centerline. The above method of feature extraction is mainly the traditional feature extraction of bone images. Although the researchers used traditional features to obtain good fractures diagnosis accuracy, since the traditional features are designed for a specific site, these features is only suitable for fractures diagnosis at a specific site and cannot be used for fractures in all parts of the bone. Therefore, traditional features for fractures diagnosis have great limitations.

Current researchers have successfully applied depth features to the diagnosis of breast cancer and lung cancer [6]. Previously limited by the open bone fractures images data set, most researchers used traditional feature extraction methods and rarely used depth feature extraction methods in the process of feature extraction. With the publication of the 2018 MURA data set, an opportunity for deep learning in the diagnosis of fractures was provided. This paper proposed the application of deep learning to feature extraction of bone fractures images. This method used pre-trained CNN to extract depth features from different layers of the network to see if these features are sufficient to compete with visual feature descriptors. This method has been proven to be a very good general-purpose image feature extraction that provides competitive results in a variety of tasks.

2. Material and method

2.1. Bone fractures dataset

MURA is currently the largest open source medical radiographic data set. The article used the MURA data set, which was a data set of musculoskeletal radiographs containing a total of 14,863 studies of 12,173 patients with a total of 40,561 multi-view radiographs. Each is one of seven standard types of upper extremity radiology studies, namely fingers, elbows, forearms, hands, humerus, shoulders and wrists. Each study was manually labeled as normal or abnormal by a radiologist certified by the Stanford
Medical School Board of Directors. These annotations were performed from 2001 to 2012 based on interpretation of clinical medical imaging by radiodiagnostics [7]. Figure 1 shows fractures images from seven different sites in the MURA data set.

![Fractures images](image1)

**Figure 1.** Fractures image samples from different regions of the MURA dataset: (a) shoulder (b) forearm (c) finger (d) humerus (e) elbow (f) hand (g) wrist

### 2.2. Feature extraction

The idea of deep features consists of extracting features from images and using them as input for a classifier. The main difference between depth features and the current standard of using CNNs [8], is that a previously-trained CNN is simply reused as feature extractor, the output of which is fed into another classifier, trained on problem-specific data.

Before the feature extraction of the bone fractures images, the images are first preprocessed and the gray images are converted into RGB images and uniform sizes processing are 227×227×3. This paper selected the commonly used deep convolutional neural network model Alexnet as the infrastructure of this model. Despite the relatively large number of layers in the Alexnet model, in this work we focus only on extracting features from the three top-most layers, i.e., fc6, fc7, and fc8, which supposedly present the three most high-level features. These layers are composed of 4,096, 4,096, and 1,000 dimensions, respectively. Figure 2 shows the flow chart of the Alexnet feature extraction model.

![Alexnet flow chart](image2)

**Figure 2.** An illustration of the AlexNet model and at the bottom right, the reference names for the top layers are listed.

In order to facilitate the understanding of the process of depth feature extraction, Figure 3 shows an example of depth feature extraction for an X-ray image of a hand fractures.
2.3. Feature selection

Given the high-dimensionality of those vectors, we considered using the principal component analysis method to reduce the dimension of the feature data. Principal component analysis is a technique for analyzing and simplifying data set, usually used to reduce the dimensionality of a data set while maintaining the characteristics of the largest difference in the data set [9]. For a sample feature $X_{mvn}$ of a given space, the matrix $X_{mvn}$ is subjected to dimensionality reduction analysis using PCA. The specific steps are as follows:

Step 1, Zero-average each row of $X_{mvn}$ (representing each attribute field), that is, subtract the mean of the row;

Step 2, calculating a covariance matrix of $X_{mvn}$;

Step 3, calculating eigenvalues and eigenvectors of the covariance matrix;

Step 4, the feature vectors are arranged into a matrix according to the size of the corresponding feature value from top to bottom, and the first k rows are formed into a matrix, that is, the data after dimension reduction to k-dimensional.

The values of the specific dimensions taken are shown in Table 1. Experiments show that the following dimensions can be selected to retain 95% of its characteristics.

| Layer | Elbow | Finger | Forearm | Hand | Humerus | Shoulder | Wrist |
|-------|-------|--------|---------|------|---------|----------|-------|
| fc6   | 209   | 189    | 231     | 217  | 289     | 314      | 231   |
| fc7   | 232   | 237    | 203     | 228  | 193     | 212      | 241   |
| fc8   | 118   | 103    | 167     | 124  | 148     | 105      | 98    |

2.4. Diagnostic model construction

Pathological images of different parts of the MURA data set were randomly divided into training sets (70%) and test sets (30%). Then used the Alexnet depth feature of the fractures image [10] as the training set, use the classifier to establish different diagnostic models, and verify on the test set. The classifiers used in the experiment mainly have Support Vector Machine (SVM), Extreme Learning Machine (ELM), Random Forest (RF). RF is an algorithm that integrates multiple trees through the idea of integrated learning [11]. The basic unit is the decision tree, which is essentially a branch of machine learning-integrated learning method. ELM is an algorithm for solving the single hidden layer neural network proposed by Huang Guangbin [12]. Its biggest feature is that for traditional neural networks,
especially the single hidden layer feed forward neural network, the method is faster than the traditional learning algorithm under the premise of ensuring learning accuracy. The SVM uses the kernel function to classify the input features from the low-dimensional space to the higher-dimensional space according to the mapping space measure [13]. The construction flow chart of the diagnostic model is shown in Figure 4.

Figure 4. Diagnostic model construction flow chart

In order to improve the classification accuracy, the parameters of each typeifier are optimized by taking the hand fractures image as an example. SVM parameter settings: The key of SVM lies in the penalty factor c and the kernel function and the kernel parameter g. The experiment used a Gaussian radial basis function with a radius of 2.0 as the kernel function, and then uses the grid search method to perform 10-fold cross-validation to obtain the optimal penalty factor c and kernel Parameter g. ELM parameter settings: The number of parameter nodes in the hidden layer was obtained by experimenting with 100~1000 nodes at intervals of 50 nodes. RF parameter settings: Decision tree setting is obtained by taking 50~1000 trees and 50 trees as experimental intervals.

The following is a simple list of some parameter optimization charts. Table 2 shows the optimal c and g selected in the SVM, figure 5 shows the impact of hidden layer nodes on performance in ELM, and figure 6 shows the effect of trees in decision trees in random forests on performance.

Table 2. Optimal c and g

| Different parts | Optimal c | Optimal g |
|----------------|-----------|-----------|
| Elbow          | 8.0000    | 0.1250    |
| Finger         | 4.9157    | 0.6825    |
| Forearm        | 2.2171    | 0.0470    |
| Hand           | 6.2299    | 0.0825    |
| Humerus        | 3.6785    | 0.0268    |
| Shoulder       | 2.8594    | 0.0710    |
| Wrist          | 8.1760    | 0.1895    |

Figure 5. Impact of hidden layer nodes on performance in ELM
Figure 6. The effect of trees in decision trees in random forests on performance

In this paper, the classification accuracy was taken as the evaluation metric of the parameter setting. The calculation formula is as follows:

\[
    ACC = \frac{\sum_{i=1}^{n} \delta(x_i, y_i)}{n}
\]

For the nth bone fractures image, its gold standard label is \(x_i\), \(y_i\) is the classification label of the classification algorithm, and \(\delta(p, q)\) is a function that satisfies \(\delta(p, q) = 1\), if \(p = q, \delta(p, q) = 0\), otherwise.

The experimental results were obtained after 100 experiments.

3. Results and discussion

In this paper, we applied the depth feature extraction method to the diagnosis of bone fractures images. First, the 4096-dimensional, 4096-dimensional and 1000-dimensional features of the fc6, fc7 and fc8 layers are extracted to represent the bone fractures images. Then these features are selected to find the most effective features and reduce the sizes from among many features. The PCA algorithm is selected for feature selection. Then a feature that can represent more than 95% of the original data is obtained. Finally, several classical classification models were evaluated and the approximate range of the results was calculated by statistical methods of mean accuracy and standard deviation. The results of the experimental methods used in this paper are shown in the table 3.

| Classifier | Number of MURA for different parts of the human body | Overall |
|------------|------------------------------------------------------|---------|
|            | Elbow | Finger | Forearm | Hand | Humerus | Shoulder | Wrist |         |
| fc6        | 71.22±1.24 | 69.97±1.83 | 73.62±2.09 | 69.09±1.56 | 72.63±2.60 | 66.33±1.33 | 66.36±2.18 | 69.89 |
| Alexnet fc7 | 72.39±1.98 | 70.57±1.78 | 75.46±1.54 | 71.40±1.38 | 73.34±1.56 | 68.31±1.10 | 67.69±2.12 | 71.92 |
| fc8        | 72.00±2.19 | 70.51±2.56 | 75.48±1.36 | 72.62±2.29 | 71.92±1.65 | 67.55±2.78 | 66.71±2.54 | 70.97 |
| fc6        | 76.89±1.45 | 82.33±1.24 | 79.56±2.43 | 77.29±3.01 | 76.98±2.56 | 74.98±3.44 | 78.05 |
| SVM fc7    | 77.87±1.56 | 75.23±1.23 | 81.69±1.89 | 80.20±2.21 | 78.45±2.77 | 74.89±3.20 | 75.87±2.56 | 77.74 |
| fc8        | 77.16±2.10 | 76.12±2.78 | 81.35±2.56 | 80.12±3.28 | 78.39±2.53 | 74.55±2.79 | 75.13±1.99 | 77.55 |
| fc6        | 68.78±3.23 | 70.28±1.94 | 76.23±2.78 | 73.66±2.09 | 70.11±2.44 | 69.13±1.76 | 67.24±1.88 | 70.78 |
| ELM fc7    | 69.92±2.10 | 70.23±1.90 | 76.75±1.67 | 74.79±1.69 | 69.39±2.08 | 69.79±3.02 | 66.58±0.27 | 70.81 |
| fc8        | 71.98±1.56 | 78.45±1.89 | 73.21±2.03 | 69.56±1.90 | 69.99±3.56 | 66.34±2.23 | 71.00 |
| fc6        | 72.10±2.12 | 72.33±1.78 | 76.09±1.53 | 75.23±1.57 | 73.67±1.75 | 72.87±1.90 | 76.91±3.02 | 74.17 |
| RF fc7     | 71.76±3.01 | 72.45±1.89 | 76.67±1.88 | 75.47±1.29 | 73.31±2.49 | 72.98±1.68 | 77.98±3.29 | 74.37 |
| fc8        | 71.81±2.67 | 72.21±2.67 | 75.87±3.06 | 75.67±1.04 | 73.28±3.27 | 72.91±1.45 | 75.56±1.67 | 73.90 |
By comparing the classification results of 7 different data set in Table 3, it could be seen that the characteristics of the fc7 layer were generally higher for the classification of the subsequent classifier than the other two layers. By comparing the four classification models, it was found that Alexnet extracted the features with the worst classification effect and the accuracy of fc6, fc7 and fc8 layers were 69.89%, 71.92% and 70.97%, respectively. However, if the Alexnet model was only used to extract features and then sent to several other classic classifier, the accuracy of the classification had been significantly improved. The classifier with the best classification effect is SVM. The accuracy of fc6, fc7, fc8 layer classification is 78.05%, 77.74%, 77.55%. Followed by RF, and the accuracy of fc6, fc7, fc8 layer classification is respectively 74.17%, 74.37% and 73.90%. The ELM is inferior to the above two classifiers and the accuracy of the fc6, fc7and fc8 features is 70.78%, 70.81% and 71.00% respectively.

The above-mentioned Alexnet classification model classifies the human body and various parts mainly by proposing its fc6, fc7 and fc8 layer features, then classifying according to the selected layer number features, and finally outputting the classification result. This paper also explored Alexnet's other classification methods, that is, not extracting its depth features but directly outputting its classification results, which expressed it below with Alexnet*. In order to further evaluate the performance of the model, the literature with the same evaluation criteria in the same data set MURA was selected, and the optimal results in the corresponding literature were compared. The experimental results are shown in Table 4.

| Method     | elbow | finger | forearm | hand | humerus | shoulder | wrist | Overall |
|------------|-------|--------|---------|------|---------|----------|-------|---------|
| Alexnet*   | 76.88%| 72.57% | 78.97%  | 77.26%| 77.15%  | 71.99%   | 72.50%| 75.33%  |
| Wu         | 71.00%| 38.90% | 73.70%  | 85.10%| 60.00%  | 72.90%   | 93.10%| 70.67%  |
| This work  | 78.29%| 76.12% | 82.33%  | 80.20%| 78.45%  | 76.98%   | 78.05%| 78.63%  |

The wu in Table 4 was the accuracy of the Wu Enda team used a 169-layer convolutional neural network to predict bone fractures in 2018. By comparing the three methods in table 4, the Alexnet* model was superior to Wu et al. in predicting the performance of fractures and the accuracy was lower than that of wu only in the data set of hand and wrist. Compared with the proposed method, the average of the prediction accuracy was higher than the Alexnet* model and the wu method. However, the accuracy rate on the hand data set is slightly lower than that of the wu method.

The experimental result in Table 4 shows that the proposed method is 7% to 8% higher than the method used by wu, indicating that the classification effect of this model is good and has certain feasibility.

This study is mainly used for the diagnosis of bone fractures images and can also be used for benign and malignant diagnosis of tumors and fractures classification of a specific part of the bone. However, this program also has certain limitations. For example, many images of the shoulders of the data set and the parts of the humerus are not clear due to the difference between the hospital and the shooting conditions. Therefore, when Alexnet is used to extract the depth features, many unclear images are used. In fact, it interferes with the extraction of features, resulting in low classification accuracy of subsequent classifiers. Later, the accuracy of classification can be improved by combining traditional features and depth features.

### 4. Conclusion

On the basis of CAD, this study extracted the depth features of the bone images where came form seven different parts of the MURA database and used PCA to select features that contribute more than 95% to the data set to reduce its dimensions. The dimensionality-reduced depth features are then entered into different classification models. The experimental results show that the classification performance of SVM
is the best and the average accuracy of classification is 78.63%. This provides a new idea for the feature extraction step of clinical diagnosis of bone fractures, and proves that the experimental model used in this paper has good feasibility and robustness by comparing with the accuracy and standard deviation of previous studies. The model can also be applied to the classification of crack images in the future, in particular to determine the location of cracks and different types of cracks.

Acknowledgments
The corresponding author of this article is Professor Hui Cao. This work was funded by the National Natural Science Foundation (No. 81973981, No. 81473708).

References
[1] Esteva, Andre, Kuprel, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542 (7639): 115–118, 2017.
[2] Torlay, Perrone-bertolotti M, Thomas E, et al. Machine learning-XGBoost analysis of language networks to classify patients with epilepsy [J]. Brain Inform, 2017, 4 (3): 159-169.
[3] M. AL-AYYOUB, D. AL-ZGHOOL. Determining the type of long bone fractures in X-ray images, Wssea Transacations on information science and applications, vol. 10, Aug. 2013, 261-270.
[4] A. T. C, Mallikarjunaswamy M. S. and Rajesh Raman. Detection of Bone fractures using Image Processing Methods, International Journal of Computer Application, National Conference on Power & Industrial Automation NCPSIA, Aug. 2015, pp. 6-9.
[5] S. Myint, A. S. Khaing and H. M. Tun, Detecting Leg Bone fractures in X-ray Images, International Journal of Scientific & Research, vol. 5, Jun. 2016, pp. 140-144.
[6] V. D. Vegi and S. L. Patibandla, S. SKavikondala and CMAK Z. Basha, Computerized fractures Detection System using x ray Images, International Journal of Control Theory and Applications, vol. 9, Nov. 2016, pp. 615-621.
[7] Rajpurkar, P. MURA dataset: towards radiologist-level abnormality detection in musculoskeletal radiographs. Preprint at https://arxiv.org/abs/1712.06957 (2017).
[8] N. Bayramoglu, J. Kannala, and J. Heikkila. Deep learning for magnification independent breast cancer histopathology image classification, in 23rd International Conference on Pattern Recognition, vol. 1, December 2016.
[9] Gale W, Oakden-Rayner L, Carneiro G, Bradley AP, Palmer LJ. Detecting hip fractures with radiologist-level performance using deep neural networks. arXiv website. archiv.org/abs/1711.06504. November 17, 2017.
[10] Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine learning for medical imaging. RadioGraphics 2017; 37: 505–515
[11] Breiman L. Random forests, machine learning 45 [J]. Journal of Clinical Microbiology, 2001, 2: 199-228
[12] Huang Guangbin, Zhou Qinyu, Siew CK.Extreme learning machine: Theory and application [J]. Neurocomputering, 2006, 70 (1): 489-501
[13] Cortes C, Vapnik V. Support-vector network [J]. Machine Learning, 1995, 20 (3): 273-297