

Osteoporosis Recognition Based on Similarity Metric with SVM

Ke Zhou¹, Jie Cai^{1,#}, Yong-Hui Xu², Tian-Xiu Wu^{3,*}

¹School of Information Engineering
Guangdong Medical University
Zhanjiang 524023, China
E-mails: kitty@gdmc.edu.cn, gdmcli@qq.com, caijie513@163.com

²School of Computer Science and Engineering
South China University of Technology
Guangzhou 510006, China
E-mail: 840411554@qq.com

³School of Basic Medical Science
Guangdong Medical University
Zhanjiang 524023, China
E-mail: wutianxiu2005@163.com

#Author contributed equally to the paper

*Corresponding author

Received: January 11, 2016

Accepted: June 1, 2016

Published: June 30, 2016

Abstract: *The purpose:* Applying different techniques of classification to osteoporotic bone tissue texture analysis, exploring the recognition rate of the different classification methods.

Methods: Using gray-level co-occurrence matrix (GLCM) and running a length matrix texture analysis to extract bone tissue slice image characteristic parameters, and to classify respectively 4× and 10× microscope images of the two groups: the sham (SHAM) and the ovariectomized (OVX) group image.

Results: The metric support vector machine (SVM) classification algorithm, based on SVM learning or recognition rate, was higher than the stand-alone measure, and the classification results were stable.

Conclusion: Measurement of the SVM classification algorithm for osteoporotic bone slices texture analysis revealed a high recognition rate.

Keywords: Osteoporosis, Texture analysis, Information-theoretic metric learning (ITML), Support vector machine (SVM).

Introduction

In 2013, there were more than two hundred million members of the population over 60 years old in China. The study showed that the morbidity rate of osteoporosis among males and females aged 60-69 was 33% and 73.8% respectively, and among 70-79 year olds it was 55.6% and 89.7% respectively. The osteoporosis fracture has become a significant social and health problem with definite pathological, physiological and economic consequences. It is therefore very urgent to carry out early warning testing and diagnosis of osteoporosis, with related effective interventions and treatments as well.

The World Health Organization (WHO) defines osteoporosis as “the system for metabolic bone diseases”, with characteristics of “low bone mass” and “degeneration of bone microarchitecture”, which leads to increasing osteoporosis [12]. This definition emphasizes the bone mass, paying more attention to the quality of the bones, such as the microarchitecture of the bones. The definition of bone microarchitecture is the connection degree between the three-dimensional construction of bone trabecular and the trabecular. The bone trabecular are highly complex, anisotropic materials, which could bear the different sized tensile and compressive stresses [13]. Studying the anisotropy of the bone trabecular is the key to the accuracy of the biomechanical analysis [6, 11], which reflects the problems of the consistency of bone trabecular. Similarly, compared with the bone mineral density (BMD), which would be the determinant of the osteoporosis [1, 8], the changes of trabecular microarchitecture are more sensitive. The traditional bone histomorphometry is one of the gold standards in quantitative analysis of bone microarchitecture [9]. On account of the principles of stereology, this method utilizes the two-dimensional image to calculate the related parameters of bone microarchitecture, which would observe and measure scientifically, the morphological changes of the quality level of bone tissue (bone structure) and the quantity (bone mass) [3]. Due to the drifting structural pattern of bones [15], traditional bone histomorphometry cannot determine the anisotropic structure, because it is difficult to reflect the stereo-chemical structure of the bone trabecular. However, the morphometry of the two-dimensional biopsy could obtain the kinetic parameters of the bone tissues, which would provide the dynamic information of the bone histology regarding the bone remodeling and construction. All of these parameters could not be achieved by the morphometry of a three-dimensional biopsy unless an invasive computed tomographic (CT) microscope were used. The use of the micro CT and magnetic resonance imaging (MRI) have been limited by the various imaging conditions, using the standard contradictions and the expensive cost. These machines have determined that bone histomorphometry is an irreplaceable field of research with many clinical applications. On that account, based on bone histomorphometry, exploring a more convenient and objective technological means that can better evaluate bone trabecular anisotropic characteristics is the most important area of study in the osteoporosis field.

The image processing techniques were used to extract the texture characteristic parameters and the quantitative or qualitative description. From this, the texture analysis could analyze the image in the region of the image (many pixels). Image texture analysis has been widely used in medical imaging, especially for the image processing of MRIs of the brain [7, 18]. Also, some scholars have incorporated texture analysis into the imaging diagnosis of osteoporosis [10, 16], but do not involve it in the field of bone histomorphometry. According to the previous research [17], texture analysis could distinguish between rats with osteoporosis, without the ovary and bone tissue, and normal rats, which demonstrates it as an effective analysis technique. At the same time, under the four-time mirror, regions of interest (ROI) were extracted (Fig. 1), and then the texture features at the ROI area of 0° , 45° , 90° , 135° were extracted by choosing the co-occurrence and run-length matrix methods. The results in [2] show that there are significant differences in the texture parameters between the rats with osteoporosis, without ovary and bone tissue, and the normal rats ($p < 0.05$). Based on previous studies, the research obtained the texture characteristics by using the co-occurrence and run-length matrix, and then classified the texture characteristics according to the different classification techniques to discuss the recognition rate of different algorithms.

Fig. 1 and Fig. 2 show bone tissue slices.

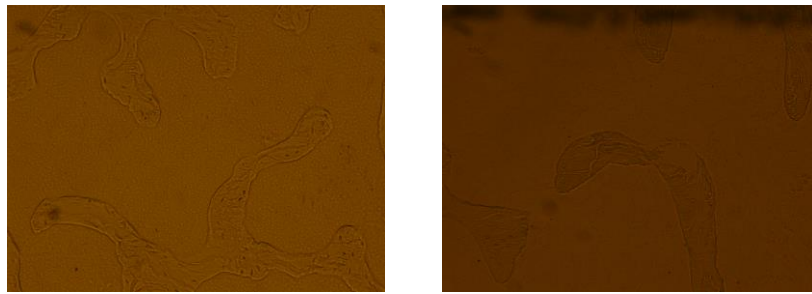


Fig. 1 Normal biopsy samples (left) and OVX slice (right) under 4× magnification

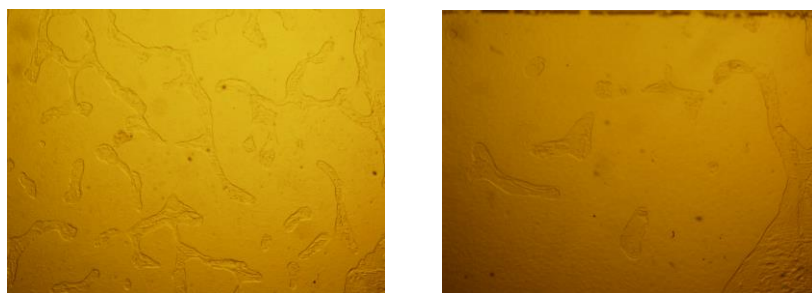


Fig. 2 Normal biopsy samples (left) and OVX slice (right) under 10× magnification

Materials and methods

Animals for the experiment and data sources

The rats used were 3-month-old female SD rats, whose average weight was 250.1 g and the test time was 90 days. The rats were randomly divided into two groups: (1) sham group (SHAM group), (2) ovariectomized (OVX group). The specific grouping of samples is shown in Table 1. At the end of the testing phase, the rats were sacrificed under sterile conditions. Soft tissues were removed and weighed and the right tibiae were removed for bone histomorphometry. A semi-automatic digitizing image analysis system (Osteometrics, Inc. Decatur, GA, USA) was used for quantitative bone histomorphometric measurements. To avoid the primary spongiosa, the proximal tibial metaphysis region of interest was cancellous bone between 1 and 4 mm distal to the growth plate-epiphyseal junction. The regions above were selected as ROI areas for texture analysis. The omitted 1 mm region to the growth plate excluded the primary spongiosa. Bone tissue sections were directly observed and photographed under the microscope at 4× and 10× magnification. Using the GLCM (co-occurrence matrix) and run length matrix (run-length matrix) methods the ROI texture features in the 0°, 45°, 90°, 135° directions were extracted to obtain 15 characteristic parameters [17], such as entropy, contrast, factors and so on. These characteristic parameters were applied to a classification metric SVM to provide a classification algorithm.

Table 1. Data sample case

Group	Microscope 4×	Microscope 10×
SHAM	16	47
OVX	11	40

Classical classification algorithms

Support vector machine

SVM aims to find an optimal hyper plane to minimize global structural risk [4, 14]. This hyper plane can be represented as $w^T x + b = 0$, where:

$$(w^*, b^*) = \arg \min_{(w,b)} \frac{1}{2} \|w\|_2^2 + C \sum_{i=1}^n \xi_i^p \quad (1)$$
$$s.t. \quad \forall i, y_i(w^T x_i + b) \geq 1 - \xi_i, \xi_i \geq 0$$

Currently, within the existing classification algorithms, SVM is the most suitable algorithm for solving binary classification problems, especially in the case of small, nonlinear and highly dimensional samples.

Due to the fact that SVM has a strong generalization capability, it is the most commonly used classification algorithm in many practical applications. This is also the basic motivation for us to choose SVM as the base classification algorithm.

SVMs have been researched and documented thoroughly in the literature. For the classification of small samples, the authors found that tested linear SVM (Linear-SVM) classification to be the best method. Therefore, in this paper, the results provided only Linear-SVM for classification.

Metric learning

Metric learning is an important branch of machine learning. It is used to calculate a metric space which can minimize the distance within the class and maximize the distance between classes.

Given a transformation matrix L , the mahalanobis distance between x_i and x_j can be represented as:

$$S_L(x_i, x_j) = (Lx_i)^T (Lx_j) = x_i^T L^T L x_j \quad (2)$$

The learning transformation matrix L is equivalent to the learning mahalanobis matrix $M = L^T L$.

The most common metric learning algorithm, named information theoretic metric learning (ITML), was proposed by Davis et al. in 2007 [5]. For a given metric A_1 , the distribution of the samples under the metric space generated by A_2 is fixed.

If A_1 is replaced by A_2 , the distribution will change. The KL -divergence between distributions generated by A_1 and A_2 can be represented as:

$$KL(p(x; m, A_1) \| p(x; m, A_2)) = \int p(x; m, A_1) \log \frac{p(x; m, A_1)}{p(x; m, A_2)} dx. \quad (3)$$

The optimal metric space generated by A can be minimized $KL(p(x; m, A) \| p(x; m, A_0))$. When calculated, A is equivalent to the minimum $D_\varphi(A_1, A_2)$, where:

$$D_\varphi(x, y) = \varphi(x) - \varphi(y) - (x - y)^T \nabla \varphi(y). \quad (4)$$

The ITML algorithm is described as follows:

Algorithm: ITML

Input: $X \in R^{d \times n}$, S : Collection of pair points within class, D : Collection of pair points between classes, e : Threshold

Output: W

1. Assume $W_0 = I_d$, and $\lambda_{ij} = 0, \forall i, j$

2. While $|W_t - W_{t-1}| > e$:

 For (x_i, x_j) in S or D

 Compute $w = (x_i - x_j)^T$

 if $(x_i, x_j) \in S$:

$$\gamma = \min \left(\lambda_{ij}, \frac{1}{\|w\|_2^2} - \frac{1}{u} \right)$$

$$\beta = \gamma / (1 - \gamma \|w\|_2^2)$$

 Else if $(x_i, x_j) \in D$

$$\gamma = \min \left(\lambda_{ij}, \frac{1}{l} - \frac{1}{\|w\|_2^2} \right)$$

$$\beta = -\gamma / (1 + \gamma \|w\|_2^2)$$

$$\lambda_{ij} = \lambda_{ij} - \gamma$$

 Compute Cholesky-decomposition: $LL^T = I + \beta w w^T$

$$W_t \leftarrow L^T W_{t-1}$$

3. Return W

Support vector machine based on metric learning

Linear-SVM was proposed in [19]. The basic motivation is to combine metric learning with SVM.

Combining Eq. (1) and Eq. (2), we can obtain the following optimization problem:

$$\begin{aligned}
 (L^*, w^*, b^*) &= \arg \min_{(L, w, b)} \frac{1}{2} R(L)^2 \|w\|_2^2 + C \sum_{i=1}^n \xi_i^2, \\
 \text{s.t. } \forall i, & y_i(w^T L x_i + b) \geq 1 - \xi_i, \xi_i \geq 0
 \end{aligned} \tag{5}$$

where L^* is the optimal metric and (w^*, b^*) is the optimal hyper plane.

This optimization problem can be solved by gradient descent:

$$w_{t+1} = w_t + \alpha \frac{\partial f}{\partial w} \Big|_{w^t}, b_{t+1} = b_t + \alpha \frac{\partial f}{\partial b} \Big|_{w^t}, L_{t+1} = [L_t + \alpha \frac{\partial f}{\partial L} \Big|_{w^t}]^+, \tag{6}$$

where:

$$\begin{aligned}
 \frac{\partial f}{\partial w} &= w - 2C \sum_{i=1}^n \max[0, 1 - y_i(w^T L x_i + b)] (y_i L x_i), \\
 \frac{\partial f}{\partial b} &= -2C \sum_{i=1}^n \max[0, 1 - y_i(w^T L x_i + b)] y_i, \\
 \frac{\partial f}{\partial L} &= -2C \sum_{i=1}^n \max[0, 1 - y_i(w^T L x_i + b)] (y_i w x_i^T).
 \end{aligned} \tag{7}$$

Linear Metric learning with SVM (L-MSVM) algorithm is described as follows:

Algorithm: L-MSVM
Input: Dataset: $\{x_i, y_i\}_{i=1}^n = 1$
Output: L^*, w^*, b^*
Initial: β ($0 < \beta < 1$), $e = 10^{-6}$, $x_i \leftarrow x_i - \frac{1}{n} \sum_{i=1}^n x_i$, L_0, w_0, b_0
for $t = 0, 1, 2, \dots$
Calculate gradient $\frac{\partial f}{\partial(L_t, w_t, b_t)}$ by Eq. (7)
for $m = 0, 1, 2, \dots$
$\alpha = \beta^m$
Update $(L_{t+1}, w_{t+1}, b_{t+1})$ by Eq. (6)
If Armijo condition is satisfied:
break
if $ \frac{f(L_t, w_t, b_t) - f(L_{t+1}, w_{t+1}, b_{t+1})}{f(L_t, w_t, b_t)} < \varepsilon$,
$L^* = L_{t+1}, w^* = w_{t+1}, b^* = b_{t+1}$
break

Results and discussion

In this experiment, firstly GLCM was used and a length matrix was run to extract texture features from the microscope image at 4× and 10× magnifications. From the texture analysis

15 characteristic parameters were obtained, such as entropy, contrast, and factors and so on. Then, the two-samples were used to screen significant differences from two sets of 15 parameters. T-test selection parameters, as well as specific differences in the number of parameters, are as shown in Table 2.

Table 2. Parameters showing differences

Group (number)	Parameter
Group 4× (13)	Angular second moment, contrast, square and deficit moments, and variance, entropy, and entropy, difference entropy, run length nonuniformity factor, difference variance, intensity inhomogeneity factor, long-run factors, short-run factors
Group 10× (6)	Deficit moment, sum entropy, entropy, difference entropy, short run factor

To study the combined results of different classification algorithms, the experimental data was grouped. In the first experiment, after obtaining the filtered texture parameters, the three methods of linear SVM, ITML and L-MSVM were used directly to classify 10 times without a screening phase. Fig. 3 shows the results obtained:

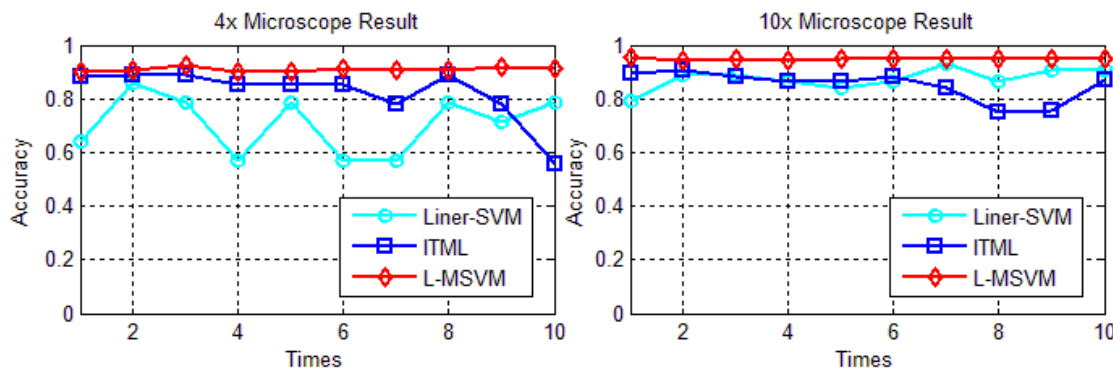


Fig. 3 The classified results without selection feature parameters

In the second case, significant difference parameters from the texture parameters were obtained with a two-sample T-test method, and then the significant difference parameters were classified 10 times, to obtain the characteristic parameters by the three methods. The classification results obtained are shown in Fig. 4.

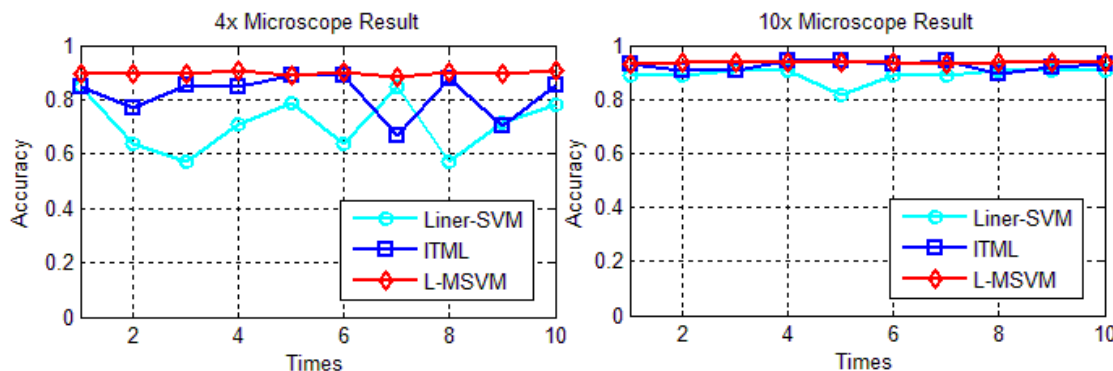


Fig. 4 The classification results of selection features parameters

From Figs. 3 and 4, it is clear that the classification accuracy based upon SVM measurement technology was the highest and most stable. The accuracy of just SVM and ITML was relatively lower and the experimental results were volatile.

Table 3 shows the mean value of the classification results before and after the selection of the characteristic parameters.

Table 3. The mean value for each experiment 10 times

	Before selection of parameters 4× magnification	After selection of parameters 4× magnification	Before selection of parameters 10× magnification	After selection of parameters 10× magnification
Linear-SVM	70.71%	71.09%	87.49%	89.08%
ITML	82.16%	82.00%	85.17%	92.52%
L-MSVM	90.89%	89.68%	94.77%	93.46%

As can be seen from Table 3, the overall accuracy rate of the results of the classification parameters under 4× magnification was lower than the result obtained under 10× magnification. In the four experiments of the three methods, the accuracy rate of the SVM classification algorithm based on the metrics was highest.

Another study of this experiment was to compare whether the image feature selection had a direct impact on the classification results. As can be seen from Table 3 and Fig. 4, the classification results from before and after the selection under 4× magnification did not exhibit great difference. However, after feature selection under 10× magnification, the recognition rate of Linear-SVM and ITML improved greatly. The author believes that 4× magnification, which before the feature selection had 15 parameters and after feature selection had only 13 parameters, showed that the number of parameters after the recognition rate was not very different. Conversely, while under 10× magnification, before feature selection it had 15 parameters, and after selection there were only six parameters which shows that there were significantly different characteristics after the selection. Therefore, the recognition rate after the selection of two algorithms improved significantly. Using the L-MSVM algorithm, the recognition rate of both before and after feature selection or selection feature, was very high, indicating that the degree of difference between the characteristic parameters for the algorithm was not sensitive and is reliable.

Furthermore, building upon the SVM classification algorithm based on the metrics, and adjusting the number of training samples of different proportions, the accuracy obtained was shown in Fig. 5.

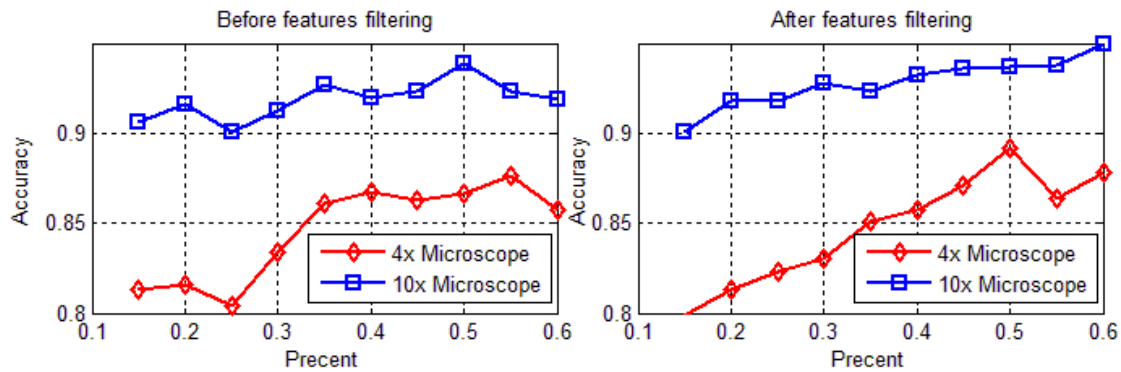


Fig. 5 The results of training samples of different proportion

As can be seen from Fig. 5, as the proportion of the samples increased, the recognition rate also gradually increased. When the proportion of the training samples was about 50%, the recognition rate was maximized and, as can be seen from Fig. 5, the recognition rate under 10× magnification was higher than under 4× magnification. Before no feature selection, the recognition rate curve fluctuated whilst after feature selection, the recognition rate curve was steadily increasing. Thus, feature selection for sample classification still plays a certain role.

Conclusion

In this study, the analysis for the different classification algorithms of osteoporosis can be seen. The recognition rate based on a measure was higher than a SVM algorithm using just learning metrics or SVM, and the classification results are more stable. On the other hand, as can be seen from the two sets of experiments, even if there was no filter for texture parameters, a high recognition rate could be still be obtained. Thus, we can conclude that, a SVM classification algorithm based on the metrics is suitable for low-dimensional, small sample data classification.

The image feature selection has a direct impact on the classification results. For the L-MSVM algorithm, the recognition rate of both before and after feature selection or selection feature was very high, indicating that the degree of difference between the characteristic parameters for the algorithm was not sensitive but was more robust.

The SVM classification algorithm based metrics can be used in the diagnosis of osteoporosis bone slice texture, thus it had a guiding significance in the diagnosis results on the limited conditions of medical experiments and clinical cases.

Acknowledgements

This study was supported by National Natural Science Foundation of China (No. 81541104), Science and Technology Project of Guangdong Province (No. 2014A030304006), Science and Technology Project of Zhanjiang (No. 2015A01039).

References

1. Brandi M. L. (2009). Microarchitecture, the Key to Bone Quality, *Rheumatology (Oxford)* 2009, 48(4), iv3-8.
2. Cai J., T. Wu, K. Zhou, W. Li (2015). Recognition of Osteoporosis based on Texture Analysis and Support Vector Machine, *International Journal Bioautomation*, 19(1), 107-118.

3. Chen J., H. Zhang, G. Yang, X. Lu, X. Jin, L. Lu, Q. Li (2014). Expert Consensus about the Current Application of Bone Histomorphometry, *Chinese Journal of Osteoporosis*, 12(10), 1031-1038 (in Chinese).
4. Cortes C., V. N. Vapnik (1995). Support Vector Networks, *Mach Learn*, 20, 273-297.
5. Davis J., B. Kulis, P. Jain, S. Sra, I. S. Dhillon (2007). Information-theoretic Metric Learning, *Proc 24th Int Conf Mach Learn*, 209-216.
6. Hazrati Marangalou J., K. Ito, B. van Rietbergen (2015). A Novel Approach to Estimate Trabecular Bone Anisotropy from Stress Tensors, *Biomech Model Mechanobiol*, 14(1), 39-48.
7. Hu L.-J., X. Li, H. Xia, L-Z. Tong (2012). 3D Texture Analysis of Hippocampus based on MR Images in Patients with Alzheimer Disease and Mild Cognitive Impairment, *Journal of Beijing University of Technology*, 38(6), 942-947 (in Chinese).
8. Koehne T., E. Vettorazzi, N. Küsters, R. Lüneburg, B. Kahl-Nieke, K. Püschel, M. Amling, B. Busse (2014). Trends in Trabecular Architecture and Bone Mineral Density Distribution in 152 Individuals Aged 30-90 Years, *Bone*, 66, 31-38.
9. Li Q. (2001). The Study of Animal Experiment on Osteopeina-bone Histomorphometry, Sichuan University Press (in Chinese).
10. Ollivier M., T. Le Corroller, G. Blanc, S. Parrattea, P. Champsaura, P. Chabrand, J.-N. Argenson (2013). Radiographic Bone Texture Analysis is Correlated with 3D Microarchitecture in the Femoral Head, and Improves the Estimation of the Femoral Neck Fracture Risk when Combined with Bone Mineral Density, *Eur J Radiol*, 82(9), 1494-1498.
11. Pei B.-Q., T.-M. Wang (2008). Comprehensive Biomechanical Analysis of Microcosimc Trabecular Structure of Cancellous Bone, *Journal of Beijing University of Aeronautics and Astronautics*, 34(2), 215-218 (in Chinese).
12. Qi J., H.-F. Xu, J. Wang et al. (2012). Application of Bone Histomorphometry and Micro CT Measurement Technology in the Study of Osteoporosis, *International Journal of Orthopaedics*, 33(3), 157-159 (in Chinese).
13. Udhayakumar G., C. M. Sujatha, S. Ramakrishnan (2013). Trabecular Architecture Analysis in Femur Radiographic Images Using Fractals, *Proc Inst Mech Eng H*, 227(4), 448-453.
14. Vapnik V. N. (1995). *The Nature of Statistical Learning Theory*, New York: Springer-Verlag.
15. Wang H. (2009). *Drug Research Methods and Technology in Osteoporosis*, Beijing People's Medical Publishing House, 134 (in Chinese).
16. Wolski M., P. Podsiadlo, G. W. Stachowiak (2014). Directional Fractal Signature Methods for Trabecular Bone Texture in Hand Radiographs: Data from the Osteoarthritis Initiative, *Medical Physics*, 41(8), doi: 10.1118/1.4890101.
17. Wu T., J. Cai (2014). Quantitative Diagnosis of Osteoporosis based on Texture Analysis, *Chinese Journal of Osteoporosis*, 20(7), 10-14 (in Chinese).
18. Zhou K., J. Cai, G.-Q. Xiong (2014). Optimized VBM and DARTEL Algorithm in Analyzing MRI of Alzheimer's Disease, *Chin J Med Imaging Technol*, 30(3), 462-466 (in Chinese).
19. Zhu X., P. Gong, Z. Zhao, C. Zhang (2012). Learning Similarity Metric with SVM, *IEEE World Congress on Computational Intelligence (WCCI)*, Brisbane, Australia, Article number 6252829.

Assoc. Prof. Ke Zhou, M.Sc.

E-mail: kitty@gdmc.edu.cn



Ke Zhou is an Associate Professor. She had her major in Computer Application. Now she is engaged in the researches on medical data mining and medical image analysis and processing.

Jie Cai, M.Sc.

E-mail: caijie513@163.com



Jie Cai is a lecturer. She had her major in Applied Mathematics. Now she is engaged in the researches on medical data mining and medical image analysis and processing.

Yong-Hui Xu, Ph.D. Student

E-mail: 840411554@qq.com



Yong-Hui Xu received his B.Sc. degree in Information and Computing Science from the Henan University, Kai Feng, China, in 2011. He is currently an undergraduate Ph.D. student with the School of Computer Science and Engineering, South China University of Technology, Guangzhou, China. His current research interests are focused on transfer learning, metric learning, multi-instance multi-label learning and their applications in computer vision and bioinformatics engineering.

Assoc. Prof. Tian-Xiu Wu, M.Sc.

E-mail: wutianxiu2005@163.com



Tian-Xiu Wu is an Associate Professor and a Postgraduate Supervisor. Her areas of expertise are bone biology and osteoporosis.