

Osteoporosis Diagnosis Based on Ultrasound Radio Frequency Signal via Multi-channel Convolutional Neural Network

Zhiwei Chen[†], Wenqiang Luo[†], Qi Zhang, Baiying Lei, Tianfu Wang, Zhong Chen, Yuan Fu, Peidong Guo, Changchuan Li, Teng Ma*, Yue Ding*, Jiang Liu*

Abstract—Osteoporosis is a metabolic osteopathy syndrome, and the incidence of osteoporosis increases significantly with age. Currently, bone quantitative ultrasound (QUS) has been considered as a potential method for screening and diagnosing osteoporosis. However, its diagnostic accuracy is quite low. By contrast, deep learning based methods have shown the great power for extracting the most discriminative features from complex data. To improve the osteoporosis diagnostic accuracy and take advantages of QUS, we devise a deep learning method based on ultrasound radio frequency (RF) signal. Specifically, we construct a multi-channel convolutional neural network (MCNN) combined with a sliding window scheme, which can enhance the number of data as well. By using speed of sound (SOS), the quantitative experimental results of our preliminary study indicate that our proposed osteoporosis diagnosis method outperforms the conventional ultrasound methods, which may assist the clinician for osteoporosis screening.

I. INTRODUCTION

Osteoporosis is a metabolic osteopathy syndrome characterized by decreased bone mass and destruction of bone microstructure, which leads to increased bone brittleness and risk of bone fracture [1]. The incidence of osteoporosis increases significantly with age. At present, the standard diagnostic methods for osteoporosis detection are dual energy X-ray absorptiometry (DXA), which provides great precision, accuracy and short-time scanning on the calculation of bone mineral density (BMD) [2]. However, the measured value of BMD depends on the size, shape, and soft tissue covering measured object to a large extent. Also, the value is sensitive to the change of human body posture when screening, and the trabecular microstructure cannot be evaluated using DXA [3].

*This work was supported partly by Bioland Laboratory (Guangzhou Regenerative Medicine and Health Guangdong Laboratory) (No. 1102101201).

Zhiwei Chen, Baiying Lei, Tianfu Wang are with the National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen, China, 518060.

Wenqiang Luo, Zhong Chen, Yuan Fu, Peidong Guo, Changchuan Li, Yue Ding are with the Department of Orthopedic Surgery, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, China.

Wenqiang Luo, Yue Ding are with the Bioland Laboratory (Guangzhou Regenerative Medicine and Health Guangdong Laboratory), Guangzhou, China, 510500.

Qi Zhang, Teng Ma are with the Institute of Biomedical and Health Engineering, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, 518055.

Jiang Liu is with the Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China, 518055, and with the Cixi Institute of Biomedical Engineering, Chinese Academy of Sciences, Zhejiang, China, 315201. Correspondence should be addressed to Jiang Liu (liuj@sustech.edu.cn), Yue Ding (dingyue@mail.sysu.edu.cn), Teng Ma (teng.ma@siat.ac.cn).

Furthermore, DXA scanners are expensive, bulky, highly ionizing, and are mostly owned by tertiary medical institutions. For these reasons, DXA is not the best technique for osteoporosis screening and fracture risk assessment in primary health care [4].

Since the first use of cancellous quantitative ultrasound (QUS) in 1984, QUS has been considered as a potential method for screening and diagnosing osteoporosis [5]. QUS technology evaluates bone health by measuring parameters related to the propagation of ultrasonic waves at different frequencies in the bone. The speed of sound (SOS) is one of the most commonly used parameters in traditional ultrasonic equipment [6]. As a noninvasive method of diagnosing osteoporosis, QUS has unique advantages, including lower operating cost, light, radiation-free, and no need for professional technicians [6-8]. However, the current diagnostic accuracy using QUS devices is too low to diagnose osteoporosis compared with DXA [8, 9]. One of the main reasons for the low diagnosis accuracy is that traditional QUS utilizes a simplified physical model of sound propagation in the bone and extracts quite few features such as SOS from ultrasound signals for BMD estimation [7, 8].

The ultrasonic radio frequency (RF) signal is a native unfiltered ultrasound signal, which contains rich information on structural details of tissues through which ultrasound passes. However, the large amount of information in the ultrasonic RF signal makes it difficult to analyze using the traditional algorithm. Since deep learning methods have shown the great power for extracting the most discriminative features from a large amount of raw data such as electrocardiograph and electroencephalogram analysis [10, 11], we devise a deep learning based model based on ultrasound RF signals from QUS detection instrument for osteoporosis diagnosis.

As a classic deep learning method, convolutional neural network (CNN) [12] can fully mine the hidden information of data in multiple fields. In order to explore the information of the RF signals of each channel, we construct a multi-channel neural network (MCNN) combined with a sliding window scheme to improve the osteoporosis diagnostic efficiency of QUS. Simultaneously, compared with conventional QUS method, traditional machine learning method and other deep learning methods, our method shows its high performance.

In this paper, we extract the original RF signals from the traditional radial QUS device, and utilize them as the input of MCNN rather than extracting several frequency or location information from raw data. Overall, our contributions are as follow:

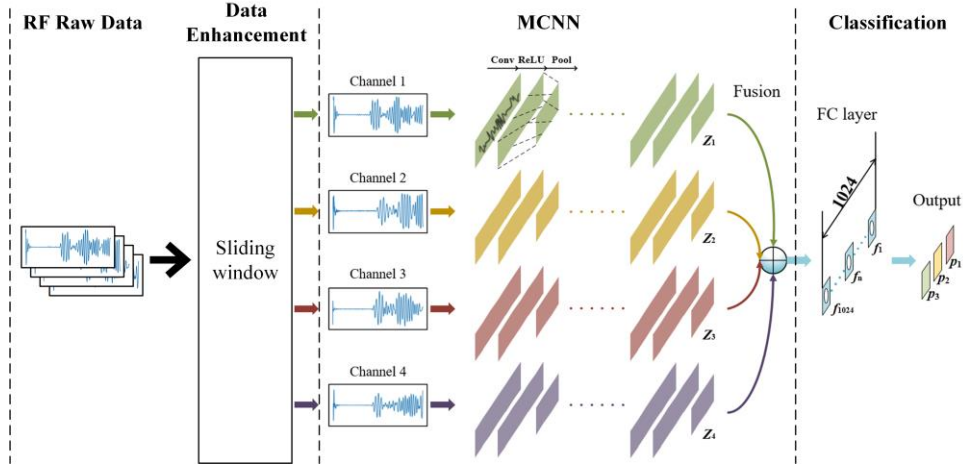


Fig. 1. Architecture of the proposed osteoporosis diagnostic model. Z_j is the feature vector from j -th channel, f_n is the output from n -th neuron of FC layer, and p_o is the prediction probability of the o -th specific class.

- We establish an osteoporosis diagnosis model via deep learning combined with ultrasound RF signals from QUS device, which is the first time as we know.
- Our method shows better osteoporosis diagnosis and bone mass loss classification performance than conventional QUS method using SOS, and has potential for osteoporosis screening.

II. METHODOLOGY

A. Data Acquisition

Ultrasound RF signals from 33 osteoporosis patients, 50 osteopenia patients and 31 normal subjects were obtained from Sun Yat-sen Memorial Hospital, Sun Yat-sen University (Guangzhou, China). Apart from DXA or QUS examination, all subjects had completed the FRAX questionnaire including age, sex, height, weight and so on.

The RF signal are time-amplitude data, which are collected at 1/3 of the distal radius of the non-dominant hand using an ultrasound bone sonometers (OSTEOKJ7000+, Kejin Industry Co., Ltd, Nanjing, China) that is capable of recording unprocessed original ultrasonic RF signal and calculating the SOS values. The ultrasound used in this study travels along the radius by axial transmission at 1MHz. SOS is the speed at which sound waves pass through the bone surface from the transmitting transducer to the receiving transducer. In addition to the SOS values, we extract the original ultrasound RF signals from the ultrasonic instrument. For each sampling frame, we can receive RF signals from each channel, and we get 4 channels data totally. In this study, each patient received the QUS examination 3 times.

B. Multi-channel Convolutional Neural Network

The MCNN was established using the Python package ‘keras’. As a complex model, the MCNN can reflect the complexity of the problem and extract the discriminative features from raw data. The architecture of osteoporosis diagnostic model based on the MCNN is presented in Fig. 1. In the context of CNNs, there are different ways that the modalities or data channels can be combined [13]. In this study, we employ the feature-level fusion scheme to fuse the feature vectors extracted from different channels. The RF data from 4 channels, which are regarded as multi-dimension data, are

provided to separate sets of convolutional layers and merge before the joint fully connected layer (FC layer). Each path learns independent sets of convolutional weights corresponding to their individual data channels, while traditional CNN can only learn shared weights among multiple channels.

Specifically, the model includes 4 input channels, 8 convolution blocks of each path, and a multi-layer perceptron which is composed of 2 FC layers and a softmax classifier for classification. As the feature extractor, each convolution block consists of a 1D convolution layer with 8 filters of length 5 and Rectified Linear Unit (ReLU) as the activation function, and following a maxpooling layer of length 2. For a multi-dimension time $T = t(i, j)$, where $t(i, j)$ denotes the value of the i -th time point of j -th dimension. Thus, the data of k -th dimension is denoted as:

$$T_{j=k} = \{t_{1,k}, t_{2,k}, \dots, t_{i,k}, \dots, t_{I,k}\}, \quad (1)$$

where I is the total time. At the time point i , the m -th element of output feature vector of the l -th convolution layer for the $T_{j=k}$ is given by

$$y^l(m, j = k) = \sum_i^{I=276210} w^l(i, j = k) x_{i-m,k}^{l-1} + b_l, \quad (2)$$

where $x_{i-m,k}^{l-1}$ is the input, w and b are the weights and biases.

After extracting the features, the output of the eighth consecutive convolution blocks for all dimensions is concatenated over the channel axis and then fed to a FC layer with 1024 neurons with ReLU as the activation function, following a dropout layer at a rate of 0.5. The concatenation can be expressed as:

$$f_n = \sum_j^{J=4} w(n, j) Z_j + b_n, \quad (3)$$

$$Z_j = \{y^8(1, j), y^8(2, j), \dots, y^8(m, j), \dots, y^8(M, j)\}, \quad (4)$$

where Z_j is the output of j -th channel’s convolution blocks, M is the length of the feature vector from each convolution pipeline, and f_n denotes the output of n -th neuron.

Finally, following the second FC layer with the number of neurons equal to the number of dataset classes, the softmax classifier is used for non-linear classification. The basic principle is that it provides a more systematic approach to data integration, and acquires an individual high-level feature representation of each data channel [14].

Same as other multi-class classification task, we utilize the categorical cross-entropy as the loss function. The loss function L is defined as:

$$L = - \sum \tilde{p}_o \log(p_o), \quad (5)$$

where \tilde{p}_o denotes the expected desired probability of o -th class, and p_o is the prediction probability of the o -th specific class. We adopt the stochastic gradient descent (SGD) to minimize the loss function, the parameters of the model can be manually optimized to achieve excellent performance.

C. Sliding window

Owing to the time translation invariance, the diagnostic result of QUS measurement for a subject won't be influenced by the start time and the end time of sampling in just a few minutes. Thus, to mitigate the lack of sufficient data in the test dataset and data imbalance, we utilize a sliding window scheme to enhance the dataset [15]. Fig. 2 presents that the sliding window can split data into a set number of sub-data according to the specific window length and stride. In this study, the window length and stride are set to 90 frames and 5 frames, respectively. 105 frames RF series data is collected per sampling. Thus, after preprocessing by sliding window, each subject has 4 RF series data of 90 frames per sampling.

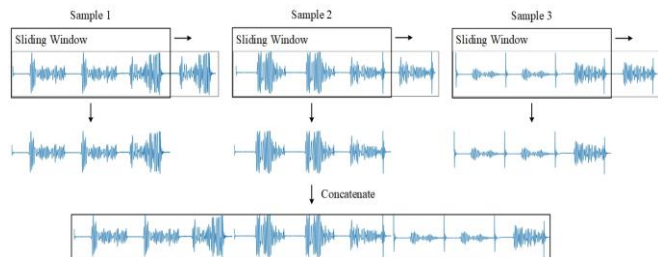


Fig. 2. The illustration of sliding window scheme.

As the method for osteoporosis screening, conventional QUS measurement uses SOS value to evaluate the bone mass loss of subjects. Each subject undergoes DXA and QUS examination 3 times, and outputs the mean of evaluation metrics as the diagnostic results, respectively. Therefore, we concatenate data from 3 times sampling after sliding window preprocessing.

III. EXPERIMENTS AND RESULTS

A. Experiment Setup and Data Preprocessing

In this study, MCNN is trained, validated and tested using RF data from 85, 14 and 15 subjects, respectively. The DXA detection results are used as true labels. To minimize the impact of high frequency noise, we employ an 8-th order low-pass filter with the cut-off frequency of 20MHz to preprocess the RF raw data for denoising [16]. Then RF signal are preprocessed by the sliding window scheme and concatenation, and each subject has 4 RF data of 270 frames in this study.

We normalize the data to the range where the variance is 1 and the average is 0. The data in both datasets are selected randomly. 10 times of 5-fold cross-validation are carried out to evaluate the model performance by averaging the results. The batch size is set to 32, and the learning rate is set to 0.0001 with the optimizer.

B. Results and discussion

In this paper, we evaluate the performance of MCNN by two tasks, osteoporosis diagnosis (OD) and classification of bone mass loss (CBML). We compare the diagnostic results of conventional QUS methods using SOS, Random Forest, CNN, Encoder [17] and MCNN of each subject for these tasks. After the sliding window preprocessing, data are transformed into 456 sequences for the experiments. We randomly divide data into 3 groups: 340 in the training group, 56 in the validation group, and 60 in the testing group. Simultaneously, to mitigate the impact of data distribution, all the groups had almost the same category ratio.

In order to validate our proposed method for the OD, we evaluate the model's performance using accuracy (ACC), specificity (SPE), sensitivity (SEN) and Kappa values. Table I shows the diagnostic results of different methods.

TABLE I. OSTEOPOROSIS DIAGNOSIS AND BONE MASS LOSS CLASSIFICATION PERFORMANCE USING DIFFERENT METHODS.

Method	Task	ACC	SPE	SEN	Kappa
SOS	OD	0.702	0.765	0.545	0.282
	CBML	0.421	0.704	0.423	0.120
Random Forest (RF signal)	OD	0.709	0.822	0.483	0.302
	CBML	0.518	0.736	0.472	0.229
CNN (RF signal)	OD	0.748	0.831	0.518	0.351
	CBML	0.594	0.715	0.550	0.340
Encoder (RF signal)	OD	0.788	0.846	0.630	0.467
	CBML	0.603	0.783	0.575	0.363
MCNN (RF signal)	OD	0.807	0.848	0.691	0.527
	CBML	0.641	0.830	0.617	0.422
MCNN (RF signal + MCRF)	OD	0.841	0.895	0.690	0.602
	CBML	0.695	0.837	0.696	0.524

Table 1 indicates that the overall accuracy based on SOS is 0.702, which is lower than the accuracy of 0.807 for RF signals using MCNN. Meanwhile, compared with the result of Random Forest, the results of deep learning methods in our study indicate that deep learning has a great power to mine RF signal which is better than the traditional machine learning method. Furthermore, the CNN and Encoder get the accuracy of 0.748 and 0.788, respectively, which shows that the feature-level fusion scheme can improve the model's performance. When adding main clinical risk factors (MCRF) of subjects to the model, such as sex, age, height and weight, the MCNN achieves the best accuracy of 0.841 while the sensitivity is basically constant.

Fig. 3 shows the receiver operating characteristic curves (ROC) and area under ROC (AUC) of different diagnostic methods for osteoporosis diagnosis. The AUC of SOS, CNN, Random Forest and Encoder based on RF signals are lower than the AUC of 0.80 when using MCNN. When adding MCRF into MCNN, the AUC increases to 0.87, which is around 0.2 higher than the conventional QUS method using SOS.

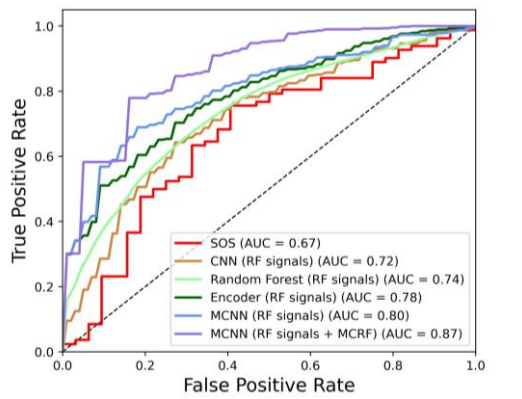


Fig. 3. ROC curve of different methods for osteoporosis diagnosis.

To further evaluate MCNN to distinguish bone mass loss, we divide the subjects into 3 groups according to DXA results, which are *normal*, *osteopenia* and *osteoporosis*, and compare the classification performance using the same method as the previous task. Table I summarizes the evaluation results. The results present that MCNN is significantly better than other methods on this task. Although the accuracy of the MCNN is only 0.641, MCNN outperforms SOS on this task using the same data. After adding main clinical risk factors (MCRF) to train the model, the ACC, SPE, SEN and Kappa value are improved to 0.695, 0.837, 0.696 and 0.524, which are around 27.2%, 13.3%, 27.3% and 40.4% higher than the conventional QUS method using SOS.

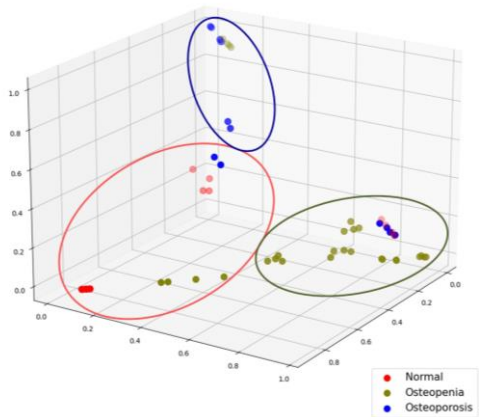


Fig. 4. The t-SNE visualization result of features extracted from MCNN.

Besides, we use the *t*-distributed stochastic neighbor embedding (t-SNE) method to visualize high level features extracted from MCNN. In Fig. 4, the t-SNE visualization result shows that the features have obvious inter-class differences, while the cycles of different colors indicate the classification results of MCNN. However, it still needs more researches to discriminate between normal and osteopenia patients in our future work. Moreover, we will explore more deep learning based methods for these tasks to acquire better performance, and the diversity of different deep learning methods may help us to better understand the ultrasound RF signals.

IV. CONCLUSION

In this study, we propose a deep learning based method to establish an osteoporosis diagnostic model based on ultrasonic RF signals of radius. We utilize a sliding window scheme to augment RF data, and explore a MCNN to extract and

integrate physiological information of different channels' data. The experimental results of our preliminary study initially indicate that our proposed method gets better performance on osteoporosis diagnosis compared with the conventional method. Our system may provide an effective screening method for people with high risk of osteoporosis. In the future, we will also try to explore other deep learning based methods to improve the performance of bone QUS apart from collecting more suitable data.

ACKNOWLEDGEMENT

This study was approved by the Sun Yat-sen Memorial Hospital Ethics Board (No: SYSEC-KY-KS-2019-159) and satisfied the criteria of the World Medical Association Declaration of Helsinki-Ethical Principles for Medical Research.

REFERENCES

- [1] J. E. Compston, M. R. McClung, and W. D. Leslie, "Osteoporosis," *Lancet*, vol. 393, no. 10169, pp. 364-376, 2019.
- [2] T. M. Link and G. Kazakia, "Update on Imaging-Based Measurement of Bone Mineral Density and Quality," *Current Rheumatology Reports*, vol. 22, no. 5, pp. 1-11, 2020.
- [3] N. R. Fuggle, *et al.*, "Fracture prediction, imaging and screening in osteoporosis," *Nature Reviews Endocrinology*, vol. 15, no. 9, pp. 525-527, 2019.
- [4] Y. Fu, *et al.*, "Fragility fracture discriminative ability of radius quantitative ultrasound: a systematic review and meta-analysis," *Osteoporosis International*, vol. 32, pp. 23-38, 2021.
- [5] C. M. Langton, S. B. Palmer, and R. W. Porter, "The measurement of broadband ultrasonic attenuation in cancellous bone," *Engineering in Medicine*, vol. 13, no. 2, pp. 89-91, 1984.
- [6] K. Thomsen, *et al.*, "Is calcaneal quantitative ultrasound useful as a prescreen stratification tool for osteoporosis?," *Osteoporosis International*, vol. 26, no. 5, pp. 1459-1475, 2015.
- [7] D. Hans and S. Baim, "Quantitative Ultrasound (QUS) in the Management of Osteoporosis and Assessment of Fracture Risk," *Journal of Clinical Densitometry*, vol. 20, no. 3, pp. 322-333, 2017.
- [8] E. V. McCloskey, *et al.*, "Predictive ability of heel quantitative ultrasound for incident fractures: an individual-level meta-analysis," *Osteoporosis International*, vol. 26, no. 7, pp. 1979-1987, 2015.
- [9] S. Nayak, *et al.*, "Meta-analysis: accuracy of quantitative ultrasound for identifying patients with osteoporosis," *Annals of Internal Medicine*, vol. 144, no. 11, pp. 832-841, 2006.
- [10] J. Jin, *et al.*, "Bispectrum-based channel selection for motor imagery based brain-computer interfacing," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 28, no. 10, pp. 2153-2163, 2020.
- [11] M. M. Al Rahhal, *et al.*, "Deep learning approach for active classification of electrocardiogram signals," *Information Sciences*, vol. 345, pp. 340-354, 2016.
- [12] Y. Lecun, Y. Bengio, and G. J. N. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.
- [13] B. K. Iwana and S. Uchida, "Time series classification using local distance-based features in multi-modal fusion networks," *Pattern Recognition*, vol. 97, pp. 1-12, 2019.
- [14] M. Liu, J. Zhang, E. Adeli, and D. Shen, "Joint Classification and Regression via Deep Multi-Task Multi-Channel Learning for Alzheimer's Disease Diagnosis," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 5, pp. 1195-1206, 2019.
- [15] Z. Shen, Y. Zhang, J. Lu, J. Xu, and G. Xiao, "A novel time series forecasting model with deep learning," *Neurocomputing*, vol. 396, pp. 302-313, 2020.
- [16] H. Fujisaka, T. Kamio, C. Ahn, M. Sakamoto, and K. Haeiwa, "A Sigma-Delta Domain Lowpass Wave Filter," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 62, no. 1, pp. 167-176, 2015.
- [17] J. Serrà, S. Pascual and A. Karatzoglou, "Towards a universal neural network encoder for time series," *arXiv preprint arXiv: 1805.03908v1*, 2018.