MALIGNANCY CHARACTERIZATION OF HEPATOCELLULAR CARCINOMA USING HYBRID TEXTURE AND DEEP FEATURES

Qiyao Wang, Lijuan Zhang, Yaoqin Xie, Hairong Zheng, Wu Zhou

Key Laboratory for Health Informatics, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

ABSTRACT

Malignancy of hepatocellular carcinoma (HCC) is significant to establish a therapeutic strategy preoperatively for liver cancer and is one of critical issues that influence recurrence and patient survival. Recently, quantitative texture feature of HCC in arterial phase of Contrast-enhanced MR has been shown to be promising for malignancy characterization of HCC. However, such texture feature is low-level, which is usually insufficient to capture the complicated characteristics of HCC. In this work, we propose a systematic method to automatically extract deep feature from the arterial phase of Contrast-enhanced MR using convolution neural network (CNN) in order to characterize malignancy of HCC. Specifically, we resample each 3D tumor in three orthogonal views (Axial, Coronal and Sagittal) independently to increase training sets, and train one CNN for each view to generate its corresponding deep feature. We investigate a multi-kernel feature fusion method that can fuse deep features derived from three views or fuse deep feature and texture feature in a kernel space. Our experimental results demonstrate several interesting conclusions: (1) deep feature significantly outperforms previous texture feature for malignancy characterization of HCC, (2) fusion of deep feature and texture feature yields best results for malignancy characterization of HCC.

***Index Terms*—** hepatocellular carcinoma, deep feature, texture feature, multiple kernel learning, feature fusion

1. INTRODUCTION

Hepatocellular carcinoma (HCC) is the third largest cancer in the world [1]. The histological grade of HCC differentiation is significant in establishing good treatment strategy and predicting the patient survival in clinical practice [2]. Generally, the histological grade of HCC is determined by the pathologic features of the lesion and assessed by a small preoperative tumor tissue obtained by biopsy [3]. However, biopsy is not necessarily reliable because of the sample error. Existing methods of medical imaging for malignancy characterization of HCC are usually performed by identifying reliable imaging features related to the malignancy nature of HCCs in terms of vascularity, cellular differentiation and capillary perfusion [4]. However, such image features for malignancy characterization are mostly selected according to the radiologists' experience, which may be insufficient to characterize the heterogeneity of the neoplasm.

This research is supported by the grant from National Natural Science Foundation of China (NSFC: U1301258), in part by grants from National Natural Science Foundation of China (NSFC: 61302171), Shenzhen Basic research project (No. JCYJ20150630114942291)and Guangdong Innovative Research Team Program (No. 2011S013).

Recently, learning deep features in a data-driven way has become an emerging technique for lesion characterization. It has been shown that deep features are better than some typical handcrafted features for cancer [5]. Suk H et al. used Deep Boltzmann Machine to find a latent feature representation from a volumetric patch and discover a joint feature representation from multi-modality [6]. Anthimopoulos M et al. proposed a deep CNN network to capture the low-level textural features of the lung tissue and classify lung CT image patches into 7 classes [7]. Inspired by the promising performance of deep feature, we believe that deep feature may have potential ability in biological characterization of HCC, especially for malignancy characterization. Meanwhile, previous studies also showed that the combination of low-level texture feature and high-level deep feature obtained performance improvement for lesion characterization [8]. As texture feature has been shown to be promising in malignancy characterization of HCCs, we can hypothesize that the fusion of deep feature and previous texture features may also yield better results for malignancy characterization of HCCs.

To this end, a systematic approach was proposed to extract deep feature and fuse low-level and high-level features of HCC from the arterial phase of Contrast-enhanced MR. Firstly, each 3D tumor was resampled in three orthogonal views (Axial, Coronal and Sagittal) independently to increase training sets. Then, each view was performed with CNN to generate corresponding deep feature. Furthermore, a multi-kernel feature fusion method was investigated to fuse deep features derived from the three orthogonal views or fuse deep features and texture features in a kernel space. Finally, experiments were conducted to evaluate the performance of the proposed method for malignancy characterization of HCC.

2. METHODS

2.1. Data and preprocessing

Consecutive 46 patients with 46 histopathologically proved HCC were included in the study from September 2011 to October 2015 (43male, 3 female, aged 53.09$\pm$12.45 years ranged 27 to 76 years). Gd-DTPA enhanced MR imaging were conducted with a 3T MR scanner and the precontrast, arterial phase, portal vein phase and delayed phased images were available. The malignancy of the HCC was acquired from the archived histology report, which histopathologically divided into Edmonson I,II,III and IV. Of the 46 HCCs, one was Edmonson I, twenty were Edmonson II, twenty-four were Edmonson III and one was Edmonson IV. Clinically, Edmonson I and II correspond to low-grade HCC, while Edmonson III and IV correspond to high-grade HCC. The objective of the study is to differentiate the twenty-one low-grade HCCs and the twenty-five high-grade HCCs. The general region of interests (ROIs) of all HCC in MR images were delineated by a radiologist who had 10 years of

experience in abdomen radiology.

Generally, CNN requires at least thousands of samples for training the network. However, the number of lesion is limited in the present study. In order to increase the training sample for deep learning, a multiview resampling technique was performed within the 3D tumor region similar to the work in [9]. As shown in Fig.1, multiple orthogonal views (Axial, Coronal and Sagittal) within ROI of a 3D tumor were considered as the samples. In the 3D tumor, each view of the axial, coronal and sagittal carries out N translations within the ROIs, resulting in $N = Na * Nc * Ns$ observations for each view, where Na is the translations for the Axial view, and Nc is the translations for the Coronal view and Ns is the translations for the Sagittal view. In this work, Na, Nc and Ns were optionally considered to be 5, and this procedure generated 125 samples for each view (Axial, Coronal or Sagittal), and finally results in 375 samples from a 3D tumor.

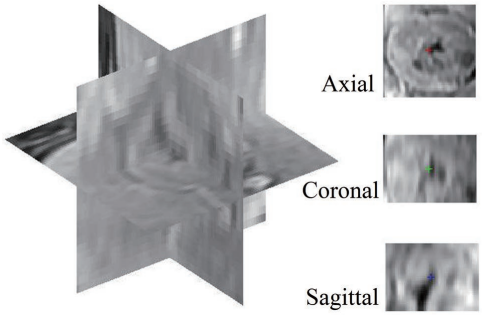

Fig. 1. Orthogonal views of a 3D tumor.

2.2. Convolutional Neural Networks

Deep learning is the most rapid development of machine learning method in recent years, and the convolutional neural network (CNN) is the most widely used method of deep learning in image processing [10]. It can use the output of a layer in the middle of the network as another representation of the data, which can be considered to be deep feature for object characterization [11]. The CNN convolves the original image x with convolution kernel w to get the convolution feature maps of the image:

$$y_n^i = F(\sum_{j=0}^{C_{n-1}} w_n^{ij} * x_{n-1}^j + b_n) \tag{1}$$

Where, y_n^i is the output of the i-th neuron of the n-th layer, and x_{n-1}^j is the input of the j-th neuron of the $(n-1)$-th layer. C_{n-1} is the number of $n-1$ neurons, b is the bias parameter in the convolution process, and $F(\cdot)$ is the squashing function. Pooling adopts the max-pooling:$y(i,j) = max_{0\leq m,n\leq s}\{x_{i\cdot s+m,j\cdot s+n}\}$. Where s is the size of the cell layer, x is the output of the convolution layer, and y is the output of the max-pooling layer.

The structure of CNN is shown in Fig.2. In this study, our model has two convolutions, and the convolution kernel is 5×5. The first convolution layer output 20 feature maps, and the second convolution output 50 feature maps. Each convolution layer is connected with a 2×2 layer of the max-pooling, followed by connected with the two-layer fully connected layer, in which the number of neurons was 500 and 2, respectively. The penultimate fully connected layer

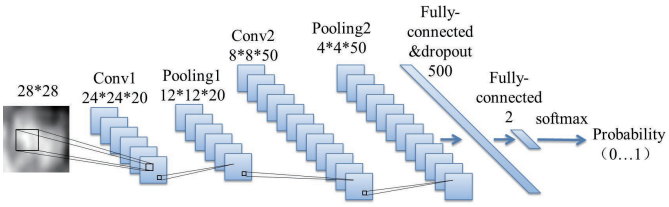

Fig. 2. Convolutional neural network architecture.

is the deep feature with dimension 500. Finally, the "softmax" layer output classification of the probability of identification results. The "dropout" technique [12] was used between the two fully connected layers to make the model have better generalization ability and prevent over-fitting. An open source deep learning framework "caffe" [13] was used to implement our model, which provided the interface with the NVIDIA Graphics Processing Unit. Cuda and cuDNN achieve GPU acceleration were used to improve training speed of the model. In addition, the neuron activation function "ReLU" [14], and $f(x) = max(0, x)$ was used to accelerate the convergence and learning speed of the model.

2.3. Texture feature

Texture analysis has been proved to be a technique to quantify the information of image pixels, which is widely used in the classification of tumor images [15]. Texture is the quantitative feature that relies on manual design. 141 texture feature were extracted for each image in the study, including Histogram features, Gray-level co-occurrence matrix (GLCM) [16], which consists four dimensions(horizontal, vertical, 45° and 135°) of the 28-dimensional features, and Gray-level run-length matrix (GLRLM) [17] with the same four directions of the 11-dimensional features. List of texture features was shown in Table 1.

2.4. Multiple kernel learning

Multiple kernel learning (MKL) method [18] comes from kernel learning methods, such as SVM classifier. MKL mainly solves the problem of multiple heterogeneous data sources and multiple kernel space corresponding to multiple data information. MKL has been successfully applied to fuse features derived from multimodalities [19]. Therefore, we consider fusing the deep features corresponding to three orthogonal views or fusing deep features and texture features using the MKL method, and then use the Support Vector Machines (SVM) as the classifier of the kernel learning method.

$$min\frac{1}{2}(\sum_{k=1}^{K}\|w_k\|_2)^2 + C\sum_{i=1}^{N}\xi_i \tag{2}$$

$$s.t. \quad y_i(\sum_{k=1}^{K}\langle w_k\Phi_k(X_i)\rangle + b) \geq 1 - \xi_i$$
$$\forall i = 1, \ldots, N, \xi_i \geq 0$$

Where, w_k is the weight set of the k-th sub-kernel's feature space, and ξ_i is the relaxation variable, y_i is the category label, $\Phi_k(\cdot)$ is the k-th sub-kernel feature space, and b is the weighting coefficient. MKL method gets the weight of each kernel by adjusting w_k. The larger w_k is, the greater the contribution of kernel to the final

Table 1. List of texture features used in the study

Category	Feature	N
Histogram	Mean, variance, skewness, kurtosis, energy	5
GLCM	Autocorrelation, Contrast, Correlation, Correlation, Cluster Prominence, Cluster Shade, Dissimilarity, Energy, Entropy, Homogeneity, Homogeneity, Maximum probability, Sum of squares, Sum average, Sum variance , Sum entropy, Difference variance, Difference entropy, Information measure of correlation1,Information measure of correlation2,Inverse difference, Inverse difference normalized, Inverse difference moment normalized. Matrices were calculated for one inter-pixel distance in four directions: horizontal, vertical, 45° and 135°.	28∗ 4
GLRLM	Short run emphasis, Long run emphasis, Grey level nonuniformity, Run length nonuniformity, Run percentage,Low grey level run emphasis, High grey level run emphasis , Short run low grey level emphasis, Short run high grey level emphasis, Long run low grey level emphasis, and Long run high grey level emphasis. Each feature was calculated in four directions: horizontal, vertical, 45° and 135°.	11∗4

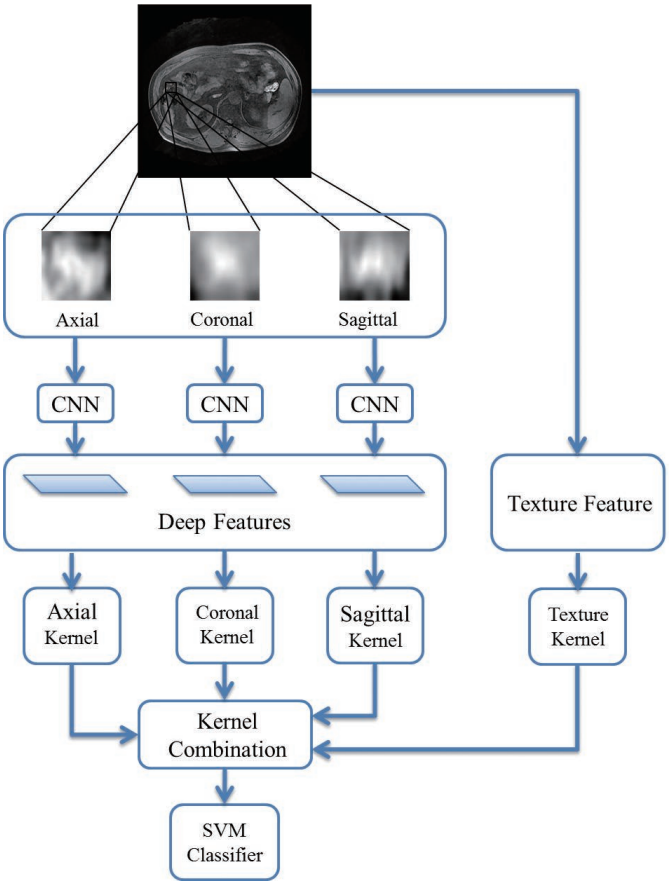

Fig. 3. Flowchart of texture and deep feature fusion.

classifier result. The deep feature based on three orthogonal views corresponds to three different sub-kernels, and the weight assigned to them can be interpreted as their corresponding contribution size.

2.5. The proposed framework of feature fusion

A multi-kernel feature fusion method was used to fuse deep features derived from three orthogonal views or fuse deep and texture feature for malignancy characterization in a kernel space, respectively. The flowchart of the method was shown in Fig.3. First, three deep features with 500-dimension from the three orthogonal views were extracted using CNN, respectively. Then, texture feature with 141-dimension were generated for the axial view of HCCs. Finally, deep and texture features were trained independently to yield optimal kernel functions, while weights of kernel functions associated with deep and texture features were determined by the MKL method that could map deep and texture features into high dimension to achieve optimal classification performance.

3. EXPERIMENTS AND RESULTS

Experiments were conducted on the 46 clinical HCCs, which consisted of 21 low-grade and 25 high-grade HCCs. The accuracy, sensitivity and specificity were used to quantitatively evaluate the performance of the characterization for high-grade HCCs. In order to objectively evaluate the characterization performance of different features for malignancy estimation of HCCs, 4-fold cross-validation with 10 repetitions on the 46 HCCs was used to test the proposed framework, in which 33 were randomly selected for training and the remaining 13 for testing.

3.1. Texture features VS deep features

Texture features with 141-dimension were extracted from axial view, including histogram, gray level co-occurrence matrix and the gray-level run-length matrix, and combined with the SVM-RFE feature selection method [20] to obtain the best classification results. Deep features for the three orthogonal views were derived from multiple samples for each view with CNN, respectively. The deep feature used a 500-dimensional fully-connected layer matrix. All features were classified using the LibSVM [21] classifier. 10-times 4-fold cross validation was used for the performance comparison and the test results were shown in Table 2.

As shown in Table 2, deep feature clearly outperformed texture feature in Axial view in terms of accuracy, sensitivity and specificity. Deep feature was high-level, which was verified to be more powerful to characterize malignancy of HCCs than the low-level texture feature. Specifically, deep feature in Axial view yielded best performance for malignancy characterization, which could be appropriately explained that the axial view had the highest resolution for lesion imaging, resulting in higher performance for malignancy characterization compared with other two orthogonal views.

Table 2. Comparison of texture and deep features (%)

Feature		Accuracy	Sensitivity	Specificity
Texture	Axial	82.50±12.70	88.89±11.15	74.55±22.01
Deep	Axial	96.92±3.97	98.75±3.95	95.46±7.61
	Coronal	76.15±10.54	76.39±17.88	79.96±13.38
	Sagittal	85.38±5.68	86.25±8.96	86.89±14.64

Table 3. Comparison results of feature fusion (%)

Methods	Accuracy	Sensitivity	Specificity
CNN	86.15±7.07	85.00±9.46	87.14±10.54
Deep-MKL	97.95±5.12	95.71±7.12	1.00±0.00
Deep+Texture-MKL	99.23±2.43	98.89±3.51	1.00±0.00

3.2. Fusion of deep and texture feature

Three kinds of feature fusion were considered for the comparison: (1) Directly three orthogonal views ensemble for CNN to generate a deep feature, (2) fusion of deep features from three orthogonal views, (3) fusion of deep and texture. Experimental results were shown in Table 3. Deep-MKL represented the fusion of deep features from three orthogonal views using MKL method, while Deep+Texture-MKL denoted the fusion of three deep features from three orthogonal views and texture feature from the axial view using MKL method. As shown in Table 3, the method of CNN that using all three orthogonal views as the input to generate a deep feature for classification did not generate promising result, which was much lower than the case of using the axial view. It might be indicated that directly multiple views ensemble for CNN did not result in better results due to the heterogeneity of tumor in three orthogonal views. It could also be easily found that the accuracy of the Deep-MKL ($97.95 \pm 5.12\%$) was better than that of deep feature in single orthogonal views (axial: $96.92 \pm 3.97\%$, Coronal: $76.15 \pm 10.54\%$, Sagittal: $85.38 \pm 5.68\%$). It clearly demonstrated that fusion of deep features in three orthogonal views using MKL was remarkably effective to obtain better results for malignancy characterization. Furthermore, the Deep+Texture-MKL yielded the best result ($99.23 \pm 2.43\%$) for malignancy characterization, which verified that low-level texture feature could also provide additional information to improve performance of deep feature for malignancy characterization.

4. CONCLUSION

Our study demonstrates that deep feature outperform texture feature for malignancy characterization, and fusion of deep and texture feature yields better results. Meanwhile, our study also suggests that fusion of deep features derived from three orthogonal views outperforms deep feature from a single view for malignancy characterization of HCC. We believe that our proposed approach can be extensively used for lesion characterization in clinical practice.

5. REFERENCES

[1] J.M. Llovet, A. Burroughs, and J. Bruix, "Hepatocellular carcinoma," *Lancet*, vol. 362, pp. 1907–1917, December 2003.

[2] S. Jonas, W.O. Bechstein, T. Steinmller, M. Herrmann, C. Radke, T. Berg, U. Settmacher, and P. Neuhaus, "Vascular invasion and histopathologic grading determine out-come after liver transplantation for hepatocellular carcinoma in cirrhosis," *Hepatology*, vol. 33, pp. 1080–1086, May 2001.

[3] M. Robert, A.N. Sofair, A. Thomas, B. Bell, S. Bialek, C. Corless, G. Van, S. Hu ie White, N. Stabach, and A. Zaman, "A comparison of hepatopathologists' and community pathologists' review of liver biopsy specimens from patients with hepatitis c," *Clinical Gas-troenterology Hepatology*, vol. 7, pp. 335–338, March 2009.

[4] W. Zhou, L. Zhang, K. Wang, S. Chen, G. Wang, Z. Liu, and C. Liang, "Malignancy characterization of hepatocellular carcinomas based on texture analysis of contrast-enhanced mr images," *Journal of Magnetic Resonance Imaging*, vol. 45, pp. 1476–1484, May 2017.

[5] Y. Xu, T. Mo, Q. Feng, and P. Zhong, "Deep learning of feature representation with multiple instance learning for medical image analysis," in *International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2014, pp. 1626–1630.

[6] H.I. Suk, S.W. Lee, and D. Shen, "Hierarchical feature representation and multimodal fusion with deep learning for ad/mci diagnosis," *Neuroimage*, vol. 101, pp. 569–582, November 2014.

[7] M. Anthimopoulos, S. Christodoulidis, L. Ebner, and A. Christe, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 35, pp. 1207–1206, May 2016.

[8] T. Xu, H. Zhang, X. Huang, S. Zhang, and D.N. Metaxas, "Multimodal deep learning for cervical dysplasia diagnosis," in *International Conference on Medical Image Computing and Computer-assisted Intervention*, 2016, pp. 115–123.

[9] H. Roth, L. Lu, J. Liu, J. Yao, A. Seff, K. Cherry, L. Kim, and R.M. Summers, "Improving computer-aided detection using convolutional neural networks and random view aggregation," *IEEE Transactions on Medical Imaging*, vol. 35, pp. 1170–1181, May 2016.

[10] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "Multimodal deep learning for cervical dysplasia diagnosis," in *International Conference on Neural Information Processing Systems*, 2012, vol. 25, pp. 1097–1105.

[11] A.S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "Cnn features off-the-shelf: An astounding baseline for recognition," in *IEEE Conferenc on Computer Vision and Pattern Recognition Workshops*. IEEE, 2014, pp. 512–519.

[12] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.

[13] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: convolutional architecture for fast feature embedding," *Eprint Arxiv*, pp. 675–678, 2014.

[14] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *International Conference on Artificial Intelligence and Statistics*, 2011, pp. 315–323.

[15] G.M. Stavroula, K.V. Ioannis, N. Alexandra, and S.N. Konstantina, "Differential diagnosis of ct focal liver lesions using texture features, feature selection and ensemble driven classifiers," *Artificial Intelligence in Medicine*, vol. 41, pp. 25–37, September 2007.

[16] R.M. Haralick, K. Shanmugan, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems Man and Cybernetics*, pp. 610–621, 1973.

[17] X. Tang, "Texture information in run-length matrices," *IEEE Transactions on Image Processing*, vol. 7, pp. 1602–1609, 1998.

[18] F. R. Bach, G. R. G. Lanckriet, and M. I. Jordan, "Multiple kernel learning, conic duality, and the smo algorithm," in *International Conference on Machine learning*, 2004, p. 6.

[19] C. Hinrichs, V. Singh, G. Xu, and S.C. Johnson, "Predictive markers for ad in a multi-modality framework: An analysis of mci progression in the adni population," *Neuroimage*, vol. 55, pp. 574–589, March 2011.

[20] K. Yan and D. Zhang, "Feature selection and analysis on correlated gas sensor data with recursive feature elimination," *Sensors and Actuators B:Chemical*, vol. 212, pp. 353–363, June 2015.

[21] C.C. Chang and C.J. Lin, "www.csie.ntu.edu.tw/ cjlin/libsvm," .