A Deep Learning Trial on Transient Elastography for Assessment of Liver Fibrosis

Yongshuai Li, Qiong He and Jianwen Luo
Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing, China
Email: luo_jianwen@tsinghua.edu.cn

***Abstract*—Transient elastography (TE) is a non-invasive, rapid and reproducible technology, which performs liver stiffness measurement (LSM) for staging the liver fibrosis in the clinic. In the procedure of LSM, an M-mode strain image (MSI) is estimated from the M-mode radiofrequency (RF) data. On the hypothesis that the MSI contains richer information than LSM, we propose a deep learning (DL) method based on the MSI to improve the performance of fibrosis staging. A multicenter study was conducted where both TE and liver biopsy were performed on 421 patients and finally 245 patients with 2,713 MSIs were qualified. An optimal deep learning model with 8 layers (5 convolutional layers and 3 fully-connected layers) was built and evaluated on the dataset. LSMs were also used to assess liver fibrosis for comparison, while liver biopsy acted as the gold standard. The receiver operating characteristic (ROC) curve was plotted to evaluate the performance for assessing significant fibrosis ($\geq$ F2), advanced fibrosis ($\geq$ F3) and cirrhosis (F4). The AUC of the DL method was 0.850 for $\geq$ F2. The AUCs of the DL method were 0.948 for $\geq$ F3 and 0.934 for F4, respectively, slightly better than those of the LSM method (0.935 for $\geq$ F3 and 0.908 for F4, respectively). To conclude, the DL method performs better than the LSM method for advanced fibrosis and cirrhosis and could be used in transient elastography for assessment of liver fibrosis.**

***Keywords*—deep learning, liver fibrosis, transient elastography**

I. Introduction

Liver fibrosis is an injury healing response caused by various chronic liver diseases (CLDs), such as virus hepatitis (mostly hepatitis B and hepatitis C), non-alcoholic fatty liver disease (NAFLD), drug-induced liver injury (DILI), and immunological liver injury [1]. Accurate assessment of the degree of liver fibrosis is important for treatment arrangement and fibrosis reversal [2], [3]. For liver fibrosis staging, liver biopsy is considered as the gold standard. However, it may fail due to sampling error and cause other complications because of its invasiveness [3]. Recently, transient elastography (TE) was developed for assessment of liver fibrosis as a non-invasive, rapid and reproducible modality. In clinical practice with TE, liver stiffness measurement (LSM) is performed and certain thresholds are adopted to stage liver fibrosis.

However, LSM method has obvious limitations: (1) Only the liver stiffness value is used, leading to inefficient use of information; (2) The thresholds vary in different studies [4], [5]. In the procedure of LSM, an MSI is obtained from the M-mode radiofrequency (RF) data of the liver acquired after an external vibration. On the hypothesis that the M-mode strain image (MSI) contains richer information than LSM, we utilize image analysis technique based on the MSI to improve the performance of fibrosis staging.

Deep learning (DL) has greatly improved the performance of image detection, segmentation and classification [6], [7]. In essence, liver fibrosis staging based on MSIs is an image classification task. Currently, DL has been used to assess liver fibrosis on liver B-mode ultrasound images [8] and liver shear wave elastography (SWE) [9]. Convelutional neural networks (CNNs) were used to automatically extract high-level features of images. They both achieved encouraging improvement. Inspired by these two studies, we propose a DL method of TE based on the MSI to improve the performance of fibrosis staging.

II. Methods

A. Recruitment and Processing

A multicenter study involved with 13 Chinese hospitals was conducted where both TE and liver biopsy were performed on patients. The patient inclusion criteria were as follows: (1) older than 18 year; (2) without ascites; (3) without space-occupying lesions. The patient exclusion criteria were as follows: (1) unqualified histology; (2) incomplete data or information; (3) less than 3 successful TE measurements. All the enrolled patients signed the informed consent. From Mar 2014 to Jun 2017, 421 patients were enrolled and 245 patients with 2,713 data were qualified.

In this study, all the staging results were recorded using Metavir scoring system, from normal, significant fibrosis to cirrhosis as F0-F4. Liver histological samples were read by three experts and the staging results acted as the gold standard. The procedure of obtaining LSMs and MSIs were as follows. First, the M-mode RF data were acquired with a commercial TE system (FibroTouch, Wuxi Hisky, China. vibration freq. = 50 Hz, f_0 = 2.5 MHz, f_s = 25 MHz). Then, a normalized cross-correlation method (window length = 2 mm) was applied on the RF data for displacement estimation [10]. Finally, the MSI was obtained by using Savitzky-Golay (SG) filter [11]. Examples of MSIs of F0-F4 are shown in

This study was supported in part by the National Key R&D Program of China (2016YFC0102200), the National Natural Science Foundation of China (81471665, 81561168023, 61871251, and 61801261) and the China Postdoctoral Science Foundation (2017M620802).

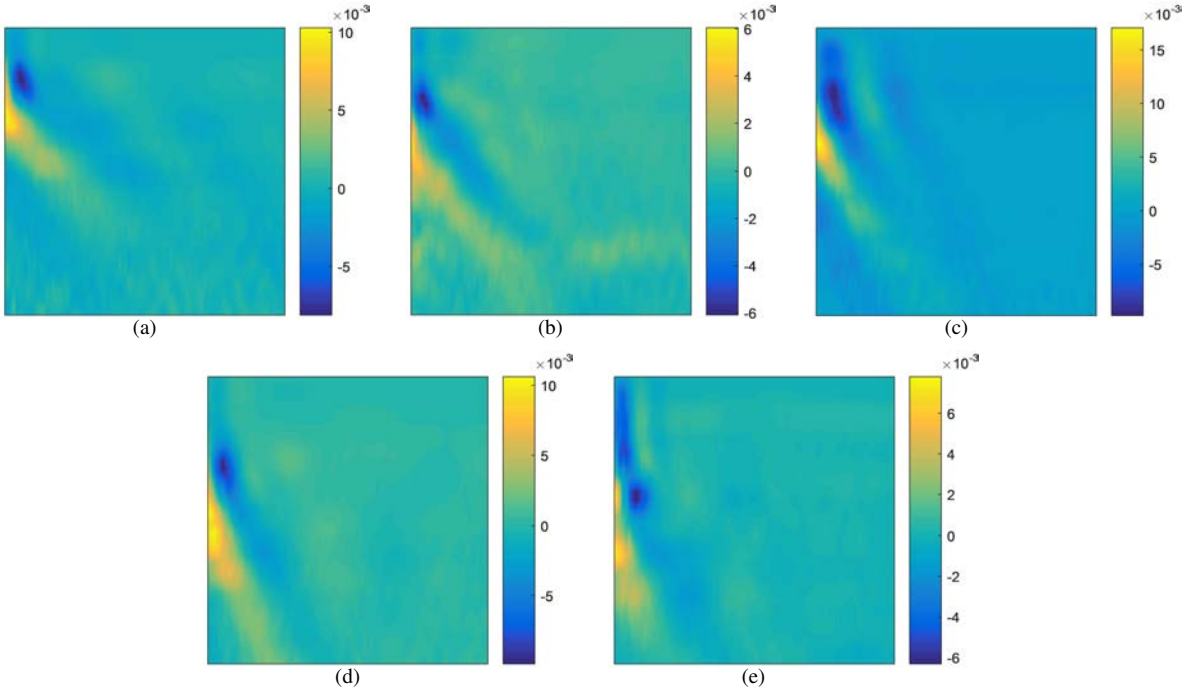


Fig. 1. Examples of M-mode strain images at (a) F0, (b) F1, (c) F2, (d) F3 and (e) F4, respectively.

Fig. 1.

To suppress overfitting due to small dataset, data augmentation was implemented by cropping the MSIs randomly. The cropping time of MSI varies with fibrosis stage to compensate for the unbalanced distribution of data (as shown in Table I).

TABLE I
NUMBERS OF PATIENTS, MSIS AND CROPPING TIMES AT EACH FIBROSIS STAGE

Fibrosis stages	Patients	MSIs	Cropping times
F0	24	266	11
F1	75	727	4
F2	57	577	5
F3	74	939	3
F4	15	204	14
Total	**245**	**2713**	**–**

B. Deep Learning Model

Considering the trade-off between prediction losses and overfitting, an optimal DL model was built. The architecture of the DL model was as follows (as shown in Fig. 2): 5 convolutional units, 2 fully-connected layers with dropout and 1 fully-connected layer for output. Each convolutional unit consisted of a convolutional layer (filter kernel size: 3 $\times$ 3, stride: 3, filter number: $16k$, where k was the kth convolutional unit) with a Relu activation function followed by batch normalization and max-pooling layer (kernel size: 2 $\times$ 2, stride: 2). Using this structure, the input images/feature maps were subsampled to half in length and width while doubled in depth through each convolutional unit. The hidden fully-connected layers transformed feature maps to vectors (with a length of 512) and the dropout probability was 0.5. It took the MSIs (80 $\times$ 80 pixels) as the input and output the probablityprobability of fibrosis stages. Biopsy results acted as the gold standard and labels for the DL model.

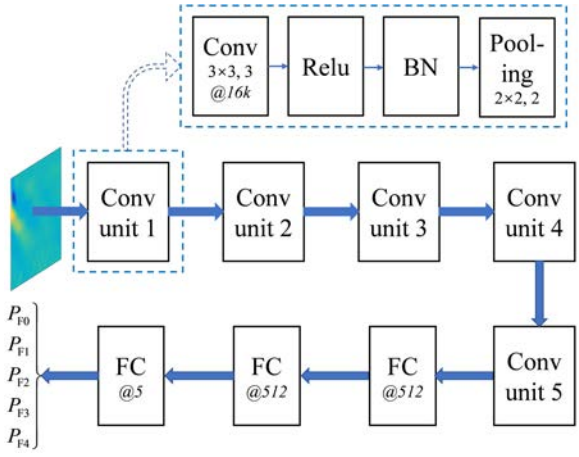


Fig. 2. Diagram of the DL model. @$16k$ denotes the output number of the kth Conv unit is $16k$; BN: batch normalization layer; FC: fully connected layer.

The dataset was randomly divided into 3 cohorts, each containing 1/3 of the patients at each stage. The training set consisted of two cohorts and the testing set consisted of the rest one. The DL model was trained on the training set to optimize the hyperparameters, then the testing set was used to validate the generated model. LSMs were also used to assess

liver fibrosis on the testing set for comparison, while liver biopsy acted as the gold standard.

III. Results

A. Parameters of DL Model

In the training input pipelines, the input data were shuffled and the mini-patch size was 64. Softmax cross entropy cost function was adopted. AdaDelt optimizer with adaptive learning rate was used and the initial learning rate was 0.1. The DL model was trained on the training set for 100 epochs on Tensorflow-gpu (vesion 1.10.0, Google Inc., Santa Clara, California) using TITAN V GPUs (NVIDIA, Santa Clara, California) with 12-GB memory.

B. Distribution of Data

Before liver fibrosis staging, we first analyzed the separability of LSMs and MSIs. The distributions of LSMs from F0 to F4 are shown in Fig. 3 (a). The mean value of LSMs is in direct proportion to the degree of liver fibrosis. The distribution of F2 and F3 had obvious difference. However, there is large overlaps between the distributions of F1 and F3, F3 and F4. This indicates that LSM is an effective but not ideal indicator for liver fibrosis staging. To visualize the distributions of MSIs, t-distributed stochastic neighbor embedding (t-SNE) was performed (number of components: 2, perplexity: 50) and the MSIs were reduced to 2 points [12], [13]. As shown in Fig. 3 (b), the overlay between different stages still exists. However, the MSIs may be more separable in higher dimensions and the features could be captured by the DL method.

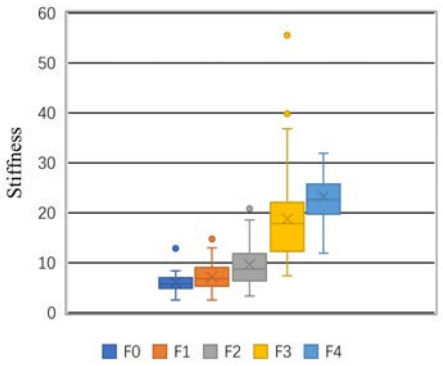
C. Diagnostic Accuracy of DL Method

The receiver operating characteristic (ROC) curve was plotted to evaluate the performance for assessing significant fibrosis ($\geq$ F2), advanced fibrosis ($\geq$ F3) and cirrhosis (F4). 3-fold cross-validation was implemented, the ROC curves were averaged (Fig. 4), and the area under ROC curve (AUC) was calculated from the average curve.

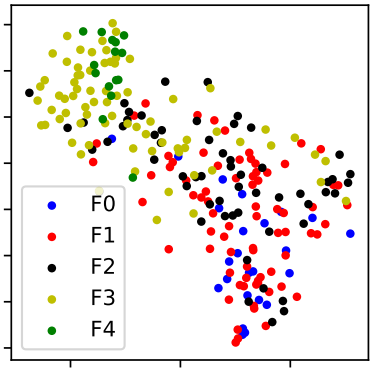
The AUC of the DL method was 0.850 for $\geq$ F2, slightly less than that of LSM method (0.865). The AUCs of the DL method were 0.948 for $\geq$ F3 and 0.934 for F4, respectively, slightly better than those of the LSM method (0.935 for $\geq$ F3 and 0.908 for F4, respectively), as shown in Table II.

TABLE II
PERFORMANCE OF DL METHOD OF MSIS ON TESTING SET

	Method	AUC	Sensitivity (%)	Specificity (%)	PPV (%)	FPV (%)
≥F2	DL	0.850	76.64	71.72	80.70	68.32
	LSM	0.865	77.55	78.79	84.44	70.27
≥F3	DL	0.948	81.92	89.74	82.22	89.90
	LSM	0.935	80.00	88.46	80.00	88.46
F4	DL	0.934	40.00	94.79	53.97	96.15
	LSM	0.908	40.00	94.81	33.33	96.05



(a)



(b)

Fig. 3. Distributions of (a) LSMs, and (b) t-SNE of M-mode strain images.

IV. Discussion

The DL method obtains a slightly better performance than LSM method for assessment of advanced fibrosis ($\geq$ F3) and cirrhosis (F4), but its performance is close to that of the DLRE method proposed in [9]. In the DL method, MSIs are estimated in the procedure of LSM and no more data are required. With the powerful image analysis capability of DL, better liver fibrosis assessment results are obtained. Thus, the DL method is feasible to implement on existing TE systems.

Both the DL method and LSM method have relatively low performance on assessing significant fibrosis ($\geq$ F2) with AUCs of around 0.85. This could be explained by the distributions of the data (Fig. 3). There is a large overlap between F2 and F3 in the distributions of both LSMs and MSIs. More information or indicators are required to further improve the performance of liver fibrosis assessment.

V. Conclusion

In this paper, a multicenter study with 13 hospitals and 421 patients is conducted to evaluate the performance of the proposed deep learning method in the assessment of liver fibrosis. The M-mode strain images estimated in the procedure of liver stiffness measurement are used as input. Without

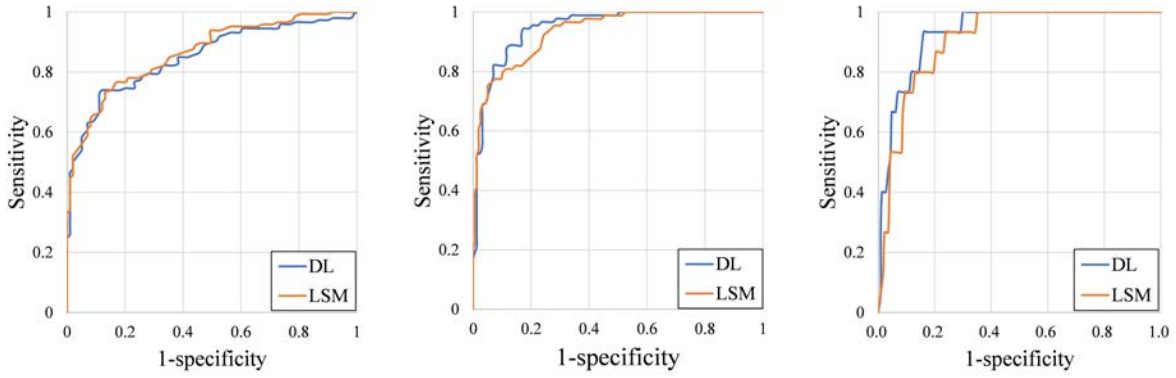


Fig. 4. Comparison of ROC curves of DL method and LSM method on for assessment of (a) significant fibrosis ($\geq$ F2), (b) advanced fibrosis ($\geq$ F3) and (c) cirrhosis (F4) testing set, respectively.

requirement of more data, the deep learning method obtains better liver fibrosis assessment results than liver stiffness measurement method. The deep learning method could be used in transient elastography for assessment of liver fibrosis.

REFERENCES

[1] M. Pinzani, K. Rombouts, and S. Colagrande, "Fibrosis in chronic liver diseases: diagnosis and management," *Journal of Hepatology*, vol. 42, no. 1, pp. S22–S36, 2005.

[2] S. L. Friedman and M. B. Bansal, "Reversal of hepatic fibrosis —fact or fantasy?" *Hepatology*, vol. 43, no. S1, pp. S82–S88, 2006.

[3] D. Rockey, "Noninvasive assessment of liver fibrosis and portal hypertension with transient elastography," *Gastroenterology*, vol. 134, no. 1, pp. 8–14, 2008.

[4] L. Castera, X. Forns, and A. Alberti, "Non-invasive evaluation of liver fibrosis using transient elastography," *Journal of Hepatology*, vol. 48, no. 5, pp. 835–847, 2008.

[5] J. Foucher, "Diagnosis of cirrhosis by transient elastography (FibroScan): a prospective study," *Gut*, vol. 55, no. 3, pp. 403–408, 2006.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[7] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[8] D. Meng, L. Zhang, G. Cao, W. Cao, G. Zhang, and B. Hu, "Liver fibrosis classification based on transfer learning and FCNet for ultrasound images," *IEEE Access*, pp. 5804–5810, 2017.

[9] K. Wang, X. Lu, H. Zhou, Y. Gao, J. Zheng, M. Tong, C. Wu, C. Liu, L. Huang, T. Jiang, F. Meng, Y. Lu, H. Ai, X.-Y. Xie, L. Yin, P. Liang, J. Tian, and R. Zheng, "Deep learning radiomics of shear wave elastography significantly improved diagnostic performance for assessing liver fibrosis in chronic hepatitis b: a prospective multicentre study," *Gut*, pp. gutjnl–2018–316 204, 2018.

[10] J. Luo and E. E. Konofagou, "A fast normalized cross-correlation calculation method for motion estimation," *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 57, no. 6, pp. 1347–1357, 2010.

[11] J. Luo, J. Bai, P. He, and K. Ying, "Axial strain calculation using a low-pass digital differentiator in ultrasound elastography," *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, vol. 51, no. 9, pp. 1119–1127, 2004.

[12] L. V. D. Maaten and G. Hinton, "Visualizing data using t-sne," *Journal of Machine Learning Research*, vol. 9, no. 2605, pp. 2579–2605, 2008.

[13] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. J. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.