

Imaging-based deep learning in liver diseases

Enyu Yuan^{1,2}, Zheng Ye^{1,2}, Bin Song^{1,2,3}

¹West China School of Medicine, Sichuan University, Chengdu, Sichuan 610041, China;

²Department of Radiology, West China Hospital, Chengdu, Sichuan 610041, China;

³Department of Radiology, Sanya People's Hospital, Sanya/West China (Sanya) Hospital, Sanya, Hainan 572022, China.

Liver disease is a collective term for hepatic damage caused by heterogeneous etiologies and poses a heavy health burden worldwide. Modern medical imaging, such as computed tomography (CT), magnetic resonance imaging (MRI), and ultrasonography, plays an essential role in the diagnosis and management of liver diseases. Imaging-based deep learning (DL) is one of the most actively investigated techniques in recent decade. It enables high-dimensional features extraction through the layered structures and can be used in a variety of clinical scenarios, especially with the convolutional neural network in computer vision.

Most DL tasks can be divided into image segmentation, image classification, and lesion detection. In terms of segmentation, unlike traditional semi-automatic segmentation algorithms that need manual tracing or correction, most DL segmentation algorithms perform end-to-end segmentations, which can reduce labor and time cost. The algorithms output an array of the same size as the input images. Pixels (voxels) in the target usually demonstrate higher values in the corresponding location on the output array. By setting a cutoff value, a mask of the target region is then generated. The segmentation algorithms are often evaluated on spatial overlapping by Dice Similarity Coefficient (DSC) and Hausdorff distance, with higher DSC and lower Hausdorff distance indicating better segmentation performance.

In classification tasks, DL algorithms can classify the input image with or without a lesion mask into a specific and output a probability from zero to one. DL classification is usually assessed by discrimination, calibration, and clinical utility. Discrimination refers to the ability to differentiate between those with or without the outcome event, and is usually estimated by area under receiver operating characteristic curve (AUC). Calibration reflects the agreement between model-derived outcomes and actual outcomes, which is preferably analyzed by calibration curve. In addition, the clinical utility is often evaluated by decision

curve analysis, which allows evaluating and comparing prediction models that incorporate clinical consequences.

DL detection task refers to lesion localization and classification. The algorithms output both the lesion location (a precise location array or a rough bounding box) and the type of lesions. Different from image segmentation algorithms, lesion detection algorithms focus more on the identification of real lesions rather than the precise spatial overlapping between the outputs and the ground-truth regions. The results of detection algorithms are divided into true-positive (lesions detected), true-negative (no lesion in the image and no positive output), false-positive (no lesion in the image but a positive output), and false-negative (lesion undetected), and detection performance can be measured by true-positive rate and false-positive rate.

Application in diffuse liver diseases: Perez *et al*^[1] used a CT-based DL segmentation algorithm to automatically segment the liver and identify its volume, aiming to explore a more direct measurement for organ enlargement. They found the automated algorithm-derived liver volume demonstrated a normal distribution and increased linearly with patient weight, which matched well with manual and semi-automated methods in a subset analysis of 189 patients. Their results provided an objective and more accurate assessment of liver size than linear measures with added value in the screening of hepatomegaly.

Martí-Aguado *et al*^[2] conducted a prospective multicenter study to compare automated whole-liver segmentation and manual region of interest (ROI) for proton density fat fraction and iron estimation in MRI for patients with chronic liver disease. Measurements derived from whole-liver segmentation were found accurate for liver steatosis grading, and strongly correlated with pathological fat ratio. Furthermore, the DL method provided similar diagnostic accuracy, as well as demonstrating less variability and time cost than manual segmentation.

Access this article online

Quick Response Code:



Website:

www.cmj.org

DOI:

10.1097/CM9.0000000000002199

Enyu Yuan and Zheng Ye contributed equally to this work.

Correspondence to: Dr. Bin Song, Department of Radiology, West China Hospital, Sichuan University, Chengdu, Sichuan 610041, China
E-Mail: songlab_radiology@163.com

Copyright © 2022 The Chinese Medical Association, produced by Wolters Kluwer, Inc. under the CC-BY-NC-ND license. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Chinese Medical Journal 2022;135(11)

Received: 14-09-2021; Online: 13-07-2022 Edited by: Yuanyuan Ji

In CT images, Yin *et al*^[3] used Gradient-weighted Class Activation Mapping (Grad-CAM) method to provide a visual-based explanation of DL in predicting liver fibrosis. The location map demonstrated the DL algorithm focused more on the liver surface in patients without liver fibrosis (F0), while it had more focus on the parenchyma of the liver and spleen in cases of cirrhosis (F4), which was logical from a clinical point of view in some degree. Nowak *et al*^[4] utilized the same DL technique in MRI, and found the caudate lobe area was important for the DL algorithms in detecting liver cirrhosis. In the future, the Grad-CAM method can be used for the quality control of DL fibrosis stage prediction with prior knowledge of specific location map patterns.

Application in focal liver diseases: Quantitative tumor image analysis begins with accurate lesion segmentation. Conventional radiomics features are extracted from manually drawn ROI, which often leads to segmentation variability. Khan *et al*^[5] proposed a residual U-Net with the dilated convolution and a new loss function based on a combination of DSC and absolute volumetric difference. Their model achieved mean DSC of 91.92% and 86.70% for liver tumor segmentation in 3D Image Reconstruction for Comparison of Algorithm Database and Liver Tumor Segmentation Challenge dataset, respectively.

Based on the images from gadoteric acid-enhanced MRI, Takenaga *et al*^[6] proposed a multichannel three-dimensional fully convolutional residual network for focal liver lesions (FLLs) detection and classification. The model reached a true-positive ratio of 0.6 at an average of 25 false-positives per case and classification accuracy of 0.790. However, the detection accuracy of hemangiomas, one of the most common benign FLLs, was low, mostly owing to their rarity in the dataset. Thus, this algorithm required further improvement to help reduce the workload of radiologists.

In the study by Wang *et al*,^[7] a DL-based hepatocellular carcinoma (HCC) diagnosis system was developed on a training dataset of CT images from 7512 patients. This system firstly predicted the presence or absence of liver nodules, and classified them into HCCs or non-HCCs if presented. The system was validated internally ($n = 385$) and externally ($n = 556$), which achieved good diagnostic performance with AUC values of 0.887 and 0.883 on each dataset. In addition, the diagnostic accuracy of HCC was significantly improved in radiologists with DL assisted.

Liu *et al*^[8] developed a DL model for predicting microvascular invasion (MVI) in patients with HCC. They firstly tested proposed models on 309 patients, and validated them on 164 patients from 54 different hospitals. The model incorporated with DL image features and clinical factors was superior to traditional radiomics models, and achieved the highest AUC values of 0.845 and 0.777 in internal validation and external dataset, respectively. Moreover, by using the Grad-CAM method, the authors found that the DL model identified MVI was similar to the logic of a previously reported image biomarker. This may help to build a more interpretable predictive model for MVI.

Careful surgery planning with accurate segmentation of major vessels is of great importance before hepatectomy and liver transplantation. Kazami *et al*^[9] developed a DL algorithm for fast portal vein and hepatic vein segmentation on CT images. The sensitivity and DSC of the DL algorithm were significantly higher than traditional tracking-based algorithm. This technique may help predict future liver remnant volume and optimize decision-making regarding liver graft selection and venous reconstruction in liver transplantation. However, the feasibility of this DL algorithm on assisting liver malignancy surgery remains further validations.

Another study^[10] developed a DL strategy to optimize the treatment selection in patients with very-early or early-stage HCC by pre-operatively predicting the progression-free survival of radiofrequency ablation and surgical resection. Nomograms incorporated with DL prediction provided good 2-year progression-free survival prediction accuracy and good calibrations. As reported, 17.3% patients treated with radiofrequency ablation and 27.3% patients treated with surgery were identified and suggested to swap their treatment for a higher 2-year progression-free survival. In terms of trans-arterial chemoembolization (TACE) treatment, Peng *et al*^[11] and Liu *et al*^[12] developed DL models for the prediction of treatment response after the first TACE session. Although they used different image modalities, both models reached relatively high AUCs ranging from 0.80 to 0.93 in the validation cohorts.

Moawad *et al*^[13] evaluated the correlation between DL automated volumetric assessment and unidimensional modified Response Evaluation Criteria In Solid Tumors measurements for response evaluation of HCC after TACE treatment. Their study showed a good correlation between unidimensional modified Response Evaluation Criteria In Solid Tumors and the automated volumetric RECIST with a correlation coefficient of 0.774, indicating that DL-based automated volumetric measurements may be good substitutes for manual volumetric measurements.

Plentiful studies witnessed the dramatic potential and clinical value of DL in a variety of liver diseases in recent years. There is no doubt that DL-assisted medical decision-making will become a trend in the coming decades, thanks to its great power in data processing. However, several challenges remain to be solved before the DL technique can be widely applied in clinical practice.

Firstly, as DL algorithms succeed on big data, large datasets on liver images are needed to support its development. Although several public medical imaging datasets were recently established, most of them failed to provide detailed demographic characteristics of patients due to the privacy protocol, and thus cannot be used in some classification tasks.

Secondly, the interpretability of DL framework is to be improved, which may increase the usefulness, reliability, and effectiveness of DL models in clinical environment. Now some researchers have shifted their focus from the development of much more sophisticated DL models to the medical explanation of the DL output.^[3,4,8] New

medical knowledge may be generated from DL algorithms in the future. On the other hand, the integration of prior knowledge when designing the DL architecture can gather more confidence from doctors in machine decisions.

Thirdly, current DL models should be validated externally on a large scale. The algorithm outcomes are reliable only when they can generalize to data in other medical centers, especially in classification models. The sizes of DL models will become larger along with the ever-growing computing power. In this condition, research into interpretability is becoming much harder due to the complexity of models. Multicenter large-scale validation provides a solution in a black-box manner to assess the model's performance. If the model can reach similar satisfactory performance in external validation, doctors can choose to trust this DL model in clinical condition.

Finally, the study reporting guideline and risk of bias tool are urgently needed. Studies with artificial intelligence face some unique challenges when applying current risk of bias assessment tools.^[14] For example, a predictive model usually demonstrated low risk of bias when the outcome events per variable ratio is over 20. However, this part of bias cannot be evaluated in DL models since they do not need predictors when calculating. Furthermore, it is almost impossible to present the whole parameters of a DL model and compare them between different models. Recently, multi-disciplinary experts are working on developing extensions to the report guideline and risk of bias assessment tool for studies based on machine learning techniques, which may help report key details specifically focused in DL, reduce research waste, and evaluate the bias with a robust standardized tool.^[15]

In conclusion, DL shows promising performance in data mining and quantitative image assessment. Recent researches have unveiled the great potential of DL in both diffuse and focal liver diseases [Supplementary Figure 1, <http://links.lww.com/CM9/B129>]. However, current studies mostly focused on the feasibility of DL, and whether these models can be utilized to handle the sophisticated clinical practice remains unknown. More accurate, interpretable, and robust DL models with large-scale validation are warranted before they can be widely accepted for medical use. Reporting guidelines and risk of bias tools are also needed to improve the standardization of proposed algorithms.

Funding

This study was supported by Research Grant of National Nature Science Foundation of China (No. 81971571), Science and Technology Support Program of Sichuan Province (No. 2021YFS0021), and the 1.3.5 Project for Disciplines of Excellence, West China Hospital, Sichuan University (No. ZYJC21012).

Conflicts of interest

None.

References

- Perez AA, Noe-Kim V, Lubner MG, Graffy PM, Garrett JW, Elton DC, *et al.* Deep learning CT-based quantitative visualization tool for liver volume estimation: defining normal and hepatomegaly. *Radiology* 2022;302:336–342. doi: 10.1148/radiol.2021210531.
- Martí-Aguado D, Jiménez-Pastor A, Alberich-Bayarri Á, Rodríguez-Ortega A, Alfaro-Cervello C, Mestre-Alagarda C, *et al.* Automated whole-liver MRI segmentation to assess steatosis and iron quantification in chronic liver disease. *Radiology* 2022;302:345–354. doi: 10.1148/radiol.2021211027.
- Yin Y, Yakar D, Dierckx R, Mouridsen KB, Kwee TC, de Haas RJ. Liver fibrosis staging by deep learning: a visual-based explanation of diagnostic decisions of the model. *Eur Radiol* 2021;31:9620–9627. doi: 10.1007/s00330-021-08046-x.
- Nowak S, Mesropyan N, Faron A, Block W, Reuter M, Attenberger UI, *et al.* Detection of liver cirrhosis in standard T2-weighted MRI using deep transfer learning. *Eur Radiol* 2021;31:8807–8815. doi: 10.1007/s00330-021-07858-1.
- Khan RA, Luo Y, Wu FX. RMS-UNet: residual multi-scale UNet for liver and lesion segmentation. *Artif Intell Med* 2022;124:102231. doi: 10.1016/j.artmed.2021.102231.
- Takenaga T, Hanaoka S, Nomura Y, Nakao T, Shibata H, Miki S, *et al.* Multichannel three-dimensional fully convolutional residual network-based focal liver lesion detection and classification in Gd-EOB-DTPA-enhanced MRI. *Int J Comput Assist Radiol Surg* 2021;16:1527–1536. doi: 10.1007/s11548-021-02416-y.
- Wang M, Fu F, Zheng B, Bai Y, Wu Q, Wu J, *et al.* Development of an AI system for accurately diagnose hepatocellular carcinoma from computed tomography imaging data. *Br J Cancer* 2021;125:1111–1121. doi: 10.1038/s41416-021-01511-w.
- Liu SC, Lai J, Huang JY, Cho CF, Lee PH, Lu MH, *et al.* Predicting microvascular invasion in hepatocellular carcinoma: a deep learning model validated across hospitals. *Cancer Imaging* 2021;21:56. doi: 10.1186/s40644-021-00425-3.
- Kazami Y, Kaneko J, Keshwani D, Takahashi R, Kawaguchi Y, Ichida A, *et al.* Artificial intelligence enhances the accuracy of portal and hepatic vein extraction in computed tomography for virtual hepatectomy. *J Hepatobiliary Pancreat Sci* 2022;29:359–368. doi: 10.1002/jhbp.1080.
- Liu F, Liu D, Wang K, Xie X, Su L, Kuang M, *et al.* Deep Learning Radiomics Based on Contrast-Enhanced Ultrasound Might Optimize Curative Treatments for Very-Early or Early-Stage Hepatocellular Carcinoma Patients. *Liver Cancer* 2020;9:397–413. doi: 10.1159/000505694.
- Peng J, Kang S, Ning Z, Deng H, Shen J, Xu Y, *et al.* Residual convolutional neural network for predicting response of transarterial chemoembolization in hepatocellular carcinoma from CT imaging. *Eur Radiol* 2020;30:413–424. doi: 10.1007/s00330-019-06318-1.
- Liu D, Liu F, Xie X, Su L, Liu M, Xie X, *et al.* Accurate prediction of responses to transarterial chemoembolization for patients with hepatocellular carcinoma by using artificial intelligence in contrast-enhanced ultrasound. *Eur Radiol* 2020;30:2365–2376. doi: 10.1007/s00330-019-06553-6.
- Moawad AW, Fuentes D, Khalaf AM, Blair KJ, Szklaruk J, Qayyum A, *et al.* Feasibility of Automated Volumetric Assessment of Large Hepatocellular Carcinomas' Responses to Transarterial Chemoembolization. *Front Oncol* 2020;10:572. doi: 10.3389/fonc.2020.00572.
- Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, *et al.* PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann Intern Med* 2019;170:51–58. doi: 10.7326/M18-1376.
- Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooff L, Reitsma JB, *et al.* Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ Open* 2021;11:e048008. doi: 10.1136/bmjopen-2020-048008.

How to cite this article: Yuan E, Ye Z, Song B. Imaging-based deep learning in liver diseases. *Chin Med J* 2022;135:1325–1327. doi: 10.1097/CM9.00000000000002199