Journal of Traditional and Complementary Medicine 14 (2024) 544–549

Contents lists available at ScienceDirect

Journal of Traditional and Complementary Medicine

journal homepage: www.elsevier.com/locate/jtcme

Exploring hepatic fibrosis screening via deep learning analysis of tongue images

Xiao-zhou Lu [a,1], Hang-tong Hu [b,1], Wei Li [b], Jin-feng Deng [a], Li-da Chen [b], Mei-qing Cheng [b], Hui Huang [b], Wei-ping Ke [b], Wei Wang [b,**], Bao-guo Sun [a,*]

[a] Department of Traditional Chinese Medicine, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China
[b] Department of Medical Ultrasonics, Institute of Diagnostic and Interventional Ultrasound, MedAI Collaborative Lab, The First Affiliated Hospital of Sun Yat-Sen University, Guangzhou, China

ARTICLE INFO

Keywords:
Chinese medicine
Tongue images
Artificial intelligence
DenseNet
Hepatic fibrosis

ABSTRACT

Background: Tongue inspection, an essential diagnostic method in Traditional Chinese Medicine (TCM), has the potential for early-stage disease screening. This study aimed to evaluate the effectiveness of deep learning-based analysis of tongue images for hepatic fibrosis screening.

Methods: A total of 1083 tongue images were collected from 741 patients and divided into training, validation, and test sets. DenseNet-201, a convolutional neural network, was employed to train the AI model using these tongue images. The predictive performance of AI was assessed and compared with that of FIB-4, using real-time two-dimensional shear wave elastography as the reference standard.

Results: The proposed AI model achieved an accuracy of 0.845 (95% CI: 0.79–0.90) and 0.814 (95% CI: 0.76–0.87) in the validation and test sets, respectively, with negative predictive values (NPVs) exceeding 90% in both sets. The AI model outperformed FIB-4 in all aspects, and when combined with FIB-4, the NPV reached 94.4%.

Conclusion: Tongue inspection, with the assistance of AI, could serve as a first-line screening method for hepatic fibrosis.

1. Introduction

Chronic liver disease (CLD) is a major public health problem, accounting for significant morbidity and mortality worldwide. Cirrhosis is currently the 11th most common cause of death, and liver cancer is the 5th leading cause of cancer-associated death globally.[1] As a pre-stage of cirrhosis, fibrosis is closely related to liver function. If not dealt with promptly, it can progress to cirrhosis and even carcinoma.[2] Thus, one approach to preventing liver-related mortality is to prevent the progression of fibrogenesis. As a result, liver fibrosis screening is critical for evaluating patients with CLD, including clinical diagnosis, monitoring, treatment, and prognosis.

In the quest to screen for hepatic fibrosis, it is clear that non-invasive, repeatable, and cost-effective methods are highly desirable. Currently, non-invasive tests for the assessment of CLD can be classified into blood-based tests (serum markers of fibrosis) and imaging methods (e.g., elastography).[3] Serum biomarkers are effective at ruling out the presence of advanced fibrosis and cirrhosis with high negative predictive values.[4] Elastography techniques provide a quantitative estimate of tissue stiffness and a more accurate result of the fibrosis stage. It has proven to be efficient in assessing significant fibrosis, which allows it to serve as a reference standard in screening for fibrosis. However, most non-invasive tests require professional equipment and inspectors. Also, they may not be available in primary care facilities, which becomes an obstacle in the screening process.

Traditional Chinese Medicine (TCM) has provided four important diagnostic methods: inspection, listening and smelling, inquiring, and taking the pulse. Tongue diagnosis is one of the inspection methods to use when a TCM practitioner differentiates the syndromes, makes diagnoses, and delivers prescriptions. An important theory in TCM is that

Peer review under responsibility of The Center for Food and Biomolecules, National Taiwan University.
* Corresponding author.
** Corresponding author.
E-mail addresses: wangw73@mail.sysu.edu.cn (W. Wang), sunbaog@mail.sysu.edu.cn (B.-g. Sun).
[1] These authors have contributed equally to this work.

https://doi.org/10.1016/j.jtcme.2024.03.010
Received 29 July 2023; Received in revised form 1 February 2024; Accepted 5 March 2024
Available online 6 March 2024

the tongue's appearance reflects the conditions of the organs and the severity of illnesses. A study[5] found that some tongue features strongly correlated with aminotransferases, suggesting the possible use of tongue features to provide an early warning of liver diseases. Tongue inspection could be easily and quickly done and barely cost anything compared to serology tests and imaging examinations. With certain features, a tongue could provide much information about the state of the human body. Thus, tongue inspection is a potential screening method for CLD.

In recent years, there have been increasing studies about applying Artificial Intelligence (AI) using tongue images to predict and identify diseases. For example, Jiang et al.[6] used convolutional neural networks to recognize and classify tongue images, revealing that fissured tongue and toothmarks were closely related to hypertension, dyslipidemia, and nonalcoholic fatty liver disease. Since deep learning models can learn the most predictive features directly from raw image pixels,[7,8] we believe that with the assistance of AI, tongue images could offer important clues for diseases.[9] Previous studies have shown that tongue diagnosis could serve as a primary screening tool in the early detection of several diseases, such as diabetes,[10] breast cancer,[11] and nonalcoholic fatty liver disease (NAFLD).[12]

In this study, we aimed to construct a deep-learning model based on tongue images to screen hepatic fibrosis. The predictive performance of AI was assessed and compared with that of FIB-4, an indirect biomarker panel based on four factors, including age, aspartate aminotransferase (AST), alanine aminotransferase (ALT), and platelet.[13]

2. Methods

2.1. Participants

This prospective study was approved by the ICE for Clinical Research and Animal Trials of the First Affiliated Hospital of Sun Yat-sen University (No.2021464), and the informed consent forms were obtained from the patients.

We recruited patients diagnosed with CLD in the Gastroenterology Department of the First Affiliated Hospital of Sun Yat-sen University. From April 2021 to February 2022, we collected 1083 tongue images from 741 patients after eliminating ambiguous or low-quality images.

2.2. Tongue images collection

Patients were required to have fasted for 4 h before the ultrasound examination. Meanwhile, they were told to avoid colored drinks ahead in case of stains left on their tongues. After accepting ultrasound elastography, patients extended and stretched their tongues naturally outside for the tongue images. Photos were all captured under the same conditions (in terms of brightness, a distance of 15 cm, and an angle of 45° between the camera and tongue surface) with the same camera (Sony DSC-RX100, no flash mode). Patients should stretch out their tongues for no more than 5 s. Patients were asked to rest for a few seconds between the shots if more than one picture was needed.

2.3. Reference standard: ultrasound elastography examination

Real-time two-dimensional shear wave elastography (2D SWE) was performed on all the patients by an experienced doctor (more than ten years of working experience) before the acquisition of tongue images. The patients lay supine, and the scans were performed through the right intercostal region. Patients were told to hold the breadth for 4–5 s when liver stiffness was measured. A trapezoidal color box (3.5 cm × 2.5 cm) was positioned in the liver parenchyma and acquired the elasticity signals. A round ROI was located in a homogenous elastogram in the liver parenchyma, where large vessels or hepatic nodules were avoided. Measured elasticity values were expressed in kilopascal (kPa). In this study, ultrasound elastography was the reference standard method to detect the degree of fibrosis in the patients. The cutoff value of 7[14] would divide the patients into two groups, which were the "fibrosis group" and the "non-fibrosis group".

2.4. Data preparation

2.4.1. Image preprocessing

The 1083 tongue pictures were divided into training, validation, and test sets by 6.5: 1.5: 2. Firstly, photos in the training set were randomly cropped to remove unnecessary information. One patient's tongue picture could be derived into several images, and we only selected the images in which the tongue could be entirely seen. Secondly, data augmentation techniques, such as rotation, flipping, and scaling, were applied to enhance the model's ability to generalize and recognize patterns across diverse instances. After the augmentation, 13,381 images were generated for AI training.

The numbers of images in the validation and the test sets were 167 and 209, which meant there was only one picture for each patient. Our training set consisted of 8847 tongue images from non-fibrosis patients and 4534 from fibrosis patients. Meanwhile, the validation and test set contained 112 and 142 tongue images from non-fibrosis patients, and 55 and 68 images from fibrosis patients.

3. The training process

3.1. Network architectures

The deep learning structure for the classification task employed in the study was DenseNet-201. In traditional Chinese medicine, tongue diagnosis is a complex system where the tongue's shape, color, coating, and other features reflect visceral functions through intricate relationships. The intricate nature of these features necessitates a sophisticated approach to capture and analyze the subtle details in tongue images effectively. In pre-experiments, we found that DenseNet-201, distinguished by its densely interconnected layers,[15] outperformed alternative networks, such as ResNet, VGG, Inception, etc. It offered significant advantages, including enhanced feature reuse and a reduced parameter count, which were crucial for efficiently processing the intricate features in tongue images. Consequently, we considered DenseNet a suitable model for our study.

In our study, the DenseNet-201 comprised four dense blocks, three transition layers, and a global average pooling layer. Each dense block included a batch normalization (BN), followed by a rectified linear unit (ReLU) and a 3 × 3 convolution. The transition layers, layers between blocks, consisted of a BN layer and a 1 × 1 convolutional layer followed by a 2 × 2 average pooling layer. The growth rate of the network was $k = 32$.[16]

3.2. Training protocol

The model leverages the comprehensive "densenet" package integrated within "torchvision", specifically opting for the "densenet201" module. Due to the scarcity of training data, a transfer learning approach was employed to capitalize on the knowledge acquired by the pre-existing parts of the network. The "pretrained = False" setting was specified, prompting the download of the pre-trained model from "https://download.pytorch.org/models/densenet201-c1103571.pth". Subsequent model training was conducted based on this pre-trained foundation. The input images were resized to a resolution of 224 × 224 pixels to standardize the input format. After the input, we trained the model for 300 epochs with a batch size of 30 for each training iteration. To optimize the learning process, an initial learning rate of 0.001 was employed, and a weight decay of 5×10^4 was implemented to facilitate gradual learning rate reduction. The model aimed to predict the patient's diagnosis as either "fibrosis" or "non-fibrosis" based on the features extracted from the input images.

The training was performed on a workstation with an NVIDIA A100

Tensor Core GPU, a Core i7-6700 K (Intel) central processing unit, and 64 GB of random-access memory. Python 3.5 (https://www.python.org) and the Torch (http://torch.ch) framework for neural networks were used for this purpose. Augmentation was performed using the Python imaging library of Pillow 3.3.1 (https://pypi.python. org/pypi/Pillow/3.3.1).

4. Artificial intelligence performance analysis

4.1. Performance of the AI model versus FIB-4

By applying the AI to the test set, each case was evaluated by the same input method used in the training process. The output was presented as the corresponding diagnosis of fibrosis or non-fibrosis. On the other hand, we predicted each patient's diagnosis in the test set by calculating their FIB-4 score and observed FIB-4's diagnostic performance before comparing it with that of the AI model.

FIB-4 was first developed to access liver fibrosis in HIV/hepatitis C virus (HCV) coinfection and was calculated by the formula of **(age [years] × AST [U/L]) / ((PLT [10^9/L]) × (ALT [U/L])$^{1/2}$)**.[17] Previous studies have proved that FIB-4 performed better in classifying different stages of liver fibrosis when compared to the AST-to Platelet ratio index (APRI), AST-to-ALT ratio index, and age-spleen platelet ratio index [18,19]. After calculating the FIB-4 scores of the patients in the test set, we put them into three groups with cut-off values of 1.45 and 3.25 (group A: FIB-4<1.45; group B: 1.45≤FIB-4<3.25; group C: FIB-4≥3.25), which were initially defined by Sterling et al. [17] to rule out and rule in advanced fibrosis.

4.2. Heatmap of the model's attention

The Gradient-weighted Class Activation Mapping (Grad-CAM) technique was employed to produce visual explanations for the decisions from the system by using gradient information flowing into the last convolutional layer of the CNN to generate a localization map highlighting the critical regions in the image for predicting the concept.[20] With Grad-CAM, the heatmap was generated to interpret the AI model's attention when discriminating between fibrosis and non-fibrosis.

5. Statistical methods

Descriptive statistics were summarized as the mean ± standard deviation (SD) or median and interquartile range. Continuous variables were compared by the *t*-test or Mann–Whitney *U* test, and categorical variables were compared using the $\chi 2$ test. The performances of the AI model and FIB-4 were mainly evaluated in terms of the AUC (Area Under Curve), accuracy (ACC), sensitivity (Se), specificity (Sp), positive predictive value (PPV), negative predictive value (NPV), and error rates. R software (version 3.4.1; https://www.r-project.org) was used for statistical analysis. Results with two-sided *P*-values of less than 0.05 indicated a statistically significant difference.

6. Result

6.1. Clinical characteristics

This study included 741 patients, of whom 249 were females (mean age, 46.3 ± 13.0 years old), and 492 were males (mean age, 43.6 ± 12.6). According to the reference standard, patients were categorized into a fibrosis group (training: 236 patients; validation: 55 patients; test: 68 patients) and a non-fibrosis group (training: 471 patients; validation: 112 patients; test: 142 patients). The performance of AI and FIB-4 were compared on the test set. In the test set, 174 patients had undergone serum tests right before participating in the study, of whom 115 were diagnosed with non-fibrosis and 59 with fibrosis. The clinical characteristics of the 174 patients are listed in Table 1.

6.2. Diagnostic performance of the AI model

The AI model achieved an ACC of 0.845 (95% CI: 0.79–0.90) in the validation set and 0.814 (95% CI: 0.76–0.87) in the test set. The validation and test sets achieved a considerably high NPV of 0.907 (95% CI: 0.85–0.96) and 0.906 (95% CI: 0.86–0.96), respectively. The performance measurements of both validation and test sets are shown in Table 2, and their receiver operating characteristic curves (ROC curves) are shown in Fig. 1.

6.3. Performance comparison between the AI model and FIB-4

We evaluated the serological tests of 174 patients in the test set and determined their FIB-4 scores based on laboratory results. Of these patients, 104 patients had FIB-4 scores less than 1.45, indicating METAVIR F0–F1 stage, and were labeled as non-fibrosis. And the rest of the 70 patients who got FIB-4 scores equal to or greater than 1.45 were labeled as fibrosis.

Comparing the FIB-4 label with the corresponding patient's actual diagnosis, the accuracy, sensitivity, specificity, PPV, and NPV of FIB-4 in the test set were 71.8%, 67.8%, 73.9%, 57.1%, and 81.7%, which were inferior to that of the AI model. When combining the AI model and FIB-4, we got a higher NPV than both the AI model and FIB-4 respectively. Details are shown in Table 3.

A previous study showed that patients with FIB-4≥1.45 were classified as having significant fibrosis.[21] And since FIB-4 serves as a screening tool with high NPVs in ruling out fibrosis,[13] here we propose a pathway consisting of a 2-step non-invasive assessment starting with tongue image analysis followed by FIB-4. On the basis of test set data, of 174 patients who received tongue image analysis, 102 (58.6%) patients were predicted as negative and recommended staying in the primary care setting. The rest 72 patients then received serological tests, resulting in 32 (18.4%) patients with FIB-4<1.45 and 40 (23%) patients with FIB-4≥1.45. According to Fig. 2, which displays the flowchart of the 2-step diagnostic pathway, 77% of patients are deemed at low risk of advanced fibrosis, while the remaining 23% are referred to liver specialists.

To observe the AI model's performance in different situations, we built three confusion matrices for Group A, B, and C. In Group A, when FIB-4 was less than 1.45, the AI model correctly labeled more than half of the images as "non-fibrosis" (68 out of 104). Conversely, in Group C (FIB-4≥3.25), 16 out of 20 images were identified as "true fibrosis" by the AI model. Fig. 3 below provides an intuitive illustration of the AI's performance.

6.4. Interpretability of the AI model

To visualize and interpret the model predictions, we generated

Table 1
Characteristics of the test set.

Characteristic	Fibrosis (N = 59)	Non-fibrosis (N = 115)	*p* value
Age (y)[a]	50.3 ± 13.2	43.0 ± 12.9	<0.001
Gender	46 (78.0%)	60 (52.2%)	<0.05
Male	13 (22.0%)	55 (47.8%)	
Female			
Laboratory tests			
Alanine aminotransferase (U/L)[a]	44.7 ± 49.2	31.8 ± 29.4	<0.05
Aspartate aminotransferase (U/L)[a]	45.2 ± 45.4	29.9 ± 16.9	<0.05
Platelet count (× 10^9/L)[a]	165.7 ± 72.3	227.0 ± 55.8	<0.001

[a] Data are means ± standard deviations.

Table 2
Performance measurements of the validation and test sets.

	ACC	Se	Sp	PPV	NPV
Validation	0.845 (0.791, 0.900)	0.818 (0.716, 0.920)	0.858 (0.794, 0.923)	0.738 (0.627, 0.848)	0.907 (0.851, 0.962)
Test	0.814 (0.762, 0.867)	0.824 (0.733, 0.914)	0.810 (0.745, 0.874)	0.675 (0.574, 0.775)	0.906 (0.855, 0.956)

ACC, accuracy; PPV, positive predictive value; NPV, negative predictive value; Se, sensitivity; Sp, specificity.

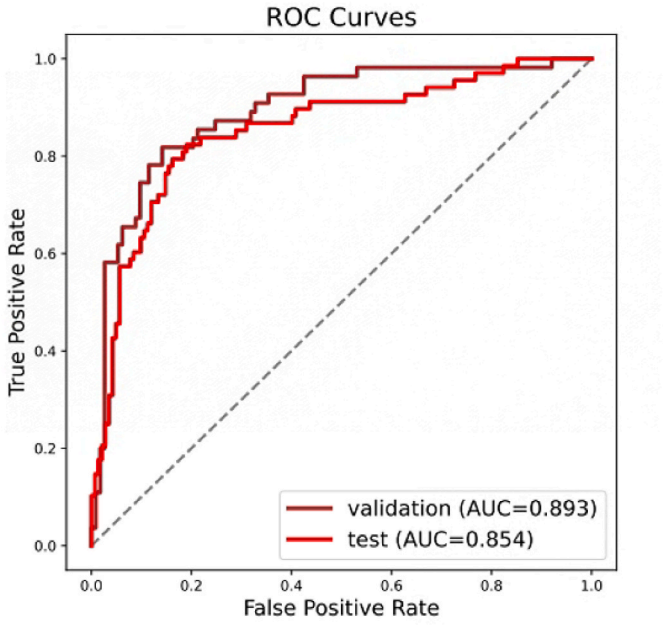

Fig. 1. Receiver operating characteristic curves of the validation group (brown curve) and the test group (red curve). The areas under the curves are 0.893 and 0.854, respectively.

Table 3
Comparisons among the AI model, FIB-4, and the combination.

	AI model	FIB-4	Combined
Accuracy (%)	81.4	71.8	79.8
Sensitivity (%)	82.4	67.8	78.9
Specificity (%)	81.0	73.9	80.0
PPV (%)	67.5	57.1	46.9
NPV (%)	90.6	81.7	94.4

PPV, positive predictive value; NPV, negative predictive value.

attention heatmaps, as shown in Fig. 4. By analyzing the heatmaps along with the corresponding images, we observed that the AI model focused on the lateral sides of the tongue in cases that were diagnosed with fibrosis. These regions were particularly prominent in the heatmaps, indicating their importance in the model's decision-making process. The intensity of the color in the heatmap corresponds to the degree of contribution of the corresponding region to the recognition of liver fibrosis. According to TCM theory, different tongue regions represent different organs in the body. Fig. 5 shows the subdivisions of the tongue, including spleen-stomach on the middle part, liver-gallbladder on the lateral sides, heart-lung on the tip, and kidney on the root.[11]

7. Discussion

In this study, we applied a deep learning approach to analyze tongue images and established a classification task to validate its performance of screening for hepatic fibrosis. Our analysis demonstrates that tongue images could be of great value in estimating liver fibrosis and provide vital information for physicians with the assistance of AI. To the best of our knowledge, no previous works have applied deep-learning models to predict hepatic fibrosis using only tongue images.

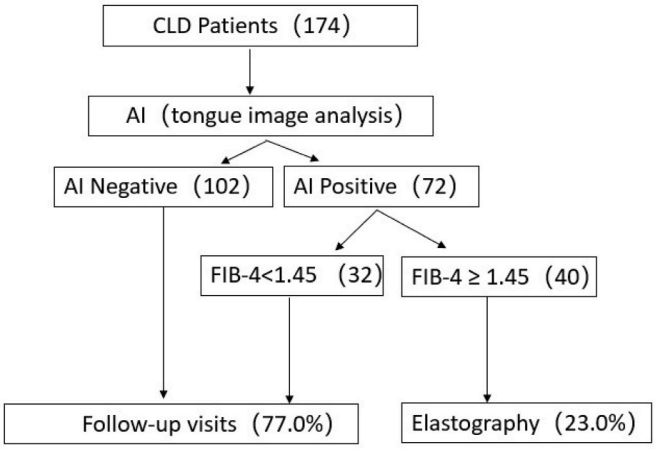

Fig. 2. Diagnostic pathway consisted of a 2-step assessment.

Traditional tongue inspection inevitably involves the observer's subjectivity, which could lead to bias in diagnosis.[22] With the development of AI, such a problem seems to be potentially solvable. In this study, our proposed AI model achieved an ACC of 84.5% and 81.4% in the validation and test sets, respectively, while NPVs were over 90% in both sets. Compared with FIB-4, our model exhibited better performance in all aspects. Furthermore, when combining the AI model and FIB-4, NPV could reach 94.4%, exceeding any of the two.

A model with high NPV has the advantage of excluding diseases in the screening process. Following the 2-step assessment we proposed resulted in a 77% reduction of unnecessary referrals. It suggests that when integrating the method into clinical practice, the majority of referrals made to hepatologists or tertiary hospitals could have been managed in primary care, which has the benefits of easing patients' economic and emotional burdens, relieving the pressure on secondary or tertiary care services, and reducing the costs for the healthcare system. In addition, it's possible to install the deep learning model into smartphone applications, enabling patients to capture pictures of their tongue and make a preliminary diagnosis even before going out to seek physicians. All in all, we believe promoting this method in clinical practice is beneficial to facilitating the screening process for hepatic fibrosis.

It should be noted that there were 24 false positives (13.8%) and 11 false negatives (6.3%) when the AI model made the diagnosis prediction. Among those falsely predicted to be fibrosis by the AI model, more than half were middle-aged with a rather long history of hepatitis. Their ALT and AST were normal, as well as the liver elasticity measurement (LSM). However, we could not rule out the possibility that dysfunction of the qi and blood had happened and manifested on the tongue surfaces. Therefore, we consider that even if the LSM is within normal range, it should not be treated lightly for patients with chronic hepatitis history. For the false negative patients, their aminotransferase values were relatively higher, which indicated their livers were under an inflammatory state. In this situation, the result of ultrasound elastography had the possibility of being false positive on the contrary.

Deep learning is often regarded as a black box. However, with the Grad-CAM technique, the AI model's decision-making is more transparent and explainable. The TCM theory emphasizes the unity of the human body and demonstrates that the tongue is connected to the internal organs through meridians. The condition of *qi* and blood of the organs are usually manifested in the changes of the tongue body and

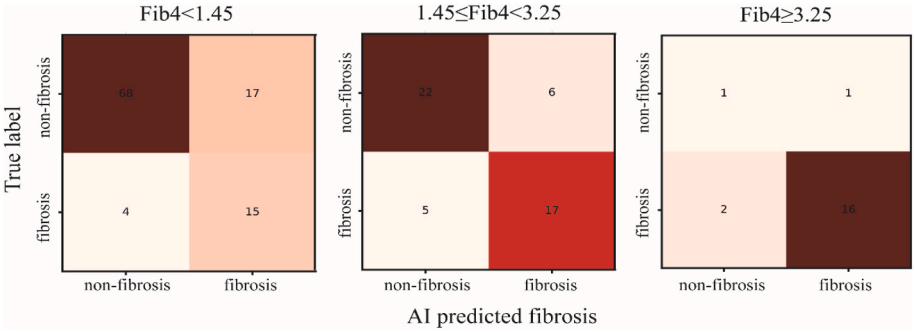

Fig. 3. The confusion matrices of Group A (FIB-4<1.45), B (1.45 ≤ FIB-4<3.25), and C (FIB-4≥3.25).

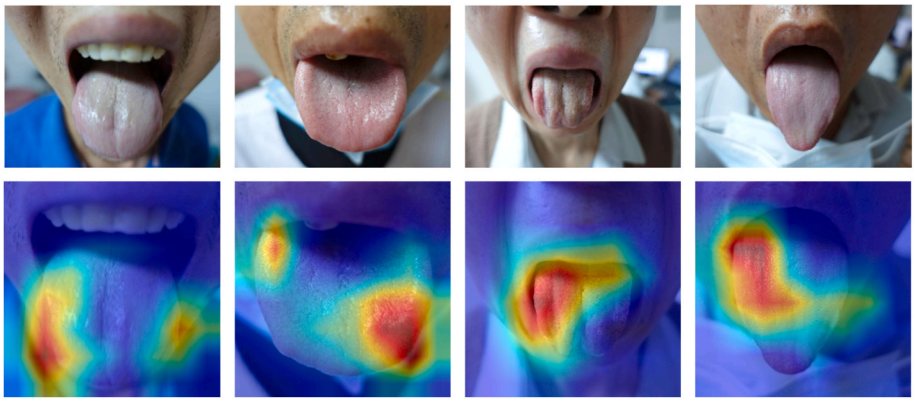

Fig. 4. Images of patients' tongues (fibrosis group), and the corresponding heatmaps.

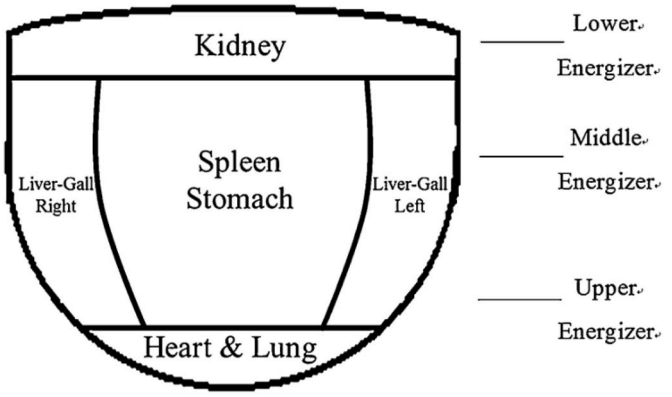

Fig. 5. The tongue is sub-divided into areas corresponding to different internal organs.

tongue coating. The heatmap revealed that the lateral sides of the tongue body were identified as the critical regions for the AI model when discriminating between fibrosis and non-fibrosis. The lateral sides of the tongue represent the liver and gallbladder, indicating that the model's result conforms to the TCM tongue subdivision theory and provides objective evidence of such theory.

This study has several limitations. First, in consideration of the feasibility of the study, we selected ultrasound elastography as the reference standard when analyzing the performance of both the proposed AI model and FIB-4. Ultrasound elastography has a high performance in diagnosing and staging hepatic fibrosis. However, the imaging results could be affected by the operator's experience and the patient's condition (e.g., obesity, liver function), leading to biased results to some degree. Magnetic resonance elastography (MRE), which performs better for the earlier fibrosis stage, could be considered a substitute to improve reliability. Second, even though a tongue image could potentially screen hepatic fibrosis with the help of AI, the critical features that determine the result are still unclear. Images were labeled at the pixel level for deep neural networks. Still, details of tongue information, such as the shape, the color, and the tongue coating, were unknown when the AI model classified the tongue images. Exploring the characteristics of the tongues that assist in the diagnosis of hepatic fibrosis is the direction of future research. Lastly, the deep learning methods are data-driven, meaning the robustness of the models is closely related to the amount and diversity of training data. Future studies using a larger dataset that includes data from multiple centers for training may improve the AI model's performance.

Summary

In summary, we developed an AI model based on tongue images for differentiating between fibrosis and non-fibrosis, which outperformed the FIB-4. By combining our AI model with the FIB-4, we proposed a clinically applicable screening strategy for hepatic fibrosis that could potentially reduce unnecessary referrals in primary care.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by the Institutional Review Board of the First Affiliated Hospital of Sun Yat-Sen University. The patients/participants provided their written informed consent to participate in this study.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This study was supported by National Nature Science Foundation of China (No.82272076 and No.82202156), Guangdong Regional Joint Foundation (No. 2021B1515120030), and the Kelin Outstanding Young Scientist of the First Affiliated Hospital of Sun Yat-sen University (No. R08029).

References

1. Asrani SK, Devarbhavi H, Eaton J, Kamath PS. Burden of liver diseases in the world. *J Hepatol*. 2019;70:151–171.
2. Iredale JP. Models of liver fibrosis: exploring the dynamic nature of inflammation and repair in a solid organ. *J Clin Invest*. 2007;117:539–548.
3. Archer AJ, Belfield KJ, Orr JG, Gordon FH, Abeysekera KWM. EASL clinical practice guidelines: non-invasive liver tests for evaluation of liver disease severity and prognosis. *Frontline Gastroenterol*. 2022;13:436–439.
4. Anstee QM, Castera L, Loomba R. Impact of non-invasive biomarkers on hepatology practice: past, present and future. *J Hepatol*. 2022;76:1362–1378.
5. Hu MC, Lan KC, Fang WC, et al. Automated tongue diagnosis on the smartphone and its applications. *Comput Methods Progr Biomed*. 2019;174:51–64.
6. Jiang T, Lu Z, Hu XJ, et al. Deep learning multi-label tongue image analysis and its application in a population undergoing routine medical checkup. *Evid base Compl Alternative Med*. 2022, 2022.
7. LeCun Y, Bengio Y, Hinton G. *Deep learning, Nature*. 2015;521:436–444.
8. Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60–88.
9. Yuan L, Yang L, Zhang S, et al. Development of a tongue image based machine learning tool for the diagnosis of gastric cancer a prospective multicentre clinical cohort study. *Clin Med*. 2023;57, 101834.
10. Li J, Yuan P, Hu XJ, et al. A tongue features fusion approach to predicting prediabetes and diabetes with machine learning. *J Biomed Inf*. 2021;115.
11. Lo LC, Cheng TL, Chen YJ, Natsagdorj S, Chiang JY. TCM tongue diagnosis index of early-stage breast cancer. *Compl Ther Med*. 2015;23:705–713.
12. Jiang T, Guo XJ, Tu LP, et al. Application of computer tongue image analysis technology in the diagnosis of NAFLD. *Comput Biol Med*. 2021;135, 104622.
13. Xu XL, Jiang LS, Wu CS, et al. The role of fibrosis index FIB-4 in predicting liver fibrosis stage and clinical prognosis: a diagnostic or screening tool? *J Formos Med Assoc*. 2022;121:454–466.
14. Ferraioli G, Tinelli C, Dal Bello B, Zicchetti M, Filice G, Filice GC. Liver fibrosis study, accuracy of real-time shear wave elastography for assessing liver fibrosis in chronic hepatitis C: a pilot study. *Hepatology*. 2012;56:2125–2133.
15. Yilmaz F, Kose O, Demir A. *Ieee, Comparison of Two Different Deep Learning Architectures on Breast Cancer, Medical Technologies Congress*. TURKEY: TIPTEKNO) Izmir; 2019:521–524.
16. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. *IEEE Conference on Computer Vision and Pattern Recognition*. 2017:2261–2269. CVPR), 2017.
17. Sterling RK, Lissen E, Clumeck N, et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology*. 2006;43:1317–1325.
18. Kim BK, Kim DY, Park JY, et al. Validation of FIB-4 and comparison with other simple noninvasive indices for predicting liver fibrosis and cirrhosis in hepatitis B virus-infected patients. *Liver Int*. 2010;30:546–553.
19. Castera L, Friedrich-Rust M, Loomba R. Noninvasive assessment of liver disease in patients with nonalcoholic fatty liver disease. *Gastroenterology*. 2019;156:1264–+.
20. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. *Ieee, Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization, 16th IEEE International Conference on Computer Vision*. ITALY: ICCV)Venice; 2017:618–626.
21. Li Y, Cai Q, Zhang YF, et al. Development of algorithms based on serum markers and transient elastography for detecting significant fibrosis and cirrhosis in chronic hepatitis B patients: significant reduction in liver biopsy. *Hepatol Res*. 2016;46:1367–1379.
22. Wang X, Liu JW, Wu CY, et al. Artificial intelligence in tongue diagnosis: using deep convolutional neural network for recognizing unhealthy tongue with tooth-mark. *Comput Struct Biotechnol J*. 2020;18:973–980.