Journal of Clinical and Translational Hepatology **2026** vol. 14(4) | 416–429
DOI: 10.14218/JCTH.2025.00631

Original Article

Tongue Image Analysis and Clinical Data Fusion: A Novel Approach for Non-invasive Diagnosis of Metabolic Dysfunction-associated Fatty Liver Disease

Chen-Xia Lu[1,2#], Chuan-Xi Tian[2#], Yi-Bo Jiao[3], Hui Zhu[1,4,5], Hai-Yan Yu[3], Zi-Xin Shu[1,4,5], Ling-Han Zhang[6], Jia Zhang[1,4,5], Lan Wang[3], Qi Hao[1,4,5], Wen-Bin Zou[6], Ming-Zhong Xiao[1,4,5], Cheng-Hai Liu[1,4,5], Qiu-Yang He[6], Bee Luan Khoo[6,7,8*] and Xiao-Dong Li[1,4,5*]

[1]Institute of Liver Diseases, Hubei Provincial Hospital of Traditional Chinese Medicine, Wuhan, Hubei, China; [2]Guang'an men Hospital, China Academy of Chinese Medical Sciences, Beijing, China; [3]University of Electronic Science and Technology of China, Chengdu, Sichuan, China; [4]Hubei Provincial Key Laboratory of Traditional Chinese Medicine for Kidney and Liver Diseases, Wuhan, Hubei, China; [5]Hubei Shizhen Laboratory, Wuhan, Hubei, China; [6]Department of Biomedical Engineering, City University of Hong Kong, Hong Kong, China; [7]Institute of Digital Medicine, City University of Hong Kong, Hong Kong, China; [8]City University of Hong Kong Futian-Shenzhen Research Institute, Shenzhen, China

Received: November 24, 2025 | **Revised:** January 20, 2026 | **Accepted:** March 02, 2026 | **Published online:** April 08, 2026

Abstract

Background and Aims: Metabolic dysfunction-associated fatty liver disease (MAFLD) represents a predominant cause of chronic liver disease, underscoring the demand for accessible, non-invasive diagnostic tools. Tongue diagnosis in Traditional Chinese Medicine provides a distinctive perspective on systemic health, though it remains largely subjective. This study aimed to develop an interpretable multimodal deep learning model for MAFLD screening by integrating quantitative tongue image features with routine clinical data. ***Methods:*** From 904 screened candidates, 477 subjects (157 healthy, 320 MAFLD) were included and randomly allocated to training, validation, and test sets in an 8:1:1 ratio. All participants underwent standardized tongue imaging (International Commission on Illumination *L*a*b* color features) and comprehensive clinical evaluation. We constructed a dual-stream deep learning model, combining a ConvNeXt-Tiny network for tongue images and a multilayer perceptron for clinical variables. Feature fusion was achieved via a Dynamic Affine Feature Transformation module, and the model was trained using weighted cross-entropy loss. ***Results:*** MAFLD patients showed significant metabolic abnormalities compared to healthy controls. A progressive decrease in tongue yellowness (b* value) was observed with advancing fibrosis. On an independent test set (n = 48), the multimodal model achieved 97.92% accuracy, Quadratic Weighted Kappa of 0.9538, and 96.88% sensitivity, and 100% specificity, outperforming single-modality and serological models. Interpretability analyses confirmed the model's focus on clinically relevant tongue regions and key metabolic drivers. ***Conclusions:*** We developed an accurate and interpretable multimodal model that synergizes tongue image features with metabolic indicators for MAFLD screening. This approach presents a promising, low-cost tool potentially well-suited for resource-limited settings.

Citation of this article: Lu CX, Tian CX, Jiao YB, Zhu H, Yu HY, Shu ZX, *et al*. Tongue Image Analysis and Clinical Data Fusion: A Novel Approach for Non-invasive Diagnosis of Metabolic Dysfunction-associated Fatty Liver Disease. J Clin Transl Hepatol 2026;14(4):416–429. doi: 10.14218/JCTH.2025.00631.

Introduction

Metabolic dysfunction-associated fatty liver disease (MAFLD) has emerged as the predominant hepatic manifestation of systemic metabolic dysregulation, intricately linked to obesity and metabolic syndrome (MetS).[1] It represents a critical and growing global public health challenge. The prevalence of MAFLD is rising at an alarming rate worldwide, a trajectory that closely parallels the concurrent epidemics of obesity, MetS, and type 2 diabetes mellitus (T2DM).[2] Current projections estimate that by 2040, over 55% of the global adult population could be affected.[3] This trend is particularly pronounced in China, where recent epidemiological studies report an adult prevalence as high as 44.39%,[4] underscoring the urgency of addressing this condition within national and global health agendas.

The clinical significance of MAFLD extends far beyond the spectrum of progressive liver disease, including steatohepatitis, fibrosis, cirrhosis, and hepatocellular carcinoma. It is

Keywords: Metabolic dysfunction-associated fatty liver disease; Tongue image analysis; Non-invasive prediction; Multi-modal data fusion; Deep learning; ConvNeXt-Tiny network; Non-invasive Diagnosis.

[#]Contributed equally to this work.

[*]**Correspondence to:** Xiao-Dong Li, Hubei Provincial Hospital of Traditional Chinese Medicine, 4 Huayuan Shan Road, Wuchang District, Wuhan, Hubei 430061, China. ORCID: https://orcid.org/0000-0002-6406-9416. Tel: +86-13908658127, Fax: +86-27-88844689, E-mail: lixiaodong555@126.com; Bee Luan Khoo, Institute of Digital Medicine, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon Tong, Hong Kong 999077, China. ORCID: https://orcid.org/0000-0003-1100-9994. Tel/Fax: +852-34429423, E-mail: blkhoo@cityu.edu.hk.

now recognized as a multisystem disorder that engages in a complex, bidirectional interplay with various metabolic aberrations. This synergy significantly amplifies the risk of severe extrahepatic complications.[5] Compelling evidence indicates that the confluence of MAFLD with MetS and T2DM can escalate the risk of hepatocellular carcinoma by up to five-fold.[6] While the recent advent of effective pharmacotherapy underscores the need for early detection, the current diagnostic paradigm remains suboptimal. Therefore, the timely diagnosis and effective monitoring of MAFLD are paramount not only for mitigating hepatic outcomes but also for the primary and secondary prevention of life-threatening cardiovascular and malignant diseases.

In response to this clinical imperative, the diagnostic paradigm for MAFLD is shifting from reliance on invasive liver biopsy toward the use of non-invasive tests (NITs). A range of NITs, which integrate imaging modalities, clinical parameters, and serum biomarkers into predictive models, are increasingly employed to assess disease severity and stratify patients according to their risk of liver-related events.[7–9] However, while tools like the Enhanced Liver Fibrosis test have gained regulatory approval for prognostic staging,[10] a substantial unmet clinical need remains. There is a critical shortage of validated, accessible, and accurate NITs for other essential clinical applications in MAFLD, including early detection in primary care, precise phenotypic differentiation, and guidance for personalized management strategies. This gap underscores the necessity for developing novel, cost-effective, and clinically integrative diagnostic approaches.

Concurrently, there is a resurgence of interest in tongue diagnosis, a cornerstone of traditional Chinese medicine (TCM), now augmented by modern technology.[11] Recent studies have indicated that characteristics such as the microbial composition of tongue coating, tongue color, and tongue morphology are associated with systemic metabolism and circulatory disorders.[12–14] Advances in digital imaging and computational analysis have modernized this practice, transforming it from a subjective art into an objective, quantifiable discipline. Standardized image acquisition and automated feature extraction techniques have minimized environmental bias and inter-observer variability. Consequently, objective tongue diagnostic platforms have shown promising utility in screening various hepatic conditions, including viral hepatitis, cirrhosis, and MAFLD, demonstrating enhanced diagnostic reproducibility.

Recent studies leveraging machine learning to analyze quantitative features from tongue images have reported encouraging accuracy in identifying MAFLD.[15–17] Nonetheless, a significant translational chasm remains. Most existing computational models operate as "black boxes", relying on high-dimensional optical data that are decoupled from the clinically interpretable visual features used by physicians. This lack of interpretability severely limits the practical integration of tongue diagnosis into routine metabolic assessment workflows and hinders clinician trust and adoption.

To bridge this gap, we hypothesize that a clinically intuitive framework, which directly incorporates visual tongue characteristics understood by practitioners, can enhance diagnostic utility. This study therefore proposes to develop and validate a clinician-oriented, multimodal fusion model. This model will uniquely integrate intuitively discernible visual tongue features with a panel of readily accessible clinical metabolic indicators. We aimed to create a tool that not only improves the accuracy and interpretability of MAFLD risk assessment but also serves to elucidate the connections between external clinical signs and the internal multisystem metabolic dysregulation that defines MAFLD. By doing so, this work sought to translate a traditional diagnostic method into a validated, modern tool for stratified hepatology care.

Methods

Study design

This prospective, observational, single-center cohort study was conducted in China and utilized data from the cohort "A prospective cohort study of real-world clinical diagnosis and treatment of MAFLD", which is registered with the Chinese Clinical Trial Registry (ChiCTR: https://www.chictr.org.cn; No. ChiCTR2200063127).

Ethical statement

The study protocol adhered strictly to the ethical principles of the Declaration of Helsinki and the relevant regulations of China's "Ethical Review Measures for Biomedical Research Involving Humans." It was approved by the Ethics Committee of the Hubei Provincial Hospital of Traditional Chinese Medicine, the lead institution (Approval No. HBZY2022-C08-01). All participants were thoroughly informed about the study's purpose, procedures, potential risks, and benefits. Ample time was provided for consideration, and written informed consent was obtained from each subject prior to enrollment.

Study population

Inclusion and exclusion criteria: Participants were required to meet all the following criteria: (1) Aged between 18 and 75 years, regardless of gender; (2) Diagnosed with MAFLD based on FibroTouch® transient elastography, defined by evidence of hepatic steatosis: Controlled Attenuation Parameter ≥ 245 dB/m, and meeting at least one of the following three criteria, as per the International Expert Consensus on the New Definition of MAFLD (2020)[18]; (3) For the healthy control group: absence of MAFLD diagnostic components, such as overweight/obesity, T2DM, abnormal blood pressure, dyslipidemia (including triglycerides and total cholesterol), and insulin resistance; (4) Ability to understand and willingness to comply with the study protocol, and provision of voluntary written informed consent; (5) Ability to cooperate with tongue image acquisition, anthropometric measurements, questionnaire survey, and blood sample collection. Participants were excluded if they met any of the following conditions: (1) Liver-related conditions: confirmed diagnosis of viral hepatitis (Hepatitis B surface antigen positive or Hepatitis C virus antibody positive), autoimmune liver disease, drug-induced liver injury, genetic metabolic liver diseases (e.g., Wilson's disease, alpha-1-antitrypsin deficiency), or other specific etiologies of chronic liver disease; (2) Excessive alcohol consumption or diagnosed decompensated hepatocellular carcinoma; (3) Severe comorbidities: presence of life-threatening or study compliance-affecting cardiovascular diseases (e.g., heart failure NYHA class III-IV, uncontrolled hypertension), respiratory diseases, renal failure [estimated glomerular filtration rate < 30 mL/min/1.73 m^2], hematological diseases, or active malignancy (within 5 years); (4) Factors affecting tongue imaging: major oral diseases, such as oral ulcers, oral cancer, history of tongue surgery, or severe tongue markings distorting tongue morphology, as well as congenital tongue anomalies, such as geographic tongue or fissured tongue, which may interfere with image analysis; (5) Medication use: long-term use of medications known to affect hepatic fat metabolism or fibrosis (e.g., glucocorticoids, amiodarone, tamoxifen, specific antiviral drugs) within 3 months prior to enrollment; (6) Special populations: pregnant or lactating women; (7) Concur-

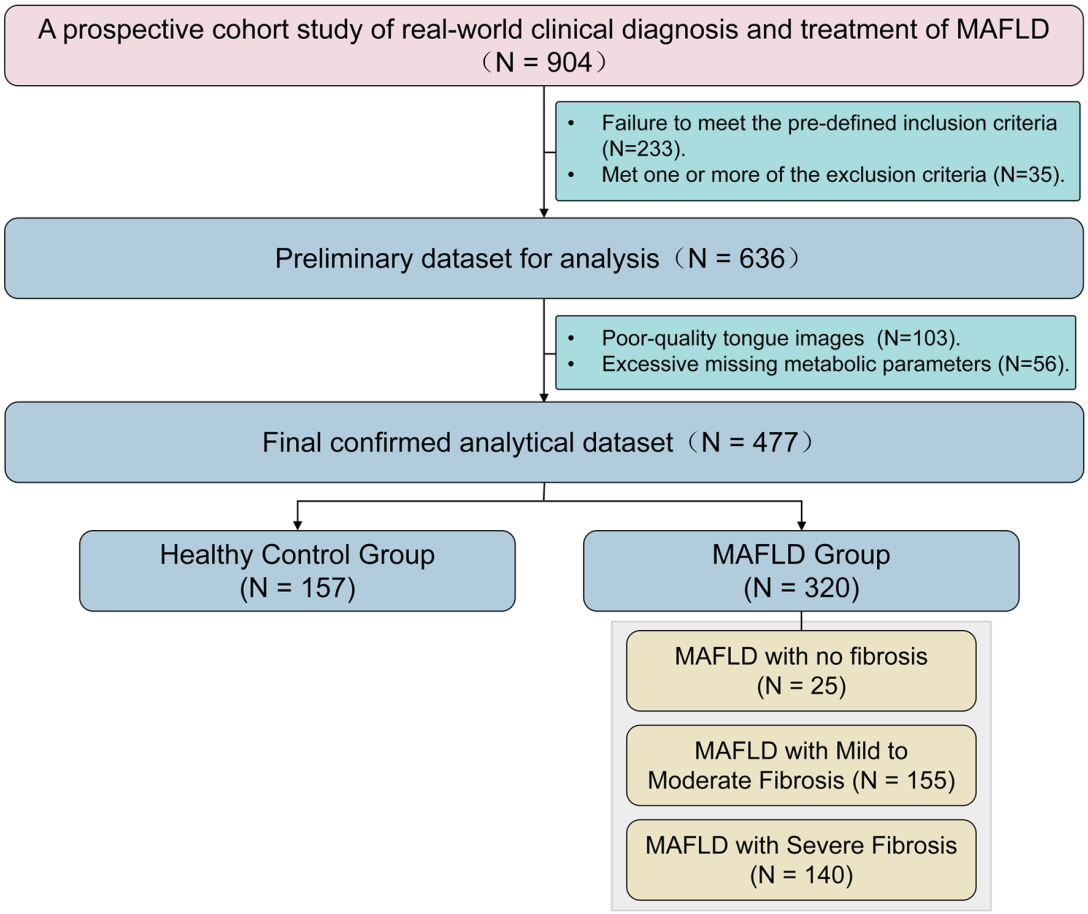

Fig. 1. Flowchart of participant selection. A total of 904 candidates were initially screened. Of these, 268 were excluded for not meeting the eligibility criteria, 103 were excluded due to poor-quality tongue images, and 56 were excluded because of extensive missing clinical data. After applying these exclusion criteria, the final analytical dataset consisted of 477 participants, including 157 healthy controls and 320 MAFLD patients. The MAFLD group was further stratified by fibrosis stage: no fibrosis progression (N = 25), mild-to-moderate fibrosis progression (N = 155), and severe fibrosis progression (N = 140). MAFLD, Metabolic dysfunction-associated fatty liver disease.

rent participation in another interventional clinical trial; or any other condition deemed by the investigators as unsuitable for study participation.

Participant grouping: Healthy control group (non-MAFLD): defined by Controlled Attenuation Parameter < 245 dB/m, normal liver function tests, and absence of MAFLD-related metabolic risk factors and relevant medical history.

MAFLD group and fibrosis stratification definition: all patients meeting MAFLD diagnostic criteria and hepatic fibrosis progression were stratified by Liver Stiffness Measurement (LSM) into: MAFLD-No Fibrosis (F0-F1): LSM < 7.3 kPa; MAFLD-Mild to Moderate Fibrosis (F2-F3): 7.3 kPa ≤ LSM < 12.4 kPa; MAFLD-Severe Fibrosis (F4): LSM ≥ 12.4 kPa. A flowchart detailing the participant inclusion process is depicted in Figure 1.

Data collection and statistical analysis

All data were collected by uniformly trained and certified research personnel following standard operating procedures to ensure consistency and accuracy. The methodology regarding the collection of tongue manifestation and clinical feature data was described in our earlier work.[19]

Clinical data analysis was conducted using Python (executed in PyCharm 2025.2; JetBrains, Prague, Czech Republic) and R (executed in RStudio). Key Python libraries included pandas (v2.1.4), numpy (v1.26.2), scipy (v1.11.4), and matplotlib (v3.8.2). Key R packages included gtsummary and dplyr. All indicators were converted to numerical variables, and missing and invalid values were uniformly set to zero without additional standardization or normalization to preserve the original numerical characteristics and clinical significance of the indicators.

The Kruskal-Wallis H test was used to verify the overall distribution differences of each metabolic indicator among different fibrosis grades of MAFLD, with a significance level of $\alpha = 0.05$. Spearman rank correlation analysis was further used to quantify the association strength and trend direction between the indicators and fibrosis grades, defining a strong association as $|r| \geq 0.2$ and $P < 0.05$, and a weak association as $0 < |r| < 0.2$ and $P < 0.05$.

Multimodal medical AI model

Multimodal learning has demonstrated significant advantages in medical diagnostics through the integration of complementary data sources. However, in the specific field of computer-aided tongue diagnosis, research on multimodal fusion remains in its early stages. Current approaches attempting to combine tongue image features with clinical in-

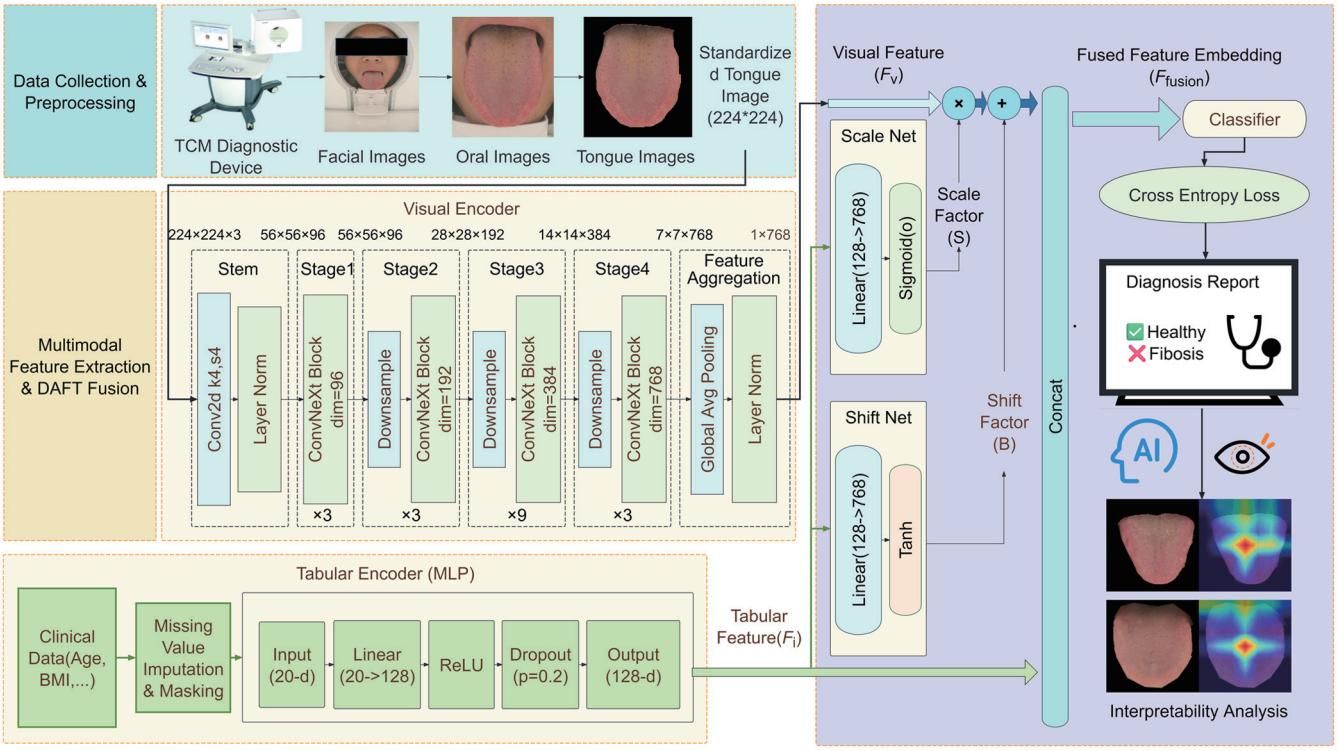

Fig. 2. Overall system architecture diagram. The framework consists of a standardized preprocessing pipeline, a dual-stream feature extraction module that processes both engineered features and deep representations, an interactive fusion module designed to model cross-modal dependencies, and a final classifier optimized for robust healthy-versus-MAFLD discrimination. MAFLD, Metabolic dysfunction-associated fatty liver disease; TCM, traditional Chinese medicine; BMI, Body Mass Index.

dicators often rely on elementary fusion strategies such as feature concatenation or late fusion. These methods fail to capture the intricate and interactive relationships between visual features of the tongue and systemic physiological or metabolic parameters. Moreover, while many existing models treat disease staging as a nominal classification task, the present study focuses specifically on the binary discrimination between healthy individuals and patients with MAFLD. This streamlined objective prioritized accurate screening and early detection, aligning with the practical need for accessible, non-invasive diagnostic tools in population health. The architecture of the proposed "Multimodal-aided Diagnostic System" in this study is illustrated in Figure 2.

The proposed multimodal auxiliary diagnostic system is designed to emulate the comprehensive diagnostic logic employed by clinical experts, akin to the integrative approach of "four diagnostic methods combined" in TCM. By leveraging deep learning techniques, the framework achieves a systematic integration of macro-level visual representations derived from tongue images with micro-level metabolic data obtained from clinical biochemical indicators. This section details the core algorithms of the system across five key components: data preprocessing, overall network architecture, feature extraction mechanisms, multimodal fusion strategy, and classifier design.

Image data preprocessing: High-quality, standardized input data are fundamental to the performance and generalizability of any computational model, particularly in multimodal medical artificial intelligence, where heterogeneous data sources must be integrated. To address pervasive challenges in real-world clinical datasets—such as inconsistent image acquisition protocols, confounding background noise, and missing values in electronic health records—we established a rigorous preprocessing pipeline for both visual and tabular modalities.

Image preprocessing: For tongue images, preserving morphologically and diagnostically significant features was paramount. To avoid geometric distortion of the tongue body, which could obscure critical signs such as swelling or tooth marks, we rejected conventional fixed-size resizing. Instead, we implemented an aspect ratio-preserving Letterbox strategy. The longer side of each image was rescaled to 224 pixels, maintaining the original proportions, while zero-padding (RGB = [0,0,0]) was applied to the shorter side to generate a standardized 224 × 224 pixel input. During resizing, bicubic interpolation was employed to conserve high-frequency texture details relevant to coating analysis. Subsequently, pixel intensities were normalized using Z-score standardization based on ImageNet statistics (mean [0.485, 0.456, 0.406]; standard deviation [0.229, 0.224, 0.225]), ensuring compatibility with pre-trained vision models. To mitigate overfitting and enhance robustness given limited sample sizes, moderate data augmentation was applied stochastically during training only, including horizontal flipping (P = 0.5), minor affine transformations (rotation ± 15°, translation ± 10%), and controlled color jitter (brightness/contrast variation ± 20%).

Tabular data processing: Clinical and laboratory variables (e.g., Age, BMI, biochemical markers) underwent systematic cleaning and encoding. Missing values, an inevitable feature of retrospective clinical data, were imputed with zero. Crucially, to prevent the loss of information regarding data completeness, a binary mask vector indicating the presence or absence of each feature was concatenated with the imputed dataset, allowing the model to explicitly learn from missingness patterns. Label normalization was performed to ensure

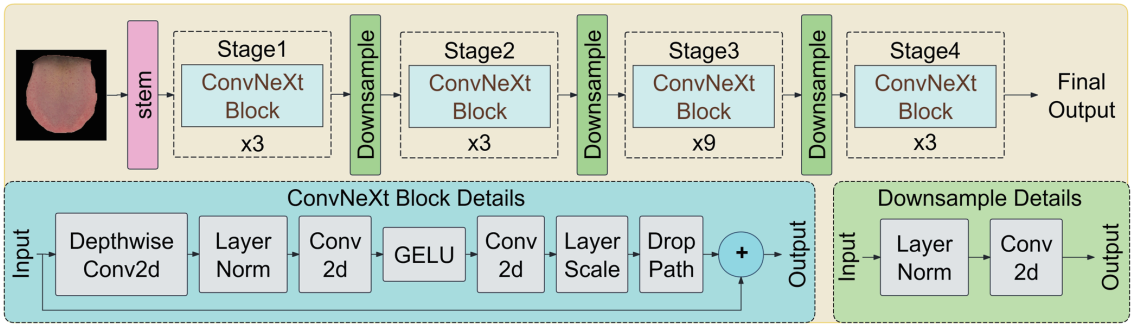

Fig. 3. Schematic of the ConvNeXt-Tiny backbone architecture for visual feature extraction. This figure outlines the overall structure of the ConvNeXt-Tiny model employed in our study. The left panel details the internal design of a single ConvNeXt block, which incorporates modern architectural elements such as depthwise separable convolutions. This design enhances representational capacity while preserving the efficiency and stability inherent to convolutional networks. The right panel illustrates the main network pipeline: an input tongue image is processed sequentially through four hierarchical stages, each comprising multiple ConvNeXt blocks followed by downsampling layers to progressively extract and condense visual features. The model is tasked with autonomously learning local textural and morphological patterns from raw tongue images that correlate with metabolic dysfunction.

consistency; numerical annotations (e.g., fibrosis stages) were extracted from unstructured text using regular expressions, and outliers were rigorously filtered to constrain all labels to a validated, predefined range.

This comprehensive preprocessing framework ensures that both imaging and clinical data streams are transformed into a coherent, analysis-ready format, forming a reliable foundation for the subsequent multimodal fusion and modeling stages.

Visual perception stream: This branch handles structured clinical biochemical indicators ($X_t \in R^{H*}W^{*3}$). A multilayer perceptron maps discrete and continuous physiological parameters into a high-dimensional latent space, constructing a panoramic metabolic profile of the patient. The resulting feature vector provides a compact representation of systemic inflammation and fibrosis risk, extending beyond a mere numerical transformation.

Backbone network: Given that our dataset size represents a typical small-sample scenario in medical research, we selected ConvNeXt-Tiny as the visual backbone. Unlike Vision Transformers (ViTs), which rely heavily on large-scale pre-training to establish global dependencies, ConvNeXt retains the inherent inductive biases of convolutional neural networks, such as translation equivariance and locality. This characteristic enables it to effectively capture subtle textures (e.g., coating granularity) and local morphological features in tongue images even without extensive pre-training on massive datasets, thereby significantly mitigating the risk of overfitting (Fig. 3).

Metabolic feature stream: Serving as the "logical core" of the system, this branch handles structured clinical biochemical features ($X_t \in R^N$). A multilayer perceptron maps discrete and continuous physiological parameters into a high-dimensional latent space, constructing a panoramic metabolic profile of the patient. The resulting feature vector provides a compact representation of systemic inflammation and fibrosis risk, extending beyond a mere numerical transformation.

A multilayer perceptron-based encoder is designed to process continuous clinical features for effective embedding. This component maps heterogeneous physiological and biochemical parameters into a latent space that is semantically aligned with the visual feature representations.

The network architecture comprises a linear layer (20 $\rightarrow$ 128), followed by one-dimensional batch normalization (BatchNorm1d), a rectified linear unit activation function, and a Dropout layer (P = 0.2). This encoder transforms the raw 20-dimensional clinical data into a 128-dimensional tabular feature embedding F_t, facilitating cross-modal alignment at a semantic level.

The inclusion of BatchNorm1d standardizes the activation distributions across layers, accelerating training convergence. The Dropout layer introduces stochasticity by randomly deactivating neurons, thereby enhancing the model's robustness to potential missing indicators or measurement errors commonly encountered in clinical data. The resultant metabolic feature vector F_t serves as the conditioning input for the subsequent Dynamic Affine Feature Transformation (DAFT) module, enabling it to actively modulate the visual feature stream. (in Fig. 4)

Dynamic interaction and feature modulation: This is the core innovation of this architecture. Unlike traditional "post-fusion" strategies that simply concatenate the outputs of two streams, we designed a deep interaction mechanism based on DAFT.

In this mechanism, the metabolic feature stream is no longer a passive input but acts as an active "conditional regulator." It dynamically recalibrates the feature channels in the visual stream by learning the generated affine transformation parameters (scaling factor α and translation factor β).

Mathematically, this mechanism simulates a Bayesian inference process: clinical indicators provide prior probabilities that guide the model to focus on a more discriminative posterior distribution within the visual feature space, thereby significantly suppressing non-pathological visual noise (e.g., illumination variation or physiological tongue enlargement). The modulated multimodal features are subsequently fed into a classifier to output a probability distribution indicating whether the person is healthy or has MAFLD.

Conventional feature fusion methods, such as simple vector concatenation or element-wise summation, often fail to account for the substantial disparities in data distribution and semantic meaning between heterogeneous modalities. To enable clinically informative, low-dimensional metabolic indicators to effectively guide the interpretation of high-dimensional, semantically sparse visual tongue features, we introduce a DAFT module. This component establishes a deep fusion mechanism based on conditional feature recalibration.

The central premise of the DAFT module is to treat the tabular (metabolic) modality as a dynamic "controller" that generates affine transformation parameters for the visual modality. This mechanism allows the model to adaptively enhance or suppress specific channels within the visual feature maps according to the patient's metabolic profile. For instance, a

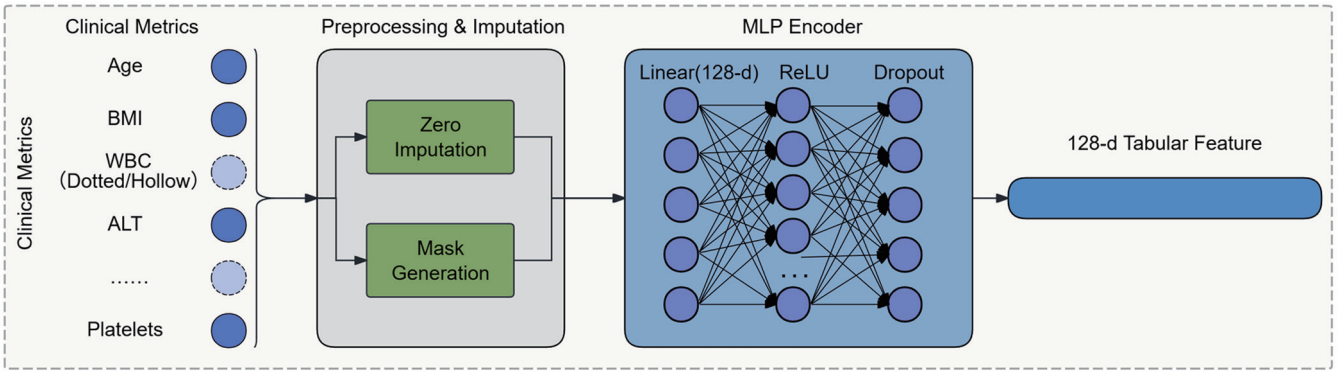

Fig. 4. Architecture of the clinical data encoder. This schematic shows the multilayer perceptron-based encoder that processes structured clinical parameters (20-dimensional input) into a compact 128-dimensional feature embedding (F_t). The encoder consists of a linear layer, batch normalization, rectified linear unit activation, and dropout. The resulting metabolic feature vector Ft is then used to modulate the visual feature stream in the downstream fusion module. BMI, Body Mass Index; WBC, White Blood Cell Count; ALT, Alanine Aminotransferase; MLP, Multilayer Perceptron.

clinical indicator such as "low platelet count" can guide the model to amplify features corresponding to a "purplish-dark tongue body." This facilitates deep, interactive integration of multimodal information at the feature-extraction stage.

First, the metabolic features are projected to the dimensionality of the visual feature space:

Let $F_v \in R^{D_v}$ and $F_v \in R^{D_v}$ denote the visual and metabolic feature vectors, respectively. The DAFT module employs two lightweight auxiliary networks: a Scale Generator (*Gs*) and a Shift Generator (*Gb*). The fusion process is defined as follows:

$$S = \sigma\left(G_S\left(F_t\right)\right) \in (0,\ 1)^{D_v}$$

$$B = \tanh\left(G_b\left(F_t\right)\right) \in (0,\ 1)^{D_v}$$

where σ is the Sigmoid activation function, ensuring that the scaling factor S acts as a gating signal, and tanh is used to generate the bidirectional feature offset B.

Subsequently, a channel-based affine transformation is performed to generate the fused features F_{refined}:

$$F_{\text{refined}} = S \odot F_v + B$$

Here, $\odot$ represents the Hadamard product, i.e., element-wise multiplication. Geometrically, this transformation is equivalent to a dynamic distortion and correction of the visual feature space based on metabolic states.

Finally, to preserve the original metabolic information, residual joins or concatenation were performed between the recalibrated visual features and the original tabular features:

$$F_{\text{fusion}} = \text{Concat}\left(F_{\text{refined}}, F_t\right)$$

This design offers inherent clinical interpretability. The scaling factor S mimics a clinician's "attentional focus," amplifying the weights of visual feature channels relevant to specific pathological states. Conversely, the shifting factor B acts as a "baseline calibrator," adjusting the decision threshold for disease severity based on contextual factors such as age or sex, even when visual presentations are similar. This proactive, context-aware fusion strategy demonstrably outperforms passive data concatenation, yielding diagnostic synergy where the integrated whole is greater than the sum of its individual parts (in Fig. 5).

Training strategy and evaluation metrics

Dataset partitioning: To ensure a rigorous, unbiased, and reproducible evaluation of our multimodal framework, we implemented a strict data partitioning protocol. From the final curated cohort of 477 eligible participants, the dataset was randomly split into training, validation, and a held-out independent test set using an 8:1:1 ratio. This resulted in 381

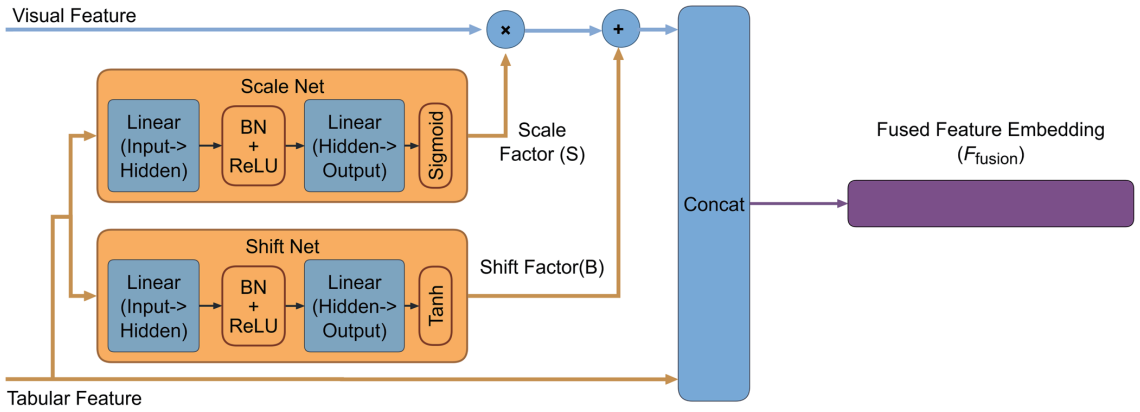

Fig. 5. Dynamic multimodal fusion via the DAFT module. This schematic illustrates the DAFT, the core of our interactive fusion strategy. Clinical metabolic features act as a dynamic controller to recalibrate the visual feature map extracted from tongue images. DAFT learns to apply channel-specific scaling and shifting to the visual features, amplifying or suppressing patterns based on the patient's metabolic profile. This mimics clinical reasoning by allowing systemic biomarkers to guide the interpretation of anatomical signs, leading to a context-aware and diagnostically synergistic multimodal representation. DAFT, Dynamic Affine Feature Transformation.

subjects allocated for model training, 48 for validation (used for hyperparameter tuning and early stopping to prevent overfitting), and a final, completely independent set of 48 subjects for testing. The random partitioning was performed at the patient level using stratified sampling to preserve the approximate distribution of key classes (healthy vs. MAFLD) across all splits, thereby mitigating potential evaluation bias.

All model development, training, and evaluation were conducted within the PyTorch 2.8 deep learning ecosystem. Computations were performed on dedicated NVIDIA RTX 5090 GPUs (32 GB memory), with software environments containerized to ensure consistency.

Loss function: In clinical practice, the task of early screening, accurately identifying the presence or absence of significant liver fibrosis, holds greater immediate relevance. Therefore, this study formulates the problem as a binary classification task, aiming to determine whether a patient has fibrotic lesions.

Weighted cross-entropy loss: For the binary classification objective, we employ cross-entropy as the base loss function. To address the potential class imbalance commonly encountered in medical datasets, where healthy (negative) samples may outnumber diseased (positive) ones, we introduce a weighting mechanism. This enhances the model's sensitivity to the minority class (typically the diseased cases), which is critical for reducing false negatives in a screening context.

The loss function is defined as follows:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\left[\alpha y_i \log\left(p_i\right) + \left(1-y_i\right)\log\left(1-p_i\right)\right]$$

where $y_i \in \{0,1\}$ is the true label of the i-th sample (1 indicates the presence of fibrosis, 0 indicates health), and p_i is the probability that the model predicts it as positive. α is the weighting coefficient for positive samples. When $\alpha > 1$, the model pays more attention to the classification error of positive samples, effectively avoiding the clinical risks caused by "missed diagnoses." This design ensures that while pursuing high accuracy, the model prioritizes the high sensitivity required for the screening task.

Parameter optimization strategy based on transfer learning: To address the pervasive challenges of data scarcity and costly annotation in medical imaging, this study adopts a backbone-freezing transfer learning strategy.

Rationale: The shallow filters of deep convolutional neural networks typically learn general visual features—such as edges and textures—that exhibit high transferability between natural image domains (e.g., ImageNet) and medical images. For small-scale datasets like ours (≈500 samples), fine-tuning all parameters is prone to overfitting dataset-specific noise and may trigger catastrophic forgetting, thereby degrading the robust feature-extraction capability already acquired during pre-training.

Implementation: Source domain: The visual encoder was initialized with ConvNeXt-Tiny weights pre-trained on the large-scale ImageNet 1K dataset (1.2 million images).

Target domain: On our tongue image dataset, we strictly froze all parameters of the backbone network. Gradient updates were allowed only for the newly added tabular encoder, the DAFT fusion module, and the classification head.

Advantages: This "frozen-backbone" strategy reduces the number of trainable parameters by over 90%, substantially lowering computational overhead. More importantly, it forces the model to learn how to use clinical priors to modulate general visual features, rather than relearning visual representations from scratch. Experiments confirm that this approach achieves better generalization and more stable convergence under limited data compared to full fine-tuning.

Evaluation metrics: Model performance was comprehensively assessed using the following metrics:

Accuracy: Measures the overall proportion of correctly classified samples.

$$\text{Accuracy} = \frac{\sum_{i=1}^{C} TP_i}{N}$$

where TP_i is the number of samples correctly predicted for class i, N is the total number of samples, and C is the total number of classes.

Quadratic weighted kappa (QWK)—The primary metric for this study: QWK quantifies the agreement between predicted and true ordinal labels by applying a quadratic weight to the distance of disagreement. It is more sensitive than accuracy in evaluating a model's ability to correctly assess disease severity progression.

$$k = 1 - \frac{\sum_{i,j}\omega_{i,j}O_{i,j}}{\sum_{i,j}\omega_{i,j}E_{i,j}}$$

$$\omega_{i,j} = \frac{\left(i-j\right)^2}{\left(C-1\right)^2}$$

where $O_{i,j}$ is the observed confusion matrix, $E_{i,j}$ is the expected random consistency matrix, and $\omega_{i,j}$ is the quadratic weight matrix.

Sensitivity (Recall): Measures the model's ability to correctly identify all positive cases (e.g., patients with significant fibrosis ≥ F2).

$$\text{Sensitivity}_k = \frac{TP_k}{TP_k + FN_k}$$

where TP_k represents the positive examples correctly predicted as belonging to this category, and FN_k represents the false negative examples incorrectly predicted as belonging to other categories.

Specificity: Measures the model's ability to correctly identify all negative cases (e.g., healthy controls or F0 patients).

$$\text{Specificity}_k = \frac{TN_k}{TN_k + FP_k}$$

where TN_k represents the negative examples correctly excluded from the category, and FP_k represents the falsely predicted negative examples incorrectly predicted to belong to the category.

Results

Trend-based analysis of metabolic indicators

This study enrolled 477 participants. MAFLD patients exhibited significantly higher levels across all measured parameters, including liver function, lipid profiles, uric acid, anthropometric measures, and hepatic fat deposition, compared to healthy subjects, as detailed in Table 1. The distributions of all 20 metabolic indicators differed significantly across the healthy group and MAFLD groups with different fibrosis grades. Spearman rank correlation analysis further demonstrated that the trends of the aforementioned indicators showed a strong positive correlation with fibrosis grade ($r > 0.2$, $P < 0.001$). Arranged in descending order of correlation coefficient, the core indicators were BMI (r = 0.7904), VFA (r = 0.7279), WHR (r = 0.6889), GGT (r = 0.5121), ALT (r = 0.5106), and AST (r = 0.498), indicating that the values of these indicators significantly increased with the severity of fibrosis, with no weak positive trends or other trends observed (Fig. 6). Among the lipid profiles, TG (r = 0.4587), CHOL (r

Table 1. Comparison of clinical characteristics between healthy group and MAFLD group

Characteristic	Healthy Group (N = 157)	MAFLD Group (N = 320)	*P*[a]
Gender, n (%)			<0.001
Female	144 (92)	118 (37)	
Male	13 (8.3)	202 (63)	
Age (y)	31.00 (27.00, 39.00)	28.00 (23.00, 34.00)	<0.001
CAP (dB/m)	230.02 (219.71, 239.43)	345.21 (331.27, 357.59)	<0.001
LSM (kPa)	5.84 (5.28, 6.53)	11.53 (8.97, 15.23)	<0.001
BMI (kg/m^2)	25.30 (24.00, 26.30)	36.90 (34.50, 40.10)	<0.001
ALT (U/L)	15.00 (12.00, 18.00)	80.00 (52.00, 112.00)	<0.001
AST (U/L)	18.00 (14.50, 20.50)	40.50 (29.00, 60.00)	<0.001
GGT (U/L)	14.00 (10.00, 20.50)	39.00 (27.00, 57.00)	<0.001
CHOL (mmol/L)	4.74 ± 0.75	4.95 ± 0.91	0.14
TG (mmol/L)	1.14 (0.85, 1.65)	1.81 (1.43, 2.48)	<0.001
HDL-C (mmol/L)	1.24 (1.04, 1.37)	0.98 (0.88, 1.09)	<0.001
LDL-C (mmol/L)	2.67 ± 0.60	3.15 ± 0.69	<0.001
UA (μmol/L)	322.50 (269.00, 372.00)	484.00 (426.00, 593.00)	<0.001

[a]Statistical analysis was performed as follows: Continuous variables following a normal distribution were compared using the two-sample t-test, while non-normally distributed continuous variables were analyzed using the Wilcoxon rank-sum test. The categorical variable (age) was compared using Pearson's chi-squared test. MAFLD, Metabolic dysfunction-associated fatty liver disease; CAP, Controlled attenuation parameter; LSM, Liver stiffness measurement; BMI, Body mass index; ALT, Alanine aminotransferase; AST, Aspartate aminotransferase; GGT, Gamma-glutamyl transferase; CHOL, Cholesterol; TG, Triglycerides; HDL-C, High-density lipoprotein cholesterol; LDL-C, Low-density lipoprotein cholesterol; UA, Uric acid.

= 0.3879), and LDL-C (r = 0.4198) were all significantly and strongly positively correlated with fibrosis grade.

Tongue image feature analysis

We performed colorimetric analysis by measuring the *Lab* values of the tongue and coating across predefined regions (including the tip, root, left side, right side, and entire area). Trend analysis was subsequently conducted on the resulting 36 *Lab* color features (in Fig. 7). The results showed that the *Lab-b** value (yellowness) of the tongue image was the most significant color indicator changing with the progression of MAFLD fibrosis, exhibiting a clear negative trend in both the tongue and tongue coating areas.

Regarding trend strength, strong negative trends were displayed by ten *Lab-b** value features covering multiple regions of both the tongue and coating (entire, center, tip, right side, left side). Only the *Lab-b** value at the coating root showed a weak negative trend. No positive trends were observed in any *Lab-b** features, confirming a consistent decrease in yellowness across all regions. Linear fitting quantified the rate of yellowness (*Lab-b**) decrease. The coating exhibited a significantly faster decline (slope of *Lab-b** at the tip: −1.1675) compared to the tongue (slope of *Lab-b**: −0.4715). This consistent negative trend was further supported by a stable negative Spearman correlation with fibrosis grade and significant Kruskal-Wallis test results ($P < 0.05$). In contrast, trends for *Lab-a** and *Lab-L** values were minimal and non-specific.

A dual-ranking analysis identified the most altered tongue regions. Based on the mean absolute Spearman correlation coefficient (|r|), the top regions were the left side (0.3179), tip (0.3129), and whole area (0.3014). A composite index (F+d) combining ANOVA F-value and Cohen's d yielded a congruent ranking: left side (1.8694), tip (1.867), and whole area (1.8321) (Fig. 8). Both methods indicated that the coating was generally more sensitive than the tongue, and the lateral edges (particularly the left side) showed the most pronounced inter-group differences.

The observed pattern, where the most significant changes localizes to the lateral tongue edges, aligns with the TCM theory that this region corresponds to the liver and gallbladder. This correlation provides supportive evidence linking the topographic findings of tongue diagnosis to the modern pathological understanding of MAFLD. The heightened sensitivity of the tongue coating also suggests its potential utility as a superficial marker for early screening.

Predictive model quantitative analysis

The multimodal fusion network was implemented using the PyTorch framework. Input images were standardized to 224 × 224 pixels. To prevent overfitting given the limited sample size, the ConvNeXt-Tiny backbone was frozen, and only the DAFT fusion module and classification head were optimized. The trainable parameters were updated using the AdamW optimizer with an initial learning rate of 1×10^{-3} and a weight decay of 0.05. A Cosine Annealing scheduler was employed to dynamically adjust the learning rate, decaying it to a minimum of 1×10^{-6} over the course of 100 training epochs. The model was trained with a batch size of 32. For the fibrosis diagnosis task, we utilized the Binary Cross-Entropy Loss function. To ensure reproducibility, the random seed was fixed at 42. We adopted a best-model checkpointing strategy, retaining the model state that achieved the highest QWK on the validation set.

Given the retrospective nature of the study, incomplete clinical records were unavoidable. To address this, we implemented a zero-imputation combined with masking strategy for all tabular covariates, including demographic (e.g., age, BMI) and biochemical indices. Specifically, missing entries were replaced with zeros to maintain dimensional consistency. To prevent the model from interpreting these placeholders as clinically meaningful low values, we simultaneously generated a binary mask vector corresponding to the input

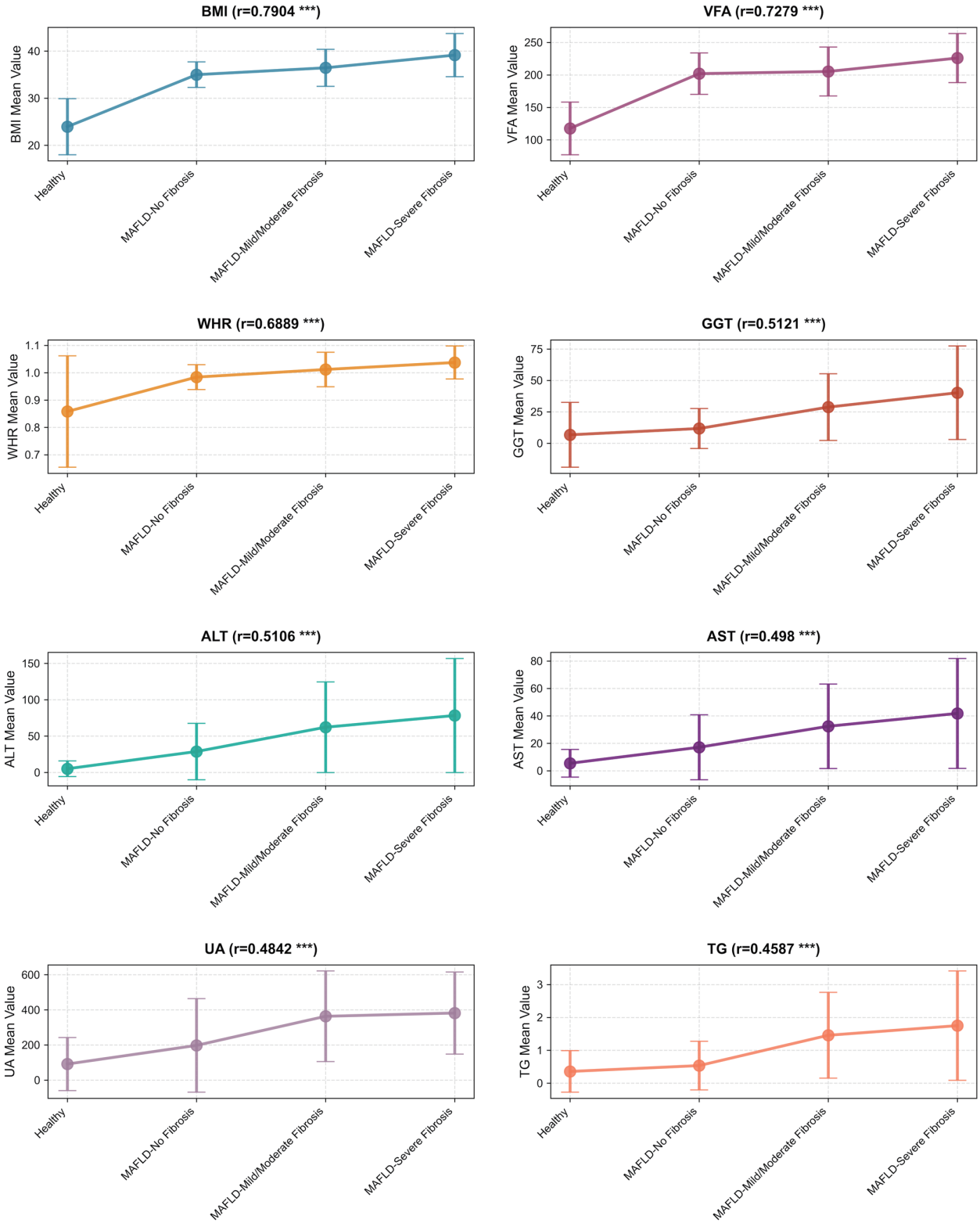

Fig. 6. Trends of metabolic indicators in healthy controls and MAFLD patients with different fibrosis progression. The top eight metabolic indicators with the strongest correlation are shown above, including anthropometric measurements (BMI, VFA, WHR), liver enzymes (GGT, ALT, AST), uric acid, and TG, all demonstrating a gradual increasing trend across advancing fibrosis stages in MAFLD. In the figure, *r* denotes the correlation coefficient; ***$P < 0.001$. MAFLD, Metabolic dysfunction-associated fatty liver disease; BMI, Body Mass Index; VFA, Visceral Fat Area; WHR, Waist-to-Hip Ratio; GGT, Gamma-Glutamyl Transferase; ALT, Alanine Aminotransferase; AST, Aspartate Aminotransferase; TG, Triglycerides.

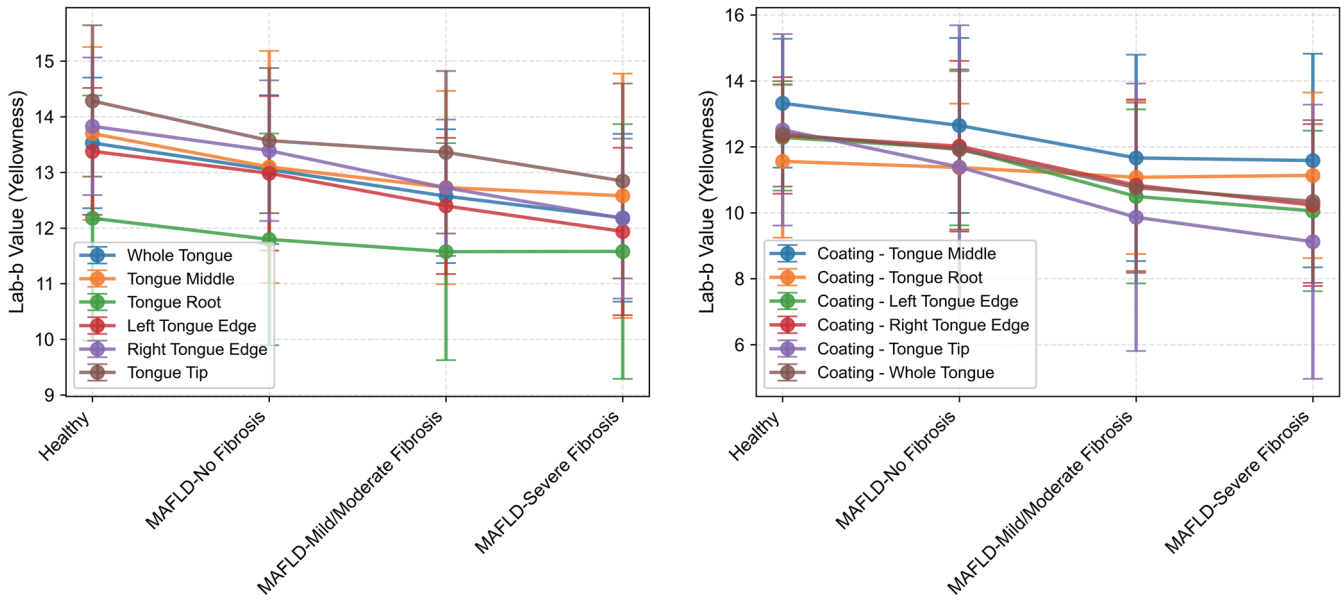

Fig. 7. Tongue and coating *Lab-b values Trend with MAFLD and Fibrosis Progression.** The left panel shows the *Lab-b** values for specified regions of the tongue, while the right panel shows the corresponding *Lab-b** values for the coating regions. A consistent declining trend in yellowness (*Lab-b** value) was observed across nearly all regions with advancing fibrosis, with the exception of the coating root area, which exhibited a minimal or non-significant trend. MAFLD, Metabolic dysfunction-associated fatty liver disease.

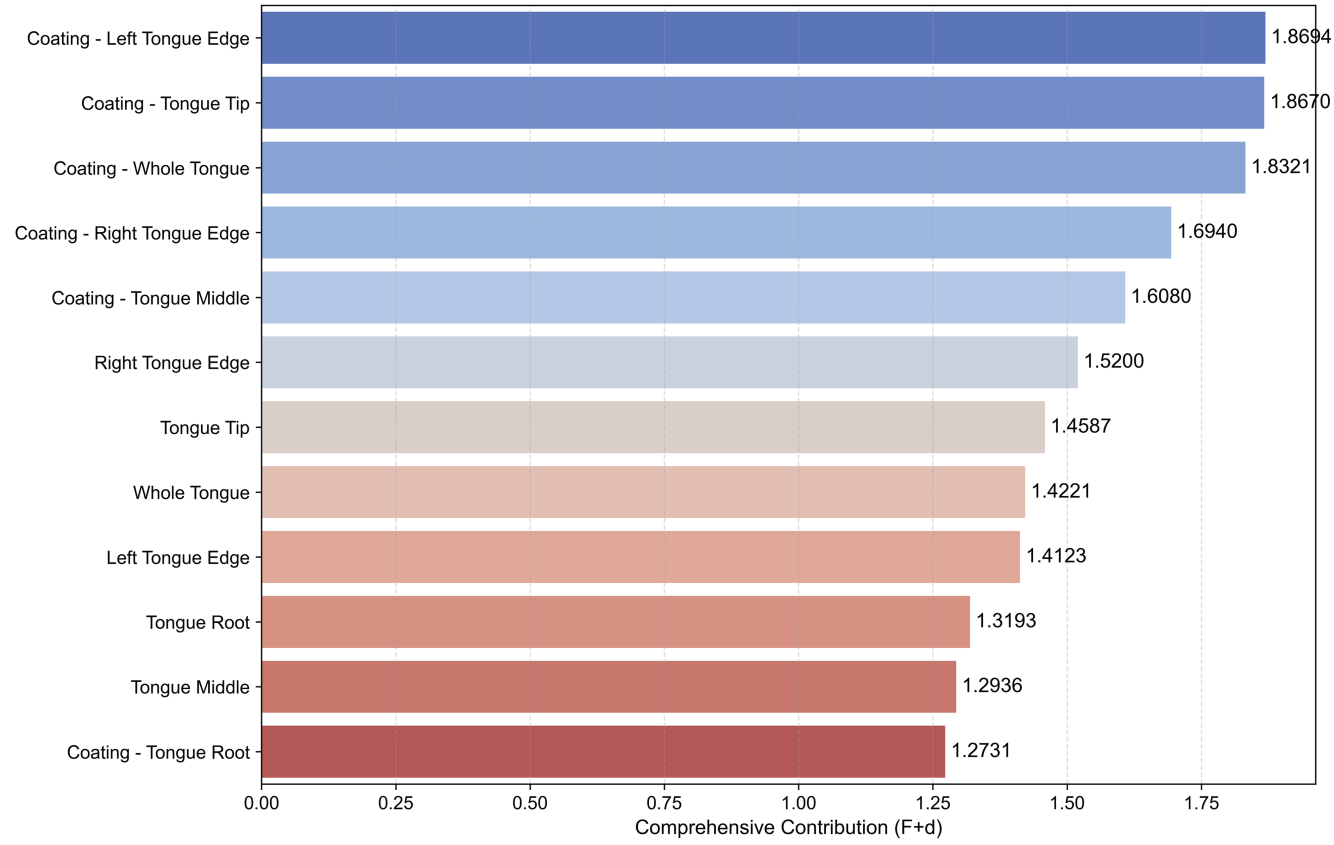

Fig. 8. Ranking of differential contributions of regional tongue and coating Lab-b* values to MAFLD and fibrosis progression. The differential contribution of each tongue and coating region was evaluated using a composite index (F + d) derived from the ANOVA F-value and Cohen's d effect size. The coating of the left side (F + d = 1.8694), tip (F + d = 1.867), and entire area (F + d = 1.8321) all showed F+d scores > 1.8, validating the significant role of the tongue coating in distinguishing MAFLD and its fibrosis stages. MAFLD, Metabolic dysfunction-associated fatty liver disease; ANOVA, Analysis of Variance.

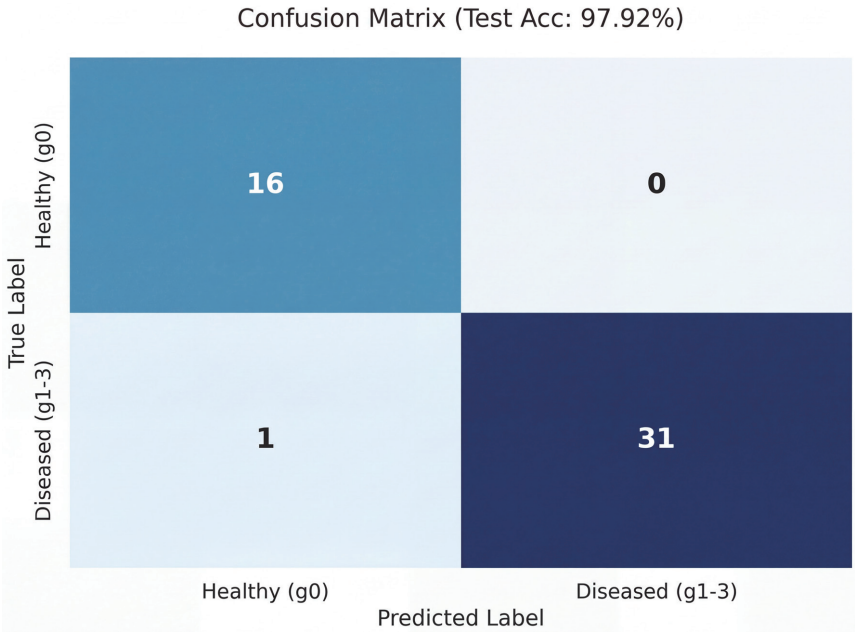

Fig. 9. Confusion matrix of the proposed model on the independent test set. The matrix summarizes the binary classification performance for distinguishing healthy controls from MAFLD patients. Out of 48 independent test samples, the model correctly classified all 16 healthy subjects (specificity = 100%) and 31 of 32 MAFLD cases (sensitivity = 96.88%), yielding an overall accuracy of 97.92% and a Quadratic Weighted Kappa of 0.9538. These results demonstrate the model's high discriminative power and robust generalization to unseen data. MAFLD, Metabolic dysfunction-associated fatty liver disease.

features. In this masking scheme, a value of 1 denotes an observed measurement, while 0 indicates a missing value. This mask vector is concatenated with the feature vector and fed into the tabular encoder, thereby enabling the neural network to explicitly distinguish between observed data and imputed values, and to adaptively attenuate the noise introduced by missing data during the feature fusion process.

To comprehensively evaluate the effectiveness and robustness of our proposed multimodal binary classification model for diagnosing MAFLD, we conducted a quantitative assessment on an independent test set (n = 48). The model demonstrated strong diagnostic performance, achieving an overall accuracy of 97.92% and a QWK of 0.9538. Further analysis of the confusion matrix (Fig. 9) revealed its discriminative capability across classes. The model correctly identified all 16 healthy control samples, yielding a specificity of 100%. Among the 32 MAFLD-positive samples, it successfully recognized 31 cases, resulting in a sensitivity of 96.88%. This balance between high specificity and sensitivity indicated robust generalization on unseen data.

Comparative evaluation of unimodal versus multimodal diagnostic approaches

To rigorously evaluate the diagnostic contribution of tongue imaging and to benchmark the added value of multimodal integration, we conducted a systematic comparative analysis. This involved training and validating state-of-the-art deep learning models exclusively on tongue image data, using the identical dataset partition (training, validation, and independent test sets) as employed for our primary multimodal framework.

For this unimodal assessment, we implemented two distinct architectural paradigms: a YOLO-based model optimized for efficient localization and feature extraction from tongue regions, and an ViT model to capture long-range dependencies and global contextual information within the images. Both models were tasked with the binary classification of healthy controls versus MAFLD patients, mirroring the primary objective of our multimodal system.

The performance of these image-only models was quantitatively inferior to that of our multimodal architecture (Table 2). While achieving non-trivial accuracy, both the YOLO and ViT baselines exhibited consistently lower sensitivity, specificity, and overall balanced accuracy on the held-out test set. Specifically, the unimodal models demonstrated a notable decrease in sensitivity, indicating a higher rate of missed MAFLD cases compared to our integrated approach.

This performance gap underscores a key finding: although tongue imagery contains discriminative signals related to MAFLD, these features are insufficient in isolation for robust screening. The superior diagnostic accuracy of our multimodal model arises from the synergistic integration of visual tongue features with core metabolic and clinical parameters. The comparative results provide empirical validation that the

Table 2. Comparison of performance metrics for the MAFLD multimodal diagnosis model

Models	QWK	Accuracy	Sensitivity	Specificity
Vit	0.3662	0.6875	0.6562	0.7500
Yolov8	0.3836	0.6875	0.6250	0.8125
Our proposed model	0.9538	0.9792	0.9688	1

MAFLD, Metabolic dysfunction-associated fatty liver disease; QWK, Quadratic weighted kappa.

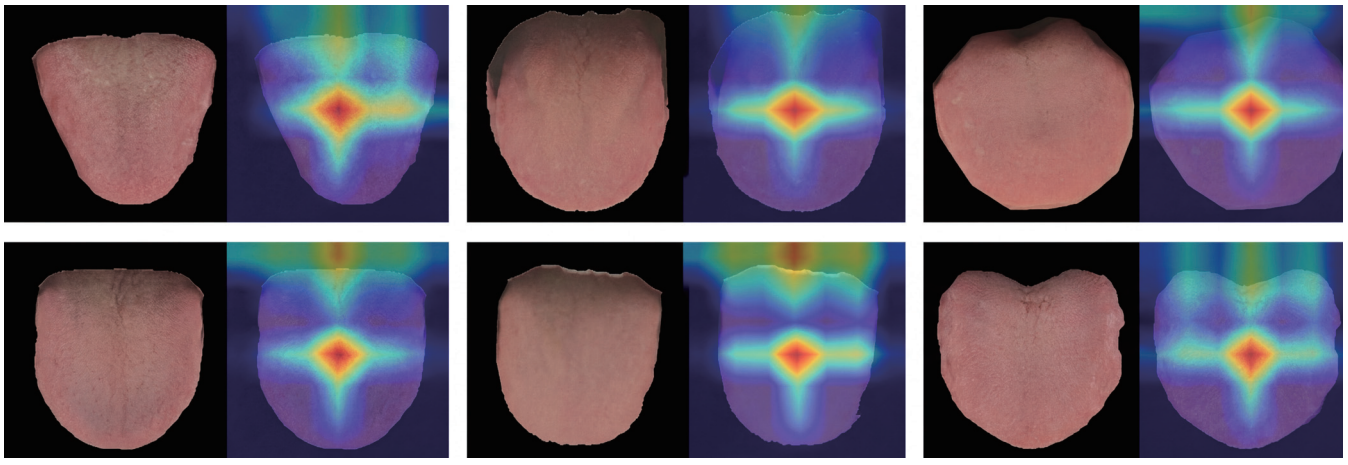

Fig. 10. Visualization of tongue features based on EigenCAM. Based on EigenCAM, tongue features are visualized. The red areas represent areas of high interest during model decision-making, mainly concentrated in the middle of the tongue (spleen and stomach area) and the sides of the tongue (liver and gallbladder area).

complex pathophysiology of MAFLD is more completely captured by the confluence of phenotypic (tongue appearance) and physiological (clinical biomarkers) data domains than by either modality alone. This evidence firmly establishes the necessity and advantage of the proposed multimodal fusion framework for improving non-invasive MAFLD screening.

Interpretability analysis

To elucidate the decision logic of our multimodal model and establish its clinical relevance, we employed complementary interpretability techniques across both visual and feature domains. Attention visualization using EigenCAM revealed that the model consistently attended to anatomically and diagnostically meaningful regions of the tongue. As shown in the generated heatmaps (Fig. 10), high-response areas were primarily concentrated in the central tongue body and the lateral edges. This focus aligns precisely with core tenets of TCM diagnosis: the central region corresponds to the spleen and stomach, often affected by dampness accumulation in MAFLD, while the lateral edges correspond to the liver and gallbladder, where signs of qi stagnation manifest.

This visual interpretability was further quantified and extended through SHapley Additive exPlanations (SHAP) analysis. SHAP analysis quantified the contribution of each feature to the model's predictions, identifying the top drivers (Fig. 11).

Feature importance ranking: Bars represent the mean absolute SHAP value for each clinical variable, indicating its average magnitude of contribution to the model's decision. Body composition and liver function markers, notably BMI and ALT, emerge as the most influential predictors, consistent with the central role of adiposity and hepatocellular injury in MAFLD pathogenesis.

Directional impact of individual features: Each point represents the SHAP value for a feature in a single sample; the horizontal position indicates whether the feature value increased (positive SHAP) or decreased (negative SHAP) the predicted likelihood of MAFLD. Features are ordered vertically as in (a). This visualization reveals consistent directional effects: for example, higher BMI and ALT values systematically increase the predicted risk, whereas higher albumin levels exhibit a protective effect. The analysis confirms that the model's decisions are driven by clinically meaningful and biologically interpretable feature interactions.

Collectively, these interpretability analyses demonstrate that our data-driven model does not merely learn statistical correlations but discovers and leverages biomedical features with established diagnostic significance. The convergence of its attention mechanisms with TCM theory and its reliance on key metabolic and morphological indicators provide a transparent, clinically grounded rationale for its predictions, effectively bridging computational analysis with traditional diagnostic wisdom.

Discussion

Against the backdrop of global lifestyle shifts, metabolic diseases, particularly MAFLD, have emerged as a major public health challenge.[20] The MAFLD criteria enable better identification of individuals with hepatic steatosis and significant fibrosis, as assessed by NITs.[21] Given that MAFLD with clinically significant fibrosis markedly elevates the risk of liver-related complications and mortality,[22] early screening is of critical importance. In TCM theory, tongue diagnosis serves as a key indicator of the functional state of zang-fu organs, qi, and blood. However, conventional TCM tongue assessment has long relied on subjective clinical evaluation, lacking objective standardization.[23] Meanwhile, current non-invasive tools for MAFLD, such as the Fatty Liver Index, NAFLD Fibrosis Score, and Fibrosis-4 Index, have advanced fibrosis risk stratification,[24] yet more accessible early screening models remain needed. TCM tongue diagnosis offers a convenient and promising avenue for this purpose.

A key methodological strength lies in the model's interactive fusion design. We built a dual-stream architecture where a ConvNeXt-Tiny network processes tongue images and a multilayer perceptron encodes clinical variables. The DAFT module enables context-aware fusion—metabolic features dynamically recalibrate visual feature maps, simulating clinician reasoning. Using a weighted cross-entropy loss and a frozen-backbone transfer-learning strategy, the model prioritized sensitivity (96.88%) and robustness, which are critical for a screening tool.

On an independent test set, the model achieved an accuracy of 97.92% and a QWK of 0.9538, with 96.88% sensitivity and 100% specificity, outperforming single-modality and conventional serological models. Interpretability analyses confirmed that the model focused on tongue regions aligned with TCM theory and was driven by key metabolic features such as visceral fat area. Notably, a progressive decline in tongue yellowness (*Lab-b** value) was observed with fibrosis

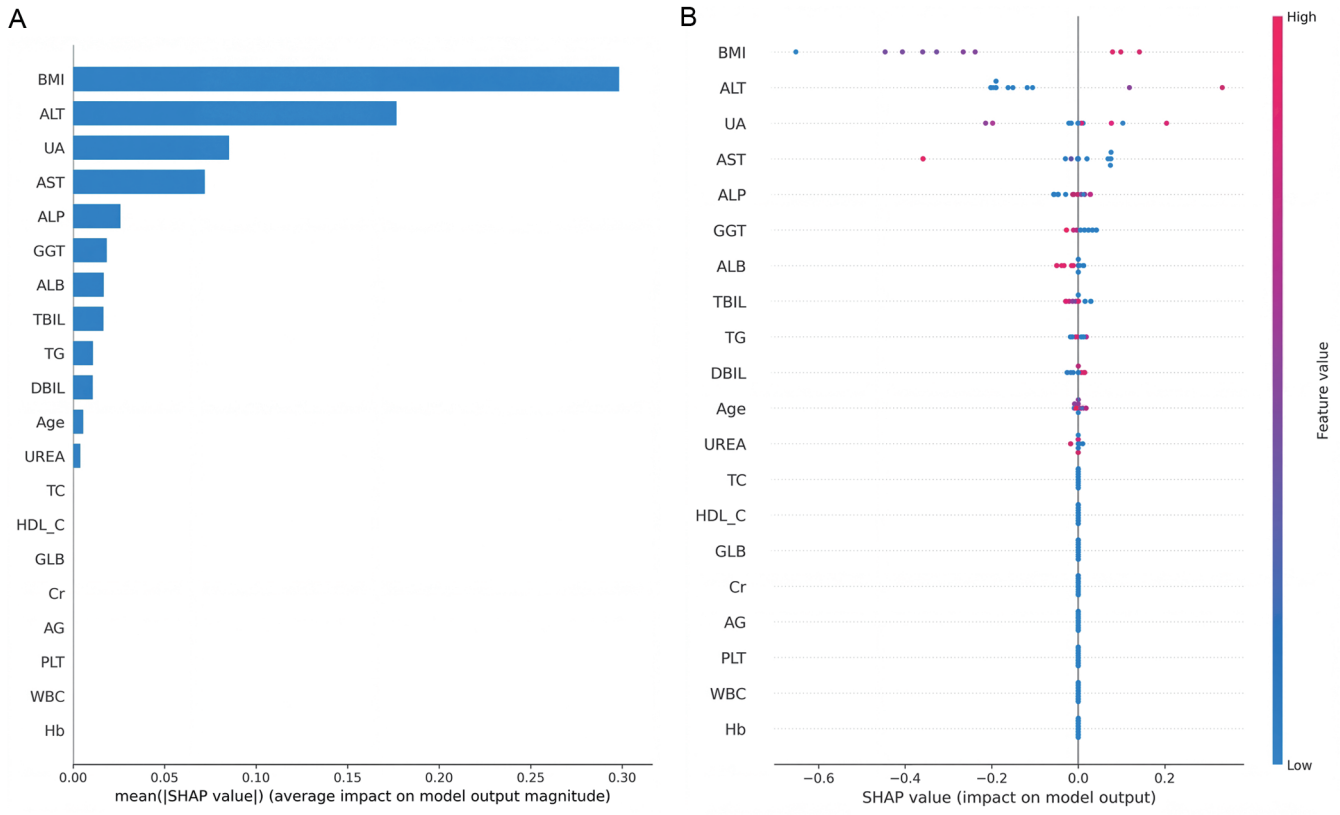

Fig. 11. Clinical feature importance analysis based on SHAP. This figure presents a comprehensive interpretation of how clinical and biochemical features contribute to the multimodal model's prediction of MAFLD status. The SHAP framework quantifies both the magnitude and direction of each feature's influence on the model output, providing insight into the learned diagnostic logic and its alignment with clinical pathophysiology: (A) Mean absolute SHAP values, ranking clinical features by their average contribution magnitude to the model's prediction. (B) Distribution of SHAP values for each feature, where positive values increase the predicted risk of MAFLD and negative values decrease it. MAFLD, Metabolic dysfunction-associated fatty liver disease; SHAP, SHapley additive exPlanations.

progression, most pronounced at the lateral edges—consistent with the TCM principle linking the tongue's left side to the liver and gallbladder. This finding provides objective, modern evidence supporting the physiologic relevance of TCM tongue inspection.

Clinically, the model is designed as a practical binary (Healthy vs. MAFLD) screening tool for primary-care or community settings. Its purpose is efficient triage—identifying individuals needing further specialist assessment—rather than replicating detailed fibrosis staging. This focus enhances sensitivity, reduces missed cases, and improves generalizability, making it suitable for resource-limited environments.

Several limitations of the current study should be acknowledged. First, the model in its present form performs binary screening for MAFLD and does not provide a stratification of fibrosis severity. Second, the single-center, cross-sectional design necessitates external validation in multi-center, prospective cohorts. Although the 8:1:1 data partitioning ensured internal validity, the independent test size remains modest. Future studies with expanded cohorts are needed to enhance statistical power and to evaluate model generalizability across diverse populations and imaging conditions.

Methodologically, while LSM by vibration-controlled transient elastography is a widely validated, non-invasive surrogate for liver biopsy in routine practice,[25,26] an ideal diagnostic tool should also demonstrate the capacity for specific fibrosis staging, prognosis prediction, progression monitoring, and treatment response assessment.[27] To advance towards non-invasive fibrosis grading, particularly in cohorts without biopsy confirmation, future work must develop more precise image recognition techniques, robust multimodal data fusion frameworks, and rigorous ordinal regression methods. For example, the key technical challenges to address include mitigating potential domain shifts in tongue image acquisition and clinical covariates,[28] as well as enhancing the robustness of core fusion modules to variations in input data quality.

In summary, this study presents a novel, interpretable multimodal framework that synergizes TCM tongue diagnostics with metabolic indicators for non-invasive MAFLD screening. By combining methodological innovation with clinically meaningful interpretation, the model offers a low-cost, practical tool for large-scale risk stratification, particularly in settings where specialist resources are limited.

Conclusions

This study successfully developed and validated a clinically oriented, multimodal auxiliary screening and diagnostic model for MAFLD that integrates objective TCM tongue appearance features with conventional metabolic indicators. Through interpretable fusion analysis, it provides a practical application for using TCM tongue appearance as a "window" for the non-invasive assessment of MAFLD. This model holds promise as a new non-invasive, low-cost, and efficient tool for MAFLD screening, particularly in regions with limited healthcare resources.

Funding

This work was supported by the National Natural Science Foundation of China (No. 82274352 to XDL), the Hubei Provincial Key Research and Development Program (No. 2024BCB 038 to XDL), and City University of Hong Kong (7006082, 7020073, 9609332, 9609333, 9678292, 7020002), and the Research Grants Council (9048206, 8799020 to BLK).

Conflict of interest

The authors have no conflict of interests related to this publication.

Author contributions

Study concept and design, methodology, funding acquisition, project administration, and review & editing of the manuscript (CXL, MZX, CHL, BLK, XDL), data curation, formal analysis, visualization, and writing – original draft (CXL, CXT, QYH), investigation (tongue image acquisition and data processing) (ZXS, QH), investigation (clinical data collection and validation) (YBJ, LW, LHZ, HYY, WBZ), AI algorithm design, model architecture development, model training, illustration of AI model architecture and related figures (QYH, YBJ), AI model validation and optimization (LW), bioinformatics analysis and statistical analysis (QYH, HYY), resources and patient recruitment (HZ, JZ). All authors approved the final manuscript.

Ethical statement

The study protocol adhered strictly to the ethical principles of the Declaration of Helsinki (as revised in 2024) and the relevant regulations of China's "Ethical Review Measures for Biomedical Research Involving Humans." It was approved by the Ethics Committee of the Hubei Provincial Hospital of Traditional Chinese Medicine, the lead institution (Approval No. HBZY2022-C08-01). All participants were thoroughly informed about the study's purpose, procedures, potential risks, and benefits. Ample time was provided for consideration, and written informed consent was obtained from each subject prior to enrollment.

Data sharing statement

The data that support the findings of this study are available on request from the corresponding authors. The data are not publicly available due to privacy and ethical restrictions, as they contain sensitive participant information, including clinical health records and identifiable tongue images.

References

[1] Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. Hepatology 2016;64(1):73–84. doi:10.1002/hep.28431, PMID:26707365.
[2] Teng ML, Ng CH, Huang DQ, Chan KE, Tan DJ, Lim WH, *et al*. Global incidence and prevalence of nonalcoholic fatty liver disease. Clin Mol Hepatol 2023;29(Suppl):S32–S42. doi:10.3350/cmh.2022.0365, PMID:36517002.
[3] Tilg H, Petta S, Stefan N, Targher G. Metabolic Dysfunction-Associated Steatotic Liver Disease in Adults: A Review. JAMA 2026;335(2):163–174. doi:10.1001/jama.2025.19615, PMID:41212550.
[4] Man S, Deng Y, Ma Y, Fu J, Bao H, Yu C, *et al*. Prevalence of Liver Steatosis and Fibrosis in the General Population and Various High-Risk Populations: A Nationwide Study With 5.7 Million Adults in China. Gastroenterology 2023;165(4):1025–1040. doi:10.1053/j.gastro.2023.05.053, PMID: 37380136.
[5] Abdelmalek MF. Nonalcoholic fatty liver disease: another leap forward. Nat Rev Gastroenterol Hepatol 2021;18(2):85–86. doi:10.1038/s41575-020-00406-0, PMID:33420415.
[6] Berkan-Kawińska A, Piekarska A. Hepatocellular carcinoma in non-alcohol fatty liver disease - changing trends and specific challenges. Curr Med Res Opin 2020;36(2):235–243. doi:10.1080/03007995.2019.1683817, PMID:31631714.
[7] Masoodi M, Gastaldelli A, Hyötyläinen T, Arretxe E, Alonso C, Gaggini M, *et al*. Metabolomics and lipidomics in NAFLD: biomarkers and non-invasive diagnostic tests. Nat Rev Gastroenterol Hepatol 2021;18(12):835–856. doi:10.1038/s41575-021-00502-9, PMID:34508238.
[8] Vali Y, Lee J, Boursier J, Petta S, Wonders K, Tiniakos D, *et al*. Biomarkers for staging fibrosis and non-alcoholic steatohepatitis in non-alcoholic fatty liver disease (the LITMUS project): a comparative diagnostic accuracy study. Lancet Gastroenterol Hepatol 2023;8(8):714–725. doi:10.1016/S2468-1253(23)00017-1, PMID:36958367.
[9] Loomba R, Adams LA. Advances in non-invasive assessment of hepatic fibrosis. Gut 2020;69(7):1343–1352. doi:10.1136/gutjnl-2018-317593, PMID:32066623.
[10] Sanyal AJ, Williams SA, Lavine JE, Neuschwander-Tetri BA, Alexander L, Ostroff R, *et al*. Defining the serum proteomic signature of hepatic steatosis, inflammation, ballooning and fibrosis in non-alcoholic fatty liver disease. J Hepatol 2023;78(4):693–703. doi:10.1016/j.jhep.2022.11.029, PMID:36528237.
[11] Huang L, Wang S, Zhang H, Feng S, Zhong H, Chen J, *et al*. Clinical efficacy evaluation of washed microbiota transplantation treatment for metabolic related fatty liver disease and its impact on tongue coating microorganisms. Front Endocrinol (Lausanne) 2025;16:1684173. doi:10.3389/fendo.2025.1684173, PMID:41234231.
[12] Wziątek-Kuczmik D, Świątkowski A, Cholewka A, Mrowiec A, Niedzielska I, Stanek A. Thermal Imaging of the Tongue Surface as a Predictive Method in the Diagnosis of Type 2 Diabetes Mellitus. Sensors (Basel) 2024;24(8):2447. doi:10.3390/s24082447, PMID:38676064.
[13] Li J, Xiong D, Hong L, Lim J, Xu X, Xiao X, *et al*. Tongue color parameters in predicting the degree of coronary stenosis: a retrospective cohort study of 282 patients with coronary angiography. Front Cardiovasc Med 2024;11:1436278. doi:10.3389/fcvm.2024.1436278, PMID:39280030.
[14] Shi H, Huang KC. Pictures of Tongues Sticking Out. Trends Endocrinol Metab 2020;31(11):805–807. doi:10.1016/j.tem.2020.05.003, PMID:32475653.
[15] Jiang T, Guo XJ, Tu LP, Lu Z, Cui J, Ma XX, *et al*. Application of computer tongue image analysis technology in the diagnosis of NAFLD. Comput Biol Med 2021;135:104622. doi:10.1016/j.compbiomed.2021.104622, PMID:34242868.
[16] Zhang Q, Wen J, Zhou J, Zhang B. Missing-view completion for fatty liver disease detection. Comput Biol Med 2022;150:106097. doi:10.1016/j.compbiomed.2022.106097, PMID:36244304.
[17] Gao J, Chen T, Xu Y, Wu Y, Liu K, Qiu W, *et al*. Accurate fatty liver disease diagnosis with a multi-source feature fusion model on the segmented tongue image dataset. J Adv Res 2025. doi:10.1016/j.jare.2025.10.003, PMID:41047021.
[18] Eslam M, Newsome PN, Sarin SK, Anstee QM, Targher G, Romero-Gomez M, *et al*. A new definition for metabolic dysfunction-associated fatty liver disease: An international expert consensus statement. J Hepatol 2020;73(1):202–209. doi:10.1016/j.jhep.2020.03.039, PMID:32278004.
[19] Lu C, Zhu H, Zhao D, Zhang J, Yang K, Lv Y, *et al*. Oral-Gut Microbiome Analysis in Patients With Metabolic-Associated Fatty Liver Disease Having Different Tongue Image Feature. Front Cell Infect Microbiol 2022;12:787143. doi:10.3389/fcimb.2022.787143, PMID:35846747.
[20] Patil A, Mungase SB, Nadella M, Adela R. Diagnostic performance of non-invasive markers for distinguishing MASLD/MASH: insights from meta-analysis and real-world data. Clin Chim Acta 2026;582:120787. doi:10.1016/j.cca.2025.120787, PMID:41391581.
[21] Xue Y, Xu J, Li M, Gao Y. Potential screening indicators for early diagnosis of NAFLD/MAFLD and liver fibrosis: Triglyceride glucose index-related parameters. Front Endocrinol (Lausanne) 2022;13:951689. doi:10.3389/fendo.2022.951689, PMID:36120429.
[22] Chen Y, Liang X, Qi Y, Liu C, Dong B, Li X, *et al*. Prevalence of Metabolic Dysfunction-Associated Steatotic Liver Disease with Clinically Significant Fibrosis in Obese Patients with Type 2 Diabetes Mellitus - China, 2017-2024. China CDC Wkly 2025;7(47):1483–1490. doi:10.46234/ccdcw2025.248, PMID:41321337.
[23] Lin H, Ning Z, Zhang C, Men S, Zhang D. Computerized tongue image analysis for non-invasive disease screening: a review. Chin Med 2025;20(1):196. doi:10.1186/s13020-025-01242-7, PMID:41267018.
[24] Younossi ZM, de Avila L, Petta S, Hagström H, Kim SU, Nakajima A, *et al*. Global performance of non-invasive tests in MASLD: Insights from the G-MASLD study. Hepatology 2025. doi:10.1097/HEP.0000000000001564, PMID:41100867.
[25] Mózes FE, Lee JA, Selvaraj EA, Jayaswal ANA, Trauner M, Boursier J, *et al*, LITMUS Investigators. Diagnostic accuracy of non-invasive tests for advanced fibrosis in patients with NAFLD: an individual patient data meta-analysis. Gut 2022;71(5):1006–1019. doi:10.1136/gutjnl-2021-324243, PMID:34001645.
[26] Eddowes PJ, Sasso M, Allison M, Tsochatzis E, Anstee QM, Sheridan D, *et al*. Accuracy of FibroScan Controlled Attenuation Parameter and Liver Stiffness Measurement in Assessing Steatosis and Fibrosis in Patients With Nonalcoholic Fatty Liver Disease. Gastroenterology 2019;156(6):1717–1730. doi:10.1053/j.gastro.2019.01.042, PMID:30689971.
[27] Lin H, Lee HW, Yip TC, Tsochatzis E, Petta S, Bugianesi E, *et al*, VCTE-Prognosis Study Group. Vibration-Controlled Transient Elastography Scores to Predict Liver-Related Events in Steatotic Liver Disease. JAMA 2024;331(15):1287–1297. doi:10.1001/jama.2024.1447, PMID:38512249.
[28] Lu PH, Chiang CC, Yu WH, Yu MC, Hwang FN. Machine Learning-Based Technique for the Severity Classification of Sublingual Varices according to Traditional Chinese Medicine. Comput Math Methods Med 2022;2022:3545712. doi:10.1155/2022/3545712, PMID:36388160.