Machine Intelligence Research
www.mi-research.net
DOI: 10.1007/s11633-024-1506-4

Multimodal Learning-based Prediction for Nonalcoholic Fatty Liver Disease

Yaran Chen[1,2†] Xueyu Chen[3†] Yu Han[1,2†] Haoran Li[1,2]
Dongbin Zhao[1,2] Jingzhong Li[4] Xu Wang[4] Yong Zhou[5]

[1] State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

[2] College of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, China

[3] Department of Biostatistics, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan 205100, China

[4] School of Life Sciences, Beijing University of Chinese Medicine, Beijing 100029, China

[5] Clinical Research Institute, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 201620, China

Abstract: Nonalcoholic fatty liver disease (NAFLD) is the most common cause of chronic liver disease, and if it is accurately predicted, severe fibrosis and cirrhosis can be prevented. While liver biopsies, the gold standard for NAFLD diagnosis, is intrusive, expensive, and prone to sample errors, noninvasive studies are extremely promising but are still in their infancy due to a dearth of comprehensive study data and sophisticated multimodal data methodologies. This paper proposes a novel approach for diagnosing NAFLD by integrating a comprehensive clinical dataset with a multimodal learning-based prediction method. The dataset comprises physical examinations, laboratory and imaging studies, detailed questionnaires, and facial photographs of a substantial number of participants, totaling more than 6 000. This comprehensive collection of data holds significant value for clinical studies. The dataset is subjected to quantitative analysis to identify which clinical metadata, such as metadata and facial images, has the greatest impact on the prediction of NAFLD. Furthermore, a multimodal learning-based prediction method (DeepFLD) is proposed that incorporates several modalities and demonstrates superior performance compared to the methodology that relies only on metadata. Additionally, satisfactory performance is assessed through verification of the results using other unseen data. Inspiringly, the proposed DeepFLD prediction method can achieve competitive results by solely utilizing facial images as input rather than relying on metadata, paving the way for a more robust and simpler noninvasive NAFLD diagnosis.

Keywords: Nonalcoholic fatty liver disease detection (NAFLD), disease diagnosis, convolutional neural networks, multimodal data, multimodal learning-based prediction.

Citation: Y. Chen, X. Chen, Y. Han, H. Li, D. Zhao, J. Li, X. Wang, Y. Zhou. Multimodal learning-based prediction for nonalcoholic fatty liver disease. *Machine Intelligence Research*. http://doi.org/10.1007/s11633-024-1506-4

1 Introduction

Chronic liver diseases are prevalent factors contributing to morbidity and mortality on a global scale, with liver-related ailments responsible for more than 2 million annual fatalities worldwide[1]. Nonalcoholic fatty liver disease (NAFLD), also known as metabolic-associated fatty liver disease, is one of the most common chronic diseases and metabolic complications of obesity[2]. As obesity rapidly increases, the prevalence of NAFLD is increasing globally, ranging from approximately 30% in the general population to approximately 80% in morbidly obese individuals[3]. NAFLD, a spectrum of liver abnormalities ranging from NAFLD to nonalcoholic steatohepatitis (NASH), is predicted to be the most common indication for liver transplantation by 2030[4, 5]. NAFLD is characterized by excessive fat accumulation and is a major risk factor for the development of NASH, liver fibrosis, and cirrhosis[6]. Early diagnosis and treatment are critical for reducing associated complications and mortality.

For centuries, physicians in the clinic have employed several techniques to detect NAFLD, of which liver biopsy has been evaluated as the gold standard, yet it is considered intrusive and costly[7]. Some radiological techniques and ultrasonography are effective alternatives for liver biopsy, but they have limited access to remote areas due to the high cost of instruments and tests[8]. Therefore, noninvasive and inexpensive NAFLD diagnostic techniques are extremely promising.

Research Article
Manuscript received on September 4, 2023; accepted on March 29, 2024
Recommended by Associate Editor Daoqiang Zhang
Colored figures are available in the online version at https://link.springer.com/journal/11633
†These authors contributed equally to this work

In previous studies, there have been several alternative methods for detecting NAFLD. Leung et al.[9] used a machine learning model to classify NAFLD according to human serum and stool. In addition, Fibrosis-4 (FIB-4), nonalcoholic fatty liver disease fibrosis score (NFS), and neck circumference have also been studied for the diagnosis of NAFLD[10, 11]. However, most of the above methods are concentrated only on unilateral factors, which may be due to the lack of comprehensive data. Considering multifaceted perspectives for NAFLD prediction could be effective.

To predict NAFLD from multiple perspectives, we intend to create a comprehensive clinical database that includes questionnaires, physical examinations, laboratory tests, and imaging examinations (routine blood examination, urinalysis, and so on). In fact, the image of the face is a convenient window into the internal organs′ function. Facial images have been used as an important diagnostic tool in traditional Chinese medicine and Western medicine clinical fields[12–14]. At present, studies have proven that human facial features can reflect developmental syndromes, biological age, and the aging degree of organs[15–17]. Many studies have fully proven the auxiliary value of facial images in disease diagnosis, which can easily and conveniently help clinicians make disease judgments, especially in traditional Chinese medicine. Recently, deep convolutional neural networks (CNNs), as one of the most efficient networks in computer vision[18–20], have been widely used for image-based disease diagnosis such as heart disease and small-bowel disease[21–23]. CNN-based deep-learning algorithms have achieved near-human-level performance in disease classification, and even surpassed humans in subtle points that humans cannot observe. Therefore, facial images, which can be acquired rapidly, noninvasively, and freely, may be potential and essential information for the screening and prediction of NAFLD. In this study, we aim to build comprehensive clinical data, including facial images, physical examinations, and so on. To the best of our knowledge, no prior studies have exploited the association between facial images and NAFLD.

In this paper, an NAFLD diagnosis system is developed to distinguish NAFLD using multimodal input, encompassing facial images and metadata. First, we compile a comprehensive medical dataset FLDData, by gathering physical tests, laboratory and imaging studies, questionnaires, and facial images from a pool of volunteers. Next, we employ a collaborative approach utilizing a joint indicator-based data analysis to quantitatively examine and identify the clinical metadata that holds the most relevance to NAFLD within the medical dataset. Based on the selected data, we propose a multimodal-based NAFLD prediction method DeepFLD, which incorporates both facial images and metadata. Due to the intricate nature of NAFLD, it is difficult to extract effective features directly from facial images. In DeepFLD, a medical constraint-based auxiliary task is designed to extract valid image features. Compared with the indicators selected by considering only the Pearson correlation coefficient, the indicators we selected can improve the classification accuracy of NAFLD. The proposed DeepFLD model with multimodal input exhibits superior performance compared to models that rely solely on metadata as input. The DeepFLD model demonstrated satisfactory performance when applied to previously unseen data. It can achieve competitive results using only facial images as input rather than metadata, which is encouraging.

In summary, our main contributions of this paper are as follows:

1) A comprehensive human clinical dataset is constructed by aggregating facial images, physical examination data, laboratory test data, imaging information, and questionnaires. We employed a joint indicator-based data analysis method to quantitatively examine the key medical indicators associated with NAFLD.

2) We propose an NAFLD prediction method DeepFLD with multimodal input and medical constraints, that facilitates valid feature extraction from the high-dimensional facial images. As the first to introduce facial images into NAFLD prediction, DeepFLD with multimodal input outperforms other methods with only metadata as input, and verifies satisfactory performance on other unseen datasets. Furthermore, compared to metadata, DeepFLD can achieve competitive results with only facial images as input, providing a viable method for a more robust and simpler noninvasive diagnosis of NAFLD.

3) We analyze the NAFLD prediction results by exploring the facial characteristics of people with NAFLD. Among these characteristics, dark skin color and the presence of melasma can be supported by previous medical studies.

The remainder of this paper is organized as follows: In Section 2, we give a review of the related work; Section 3 introduces the prediction system, including the collected dataset and the prediction models; and Section 4 presents our experiments and results from the collected dataset and analyses of lifestyles. Finally, we conclude the study in Section 5.

2 Related works

2.1 NAFLD diagnosis

For decades, there have been many ways to detect NAFLD. Liver biopsy has been evaluated as the gold standard for diagnosing the presence and extent of liver inflammation and fibrosis, but it has the disadvantage of being invasive[7]. Several radiological techniques and ultrasonography are effective alternatives to liver biopsy for quantification of hepatic steatosis (computed tomo-

graphy, magnetic resonance imaging, or magnetic resonance spectroscopy) and fibrosis (transient elastography), but screening of fatty liver populations is more problematic due to expensive or limited access to remote areas[8, 24]. Invasive nature, high cost, and inconvenience hinder their detection in large populations and screening in remote populations.

Several alternative methods for detecting NAFLD have been identified in previous studies, but the search for alternatives for sensitive NAFLD detection is still ongoing. Leung et al.[9] explored human serum and feces for metabolomics studies to predict NAFLD. A machine learning method was used to predict NAFLD in 180 participants with a final area under the curve (AUC) of 0.72–0.80. Furthermore, several studies of alternative diagnostic methods have investigated the expression status of genes cholesterol, triglyceride, and other lipid metabolism and identified NAFLD-related genes[25]. Stender et al.[26] showed that obesity significantly amplifies the effects of three gene sequence variants associated with NAFLD. The synergistic effect between adiposity and genotype promotes NAFLD pathology, from steatosis to liver inflammation to cirrhosis. At the same time, there are also studies on the diagnosis and prediction of NAFLD by scoring. Fojas et al.[10] explored the effect of neck circumference on NAFLD by recruiting 674 patients. Neck circumference is found to be associated with fatty liver prevalence, but it cannot predict NAFLD on its own. By recruiting 5 129 volunteers and measuring their FIB-4 and NFS, Graupera et al.[11] explored whether FIB-4 and NFS could be used to screen for NAFLD, while the results are not satisfactory. However, although these indicators are convenient, they cannot be used to accurately evaluate and screen for disease. Diagnosis by genomics or metabolomics is relatively reliable but difficult for large-scale screening and monitoring in community-based populations.

2.2 Multimodal learning

In the realm of multimodal deep learning, fusion techniques play a crucial role in harmonizing data from diverse sources. Current research has predominantly focused on three fusion paradigms: early, late and deep fusion[27]. Early fusion integrates features immediately after extraction, capitalizing on the inherent correlations between modalities[28]. This method is particularly effective when the modalities share a strong relationship, yet it can be challenging to capture the nuanced correlations required for optimal fusion[29]. Late fusion, on the other hand, merges the outputs from models trained independently on each modality[30]. This approach is favored for its agnosticism to feature extraction, making it ideal for scenarios where modalities exhibit low correlation. It harnesses the strengths of individual models by combining their predictions post-training[31]. Deep fusion represents a more sophisticated approach, blending high-level features with raw data to harness the full spectrum of information provided by each modality[32]. The TransFG[33] method converts images into image patches, processes them through an embedding layer, and then concatenates the features with those of blood indicators after embedding. These combined features are then input into a vision transformer to extract and fuse features for clinical diagnostics. This technique is especially valuable when modalities offer complementary insights, as it allows for the integration of both abstracted and detailed data[34]. Our proposed approach, which integrates facial and metadata features, plans to adopt deep fusion to capitalize on the synergistic benefits of both feature levels. This strategy enables our model to achieve enhanced performance in the classification task, effectively leveraging the rich, multifaceted nature of the data.

3 The DeepFLDDiag system

In this section, we present the proposed NAFLD diagnosis system, called DeepFLDDiag, which comprises three main components: data collection, data processing, and disease prediction, as shown in Fig. 1. First, the dataset FLDData is constructed, encompassing the facial image data and medical indicators. Second, a quantitative data analysis method is employed to examine and identify the clinical information that exhibits the strongest correlation with NAFLD in the medical datasets. Finally, we design an NAFLD prediction method called DeepFLD. In particular, it is crucial to note the following points: 1) The data we collected contains very comprehensive medical information about volunteers: face data and a set of 480–dimensional indicators, including physical examination data and questionnaire data; 2) during the data analysis process, we not only account for the influence of individual factors on the data, but also employ a joint indicators approach to uncover the combined impact of multiple factors on the outcomes; and 3) in DeepFLD, we designed an auxiliary task based on multimodal data fusion and, medical constraints to extract valid image features. Sections 3.1–3.4 specifically describe each component.

3.1 FLDData

FLDData is compiled by gathering information from more than 6 000 volunteers. The present study used an observational, prospective, community-based cohort dataset. All participants underwent a physical examination, laboratory and imaging studies, extensive questionnaires, and facial image acquisition. Ethics approval for this study is obtained from the Medical Ethics Committee, Staff Hospital. All participants originally signed a written informed consent form.

3.1.1 Data acquisition

The physical examination included measurements of

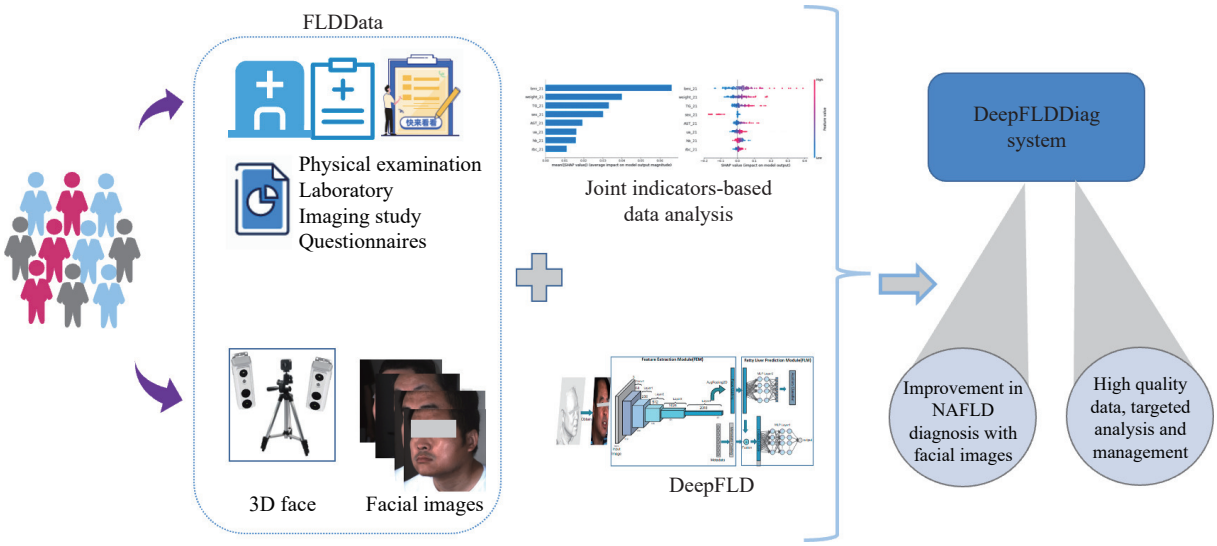


Fig. 1 The AI system DeepFLDDiag for NAFLD prediction. We captured volunteers′ facial image data using a 3D face instrument and collected their medical metadata from their physical examinations, extensive questionnaires, and so on. Through the proposed data analysis and machine learning method, we can improve NAFLD diagnosis with facial images and high-quality data, targeted analysis and management. (Colored figures are available in the online version at https://link.springer.com/journal/11633)

the patient′s height, weight, waist circumference, hip circumference, and blood pressure. Blood samples for laboratory-based examinations were collected after an overnight fast. Imaging results included ultrasonography and computed tomography. Moreover, information on medical history, medication use, lifestyle, and other sociodemographic factors was obtained via interviews and questionnaires. Including the information mentioned above, each participant′s 480-dimensional metadata recorded in FLDData. To explore the relationship between facial images and NAFLD, we employed a camera to capture the RGB and depth images and 3D images of the participants. Participants stood in the assigned spots for 1 minute while looking directly into the camera for the facial images to be taken. To achieve a balanced distribution of data, efforts are made to collect data that are as evenly distributed as possible. The cross-sectional analysis included a total of 6 760 participants, 49.9% of whom had NAFLD and 68.9% of whom were male. The average age of the participants was 44 years old.

3.1.2 Assessment of NAFLD

NAFLD was diagnosed via abdominal ultrasonography (ACUSON X300, Siemens, Munich, Germany) using a 3.5-MHz probe that is performed on all participants by skilled sonographers following a standardized protocol. We assessed NAFLD according to the standard criteria established by the Asia-Pacific Working Party on NAFLD and the Chinese Association for the study of liver disease[35, 36]. After the exclusion of subjects with excessive alcohol intake or other hepatic diseases, participants with NAFLD must present two or more of the following abnormal characteristics: 1) diffusely increased liver near-field echogenicity relative to the kidney; 2) ultrasound beam attenuation; and 3) poor visualization of intrahepatic structures.

3.2 Joint indicator-based data analysis

In FattyLiverData, we have collected 480 indicators for each participant, including the content of physical examinations such as height and weight, and the content of questionnaires on living habits. We intend to select the most relevant data for NAFLD from these indicators because they are redundant and repetitive.

The correlation coefficient is a basic similarity metric between variables in statistics and is used to analyze FLDData in this paper. Although we can obtain the correlation between a certain indicator and NAFLD, we cannot obtain the correlations between multiple indicators and NAFLD. NAFLD is known to be a complex disease caused by multiple factors. So, these indicators selected by the correlation coefficient do not certainly contribute to NAFLD prediction. In order to obtain effective indicators for the prediction of NAFLD, we consider the impact of multiple indicators on NAFLD. Specifically, we designed a joint indicator-based data analysis method that combines the correlation coefficient (Pearson correlation coefficient) and the model analysis method (SHAP) to obtain the effect of a combination of indicators on NAFLD. Then the most helpful indicators for NAFLD prediction were selected from FLDData.

3.2.1 Pearson correlation coefficient

As a common correlation coefficient, the Pearson correlation coefficient, which ranges from −1 to 1, can measure the correlation between two variables. First, we used the Pearson correlation coefficient to analyze FLDData and identify the indicators that are most closely related to NAFLD. In mathematical form, Pearson correlation

coefficient of two variables (p, q) can be described as follows:

$$\rho(p,q) = \frac{\sum pq - \frac{\sum p \sum q}{N}}{\sqrt{(\sum p^2 - \frac{(\sum p)^2}{N})(\sum q^2 - \frac{(\sum q)^2}{N})}} \quad (1)$$

where N is the total number of attributes. The symbol $|\rho(p,q)| = 1$ indicates that p and q are perfectly correlated, corresponding to the modified Euclidean distance. The stronger the correlation between (p, q) is, the greater the value of $|\rho(p,q)|$ is.

Considering that many indicators in our data are very redundant for NAFLD, we selected 21 indicators by calculating the Pearson correlation coefficient between NAFLD and all other indicators, as shown in Table 1. In Table 1, continuous variables are expressed as means and standard deviations, and categorical variables are expressed as frequencies and percentages. In addition, the 21 selected indicators are analyzed by statistical analysis (SAS 9.4). The 21 selected indicators are compared using two-sample independent t-tests, and $P < 0.05$ is set as the level of significance. The NAFLD group's indicators have significantly higher means than the NON-NAFLD group's. Meanwhile, significant differences are observed in the indicators of the NAFLD / NON-NAFLD groups by statistical analysis ($P < 0.05$).

3.2.2 Shapley additive explanations (SHAP)

SHAP[37] quantifies the impact of each indicator on the results in the model and is often used to improve the interpretability of black box models. For the selected 21 indicators, we adopted SHAP to further analyze the contribution of each indicator to the NAFLD prediction. According to expert experience, people who smoke and drink are also prone to acquiring NAFLD. However, they are not selected by Pearson correlation coefficient. Therefore, we added the two indicators to the 21 selected indicators to be considered together. First, we built a neural network for NAFLD prediction with the 23 indicators (denoted as $Metadata_{23}$) as inputs. Then we calculated the SHAP value for each indicator, which contributes to the NAFLD prediction. For the set $U \subseteq \{x_1, x_2, \cdots, x_p\}$ of selected indicators, where x_j denotes the j-th indicator, the SHAP value of x_j is calculated as follows:

Table 1 Distribution and order of some indicators according to Pearson correlation coefficient

Variables	Total	NAFLD	NON-NAFLD	ρ_1	Note
BMI (kg/m²)[2]	24.4 ± 3.7	26.3 ± 3.2	22.0 ± 2.4	0.602	A body mass index
WEIGHT (kg)[2]	67.3 ± 16.9	76.0 ± 12.8	61.0 ± 9.3	0.547	Weight
HLP n (%)	3 024.0 (46.3)	2 346.0 (63.5)	678.0 (24.0)	0.392	Hyperlipidaemia
UA (umol/L)[2]	354.0 ± 95.6	385.2 ± 95.2	312.0 ± 78.1	0.381	Uric acid
TG (mmol/L)[2]	1.8 ± 1.6	2.3 ± 1.9	1.2 ± 0.6	0.362	Triglyceride
OBE n (%)	942.0 (15.4)	911.0 (26.3)	31.0 (1.2)	0.346	Obesity
DBP (mmHg)[2]	75.2 ± 12.3	78.5 ± 12.3	70.3 ± 10.4	0.332	Diastolic blood pressure
SBP (mmHg)[2]	124.1 ± 18.8	129.6 ± 18.8	117.1 ± 16.0	0.329	Systolic blood pressure
APOB (g/L)[2]	0.9 ± 0.2	1.0 ± 0.2	0.8 ± 0.2	0.328	Apolipoprotein B
HGB(g/L)[2]	140.1 ± 15.9	145.2 ± 14.8	134.8 ± 15.4	0.324	Hemoglobin
RBC (10^{12}/L)[2]	4.5 ± 0.4	4.6 ± 0.4	4.4 ± 0.4	0.321	Red blood cell
MALE n (%)	3 351.0 (51.0)	2 394.0 (64.4)	957.0 (33.5)	0.306	Male
AST (U/L)[2]	23.6 ± 11.7	25.5 ± 12.5	20.9 ± 8.8	0.304	Aspartate aminotransferase
HUA n (%)	1 828.0 (28.0)	1 431.0 (38.7)	397.0 (14.0)	0.272	Hyperuricemia
HPT n (%)	1 441.0 (23.0)	1 149.0 (32.4)	292.0 (10.7)	0.255	Hypertension
HDL (mmol/L)[2]	1.4 ± 0.4	1.3 ± 0.3	1.5 ± 0.4	0.252	High density lipoprotein cholestero
WBC (10^9/L)[2]	6.6 ± 1.6	6.9 ± 1.7	6.1 ± 1.5	0.242	White blood cell
LDL(mmol/L)[2]	2.0 ± 0.7	2.2 ± 0.7	1.8 ± 0.7	0.238	Low density lipoprotein cholestero
APOA (g/L)	1.4 ± 0.3	1.3 ± 0.3	1.5 ± 0.3	0.223	Apolipoprotein A
ALP (U/L)[2]	71.8 ± 23.6	75.7 ± 20.9	66.0 ± 22.2	0.219	Alkaline phosphatase
FBG (mmol/L)[2]	5.8 ± 1.5	6.0 ± 1.7	5.4 ± 1.0	0.217	Fasting blood glucose

[1] ρ denotes the absolute value of Pearson correlation coefficient. The value of ρ indicates the strength of the relationship between the indicator and the disease. As the value of ρ increases, the correlation between the indicator and the disease increases.
[2] Results are presented as mean ± s.d.

$$\phi_j = \sum_{S \subseteq \{x_1, \cdots, x_p\} \setminus x_j} \frac{\|S\|!(p - \|S\| - 1)!}{p!} \times (f_x(S \bigcup x_j) - f_x(S)) \quad (2)$$

where $\{x_1, \cdots, x_p\} \setminus x_j$ denote the subset V without the indicator x_j, S is a subset of set V, and $\frac{\|S\|!(p-\|S\|-1)!}{p!}$ refers to the probability of S occurring. The NAFLD prediction model $f_x(S)$ refers to the prediction performance when the indicator set S is used as the input.

Given the current set of indicators, the estimated Shapley value is calculated based on the actual prediction performance and the average prediction performance (shown in (2)).

By calculating the SHAP value, we obtained the indicator contribution ranking, shown in Fig. 2. It shows that body mass index (BMI) has the largest contribution to NAFLD prediction. Compared to the ranking of Pearson correlation coefficients (Table 1), the ranking of the SHAP values changed significantly. The two added indicators smoke and drink play an important role in the network. Therefore, through the data analysis, we can choose the indicators with the greatest effectiveness.

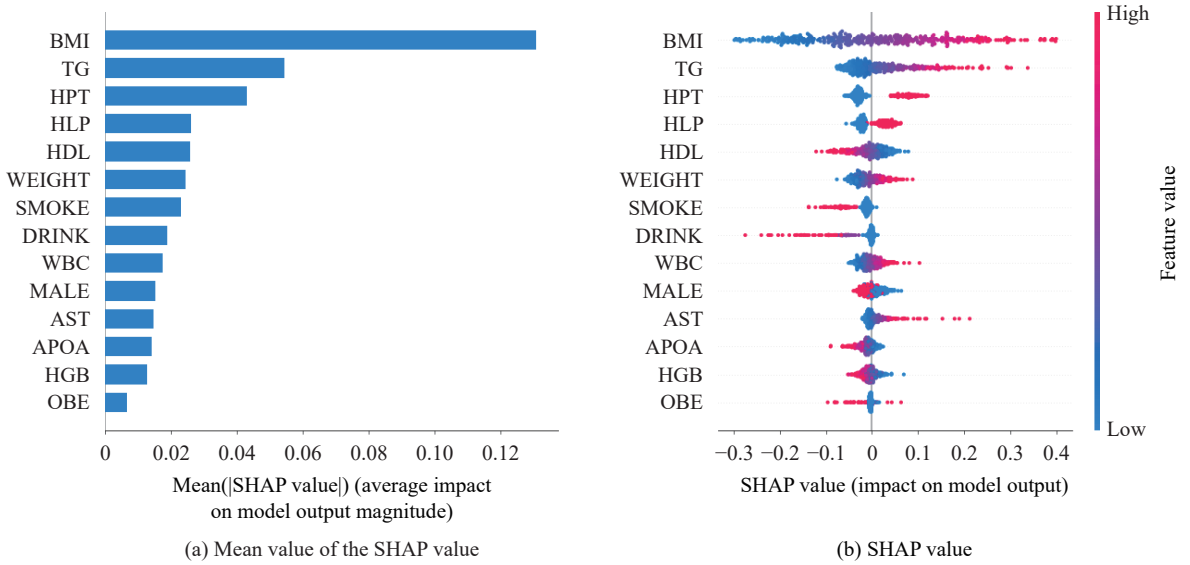
3.3 DeepFLD prediction method

Both metadata and face data contain crucial information for diagnosing NAFLD, although face data has not been utilized to the best of our knowledge. A multimodal fusion neural network (called DeepFLD) is used to integrate the metadata and face data, as shown in Fig. 3. The proposed DeepFLD model contains the feature extraction module (FEM) and the fatty liver prediction module (FLM). In the feature extraction module FEM, the facial images are input into a multilayer convolutional neural network to obtain the feature coding from facial images. An embedding coder layer is adopted for extracting feature coding from metadata. The prediction module FLM then combines facial and metadata coding to predict NAFLD. We also designed an auxiliary task to extract effective features and accelerate convergence during the feature extraction module FEM training process.

3.3.1 Feature extraction module

Generally, the facial appearance, features and expressions of patients are used by physicians to assess their health status. Theoretically, the facial images can contain all the features a physician requires to determine health status. So, we put the facial images into the feature extraction module. The feature extraction module is responsible for extracting high-dimensional features from the input facial images.

Fig. 3 shows the architecture of $\mathbf{FEM}_{CNN}$. A three-channel facial image $\mathbf{I}_{image} \in \mathbf{R}^{m \times n \times c_f}$, $m = 512$, $n = 512$, $c_f = 3$, is fed into multiple convolutional blocks and an average pooling layer ($\mathbf{FEM}_{CNN}$), which gradually reduces the resolution and increases the channel dimension, and finally outputs a one-dimensional feature vector $\mathbf{Z}_{image} \in \mathbf{R}^{1 \times 1 \times c_{fc}}$. During the extraction of the feature vector $\mathbf{Z}_{image}$ from facial image, it is called face coding. The $\mathbf{FEM}_{CNN}$ contains Conv1 layer (7×7 convolutional operation) and 4 layers. Each layer contains two 3×3 convolutional operations and a downsampling operation. After each layer, the spatial dimensions are halved, while the channel dimensions are doubled. The input image, initially with dimensions (512, 512, 3), undergoes a series of layers that reduce the spatial dimensions by half, while simultaneously doubling the number of channels. The final feature map, with dimensions (32, 32, 2 048), is



(a) Mean value of the SHAP value

(b) SHAP value

Fig. 2 Indicator rankings based on SHAP value. The indicator's contribution to the prediction of NAFLD increases with its SHAP value, with BMI making up the majority of that contribution. (Colored figures are available in the online version at https://link.springer.com/journal/11633)

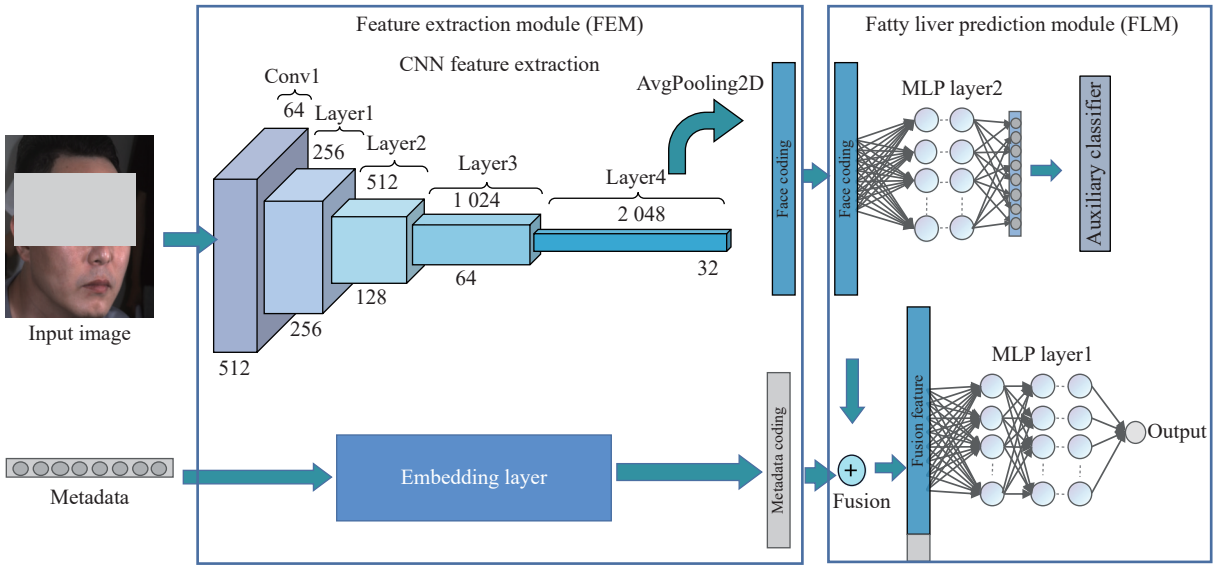


Fig. 3 The architecture of DeepFLD. It contains the feature extraction module (FEM) and the fatty liver prediction module (FLM). First, we adopt the FEM to extract the face coding and metadata coding. The FLM is used to fuse the two codings and predict NAFLD through an auxiliary task. (Colored figures are available in the online version at https://link.springer.com/journal/11633)

converted into a 1-dimensional vector $\mathbf{Z}_{image}$ with dimensions (1, 1, 2 048) through a pooling operation. The dropout value is set to 0.4.

The metadata including age, weight, height, and BMI, were extracted from the volunteers' physical examinations and clinical laboratory assays. Through the data analysis (3.2), we obtain the relevant and important metadata $\mathbf{I}_{metadata} \in \mathbf{R}^{1\times1\times c_m}$. An embedding layer is designed to normalize each metadata element to an N(0, 1) distribution and compute the metadata coding $\mathbf{Z}_{metadata} \in \mathbf{R}^{1\times1\times c_{mc}}$. Through the feature extraction module FEM, the inputs with different attributes in varying dimensions are extracted into features with the same dimension, which is facilitated by feature fusion on the fatty liver prediction module.

$$\begin{aligned}\mathbf{Z}_{image} &= \mathbf{FEM}_{CNN}(\mathbf{I}_{image})\\ \mathbf{Z}_{metadata} &= \mathbf{FEM}_{embedding}(\mathbf{I}_{metadata}).\end{aligned} \tag{3}$$

3.3.2 Fatty liver prediction module

The main task of the model is to predict NAFLD, so we built a multilayer perceptron (MLP) to diagnose whether a person has NAFLD. The whole system, especially the convolutional neural network in the feature extraction module FEM, contains many trainable parameters. The leading result is that it may not be possible to extract effective information well just by determining whether the sample has NAFLD or not. We have also conducted some experiments in which we input facial images and metadata to directly predict NAFLD. However, the results were not good enough to extract valid features related to NAFLD. Therefore, we created an auxiliary task to help train the model. The auxiliary task aims to facilitate neural network training and convergence, so some indicators related to NAFLD are considered to be selected for prediction. In this paper, the auxiliary task is to predict the three indicators: gender, BMI and weight.

We used MLP to build the main task network $\mathbf{FLM}_{MLP_1}$ for the fatty liver prediction and the auxiliary task network $\mathbf{FLM}_{MLP_2}$ for the facial image feature extraction in the fatty liver prediction module FLM. Each layer of the MLP is composed of multiple neurons, and the output of the previous layer is the input of the next layer. It is also called the fully connected (FC) layer. In order to improve the performance of the MLP and prevent overfitting, we added a nonlinear activation layer (ReLU function) and a dropout function after each FC layer. $\mathbf{FLM}_{MLP_1}$ (fatty liver prediction) uses the fusion data $\mathbf{Z}_{fusion} \in \mathbf{R}^{(c_{fc}+c_{mc})\times1}$ as input, which includes face and metadata coding. $\mathbf{FLM}_{MLP_2}$ (image feature extraction) on the other hand, only accepts face coding $\mathbf{Z}_{image}$.

$$\mathbf{Z}_{fusion} = \mathbf{Concat}(\mathbf{Z}_{image}, \mathbf{Z}_{metadata}) \tag{4}$$

$$\hat{\mathbf{Z}}_{fusion} = \frac{(\mathbf{Z}_{fusion} - \mu)}{\theta}. \tag{5}$$

We first normalize all the input elements, subtract their mean μ and divide their variance θ to make them into a positive-terminus distribution. Then, based on the fusion data $\hat{\mathbf{Z}}_{fusion}$, set the number of the first channels in $\mathbf{FLM}_{MLP_1}$ (fatty liver prediction) and $\mathbf{FLM}_{MLP_2}$ (image feature extraction). $\mathbf{FLM}_{MLP_1}$ (fatty liver prediction) has 6 FC layers, each with 2 056, 1 024, 1 024, 512 256, and 128 nodes. $\mathbf{FLM}_{MLP_2}$ (image feature extraction) has three FC layers, each with 2 048, 1 024, and 1 024 nodes.

After the fusion data is subjected to multiple nonlinear mappings in the $\mathbf{FLM}_{MLP_1}$ (fatty liver prediction) hidden layer, a regression value y_{fat} is obtained and then normalized by the sigmoid function, and the final output is the prediction of fatty liver. Meanwhile, face coding Z_{image} is fed into $\mathbf{FLM}_{MLP_2}$ (image feature extraction), and the auxiliary task regression results $\boldsymbol{y}$ are output.

$$\begin{aligned} y_{fat} &= \mathbf{FLM}_{MLP_1}(\hat{\mathbf{Z}}_{fusion}) \\ \boldsymbol{y} &= \mathbf{FLM}_{MLP_2}(\mathbf{Z}_{image}). \end{aligned} \tag{6}$$

Convolutional neural networks trained by auxiliary tasks can also achieve NAFLD diagnosis when the metadata is removed and only images are input. However, the network without auxiliary task training cannot diagnose NAFLD very well. It shows that the designed auxiliary task is very beneficial for image training.

3.3.3 Training

The entire network is trained end-to-end, and the loss function is a joint loss $\boldsymbol{L}$, which includes the NAFLD task loss $\boldsymbol{L}_{NAFLD}$ as well as the auxiliary task loss $\boldsymbol{L}_{Auxi}$,

$$\boldsymbol{L} = \alpha \boldsymbol{L}_{NAFLD} + (1-\alpha)\boldsymbol{L}_{Auxi}$$

where α is 0.7 in this paper. The main task is to predict whether the person has NAFLD or not, which is a binary classification task. Therefore, we adopted classical cross-entropy. The output $y_{fat} \in [0,1]$ of module $\mathbf{FLM}_{MLP_1}$ (fatty liver prediction) represents the probability that the sample has NAFLD. The NAFLD task loss $\boldsymbol{L}_{NAFLD}$ can be expressed as follows:

$$\boldsymbol{L}_{NAFLD} = -(\hat{y}_{fat}\log(y_{fat}) + (1-\hat{y}_{fat})\log(1-y_{fat})) \tag{7}$$

where $\hat{y}_{fat} \in \{0,1\}$ denotes the ground truth indicating whether the person has NAFLD or not. When $\hat{y}_{fat}=1$, $\boldsymbol{L}_{NAFLD} = -\log(y_{fat})$. Minimizing the loss $\boldsymbol{L}_{NAFLD}$ is equivalent to increasing y_{fat}, i.e., increasing the likelihood of having a nonalcohol fatty liver. When $\hat{y}_{fat}=0$, $\boldsymbol{L}_{NAFLD} = -\log(1-y_{fat})$. Minimizing the loss $\boldsymbol{L}_{NAFLD}$ is equivalent to making y_{fat} decrease, namely decreasing the probability of having NAFLD.

4 Experiments

For the proposed NAFLD model, there are several aspects to be asked about: 1) The NAFLD prediction results with indicators input, and whether adding images can improve the prediction performance? 2) Good prediction results are obtained when migrated to other data? 3) What are the prediction results of images as input alone? and 4) do the method prediction results have any explanations? Hence, we performed a series of experiments.

We compare the performance of the proposed DeepFLD method with multimodal input and the models with metadata as input, to answer the Question 1) (see Section 4.3.1). To answer Question 2), we migrate the trained model to an unseen dataset collected in other years (see Section 4.3.3). And conducting experiments only using facial images as input to answer Question 3), (see Section 4.3.4). Moreover, ablation studies are conducted to validate 1) the significance of the clinical metadata selected by the joint indicator-based data analysis method and 2) the effectiveness of the proposed DeepFLD method with facial image input. Finally, we analyze the relationship between the face and the NAFLD through visualization to answer Question 4) (see Section 4.5).

4.1 Experimental setup

Through data analysis, we selected eight clinical indicators that contributed to NAFLD prediction. Since gender affects the average level of some indicators and is easily obtained, we selected the top 7 indicators based on SHAP value along with gender to form 8 indicators for NAFLD prediction, denoted by $\mathbf{MetaData}_8^*$ = [BMI, TG, HPT, HLP, HDL, WEIGHT, DRINK, MALE]. And three non-traumatic indicators are chosen: $\mathbf{MetaData}_3^*$ = [BMI, WEIGHT, MALE]. For real-world clinical applications, we performed 5 different combinations of clinical metadata and facial images: $\mathbf{MetaData}_8^*$ (8 indicators), clinical metadata $\mathbf{MetaData}_3^*$ (3 indicators), the combination of clinical metadata $\mathbf{MetaData}_8^*$ (8 indicators) and facial images, the combination of clinical metadata $\mathbf{MetaData}_3^*$ (3 indicators) and facial images, and the facial images only.

To predict NAFLD under different types of data inputs, we build prediction methods corresponding to indicator input, image input, and joint input. We compare three models (traditional machine learning, deep machine learning, and the proposed DeepFLD) corresponding to clinical metadata as input, facial image input, and joint input. Random forest (RF) and MLP are the common and effective machine learning methods for many classification tasks with one-dimensional vector input and are adopted in NAFLD[38, 39]. To fit our FLDData, we modified the methods in [38] and [39]. In this paper, RF is a combination of 50 simple decision trees, of which the input is the medical indicators and the output is determined by the voting method. The input of MLP is the clinical metadata, and the output is the probability of the NAFLD prediction. Specifically, we first normalize all the clinical metadata, subtract their mean and divide their variance to make them into a (0, 1) distribution. Then set the number of channels in the input layer according to the input clinical metadata type (3 indicators or 8 indicators). After the input indicator is subjected to mul-

tiple nonlinear mappings in the hidden layer, a regression value is obtained and then normalized by the sigmoid function, and the final output is the prediction of NAFLD. The RF and MLP with 8 indicators are denoted as RF ($\mathbf{MetaData}_8^*$) and MLP ($\mathbf{MetaData}_8^*$), and those with 3 indicators are denoted as RF ($\mathbf{MetaData}_3^*$) and MLP ($\mathbf{MetaData}_3^*$), respectively. It is crucial to highlight that when the DeepFLD model operates solely on metadata input, it essentially reverts to an MLP architecture, meaning that DeepFLD ($\mathbf{MetaData}_8^*$) is equivalent to MLP ($\mathbf{MetaData}_8^*$).

In DeepFLD, we have employed a multimodal disease diagnosis approach that utilizes CNNs for feature extraction. To further validate our method, we conduct comparative experiments with state-of-the-art multimodal disease diagnosis models, specifically one that relies on vision transformers (ViT) for feature extraction named TransFG[33]. In this paper, we adopted a configuration same to that of [33], where TransFG is utilized to process images and metadata. The images are first sliced into patches, which are then converted into tokens. These image tokens, in concat with metadata embeddings, are subsequently input into the vision transformer (ViT) model with a 9 transformer layers for extracting fusion features.

For the proposed model DeepFLD, we conduct experiments with three types of inputs: images only, the combination of clinical metadata (8 indicators) and facial images, and the combination of clinical metadata (3 indicators) and facial images. Different from the CNN model, the DeepFLD model for facial image input has an auxiliary task that is to help the network learn to extract useful features. So we also conduct the experiment to compare the performance of CNN and DeepFLD with images as the only input. Due to missing data, we selected 676 samples containing images and 8 indicators for the following NAFLD prediction experiments.

4.2 Metric

To meet various screening requirements or clinical applications, we take the common metric containing classification accuracy (ACC), specificity (SP), positive predictive value (PPV), and area under curve (AUC) of different combined models for NAFLD detection. ACC describes the classification accuracy of the classifier as follows:

$$\mathrm{ACC} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN} + \mathrm{TN}}$$

where TP, TN, FP and FN denote true positive samples, true negative samples, false positive samples, and false negative samples, respectively. SP describes the ratio of the negative samples predicted by models to all negative samples in the dataset:

$$\mathrm{SP} = \frac{\mathrm{TN}}{\mathrm{FP} + \mathrm{TN}}.$$

PPV describes the ratio of the predicted true positive samples to all predicted positive samples:

$$\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}.$$

AUC is defined as the area enclosed by the receiver operating characteristic curve (ROC) and the coordinate axis, and its value ranges from 0 to 1. It indicates the probability that the predicted positive sample is ahead of the negative sample. For each of the aforementioned metric values, the larger the value is, the better the classification effect is.

4.3 Experimental results

4.3.1 NAFLD prediction with multimodal input

In this paper, we compare the NAFLD prediction results of DeepFLD with multimodal input and the models with metadata as input. For the fairness of experiments, we adopted K-fold cross-validation, which means the dataset is divided into K parts for cross-validation. Because the feature extraction network requires enough training data, if K is too small, the network will not converge sufficiently. If K is too large, the ratio of training sets to testing sets will be high, and data for testing will be less. Therefore, we choose K as 7 here. Every cross-validation is performed seven times. Each cross-validation is reiterated 7 times. Since AUC is a comprehensive metric to evaluate the effectiveness of models, we focus on considering this metric as the basis for model comparison.

Fig. 4 shows the experimental performance, which is assessed by 7-fold internal cross-validated metric. To predict NAFLD with metadata as input, we selected two classical models: RF[38] and MLP[39], which are effectively adapted for classification tasks with vector inputs. And there are 2 types of metadata as inputs: the selected 3 indicators $\mathbf{MetaData}_3^*$ and the selected 8 indicators $\mathbf{MetaData}_8^*$. Among them, $\mathbf{MetaData}_3^*$ contains gender, BMI and weight, which are easily available and noninvasive. Their results range from 87.4% to 91.3%, with MLP outperforming RF.

To obtain the NAFLD prediction results with multimodal data, including facial images and metadata, we performed two experiments involving DeepFLD (image + $\mathbf{MetaData}_3^*$) and DeepFLD (image + $\mathbf{MetaData}_8^*$). Compared with the indicator-input models, the proposed DeepFLD achieves 92.3% in AUC, 87.5% in ACC metric, 82.5% in SP and 83.9% in PPV, which exceeds at least 1% in AUC, 1.1% in ACC, 5.4% in SP, 2.4% in PPV, respectively. These findings illustrate that facial images actually contribute to the diagnosis of NAFLD.

From Fig. 4, we can see that RF ($\mathbf{MetaData}_8^*$) ex-

ceeds RF ($\mathbf{MetaData}_3^*$) 2.6%, MLP ($\mathbf{MetaData}_8^*$) exceeds MLP ($\mathbf{MetaData}_3^*$) 0.5% and DeepFLD (image + $\mathbf{MetaData}_8^*$) exceeds DeepFLD (image + $\mathbf{MetaData}_3^*$) 0.8% in the comprehensive metric AUC. It may be due to that more indicators can provide more information and improve model performance. Moreover, DeepFLD ($\mathbf{MetaData}_3^*$) also beyonds MLP ($\mathbf{MetaData}_8^*$) in AUC, to show that NAFLD can be diagnosed only through facial image and three noninvasive indicators [AGE MALE BMI]. It may be due to the fact that facial images can provide more information than the indicators [TG HPT HLP HDL DRINK].

Fig. 5 presents a plot of the training and validation loss over the course of the training process. The plot reveals that the loss values exhibit a sharp decline during the initial 50 epochs, signifying the model's rapid learning phase and its progressive convergence. The validation loss, which mirrors the trend of the training loss, also demonstrates a steady reduction. This parallel reduction in both loss metric is indicative of the model's convergence and its ability to generalize effectively. The validation loss, serving as a benchmark for the model's performance on unseen data, corroborates the model's learning efficacy and its capacity to adapt to new information.

4.3.2 Comparison with ViT model

We have compared DeepFLD with the state-of-the-art multimodal medical diagnostic approach TransFG[33] which uses ViT, as shown in Table 2. To ensure a comprehensive comparison, we evaluate the performance of both models under different input conditions: images, im-

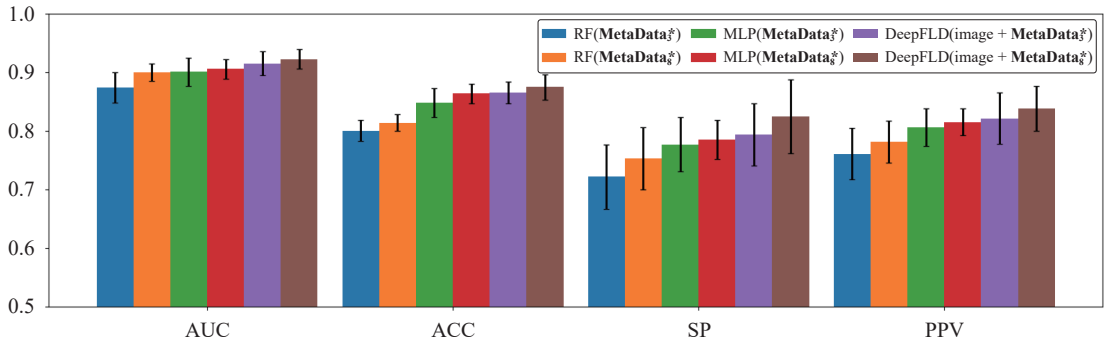


Fig. 4 The K-fold validation results of NAFLD prediction with different models, in which RF ($\mathbf{MetaData}_3^*$), MLP ($\mathbf{MetaData}_3^*$), RF ($\mathbf{MetaData}_8^*$), and MLP ($\mathbf{MetaData}_8^*$) denote the random forest and MLP models with the selected 3 indicators $\mathbf{MetaData}_3^*$ and 8 indicators $\mathbf{MetaData}_8^*$ as input, respectively. DeepFLD (image + $\mathbf{MetaData}_3^*$) and DeepFLD (image + $\mathbf{MetaData}_8^*$) denote the proposed DeepFLD model with the selected 3 indicators $\mathbf{MetaData}_3^*$ and facial images as joint inputs, and the selected 8 indicators $\mathbf{MetaData}_8^*$ and facial images as joint inputs, respectively. (Colored figures are available in the online version at https://link.springer.com/journal/11633)

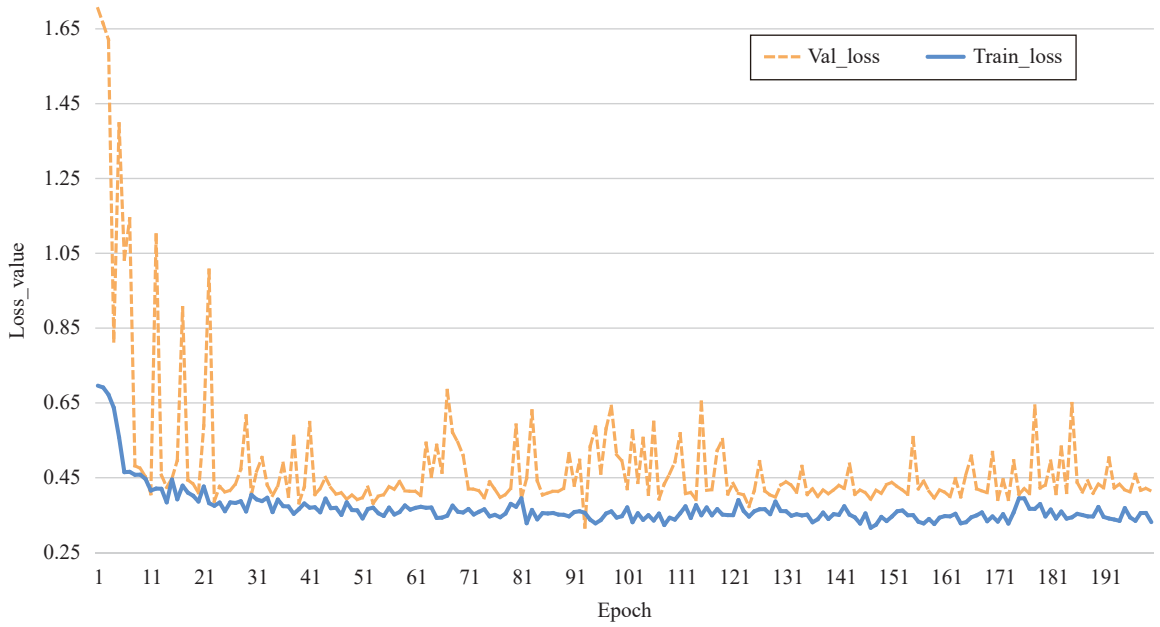


Fig. 5 The loss on training set and validation set. The convergence of the network is evident as the training and validation losses decrease and stabilize over time. (Colored figures are available in the online version at https://link.springer.com/journal/11633)

Table 2 Results of DeepFLD and ViT model comparison. The proposed DeepFLD outperforms the TransFG[33] (ViT) model in both image input and multimodal input scenarios.

Methods	AUC	ACC	SP	PPV
TransFG (images)	0.796 ± 0.030	0.742 ± 0.020	0.675 ± 0.080	0.701 ± 0.030
DeepFLD (images)	**0.863** ± 0.020	**0.818** ± 0.022	**0.738** ± 0.100	**0.782** ± 0.040
TransFG (images + $\mathbf{MetaData}_3^*$)	0.833 ± 0.040	0.784 ± 0.030	0.706 ± 0.050	0.715 ± 0.060
DeepFLD (images + $\mathbf{MetaData}_3^*$)	**0.915** ± 0.020	**0.865** ± 0.010	**0.849** ± 0.080	**0.858** ± 0.073
TransFG (images + $\mathbf{MetaData}_8^*$)	0.836 ± 0.080	0.805 ± 0.060	0.710 ± 0.100	0.765 ± 0.050
DeepFLD (images + $\mathbf{MetaData}_8^*$)	**0.923** ± 0.010	**0.875** ± 0.020	**0.825** ± 0.060	**0.845** ± 0.040

ages plus 3 indicators, and image plus 8 indicators. The results indicate that the DeepFLD method consistently outperforms the TransFG model across all the metric, both with image-only and multimodal inputs. DeepFLD achieves higher AUC, ACC, SP and PPV scores, demonstrating its effectiveness in disease diagnosis. The addition of metadata further enhances the performance of both models, with DeepFLD maintaining a clear advantage.

The results also suggest that for the task of NAFLD, where local features are important and the amount of data is limited, CNNs are more effective at feature extraction compared to ViTs. CNNs are able to learn the necessary patterns from the limited dataset, which is crucial for accurate diagnosis.

Fig. 6 illustrates the regions of interest (ROIs) identified by the DeepFLD and TransFG methods on original facial images. The DeepFLD method appears to focus on specific local areas of the face, which are more conducive to diagnosing NAFLD. In contrast, the TransFG method seems to capture a broader range of features, emphasizing global facial characteristics. This suggests that DeepFLD may be more effective in identifying the subtle local features that are indicative of NAFLD, whereas TransFG′s approach to feature extraction may not be as advantageous for this particular diagnostic task.

4.3.3 Results on the unseen dataset

In order to analyze whether the model overfits the current dataset, we migrate the trained model to the data which is collected in another year. The above results are based on data collected in 2021. The trained model on the 2021 data is migrated to the 2020 data without any fine-tuning, and the results are shown in Table 3. Since the datasets were collected in different years (2020 and

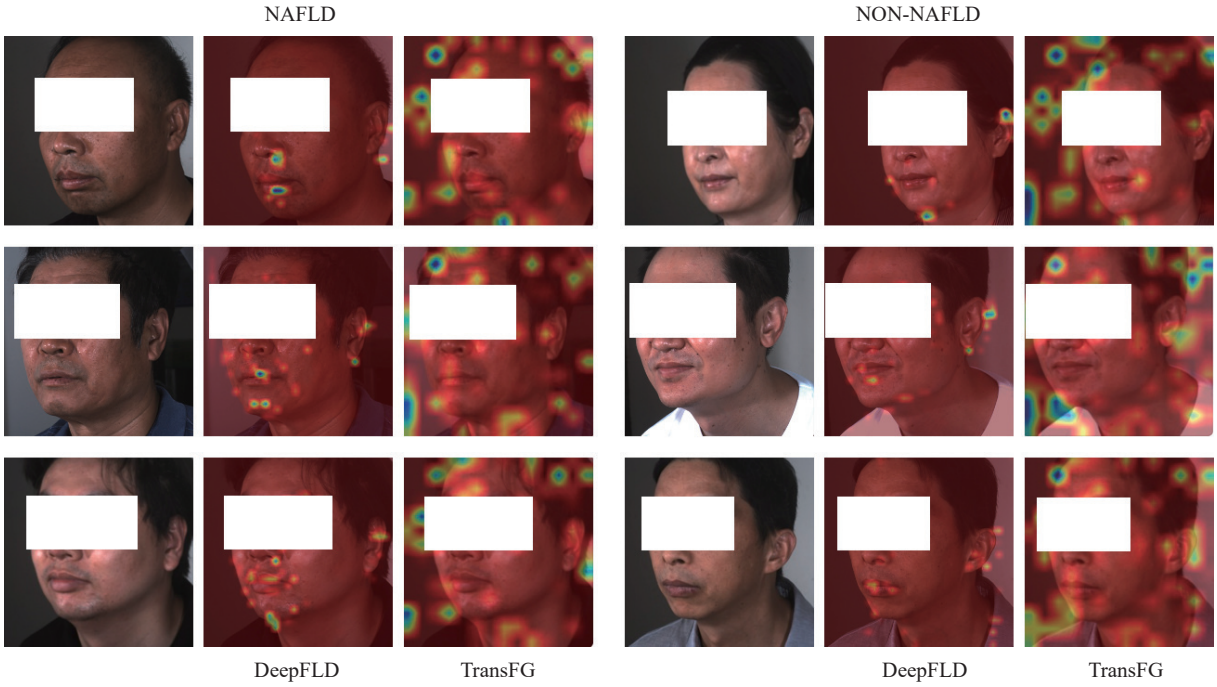


Fig. 6 Visualization of patients with NAFLD and healthy people (NON-NAFLD). We show the ROIs identified by the DeepFLD and TransFG methods on original facial images. (Colored figures are available in the online version at https://link.springer.com/journal/11633)

2021), the differences in the collection instruments lead to some differences in the distribution of indicators and facial images. Although the performance in the 2020 data is not as good as that in the 2021 data, most models are still satisfactory, achieving an 80% in AUC. These methods do not overfit the training dataset. Moreover, the proposed DeepFLD with multimodal input has the best performance, with an improvement of approximately 2%, which indicates that it has satisfactory generalizability.

Table 3 Results of different models on new data. The second column shows the results of cross-validations on the 2021 data. The third column shows the results of models tested on the 2020 data, which are trained on the 2021 data.

Methods	AUC(2021)	AUC(2020)
RF ($\mathbf{MetaData}_8^*$)	0.900 ± 0.010	0.768
MLP ($\mathbf{MetaData}_8^*$)	0.906 ± 0.010	0.806
DeepFLD (images + $\mathbf{MetaData}_8^*$)	0.923 ± 0.010	**0.825**
DeepFLD (image)	0.863 ± 0.020	0.801

4.3.4 NAFLD prediction with only image input

To obtain NAFLD prediction results with only images as input, we performed two experiments: A cross-validation of DeepFLD with image input only on the 2021 data, and a migration of the trained DeepFLD model from the 2021 data to the 2020 data. The experimental results are shown in the last row of Table 3. The experimental result of migration decreases compared to the one before migration, but it is still acceptable with an AUC of 80.1%. That may because the brightness, tone and exposure of the images taken in different years are different. Inspiringly, DeepFLD with images only has competitive results compared to those with metadata. Moreover, the result AUC of the RF model after migration (80.1%) exceeded that of the RF model with metadata (76.8%) by 3.3% in AUC, paving the way for a more robust and simpler NAFLD diagnosis.

4.4 Ablation study

4.4.1 Effectiveness of the joint indicator-based data analysis

In the data analysis section (Section 3.2), we adopted Pearson correlation coefficient and SHAP to identify the most important clinical indicators for NAFLD diagnosis. Therefore, we conducted experiments with different clinical indicators. We selected the top 8 indicators ranked by Pearson coefficient, denoted as $\mathbf{MetaData}_8$ = [BMI, WEIGHT, HLP, UA, TG, OBE, DBP, SBP]. We compared the performances of the models under $\mathbf{MetaData}_8$ and $\mathbf{MetaData}_8^*$, as seen in Fig. 7. The DeepFLD and MLP models with $\mathbf{MetaData}_8^*$ as input exceed the models with $\mathbf{MetaData}_8$ as input in all the metric. $\mathbf{MetaData}_8$ contains 6 invasive indicators and $\mathbf{MetaData}_8^*$ only contains 4 invasive indicators. It may be due to the joint indicator-based data analysis by Pearson correlation coefficient and SHAP that the correlation between NAFLD and joint indicators can be obtained. While the data analysis by Pearson correlation coefficient only obtains the correlation between NAFLD and a single indicator.

4.4.2 Effectiveness of the medical constraints

To illustrate the effectiveness of the constraints added as auxiliary tasks, we performed comparison experiments, DeepFLD (image) without auxiliary task and DeepFLD (image) with the auxiliary task, which are denoted as DeepFLD (image) w/o auxiliary and DeepFLD (image) w/ auxiliary, as shown in Fig. 8. The poor results of DeepFLD (image) without the auxiliary task indicate that the auxiliary task can help the network converge and learn the effective features to help NAFLD identification. The proposed DeepFLD achieves satisfactory diagnosis results with the designed auxiliary task. NAFLD screening can be achieved with just a single facial image through DeepFLD, which has a high practical

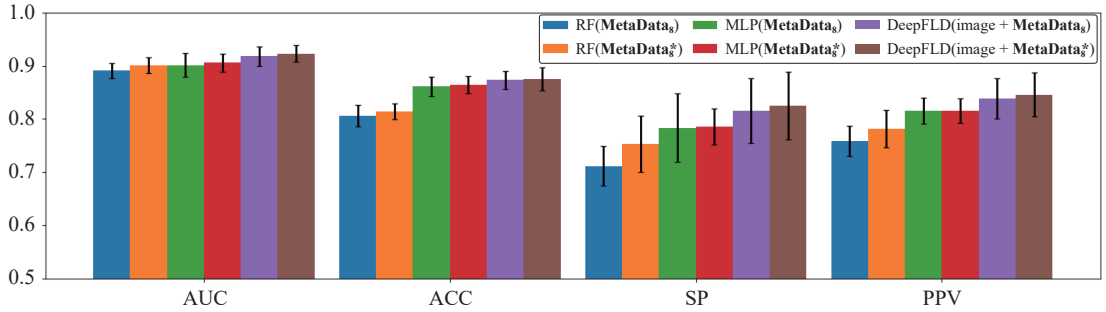


Fig. 7 Performance of models with different indicators, in which RF ($\mathbf{MetaData}_8$), MLP ($\mathbf{MetaData}_8$), RF ($\mathbf{MetaData}_8^*$), and MLP ($\mathbf{MetaData}_8^*$) denote random forest (RF) and MLP models, respectively, with the input of the 8 selected indicators $\mathbf{MetaData}_8$, by Pearson correlation coefficient, and the 8 selected indicators $\mathbf{MetaData}_8^*$ by the joint indicators data analysis method. DeepFLD (image + $\mathbf{MetaData}_8$) and DeepFLD (image + $\mathbf{MetaData}_8^*$) denote the proposed DeepFLD model with the 8 selected indicators $\mathbf{MetaData}_8$ by Pearson correlation coefficient and facial images as joint inputs, and the 8 selected indicators $\mathbf{MetaData}_8^*$ by the joint indicators data analysis method and facial images used as joint input, respectively. (Colored figures are available in the online version at https://link.springer.com/journal/11633)

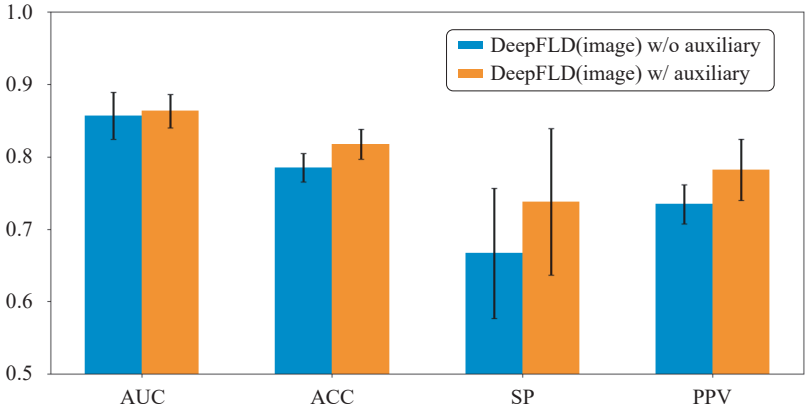


Fig. 8 Performance of DeepFLD without auxiliary task (DeepFLD (image) w/o auxiliary) and DeepFLD with auxiliary task (DeepFLD (image) w/ auxiliary) using only facial images as input. NAFLD screening can be achieved with just a single facial image through DeepFLD, which has a high practical application value. (Colored figures are available in the online version at https://link.springer.com/journal/11633)

application value.

4.5 Visualization and analysis

To explain the underlying mechanism of our study and minimize the black box effect, the visualization technique is adopted to highlight the abnormal areas recognized by the algorithms. Fig. 6 shows the visualization of patients with NAFLD and healthy people.

In this study, we observe that: 1) Compared with healthy people, the facial skin color of patients with NAFLD is darker and yellower; and 2) patients with NAFLD have more melasma. The phenomenon of a difference in skin color may be related to bilirubin[40]. Traditionally, serum bilirubin has been used as a diagnostic marker for hepatobiliary disease, and total bilirubin elevation can occur in liver diseases[41, 42]. NAFLD is characterized by lipid accumulation in hepatocytes, which usually have hepatic steatosis[4, 43]. Antioxidative bilirubin and bilirubin-secreting biliverdin reductase inhibit the inflammatory response of hepatocytes, intervene in the process of hepatic steatosis, and reduce liver damage[44]. Elevated levels of bilirubin binding to the epidermis do not significantly increase skin yellowing. Moreover, bilirubin produced by cells may directly contribute to the dull appearance of facial skin[45]. Previous studies have found that metabolic abnormalities in the liver cause hyperpigmentation and melasma[46, 47]. Melasma is a functional chronic disease that manifests in the face[48]. Traditional Chinese medicine believes that the abnormal metabolism of the liver causes the metabolic waste to be effectively discharged and deposited under the skin, causing hyperpigmentation. Moreover, Fig. 8 shows that the model focuses on the facial modiolus near the mouth, which may be related to NAFLD, and needs further investigation.

In order to further find the differences between people with NAFLD and healthy people and explain why the model can distinguish them, we find the same people from 2020 and 2021 data, who had a change in their liver health, namely, they had NAFLD in one year and did not have NAFLD in the other year. From Fig. 9, we can see that the skin color of the people with NAFLD is darker and yellower, especially in the first row in Fig. 9. The second row shows the people with NAFLD have more melasma, and the last row shows the people with NAFLD are fatter. The visualization results are consistent with the findings of our analysis above, further validating the conclusions of the above analysis.

5 Conclusions

This paper presents an intelligent NAFLD diagnosis system, DeepFLDDiag, with a comprehensive clinical dataset, FLDData, and a novel NAFLD classification algorithm to investigate whether the facial image contributes to the prediction of NAFLD. FLDData includes faces and 480 physical examination and lifestyle indicators for participants. Through joint indicator-based data analysis, we determine the eight most useful indicators for NAFLD classification. With multimodal input and medical constraints, the proposed DeepFLD method can facilitate the extraction of image features from high-dimensional facial images. With multimodal input, the proposed DeepFLD achieves better performance than with metadata and acceptable performance on unseen data. Inspiringly, DeepFLD can achieve competitive results using only facial images as input rather than metadata, paving the way for a more robust and less invasive approach to NAFLD diagnosis.

Despite the promising results of our DeepFLD method in diagnosing NAFLD using facial images and metadata, there are inherent limitations to consider. Currently, the model is predominantly based on 2D facial images, which may not fully capture the complexity of facial structures and underlying health conditions. The integration of three-dimensional facial data could offer significant advantages by providing spatial linkages and po-

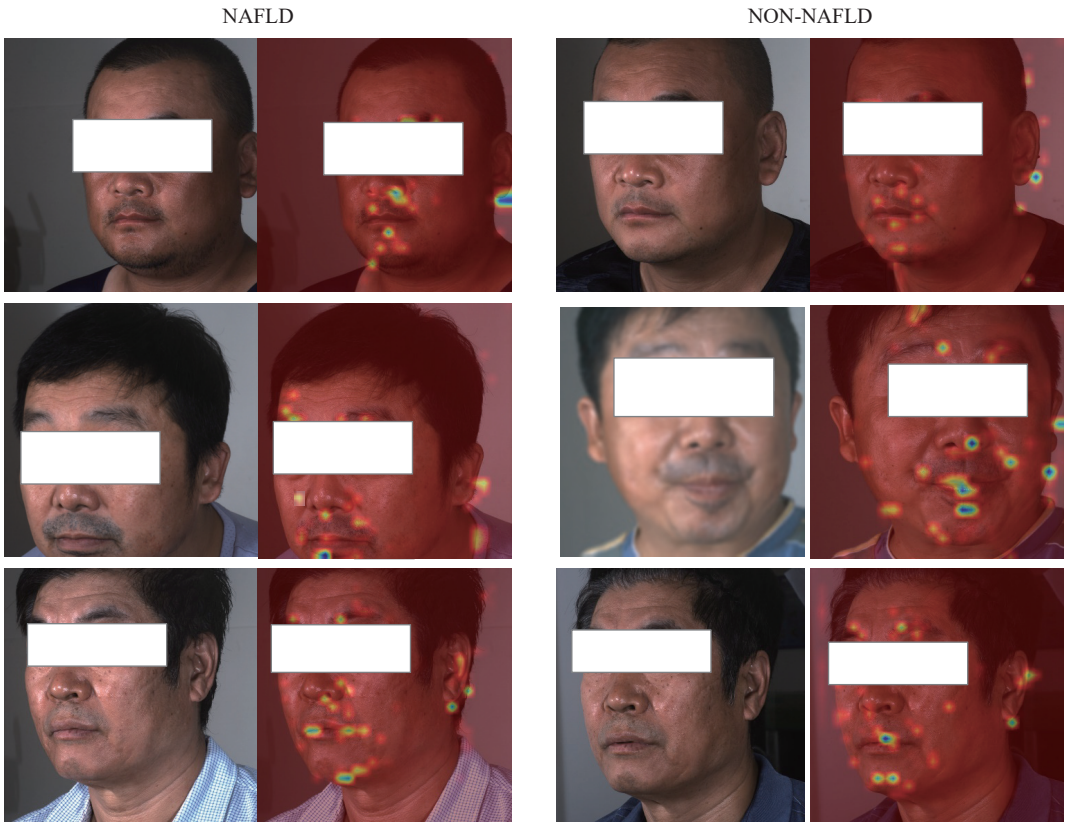


Fig. 9 Visualization of patients with NAFLD and healthy people (NON-NAFLD), who have a change in their liver health. Each row is the same person. (Colored figures are available in the online version at https://link.springer.com/journal/11633)

tentially enhancing the identification of NAFLD. Future research directions may explore the application of three-dimensional facial data for the direct prediction of NAFLD, aiming to further refine and advance the diagnostic capabilities of our system.

Declarations of conflict of interest

The authors declared that they have no conflicts of interest to this work.

References

[1] S. K. Asrani, H. Devarbhavi, J. Eaton, P. S. Kamath. Burden of liver diseases in the world. *Journal of Hepatology*, vol. 70, no. 1, pp. 151–171, 2019. DOI: 10.1016/j.jhep.2018.09.014.

[2] J. J. Yu, C. Y. Zhu, X. B. Wang, K. Kim, A. Bartolome, P. Dongiovanni, K. P. Yates, L. Valenti, M. Carrer, T. Sadowski, L. Qiang, I. Tabas, J. E. Lavine, U. B. Pajvani. Hepatocyte TLR4 triggers inter-hepatocyte Jagged1/Notch signaling to determine NASH-induced fibrosis. *Science Translational Medicine*, vol. 13, no. 599, Article number eabe1692, 2021. DOI: 10.1126/SCITRANSLMED.ABE1692.

[3] G. Weinstein, S. Zelber-Sagi, S. R. Preis, A. S. Beiser, C. DeCarli, E. K. Speliotes, C. L. Satizabal, R. S. Vasan, S. Seshadri. Association of nonalcoholic fatty liver disease with lower brain volume in healthy middle-aged adults in the Framingham study. *JAMA Neurology*, vol. 75, no. 1, pp. 97–104, 2018. DOI: 10.1001/jamaneurol.2017.3229.

[4] C. D. Byrne, G. Targher. NAFLD: A multisystem disease. *Journal of Hepatology*, vol. 62, no. S1, pp. S47–S64, 2015. DOI: 10.1016/j.jhep.2014.12.012.

[5] S. L. Friedman, B. A. Neuschwander-Tetri, M. Rinella, A. J. Sanyal. Mechanisms of NAFLD development and therapeutic strategies. *Nature Medicine*, vol. 24, no. 7, pp. 908–922, 2018. DOI: 10.1038/s41591-018-0104-9.

[6] S. M. Armour, J. R. Remsberg, M. Damle, S. Sidoli, W. Y. Ho, Z. H. Li, B. A. Garcia, M. A. Lazar. An HDAC3-PROX1 corepressor module acts on HNF4α to control hepatic triglycerides. *Nature Communications*, vol. 8, no. 1, Article number 549, 2017. DOI: 10.1038/s41467-017-00772-5.

[7] N. Chalasani, Z. Younossi, J. E. Lavine, M. Charlton, K. Cusi, M. Rinella, S. A. Harrison, E. M. Brunt, A. J. Sanyal. The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American association for the study of liver diseases. *Hepatology*, vol. 67, no. 1, pp. 328–357, 2018. DOI: 10.1002/hep.29367.

[8] L. Castera, M. Friedrich-Rust, R. Loomba. Noninvasive assessment of liver disease in patients with nonalcoholic fatty liver disease. *Gastroenterology*, vol. 156, no. 5, pp. 1264–1281, 2019. DOI: 10.1053/j.gastro.2018.12.036.

[9] H. Leung, X. X. Long, Y. Q. Ni, L. L. Qian, E. Nychas, S. L. Siliceo, D. Pohl, K. Hanhineva, Y. Liu, A. M. Xu, H. B. Nielsen, E. Belda, K. Clément, R. Loomba, H. T. Li, W. P. Jia, G. Panagiotou. Risk assessment with gut microbiome and metabolite markers in NAFLD development. *Science Translational Medicine*, vol. 14, no. 648, Article number

eabk0855, 2022. DOI: 10.1126/scitranslmed.abk0855.

[10] E. G. F. Fojas, A. J. Buckley, N. Lessan. Associations between neck circumference and markers of dysglycemia, non-alcoholic fatty liver disease, and dysmetabolism independent of body mass index in an Emirati population. *Frontiers in Endocrinology*, vol. 13, Article number 929724, 2022. DOI: 10.3389/fendo.2022.929724.

[11] I. Graupera, M. Thiele, M. Serra-Burriel, L. Caballeria, D. Roulot, G. L. H. Wong, N. Fabrellas, I. N. Guha, A. Arslanow, C. Expósito, R. Hernández, G. P. Aithal, P. R. Galle, G. Pera, V. W. S. Wong, F. Lammert, P. Ginès, L. Castera, A. Krag, Investigators of the LiverScreen Consortium. Low accuracy of FIB-4 and NAFLD fibrosis scores for screening for liver fibrosis in the population. *Clinical Gastroenterology and Hepatology*, vol. 20, no. 11, pp. 2567–2576, 2022. DOI: 10.1016/j.cgh.2021.12.034.

[12] G. Bai, T. J. Zhang, Y. Y. Hou, G. Y. Ding, M. Jiang, G. A. Luo. From quality markers to data mining and intelligence assessment: A smart quality-evaluation strategy for traditional Chinese medicine based on quality markers. *Phytomedicine*, vol. 44, pp. 109–116, 2018. DOI: 10.1016/j.phymed.2018.01.017.

[13] J. Cox-Brinkman, A. Vedder, C. Hollak, L. Richfield, A. Mehta, K. Orteu, F. Wijburg, P. Hammond. Three-dimensional face shape in Fabry disease. *European Journal of Human Genetics*, vol. 15, no. 5, pp. 535–542, 2007. DOI: 10.1038/sj.ejhg.5201798.

[14] S. Wang, Y. Cong, H. C. Zhu, X. Y. Chen, L. Q. Qu, H. J. Fan, Q. Zhang, M. X. Liu. Multi-scale context-guided deep network for automated lesion segmentation with endoscopy images of gastrointestinal tract. *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 2, pp. 514–525, 2021. DOI: 10.1109/JBHI.2020.2997760.

[15] W. Y. Chen, X. Xia, Y. Huang, X. W. Chen, J. D. J. Han. Bioimaging for quantitative phenotype analysis. *Methods*, vol. 102, pp. 20–25, 2016. DOI: 10.1016/j.ymeth.2016.01.017.

[16] X. Xia, X. W. Chen, G. Wu, F. Li, Y. Y. Wang, Y. Chen, M. X. Chen, X. Y. Wang, W. Y. Chen, B. Xian, W. Z. Chen, Y. Q. Cao, C. Xu, W. X. Gong, G. Y. Chen, D. H. Cai, W. X. Wei, Y. Z. Yan, K. P. Liu, N. Qiao, X. H. Zhao, J. Jia, W. Wang, B. K. Kennedy, K. Zhang, C. V. Cannistraci, Y. Zhou, J. D. J. Han. Three-dimensional facial-image analysis to predict heterogeneity of the human ageing rate and the impact of lifestyle. *Nature Metabolism*, vol. 2, no. 9, pp. 946–957, 2020. DOI: 10.1038/s42255-020-00270-x.

[17] Y. B. Rong, D. H. Xiang, W. F. Zhu, K. Yu, F. Shi, Z. Fan, X. J. Chen. Surrogate-assisted retinal OCT image classification based on convolutional neural networks. *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 1, pp. 253–263, 2019. DOI: 10.1109/JBHI.2018.2795545.

[18] D. B. Zhao, Y. R. Chen, L. Lv. Deep reinforcement learning with visual attention for vehicle classification. *IEEE Transactions on Cognitive and Developmental Systems*, vol. 9, no. 4, pp. 356–367, 2017. DOI: 10.1109/TCDS.2016.2614675.

[19] Y. R. Chen, D. B. Zhao, L. Lv, Q. C. Zhang. Multi-task learning for dangerous object detection in autonomous driving. *Information Sciences*, vol. 432, pp. 559–571, 2018. DOI: 10.1016/j.ins.2017.08.035.

[20] Y. R. Chen, R. Y. Gao, F. G. Liu, D. B. Zhao. ModuleNet: Knowledge-inherited neural architecture search. *IEEE Transactions on Cybernetics*, vol. 52, no. 11, pp. 11661–11671, 2022. DOI: 10.1109/TCYB.2021.3078573.

[21] S. Bhattacharyya, S. Majumder, P. Debnath, M. Chanda. Arrhythmic heartbeat classification using ensemble of random forest and support vector machine algorithm. *IEEE Transactions on Artificial Intelligence*, vol. 2, no. 3, pp. 260–268, 2021. DOI: 10.1109/TAI.2021.3083689.

[22] M. Ghafourian, J. Fierrez, R. Vera-Rodriguez, A. Morales, I. Serna. OTB-morph: One-time biometrics via morphing. *Machine Intelligence Research*, vol. 20, no. 6, pp. 855–871, 2023. DOI: 10.1007/s11633-023-1432-x.

[23] W. C. Wang, E. Ahn, D. G. Feng, J. Kim. A review of predictive and contrastive self-supervised learning for medical images. *Machine Intelligence Research*, vol. 20, no. 4, pp. 483–513, 2023. DOI: 10.1007/s11633-022-1406-4.

[24] V. Nobili, A. Alisi, L. Valenti, L. Miele, A. E. Feldstein, N. Alkhouri. Nafld in children: New genes, new diagnostic modalities and new drugs. *Nature Reviews Gastroenterology & Hepatology*, vol. 16, no. 9, pp. 517–530, 2019. DOI: 10.1038/s41575-019-0169-z.

[25] H. K. Min, A. Kapoor, M. Fuchs, F. Mirshahi, H. P. Zhou, J. Maher, J. Kellum, R. Warnick, M. J. Contos, A. J. Sanyal. Increased hepatic synthesis and dysregulation of cholesterol metabolism is associated with the severity of nonalcoholic fatty liver disease. *Cell Metabolism*, vol. 15, no. 5, pp. 665–674, 2012. DOI: 10.1016/j.cmet.2012.04.004.

[26] S. Stender, J. Kozlitina, B. G. Nordestgaard, A. Tybjærg-Hansen, H. H. Hobbs, J. C. Cohen. Adiposity amplifies the genetic risk of fatty liver disease conferred by multiple loci. *Nature Genetics*, vol. 49, no. 6, pp. 842–847, 2017. DOI: 10.1038/ng.3855.

[27] I. Afyouni, Z. A. Aghbari, R. A. Razack. Multi-feature, multi-modal, and multi-source social event detection: A comprehensive survey. *Information Fusion*, vol. 79, pp. 279–308, 2022. DOI: 10.1016/j.inffus.2021.10.013.

[28] A. Kedia, M. A. Zaidi, H. Lee. FiE: Building a global probability space by leveraging early fusion in encoder for open-domain question answering. In *Proceedings of Conference on Empirical Methods in Natural Language Processing*, Abu Dhabi, UAE, pp. 4246–4260, 2022. DOI: 10.18653/v1/2022.emnlp-main.285.

[29] M. Haris, A. Glowacz. Navigating an automated driving vehicle via the early fusion of multi-modality. *Sensors*, vol. 22, no. 4, Article number 1425, 2022. DOI: 10.3390/s22041425.

[30] D. Salvati, C. Drioli, G. L. Foresti. A late fusion deep neural network for robust speaker identification using raw waveforms and gammatone cepstral coefficients. *Expert Systems with Applications*, vol. 222, Article number 119750, 2023. DOI: 10.1016/j.eswa.2023.119750.

[31] T. J. Zhang, X. W. Liu, L. Gong, S. W. Wang, X. Niu, L. Shen. Late fusion multiple kernel clustering with local kernel alignment maximization. *IEEE Transactions on Multimedia*, vol. 25, pp. 993–1007, 2023. DOI: 10.1109/TMM.2021.3136094.

[32] A. Batool, Y. P. Dai, H. B. Ma, S. J. Yin. Deep feature fusion classification model for identifying machine parts. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 27, no. 5, pp. 876–885, 2023. DOI: 10.20965/jaciii.2023.p0876.

[33] L. Yuan, L. Yang, S. C. Zhang, Z. Y. Xu, J. J. Qin, Y. F. Shi, P. C. Yu, Y. Wang, Z. H. Bao, Y. H. Xia, J. C. Sun, W. Y. He, T. H. Chen, X. L. Chen, C. Hu, Y. L. Zhang, C. W. Dong, P. Zhao, Y. N. Wang, N. Jiang, B. Lv, Y. W. Xue, B. P. Jiao, H. Y. Gao, K. Q. Chai, J. Li, H. Wang, X. B. Wang, X. Q. Guan, X. Liu, G. Zhao, Z. C. Zheng, J. Yan, H. Y. Yu, L. C. Chen, Z. S. Ye, H. Q. You, Y. Bao, X. Cheng, P. Z. Zhao, L. Wang, W. T. Zeng, Y. F. Tian, M. Chen, Y. You, G. H. Yuan, H. Ruan, X. L. Gao, J. L. Xu, H. D. Xu, L. B. Du, S. J. Zhang, H. Y. Fu, X. D. Cheng. Development of a tongue image-based machine learning tool for the diagnosis of gastric cancer: A prospective multicentre clinical cohort study. *eClinicalMedicine*, vol. 57, Article number 101834, 2023. DOI: 10.1016/j.eclinm.2023.101834.

[34] Y. S. Li, W. Chen, X. Huang, Z. Gao, S. W. Li, T. He, Y. J. Zhang. MFVNet: A deep adaptive fusion network with multiple field-of-views for remote sensing image semantic segmentation. *Science China Information Sciences*, vol. 66, no. 4, Article number 140305, 2023. DOI: 10.1007/s11432-022-3599-y.

[35] X. Y. Chen, F. X. Shi, J. Xiao, F. Y. Huang, F. Cheng, L. H. Wang, Y. L. Ju, Y. Zhou, H. Y. Jia. Associations between abdominal obesity indices and nonalcoholic fatty liver disease: Chinese visceral adiposity index. *Frontiers in Endocrinology*, vol. 13, Article number 831960, 2022. DOI: 10.3389/fendo.2022.831960.

[36] S. Chitturi, G. C. Farrell, E. Hashimoto, T. Saibara, G. K. K. Lau, J. D. Sollano, The Asia–Pacific Working Party on NAFLD. Non-alcoholic fatty liver disease in the Asia–Pacific region: Definitions and overview of proposed guidelines. *Journal of Gastroenterology and Hepatology*, vol. 22, no. 6, pp. 778–787, 2007. DOI: 10.1111/j.1440-1746.2007.05001.x.

[37] S. M. Lundberg, S. I. Lee. A unified approach to interpreting model predictions. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, USA, pp. 4768–4777, 2017.

[38] R. García-Carretero, R. Holgado-Cuadrado, Ó. Barquero-Pérez. Assessment of classification models and relevant features on nonalcoholic steatohepatitis using random forest. *Entropy*, vol. 23, no. 6, Article number 763, 2021. DOI: 10.3390/E23060763.

[39] C. C. Wu, W. C. Yeh, W. D. Hsu, M. M. Islam, P. A. Nguyen, T. N. Poly, Y. C. Wang, H. C. Yang, Y. C. Li. Prediction of fatty liver disease using machine learning algorithms. *Computer Methods and Programs in Biomedicine*, vol. 170, pp. 23–29, 2019. DOI: 10.1016/j.cmpb.2018.12.032.

[40] H. Yamaza, S. Sonoda, K. Nonaka, T. Kukita, T. Yamaza. Pamidronate decreases bilirubin-impaired cell death and improves dentinogenic dysfunction of stem cells from human deciduous teeth. *Stem Cell Research & Therapy*, vol. 9, no. 1, Article number 303, 2018. DOI: 10.1186/s13287-018-1042-7.

[41] P. Y. Kwo, S. M. Cohen, J. K. Lim. ACG clinical guideline: Evaluation of abnormal liver chemistries. *American Journal of Gastroenterology*, vol. 112, no. 1, pp. 18–35, 2017. DOI: 10.1038/ajg.2016.517.

[42] K. H. Wagner, R. G. Shiels, C. A. Lang, N. Seyed Khoei, A. C. Bulmer. Diagnostic criteria and contributors to Gilbert's syndrome. *Critical Reviews in Clinical Laboratory Sciences*, vol. 55, no. 2, pp. 129–139, 2018. DOI: 10.1080/10408363.2018.1428526.

[43] S. Tanaka, H. Hikita, T. Tatsumi, R. Sakamori, Y. Nozaki, S. Sakane, Y. Shiode, T. Nakabori, Y. Saito, N. Hiramatsu, K. Tabata, T. Kawabata, M. Hamasaki, H. Eguchi, H. Nagano, T. Yoshimori, T. Takehara. Rubicon inhibits autophagy and accelerates hepatocyte apoptosis and lipid accumulation in nonalcoholic fatty liver disease in mice. *Hepatology*, vol. 64, no. 6, pp. 1994–2014, 2016. DOI: 10.1002/hep.28820.

[44] L. Weaver, A. R. Hamoud, D. E. Stec, T. D. Jr Hinds. Biliverdin reductase and bilirubin in hepatic disease. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, vol. 314, no. 6, pp. G668–G676, 2018. DOI: 10.1152/ajpgi.00026.2018.

[45] B. Fang, P. D. Card, J. J. Chen, L. J. Li, T. Laughlin, B. Jarrold, W. Z. Zhao, A. M. Benham, A. T. Määttä, T. J. Hawkins, T. Hakozaki. A potential role of keratinocyte-derived bilirubin in human skin yellowness and its amelioration by sucrose laurate/dilaurate. *International Journal of Molecular Sciences*, vol. 23, no. 11, Article number 5884, 2022. DOI: 10.3390/IJMS23115884.

[46] S. H. Kwon, Y. J. Hwang, S. K. Lee, K. C. Park. Heterogeneous pathology of melasma and its clinical implications. *International Journal of Molecular Sciences*, vol. 17, no. 6, Article number 824, 2016. DOI: 10.3390/ijms17060824.

[47] G. Su, X. Zhou. Clinical efficacy of self-designed Xiaoban Huoxue prescription on chloasma derived from liver stagnation and blood stasis. *American Journal of Translational Research*, vol. 13, no. 12, pp. 14031–14038, 2021.

[48] E. Bronzina, A. Clement, B. Marie, K. T. Fook Chong, P. Faure, T. Passeron. Efficacy and tolerability on melasma of a topical cosmetic product acting on melanocytes, fibroblasts and endothelial cells: A randomized comparative trial against 4% hydroquinone. *Journal of the European Academy of Dermatology and Venereology*, vol. 34, no. 4, pp. 897–903, 2020. DOI: 10.1111/jdv.16150.

Yaran Chen (Member, IEEE) received the Ph. D. degree in control theory and control engineering from the Institute of Automation, Chinese Academy of Sciences, China in 2018. She is currently an associate professor with the State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, and the College of Artificial Intelligence, University of Chinese Academy of Sciences, China.

Her research interests include deep learning, deep reinforcement learning, and embodied AI.

E-mail: chenyaran2013@ia.ac.cn

ORCID iD: 0000-0001-9356-0610

Xueyu Chen received the M.D. degree in epidemiology and health statistics from Shandong University, China in 2025. He is dedicated to analysing liver disease classification tasks using imaging omics data, such as facial and retinal images.

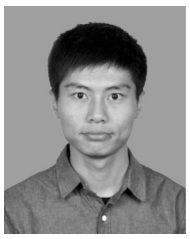
His research interest is studying various obesity assessment indices for fatty liver disease.

E-mail: chenxueyv@163.com
ORCID iD: 0000-0001-6104-4240

Yu Han received the B. Eng. degree in electronic & information from Xidian University, China in 2021. He is currently a master student in electronic and information engineering, College of Artificial Intelligence, Chinese Academy of Sciences, China.

His research interests include computer vision, image processing, and robot perception.

E-mail: hanyu2021@ia.ac.cn
ORCID iD: 0000-0002-5415-7535



Haoran Li received the B. Sc. degrees in detection, guidance and control technology from Central South University, China in 2015, and the Ph. D. degree in control theory and control engineering at the Institute of Automation, Chinese Academy of Sciences, China in 2020. He is currently an associate professor of the Institute of Automation, Chinese Academy of Sciences, China, and also with the College of Artificial Intelligence, University of Chinese Academy of Sciences, China.

His research interests include deep learning, deep reinforcement learning and robotic learning.

E-mail: lihaoran2015@ia.ac.cn

Dongbin Zhao received the B. Sc. degree in robotics, the M. Sc. degree in intelligent control theory and applications, and the Ph. D. degree in computational intelligence from Harbin Institute of Technology, China in 1994, 1996 and 2000 respectively. He is now a professor with the Institute of Automation, Chinese Academy of Sciences, China, and with the School of Artificial Intelligence, University of Chinese Academy of Sciences, China. He is an IEEE Fellow. He has published 7 books, and over 100 international journal papers. He is a recipient of the 2022 Best Paper Reward of *IEEE Transactions on Automation Science and Engineering*, the 2022 Outstanding Paper Reward of *IEEE Transactions on Emerging Topics in Computational Intelligence*, and the 2020 Outstanding Paper Reward of *IEEE Transactions on Cognitive and Developmental Systems*.

His research interests include deep reinforcement learning, autonomous driving, game artificial intelligence, and robotics.

E-mail: dongbin.zhao@ia.ac.cn
ORCID iD: 0000-0001-8218-9633

Jingzhong Li received the B. Eng. degree in computer science and technology from Beijing University of Chemical Technology, China in 2005. He is a senior experimentalist at the Information and Educational Technology Center, Beijing University of Chinese Medicine, China.

His research interest is AI in disease recognition and drug discovery.

E-mail: lijz@bucm.edu.cn

Xu Wang received the B. Eng. degree in artificial intelligence science and technology from Beijing University of Posts and Telecommunications, China in 2010, and the Ph. D. degree in cognitive neuroscience at State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, China in 2016. Currently, she is an associate professor at the School of Life Sciences, Beijing University of Chinese Medicine, China.

Her research interest is AI in disease recognition and efficacy predication based on multimodal medical images (e.g., face, tongue, and brain images).

E-mail: wangx@bucm.edu.cn(Corresponding author)
ORCID iD: 0000-0003-0593-8615

Yong Zhou received the M.D. degree in internal medicine from Capital Medical University, China in 2007. He specializes in interdisciplinary research that bridges epidemiology with artificial intelligence, with a focus on chronic liver and cerebrovascular diseases. Using machine learning and deep learning algorithms to promote research on human health and diseases, numerous early warning models for various diseases have been developed. The expertise in constructing predictive models with prospective cohort data demonstrates the generalizability and reliability of these models.

His research interests include artificial intelligence for the screening and prediction of various chronic diseases.

E-mail:yongzhou78214@163.com(Corresponding author)
ORCID iD: 0000-0001-5221-8026