Hard Sample Mining-based Tongue Diagnosis for Fatty Liver Disease Severity Classification

Tao Chen[1], Jie Gao[2], Yong Xu[3], Weihong Qiu[2], Yijie Wu[1(✉)], Weimin Ye[2,4(✉)], and Kunhong Liu[5,6(✉)]

[1] State Key Laboratory of Fluid Power and Mechatronic Systems, School of Mechanical Engineering, Zhejiang University, Hangzhou, 310027, China
wyj1116@zju.edu.cn
[2] Department of Epidemiology and Health Statistics, School of Public Health, Fujian Medical University, Fuzhou, 350122, China
ywm@fjmu.edu.cn
[3] Xiamen Key Laboratory of Intelligent Fishery, Xiamen Ocean Vocational College, Xiamen, 361100, China
[4] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, 17177, Sweden
[5] Department of Digital Media, School of Film, Xiamen University, Xiamen, 361005, China
[6] Xiamen Key Laboratory of Intelligent Storage and Computing, School of Informatics, Xiamen University, Xiamen, 361005, China
lkhqz@xmu.edu.cn

Abstract. Fatty liver disease (FLD) negatively affects over 30% of the global population and can ultimately lead to cirrhosis and death. Early detection and intervention on the severity of FLD help control its progression. However, facilities for assessing the severity of FLD are lacking in economically disadvantaged regions, highlighting an urgent need for a cost-effective and scalable screening method. Traditional Chinese Medicine (TCM) suggests a strong correlation between tongue characteristics and liver health, positioning tongue diagnosis as a non-invasive means for assessing FLD severity. Establishing an automated tongue diagnosis method holds promise for large-scale and rapid classification of FLD severity among rural populations. In this paper we present a Hard sample Mining-based Tongue Diagnosis Framework (HM-TDF) for multi-class classification of FLD severity. The HM-TDF identifies hard samples using a novel uncertainty estimation approach and addresses them through a multi-expert classifier. We introduce a Multi-source Feature Fusion Kolmogorov-Arnold Network (MFF-KAN) to model the relationship between tongue images plus basic physiological indicators and FLD severity. We propose a three-step training strategy to train this heterogeneous model. We construct and release a novel tongue diagnosis dataset for FLD severity classification, named Tongue-FLD, to advance research in automated tongue diagnosis. Experimental results on this dataset indicate that the proposed method surpasses existing automated tongue diagnosis methods in the classification of FLD severity. Moreover, MFF-KAN effectively visualizes the key pathways from input to output, providing strong

T. Chen and J. Gao—Contributed equally.

interpretability. The dataset and code are available at https://github.com/MLDMXM2017/HM-TDF.

Keywords: Hard mining, Kolmogorov-Arnold network (KAN), tongue diagnosis, net-work interpretability.

1 Introduction

Fatty liver disease (FLD) is one of the most common liver disorders worldwide [1]. FLD can be categorized into alcoholic and non-alcoholic types [2]. Both types have the potential to progress to steatohepatitis and hepatic fibrosis [3]. Early detection and intervention on the severity of FLD help control its progression. Therefore, there is an urgent need to develop an inexpensive, convenient and scalable approach for assessing the severity of FLD, especially in medically underserved populations.

Tongue diagnosis is an essential method in Traditional Chinese Medicine (TCM), which itself plays a significant role in Chinese healthcare [4]. The superficial capillaries in the tongue provide non-invasive access to internal body information in a straightforward manner [5]. Both TCM classical texts and contemporary medical research have highlighted a significant correlation between liver health status and tongue characteristics [6, 7]. Based on observed characteristics, TCM practitioners derive diagnostic and treatment decisions using their expertise and knowledge. However, for assessing the severity of FLD, TCM lacks a standardized tongue diagnosis system to guide inexperienced practitioners, which hinders its widespread application. Therefore, studies have attempted to model the relationship between tongue images and specific diseases to facilitate automated tongue diagnosis [5]. These studies have demonstrated the feasibility of automated tongue diagnosis for diseases such as FLD [8-11] and liver cancer [6]. However, the majority of existing studies focus on binary classification to diagnose the presence or absence of diseases. This may be attributed to the complex and diverse nature of the tongue characteristics, leading to hard samples. The challenges are even severer if only a limited number of training samples are available. Additionally, many neural network-based tongue diagnosis models contain a vast number of learnable weights, enhancing the difficulty of model interpretation.

To address the above challenges, we advance both methodological and data aspects. Our contributions can be summarized as: (1) We propose a Hard sample Mining-based Tongue Diagnosis Framework (HM-TDF) for multi-class classification of FLD severity, which identifies hard samples based on uncertainty and applies targeted classifier; (2) We design a Multi-source Feature Fusion Kolmogorov-Arnold Network (MFF-KAN) for modeling the relationships between complex inputs and FLD severity; (3) We construct and release a novel dataset (Tongue-FLD), which is the largest publicly available tongue diagnosis dataset for FLD severity classification. (4) We interpret the network behavior based on visualization of key pathways within the network.

2 Related work

Kolmogorov-Arnold Network (KAN) is a network derived from the Kolmogorov-Arnold Representation Theorem (KART), which states that any continuous multivariate function can be represented as a finite sum and composition of univariate functions:

$$f(\boldsymbol{x}) = f(x_1, \ldots, x_n) = \sum_{q=0}^{2n} \psi_q \left(\sum_{p=1}^{n} \phi_{p,q}(x_p) \right), \tag{1}$$

where $f: [0,1]^n \to \mathbb{R}$, $\phi_{p,q}: [0,1] \to \mathbb{R}$, and $\psi_q: \mathbb{R} \to \mathbb{R}$. Based on this theorem, Liu et al. [12] proposed KAN by replacing each linear weight matrix in an MLP with a learnable activation function matrix $\boldsymbol{\Phi}_l = \{\phi_{l,p,q}\}$. Here, $\phi_{l,p,q}$ represents the activation function corresponding to the p-th input and q-th output in l-th layer. Due to its structural similarity to MLP, KAN is also referred to as KAN Linear.

The architecture design based on KART endows KAN with inherently strong representation capabilities. In recent studies, the concept of learnable activation functions has been utilized to enhance convolutional and transformer architectures, leading to the development of KAN Conv [13] and Kolmogorov-Arnold Transformer [14]. The new models achieve comparable or even superior performance with significantly fewer parameters, highlighting the substantial potential of KAN-based architectures.

3 Method

3.1 Overview

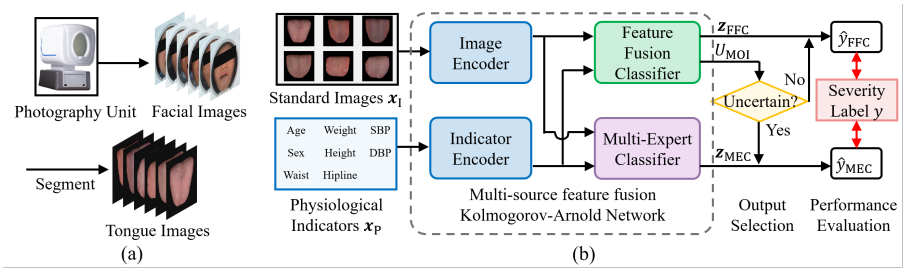

Fig. 1. Overview of the method. (a) The process of acquisition and segmentation of tongue images. (b) The Hard sample Mining-based Tongue Diagnosis Framework (HM-TDF).

To enable automated tongue diagnosis for assessing FLD severity, we first collect facial images with protruded tongues under standard lighting conditions. These images were then segmented to obtain black-background images containing only the tongue, as shown in Fig. 1(a). The main challenge is to fully extract the diverse features from the surface of the tongue. At the same time, there are hard samples that are of little features to distinguish between different types. To this end, we propose a Hard sample Mining-based Tongue Diagnosis Framework, termed HM-TDF. As shown in Fig. 1(b), the framework takes segmented standard tongue images $\boldsymbol{x}_I$ and basic physiological

indicators $\boldsymbol{x}_{\mathrm{P}}$ as inputs. The inputs are processed by the Multi-source Feature Fusion Kolmogorov-Arnold Network (MFF-KAN), whose KAN architecture is well-suited for modeling complex feature associations. For each sample, the MFF-KAN produces two prediction outputs, generated respectively by the feature fusion classifier (FFC) and the multi-expert classifier (MEC). When the uncertainty of the FFC is low, the current input is considered a general sample, and $\hat{\mathrm{y}}_{\mathrm{FFC}}$ is selected as final prediction. Otherwise, $\hat{\mathrm{y}}_{\mathrm{MEC}}$ is used as the final prediction for a hard sample. The detailed network architecture and training process are introduced in the next subsection.

3.2 Network architecture

The detailed architecture of MFF-KAN is illustrated in Fig. 2(a). To effectively extract feature information from tongue images, we design the image encoder (IE) to incorporate both traditional convolutions (CNN) and KAN Conv. The role of the CNN is to generate feature maps with significantly compressed spatial dimensions, while retaining as much of the original image information as possible. Then, KAN Conv leverage their powerful capacity for capturing complex relationships to extract feature vector $\boldsymbol{f}_{\mathrm{I}}$ from the feature maps. The indicator encoder (DE) is a KAN and performs extensive functional mapping of the fundamental physiological indicators to obtain the indicator feature vector $\boldsymbol{f}_{\mathrm{D}}$. The concatenated fusion feature vector is subsequently used by the classifiers. The FFC and MEC are both KANs with a progressively shrinking architecture. The output of FFC has a dimensionality equal to the number of classes n_{L}. In contrast, MEC essentially consists of multiple binary classifiers, each acting as an expert specialized in recognizing specific class group. Therefore, MEC arranges its output based on the number of experts n_{E}.

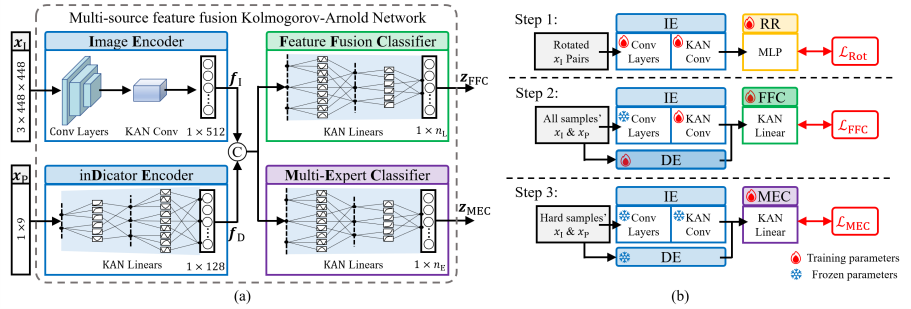

Fig. 2. Illustration of MFF-KAN. (a) Architecture of MFF-KAN. (b) Three-step training strategy.

A key aspect of the MEC its training with ordinal multi-expert labels (OMELs), which effectively diversify the grouping of different severity classes. From the perspective of an ordinal classification, class groups should not overlap. The grouping method that satisfies the aforementioned requirements is limited to $n_{\mathrm{L}} - 1$ possible combinations. To increase the diversity of groupings, we mask a variable set of classes to derive additional grouping configurations. Ultimately, the grouping information for all classes is consolidated into the ordinal multi-expert matrix $\boldsymbol{M}_O \in \mathbb{Z}^{n_{\mathrm{E}} \times n_{\mathrm{L}}}$, where each column

represents a FLD severity level, and each row represents the grouping target. Specifically, each row in $\boldsymbol{M}_O$ contains at least one 1 and one -1, which are treated as positive and negative classification targets, respectively. The value 0 indicates that the class is ignored by the expert. After training with OMELs, the experts within MEC only need to focus on their respective binary classification tasks, thereby improving training efficiency. Meanwhile, the integration of multi-expert decisions contributes to enhancing the robustness of our model. During testing, MEC selects the class corresponding to the column in $\boldsymbol{M}_O$ with the minimum distance to the output vector $\boldsymbol{z}_{\mathrm{MEC}}$.

3.3 Training strategy

MFF-KAN comprises multiple networks with varying architectures, making simultaneous training unlikely to achieve good convergence. Therefore, we designed a three-step training strategy, as shown in Fig. 2(b).

Step 1 aims to pre-train the CNN within the IE. We first connect the IE to a rotation regressor (RR), forming a new network for rotation detection training. After completing the rotation detection training, we deploy the CNN component to develop a robust dimensionality reduction capability, and froze its parameters in subsequent training steps.

Step 2 aims to train the IE, DE and FFC for FLD severity classification. We treat the raw output of FFC as the prior probability predictions, and propose a moment of inertia (MOI) loss for this ordinal classification, formulated as follows:

$$\mathcal{L}_{\mathrm{MOI}} = \sum_{k=0}^{K} \left(c_y - k\right)^2 p_k, \tag{2}$$

Here c_y represents the rotational center of the label, which can be defined as the class index. p_k denotes the probability of the k-th class. The MOI loss is designed to concentrate the output probability towards the labeled target. By simultaneously optimizing both MIO loss and cross entropy loss, the network can flexibly adapt to classification problems with varying degrees of ordinal relationships. After this training step, the encoders are able to effectively extract disease-related features but are still inadequate in handling hard samples. To address this, we select hard samples according to the MIO uncertainty, where the c_y in $\mathcal{L}_{\mathrm{MOI}}$ is replaced by rotational center of the probability. Samples with uncertainty exceeding a given threshold (50% of MOI uncertainty under the assumption that p_k follows a uniform distribution) are selected as hard samples.

At step 3, MEC learns the mapping from features in hard samples to FLD severity using OMELs. During this train step, the parameters of the encoders are kept fixed and a progressively-balanced sampling method is employed to mitigate class imbalance.

4 Experiments

4.1 Data description

The data used in this study were obtained from a cohort study approved by the Ethics Review Committee of Fujian Medical University (Approval Nos. [2017-07] and [2020-

58]). The participants were residents of Fuqing City, Fujian Province, China, aged 35 to 75 years. For each participant, a facial image with the tongue extended was captured and basic physiological indicators were measured. Subsequently, participants underwent ultrasound examinations, and FLD severity was assessed by radiologists [15]. Following experiments are conducted based on this newly released dataset, named Tongue-FLD. The dataset includes 5,717 samples: 3,690 with non-FLD, 1,512 with mild FLD, and 515 with moderate/severe FLD, resulting in an imbalance ratio of 7.17/2.94/1.00.

4.2 Comparison experiments

Comparison setting. All methods were evaluated using the same input modalities, including both tongue images and physiological indicators. Under this setting, we compared our proposed method with 11 competing algorithms, including: a) five tongue diagnosis algorithms: (M1) Wang et al. [16], (M2) Dai et al. [11], (M3) Li et al. [17] , (M4) Qiu et al. [18], and (M5) Yuan et al. [19]; b) three state-of-the-art image classification models: (M6) Vision Transformer (ViT) [20], (M7) ConvNeXt [21], and (M8) Vision Mamba (Vim) [22]. c) three widely-used classifiers: (M9) Support Vector Machine (SVM), (M10) Deep Forest (DF), and (M11) XGBT. We performed five-fold cross-validation and calculated the mean values for each evaluation metric, including Accuracy, and Macro-averaged Precision, Recall, and F1-score [23]. Considering the ordinal nature of FLD severity classification, Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) [23] are also included.

Table 1. Comparison of tongue diagnosis methods in FLD severity classification.

Method	Encoder	Classifier	Accuracy	Precision	Recall	F1-score	MAE ↓	RMSE ↓	Rank ↓
M1 [16]	ResNet34	MLP	$0.7260_{(9)}$	$0.5812_{(11)}$	$0.5632_{(8)}$	$0.5670_{(10)}$	$0.3097_{(10)}$	$0.6167_{(10)}$	9.7
M2 [11]	GAN	XGBT	$0.7305_{(6)}$	$0.6215_{(3)}$	$0.5567_{(10)}$	$0.5738_{(7)}$	$0.2919_{(5)}$	$0.5801_{(3)}$	6.2
M3 [17]	ResNet50	GA_XGBT	$0.7498_{(2)}$	$0.6123_{(6)}$	$0.5583_{(9)}$	$0.5711_{(8)}$	$0.2758_{(2)}$	$0.5716_{(2)}$	5.2
M4 [18]	MobileNet	MLP	$0.7270_{(8)}$	$0.5941_{(7)}$	$0.6006_{(4)}$	$0.5906_{(4)}$	$0.3024_{(8)}$	$0.6006_{(7)}$	4.7
M5 [19]	TransFG	MLP	$0.6490_{(12)}$	$0.5426_{(12)}$	$\mathbf{0.6114}_{(1)}$	$0.5498_{(12)}$	$0.4045_{(12)}$	$0.7152_{(12)}$	10.5
M6[20]	ViT	MLP	$0.7172_{(10)}$	$0.5919_{(8)}$	$0.5876_{(6)}$	$0.5877_{(5)}$	$0.3088_{(9)}$	$0.6008_{(8)}$	7.5
M7 [21]	ConvNeXt	MLP	$0.7463_{(3)}$	$0.6133_{(5)}$	$0.6111_{(2)}$	$0.6047_{(2)}$	$0.2843_{(3)}$	$0.5879_{(5)}$	2.7
M8 [22]	Mamba	MLP	$0.7305_{(6)}$	$0.5903_{(9)}$	$0.5990_{(5)}$	$0.5932_{(3)}$	$0.3018_{(7)}$	$0.6055_{(9)}$	6.8
M9	IE+DE	SVM	$0.7305_{(6)}$	$0.6326_{(2)}$	$0.5535_{(11)}$	$0.5700_{(9)}$	$0.2957_{(6)}$	$0.5901_{(6)}$	7.2
M10	IE+DE	DF	$0.7020_{(11)}$	$0.5887_{(10)}$	$0.5533_{(12)}$	$0.5614_{(11)}$	$0.3330_{(11)}$	$0.6346_{(11)}$	11.0
M11	IE+DE	XGBT	$0.7361_{(4)}$	$0.6195_{(4)}$	$0.5646_{(7)}$	$0.5815_{(6)}$	$0.2899_{(4)}$	$0.5847_{(4)}$	5.3
HM-TDF	IE+DE	FFC+MEC	$\mathbf{0.7524}_{(1)}$	$\mathbf{0.6570}_{(1)}$	$0.6070_{(3)}$	$\mathbf{0.6237}_{(1)}$	$\mathbf{0.2694}_{(1)}$	$\mathbf{0.5596}_{(1)}$	**1.3**

Comparison results. The results are presented using rank order derived from the Friedman test [24], where a lower average rank signifies a superior overall performance. In Table 1, detailed results are presented, with numbers in parentheses indicating the rank order of algorithms for that metric. Our proposed method, with an average rank of 1.3,

is the lowest among all compared methods. In comparison, ConvNeXt and the method proposed by Qiu et al. [18] achieve average ranks of 2.7 and 4.7, respectively, placing second and third in average performance. Directly feeding the extracted features into a single classifier is not an effective approach, as even the best-performing XGBT only achieved an average rank of 5.3. Thirdly, our proposed method achieves the best performance on the MAE and RMSE metrics. This indicates that severe misclassifications between moderate/severe cases and non-FLD cases are reduced. Consequently, our method demonstrates superior performance in classifying the FLD severity.

Table 2. The ablation study results for the innovations in the proposed method

Classifier	Loss	RR-Pretrain	Accuracy	Precision	Recall	F1-score	MAE ↓	RMSE ↓
MLP	Cross-entropy	-	0.7162	0.5789	0.5872	0.5808	0.3181	0.6216
FFC	Cross-entropy	-	0.7209	0.5864	0.5854	0.5855	0.3024	0.5908
FFC	$\mathcal{L}_{\mathrm{FFC}}$	-	0.7262	0.6285	0.5968	0.5892	0.2843	0.5749
FFC	$\mathcal{L}_{\mathrm{FFC}}$	✓	0.7233	0.6298	0.6037	0.6060	0.2716	0.5738
FFC, MEC	$\mathcal{L}_{\mathrm{FFC}}$, $\mathcal{L}_{\mathrm{MEC}}$	✓	**0.7524**	**0.6570**	**0.6070**	**0.6237**	**0.2694**	**0.5596**

4.3 Ablation experiments

Table 2 provides a detailed overview of the ablation study results. A commonly employed MLP classifier with cross-entropy loss is presented as the baseline. Through comparison of results under different experimental settings, we found that: 1) The FFC classifier demonstrates superior performance compared to the MLP classifier, with a notable decrease in RMSE by approximately 3%; 2) the replacement of the cross-entropy loss function with FFC leads to improvements across all metrics; 3) pre-training on the rotation regression task facilitates better parameter initialization, benefiting all performance metrics except Accuracy; 4) MEC aids in better handling of hard samples, resulting in improvements across all metrics.

4.4 Interpretability

The proposed MFF-KAN exhibits strong interpretability. We enhance the understanding of the model's diagnostic process through key pathway selection and visualization in the trained MFF-KAN. Specifically, starting from the final output layer of FFC, we consider all the classes as key output nodes. Then, we select key input nodes with the most significant influence on the key output nodes based on their significance scores. The significance score of the i-th input node in l-th layer is computed as follows:

$$score_{l,i} = \frac{1}{n_{\mathrm{key}}} \sum_j \left|\phi_{l,i,j}\right|_1. \tag{3}$$

Here n_{key} is the number of key output nodes. j refers to the index of the key output nodes rather than all nodes. $\left|\phi_{l,i,j}\right|_1$ represents the average magnitude of the activation function on the corresponding edge. These key input nodes are then set as the key output

nodes for the preceding layer, and the process is iterated until reaching the image feature vector $\boldsymbol{f}_{\mathrm{I}}$ or physiological indicators. Through this process, we achieve selection of the key pathways from input to output in MFF-KAN, as shown in Fig. 3.

For each key input node in the image feature vector, we aim to identify which tongue characteristics are associated with it. By examining the representative tongue images in each node group, we observed consistent characteristics. For instance, most images in the group for node 210 tend to show a white tongue coating, while those in the group for node 275 often display a purple tongue. These tongue characteristics are highly correlated with the values of the key nodes, indicating that they are crucial features for our FLD severity tongue diagnosis task.

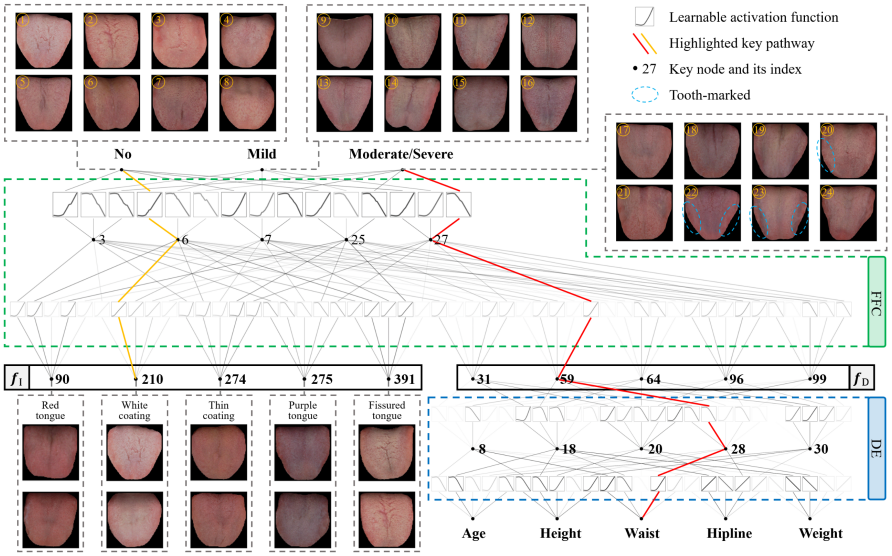

Fig. 3. Visualization of input-output relationships in MFF-KAN. For each layer, the top 5 key nodes are selected to visualize based on their importance scores. Two key pathways are highlighted to demonstrate how waist circumference and white tongue coating influence the output. For each level of FLD severity, tongue images with the highest output probability are shown.

Using the visualized key pathways, we can easily understand how each input affects the final outputs. Specifically, we first select an input node of interest and follow its primary propagation pathway (represented by opaque lines), recording the correlation of activation functions along the way. For instance, in the red-highlighted pathway in Fig. 3, the "Waist" input value undergoes two positive and two negative correlations, eventually contributing to the "Moderate/Severe" output. This implies that a larger waist circumference positively influences the prediction for Moderate/Severe FLD. Meanwhile, in the yellow-highlighted pathway, a similar analytical process indicates that the "white coating" characteristic contributes to the "No" output. Hence, the relative relationship between inputs and outputs can be interpreted and cross-referenced against medical domain knowledge.

We further present samples with the highest output logits for each category at the top of Fig. 3, providing a more intuitive understanding of which tongue images are considered representative for each FLD severity level. It is evident that in terms of tongue color, non-FLD cases appear redder, while FLD patients tend to exhibit a more purplish hue. Additionally, non-FLD samples predominantly show a white tongue coating, while some mild samples exhibit a greasy, thick coating (annotated as Nos. 10, 11, and 14 in the figure). Furthermore, tooth marks are more frequently observed in patients with moderate to severe FLD (Nos. 20, 22, and 23). Overall, a red tongue and white coating are indicative of lower FLD risk, whereas a purplish tongue and greasy coating are risk factors. Additionally, tooth marks are associated with greater disease severity.

5 Conclusion

In this paper we presented a Hard sample Mining-based Tongue Diagnosis Framework (HM-TDF) for multi-class classification of FLD severity. We introduced a Multi-source Feature Fusion Kolmogorov-Arnold Network (MFF-KAN) along with an effective training strategy. We released a novel tongue diagnosis dataset (Tongue-FLD) for FLD severity classification. Experimental results showed that the proposed method outperforms existing tongue diagnosis approaches. Furthermore, MFF-KAN offers a clear interpretation by visualizing key pathways within the network.

Acknowledgments. This study was funded by National Key R&D Program of China (No. 2023YFC2604400), National Natural Science Foundation of China (No. 82304086), Natural Science Foundation of Fujian Province (Nos. 2022J01707 and 2023J01630), Fujian Science and Technology Plan Industry-University-Research Cooperation Project of China (No. 2021H6015), Public Technology Service Platform Project of Xiamen City of China (No. 3502Z20231043), Government Funding of Fuqing City (No. 2019B003), High-level Talents Research Start-up Project of Fujian Medical University (Nos. XRCZX2020034, XRCZX2021025, XRCZX2022018).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Le, M.H., et al.: Global incidence of non-alcoholic fatty liver disease: A systematic review and meta-analysis of 63 studies and 1,201,807 persons. Journal of Hepatology 79, 287–295 (2023)
2. Han, S.K., Baik, S.K., Kim, M.Y.: Non-alcoholic fatty liver disease: Definition and subtypes. Clinical and Molecular Hepatology 29, S5–S16 (2023)
3. Tarantino, G., Citro, V.: What are the common downstream molecular events between alcoholic and nonalcoholic fatty liver? Lipids in Health and Disease 23, (2024)
4. Solos, I., Liang, Y.: A historical evaluation of Chinese tongue diagnosis in the treatment of septicemic plague in the pre-antibiotic era, and as a new direction for revolutionary clinical research applications. Journal of Integrative Medicine 16, 141–146 (2018)

5. Liu, Q., et al.: A survey of artificial intelligence in tongue image for disease diagnosis and syndrome differentiation. Digital Health 9, (2023)
6. Zhang, Y., et al.: Relationship between thick or greasy tongue-coating microbiota and tongue diagnosis in patients with primary liver cancer. Frontiers in Microbiology 13, (2022)
7. Zhao, C.-Q., Zhou, Y., Ping, J., Xu, L.-M.: Traditional Chinese medicine for treatment of liver diseases: progress, challenges and opportunities. Journal of Integrative Medicine 12, 401–408 (2014)
8. Jiang, T., et al.: Application of computer tongue image analysis technology in the diagnosis of NAFLD. Computers in Biology and Medicine 135, (2021)
9. Wang, R.-R., et al.: Noninvasive Diagnostic Technique for Nonalcoholic Fatty Liver Disease Based on Features of Tongue Images. Chinese Journal of Integrative Medicine 30, 203–212 (2024)
10. Zhang, Q., Wen, J., Zhou, J., Zhang, B.: Missing-view completion for fatty liver disease detection. Computers in Biology and Medicine 150, (2022)
11. Dai, S., et al.: Application of intelligent tongue image analysis in Conjunction with microbiomes in the diagnosis of MAFLD. Heliyon 10, (2024)
12. Liu, Z., et al.: Kan: Kolmogorov-arnold networks. arXiv:2404.19756. arXiv preprint (2024)
13. Bodner, A.D., Tepsich, A.S., Spolski, J.N., Pourteau, S.: Convolutional Kolmogorov-Arnold Networks. arXiv:2406.13155. arXiv preprint (2024)
14. Yang, X., Wang, X.: Kolmogorov-Arnold Transformer. arXiv:2409.10594. arXiv preprint (2024)
15. Fan, J., et al.: Guidelines for the diagnosis and management of nonalcoholic fatty liver disease: update 2010. Journal of Digestive Diseases 12, 38–44 (2011)
16. Wang, X., et al.: Constructing tongue coating recognition model using deep transfer learning to assist syndrome diagnosis and its potential in noninvasive ethnopharmacological evaluation. Journal of Ethnopharmacology 285, (2022)
17. Li, J., et al.: A tongue features fusion approach to predicting prediabetes and diabetes with machine learning. Journal of Biomedical Informatics 115, (2021)
18. Qiu, D., et al.: A novel tongue feature extraction method on mobile devices. Biomedical Signal Processing and Control 80, (2023)
19. Yuan, L., et al.: Development of a tongue image-based machine learning tool for the diagnosis of gastric cancer: a prospective multicentre clinical cohort study. EClinicalMedicine 57, (2023)
20. Dosovitskiy, A., et al.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv:2010.11929. arXiv preprint (2020)
21. Liu, Z., Mao, H., Wu, C., Feichtenhofer, C., Darrell, T., Xie, S.: A ConvNet for the 2020s. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11966–11976. (2022)
22. Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W., Wang, X.: Vision Mamba: Efficient Visual Representation Learning with Bidirectional State Space Model. arXiv:2401.09417. arXiv preprint (2024)
23. Yilmaz, A.E., Demirhan, H.: Weighted kappa measures for ordinal multi-class classification performance. Applied Soft Computing 134, (2023)
24. Demsar, J.: Statistical comparisons of classifiers over multiple data sets. Journal of Machine Learning Research 7, 1–30 (2006)