MLLM-Enhanced Region-Aware Bidirectional Evidence-Based Model for Tongue Diagnosis

Yiwei Du[1†], Chenyang Lu[1†], Baochen Fu[3], Yuqi Fang[1,2*], and Caifeng Shan[1*]

[1] School of Intelligence Science and Technology, Nanjing University, Suzhou, Jiangsu, 215163, China
[2] National Institute of Healthcare Data Science at Nanjing University, Nanjing University, Nanjing, Jiangsu, 210093, China
[3] School of Software, Shandong University, Jinan, Shandong, 250101, China
{yqfang,cfshan}@nju.edu.cn

[†] Equal contribution [*] Co-corresponding authors

Abstract. Tongue diagnosis, a convenient and noninvasive traditional diagnostic method in Traditional Chinese Medicine (TCM), provides an important tool for early health screening. Tongue images not only reveal TCM syndrome patterns but also allow a preliminary assessment of relevant organ health. However, most existing methods face three major limitations: (i) syndrome patterns and organ states prediction are often treated as independent tasks, ignoring their coupled mechanisms; (ii) tongue region-dependent signs are insufficiently integrated with global tongue appearance, resulting in inadequate attention to salient local cues; and (iii) clinical priors and TCM knowledge are underutilized, constraining clinically grounded reasoning and interpretability. To address these limitations, we propose an MLLM-Enhanced Region-Aware Bidirectional Evidence-Based Model for tongue diagnosis. Specifically, first, an Augmented Global-Local Feature Fusion (AGLFF) module is introduced to reconcile holistic tongue context with regional cues by mutually refining global and local representations, strengthening region-sensitive evidence extraction. Second, an Uncertainty-Weighted Bidirectional Mixture-of-Experts (UWBMoE) module is designed to propagate syndrome-level and organ-level information, thereby stabilizing multi-task learning and cross-level reasoning. Moreover, a multimodal large language model (MLLM) is incorporated to enrich semantic representations and improve alignment between visual evidence and clinically meaningful concepts. Experiments demonstrate that the proposed approach outperforms state-of-the-art baselines on both syndrome patterns and organ states prediction tasks. The framework provides clinically relevant, evidence-based decision support for noninvasive screening by jointly modeling syndrome-organ associations. The source code is publicly available at https://github.com/NJU-MedAI-Lab/CycleTCM.

Keywords: Tongue Diagnosis · Syndrome-Organ Association · Evidence-Based Model.

1 Introduction

Noncommunicable diseases (NCDs) are now the leading cause of death globally, responsible for 41 million deaths annually, about 74% of all deaths [24,5,25]. In this context, efficient early screening is critical for risk identification, enabling timely interventions to slow disease progression and reduce mortality and healthcare burdens [16,12]. However, comprehensive invasive examinations are impractical due to their high cost, time consumption, and resource requirements [22]. Additionally, invasive procedures are not suitable for all patients, particularly older adults, due to reduced tolerance and higher complication risks [9]. Thus, there is a clear need for noninvasive, rapid, and convenient screening methods for early lesion detection and precision management.

Tongue diagnosis, a core component of inspection in Traditional Chinese Medicine (TCM), supports clinical decision-making through visual assessment of tongue features [2]. Grounded in the holistic theory linking *zang-fu organs, meridians, and qi-blood-body fluids*, it holds that internal physiological and pathological conditions are reflected in observable tongue characteristics [17,14,23]. Traditional theory further associates specific tongue regions with corresponding organ systems [1,19]. Recently, tongue zoning formalizes region partitioning to enable automated, organ-oriented quantitative assessment [10].

Advances in AI for medical imaging increasingly support intelligent tongue-image analysis [7]. Existing studies, such as HWmixer [26], TDFnet [8], and TV-MoE [15] target TCM syndrome pattern recognition, whereas ILPnet [18] and FCRAN [11] focus on single-dimensional modeling of organ status or constitution. In addition, Signnet [3] further learns structured local features by separating tongue border and body, and MIRnet [13] improves multi-label robustness via MAE pretraining and graph-attention reasoning. Despite these advances, several limitations remain. (i) Syndrome identification and organ-state assessment are typically pursued as separate problems, failing to explicitly model their intrinsic relationship. (ii) Existing approaches often under-exploit the complementarity between regional tongue signs and global appearance, diluting the representation of salient local cues. (iii) TCM-informed clinical priors are seldom incorporated in a rigorous manner, limiting meaningful reasoning and interpretability.

To address these challenges, we propose an MLLM-Enhanced Region-Aware Syndrome-Organ Bidirectional Evidence-Based Model for tongue diagnosis. The proposed model follows a three-stage design. First, we construct coupled global-regional representations by fusing whole-tongue global features with region-level tokens via an Augmented Global-Local Feature Fusion (AGLFF) module. Second, we model the tasks through an Uncertainty-Weighted Bidirectional Mixture-of-Experts (UWBMoE) module that routes and propagates evidence between syndrome and organ branches in both directions, while consistency constraints regularize cross-branch interactions. Third, we extract high-level semantic features from an MLLM and inject them into visual representations through adaptive fusion to support joint inference. Our contributions are summarized below: (i) **Region-aware global-local evidence aggregation**: AGLFF improves the capture of region-specific tongue signs and their global context, pro-

ducing richer fused representations and enhancing the utilization of global-local complementarity. (ii) **Cross-level coupling for joint inference**: UWBMoE strengthens syndrome-organ co-modeling by enabling bidirectional evidence interaction and complementation, enhancing cross-level reasoning and improving joint decision consistency. (iii) **Clinically grounded semantic enrichment**: MLLM-derived priors enhance semantic alignment and produce more clinically meaningful predictions, improving model utility in screening-oriented settings.

2 Methodology

As illustrated in Fig. 1, our approach consists of an augmented global-local feature fusion (AGLFF) module, an uncertainty-weighted bidirectional MoE (UWBMoE) module, and a clinical-prior-driven semantic enhancement module. Our framework first enhances global-local visual representations via AGLFF, then enables uncertainty-aware bidirectional interaction between branches through UWBMoE, and finally integrates clinical-prior-driven high-level semantic features for complementary multimodal fusion.

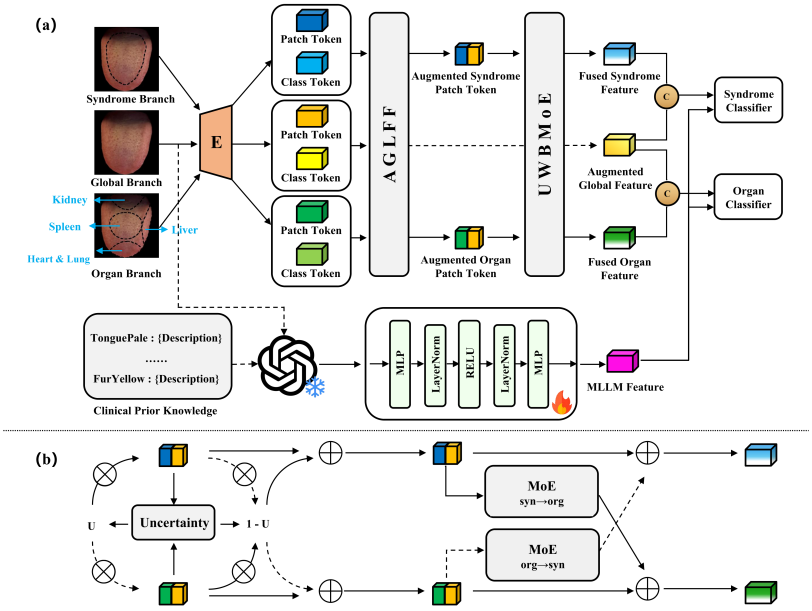

Fig. 1. Overview of the proposed framework. (a) Global-regional features are constructed by AGLFF, followed by uncertainty-weighted bidirectional routing via UWBMoE. MLLM priors are adaptively fused for joint prediction. (b) Illustration of UWBMoE: uncertainty-guided complementary weighting enables bidirectional MoE interaction (Syndrome-Organ) for consistency-aware inference.

2.1 Augmented Global-Local Feature Fusion

The global image of the tongue captures the overall structure, while region-specific inputs highlight local diagnostic areas. Regional inputs are obtained by a simple rule-based partition using the image center as reference, cropping upper/lower, left/right strip regions, and a central region. To integrate global features and fine-grained regional evidence, we propose the AGLFF module to collaboratively enhance global tokens and patch-level tokens across global and local branches, producing multi-scale representations for subsequent fusion.

Specifically, as shown in Fig. 1(a), AGLFF takes six types of tokens generated by the three-branch image encoder as input: the global branch produces a mean token $\mathbf{T}^{g}_{cls}$ and patch tokens $\mathbf{T}^{g}_{patch}$, the syndrome-local branch produces $\mathbf{T}^{syn}_{cls}$ and $\mathbf{T}^{syn}_{patch}$, and the organ-local branch produces $\mathbf{T}^{org}_{cls}$ and $\mathbf{T}^{org}_{patch}$. Through fusion, the module outputs three tokens: $\mathbf{T}^{g*}_{cls}$ for enhanced global representation, $\mathbf{T}^{syn}_{fused}$ for enhanced patch-level syndrome representation, and $\mathbf{T}^{org}_{fused}$ for enhanced patch-level organ representation. At the cls token enhancement level, we update the global representation by attending to the syndrome and organ cls tokens:

$$\tilde{\mathbf{T}}^{g\leftarrow i}_{cls} = \mathbf{T}^{g}_{cls} + \mathrm{CA}\big(\mathbf{Q}=\mathbf{T}^{g}_{cls},\, \mathbf{K}=\mathbf{T}^{i}_{cls},\, \mathbf{V}=\mathbf{T}^{i}_{cls}\big), \quad i \in \{\text{syn, org}\}, \tag{1}$$

where $\mathrm{CA}(\mathbf{Q},\mathbf{K},\mathbf{V})$ denotes a multi-head cross-attention operator. We then obtain the final global representation $\mathbf{T}^{g*}_{cls}$ by averaging the two enhanced tokens $\tilde{\mathbf{T}}^{g\leftarrow\text{syn}}_{cls}$ and $\tilde{\mathbf{T}}^{g\leftarrow\text{org}}_{cls}$. At the patch token enhancement level, we apply a convolutional projector to each branch to obtain the refined patch tokens $\tilde{\mathbf{T}}^{i}_{patch}$, where $i \in \{\text{g, syn, org}\}$. For each local branch $j \in \{\text{syn}, \text{org}\}$, we compute a gating map and fuse the enhanced local features with the global patch tokens:

$$\mathbf{T}^{j}_{fused} = \mathbf{G}_j \odot \left(\mathbf{T}^{j*}_{patch} + \tilde{\mathbf{T}}^{j}_{patch}\right) + (\mathbf{1}-\mathbf{G}_j) \odot \tilde{\mathbf{T}}^{g}_{patch}, \tag{2}$$

where $\mathbf{G}_j = \sigma\Big(\mathrm{Conv}_{gate}\Big(\mathbf{T}^{j*}_{patch} + \tilde{\mathbf{T}}^{j}_{patch}\Big)\Big)$, $\mathbf{T}^{j*}_{patch}$ is obtained by global-average pooling of $\tilde{\mathbf{T}}^{j}_{patch}$, and $\sigma(\cdot)$ denotes the Sigmoid function. This gated fusion selectively preserves discriminative regional responses while injecting complementary global context to suppress redundant or spurious activations.

AGLFF enhances the representational capacity of the global cls token and the fine-grained spatial cues relevant to syndrome or organ discrimination, leading to more robust downstream reasoning.

2.2 Uncertainty-Weighted Bidirectional Mixture of Experts

Syndrome patterns and organ states prediction are often treated as independent tasks, ignoring their coupled mechanisms. UWBMoE module couples two components: an uncertainty-weighted fusion step and a conditional Top-K MoE interaction, aiming to achieve bidirectional complementarity between syndrome and organ information through reliability-aware evidence correction and context-conditioned expert selection.

In the uncertainty-weighted fusion interaction, we denote the patch embeddings from the syndrome and organ branches by $\{\mathbf{e}_n^{syn}\}_{n=1}^N$ and $\{\mathbf{e}_n^{org}\}_{n=1}^N$, respectively, where $\mathbf{e}_n^{syn}, \mathbf{e}_n^{org} \in \mathbb{R}^d$, N is the number of patches, and d denotes the embedding dimension. To quantify the consistency of local evidence across syndrome and organ branches, we compute the cosine similarity s_n between the paired patch embeddings and map it to a branch uncertainty score u_n:

$$s_n = \cos(\mathbf{e}_n^{syn}, \mathbf{e}_n^{org}) = \frac{\langle \mathbf{e}_n^{syn}, \mathbf{e}_n^{org} \rangle}{\|\mathbf{e}_n^{syn}\|_2 \, \|\mathbf{e}_n^{org}\|_2 + \varepsilon}, \quad u_n = (1 - s_n)/2. \tag{3}$$

Considering that uncertainty magnitudes may vary across samples due to differences in imaging conditions and individual variability, we perform sample-wise min-max normalization for the two uncertainty sets and take their complements α_n^{syn}, α_n^{org} as confidence weights. These confidence weights are then used to realize bidirectional fusion and mutual calibration:

$$\mathbf{F}_n^{syn} = \alpha_n^{syn} \, \mathbf{e}_n^{syn} + (1 - \alpha_n^{syn}) \, \mathbf{e}_n^{org}, \quad \mathbf{F}_n^{org} = \alpha_n^{org} \, \mathbf{e}_n^{org} + (1 - \alpha_n^{org}) \, \mathbf{e}_n^{syn}. \tag{4}$$

The uncertainty-weighted fusion part yields the syndrome and organ patch tokens, $\mathbf{F}^{syn}$ and $\mathbf{F}^{org}$, which are then fed into the bidirectional MoE network for interaction refinement via conditional Top-K mixture-of-experts in both directions. For either direction, we denote the main-branch input as $\mathbf{F}^{main}$ and the conditioning-branch input as $\mathbf{F}^{cond}$. Routing mechanism produces the expert gating distribution by combining a token-wise router and a condition-wise router. The refined representation $\mathbf{F}^{ref}$ is computed as:

$$\mathbf{F}^{ref} = \mathbf{F}^{main} + \sum\nolimits_{e \in \mathrm{TopK}(\mathbf{g})} \hat{\mathbf{g}}_e \odot \mathcal{E}_e\big(\mathbf{F}^{main}\big), \tag{5}$$

where $\mathbf{g} = \mathrm{softmax}\big(\mathbf{W}_x \mathbf{F}^{main} + \mathbf{W}_c \, \mathbf{F}^{cond}\big)$, $\mathbf{W}_x$ and $\mathbf{W}_c$ denote the routing projections for the main branch and the conditioning branch, respectively. $\mathcal{E}_e(\cdot)$ denotes the e-th expert MLP, and $\hat{\mathbf{g}}_e$ is the normalized Top-K gating weight. This bidirectional MoE ensures that each patch-token transformation is jointly constrained by its own representation and cross-dimension evidence. The $\mathbf{F}^{ref}$ obtains features from both syndrome and organ, leading to improved accuracy and generalization in downstream prediction tasks.

2.3 MLLM-Enhanced Clinical Prior Embedding

To further improve the accuracy and reliability of classification, we introduce an MLLM-driven clinical-prior semantic enhancement module. Our system prompt is used to guide the MLLM to act as a TCM tongue-diagnosis expert, focusing on tongue attributes and organ-related abnormalities. The whole tongue image and TCM semantic syndrome descriptions are fed into the frozen **Qwen3-VL-4B-Instruct** model together with the prompt, from which we extract the last hidden-state feature H_{MLLM}. A lightweight trainable adapter then projects it into the visual feature space: $\tilde{H} = \mathcal{A}(H_{\mathrm{MLLM}})$, where $\mathcal{A} = \mathrm{MLP}_2 \circ \mathrm{LN} \circ \mathrm{ReLU} \circ \mathrm{LN} \circ \mathrm{MLP}_1$. And the projected output is injected into both syndrome and organ branches.

3 Experiments and Results

3.1 Datasets and Implementation Details

We conduct experiments on the TongueDx dataset [3]. Designed for remote tongue-diagnosis scenarios, TongueDx contains 5109 tongue images from 4650 subjects, collected under diverse environments and imaging conditions using smartphones and laptop cameras. In our settings, we use eight syndrome-related attribute labels (*TonguePale, TipSideRed, RedSpot, Ecchymosis, Crack, ToothMark, FurThick, FurYellow*) and five organ-abnormality labels (*Heart, Lung, Spleen, Liver, Kidney*). These labels are annotated by expert TCM physicians and serve as ground-truth TCM diagnoses. The dataset is split at the subject level into 3,371/843/895 samples for training/validation/testing, respectively, avoiding data leakage.

Our method is implemented in PyTorch and optimized with Adam. All experiments are conducted on an NVIDIA H20 GPU. Images are resized to 224×224, and the batch size is set to 32. We train for 200 epochs with early stopping (patience=50). The initial learning rate is set to $2e-4$ and is reduced by a factor of 0.3 using ReduceLROnPlateau [21]. The weight decay is set to $1e-4$.

3.2 Results and Discussion

Comparison with Existing SOTA Methods. To evaluate the effectiveness of our method, we compare it with advanced tongue classification models including: Signnet [3], MIRnet [13], ILPnet [18], HWmixer [26], TDFnet [8], and TVMoE [15]. As shown in Tables 1 and 2, our method achieves the highest or near-highest scores across metrics including Accuracy (Acc) and F1. Quantitatively, for syndrome prediction, our method improves the average Acc and average F1 over the second-best method by 1.36% and 1.60%; for organ prediction, our method improves the average Acc and average F1 over the second-best method by 2.57% and 0.47%. Subject-level bootstrap analysis further shows statistically reliable gains in average Acc, with 95% confidence intervals of +1.54 percentage points [+0.28, +2.74] for syndrome prediction and +2.57 percentage points [+0.80, +4.34] for organ prediction, compared with the second-best model, TVMoE. In addition, the GradCAM [20] visualizations in Fig. 2 show that our method directs the model's attention to tongue regions that align with syndrome and organ cues, providing more clinically plausible evidence to support model decisions.

Ablation Study. Table 3 demonstrates the contribution of key components to the model's performance. Backbone-based, we introduce the AGLFF, UWBMoE, and MLLM-Enhanced module respectively, with the average accuracy improvement of 0.53%, 1.30% and 0.62% for syndrome prediction, and 1.25%, 1.36% and 1.09% for organ prediction. Furthermore, enabling AGLFF and UWBMoE jointly yields more pronounced gains on both tasks, improving syndrome accuracy by 1.62% and organ accuracy by 1.96%. The full model integrating all

Table 1. Comparison among methods for syndrome classification (values in mean%).

Model	Pale		TipSideRed		Spot		Ecchymosis		Crack	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
Signnet	88.38	29.73	70.50	64.99	77.09	75.27	89.27	30.43	84.58	90.87
MIRnet	**89.61**	00.00	68.38	63.01	66.03	66.00	**90.84**	00.00	83.47	90.84
ILPnet	82.79	**45.39**	66.03	67.66	76.54	78.03	85.70	34.02	85.36	91.79
HWmixer	84.58	39.47	70.06	63.69	78.77	76.60	88.72	21.71	86.03	**91.84**
TDFnet	88.04	33.54	68.04	65.12	75.64	75.23	87.26	26.92	81.68	89.08
TVMoE	86.93	38.74	71.51	67.68	79.11	76.36	88.38	24.64	84.13	90.55
Ours	87.93	41.94	**74.64**	**70.48**	**81.23**	**81.12**	90.06	**35.04**	**86.15**	91.81

Model	ToothMark		FurThick		FurYellow		**Average**	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1
Signnet	77.21	81.18	97.32	98.64	**93.18**	79.60	84.69	68.84
MIRnet	68.49	75.61	97.32	98.63	86.48	43.19	81.32	54.66
ILPnet	74.19	80.80	97.21	98.58	88.49	72.97	82.04	71.15
HWmixer	74.53	79.39	96.54	98.20	90.06	73.11	83.66	68.00
TDFnet	74.19	80.67	96.98	98.47	90.84	75.00	82.84	68.00
TVMoE	75.64	81.08	97.21	98.58	**93.18**	**80.39**	84.51	69.75
Ours	**77.77**	**82.74**	**97.54**	**98.75**	93.07	80.13	**86.05**	**72.75**

Table 2. Comparison among methods for organ classification (values in mean%).

Model	Heart		Lung		Spleen		Liver		Kidney		**Average**	
	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1	Acc	F1
Signnet	–	–	–	–	–	–	–	–	–	–	–	–
MIRnet	69.05	66.26	65.59	72.60	99.33	99.66	70.61	77.88	73.18	81.34	75.55	79.55
ILPnet	66.59	69.33	67.60	75.21	99.33	99.66	73.07	81.04	76.31	82.85	76.58	81.62
HWmixer	69.27	65.23	61.23	73.57	99.33	99.66	72.18	78.14	71.17	81.03	74.64	79.53
TDFnet	70.73	67.89	65.47	75.34	99.33	99.66	66.93	78.33	73.52	82.48	75.20	80.74
TVMoE	65.25	68.68	69.61	76.63	99.33	99.66	75.08	81.59	78.32	84.46	77.52	82.20
Ours	**73.41**	**69.64**	**70.61**	**76.45**	**99.33**	**99.66**	**77.88**	**82.84**	**79.22**	**84.75**	**80.09**	**82.67**

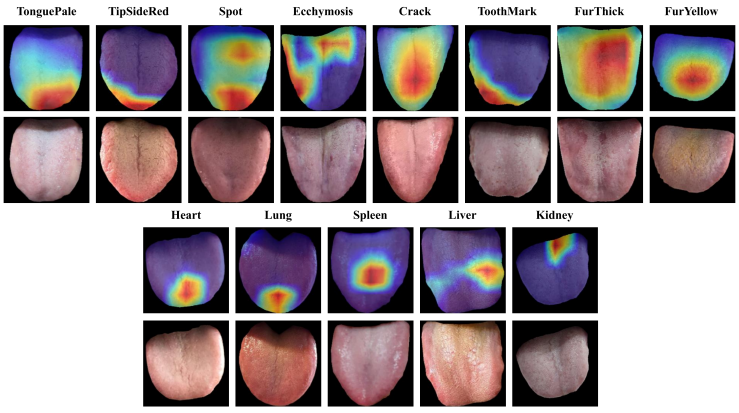

Fig. 2. GradCAM visualizations for eight syndromes and five organs.

Table 3. Average performance across syndrome and organ classification tasks in the ablation study (values in mean%).

Experimental Setting				Syndrome Classification				Organ Classification			
Backbone	AGLFF	UWBMoE	MLLM-Enhanced	Acc	F1	SEN	PRE	Acc	F1	SEN	PRE
✓				83.56	68.35	69.71	67.75	77.50	78.82	79.17	79.09
✓	✓			84.09	70.47	72.92	68.50	78.75	81.64	84.81	78.85
✓		✓		84.86	69.83	69.99	70.11	78.86	80.77	81.13	80.50
✓			✓	84.18	70.69	72.13	70.99	78.59	81.33	83.79	79.45
✓	✓	✓		85.18	71.60	72.93	70.46	79.46	81.83	83.24	**80.51**
✓	✓	✓	✓	**86.05**	**72.75**	**73.91**	**72.11**	**80.09**	**82.67**	**85.41**	80.24

components achieves the best overall performance, confirming the synergistic contributions and necessity of each module within the unified framework.

Discussion for Syndrome-Organ Associations. To further delineate potential coupling between syndrome representations and organ-specific attribution, we employed a Sankey diagram [6] to visualize the weight propagation from syndromes to organs. Fig. 3 indicates substantial non-uniformity and selectivity in organ attribution across syndrome patterns. Specifically, the *TonguePale*→ *Liver* connection carries the highest weight, suggesting that the "pale tongue" feature preferentially co-occurs with liver-related phenotypic signatures [4]. In addition, *Ecchymosis* retains appreciable weights to the *Heart*, *Lung*, and *Kidney*, implying that this feature is unlikely to be confined to a single-organ signal but may instead reflect a shared cross-organ, system-level pathological background. Collectively, these associations support the notion that syndrome features may encode systemic pathophysiological information, thereby strengthening the clinical interpretability of the inferred syndrome-organ associations.

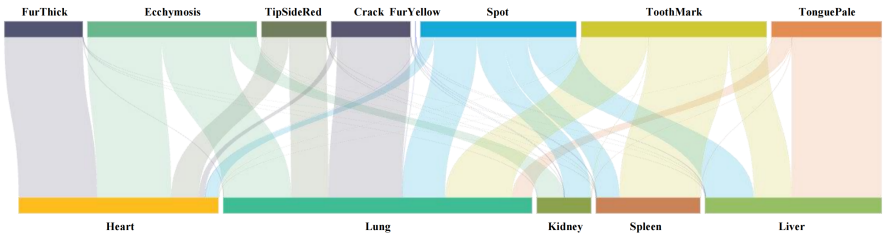

Fig. 3. Sankey diagram of syndrome-to-organ weight propagation.

4 Conclusions

In conclusion, we propose an MLLM-Enhanced Region-Aware Syndrome-Organ Bidirectional Evidence-Based Model for Tongue Diagnosis. Unlike prior methods

that model syndrome patterns or organ status in isolation, our approach performs unified representation learning and coupled inference to more fully characterize their intrinsic association. Experimental results show consistent improvements over single-task or independently trained baselines for both syndrome and organ prediction, yielding more robust overall performance. These findings suggest that our framework provides a practical, noninvasive, and scalable option for screening-oriented clinical decision support.

Acknowledgments. This work is funded by Young Scientists Fund of the National Natural Science Foundation of China (13001024).

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

References

1. Akrami, M., Jafari, A., Pedrammehr, S.: Disease Diagnosis and Disease Patterns by Visual Inspection of the Tongue. Cambridge Scholars Publishing (2024)
2. Bing, Z., Hongcai, W.: Diagnostics of traditional Chinese medicine. Singing Dragon (2010)
3. Chen, Y., Ho, S.S., Xu, C., Xie, Y.J., Yeung, W.F., He, S., Qin, J.: Dr. tongue: Sign-oriented multi-label detection for remote tongue diagnosis. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39, pp. 2302–2310 (2025)
4. Deng, J., Dai, S., Liu, S., Tu, L., Cui, J., Hu, X., Qiu, X., Lu, H., Jiang, T., Xu, J.: Clinical study of intelligent tongue diagnosis and oral microbiome for classifying tcm syndromes in masld. Chinese Medicine **20**(1), 78 (2025)
5. Dong, D., Huang, H., Zhou, A., Dong, C.: A systematic analysis of china's non-communicable disease policies: policy themes, policy instruments, and policy effectiveness from 2009 to 2025. Frontiers in Public Health **13**, 1643400 (2025)
6. Foundation, A.S.: Echarts - sankey diagram. https://echarts.apache.org/en/option.html#series-sankey, accessed: 2023-02-27
7. Fu, B., Duan, M., Li, Z., Zuo, X., Qiao, X.: Automatic diagnosis model of gastrointestinal diseases based on tongue images. In: International conference on pattern recognition. pp. 290–301. Springer (2024)
8. Gao, J., Chen, T., Xu, Y., Wu, Y., Liu, K., Qiu, W., Ye, W.: Accurate fatty liver disease diagnosis with a multi-source feature fusion model on the segmented tongue image dataset. Journal of Advanced Research (2025)
9. Grecula, M.J., Caban, M.E.: Common orthopaedic problems in the elderly patient. Journal of the American College of Surgeons **200**(5), 774–783 (2005)
10. Ho, S.C., Chen, Y., Xie, Y.J., Yeung, W.F., Chen, S.C., Qin, J.: Visceral condition assessment through digital tongue image analysis. Frontiers in artificial intelligence **7**, 1501184 (2025)
11. Hu, Y., Wen, G., Luo, M., Yang, P., Dai, D., Yu, Z., Wang, C., Hall, W.: Fully-channel regional attention network for disease-location recognition with tongue images. Artificial Intelligence in Medicine **118**, 102110 (2021)
12. Hussain, S.: Advances in cancer screening and early diagnosis: A focus on low-resource settings. Non-Communicable Diseases and Care **1**(2), 28–39 (2024)

13. Kong, S., Wang, Z., Cui, N., Tang, H., Meng, Y., Wei, Y., Chen, F., Wang, Y., Cai, Z., Wang, Y., et al.: Mirnet: Integrating constrained graph-based reasoning with pre-training for diagnostic medical imaging. arXiv preprint arXiv:2511.10013 (2025)
14. Lozano, F.: Basic theories of traditional chinese medicine. In: Acupuncture for pain management, pp. 13–43. Springer (2013)
15. Lu, X.Z., Liu, S., Lin, X.X., Zeng, Y., Chen, J.H., Ke, W.P., Deng, J.F., Cheng, M.Q., Li, W., Chen, L.D., et al.: An ai-powered tongue image model for home-based monitoring of liver fibrosis. npj Digital Medicine (2025)
16. Makhmirzaeva, G.G., Golubkina, E.S., Airapetyan, A.D., Root, V.A., Uranoolovich, O.K., Kuzhuget, A.A., Aliyeva, K.M.: The importance of early screening in preventing chronic diseases. Revista Latinoamericana de Hipertension (1) (2025)
17. Mize, A., Welden, J.: Toward a more accurate tongue analysis in chinese medicine (2023)
18. Qiao, X., Lu, C., Duan, M., Liu, Z., Liu, Y., Chen, W., Gao, R., Li, Z., Zuo, X.: Intelligent tongue diagnosis model for gastrointestinal diseases based on tongue images. Biomedical Signal Processing and Control **96**, 106643 (2024)
19. Sakr, M.F.: Tongue Lesions: Diagnostic Challenges and Therapeutic Strategies. Springer Nature (2022)
20. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision. pp. 618–626 (2017)
21. Team, P.: Reducelronplateau — pytorch 1.10.0 documentation. https://pytorch.org/docs/stable/generated/torch.optim.lr_scheduler.ReduceLROnPlateau.html (2021), accessed: 2023-02-27
22. Wankhede, P., Bhuyar, D., Zanwar, S., Pawar, R., Jadhav, M.R., Gandhewar, N., Kulkarni, M.B., Bhaiyya, M., Haick, H.: Artificial intelligence for noninvasive health diagnostics. ACS sensors **10**(11), 8217–8255 (2025)
23. Yu, Q., Wang, W., Gong, M., Lin, Y., Yu, Z., Du, S.: A review of theoretical research on traditional chinese medicine diagnostic methods in 2024. Guidelines and Standards in Chinese Medicine pp. 10–1097 (2026)
24. Yuanfeng, L., Xu, Z.: Active aging and health among older adults in china: a perspective based on downward intergenerational economic support. Frontiers in Public Health **12**, 1337829 (2024)
25. Zhang, J., Wu, Y., Tian, R., Zhao, D.: Developing an integrated evaluation indicator system for non-communicable disease management in china's primary care: a modified delphi-analytic hierarchy process study based on the structure-process-outcome framework. BMC Public Health (2025)
26. Zhang, M., Wen, G., Yang, P., Wang, C., Chen, C.: Multi-label body constitution recognition via hwmixer-mlp for facial and tongue images. Expert Systems with Applications **269**, 126383 (2025)