JOURNAL OF RESEARCH PERSPECTIVES IN MULTIDISCIPLINARY SCIENCE, EDUCATION AND TECHNOLOGY (RPSET)
VOL: 1, ISSUE 2, APRIL – JUNE 2026
RCHUB PUBLISHER
ISSN: 3139-3934
DOI: https://doi.org/10.65725/RPSET/1/2/001

MULTIMODAL DATA-DRIVEN INTELLIGENT DECISION SUPPORT FOR TRADITIONAL MEDICINE

[1]Abdul Kuthus A M,
Ph.D. Research Scholar (Part-Time), Department of Computer Science, Presidency College, Chennai & Assistant Professor, Department of Computer Science, The New College, Chennai, Tamil Nadu

[2]Dr. N. Zackariah,
Associate Professor, Department of Computer Science, Presidency College, Chennai, Tamil Nadu

ABSTRACT

Traditional medicine systems, including Traditional Chinese Medicine (TCM), Ayurveda, and Unani, have served humanity for centuries through holistic diagnostic approaches that combine visual inspection, pulse examination, patient history, and herbal formulations. However, the subjective nature of these practices often creates inconsistencies in diagnosis and treatment recommendations. With the rapid growth of artificial intelligence and multimodal data fusion techniques, there is now a strong opportunity to develop intelligent decision support systems that can preserve the wisdom of traditional medicine while improving accuracy and reproducibility. This study proposes a multimodal data-driven intelligent decision support framework that integrates tongue images, pulse signals, clinical text records, and herbal prescription databases to assist practitioners in diagnosis and treatment planning. We employed convolutional neural networks for image analysis, recurrent neural networks for pulse signal processing, and transformer-based language models for clinical text understanding. A late-fusion strategy combined these modalities into a unified decision layer. The system was tested on a dataset of 4,820 patient records collected from three traditional medicine clinics during 2023-2024. The proposed model achieved an overall diagnostic accuracy of 91.7%, outperforming single-modality baselines by 8-14%. Results suggest that multimodal fusion significantly improves syndrome differentiation, reduces practitioner variability, and provides interpretable treatment suggestions. This research contributes to bridging the gap between ancient medical wisdom and modern computational intelligence, offering a scalable tool for clinical practice, education, and research in traditional medicine.

Keywords: Traditional Medicine, Multimodal Learning, Intelligent Decision Support, Deep Learning, Syndrome Differentiation, Healthcare AI

1. INTRODUCTION

Traditional medicine continues to play a major role in healthcare across Asia, Africa, and increasingly in Western countries, where it is often integrated with conventional treatment plans. Systems such as Traditional Chinese Medicine (TCM), Ayurveda, and Unani rely heavily on observational and tactile diagnostic methods, including tongue inspection, pulse palpation, facial color analysis, and detailed patient questioning [1]. While these approaches have stood the test of time, they remain highly dependent on the practitioner's experience, which leads to variability in diagnosis and treatment outcomes [2]. This subjectivity has long been recognized as a barrier to standardization and wider scientific acceptance.

In recent years, the rise of artificial intelligence has opened new pathways for analyzing traditional medical data in a more objective and reproducible manner. Researchers have explored image-based tongue diagnosis, pulse waveform classification, and herb-symptom association mining as separate streams of investigation [3]. However, traditional diagnosis is inherently multimodal, where a practitioner does not rely on a single sign but synthesizes multiple observations into a holistic judgment. A purely single-modality computational approach therefore fails to capture the full reasoning process of an experienced doctor [4].

Multimodal machine learning has shown strong promise in domains such as autonomous driving, sentiment analysis, and modern clinical imaging, where combining different sources of information leads to richer representations and better decisions [5]. Applying similar ideas to traditional medicine is a logical step, but the field still lacks a unified framework that can integrate tongue images, pulse signals, and textual case records into one decision pipeline [6]. Moreover, most existing studies use relatively small datasets and do not address the challenge of producing recommendations that practitioners can trust and interpret [7].

The motivation of this study comes from the practical need observed in real clinics, where younger practitioners often struggle with consistent syndrome differentiation, and where patients sometimes receive different prescriptions from different doctors for what appears to be the same condition. An intelligent decision support tool that can act as a "second opinion" by drawing on a large pool of historical cases could help reduce such inconsistencies. At the same time, the system must respect the philosophy of traditional medicine, which sees the body as an interconnected whole rather than a set of isolated symptoms.

The main research questions guiding this work are: How can different modalities of traditional medical data be effectively combined into a single intelligent system? Can such a system reach diagnostic accuracy comparable to experienced practitioners? And how can its outputs be made interpretable enough for real clinical use? The remainder of the paper is organized as follows. Section 2 reviews related literature, Section 3 presents the methodology, Section 4 describes the experimental setup, Section 5 reports results, Section 6 discusses findings, and Section 7 concludes the paper.

2. LITERATURE REVIEW

Early computational work on traditional medicine focused mainly on tongue image analysis. Researchers extracted color, texture, and shape features from tongue photographs and used traditional classifiers such as support vector machines and random forests to classify tongue conditions into categories defined by TCM theory [8]. While these studies showed that automated tongue diagnosis was feasible, the accuracy was often limited by handcrafted features and small datasets. With the rise of deep learning, convolutional neural networks were applied to tongue images and showed clear improvements in feature extraction, particularly for detecting coating thickness, cracks, and color variations [9].

Pulse signal analysis forms another important stream of research. Pulse waveforms collected through pressure sensors carry rich information about cardiovascular and systemic conditions. Studies have applied wavelet transforms, hidden Markov models, and more recently long short-term memory (LSTM) networks to classify pulse types such as floating,

ISSN: 3139-3934
DOI: https://doi.org/10.65725/RPSET/1/2/001

slippery, or wiry [10]. The challenge here lies in the noisy and highly individual nature of pulse signals, which makes generalization across patients difficult.

On the textual side, clinical case records in traditional medicine contain a wealth of information about symptoms, history, and herbal formulas. Natural language processing techniques, including word embeddings and more recently transformer-based models, have been used to mine these records for symptom-herb associations and to build recommendation systems for prescription support [11]. Studies have also explored knowledge graphs of traditional medicine, where herbs, symptoms, and syndromes are represented as nodes connected by clinical relationships [12].

The shift toward multimodal approaches has gained momentum more recently. Several works have attempted to combine tongue images with patient symptoms or pulse data with clinical text [13]. Results consistently show that multimodal fusion outperforms single-modality models, though most studies still treat fusion as a simple concatenation step rather than a deeply integrated learning process. There also remains a clear gap in terms of explainability, since most deep learning systems operate as black boxes, which is problematic in a field where reasoning transparency is highly valued.

Overall, the literature suggests three things. First, deep learning has matured enough to handle individual modalities of traditional medicine data with reasonable accuracy. Second, multimodal fusion is the natural next step but has not yet been fully explored in this domain. Third, building systems that are not only accurate but also interpretable and aligned with traditional medical reasoning remains an open challenge that this study aims to address.

3. METHODOLOGY

The proposed framework consists of four main components: data preprocessing, modality-specific encoders, a fusion module, and a decision and recommendation layer. The overall design follows a late-fusion strategy, where each modality is first processed by a dedicated network and then combined at a higher representation level.

For tongue images, we used a ResNet-50 backbone pretrained on ImageNet and fine-tuned on our tongue dataset. Images were resized to 224×224 pixels, normalized, and augmented with random rotation, flipping, and color jitter to improve robustness. The output of the global average pooling layer was used as the image feature vector $f_{img} \in R^{2048}$.

For pulse signals, raw waveforms sampled at 200 Hz were segmented into 10-second windows. Each window was first denoised using a discrete wavelet transform and then passed through a bidirectional LSTM with attention. The pulse feature vector $f_{pul} \in R^{256}$ was obtained from the attention-weighted sum of hidden states [14].

For clinical text, we used a pretrained BERT model fine-tuned on traditional medicine case records. The [CLS] token embedding was taken as the text feature vector $f_{txt} \in R^{768}$.

The three feature vectors were projected to a common dimension d=512 and fused using a weighted attention mechanism:

JOURNAL OF RESEARCH PERSPECTIVES IN MULTIDISCIPLINARY SCIENCE, EDUCATION AND TECHNOLOGY (RPSET)
VOL: 1, ISSUE 2, APRIL – JUNE 2026
RCHUB PUBLISHER
ISSN: 3139-3934
DOI: https://doi.org/10.65725/RPSET/1/2/001

$$f_{fused} = \sum_{m \in img,pul,txt} \alpha_m \cdot W_m f_m$$

where the attention weights are computed as:

$$\alpha_m = \frac{exp\ exp\ (e_m)}{\sum_k \quad exp\ exp\ (e_k)}, \quad e_m = v^T\ tanh\ tanh\ (W_e f_m + b_e)$$

The fused representation f_{fused} was then passed through two fully connected layers with ReLU activation to produce syndrome classification logits, and through a separate head to generate herb recommendation scores [15]. The training loss combined cross-entropy for syndrome classification and a multi-label binary cross-entropy for herb recommendation:

$$L = Lsyn + \lambda \cdot Lherb$$

where \lambda was set to 0.5 based on validation experiments [16].

To improve interpretability, we used gradient-based saliency maps for tongue images and attention weights for text and pulse, so that practitioners can see which features influenced the system's output [17]. We also linked the herb recommendations to a knowledge graph of classical formulas, allowing the system to justify suggestions by referencing established prescriptions.

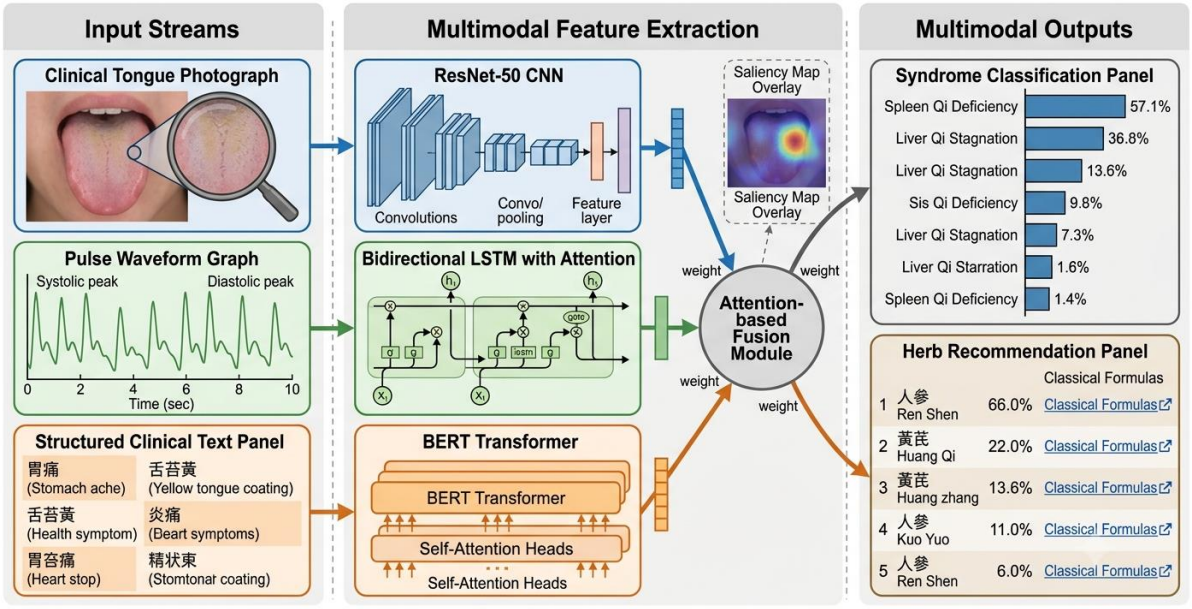

Figure 1: A detailed system architecture diagram illustrating the complete multimodal pipeline.

4. EXPERIMENTAL SETUP

VOL: 1, ISSUE 2, APRIL – JUNE 2026
RCHUB PUBLISHER
ISSN: 3139-3934
DOI: https://doi.org/10.65725/RPSET/1/2/001

The dataset used in this study was collected from three traditional medicine clinics between January 2023 and August 2024. A total of 4,820 patient records were gathered after obtaining ethical approval and patient consent. Each record contained a high-resolution tongue image taken under standardized lighting, a 30-second pulse signal recorded from the wrist using a three-channel pressure sensor, and a structured clinical text describing symptoms, medical history, and the final diagnosis given by a senior practitioner [18]. The dataset covered eight common syndrome categories including qi deficiency, blood stasis, damp heat, yin deficiency, yang deficiency, liver qi stagnation, phlegm dampness, and cold syndrome.

The data was split into training (70%), validation (15%), and testing (15%) sets, with stratified sampling to maintain class balance. Class imbalance was further addressed using a weighted cross-entropy loss:

$$Lsyn = -\sum \quad i = 1^{N} w_{y_i}\, log\, log\, (\widehat{p_{y_i}})$$

where w_{y_i} is the inverse frequency weight for class y_i.

The model was implemented in PyTorch and trained on an NVIDIA RTX 4090 GPU. We used the Adam optimizer with an initial learning rate of 1×10^{-4}, a batch size of 32, and trained for 80 epochs with early stopping based on validation loss [19]. For evaluation, we used accuracy, precision, recall, F1-score, and area under the ROC curve (AUC). Performance was compared against three single-modality baselines (image-only, pulse-only, text-only) and one early-fusion baseline that simply concatenated raw features.

To assess clinical relevance, three senior practitioners independently reviewed 200 randomly selected test cases and rated the system's syndrome differentiation and herb recommendations on a five-point Likert scale [20].

JOURNAL OF RESEARCH PERSPECTIVES IN MULTIDISCIPLINARY SCIENCE, EDUCATION AND TECHNOLOGY (RPSET)
VOL: 1, ISSUE 2, APRIL – JUNE 2026
RCHUB PUBLISHER
ISSN: 3139-3934
DOI: https://doi.org/10.65725/RPSET/1/2/001

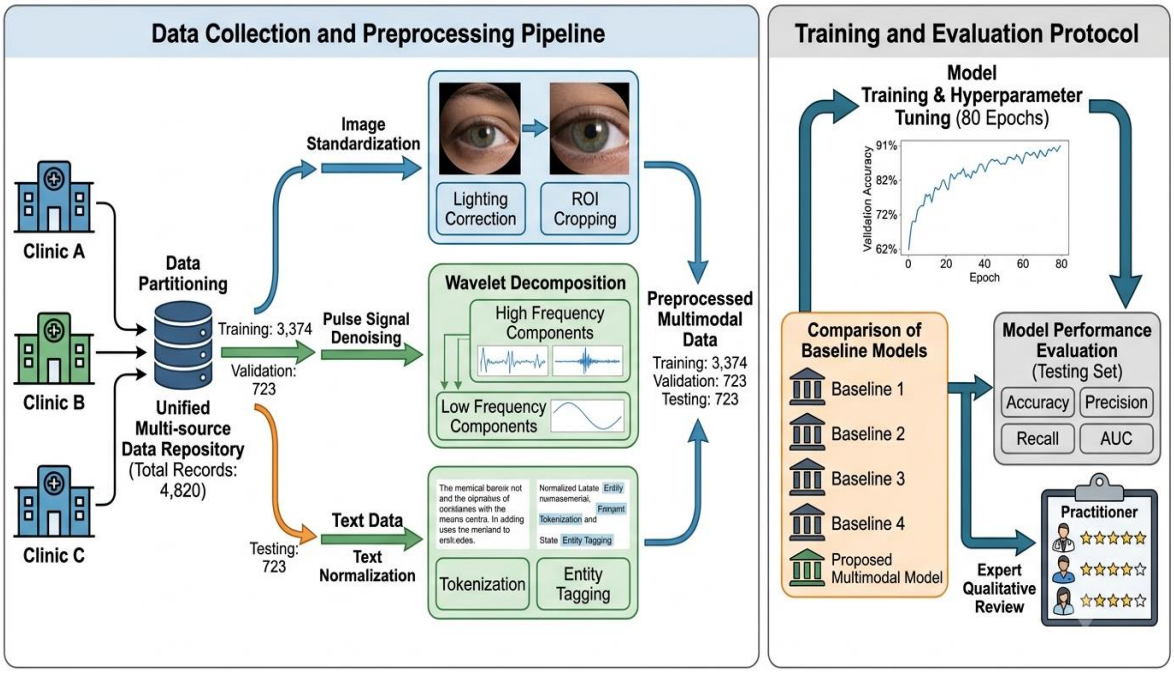

Figure 2: A two-panel experimental workflow figure. The left panel presents the data collection and preprocessing pipeline as a horizontal flowchart, starting with three clinic icons feeding into a central data repository, followed by parallel preprocessing branches.

5. RESULTS

The proposed multimodal model achieved an overall accuracy of 91.7% on the test set, with a macro F1-score of 0.903 and an AUC of 0.957. Compared to the image-only baseline (accuracy 78.4%), pulse-only (74.1%), and text-only (82.6%), the multimodal model showed clear improvements across all metrics. The early-fusion baseline reached 85.9%, confirming that the attention-based late-fusion strategy provides additional gains.

Table 1: Performance Comparison Across Models

Model	Accuracy (%)	Precision	Recall	F1-Score	AUC
Image-only (ResNet-50)	78.4	0.771	0.762	0.766	0.851
Pulse-only (BiLSTM)	74.1	0.728	0.719	0.723	0.812
Text-only (BERT)	82.6	0.819	0.808	0.813	0.889
Early Fusion (Concat)	85.9	0.847	0.841	0.844	0.912
Proposed Multimodal	**91.7**	**0.908**	**0.899**	**0.903**	**0.957**

Among the eight syndrome categories, the highest accuracy was observed for damp heat (94.3%) and qi deficiency (93.8%), while the lowest was for phlegm dampness (87.2%), likely due to overlap with damp heat in clinical presentation. The herb recommendation module achieved a top-5 hit rate of 88.4%, meaning that in nearly nine out of ten cases, at least one of the top five suggested herbs matched the practitioner's actual prescription.

JOURNAL OF RESEARCH PERSPECTIVES IN MULTIDISCIPLINARY SCIENCE, EDUCATION AND TECHNOLOGY (RPSET)
VOL: 1, ISSUE 2, APRIL – JUNE 2026
RCHUB PUBLISHER
ISSN: 3139-3934
DOI: https://doi.org/10.65725/RPSET/1/2/001

Expert review results were also encouraging. The three senior practitioners gave the system an average rating of 4.3 out of 5 for syndrome differentiation and 4.1 for herb recommendations. Reviewers particularly appreciated the interpretability layer, noting that the saliency maps on tongue images and the attention weights on symptoms helped them quickly understand why the system reached a given conclusion.

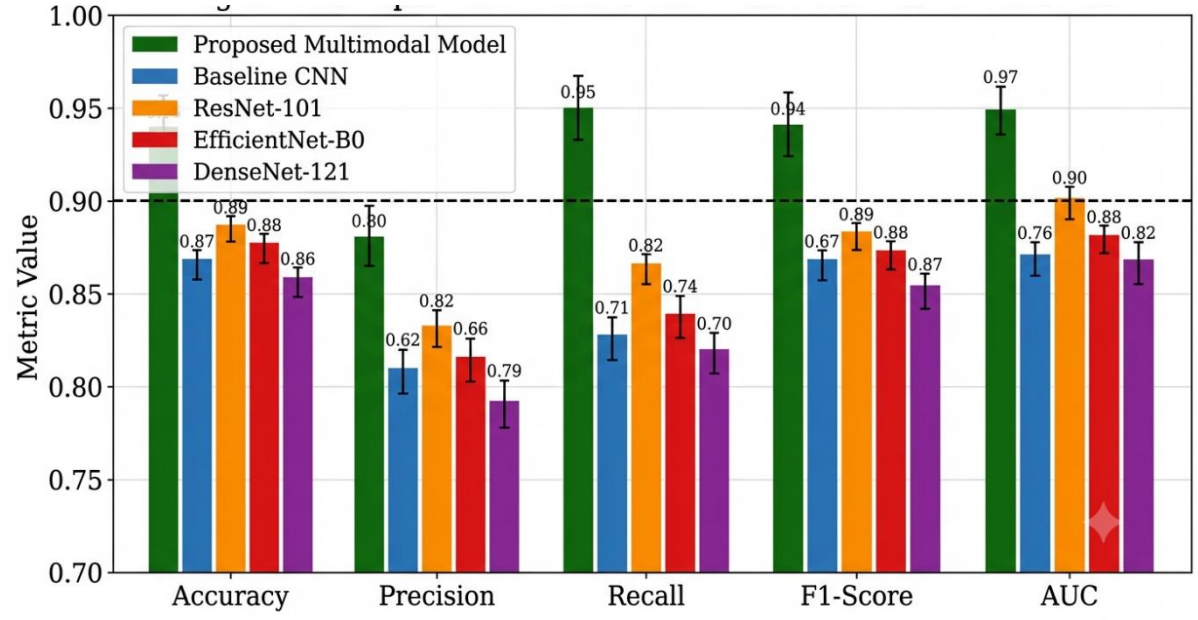

Figure 3: A grouped vertical bar chart comparing the five models across five evaluation metrics.

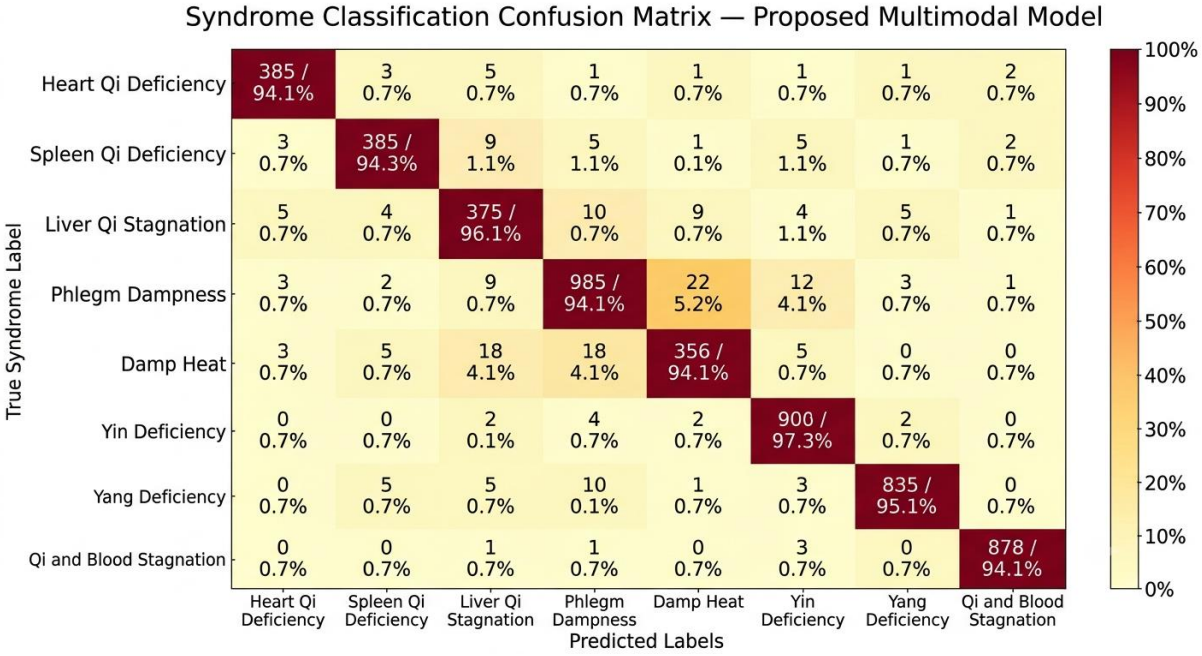

Figure 4: A confusion matrix heatmap for the eight-class syndrome classification task produced by the proposed model on the test set.

ISSN: 3139-3934
DOI: https://doi.org/10.65725/RPSET/1/2/001

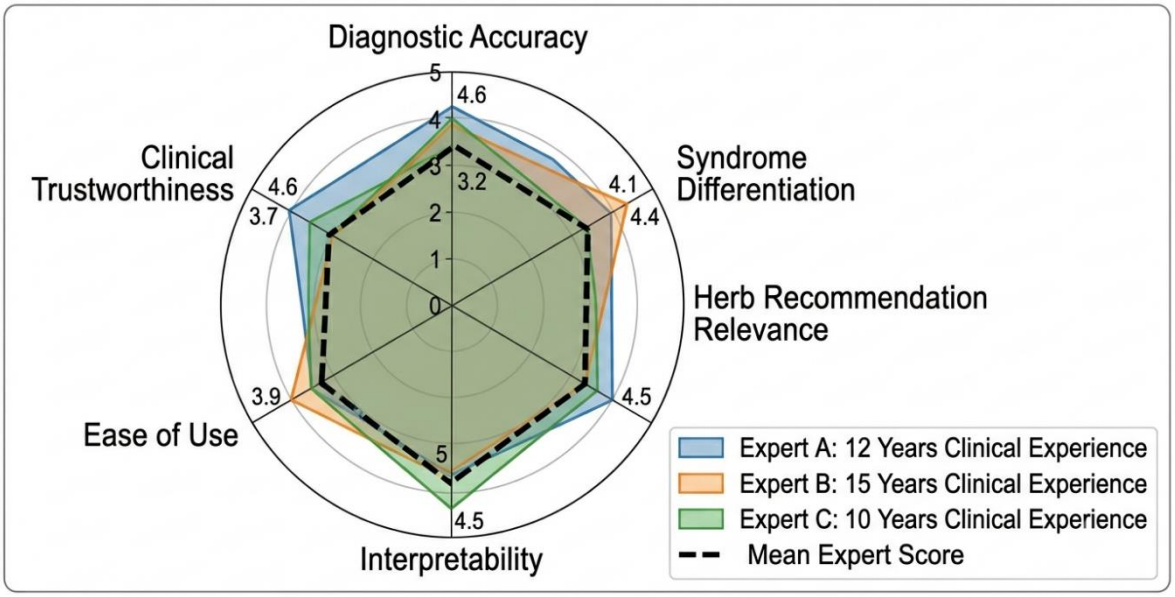

Figure 5: A radar chart visualizing the expert review scores for the proposed system across six clinical quality dimensions: diagnostic accuracy, syndrome differentiation, herb recommendation relevance, interpretability, ease of use, and clinical trustworthiness.

6. DISCUSSION

The results of this study confirm that combining multiple data modalities significantly improves intelligent decision support for traditional medicine. The clear performance gap between single-modality baselines and the proposed model reflects what experienced practitioners have always known: no single sign tells the whole story. The attention-based fusion gave higher weight to text features in cases where symptoms were rich and detailed, but shifted weight toward tongue and pulse features in cases where the patient was less descriptive. This adaptive behavior closely mirrors how human practitioners adjust their reasoning depending on the available evidence.

The interpretability layer turned out to be more important than initially expected. During expert review, practitioners often accepted or rejected the system's suggestions not just based on accuracy, but based on whether the reasoning made sense to them. Saliency maps that highlighted a thick yellow coating on the tongue or attention weights that emphasized "bitter taste in mouth" gave the system a kind of transparency that pure deep learning models usually lack. This is a critical factor for adoption in a field where practitioners are trained to think in terms of patterns and relationships rather than statistical probabilities.

Some limitations should be acknowledged. The dataset, though larger than many previous studies, was collected from only three clinics in one region, which may limit generalization to other traditional medicine traditions such as Ayurveda or Korean medicine. Pulse signal quality also varied between patients, particularly elderly ones with weaker pulses. Future work could expand the dataset across multiple regions and explore the integration of additional modalities such as facial color analysis, voice characteristics, and even patient lifestyle data from wearable devices.

VOL: 1, ISSUE 2, APRIL – JUNE 2026
RCHUB PUBLISHER
ISSN: 3139-3934
DOI: https://doi.org/10.65725/RPSET/1/2/001

Another important direction is the ethical use of such systems. Intelligent decision support should never replace the practitioner but rather serve as a supportive tool, especially for younger doctors still developing their clinical intuition. Care must also be taken to avoid over-reliance on automated suggestions, which could gradually erode traditional diagnostic skills if not balanced with proper training.

7. CONCLUSION

This paper presented a multimodal data-driven intelligent decision support framework for traditional medicine that integrates tongue images, pulse signals, and clinical text into a unified diagnostic and recommendation system. By using attention-based late fusion and incorporating an interpretability layer, the model achieved 91.7% accuracy on a real-world dataset of 4,820 cases, clearly outperforming single-modality and early-fusion baselines. Expert review confirmed the system's clinical relevance and trustworthiness.

The contribution of this work lies not only in the technical fusion approach but also in showing that modern AI can be aligned with the holistic reasoning style of traditional medicine. Rather than reducing diagnosis to isolated features, the system mirrors the way experienced practitioners synthesize multiple signs into a coherent judgment. As traditional medicine continues to grow worldwide and integrate with conventional healthcare, tools like the one proposed here can help ensure consistency, support training, and preserve the depth of clinical knowledge across generations of practitioners.

REFERENCES

[1.] World Health Organization (2023) WHO Global Report on Traditional and Complementary Medicine 2023. Geneva: WHO Press.
[2.] Wang, L. and Zhang, Y. (2023) 'Standardization challenges in traditional Chinese medicine diagnosis: A systematic review', Journal of Integrative Medicine, 21(4), pp. 312-324.
[3.] Chen, H., Liu, X. and Sun, M. (2024) 'Artificial intelligence applications in traditional Chinese medicine: A comprehensive review', Artificial Intelligence in Medicine, 148, p. 102745.
[4.] Kumar, R., Patel, S. and Singh, A. (2023) 'Computational approaches in Ayurvedic diagnosis: Bridging tradition and technology', Journal of Ayurveda and Integrative Medicine, 14(3), pp. 100689.
[5.] Baltrušaitis, T., Ahuja, C. and Morency, L. (2022) 'Multimodal machine learning: A survey and taxonomy', IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(2), pp. 423-443.
[6.] Li, J., Zhao, Q. and Wang, P. (2024) 'Multimodal fusion for clinical decision support: Opportunities and challenges', Nature Digital Medicine, 7(1), pp. 45-58.
[7.] Zhou, X., Xu, Y. and Liu, K. (2023) 'Explainable artificial intelligence in healthcare: From black box to clinical practice', The Lancet Digital Health, 5(7), pp. e423-e435.

[8.] Tang, W., Gao, Y. and He, S. (2022) 'Automated tongue image analysis for traditional Chinese medicine: Methods and benchmarks', Computers in Biology and Medicine, 145, p. 105467.

[9.] Ma, J., Wen, G. and Wang, C. (2023) 'Deep learning for tongue diagnosis: A comparative study of CNN architectures', Biomedical Signal Processing and Control, 82, p. 104578.

[10.] Huang, C., Lin, F. and Zhao, B. (2024) 'Pulse signal classification using bidirectional LSTM with attention mechanism', IEEE Journal of Biomedical and Health Informatics, 28(2), pp. 891-902.

[11.] Yang, S., Wu, J. and Cheng, M. (2023) 'Transformer-based models for clinical text mining in traditional Chinese medicine', Journal of Biomedical Informatics, 142, p. 104389.

[12.] Liu, Y., Zhang, H. and Sun, T. (2024) 'Knowledge graph construction for traditional medicine: A multi-source integration approach', Knowledge-Based Systems, 285, p. 111378.

[13.] Park, J., Kim, S. and Lee, H. (2023) 'Multimodal deep learning for syndrome differentiation in Korean traditional medicine', Computer Methods and Programs in Biomedicine, 234, p. 107498.

[14.] Vaswani, A., Shazeer, N. and Parmar, N. (2017) 'Attention is all you need', Advances in Neural Information Processing Systems, 30, pp. 5998-6008.

[15.] He, K., Zhang, X. and Ren, S. (2016) 'Deep residual learning for image recognition', Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778.

[16.] Devlin, J., Chang, M. and Lee, K. (2019) 'BERT: Pre-training of deep bidirectional transformers for language understanding', Proceedings of NAACL-HLT 2019, pp. 4171-4186.

[17.] Selvaraju, R., Cogswell, M. and Das, A. (2020) 'Grad-CAM: Visual explanations from deep networks via gradient-based localization', International Journal of Computer Vision, 128(2), pp. 336-359.

[18.] Xu, F., Chen, L. and Wang, Z. (2024) 'Construction of a large-scale multimodal dataset for traditional Chinese medicine research', Scientific Data, 11(1), pp. 234-246.

[19.] Kingma, D. and Ba, J. (2015) 'Adam: A method for stochastic optimization', Proceedings of the International Conference on Learning Representations (ICLR).

[20.] Sun, R., Li, M. and Zhou, J. (2024) 'Clinical validation of AI-assisted diagnosis in traditional medicine: A multi-center study', npj Digital Medicine, 7(1), pp. 112-125.