

A Research Report on Interpretable Knowledge-Based Reasoning for Traditional Chinese Medicine Diagnosis

Author: Tak Heung Sze

Affiliation: Independent Researcher

Document Type: Research Report

Companion Technical Report: *A Technical Report on the Design, Implementation, and Application of a Knowledge-Based Reasoning System for Traditional Chinese Medicine* (Sze, 2025)

Abstract

This research report examines an interpretable, knowledge-based reasoning paradigm for Traditional Chinese Medicine (TCM) diagnosis. Unlike data-driven or end-to-end predictive approaches, TCM diagnostic practice is inherently structured, multi-layered, and explanatory, involving relationships among symptoms, mechanisms, patterns, and therapeutic principles. This work formalizes TCM diagnosis as a reasoning problem rather than a classification task and analyzes a framework based on explicit knowledge representation and rule- and graph-based inference. The report focuses on the conceptual assumptions underlying this reasoning paradigm, its interpretability properties, and its suitability for educational, clinical reasoning, and decision-support contexts. The primary contribution lies in articulating a transparent and auditable approach to TCM diagnostic reasoning rather than optimizing predictive performance.

Keywords

Traditional Chinese Medicine; Knowledge Representation; Interpretability; Clinical Reasoning; Graph-Based Inference; Explainable Artificial Intelligence

1. Introduction

Traditional Chinese Medicine diagnosis is fundamentally a reasoning-intensive process. Clinical conclusions are derived through structured interpretation of observed symptoms and signs, guided by established conceptual frameworks such as pathogenic mechanisms, organ systems, and pattern differentiation. These processes emphasize explanation, comparison, and contextual judgment rather than direct prediction.

Recent computational approaches in medical artificial intelligence often frame diagnosis as a classification or regression problem. While effective in certain biomedical domains, such approaches are poorly aligned with the epistemic structure of TCM, where intermediate reasoning steps are clinically meaningful and must remain interpretable.

This research report argues that TCM diagnosis should be modeled as an explicit reasoning process rather than a black-box predictive task. It analyzes a knowledge-based reasoning framework designed to preserve interpretability, traceability, and conceptual coherence.

2. TCM Diagnosis as a Reasoning Problem

2.1 Multi-Layered Diagnostic Structure

TCM diagnostic reasoning typically involves multiple conceptual layers:

- Observable symptoms and signs
- Underlying pathogenic mechanisms
- Diagnostic patterns (證型)
- Therapeutic principles

These layers are interdependent, and reasoning proceeds through iterative association, comparison, and revision rather than linear deduction.

2.2 Diagnostic Plurality

A defining feature of TCM reasoning is the coexistence of multiple plausible diagnostic interpretations. Different patterns may explain overlapping symptom sets, and clinical judgment involves weighing and comparing alternatives rather than selecting a single ground-truth label.

This diagnostic plurality motivates a reasoning-oriented computational model rather than a purely predictive one.

3. Conceptual Limitations of Data-Driven Models

3.1 Opaqueness of End-to-End Prediction

End-to-end machine learning models typically compress diagnostic logic into latent representations. Although such models may achieve high predictive accuracy, they obscure intermediate reasoning steps and offer limited explanatory value.

In TCM contexts, where explanation and pedagogical clarity are central, this opacity represents a fundamental limitation.

3.2 Mismatch of Evaluation Criteria

Standard evaluation metrics such as accuracy or precision presuppose objective labels and unambiguous ground truth. In TCM diagnosis, however, multiple interpretations may be clinically acceptable depending on context, practitioner emphasis, and therapeutic intent.

This mismatch suggests that interpretability and reasoning transparency constitute more appropriate evaluation dimensions.

4. Knowledge Representation Assumptions

4.1 Explicit Conceptual Encoding

The framework assumes that core TCM concepts can be represented as explicit, discrete knowledge units rather than implicit numerical features. Symptoms, mechanisms, patterns, and therapeutic principles are treated as semantically meaningful entities.

4.2 Unified Knowledge Schema

All concepts are embedded within a unified schema that preserves traditional hierarchical and associative relationships. This structure supports systematic traversal and inspection of reasoning paths while maintaining conceptual fidelity.

5. Reasoning Framework

5.1 Rule-Based Associations

Reasoning is grounded in explicit associations derived from classical theory and established clinical conventions. These associations define how symptoms relate to mechanisms, how mechanisms suggest patterns, and how patterns imply therapeutic principles.

5.2 Graph-Based Traversal

The reasoning process is formalized as traversal through a directed knowledge graph. Multiple paths may be explored concurrently, and no single path is privileged as the definitive outcome. This conceptual formulation is instantiated and visualized in the NeoTCM reasoning system (Sze, 2025), as documented in a companion technical report.

5.3 Preservation of Reasoning Trace

At each inference step, the framework preserves the trace of associations used. This traceability enables inspection, comparison, and critique of diagnostic reasoning.

6. Interpretability and Clinical Reasoning

6.1 Interpretability as a Core Property

In this framework, interpretability is not an auxiliary feature but a foundational design objective. Each inference must be explainable in terms of explicit knowledge relationships recognizable to TCM practitioners.

6.2 Support for Comparative Reasoning

By presenting alternative reasoning paths, the framework supports comparative evaluation of diagnostic hypotheses. This reflects real-world TCM practice, where differential diagnosis and contextual judgment are central.

7. Scope and Limitations

This research framework does not claim to automate diagnosis or replace clinical expertise. It does not aim to optimize predictive accuracy or generate probabilistic confidence scores.

Its scope is limited to supporting structured reasoning, learning, and diagnostic exploration. Knowledge completeness and reasoning quality remain dependent on the underlying knowledge base and curation standards.

8. Discussion

8.1 Implications for Education and Clinical Training

The proposed reasoning paradigm has direct implications for TCM education and clinical training. By making diagnostic reasoning explicit and inspectable, the framework supports structured learning of pattern differentiation and mechanism-based thinking. Trainees can examine how specific symptoms lead to different diagnostic interpretations and compare alternative reasoning paths, aligning traditional case-based teaching approaches with computationally organized reasoning support.

8.2 Relationship to Cognitive Modeling and Educational Technology

From a broader perspective, the framework shares conceptual similarities with cognitive models of expert reasoning, in which knowledge is organized into structured representations and inference proceeds through explainable transitions. The explicit encoding of reasoning steps positions the framework as a potential bridge between TCM education and educational technology systems that emphasize transparency, reflection, and reasoning traceability.

8.3 Comparison with Western Clinical Decision-Support Systems

In contrast to many contemporary Western clinical decision-support systems (CDSS), which often emphasize probabilistic risk estimation and guideline-driven recommendations, the proposed framework prioritizes interpretability and the representation of diagnostic plurality. Rather than optimizing for a single recommended outcome, the system supports exploration and comparison of competing diagnostic hypotheses. This distinction highlights a complementary role for knowledge-based reasoning systems in contexts where explanation and clinical reasoning are emphasized alongside automated decision-support tools.

9. Relationship to Practical Implementation

The conceptual framework described in this report is instantiated in a separate technical implementation documented in a companion technical report. That document focuses on system design, implementation, and application case studies, while the present report concentrates on methodological assumptions and reasoning structure.

This separation preserves clarity between research analysis and system engineering.

10. Conclusion

This research report articulates an interpretable, knowledge-based reasoning paradigm for Traditional Chinese Medicine diagnosis. By reframing diagnosis as a reasoning process rather than a prediction task, the framework aligns computational modeling with the epistemic foundations of TCM.

The primary contribution of this work lies in formalizing an explicit and auditable approach to diagnostic reasoning that supports explanation, comparison, and educational use. This paradigm provides a foundation for future research on interpretable medical reasoning systems and for further methodological refinement.

References

Sze, T. H. (2025). *A Technical Report on the Design, Implementation, and Application of a Knowledge-Based Reasoning System for Traditional Chinese Medicine*.

End of Research Report