

HỆ THỐNG CHẨN ĐOÁN THÔNG MINH TRONG Y HỌC CỔ TRUYỀN: TỔNG QUAN PHẠM VI

Nguyễn Trường Nam¹, Lê Huyền Trang², Nguyễn Thị Như Quỳnh¹,
Hoàng Thúy Hồng¹, Nguyễn Thị Ngọc Hà³, Nguyễn Đức Nhân¹, Bùi Minh Đức¹,
Nguyễn Huy Nhật¹, Dương Hằng Song Thu¹, Mai Xuân Tráng¹

TÓM TẮT

Mục tiêu: Tổng hợp và đánh giá xu hướng ứng dụng trí tuệ nhân tạo (AI) trong khách quan hóa chẩn đoán Y học cổ truyền (YHCT) giai đoạn 2015–2025.

Phương pháp: Nghiên cứu tổng quan phạm vi được thực hiện trên các cơ sở dữ liệu PubMed và CNKI. Các nghiên cứu thực chứng sử dụng mô hình AI trong chẩn đoán YHCT được lựa chọn theo tiêu chuẩn xác định trước và phân tích định tính kết hợp định lượng mô tả.

Kết quả: Tổng cộng 25 nghiên cứu được phân tích. AI chủ yếu ứng dụng trong các lĩnh vực đơn phương diện, nổi bật là thiết chẩn (16%). Học sâu, đặc biệt CNN và RNN, là phương pháp chủ đạo; trong khi mô hình đa phương thức (4%), AI giải thích được (12%) và mô hình ngôn ngữ lớn (8%) còn ở giai đoạn sớm. Tích hợp dữ liệu Đông – Tây y chiếm 20%, cho thấy tiềm năng liên kết chỉ số sinh học với hội chứng YHCT.

Kết luận: AI có tiềm năng hỗ trợ định lượng hóa và chuẩn hóa chẩn đoán YHCT. Tuy nhiên, bằng chứng còn hạn chế do cỡ mẫu nhỏ, dữ liệu chưa chuẩn hóa và thiếu kiểm chứng lâm sàng. Cần nghiên cứu đa phương thức quy mô lớn để tăng khả năng ứng dụng thực tiễn.

Từ khóa: Trí tuệ nhân tạo; Y học cổ truyền; Khách quan hóa chẩn đoán; Biện chứng luận trị; Học sâu; Mô hình ngôn ngữ lớn.

INTELLIGENT DIAGNOSTIC METHOD SYSTEM IN TRADITIONAL MEDICINE: A SCOPING REVIEW

ABSTRACT

Objective: To synthesize and evaluate trends in the application of artificial intelligence (AI) for the objectification of Traditional Medicine diagnosis during the period 2015–2025.

Methods: A scoping review was conducted using the PubMed and CNKI databases. Empirical studies applying AI models in Traditional Medicine diagnosis were selected according to predefined criteria and analyzed using qualitative synthesis combined with descriptive quantitative analysis.

Results: A total of 25 studies were included in the analysis. AI was mainly applied to single-modality diagnostic domains, with tongue diagnosis being the most prominent application (16%). Deep learning, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), was the predominant approach, whereas multimodal models (4%), explainable AI (12%), and large language models (8%) remained at an early stage of development. The integration of Eastern and Western medical data accounted for 20%, indicating the potential to link biological indicators with Traditional Medicine syndromes.

1. Đại học Phenikaa
2. Phòng khám Y học cổ truyền Việt Nhật
3. Học viện Y Dược học cổ truyền Việt Nam
Tác giả liên hệ: Nguyễn Trường Nam
Email: trang.xuanmai@phenikaa-uni.edu.vn
Ngày nhận bài : 6.5.2026
Ngày được chấp nhận: 20.5.2026



Conclusion: AI shows potential in supporting the quantification and standardization of Traditional Medicine diagnosis. However, current evidence remains limited due to small sample sizes, insufficient data standardization, and a lack of clinical validation. Large-scale multimodal studies are needed to enhance practical applicability.

Keywords: Artificial intelligence; Traditional Medicine; Diagnostic objectification; Syndrome differentiation and treatment; Deep learning; Large language models.

I. ĐẶT VẤN ĐỀ

Y học cổ truyền (YHCT) vận hành dựa trên nguyên lý “biện chứng luận trị”, trong đó quá trình thu thập dữ liệu lâm sàng thông qua Tứ chẩn (vọng, văn, vấn, thiết) đóng vai trò quyết định. Tuy nhiên, phương thức chẩn đoán truyền thống này bộc lộ hạn chế lớn về tính chủ quan, phụ thuộc chặt chẽ vào kinh nghiệm cá nhân và thiếu các tiêu chuẩn định lượng thống nhất [1]. Sự thiếu hụt các chỉ số khách quan không chỉ gây khó khăn trong việc chuẩn hóa kết quả điều trị mà còn là rào cản chính trong nỗ lực hội nhập của YHCT vào hệ thống y học hiện đại dựa trên bằng chứng.

Trong thập kỷ qua (2015–2025), sự phát triển của Trí tuệ nhân tạo (AI) đã mở ra hướng đi mới trong việc khách quan hóa dữ liệu YHCT. Các mô hình Học máy (Machine Learning) và Học sâu (Deep Learning) như Mạng thần kinh tích chập (CNN), Mạng thần kinh tái phát (RNN) và Đồ thị tri thức (Knowledge Graph) đã được ứng dụng rộng rãi để phân tích hình ảnh lưỡi, sắc mặt, tín hiệu mạch tượng và mô phỏng tư duy biện chứng. Đặc biệt, sự ra đời của các Mô hình ngôn ngữ lớn (LLM) gần đây đã nâng cao khả năng xử lý ngôn ngữ tự nhiên và suy luận lâm sàng phức tạp [2][3].

Mặc dù đạt được những bước tiến đáng kể, việc ứng dụng AI trong chẩn đoán YHCT vẫn tồn tại nhiều thách thức kỹ thuật. Do đó,

nghiên cứu tổng quan phạm vi này được thực hiện nhằm hệ thống hóa các ứng dụng của các mô hình AI tiên tiến trong khách quan hóa chẩn đoán YHCT giai đoạn 2015–2025, từ đó tạo nền tảng định hướng cho việc phát triển các công cụ hỗ trợ chẩn đoán chính xác và khách quan hơn trong tương lai.

II. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Đối tượng và thời gian nghiên cứu

Tiêu chuẩn lựa chọn:

Các bài báo có nội dung trí tuệ nhân tạo ứng dụng trong chẩn đoán y học cổ truyền.

Bài báo toàn văn, viết bằng tiếng Anh hoặc tiếng Trung, công bố trên cơ sở dữ liệu học thuật như PubMed, CNKI phù hợp với mục tiêu tổng quan phạm vi (scoping review). Các bài báo được xuất bản trong thời gian từ tháng 01/2015 đến tháng 12/2025.

Tiêu chuẩn loại trừ:

Các bài báo trùng lặp; các bài báo không liên quan trực tiếp ứng dụng của trí tuệ nhân tạo trong chẩn đoán y học cổ truyền; không toàn văn, không đủ thông tin trích dẫn.

Các bài báo sử dụng thuật toán khai phá dữ liệu (Data Mining) như luật kết hợp (Association Rules), thuật toán Apriori, hoặc phân cụm (Cluster Analysis)

2.2. Phương pháp và chiến lược tìm kiếm

Nghiên cứu được thực hiện theo phương pháp nghiên cứu tổng quan phạm vi, tuân thủ hướng dẫn PRISMA-ScR. Việc tìm kiếm tài liệu được tiến hành vào tháng 4/2026 trên các cơ sở dữ liệu bao gồm PubMed, CNKI.

Đối với PubMed, các tổ hợp từ khóa bằng tiếng Anh gồm: "Artificial Intelligence" AND "TCM" AND "diagnosis"

Đối với CNKI, các tổ hợp từ khóa bằng tiếng Trung gồm: “(人工智能 + AI + 机器学习 + 深度学习) * (中医诊断)” .

Các từ khóa được tìm kiếm trong trường tiêu đề và tóm tắt, giới hạn ngôn ngữ tiếng Anh và tiếng Trung.

Các tài liệu thu thập được nhập vào phần mềm quản lý Zotero để loại bỏ trùng lặp trước khi tiến hành sàng lọc. Các nghiên cứu viên độc lập sàng lọc theo tiêu đề, tóm tắt và toàn văn theo tiêu chuẩn lựa chọn và loại trừ. Trường hợp có sự mâu thuẫn giữa các nghiên cứu viên trong nhóm nghiên cứu, cả nhóm sẽ cùng nhau thảo luận và thống nhất quyết định lựa chọn.

2.4. Trích xuất dữ liệu

Dữ liệu được trích xuất theo mẫu thống nhất bao gồm

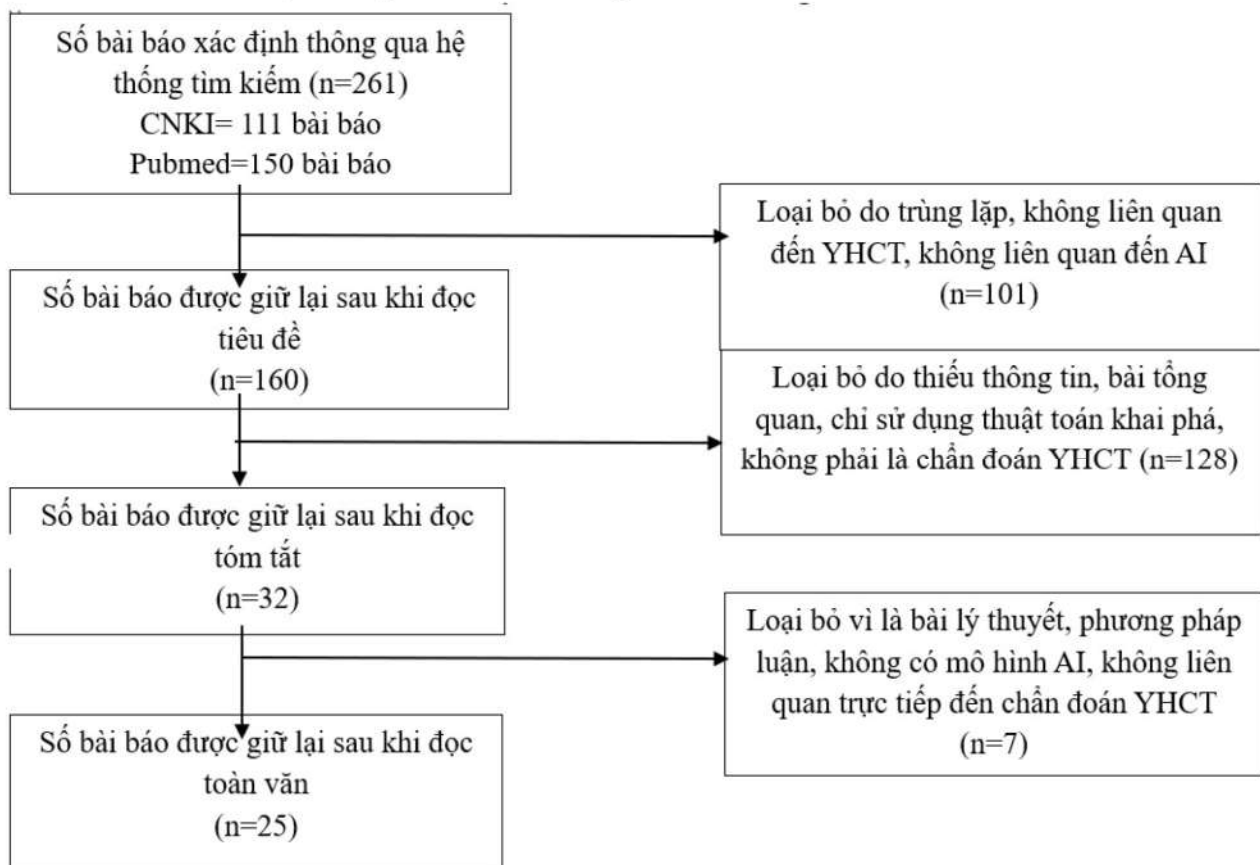
- Thông tin bài báo: tên bài báo, tác giả, năm công bố, địa điểm nghiên cứu, tạp chí
- Thông tin nghiên cứu: đối tượng nghiên cứu, thiết kế nghiên cứu, phương pháp nghiên cứu, kết quả nghiên cứu.

III. KẾT QUẢ VÀ BÀN LUẬN

1. Kết quả nghiên cứu

Trong tổng số 250 tiêu đề và tóm tắt bài báo được tìm thấy trên hệ thống Pubmed và CNKI, chúng tôi lọc ra được 32 tiêu đề và tóm tắt phù hợp để đưa vào đọc toàn văn. Sau khi đọc toàn văn, 25 bài báo đáp ứng tiêu chuẩn lựa chọn và tiêu chuẩn loại trừ được đưa vào nghiên cứu (sơ đồ 1).

Sơ đồ 1. Sơ đồ tuyển chọn bài báo vào nghiên cứu



**Bảng 3.1. Tóm tắt kết quả nghiên cứu theo năm**

STT	Tên tác giả	Năm	Đối tượng nghiên cứu	Thiết kế nghiên cứu	Phương pháp nghiên cứu	Kết quả chính
1	Xiaojie Jin et al.	2025	Hội chứng Hàn – Nhiệt trong viêm phổi virus	Phát triển, xác nhận mô hình	GBM, LightGBM, RF, LR, SVM	GBM tốt nhất: AUC 0,8329; độ chính xác 0,8000
2	Hulin Long et al.	2025	Biện chứng đột quy theo YHCT	Đánh giá định lượng	LLM: GPT-4o, DeepSeek-R1	LLM chuyên biệt vượt trội hơn LLM thông thường
3	Xionghui Wang et al.	2025	Gan nhiễm mỡ liên quan chuyển hóa	So sánh, đánh giá hiệu suất	ChatGPT-4.0, ChatGLM + Knowledge Base	ChatGLM4 + KB có tỷ lệ đồng thuận cao
4	Luo Guancong et al.	2025	Phân loại sắc mặt trong vọng chẩn	Phát triển, tối ưu thuật toán	Meta-learning kết hợp CNN	Đạt độ chính xác cao trên tập dữ liệu nhỏ
5	Chen Lingxuan et al.	2025	Nhĩ chẩn	Thiết kế, ứng dụng hệ thống	AI xử lý hình ảnh và ngôn ngữ	Xây dựng hệ thống nhĩ chẩn thông minh
6	Jiang Qiyu et al.	2025	6 hội chứng bệnh huyễn vựng	Phát triển mô hình suy luận	BERT kết hợp thuật toán suy luận	Độ chính xác trung bình 0,990
7	Zhang Pengfei et al.	2025	Tứ chẩn hợp tham	Phát triển khung tích hợp	Học sâu đa phương thức	Đề xuất khung tích hợp dữ liệu đa phương thức
8	Ziming Yin et al.	2024	Biện chứng 5 loại ù tai	Phát triển, xác nhận mô hình	Đồ thị tri thức + k-NN	Độ chính xác 0,994; tăng khả năng giải thích
9	Wang Fei et al.	2024	Nhận diện bất thường vùng lưới	Thiết kế thuật toán	YOLOv10 cải tiến	Phát hiện tự động vùng bệnh lý lưới
10	J Liu et al.	2023	Đàm ú tương kết ở rối loạn mỡ máu	Phát triển mô hình dự đoán	Mạng thần kinh học sâu	Hiệu suất dự đoán tốt, AUC cao
11	Li Renliang et al.	2023	Hội chứng xương khớp / Tý chứng	Mô hình hóa chẩn đoán	CART, mạng BP	Mạng BP đạt độ chính xác >83%
12	Du Fangzhen et al.	2023	28 hội chứng tạng phủ	So sánh, phát triển thuật toán	ML-kNN, FCN, CNN 1D	FCN đạt độ chính xác 84,48%
13	Yu Xiaohan et al.	2023	Hội chứng NAFLD	Phân tích đặc trưng hình ảnh	U-Net cải tiến + cây quyết định	Hỗ trợ chẩn đoán hội chứng qua ảnh đáy mắt
14	Zonghai Huang et al.	2022	Hội chứng đau bụng kinh	Phát triển, xác nhận mô hình	Cross-FGCNN	Độ chính xác và F1-score đạt 0,9621
15	Yue Zhao, Yuandi Huang	2022	Tắc tĩnh mạch võng mạc	Phát triển thuật toán định lượng	ART kết hợp CNN	Tỷ lệ nhận dạng tăng 13,09%

16	Minghuan Li et al.	2022	Biện chứng qua vấn chẩn	Phát triển, đánh giá mô hình	Mô hình biện chứng thông minh	Số câu hỏi trung bình 4,6–6,4 câu
17	Heng Weng et al.	2022	Chẩn đoán đa bệnh/hội chứng	Phát triển framework	Đồ thị tri thức, CoKE, BiLSTM-KGE	Xây dựng 59.882 thực thể
18	Meng-Yi Li et al.	2021	Lưới nút	Phát triển thuật toán	U-Net cải tiến + Global Convolution	Phân đoạn, nhận diện vết nứt hiệu quả
19	Hong Zhang et al.	2020	187 bệnh và 466 hội chứng YHCT	Xác nhận trên dữ liệu hỏi cứu	RNN, CNN, học tích hợp	Top-1: 80,5%; Top-3: 91,6%; Top-5: 94,2%
20	Ziqing Liu et al.	2020	Hội chứng ung thư phổi	Phát triển, xác nhận mô hình	BERT, Text-CNN/RNN/HAN	AUC 0,9618; F1-score 0,8884
21	Shujie Xia et al.	2020	Hội chứng chuyển hóa	Phát triển mô hình	ML-kNN + chỉ số hóa lý	Độ chính xác trung bình 0,714
22	Xu Wang et al.	2020	Lưới có dấu răng / Tỷ hư	Phát triển thuật toán	Deep CNN – ResNet34	Phân loại lưới có dấu răng hiệu quả
23	Dong Jingfang	2020	Đặc trưng lưới ở bệnh nhân ung thư	Ứng dụng thuật toán lâm sàng	CNN – ResNet50	Nhận diện tổn thương với độ chính xác, AUC cao
24	An Weitao et al.	2019	8 hội chứng Bát cương	Phát triển thuật toán	DBN dựa trên RBM	Xử lý quan hệ một-nhiều giữa triệu chứng và hội chứng
25	Wang Yiqun et al.	2015	5 loại mạch tượng lâm sàng	Thiết kế, thử nghiệm hệ thống	Mạng thần kinh BP 4 lớp	Nhận diện mạch tượng đạt >90%

2. Bàn luận

Sự phát triển của trí tuệ nhân tạo (AI) trong chẩn đoán YHCT giai đoạn 2015–2025 cho thấy một chuyển dịch đáng chú ý về mặt công nghệ. Trong giai đoạn đầu, các nghiên cứu chủ yếu ứng dụng học máy truyền thống nhằm giải quyết các bài toán phân loại hội chứng đơn lẻ. Chẳng hạn, Du Fangzhen (2023) đã sử dụng thuật toán phân loại đa nhãn để chẩn đoán hội chứng [4]. Sau đó, các nghiên cứu nhanh chóng chuyển sang học sâu (Deep Learning) nhằm tự động hóa quá trình trích xuất đặc trưng từ dữ liệu hình ảnh và văn bản. Đáng chú ý, giai đoạn 2022–2025 ghi nhận sự xuất hiện của các phương pháp như Đồ thị tri thức (Knowledge Graphs) và Mô hình ngôn ngữ lớn (LLMs), mặc dù hiện mới chỉ chiếm tỷ lệ hạn chế 8%. Các phương pháp này bước đầu cho thấy khả năng tiệm cận việc mô phỏng một số khía cạnh của quá trình biện chứng luận trị, đặc biệt trong xử lý suy luận đa chiều.

Song song với sự tiến hóa của thuật toán, phạm vi ứng dụng chẩn đoán cũng mở rộng từ đơn phương diện sang đa phương thức. Phân tích dữ liệu cho thấy các nghiên cứu hiện vẫn chủ yếu tập trung vào việc số hóa các khía cạnh riêng lẻ của Tứ chẩn, trong đó thiết chẩn chiếm tỷ trọng cao nhất với 16%. Các nghiên cứu tiêu biểu như Wang Fei (2024) với mô hình YOLOv10 cải tiến [5] hay Li MY (2021) với U-Net tích hợp Global Convolution cho thấy hiệu quả trong việc phân đoạn và nhận diện đặc trưng hình ảnh lưới [6]. Các phương diện khác như sắc mặt xuất hiện với tần suất thấp hơn, mặc dù một số phương pháp như meta-learning (Luo Guancong, 2025 [7]) đã được đề xuất nhằm



khắc phục hạn chế về dữ liệu. Sự ưu thế của thiết chẩn có thể được lý giải bởi tính trực quan và khả năng chuẩn hóa cao của dữ liệu hình ảnh, tạo điều kiện thuận lợi cho các mô hình học sâu. Ngược lại, các dữ liệu như mạch tượng hoặc vấn chẩn có độ biến thiên lớn và khó chuẩn hóa hơn.

Quá trình số hóa này không chỉ hỗ trợ thu thập dữ liệu mà còn góp phần thúc đẩy chuyên dịch từ tiếp cận định tính sang định lượng trong chẩn đoán YHCT. Tuy nhiên, một trong những rào cản chính của các mô hình học sâu là tính “hộp đen”, hạn chế khả năng giải thích. Trong bối cảnh đó, các phương pháp AI có khả năng giải thích (XAI) đang được quan tâm hơn, dù hiện chỉ chiếm 3/25 nghiên cứu. Một số nghiên cứu như Jiang Qiyu (2025) hay Yin Z (2024) cho thấy việc kết hợp mô hình học sâu với các cơ chế suy luận có thể tạo ra các biểu diễn có khả năng diễn giải, góp phần nâng cao tính minh bạch.

Bên cạnh đó, xu hướng tích hợp giữa YHCT và y học hiện đại cũng ngày càng rõ nét. Các nghiên cứu theo hướng này chiếm khoảng 20% và sử dụng các dữ liệu cận lâm sàng như xét nghiệm sinh hóa hoặc hình ảnh y khoa để hỗ trợ phân loại hội chứng. Như Xia S (2020) [8] đã kết hợp thuật toán ML-kNN với các chỉ số hóa lý để phân tích hội chứng chuyên hóa, trong khi Jin X (2025) [9] sử dụng mô hình Gradient Boosting Machine để phân biệt hội chứng Hàn – Nhiệt trong viêm phổi do virus. Cách tiếp cận này bước đầu cung cấp bằng chứng định lượng hỗ trợ việc liên kết giữa các chỉ số sinh học và hội chứng YHCT, góp phần thu hẹp khoảng cách giữa hai hệ thống y học. Tuy nhiên, phần lớn các nghiên cứu vẫn mang tính hồi cứu và chưa đủ cơ sở để khẳng định khả năng ứng dụng rộng rãi.

IV. KẾT LUẬN

AI ngày càng được ứng dụng trong chẩn đoán YHCT, với xu hướng chuyển từ học máy truyền thống sang học sâu và mô hình tích hợp, góp phần định lượng hóa chẩn đoán, đặc biệt

trong thiết chẩn. Tuy nhiên, bằng chứng hiện còn hạn chế do đa số nghiên cứu có quy mô nhỏ, dữ liệu chưa chuẩn hóa, tính giải thích và khả năng khái quát hóa chưa cao. Do đó, cần phát triển dữ liệu chuẩn hóa quy mô lớn và tăng cường kiểm chứng lâm sàng để nâng cao khả năng ứng dụng AI trong thực hành YHCT.

TÀI LIỆU THAM KHẢO

1. Jiang M, Lu C, Zhang C, Yang J, Tan Y, Lu A, Chan K. Syndrome differentiation in modern research of traditional Chinese medicine. *Journal of Ethnopharmacology*. 2012;140:634-642. doi: 10.1016/j.jep.2012.01.033.
2. Pan D, Guo Y, Fan Y, Wan H. Development and Application of Traditional Chinese Medicine Using AI Machine Learning and Deep Learning Strategies. *Am J Chin Med*. 2024;52(3):605-623. doi: 10.1142/S0192415X24500265.
3. Ren Y, Luo X, Wang Y, Li H, Zhang H, Li Z, et al. Large Language Models in Traditional Chinese Medicine: A Scoping Review. *J Evid Based Med*. 2025;18(1):e12658. doi: 10.1111/jebm.12658.
4. 杜昉臻, 何圆姣, 冯西贝, 刘国华. 基于人工智能的中医证候分类算法研究. *南开大学学报(自然科学版)*. 2023;56(2):12-16.
5. 王菲, 曲强, 刘明阳, 等. 基于改进YOLOv10的舌象检测辨识系统. 2024.
6. Li MY, Zhu DJ, Xu W, Lin YJ, Yung KL, Ip AWH. Application of U-Net with Global Convolution Network Module in Computer-Aided Tongue Diagnosis. *J Healthc Eng*. 2021;2021:5853128. doi: 10.1155/2021/5853128.
7. 罗冠聪, 冯跃, 徐红, 秦传波, 李福凤, 钱鹏, 等. 元学习优化的面部色诊分类方法. *计算机与现代化*. 2025;(12):115-122.
8. Xia S, Zhang J, Du G, Li S, Vong CT, Yang Z, et al. A Microcosmic Syndrome Differentiation Model for Metabolic Syndrome with Multilabel Learning. *Evid Based Complement Alternat Med*. 2020;2020:9081641. doi: 10.1155/2020/9081641.
9. Jin X, Wang Y, Wang J, Gao Q, Huang Y, Shao L, et al. A Machine Learning Approach to Differentiate Cold and Hot Syndrome in Viral Pneumonia Integrating Traditional Chinese Medicine and Modern Medicine: Machine Learning Model Development and Validation. *JMIR Med Inform*. 2025;13:e64725. doi: 10.2196/64725.

INTELLIGENT DIAGNOSTIC METHOD SYSTEM IN TRADITIONAL MEDICINE: A SCOPING REVIEW

Nguyễn Trường Nam¹, Lê Huyền Trang², Nguyễn Thị Như Quỳnh¹,
Hoàng Thúy Hồng¹, Nguyễn Thị Ngọc Hà³, Nguyễn Đức Nhân¹, Bùi Minh Đức¹,
Nguyễn Huy Nhật¹, Dương Hằng Song Thu¹, Mai Xuân Tráng¹

ABSTRACT

Objective: To synthesize and evaluate trends in the application of artificial intelligence (AI) for the objectification of Traditional Medicine diagnosis during the period 2015–2025.

Methods: A scoping review was conducted using the PubMed and CNKI databases. Empirical studies applying AI models in Traditional Medicine diagnosis were selected according to predefined criteria and analyzed using qualitative synthesis combined with descriptive quantitative analysis.

Results: A total of 25 studies were included in the analysis. AI was mainly applied to single-modality diagnostic domains, with tongue diagnosis being the most prominent application (16%). Deep learning, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), was the predominant approach, whereas multimodal models (4%), explainable AI (12%), and large language models (8%) remained at an early stage of development. The integration of Eastern and Western medical data accounted for 20%, indicating the potential to link biological indicators with Traditional Medicine syndromes.

Conclusion: AI shows potential in supporting the quantification and standardization of Traditional Medicine diagnosis. However, current evidence remains limited due to small sample sizes, insufficient data standardization, and a lack of clinical validation. Large-scale multimodal studies are needed to enhance practical applicability.

Keywords: Artificial intelligence; Traditional Medicine; Diagnostic objectification; Syndrome differentiation and treatment; Deep learning; Large language models.

I. INTRODUCTION

Traditional Chinese Medicine (TCM) operates on the principle of “pattern differentiation and treatment,” in which the process of collecting clinical data through the Four Examinations - inspection, listening and smelling, inquiry, and palpation - plays a decisive role. However, this traditional diagnostic approach has major limitations in terms of subjectivity, strong dependence on individual clinical experience, and the lack of unified quantitative standards [1]. The absence of objective indicators not only creates difficulties in standardizing treatment outcomes but also represents a major barrier to the integration of TCM into the modern evidence-based medical system.

Over the past decade (2015–2025), the development of Artificial Intelligence (AI) has opened up new directions for objectifying diagnostic data in TCM. Machine Learning and Deep Learning models, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Knowledge Graphs, have been widely applied to analyze tongue images, facial complexion, pulse signals, and to simulate TCM pattern differentiation reasoning. In particular, the recent emergence of Large Language Models (LLMs) has enhanced the ability to process natural language and perform complex clinical reasoning [2][3].

Despite considerable progress, the application of AI in TCM diagnosis still faces various



technical challenges. Therefore, this scoping review was conducted to systematize the applications of advanced AI models in objectifying TCM diagnosis during the period 2015 - 2025, thereby providing a foundation for the future development of more accurate and objective diagnostic support tools.

II. SUBJECTS AND METHODS

2.1. Study subjects and study period

Inclusion criteria:

Articles focusing on the application of artificial intelligence in TCM diagnosis.

Full-text articles written in English or Chinese and published in academic databases such as PubMed and CNKI, in accordance with the objectives of this scoping review. Articles published from January 2015 to December 2025 were included.

Exclusion criteria:

Duplicate articles; articles not directly related to the application of artificial intelligence in TCM diagnosis; articles without full text or with insufficient citation information.

Articles using data mining algorithms such as Association Rules, the Apriori algorithm, or Cluster Analysis.

2.2. Methods and search strategy

This study was conducted as a scoping review in accordance with the PRISMA-ScR guidelines. The literature search was performed in April 2026 using databases including PubMed and CNKI.

For PubMed, the English keyword combination was: “Artificial Intelligence” AND “TCM” AND “diagnosis”

For CNKI, the Chinese keyword combination was: “(人工智能 + AI + 机器学习 + 深度学习) * (中医诊断)”

Keywords were searched in the title and abstract fields, with language restrictions limited to English and Chinese.

The retrieved documents were imported into Zotero reference management software to remove duplicates before screening. Independent researchers screened the titles, abstracts, and full texts according to the inclusion and exclusion criteria. In cases of disagreement among researchers, the research team discussed and reached a consensus on the final selection.

2.4. Data extraction

Data were extracted using a standardized form, including:

- Article information: title, authors, year of publication, study location, and journal;
- Study information: study subjects, study design, research methods, and study results.

III. RESULTS AND DISCUSSION

1. Study results

Among a total of 250 article titles and abstracts identified from PubMed and CNKI, 32 titles and abstracts were screened as potentially eligible for full-text review. After full-text assessment, 25 articles met the inclusion and exclusion criteria and were included in the study (Figure 1).

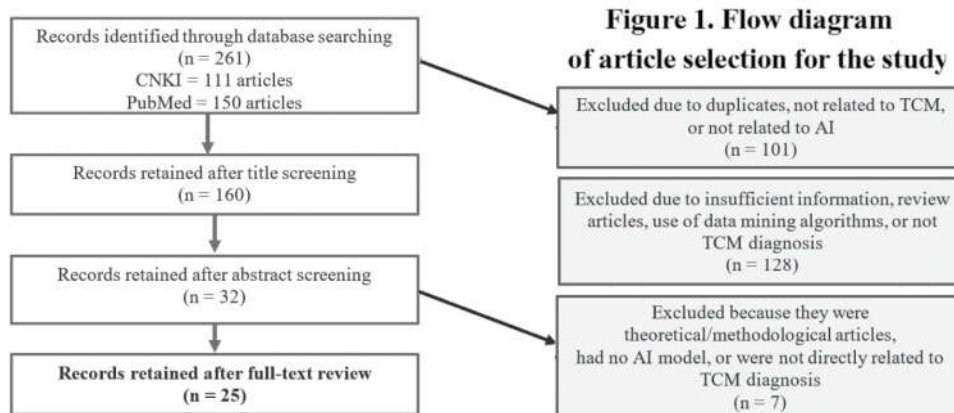


Table 3.1. Summary of study results by year

No.	Author	Year	Study subject	Study design	Research method	Main results
1	Xiaojie Jin et al.	2025	Cold - Heat syndrome in viral pneumonia	Model development and validation	GBM, LightGBM, RF, LR, SVM	GBM performed best: AUC 0.8329; accuracy 0.8000
2	Hulin Long et al.	2025	TCM pattern differentiation in stroke	Quantitative evaluation	LLMs: GPT - 4o, DeepSeek-R1	Specialized LLMs outperformed general - purpose LLMs
3	Xionghui Wang et al.	2025	Metabolic dysfunction - associated fatty liver disease	Comparative performance evaluation	ChatGPT - 4.0, ChatGLM + Knowledge Base	ChatGLM4 + KB achieved a high agreement rate
4	Luo Guancong et al.	2025	Facial complexion classification in inspection diagnosis	Algorithm development and optimization	Meta - learning combined with CNN	Achieved high accuracy on small datasets
5	Chen Lingxuan et al.	2025	Auricular diagnosis	System design and application	AI - based image and language processing	Developed an intelligent auricular diagnosis system
6	Jiang Qiyu et al.	2025	Six TCM patterns of vertigo	Reasoning model development	BERT combined with reasoning algorithms	Mean accuracy reached 0.990
7	Zhang Pengfei et al.	2025	Integrated Four Examinations	Integrated framework development	Multimodal deep learning	Proposed a framework for multimodal data integration
8	Ziming Yin et al.	2024	Pattern differentiation of five types of tinnitus	Model development and validation	Knowledge Graph + k-NN	Accuracy 0.994; improved interpretability
9	Wang Fei et al.	2024	Detection of abnormal tongue regions	Algorithm design	Improved YOLOv10	Automatically detected pathological tongue regions
10	J Liu et al.	2023	Phlegm-stasis binding pattern in dyslipidemia	Prediction model development	Deep neural network	Good predictive performance with high AUC
11	Li Renliang et al.	2023	Musculoskeletal syndromes / Bi syndrome	Diagnostic modeling	CART, BP neural network	BP neural network achieved accuracy >83%
12	Du Fangzhen et al.	2023	28 zang - fu organ patterns	Algorithm comparison and development	ML-kNN, FCN, 1D-CNN	FCN achieved an accuracy of 84.48%
13	Yu Xiaohan et al.	2023	NAFLD patterns	Image feature analysis	Improved U-Net + decision tree	Supported pattern diagnosis using fundus images
14	Zonghai Huang et al.	2022	Dysmenorrhea patterns	Model development and validation	Cross - FGCNN	Accuracy and F1 - score reached 0.9621
15	Yue Zhao, Yuandi Huang	2022	Retinal vein occlusion	Quantitative algorithm development	ART combined with CNN	Recognition rate increased by 13.09%



16	Minghuan Li et al.	2022	Pattern differentiation through inquiry	Model development and evaluation	Intelligent pattern differentiation model	Average number of questions ranged from 4.6 to 6.4
17	Heng Weng et al.	2022	Diagnosis of multiple diseases/patterns	Framework development	Knowledge Graph, CoKE, BiLSTM-KGE	Constructed 59,882 entities
18	Meng-Yi Li et al.	2021	Fissured tongue	Algorithm development	Improved U-Net + Global Convolution	Effective segmentation and recognition of tongue fissures
19	Hong Zhang et al.	2020	187 diseases and 466 TCM patterns	Validation using retrospective data	RNN, CNN, ensemble learning	Top - 1: 80.5%; Top - 3: 91.6%; Top - 5: 94.2%
20	Ziqing Liu et al.	2020	Lung cancer patterns	Model development and validation	BERT, Text-CNN/RNN/HAN	AUC 0.9618; F1 - score 0.8884
21	Shujie Xia et al.	2020	Metabolic syndrome	Model development	ML - kNN + physicochemical indicators	Mean accuracy 0.714
22	Xu Wang et al.	2020	Tooth-marked tongue / Spleen deficiency	Algorithm development	Deep CNN - ResNet34	Effective classification of tooth-marked tongue
23	Dong Jingfang	2020	Tongue features in cancer patients	Clinical algorithm application	CNN - ResNet50	Identified lesions with high accuracy and AUC
24	An Weitao et al.	2019	Eight - principle pattern differentiation	Algorithm development	DBN based on RBM	Addressed many - to - one relationships between symptoms and patterns
25	Wang Yiqun et al.	2015	Five clinical pulse types	System design and testing	Four - layer BP neural network	Pulse recognition accuracy exceeded 90%

2. Discussion

The development of artificial intelligence in TCM diagnosis during the period 2015 - 2025 demonstrates a remarkable technological transition. In the early stage, studies mainly applied traditional machine learning methods to solve classification problems involving individual TCM patterns. For example, Du Fangzhen (2023) used multi-label classification algorithms for pattern diagnosis [4]. Subsequently, research rapidly shifted toward deep learning to automate the extraction of features from image and text data. Notably, the period from 2022 to 2025 witnessed the emergence of methods such as Knowledge Graphs and Large Language Models, although these approaches currently account for only a limited proportion, approximately 8%. These methods have initially demonstrated the potential to simulate certain aspects of TCM pattern differentiation reasoning, particularly in multidimensional reasoning tasks.

Alongside the evolution of algorithms, the scope of diagnostic applications has expanded from single-modality approaches to multimodal approaches. Data analysis shows that current studies still mainly focus on digitizing individual aspects of the Four Examinations, among which tongue diagnosis accounts for the highest proportion, at 16%. Representative studies, such as Wang Fei (2024) using an improved YOLOv10 model [5] and Li MY (2021) using U-Net integrated with Global Convolution, demonstrated effectiveness in segmenting and identifying tongue image features [6]. Other diagnostic aspects, such as facial complexion, appeared less frequently, although

methods such as meta-learning proposed by Luo Guancong (2025) [7] have been introduced to address the limitation of small datasets. The predominance of tongue diagnosis may be explained by the visual nature and relatively high standardizability of image data, which facilitates the application of deep learning models. In contrast, data such as pulse signals and inquiry-based information exhibit greater variability and are more difficult to standardize.

This digitization process not only supports data collection but also promotes the transition from qualitative to quantitative approaches in TCM diagnosis. However, one of the main barriers of deep learning models is their “black - box” nature, which limits interpretability. In this context, explainable AI (XAI) methods have gained increasing attention, although they currently appear in only 3 of the 25 studies. Some studies, such as those by Jiang Qiyu (2025) and Yin Z (2024), indicate that combining deep learning models with reasoning mechanisms can generate interpretable representations, thereby improving transparency.

In addition, the trend toward integrating TCM with modern medicine has become increasingly evident. Studies in this direction account for approximately 20% and utilize paraclinical data, such as biochemical tests or medical imaging, to support pattern classification. For example, Xia S (2020) [8] combined the ML - kNN algorithm with physicochemical indicators to analyze metabolic syndrome, while Jin X (2025) [9] applied a Gradient Boosting Machine model to distinguish Cold–Heat patterns in viral pneumonia. This approach initially provides quantitative evidence supporting the linkage between biological indicators and TCM patterns, thereby helping to bridge the gap between the two medical systems. However, most studies remain retrospective in nature and do not yet provide sufficient evidence to confirm broad clinical applicability.

IV. CONCLUSION

AI is increasingly being applied in TCM

diagnosis, with a trend shifting from traditional machine learning toward deep learning and integrated models. These developments contribute to the quantification of diagnosis, particularly in tongue diagnosis. However, current evidence remains limited because most studies have small sample sizes, lack standardized data, and still face challenges related to interpretability and generalizability. Therefore, it is necessary to develop large-scale standardized datasets and strengthen clinical validation to enhance the applicability of AI in TCM practice.

REFERENCES

1. Jiang M, Lu C, Zhang C, Yang J, Tan Y, Lu A, Chan K. Syndrome differentiation in modern research of traditional Chinese medicine. *Journal of Ethnopharmacology*. 2012;140:634-642. doi: 10.1016/j.jep.2012.01.033.
2. Pan D, Guo Y, Fan Y, Wan H. Development and Application of Traditional Chinese Medicine Using AI Machine Learning and Deep Learning Strategies. *Am J Chin Med*. 2024;52(3):605-623. doi: 10.1142/S0192415X24500265.
3. Ren Y, Luo X, Wang Y, Li H, Zhang H, Li Z, et al. Large Language Models in Traditional Chinese Medicine: A Scoping Review. *J Evid Based Med*. 2025;18(1):e12658. doi: 10.1111/jebm.12658.
4. 杜昉臻, 何圆姣, 冯西贝, 刘国华. 基于人工智能的中医证候分类算法研究. *南开大学学报(自然科学版)*. 2023;56(2):12-16.
5. 王菲, 曲强, 刘明阳, 等. 基于改进YOLOv10的舌象检测辨识系统. 2024.
6. Li MY, Zhu DJ, Xu W, Lin YJ, Yung KL, Ip AWH. Application of U-Net with Global Convolution Network Module in Computer-Aided Tongue Diagnosis. *J Healthc Eng*. 2021;2021:5853128. doi: 10.1155/2021/5853128.
7. 罗冠聪, 冯跃, 徐红, 秦传波, 李福凤, 钱鹏, 等. 元学习优化的面部色诊分类方法. *计算机与现代化*. 2025; (12): 115-122.
8. Xia S, Zhang J, Du G, Li S, Vong CT, Yang Z, et al. A Microcosmic Syndrome Differentiation Model for Metabolic Syndrome with Multilabel Learning. *Evid Based Complement Alternat Med*. 2020;2020:9081641. doi: 10.1155/2020/9081641.
9. Jin X, Wang Y, Wang J, Gao Q, Huang Y, Shao L, et al. A Machine Learning Approach to Differentiate Cold and Hot Syndrome in Viral Pneumonia Integrating Traditional Chinese Medicine and Modern Medicine: Machine Learning Model Development and Validation. *JMIR Med Inform*. 2025;13:e64725. doi: 10.2196/64725.