

## Artificial intelligence in traditional Chinese medicine: from systems biological mechanism discovery, real-world clinical evidence inference to personalized clinical decision support

Dengying Yan, Qiguang Zheng, Kai Chang, Rui Hua, Yiming Liu, Jingyan Xue, Zixin Shu, Yunhui Hu, Pengcheng Yang, Yu Wei, Jidong Lang, Haibin Yu, Xiaodong Li, Runshun Zhang, Wenjia Wang, Baoyan Liu, Xuezhong Zhou

**Citation:** Dengying Yan, Qiguang Zheng, Kai Chang, Rui Hua, Yiming Liu, Jingyan Xue, Zixin Shu, Yunhui Hu, Pengcheng Yang, Yu Wei, Jidong Lang, Haibin Yu, Xiaodong Li, Runshun Zhang, Wenjia Wang, Baoyan Liu, Xuezhong Zhou, Artificial intelligence in traditional Chinese medicine: from systems biological mechanism discovery, real-world clinical evidence inference to personalized clinical decision support, *Chinese Journal of Natural Medicines*, 2025, 23(11), 1310–1328. doi: [10.1016/S1875-5364\(25\)60983-6](https://doi.org/10.1016/S1875-5364(25)60983-6).

View online: [https://doi.org/10.1016/S1875-5364\(25\)60983-6](https://doi.org/10.1016/S1875-5364(25)60983-6)

---

## Related articles that may interest you



Wechat



Contents lists available at ScienceDirect

## Chinese Journal of Natural Medicines

journal homepage: [www.cjnmcpu.com/](http://www.cjnmcpu.com/)

## Review

# Artificial intelligence in traditional Chinese medicine: from systems biological mechanism discovery, real-world clinical evidence inference to personalized clinical decision support



Dengying Yan<sup>a</sup>, Qiguang Zheng<sup>a</sup>, Kai Chang<sup>a</sup>, Rui Hua<sup>a</sup>, Yiming Liu<sup>a</sup>, Jingyan Xue<sup>a</sup>, Zixin Shu<sup>b</sup>, Yunhui Hu<sup>c</sup>, Pengcheng Yang<sup>c</sup>, Yu Wei<sup>c</sup>, Jidong Lang<sup>c</sup>, Haibin Yu<sup>d</sup>, Xiaodong Li<sup>b</sup>, Runshun Zhang<sup>e</sup>, Wenjia Wang<sup>c,\*</sup>, Baoyan Liu<sup>f,\*</sup>, Xuezhong Zhou<sup>a,\*</sup>

<sup>a</sup> Institute of Medical Intelligence, School of Computer Science & Technology, Beijing Jiaotong University, Beijing 100044, China

<sup>b</sup> Institute of Liver Diseases, Hubei Provincial Hospital of Traditional Chinese Medicine, Wuhan, Hubei 430061, China

<sup>c</sup> Tianjin Tasly Digital Intelligence Chinese Medicine Technology Co., Ltd., Tianjin 300410, China

<sup>d</sup> Department of Respiratory Diseases, the First Affiliated Hospital of Henan University of Chinese Medicine, Zhengzhou 450003, China

<sup>e</sup> Guang'anmen Hospital, China Academy of Chinese Medical Sciences, Beijing 100053, China

<sup>f</sup> China Academy of Chinese Medical Sciences, Beijing, 100700, China

## ARTICLE INFO

## Article history:

Received 25 March 2025

Revised 14 August 2025

Accepted 23 August 2025

Available online 20 November 2025

## Keywords:

Artificial intelligence

Systems biological mechanism

Real-world clinical evidence

Clinical decision support

## ABSTRACT

Traditional Chinese medicine (TCM) represents a paradigmatic approach to personalized medicine, developed through the systematic accumulation and refinement of clinical empirical data over more than 2000 years, and now encompasses large-scale electronic medical records (EMR) and experimental molecular data. Artificial intelligence (AI) has demonstrated its utility in medicine through the development of various expert systems (e.g., MYCIN) since the 1970s. With the emergence of deep learning and large language models (LLMs), AI's potential in medicine shows considerable promise. Consequently, the integration of AI and TCM from both clinical and scientific perspectives presents a fundamental and promising research direction. This survey provides an insightful overview of TCM AI research, summarizing related research tasks from three perspectives: systems-level biological mechanism elucidation, real-world clinical evidence inference, and personalized clinical decision support. The review highlights representative AI methodologies alongside their applications in both TCM scientific inquiry and clinical practice. To critically assess the current state of the field, this work identifies major challenges and opportunities that constrain the development of robust research capabilities—particularly in the mechanistic understanding of TCM syndromes and herbal formulations, novel drug discovery, and the delivery of high-quality, patient-centered clinical care. The findings underscore that future advancements in AI-driven TCM research will rely on the development of high-quality, large-scale data repositories; the construction of comprehensive and domain-specific knowledge graphs (KGs); deeper insights into the biological mechanisms underpinning clinical efficacy; rigorous causal inference frameworks; and intelligent, personalized decision support systems.

## 1. Introduction

Traditional Chinese medicine (TCM) represents an ancient medical system founded on empirical observations and accumulated clinical knowledge spanning over 2000 years of Chinese history. Its distinctive theoretical framework stems from fundamental Chinese philosophical concepts, particularly the “Yin-Yang” and “Five Elements” theory. Clinical practice in TCM adheres to the principle of “Treatment based on Syndrome Differentiation” (辨证论治), a comprehensive and personalized approach encompassing diagnosis, treatment formulation, and outcome evaluation.

Substantial differences exist between TCM's theoretical foundations and diagnostic-therapeutic methodologies compared to modern biomedicine<sup>3</sup>. Since 1950, TCM modernization has steadily progressed toward scientific systematization<sup>4</sup>. Qian Xuesen, a prominent Chinese scientist, notably asserted: “The future of medicine is the modernization of TCM, not in any other way ...<sup>5</sup>” Post-2011, coordinated efforts have strengthened TCM's scientific foundation and clinical applications through interdisciplinary, interregional, and intersectoral collaborations<sup>6</sup>. The advent of the big data era has particularly highlighted the advantages of computational methods, accelerating their implementation in TCM research<sup>7</sup>. For instance, in 2010, Li et al.<sup>8</sup> utilized network pharmacology and computational modeling techniques to elucidate the combinatorial mechanisms of herbal formulations, using the classic Liu-wei-di-huang (LWDH) formula as

\* Corresponding author.

E-mail addresses: [tsl-wangwenjia@tasly.com](mailto:tsl-wangwenjia@tasly.com) (W. Wang); [liubaoyan@139.com](mailto:liubaoyan@139.com) (B. Liu); [xzzhou@bjtu.edu.cn](mailto:xzzhou@bjtu.edu.cn) (X. Zhou)

a case study. TCM's clinical value received additional recognition during the COVID-19 pandemic<sup>9</sup>. However, the biological mechanisms of TCM diagnosis and treatment require further elucidation, and sophisticated clinical capabilities and high-quality evidence need additional development due to the complex systems involved in both TCM clinical decisions and the network systems of medical entities (e.g., herb prescriptions and TCM theoretical concepts).

Artificial intelligence (AI) was formally established as an academic discipline at the Dartmouth Summer Research Project in June 1956, where John McCarthy—widely recognized as the father of AI<sup>10</sup>, defined it as “the science and engineering of creating intelligent machines”. Following this foundational moment, sub-fields such as machine learning (ML) and knowledge engineering (KE) emerged as core components of AI, endowing machines with capabilities for logical inference and adaptive, human-like intelligent behavior<sup>11,12</sup>. Over the past several decades, AI has evolved at an unprecedented pace, becoming one of the most transformative general-purpose technologies in modern history. Its cross-disciplinary impact is particularly evident in medicine, where AI has played a pivotal role in advancing clinical decision-support systems<sup>13,14</sup>. A seminal milestone occurred in 1972 with MYCIN, an expert system for diagnosing bacterial infections and recommending anti-microbial therapies<sup>15</sup>. Since then, AI applications in healthcare have expanded significantly, encompassing clinical decision support for diagnosis and treatment, elucidation of disease mechanisms, and the advancement of precision medicine. Key achievements include breakthroughs in medical image analysis<sup>16</sup>, medical knowledge question answering<sup>17</sup>, disease risk prediction<sup>18</sup>, clinical research<sup>19</sup>, pharmacological mechanism research (such as Network Pharmacology<sup>20</sup>), and protein structure prediction (such as AlphaFold<sup>21</sup>). Recent years have witnessed important technical explorations in medical large language model (LLM) research<sup>22</sup>. TCM entered the AI domain in 1978 through Guan Youbo's computer program for diagnosis and treatment of liver diseases<sup>23</sup>. This early TCM expert system represented a significant achievement, serving patients across more than 30 Chinese medical institutions while effectively disseminating specialist knowledge to primary care settings, thereby improving healthcare quality. Contemporary AI applications continue revolutionizing TCM practice by optimizing diagnostic accuracy, treatment personalization, and healthcare delivery<sup>24,25</sup>.

In recent years, artificial intelligence in traditional Chinese medicine (AITCM) has emerged as a transformative research frontier for both clinical decision-making and biological research<sup>26</sup>. By synergizing AI with large-scale clinical data and rich theoretical knowledge of TCM, AITCM addresses longstanding clinical and biological challenges in TCM fields, including scientific systems biological mechanism discovery, real-world clinical evidence inference, and personalized clinical decision support. This convergence promises to catalyze TCM's scientific valida-

tion, global acceptance, and modernization, propelling the field toward an era of intelligent and precision healthcare.

Therefore, with the surge of LLMs and medical intelligence, this review systematically examines current progress of AITCM from three key perspectives, while critically evaluating the data sources and knowledge databases, technical achievements, persistent challenges, and future directions to develop high-value AITCM techniques and applications. (Fig. 1 shows the outline of this article)

## 2. TCM databases and knowledge graphs (KGs)

Due to the data-driven paradigm and knowledge-oriented discipline characteristics of contemporary AI techniques, both high-quality data resources and KGs constitute key foundations for AITCM.

TCM fields encompass diverse data sources, including real-world data (RWD) (e.g., electronic medical records (EMR)), scientific literature, and TCM ancient books. Additionally, high-throughput biomedical data (e.g., clinical genomics and proteomics, pharmacological data on cell lines and animals) have accumulated rapidly<sup>27,28</sup>. Modern biomedical public repositories, such as DrugBank<sup>29</sup> and ChEMBL<sup>30</sup>, provide additional valuable resources. Substantial efforts have been made to integrate these data resources to advance AITCM research. Several types of databases and datasets have been developed, including TCM pharmacology databases (TCM Database @Taiwan<sup>31</sup>, TCMID<sup>32</sup>, TCMSP<sup>33</sup>, TCM-Mesh<sup>34</sup>, YaTCM<sup>35</sup>, ETCM<sup>36</sup>, HERB<sup>37</sup>, TCMBank<sup>38</sup>, HIT2.0<sup>39</sup>, BATMAN-TCM 2.0<sup>40</sup>, CPMCP<sup>41</sup>), TCM phenotype-related databases (SymMap<sup>42</sup>, LTM-TCM<sup>43</sup>, TCMM<sup>44</sup>, TCMSSD<sup>45</sup>), and disease-specific TCM databases (e.g. TCMIO<sup>46</sup>). For trustworthy AI, high-quality datasets with multiple quality dimensions (e.g., accuracy, completeness, and consistency) have been constructed for specific AITCM tasks. For instance, Dong et al.<sup>47</sup> developed a clinical TCM-Lung dataset incorporating symptoms, syndromes, treatment methods, and herbs, addressing previous datasets' limitations and achieving significant improvements in clinical diagnosis and treatment. Hua et al.<sup>48</sup> created an extensive TCM dataset, including of TCM pre-trained (TCMPT) dataset, traditional Chinese patent medicine (TCPM) question answering dataset, and outpatient data of TCM spleen and stomach herbal prescription recommendation (SSHPR), encompassing TCM ancient books, textbooks, case records, the Chinese Pharmacopoeia (2020), and package inserts, along with publicly available data resources, providing a comprehensive foundation for LLM training in TCM. Table 1 presents a comprehensive summary of TCM databases. Fig. 2 provides a visual comparison of these databases. Among them, SymMap, which bridges TCM and modern medicine, represents an integrative TCM database enhanced by symptom mapping<sup>42</sup>. Its version 2.0 contains expanded data including 14 086 diseases associated with 2518 TCM symptoms, 698 herbs,

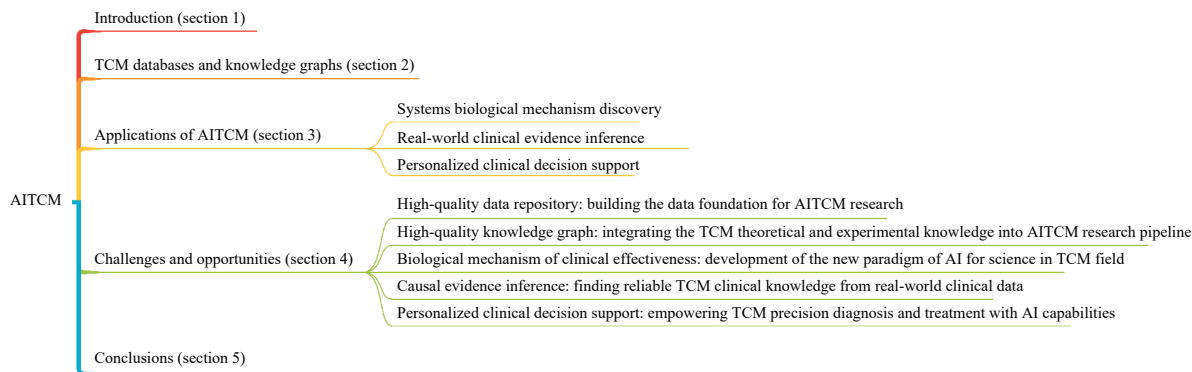


Fig. 1 The outline of this article

**Table 1** Public databases related to TCM

Year	Name	Entity Types	Total Amount	Advantages	Limitations	Database URL	Website Status
2011	TCM Database@Taiwan	Herb, Ingredient/Compound	20453	TCM Database@Taiwan is a milestone for modern TCM researches, with each pure compound are available for download and virtual screening in both cdx (2D) and Tripos mol2 (3D) formats.	Released in 2011, the TCM Database@Taiwan has become outdated and limited in coverage, and has since been succeeded by the more comprehensive TCMBank.	<a href="http://tcm.cmu.edu.tw/">http://tcm.cmu.edu.tw/</a>	No access
2013	TCM-ID	Herb, Ingredient/Compound, Prescription, Drug, Disease, Target	108421	TCMID bridge TCM and modern Western medicine at the level of herbal ingredients/compounds with their targets.	Limitations of TCMID include uncertainty in ingredient-target associations and the lack of comprehensive omics and ethnomedicine data. Version 2.0 was released in 2018.	<a href="http://www.megabionet.org/tcmid/">http://www.megabionet.org/tcmid/</a>	No access
2014	TCMSP	Herb, Ingredient/Compound, Disease, Target	34031	TCMSP is more than a data repository, featuring herbal entries with ADME properties, and also contains tools for visualization and analysis of TCM results on the network level.	TCMSP is currently limited by the lack of detailed pharmacological data, such as drug action modes and combinations, and does not yet incorporate physiologically based pharmacokinetic (PBPK) modeling for realistic in vivo behavior prediction.	<a href="https://old.tcmsp-e.com/load_intro.php?id=40">https://old.tcmsp-e.com/load_intro.php?id=40</a>	Normal access
2017	TCM-Mesh	Herb, Ingredient/Compound, Disease, Gene, Side Effect Record, Toxic Record	573869	TCM-Mesh is a platform designed as an integration of database (providing toxic and side effect records) and a data-mining system for network pharmacology analysis of TCM preparations.	TCM-Mesh is limited by its use of a rudimentary combined scoring system that lacks precision in representing compound-target-disease pathways, highlighting the need for a more refined scoring function.	<a href="http://mesh.tcm.microbioinformatics.org/">http://mesh.tcm.microbioinformatics.org/</a>	No access
2019	SymMap	Herb, Ingredient/Compound, Disease, Target, MM symptom, TCM symptom	32309	SymMap is the first publically available database containing comprehensive information regarding the relationships between TCM symptoms, TCM herbs and MM symptoms	TCM symptom and herb associations are derived from the Chinese Pharmacopoeia (CHPH, 2015 edition) and manually curated, but lack clinical validation.	<a href="http://www.symmap.org/">http://www.symmap.org/</a>	Normal access
2021	HERB	Herb, Ingredient/Compound, Disease, Target	97666	HERB is a high-throughput experiment- and reference-guided database of TCM, establishing connections between TCM herbs/ingredients and modern drugs.	HERB is currently limited by the insufficient quantity and diversity of pharmacotranscriptomic datasets, lacking integration of multi-omics data such as proteomics, metabolomics, and metagenomics.	<a href="http://herb.ac.cn/">http://herb.ac.cn/</a>	Normal access
2022	LTM-TCM	Herb, Ingredient/Compound, Prescription, Target, TCM symptom	107252	LTM-TCM integrates 14 authoritative TCM authoritative databases, provides online computational pipelines that allow users to perform virtual docking calculations on potential targets of ingredients and predictive analysis services for ADME.	LTM-TCM is limited by incomplete translation of ancient data, insufficient integration with modern medical resources, limited validation of herb-related associations, and partial disclosure of safety information.	<a href="http://cloud.tasly.com/#/tcm/home">http://cloud.tasly.com/#/tcm/home</a>	No access
2022	HIT2.0	Herb, Ingredient/Compound, Target	4695	HIT 2.0 is the latest and the largest dataset regarding curated ingredient-target pairs for herbs.	HIT 2.0 is limited by its results collected mostly from in vitro rather than in vivo experiments.	<a href="http://hit2.badd-cao.net">http://hit2.badd-cao.net</a>	Normal access
2023	TCMBank	Herb, Ingredient/Compound, Disease, Target	118866	TCMBank is an extension of TCM Database@Taiwan, provides a structure-based ingredient compound search function and simple batch download of molecular 3D structures.	TCMBank is limited by relatively low gene data coverage and currently relies on external public databases for future data expansion and deeper exploration of herb-ingredient-target-disease associations.	<a href="https://TCMBank.cn/">https://TCMBank.cn/</a>	Normal access
2023	ETCM 2.0	Herb, Ingredient/Compound, Prescription, Chinese Patent Drug, Disease, Target	107776	ETCM 2.0 adds a new data field "Traditional Chinese Medicine Formulas" based on ancient Chinese medical books, provides functional applications such as drug like evaluation and potential toxicity evaluation for drug discovery.	ETCM 2.0 serves as an encyclopedia of TCM, but its data needs further evaluation.	<a href="http://www.tcmip.cn/ETCM2/front/#/">http://www.tcmip.cn/ETCM2/front/#/</a>	Normal access

Continued

Year	Name	Entity Types	Total Amount*	Advantages	Limitations	Database URL	Website Status
2024	TCMM	Herb, Ingredient/Compound, Prescription, Disease, Target, MM symptom, TCM symptom, Syndrome	244730	TCMM contains the most comprehensive entity type and introduces the prescription-syndrome associations for the first time.	The relations of prescription in TCMM require further validation, and both the accuracy and interpretability of the data need to be improved, so it is only used for educational and scientific research purposes only.	<a href="https://www.tcm.net.cn/">https://www.tcm.net.cn/</a>	Normal access
2024	TCMSSD	Herb, Ingredient/Compound, Prescription, Drug, Disease, Target, Syndrome	219671	TCMSSD highlights the content and modules of TCM syndrome differentiation (characteristically including TCM-specific diseases) and is the first comprehensive database exclusively designed for TCM syndrome differentiation	The limitation of TCMSSD is its inability to accurately link syndromes with diseases due to biological complexity and syndrome heterogeneity.	<a href="http://tcmssd.ratcm.cn">http://tcmssd.ratcm.cn</a>	Normal access
2024	BATMAN-TCM 2.0	Herb, Ingredient/Compound, Prescription, Targets	112334	BATMAN-TCM 2.0 has significantly expanded TCM ingredient-target protein interaction (TTI) coverage, can provide more information on potential target proteins for drug development.	BATMAN-TCM 2.0 is limited by its lack of integration with omics databases and limited coverage of experimental species in the TTI dataset.	<a href="http://bionet.ncpsb.org.cn/batman-tcm/">http://bionet.ncpsb.org.cn/batman-tcm/</a>	Normal access
2024	CPMCP	Herb, Ingredient/Compound, Prescription, Chinese Patent Drug, Disease, Target, MM symptom, TCM symptom	68507	CPMCP maps TCM prescriptions with TCM symptoms manually and summarizes the common combination principles of prescriptions.	CPMCP has a limited data scope, covering only Chinese patent medicines and ancient Chinese medicine prescriptions.	<a href="http://cpmcp.top">http://cpmcp.top</a>	No access

\* The Total Amount refers solely to the number of entities listed under the Entity Type column.

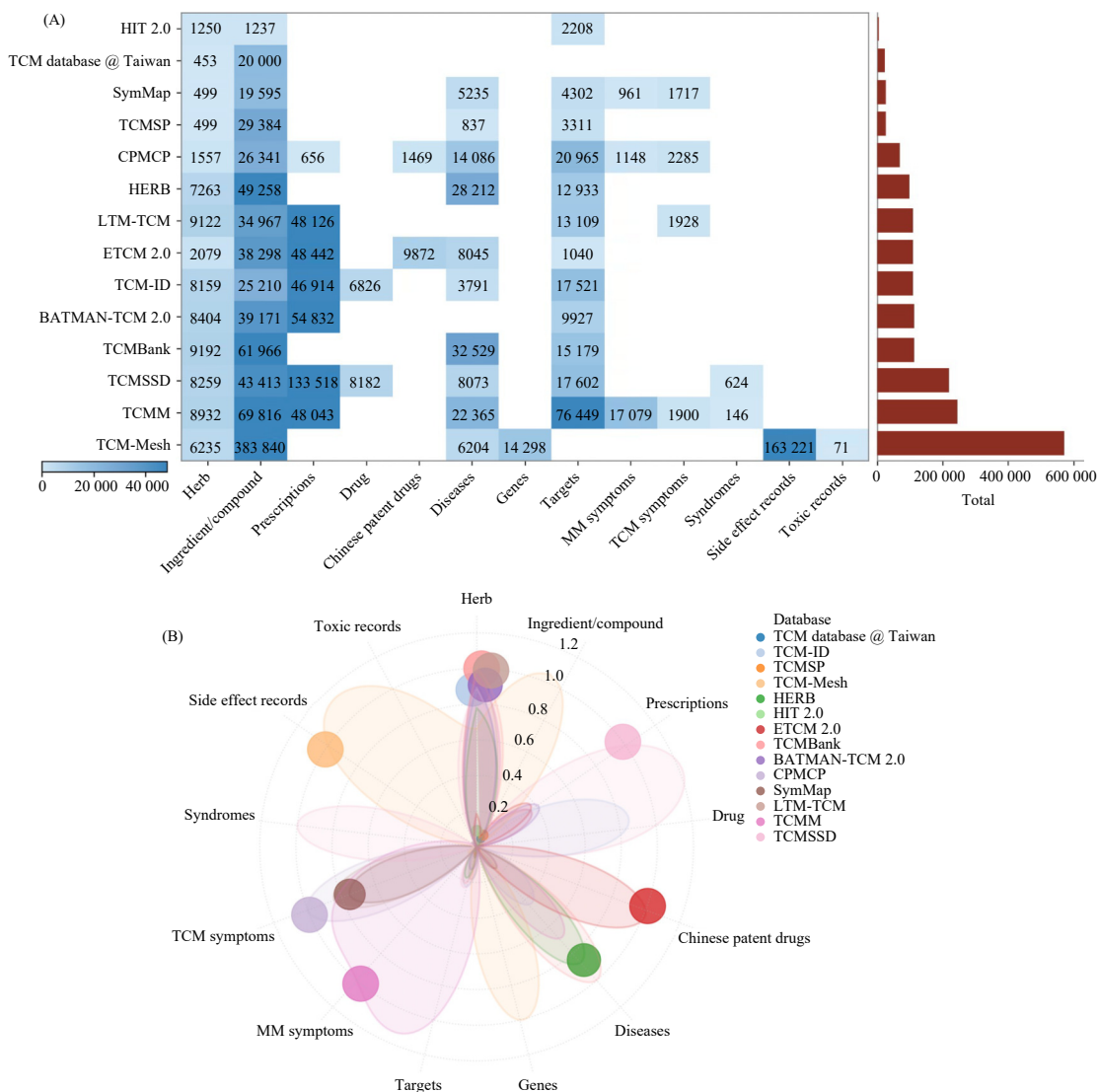
26 035 herbal ingredients and 20 965 target genes<sup>42</sup>, supporting numerous TCM network pharmacology studies<sup>49,50</sup>. While these databases complement each other and maintain distinct research focuses<sup>51</sup>, they face challenges regarding data accuracy and authenticity<sup>52</sup>. Furthermore, inconsistent terminology and diverse information across TCM databases limit data reusability and integration, hindering the development of findable, accessible, interoperable, and reusable TCM databases<sup>53</sup>. Researchers requiring complex data manipulation must often conduct multiple searches across various databases, highlighting the persistent need for a comprehensive, authoritative, high-quality database.

KGs have emerged as an effective solution for managing and integrating diverse data sources<sup>54</sup>. Furthermore, they enable transparent reasoning and interpretation, which are crucial for advancing AI-driven insights<sup>55</sup>. Google introduced the concept of KG in 2012 to describe and structure large-scale semantic relationships between concepts and entities in a graph format<sup>56,57</sup>. YAGO represents a prominent example<sup>58</sup>. KG formats can be expressed through semantic triples using resource description framework (RDF) or description logic ontologies<sup>56</sup>. An ontology represents a formal specification of conceptualization that facilitates knowledge sharing and reuse<sup>59-61</sup>. It functions as a crucial tool for data integration, enabling semantic knowledge modeling that encompasses classes and properties (e.g., owl:ObjectProperty and owl:DatatypeProperty) as well as instances<sup>57,62</sup>.

In the medical field, numerous KGs have been developed, including notable examples such as the unified medical language system (UMLS)<sup>63</sup>, clinical knowledge graph (CKG)<sup>64</sup>, PrimeKG<sup>65</sup>, PharmKG<sup>66</sup>, and the scalable precision medicine open knowledge engine (SPOKE)<sup>64</sup>. UMLS, developed in the 1980s by the US National Library of Medicine, stands as a prominent medical KG with diverse applications, integrating over 200 medical terminologies and vocabularies worldwide<sup>63</sup>. Recent medical KGs like PrimeKG have incorporated multi-omic data, enabling the integration of disparate information to advance precision medicine in disease understanding and treatment optimization. Additionally,

researchers increasingly focus on developing interpretable causal KGs to enhance clinical decision-making and medical discovery<sup>67</sup>. Causal KGs leverage recent advances in causality and KG technology to improve explainability and facilitate causal knowledge integration<sup>68</sup>. Various data sources, including RWD, can be utilized for information extraction to develop high-quality causal KGs supporting clinical decisions for diabetic nephropathy<sup>69</sup>.

TCM represents a knowledge-intensive discipline, distinguished by its complex theoretical framework and extensive empirical knowledge accumulated over centuries<sup>70</sup>. The theoretical knowledge in TCM has historically been characterized by vagueness, uncertainty, uninterpretability and unverifiability<sup>71</sup>. These challenges arise from the inherent complexity and non-standardized nature of TCM concepts<sup>72</sup>. Moreover, TCM encompasses intricate semantic relationships, with multiple connections emerging throughout the system<sup>73</sup>. Consequently, modern medicine often questions TCM's analytical and inferential processes<sup>74</sup>. To address these issues, initial research focused on scientific representation and formalization of TCM knowledge, particularly through ontology-based semantic models. These models reflect human cognitive organization, making them suitable for complex knowledge system representation<sup>73</sup>. TCMLS (originally UTCMLS<sup>72</sup>), established in 2001, represents a significant TCM ontology. Developed through collaborative efforts of over 300 experts, TCMLS provides comprehensive TCM knowledge integration, establishing core top-level categories and fundamental structures analogous to UMLS architecture. Following a decade of implementation and enhancement, TCMLS has become the largest TCM domain ontology<sup>75</sup>. Its impact has inspired additional TCM subfield ontologies, including the 2018 General Formal Ontology (GFO)-TCM, which bridges traditional and modern medical terminology<sup>76</sup>. The Traditional Chinese Drugs Ontology (TCDO), developed in 2019, defines Chinese drugs and their classifications<sup>77</sup>. Recent additions include SoFDA, a manually curated TCM syndrome ontology database<sup>78</sup>, and ISPO, designed for TCM symptom data semantic integration<sup>79</sup>. Given TCM's specialized terminology, continued ontology development for areas such as Chinese herbal



**Fig. 2** Comparison of TCM databases (A) Comparison of the number of entities and total amount across different databases. (B) Normalization of internal data within each database, represented in the radar bubble chart to reflect the unique characteristics of each database.

medicine (CHM) efficacy and classical prescriptions remains essential for standardizing knowledge representation.

With advancements in the storage and transmission of molecular biological information, there is increasing interest in integrating TCM with molecular biology. KGs have become prominent by encompassing a broad range of entities and semantic connections, facilitating the integration of medical knowledge across multiple domains<sup>80</sup>. As shown in Table 1, an extensive collection of TCM data and associated databases or KGs centers around key elements in TCM research, including prescriptions, herbs, components, targets, genes, phenotypes, and TCM syndromes. Lu et al.<sup>81</sup> introduced SympGAN, a comprehensive, high-quality KG focused on symptom-gene associations. This network, structured around four core entities (symptoms, drugs, genes, and diseases), distinctively incorporates inferred relationships between symptoms and genes. This research demonstrates significant potential for advancing TCM symptom science. While TCM's complexity maintains its "black box" nature<sup>82</sup>, the aforementioned databases, ontologies, and KGs provide essential resources for understanding its theories and training AI models to support TCM diagnosis and treatment<sup>83,84</sup>.

### 3. Applications of AITCM

AITCM applications can be broadly categorized into two top-

ics: clinical and biological research (Fig. 3 shows biological and clinical research scenarios of AITCM). Clinical research topics primarily focus on enhancing clinical evidence discovery and decision capabilities, specifically for real-world clinical evidence generation and personalized clinical decision support. In recent decades within TCM fields, researchers have conducted various clinical research tasks, including the development of acquisition devices for the four diagnostic methods of TCM<sup>85</sup>, intelligent diagnosis<sup>86</sup>, herbal or acupuncture treatment recommend systems<sup>87,88</sup>, clinical data mining algorithms<sup>89,90</sup> and decision support systems based on data warehouse<sup>91,92</sup>. Regarding biological research topics, extensive investigation has focused on the underlying systems-level biological mechanism of syndrome phenotypes and various types of herbs or herbal prescriptions, with network pharmacology representing a typical task (Table 2 shows the open-source prediction tools for AI applications in TCM in the past five years).

#### 3.1. Systems biological mechanism discovery

The primary objective of elucidating the systems biological mechanisms underlying TCM is to understand how TCM regulates human life systems and disease pathologies, thus establishing a connection between TCM and modern medicine. The main tasks involve identifying associated genes and network mechan-

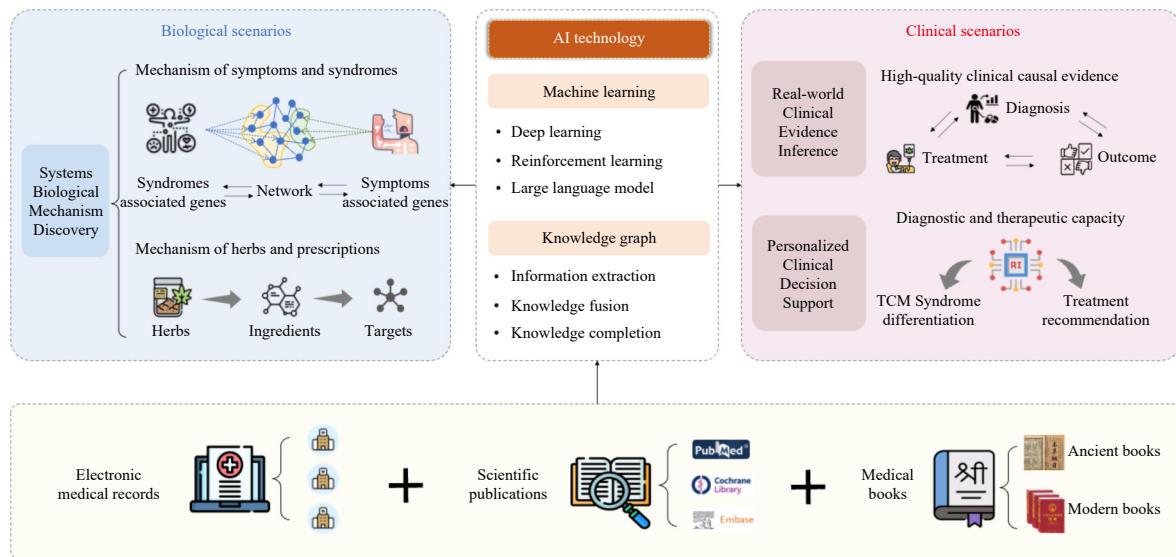


Fig. 3 Biological and clinical research scenarios of AITCM.

isms of symptoms and syndromes, as well as associated drug targets and network mechanisms of herbs and/or TCM treatments (e.g., herbal prescriptions and acupuncture). Given the network structure of related data sources, the principal AI approaches employed include network medicine (e.g., complex network methods using biomedical networks), network pharmacology, link prediction, network community detection, and graph representation learning (i.e., deep learning on graph/network structures). Additionally, the emergence of high-throughput technologies, such as genomics, proteomics, metabolomics, and transcriptomics, has transformed biological research. Multi-omics strategies utilize molecular profiles at multiple levels to demonstrate the dynamic changes during therapeutic responses, providing novel perspectives on TCM's biological mechanisms<sup>103,104</sup>. (Fig. 4 shows the research framework of systems biological mechanism discovery)

Disease genes cause or contribute genetically to the development of complex diseases<sup>105</sup>. In biomedical research, identifying genes underlying disease susceptibility in humans has been a primary objective. Recent studies demonstrate that disease genes serve as promising drug targets for human disorders<sup>106</sup>. The identification of disease-associated genes and their corresponding targets has become fundamental to drug discovery<sup>107-109</sup>. Network medicine has emerged as an essential tool for elucidating these relationships by integrating multi-omics data with complex network analysis. Several computational approaches are employed to identify or predict relationships between targets and genes. These include constructing gene regulatory networks and protein-protein interaction (PPI) networks to explore functional connections between genes and their potential as drug targets. The topological properties of networks, such as node centrality, connectivity, and modularity, provide significant insights into the biological significance of specific genes<sup>110</sup>. Analysis of these properties enables identification of key regulatory genes that may serve as potential therapeutic targets. Experimental validation methods, including gene expression profiling and ribonucleic acid interference (RNAi), complement computational network analysis to confirm predicted relationships. These experimental approaches refine initial network model predictions, providing robust evidence for the biological relevance of identified targets. For instance, Chen et al.<sup>111</sup> constructed a liver-specific network for type 2 diabetes by integrating RNA sequencing data with network topological properties, identifying key nodes subsequently validated through network pharmacology and molecular docking. This approach demonstrates how combining computational pre-

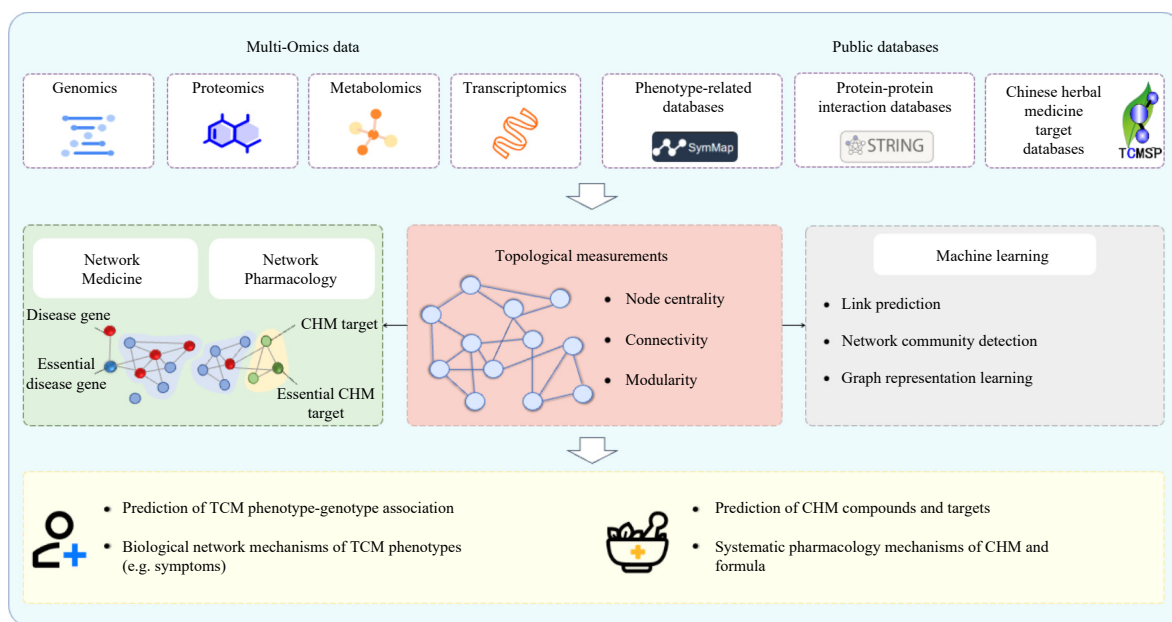
dictions with experimental validation enhances target identification accuracy and efficiency. The integration of these methodologies enables both prediction of gene-target relationships and deeper understanding of underlying systems biology mechanisms, potentially leading to more effective therapeutic strategies.

Analogous to disease-gene prediction for disease mechanism identification, symptom-gene or syndrome-gene association prediction is crucial for TCM phenotype-genotype associations. Over the past two decades in TCM research, numerous studies have investigated associated genes of symptom phenotypes using clinical data and biomedical networks<sup>112,113</sup>. Symptom phenotypes represent observable disease manifestations recognized by physicians and patients. Understanding their molecular mechanisms, particularly network mechanisms, is essential for investigating the pathogenesis of symptom phenotype non-specificity. Shu et al.<sup>114</sup> developed a phenotype network of symptoms by integrating clinical co-occurrence, symptom-gene associations, and PPIs to explore molecular mechanisms, applying a network diversity measure to quantify phenotypic and molecular diversity. TCM syndrome, representing patient manifestation classification, forms the foundation of "Treatment based on Syndrome Differentiation". Substantial cellular and molecular studies have investigated biomedical explanations of TCM syndromes. Dynamic syndrome changes present a significant challenge in elucidating their biological basis. In 2009, Li et al.<sup>115</sup> introduced the initial "Phenotype network-Biological network-Herb network" map to explore underlying network systems and identify TCM syndrome biomarkers. This multi-layered map initiated systematic interpretation of TCM theory and practice, bridging ancient TCM with future systems biology-based medicine at system and molecular levels. In 2013, Li et al.<sup>116</sup> proposed a network balance model successfully applied to reveal the biological basis of Cold Syndrome and Hot Syndrome in chronic gastritis. However, additional AI methods are required to fully understand TCM syndromes' biological basis.

Drug compounds interact with their targets as primary therapeutic approaches for complex diseases<sup>117</sup>. Predicting drug compounds and targets in CHM remains fundamental to AITCM research<sup>118</sup>. The extraction of bioactive compounds from natural plants and animals is essential for understanding CHM's pharmacological mechanisms<sup>119</sup>. The discovery of artemisinin by Tu Youyou exemplifies CHM's contribution to modern drug development<sup>120</sup>, demonstrating its potential and adaptability to biomedical advances<sup>121,122</sup>. While thousands of therapeutic herbs have been identified, the complexity of isolating and validating indi-

**Table 2** Open-source prediction tools for AI applications in TCM in the past five years

Prediction Task	Name	Publication Year	Methods	Description	Web link
Prediction of TCM phenotype-genotype association	SympGAN <sup>81</sup>	2023	Heterogeneous network embedding representation algorithm	SympGAN is a high-quality, symptom-centric knowledge base that enables advanced symptom knowledge inference through integrated KGs and interactive online services.	<a href="http://www.sympgan.org/">http://www.sympgan.org/</a>
Prediction of Chinese medicine components and targets	SYSTEM <sup>93</sup>	2024	GoogLeNet and Support Vector Machines	SYSTEM is a systemic platform for the identification of pharmacological effects of TCM and TCM components.	<a href="http://system.cn">http://system.cn</a>
Prediction of Chinese medicine components and targets	NP-TCMtarget <sup>94</sup>	2025	Deep learning binary classification model	NP-TCMtarget can identify direct targets that directly bind with the drug to produce biological effects and indirect targets that are subsequently influenced by the direct targets in the biological molecular network.	<a href="http://www.bcxnfz.top/NP-TCMtarget">http://www.bcxnfz.top/NP-TCMtarget</a>
Discovery of TCM prescription compatibility rules	TCMNPAS <sup>95</sup>	2024	Binomial formula-target identification model, BK algorithm for prescription mining, network-based proximity measures and molecular docking	TCMNPAS is a comprehensive platform that enables the analysis of "core formulas with core herbs and core targets" using a network approach.	<a href="https://github.com/yangpluszhu/tcmnpas">https://github.com/yangpluszhu/tcmnpas</a>
Discovery of TCM diagnosis and treatment knowledge	FangNet <sup>96</sup>	2021	PageRank algorithm	FangNet, a tool for retrospective analysis to a collection of empirical prescriptions for particular diseases, which can rank each herb's relative importance, determine the core herbs and find the associations to specific symptoms.	<a href="http://fangnet.org">http://fangnet.org</a>
Prediction of the efficacy of TCM	TCM2Vec <sup>97</sup>	2023	Deep learning model	TCM2Vec can predict the efficacy of formula.	<a href="https://github.com/Gao-WQ/TCM2Vec">https://github.com/Gao-WQ/TCM2Vec</a>
TCM Clinical Decision Support	TCMFP <sup>98</sup>	2023	Network target's score and semi-supervised learning genetic algorithms	TCMFP combines therapy experience of TCM, AI and network science algorithms to screen optimal herbal formula for diseases efficiently.	<a href="https://github.com/niuqikai/formula_study">https://github.com/niuqikai/formula_study</a>
TCM Clinical Decision Support	PresRecST <sup>47</sup>	2024	Residual-like neural networks	PresRecST implements a progressive diagnostic and treatment workflow and effectively combines the components of syndrome differentiation, treatment method determination, and herb recommendation.	<a href="https://github.com/2020MEAI/PresRecST">https://github.com/2020MEAI/PresRecST</a>
TCM Clinical Decision Support	ShenNong-TCM <sup>99</sup>	2023	LLMs	ShenNong-TCM is the first TCM LLM and can effectively recommend prescriptions/CHMs based on symptoms.	<a href="https://github.com/michael-wzhu/ShenNong-TCM-LLM">https://github.com/michael-wzhu/ShenNong-TCM-LLM</a>
TCM Clinical Decision Support	Lingdan <sup>48</sup>	2024	LLMs	Lingdan demonstrates state-of-the-art performances for the tasks of TCM clinical knowledge answering and herbal prescription recommendation.	<a href="https://github.com/TCMAI-BJTU/LingdanLLM">https://github.com/TCMAI-BJTU/LingdanLLM</a>
TCM Clinical Decision Support	Zhongjing <sup>100</sup>	2024	LLMs	Zhongjing significantly enhances the model's capability for complex dialogue and proactive inquiry initiation.	<a href="https://github.com/SupritYoung/Zhongjing">https://github.com/SupritYoung/Zhongjing</a>
TCM Clinical Decision Support	BianCang <sup>101</sup>	2024	LLMs	BianCang first injects domain-specific knowledge to enhance TCM diagnostic and differentiation capabilities.	<a href="https://github.com/QLU-NLP/BianCang">https://github.com/QLU-NLP/BianCang</a>
Comprehensive platform for TCM knowledge discovery	TCMKD <sup>102</sup>	2025	Association rule analysis, Bon-Kerbosch analysis, cluster analysis and decision tree analysis	TCMKD facilitate the exploration of relationships between different data categories, uncovering the intrinsic patterns within TCM data.	<a href="https://cbcb.edutcm.edu.cn/TCMKD/">https://cbcb.edutcm.edu.cn/TCMKD/</a>



**Fig. 4** Research framework of systems biological mechanism discovery.

vidual ingredients has increased emphasis on computational approaches, which have achieved substantial progress<sup>123</sup>. Li et al.<sup>124</sup> introduced the TCM network pharmacology framework, which analyzes molecular interactions between herbs and therapeutic targets to elucidate the biological mechanisms of herbs and herbal prescriptions<sup>125</sup>. Network pharmacology, a crucial component of systems biology, addresses the multi-component, multi-target aspects of pharmacology. Over the past decade, network pharmacology has been extensively applied in TCM research to investigate the pharmacological mechanisms of numerous herbal treatments and natural products<sup>126,127</sup>. Huang et al.<sup>128</sup> employed network pharmacology to identify five key compounds of Ephedra and determine targets associated with its anti-asthmatic effects. Additionally, Huang et al.<sup>129</sup> combined network pharmacology with experimental evaluation to investigate Huanglian Jiedu Decoction's mechanism in treating hepatocellular carcinoma. Recent developments include AI application platforms such as SYST-CM, developed in 2024, which utilizes deep learning and ML to identify pharmacological effects of TCM components<sup>93</sup>. A comparable platform is NP-TCMtarget<sup>94</sup>. Specialized network algorithms, including distance-based mutual information model (DMIM) and multi-layer co-modules, were developed to explore herb interactions within formulas and analyze heterogeneous networks connecting herbs, biomolecules, and diseases. These network pharmacology applications provide crucial mechanistic insights into herbal combinations and support modern TCM research. Given that herb-target interactions typically involve complex phenotype-genotype multi-scale networks, utilizing heterogeneous pharmacological networks is essential for obtaining high-quality representations of herbs and targets. Recent developments in heterogeneous link prediction and graph representation learning methods have enhanced herb-target predictions<sup>130-134</sup>.

Herbal medicine constitutes the foundation of TCM therapies. The historical lack of genomic and molecular information on traditional medicinal plants has presented a significant challenge to TCM research. Through multi-omics technologies, Chinese scientists have successfully decoded the genetic backgrounds of key medicinal species including *Perilla*<sup>135</sup> and *Coptis chinensis*<sup>136</sup>. These developments have accelerated TCM's integration into the modern "omics era", establishing a molecular foundation for deeper investigation of TCM resources and mechanisms. Multi-omics strategies have demonstrated widespread success in elucidating TCM mechanisms for treating various diseases, including cardiovascular, infectious, metabolic, and neurodegenerative conditions<sup>137</sup>. As omics datasets become increasingly large-scale, high-dimensional, and heterogeneous, often exceeding the capabilities of conventional statistical methods, AI emerges as the most effective analytical tool<sup>138,139</sup>. Li et al.<sup>140</sup> integrated high-resolution mass spectrometry with transcriptomics, metabolomics, bioinformatics, and molecular docking to identify connections between active ingredients in *M. tenacissima* extracts (MTE) and key targets of MTE's anti-tumor action. This research represents the first comprehensive investigation of anti-tumor molecular mechanisms while maintaining TCM's integrity. Similar research approaches will continue advancing TCM research<sup>141</sup>.

With the increasing global utilization of herbal medicines, concerns regarding their safety have intensified, primarily due to the limited availability of comprehensive pharmacological and toxicological data<sup>142</sup>. This has catalyzed the implementation of AI models to enhance the accuracy of predicting potential herbal medicine toxicity and evaluating risks associated with herb-drug interactions<sup>143-146</sup>. AI approaches encompass traditional ML models, advanced deep learning architectures, and graph neural networks (GNNs). Recent research demonstrates that incorporating multi-source datasets, including molecular geometric properties and electron ionization mass spectra (EI-MS) data, substantially

improves predictive accuracy for herbal-induced toxicities<sup>144,145</sup>. For accurate drug-herb interaction prediction, Lv et al.<sup>147</sup> developed a similarity matrix fusion (SMF) method that incorporates multi-source drug information to enhance drug-drug interaction network clustering, with the resulting networks utilized to predict novel interactions. Various deep learning models have been developed to identify latent patterns and achieve enhanced predictive accuracy by incorporating heterogeneous data sources such as chemical structures and pharmacological profiles<sup>148</sup>. GNNs have emerged as prominent deep learning frameworks<sup>149,150</sup>. By combining KG embedding methods with GNNs, researchers can capture global structural features of interaction networks, thereby improving drug-drug interaction prediction performance<sup>151</sup>. Collectively, these AI-driven approaches provide a robust and comprehensive framework for evaluating herbal medicine safety profiles.

### 3.2. Real-world clinical evidence inference

It is a fundamental AI task to extract and identify reliable TCM clinical evidence (e.g., diagnosis differentiation regularities or symptom clusters, herbal combination regularities, effective herbal treatment and herb-symptom associations) from real-world clinical data<sup>152</sup>. RWD incorporates EMR as a primary data source. Due to the complexity and various processing challenges associated with RWD<sup>91</sup>, generating high-quality clinical evidence requires several sequential key tasks. These tasks include clinical data warehouse (CDW), information extraction from EMR text, clinical data mining, and clinical evidence generation<sup>91,153</sup>. (Fig. 5 shows the research framework of real-world clinical evidence inference)

CDW focuses on establishing data management platforms to integrate and store large-scale clinical data. In 2010, a TCM CDW system was constructed containing 20 000 inpatient data and 20 000 outpatient data, representing a promising approach for large-scale clinical data integration and pre-processing tasks<sup>91,92,154</sup>. Clinical information extraction is essential for identifying relevant biomedical entities and their relationships within clinical data<sup>155</sup>. Several research efforts<sup>156-158</sup> have developed sophisticated methods for extracting and structuring clinical data. Their work encompasses automatic extraction of TCM clinical symptoms from EMR, development of a precision-focused system for TCM article data extraction, and creation of a human-machine cooperative phenotypic spectrum annotation system for biomedical named entity recognition (BioNER) from clinical texts. These advancements facilitate subsequent data analysis by providing well-organized, structured clinical data.

Clinical data mining methods including clustering, association rule mining, complex network analysis, and prognosis regression have been widely applied in TCM to uncover diagnosis differentiation patterns, symptom clusters, and herbal combination rules<sup>159,160</sup>. In diagnosis differentiation, clustering methods are extensively utilized to facilitate accurate syndrome differentiation. In 2008, Zhang et al.<sup>161</sup> developed a clustering method using latent tree models and proposed a syndrome differentiation standard for TCM, validated through a kidney deficiency study. Association rule mining identifies patterns of co-occurring symptoms or herbs, providing insights into herbal treatments and herb-symptom associations. Algorithms such as the FP-growth algorithm<sup>162</sup> and Apriori algorithm<sup>163</sup> have been implemented to identify herb associations, revealing core herbs and combination patterns in treating asthma<sup>164</sup>, chronic diseases<sup>165</sup>, chronic kidney disease<sup>166</sup>, lung cancer<sup>167</sup> and COVID-19<sup>168</sup>. Complex network analysis is crucial for mapping herbal interactions and optimizing herbal formulas. In 2009, a complex network analysis system for TCM clinical data was introduced<sup>169</sup>. This system incorporated network analysis components including shortest path

computation, vertex/edge betweenness, and vertex degree to extract core herb combinations. Prognostic regression models increasingly predict treatment outcomes. For example, Liu et al.<sup>170</sup> employed logistic regression modeling to examine correlations between TCM syndrome and short-term prognosis of ischemic stroke.

Real-world clinical research in TCM demonstrates significant potential for generating robust empirical evidence<sup>171</sup>. AI techniques enhance the processing and refinement of large datasets, establishing high-quality data sources for real-world clinical investigations. A notable study examining hospitalized patients with acute exacerbations of chronic obstructive pulmonary disease (COPD) analyzed over 7 000 inpatient cases and revealed that individualized CHM treatments substantially reduced mortality rates<sup>172</sup>. Additionally, semi-individualized CHM treatments administered as adjunct therapy for COVID-19 patients demonstrated decreased mortality rates<sup>173</sup>. These outcomes provide compelling clinical evidence supporting TCM's integration as a complementary approach in managing complex diseases and validate its therapeutic efficacy.

### 3.3. Personalized clinical decision support

Personalized clinical decision support constitutes a vital application domain of AITCM, primarily encompassing TCM syndrome differentiation/diagnosis and treatment planning/recommendation as fundamental tasks<sup>25,174</sup>. These tasks are typically addressed through supervised learning AI models or expert systems based on production rule inference. TCM's computational journey began with disease diagnosis, marked by China's first TCM expert system in 1979: Guan Youbo's computer program for

diagnosis and treatment of liver diseases<sup>23</sup>. In 2007, Chen et al.<sup>70</sup> developed the TCM e-Science system utilizing UTCMLS, providing semantically enhanced services including browsing, searching, querying, and knowledge discovery. In 2020, Long et al.<sup>175</sup> introduced the ONTOPV system, which implements a domain-specific ontology and employs fuzzy logic reasoning alongside case-based reasoning to facilitate syndrome differentiation and treatment of Psoriasis Vulgaris in TCM. (Fig. 6 shows the research framework of personalized clinical decision support)

TCM syndrome differentiation represents a distinctive prediction task in AITCM, as TCM practitioners consider syndrome diagnosis fundamental for herbal treatment. TCM syndrome diagnosis incorporates comprehensive information gathering through the Four Diagnostic Methods of "looking, listening, questioning, and feeling the pulse", along with the clinical expertise of TCM practitioners<sup>176</sup>. Recent advances in digitalization and instrumentation, particularly in tongue and pulse diagnosis<sup>177</sup>, have enabled the collection of heterogeneous data including images, texts, audio signals, and other sensory modalities, though these present considerable challenges for intelligent TCM diagnosis<sup>178</sup>. Since the 1990s, syndrome differentiation research has primarily employed traditional supervised learning models, such as support vector machine, decision tree, or Bayesian network, to extract decision rules from TCM physicians' clinical data or specific disease categories including coronary heart disease<sup>179,180</sup>, hypertension<sup>181</sup>, depression<sup>182</sup>, type 2 diabetes<sup>183</sup>, chronic hepatitis B<sup>184</sup>, and chronic fatigue<sup>185</sup>. Recent developments have introduced deep learning-based models and applications for TCM syndrome differentiations, demonstrating enhanced performance and integration of multi-modal data sources (e.g., pulse information from wearable devices and tongue images<sup>186</sup>) across

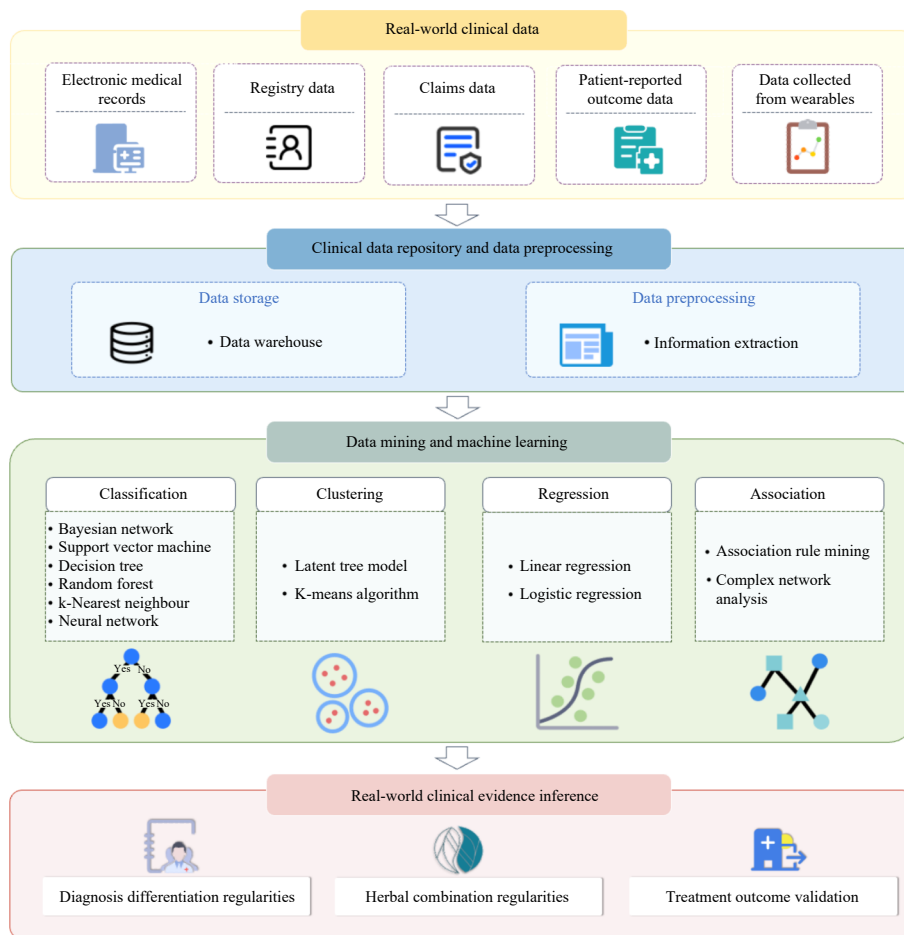


Fig. 5 Research framework of real-world clinical evidence inference.

various clinical conditions and diseases<sup>187-190</sup>. Multi-modal learning enables the utilization of complementary information from diverse modalities, enhancing diagnostic accuracy and reliability<sup>191</sup>. Given the rich clinical features in tongue pathologies, several studies have implemented deep learning models using tongue images for chronic disease diagnosis (e.g., gastric cancer<sup>192</sup>) and herbal prescriptions<sup>193</sup>. Dai et al.<sup>178</sup> developed a multi-modal deep learning framework that emulates TCM practitioners' cognitive processes by integrating multi-modal perceptions of "looking, listening, questioning, and feeling the pulse." This methodology combines representation learning from high-dimensional sensory data with conceptual alignment neural networks to incorporate prior knowledge, yielding interpretable latent features and superior diagnostic outcomes compared to uni-modal approaches.

Prescription recommendation or generation constitutes another crucial application domain. Various ML techniques integrate symptom information for prescription generation, consistent with TCM principles of prescribing herbal formulas based on individual symptom profiles. A groundbreaking study in 2018 introduced a prescription topic model based on TCM theory, representing an early approach to herbal recommendation and symptom suggestion<sup>194</sup>. In 2021, Zhou et al.<sup>195</sup> introduced FordNet, an intelligent prescription recommendation system utilizing deep learning. The system enhanced the hit ratio by 17.3% through the integration of phenotypic and molecular information, compared to models using only macroscopic data. The Attentive-Herb model<sup>196</sup> implemented an innovative two-stage attention mechanism for sequential CHM generation based on symptom inputs. Dong et al.<sup>197</sup> developed a subnetwork-based symptom term mapping (SSTM) method, leading to the traditional Chinese medicine prescription recommendation (TCMPR) model. This approach effectively utilizes underlying symptom feature information for more accurate prescription generation, advancing precision medicine applications in TCM.

The emergence of LLMs has created significant opportunities

for advancing TCM syndrome differentiation and treatment, particularly in clinical prescription recommendation. Several TCM-specific LLMs, including ShenNong-TCM<sup>99</sup>, BianCang<sup>101</sup> and Zhongjing<sup>100</sup>, have emerged, typically involving fine-tuned pre-trained models<sup>198</sup>. However, their clinical effectiveness requires additional validation. In 2024, our research team assessed LLMs' prescription recommendation capabilities using extensive real-world clinical data<sup>199</sup>, and developed the TCM pre-trained model Lingdan<sup>48</sup> by incorporating ancient TCM texts, KGs, and clinical case data. This development improved clinical prescription recommendation accuracy. Specifically, utilizing over 60 000 outpatient records for fine-tuning, the TCM-FTP prescription recommendation algorithm for spleen-stomach diseases achieved an F1@20 score exceeding 80%, demonstrating the potential for developing clinically viable TCM prescription recommendation systems<sup>197</sup>.

#### 4. Challenges and opportunities

Over the past three decades, AITCM has addressed numerous valuable topics and tasks related to biological mechanisms of syndrome and treatment, clinical evidence discovery, and clinical decision support through various traditional and emerging AI techniques. However, several factors limit the clinical applicability and performance of most AITCM applications and models, including insufficient high-quality datasets and benchmarks, incomplete KGs, complex TCM real-world clinical data and herbal therapies, and variable "Treatment based on Syndrome Differentiation" clinical procedures. This section examines several promising research directions that warrant systematic investigation to realize AITCM's full potential.

##### 4.1. High-quality data repository: building the data foundation for AITCM research

The primary challenge in AITCM development stems from

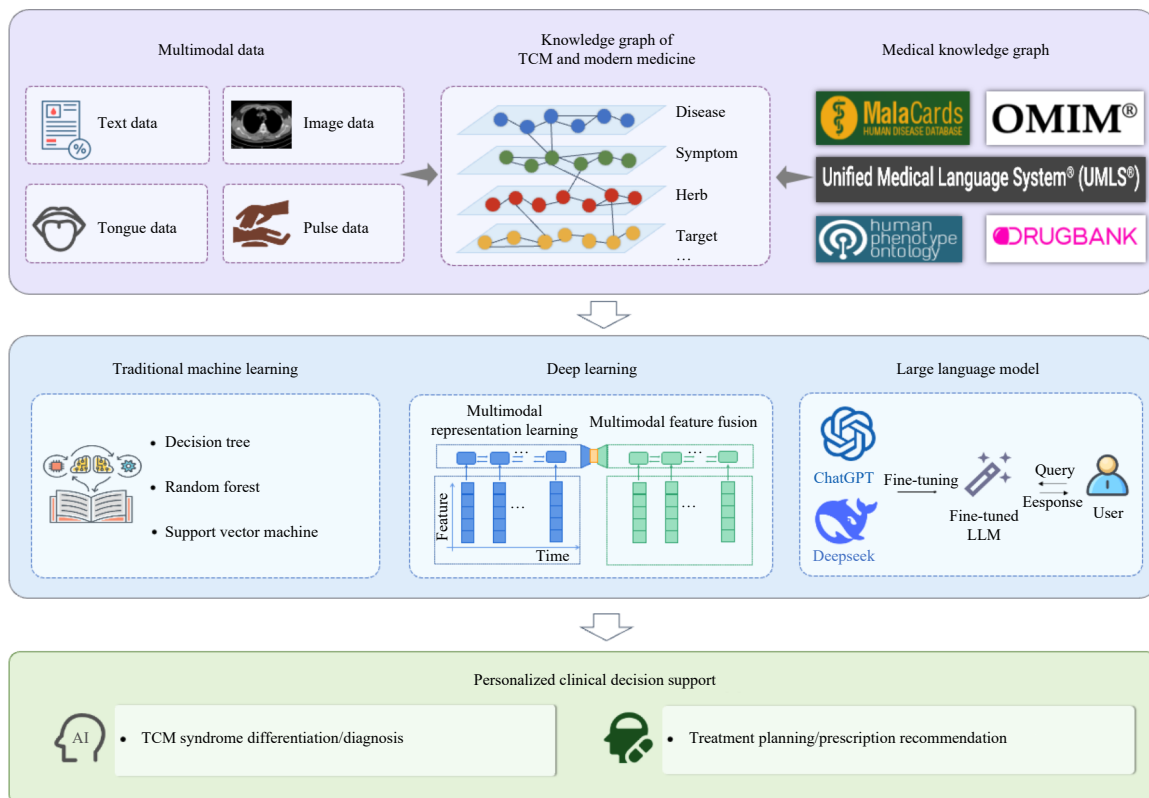


Fig. 6 Research framework of personalized clinical decision support.

the lack of high-quality, large-scale, comprehensive, and authoritative databases, which significantly impedes the advancement of reliable AI techniques. RWD sources frequently exhibit fragmentation, incompleteness, and heterogeneity, resulting in substantially lower data quality compared to controlled trials. Therefore, constructing high-quality databases from such diverse data sources remains an ongoing challenge.

A solid data foundation relies on three critical components: data collection, data governance, and standardized dataset development. Regarding data collection, ensuring efficient, accurate, and comprehensive clinical data capture is essential. Since January 2025, DeepSeek has achieved widespread adoption across tertiary hospitals in China<sup>200</sup>. Concurrently, AI-driven applications such as Freed<sup>201</sup> and Saykara<sup>202</sup> are emerging to capture multi-modal clinical data, particularly conversational voice data during medical encounters. Additionally, systems like Gator-Tron<sup>203</sup> utilize fine-grained clinical narratives to extract detailed patient information. These advancements collectively enable more comprehensive and precise clinical data acquisition, thereby enhancing data quality at the source and establishing the foundation for high-quality datasets.

Data governance serves a crucial role in enhancing data quality. The convergence of data management and ML has created new research opportunities in big data and AI integration. ML-Clean exemplifies this by combining traditional data cleaning, bias mitigation, and data sanitization within a unified framework to meet the requirements of accurate, fair, and robust model training in modern ML environments<sup>204</sup>. Additionally, data normalization in data management corresponds to entity linking in AI. DNorm<sup>205</sup> pioneered ML approaches to biomedical entity linking, and recent developments have shown promising results in encoding Chinese clinical terms<sup>206</sup>. This process is enhanced through the adoption of internationally recognized frameworks, such as the ISO/TC 249 standards for TCM terminology and coding, which provide structured definitions and classification schemes for key clinical entities<sup>207</sup>.

Furthermore, numerous data-centric, full-process management platforms have emerged recently<sup>208,209</sup>. Supported by national initiatives, several countries have initiated the development of healthcare big data platforms<sup>210</sup>. China has particularly emphasized the construction of healthcare big data infrastructure to address the challenges of integrating, processing, storing, and managing multi-source heterogeneous data systematically<sup>211</sup>. The West China Hospital of Sichuan University exemplifies this progress, having established a big data platform capable of effectively managing massive, multi-dimensional datasets from various sources, thereby creating accessible data assets with substantial healthcare application value<sup>212</sup>. Although these platforms represent significant progress in large-scale health data aggregation and management, advancing AI in TCM faces the persistent challenge of facilitating multi-institutional data sharing while maintaining stringent privacy and data protection. Federated learning presents a promising solution by enabling AI models to learn from diverse datasets across various locations without data leaving its original source, thus protecting privacy and reducing data transmission costs and latency<sup>213,214</sup>. This federated environment provides the advantage of model training on broader landscape data<sup>215</sup>. Consequently, this approach enhances model robustness and generalizability across institutions and geographic regions. The development of multi-center, publicly accessible big data platforms integrating data across hospital operation systems represents an inevitable future trend. Such platforms would facilitate high-quality dataset generation and massive, high-dimensional medical database development, supporting clinical practice, scientific research, and hospital management.

Beyond rigorous data acquisition and governance, emerging

generative AI techniques, particularly diffusion models, offer novel solutions to address persistent TCM dataset limitations, including small sample sizes, class imbalance, and incomplete multi-modal records<sup>216,217</sup>. These models learn the underlying dataset structure to generate new samples resembling original data, such as realistic synthetic clinical records and molecular structures for novel compound discovery<sup>218,219</sup>. Such synthetic data can enhance real-world datasets through multi-dimensional evaluation, including utility, fidelity, diversity, qualitative assessment, clinical validation, and privacy considerations, thereby enabling more robust model training and improving generalization capabilities<sup>220</sup>. Future efforts should incorporate multi-modal datasets, such as tongue images and pulse waveforms, and expand coverage to include both ancient and modern literature texts, advancing AITCM research and AI technology iteration. Generative AI facilitates multi-modal data integration by generating professional medical textual reports from medical images or synthesizing missing imaging data from textual records (e.g., PathLDM)<sup>221-224</sup>. In TCM, GADM represents the first diffusion model proposed for synthesizing pulse data. However, its application requires comprehensive assessment covering specific fidelity metrics, proper clinical validation and human assessment, and well-defined diversity metrics<sup>225</sup>. Incorporating generative AI into TCM dataset construction may thus play a crucial role in addressing data scarcity while maintaining AITCM research rigor.

#### 4.2. High-quality KG: integrating the TCM theoretical and experimental knowledge into AITCM research pipeline

High-quality KGs play a vital role in converting big data into actionable knowledge, thereby enhancing the effectiveness of clinical decision support systems and scientific discovery in AI applications<sup>226</sup>. The construction and utilization of KGs encompass several essential processes: knowledge acquisition (e.g., named entity recognition and relationship extraction), knowledge fusion (e.g., entity linking and alignment), knowledge complementation (e.g., inference of tacit knowledge), and knowledge application (e.g., intelligent question answering)<sup>227</sup>. The development of KGs represents a long-term, systematic endeavor, as demonstrated by foundational KGs such as WordNet<sup>228</sup>, Dbpedia<sup>229</sup>, UMLS<sup>63</sup>, and Gene Ontology (GO)<sup>230</sup>, which have undergone decades of continuous updates and refinements. However, a significant gap persists in developing high-quality, large-scale KGs in the field of TCM, both in terms of knowledge reliability and coverage. These may contain contradictory, erroneous, or incomplete knowledge<sup>231</sup>. For instance, recent TCM network pharmacology applications frequently face scrutiny regarding the non-specificity of associated targets, enriched biomedical pathways, and molecular functions. The primary contributing factor is the quality of network pharmacological databases. The ingredients of herbs and herbal compound-target associations in current network pharmacological databases require further evaluation and refinement.

Knowledge graph completion (KGC) has demonstrated effectiveness in enhancing the completeness, accuracy, and usability of KGs. To support this process, numerous KG reasoning algorithms have been proposed, such as TransE<sup>232</sup>, RGhat<sup>233</sup>, etc. Most current TCM KGs are based on explicit knowledge extracted from diverse sources, providing a solid foundation. However, much of TCM's value resides in tacit knowledge<sup>234</sup>, embedded in experiential reasoning and clinical decision-making accumulated by TCM experts. This knowledge is often individualized, implicit, and difficult to articulate. Capturing and representing such tacit knowledge remains a major challenge but is essential for developing comprehensive and clinically relevant KGs. Knowledge discovery methods have long been applied in medical information analysis<sup>235</sup>, with approaches such as decision trees employed to

uncover valid, novel, and interpretable patterns in tacit knowledge. A major challenge in this process is validation: expert confirmation is theoretically required<sup>236</sup>, yet individually validating each rule is impractical<sup>237</sup>. To address this, human-computer interactive strategies have been proposed to facilitate KG construction and enable experts to efficiently detect and correct inaccuracies<sup>238</sup>. Recent advances highlight the growing role of human-in-the-loop paradigms in improving KG quality. Tools such as Kyurem<sup>239</sup>, an interactive widget library, enable continuous curation by allowing experts to explore, validate, and integrate new knowledge efficiently. Similarly, frameworks like HC-COVID leverage hierarchical crowdsourcing to incorporate expertise at different levels, facilitating error detection and misinformation identification<sup>240</sup>. Together, these approaches are steering KG construction toward a more collaborative and intelligent paradigm. Beyond these, interactive machine learning (iML) integrates the strengths of human-computer interaction and knowledge discovery/data mining by embedding human feedback into the learning loop, enabling more effective, transparent, and adaptive KG construction<sup>241-243</sup>. In this process, human experts are not merely data annotators but active participants in data exploration and feature extraction, engaging with the model through intuitive interfaces and iterative feedback<sup>244-246</sup>. Such collaboration enhances both the accuracy and interpretability of AI systems, potentially improving the interpretability and clinical relevance of AI-assisted TCM applications.

Another aspect, an integrated TCM KG serves to formalize TCM knowledge and bridge macro-level (phenotypic) and micro-level (molecular) perspectives between TCM and modern medicine. As such, it establishes a solid knowledge foundation for tasks such as clinical causal inference, therapeutic mechanism elucidation, and evidence-based clinical decision-making. However, TCM data is typically multi-sourced and heterogeneous. The integration of multi-modal data, such as images, texts, and waveforms, into KGs has become a critical trend in precision medicine<sup>65,247</sup>, and is equally applicable and urgently needed in the context of TCM.

In the era of LLMs, the integration of KGs with these models has emerged as a cutting-edge research frontier. This synergy is poised to build interpretable, transparent, and reliable systems for new drug discovery and clinical decision support in TCM<sup>248</sup>. By leveraging the structured, semantically rich data embedded in KGs, covering drug properties, disease mechanisms, and clinical outcomes, and combining them with the advanced reasoning and generative capabilities of LLMs, researchers can gain deeper insights into TCM. This integration facilitates the identification of novel therapeutic targets, the prediction of treatment efficacy, and the personalization of TCM therapies. Furthermore, it enhances the interpretability of AI models, ensuring that their outputs are transparent, evidence-based and clinically meaningful. Ultimately, this convergence will promote the modernization and global acceptance of TCM, strengthening both its scientific foundation and clinical applicability.

#### 4.3. Biological mechanism of clinical effectiveness: development of the new paradigm of AI for science in TCM field

The fundamental scientific challenge in both clinical and basic TCM research centers on definitively establishing its therapeutic effectiveness, encompassing both clinical outcome validation and the investigation of underlying biological mechanisms. This necessitates analyzing the microscopic biological causal relationships between interventions, patient characteristics, and therapeutic outcomes. Such analysis requires identifying active substances or functional components within interventions, as well as the regulatory pathways through which they function to treat disease and restore systemic balance in the human

body<sup>249-251</sup>.

Herbal medicine remains the cornerstone therapeutic modality within TCM. Recent years have witnessed significant progress in herbal compound analysis, driven by improvements in sample pretreatment optimization, advanced analytical instrumentation development, and the integration of AI-based data processing with online bioactivity evaluation<sup>252</sup>. These developments have expanded both the scope and precision of compound analysis in TCM, enabling more thorough investigations into herbal medicines' complex constituents. Significantly, these advances have enhanced the quality and diversity of datasets available for AI models, providing an essential foundation for utilizing AI to elucidate TCM therapeutic mechanisms and accelerate their translation into clinical applications.

Modern biomedical technologies, including high-throughput multi-omics, single-cell analysis, and spatial transcriptomics, have created unprecedented opportunities and data resources for understanding the mechanistic basis of clinical efficacy<sup>253,254</sup>. Additionally, emerging computational biology methods, particularly network medicine<sup>255</sup> and network pharmacology<sup>124</sup>, enable researchers to examine disease-symptom relationships and prescription mechanisms through complex biological network interactions and dynamic system processes<sup>127,256</sup>. Combined with high-quality clinical data and KGs, these approaches provide novel perspectives for understanding TCM syndromes' biological foundations and CHM pharmacological mechanisms.

Future research must clearly define the roles of AI and experimental modules in TCM while maximizing their synergistic potential. Experimental studies provide AI models with validated, high-quality biological data, establishing a foundation for robust model development<sup>257</sup>. Conversely, AI generates data-driven hypotheses, reveals hidden patterns, and provides prior knowledge to guide experimental design and target selection<sup>258</sup>. Establishing an iterative feedback loop, where AI-driven predictions undergo experimental validation and subsequently refine the models, will enhance TCM research efficiency and precision, facilitating progression toward an integrated, mechanism-informed, data-driven discovery paradigm.

The integration of real-world clinical data with network medicine and the development of an AI-driven research framework focused on clinical effectiveness mechanism investigation shows considerable promise. Recent proof-of-concept studies demonstrate the viability of combining AI techniques with network-based approaches to explain efficacy within TCM's syndrome differentiation and treatment framework. For example, the mechanistic relationship between drugs and symptoms in TCM can be represented through network coupling between herbal target networks and symptom-related mechanism networks. The development of a network topology coupling index demonstrates strong predictive capability in assessing whether specific herbs may alleviate particular symptoms<sup>259</sup>. While these studies represent significant progress, they mark only the initial phase of AI-based exploration into clinical efficacy mechanisms.

A promising direction emerges with the establishment of large-scale, high-quality, closed-loop RWD resources and comprehensive KG infrastructures, where the integration of KGs between TCM and modern medicine will enable macro-to-micro knowledge domain linkage. Through the incorporation of network medicine, reinforcement learning, and advanced causal learning methodologies, the development of a dual-validation and optimization system for both clinical effectiveness mechanisms and therapeutic prescription evaluation becomes feasible. This approach will support the development of a clinically grounded, mechanistically informed AI system for TCM with iterative optimization capabilities. This integrative strategy represents a significant innovation at the intersection of AI and TCM, potentially enhancing scientific understanding and global acceptance of tradi-

tional practices.

#### 4.4. Causal evidence inference: finding reliable TCM clinical knowledge from real-world clinical data

The identification of high-quality causal evidence remains central to clinical research. Such evidence guides healthcare practitioners in decision-making and provides a foundation for translating RWD into actionable clinical insights and decisions<sup>260-262</sup>. However, current mainstream ML approaches primarily rely on statistical associations between features, with conditional probability underlying many deep learning methods, including LLMs<sup>263</sup>. Given RWD's complexity and heterogeneity, which often contains latent pseudo-correlations or inconsistent associations (e.g., Simpson's paradox<sup>264</sup>), existing AI techniques face challenges in reliably identifying true causal relationships, particularly in distinguishing confounding factors from genuine causality.

A significant challenge in AI-driven real-world clinical research on TCM lies in extracting high-quality clinical causal knowledge and robust clinical effectiveness evidence from heterogeneous and noisy clinical data. The development of novel causal learning methodologies (e.g., causal representation learning and counterfactual inference) has emerged as a crucial technical pathway for advancing causal evidence discovery in both general medicine and TCM clinical studies<sup>265,266</sup>. However, the individualized and holistic nature of TCM diagnosis and treatment presents additional complexities. Critical aspects such as the Four Diagnostics, etiology and pathogenesis, and therapeutic principles are frequently subjective or incompletely documented in clinical records<sup>267</sup>. Moreover, TCM interventions primarily involve herbal formulas comprising thousands of ingredients with complex pharmacological mechanisms across multiple biological targets and pathways.

This complexity creates a substantial barrier to identifying comprehensive causal pathways that span both macro (clinical outcome) and micro (biological mechanism) levels. Despite AI advancements, uncovering multi-layered causal evidence in TCM remains a significant scientific and computational challenge. Nevertheless, this complexity offers a unique experimental domain for developing and applying causal inference and representation learning techniques. Grounded in a systematic and networked medical paradigm, including theories of Yin-Yang, Five Elements, meridians, Zang-Fu organs, and Chinese materia medica, TCM demonstrates an intricate interplay between therapeutic interventions, physiological systems, and disease processes. These conceptual frameworks correspond to extensive, non-linear causal networks involving both abstract medical constructs and biochemical entities. Given the sparsity and incompleteness of many RWD datasets, discovering causal structures and assessing treatment effects in TCM necessitate not only novel algorithms but also substantial advances in computational efficiency and robustness.

Promising advances have emerged in this domain. For instance, in 2020, Richens et al.<sup>266</sup> developed a counterfactual-based algorithm and evaluated it against 44 physicians using 1671 clinical vignettes, achieving expert-level diagnostic accuracy—particularly for rare and very rare diseases. In 2022, Charpignon et al.<sup>268</sup> implemented a rigorous causal inference framework addressing competing risks in emulated clinical trials using data from two electronic health record systems. Their research indicated that metformin may reduce dementia risk in diabetic patients through non-glycemic mechanisms. These studies demonstrate the feasibility of discovering high-quality causal evidence within complex RWD environments. Additionally, the integration of KGs with causal learning represents an emerging research frontier. KGs can encode and provide rich prior know-

ledge structures, from herbal functions to disease mechanisms, potentially supporting causal discovery and effect estimation in TCM. This combination of structured knowledge and advanced causal inference techniques offers the potential to establish a novel and systematic framework for identifying and validating clinical effectiveness in real-world TCM studies<sup>269</sup>.

#### 4.5. Personalized clinical decision support: empowering TCM precision diagnosis and treatment with AI capabilities

Enhancing clinical diagnostic and therapeutic capabilities represents a fundamental goal for high-quality TCM development. Consequently, developing AI technologies for clinical decision support based on syndrome differentiation and treatment constitutes a critical focus in real-world clinical decision and services. Throughout the past five decades, numerous TCM syndrome diagnosis models have been developed using supervised learning techniques such as decision trees, Bayesian networks, and support vector machines<sup>161,270,271</sup>. Despite these developments, current AI models for syndrome differentiation remain restricted in scope, typically addressing fewer than 10 different syndromes, whereas real-world clinical practice encompasses hundreds of distinct syndromes, particularly in high-incidence chronic diseases such as diabetes, cardiovascular diseases, and cancer. Moreover, enhancing the accuracy, robustness, and generalizability of AI models in real-world, multi-center clinical settings remains a significant challenge for future research. Research on prescription recommendation has remained relatively limited, primarily due to TCM prescriptions' inherent complexity. A critical challenge in developing prescription recommendation algorithms is the scarcity of real-world clinical prescription datasets<sup>47</sup>.

AI models based on deep learning or even LLMs encounter significant challenges in clinical diagnostics and treatment. These include deficiencies in basic professional knowledge, limitations in real-world diagnostic capabilities, and critical constraints such as explainability and generalizability<sup>272</sup>. TCM clinical decision support systems face similar challenges. In an open clinical environment, practitioners require both accurate and reliable decision support, as well as a credible rationale for their decisions<sup>273,274</sup>. This dual requirement is essential for enhancing the efficiency and practicality of human-machine collaborative decision-making. TCM clinical decision-making is characterized by the long-term integration of individual patient data and theoretical knowledge. TCM KGs through the fusion of structured knowledge with empirical data demonstrate robust reasoning capabilities<sup>275</sup>. When combined with advanced reasoning technologies, their potential for clinical decision support is further enhanced<sup>276</sup>.

The inherent "black box" nature of deep learning algorithms has drawn increasing attention to the interpretability of the knowledge reasoning process. Significant advances have emerged in this domain, with techniques such as SHapley Additive exPlanations (SHAP) and attention mechanisms offering insights into model decision pathways<sup>277,278</sup>. For example, Lin et al.<sup>279</sup> developed a textual inference framework that effectively utilizes structured commonsense KGs alongside a hierarchical path-based attention mechanism, generating more transparent and trustworthy inferences. Interpretable models facilitate human-AI collaboration and enable practitioners to make better-informed decisions beyond simple class labels<sup>280</sup>. Additionally, enhanced interpretability promotes accountability and builds trust in ML systems. Thus, developing clinical decision support technologies based on high-quality real-world clinical datasets, enriched by clinical evidence and KGs, and designed with strong interpretability, shows significant potential for advancing medical AI through improved human-AI collaboration, informed decision-making, and trust. This approach remains fundamental for developing

reliable clinical decision-making tools and translating real-world TCM clinical evidence into personalized treatment strategies<sup>281,282</sup>.

For AI models in TCM to achieve true global applicability, validation across diverse ethnicities, geographic regions, and cultural contexts is essential. Evidence indicates that models trained primarily on datasets from a single population may significantly limit generalizability and risk producing biased results when applied to other demographic populations<sup>283-285</sup>. These disparities highlight the necessity for systematic cross-population evaluation. Methodological approaches, including transfer learning and domain adaptation, present viable solutions to enhance model generalizability while accounting for regional variations in clinical presentation and diagnostic patterns<sup>286-288</sup>. Implementation of these strategies will be crucial for the equitable, trustworthy, and effective globalization of AITCM in the future.

## 5. Conclusions

In the AI era, establishing a new data-driven and knowledge-enhanced paradigm for TCM clinical services and basic research is both promising and essential. The application of AI technology to analyze and optimize individualized TCM therapies, form precise syndrome classifications, and investigate the underlying system biological mechanisms of clinical effectiveness, addresses long-standing challenges in TCM. AI's extensive development history demonstrates that knowledge and data function as dual engines powering intelligent systems, integrating both deductive and inductive reasoning methodologies. Therefore, developing high-quality real-world clinical datasets and integrated KGs that bridge TCM and western medicine is crucial for creating a clinically applicable and scientifically efficient AITCM technological system. Moreover, TCM presents a complex and challenging domain for AI techniques and systems. The development of novel causal reasoning/learning algorithms and KG enhanced clinical decision support techniques represents a major technical requirement for AITCM. If these challenges are properly addressed, this historical opportunity for TCM through the AITCM approach will promote its high-quality development in the AI era.

## Funding

This work is supported by the National Key Research and Development Program (No. 2023YFC3502604), the National Natural Science Foundation of China (Nos. U23B2062, 82274352, 82174533, 82374302, 82204941), the Noncommunicable Chronic Diseases-National Science and Technology Major Project (No. 2023ZD0505700), the Beijing-Tianjin-Hebei Basic Research Cooperation Project (No. 22JZJXJC00070), the State Key Laboratory on Technologies for Chinese Medicine Pharmaceutical Process Control and Intelligent Manufacture (No. SKL2024Z0102), and Key R&D project of Ningxia Autonomous Region (No. 2022BEG02036).

## Supporting Information

Supporting information for this study is available upon request via email to the corresponding author.

## Declaration of competing interest

These authors have no conflict of interest to declare.

## References

- Tang J, Liu B, Ma K. Traditional Chinese medicine. *Lancet*. 2008;372(9654):1938-1940. [https://doi.org/10.1016/S0140-6736\(08\)61354-9](https://doi.org/10.1016/S0140-6736(08)61354-9).
- Li Shao, Xiao Wei, et al. General expert consensus on the application of network pharmacology in the research and development of new traditional Chinese medicine drugs. *Chin J Nat Med*. 2025;23(2):129-142. [https://doi.org/10.1016/S1875-5364\(25\)60802-8](https://doi.org/10.1016/S1875-5364(25)60802-8).
- Jiang WY. Therapeutic wisdom in traditional Chinese medicine: a perspective from modern science. *Trends Pharmacol Sci*. 2005;26(11):558-563. <https://doi.org/10.1016/j.tips.2005.09.006>.
- Qiu J. China plans to modernize traditional medicine. *Nature*. 2007;446(7136):590. <https://doi.org/10.1038/446590a>.
- Wei G, Wang M, Yang Y, et al. Combination of Chinese traditional and western medicine focuses on doctors and shows in patients. *J Innov Med Res*. 2023;2(9):24-28. <https://www.paradigmppress.org/jimr/article/view/786>.
- Xu Q, Bauer R, Hendry BM, et al. The quest for modernisation of traditional Chinese medicine. *BMC Complement Altern Med*. 2013;13(1):132. <https://doi.org/10.1186/1472-6882-13-132>.
- Wamba SF, Akter S, Edwards A, et al. How 'big data' can make big impact: findings from a systematic review and a longitudinal case study. *Int J Prod Econ*. 2015;165:234-246. <https://doi.org/10.1016/j.ijpe.2014.12.031>.
- Li S, Zhang B, Jiang D, et al. Herb network construction and co-module analysis for uncovering the combination rule of traditional Chinese herbal formulae. *BMC Bioinf*. 2010;11(11):S6. <https://doi.org/10.1186/1471-2105-11-S11-S6>.
- Gajewski A, Kośmider A, Nowacka A, et al. Potential of herbal products in prevention and treatment of COVID-19. Literature review. *Biomed Pharmacother*. 2021;143:112150. <https://doi.org/10.1016/j.biopha.2021.112150>.
- Andresen SL, John McCarthy: father of AI. *IEEE Intell Syst*. 2002;17(5):84-85. <https://doi.org/10.1109/MIS.2002.1039837>.
- Das S, Dey A, Pal A, et al. Applications of artificial intelligence in machine learning: review and prospect. *Int J Comput Appl*. 2015;115(9):31-41. <https://doi.org/10.5120/20182-2402>.
- Feigenbaum EA. Knowledge engineering: the applied side of artificial intelligence. In Proc. of a Symposium on Computer Culture: The Scientific, Intellectual, and Social Impact of the Computer. New York Academy of Sciences;1984;91-107. [10.5555/4959.4968](https://doi.org/10.5555/4959.4968).
- Brynjolfsson E, McAfee A. Artificial intelligence, for real. *Harv Bus Rev*. 2017;1(1):1-31. <https://starlab-alliance.com/wp-content/uploads/2017/09/AI-Article.pdf>.
- He J, Baxter SL, Xu J, et al. The practical implementation of artificial intelligence technologies in medicine. *Nat Med*. 2019;25(1):30-36. <https://doi.org/10.1038/s41591-018-0307-0>.
- Van Melle W. MYCIN: a knowledge-based consultation program for infectious disease diagnosis. *Int J Man Mach Stud*. 1978;10(3):313-322. [https://doi.org/10.1016/S0020-7373\(78\)80049-2](https://doi.org/10.1016/S0020-7373(78)80049-2).
- Lambin P, Leijenaar RT, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14(12):749-762. <https://doi.org/10.1038/nrclinonc.2017.141>.
- Singhal K, Azizi S, Tu T, et al. Large language models encode clinical knowledge. *Nature*. 2023;620(7972):172-180. <https://doi.org/10.1038/s41586-023-06291-2>.
- Nilanjan C, Shi JX, Montserrat GC. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat Rev Genet*. 2016;17(7):392-406. <https://doi.org/10.1038/nrg.2016.27>.
- Jin Q, Wang Z, Floudas CS, et al. Matching patients to clinical trials with large language models. *Nat Commun*. 2024;15(1):9074. <https://doi.org/10.1038/s41467-024-53081-z>.
- Hopkins AL. Network pharmacology: the next paradigm in drug discovery. *Nat Chem Biol*. 2008;4(11):682-690. <https://doi.org/10.1038/nchembio.118>.
- Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596(7873):583-589. <https://doi.org/10.1038/s41586-021-03819-2>.
- Lu Z, Peng Y, Cohen T, et al. Large language models in biomedicine and health: current research landscape and future directions. *J Am Med Inf Assoc*. 2024;31(9):1801-1811. <https://doi.org/10.1093/jamia/ocae202>.
- Zong X, Dai L. Analysis of 196 cases of liver disease treated by computer. *Liaoning J Tradit Chin Med*. 1992(06):26-27. <https://doi.org/10.13192/j.ljtc.1992.06.28.zongxq.016>.
- Ke Z, Liu M, Liu J, et al. The application of artificial intelligence in the research and development of traditional Chinese medicine. *Int J Drug Discov Pharmacol*. 2024;100001. <https://doi.org/10.53941/ijddp.2024.100001>.
- Hong Y, Zhu S, Liu Y, et al. The integration of machine learning into traditional Chinese medicine. *J Pharm Anal*. 2024;101157. <https://doi.org/10.1016/j.jpha.2024.101157>.
- Song Z, Chen G, Chen Y. AI empowering traditional Chinese medicine? *Chem Sci*. 2024;15(41):16844-16886. <https://doi.org/10.1039/D4SC04107K>.
- Johnson KB, Wei WQ, Weeraratne D, et al. Precision medicine, AI, and the future of personalized health care. *Clin Transl Sci*. 2021;14(1):86-93. <https://doi.org/10.1111/cts.12884>.
- Jurisica I. Explainable biology for improved therapies in precision medicine: AI is not enough. *Best Pr Res Clin Rheumatol*. 2024;102006. <https://doi.org/10.1016/j.berh.2024.102006>.
- Wishart DS, Knox C, Guo AC, et al. DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res*. 2008;36(suppl\_1):D901-D906. <https://doi.org/10.1093/nar/gkm958>.
- Gaulton A, Bellis LJ, Bento AP, et al. ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res*. 2012;40(D1):D1100-D1107. <https://doi.org/10.1093/nar/gkr777>.
- Chen CYC. TCM Database@Taiwan: the world's largest traditional Chinese medicine database for drug screening in silico. *PLoS One*. 2011;6(1):e15939. <https://doi.org/10.1371/journal.pone.0015939>.
- Xue R, Fang Z, Zhang M, et al. TCMID: traditional Chinese medicine

- integrative database for herb molecular mechanism analysis. *Nucleic Acids Res.* 2013;41(D1):D1089-D1095. <https://doi.org/10.1093/nar/gks1100>.
- 33 Ru J, Li P, Wang J, et al. TCMPSP: a database of systems pharmacology for drug discovery from herbal medicines. *J Cheminf.* 2014;6(1):13. <https://doi.org/10.1186/1758-2946-6-13>.
  - 34 Zhang R, Yu S, Bai H, et al. TCM-Mesh: the database and analytical system for network pharmacology analysis for TCM preparations. *Sci Rep.* 2017;7(1):2821. <https://doi.org/10.1038/s41598-017-03039-7>.
  - 35 Li B, Ma C, Zhao X, et al. YaTCM: yet another traditional Chinese medicine database for drug discovery. *Comput Struct Biotechnol J.* 2018;16:600-610. <https://doi.org/10.1016/j.csbj.2018.11.002>.
  - 36 Zhang Y, Li X, Shi Y, et al. ETCM v2.0: an update with comprehensive resource and rich annotations for traditional Chinese medicine. *Acta Pharm Sin B.* 2023;13(6):2559-2571. <https://doi.org/10.1016/j.apsb.2023.03.012>.
  - 37 Fang S, Dong L, Liu L, et al. HERB: a high-throughput experiment- and reference-guided database of traditional Chinese medicine. *Nucleic Acids Res.* 2021;49(D1):D1197-D1206. <https://doi.org/10.1093/nar/gkaa1063>.
  - 38 Lv Q, Chen G, He H, et al. TCMBank: bridges between the largest herbal medicines, chemical ingredients, target proteins, and associated diseases with intelligence text mining. *Chem Sci.* 2023;14(39):10684-10701. <https://doi.org/10.1039/D3SC02139D>.
  - 39 Yan D, Zheng G, Wang C, et al. HIT 2.0: an enhanced platform for herbal ingredients' targets. *Nucleic Acids Res.* 2022;50(D1):D1238-D1243. <https://doi.org/10.1093/nar/gkab1011>.
  - 40 Kong X, Liu C, Zhang Z, et al. BATMAN-TCM 2.0: an enhanced integrative database for known and predicted interactions between traditional Chinese medicine ingredients and target proteins. *Nucleic Acids Res.* 2024;52(D1):D1110-D1120. <https://doi.org/10.1093/nar/gkad926>.
  - 41 Sun C, Huang J, Tang R, et al. CPMCP: a database of Chinese patent medicine and compound prescription. *Database.* 2022;2022:baac073. <https://doi.org/10.1093/database/baac073>.
  - 42 Wu Y, Zhang F, Yang K, et al. SymMap: an integrative database of traditional Chinese medicine enhanced by symptom mapping. *Nucleic Acids Res.* 2019;47(D1):D1110-D1117. <https://doi.org/10.1093/nar/gky1021>.
  - 43 Li X, Ren J, Zhang W, et al. LTM-TCM: a comprehensive database for the linking of traditional Chinese medicine with modern medicine at molecular and phenotypic levels. *Pharmacol Res.* 2022;178:106185. <https://doi.org/10.1016/j.phrs.2022.106185>.
  - 44 Ren Z, Ren Y, Li Z, et al. TCMM: a unified database for traditional Chinese medicine modernization and therapeutic innovations. *Comput Struct Biotechnol J.* 2024;23:1619-1630. <https://doi.org/10.1016/j.csbj.2024.04.016>.
  - 45 Huang L, Wang Q, Duan Q, et al. TCMSDD: a comprehensive database focused on syndrome standardization. *Phytomedicine.* 2024;128:155486. <https://doi.org/10.1016/j.phymed.2024.155486>.
  - 46 Liu Z, Cai C, Du J, et al. TCMIO: a comprehensive database of traditional Chinese medicine on immuno-oncology. *Front Pharmacol.* 2020;11. <https://doi.org/10.3389/fphar.2020.00439>.
  - 47 Dong X, Zhao C, Song X, et al. PresRecST: a novel herbal prescription recommendation algorithm for real-world patients with integration of syndrome differentiation and treatment planning. *J Am Med Inf Assoc.* 2024;31(6):1268-1279. <https://doi.org/10.1093/jamia/ocae066>.
  - 48 Hua R, Dong X, Wei Y, et al. Lingdan: enhancing encoding of traditional Chinese medicine knowledge for clinical reasoning tasks with large language models. *J Am Med Inf Assoc.* 2024;31(9):2019-2029. <https://doi.org/10.1093/jamia/ocae087>.
  - 49 Kong Y, Hao M, Chen A, et al. SymMap database and TMNP algorithm reveal huangqi tongqiao granules for allergic rhinitis through IFN-mediated neuroimmuno-modulation. *Pharmacol Res.* 2022;185:106483. <https://doi.org/10.1016/j.phrs.2022.106483>.
  - 50 Yang R, Liu H, Bai C, et al. Chemical composition and pharmacological mechanism of qingfei paidu decoction and ma xing shi gan decoction against coronavirus disease 2019 (COVID-19): *in silico* and experimental study. *Pharmacol Res.* 2020;157:104820. <https://doi.org/10.1016/j.phrs.2020.104820>.
  - 51 Wang Y, Liu M, Jafari M, et al. A critical assessment of traditional Chinese medicine databases as a source for drug discovery. *Front Pharmacol.* 2024;15. <https://doi.org/10.3389/fphar.2024.1303693>.
  - 52 Li X, Zhang J, Shen X, et al. Overview and limitations of database in global traditional medicines: a narrative review. *Acta Pharmacol Sin.* 2025;46(2):235-263. <https://doi.org/10.1038/s41401-024-01353-1>.
  - 53 Almada M, Midão L, Portela D, et al. A new paradigm in health research: FAIR data (findable, accessible, interoperable, reusable). *Acta Med Port.* 2020;33(12):828-834. <https://doi.org/10.20344/amp.12910>.
  - 54 Morris JH, Soman K, Akbas RE, et al. The scalable precision medicine open knowledge engine (SPOKE): a massive knowledge graph of biomedical information. *Bioinformatics.* 2023;39(2):btad080. <https://doi.org/10.1093/bioinformatics/btad080>.
  - 55 Li X, Chen CH, Zheng P, et al. A knowledge graph-aided concept-knowledge approach for evolutionary smart product-service system development. *J Mech Des.* 2020;142(10):101403. <https://doi.org/10.1115/1.4046807>.
  - 56 Paulheim H. Knowledge graph refinement: a survey of approaches and evaluation methods. *Semant Web.* 2017;8(3):489-508. <https://doi.org/10.3233/SW-160218>.
  - 57 Ehrlinger L, Wöfl W. Towards a definition of knowledge graphs. *Semant Posters Demos Success.* 2016;48(1-4):2. [https://ceur-ws.org/Vol-1695/paper4.pdf?utm\\_source=gradientflow&utm\\_medium=newsletter&utm\\_campaign=issues20](https://ceur-ws.org/Vol-1695/paper4.pdf?utm_source=gradientflow&utm_medium=newsletter&utm_campaign=issues20)
  - 58 Suchanek FM, Kasneci G, Weikum G. Yago: a core of semantic knowledge. Proceedings of the 16th International Conference on World Wide Web. 2007;697-706. <https://doi.org/10.1145/1242572.1242667>
  - 59 Buitelaar P, Cimiano P, Frank A, et al. Ontology-based information extraction and integration from heterogeneous data sources. *Int J Hum Comput Stud.* 2008;66(11):759-788. <https://doi.org/10.1016/j.ijhcs.2008.07.007>.
  - 60 Gruber TR. Toward principles for the design of ontologies used for knowledge sharing? *Int J Hum Comput Stud.* 1995;43(5):907-928. <https://doi.org/10.1006/ijhc.1995.1081>
  - 61 Roldán ML, Gonnet S, Leone H. An ontology-based approach for sharing, integrating, and retrieving architectural knowledge. *Electron Notes Theor Comput Sci.* 2018;339:43-62. <https://doi.org/10.1016/j.entcs.2018.06.004>.
  - 62 Buccella A, Cechich A, Rodríguez BN. An ontology approach to data integration. *J Comput Sci Tech.* 2003;3(2):62-68. <http://sedici.unlp.edu.ar/handle/10915/9470>
  - 63 Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic Acids Res.* 2004;32(suppl\_1):D267-D270. <https://doi.org/10.1093/nar/gkh061>.
  - 64 Santos A, Colaço AR, Nielsen AB, et al. A knowledge graph to interpret clinical proteomics data. *Nat Biotechnol.* 2022;40(5):692-702. <https://doi.org/10.1038/s41587-021-01145-6>.
  - 65 Chandak P, Huang K, Zitnik M. Building a knowledge graph to enable precision medicine. *Sci Data.* 2023;10(1):67. <https://doi.org/10.1038/s41597-023-01960-3>.
  - 66 Zheng S, Rao J, Song Y, et al. PharmKG: a dedicated knowledge graph benchmark for biomedical data mining. *Brief Bioinf.* 2021;22(4):bbaa344. <https://doi.org/10.1093/bib/bbaa344>
  - 67 Yang Z. Biomedical information retrieval incorporating knowledge graph for explainable precision medicine. Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR '20. Association for Computing Machinery; 2020;2486. <https://doi.org/10.1145/3397271.3401458>
  - 68 Jaimini U, Sheth A. CausalKG: causal knowledge graph explainability using interventional and counterfactual reasoning. *IEEE Internet Comput.* 2022;26(1):43-50. <https://doi.org/10.1109/MIC.2021.3133551>.
  - 69 Lyu K, Tian Y, Shang Y, et al. Causal knowledge graph construction and evaluation for clinical decision support of diabetic nephropathy. *J Biomed Inf.* 2023;139:104298. <https://doi.org/10.1016/j.jbi.2023.104298>.
  - 70 Chen H, Mao Y, Zheng X, et al. Towards semantic e-Science for traditional Chinese medicine. *BMC Bioinf.* 2007;8(3):S6. <https://doi.org/10.1186/1471-2105-8-S3-S6>.
  - 71 Chu X, Sun B, Huang Q, et al. Quantitative knowledge presentation models of traditional Chinese medicine (TCM): a review. *Artif Intell Med.* 2020;103:101810. <https://doi.org/10.1016/j.artmed.2020.101810>.
  - 72 Zhou X, Wu Z, Yin A, et al. Ontology development for unified traditional Chinese medical language system. *Artif Intell Med.* 2004;32(1):15-27. <https://doi.org/10.1016/j.artmed.2004.01.014>.
  - 73 Gao M, Wang L, Cui M. Expression model for multiple relationships in the ontology of traditional Chinese medicine knowledge. *J Tradit Chin Med Sci.* 2016;3(1):59-65. <https://doi.org/10.1016/j.jtcms.2016.07.001>.
  - 74 Ji W, Zhang Y, Wang X, et al. Latent semantic diagnosis in traditional Chinese medicine. *World Wide Web.* 2017;20(5):1071-1087. <https://doi.org/10.1007/s11280-017-0443-3>.
  - 75 Cui M, Jia L, Yu T, et al. Current status of traditional Chinese medicine language system. In *Frontier and Future Development of Information Technology in Medicine and Education*. Netherlands: Springer; 2014:2287-2292. [https://doi.org/10.1007/978-94-007-7618-0\\_280](https://doi.org/10.1007/978-94-007-7618-0_280)
  - 76 Long H, Zhu Y, Jia L, et al. An ontological framework for the formalization, organization and usage of TCM-knowledge. *BMC Med Inform Decis Mak.* 2019;19(2):53. <https://doi.org/10.1186/s12911-019-0760-9>.
  - 77 Xing AL, Wang F, Liu JZ, et al. The prospect and underlying mechanisms of Chinese medicine in treating periodontitis. *Chin J Nat Med.* 2025;23(3):269-285. [https://doi.org/10.1016/S1875-5364\(25\)60842-9](https://doi.org/10.1016/S1875-5364(25)60842-9).
  - 78 Zhang Y, Wang N, Du X, et al. SoFDA: an integrated web platform from syndrome ontology to network-based evaluation of disease-syndrome-formula associations for precision medicine. *Sci Bull.* 2022;67(11):1097-1101. <https://doi.org/10.1016/j.scib.2022.03.013>.
  - 79 Shu Z, Hua R, Yan D, et al. ISPO: an integrated ontology of symptom phenotypes for semantic integration of traditional Chinese medical data. *Methods Inf Med.* 2025;63:164-175. <https://doi.org/10.1055/a-2576-1847>.
  - 80 Huan JM, Wang XJ, Li Y, et al. The biomedical knowledge graph of symptom phenotype in coronary artery plaque: machine learning-based analysis of real-world clinical data. *Biodata Min.* 2024;17(1):13. <https://doi.org/10.1186/s13040-024-00365-1>.
  - 81 Lu K, Yang K, Sun H, et al. SympGAN: a systematic knowledge integration system for symptom-gene associations network. *Knowl-Based Syst.* 2023;276:110752. <https://doi.org/10.1016/j.knsys.2023.110752>.
  - 82 Theodorou M, Fleckenstein J. The chinese black box—a scientific model of traditional Chinese medicine. *J Acupunct Res.* 2019;36(1):1-11. <https://doi.org/10.13045/jar.2018.00297>.
  - 83 Mohamed SK, Nouu A, Nováček V. Biological applications of knowledge graph embedding models. *Brief Bioinform.* 2021;22(2):1679-1693. <https://doi.org/10.1093/bib/bbaa012>.
  - 84 Gu Y, Wu GS, Li HK, et al. Strategy of systems biology for visualizing the "black box" of traditional Chinese medicine. *World J Tradit Chin Med.* 2020;6(3):260. [https://doi.org/10.4103/wjtc.wjtc\\_31\\_20](https://doi.org/10.4103/wjtc.wjtc_31_20).
  - 85 Tianyu C, Tingli N, Xin N, et al. Application of traditional Chinese medicine four-diagnostic auxiliary apparatus in evaluation of health status and clinical treatment. *J Tradit Chin Med.* 2018;38(3):447-451. [https://doi.org/10.1016/S0254-6272\(18\)30637-X](https://doi.org/10.1016/S0254-6272(18)30637-X).
  - 86 Q Chen. Analysis of TCM syndrome and treatment model based on artificial neural network. *Chin Arch Traditonal Chin Med.* 2009;27(07):1517-1520. <https://doi.org/10.13193/j.archctcm.2009.07.174.chenqw.010>.
  - 87 Lam CFD, Leung KS, Heng PA, et al. Chinese acupuncture expert system (CAES)—a useful tool to practice and learn medical acupuncture. *J Med Syst.*

- 2012;36(3):1883-1890. <https://doi.org/10.1007/s10916-010-9647-0>.
- 88 Qiao S, Tang C, Jin H, et al. KISTCM: knowledge discovery system for traditional Chinese medicine. *Appl Intell*. 2010;32(3):346-363. <https://doi.org/10.1007/s10489-008-0149-4>.
- 89 Zhang Z, Zhang Y, Yao L, et al. A sensor-based wrist pulse signal processing and lung cancer recognition. *J Biomed Inf*. 2018;79:107-116. <https://doi.org/10.1016/j.jbi.2018.01.009>.
- 90 Tian Z, Wang D, Sun X, et al. Current status and trends of artificial intelligence research on the four traditional Chinese medicine diagnostic methods: a scientometric study. *Ann Transl Med*. 2023;11(3):145. <https://doi.org/10.21037/atm-22-6431>.
- 91 Liu B, Zhou X, Wang Y, et al. Data processing and analysis in real-world traditional Chinese medicine clinical data: challenges and approaches. *Stat Med*. 2012;31(7):653-660. <https://doi.org/10.1002/sim.4417>.
- 92 Zhou X, Liu B, Wang Y, et al. Building clinical data warehouse for traditional Chinese medicine knowledge discovery. In: 2008 International Conference on BioMedical Engineering and Informatics. Vol 1.; 2008:615-620. <https://doi.org/10.1109/BMEI.2008.83>.
- 93 Wang Z, Huo M, Qiao L, et al. SYSTCM: a systemic web platform for objective identification of pharmacological effects based on interplay of "traditional Chinese medicine-components-targets". *Comput Biol Med*. 2024;179:108878. <https://doi.org/10.1016/j.compbiomed.2024.108878>.
- 94 Wang A, Peng H, Wang Y, et al. NP-TCMtarget: a network pharmacology platform for exploring mechanisms of action of traditional Chinese medicine. *Brief Bioinform*. 2025;26(1):bbaf078. <https://doi.org/10.1093/bib/bbaf078>.
- 95 Liu Y, Li X, Chen C, et al. TCMNPAS: a comprehensive analysis platform integrating network pharmacology and network pharmacology for exploring traditional Chinese medicine. *Chin Med*. 2024;19(1):50. <https://doi.org/10.1186/s13020-024-00924-y>.
- 96 Bu D, Xia Y, Zhang J, et al. FangNet: mining herb hidden knowledge from TCM clinical effective formulas using structure network algorithm. *Comput Struct Biotechnol J*. 2021;19:62-71. <https://doi.org/10.1016/j.csbj.2020.11.036>.
- 97 Gao W, Cheng N, Xin G, et al. TCM2Vec: a detached feature extraction deep learning approach of traditional Chinese medicine for formula efficacy prediction. *Multimed Tools Appl*. 2023;82(17):26987-27004. <https://doi.org/10.1007/s11042-023-14701-w>.
- 98 Niu Q, Li H, Tong L, et al. TCMFP: a novel herbal formula prediction method based on network target's score integrated with semi-supervised learning genetic algorithms. *Brief Bioinform*. 2023;24(3):bbad102. <https://doi.org/10.1093/bib/bbad102>.
- 99 Zhu W, Wang X. Shennong-tcm: a traditional Chinese medicine large language model. *Github*. <https://github.com/michael-wzhu/ShenNong-TCM-LLM>.
- 100 Yang S, Zhao H, Zhu S, et al. Zhongjing: enhancing the Chinese medical capabilities of large language model through expert feedback and real-world multi-turn dialogue. *Proc AAAI Conf Artif Intell*. 2024;38(17):19368-19376. <https://doi.org/10.1609/aaai.v38i17.29907>.
- 101 Wei S, Peng X, Wang Y, et al. BianCang: a traditional Chinese medicine large language model. *arXiv*. 2024. <https://doi.org/10.48550/arXiv.2411.11027>.
- 102 Xiao W, Zhang M, Zhao D, et al. TCMKD: from ancient wisdom to modern insights—a comprehensive platform for traditional Chinese medicine knowledge discovery. *J Pharm Anal*. 2025;15(6):101297. <https://doi.org/10.1016/j.jpba.2025.101297>.
- 103 Qiao C, Zhang HX, Tian XT, et al. Harnessing multi-omics approaches to elucidate the role of Chinese herbal compounds in chemotherapy-induced gastrointestinal damage. *World J Gastrointest Oncol*. 2025;17(2):101500. <https://doi.org/10.4251/wjgo.v17.i2.101500>.
- 104 Buriani A, Garcia-Bermejo ML, Bosisio E, et al. Omic techniques in systems biology approaches to traditional Chinese medicine research: present and future. *J Ethnopharmacol*. 2012;140(3):535-544. <https://doi.org/10.1016/j.jep.2012.01.055>.
- 105 Momozawa Y, Mizukami K. Unique roles of rare variants in the genetics of complex diseases in humans. *J Hum Genet*. 2021;66(1):11-23. <https://doi.org/10.1038/s10038-020-00845-2>.
- 106 Plenge RM, Scolnick EM, Altshuler D. Validating therapeutic targets through human genetics. *Nat Rev Drug Discov*. 2013;12(8):581-594. <https://doi.org/10.1038/nrd4051>.
- 107 Pun FW, Ozerov IV, Zhavoronkov A. AI-powered therapeutic target discovery. *Trends Pharmacol Sci*. 2023;44(9):561-572. <https://doi.org/10.1016/j.tips.2023.06.010>.
- 108 Patel MN, Halling-Brown MD, Tym JE, et al. Objective assessment of cancer genes for drug discovery. *Nat Rev Drug Discov*. 2013;12(1):35-50. <https://doi.org/10.1038/nrd3913>.
- 109 Sanseau P, Agarwal P, Barnes MR, et al. Use of genome-wide association studies for drug repositioning. *Nat Biotechnol*. 2012;30(4):317-320. <https://doi.org/10.1038/nbt.2151>.
- 110 Wang M, Wang H, Zheng H. A mini review of node centrality metrics in biological networks. *Int J Netw Dyn Intell*. 2022;1(1):99-110. <https://doi.org/10.53941/ijndi0101009>.
- 111 Chen Q, Zhang S, Jiang X, et al. The transcriptomic-based disease network reveals synergistic therapeutic effect of total alkaloids from *Coptis chinensis* and total ginsenosides from *Panax ginseng* on type 2 diabetes mellitus. *Chin J Nat Med*. 2025;23(8):997-1008. [https://doi.org/10.1016/S1875-5364\(24\)60689-8](https://doi.org/10.1016/S1875-5364(24)60689-8).
- 112 Yang K, Wang N, Liu G, et al. Heterogeneous network embedding for identifying symptom candidate genes. *J Am Med Inform Assoc JAMIA*. 2018; 25(11):1452-1459. <https://doi.org/10.1093/jamia/ocy117>.
- 113 Lu K, Yang K, Niyongabo E, et al. Integrated network analysis of symptom clusters across disease conditions. *J Biomed Inform*. 2020;107:103482. <https://doi.org/10.1016/j.jbi.2020.103482>.
- 114 Shu Z, Wang J, Sun H, et al. Diversity and molecular network patterns of symptom phenotypes. *npj Syst Biol Appl*. 2021;7(1):1-14. <https://doi.org/10.1038/s41540-021-00206-5>.
- 115 Li S. Network systems underlying traditional Chinese medicine syndrome and herb formula. *Curr Bioinforma*. 2009;4(3):188-196. <https://doi.org/10.2174/157489309789071129>.
- 116 Li R, Ma T, Gu J, et al. Imbalanced network biomarkers for traditional Chinese medicine syndrome in gastritis patients. *Sci Rep*. 2013;3(1):1543. <https://doi.org/10.1038/srep01543>.
- 117 Sun J, Zhu K, Zheng WJ, et al. A comparative study of disease genes and drug targets in the human protein interactome. *BMC Bioinformatics*. 2015;16(5):S1. <https://doi.org/10.1186/1471-2105-16-S5-S1>.
- 118 Li X, Liu ZQ, Liao J, et al. Network pharmacology approaches for research of Traditional Chinese Medicines. *Chin J Nat Med*. 2023;21(5):323-332. [https://doi.org/10.1016/S1875-5364\(23\)60429-7](https://doi.org/10.1016/S1875-5364(23)60429-7).
- 119 Thomford NE, Sentehebane DA, Rowe A, et al. Natural products for drug discovery in the 21st century: innovations for novel drug discovery. *Int J Mol Sci*. 2018;19(6):1578. <https://doi.org/10.3390/ijms19061578>.
- 120 Tu Y. The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine. *Nat Med*. 2011;17(10):1217-1220. <https://doi.org/10.1038/nm.2471>.
- 121 Howells LM, Berry DP, Elliott PJ, et al. Phase I randomized, double-blind pilot study of micronized resveratrol (SRT501) in patients with hepatic metastases—safety, pharmacokinetics, and pharmacodynamics. *Cancer Prev Res*. 2011;4(9):1419-1425. <https://doi.org/10.1158/1940-6207.CAPR-11-0148>.
- 122 Jakubikova J, Cervi D, Ooi M, et al. Anti-tumor activity and signaling events triggered by the isothiocyanates, sulforaphane and phenethyl isothiocyanate, in multiple myeloma. *Haematologica*. 2011;96(8):1170. <https://doi.org/10.3324/haematol.2010.029363>.
- 123 Liang X, Lai G, Yu J, et al. Herbal ingredient-target interaction prediction via multi-modal learning. *Inf Sci*. 2025;711:122115. <https://doi.org/10.1016/j.ins.2025.122115>.
- 124 Li S, Zhang B. Traditional Chinese medicine network pharmacology: theory, methodology and application. *Chin J Nat Med*. 2013;11(2):110-120. [https://doi.org/10.1016/S1875-5364\(13\)60037-0](https://doi.org/10.1016/S1875-5364(13)60037-0).
- 125 Liu Y, Li X, Chen C, et al. Exploration of compatibility rules and discovery of active ingredients in TCM formulas by network pharmacology. *Chin Herb Med*. 2024;16(4):572-588. <https://doi.org/10.1016/j.chmed.2023.09.008>.
- 126 Ding LL, Xu L, Hu N, et al. Deciphering the therapeutic potential and mechanisms of Artemisia argyrit essential oil on flagellum-mediated Salmonella infections. *Chin J Nat Med*. 2025;23(6):714-726. [https://doi.org/10.1016/S1875-5364\(25\)60890-9](https://doi.org/10.1016/S1875-5364(25)60890-9).
- 127 Wang X, Wang ZY, Zheng JH, et al. TCM network pharmacology: a new trend towards combining computational, experimental and clinical approaches. *Chin J Nat Med*. 2021;19(1):1-11. [https://doi.org/10.1016/S1875-5364\(21\)60001-8](https://doi.org/10.1016/S1875-5364(21)60001-8).
- 128 Huang XF, Cheng WB, Jiang Y, et al. A network pharmacology-based strategy for predicting anti-inflammatory targets of ephedra in treating asthma. *Int Immunopharmacol*. 2020;83:106423. <https://doi.org/10.1016/j.intimp.2020.106423>.
- 129 Huang J, Guo W, Cheung F, et al. Integrating network pharmacology and experimental models to investigate the efficacy of coptidis and scutellaria containing huanglian jiedu decoction on hepatocellular carcinoma. *Am J Chin Med*. 2020;48(1):161-182. <https://doi.org/10.1142/S0192415X20500093>.
- 130 Yang K, Liu G, Wang N, et al. Heterogeneous network propagation for herb target identification. *BMC Med Inform Decis Mak*. 2018;18(S1):17. <https://doi.org/10.1186/s12911-018-0592-z>.
- 131 Wang N, Li P, Hu X, et al. Herb target prediction based on representation learning of symptom related heterogeneous network. *Comput Struct Biotechnol J*. 2019;17:282-290. <https://doi.org/10.1016/j.csbj.2019.02.002>.
- 132 Duan P, Yang K, Su X, et al. HTNet2: herb-target prediction via knowledge graph embedding and residual-like graph neural network. *Brief Bioinform*. 2024;25(5):bbae414. <https://doi.org/10.1093/bib/bbae414>.
- 133 Zhao W, Wu H, He J. HGNA-HTI: heterogeneous graph neural network with attention mechanism for prediction of herb-target interactions. In: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE; 2021:3949-3956. <https://doi.org/10.1109/BIBM52615.2021.9669308>.
- 134 Zhu Y, Ren L, Sun R, et al. Herb-target interaction prediction by multi-instance learning. *IEEE Trans Artif Intell*. Published online 2024:1-10. <https://doi.org/10.1109/TAI.2024.3515932>.
- 135 Zhang Y, Shen Q, Leng L, et al. Incipient diploidization of the medicinal plant perilla within 10,000 years. *Nat Commun*. 2021;12(1):5508. <https://doi.org/10.1038/s41467-021-25681-6>.
- 136 Liu Y, Wang B, Shu S, et al. Analysis of the coptis chinensis genome reveals the diversification of protoberberine-type alkaloids. *Nat Commun*. 2021;12(1):3276. <https://doi.org/10.1038/s41467-021-23611-0>.
- 137 Guo R, Luo X, Liu J, et al. Omics strategies decipher therapeutic discoveries of traditional Chinese medicine against different diseases at multiple layers molecular-level. *Pharmacol Res*. 2020;152:104627. <https://doi.org/10.1016/j.phrs.2020.104627>.
- 138 Biswas N, Chakrabarti S. Artificial intelligence (AI)-based systems biology approaches in multi-omics data analysis of cancer. *Front Oncol*. 2020;10. <https://doi.org/10.3389/fonc.2020.588221>.
- 139 Yetgin A. Revolutionizing multi-omics analysis with artificial intelligence and data processing. *Quant Biol*. 2025;13(3):e70002. <https://doi.org/10.1002/qub.2.70002>.
- 140 Li S, Pei W, Yuan W, et al. Multi-omics joint analysis reveals the mechanism of action of the traditional Chinese medicine *marsdenia tenacissima* (Roxb)

- moon in the treatment of hepatocellular carcinoma. *J Ethnopharmacol.* 2022;293:115285. <https://doi.org/10.1016/j.jep.2022.115285>.
- 141 Chen W, Li Y, Zhang C, et al. Multi-omics and experimental validation reveal anti-HCC mechanisms of tibetan liuwei muxiang pill and quercetin. *Pharmaceuticals.* 2025;18(6):900. <https://doi.org/10.3390/ph18060900>.
- 142 Woo CSJ, Lau JSH, El-Nezami H. Herbal medicine: toxicity and recent trends in assessing their potential toxic effects. *Adv Bot Res.* 2012;62:365-384. <https://doi.org/10.1016/B978-0-12-394591-4.00009-X>
- 143 Spanakis M, Tzamali E, Tzedakis G, et al. Artificial intelligence models and tools for the assessment of drug-herb interactions. *Pharmaceuticals.* 2025;18(3):282. <https://doi.org/10.3390/ph18030282>.
- 144 Jia C, Li X, Hu S, et al. Advanced mass-spectra-based machine learning for predicting the toxicity of traditional Chinese medicines. *Anal Chem.* 2025;97(1):783-792. <https://doi.org/10.1021/acs.analchem.4c05311>.
- 145 Wu W, Qian J, Liang C, et al. GeoDILL: a robust and interpretable model for drug-induced liver injury prediction using graph neural network-based molecular geometric representation. *Chem Res Toxicol.* 2023;36(11):1717-1730. <https://doi.org/10.1021/acs.chemrestox.3c00199>.
- 146 Wang R, Liu Z, Gong J, et al. An uncertainty-guided deep learning method facilitates rapid screening of CYP3A4 inhibitors. *J Chem Inf Model.* 2023;63(24):7699-7710. <https://doi.org/10.1021/acs.jcim.3c01241>.
- 147 Lv J, Liu G, Ju Y, et al. Integrating multi-source drug information to cluster drug-drug interaction network. *Comput Biol Med.* 2023;162:107088. <https://doi.org/10.1016/j.combiomed.2023.107088>.
- 148 Li X, Xiong Z, Zhang W, et al. Deep learning for drug-drug interaction prediction: a comprehensive review. *Quant Biol.* 2024;12(1):30-52. <https://doi.org/10.1002/qub.2.32>.
- 149 Pang S, Zhang Y, Song T, et al. AMDE: a novel attention-mechanism-based multidimensional feature encoder for drug-drug interaction prediction. *Brief Bioinform.* 2022;23(1):1. <https://doi.org/10.1093/bib/bbab545>.
- 150 Zhu X, Shen Y, Lu W. Molecular substructure-aware network for drug-drug interaction prediction. In: Proceedings of the 31st ACM International Conference on Information & Knowledge Management. CIKM '22. Association for Computing Machinery; 2022:4757-4761. <https://doi.org/10.1145/3511808.3557648>
- 151 Ren ZH, You ZH, Yu CQ, et al. A biomedical knowledge graph-based method for drug-drug interactions prediction through combining local and global features with deep neural networks. *Brief Bioinform.* 2022;23(5):bbac363. <https://doi.org/10.1093/bib/bbac363>.
- 152 Bartlett VL, Dhruva SS, Shah ND, et al. Feasibility of using real-world data to replicate clinical trial evidence. *JAMA Netw Open.* 2019;2(10):e1912869. <https://doi.org/10.1001/jamanetworkopen.2019.12869>.
- 153 Zhou X, Peng Y, Liu B. Text mining for traditional Chinese medical knowledge discovery: a survey. *J Biomed Inform.* 2010;43(4):650-660. <https://doi.org/10.1016/j.jbi.2010.01.002>.
- 154 Zhou X, Chen S, Liu B, et al. Development of traditional Chinese medicine clinical data warehouse for medical knowledge discovery and decision support. *Artif Intell Med.* 2010;48(2-3):139-152. <https://doi.org/10.1016/j.artmed.2009.07.012>.
- 155 Luo ZW, Yin FC, Wang XB, et al. Progress in approved drugs from natural product resources. *Chin J Nat Med.* 2024;22(3):195-211. [https://doi.org/10.1016/S1875-5364\(24\)60582-0](https://doi.org/10.1016/S1875-5364(24)60582-0).
- 156 Zou Q, Yang K, Shu Z, et al. Phenonizer: a fine-grained phenotypic named entity recognizer for Chinese clinical texts. *Biomed Res Int.* 2022;2022(1):3524090. <https://doi.org/10.1155/2022/3524090>.
- 157 Du Z, Tang D, Xie D. Automatic extraction of clinical symptoms in traditional Chinese medicine for electronic medical records. In: 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2021: 3784-3790. <https://doi.org/10.1109/BIBM52615.2021.9669345>
- 158 Xia Y, Cai J, Li Y, et al. A precision-preferred comprehensive information extraction system for clinical articles in traditional Chinese Medicine. *Int J Intell Syst.* 2022;37(8):4994-5010. <https://doi.org/10.1002/int.22748>.
- 159 Jafari M, Wang Y, Amiryousefi A, et al. Unsupervised learning and multipartite network models: a promising approach for understanding traditional medicine. *Front Pharmacol.* 2020;11. <https://doi.org/10.3389/fphar.2020.01319>
- 160 Denny JC, Collins FS. Precision medicine in 2030—seven ways to transform healthcare. *Cell.* 2021;184(6):1415-1419. <https://doi.org/10.1016/j.cell.2021.01.015>.
- 161 Zhang NL, Yuan S, Chen T, et al. Latent tree models and diagnosis in traditional Chinese medicine. *Artif Intell Med.* 2008;42(3):229-245. <https://doi.org/10.1016/j.artmed.2007.10.004>.
- 162 Xu F, Lu H. The application of FP-growth algorithm based on distributed intelligence in wisdom medical treatment. *Int J Pattern Recognit Artif Intell.* 2017;31(04):1759005. <https://doi.org/10.1142/S0218001417590054>.
- 163 Li JF, Guo HL, Dong Y, et al. Polysaccharides from Chinese herbal medicine: a review on the hepatoprotective and molecular mechanism. *Chin J Nat Med.* 2024;22(1):4-14. [https://doi.org/10.1016/S1875-5364\(24\)60558-3](https://doi.org/10.1016/S1875-5364(24)60558-3).
- 164 Hu Y, Wang MQ, Xie J, et al. Exposure to ephedrine attenuates Th1/Th2 imbalance underlying OVA-induced asthma through airway epithelial cell-derived exosomal lnc-TRPM2-AS. *Chin J Nat Med.* 2024;22(6):530-540. [https://doi.org/10.1016/S1875-5364\(24\)60554-6](https://doi.org/10.1016/S1875-5364(24)60554-6).
- 165 Wang R, Li J, Wang Y. Research on the medication regularity of traditional Chinese medicine for common chronic diseases based on association rules. In: Proceedings of the 2022 5th International Conference on Algorithms, Computing and Artificial Intelligence. ACAI '22. Association for Computing Machinery; 2023:1-6. <https://doi.org/10.1145/3579654.3579664>
- 166 Lin YY, Wei JL, Zhang YH, et al. Shen Qi Wan attenuates renal interstitial fibrosis through upregulating AQP1. *Chin J Nat Med.* 2023;21(5):359-370. [https://doi.org/10.1016/S1875-5364\(23\)60453-4](https://doi.org/10.1016/S1875-5364(23)60453-4).
- 167 Wang M, Li J, Chen L, et al. The study of the compatibility rules of traditional Chinese medicine based on apriori and HMETIS hypergraph partitioning algorithm. In: Biomedical Data Management and Graph Online Querying. Springer International Publishing; 2016:16-31. <https://doi.org/10.1007/978-3-319-41576-2>
- 168 Zheng Y, Chen Y. The identification of Chinese herbal medicine combination association rule analysis based on an improved apriori algorithm in treating patients with COVID-19 disease. *J Healthc Eng.* 2022;2022(1):6337082. <https://doi.org/10.1155/2022/6337082>.
- 169 Zhou X, Liu B. Network analysis system for traditional Chinese medicine clinical data. In: 2009 2nd International Conference on Biomedical Engineering and Informatics; 2009:1-5. <https://doi.org/10.1109/BMEI.2009.5302924>
- 170 Liu L, Gao Y. Study on the correlation between traditional Chinese medicine syndrome and short-term prognosis of ischemic stroke using logistic regression model and repeated-measures analysis of variance. *J Chin Integr Med.* 2012;10(9):983-990. <https://doi.org/10.3736/jcim20120906>.
- 171 Egger M, Moons KGM, Fletcher C, et al. GetReal: from efficacy in clinical trials to relative effectiveness in the real world. *Res Synth Methods.* 2016;7(3):278-281. <https://doi.org/10.1002/jrsm.1207>.
- 172 Xu N, Zhong K, Yu H, et al. Add-on Chinese medicine for hospitalized chronic obstructive pulmonary disease (CHOP): a cohort study of hospital registry. *Phytomedicine.* 2023;109:154586. <https://doi.org/10.1016/j.phymed.2022.154586>.
- 173 Shu Z, Chang K, Zhou Y, et al. Add-on Chinese medicine for coronavirus disease 2019 (ACCORD): a retrospective cohort study of hospital registries. *Am J Chin Med.* 2021;49(03):543-575. <https://doi.org/10.1142/S0192415X21500257>.
- 174 Song Y, Ma S, Dai Y, et al. AI-assisted TCM syndrome differentiation: key issues and technical challenges. *Strateg Study Chin Acad Eng.* 2024;26(2):234-244. <https://doi.org/10.15302/j-SSCAE-2024.02.010>.
- 175 Long H, Wang Z, Cui Y, et al. Towards an ontology-based decision support system for syndrome-differentiation and treatment of psoriasis vulgaris in traditional Chinese medicine. *Res Sq.* Published online 2020. <https://doi.org/10.21203/rs.3.rs-38383/v2>
- 176 Wang Z, Wang D, Liu W, et al. Traditional Chinese medicine diagnosis and treatment based on systematics. *Iliver.* 2023;2(4):181-187. <https://doi.org/10.1016/j.iliver.2023.08.004>.
- 177 Zhao Y, Sun Q, Mei S, et al. Wearable multichannel-active pressurized pulse sensing platform. *Microsyst Nanoeng.* 2024;10(1):77. <https://doi.org/10.1038/s41378-024-00703-7>.
- 178 Dai Y, Wang G, Dai J, et al. A multimodal deep architecture for traditional Chinese medicine diagnosis. *Concurr Comput Pract Exp.* 2020;32(19):e5781. <https://doi.org/10.1002/cpe.5781>.
- 179 Liu GP, Li GZ, Wang YL, et al. Modelling of inquiry diagnosis for coronary heart disease in traditional Chinese medicine by using multi-label learning. *BMC Complement Altern Med.* 2010;10(1):37. <https://doi.org/10.1186/1472-6882-10-37>.
- 180 Wang J, He Q, Yao K, et al. Support vector machine (SVM) and traditional Chinese medicine: syndrome factors based an SVM from coronary heart disease treated by prominent traditional Chinese medicine doctors. In: 2009 Fifth International Conference on Natural Computation. 2009. 2: 176-180. <https://doi.org/10.1109/ICNC.2009.735>
- 181 Ouyang WW, Lin X, Ren Y, et al. TCM syndromes diagnostic model of hypertension: study based on tree augmented naive bayes. In: 2011 IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW). 2011:834-837. <https://doi.org/10.1109/BIBMW.2011.6112481>
- 182 Zhu W, Yan J, Huang B. Application of bayesian network in syndrome differentiation system of traditional Chinese medicine. *J Chin Integr Med.* 2006;4(6):567-571. <https://doi.org/10.3736/jcim20060604>.
- 183 Feng YT, Liu J, Gong L, et al. Inonotus obliquus (Chaga) against HFD/STZ-induced glucolipid metabolism disorders and abnormal renal functions by regulating NOS-cGMP-PDE5 signaling pathway. *Chin J Nat Med.* 2024;22(7):619-631. [https://doi.org/10.1016/S1875-5364\(24\)60616-3](https://doi.org/10.1016/S1875-5364(24)60616-3).
- 184 Chen X, Ma L, Chu N, et al. Diagnosis based on decision tree and discrimination analysis for chronic hepatitis B in TCM. In: 2011 IEEE International Conference on Bioinformatics and Biomedicine Workshops (BIBMW). 2011:817-822. <https://doi.org/10.1109/BIBMW.2011.6112478>
- 185 Wang H, Liu X, Lv B, et al. Reliable multi-label learning via conformal predictor and random forest for syndrome differentiation of chronic fatigue in traditional Chinese medicine. *PLoS One.* 2014;9(6):e99565. <https://doi.org/10.1371/journal.pone.0099565>.
- 186 Xu J, Jiang T, Liu S. Research status and prospect of tongue image diagnosis analysis based on machine learning. *Digit Chin Med.* 2024;7(1):3-12. <https://doi.org/10.1016/j.dcm.2024.04.002>.
- 187 Chen Z, Zhang D, Liu C, et al. Traditional Chinese medicine diagnostic prediction model for holistic syndrome differentiation based on deep learning. *Integr Med Res.* 2024;13(1):101019. <https://doi.org/10.1016/j.imr.2023.101019>.
- 188 Hu Q, Yu T, Li J, et al. End-to-end syndrome differentiation of Yin deficiency and Yang deficiency in traditional Chinese medicine. *Comput Methods Programs Biomed.* 2019;174:9-15. <https://doi.org/10.1016/j.cmpb.2018.10.011>.
- 189 Huang Z, Miao J, Chen J, et al. A traditional Chinese medicine syndrome classification model based on cross-feature generation by convolution neural network: model development and validation. *JMIR Med Inform.* 2022;10(4):e29290. <https://doi.org/10.2196/29290>.
- 190 Xu Q, Guo Q, Wang CX, et al. Network differentiation: a computational method of pathogenesis diagnosis in traditional Chinese medicine based on systems science. *Artif Intell Med.* 2021;118:102134. <https://doi.org/10.1016/j.artmed.2021.102134>.
- 191 Waqas A, Tripathi A, Ramachandran RP, et al. Multimodal data integration for oncology in the era of deep neural networks: a review. *Front Artif Intell.* 2024;7. <https://doi.org/10.3389/frai.2024.1408843>

- 192 Yuan L, Yang L, Zhang S, et al. Development of a tongue image-based machine learning tool for the diagnosis of gastric cancer: a prospective multicentre clinical cohort study. *eClinicalMedicine*. 2023;57. <https://doi.org/10.1016/j.eclinm.2023.101834>
- 193 Hu Y, Wen G, Liao H, et al. Automatic construction of Chinese herbal prescriptions from tongue images using CNNs and auxiliary latent therapy topics. *IEEE Trans Cybern*. 2021;51(2):708-721. <https://doi.org/10.1109/TCYB.2019.2909925>
- 194 Yao L, Zhang Y, Wei B, et al. A topic modeling approach for traditional Chinese medicine prescriptions. *IEEE Trans Knowl Data Eng*. 2018;30(6):1007-1021. <https://doi.org/10.1109/TKDE.2017.2787158>
- 195 Zhou W, Yang K, Zeng J, et al. FordNet: recommending traditional Chinese medicine formula via deep neural network integrating phenotype and molecule. *Pharmacol Res*. 2021;173:105752. <https://doi.org/10.1016/j.phrs.2021.105752>
- 196 Liu Z, Zheng Z, Guo X, et al. AttentiveHerb: a novel method for traditional Chinese medicine prescription generation. *IEEE Access*. 2019;7:139069-139085. <https://doi.org/10.1109/ACCESS.2019.2941503>
- 197 Dong X, Zheng Y, Shu Z, et al. TCMPr: TCM prescription recommendation based on subnetwork term mapping and deep learning. *Biomed Res Int*. 2022;2022(1):4845726. <https://doi.org/10.1155/2022/4845726>
- 198 Liu L, Yang X, Lei J, et al. A survey on medical large language models: technology, application, trustworthiness, and future directions. *arXiv*. 2024. <https://doi.org/10.48550/arXiv.2406.03712>
- 199 Tian H, Yang K, Dong X, et al. TCMLLM-PR: evaluation of large language models for prescription recommendation in traditional Chinese medicine. *Digit Chin Med*. 2024;7(4):343-355. <https://doi.org/10.1016/j.dcm.2025.01.007>
- 200 Chen J, Miao C. DeepSeek deployed in 90 chinese tertiary hospitals: how artificial intelligence is transforming clinical practice. *J Med Syst*. 2025;49(1):53. <https://doi.org/10.1007/s10916-025-02181-4>
- 201 Mess SA, Mackey AJ, Yarowsky DE. Artificial intelligence scribe and large language model technology in healthcare documentation: advantages, limitations, and recommendations. *Plast Reconstr Surg - Glob Open*. 2025;13(1):e6450. <https://doi.org/10.1097/GOX.00000000000006450>
- 202 Akinwande V, Cintas C, Speakman S, et al. Identifying audio adversarial examples via anomalous pattern detection. *arXiv*. 2020. <https://doi.org/10.48550/arXiv.2002.05463>
- 203 Yang X, Chen A, PourNejatian N, et al. A large language model for electronic health records. *npj Digit Med*. 2022;5(1):1-9. <https://doi.org/10.1038/s41746-022-00742-2>
- 204 Tae KH, Roh Y, Oh YH, et al. Data cleaning for accurate, fair, and robust models: a big data - AI integration approach. In: Proceedings of the 3rd International Workshop on Data Management for End-to-End Machine Learning. DEEM'19. Association for Computing Machinery; 2019:1-4. <https://doi.org/10.1145/3329486.3329493>
- 205 Leaman R, Islamaj Doğan R, Lu Z. DNorm: disease name normalization with pairwise learning to rank. *Bioinformatics*. 2013;29(22):2909-2917. <https://doi.org/10.1093/bioinformatics/btt474>
- 206 Wang Q, Ji Z, Wang J, et al. A study of entity-linking methods for normalizing Chinese diagnosis and procedure terms to ICD codes. *J Biomed Inf*. 2020;105:103418. <https://doi.org/10.1016/j.jbi.2020.103418>
- 207 Xu X, Shi Y, Sang Z, et al. Features and development trends in international standardization of Chinese materia medica in ISO/TC 249. *Pharmacol Res*. 2021;167:105519. <https://doi.org/10.1016/j.phrs.2021.105519>
- 208 Semenov I, Osenev R, Gerasimov S, et al. Experience in developing an FHIR medical data management platform to provide clinical decision support. *Int J Env Res Public Health*. 2020;17(1):73. <https://doi.org/10.3390/ijerph17010073>
- 209 Yang C, Liu J, Chen S, et al. Implementation of a big data accessing and processing platform for medical records in cloud. *J Med Syst*. 2017;41(10):149. <https://doi.org/10.1007/s10916-017-0777-5>
- 210 Štufi M, Bačić B, Stoimenov L. Big data analytics and processing platform in Czech Republic healthcare. *Appl Sci Basel*. 2020;10(5):1705. <https://doi.org/10.3390/app10051705>
- 211 Zhou G, Xu X, Zhang X, et al. Design of data governance system based on national health information platform construction. *Chin J Health Inf Manag*. 2019;16(2):131-134. <https://doi.org/10.3969/j.issn.1672-5166.2019.02.02>
- 212 Wang M, Li S, Zheng T, et al. Big data health care platform with multisource heterogeneous data integration and massive high-dimensional data governance for large hospitals: design, development, and application. *JMIR Med Inf*. 2022;10(4):e36481. <https://doi.org/10.2196/36481>
- 213 Li N, Lewin A, Ning S, et al. Privacy-preserving federated data access and federated learning: improved data sharing and AI model development in transfusion medicine. *Transfusion (Paris)*. 2025;65(1):22-28. <https://doi.org/10.1111/trf.18077>
- 214 Rieke N, Hancox J, Li W, et al. The future of digital health with federated learning. *npj Digit Med*. 2020;3(1):119. <https://doi.org/10.1038/s41746-020-00323-1>
- 215 Aouedi O, Sacco A, Piamrat K, et al. Handling privacy-sensitive medical data with federated learning: challenges and future directions. *IEEE J Biomed Health Inform*. 2023;27(2):790-803. <https://doi.org/10.1109/JBHI.2022.3185673>
- 216 Huang H, Chen R, Lin Y, et al. Research on chronic kidney disease staging prediction and prescription recommendation model based on multimodal data fusion. In: 2025 2nd International Conference on Electronic Engineering and Information Systems (EEISS). 2025:1-5. <https://doi.org/10.1109/EEISS65394.2025.11085645>
- 217 Zhan Z, Qinghua P, Xiaoxia X, et al. An interpretability model for syndrome differentiation of HBV-ACLF in traditional Chinese medicine using small-sample imbalanced data. *Digit Chin Med*. 2024;7(2):137-147. <https://doi.org/10.1016/j.dcm.2024.09.005>
- 218 Yan Y, Li C, Huang Y, et al. TCDiff: triplex cascaded diffusion for high-fidelity multimodal EHRs generation with incomplete clinical data. *arXiv*. 2025. <https://doi.org/10.48550/arXiv.2508.01615>
- 219 Yim J, Stärk H, Corso G, et al. Diffusion models in protein structure and docking. *WIREs Comput Mol Sci*. 2024;14(2):e1711. <https://doi.org/10.1002/wcms.1711>
- 220 Ibrahim M, Khalil YA, Amirrajab S, et al. Generative AI for synthetic data across multiple medical modalities: a systematic review of recent developments and challenges. *Comput Biol Med*. 2025;189:109834. <https://doi.org/10.1016/j.compbiomed.2025.109834>
- 221 Yellapragada S, Graikos A, Prasanna P, et al. PathLDM: text conditioned latent diffusion model for histopathology. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2024:5182-5191.
- 222 Molino D, Feola FD, Faiella E, et al. XGeM: a multi-prompt foundation model for multimodal medical data generation. *arXiv*. 2025. <https://doi.org/10.48550/arXiv.2501.04614>
- 223 Fan L, Chen T, He L, et al. GADM: data augmentation using generative adversarial diffusion model for pulse-based disease identification. *Biomed Signal Process Control*. 2025;100:107005. <https://doi.org/10.1016/j.bspc.2024.107005>
- 224 Sun S, Su Z, Meizhou J, et al. Optimizing medical image report generation through a discrete diffusion framework. *J Supercomput*. 2025;81(5):637. <https://doi.org/10.1007/s11227-025-07111-2>
- 225 Lena MH, Reinke A, Godau P, et al. Metrics reloaded: recommendations for image analysis validation. *Nat Methods*. 2024;21(2):195-212. <https://doi.org/10.1038/s41592-023-02151-z>
- 226 Nguyen HL, Vu DT, Jung JJ. Knowledge graph fusion for smart systems: a survey. *InfFusion*. 2020;61:56-70. <https://doi.org/10.1016/j.infuss.2020.03.014>
- 227 Peng C, Xia F, Naseriparsa M, et al. Knowledge graphs: opportunities and challenges. *Artif Intell Rev*. 2023;56(11):13071-13102. <https://doi.org/10.1007/s10462-023-10465-9>
- 228 Fellbaum C. WordNet. In: Poli R, Healy M, Kameas A, eds. Theory and applications of ontology: computer applications. Berlin: Springer Netherlands; 2010:231-243. [https://doi.org/10.1007/978-90-481-8847-5\\_10](https://doi.org/10.1007/978-90-481-8847-5_10)
- 229 Auer S, Bizer C, Kobilarov G, et al. DBpedia: a nucleus for a web of open data. In: Aberer K, Choi KS, Noy N, et al., eds. The Semantic Web. Berlin: Springer; 2007:722-735. [https://doi.org/10.1007/978-3-540-76298-0\\_52](https://doi.org/10.1007/978-3-540-76298-0_52)
- 230 Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology. *Nat Genet*. 2000;25(1):25-29. <https://doi.org/10.1038/75556>
- 231 Huang H, Wang X, Gu Z, et al. Research on medical knowledge graph construction technology and development status. *J Comput Eng Appl*. 2023;59(13):33-48. <https://doi.org/10.3778/j.issn.1002-8331.2209-0475>
- 232 Lin Y, Liu Z, Sun M, et al. Learning entity and relation embeddings for knowledge graph completion. *Proc AAAI Conf Artif Intell*. 2015;29(1). <https://doi.org/10.1609/aaai.v29i1.9491>
- 233 Zhang Z, Zhuang F, Zhu H, et al. Relational graph neural network with hierarchical attention for knowledge graph completion. *Proc AAAI Conf Artif Intell*. 2020;34(5):9612-9619. <https://doi.org/10.1609/aaai.v34i05.6508>
- 234 Lan X, Zhao J, Zhang Y, et al. Tacit knowledge mining: the key traditional Chinese medical inheritance. *Chin Med Cult*. 2020;3(1):15. [https://doi.org/10.4103/CMAC.CMAC\\_2\\_20](https://doi.org/10.4103/CMAC.CMAC_2_20)
- 235 Feng Y, Wu Z, Zhou X, et al. Knowledge discovery in traditional Chinese medicine: state of the art and perspectives. *Artif Intell Med*. 2006;38(3):219-236. <https://doi.org/10.1016/j.artmed.2006.07.005>
- 236 Wang S, Du X, Liu G, et al. An interpretable data-driven medical knowledge discovery pipeline based on artificial intelligence. *IEEE J Biomed Health Inform*. 2023;27(10):5099-5109. <https://doi.org/10.1109/JBHI.2023.3299339>
- 237 Xue B, Zou L. Knowledge graph quality management: a comprehensive survey. *IEEE Trans Knowl Data Eng*. 2023;35(5):4969-4988. <https://doi.org/10.1109/TKDE.2022.3150080>
- 238 Qi Y, Zheng W, Hong L, et al. Evaluating knowledge graph accuracy powered by optimized human-machine collaboration. In: Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. KDD '22. Association for Computing Machinery; 2022:1368-1378. <https://doi.org/10.1145/3534678.3539233>
- 239 Rahman S, Choi F, Kim H, et al. Knowledge acquisition and integration with expert-in-the-loop. *arXiv*. 2024. <https://doi.org/10.48550/arXiv.2402.03291>
- 240 Kou Z, Shang L, Zhang Y, et al. HC-COVID: a hierarchical crowdsourcing knowledge graph approach to explainable COVID-19 misinformation detection. *Proc ACM Hum-Comput Interact*. 2022;6(GROUP):361-36:25. <https://doi.org/10.1145/3492855>
- 241 Fülll A, Nissen V, Heringkle SH. Interactive machine learning of knowledge graph-based explainable process analysis. In: Ruiz M, Soffer P, eds. Advanced Information Systems Engineering Workshops. Springer International Publishing; 2023:112-124. [https://doi.org/10.1007/978-3-031-34985-0\\_12](https://doi.org/10.1007/978-3-031-34985-0_12)
- 242 Eduardo MR, Elena HP, David AR, et al. Human-in-the-loop machine learning: a state of the art. *Artif Intell Rev*. 2023;56(4):3005-3054. <https://doi.org/10.1007/s10462-022-10246-w>
- 243 Holzinger A. Interactive machine learning for health informatics: when do we need the human-in-the-loop? *Brain Inform*. 2016;3(2):119-131. <https://doi.org/10.1007/s40708-016-0042-6>
- 244 Monarch RM. Human-in-the-loop machine learning: active learning and annotation for human-centered AI. Shelton: Manning; 2021. <https://ieeexplore.ieee.org/document/10280384>
- 245 Jarrahi MH, Davoudi V, Haeri M. The key to an effective AI-powered digital pathology: establishing a symbiotic workflow between pathologists and

- machine. *J Pathol Inform.* 2022;13:100156. <https://doi.org/10.1016/j.jpi.2022.100156>.
- 246 Wu X, Xiao L, Sun Y, et al. A survey of human-in-the-loop for machine learning. *Future Gener Comput Syst.* 2022;135:364-381. <https://doi.org/10.1016/j.future.2022.05.014>.
- 247 Yang J, Zhuang X, Li Z, et al. CPMKG: a condition-based knowledge graph for precision medicine. *Database.* 2024;2024:baae102. <https://doi.org/10.1093/database/baae102>.
- 248 Vidal ME, Chudasama Y, Huang H, et al. Integrating knowledge graphs with symbolic AI: the path to interpretable hybrid AI systems in medicine. *J Web Semant.* 2025;84:100856. <https://doi.org/10.1016/j.websem.2024.100856>.
- 249 Li M, Xiao J, Chen B, et al. Loganin inhibits the ROS-NLRP3-IL-1 $\beta$  axis by activating the NRF2/HO-1 pathway against osteoarthritis. *Chin J Nat Med.* 2024;22(11):977-990. [https://doi.org/10.1016/S1875-5364\(24\)60555-8](https://doi.org/10.1016/S1875-5364(24)60555-8).
- 250 Xiu Y, Wang S, Zhang P, et al. Total glucosides of paeony alleviates cGAS-STING-mediated diseases by blocking the STING-IRF3 interaction. *Chin J Nat Med.* 2024;22(5):402-415. [https://doi.org/10.1016/S1875-5364\(24\)60572-8](https://doi.org/10.1016/S1875-5364(24)60572-8).
- 251 Huang Q, Wang M, Wang M, et al. *Scutellaria baicalensis*: a promising natural source of antiviral compounds for the treatment of viral diseases. *Chin J Nat Med.* 2023;21(8):563-575. [https://doi.org/10.1016/S1875-5364\(23\)60401-7](https://doi.org/10.1016/S1875-5364(23)60401-7).
- 252 Zhou T. Unveiling secrets of traditional Chinese medicine: cutting-edge techniques in component analysis. *Chin Herb Med.* 2025;17(3):484-487. <https://doi.org/10.1016/j.chmed.2025.05.006>.
- 253 Qian J, Shao X, Bao H, et al. Identification and characterization of cell niches in tissue from spatial omics data at single-cell resolution. *Nat Commun.* 2025;16(1):1693. <https://doi.org/10.1038/s41467-025-57029-9>.
- 254 Yang P, Jin K, Yao Y, et al. Spatial integration of multi-omics single-cell data with SIMO. *Nat Commun.* 2025;16(1):1265. <https://doi.org/10.1038/s41467-025-56523-4>.
- 255 Barabási AL, Gulbace N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet.* 2011;12(1):56-68. <https://doi.org/10.1038/nrg2918>.
- 256 Guo R, Zhang S, Li A, et al. Ginsenoside Rb1 and berberine synergistically protect against type 2 diabetes mellitus via GDF15/HAMP pathway throughout the liver lobules: insights from spatial transcriptomics analysis. *Pharmacol Res.* 2025;215:107711. <https://doi.org/10.1016/j.phrs.2025.107711>.
- 257 Gao F, Huang K, Xing Y. Artificial intelligence in omics. *Genomics Proteomics Bioinformatics.* 2022;20(5):811-813. <https://doi.org/10.1016/j.gpb.2023.01.002>.
- 258 Nam Y, Kim J, Jung SH, et al. Harnessing artificial intelligence in multimodal omics data integration: paving the path for the next frontier in precision medicine. *Annu Rev Biomed Data Sci.* 2024;7:225-250. <https://doi.org/10.1146/annurev-biodatasci-102523-103801>.
- 259 Gan X, Shu Z, Wang X, et al. Network medicine framework reveals generic herb-symptom effectiveness of traditional Chinese medicine. *Sci Adv.* 2023;9(43):eadh0215. <https://doi.org/10.1126/sciadv.adh0215>.
- 260 Amann J, Blasimme A, Vayena E, et al. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Med Inf Decis Mak.* 2020;20:1-9. <https://doi.org/10.1186/s12911-020-01332-6>.
- 261 Ghassemi M, Naumann T, Schulam P, et al. A review of challenges and opportunities in machine learning for health. *AMIA Jt Summits Transl Sci Proc.* 2020;2020:191. <https://pubmed.ncbi.nlm.nih.gov/32477638/>.
- 262 Hasan U, Gani MO. Krcr: a prior knowledge based causal discovery framework with reinforcement learning. In: Machine Learning for Healthcare Conference; PMLR. 2022:691-714. <https://proceedings.mlr.press/v182/hasan22a.html>.
- 263 Esmaili P, Roshanravan N, Ghaffari S, et al. Unraveling atherosclerotic cardiovascular disease risk factors through conditional probability analysis with Bayesian networks: insights from the AZAR cohort study. *Sci Rep.* 2024;14(1):4361. <https://doi.org/10.1038/s41598-024-55141-2>.
- 264 Pearl J. Comment: understanding simpson's paradox. In: Probabilistic and Causal Inference: The Works of Judea Pearl. Shelton: Manning Publications. 2022:399-412. <https://doi.org/10.1145/3501714.3501738>.
- 265 Cheng F, Kovács IA, Barabási AL. Network-based prediction of drug combinations. *Nat Commun.* 2019;10(1):1197. <https://doi.org/10.1038/s41467-019-09186-x>.
- 266 Richens JG, Lee CM, Johri S. Improving the accuracy of medical diagnosis with causal machine learning. *Nat Commun.* 2020;11(1):3923. <https://doi.org/10.1038/s41467-020-17419-7>.
- 267 Chu X, Wu S, Sun B, et al. Data-driven quantification and intelligent decision-making in traditional Chinese medicine: a review. *Int J Mach Learn Cybern.* 2024;15(8):3455-3470. <https://doi.org/10.1007/s13042-024-02103-9>.
- 268 Charpignon M, Vakulenko B, Zheng B, et al. Causal inference in medical records and complementary systems pharmacology for metformin drug repurposing towards dementia. *Nat Commun.* 2022;13(1):7652. <https://doi.org/10.1038/s41467-022-35157-w>.
- 269 Xian Y, Fu Z, Muthukrishnan S, et al. Reinforcement knowledge graph reasoning for explainable recommendation. In: Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval; SIGIR'19. 2019:285-294. <https://doi.org/10.1145/3331184.3331203>.
- 270 Wang H, Wang J. A quantitative diagnostic method based on bayesian networks in traditional Chinese medicine. In: Neural Information Processing. Springer Berlin. 2006:176-183. [https://doi.org/10.1007/11893295\\_20](https://doi.org/10.1007/11893295_20).
- 271 Xia C, Deng F, Wang Y, et al. Classification research on syndromes of TCM based on SVM. In: 2009 2nd International Conference on Biomedical Engineering and Informatics. 2009:1-4. <https://doi.org/10.1109/BMEI.2009.5305418>.
- 272 Hager P, Jungmann F, Holland R, et al. Evaluation and mitigation of the limitations of large language models in clinical decision-making. *Nat Med.* 2024;30(9):2613-2622. <https://doi.org/10.1038/s41591-024-03097-1>.
- 273 Bussone A, Stumpf S, O'Sullivan D. The role of explanations on trust and reliance in clinical decision support systems. In: 2015 International Conference on Healthcare Informatics. IEEE. 2015:160-169. <https://doi.org/10.1109/ICHI.2015.26>.
- 274 Tonekaboni S, Joshi S, McCradden MD, et al. What clinicians want: contextualizing explainable machine learning for clinical end use. In: Machine Learning for Healthcare Conference. PMLR. 2019:359-380. <https://proceedings.mlr.press/v106/tonekaboni19a.html>.
- 275 Yin Z, Kuang Z, Zhang H, et al. Explainable AI method for tinnitus diagnosis via neighbor-augmented knowledge graph and traditional Chinese medicine: Development and validation study. *JMIR Med Inform.* 2024;12(1):e57678. <https://doi.org/10.2196/57678>.
- 276 Chen Y, Li H, Li H, et al. An overview of knowledge graph reasoning: key technologies and applications. *J Sens Actuator Netw.* 2022;11(4):78. <https://doi.org/10.3390/jsan11040078>.
- 277 Tutek M, Śnajder J. Toward practical usage of the attention mechanism as a tool for interpretability. *IEEE Access.* 2022;10:47011-47030. <https://doi.org/10.1109/ACCESS.2022.3169772>.
- 278 Nohara Y, Matsumoto K, Soejima H, et al. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. *Comput Methods Programs Biomed.* 2022;214:106584. <https://doi.org/10.1016/j.cmpb.2021.106584>.
- 279 Lin BY, Chen X, Chen J, et al. KagNet: knowledge-aware graph networks for commonsense reasoning. *arXiv.* 2019. <https://doi.org/10.48550/arXiv.1909.02151>.
- 280 Singh J, Rani S, Sriakashmi G. Towards explainable AI: interpretable models for complex decision-making. In: 2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS). 2024:1-5. <https://doi.org/10.1109/ICKECS61492.2024.10616500>.
- 281 Sutton RT, Pincok D, Baumgart D, et al. An overview of clinical decision support systems: benefits, risks, and strategies for success. *npj Digit Med.* 2020;3(1):17. <https://doi.org/10.1038/s41746-020-0221-y>.
- 282 Cai CJ, Reif E, Hegde N, et al. Human-centered tools for coping with imperfect algorithms during medical decision-making. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. CHI '19. 2019:1-14. <https://doi.org/10.1145/3290605.3300234>.
- 283 Silvia DA, Norajitra T, Lüth CT, et al. How do deep-learning models generalize across populations? cross-ethnicity generalization of COPD detection. *Insights Imaging.* 2024;15(1):198. <https://doi.org/10.1186/s13244-024-01781-x>.
- 284 Futoma J, Simons M, Panch T, et al. The myth of generalisability in clinical research and machine learning in health care. *Lancet Digit Health.* 2020;2(9):e489-e492. [https://doi.org/10.1016/S2589-7500\(20\)30186-2](https://doi.org/10.1016/S2589-7500(20)30186-2).
- 285 Arcadu F, Benmansour F, Maunz A, et al. Deep learning algorithm predicts diabetic retinopathy progression in individual patients. *npj Digit Med.* 2019;2(1):92. <https://doi.org/10.1038/s41746-019-0172-3>.
- 286 Gu Y, Ge Z, Bonnington CP, et al. Progressive transfer learning and adversarial domain adaptation for cross-domain skin disease classification. *IEEE J Biomed Health Inform.* 2020;24(5):1379-1393. <https://doi.org/10.1109/JBHI.2019.2942429>.
- 287 Chaddad A, Lu Q, Li J, et al. Explainable, domain-adaptive, and federated artificial intelligence in medicine. *IEEECAA J Autom Sin.* 2023;10(4):859-876. <https://doi.org/10.1109/JAS.2023.123123>.
- 288 Albalawi E, Mahesh T, Thakur A, et al. Integrated approach of federated learning with transfer learning for classification and diagnosis of brain tumor. *BMC Med Imaging.* 2024;24(1):110. <https://doi.org/10.1186/s12880-024-01261-0>.