Contents lists available at ScienceDirect

Digital Chinese Medicine

journal homepage: http://www.keaipublishing.com/dcmed

Advances and prospects of the integration of multi-omics and artificial intelligence in traditional Chinese medicine research

Guicheng LIU[a], Xi LONG[a], Qinghua PENG[a, b*], Sainan TIAN[a*], Shujuan HU[a*]

a. School of Traditional Chinese Medicine, Hunan University of Chinese Medicine, Changsha, Hunan 410208, China

b. Institute of Traditional Chinese Medicine Diagnostics, Hunan University of Chinese Medicine, Changsha, Hunan 410208, China

ARTICLE INFO

Article history
Received 14 April 2025
Accepted 27 June 2025
Available online 25 September 2025

Keywords
Traditional Chinese medicine
Multi-omics
Artificial intelligence
Bibliometrics
CiteSpace

ABSTRACT

Objective To map the research hotspots, developmental trends, and existing challenges in the integration of artificial intelligence (AI) with multi-omics in traditional Chinese medicine (TCM) through comprehensive bibliometric analysis.

Methods China National Knowledge Infrastructure (CNKI), Wanfang Data, China Science and Technology Journal Database (VIP), Chaoxing Journal Database, PubMed, and Web of Science were searched to collect literature on the theme of AI in TCM multi-omics research from the inception of each database to December 31, 2024. Eligible records were required to simultaneously address AI, TCM, and multi-omics. Quantitative and visual analyses of publication growth, core authorship networks, institutional collaboration patterns, and keyword co-occurrence were performed using Microsoft Excel 2021, NoteExpress v4.0.0, and CiteSpace 6.3.R1. AI application modes in TCM multi-omics research were also categorized and summarized.

Results A total of 1 106 articles were enrolled (932 Chinese and 174 English). Publication output has increased continuously since 2010 and accelerated after 2016. Region-specific collaboration clusters were identified, dominated by Beijing University of Chinese Medicine, China Academy of Chinese Medical Sciences, Shanghai University of Traditional Chinese Medicine, and Nanjing University of Chinese Medicine. Keyword co-occurrence analysis revealed that current AI applications predominantly centered on metabolomics and algorithms such as cluster analysis and data mining. Research foci mainly ranked as follows: single herbs, herbal formulae, and disease-syndrome differentiation.

Conclusion Machine learning methods are the predominant integrative modality of AI in the realm of TCM multi-omics research at present, utilized for processing omics data and uncovering latent patterns therein. The domain of TCM, in addition to investigating omics information procured through high-throughput technologies, also integrates data on traditional Chinese medicinal substances and clinical phenotypes, progressing towards joint analysis of multi-omics, high-dimensionality of data, and multi-modality of information. Deep learning approaches represent an emerging trend in the field.

1 Introduction

For decades, traditional Chinese medicine (TCM) has mainly confronted three challenges: (i) reconciling TCM theory with modern science; (ii) objective and standardized criteria establishment for assessing the clinical efficacy; (iii) updating diagnostic and therapeutic methodologies. Multi-omics, an analytical framework derived

---

*Corresponding author: Shujuan HU, E-mail: hsj2012@163.com. Sainan TIAN, E-mail: tsn004826@163.com. Qinghua PENG, E-mail: pqh410007@126.com.

Peer review under the responsibility of Hunan University of Chinese Medicine.

DOI: 10.1016/j.dcmed.2025.09.003

Citation: LIU GC, LONG X, PENG QH, et al. Advances and prospects of the integration of multi-omics and artificial intelligence in traditional Chinese medicine research. Digital Chinese Medicine, 2025, 8(3): 300-312.

from the central dogma of biology, offers a hierarchical and multidimensional platform that simultaneously interrogates genes, transcripts, proteins, metabolites, phenotypes, images, and microbiota [1, 2]. Such holistic profiling can illuminate molecular interactions and regulatory networks with unprecedented depth, providing multi-omics as an indispensable tool for elucidating organismal functions and addressing the above challenges. Beyond canonical omics layers, TCM community has further pioneered syndrome phenomics, herbal phenomics, and formula omics, thereby expanding the conceptual and technical frontiers of life-science research [3]. As an emerging interdisciplinary field, TCM multi-omics has become pivotal for TCM theory and herbal drug discovery. Yet, the integration of these heterogeneous big-data streams is inevitably hampered by the curse of dimensionality. AI excels at handling massive datasets, enhancing both efficiency and precision. Coupling AI with multi-omics not only maximizes data utility, but also refines conclusions that might be missed by single-omics or non-AI approaches [4]. Consequently, the convergence of multi-omics and AI constitutes a powerful catalyst for TCM modernization [5, 6], underscoring the urgency of systematically mapping its current status and developmental trajectories.

Bibliometric analysis represents a quantitative approach that extract patterns and insights through systematic evaluation of scholarly output, integrating mathematics, statistics, and information science. This approach integrates publication metadata—including citation networks and temporal publication patterns—as quantifiable indicators to map research evolution and knowledge diffusion dynamics. This study seeks to delineate the intellectual topography, collaborative architecture, and prospective trajectories of artificial intelligence-driven multi-omics investigations within the domain of TCM, thereby remedying the conspicuous absence of scientometric and panoramic analyses in this emerging field.

2 Data and methods

2.1 Literature search strategy

During database foundation to December 31, 2024, six databases were included: China National Knowledge Infrastructure, Wanfang Data, China Science and Technology Journal Database (VIP), Chaoxing Journal Database, PubMed, and Web of Science. A three-step Boolean strategy was used, combining medical subject headings (MeSH) terms and free-text synonyms in each step.

(i) AI technologies: artificial intelligence OR machine learning OR deep learning OR data mining OR cluster analysis OR Bayesian network OR neural networks OR algorithm OR association rule.

(ii) TCM domain: traditional Chinese medicine OR Chinese herbal medicine OR syndrome differentiation OR Zheng OR acupuncture OR electroacupuncture.

(iii) Multi-omics: omics OR genomics OR transcriptomics OR proteomics OR metabolomics OR microbiome OR microbiomics OR metagenomics OR radiomics.

Within each step, terms were linked with OR. The three steps were combined with AND. In PubMed and Web of Science, both MeSH/topic terms and title/abstract fields were searched. Chinese databases were queried on “subject” or “all fields” without language restriction.

2.2 Data screening

Studies were included if: (i) articles were publicly available in the searched databases; (ii) they explicit coverage the three thematic areas—AI, TCM, and multi-omics; (iii) they contained complete bibliographic elements, including keywords, abstract, and author information.

Studies were excluded if: (i) they were studies unrelated to human biomedicine or to drug research involving animals or cell experiments; (ii) they were conference abstracts, calls for papers, popular science articles, consensus statements, and guidelines; (iii) they were reports lacking a clearly defined omics layer or whose experimental methods could not be classified within any omics domain; (iv) they were literature on ethnic minority medicines not grounded in TCM theory.

2.3 Data processing and standardization

All citations were imported into NoteExpress v4.0.0 and deduplicated. Two reviewers independently screened titles, abstracts, and keywords for inclusion or exclusion. Disagreements were resolved by a third reviewer.

Prior to analysis, we standardized the remaining records. Institutional names were updated. For example, “Henan College of Traditional Chinese Medicine” was merged under “Henan University of Chinese Medicine”. Affiliations located in Hong Kong and Macao were recorded as “China”. Keywords and other metadata were harmonised, for example, “data mining” “big-data mining” and “data-mining method” were merged under “data mining”.

2.4 Statistical analysis

In this study, Excel 2021 was used for data organization and basic statistical analysis. NoteExpress v4.0.0 managed the references and performed preliminary classification. CiteSpace 6.3.R1 analyzed keyword co-occurrence, author collaboration networks, and institutional cooperation patterns. Descriptive statistics and visualization techniques were employed to identify research hotspots and collaboration patterns.

3 Results

3.1 Temporal dynamics of publication output

The initial search yielded 2 204 Chinese and 297 English records. After reconciliation, 932 Chinese and 174 English articles were retained. As shown in Figure 1, the earliest relevant records in Chinese databases dated back to 2001. Publications began to rise around 2010, increased sharply after 2016, and continued an upward trajectory until the end of 2024. A similar pattern was observed in English-language articles.

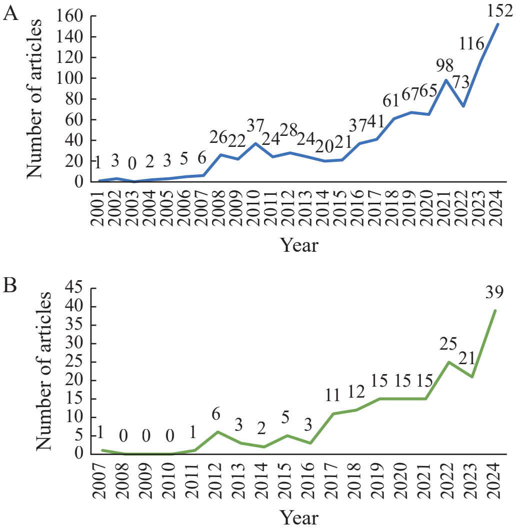

Figure 1 Annual publication trends in Chinese and English publications

A, Chinese publications. B, English publications.

3.2 Author collaboration networks and thematic affinities

A total of 604 Chinese-language authors and 1 143 international authors were identified. In the Chinese dataset ($M_{\max} = 17$), the threshold was $N = 3.09$, yielding 20 core authors. In the Web of Science/PubMed dataset ($M_{\max} = 13$), the threshold was $N = 3$, yielding 100 core authors. CiteSpace 6.3.R1 visualisation revealed that these core authors formed small, loosely connected clusters with limited inter-group linkage. Clustering the most strongly collaborating teams (Figure 2) showed that research on AI-assisted TCM multi-omics is most cohesive in the ophthalmic areas, cardiovascular and rheumatic-immune diseases, employing metabolomics, radiomics, and proteomics. Methodologically, network analyses and machine learning algorithms predominated in processing the underlying omics data.

3.3 Inter-institutional synergy maps and geospatial concentration of research hubs

We identified 476 institutions in Chinese corpus and 219 across the English corpus. The resulting collaboration networks were shown in Figure 3. The field was dominated by TCM universities and their affiliated hospitals, with Beijing University of Chinese Medicine as the most prolific contributor. Dense inter-institutional links were evident between Beijing University of Chinese Medicine, the China Academy of Chinese Medical Sciences, Shanghai University of Traditional Chinese Medicine, and several other key centers, both in Chinese and English publications.

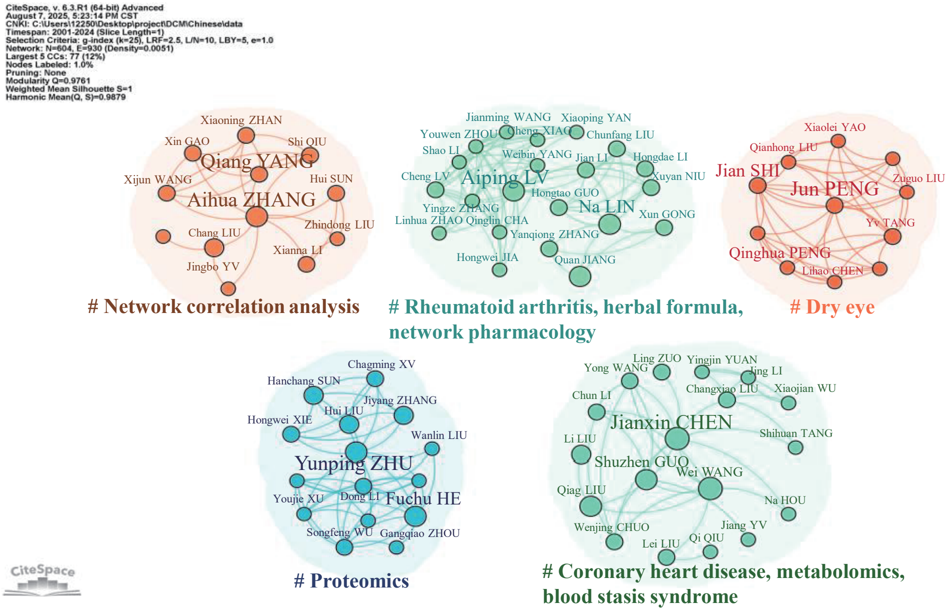
```
CiteSpace, v. 6.3.R1 (64-bit) Advanced
August 7, 2025, 5:23:14 PM CST
CNKI: C:\Users\12250\Desktop\project\DCM\Chinese\data
Timespan: 2001-2024 (Slice Length=1)
Selection Criteria: g-index (k=25), LRF=2.5, L/N=10, LBY=5, e=1.0
Network: N=604, E=930 (Density=0.0051)
Largest 5 CCs: 77 (12%)
Nodes Labeled: 1.0%
Pruning: None
Modularity Q=0.9761
Weighted Mean Silhouette S=1
Harmonic Mean(Q, S)=0.9879
```

Figure 2 Core author collaboration network and thematic clusters in AI-driven TCM multi-omics research

Co-authorship clusters were based on research keywords. Labels beneath each cluster denote collaborative thematic foci. Color blocks differentiate distinct keyword clusters. Nodes represent authors within each cluster, with node and font sizes proportional to each author's publication count. Edges between nodes indicate collaborative relationships among authors.

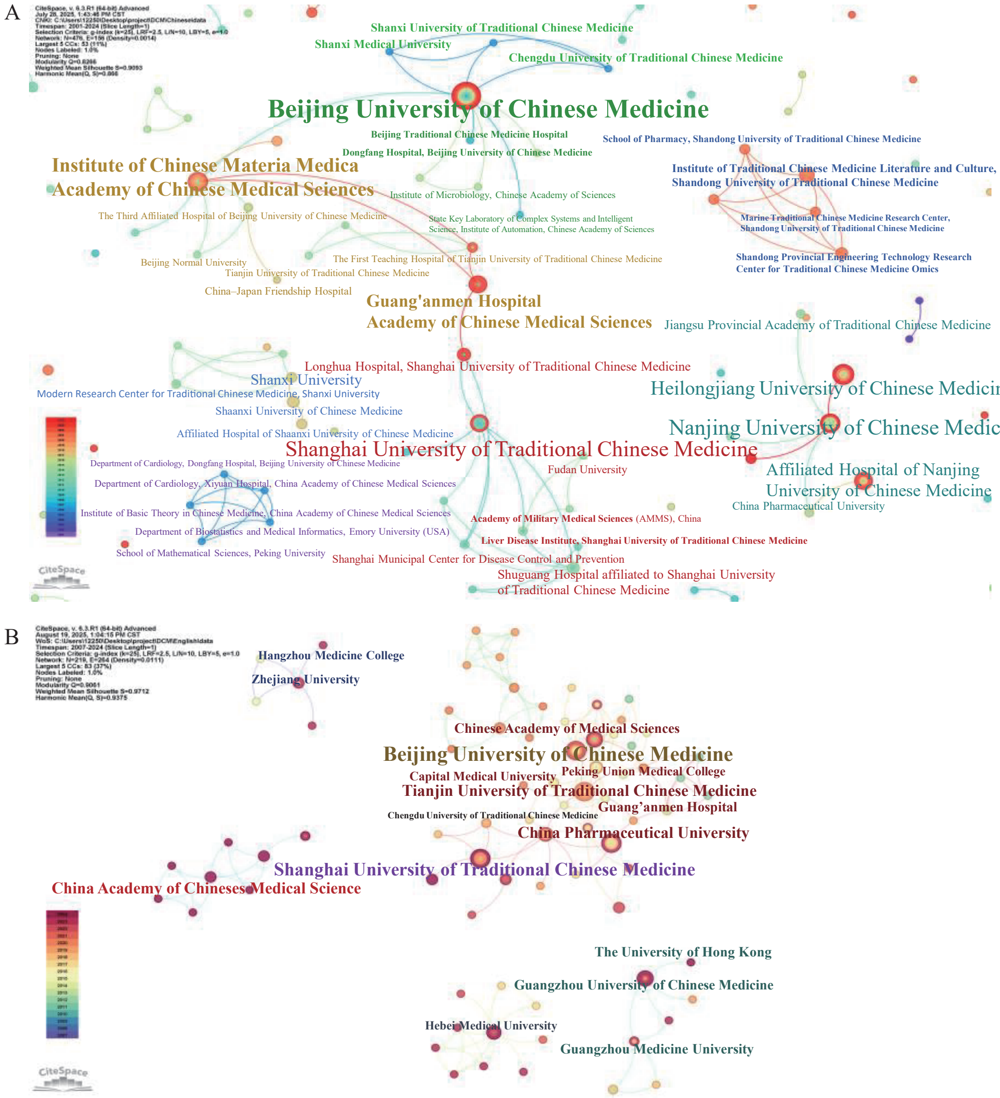

Figure 3 Institutional collaboration networks in AI-driven TCM multi-omics research

A, institutional collaboration in Chinese literature. B, institutional collaboration in English literature. Node color indicates publication year; closer to red means closer to 2024.

Keyword-driven cluster analysis beneath each institutional network revealed distinct research profiles. Beijing University of Chinese Medicine's collaboration web was broad, centring on metabolomics and employing AI-based clustering and dimensionality-reduction algorithms to interrogate omics data. China Academy of Chinese Medical Sciences focused on rheumatoid arthritis and other metabolic-immune disorders, prioritising transcriptomics and metabolomics, while integrating machine-learning approaches in data-mining and network-pharmacology studies. Shanghai University of Traditional Chinese Medicine's cluster was anchored in herbal-medicine research, emphasising metabolomics and proteomics. It has also pioneered phenomics adoption in TCM, applying hierarchical clustering, support-vector machines and decision-tree classifiers to omics datasets. Shandong University of Traditional Chinese Medicine's network is distinguished by investigations into marine-derived TCM, with frequent deployment of association-rule-mining algorithms. Institutions, such as Shanxi University of Chinese Medicine, Shaanxi University of Chinese Medicine, and Shanxi University concentrated on biomarker discovery for clinical syndromes, predominantly via metabolomics coupled with clustering-based machine-learning techniques. Nanjing University of Chinese Medicine's collaboration hub excelled in studies on crude drugs, single compounds and herbal formulae, integrating metabolomics, proteomics and genomics and routinely employing clustering algorithms to interface with AI. Guangzhou University of Chinese Medicine's cluster was notable for transcriptomic research, leveraging AI to orchestrate multi-omics integration and complex-network construction within network-pharmacology-based investigations of herbal formulae.

3.4 Funding support from the National Natural Science Foundation of China (NNSFC) and thematic distribution of projects

Funding sources served as a proxy for project impact and disciplinary priority. Among the 932 Chinese-language articles, 96 (10.3%) were supported by NNSFC (Figure 4A). Content analysis of these 96 grants revealed four principal research themes (Figure 4B). The largest share is devoted to herbal-medicine studies—covering resource exploitation, quality evaluation, active-ingredient identification, and pharmacological mechanisms—where AI is primarily applied to omics-based clustering and association-rule mining.

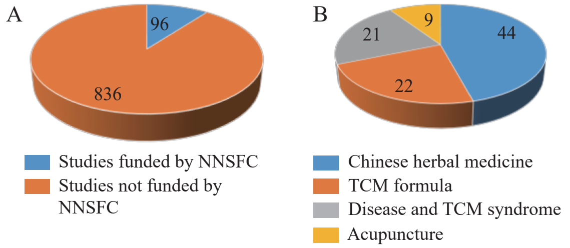

Figure 4 Funding landscape of AI-driven TCM multi-omics research in Chinese publications

A, number of studies funded by NNSFC vs. not funded by NNSFC among Chinese publications. B, distribution of research directions of projects funded by NNSFC.

3.5 Keyword evolution and burst dynamics across linguistic corpora

Keywords offered a concise distillation of article content, with the frequency and co-occurrence patterns revealing prevailing research foci [7]. The top ten keywords ranked by betweenness centrality are presented in Table 1. In the Chinese corpus, "metabolomics" displayed the highest centrality, indicating that metabolomics was the principal omics layer interfacing with AI. Remaining keywords confirmed that machine learning remained the dominant AI modality in this literature. In English-language articles, TCM research coupled AI primarily with metabolism, network pharmacology, and molecular docking, while the "artificial neural network" signaled the incipient contribution of deep learning approaches. Keyword co-occurrence networks are presented in Figure 5.

Clustering keywords along a temporal axis revealed the interconnections among research hotspots and their evolution, thereby mapping past trends and future directions in AI-assisted TCM multi-omics studies [7]. These

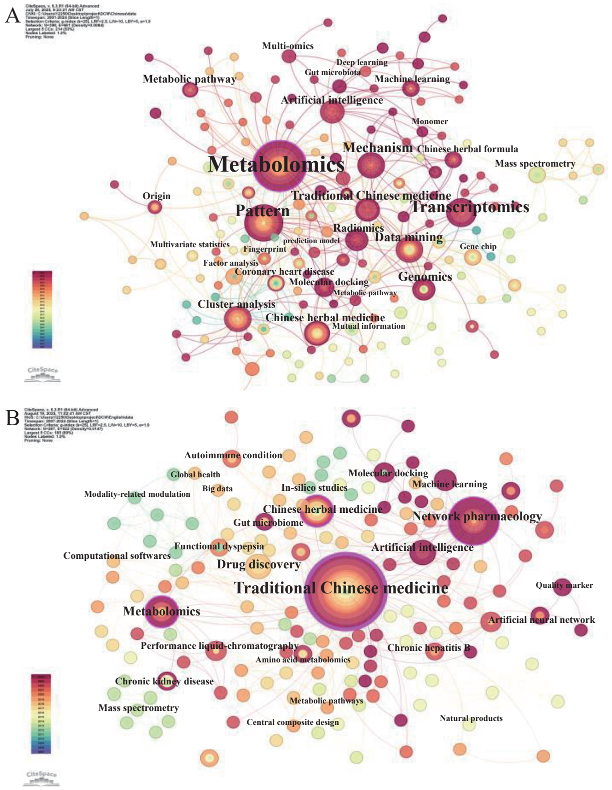

Figure 5 Temporal keyword co-occurrence networks delineating the thematic evolution of AI-driven TCM multi-omics research

A, keywords derived from Chinese literature. B, keywords derived from English literature. The larger the node, the greater the keyword's betweenness centrality. Nodes closer to red indicate keywords whose active years are nearer to 2024. An edge between two nodes signifies their co-occurrence.

Table 1 Top 10 keywords by betweenness centrality in Chinese and English publications

Rank	Keyword in Chinese publications (centrality, count)	Keyword in English publications (centrality, count)
1	Metabolomics (0.20, 57)	Traditional Chinese medicine (0.49 36)
2	Traditional Chinese medicine pattern (0.08, 27)	Network pharmacology (0.20, 17)
3	Cluster analysis (0.08, 17)	Metabolomics (0.14, 6)
4	Data mining (0.06, 15)	Chinese herbal medicine (0.12, 4)
5	Machine learning (0.05, 5)	drug discovery (0.08, 4)
6	Traditional Chinese medicine (0.04, 11)	Chronic kidney disease (0.07, 3)
7	Genomics (0.04, 8)	Functional dyspepsia (0.06, 2)
8	Transcriptomics (0.03, 17)	Molecular docking (0.02, 3)
9	Chinese herbal medicine (0.03,15)	Artificial neural network (0.02, 2)
10	Artificial intelligence (0.03,11)	Asari radix et rhizoma (0.02, 2)

clusters against publication time were presented in Figure 6. Eleven clusters (Figure 6A, Chinese literature) were presented as follows. Metabolomics, data mining, and cluster analysis exhibited long temporal spans, underscoring their status as classic applications of AI in TCM omics. The "artificial intelligence" cluster first dispersed terms such as machine learning and herbal formulae around 2016, followed by deep learning and multimodal nodes near 2022, reflecting the discipline's technological trajectory. Seven clusters (Figure 6B, English literature) were presented. Metabolomics was the earliest and most enduring, whereas network pharmacology remained an active and expanding focus.

A keyword burst was denoted as a sudden, marked increase in frequency within a short period, signaling intensified scholarly attention and the shifting of research hotspots. Among the Chinese-language records, eighteen burst terms were detected (Figure 7A). Early bursts (2006 – 2012) included "mutual information" "cluster analysis" "mass spectrometry" and "data mining", indicating that machine-learning approaches were already being applied to extract features and build models from TCM multi-omics data. During 2010 – 2020, the concept of multi-omics gained prominence, evidenced by "gene chip" bursts, "metabolomics" and context-specific terms, such as "cerebral hemorrhage" "rat" and "producing area". This reflected progress in disease diagnosis, animal experimentation, and herbal medicine research that were accelerated by bioinformatics and computational biology. Since 2020, "artificial intelligence" "transcriptomics" "molecular docking" and "radiomics" have emerged as new bursts. Molecular docking now was benefited from graph neural networks and geometric deep learning methods that accurately

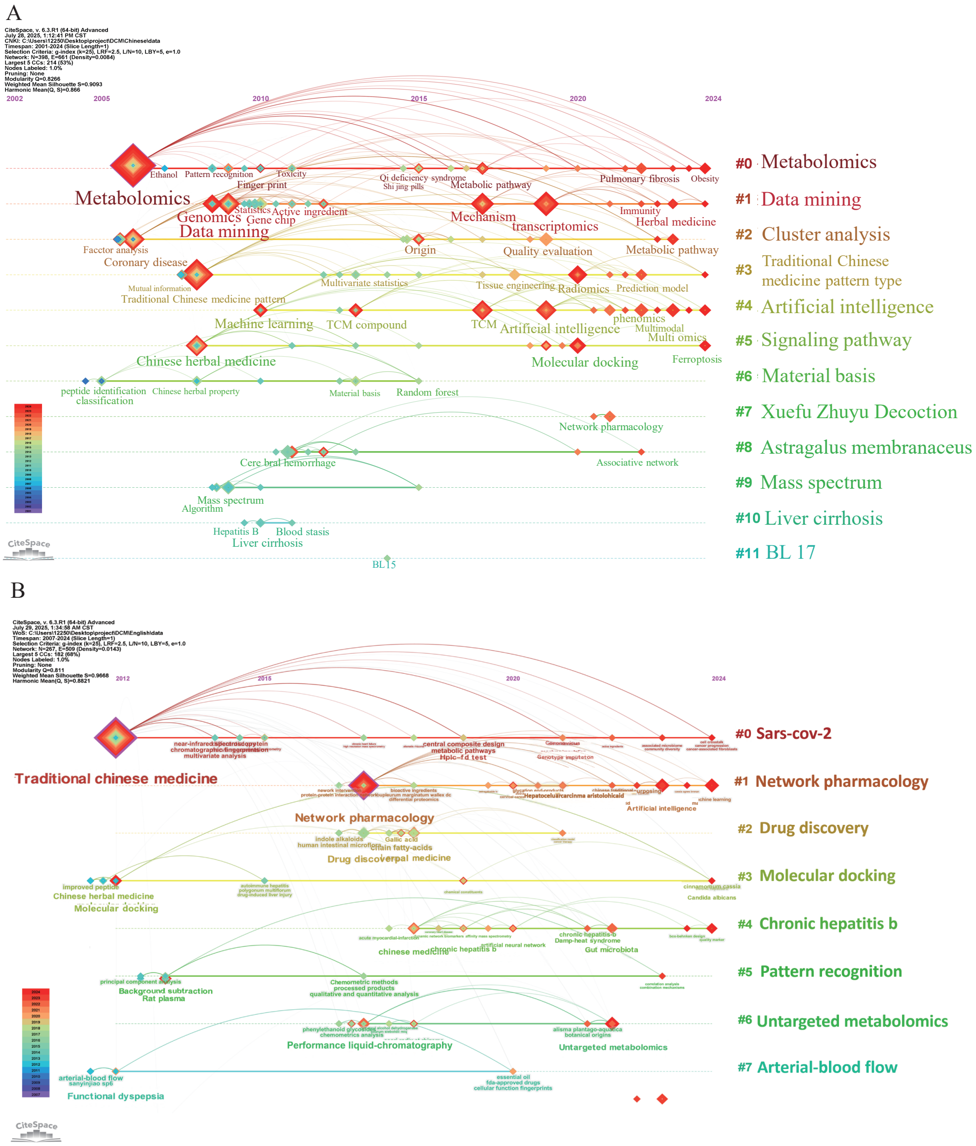

Figure 6 Timeline-based keyword cluster analysis of AI-driven TCM multi-omics research

A, keywords derived from Chinese publications. B, keywords derived from English publications. Nodes and clusters are colored by their latest occurrence year: the closer to red, the nearer to 2024.

predicted molecular interactions. Meanwhile, the incorporation of radiomics greatly enriched data dimensionality and modalities, necessitating deep learning frameworks for integrative multi-omics analysis. In the English-language corpus (Figure 7B), gut-microbiome studies became the latest international focus, and random forest highlighted as a key algorithm linking AI to the discipline. In summary, unsupervised machine learning techniques, especially clustering and dimension-reduction algorithms, remained central for pattern discovery in TCM omics, yet deep learning and multimodal approaches rapidly gained traction in integrative multi-omics investigations.

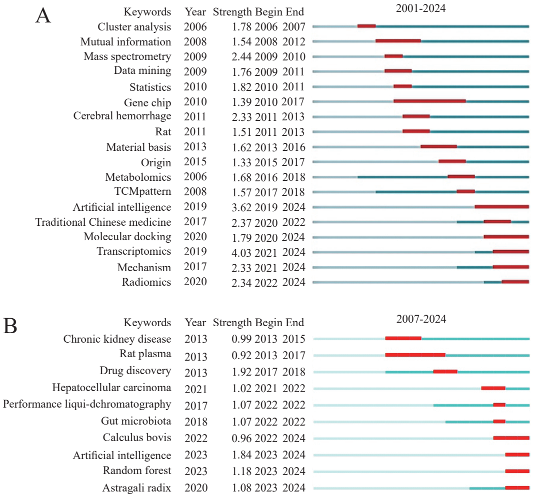

Figure 7 Temporal dynamics of emergent research foci revealed by keyword bursts in AI-driven TCM multi-omics literature

A, keywords derived from Chinese publications. B, keywords derived from English publications. Red areas indicate the active time periods of keywords.

4 Discussion

4.1 Current status of AI in TCM omics research

4.1.1 Advances in AI-multi-omics integration Driven by rapid improvements in high-throughput technologies, omics approaches have become indispensable throughout the biomedical spectrum, with their pivotal roles in defining disease mechanisms, biomarkers, drug targets, and refining diagnostic strategies, thereby creating unprecedented opportunities for translational medicine [8, 9]. Contemporary studies increasingly recognize that complex disorders arise from the interplay of multiple environmental and genetic factors. Integrative analyses indicated that diverse omics layers can fully capture the cross-level molecular circuitry underlying these conditions. Consequently, the transition from single-omics investigations to multi-omics data fusion has become inevitable.

Multi-omics integration presented formidable challenges. The datasets are massive, heterogeneous and intrinsically noisy, rendering conventional analytical workflows—largely reliant on manual curation and prior knowledge—insufficient for causal relationships or potential mechanisms. The central task, therefore, is to distil interpretable biological insights from these high-dimensional, discordant data, a goal that depends on the advanced AI algorithms [10].

In recent years, AI has achieved breakthroughs across the entire bench-to-bedside continuum of biomedical multi-omics research. In disease-mechanism discovery, graph neural networks (GNN) and Transformer architectures are deployed for molecular-property prediction [11]. Examples included DeepVariant, a deep learning model that accurately calls genomic variants [12], and TranSVAE, a variational autoencoder that enables cross-omics dynamic modelling [13]. These tools have shifted the analytical paradigm from static association to dynamic causal inference. For biomarker screening, XGBoost can automatically select high-dimensional features from multi-omics profiles and efficiently pinpoints candidate markers [14, 15], whereas random forest mitigated the “small-sample, high-noise” problem typical of clinical datasets [16]. Multimodal frameworks such as mixture-of-experts multimodal variational autoencoder (MMVAE) and attention-based models like TabTransformer further enhance biomarker specificity [17], accelerating risk prediction and precision-medicine applications. In drug discovery, the synergistic use of AlphaFold2 for protein-structure prediction and AutoDock-GPU for molecular docking—augmented by knowledge graphs and graph attention networks—has markedly increased the efficiency of target identification.

4.1.2 Omics platforms frequently applied in TCM and their trajectories Modernisation of TCM increasingly relies on elucidating molecular disease pathogenesis and the objective therapeutic efficacy of herbal compounds. The systems-biology framework powered by multi-omics technologies, aligns closely with the TCM holistic philosophy, offering innovative and efficient strategies for evidence-based TCM research, herbal-resource exploitation, and digital transformation in diagnostics and therapeutics [18]. Subsequent sections reviewed the current applications and emerging trends of three widely adopted omics platforms—metabolomics, proteomics, and transcriptomics—in advancing TCM research.

Metabolomics has emerged as a central pillar of TCM research, informing the biological substrates of syndrome differentiation, the mechanisms of herbal formulae, and the rationale underlying acupuncture. In syndrome-based investigations, metabolomic profiling has revealed discrete metabolic disturbances that map to classical TCM patterns. For example, blood-stasis syndrome is underpinned by aberrant amino acid metabolism, whereas Qi-deficiency syndrome disrupts energy and oxidative stress pathways [19]. Studies of multi-herb prescriptions use metabolomics to identify blood-borne active constituents, dissect pharmacokinetics, and elucidate the systems-level orchestration of therapeutic effects [20]. Comparative metabolomic profiling of local and systemic biofluids demonstrates differential metabolic signatures that underpin acupuncture's therapeutic efficacy. Emerging studies couple spatial metabolomics, single-cell metabolomics, and AI-driven network modelling to acupuncture, transitioning the field from macroscopic symptom description to multi-scale mechanistic insights spanning cells, molecules and pathways [21]. Despite these advancements, metabolomics remains largely descriptive, with limited integration into deeper TCM frameworks, such as Yin-Yang theory, Zang-Fu interplay and dynamic syndrome evolution. Consequently, a conceptual gap persists between metabolomic signatures and classical TCM doctrine, and causal links between metabolite patterns and clinical syndromes remain underexplored.

Proteomics occupies a pivotal role in TCM research, as proteins serve as primary effectors of biological function, mediating constitutional, pathological, and syndrome-related phenotypes central to TCM diagnostics. Comparative proteomic surveys of biofluids or tissues have identified syndrome-specific protein signatures, establishing molecular foundations for objective and standardised diagnosis while elucidating dynamic proteomic trajectories underlying syndrome evolution [21, 22]. Notably, proteomics captures post-translational modifications on minute-to-hour time-scales, enabling detection of acute disease shifts and early pharmacological responses to acupuncture or herbal interventions with high sensitivity. Unlike metabolomics—where altered metabolites required functional validation, so proteomic signatures directly reflect effector mechanisms, offering function-oriented biomarkers for clinical translation. In herbal-medicine studies, pre- and post-treatment proteomic comparisons map to systemic impacts of multi-compound prescriptions. Network-based reconstruction of drug-modulated protein interactions reveals how individual herbs within a formula cooperatively regulate core protein modules, thereby furnishing molecular rationales for active-ingredient discovery and formula-combination principles [23].

Transcriptomics deciphers the systemic regulatory mechanisms of TCM via gene expression, serving as a pivotal tool for disentangling the intricate relationships among disease pathogenesis, TCM syndromes, and pharmacotherapeutic interventions. In TCM syndrome studies, comparative transcriptome profiling of tissues or blood from patients with distinct syndrome patterns has uncovered syndrome-specific gene modules that furnish objective molecular signatures [24, 25]. For complex herbal formulae, transcriptomics delineates multi-target modes of action for both whole prescriptions and individual constituents, pinpointing critical pathway genes and extending the analysis to epitranscriptomic layers [26, 27]. Key advantages include the ability to capture early transcriptional events, preceding downstream protein or metabolic alterations, and offering cost-effective scalability. Beyond mechanistic studies, transcriptomics supports the sustainable development of medicinal resources by identifying key genes involved in the biosynthesis of active ingredients in geo-authentic herbs, thereby guiding the high-quality botanical drug production [28, 29]. Nevertheless, a transcript-only snapshot risks decoupling mRNA abundance from functional protein output. Future paradigms in TCM transcriptomics are therefore shifting toward integrative strategies that combine proteomic validation, single-cell and spatial transcriptomics to map therapeutic micro-domains, and AI-driven network models that embed "disease–syndrome–formula" relationships within predictive systems [30, 31].

The holistic, complex, and spatiotemporally dynamic nature of TCM naturally aligns with integrative multi-omics strategies [32], yet its theory cannot be mapped one-to-one onto any single omics layer. Driven by high-throughput technologies and a conceptual shift in biomedicine from single-trait observations to systematic phenotypic analyses, the notion of the "phenome" has emerged. In 2010, David Houle's review established phenomics as a discipline leveraging large-scale, high-throughput techniques to dissect phenotypic traits and their interactions with genotype and environment, thereby constructing comprehensive genotype–phenotype–environment landscapes [33]. Medical phenomics integrates upstream molecular multi-omics data with external environmental exposures to yield high-dimensional insights into disease mechanisms and individual variability. Building on this framework, Chinese scholars such as Dayue DUAN and Xiaolin TONG proposed "TCM phenomics" in 2016, creating a translational bridge between TCM theory and molecular-level multi-omics technologies [10, 34, 35]. TCM phenomics retains the core attributes of general phenomics: (i) its central mission is to fuse multi-dimensional measurements—from molecules to the whole organism—to capture the full spectrum of gene-environment interactions, rendering the field intrinsically multi-modal [36]. (ii) Its outputs simultaneously illuminate disease commonalities and individual differences, offering exceptional potential for precision medicine. Moreover, TCM phenomics embeds classical TCM concepts, subdividing into clinical TCM phenomics and herbal TCM phenomics, and incorporates diverse data types—symptoms, tongue appearance, pulse patterns, herbal decoction characteristics, and prescription rules, to portray organismic states comprehensively. The resulting high-dimensional datasets necessitate sophisticated analytical and network construction capabilities, making the infusion of AI algorithms imperative for discipline advancement.

4.1.3 Current landscape of AI applications in TCM multi-omics The integration of heterogeneous multi-omics represents the most significant challenge in TCM research. AI offers a powerful suite of tools to overcome this challenge, enabling the harmonization of disparate datasets, the reconstruction of interaction networks, and the development and validation of predictive models. Through these processes, AI transforms extensive, high-dimensional modern datasets into coherent representations of TCM's continuous theoretical framework, bridging the holistic TCM paradigm with computable network-based scientific models. At the algorithmic level, both machine learning and deep learning have made decisive contributions. A diverse array of algorithms is applied to standardise and extract features from heterogeneous TCM multi-omics data—spanning genomics, metabolomics, tongue images, pulse signals, and more thereby elucidating the dynamic regulatory relationships among "herbal formula/intervention-biological network-clinical phenotype". This enables precise forecasting of syndrome evolution, therapeutic efficacy and safety. Representative algorithms and their use-cases are summarized in Table 2. In broad terms, machine learning techniques excel with small samples and furnish interpretable links between omics data and TCM theory [37], whereas deep learning approaches are adept at uncovering complex, high-dimensional interactions. A flexible, task-specific deployment of these complementary algorithms has become the dominant mode of AI-assisted TCM multi-omics research.

A central mission of TCM multi-omics is to render objectivity and quantification of TCM theoretical construction [5]. The nascent discipline of TCM phenomics—integrated with AI—seeks to construct evidence networks that map microscopic omics data onto macroscopic phenotypic traits while remaining firmly anchored in TCM theoretical principles. TCM phenomics can be further subdivided into Chinese herbal medicine phenomics and clinical TCM phenomics, both demonstrating tangible advancements through AI integration.

Chinese herbal medicine phenomics focuses on how environmental factors and growth cycles dynamically influence the active constituents of medicinal plants. Researchers typically combine multi-omics with AI to evaluate crude-drug quality and conduct pharmacological investigations [38, 39]. By integrating multimodal data—images, mass spectra, chromatograms and spectroscopic profiles—AI assists in rapid classification, geographical traceability, detection of toxic adulterants, and qualitative/quantitative analysis of bioactive compounds [40, 41]. Indeed, pharmacological studies of Chinese materia medica are moving beyond single-compound analyses

Table 2 Tasks and algorithm utilization in multi-omics research of TCM

Task	Models class	Representative algorithm/architecture	Typical scenario	Key advantage
Multi-omics data integration & alignment	Machine learning	PCA/t-SNE/UMAP	High-dimensional reduction & visualisation	Fast computation; suitable for small samples
		Feature engineering/ multiple-kernel learning	Fusion of heterogeneous data (e.g., tongue images and pulse signals)	Preserves original data structure
	Deep learning	Autoencoder	Joint embedding of multi-omics & multi-modal data	Learns shared features across omics & modalities automatically
Cross-omics association network construction	Machine learning	Random forest/XGBoost	Key pathway identification	High interpretability; easily maps omics features to TCM theory
	Deep learning	GNN	Herb-compound-target-pathway-phenotype networks	Multi-layer modelling in a single framework
	Hybrid	GNN + attention mechanism	Pinpointing core therapeutic targets	Visualises synergistic weights of formula constituents; aligns with TCM concepts
Clinical prediction modelling	Deep learning	Transformer	Dynamic pattern tracking	Handles temporal & textual data via NLP
		Generative adversarial network (GAN)	Novel formula generation	Multi-objective optimisation under TCM theoretical constraints
	Machine learning	XGBoost	Efficacy assessment	Gradient-boosted trees handle small, noisy clinical datasets
		Cox proportional-hazards model	Survival analysis	Quantifies TCM protective effects with censoring

toward integrative investigations that capture dynamic compositional changes and network principles governing formula compatibility.

Clinical TCM phenomics, encompassing macroscopic phenotypic data from the four canonical diagnostic methods (inspection, auscultation and olfaction, inquiry, and palpation), exhibits an inherently multimodal structure, becoming a focal point for AI innovation. Within this framework, AI employs multimodal sensors coupled with deep learning algorithms to achieve objective quantification of traditional indicators. Computer vision converts tongue and facial images into numerical descriptors [42, 43]. Voice-print recognition combined with Transformer architectures extracts features from cough and vocal tone [44]. Large-language models (LLMs) construct structured frameworks for TCM inquiry [45]. Multi-sensor pulse analyzers capture spatiotemporal pulse patterns whose 3D (time-frequency-region) signatures are classified via machine learning annotation [46]. Across these applications, machine learning enhances the efficiency and accuracy of heterogeneous TCM multi-omics data and of identifying critical network nodes. To accommodate the nonlinear, complex-system nature of TCM theory, deep learning algorithms are particularly advantageous for multidimensional mapping networks between TCM concept construction and biomedical entities, as well as the potential associations. The convergence of phenomics and AI is emerging as an innovation engine for TCM research, accelerating the objectification and modernization of TCM theory through precise characterisation of herbal constituents and clinical phenotypes, and charting new avenues for TCM research and application.

4.2 Prospects for AI application in TCM multi-omics studies

In 2016, the Chinese government introduced several key policy documents, including the 13th Five-Year Plan for National Strategic Emerging Industries Development [47] and the Development Plan for New Generation Artificial Intelligence [48]. These initiatives have significantly accelerated the AI application across multiple sectors, including healthcare sectors. Advances in high-throughput and mass spectrometry technologies have enabled comprehensive, multi-dimensional biological systems analysis, and have been successfully commercialized in China. For instance, Beijing Genomics Institute’s launch of next-generation sequencing platforms in 2016 has spurred related scientific, translational, and clinical research [6]. Overall, these factors—policy support, technological innovation, and disciplinary evolution—collectively drove the surge in TCM multi-omics and AI research output around 2016. The Healthy China 2030 Outline issued in 2016 by the Central Committee of the Communist Party of China and the State Council [49], has also placed a strong emphasis on integrating TCM with modern biology, medicine, omics, and computer science. To meet the contemporary demands for objective and personalized healthcare, TCM must continuously adapt to modern societal developments by actively engaging with these disciplines. The natural progression of TCM’s development requires deeper cross-disciplinary research.

Bibliometric analysis revealed that current institutional integration of AI into TCM multi-omics research predominantly involves top-tier TCM universities and tertiary hospitals located in first-tier cities, highlighting issues of institutional homogeneity, limited collaboration and data bias. Moreover, the depth of cross-disciplinary integration remains superficial, with the majority of studies employing basic AI techniques, such as data mining and cluster analysis to address straightforward data stratification problems within the overlap of AI and mathematical statistics. To advance the field, targeted institutional support should be strengthened for primary research entities, while multi-center and cross-regional collaborations should be systematically facilitated. And more industry standards and norms for TCM should be established. Additionally, the establishment of information-sharing platforms and the release of open-source TCM datasets are essential to promote the integration of TCM multi-omics. These measures are fundamental for achieving objective quantification of TCM therapeutic mechanisms and establishing globally recognized evaluation paradigms.

The rapid advancement of high-throughput multi-omics platforms has precipitated a paradigm shift in biomedical research, transitioning from single-omics analyses to integrative systems-level approaches, with AI for high-efficiency, multi-dimensional data analysis becoming integral to this shift. Given TCM’s reliance on omics technologies for modernisation, the AI integration is equally critical. Bibliometric analysis showed that pre-2015 research largely focused on single-omics data analysis, such as genomics-based predictions of herbal compound properties, and was dominated by theoretical exploration and small-sample validation. Since 2018, three key transformations have occurred in multi-omics integration: (i) transition from single-omics to multi-omics data fusion, leveraging AI to align TCM’s holistic philosophy with systems biology; (ii) extension from basic research to clinical translation, particularly in developing intelligent correlation models between TCM pattern identification and molecular subtypes; (iii) expansion from technology validation to real-world applications such as smart monitoring of herbal drug quality, personalised medicine, and clinical decision-support systems. Through bibliometric analysis and keyword frequency studies, four major research hotspots have emerged: (i) AI-assisted molecular subtyping and biomarker discovery for TCM patterns; (ii) mechanisms of herbal formulae with integrated multi-omics data; (iii) AI-driven precision matching systems for "disease-pattern-formula" relationships to support clinical decision-making; (iv) multimodal fusion models for disease prediction integrating TCM and western medicine.

5 Conclusion

Employing bibliometric mapping, this study delineated the evolutionary trajectory of AI-driven TCM multi-omics: a metabolomics-machine learning dyad dominated, while deep learning and multimodal fusion were rapidly ascending. Collaboration networks exhibited a "top-tier TCM universities-led yet regionally fragmented" topology, calling for inter-institutional, interdisciplinary synergy and open data ecosystems. Additionally, we taxonomized current algorithmic implementations, furnishing a roadmap for precision TCM research.

Fundings

General Project of Scientific Research of Hunan Provincial Education Department (22C0191), General Project of University-level Scientific Research of Hunan University of Chinese Medicine (Z2023XJYB21), and Hunan Provincial Degree and Graduate Education Reform Research Project (2024JGYB157).

Competing interests

Qinghua PENG is an editorial board member for *Digital Chinese Medicine* and was not involved in the editorial review or the decision to publish this article. All authors declare that there are no competing interests.

References

[1] LIU SY, LIN XC, CHENG S, et al. Multi-omics big data and medical advancements. Science & Technology Review, 2024, 42(12): 51–74.

[2] LIU JF, LI WL, WANG L, et al. Multi-omics technology and its applications to life sciences: a review. Chinese Journal of Biotechnology, 2022, 38(10): 3581–3593.

[3] WANG XQ, DU X, CAO Q, et al. Interpretation and application of terms related to phenomics of traditional Chinese medicine. Journal of Nanjing University of Traditional Chinese Medicine, 2023, 39(3): 210–220.

[4] WANG XY, QU HZ, FANG XD. Omics big data and medical artificial intelligence. Hereditas, 2021, 43(10): 930–937.

[5] PENG HY, LU CL, HUANG XQ, et al. Analysis of the current status and prospects of multi-omics technologies in the field of traditional Chinese medicine. Journal of Traditional Chinese Medicine, 2024, 65(8): 775–781.

[6] WANG YJ, LU ZC, CHEN JY, et al. Development of the high-throughput sequencing technology and its application in clinical detection. Journal of Xiamen University (Natural Science), 2021, 60(5): 811–820.

[7] LIU XX, CHEN SY, CAI YT, et al. Analysis of research situation on DNA barcoding of Chinese materia medica based on bibliometrics. Chinese Traditional and Herbal Drugs, 2024, 55(15): 5201–5211.

[8] KREITMAIER P, KATSOULA G, ZEGGINI E. Insights from multi-omics integration in complex disease primary tissues. Trends in Genetics, 2023, 39(1): 46–58.

[9] HASIN Y, SELDIN M, LUSIS A. Multi-omics approaches to disease. Genome Biology, 2017, 18(1): 83.

[10] RITCHIE MD, HOLZINGER ER, LI RW, et al. Methods of integrating data to uncover genotype–phenotype interactions. Nature Reviews Genetics, 2015, 16(2): 85–97.

[11] TIAN YN, WANG XR, YAO XJ, et al. Predicting molecular properties based on the interpretable graph neural network with multistep focus mechanism. Briefings in Bioinformatics, 2023, 24(1): bbac534.

[12] POPLIN R, CHANG PC, ALEXANDER D, et al. A universal SNP and small-indel variant caller using deep neural networks. Nature Biotechnology, 2018, 36(10): 983–987.

[13] CUTLER KJ, STRINGER C, LO TW, et al. Omnipose: a high-precision morphology-independent solution for bacterial cell segmentation. Nature Methods, 2022, 19(11): 1438–1448.

[14] CHEN TQ, GUESTRIN C. XGBoost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016: 785–794.

[15] SHI JH, CHEN L, YUAN XR, et al. A potential XGBoost Diagnostic Score for Staphylococcus aureus bloodstream infection. Frontiers in Immunology, 2025, 16: 1574003.

[16] MINOT M, REDDY ST. Meta learning addresses noisy and under-labeled data in machine learning-guided antibody engineering. Cell Systems, 2024, 15(1): 4-18.e4.

[17] KUMAR S, PAYNE P RO, SOTIRAS A. Normative modeling using multimodal variational autoencoders to identify abnormal brain volume deviations in alzheimer's disease. Proceedings of SPIE—the International Society For Optical Engineering, 2023, 12465: 1246503.

[18] JI XX, ZHUANG ZH, LONG HY. Application of systems biology in the modernization of Traditional chinese Medicine. Journal of Li-shizhen Traditional Chinese Medicine, 2025, 36(3): 520–524.

[19] LYU YH, XIE L, CHEN W, et al. Application of metabonomics in study of traditional Chinese medicine syndrome: a review. China Journal of Chinese Materia Medicine, 2022, 47(2): 367–375.

[20] HE Y, ZHOU ZK, LI WJ, et al. Metabolic profiling and pharmacokinetic studies of Baihu Guizhi Decoction in rats by UFLC-Q-TOF–MS/MS and UHPLC-Q-TRAP-MS/MS. Chinese Medicine, 2022, 17(1): 117.

[21] JIANG XY, LI T, ZHOU YH, et al. A new direction in metabolomics: analysis of the central nervous system based on spatially resolved metabolomics. TrAC Trends in Analytical Chemistry, 2023, 165: 117103.

[22] ZHOU Y, XU JY, LI QZ, et al. Advances in the proteomics research on traditional Chinese medicine for the prevention and treatment of coronary heart disease. World Chinese Medicine, 2024, 19(1): 109–115.

[23] MU XY, SUN P, XUE L, et al. Discussion on the research

progress of proteomics in traditional Chinese medicine based on holistic concept. China Journal of Traditional Chinese Medicine and Pharmacy, 2021, 36(11): 6585–6588.

[24] LIU YQ, ZHAN SY, RUAN YE, et al. Application progress of proteomic in pharmacological study of Chinese medicinal formulae. China Journal of Chinese Materia Medicine, 2017, 42(20): 3873–3879.

[25] HE HQ, CHEN G, GAO JL, et al. Research process of transcriptomics in traditional Chinese medicine syndrome. Modernization of Traditional Chinese Medicine and Materia Medica-World Science and Technology, 2018, 20(1): 1–6.

[26] WU Q, QIU WQ, SONG M, et al. Study on transcriptomics and traditional Chinese medicine syndrome. China Journal of Traditional Chinese Medicine and Pharmacy, 2017, 32(9): 4094–4097.

[27] MENG XT, LIU YY, WANG S, et al. Transcriptomics and superposition analysis based on formula dismantling reveal the mechanism of anti-hepatocellular carcinoma effect of huazheng sanji formula. Pharmacology and Clinics of Chinese Materia Medica, 2025. doi:10.13412/j.cnki.zyyl.20250319.001.

[28] DONG Y, LI J, ZHANG ZP, et al. Thoughts and Methods of Transcriptomics for Targets and Mechanisms of Traditional Chinese Medicine. World Chinese Medicine, 2023, 18(14): 2081–2087.

[29] SUN XQ, YANG B, LI YB. Research progress on quality evaluation of multi-source traditional Chinese medicine. Chinese Traditional and Herbal Drugs, 2024, 55(12): 4214–4224.

[30] WANG ZW, ZHANG JJ, SHAN TY, et al. Identification and analysis of the bHLH transcription factor family involved in polysaccharide biosynthesis pathway in Polygonatum cyrtonema. Journal of Biology, 2025, 42(3): 51–58.

[31] LI YR, LI RX, FAN HJ, et al. Application research advances of transecriptome technology in the field of traditional Chinese medicine. Journal of Li-shizhen Traditional Chinese Medicine, 2022, 33(4): 943–947.

[32] YANG PH, JIN LJ, LIAO J, et al. Modern research on Chinese medicine based on single-cell omics: technologies and strategies. China Journal of Chinese Materia Medicine, 2022, 47(15): 3977–3985.

[33] LI JY, WANG Y. Application of human phenomics in coronary heart disease and Chinese medicine research. Journal of Practical Traditional Chinese Internal Medicine, 2025, 39(4): 103–107.

[34] HOULE D, GOVINDARAJU DR, OMHOLT S. Phenomics: the next challenge. Nature Reviews Genetics, 2010, 11(12): 855–66.

[35] TONG XL, ZHAO LH, ZHANG BX. Phenomics and the innovative development of traditional Chinese medicine. Science, 2022, 74(1): 22–25.

[36] TONG XL, HE LS, ZHAO LH. Assisted precision medicine: the mutual reference and fusion of phenomics and traditional Chinese medicine. Chinese Journal of Integrated Traditional and Western Medicine, 2018, 38(3): 368–370.

[37] WANG N, YU CC, YANG H, et al. Preliminary exploration of multi-omics data fusion methods for high-dimensional small-sample datasets in traditional Chinese medicine. China Journal of Chinese Materia Medica, 2025, 50(1): 278–284.

[38] FAN JY, QIU HX, YU PS, et al. Exploration and thinking of new technique and method of toxicity test of traditional Chinese medicine. Journal of Shenyang Pharmaceutical University, 2025, 42(5): 406–411.

[39] ZHANG N, WANG XY, HAN B, et al. Research progress of artificial intelligence technology in pharmacology of traditional Chinese medicine. Acta Pharmaceutica Sinica, 2025, 60(3): 550–558.

[40] WANG GH, ZHANG ZH, XI MJ, et al. Construction of an artificial intelligence model and application for an automatic recognition of traditional Chinese medicine herbals based on convolutional neural networks. Chinese General Practice, 2025, 28(9): 1128–1136.

[41] WANG CC, ZHANG XC, GU ZC, et al. Applications of artificial intelligence in the detection of traditional Chinese herbal medicines and prepared slices. Strategic Study of Chinese Academy of Engineering, 2024, 26(2): 245–254.

[42] CHEN EN, LI SY, HU YQ, et al. Research progress in image acquisition and color information analysis of tongue diagnosis in traditional Chinese medicine. China Journal of Traditional Chinese Medicine and Pharmacy, 2024, 39(7): 3586–3589.

[43] JIANG TT, ZHAO YP, ZHAO YF, et al. Analysis and exploration of the application of machine vision technology in traditional Chinese medicine intelligent devices. Journal of Basic Chinese Medicine, 2024, 30(3): 407–412.

[44] MEN SY, CHEN LJ, HUANG XM, et al. Research on sound diagnosis constitution identification based on deep learning transformer and transfer learning. Modernization of Traditional Chinese Medicine and Materia Medica-World Science and Technology, 2025, 27(6): 1750–1757.

[45] LIU YQ, LI Y, YANG HJ, et al. Construction and validation of a large language model-based intelligent pre-system for traditional Chinese medicine. Journal of Traditional Chinese Medicine, 2025, 66(9): 895–900.

[46] YI K, ZHANG MD, GUO S, et al. Research review on data annotation for intelligent pulse diagnosis devices in traditional Chinese medicine. Shanghai Journal of Traditional Chinese Medicine, 2024, 58(10): 5–10

[47] State Council, P. R. China. Notice on issuing the 13th Five-Year Plan for the Development of Strategic Emerging Industries. Beijing: State Council. 2016.12.19. Available from: http://www.gov.cn/zhengce/content/2016-12/19/content_5150090.htm.

[48] State Council, P. R. China. Development plan of new generation artificial intelligence. 2017.07.20. Available from: http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm.

[49] CPC Central Committee & State Council. Healthy China 2030 Planning Outline. 2016.10.25. Available from: http://www.gov.cn/zhengce/2016-10/25/content_5124174.htm.

中医药研究中多组学与人工智能结合应用的进展与前景分析

刘桂成[a], 龙茜[a], 彭清华[a, b*], 田赛男[a*], 胡淑娟 a[a*]

a. 湖南中医药大学中医学院, 湖南 长沙 410208, 中国
b. 湖南中医药大学中医诊断研究所, 湖南 长沙 410208, 中国

【摘要】目的 利用文献计量学方法，分析人工智能（AI）结合中医药多组学研究领域的研究热点、发展趋势及面临的问题。**方法** 研究者检索了中国期刊全文数据库（CNKI）、万方数据知识服务平台、中文科技期刊数据库（维普）、超星中文学术期刊全文数据库，以及 PubMed 和 Web of Science，收集了从建库截至 2024 年 12 月 31 日的 AI 结合中医药多组学研究相关主题文献。使用 Excel 2021、NoteExpress v4.0.0、CiteSpace 6.3.R1 软件对文献数据进行可视化分析，包括发文趋势、作者、发文量、研究机构和关键词。并对中医药多组学研究中 AI 应用模式进行分类总结。**结果** 纳入文献共计 1106 篇（中文 932 篇，英文 174 篇）。自 2010 年起，相关文献发文量持续增长，且在 2016 年后增长速度明显加快。学科在国内形成了多个机构区域合作网络，核心机构包括北京中医药大学、中国中医科学院、上海中医药大学、南京中医药大学等。关键词分析显示，AI 技术在中医药多组学研究中的应用主要集中在代谢组学研究，常用聚类分析、数据挖掘等算法，以中药研究为主，其次为复方和疾病证候研究。**结论** 机器学习方法是当前 AI 在中医药结合多组学研究领域最主要的结合形式，被用于处理组学数据与探索数据潜在规律。中医药领域除了研究使用高通量技术的组学信息，还纳入中药、临床表型等数据，向多组学联合分析、数据高纬度、信息多模态方向发展，深度学习方法是学科热点趋势。

【关键词】中医药；多组学；人工智能；文献计量学；CiteSpace