



OPEN ACCESS

EDITED BY

Michael Heinrich,
University College London, United Kingdom

REVIEWED BY

Xin Chen,
Tongji University, China
Yaolei Li,
National Institutes for Food and Drug Control,
China
Ziming Yin,
University of Shanghai for Science and
Technology, China

*CORRESPONDENCE

Qingzhong Liu,
✉ liuqingzhong@shutcm.edu.cn

[†]These authors have contributed equally to this work and share first authorship

RECEIVED 07 December 2024

ACCEPTED 25 March 2025

PUBLISHED 15 April 2025

CITATION

Li Y, Liu X, Zhou J, Li F, Wang Y and Liu Q (2025) Artificial intelligence in traditional Chinese medicine: advances in multi-metabolite multi-target interaction modeling. *Front. Pharmacol.* 16:1541509. doi: 10.3389/fphar.2025.1541509

COPYRIGHT

© 2025 Li, Liu, Zhou, Li, Wang and Liu. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Artificial intelligence in traditional Chinese medicine: advances in multi-metabolite multi-target interaction modeling

Yu Li[†], Xiangjun Liu[†], Jingwen Zhou, Fengjiao Li, Yuting Wang and Qingzhong Liu^{*}

Department of Clinical Laboratory, Shanghai Municipal Hospital of Traditional Chinese Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai, China

Traditional Chinese Medicine (TCM) utilizes multi-metabolite and multi-target interventions to address complex diseases, providing advantages over single-target therapies. However, the active metabolites, therapeutic targets, and especially the combination mechanisms remain unclear. The integration of advanced data analysis and nonlinear modeling capabilities of artificial intelligence (AI) is driving the transformation of TCM into precision medicine. This review concentrates on the application of AI in TCM target prediction, including multi-omics techniques, TCM-specialized databases, machine learning (ML), deep learning (DL), and cross-modal fusion strategies. It also critically analyzes persistent challenges such as data heterogeneity, limited model interpretability, causal confounding, and insufficient robustness validation in practical applications. To enhance the reliability and scalability of AI in TCM target prediction, future research should prioritize continuous optimization of the AI algorithms using zero-shot learning, end-to-end architectures, and self-supervised contrastive learning.

KEYWORDS

artificial intelligence, algorithms, traditional Chinese medicine, active metabolites, therapeutic targets

1 Introduction

Traditional Chinese Medicine (TCM), with its millennia-old history, has demonstrated significant therapeutic efficacy across East Asia and is increasingly gaining global recognition. In recent years, natural products account for over 60% of the world's medicines (Lin et al., 2022; Zhu X. et al., 2022). Notably, several Western pharmaceuticals, such as artemisinin from *Artemisia annua* for malaria and ephedrine from *Ephedra* for asthma, trace their origins to TCM (Kong et al., 2023). Conventional drug discovery, which predominantly focuses on single-target interactions, often falls short in treating complex diseases like diabetes and cancer, frequently resulting in limited efficacy and significant side effects (Zhang R. et al., 2019). This has prompted a paradigm shift towards a multi-metabolites multi-target approach, which aligns more closely with TCM's holistic principles. In contrast to single-compound Western medicines, TCM utilizes the synergistic effects of multiple active metabolites, achieving therapeutic outcomes through complex, multi-target interactions (Heinrich et al., 2022; Liu J. et al., 2023). Nevertheless, conventional approaches—network pharmacology, experimental screening, and static

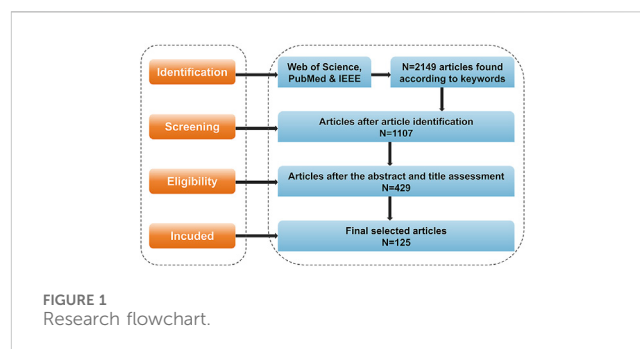
correlation analyses—are inadequate in capturing the dynamic, non-linear nature of multi-metabolite relationships, thus constraining their applicability in modern drug discovery.

Recent advancements in artificial intelligence (AI) have transformed the study of multi-metabolite interactions in TCM, with machine learning (ML) and deep learning (DL) technologies reaching sufficient maturity for analyzing complex interactions between active metabolites and their multiple targets (Wang et al., 2021; Ma et al., 2023; Zhang et al., 2023a). The unique capabilities of AI in processing large-scale data, recognizing complex patterns, and integrating multi-dimensional datasets have rendered it an indispensable tool in TCM research (Seetharam et al., 2019). ML algorithms excel at identifying potential interaction patterns from vast datasets, while DL takes this further by automatically learning higher-order features to capture complex relationships between active metabolites and their multiple targets (Calderaro et al., 2022).

Beyond data processing and pattern recognition, AI's integration into TCM research extends to the synthesis of multi-omics data, including genomics, proteomics, metabolomics, and spatial omics (Razzaq et al., 2022). Through the utilization of AI's advanced analysis capabilities, these heterogeneous data sources are integrated to construct complex network models that capture the intricate relationships between multiple metabolites and targets (Pan et al., 2024). This comprehensive integration enhances our understanding of the synergistic effects of active metabolites and significantly improves research precision, providing robust data support for investigating TCM holistic principles and efficacy mechanisms (Hua et al., 2024). The study explores the application of AI-driven biological analysis in target research, incorporating diverse TCM target databases and multi-omics approaches, including epigenetics, genomics, proteomics, metabolomics, and spatial omics. Furthermore, the study evaluates the deployment of various AI algorithms—such as ML, DL, and cross-modal data fusion—in multi-target models, assessing their suitability, advantages, and limitations in TCM research. By synthesizing current challenges, technological limitations, and emerging opportunities, this study provides valuable insights into future directions for integrating AI with TCM, particularly in understanding the complex relationships between active metabolites and their therapeutic targets.

2 Research methodology

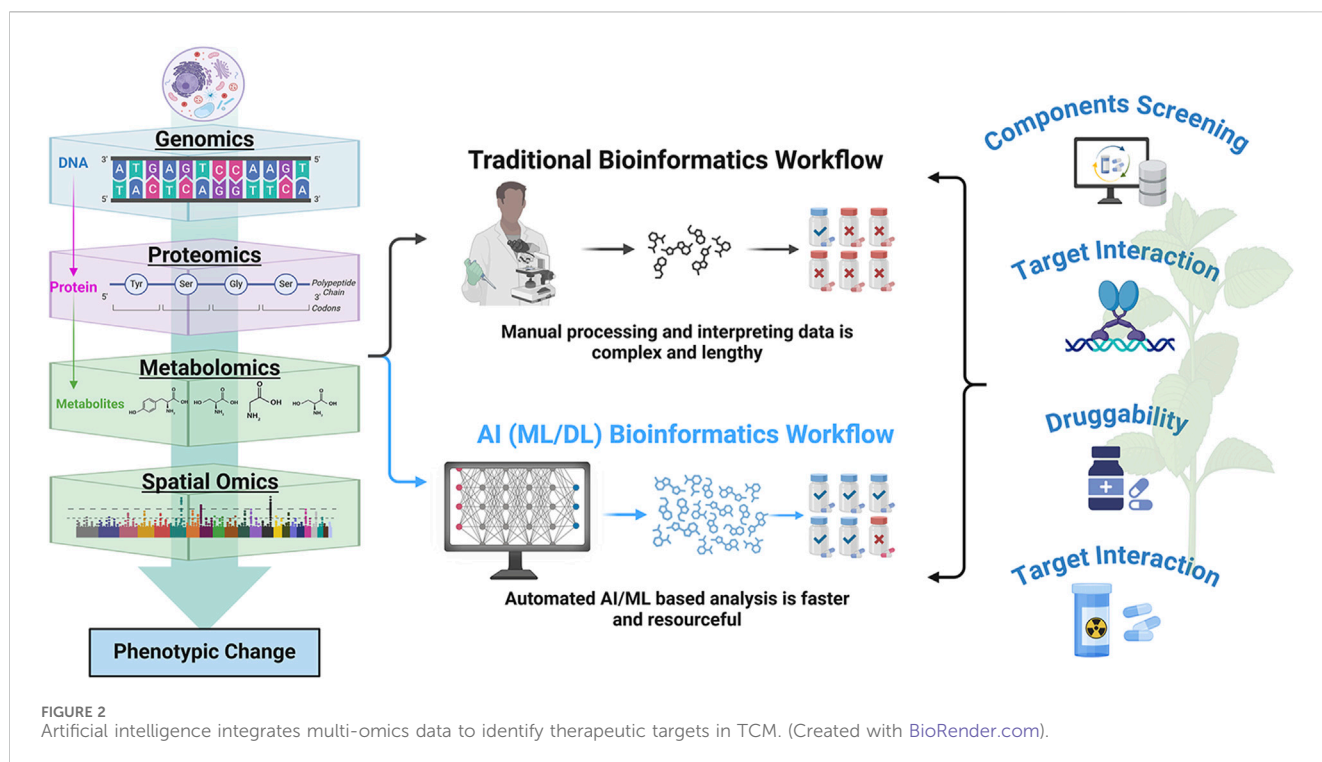
This study conducted a systematic literature review to examine the application of AI, ML, and DL technologies in TCM target research. A hybrid methodology combining the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines proposed by Moher et al. and the Systematic Literature Review (SLR) framework established by Manuel et al. was employed (Moher et al., 2015; Muhammad et al., 2021). The methodological architecture encompassed four primary procedures: formulation of research objectives, definition of scope, selection of literature, and validation. The systematic review aimed to identify and analyze current applications of AI technologies in TCM target discovery. Three databases (Web of Science, PubMed, and IEEE) were selected based on their rigorous academic standards and



established reputation as reliable sources for scholarly research. Preliminary investigations indicated that additional database searches would not significantly enhance retrieval outcomes, thus justifying this selection. Search parameters combined the following keywords: “Artificial Intelligence, Algorithm, Neural Network, Machine Learning, Deep Learning” combined with “Traditional Chinese Medicine, Target Identification, Drug Development, Botanical drugs.” The screening process entailed an initial evaluation of article titles and abstracts, followed by the elimination of duplicates and studies not related to TCM. A temporal constraint was applied to include literature published between January 2010 and January 2025, and only peer-reviewed journal articles were considered. Following a thorough evaluation of the full texts, 125 papers were deemed eligible for inclusion in the study. The complete methodology flowchart is illustrated in Figure 1.

3 Complexity of multi-metabolite multi-target interactions

The fundamental difference between TCM and Western medicine lies in their respective approaches to therapeutic formulation. TCM utilizes balanced formulations derived from multiple natural sources, including plants, animals, and minerals. These natural matrices contain active metabolites, such as alkaloids, polyphenols, polysaccharides, flavonoids, and terpenoids, that engage in multi-target biological interactions (Zhang Y. et al., 2024). The therapeutic efficacy of TCM is not derived from the activity of individual metabolites, but rather from the optimized interplay between biological metabolites (Li D. et al., 2022). This characteristic necessitates precise calibration of dosage ratios and pharmacokinetic parameters to ensure the desired therapeutic outcome. The multi-metabolite multi-target interactions have demonstrated particular clinical value in the management of complex pathologies. A notable example is the Xiangdan injection, which exemplifies multi-metabolite principles by enhancing cerebral perfusion through complementary metabolic pathways via flavonoid-saponin-polysaccharide coordination (Gao F. et al., 2024). Similarly, the Shugan Lidan Xiaoshi formulation integrates quercetin, lignans, and paeoniflorin to concurrently mitigate inflammation and oxidative stress in acute pancreatitis (J et al., 2024). Experimental studies have demonstrated the dual modulation of p38MAPK signaling and cytokine cascades (TNF- α , IL-6) in sepsis management by the Yantiao formulation (Zhu et al., 2024), thus illustrating TCM capacity for multi-pathway intervention.



However, current TCM research confronts methodological limitations. Conventional experimental paradigms inadequately characterize metabolite synergies, while clinical trial reproducibility suffers from formulation variability. Conventional reductionist approaches, which focus on single targets, fail to capture the emergent therapeutic properties of multi-metabolite systems. Integration of AI presents transformative solutions for multi-metabolite multi-target analysis. ML and DL algorithms enable systematic mapping of nonlinear relationships in multidimensional pharmacological data (Holm et al., 2021; Li X. et al., 2022). The integration of high-throughput virtual screening platforms with molecular dynamics simulations has been shown to facilitate the identification of active metabolites (Zhou E. et al., 2024). Network pharmacology tools, such as the TCMFP algorithm, have been employed to optimize formulation design through disease-specific target matching (Niu et al., 2023). Predictive pharmacokinetic models have been developed to enhance formulation optimization by simulating *in vivo* metabolic trajectories (Li et al., 2024).

These computational innovations enable rigorous analysis of TCM's complexity while preserving its holistic therapeutic framework. The integration of traditional Chinese pharmacopeia with AI-driven methodologies promises transformative advances in understanding polypharmacological systems.

4 Scope of AI biological analysis for target investigations in TCM

The exponential growth of multi-omics data, coupled with the increasing availability of comprehensive databases, has established a robust foundation for the development of sophisticated drug target

inference algorithms. This convergence of AI and innovative experimental techniques represents a highly efficient paradigm for drug discovery.

4.1 Multi-omics technologies

The comprehensive analysis of multi-omics data, encompassing epigenomics, genomics, proteomics, metabolomics, and spatial omics, offers a robust approach for elucidating drug mechanisms of action and identifying potential therapeutic targets (Figure 2). Table 1 provides a comprehensive list of commonly employed databases designed to facilitate the integration of multi-omics datasets.

Epigenomics focuses on the study of reversible chemical modifications to DNA and associated proteins that modulate gene expression without altering the underlying DNA sequence. Pharmacological agents capable of interacting with DNA can profoundly influence transcriptional processes, replication fidelity, and overall genetic expression, consequently impacting physiological functions (Chen et al., 2022). For instance, Ming et al. employed epigenomic data, encompassing DNA methylation and histone modification networks, to demonstrate that curcumin induces apoptosis and exerts anticancer effects by inhibiting DNA methyltransferase (DNMT) and histone deacetylase (HDAC) activity (Ming et al., 2022). Conversely, genomics utilizes high-throughput molecular, genetic, and cellular techniques to assess gene function. This approach finds wide application in genotype-phenotype association analysis, biomarker discovery for patient stratification, gene function prediction, and mapping of biochemically active genomic regions (McDonagh et al., 2024). For instance, Xu et al. applied a consensus clustering algorithm

TABLE 1 Commonly used repositories related to genomics, proteomics, metabolomics, and multi-omics.

Database	Full name	Description	Web link	References
Genomics				
GO	Gene Ontology	GO contains functional information for genes from over 460,000 species	http://www.geneontology.org	The Gene Ontology Consortium (2017)
GEO	Gene Expression Omnibus	GEO repository archives and freely distributes microarray, NGS and other forms of high-throughput functional genomic data	http://www.ncbi.nlm.nih.gov/geo/	Barrett et al. (2013)
GTEx	Genotype-Tissue Expression	GTEx provides gene expression profiles in different tissue types	https://gtexportal.org/home/	Consortium (2013)
ENCODE	Encyclopedia of DNA Elements	ENCODE identifies and catalogs all functional elements of the human genome previously mapped by the HGP.	https://www.encodeproject.org/	Colwell (2016)
DisGeNET	DisGeNET	DisGeNET is one of the largest collections of genes and variants involved in human disease	http://www.disgenet.org	Piñero et al. (2017)
Ensembl	Ensembl	Ensembl is unique in its flexible infrastructure for access to genomic data and annotation	https://www.ensembl.org	Cunningham et al. (2022)
Gene	Gene	Gene focuses on viral, prokaryotic, and eukaryotic NCBI RefSeq genomes	www.ncbi.nlm.nih.gov/gene/	Brown et al. (2015)
CCL	Cancer Cell Line Encyclopedia	CCL contains gene expression, chromosome copy number, and massively parallel sequencing data from 947 human cancer cell lines	www.broadinstitute.org/ccle	Barretina et al. (2012)
TCGA	The Cancer Genome Atlas	TCGA collects exome sequencing data of more than 11,000 cancer samples	https://portal.gdc.cancer.gov/	Ganini et al. (2021)
Proteomics				
PDB	Protein Data Bank	PDB focuses on ligand binding site in ligandable proteins	http://bioinfo-pharma.u-strasbg.fr/scPDB/	Desaphy et al. (2015)
STRING	STRING	STRING integrates protein-protein interactions-both physical interactions and functional associations	https://string-db.org/	Szklarczyk et al. (2023)
UniProt	Universal Protein Knowledgebase	UniProt provides a rich and accurately annotated protein sequence knowledgebase	http://www.uniprot.org	Apweiler et al. (2004)
TTD	Therapeutic Target Database	Therapeutic target database describing target druggability information	https://idrblab.org/ttd/	Zhou et al. (2024b)
Metabolomics				
HMDB	Human Metabolome Database	The world's largest and most comprehensive, organism-specific metabolomic database	https://hmdb.ca	Wishart et al. (2021)
KEGG	Kyoto Encyclopedia of Genes and Genomes	KEGG links genomic information with higher order functional information	https://www.genome.ad.jp/kegg/	Kanehisa and Goto (2000)
Reactome	Reactome	A database of reactions, pathways and biological processes	http://www.reactome.org	Croft et al. (2011)
LMPD	LIPID MAPS Proteome Database	An object-relational database of lipid-associated protein sequences and annotations	http://www.lipidmaps.org/	Cotter et al. (2006)
Multi-omics				
OmicsNet	OmicsNet	A web-based platform for multi-omics integration and network visual analytics	http://www.omicsnet.ca	Zhou et al. (2022)
Metabo Analyst	Metabo Analyst	MetaboAnalyst towards more transparent and integrative metabolomics analysis	http://metaboanalyst.ca	Chong et al. (2018)

to identify putative diabetic driver genes and showed that Nfkb1, Stat1, and Ifnrg1 may represent key targets for the anti-diabetic effects of Gegen Qinlian Decoction (Xu et al., 2020).

Proteomics is instrumental in elucidating biological processes by annotating genome sequences, quantifying protein abundance, characterizing post-translational modifications, and mapping protein-protein interactions (PPIs) (Ding et al., 2022; Xiao, 2024). For instance, Xu *et al.* developed a novel serum proteomics platform integrating data-independent acquisition mass spectrometry (dIA-MS) with customized antibody

TABLE 2 Overview of the data statistics and availability of different TCM databases.

Database	Latest update year	Prescriptions	TCM (plants)	Ingredients	Targets	Diseases	Websites	References
TM-MC	2024	5,075	635	34,107	13,992	27,997	https://tm-mc.kr	Kim et al. (2024)
ITCM	2023	25,857	8454	43,430	18,851	11,180	http://itcm.biotcm.net	Tian et al. (2023)
TCM Bank	2023	NA	9,192	61,966	15,179	32,529	https://TCMBank.cn/	Lv et al. (2023b)
TCMIP (ETCM)	2023	48,442	2005	38,298	25,647	8,045	http://www.tcmip.cn/ETCM2/front/#/	Zhang et al. (2023d)
DCABM-TCM	2023	192	194	1816	3,970	4,006	http://bionet.ncpsb.org.cn/dcabm-tcm/	Liu et al. (2023b)
TCM-suite	2022	6692	7322	704,321	19,319	15,437	http://TCM-Suite.AImicrobiome.cn	Yang et al. (2022b)
TCMSID	2022	NA	499	20,015	3270	NA	https://tcm.scbdd.com	Zhang et al. (2022b)
LTM-TCM	2022	48,126	9122	34,967	13,109	NA	http://cloud.tasly.com/#/tcm/home	Li et al. (2022b)
SuperTCM	2021	NA	6516	55,772	543	8634	http://tcm.charite.de/supertcm	Chen et al. (2021)
Hit 2.0	2021	NA	1,250	1,237	2,208	NA	http://hit2.badd-cao.net	Yan et al. (2022)
HERB	2020	NA	7263	49,258	12,933	28,212	http://herb.ac.cn/	Fang et al. (2021)
TCMIO	2020	1493	618	16,437	126,972	NA	http://tcmio.xielab.net	Liu et al. (2020)
YaTCM	2018	1813	6220	47,696	18,697	1907	http://cadd.pharmacy.nankai.edu.cn/yatcm/home	Li et al. (2018)
SymMap	2018	NA	1717	19,595	4302	5235	http://www.symmap.org/	Wu et al. (2018)
TCMID	2018	46,914	8159	25,210	NA	3791	http://www.megabionet.org/tcmid/	Huang et al. (2018)
TCM-Mesh	2017	NA	6235	383,840	4,518,065	6204	http://mesh.tcm.microbioinformatics.org/	Zhang et al. (2017)
CEMTDD	2014	NA	621	4060	2163	210	http://www.cemtdd.com/index.html	Huang and Wang (2014)
TCMSP	2014	NA	499	29,384	3311	837	http://sm.nwsuaf.edu.cn/lsp/tcmsp.php	Ru et al. (2014)
CVDHD	2013	NA	3518	35,230	2395	302	http://pkuxxj.pku.edu.cn/CVDHD	Gu et al. (2013)
TCM Database@Taiwan	2011	NA	453	24,033	NA	NA	http://tcm.cmu.edu.tw/	Chen (2011)

microarrays to identify biomarkers of psoriasis activity. This study revealed a positive association between disease activity and three specific serum proteins: PI3, CCL22, and IL-12B (Xu et al., 2019). Complementary to proteomics, metabolomics enables the qualitative and quantitative analysis of low-molecular-weight metabolites under defined physiological conditions, thereby aiding biomarker discovery (Feng et al., 2020; Xing et al., 2024a). Wu *et al.* constructed a metabolite-pathway-target network using metabolomic data to investigate the effects of Shaoyao Decoction in ulcerative colitis. This analysis identified STAT3, IL-1B, IL-6, IL-2, AKT1, IL-4, ICAM1, and CCND1 as core targets of the decoction, exhibiting significant binding affinities with active metabolites such as quercetin, baicalin, kaempferol, and wogonin (Wu et al., 2022).

As a critical extension of multi-omics frameworks, spatial omics technologies (e.g., 10x Genomics Visium, Nanostring GeoMx) provide unprecedented resolution for mapping molecular distributions within tissue microenvironments, thereby bridging the gap between TCM’s systemic effects and localized target engagement (Yang et al., 2024). For instance, the integration of graph neural networks (GNNs) with spatial transcriptomics facilitates dynamic modeling of ephedrine alkaloid-target interactions across temporal and spatial dimensions (Laubscher et al., 2024). While challenges persist in cross-platform data harmonization and computational scalability, emerging tools such as STUtility and deep spatial transformers demonstrate significant potential for standardizing TCM spatial datasets. This technological

synergy elevates multi-omics research from static network mapping to spatially resolved, dynamic interaction modeling, fundamentally advancing the interpretation of TCM's holistic therapeutic principles (Xu et al., 2023; Zhao Z. et al., 2024).

4.2 TCM databases

In the contemporary landscape of pharmaceutical research and development, target identification stands as a pivotal phase, serving as the cornerstone for subsequent innovation. A multitude of databases has emerged, offering exhaustive information pertaining to both drugs and their associated targets. These databases vary in scope and focus, with some, such as Drug Bank, Drug Central, SuperDrug2, Drug Map, and DRESIS, concentrating on pharmacological data (Wang D. et al., 2016; Griesenauer et al., 2019). In contrast, resources such as Gene Cards, TTD, and DisGeNET are primarily dedicated to target research (Liu X. et al., 2023). Additionally, molecular and bioactivity data are accessible through platforms such as PubChem, ChEMBL, and Binding DB (Kim et al., 2016). Notably, the past decade has witnessed significant growth in specialized TCM databases (Table 2), which have become invaluable resources for TCM research.

These TCM-specific databases include ITCM (Tian et al., 2023), TCM Bank (Lv et al., 2023a), Hit 2.0 (Yan et al., 2022), HERB (Fang et al., 2021), TCMIO (Liu et al., 2020), and TCMIP (ETCM) (Zhang et al., 2023b), SymMap (Wu et al., 2018), TCMID (Huang et al., 2018), TCM Database@Taiwan (Chen, 2011), LTM-TCM (Li D. et al., 2022), and TCMSp (Ru et al., 2014), TCM-Mesh (Zhang et al., 2017), TM-MC 2.0 (Kim et al., 2024), YaTCM (Li et al., 2018), CVDHD (Gu et al., 2013), CEMTDD (Huang and Wang, 2014), TM-MC (Kim et al., 2024), TCM-suite (Yang P. et al., 2022), SuperTCM (Q et al., 2021), TCMSID (Zhang L.-X. et al., 2022), and DCABM-TCM (Liu Z. et al., 2023). These databases collectively provide extensive data on TCM prescriptions, active metabolites, and their associated pathways and diseases, each with distinct emphases. For instance, SymMap links TCM symptoms, botanical drugs, and modern medical symptoms, while YaTCM identifies TCM formulas, protein targets, and pathways. TCMSp provides ADME (absorption, distribution, metabolism, and excretion) data for numerous commonly used metabolites. TCMID focuses on plant-derived chemicals, including their molecular structures, targets, and pharmacological properties, and DCABM-TCM emphasizes *in vivo* metabolites. TM-MC provides information on active metabolites in Northeast Asian traditional medicine, enhancing TCM diversity through systematically curated phytochemical profiles. TCM-suite integrates advanced phytochemical profiling, multi-omics, network pharmacology, and target prediction algorithms in a unified analytical workflow. SuperTCM employs corpus linguistics to decipher botanical drugs and contemporary pathway mapping, thereby bridging the gap between the two. TCMSID provides multi-level interaction networks and detailed metabolite profiles, ensuring structural classification and data reliability through systematic verification processes. These databases offer diverse functionalities, including comprehensive datasets, advanced text mining algorithms, and integration with contemporary biomedical systems. Despite their

differences in data quality and characteristics, these databases collectively advance TCM research by providing reliable, diverse information and specialized tools for drug discovery and integration with modern medicine.

5 Application of AI algorithms in TCM

5.1 Limitations of traditional cyberpharmacology

The rapid accumulation of biological data and the increasing complexity of multidimensional, multi-target research have exposed critical limitations in traditional cyberpharmacology approaches, particularly in handling large-scale heterogeneous datasets. First, conventional methods predominantly rely on experimental data and manual annotation, rendering them time-consuming and inefficient for large-scale data processing (Ye et al., 2020). While active metabolites frequently exhibit dose-responsive effects on individual targets, their polypharmacological actions often manifest nonlinear behaviors contingent on concentration gradients and temporal exposure patterns (Li X. et al., 2022). These phenomena are poorly captured by conventional linear regression models. A critical methodological gap exists in the static modeling frameworks of conventional approaches, which inadequately represent the dynamic network interactions underlying biological systems. This methodological limitation hinders systematic investigation of essential pharmacological mechanisms, including metabolite synergy and antagonism (Y et al., 2024). Collectively, these deficiencies in computational scalability, nonlinear system analysis, and temporal resolution impede mechanistic elucidation of multi-metabolite multi-target strategies. AI integration offers paradigm-shifting solutions to these challenges, as detailed in Figure 3.

5.2 Machine learning algorithms

Machine learning (ML) algorithms demonstrate proficiency in the extraction of critical patterns from high-dimensional data and the deciphering of complex relationships, thereby enabling more precise target prediction (Figure 4). The subsequent sections delineate specific applications of prominent ML algorithms in the domain of TCM research. This encompasses an assessment of their performance in processing high-dimensional data, feature extraction, clustering capabilities, and their applicability and limitations in multi-target prediction.

5.2.1 Support vector machine

The Support Vector Machine (SVM), a widely used linear classifier for binary classification tasks, constructs optimal hyperplanes to maximize interclass margins while achieving high accuracy through distinct category discrimination (Nedaie and Najafi, 2018). The SVM facilitates metabolite classification and pattern recognition by extracting structural and functional features (Heikamp and Bajorath, 2014). In high-dimensional nonlinear interactions, kernel functions enable SVM to project data into higher-dimensional spaces, effectively capturing latent

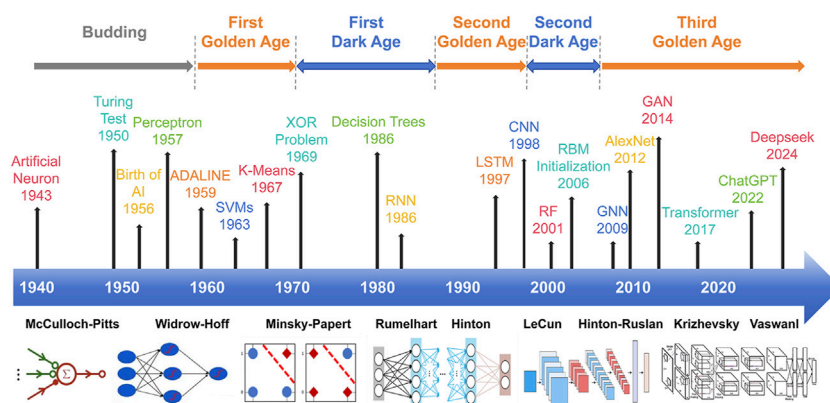


FIGURE 3

Timeline of the development of AI algorithms. Key machine learning (ML) algorithms include PCA, K-Means, Decision Trees, SVM, and RF, while significant deep learning (DL) algorithms include CNN, RNN, LSTM, GAN, and GNN. This timeline showcases the evolution and expansion of AI algorithms from traditional ML to advanced DL models.

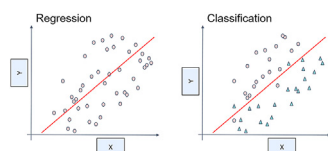


FIGURE 4

Regression and classification in machine learning (ML). The left diagram illustrates regression, where a function models the continuous relationship between the feature X and target Y ; the right diagram depicts classification, where a decision boundary separates data points into distinct categories.

nonlinear patterns (Ma et al., 2023). This approach demonstrates strong generalization and overfitting resistance in small-sample scenarios, though scalability challenges with large datasets and empirical dependency on kernel selection limit broader multi-metabolite applications.

For instance, Cong et al. developed an SVM model that achieved high predictive accuracy in identifying TNF- α converting enzyme (TACE) inhibitors (Cong et al., 2009). However, the SVM model in this study has critical limitations, including a pronounced class imbalance (443 inhibitors vs. 759 non-inhibitors), Gaussian kernel dependency without evaluating polynomial or sigmoidal alternatives, and reliance on static physicochemical descriptors (e.g., topological indices). Similarly, Zhang et al. integrated single-cell sequencing with SVM to identify core biomarkers of myocardial infarction, such as IL-1B and TLR2, and linked them to botanical drugs like *Dan shen*, *San qi*, and *Cha shugen* (Zhang Q. et al., 2022). Despite the efficiency of LASSO regression and SVM-RFE algorithms in feature selection, their reliance on single-center datasets (GSE66360, $n = 99$) that are susceptible to collinearity-driven feature selection bias is a notable limitation. These models are further hindered by their reliance on static descriptors, which lacks dynamic binding insights and inherent interpretability barriers of black-box decision boundaries. To address these limitations, mitigation strategies have been proposed, including SMOTE-

augmented class rebalancing, Bayesian-optimized kernel selection, and molecular dynamics-derived 3D interaction fingerprints. These strategies are complemented by SHAP/LIME frameworks for mechanistic interpretation (Zhang L.-X. et al., 2022). Future research must prioritize multicenter validation with ensemble architectures (e.g., random forest hybrids) and multi-omics integration to enhance biomarker discovery robustness and clinical translatability in TCM research.

5.2.2 Decision tree

Decision tree (DT) algorithms utilize a tree-like structure for classification and regression, employing “if-then” rules (Cheng et al., 2021). While individual DTs are interpretable, they are susceptible to overfitting and noise sensitivity. To address these limitations, ensemble methods have been developed, including Random Forest (RF) (Rhodes et al., 2023), Gradient Boosting Decision Tree (GBDT) (Zhang and Jung, 2021), Extreme Gradient Boosting (XGBoost) (Ching et al., 2022), and LightGBM (Yang R. et al., 2022). These methods combine multiple DTs to improve robustness and predictive accuracy. RF builds multiple independent DTs and aggregates their outcomes, effectively identifying key features and revealing metabolite-target associations (Savargiv et al., 2021). For instance, Chen et al. employed RF and SVM to predict Alzheimer's disease-related metabolites, identifying 3-O-methyl ferulic acid and cyanidanon as potential GSK3 β interactors (Chen et al., 2019). However, traditional QSAR frameworks relying on RF face limitations including dimensionality reduction artifacts from PCA/Lasso feature selection and oversimplified 2D molecular descriptors that neglect 3D steric/electronic interactions captured in CoMSIA models. Validation challenges persist, notably protein rigidity assumptions in molecular docking and insufficient conformational sampling in 100 ns MD simulations.

Conversely, RF demonstrates robustness against noise, requires minimal preprocessing, and is well-suited for high-dimensional, large-scale datasets (Jones et al., 2017). However, its interpretability diminishes with increasing complexity (Zhang Y. et al., 2019). In contrast, XGBoost improves predictive accuracy through iterative optimization, rendering it particularly effective for identifying novel

targets and pharmacological roles of active metabolites (Shin, 2022). For instance, Zheng *et al.* applied XGBoost with Bayesian optimization to identify critical biomarkers for metabolic syndrome and associated TCM indicators (Zheng *et al.*, 2023). However, the developed BO-XGBoost model relies on self-reported TCM indicators collected through questionnaires, which may introduce recall bias and subjective interpretation variability. While hybrid sampling addressed class imbalance, the original dataset's 6.6:1 class ratio might still influence model robustness for minority class predictions. Potential improvements include multicenter studies with wearable-device biometrics to augment population representativeness, longitudinal designs tracking metabolic progression, and hybrid architectures combining blood biomarkers with TCM indicators (Rhodes *et al.*, 2023). Continuous model updating mechanisms and experimental validation remain critical for clinical translation, positioning XGBoost as a powerful yet refinement-demanding tool in modern multi-metabolite multi-target research (Zheng *et al.*, 2023).

5.2.3 Clustering algorithms

Clustering algorithms, a form of unsupervised learning, are extensively utilized for data grouping and pattern recognition. These methods group active metabolites and targets based on shared features or pharmacological properties, enabling the identification of underlying patterns (Gan *et al.*, 2018). Common approaches include *k*-means and hierarchical clustering. *K*-means clustering, a method that assigns data points to a predefined number of clusters (*k*), effectively groups active metabolites with similar chemical structures or pharmacological activities (Li *et al.*, 2023a). In contrast, hierarchical clustering constructs a tree-like hierarchy of relationships through iterative merging or splitting. A notable advantage of hierarchical clustering over *k*-means is its ability to manage complex data structures, a feature particularly beneficial when analyzing such structures (Zavadlav *et al.*, 2019).

Clustering algorithms have been demonstrated to offer a unique value in identifying latent patterns from unlabeled data. However, traditional methods face critical challenges in high-dimensional datasets and noise susceptibility. Conventional approaches, such as *k*-means clustering, frequently employ empirically determined cluster numbers, which can compromise reliability through subjective parameterization. To address these limitations, Han *et al.* developed an improved artificial bee colony (IABC) algorithm that automates cluster center selection, successfully enhancing metabolite clustering (Han *et al.*, 2019). However, this method is sensitive to the choice of Gaussian kernel parameters, particularly the cutoff distance d_c , in heterogeneous density distributions, and it also exhibits premature convergence risks in complex search landscapes. To address these limitations, strategic enhancements can be made, including an adaptive d_c calibration via *k*-nearest neighbor density estimation to optimize cluster identification. Furthermore, a hybridization of IABC with quantum-inspired operators could refine the exploration-exploitation balance, thereby strengthening the algorithmic robustness of the IABC for TCM datasets characterized by variable botanical drug nomenclature and multidimensional interactions (Han *et al.*, 2019).

SVM, DT, and clustering algorithms each offer unique advantages in multi-metabolite multi-target research. SVM

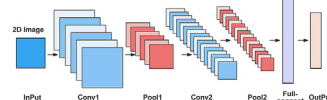


FIGURE 5

The classic LeNet architecture in CNNs is designed for 2D image feature extraction and classification. It consists of an input layer, alternating convolutional and pooling layers, a fully connected layer, and an output layer. This structure progressively extracts and maps local to global features for tasks like object detection and image classification.

demonstrates proficiency in the classification of small, high-dimensional datasets, while DT algorithms, particularly ensemble methods such as RF and XGBoost, exhibit efficacy in the extraction of features and the identification of targets in complex biological systems. Clustering algorithms, in contrast, are instrumental in the realm of unsupervised learning, facilitating the discovery of latent patterns. However, it is imperative to acknowledge the limitations inherent in these methodologies. SVM grapples with computational challenges posed by large datasets, DT models may lack interpretability due to complex trees, and clustering algorithms are sensitive to noise in high-dimensional contexts. These limitations underscore the necessity for judicious integration and optimization of these techniques. Future research should prioritize the development of hybrid approaches that synergistically leverage the strengths of these algorithms, thereby creating robust, interpretable, and multi-layered predictive models. These advancements hold great promise in deepening our understanding of multi-metabolite multi-target mechanisms in TCM and driving significant progress in pharmacological research.

5.3 Deep learning algorithms

Deep learning (DL) has been shown to outperform conventional machine learning methods in nonlinear modeling and automated feature extraction. In multi-metabolite multi-target interaction prediction, DL algorithms achieve superior accuracy by capturing intricate biological system relationships. These algorithms autonomously extract high-level molecular features, analyze complex metabolite-target interaction networks, and process dynamic biological data, enabling deeper insights into pharmacological mechanisms. Below we discuss several representative DL algorithms and their strengths in feature extraction and dynamic modeling.

5.3.1 Convolutional neural networks

Convolutional Neural Networks (CNNs), a prevalent technology in the domain of image processing (Figure 5), comprise three fundamental components: convolutional layers for local feature extraction, pooling layers for dimensionality reduction, and fully connected layers for classification or regression (Guo *et al.*, 2020). The remarkable efficacy of CNNs in processing nonlinear, high-dimensional data can be attributed to their local receptive fields, weight sharing, and pooling operations (Soffer *et al.*, 2019). In TCM research, CNNs have been employed to automatically detect

molecular features, such as spatial distributions, with the aim of predicting targets and mechanisms. For instance, Liu *et al.* developed a CNN-based drug screening platform that integrates multi-source data and topological information to predict potential therapeutic agents for Parkinson's disease and related proteins (Liu *et al.*, 2022). Similarly, Chen *et al.* combined CNNs with genetic algorithms to predict liver cancer treatment efficacy, identifying active metabolites (quercetin, kaempferol) that modulate IL-17 and TNF pathways (Chen *et al.*, 2023). However, these methods exhibit shared limitations, including increased overfitting by relying on a limited clinical data set ($n = 745$) and the risk of potential false positives from Pan-Assay interfering compounds (PAINS). Molecular dynamics (MD) simulations offer a potential solution by analyzing compound-membrane interaction patterns to effectively identify PAINS, providing enhanced specificity compared to traditional ligand-based screening approaches (Magalhães *et al.*, 2021). Future research should focus on developing hybrid graph-CNN architectures trained on MD-derived interaction fingerprints, such as halogen bond configurations, combined with ML classifiers to further improve predictive accuracy and biological relevance.

Furthermore, CNN-based drug-target interaction (DTI) models are frequently employed to predict novel targets for active metabolites. For instance, Hu *et al.* introduced SSELN-neg, a framework designed to enhance model performance through the selection of high-quality negative samples and parameter optimization via a spherical search algorithm (Hu *et al.*, 2023). In a separate investigation, Qu *et al.* utilized a CNN-based graph autoencoder to extract high-order structural information from heterogeneous networks, achieving a substantial improvement in DTI prediction accuracy (Qu *et al.*, 2024). While CNNs exhibit robust feature extraction and generalization capabilities, their applicability is constrained by reliance on grid-like data representations, challenges in distinguishing true negatives from unvalidated non-interacting pairs, and limited adaptability to time-series datasets. Future advancements in this field should prioritize the integration of geometric DL into hybrid architectures to process non-Euclidean molecular representations, the implementation of rigorous negative sample validation protocols (e.g., orthogonal experimental confirmation), and the optimization of spherical search algorithms for efficient parameter tuning in high-dimensional spaces (Guo *et al.*, 2020).

5.3.2 Recurrent neural networks

The dynamic interactions between active metabolites and their biological targets frequently exhibit significant temporal dependencies, a characteristic that CNNs often fail to accurately capture. However, recurrent neural networks (RNNs) are particularly well-suited at modelling time-series datasets, showing efficacy in applications involving sequential patterns. RNNs leverage a recurrent architecture, integrating current inputs with preceding hidden states to effectively capture dynamic features across time (Wang J. *et al.*, 2016; Mao and Sejdić, 2023). This attribute renders RNNs an ideal method for analyzing the *in vivo* metabolic transformations of active metabolites and their interactions with biological targets (Tang and Wu, 2022). For instance, Zhang *et al.* developed an RNN-based model, termed GRMC, which accurately predicts meridian

associations for active metabolites based on graph-derived neural features (Zhang P. *et al.*, 2024).

However, conventional RNNs are prone to vanishing and exploding gradients when processing long input sequences, thereby limiting their ability to model protracted temporal dependencies. This limitation spurred the development of modified RNN architectures, such as Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs) (Yu, 2022). LSTMs incorporate memory cells and sophisticated gating mechanisms to mitigate the gradient vanishing problem, thereby enabling the effective modelling of long-term dependencies (Jaihani *et al.*, 2022). GRUs, a computationally simplified version of LSTMs, merge the forget and update gates, improving efficiency while maintaining a comparable capability for modelling temporal dynamics (Kim *et al.*, 2023). Despite these advancements in capturing temporal dependencies, RNNs and their variants frequently demonstrate diminished computational efficiency when confronted with substantial datasets and intricate, nonlinear relationships. Consequently, future research endeavors should prioritize the development of hybrid architectures that seamlessly integrate attention-enhanced RNNs with graph neural networks (GNNs). These hybrid architectures should aim to concurrently model both sequential dependencies and multi-scale interaction patterns. Moreover, the utilization of parallel computing frameworks is imperative to address the computational bottlenecks inherent in these models (Wang D. *et al.*, 2016; Mao and Sejdić, 2023).

5.3.3 Graph neural networks

To address the inherent limitations of RNNs and their variants in capturing complex, non-sequential relationships, graph neural networks (GNNs) have emerged as a powerful deep learning architecture for processing graph-structured datasets. Grounded in principles of graph theory, GNNs operate by propagating and learning feature representations through connections between nodes (Pasa *et al.*, 2022). Nodes represent active metabolites or targets, while edges denote interactions. Through graph convolution, GNNs efficiently aggregate structural information to capture nonlinear relationships (Wang *et al.*, 2023). For instance, Duan *et al.* developed HTINet2, a GNN-based framework capable of extracting and representing deep metabolite-target interaction patterns (Duan *et al.*, 2024). A distinguishing feature of GNNs is their inherent independence from spatial or sequential ordering, facilitating the flexible learning of inter-node relationships and circumventing the temporal constraints of RNNs. While HTINet2 demonstrates superior performance, its limitations include dependence on knowledge graph completeness and sparse supervised signals from limited clinical data. Future directions should focus on integrating multi-omics data and experimental validation to enhance biological relevance prediction (Jin *et al.*, 2022).

5.4 Cross-modal data fusion algorithms

Cross-modal data fusion algorithms are designed to integrate information from diverse modalities, encompassing chemical structural data of active metabolites, biological target data, and

pharmacological experimental results. This approach enables a holistic analysis of metabolite-target interactions. Three primary methods are commonly used: joint embedding, attention mechanisms, and deep generative models (Liu L. et al., 2024). Joint embedding techniques create a shared feature space for multimodal data, optimizing correlations between modalities. For instance, Deep Canonical Correlation Analysis (DCCA) extracts common features from electroencephalography (EEG) and eye-tracking data to detect fatigue (Lian et al., 2024). Similarly, Zhao *et al.* developed a multimodal framework combining visual transformers and Graph Convolutional Networks (GCNs) for recommendation and prescription generation of botanical drugs (Zhao W. et al., 2024). Deep generative models, such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs), have been employed to explore metabolite-target relationships (Gao et al., 2020). GANs consist of a generator and a discriminator that work adversarially to produce realistic synthetic datasets. In TCM research, GANs generate potential active molecular structures to predict novel target interactions. In contrast, VAEs learn latent distributions from input data to generate new samples, excelling at capturing underlying feature spaces. Despite these advances, current approaches struggle with modality-specific feature misalignment and overreliance on synthetic data that has not been validated by experimental pharmacology. Future work must prioritize physics-informed generative architectures and self-supervised multimodal alignment to bridge domain gaps between computational predictions and biological plausibility (Liu M. et al., 2024).

6 Challenges

Despite significant advancements in ML and DL applications for TCM studies, persistent methodological challenges require systematic resolution. This section will therefore analyze current limitations, existing solutions, and future research directions.

6.1 Dilemma regarding input modalities

Current TCM target prediction models face fundamental limitations in processing heterogeneous data streams. Single-modality approaches inadequately capture the complexity of TCM, necessitating integration of chemical, biological, pharmacological, multi-omics (genomic, proteomic, metabolomic), and clinical data domains. Three critical barriers have been identified: First, there is technical heterogeneity from disparate database architectures and annotation protocols. Second, there are nonlinear interactions between modality-specific feature spaces. Third, there is class imbalance across disease taxonomies. These challenges collectively constrain model generalizability. Therefore, advanced multimodal fusion frameworks are necessary for robust TCM analysis. Emerging solutions demonstrate progress in multimodal integration. The Drug LAMP model enhances prediction accuracy through synergistic fusion of molecular maps and protein sequences via multimodal PLMs combined with conventional feature extraction (Luo et al., 2024). Similarly, the MKG-FENN framework achieves superior drug-drug interaction

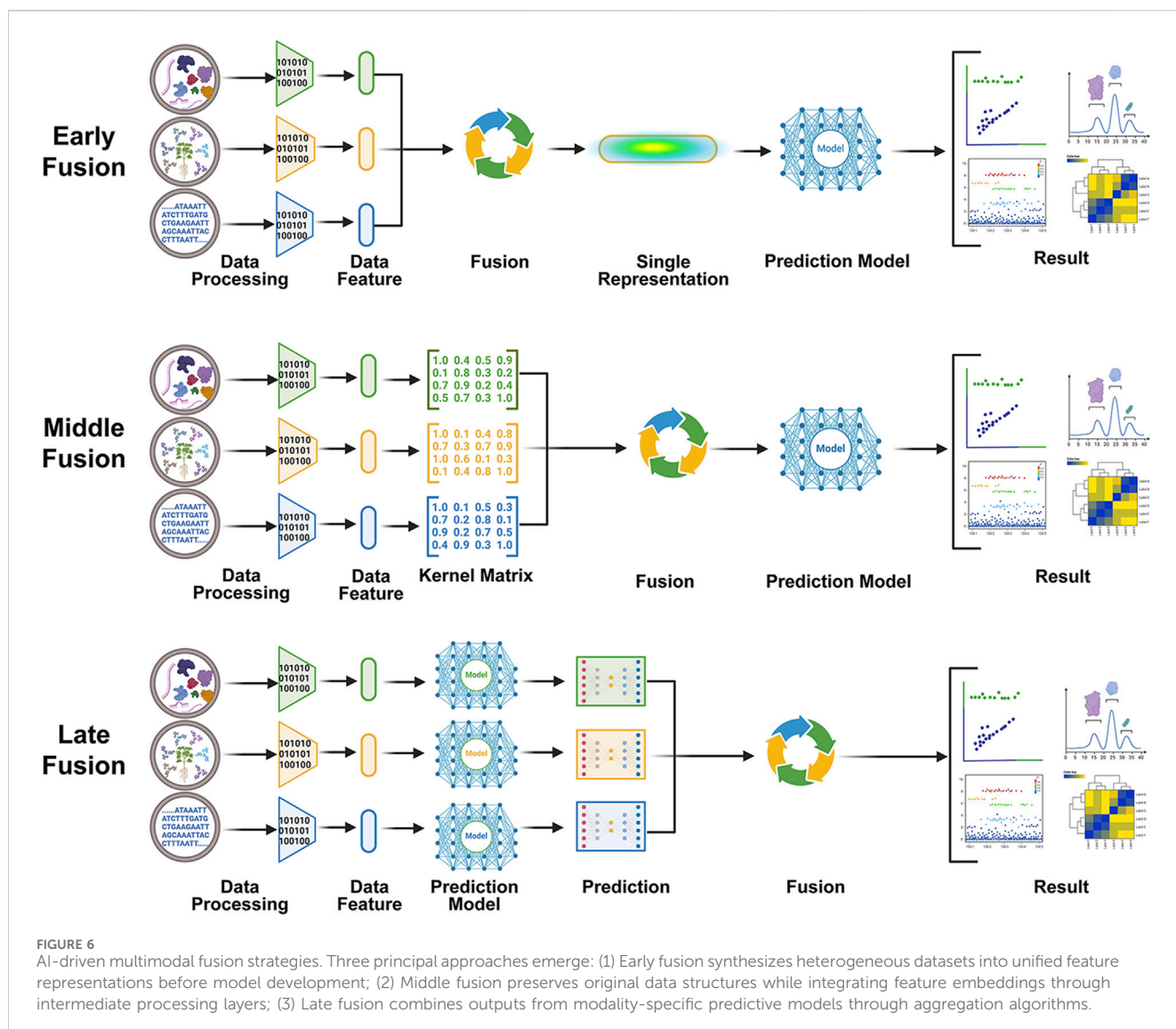
prediction by integrating neural networks with multimodal knowledge graphs, effectively modeling drug-chemical entity relationships and molecular substructure interactions (Jiao et al., 2023).

The integration of multimodal data has emerged as a prominent approach in TCM research, with the predominant strategies falling into three categories (Figure 6): early fusion (input-level concatenation), mid fusion (feature-space integration via attention mechanisms), and late fusion (output-level aggregation) (Ding et al., 2021; Hamamoto et al., 2022). Advanced implementations, such as the Drug LAMP model, employ Pocket-Guided Common Attention (PGCA) and Paired Multimodal Attention (PMMA) modules to optimize cross-modal feature alignment (Wu et al., 2021; Borse et al., 2023; Hou et al., 2024). State-of-the-art Transformer-based architectures show particular promise for TCM target prediction through their inherent capacity for contextual relationship modeling (Meyer et al., 2019; Liu J. et al., 2023). Natural language provides a rich source of fine-grained knowledge and control instructions, often used in visuomotor tasks (Lee et al., 2023; Vaid et al., 2024). Similarly, natural language processing (NLP) techniques have demonstrated potential as a means of integrating textual information associated with TCM, usage guidelines, and contraindications. For instance, Song *et al.* developed a database of adverse reactions for both Chinese and Western medicines utilizing large-scale language models (LLM) and NLP techniques, which improved prediction accuracy and utility (Song et al., 2024). However, the integration of natural language models into TCM target prediction poses challenges due to the substantial inference time, limited quantitative accuracy, and potential instability of natural language models. In addition, textual databases related to Chinese medicine may contain noise and inaccuracies. Therefore, while LLM may be suitable for specialized, complex scenarios or high-level behavioral prediction, their direct integration into TCM target prediction requires careful consideration (Lv et al., 2023b).

6.2 Dependence on feature representation

Current TCM target prediction systems face fundamental limitations in feature representation engineering. The inherent complexity of TCM formulations, characterized by polypharmacological interaction patterns, parallels the sensorimotor challenges of autonomous urban navigation systems (Hilleli and El-Yaniv, 2018). Despite methodological advances (He et al., 2016), no consensus exists for optimal TCM target representation. Emerging solutions employ heterogeneous networks integrating active metabolites, biological targets, and interaction profiles (Gao J. et al., 2024). However, these architectures require validation across diverse pharmacological contexts. A critical implementation gap persists in co-optimizing feature representations with downstream decision layers—misalignment between these stages frequently degrades prediction accuracy.

Representation learning approaches are constrained by two factors: 1) information bottleneck effects during feature compression, which eliminate contextually relevant pharmacological data; and 2) over-simplified chemical descriptors



that omit critical structural-activity relationships (Zhang S. et al., 2024). The prevalence of redundant information (e.g., inactive molecular substructures) further complicates discriminative feature extraction (Liu L. et al., 2024). Despite the potential demonstrated by self-supervised learning for TCM representation learning (Bucci et al., 2022), two fundamental challenges persist: 1) the development of pretext tasks to capture TCM's latent pharmacological signatures, and 2) the quantitative validation of learned representations in clinical prediction scenarios. Transformer-based architectures may offer solutions through their inherent capacity for context-aware feature learning, though concerns regarding computational complexity persist.

6.3 Complexity of world modeling

The application of deep reinforcement learning (DRL) to TCM target prediction is constrained by three interrelated challenges rooted in the complexity of world modeling. First, the high sample complexity inherent to DRL necessitates extensive

pharmacological datasets—a critical limitation given the polypharmacological nature and data scarcity of TCM systems (Song et al., 2024). Secondly, model-environment divergence in Model-Based Reinforcement Learning (MBRL) introduces prediction error propagation, necessitating the integration of deep neural networks with Bayesian uncertainty quantification to mitigate dynamic model inaccuracies (Guo et al., 2023). Thirdly, computational intractability arises from the combinatorial demands of multistep MBRL planning and multimodal data integration, a problem that is particularly problematic for real-time clinical applications (Lv et al., 2023a).

Current MBRL frameworks exhibit systemic biases toward established structure-activity relationships, potentially overlooking novel therapeutic targets. This limitation necessitates the implementation of entropy-driven exploration strategies to enhance solution space navigation while maintaining computational feasibility. Dimensionality reduction techniques have demonstrated efficacy in addressing high-dimensional state spaces, particularly in the context of image-based phytochemical analyses (Zhang S. et al., 2023). The development of architectural

optimizations that balance model complexity and computational tractability is imperative to bridge the gap between theoretical MBRL capabilities and the practical requirements of TCM research. However, significant challenges persist in aligning these computational frameworks with holistic pharmacological principles.

6.4 Reliance on multi-task learning

Multi-task learning (MTL) offers strategic advantages for TCM target prediction through shared representation learning across pharmacological activity, therapeutic effects, and safety profiling tasks. By leveraging inter-task correlations via task-specific heads, MTL reduces computational redundancy while enhancing model generalizability (Vandenhende et al., 2022). This approach aligns with TCM's requirement for holistic biological system modeling, where concurrent prediction of multi-target interactions benefits from shared intermediate representations. However, two critical limitations emerge: 1) Optimization challenges in balancing task-specific loss functions, particularly given TCM's sparse pharmacological annotations; and 2) Insufficient theoretical frameworks for auxiliary task selection in polypharmacological contexts (Ishihara et al., 2021; Jaeger et al., 2023).

6.5 Lack of interpretability

Despite significant advancements in the field of AI algorithms for predicting TCM metabolite-target interactions, the inherent “black box” nature of many models poses a substantial obstacle to their widespread adoption and acceptance. This opacity hinders both understanding and user trust, giving rise to significant ethical and legal concerns (Ornes, 2023). The complexity of deep neural network architectures, while often associated with high predictive accuracy, contributes significantly to a lack of model interpretability (Zhang Y. et al., 2024). A persistent trade-off exists between accuracy and interpretability, and efforts to improve model accuracy frequently necessitate more intricate architectures and algorithms, thereby compromising model transparency (Zhang P. et al., 2024). The absence of standardized evaluation metrics further exacerbates this challenge, as it prevents both the development of interpretable models and the comparative analysis of their transparency (Karim et al., 2023).

In order to address the aforementioned limitations, researchers have explored *post hoc* explainable AI (X-AI) techniques, such as generating saliency maps to highlight influential input features. However, such approaches offer limited insights, and their efficacy remains difficult for a rigorous evaluation (Solorio-Ramírez et al., 2021). Consequently, considerable attention has shifted towards the design of end-to-end frameworks that incorporate interpretability into the model architecture. Attention mechanisms, for example, offer a certain degree of interpretability by assigning weights to features, thereby highlighting their relative importance in intermediate representations. However, while attention-based visualizations provide intuitive cues, their fidelity and utility in providing comprehensive explanations remain limited (Harfouche et al., 2023). The incorporation of interpretability-focused tasks, rule integration, cost learning, natural language-

based interpretability, and uncertainty quantification holds promise for improving model reliability and transparency in TCM target prediction (Yang G. et al., 2022). However, many of these methods function primarily as auxiliary tasks, with a potentially limited impact on the final predictive outcome.

6.6 Causal confusion

Causal confounding, a persistent challenge in imitation learning for nearly 2 decades, presents a significant parallel in TCM target prediction modeling. The inherent complexity of TCM chemical compositions, coupled with potential synergistic or antagonistic interactions between active metabolites, can substantially impact predictive outcomes. Existing models may exhibit an over-reliance on readily available chemical features while neglecting other potentially important factors (Lin et al., 2022). Additionally, the inherent heterogeneity of TCM target prediction datasets, which encompass diverse data sources prone to biases and inconsistencies, introduces noise into the learning process and amplifies the effect of causal confounding (Zhu Y. et al., 2022). To address these challenges, researchers have proposed several strategies. One approach involves enhancing the model's ability to identify salient features through the incorporation of auxiliary tasks, such as semantic segmentation of active metabolites or depth estimation. However, this approach increases model complexity and necessitates high-quality annotated datasets, which are difficult to obtain (Zhang Y. et al., 2023). An alternative strategy focuses on quantifying model uncertainty modeling, enabling the identification and correction of spurious associations (Öcal et al., 2022). This strategy integrates likelihood models to capture uncertainty, providing a computationally efficient approach for quantifying uncertainty in stochastic models of gene expression.

6.7 Lack of robustness

The TCM datasets generally manifest class imbalance, characterized by the overrepresentation of a few categories while other, equally important yet less prevalent, categories exhibit a paucity of instances. This imbalanced distribution poses a substantial challenge to model generalization across diverse environments (Yang et al., 2020). To address this challenge, researchers have proposed various data processing techniques, including oversampling (Krawczyk et al., 2020), undersampling (Marin and Hedges, 2018), and data augmentation (Shorten et al., 2021), as well as weighting-based methods (Fernandes et al., 2023). Additionally, the presence of covariate bias poses a substantial obstacle. Discrepancies between the distribution of training datasets and real-world application data can lead to reduced model performance in novel testing environments (Pitt et al., 2025). Pitt et al. employed the DAgger (Dataset Aggregation) algorithm to enrich the training dataset and improve model robustness through an iterative training process involving the continuous collection and expert annotation of new data (Pitt et al., 2025).

Domain Adaptation (DA) is an alternative transfer learning methodology that aims to train a model across identical source and

target tasks but different domains. In TCM target prediction, this domain divergence may manifest as a divergence between simulated and real-world datasets (Jin et al., 2024). Addressing this divergence, studies have demonstrated the efficacy of employing image translators and discriminators to map data from disparate domains into a shared latent space or representation, such as segmentation maps (He et al., 2023). Additionally, domain randomization has been shown to enhance model robustness by randomizing the rendering and physical parameters of the simulator, thereby effectively counteracting real-world variability (Bandyopadhyay et al., 2022).

7 Future trends

In light of the aforementioned challenges and opportunities, the following key research directions are proposed to facilitate substantial advancements within the field.

7.1 Zero-shot and few-shot learning

The inherent diversity and rarity of TCM datasets pose a significant challenge for model development. Zero- and few-sample learning techniques offer a promising avenue to address this issue by enabling models to adapt to new target domains with limited or unlabeled data. For instance, the TxGNN model, developed by Huang *et al.*, efficiently predicts drug indications and contraindications by analyzing a large-scale medical knowledge graph and providing interpretable multi-pathways explanations that reveal the medical reasoning underpinning the predictions (Huang et al., 2024). This approach not only improves prediction accuracy, but also highlights the potential for drug repurposing, exhibiting a strong alignment with clinical prescribing practices.

7.2 Modular end-to-end planning

Modular end-to-end planning frameworks, which are characterized by the optimization of multiple modules while prioritizing the final planning task, offer the advantage of improved interpretability. The efficacy of this framework within the context of target prediction has also been demonstrated. By designing different perceptual modules, researchers can explore a diverse range of loss functions and training strategies to optimize both model robustness and accuracy (Lv et al., 2023b). This modular approach enables not only a deeper understanding of the model's decision-making process but also enhances its adaptability within complex environments.

7.3 Data engines

Large-scale, high-quality datasets are imperative for the advancement of target prediction in TCM. The development of an automated data labeling engine offers a significant opportunity to streamline the iterative process of data and model development. A

notable example is TCM Bank, a comprehensive TCM database that utilizes big data-driven and unsupervised learning methodologies to predict the adverse effects of both Chinese and Western medicines (Song et al., 2024). The data engine not only supports case mining and scenario generation, but also facilitates data-driven evaluation and improves model generalization.

7.4 Foundation model

Recent advancements in foundation modeling, particularly within the domains of language (Li et al., 2025) and vision (Fang et al., 2023), have demonstrated that the availability of large-scale datasets, coupled with increased model capacity, can unlock the enormous potential of AI for sophisticated reasoning tasks. These base models can be further optimized through methodologies such as self-supervised reconstruction or comparative learning (Zeng et al., 2022). To illustrate this, consider the training of a model designed to predict a plausible future state for an environment. This model can then be utilized for planning in 2D, 3D, or latent spaces to improve performance in downstream tasks (Li et al., 2023b).

7.5 Self-supervised and comparative learning

Recent advancements in ML and DL have led to the development of self-supervised and comparative learning methodologies, which have emerged as promising avenues for target prediction in TCM. For instance, the application of functional representations derived from gene signatures to metabolite-target prediction, through the use of deep learning models, has shown the ability to identify functionally similar genes and optimize gene embedding vectors (Chen et al., 2024). This approach improves predictive accuracy and reveals associations and common information across different modalities, thereby providing a novel perspective for TCM target prediction.

8 Conclusion

This review provides a comprehensive examination of the applications and advancements of AI in modelling multi-metabolite multi-target interactions within the context of TCM. AI methodologies have revolutionized the field, providing innovative tools and frameworks for the analysis and quantification of the complex interactions between active metabolites and biological targets. The integration of multi-omics datasets, advanced deep learning techniques, and knowledge graph-based frameworks has significantly improved the predictive accuracy and robustness of TCM studies, enabling more systematic metabolite screening and pharmacodynamic analysis.

However, several challenges persist. Data heterogeneity, sample imbalance, and the complexity of TCM formulations impede effective feature representation and model training. Additionally, the “black box” nature of many AI models limits their interpretability, reducing trust among researchers and practitioners. Issues such as causal confounding and insufficient

model robustness further complicate AI applications in TCM target prediction. To that end, future research should prioritize the development of zero-shot and few-shot learning paradigms, the creation of modular end-to-end planning frameworks, the development of data engines, and the integration of self-supervised learning methodologies. These approaches are designed to enhance model adaptability, interpretability, and reliability. In summary, the integration of AI into TCM represents a significant step toward the modernization of TCM and the advancement of personalized medicine. By addressing current challenges and pursuing innovative directions, the field can achieve a broader impact and global relevance. Continued interdisciplinary collaboration is essential to fully realize the potential of AI in TCM research.

Author contributions

YL: Conceptualization, Investigation, Writing – original draft, Writing – review and editing. XL: Conceptualization, Writing – original draft, Writing – review and editing. JZ: Data curation, Formal Analysis, Writing – review and editing. FL: Methodology, Software, Writing – review and editing. YW: Methodology, Software, Writing – review and editing. QL: Supervision, Visualization, Writing – original draft, Writing – review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the National Natural Science Foundation of China [82172281], Traditional Chinese Medicine Science and Technology Development Project of Shanghai Medical Innovation &

Development Foundation [WL-YBXM-2022001K and WL-HBQN-2022014K], the Cultivation Project for Medical Technology Doctoral Degree Program of Shanghai City (2021–2023) and Open Project of Shanghai Key Laboratory of Modern Optical System, University of Shanghai for Science and Technology (K241302N).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2025.1541509/full#supplementary-material>

References

- Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., et al. (2004). UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 32, D115–D119. doi:10.1093/nar/gkh131
- Bandyopadhyay, H., Deng, Z., Ding, L., Liu, S., Uddin, M. R., Zeng, X., et al. (2022). Cryo-shift: reducing domain shift in cryo-electron subtomograms with unsupervised domain adaptation and randomization. *Bioinforma. Oxf. Engl.* 38, 977–984. doi:10.1093/bioinformatics/btab794
- Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483, 603–607. doi:10.1038/nature11003
- Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., et al. (2013). NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* 41, D991–D995. doi:10.1093/nar/gks1193
- Borse, S., Klingner, M., Kumar, V. R., Cai, H., Almuzaire, A., Yogamani, S., et al. (2023). “X-Align: cross-modal cross-view alignment for bird’s-eye-view segmentation,” in *2023 IEEE/CVF winter conference on applications of computer vision (WACV)*, 3286–3296. doi:10.1109/WACV56688.2023.00330
- Brown, G. R., Hem, V., Katz, K. S., Ovetsky, M., Wallin, C., Ermolaeva, O., et al. (2015). Gene: a gene-centered information resource at NCBI. *Nucleic Acids Res.* 43, D36–D42. doi:10.1093/nar/gku1055
- Bucci, S., D’Innocente, A., Liao, Y., Carlucci, F. M., Caputo, B., and Tommasi, T. (2022). Self-supervised learning across domains. *IEEE Trans. Pattern Anal. Mach. Intell.* 44, 5516–5528. doi:10.1109/TPAMI.2021.3070791
- Calderaro, J., Seraphin, T. P., Luedde, T., and Simon, T. G. (2022). Artificial intelligence for the prevention and clinical management of hepatocellular carcinoma. *J. Hepatol.* 76, 1348–1361. doi:10.1016/j.jhep.2022.01.014
- Chen, C. Y.-C. (2011). TCM Database@Taiwan: the world’s largest traditional Chinese medicine database for drug screening *in silico*. *PLoS One* 6, e15939. doi:10.1371/journal.pone.0015939
- Chen, H., King, F. J., Zhou, B., Wang, Y., Canedy, C. J., Hayashi, J., et al. (2024). Drug target prediction through deep learning functional representation of gene signatures. *Nat. Commun.* 15, 1853. doi:10.1038/s41467-024-46089-y
- Chen, H.-Y., Chen, J.-Q., Li, J.-Y., Huang, H.-J., Chen, X., Zhang, H.-Y., et al. (2019). Deep learning and random forest approach for finding the optimal traditional Chinese medicine formula for treatment of Alzheimer’s disease. *J. Chem. Inf. Model.* 59, 1605–1623. doi:10.1021/acs.jcim.9b00041
- Chen, Q., Springer, L., Gohlke, B. O., Goede, A., Dunkel, M., Abel, R., et al. (2021). SuperTCM: a biocultural database combining biological pathways and historical linguistic data of Chinese Materia Medica for drug development. *Biomedicine Pharmacother.* 144, 112315. doi:10.1016/j.biopha.2021.112315
- Chen, S., Zhao, Y., Liu, S., Zhang, J., Assaraf, Y. G., Cui, W., et al. (2022). Epigenetic enzyme mutations as mediators of anti-cancer drug resistance. *Drug resist. updat.* 61, 100821. doi:10.1016/j.drug.2022.100821
- Chen, Z., Peng, P., Wang, M., Deng, X., and Chen, R. (2023). Bioinformatics-based and multiscale convolutional neural network screening of herbal medicines for improving the prognosis of liver cancer: a novel approach. *Front. Med.* 10, 1218496. doi:10.3389/fmed.2023.1218496

- Cheng, X., Manandhar, I., Aryal, S., and Joe, B. (2021). Application of artificial intelligence in cardiovascular medicine. *Compr. Physiol.* 11, 2455–2466. doi:10.1002/cphy.e200034
- Ching, P. M. L., Zou, X., Wu, D., So, R. H. Y., and Chen, G. H. (2022). Development of a wide-range soft sensor for predicting wastewater BOD5 using an eXtreme gradient boosting (XGBoost) machine. *Environ. Res.* 210, 112953. doi:10.1016/j.envres.2022.112953
- Chong, J., Soufan, O., Li, C., Caraus, I., Li, S., Bourque, G., et al. (2018). MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res.* 46, W486–W494–W494. doi:10.1093/nar/gky310
- Colwell, J. (2016). Expanding the scope of ENCODE. *Cancer Discov.* 6, OF4. doi:10.1158/2159-8290.CD-NB2016-020
- Cong, Y., Yang, X., Lv, W., and Xue, Y. (2009). Prediction of novel and selective TNF- α converting enzyme (TACE) inhibitors and characterization of correlative molecular descriptors by machine learning approaches. *J. Mol. Graph. Model.* 28, 236–244. doi:10.1016/j.jmglm.2009.08.001
- Consortium, T. G. (2013). The genotype-tissue expression (GTEx) project. *Nat. Genet.* 45, 580–585. doi:10.1038/ng.2653
- Cotter, D., Maer, A., Guda, C., Saunders, B., and Subramaniam, S. (2006). LMPD: LIPID MAPS proteome database. *Nucleic Acids Res.* 34, D507–D510. doi:10.1093/nar/gkj122
- Croft, D., O’Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., et al. (2011). Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.* 39, D691–D697. doi:10.1093/nar/gkq1018
- Cunningham, F., Allen, J. E., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., et al. (2022). Ensembl 2022. *Nucleic Acids Res.* 50, D988–D995. doi:10.1093/nar/gkab1049
- Desaphy, J., Bret, G., Rognan, D., and Kellenberger, E. (2015). sc-PDB: a 3D-database of ligandable binding sites—10 years on. *Nucleic Acids Res.* 43, D399–D404. doi:10.1093/nar/gku928
- Ding, Q., Sun, Y., Shang, J., Li, F., Zhang, Y., and Liu, J.-X. (2021). NMFNA: a non-negative matrix factorization network analysis method for identifying modules and characteristic genes of pancreatic cancer. *Front. Genet.* 12, 678642. doi:10.3389/fgene.2021.678642
- Ding, Z., Wang, N., Ji, N., and Chen, Z.-S. (2022). Proteomics technologies for cancer liquid biopsies. *Mol. Cancer* 21, 53. doi:10.1186/s12943-022-01526-8
- Duan, P., Yang, K., Su, X., Fan, S., Dong, X., Zhang, F., et al. (2024). HTINet2: herb-target prediction via knowledge graph embedding and residual-like graph neural network. *Brief. Bioinform.* 25, bbae414. doi:10.1093/bib/bbae414
- Fang, S., Dong, L., Liu, L., Guo, J., Zhao, L., Zhang, J., et al. (2021). HERB: a high-throughput experiment- and reference-guided database of traditional Chinese medicine. *Nucleic Acids Res.* 49, D1197–D1206. doi:10.1093/nar/gkaa1063
- Fang, Y., Wang, W., Xie, B., Sun, Q., Wu, L., Wang, X., et al. (2023). “EVA: exploring the limits of masked visual representation learning at scale,” in *2023 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, 19358–19369. doi:10.1109/CVPR52729.2023.01855
- Feng, X., Zhang, X., Chen, Y., Li, L., Sun, Q., and Zhang, L. (2020). Identification of bilobetin metabolites, *in vivo* and *in vitro*, based on an efficient ultra-high-performance liquid chromatography coupled with quadrupole time-of-flight mass spectrometry strategy. *J. Sep. Sci.* 43, 3408–3420. doi:10.1002/jssc.202000313
- Fernandes, G. J., Choi, A., Schauer, J. M., Pfammatter, A. F., Spring, B. J., Darwiche, A., et al. (2023). An explainable artificial intelligence software tool for weight management experts (PRIMO): mixed methods study. *J. Med. Internet Res.* 25, e42047. doi:10.2196/42047
- Gan, H., Huang, R., Luo, Z., Xi, X., and Gao, Y. (2018). On using supervised clustering analysis to improve classification performance. *Inf. Sci.* 454 (455), 216–228. doi:10.1016/j.ins.2018.04.080
- Ganini, C., Amelio, I., Bertolo, R., Bove, P., Buonomo, O. C., Candi, E., et al. (2021). Global mapping of cancers: the cancer genome atlas and beyond. *Mol. Oncol.* 15, 2823–2840. doi:10.1002/1878-0261.13056
- Gao, F., Zhou, Y., Yu, B., Xie, H., Shi, Y., Zhang, X., et al. (2024a). QiDiTangShen granules alleviates diabetic nephropathy podocyte injury: a network pharmacology study and experimental validation *in vivo* and *in vitro*. *Heliyon* 10, e23535. doi:10.1016/j.heliyon.2023.e23535
- Gao, J., Xiang, X., Yan, Q., and Ding, Y. (2024b). CDCS-TCM: a framework based on complex network theory to analyze the causality and dynamic correlation of substances in the metabolic process of traditional Chinese medicine. *J. Ethnopharmacol.* 328, 118100. doi:10.1016/j.jep.2024.118100
- Gao, R., Hou, X., Qin, J., Chen, J., Liu, L., Zhu, F., et al. (2020). Zero-VAE-GAN: generating unseen features for generalized and transductive zero-shot learning. *IEEE Trans. Image Process. Publ. IEEE Signal Process. Soc.* 29, 3665–3680. doi:10.1109/TIP.2020.2964429
- Griesenauer, R. H., Schillebeeckx, C., and Kinch, M. S. (2019). Assessing the public landscape of clinical-stage pharmaceuticals through freely available online databases. *Drug Discov. Today* 24, 1010–1016. doi:10.1016/j.drudis.2019.01.010
- Gu, J., Gui, Y., Chen, L., Yuan, G., and Xu, X. (2013). CVDHD: a cardiovascular disease herbal database for drug discovery and network pharmacology. *J. Cheminformatics* 5, 51. doi:10.1186/1758-2946-5-51
- Guo, X.-X., An, S., Bao, F., and Xu, T.-R. (2023). Challenges and perspectives in target identification and mechanism illustration for Chinese medicine. *Chin. J. Integr. Med.* 29, 644–654. doi:10.1007/s11655-023-3629-9
- Guo, Y., Chen, J., Du, Q., Van Den Hengel, A., Shi, Q., and Tan, M. (2020). Multi-way backpropagation for training compact deep neural networks. *Neural Netw.* 126, 250–261. doi:10.1016/j.neunet.2020.03.001
- Hamamoto, R., Takasawa, K., Machino, H., Kobayashi, K., Takahashi, S., Bolatkan, A., et al. (2022). Application of non-negative matrix factorization in oncology: one approach for establishing precision medicine. *Brief. Bioinform.* 23, bbac246. doi:10.1093/bib/bbac246
- Han, N., Qiao, S., Yuan, G., Huang, P., Liu, D., and Yue, K. (2019). A novel Chinese herbal medicine clustering algorithm via artificial bee colony optimization. *Artif. Intell. Med.* 101, 101760. doi:10.1016/j.artmed.2019.101760
- Harfouche, A. L., Nakhle, F., Harfouche, A. H., Sardella, O. G., Dart, E., and Jacobson, D. (2023). A primer on artificial intelligence in plant digital phenomics: embarking on the data to insights journey. *Trends Plant Sci.* 28, 154–184. doi:10.1016/j.tplants.2022.08.021
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). “Deep residual learning for image recognition,” in *2016 IEEE conference on computer vision and pattern recognition (CVPR)*, 770–778. doi:10.1109/CVPR.2016.90
- He, S., Feng, Y., Grant, P. E., and Ou, Y. (2023). Segmentation ability map: interpret deep features for medical image segmentation. *Med. Image Anal.* 84, 102726. doi:10.1016/j.media.2022.102726
- Heikamp, K., and Bajorath, J. (2014). Support vector machines for drug discovery. *Expert Opin. Drug Discov.* 9, 93–104. doi:10.1517/17460441.2014.866943
- Heinrich, M., Jalil, B., Abdel-Tawab, M., Echeverria, J., Kulić, Ž., McGaw, L. J., et al. (2022). Best Practice in the chemical characterisation of extracts used in pharmacological and toxicological research-The ConPhyMP-Guidelines. *Front. Pharmacol.* 13, 953205. doi:10.3389/fphar.2022.953205
- Hilleli, B., and El-Yaniv, R. (2018). Toward deep reinforcement learning without a simulator: an autonomous steering example. *Proc. AAAI Conf. Artif. Intell.* 32. doi:10.1609/aaai.v32i1.11490
- Holm, S., Stanton, C., and Bartlett, B. (2021). A new argument for no-fault compensation in health care: the introduction of artificial intelligence systems. *Health Care Anal.* 29, 171–188. doi:10.1007/s10728-021-00430-4
- Hou, J., Saad, S., and Omar, N. (2024). Enhancing traditional Chinese medical named entity recognition with Dyn-Att Net: a dynamic attention approach. *PeerJ Comput. Sci.* 10, e2022. doi:10.7717/peerj-cs.2022
- Hu, L., Fu, C., Ren, Z., Cai, Y., Yang, J., Xu, S., et al. (2023). SSELN-neg: spherical search-based extreme learning machine for drug-target interaction prediction. *BMC Bioinforma.* 24, 38. doi:10.1186/s12859-023-05153-y
- Hua, R., Dong, X., Wei, Y., Shu, Z., Yang, P., Hu, Y., et al. (2024). Lingdan: enhancing encoding of traditional Chinese medicine knowledge for clinical reasoning tasks with large language models. *J. Am. Med. Inf. Assoc. JAMIA* 31, 2019–2029. doi:10.1093/jamia/ocae087
- Huang, J., and Wang, J. (2014). CEMTDD: Chinese ethnic minority traditional drug database. *Apoptosis Int. J. Program. Cell Death* 19, 1419–1420. doi:10.1007/s10495-014-1011-2
- Huang, K., Chandak, P., Wang, Q., Havaladar, S., Vaid, A., Leskovec, J., et al. (2024). A foundation model for clinician-centered drug repurposing. *Nat. Med.* 30, 3601–3613. doi:10.1038/s41591-024-03233-x
- Huang, L., Xie, D., Yu, Y., Liu, H., Shi, Y., Shi, T., et al. (2018). TCMID 2.0: a comprehensive resource for TCM. *Nucleic Acids Res.* 46, D1117–D1120–D1120. doi:10.1093/nar/gkx1028
- Ishihara, K., Kanervisto, A., Miura, J., and Hautamäki, V. (2021). “Multi-task learning with attention for end-to-end autonomous driving,” in *2021 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW)*, 2896–2905. doi:10.1109/CVPRW53098.2021.00325
- Jaeger, B., Chitta, K., and Geiger, A. (2023). “Hidden biases of end-to-end driving models,” in *2023 IEEE/CVF international conference on computer vision (ICCV)*, 8206–8215. doi:10.1109/ICCV51070.2023.00757
- Jaihuni, M., Basak, J. K., Khan, F., Okyere, F. G., Sihlath, T., Bhujel, A., et al. (2022). A novel recurrent neural network approach in forecasting short term solar irradiance. *ISA Trans.* 121, 63–74. doi:10.1016/j.isatra.2021.03.043
- Jiao, J., Sun, H., Huang, Y., Xia, M., Qiao, M., Ren, Y., et al. (2023). GMRLNet: a graph-based manifold regularization learning framework for placental insufficiency diagnosis on incomplete multimodal ultrasound data. *IEEE Trans. Med. Imaging* 42, 3205–3218. doi:10.1109/TMI.2023.3278259
- Jin, X., Wang, Z., Ma, J., Liu, C., Bai, X., and Lan, Y. (2024). Electronic eye and electronic tongue data fusion combined with a GETNet model for the traceability and detection of Astragalus. *J. Sci. Food Agric.* 104, 5930–5943. doi:10.1002/jsfa.13450

- Jin, Y., Ji, W., Zhang, W., He, X., Wang, X., and Wang, X. (2022). A KG-enhanced multi-graph neural network for attentive herb recommendation. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 19, 2560–2571. doi:10.1109/TCBB.2021.3115489
- Jones, F. C., Plewes, R., Murison, L., MacDougall, M. J., Sinclair, S., Davies, C., et al. (2017). Random forests as cumulative effects models: a case study of lakes and rivers in Muskoka, Canada. *J. Environ. Manage.* 201, 407–424. doi:10.1016/j.jenvman.2017.06.011
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi:10.1093/nar/28.1.27
- Karim, M. R., Islam, T., Shajalal, M., Beyan, O., Lange, C., Cochez, M., et al. (2023). Explainable AI for bioinformatics: methods, tools and applications. *Brief. Bioinform.* 24, bbad236. doi:10.1093/bib/bbad236
- Kim, H., Lee, J., Moon, S., Kim, S., Kim, T., Jin, S. W., et al. (2023). Visual field prediction using a deep bidirectional gated recurrent unit network model. *Sci. Rep.* 13, 11154. doi:10.1038/s41598-023-37360-1
- Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., et al. (2016). PubChem substance and compound databases. *Nucleic Acids Res.* 44, D1202–D1213. doi:10.1093/nar/gkv951
- Kim, S.-K., Lee, M.-K., Jang, H., Lee, J.-J., Lee, S., Jang, Y., et al. (2024). TM-MC 2.0: an enhanced chemical database of medicinal materials in Northeast Asian traditional medicine. *BMC Complement. Med. Ther.* 24, 40. doi:10.1186/s12906-023-04331-y
- Kong, X., Liu, C., Zhang, Z., Cheng, M., Mei, Z., Li, X., et al. (2023). BATMAN-TCM 2.0: an enhanced integrative database for known and predicted interactions between traditional Chinese medicine ingredients and target proteins. *Nucleic Acids Res.* 52, D1110–D1120. doi:10.1093/nar/gkad926
- Krawczyk, B., Koziarski, M., and Wozniak, M. (2020). Radial-based oversampling for multiclass imbalanced data classification. *IEEE Trans. Neural Netw. Learn. Syst.* 31, 2818–2831. doi:10.1109/TNNLS.2019.2913673
- Laubscher, E., Wang, X., Razin, N., Dougherty, T., Xu, R. J., Ombelets, L., et al. (2024). Accurate single-molecule spot detection for image-based spatial transcriptomics with weakly supervised deep learning. *Cell Syst.* 15 (5), 475–482.e6. doi:10.1016/j.cels.2024.04.006
- Lee, P., Bubeck, S., and Petro, J. (2023). Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. *N. Engl. J. Med.* 388, 1233–1239. doi:10.1056/NEJMs2214184
- Li, B., Ma, C., Zhao, X., Hu, Z., Du, T., Xu, X., et al. (2018). YaTCM: yet another traditional Chinese medicine database for drug discovery. *Comput. Struct. Biotechnol. J.* 16, 600–610. doi:10.1016/j.csbj.2018.11.002
- Li, D., Hu, J., Zhang, L., Li, L., Yin, Q., Shi, J., et al. (2022a). Deep learning and machine intelligence: new computational modeling techniques for discovery of the combination rules and pharmacodynamic characteristics of Traditional Chinese Medicine. *Eur. J. Pharmacol.* 933, 175260. doi:10.1016/j.ejphar.2022.175260
- Li, H., Xu, W., Qiu, C., and Pei, J. (2023a). Fast markov clustering algorithm based on belief dynamics. *IEEE Trans. Cybern.* 53, 3716–3725. doi:10.1109/TCYB.2022.3141598
- Li, H., Zhang, R., Min, Y., Ma, D., Zhao, D., and Zeng, J. (2023b). A knowledge-guided pre-training framework for improving molecular representation learning. *Nat. Commun.* 14, 7568. doi:10.1038/s41467-023-43214-1
- Li, X., Peng, L., Wang, Y.-P., and Zhang, W. (2025). Open challenges and opportunities in federated foundation models towards biomedical healthcare. *BioData Min.* 18, 2. doi:10.1186/s13040-024-00414-9
- Li, X., Ren, J., Zhang, W., Zhang, Z., Yu, J., Wu, J., et al. (2022b). LTM-TCM: a comprehensive database for the linking of traditional Chinese medicine with modern medicine at molecular and phenotypic levels. *Pharmacol. Res.* 178, 106185. doi:10.1016/j.jphrs.2022.106185
- Li, X., Zhao, X., Yu, X., Zhao, J., and Fang, X. (2024). Construction of a multi-tissue compound-target interaction network of Qingfei Paidu decoction in COVID-19 treatment based on deep learning and transcriptomic analysis. *J. Bioinform. Comput. Biol.* 22, 2450016. doi:10.1142/S0219720024500161
- Lian, Z., Xu, T., Yuan, Z., Li, J., Thakor, N., and Wang, H. (2024). Driving fatigue detection based on hybrid electroencephalography and eye tracking. *IEEE J. Biomed. Health Inf.* 28, 6568–6580. doi:10.1109/JBHI.2024.3446952
- Lin, Y., Zhang, Y., Wang, D., Yang, B., and Shen, Y.-Q. (2022). Computer especially AI-assisted drug virtual screening and design in traditional Chinese medicine. *Phytomedicine Int. J. Phytother. Phytopharm.* 107, 154481. doi:10.1016/j.phymed.2022.154481
- Liu, J., Peng, D., Li, J., Dai, Z., Zou, X., and Li, Z. (2022). Identification of potential Parkinson's disease drugs based on multi-source data fusion and convolutional neural network. *Mol. Basel Switz.* 27, 4780. doi:10.3390/molecules27154780
- Liu, J., Shi, J.-L., Guo, J.-Y., Chen, Y., Ma, X.-J., Wang, S.-N., et al. (2023a). Anxiolytic-like effect of suanzaoren-wuweizi herb-pair and evidence for the involvement of the monoaminergic system in mice based on network pharmacology. *BMC Complement. Med. Ther.* 23, 7. doi:10.1186/s12906-022-03829-1
- Liu, L., Zhang, M., Li, C., Li, C., and Tang, J. (2024a). Cross-modal object tracking via modality-aware fusion network and a large-scale dataset. *IEEE Trans. Neural Netw. Learn. Syst.* PP, 1–14. doi:10.1109/TNNLS.2024.3406189
- Liu, M., Meng, X., Mao, Y., Li, H., and Liu, J. (2024b). ReduMixDTI: prediction of drug-target interaction with feature redundancy reduction and interpretable attention mechanism. *J. Chem. Inf. Model.* 64, 8952–8962. doi:10.1021/acs.jcim.4c01554
- Liu, X., Liu, J., Fu, B., Chen, R., Jiang, J., Chen, H., et al. (2023b). DCABM-TCM: a database of constituents absorbed into the blood and metabolites of traditional Chinese medicine. *J. Chem. Inf. Model.* 63, 4948–4959. doi:10.1021/acs.jcim.3c00365
- Liu, Z., Cai, C., Du, J., Liu, B., Cui, L., Fan, X., et al. (2020). TCMIO: a comprehensive database of traditional Chinese medicine on immuno-oncology. *Front. Pharmacol.* 11, 439. doi:10.3389/fphar.2020.00439
- Liu, Z., Tang, H., Amini, A., Yang, X., Mao, H., Rus, D. L., et al. (2023c). “BEVFusion: multi-task multi-sensor fusion with unified bird's-eye view representation,” in *2023 IEEE international conference on robotics and automation (ICRA)*, 2774–2781. doi:10.1109/ICRA48891.2023.10160968
- Luo, Z., Wu, W., Sun, Q., and Wang, J. (2024). Accurate and transferable drug-target interaction prediction with DrugLAMP. *Bioinforma. Oxf. Engl.* 40, btae693. doi:10.1093/bioinformatics/btae693
- Lv, Q., Chen, G., He, H., Yang, Z., Zhao, L., Chen, H.-Y., et al. (2023a). TCMBank: bridges between the largest herbal medicines, chemical ingredients, target proteins, and associated diseases with intelligence text mining. *Chem. Sci.* 14, 10684–10701. doi:10.1039/d3sc02139d
- Lv, Q., Chen, G., He, H., Yang, Z., Zhao, L., Zhang, K., et al. (2023b). TCMBank-the largest TCM database provides deep learning-based Chinese-Western medicine exclusion prediction. *Signal Transduct. Target. Ther.* 8, 127. doi:10.1038/s41392-023-01339-1
- Ma, S., Liu, J., Li, W., Liu, Y., Hui, X., Qu, P., et al. (2023). Machine learning in TCM with natural products and molecules: current status and future perspectives. *Chin. Med.* 18, 43. doi:10.1186/s13020-023-00741-9
- Magalhães, P. R., Reis, P. B. P. S., Vila-Viçosa, D., Machuqueiro, M., and Victor, B. L. (2021). Identification of Pan-Assay Interference compoundS (PAINS) using an MD-based protocol. *Methods Mol. Biol.* 2315, 263–271. doi:10.1007/978-1-0716-1468-6_15
- Mao, S., and Sejdíć, E. (2023). A review of recurrent neural network-based methods in computational physiology. *IEEE Trans. Neural Netw. Learn. Syst.* 34, 6983–7003. doi:10.1109/TNNLS.2022.3145365
- Marin, J., and Hedges, S. B. (2018). Undersampling genomes has biased time and rate estimates throughout the tree of life. *Mol. Biol. Evol.* 35, 2077–2084. doi:10.1093/molbev/msy103
- McDonagh, E. M., Trynka, G., McCarthy, M., Holzinger, E. R., Khader, S., Nakic, N., et al. (2024). Human genetics and genomics for drug target identification and prioritization: open targets' perspective. *Annu. Rev. Biomed. Data Sci.* 7, 59–81. doi:10.1146/annurev-biodatasci-102523-103838
- Meyer, G. P., Charland, J., Hegde, D., Laddha, A., and Vallespi-Gonzalez, C. (2019). “Sensor fusion for joint 3D object detection and semantic segmentation,” in *2019 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW)*, 1230–1237. doi:10.1109/CVPRW.2019.00162
- Ming, T., Tao, Q., Tang, S., Zhao, H., Yang, H., Liu, M., et al. (2022). Curcumin: an epigenetic regulator and its application in cancer. *Biomed. Pharmacother.* 156, 113956. doi:10.1016/j.biopha.2022.113956
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., et al. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst. Rev.* 4, 1. doi:10.1186/2046-4053-4-1
- Muhammad, K., Khan, S., Ser, J. D., and Albuquerque, V. H. C. de (2021). Deep learning for multigrade brain tumor classification in smart healthcare systems: a prospective survey. *IEEE Trans. Neural Netw. Learn. Syst.* 32, 507–522. doi:10.1109/TNNLS.2020.2995800
- Nedaie, A., and Najafi, A. A. (2018). Support vector machine with Dirichlet feature mapping. *Neural Netw. Off. J. Int. Neural Netw. Soc.* 98, 87–101. doi:10.1016/j.neunet.2017.11.006
- Niu, Q., Li, H., Tong, L., Liu, S., Zong, W., Zhang, S., et al. (2023). TCMFP: a novel herbal formula prediction method based on network target's score integrated with semi-supervised learning genetic algorithms. *Brief. Bioinform.* 24, bbad102. doi:10.1093/bib/bbad102
- Öcal, K., Gutmann, M. U., Sanguinetti, G., and Grima, R. (2022). Inference and uncertainty quantification of stochastic gene expression via synthetic models. *J. R. Soc. Interface* 19, 20220153. doi:10.1098/rsif.2022.0153
- Ornes, S. (2023). Peering inside the black box of AI. *Proc. Natl. Acad. Sci. U. S. A.* 120, e2307432120. doi:10.1073/pnas.2307432120
- Pan, Y., Zhang, H., Chen, Y., Gong, X., Yan, J., and Zhang, H. (2024). Applications of hyperspectral imaging technology combined with machine learning in quality control of traditional Chinese medicine from the perspective of artificial intelligence: a review. *Crit. Rev. Anal. Chem.* 54, 2850–2864. doi:10.1080/10408347.2023.2207652
- Pasa, L., Navarin, N., and Sperduti, A. (2022). Multiresolution reservoir graph neural network. *IEEE Trans. Neural Netw. Learn. Syst.* 33, 2642–2653. doi:10.1109/TNNLS.2021.3090503
- Piñero, J., Bravo, À., Queralt-Rosinach, N., Gutiérrez-Sacristán, A., Deu-Pons, J., Centeno, E., et al. (2017). DisGenET: a comprehensive platform integrating

- information on human disease-associated genes and variants. *Nucleic Acids Res.* 45, D833–D839. doi:10.1093/nar/gkw943
- Pitt, W. R., Bentley, J., Boldron, C., Colliandre, L., Esposito, C., Frush, E. H., et al. (2025). Real-world applications and experiences of AI/ML deployment for drug discovery. *J. Med. Chem.* 68, 851–859. doi:10.1021/acs.jmedchem.4c03044
- Qu, X., Du, G., Hu, J., and Cai, Y. (2024). Graph-DTI: a new model for drug-target interaction prediction based on heterogeneous network graph embedding. *Curr. Comput. Aided Drug Des.* 20, 1013–1024. doi:10.2174/1573409919666230713142255
- Razzaq, M., Clément, F., and Yvinec, R. (2022). An overview of deep learning applications in precocious puberty and thyroid dysfunction. *Front. Endocrinol.* 13, 959546. doi:10.3389/fendo.2022.959546
- Rhodes, J. S., Cutler, A., and Moon, K. R. (2023). Geometry- and accuracy-preserving random forest proximities. *IEEE Trans. Pattern Anal. Mach. Intell.* 45, 10947–10959. doi:10.1109/TPAMI.2023.3263774
- Ru, J., Li, P., Wang, J., Zhou, W., Li, B., Huang, C., et al. (2014). TCMSP: a database of systems pharmacology for drug discovery from herbal medicines. *J. Cheminformatics* 6, 13. doi:10.1186/1758-2946-6-13
- Savargiv, M., Masoumi, B., and Keyvanpour, M. R. (2021). A new random forest algorithm based on learning automata. *Comput. Intell. Neurosci.* 2021, 5572781. doi:10.1155/2021/5572781
- Seetharam, K., Kagiya, N., and Sengupta, P. P. (2019). Application of mobile health, telemedicine and artificial intelligence to echocardiography. *Echo Res. Pract.* 6, R41–R52–R52. doi:10.1530/ERP-18-0081
- Shin, H. (2022). XGBoost regression of the most significant photoplethysmogram features for assessing vascular aging. *IEEE J. Biomed. Health Inf.* 26, 3354–3361. doi:10.1109/JBHI.2022.3151091
- Shorten, C., Khoshgoftaar, T. M., and Furht, B. (2021). Text data augmentation for deep learning. *J. Big Data* 8, 101. doi:10.1186/s40537-021-00492-0
- Soffer, S., Ben-Cohen, A., Shimon, O., Amitai, M. M., Greenspan, H., and Klang, E. (2019). Convolutional neural networks for radiologic images: a radiologist's guide. *Radiology* 290, 590–606. doi:10.1148/radiol.2018180547
- Solorio-Ramírez, J.-L., Saldana-Perez, M., Lytras, M. D., Moreno-Ibarra, M.-A., and Yáñez-Márquez, C. (2021). Brain hemorrhage classification in CT scan images using minimalist machine learning. *Diagn. Basel Switz.* 11, 1449. doi:10.3390/diagnostics11081449
- Song, Z., Chen, G., and Chen, C. Y.-C. (2024). AI empowering traditional Chinese medicine? *Chem. Sci.* 15, 16844–16886. doi:10.1039/D4SC04107K
- Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., et al. (2023). The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* 51, D638–D646. doi:10.1093/nar/gkac1000
- Tang, Q., and Wu, B. (2022). Multilayer game collaborative optimization based on elman neural network system diagnosis in shared manufacturing mode. *Comput. Intell. Neurosci.* 2022, 6135970. doi:10.1155/2022/6135970
- The Gene Ontology Consortium (2017). Expansion of the gene ontology knowledgebase and resources. *Nucleic Acids Res.* 45, D331–D338. doi:10.1093/nar/gkw1108
- Tian, S., Zhang, J., Yuan, S., Wang, Q., Lv, C., Wang, J., et al. (2023). Exploring pharmacological active ingredients of traditional Chinese medicine by pharmacotranscriptomic map in ITCM. *Brief. Bioinform.* 24, bbad027. doi:10.1093/bib/bbad027
- Vaid, A., Duong, S. Q., Lampert, J., Kovatch, P., Freeman, R., Argulian, E., et al. (2024). Local large language models for privacy-preserving accelerated review of historic echocardiogram reports. *J. Am. Med. Inf. Assoc. JAMIA* 31, 2097–2102. doi:10.1093/jamia/ocae085
- Vandenhende, S., Georgoulis, S., Van Gansbeke, W., Proesmans, M., Dai, D., and Van Gool, L. (2022). Multi-task learning for dense prediction tasks: a survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 44, 3614–3633. doi:10.1109/TPAMI.2021.3054719
- Wang, D., He, F., Maslov, S., and Gerstein, M. (2016a). DREISS: using state-space models to infer the dynamics of gene expression driven by external and internal regulatory networks. *PLoS Comput. Biol.* 12, e1005146. doi:10.1371/journal.pcbi.1005146
- Wang, J., Wang, J., Fang, W., and Niu, H. (2016b). Financial time series prediction using elman recurrent random neural networks. *Comput. Intell. Neurosci.* 2016, 4742515. doi:10.1155/2016/4742515
- Wang, Y., Shi, X., Li, L., Efferth, T., and Shang, D. (2021). The impact of artificial intelligence on traditional Chinese medicine. *Am. J. Chin. Med.* 49, 1297–1314. doi:10.1142/S0192415X21500622
- Wang, Z., Liang, S., Liu, S., Meng, Z., Wang, J., and Liang, S. (2023). Sequence pre-training-based graph neural network for predicting lncRNA-miRNA associations. *Brief. Bioinform.* 24, bbad317. doi:10.1093/bib/bbad317
- Wishart, D. S., Guo, A., Oler, E., Wang, F., Anjum, A., Peters, H., et al. (2021). HMDB 5.0: the human metabolome database for 2022. *Nucleic Acids Res.* 50, D622–D631. doi:10.1093/nar/gkab1062
- Wu, J., Hu, R., Xiao, Z., Chen, J., and Liu, J. (2021). Vision transformer-based recognition of diabetic retinopathy grade. *Med. Phys.* 48, 7850–7863. doi:10.1002/mp.15312
- Wu, J., Luo, Y., Shen, Y., Hu, Y., Zhu, F., Wu, J., et al. (2022). Integrated metabolomics and network pharmacology to reveal the action mechanism effect of shaoyao decoction on ulcerative colitis. *Drug Des. devel. Ther.* 16, 3739–3776. doi:10.2147/DDDT.S375281
- Wu, Y., Zhang, F., Yang, K., Fang, S., Bu, D., Li, H., et al. (2018). SymMap: an integrative database of traditional Chinese medicine enhanced by symptom mapping. *Nucleic Acids Res.* 47, D1110–D1117. doi:10.1093/nar/gky1021
- Xiao, X., Ezugwu, A. L., Chukwuma, I. F., Anaduaka, E. G., and Udenigwe, C. C. (2024). Health-promoting properties of bioactive proteins and peptides of garlic (*Allium sativum*). *Food Chem.* 435 (2024), 137632–137643. doi:10.1016/j.foodchem.2023.137632
- Xing, Z., Peng, F., Chen, Y., Wan, F., Peng, C., and Li, D. (2024). Metabolomic profiling integrated with molecular exploring delineates the action of Ligusticum chuanxiong hort. on migraine. *Phytomedicine* 134, 155977. doi:10.1016/j.phymed.2024.155977
- Xu, H., Wang, S., Fang, M., Luo, S., Chen, C., Wan, S., et al. (2023). SPACEL: deep learning-based characterization of spatial transcriptome architectures. *Nat. Commun.* 14 (1), 7603. doi:10.1038/s41467-023-43220-3
- Xu, M., Deng, J., Xu, K., Zhu, T., Han, L., Yan, Y., et al. (2019). In-depth serum proteomics reveals biomarkers of psoriasis severity and response to traditional Chinese medicine. *Theranostics* 9, 2475–2488. doi:10.7150/thno.31144
- Xu, X., Gao, Z., Yang, F., Yang, Y., Chen, L., Han, L., et al. (2020). Antidiabetic effects of gegen qinlian decoction via the gut microbiota are attributable to its key ingredient berberine. *Genomics Proteomics Bioinform.* 18, 721–736. doi:10.1016/j.gpb.2019.09.007
- Yan, D., Zheng, G., Wang, C., Chen, Z., Mao, T., Gao, J., et al. (2022). HIT 2.0: an enhanced platform for herbal ingredients' targets. *Nucleic Acids Res.* 50, D1238–D1243. doi:10.1093/nar/gkab1011
- Yang, G., Ye, Q., and Xia, J. (2022c). Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: a mini-review, two showcases and beyond. *Int. J. Inf. Fusion* 77, 29–52. doi:10.1016/j.inffus.2021.07.016
- Yang, H., Cao, H., He, T., Wang, T., and Cui, Y. (2020). Multilevel heterogeneous omics data integration with kernel fusion. *Brief. Bioinform.* 21, 156–170. doi:10.1093/bib/bby115
- Yang, J., Wang, L., Liu, L., and Zheng, X. (2024). GraphPCA: a fast and interpretable dimension reduction algorithm for spatial transcriptomics data. *Genome Biol.* 25 (1), 287. doi:10.1186/s13059-024-03429-x
- Yang, P., Lang, J., Li, H., Lu, J., Lin, H., Tian, G., et al. (2022a). TCM-Suite: a comprehensive and holistic platform for Traditional Chinese Medicine component identification and network pharmacology analysis. *iMeta* 1, e47. doi:10.1002/imt2.47
- Yang, R., Yin, L., Hao, X., Liu, L., Wang, C., Li, X., et al. (2022b). Identifying a suitable model for predicting hourly pollutant concentrations by using low-cost microstation data and machine learning. *Sci. Rep.* 12, 19949. doi:10.1038/s41598-022-24470-5
- Ye, H., Gao, Y., Zhang, Y., Cao, Y., Zhao, L., Wen, L., et al. (2020). Study on intelligent syndrome differentiation neural network model of stomachache in traditional Chinese medicine based on the real world. *Med. Baltim.* 99, e20316. doi:10.1097/MD.00000000000020316
- Yu, S.-Z. (2022). Explicit duration recurrent networks. *IEEE Trans. Neural Netw. Learn. Syst.* 33, 3120–3130. doi:10.1109/TNNLS.2021.3051019
- Zavadlav, J., Marrink, S. J., and Praprotnik, M. (2019). SWINGER: a clustering algorithm for concurrent coupling of atomistic and supramolecular liquids. *Interface Focus* 9, 20180075. doi:10.1098/rsfs.2018.0075
- Zeng, X., Xiang, H., Yu, L., Wang, J., Li, K., Nussinov, R., et al. (2022). Accurate prediction of molecular targets using a self-supervised image representation learning framework. *Res. Sq.* doi:10.21203/rs.3.rs-1477870/v1
- Zhang, L.-X., Dong, J., Wei, H., Shi, S.-H., Lu, A.-P., Deng, G.-M., et al. (2022a). TCMSID: a simplified integrated database for drug discovery from traditional Chinese medicine. *J. Cheminformatics* 14, 89. doi:10.1186/s13321-022-00670-z
- Zhang, P., Wang, B., and Li, S. (2023a). Network-based cancer precision prevention with artificial intelligence and multi-omics. *Sci. Bull.* 68, 1219–1222. doi:10.1016/j.scib.2023.05.023
- Zhang, P., Zhang, D., Zhou, W., Wang, L., Wang, B., Zhang, T., et al. (2023b). Network pharmacology: towards the artificial intelligence-based precision traditional Chinese medicine. *Brief. Bioinform.* 25, bbad518. doi:10.1093/bib/bbad518
- Zhang, P., Zhang, Q., and Li, S. (2024a). Advancing cancer prevention through an AI-based integration of traditional and western medicine. *Cancer Discov.* 14, 2033–2036. doi:10.1158/2159-8290.CD-24-0832
- Zhang, Q., Guo, Y., Zhang, B., Liu, H., Peng, Y., Wang, D., et al. (2022b). Identification of hub biomarkers of myocardial infarction by single-cell sequencing, bioinformatics, and machine learning. *Front. Cardiovasc. Med.* 9, 939972. doi:10.3389/fcvm.2022.939972

- Zhang, R., Zhu, X., Bai, H., and Ning, K. (2019a). Network pharmacology databases for traditional Chinese medicine: review and assessment. *Front. Pharmacol.* 10, 123. doi:10.3389/fphar.2019.00123
- Zhang, R.-Z., Yu, S.-J., Bai, H., and Ning, K. (2017). TCM-Mesh: the database and analytical system for network pharmacology analysis for TCM preparations. *Sci. Rep.* 7, 2821. doi:10.1038/s41598-017-03039-7
- Zhang, S., Wang, W., Pi, X., He, Z., and Liu, H. (2023c). Advances in the application of traditional Chinese medicine using artificial intelligence: a review. *Am. J. Chin. Med.* 51, 1067–1083. doi:10.1142/S0192415X23500490
- Zhang, S., Zhang, X., Du, J., Wang, W., and Pi, X. (2024b). Multi-target meridians classification based on the topological structure of anti-cancer phytochemicals using deep learning. *J. Ethnopharmacol.* 319, 117244. doi:10.1016/j.jep.2023.117244
- Zhang, Y., Li, J., Lin, S., Zhao, J., Xiong, Y., and Wei, D.-Q. (2024c). An end-to-end method for predicting compound-protein interactions based on simplified homogeneous graph convolutional network and pre-trained language model. *J. Cheminformatics* 16, 67. doi:10.1186/s13321-024-00862-9
- Zhang, Y., Li, X., Shi, Y., Chen, T., Xu, Z., Wang, P., et al. (2023d). ETCM v2.0: an update with comprehensive resource and rich annotations for traditional Chinese medicine. *Acta Pharm. Sin. B* 13, 2559–2571. doi:10.1016/j.apsb.2023.03.012
- Zhang, Y., Miao, D., Wang, J., and Zhang, Z. (2019b). A cost-sensitive three-way combination technique for ensemble learning in sentiment classification. *Int. J. Approx. Reason.* 105, 85–97. doi:10.1016/j.ijar.2018.10.019
- Zhang, Z., and Jung, C. (2021). GBDT-MO: gradient-boosted decision trees for multiple outputs. *IEEE Trans. Neural Netw. Learn. Syst.* 32, 3156–3167. doi:10.1109/TNNLS.2020.3009776
- Zhao, W., Wang, B., Kong, L., Wang, Q., and Li, S. (2024a). Clinical multi-omics reveals the role of tuomin zhiti decoction intervention in allergic rhinitis from the perspective of biological network. 24303911. doi:10.1101/2024.03.10.24303911
- Zhao, Z., Qiang, Y., Yang, F., Hou, X., Zhao, J., and Song, K. (2024b). Two-stream vision transformer based multi-label recognition for TCM prescriptions construction. *Comput. Biol. Med.* 170, 107920. doi:10.1016/j.combiomed.2024.107920
- Zheng, J., Zhang, Z., Wang, J., Zhao, R., Liu, S., Yang, G., et al. (2023). Metabolic syndrome prediction model using Bayesian optimization and XGBoost based on traditional Chinese medicine features. *Heliyon* 9, e22727. doi:10.1016/j.heliyon.2023.e22727
- Zhou, E., Shen, Q., and Hou, Y. (2024a). Integrating artificial intelligence into the modernization of traditional Chinese medicine industry: a review. *Front. Pharmacol.* 15, 1181183. doi:10.3389/fphar.2024.1181183
- Zhou, G., Pang, Z., Lu, Y., Ewald, J., and Xia, J. (2022). OmicsNet 2.0: a web-based platform for multi-omics integration and network visual analytics. *Nucleic Acids Res.* 50, W527–W533. doi:10.1093/nar/gkac376
- Zhou, Y., Zhang, Y., Zhao, D., Yu, X., Shen, X., Zhou, Y., et al. (2024b). TTD: therapeutic target database describing target druggability information. *Nucleic Acids Res.* 52, D1465–D1477. doi:10.1093/nar/gkad751
- Zhu, L., Liu, D., Xu, M., Wang, W., Xiong, X., Zhou, Q., et al. (2024). Yantiao formula intervention in rats with sepsis: network pharmacology and experimental analysis. *Comb. Chem. High. Throughput Screen.* 27, 1071–1080. doi:10.2174/0113862073262718230921113659
- Zhu, X., Yao, Q., Yang, P., Zhao, D., Yang, R., Bai, H., et al. (2022a). Multi-omics approaches for in-depth understanding of therapeutic mechanism for traditional Chinese medicine. *Front. Pharmacol.* 13, 1031051. doi:10.3389/fphar.2022.1031051
- Zhu, Y., Ouyang, Z., Du, H., Wang, M., Wang, J., Sun, H., et al. (2022b). New opportunities and challenges of natural products research: when target identification meets single-cell multiomics. *Acta Pharm. Sin. B* 12, 4011–4039. doi:10.1016/j.apsb.2022.08.022