



Review paper

The integration of machine learning into traditional Chinese medicine

Yanfeng Hong^{a, b}, Sisi Zhu^{a, b}, Yuhong Liu^{a, b}, Chao Tian^{a, b}, Hongquan Xu^{a, b},
Gongxing Chen^{a, b}, Lin Tao^{a, b, **}, Tian Xie^{a, b, *}

^a School of Pharmacy, Hangzhou Normal University, Hangzhou, 311121, China

^b Key Laboratory of Elemene Class Anti-Cancer Chinese Medicines, Engineering Laboratory of Development and Application of Traditional Chinese Medicines, Collaborative Innovation Center of Traditional Chinese Medicines of Zhejiang Province, Hangzhou Normal University, Hangzhou, 311121, China



ARTICLE INFO

Article history:

Received 21 August 2024

Received in revised form

11 November 2024

Accepted 1 December 2024

Available online 4 December 2024

Keywords:

Traditional Chinese medicine

Machine learning

Quality control

Drug development

Precision medicine

ABSTRACT

Traditional Chinese medicine (TCM) is an ancient medical system distinctive and effective in treating cancer, depression, coronavirus disease 2019 (COVID-19), and other diseases. However, the relatively abstract diagnostic methods of TCM lack objective measurement, and the complex mechanisms of action are difficult to comprehend, which hinders the application and internationalization of TCM. Recently, while breakthroughs have been made in utilizing methods such as network pharmacology and virtual screening for TCM research, the rise of machine learning (ML) has significantly enhanced their integration with TCM. This article introduces representative methodological cases in quality control, mechanism research, diagnosis, and treatment processes of TCM, revealing the potential applications of ML technology in TCM. Furthermore, the challenges faced by ML in TCM applications are summarized, and future directions are discussed.

© 2024 The Authors. Published by Elsevier B.V. on behalf of Xi'an Jiaotong University. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Traditional Chinese medicine (TCM) is one of the oldest medical systems and has gained increasing recognition and application worldwide over the past century [1]. Modern research has confirmed that TCM effectively treats various diseases, including cancer [2], depression [3], and coronavirus disease 2019 (COVID-19) [4]. Additionally, numerous studies have emphasized that various natural products (NPs) derived from TCM play a crucial role in their therapeutic efficacy [5,6]. For example, artemisinin, an NP derived from the Chinese medicinal plant *Artemisia annua*, is the most widely used compound against malaria and has saved millions of lives worldwide [7]. Similarly, in tumor therapy, β -elemene (extracted from the Chinese medicinal plant *Curcuma Rhizoma*) achieves synergistic antitumor

outcomes, with no serious adverse reactions reported [8,9]. As a safe alternative to conventional drugs, TCM has attracted the attention of researchers due to its multi-component, multi-target, and multi-pathway pharmacological effects, but these characteristics render it challenging to comprehensively understand their mode of action [10]. Meanwhile, the quality of TCM materials and the prescription of medicine based on expert experience both affect the therapeutic efficacy of TCM [11,12]. Addressing these issues requires significant resources from high-quality research laboratories and the establishment of comprehensive quality standards for TCM [13,14].

With the introduction of computer technology into TCM, technologies like virtual screening, molecular docking, and high-throughput assays have driven the rapid development of TCM research [15,16]. However, these traditional computational methods have performance limitations and limited applicability, making it difficult to address the complex correlations between TCM prescriptions and their targets [17–19]. To integrate large-scale biological system features, such as gene expression and metabolic pathways, derived from experimental data, systems biology has been proposed to model the structure and function of biological systems [20]. Network pharmacology, a branch of systems biology, focuses on elucidating pharmacological mechanisms [21,22]. Recently, to clarify the compatibility principles and pharmacological mechanisms of TCM, researchers have recognized the value of combining TCM databases with network pharmacology

* Corresponding author. School of Pharmacy, Hangzhou Normal University, Hangzhou, 311121, China.

** Corresponding author. Key Laboratory of Elemene Class Anti-Cancer Chinese Medicines, Engineering Laboratory of Development and Application of Traditional Chinese Medicines, Collaborative Innovation Center of Traditional Chinese Medicines of Zhejiang Province, Hangzhou Normal University, Hangzhou, 311121, China.

E-mail addresses: taolin@hznu.edu.cn (L. Tao), xbs@hznu.edu.cn (T. Xie).

Peer review under responsibility of Xi'an Jiaotong University.

tools [23–25]. Using this knowledge system, researchers have developed various practical network pharmacology tools to explore TCM mechanisms in treating diseases such as Alzheimer's disease [26], colorectal cancer [27], and COVID-19 [28]. However, variations in data across multiple databases, weak correlations between drug components and their targets, and incomplete protein-protein interaction networks pose significant challenges to advancing network pharmacology [29].

In recent years, with the rapid development of artificial intelligence (AI), the technology has been applied to all aspects of human life [30]. Correspondingly, AI offers unique advantages in addressing current challenges in the medical field, such as severe diagnostic and treatment errors, poor quality control in drugs, and unclear pharmacological mechanisms [31]. While there are many problems and challenges in current TCM research, with the assistance of AI, particularly machine learning (ML), ancient TCM is being rejuvenated with new vitality [32,33]. In this review, we will introduce representative methodological cases and ML methods, explore the potential of ML technology, and elaborate on whether it has met researchers' expectations in the field of TCM (Fig. 1). For

ease of reading, the glossary of Chinese and English terms is shown in Table S1.

2. Application of ML in TCM quality control

2.1. Cultivating and processing of TCM

The quality of TCM is essential for its clinical efficacy and safety, which depend on two main factors: the quality of raw materials and the processing of medicines [34]. High-quality raw materials significantly enhance the therapeutic effects of TCM, while processing can modify the chemical composition of TCM, leading to diverse therapeutic properties [35–37]. However, issues such as heavy metals, soil pollution, and other contaminants in raw materials significantly impact the processing and safety of TCM [38]. Furthermore, TCM processing faces challenges including low efficiency, poor standardization, and inconsistent techniques [39,40].

The rapidly advancing field of AI is progressively addressing these dilemmas. Currently, the combination of ML and hyperspectral imaging (HSI) is frequently used for quality control in TCM

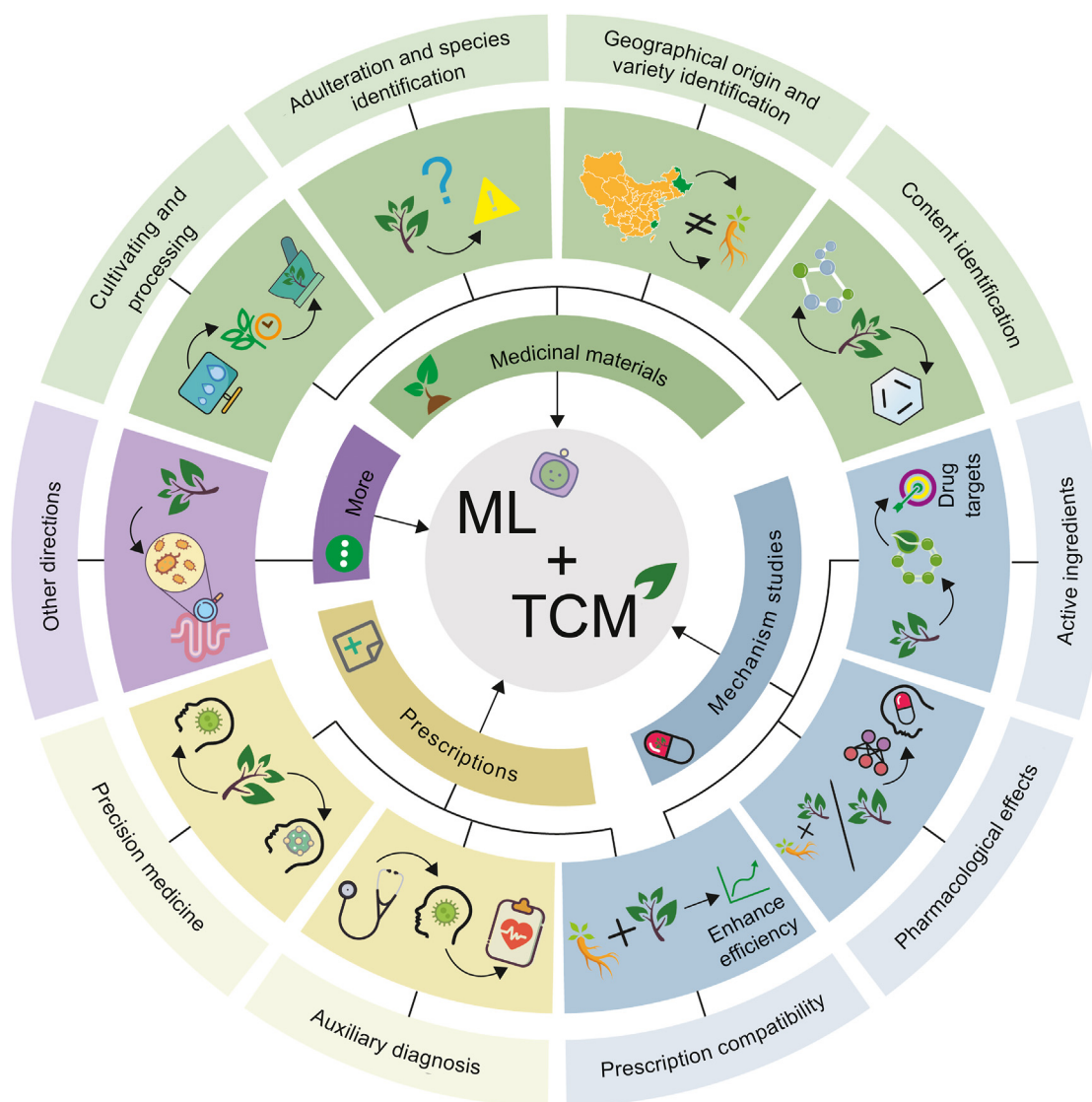


Fig. 1. The application of machine learning (ML) in traditional Chinese medicine (TCM) mainly targets three fields: quality control of medicinal materials, exploration of TCM mechanisms, and personalized prescription compatibility.

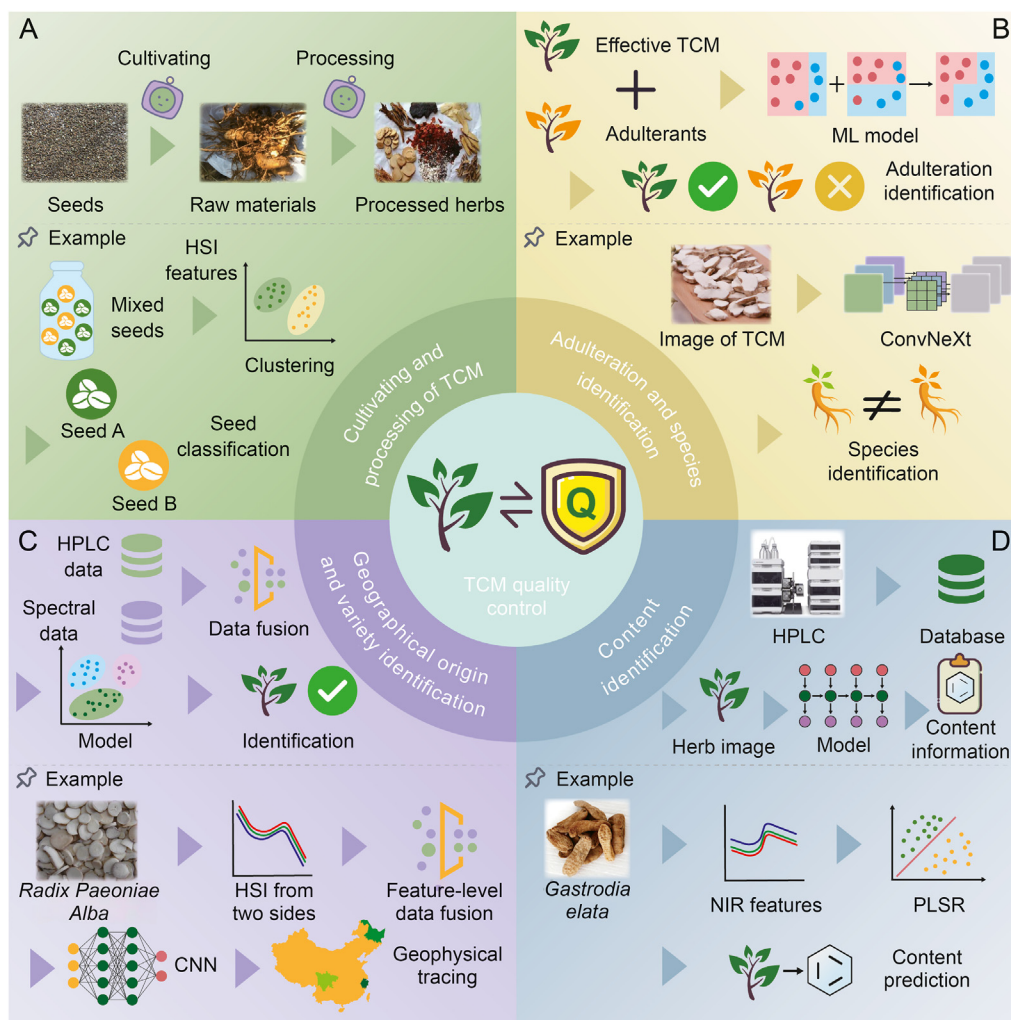


Fig. 2. Application of machine learning (ML) method in quality control of traditional Chinese medicine (TCM). (A) Cultivation and processing of TCM: For example, high-precision differentiation of *Astragalus membranaceus* seeds using Support Vector Machine (SVM) modeling based on hyperspectral imaging (HSI) data. (B) Adulteration and species identification: For example, utilizing an improved ConvNeXt network for feature extraction and classification of TCM photos to construct a classification model. (C) Geographical origin and variety identification: For example, identification of the geographical origins of *Radix Paeoniae Alba* using HSI combined with convolutional neural network (CNN)-based fusion approaches. (D) Content identification: For example, predicting gastrodin content in fresh *Gastrodia elata* using Fourier transform near-infrared (NIR) spectroscopy and partial least squares regression (PLSR) model. HPLC: high-performance liquid chromatography.

cultivation and processing (Fig. 2A). Seeds are fundamental to TCM cultivation and face specific challenges, including mixed germ-plasm resources and uneven seed quality [41]. To address this issue, Xu et al. [42] combined ML and HSI to non-destructively and accurately distinguish *Astragalus membranaceus* from similar seeds (such as *Melilotus officinalis*) by innovatively applying support vector machine (SVM) classification models to HSI data. Bai et al. [43] established five ML and deep learning (DL) models based on HSI data, and found that the model called CARS-ResNet was effective in identifying the storage age of Coix seed. Additionally, many studies have established intelligent Raman spectroscopy and near-infrared spectroscopy (NIRS) methods, often in combination with DL, for monitoring the processing of TCM, greatly improving the standardization and intelligence of the production process [44–46]. Liu et al. [47] developed three ML models using nine selected indicator compounds to accurately identify different processing levels of ginger. This fast and cost-effective strategy shows significant potential for broad application. Although ML technology has provided many practical strategies in TCM cultivation and processing, it still needs to be combined with expertise in the field of TCM to maximize its potential.

2.2. Adulteration and species identification

Adulteration in TCM has long been a major concern for researchers due to its potential to reduce treatment effects and increase toxicity [48,49]. Adulterants often do not alter the taste of herbs and are challenging to identify through visual inspection alone, leading to substantial identification errors [50]. However, ML techniques have demonstrated improvements in the accuracy of TCM adulteration identification, leading to their widespread adoption (Fig. 2B). Jiang et al. [51] introduced a partial least squares regression (PLSR) algorithm based on NIRS and HSI spectral features, resulting in accurate adulteration prediction of *Atractylodes rhizoma*, a herb of significant medicinal value. Saffron is particularly susceptible to adulteration due to its high value, and numerous studies have developed ML models to detect saffron adulteration [52–55]. For instance, Husaini et al. [52] employed saffron images and their stains to develop an effective convolutional neural network (CNN) model. This model, combined with Foldscope, swiftly identifies adulterated saffron, outperforming traditional ML methods.

Species identification is crucial for effective TCM research and management [56]. Adulterated products often involve species with

similar shapes or vernacular names, confusing species identification [57]. For example, *Ziziphi Spinosae* Semen is known for its sedative and sleep-inducing properties. However, due to their similar shapes and colors, it is often adulterated with *Hovenia Acerba* Semen and *Ziziphi Mauritanae* Semen by merchants motivated by profit [58]. Li et al. [59] successfully addressed its adulteration problem by optimizing the Fourier transform near-infrared (FT-NIR) raw spectra using a multivariate algorithm. Miao et al. [14] developed a TCM species identification model by fusing ConvNeXt and ACMix networks, achieving highly accurate recognition results on a test set of images from six herb categories (Fig. 2B). In conclusion, different ML algorithm strategies have provided an effective solution for the adulteration and species identification of TCM to a certain extent. However, this method requires ongoing refinement to enhance its widespread application.

2.3. Geographical origin and variety identification

The same species may display different chemical compositions under various geographical environments and seasonal conditions [14]. For example, the chemical composition of the underground parts of *G. rigescens* varies significantly based on the plant's geographical origin or production area [60]. This variation emphasizes the urgent need for accurate geographic origin identification of TCM, highlighting the necessity for a scientific and reliable method. Furthermore, the use of multi-source data imposes greater demands on the selection of data fusion strategies [61] (Fig. 2C). Liu et al. [62] proposed a novel geographical identification strategy, initially analyzing the content of *G. rigescens* from different regions using high-performance liquid chromatography (HPLC), followed by sequential preprocessing through orthogonalization (SPORT) for prediction. Compared with the original spectral model, the SPORT-based partial least squares discriminant analysis model (PLS-DA) improved the accuracy of the test set by 48.5% in the qualitative analysis. Cai et al. [63] employed spectral images from two sides of a *Radix Paeoniae Alba* (Baishao) sample in a CNN model to accurately determine its geographic origin.

Medicinal plants within the same genus often exhibit similar appearances [14,50]. The process of variety identification is highly complex, and relying solely on visual and experiential recognition methods cannot achieve efficient and rapid identification [64,65]. For instance, *Fritillaria thunbergii*, one of the oldest homeopathic traditional medicines in China, has diverse varieties that lead to variations in its active ingredients [66]. Kabir et al. [67] combined HSI and CNN models with cross-validation to optimize the performance of the models, enabling the successful differentiation of twelve varieties of *Fritillaria thunbergii*. Cinnamon is widely used in TCM formulas, Cruz-Tirado et al. [68] used PLS-DA and SVM models for the classification of the varieties of Cinnamon samples, and achieved classification accuracy rates exceeding 90%.

2.4. Content identification

The efficacy of TCM primarily depends on its contents, namely the various chemical and active ingredients it contains [69]. Therefore, predicting the content of herbs is crucial for evaluating the quality of TCM [50]. Numerous studies have utilized ML to tackle the challenge of content identification [70,71] (Fig. 2D). Zhu et al. [72] constructed four quantitative analysis models to optimize the flaxseed nutritional quality testing method. The Savitzky-Golay (SG)-Competitive Adaptive Reweighted Sampling (CARS)-Multiple Linear Regression (MLR) model showed the best potential to accurately derive flaxseed protein, oil content, linoleic acid, and lignin content based on HSI. Ma et al. [73] proposed combining ML with visible-near-infrared (Vis-NIR) spectroscopy to analyze the

content of physiologically active components in *Gastrodia elata* Blume and found the CNN model to be promising in this process. Additionally, due to the time-consuming nature and other defects of traditional *Puerariae Thomsonii* Radix measurement methods, Hu et al. [74] proposed an innovative and efficient method combining HSI technology and DL, offering a feasible strategy for assessing *Puerariae Thomsonii* Radix contents. These studies collectively underscore the potential of integrating HSI methods with ML. Identifying the content of TCM is crucial and warrants increased attention from researchers in the future. A detailed explanation of various application of ML in TCM quality control is shown in Table 1 [14,42,43,47,51,52,59,62,63,67,72–74].

3. Application of ML in TCM mechanism study at different levels

3.1. Exploring the mechanism of TCM at the active ingredients level

Although the discovery of drugs based on the active ingredients found in TCM (such as ginsenoside Rg3, cepharanthine, and β -elemene) shows great therapeutic potential in various diseases, the traditional drug discovery and design process is time-consuming and characterized by a low success rate [6,75,76]. Fortunately, the rapid advancement of ML technology in drug research and development has created new opportunities to address these challenges [77]. For instance, Yu et al. [78] developed a model based on ML and molecular docking methods, successfully identifying 4',5,7-trimethoxyflavonoids as potential inhibitors of the platelet-derived growth factor receptor α [78] (Fig. 3A). Using a multi-view DL model, Zhang et al. [79] successfully predicted the activation/inhibition mechanisms of NPs and drug targets. They also identified the mechanism relationships for 61 NP-target pairs beyond the training data. Yang et al. [80] introduced an ML model called forgeNet-SVM to identify potential new active candidate compounds for Alzheimer's disease. In recent years, curative effect prediction methods have also been employed to mine active molecules with strong biological activity and low side effects from TCM ingredient databases [81,82]. However, current understanding of the active ingredients in TCM is limited [83]. The ML approach may elucidate active ingredient-target interactions, facilitating the identification of potential candidate molecules [84].

3.2. Mining the mechanism of TCM at the pharmacological effects level

TCM is renowned for its multi-target and multi-pathway treatment strategies, which, while effective, complicate the understanding of its pharmacological mechanisms [8,34,85]. To address this complexity, contemporary network pharmacology employs "network targets" and "multi-component" methods to enhance target research, effectively elucidating the overall relationships between specific TCM and various diseases [86]. The advancement of network pharmacology is intrinsically linked to the exploration of systems biology, which encompasses protein interactions, cell signaling, and gene regulatory networks [21,87,88]. In this context, advanced deep learning tools such as Generalized Linear Mixed Models (GLIM), which decodes multi-level relationships in human tissue cell molecular networks [89], MoDAFold, predicting the structures of missense mutant proteins [90], and CAMP, which forecasts multi-level peptide-protein interactions [91], have significantly contributed to the development of network pharmacology.

With the advancement of algorithms, ML-based methods have offered a novel approach to addressing pharmacological challenges [92]. For example, Hu et al. [93] introduced a novel dual-channel

Table 1

Application of machine learning (ML) in traditional Chinese medicine (TCM) quality control.

Areas of TCM	Aim of study	Data	AI methods	Results	Advantages	Refs.
Cultivation and processing	Identify <i>Astragalus membranaceus</i> seeds from similar seeds.	HSI and general images of seeds	SVM; PLS-DA; ANN	ML models based on HSI data were used to accurately distinguish <i>Astragalus</i> seeds.	The first derivative-uninformative variable elimination-SVM model based on hyperspectral data achieved 100.0% accuracy in the validation set.	[42]
	Identify and classify Coix seed storage years.	HSI images of Coix seed	ResNet; SVM; KNN; RF; XGBoost	The DL model achieved can be used to accurately distinguish Coix seed storage years, exhibiting a recognition accuracy exceeding 85%	Compared to the other ML, the accuracy of the CARS-ResNet model was improved by more than 10%.	[43]
	Identify different degrees of processed ginger.	Volatile compounds in different degrees of processed ginger analyzed by GC-IMS	SVM; KNN; XGBoost; LR; RF; Linear discriminant analysis; QDA	Three ML models were developed based on GC-IMS data to distinguish different degrees of processed ginger. A TCM classification model was established, achieving good recognition results.	Compared to the other ML, the accuracy of SVM with linear kernel, LR, and QDA was improved by 2.94%–14.61%. This method incorporates a Stacked ACMix module into the second ConvNeXt block layer within the ConvNeXt network, enhancing the effectiveness of low-dimensional feature extraction from images.	[47]
Adulteration and species identification	Identify herbal varieties through images.	Plant images	CNN		When combined with CARS data and color-texture features for prediction, the model performs the best.	[14]
	Identify adulteration of <i>Atractylodes rhizome</i> .	HSI and near-infrared spectroscopy data	PLSR	An ML model was constructed for predicting the adulteration of <i>Atractylodes rhizoma</i> powder based on HSI and near-infrared spectroscopy data and image feature fusion.		[51]
	Low-cost, rapid identification of fake look-alike saffron.	Plant images and staining images	CNN; RF; SVM	The CNN model achieved 99.5% accuracy and 99.3% precision on the test dataset.	Deep neural networks (ResNet18 and DenseNet121) performed better than RF and SVM-based approaches.	[52]
Geographical origins and variety identification	Identify the adulteration of <i>Ziziphi Spinosae</i> Semen and quantitatively predict the adulteration ratio.	FT-NIR spectral data	PLSR; SVM; KNN; ANN	FT-NIR combined with ML models improved adulteration detection accuracy.	Compared to the raw spectra, the average classification accuracy of ML models increased from 77.06% to 97.58%.	[59]
	Trace the geographical origin of <i>Gentiana iridescent</i> from various regions.	HPLC-MS, fourier transform infrared spectra, and two-dimensional correlation spectra data	PLS-DA; PLSR	A chemical method was built using multi-preprocessing integration for tracing the geographical origins of wild <i>Gentiana rigescen</i> .	Compared to scatter correction modeling, the accuracy was improved by 12.8% through sequential preprocessing through orthogonalization.	[62]
	Identify the geographical origin of <i>Radix Paeoniae Alba</i> .	HSI image of the <i>Radix Paeoniae Alba</i> dried slices	KNN; LR; SVM; RF; XGBoost; CNN	Using attention mechanism-based CNN, <i>Radix Paeoniae Alba</i> samples from three geographical origins were successfully distinguished.	CNN models achieved classification accuracies ranging from 93.64% to 95.27%, outperforming conventional ML models, which achieved accuracies between 72.00% and 90.55%.	[63]
	Identify the variety of <i>Fritillaria thunbergii</i> .	HSI images	CNN; PLS-DA; SVM	The application of HSI in conjunction with the deep learning technique can be used for the classification of <i>Fritillaria thunbergii</i> varieties rapidly and non-destructively.	CNN was the most accurate with 98.88%, and 88.89% in training and test sets, followed by PLS-DA and SVM with 92.59%, 81.94%, and 99.65%, 79.17%, respectively.	[67]

(continued on next page)

Table 1 (continued)

Areas of TCM	Aim of study	Data	AI methods	Results	Advantages	Refs.
Content identification						
	Predict flaxseed protein, oil content, linoleic acid, and lignan content.	HSI images and flaxseed ingredient information	PLSR; SVR; multiple linear regression; Principal component regression	The combination of HSI technology and ML demonstrates significant potential for detecting flaxseed quality features.	The models achieve prediction accuracies exceeding 93% for protein and lignan content, and surpassing 85% for oil content.	[72]
	Predict the physiologically active ingredient of <i>Gastrodia elata</i> .	Information on <i>Gastrodia elata</i> ingredients and Vis-NIR spectroscopy data	PLS-DA; KNN; SVM; SVR; CNN	In this study, Vis-NIR spectroscopy combined with ML was applied to predict the contents of physiologically active ingredients in <i>Gastrodia elata</i> .	In the quantitative model, the CNN model exhibited smaller errors compared to the other methods.	[73]
	Non-destructive prediction of isoflavone and starch content in <i>Puerariae Thomsonii Radix</i>	Information on <i>Puerariae Thomsonii Radix</i> ingredients and HSI data	PLSR; SVR; CNN; CatBoost	The potential of combining HSI technology with DL methods to establish an efficient and non-destructive tool for predicting isoflavone and starch content in <i>Puerariae Thomsonii Radix</i> .	The combination of HSI and one-dimensional CNN yielded superior predictive performance and was not significantly dependent on feature wavelength selection.	[74]

ANN: artificial neural network; CARS: competitive adaptive reweighted sampling; CNN: convolutional neural network; DL: deep learning; FT-NIR: Fourier transform near-infrared; GC-IMS: gas chromatography-ion mobility spectrometry; HPLC-MS: high performance liquid chromatography-mass spectrometry; HSI: hyperspectral imaging; LR: logistic regression; PLS-DA: partial least squares discriminant analysis; PLSR: partial least squares regression; QDA: quadratic discriminant analysis; ResNet: residual neural network; RF: random forest; SVR: support vector regression; SVM: support vector machine; Vis-NIR: visible–near-infrared; XGBoost: extreme gradient boosting.

hypergraph convolutional network to predict herb-disease associations, accurately forecasting the most relevant indications for *Chuanxiong* and *Astragalus membranaceus*. Due to the need to investigate the relationship between TCM and targets, Fu et al. [94] proposed a DL-based network pharmacology approach to identify more potential active compounds and targets of licorice and determine the effective therapeutic target of licorice for COVID-19 (Fig. 3B). In addition, HTNet2 successfully predicted the targets of *Artemisia annua* and *Coptis chinensis* using knowledge graph embedding and residual-like graph neural network [95]. Given the potential for pharmacological toxicity, numerous researchers have focused on predicting TCM toxicity and elucidating the mechanisms of action [96,97]. Moreover, considerable attention has been directed towards the pharmacological effects of TCM prescriptions. For example, Liu et al. [98] employed an ML model to screen the feature genes of Tongmai Jiangtang capsules. In the current era of data proliferation, ML algorithms play an indispensable role in elucidating the mechanisms of action of individual TCMs and prescriptions.

3.3. Studying the mechanism of TCM at the compatibility level

Research has shown that drug combinations have a beneficial therapeutic effect on complex diseases, playing the role of “effect-enhancing and toxicity-reducing” [13,99,100]. TCM typically employs herbal formulas containing multiple herbs that systematically prevent and treat diseases through potentially synergistic interactions, aligning with modern combination therapy concepts [101–104]. However, due to the complexity of TCM with multiple components and targets, verifying effective combinations through experiments is exceedingly complicated [105]. Thus, ML methods are essential for predicting effective drug combinations and alleviating clinical burdens [106].

In contemporary drug combination research, DL methods significantly enhance prediction accuracy and processing speed compared to conventional ML, especially network-based methods [12,107,108]. The HPE-GCN predicts the therapeutic effect of tonic formulae based on graph convolutional networks (GCN), and compared to traditional ML methods, it enhances precision and recall on the test set by approximately 10% [109] (Fig. 3C). Wang et al. [107] proposed a DL method for evaluating TCM prescriptions based on Artificial Neural Networks (ANN). Utilizing known safety indicators of TCM prescriptions, this method will help more effectively explore the safety of herbal combinations. In the context of specific diseases, ML can predict effective TCM combinations. Recently, a DL method was proposed by Xue Ren’s team to identify combination drug therapies for pediatric asthma, utilizing datasets derived from the academic and clinical insights of renowned veteran TCM practitioners [12]. Given that TCM relies on multidrug synergy, predicting these synergies will be a primary research direction for understanding TCM-contraindicated combinations [110]. A detailed overview of various ML applications in TCM mechanism studies is presented in Table 2 [78,81,82,93,94,98,107–109,111].

4. Application of ML in TCM diagnosis and treatment process

4.1. Facilitating auxiliary diagnosis

The TCM diagnostic system encompasses inspection, auscultation, inquiry, and palpation, each playing a crucial role in TCM treatment [112]. However, this system is highly dependent on the expertise of practitioners and lacks standardized diagnostic criteria [113]. Advances in natural language processing (NLP) and computer image recognition technologies have facilitated progress in

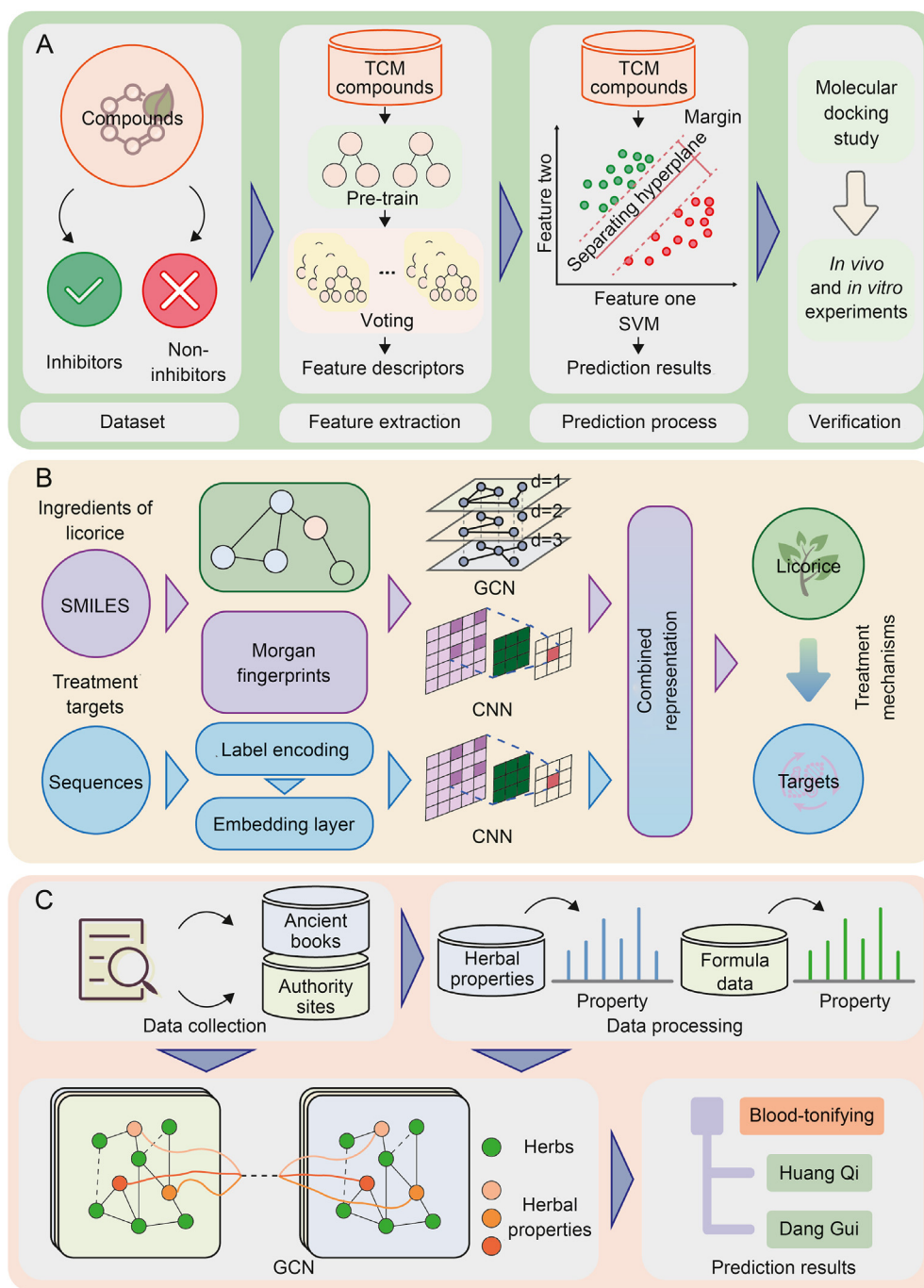


Fig. 3. Application of machine learning method in mechanism study of traditional Chinese medicine (TCM) at different levels. (A) Alzheimer-compound identification based on support vector machine (SVM) model. (B) Convolutional neural networks (CNN) and graph convolutional networks (GCN)-based network pharmacology for exploring the therapeutic mechanism of licorice. (C) Predicting the efficacy of tonic formulae via GCN model integrating traditionally defined herbal properties. SMILES: simplified molecular input line entry system.

overcoming TCM diagnostic challenges [114]. Shi et al. [115] developed four ML diagnostic models that utilize tongue manifestations and pulse conditions to differentiate between disease-related and non-disease-related fatigue. In the field of primary liver cancer (PLC) diagnosis, Ding et al. [116] constructed a neural network syndrome classification and prediction model to explore clinical data and various syndrome types in electronic medical records (EMRs), ultimately transforming these findings into clinical diagnostic tools. Wang et al. [117] introduced the use of tongue images and inquiry information combined with DL for the pre-

endoscopic screening of precancerous lesions in gastric cancer (PLGC), resulting in a 45% improvement in the area under the curve (AUC) compared to traditional PLGC screening methods. Furthermore, numerous studies have investigated the correlation between TCM syndromes and diseases using EMRs with ML techniques [118,119] (Fig. 4A). These approaches have accelerated the application of ML in TCM pattern differentiation tasks. Zhang et al. [120] developed an NLP-based auxiliary diagnostic system capable of automatically diagnosing various diseases and corresponding TCM syndromes using EMRs, achieving an accuracy rate of over 80%.

Table 2

Application of machine learning (ML) in traditional Chinese medicine (TCM) mechanism study at different levels.

Areas of TCM	Aim of study	Data	AI methods	Results	Advantages	Refs.
Exploring the mechanism of TCM in active ingredients level	Screen the PDGFRA inhibitors enhancing radioiodine sensitivity in thyroid cancer.	Dataset of PDGFRA inhibitors	RF; SVM	The study reported a prediction model for PDGFRA inhibitors and successfully identified 4',5,7-trimethoxyflavone as a potential PDGFRA inhibitor.	When applied to the test set, the optimized model achieved a prediction accuracy of 94% and an AUC of 0.95.	[78]
	Predict the hepatoprotective activity of TCM ingredients.	The dataset focused on hepatoprotection induced by TCM ingredients	RF; Naive Bayes; AdaBoost; K-star	12 TCMs were predicted to hold potential hepatoprotective activity, 9 of which have been proven to be beneficial to the liver in previous publications.	Both the training and testing sets indicate that the voting model performs approximately 3% better than the second-ranked K-star model.	[81]
	Predict the activity of potential hepatoprotective ingredients in TCM.	Binding strength scores between compounds and targets, obtained through molecular docking screening	SVR; GBR; decision tree regressor; ANN	Chelerythrine and biochanin A (model prediction components) could improve the viability of APAP-induced injury cells.	The GBR model demonstrated a higher explained variance and did not exhibit overfitting.	[82]
	Screen feature genes in SDECGs.	Expression data of Tongmai Jiangtang capsules core genes in SDECGs	RF; SVM; Generalized linear; XGBoost	ML models identified nine feature genes ranked by their importance scores.	The SVM method had the highest accuracy, with the largest area under the ROC and the lowest residuals and reverse cumulative values.	[98]
Mining the mechanism of TCM in pharmacological effects level	Investigate the correlation between herbs and disease treatment.	The herb-disease network includes components, target proteins, and other relevant elements	CNN	The study explored the potential therapeutic applications of Chuanxiong and <i>Astragalus membranaceus</i> .	Compared to other ML methods, the recall, precision, and F1-score of novel dual-channel hypergraph CNN have been improved by 5%–30% in two datasets.	[93]
	Study on the pharmacological mechanisms of licorice in treatment of COVID-19.	SMILES strings of active compounds, amino acid sequences of targets, and binding affinity values	GCN; CNN	Licorice has been proposed as an effective candidate for the treatment of COVID-19 through PTEN, MAP3K8, Mpro, S-protein, and ACE2.	Compared to other ML methods, DeepDGC achieved higher accuracy with lower mean squared error.	[94]
	Explore pharmacological mechanisms of different TCM formulas for treating CHD.	Potential bioactive compounds and targets associated with classic TCM formulae in the treatment of various types of CHD.	Network clustering; Hierarchical clustering; Random walk with restart	The study developed a rank algorithm, which integrated network topology with biological function, to identify important targets for the treatment of different types of CHD formulas.	The ratio of the literature validation target to the estimated target of the TCM formulas is between 8.7% and 24.7%.	[111]
Studying the mechanism of TCM at the compatibility level	Evaluate the TCM prescriptions officially recommended by China for the treatment of COVID-19.	Dataset of TCM prescription side effects	ANN	The study found that 10 safety indicators with values greater than 0.8 among 38 TCM prescriptions suggest their potential consideration for treating COVID-19.	The side effects and each nodal measure of 12 pharmacology networks derived from the corresponding TCM compound prescription are significantly relevant ($P < 0.05$).	[107]
	Develop a herbal recommendation system based on the pharmacological knowledge of TCM.	Publicly available TCM prescriptions	GNN; GCN; LSTM	This study developed a new method called MGAT, which recommends herbs through pre-defined meta-pathways.	When validating the model's performance in predicting the top ten prescriptions, MGAT consistently outperformed all comparative methods across all metrics, achieving improvements of 3%–8% in precision and recall.	[108]
	Predict tonic formulae based on herbal properties.	Classic tonic formulae from ancient TCM books	GCN; SVM; KNN; LR; Naive Bayes	The evaluated indices Macro-precision, Macro-recall, and Macro-F1 of HPE-GCN on the test set were 87.70%, 84.08%, and 83.51%, respectively.	The evaluated index Macro-precision, Macro-recall, and Macro-F1 of HPE-GCN increased by 7.27%, 7.41%, and 7.30%, respectively from the second-best compared models.	[109]

ACE2: angiotensin-converting enzyme 2; ANN: Artificial Neural Network; APAP: against acetaminophen; AUC: area under the curve; CHD: coronary heart disease; CNN: convolutional neural network; DTR: decision tree regressor; GBR: gradient boost regression; GCN: graph convolutional network; LSTM: long short-term memory networks; Mpro: main protease; PDGFRA: platelet-derived growth factor receptor α ; RF: random forest; ROC: receiver operating characteristic curve; SDECGs: significantly differentially expressed core genes; SVM: support vector machine; SVR: support vector regression; S-protein: spike protein; XGBoost: extreme gradient boosting.

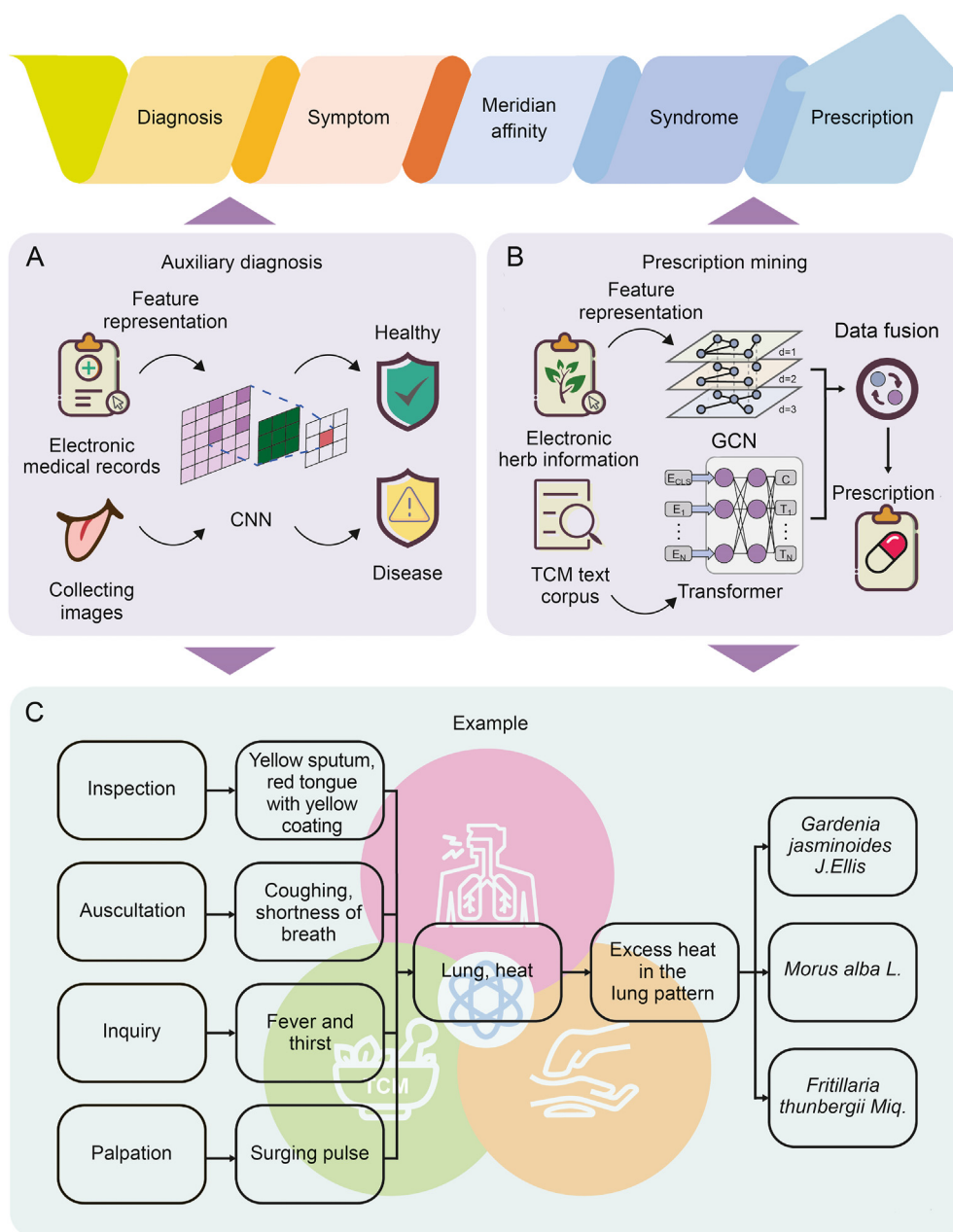


Fig. 4. Application of machine learning (ML) method in diagnosis and treatment process of traditional Chinese medicine (TCM). (A) Application of convolutional neural networks (CNN) in the auxiliary diagnosis. (B) Predicting through the prescription of graph convolutional networks (GCN) or Transformers (based on diagnostic results). (C) Case study of prescription prediction for excess heat in the lung pattern.

While ML cannot completely replace clinical experience, its integration with TCM diagnosis creates new opportunities for aligning TCM diagnosis with international standards.

4.2. Accelerating precision medicine

Since the “Precision Medicine Initiative” was proposed, precision medicine has received growing attention from researchers worldwide [121]. TCM has long advocated personalized medical treatment of precision medicine, emphasizing the combination of “disease-syndrome” and “disease-gene” [122,123]. To achieve precision medicine goals, TCM physicians utilize concepts such as temperament and meridians to describe diseases and prescribe targeted herbal treatments [124]. ML also has applications in TCM

precision medicine, such as personalized prescription recommendations, and pharmacodynamic predictions [125–127] (Figs. 4B and C). ML can provide appropriate prescription recommendations by training and learning to simulate the decision-making processes of TCM experts.

TCM places high value on the preservation of knowledge and experience, making the mining of prescriptions from renowned doctors and ancient Chinese medicine literature a practical approach to achieving precision medicine [128]. Digitizing TCM experiences greatly enhances the reference value of prescription recommendations. Niu et al. [129] developed an algorithm named TCMFP based on herb validity and compatibility validity, successfully predicting prescriptions for treating three common diseases: Alzheimer’s disease, asthma, and atherosclerosis. FordNet, a CNN-

Table 3

Application of machine learning (ML) in traditional Chinese medicine (TCM) diagnosis and treatment process.

Areas of TCM	Aim of study	Data	AI methods	Results	Advantages	Refs.
Facilitate auxiliary diagnosis	Develop a tongue images and inquiry information-based DL tool for the diagnosis for pre-endoscopic screening of PLGC.	Tongue images, inquiry information	Segment Anything models; ResNet; Binary classification networks	Validation analysis revealed that the AUC values of this method were 0.74 for overall PLGC and 0.82 for high-risk PLGC. In the independent external verification, the AUC values were 0.69 for PLGC and 0.76 for high-risk PLGC.	Compared with existing PLGC screening methods, the AUC of this method has increased by 45%, especially in high-risk PLGC screening where it has increased by 52%.	[117]
	Explore TCM treatment based on pattern differentiation in rare diseases.	TCM electronic medical records	Transformer; DT; SVM; LSTM; XGBoost	The model called TLDA achieves the highest precision on datasets, outperforming all other algorithms.	The TLDA model can be extended to other medical tasks, particularly when electronic medical records for diagnosis are limited in samples.	[119]
	Develop a wrist pulse diagnosis-based ML tool for diagnosing various common diseases.	Wrist pulse signal data	SVM; KNN; RF; CNN; LSTM; GCN; Linear discriminant analysis	The model called GBMMF can integrate the correlation of different features from multichannel wrist pulse samples.	Compared to other ML methods, the accuracy of GBMMF improved by 10%–20% for diagnosing three diseases.	[135]
Accelerate precision medicine	Recommend TCM classic tonic formulae by symptom analysis.	Classic tonic formulae from ancient TCM books	GCN; SVM; LR	The model called MGCN simulates n-ary relationships among symptoms, state elements, syndrome types, and herbs.	When the algorithm was configured to predict the number of herbs as 5, MGCN outperformed the best baseline method (SVM) by 4.51%, 6.45%, and 5.31% in precision, recall, and F1-score, respectively.	[125]
	Establish a model for recommending TCM formula by electronic medical records.	TCM electronic medical records	CNN; RF; SVM; LR	The article proposed an intelligent formula recommendation system based on the DL model, named FordNet, which integrates phenotypic and molecular information.	FordNet performs significantly better than baseline methods (the hit ratio of the top 10 improved by 46.9% compared with the best baseline RF method).	[130]
	Recommend TCM formula for the treatment of coronary heart disease with depression.	High-quality TCM formula dataset (e.g., HERB, DisGeNet)	Transformer	The model predicts that Huangjin Shuang Shen Jiawei has a protective effect on clinical symptoms of depression in mouse models of TCM with depression.	The model achieved a higher average semantic consistency score of 0.781 (Seq2Seq: 0.708, BART: 0.729), and lower average semantic ambiguity and semantic deficiency scores of 0.189 and 0.240, respectively (Seq2Seq: 0.267 and 0.293, BART: 0.236 and 0.269).	[136]

AUC: area under the curve; CNN: convolutional neural network; DL: deep learning; DT: decision tree; GCN: graph convolutional network; KNN: k-nearest node neighbors; LR: logistic regression; LSTM: long short-term memory networks; MGCN: multigraph convolutional network; PLGC: precancerous lesions of gastric cancer; RF: random forest; TLDA: Transfer Learning based Dual-Augmentation; XGBoost: extreme gradient boosting.

based prescription recommendation model, has achieved remarkable results in integrating TCM phenotype (i.e., symptoms and diseases) and molecule [130]. Moreover, the multigraph convolutional network (MGCN) and transfer learning model have performed well in mining prescriptions [125,131].

ML exhibits extraordinary potential not only in prescription recommendations but also in other aspects of precision medicine, such as dose–effect relationship analysis of TCM [132], predicting the severity of insomnia based on TCM constitution [133], and developing prognostic models [134]. However, when applying ML to the complexities of individual differences and comprehensive judgments in precision medicine, integrating the expertise and insights of TCM practitioners remains essential. A detailed explanation of the different applications of ML in TCM diagnosis and treatment process studies is shown in Table 3 [117,119,125,130,135,136].

5. The application of ML in other directions of TCM

The responsive cell types and intestinal microbiota play crucial roles in the regulatory mechanisms of TCM, providing

valuable insights for developing diverse treatment strategies [137,138]. Li et al. [138] employed target-perturbed gene regulatory networks to model and score computer-simulated drug perturbation data from untreated single-cell RNA sequencing, and accurately uncover the macrophage subpopulation responsive to tanshinone IIA and its potential targets in myocardial infarction, with subsequent experimental validation. In research on intestinal microbiota, Zeng et al. [139] used an RF model to compare the abundance of *Bifidobacterium longum* between patients with spleen qi deficiency and those with spleen qi deficiency with stomach heat, and analyzed its relationship with plasma interferon- α (IFN- α) concentration, revealing different microbial community characteristics associated with different TCM syndrome types. Addressing cognitive impairment in diabetes, Zhang et al. [140] developed a Graph Neural Network (GNN) model to elucidate how herbal medicine interventions impact gut microbiota. Additionally, ML extensively addresses various aspects of TCM, including the prediction of NP structures [141], acupoint recognition [142], and TCM dialogue systems [143]. These advancements have

accelerated researchers' understanding and application of TCM practices.

6. Conclusion and prospect

As TCM continues to globalize, aligning its standards with international benchmarks becomes essential [144]. ML methods provide diverse solutions to the current challenges of TCM. This article systematically explores the application of ML algorithms in quality control, mechanism research, diagnosis, and treatment processes of TCM. Traditional ML algorithms such as SVM, random forest (RF), Decision Tree (DT), and MLP are commonly employed in TCM. As datasets expand and ML algorithms advance, DL techniques such as CNN, GNN, and ANN exhibit superior predictive performance compared to traditional ML algorithms [145]. In recent years, large language models have demonstrated unique value in the healthcare field by further compensating for the speed limitations of ML when handling large volumes of text data [146].

Despite these advancements, the standardization of ML in TCM faces numerous challenges [147]. Incomplete pharmacological evaluation systems, inconsistent TCM databases, and unstable ML prediction results are all undermining the accuracy and reliability of ML applications. To overcome these obstacles, database development should emphasize the standardization of information extraction and conversion, to standardized and high-quality TCM database. Research on TCM related algorithm development should focus on the correctness and stability of the methods, and corresponding case results verification is also essential. Additionally, in mechanism research, it is essential to validate the predicted results with corresponding *in vitro* and *in vivo* experimental verification.

Therefore, it is essential to continually adopt cutting-edge algorithmic frameworks and integrate advanced experimental methods to advance the application of ML in TCM. For instance, the development of HSI, in conjunction with other sensor technologies, has significantly enhanced the speed of herbal quality detection, enabling real-time and portable monitoring [148]. The integration of advanced ML techniques into TCM network pharmacology analysis has yielded numerous new targets and unique compounds [149]. In addition, multi-source data fusion and collaborative DL from patient diagnosis can improve the accuracy of data analysis while ensuring excellent patient human-computer interaction experience [150]. In summary, we believe that current ML algorithms are a promising technique for TCM data analysis and mining. The continuous development of AI assisted TCM indicates broad prospects in this field.

CRediT authorship contribution statement

Yanfeng Hong: Writing – original draft, Visualization, Investigation, Data curation, Conceptualization. **Sisi Zhu:** Investigation, Data curation. **Yuhong Liu:** Data curation. **Chao Tian:** Data curation. **Hongquan Xu:** Investigation. **Gongxing Chen:** Investigation. **Lin Tao:** Writing – review & editing, Supervision, Resources, Methodology, Conceptualization. **Tian Xie:** Supervision, Resources, Funding acquisition.

Declaration of competing interest

The authors declare that there are no conflicts of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jpha.2024.101157>.

References

- [1] J. Tang, B. Liu, K. Ma, Traditional Chinese medicine, *Lancet* 372 (2008) 1938–1940.
- [2] Y. Li, X. Sui, Z. Su, et al., Meta-analysis of paclitaxel-based chemotherapy combined with traditional Chinese medicines for gastric cancer treatment, *Front. Pharmacol.* 11 (2020), 132.
- [3] X. Chi, S. Wang, Z. Baloch, et al., Research progress on classical traditional Chinese medicine formula Lily Bulb and Rehmannia Decoction in the treatment of depression, *Biomed. Pharmacother.* 112 (2019), 108616.
- [4] M. Lyu, G. Fan, G. Xiao, et al., Traditional Chinese medicine in COVID-19, *Acta Pharm. Sin.* B 11 (2021) 3337–3363.
- [5] H. Xu, W. Zhang, Y. Zhou, et al., Systematic description of the content variation of natural products (NPs): To prompt the yield of high-value NPs and the discovery of new therapeutics, *J. Chem. Inf. Model.* 63 (2023) 1615–1625.
- [6] L. Tao, F. Zhu, C. Qin, et al., Nature's contribution to today's pharmacopeia, *Nat. Biotechnol.* 32 (2014) 979–980.
- [7] Y. Tu, Artemisinin—a gift from traditional Chinese medicine to the world (Nobel lecture), *Angew. Chem. Int. Ed.* 55 (2016) 10210–10226.
- [8] B. Zhai, N. Zhang, X. Han, et al., Molecular targets of β -elemene, a herbal extract used in traditional Chinese medicine, and its potential role in cancer therapy: A review, *Biomed. Pharmacother.* 114 (2019), 108812.
- [9] C. Liu, S. Sun, Q. Feng, et al., Arsenene nanodots with selective killing effects and their low-dose combination with β -elemene for cancer therapy, *Adv. Mater. Deerfield Beach Fla* 33 (2021), e2102054.
- [10] L. Ji, T. Song, C. Ge, et al., Identification of bioactive compounds and potential mechanisms of scutellariae Radix-coptidis rhizoma in the treatment of atherosclerosis by integrating network pharmacology and experimental validation, *Biomed. Pharmacother.* 165 (2023), 115210.
- [11] P. Wu, X. Dong, G. Song, et al., Bioactivity-guided discovery of quality control markers in rhizomes of *Curcuma wenyujin* based on spectrum-effect relationship against human lung cancer cells, *Phytomedicine* 86 (2021), 153559.
- [12] X. Ren, Y. Guo, H. Wang, et al., The intelligent experience inheritance system for Traditional Chinese Medicine, *J. Evid. Base Med.* 16 (2023) 91–100.
- [13] D. Li, J. Hu, L. Zhang, et al., 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 (2022), 175260.
- [14] J. Miao, Y. Huang, Z. Wang, et al., Image recognition of traditional Chinese medicine based on deep learning, *Front. Bioeng. Biotechnol.* 11 (2023), 1199803.
- [15] X. Tan, X. Ma, Y. Dai, et al., A large-scale transcriptional analysis reveals herb-derived ginsenoside F₂ suppressing hepatocellular carcinoma via inhibiting STAT3, *Phytomedicine* 120 (2023), 155031.
- [16] A.V. Sadybekov, V. Katritch, Computational approaches streamlining drug discovery, *Nature* 616 (2023) 673–685.
- [17] S. Tian, J. Wang, Y. Li, et al., The application of *in silico* drug-likeness predictions in pharmaceutical research, *Adv. Drug Deliv. Rev.* 86 (2015) 2–10.
- [18] S. Fang, L. Dong, L. Liu, et al., Herb: A high-throughput experiment- and reference-guided database of traditional Chinese medicine, *Nucleic Acids Res.* 49 (2021) D1197–D1206.
- [19] Y. Gao, H. Zhang, F. Lirussi, et al., Dual inhibitors of histone deacetylases and other cancer-related targets: A pharmacological perspective, *Biochem. Pharmacol.* 182 (2020), 114224.
- [20] J. Nielsen, Systems biology of metabolism, *Annu. Rev. Biochem.* 86 (2017) 245–275.
- [21] A.L. Hopkins, Network pharmacology: The next paradigm in drug discovery, *Nat. Chem. Biol.* 4 (2008) 682–690.
- [22] S. Lv, Q. Wang, X. Zhang, et al., Mechanisms of multi-omics and network pharmacology to explain traditional Chinese medicine for vascular cognitive impairment: A narrative review, *Phytomedicine* 123 (2024), 155231.
- [23] C. Nogales, Z.M. Mamdouh, M. List, et al., Network pharmacology: Curing causal mechanisms instead of treating symptoms, *Trends Pharmacol. Sci.* 43 (2022) 136–150.
- [24] D. Yan, G. Zheng, C. Wang, et al., Hit 2.0: An enhanced platform for herbal ingredients' targets, *Nucleic Acids Res.* 50 (2022) D1238–D1243.
- [25] Y. Wu, F. Zhang, K. Yang, et al., SymMap: An integrative database of traditional Chinese medicine enhanced by symptom mapping, *Nucleic Acids Res.* 47 (2019) D1110–D1117.
- [26] J. Zhi, L. Yin, Z. Zhang, et al., Network pharmacology-based analysis of Jin-Si-Wei on the treatment of Alzheimer's disease, *J. Ethnopharmacol.* 319 (2024), 117291.

- [27] Y. Wang, T. Zhao, C. Huang, et al., Effect and mechanism of Banxia Xiexin decoction in colorectal cancer: A network pharmacology approach, *Phyto-medicine* 123 (2024), 155174.
- [28] H. Xu, S. Li, J. Liu, et al., Bioactive compounds from Huashi Baidu decoction possess both antiviral and anti-inflammatory effects against COVID-19, *Proc. Natl. Acad. Sci. U.S.A.* 120 (2023), e2301775120.
- [29] L. Zhao, H. Zhang, N. Li, et al., Network pharmacology, a promising approach to reveal the pharmacology mechanism of Chinese medicine formula, *J. Ethnopharmacol.* 309 (2023), 116306.
- [30] F. Aydin Temel, O. Cagcag Yolcu, N.G. Turan, Artificial intelligence and machine learning approaches in composting process: A review, *Bioresour. Technol.* 370 (2023), 128539.
- [31] E.J. Topol, High-performance medicine: The convergence of human and artificial intelligence, *Nat. Med.* 25 (2019) 44–56.
- [32] H. Wang, T. Fu, Y. Du, et al., Scientific discovery in the age of artificial intelligence, *Nature* 620 (2023) 47–60.
- [33] S. Li, B. Zhang, Traditional Chinese medicine network pharmacology: Theory, methodology and application, *Chin. J. Nat. Med.* 11 (2013) 110–120.
- [34] E. Zhou, Q. Shen, Y. Hou, Integrating artificial intelligence into the modernization of traditional Chinese medicine industry: A review, *Front. Pharmacol.* 15 (2024), 1181183.
- [35] L. Li, L. Wang, W. Fan, et al., The application of fermentation technology in traditional Chinese medicine: A review, *Am. J. Chin. Med.* 48 (2020) 899–921.
- [36] Y. Jia, M. Sun, Y. Shi, et al., A comparative study of aged and contemporary Chinese herbal materials by using delayed luminescence technique, *Chin. Med.* 15 (2020), 6.
- [37] Y. Pan, H. Zhang, Y. Chen, et al., 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.* (2023) 1–15.
- [38] C. Meng, P. Wang, Z. Hao, et al., Ecological and health risk assessment of heavy metals in soil and Chinese herbal medicines, *Environ. Geochem. Health* 44 (2022) 817–828.
- [39] W. Zhang, C. Zhang, L. Cao, et al., Application of digital-intelligence technology in the processing of Chinese materia Medica, *Front. Pharmacol.* 14 (2023), 1208055.
- [40] Y. Shi, Q. Huang, S. Zhu, et al., Chemical profiling of fritillariae thunbergii miq prepared by different processing methods reveals two new quality markers: Zhebeinoside and imperialine-3- β -D-glucoside, *J. Ethnopharmacol.* 283 (2022), 114670.
- [41] S. Wang, Y. Yu, M. Cui, et al., Seed oil quality and cultivation of *Sambucus williamsii* hance as a new oil crop, *Front. Nutr.* 8 (2021), 796175.
- [42] Y. Xu, W. Wu, Y. Chen, et al., Hyperspectral imaging with machine learning for non-destructive classification of *Astragalus membranaceus* var. *mongolicus*, *Astragalus membranaceus*, and similar seeds, *Front. Plant Sci.* 13 (2022), 1031849.
- [43] R. Bai, J. Zhou, S. Wang, et al., Identification and classification of Coix seed storage years based on hyperspectral imaging technology combined with deep learning, *Foods* 13 (2024), 498.
- [44] Y. Tao, J. Bao, Q. Liu, et al., Application of deep-learning algorithm driven intelligent Raman spectroscopy methodology to quality control in the manufacturing process of guanxinling tablets, *Molecules* 27 (2022), 6969.
- [45] X. Yan, S. Zhang, H. Fu, et al., Combining convolutional neural networks and on-line Raman spectroscopy for monitoring the Cornu Caprae Hircus hydrolysis process, *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* 226 (2020), 117589.
- [46] Y. Shi, T. He, J. Zhong, et al., Classification and rapid non-destructive quality evaluation of different processed products of *Cyperus rotundus* based on near-infrared spectroscopy combined with deep learning, *Talanta* 268 (2024), 125266.
- [47] S. Liu, H. Dong, M. Zhang, et al., Identification of different degrees of processed ginger using GC-IMS combined with machine learning, *J. Pharm. Anal.* 14 (2024) 149–151.
- [48] W. Xu, P. Zhu, T. Xin, et al., Droplet digital PCR for the identification of plant-derived adulterants in highly processed products, *Phytomedicine* 105 (2022), 154376.
- [49] X. Zhang, L. Wang, R. Li, et al., Identification strategy of *Fructus Gardeniae* and its adulterant based on UHPLC/Q-orbitrap-MS and UHPLC-QTRAP-MS/MS combined with PLS regression model, *Talanta* 267 (2024), 125136.
- [50] Y. Zhang, Y. Wang, Recent trends of machine learning applied to multi-source data of medicinal plants, *J. Pharm. Anal.* 13 (2023) 1388–1407.
- [51] Z. Jiang, A. Lv, L. Zhong, et al., Rapid prediction of adulteration content in *Atractylodes rhizoma* based on data and image features fusions from near-infrared spectroscopy and hyperspectral imaging techniques, *Foods* 12 (2023), 2904.
- [52] A.M. Husaini, S.A.U. Haq, A. Shabir, et al., The menace of saffron adulteration: Low-cost rapid identification of fake look-alike saffron using foldscope and machine learning technology, *Front. Plant Sci.* 13 (2022), 945291.
- [53] A. Amirvaresi, N. Nikounezhad, M. Amirahmadi, et al., Comparison of near-infrared (NIR) and mid-infrared (MIR) spectroscopy based on chemometrics for saffron authentication and adulteration detection, *Food Chem.* 344 (2021), 128647.
- [54] L. Zhang, C. Zhang, W. Li, et al., Rapid identification of auramine O dyeing adulteration in *Dendrobium officinale*, Saffron and Curcuma by SERS Raman spectroscopy combined with SSA-BP neural networks model, *Foods* 12 (2023), 4124.
- [55] B. Xing, S. Li, J. Yang, et al., Phytochemistry, pharmacology, and potential clinical applications of saffron: A review, *J. Ethnopharmacol.* 281 (2021), 114555.
- [56] O.A. Malik, N. Ismail, B.R. Hussein, et al., Automated real-time identification of medicinal plants species in natural environment using deep learning models—a case study from Borneo Region, *Plants* 11 (2022), 1952.
- [57] S. Zhu, Q. Liu, S. Qiu, et al., DNA barcoding: An efficient technology to authenticate plant species of traditional Chinese medicine and recent advances, *Chin. Med.* 17 (2022), 112.
- [58] A. Ren, T. Wu, Y. Wang, et al., Integrating animal experiments, mass spectrometry and network-based approach to reveal the sleep-improving effects of *Ziziphi Spinosa* Semen and γ -aminobutyric acid mixture, *Chin. Med.* 18 (2023), 99.
- [59] M. Li, Y. Shi, J. Zhang, et al., Rapid evaluation of *Ziziphi Spinosa* Semen and its adulterants based on the combination of FT-NIR and multivariate algorithms, *Food Chem.* 20 (2023), 101022.
- [60] X. Li, W. Kong, X. Liu, et al., Application of laser-induced breakdown spectroscopy coupled with spectral matrix and convolutional neural network for identifying geographical origins of *Gentiana rigescens* franch, *Front. Artif. Intell.* 4 (2021), 735533.
- [61] R. Ding, L. Yu, C. Wang, et al., Quality assessment of traditional Chinese medicine based on data fusion combined with machine learning: A review, *Crit. Rev. Anal. Chem.* 54 (2023) 2618–2635.
- [62] Z. Liu, T. Shen, J. Zhang, et al., A novel multi-preprocessing integration method for the qualitative and quantitative assessment of wild medicinal plants: *Gentiana rigescens* as an example, *Front. Plant Sci.* 12 (2021), 759248.
- [63] Z. Cai, Z. Huang, M. He, et al., Identification of geographical origins of *Radix Paeoniae Alba* using hyperspectral imaging with deep learning-based fusion approaches, *Food Chem.* 422 (2023), 136169.
- [64] S. Chen, X. Li, J. Wu, et al., *Plumula nelumbinis*: A review of traditional uses, phytochemistry, pharmacology, pharmacokinetics and safety, *J. Ethnopharmacol.* 266 (2021), 113429.
- [65] G. Li, D. Nikolic, R.B. van Breemen, Identification and chemical standardization of licorice raw materials and dietary supplements using UHPLC-MS/MS, *J. Agric. Food Chem.* 64 (2016) 8062–8070.
- [66] S.H. Nile, J. Su, D. Wu, et al., *Fritillaria thunbergii* miq. (Zhe Beimu): A review on its traditional uses, phytochemical profile and pharmacological properties, *Food Chem. Toxicol.* 153 (2021), 112289.
- [67] M.H. Kabir, M.L. Guindo, R. Chen, et al., Deep learning combined with hyperspectral imaging technology for variety discrimination of *Fritillaria thunbergii*, *Molecules* 27 (2022), 6042.
- [68] J.P. Cruz-Tirado, Y. Lima Brasil, A. Freitas Lima, et al., Rapid and non-destructive cinnamon authentication by NIR-hyperspectral imaging and classification chemometrics tools, *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* 289 (2023), 122226.
- [69] J. Jing, C. Zhu, R. Gong, et al., Research progress on the active ingredients of traditional Chinese medicine in the intervention of atherosclerosis: A promising natural immunotherapeutic adjuvant, *Biomed. Pharmacother.* 159 (2023), 114201.
- [70] J. Wan, Z. Zhang, C. Wu, et al., Astragaloside IV derivative HHQ16 ameliorates infarction-induced hypertrophy and heart failure through degradation of lncRNA4012/9456, *Signal Transduct. Targeted Ther.* 8 (2023), 414.
- [71] H. Xu, H. Zhao, C. Ding, et al., Celastrol suppresses colorectal cancer via covalent targeting peroxiredoxin 1, *Signal Transduct. Targeted Ther.* 8 (2023), 51.
- [72] D. Zhu, J. Han, C. Liu, et al., Modeling of flaxseed protein, oil content, linoleic acid, and lignan content prediction based on hyperspectral imaging, *Front. Plant Sci.* 15 (2024), 1344143.
- [73] J. Ma, X. Zhou, B. Xie, et al., Application for identifying the origin and predicting the physiologically active ingredient contents of *Gastrodia elata* blume using visible-near-infrared spectroscopy combined with machine learning, *Foods* 12 (2023), 4061.
- [74] H. Hu, T. Wang, Y. Wei, et al., Non-destructive prediction of isoflavone and starch by hyperspectral imaging and deep learning in *Puerariae Thomsonii* Radix, *Front. Plant Sci.* 14 (2023), 1271320.
- [75] Y. Zhu, Z. Ouyang, H. Du, et al., New opportunities and challenges of natural products research: When target identification meets single-cell multiomics, *Acta Pharm. Sin. B* 12 (2022) 4011–4039.
- [76] W. Chen, C. Liu, X. Ji, et al., Stanene-based nanosheets for β -elemene delivery and ultrasound-mediated combination cancer therapy, *Angew. Chem. Int. Ed.* 60 (2021) 7155–7164.
- [77] Y. Lin, Y. Zhang, D. Wang, et al., Computer especially AI-assisted drug virtual screening and design in traditional Chinese medicine, *Phytomedicine* 107 (2022), 154481.
- [78] X. Yu, X. Zhu, L. Zhang, et al., In silico screening and validation of PDGFRA inhibitors enhancing radioiodine sensitivity in thyroid cancer, *Front. Pharmacol.* 13 (2022), 883581.
- [79] S. Zhang, K. Yang, Z. Liu, et al., DrugAI: A multi-view deep learning model for predicting drug-target activating/inhibiting mechanisms, *Briefings Bioinf.* 24 (2023), bbac526.
- [80] B. Yang, W. Bao, S. Hong, Alzheimer-compound identification based on data fusion and forgeNet_SVM, *Front. Aging Neurosci.* 14 (2022), 931729.
- [81] S. He, Y. Yi, D. Hou, et al., Identification of hepatoprotective traditional Chinese medicines based on the structure-activity relationship, molecular network, and machine learning techniques, *Front. Pharmacol.* 13 (2022), 969979.

- [82] K. Liu, X. Chen, Y. Ren, et al., Multi-target-based polypharmacology prediction (mTPP): An approach using virtual screening and machine learning for multi-target drug discovery, *Chem. Biol. Interact.* 368 (2022), 110239.
- [83] J.Y. Fu, C.X. Jiang, M.Y. Wu, et al., Theabrownin induces cell apoptosis and cell cycle arrest of oligodendroglioma and astrocytoma in different pathways, *Front. Pharmacol.* 12 (2021), 664003.
- [84] C. Geng, Z. Wang, Y. Tang, Machine learning in Alzheimer's disease drug discovery and target identification, *Ageing Res. Rev.* 93 (2024), 102172.
- [85] X.X. Li, J. Yin, J. Tang, et al., Determining the balance between drug efficacy and safety by the network and biological system profile of its therapeutic target, *Front. Pharmacol.* 9 (2018), 1245.
- [86] X. Gan, Z. Shu, X. Wang, et al., Network medicine framework reveals generic herb-symptom effectiveness of traditional Chinese medicine, *Sci. Adv.* 9 (2023), eadh0215.
- [87] R. Iyengar, S. Zhao, S. Chung, et al., Merging systems biology with pharmacodynamics, *Sci. Transl. Med.* 4 (2012), 126ps7.
- [88] F. Li, J. Yin, M. Lu, et al., ConSIG: Consistent discovery of molecular signature from OMIC data, *Briefings Bioinf.* 23 (2022), bbac253.
- [89] S. Hou, P. Zhang, K. Yang, et al., Decoding multilevel relationships with the human tissue-cell-molecule network, *Briefings Bioinf.* 23 (2022), bbac170.
- [90] L. Zheng, S. Shi, X. Sun, et al., MoDAFold: A strategy for predicting the structure of missense mutant protein based on AlphaFold2 and molecular dynamics, *Briefings Bioinf.* 25 (2024), bbae006.
- [91] Y. Lei, S. Li, Z. Liu, et al., A deep-learning framework for multi-level peptide-protein interaction prediction, *Nat. Commun.* 12 (2021), 5465.
- [92] A. Rajkomar, J. Dean, I. Kohane, Machine learning in medicine, *N. Engl. J. Med.* 380 (2019) 1347–1358.
- [93] L. Hu, M. Zhang, P. Hu, et al., Dual-channel hypergraph convolutional network for predicting herb-disease associations, *Briefings Bioinf.* 25 (2024), bbae067.
- [94] Y. Fu, Y. Fang, S. Gong, et al., Deep learning-based network pharmacology for exploring the mechanism of licorice for the treatment of COVID-19, *Sci. Rep.* 13 (2023), 5844.
- [95] P. Duan, K. Yang, X. Su, et al., HTINet2: Herb-target prediction via knowledge graph embedding and residual-like graph neural network, *Briefings Bioinf.* 25 (2024), bbae414.
- [96] Q. Wu, C. Cai, P. Guo, et al., In silico identification and mechanism exploration of hepatotoxic ingredients in traditional Chinese medicine, *Front. Pharmacol.* 10 (2019), 458.
- [97] Z. Chen, M. Zhao, L. You, et al., Developing an artificial intelligence method for screening hepatotoxic compounds in traditional Chinese medicine and Western medicine combination, *Chin. Med.* 17 (2022), 58.
- [98] Y. Liu, X. Cui, X. Zhang, et al., Exploring the potential mechanisms of Tongmai Jiangtang capsules in treating diabetic nephropathy through multi-dimensional data, *Front. Endocrinol.* 14 (2023), 1172226.
- [99] C. Knox, M. Wilson, C.M. Klinger, et al., DrugBank 6.0: The DrugBank knowledgebase for 2024, *Nucleic Acids Res.* 52 (2024) D1265–D1275.
- [100] N.P. Tatonetti, P.P. Ye, R. Daneshjou, et al., Data-driven prediction of drug effects and interactions, *Sci. Transl. Med.* 4 (2012), 125ra31.
- [101] Y. Jia, C. Yun, E. Park, et al., Overcoming EGFR(T790M) and EGFR(C797S) resistance with mutant-selective allosteric inhibitors, *Nature* 534 (2016) 129–132.
- [102] W. Zhao, X. Zheng, P.Y. Tang, et al., Advances of antitumor drug discovery in traditional Chinese medicine and natural active products by using multi-active components combination, *Med. Res. Rev.* 43 (2023) 1778–1808.
- [103] Y. Hong, H. Xu, Y. Liu, et al., Ddid: A comprehensive resource for visualization and analysis of diet-drug interactions, *Briefings Bioinf.* 25 (2024), bbae212.
- [104] J. Lou, L. Zhao, Z. Huang, et al., Ginkgetin derived from Ginkgo biloba leaves enhances the therapeutic effect of cisplatin via ferroptosis-mediated disruption of the Nrf2/HO-1 axis in EGFR wild-type non-small-cell lung cancer, *Phytomedicine* 80 (2021), 153370.
- [105] G. Shtar, L. Azulay, O. Nizri, et al., Cdcdb: A large and continuously updated drug combination database, *Sci. Data* 9 (2022), 263.
- [106] B.M. Kuenzi, J. Park, S.H. Fong, et al., Predicting drug response and synergy using a deep learning model of human cancer cells, *Cancer Cell* 38 (2020) 672–684.e6.
- [107] Z. Wang, L. Li, M. Song, et al., Evaluating the traditional Chinese medicine (TCM) officially recommended in China for COVID-19 using ontology-based side-effect prediction framework (OSPF) and deep learning, *J. Ethnopharmacol.* 272 (2021), 113957.
- [108] Y. Jin, W. Ji, Y. Shi, et al., Meta-path guided graph attention network for explainable herb recommendation, *Health Inf. Sci. Syst.* 11 (2023), 5.
- [109] J. Liu, Q. Huang, X. Yang, et al., HPE-GCN: Predicting efficacy of tonic formulae via graph convolutional networks integrating traditionally defined herbal properties, *Methods* 204 (2022) 101–109.
- [110] T. Xue, Synergy in traditional Chinese medicine, *Lancet Oncol.* 17 (2016), e39.
- [111] J. Yang, S. Tian, J. Zhao, et al., Exploring the mechanism of TCM formulae in the treatment of different types of coronary heart disease by network pharmacology and machining learning, *Pharmacol. Res.* 159 (2020), 105034.
- [112] D. Tian, W. Chen, D. Xu, et al., A review of traditional Chinese medicine diagnosis using machine learning: Inspection, auscultation-olfaction, inquiry, and palpation, *Comput. Biol. Med.* 170 (2024), 108074.
- [113] Q. Zhang, J. Zhou, B. Zhang, Computational traditional Chinese medicine diagnosis: A literature survey, *Comput. Biol. Med.* 133 (2021), 104358.
- [114] T. Burki, Natural language processing and detecting delirium, *Lancet Respir. Med.* 10 (2022), 639.
- [115] Y. Shi, X. Yao, J. Xu, et al., A new approach of fatigue classification based on data of tongue and pulse with machine learning, *Front. Physiol.* 12 (2021), 708742.
- [116] L. Ding, X. Zhang, D. Wu, et al., Application of an extreme learning machine network with particle swarm optimization in syndrome classification of primary liver cancer, *J. Integr. Med.* 19 (2021) 395–407.
- [117] L. Wang, Q. Zhang, P. Zhang, et al., Development of an artificial intelligent model for pre-endoscopic screening of precancerous lesions in gastric cancer, *Chin. Med.* 19 (2024), 90.
- [118] L. Gong, J. Jiang, S. Chen, et al., A syndrome differentiation model of TCM based on multi-label deep forest using biomedical text mining, *Front. Genet.* 14 (2023), 1272016.
- [119] X. Li, K. Chen, J. Yang, et al., Tlda: A transfer learning based dual-augmentation strategy for traditional Chinese medicine syndrome differentiation in rare disease, *Comput. Biol. Med.* 169 (2024), 107808.
- [120] H. Zhang, W. Ni, J. Li, et al., Artificial intelligence-based traditional Chinese medicine assistive diagnostic system: Validation study, *JMIR Med. Inform.* 8 (2020), e17608.
- [121] E. Wong, N. Bertin, M. Hebrard, et al., The Singapore national precision medicine strategy, *Nat. Genet.* 55 (2023) 178–186.
- [122] P. Zhang, D. Zhang, W. Zhou, et al., Network pharmacology: Towards the artificial intelligence-based precision traditional Chinese medicine, *Briefings Bioinf.* 25 (2023), bbad518.
- [123] W. Wang, T. Zhang, Integration of traditional Chinese medicine and Western medicine in the era of precision medicine, *J. Integr. Med.* 15 (2017) 1–7.
- [124] S. Zhang, X. Zhang, J. Du, et al., Multi-target meridians classification based on the topological structure of anti-cancer phytochemicals using deep learning, *J. Ethnopharmacol.* 319 (2024), 117244.
- [125] W. Zhao, W. Lu, Z. Li, et al., TCM herbal prescription recommendation model based on multi-graph convolutional network, *J. Ethnopharmacol.* 297 (2022), 115109.
- [126] Y. Tang, Z. Li, D. Yang, et al., Research of insomnia on traditional Chinese medicine diagnosis and treatment based on machine learning, *Chin. Med.* 16 (2021), 2.
- [127] H. Chen, J. Chen, J. Li, et al., 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 (2019) 1605–1623.
- [128] M. Chung, L. Su, C. Chen, et al., AI-assisted literature exploration of innovative Chinese medicine formulas, *Front. Pharmacol.* 15 (2024), 1347882.
- [129] Q. Niu, H. Li, L. Tong, et al., Tcmfp: A novel herbal formula prediction method based on network target's score integrated with semi-supervised learning genetic algorithms, *Briefings Bioinf.* 24 (2023), bbad102.
- [130] W. Zhou, K. Yang, J. Zeng, et al., FordNet: Recommending traditional Chinese medicine formula via deep neural network integrating phenotype and molecule, *Pharmacol. Res.* 173 (2021), 105752.
- [131] Z. Liu, C. Luo, D. Fu, et al., A novel transfer learning model for traditional herbal medicine prescription generation from unstructured resources and knowledge, *Artif. Intell. Med.* 124 (2022), 102232.
- [132] W. Xiong, Y. Zhu, Q. Zeng, et al., Dose-effect relationship analysis of TCM based on deep Boltzmann machine and partial least squares, *Math. Biosci. Eng.* 20 (2023) 14395–14413.
- [133] S. Li, P. Zhu, G. Cai, et al., Application of machine learning models in predicting insomnia severity: An integrative approach with constitution of traditional Chinese medicine, *Front. Med.* 10 (2023), 1292761.
- [134] M. Tang, L. Gao, B. He, et al., Machine learning based prognostic model of Chinese medicine affecting the recurrence and metastasis of I-III stage colorectal cancer: A retrospective study in China, *Front. Oncol.* 12 (2022), 1044344.
- [135] Q. Zhang, J. Zhou, B. Zhang, Graph based multichannel feature fusion for wrist pulse diagnosis, *IEEE J. Biomed. Health Inform.* 25 (2021) 3732–3743.
- [136] Y. Qian, X. Wang, L. Cai, et al., Model informed precision medicine of Chinese herbal medicines formulas-a multi-scale mechanistic intelligent model, *J. Pharm. Anal.* 14 (2024), 100914.
- [137] L. Zhu, S. Li, W. Zheng, et al., Targeted modulation of gut microbiota by traditional Chinese medicine and natural products for liver disease therapy, *Front. Immunol.* 14 (2023), 1086078.
- [138] C. Li, X. Shao, S. Zhang, et al., scRank infers drug-responsive cell types from untreated scRNA-seq data using a target-perturbed gene regulatory network, *Cell Rep. Med.* 5 (2024), 101568.
- [139] H. Zeng, J. Xu, L. Zheng, et al., Traditional Chinese herbal formulas modulate gut microbiome and improve insomnia in patients with distinct syndrome types: Insights from an interventional clinical study, *Front. Cell. Infect. Microbiol.* 14 (2024), 1395267.
- [140] X. Zhang, J. Xie, X. You, et al., Risk factors and drug discovery for cognitive impairment in type 2 diabetes mellitus using artificial intelligence interpretation and graph neural networks, *Front. Endocrinol.* 14 (2023), 1213711.
- [141] D. Zhao, C. Wang, H. Qu, et al., Glycomol: A pervasive tool for structure prediction of natural saponin products basing on ms data, *J. Pharm. Anal.* 14 (2024), 100897.
- [142] F. Shi, H. Ning, R. Xiao, et al., Skin conductance-based acupoint and non-acupoint recognition using machine learning, *IEEE J. Biomed. Health Inform.* 28 (2024) 2569–2580.
- [143] Y. Tan, Z. Zhang, M. Li, et al., MedChatZH: A tuning LLM for traditional Chinese medicine consultations, *Comput. Biol. Med.* 172 (2024), 108290.

- [144] A.X. Lin, G. Chan, Y. Hu, et al., Internationalization of traditional Chinese medicine: Current international market, internationalization challenges and prospective suggestions, *Chin. Med.* 13 (2018), 9.
- [145] Y. Zhang, X. Li, Y. Shi, et al., ETCM v2.0: An update with comprehensive resource and rich annotations for traditional Chinese medicine, *Acta Pharm. Sin. B* 13 (2023) 2559–2571.
- [146] T. Han, L.C. Adams, K.K. Bressem, et al., Comparative analysis of multimodal Large Language model performance on clinical vignette questions, *JAMA* 331 (2024) 1320–1321.
- [147] Z. Wang, X. Wang, D. Zhang, et al., Traditional Chinese medicine network pharmacology: Development in new era under guidance of network pharmacology evaluation method guidance, *Zhongguo Zhongyao Zazhi* 47 (2022) 7–17.
- [148] S. Ma, J. Liu, W. Li, et al., Machine learning in TCM with natural products and molecules: Current status and future perspectives, *Chin. Med.* 18 (2023), 43.
- [149] F. Noor, M. Asif, U.A. Ashfaq, et al., Machine learning for synergistic network pharmacology: A comprehensive overview, *Briefings Bioinf.* 24 (2023), bbad120.
- [150] H. Wang, W. Wu, Z. Dou, et al., Performance and exploration of ChatGPT in medical examination, records and education in Chinese: Pave the way for medical AI, *Int. J. Med. Inf.* 177 (2023), 105173.