Journal of Multidisciplinary Healthcare

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Diagnostic Features and Prescription Rules of Influenza-Like Illnesses in Traditional Chinese Medicine: A Data Mining Approach

Xiao Liu[1], Ruiyu Chang[1], Shengyi Feng[2], Guoying Deng[3], Lin Zhou[4,5]

[1]Shanghai Institute of Tourism, Shanghai Normal University, Shanghai, People's Republic of China; [2]Center of Traumatology and Orthopedics, Yueyang Hospital of Integrated Traditional Chinese and Western Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai, People's Republic of China; [3]Trauma Center, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, People's Republic of China; [4]Management School, Guangdong Polytechnic Normal University, Guangzhou, People's Republic of China; [5]Business and Management School, Jilin University, Changchun, People's Republic of China

Correspondence: Lin Zhou, Email zhoulin628@126.com

Background: Influenza-like illness (ILI) is a syndromic diagnosis characterized by symptoms such as fever, cough, and sore throat, and may be caused by various respiratory pathogens, including influenza viruses, adenoviruses, parainfluenza viruses, and respiratory syncytial viruses. In Traditional Chinese Medicine (TCM), ILI is treated using syndrome differentiation and individualized herbal prescriptions. However, current prescription recommendations often rely on expert experience, with limited systematic analysis of clinical patterns. This study applies data mining techniques to analyze ILI-related TCM prescriptions, aiming to describe diagnostic features, prescription patterns, and herb compatibility.

Methods: Electronic medical records from April to December 2023 were collected from a TCM clinic, comprising ILI cases with documented therapeutic outcomes. Unstructured TCM inquiry text was transformed into structured diagnostic features using a rule-based keyword extraction method, encompassing 22 categories and over 70 specific indicators. Prescription data were analyzed using the Apriori algorithm with permutation-based significance testing to identify statistically significant herb combinations. Complex network analysis was applied to visualize and examine the global structure of herb compatibility.

Results: A total of 457 electronic medical records of influenza-like illness (ILI) were analyzed, yielding 105 distinct herbal ingredients with detailed characterization of thermal properties and flavor profiles. Cold- and warm-natured herbs predominated, with sweet, bitter, and acrid flavors most frequently observed. Using a rule-based extraction method, 22 diagnostic feature categories encompassing over 70 specific indicators—such as Personal Information, Mental State, Facial complexion and tongue coating, Fever, Pain, Sweating Condition, Nasal Discharge, Cough, Bowel Movement, Urination Status, Thirst—were systematically identified from TCM inquiry texts. Association rule mining initially yielded 101 statistically significant rules, of which 77 persisted under a stricter criterion, highlighting the robustness of both core and extension patterns in the treatment of influenza-like illness. The core structure of Maxing Shigan Decoction was highly stable across different herb combinations. Several extension modules, including Jiegeng, Houpo, Zhishi, and Guizhi, frequently co-occurred with the core, supporting the flexible application of classical prescriptions. Permutation testing, Z-score estimation, and effect size analysis confirmed the statistical robustness of these associations, and complex network analysis revealed a highly connected core comprising key herbs, emphasizing the stability and modularity of traditional TCM prescriptions.

Conclusion: This study systematically analyzed the diagnostic features and prescription patterns of ILI in TCM using data mining methods. The results identified Maxing Shigan Decoction as the core prescription, around which corresponding formulations and combinations were developed, quantitatively supporting classical prescription principles. Despite limitations in data volume, the findings demonstrate that integrating data-driven analysis with traditional knowledge can promote prescription standardization, enhance clinical decision-making, and contribute to the modernization of TCM within multidisciplinary healthcare contexts.

Keywords: influenza-like illness, traditional Chinese medicine, data mining, prescription patterns, herb compatibility

Received: 8 February 2025
Accepted: 3 September 2025
Published: 13 October 2025

Introduction

Influenza, commonly known as the flu, is an acute respiratory infection caused by influenza viruses,[1] characterized by sudden onset, rapid transmission, and the potential to reach epidemic levels. Influenza-like illness (ILI) is a syndromic diagnosis encompassing infections caused by multiple pathogens beyond influenza viruses, including adenoviruses, parainfluenza, and respiratory syncytial virus.[2] Due to its high contagiousness and viral mutability, influenza remains a persistent global health concern, particularly when novel strains emerge.[3]

Conventional medical management of ILI relies primarily on antiviral agents, vaccination, and symptomatic treatment.[3] While effective for specific viral strains, these approaches face notable challenges, including antiviral resistance, strain-specific vaccine limitations, and suboptimal symptom relief in non-influenza ILI cases.[4] Traditional Chinese Medicine (TCM), with its long history of combating infectious diseases, adopts a holistic approach aimed at restoring the body's equilibrium.[2,3] Within TCM theory, ILI is typically attributed to external pathogenic factors such as wind-cold or wind-heat, and its onset may be influenced by environmental elements—cold, heat, dryness, and dampness.[5] Treatment is guided by syndrome differentiation and individualized prescriptions tailored to the patient's symptoms and constitution.[6–10] By regulating the balance of yin and yang, TCM aims to expel pathogens and strengthen vital energy, offering a personalized therapeutic strategy distinct from conventional medicine.

A key distinction between TCM and conventional medicine lies in treatment philosophy: conventional medicine focuses on pathogen-specific interventions, while TCM targets the overall functional state of the patient, aiming to correct internal imbalances and strengthen resistance. This patient-centered and individualized approach gives TCM unique advantages in addressing the diverse clinical presentations of ILI, including cases where pathogen-specific treatments are unavailable or ineffective.Recent studies have demonstrated TCM's potential in managing respiratory infections, including its adjunctive role in COVID-19 treatment.[11–13] However, there are also challenges in the treatment of ILI by TCM practitioners. Prescription decisions often rely on personal experience, which can affect the systematic and objective nature of treatment. Moreover, traditional prescription analysis mostly relies on manual reading and collation, which is inefficient and may fail to fully reveal the underlying prescribing rules and characteristics.

To address these limitations, modern information technology—especially data mining—has been introduced into TCM prescription analysis. Such methods can uncover deeper patterns in TCM treatment of ILI, providing a more scientific basis for clinical application. For instance, studies analyzing large-scale TCM prescriptions have identified frequent herb pairs and core formulations for ILI.[11,12] Yet, comprehensive mining of multi-dimensional patterns remains underexplored, limiting clinical translation.

Although prior studies have explored general TCM prescription patterns,[14,15] research specifically targeting ILI remains limited, particularly in leveraging real-world clinical data and the prescription patterns. Existing works often focus on manual herb-pair analysis[16] or small-sample case reviews,[17] lacking systematic methodologies to decode syndrome-herb-dose relationships at scale. This study applies employs a hybrid analytical approach and data mining techniques to systematically analyze real electronic medical records from April to December 2023.

In this study, we constructed a real-world TCM clinical database of ILI cases with confirmed therapeutic outcomes and applied a multi-step data mining framework to uncover prescribing patterns. (1) First, we developed a rule-based approach to extract core diagnostic features reflecting key TCM syndrome characteristics. (2) Second, we applied association rule mining combined with permutation-based significance testing to identify statistically robust herb compatibility patterns. (3) Finally, we employed network analysis (Gephi and K-core) to visualize prescription structures and pinpoint central herbal combinations. By integrating statistical rigor with network-based visualization, this work provides a data-driven foundation to support more precise and evidence-based clinical decision-making for TCM treatment of ILI.

Literature Review

TCM played a notable role in the treatment of COVID-19 virus infection during the global pandemic,[14–18] and this has led to extensive attention from scholars around the world. An increasing amount of research has been conducted in the

TCM field as a result. Data mining technology is an important tool that has also been more widely used. The previous scholarly research has used data mining technology in the TCM field and for the treatment of COVID-19. In addition, it has been used to research TCM treatments for a variety of other common diseases that include headache, heart disease, cancer, chronic inflammation, diabetes, epilepsy, retinal diseases, insomnia, and depression.[19–25] Wang et al conducted a study utilizing data mining techniques to analyze the medication patterns of traditional Chinese medicine (TCM) in the treatment of bradyarrhythmia.[26] The researchers selected 48 clinical studies from the PubMed and China National Knowledge Infrastructure databases, encompassing 99 different Chinese medicinal herbs.

Xue et al analyzed 37 studies and employed data mining techniques to identify effective tonic prescriptions.[27] Zhang et al organized cases of primary epilepsy from the “Clinical Experience Collection of Epilepsy by Professor Xiaocheng Yan” and utilized the data mining functions of the Ancient and Modern Medical Case Cloud Platform to identify commonly used herbs and core herb pairs in Professor Yan’s treatment of primary epilepsy.[28] Zhang et al utilized frequency statistics and association rule mining to identify Radix astragali (Huangqi) and Rhizoma atractylodis macrocephalae (Baizhu) as core components in traditional Chinese medicine. Subsequently, they conducted animal experiments to validate the efficacy of these two herbs and their effects on relevant indicators in mice.[29] Bi S et al collected 715 traditional Chinese medicine prescriptions for the treatment of coronary heart disease and analyzed the usage patterns of herbs in these prescriptions using data mining methods, including frequency statistics and association rule analysis.[30]

Data mining technology has become increasingly mature in recent years, and it is widely used in many fields, including TCM,[31] as previously stated. Data mining has unique advantages in clinical big data research, especially the use of large medical public databases.[32] Data mining technology can be used to explore the potential laws of TCM that include the study of TCM formulas, TCM medicinal properties, Chinese medicine evidence, the prevention, diagnosis, and treatment of diseases using TCM, TCM special therapies, and other topics.

Remarkable results have already been achieved, especially for Chinese medicine prescription and medication laws. Descriptive statistics, association rule analyses, cluster analyses, complex network analyses, and other data analyses methods have been used by scholars[33–35] to analyze a large number of TCM prescriptions in depth and to reveal the laws of Chinese medicine prescriptions and drug combinations. These studies have provided powerful support for clinical practice.

Although some studies have explored TCM prescription patterns, data mining research specifically on influenza remains limited. Data mining can reveal herb usage frequency, combination patterns, and syndrome correlations, supporting personalized treatment and prescription standardization. This study analyzed clinically effective ILI cases using data mining to identify prescription patterns, aiming to optimize treatment strategies and promote the modern application of TCM in influenza management.

Data and Methods

Data Sources

Actual clinical cases were collected from the Chinese medical clinic, Renhe Tang, where patients provided doctors with detailed questionnaires and images of their tongues, and doctors prescribed medicines based on the answers to their questions. 457 Clinical cases with clear treatment results from April 2023 to December 2023 were collected. The consultation form contains 22 questions covering personal information, mental state, facial complexion, fever status, pain status, sweating condition, runny nose, coughing, phlegm, bowel and urination conditions, thirst, bad breath, nausea and vomiting, shortness of breath, appetite, dinner situation, food preferences, limb condition, sleep quality, medical history, current symptoms, and differences in symptoms between day and night.

Data Inclusion Criteria

Influenza-like cases were selected according to the Guidelines for Diagnosis and Treatment of Influenza (2011 edition), and other cases were excluded. The scope of the search was described as seasonal cold, influenza-related diseases, and those diagnosed as wind-warm, summer-warm, damp-warm, external, and high fever whose symptoms were consistent with influenza. A total of 457 clinical medical cases with clear therapeutic effects were included in this study.

Exclusion Criteria

i. Patients who lacked primary symptoms or whose primary symptoms did not meet the inclusion criteria;
ii. Patients who failed to demonstrate significant clinical improvement;
iii. Patients who received other medications and did not undergo a full course of Chinese medicine treatment.

Database Creation and Data Normalization

The screened and organized prescriptions were entered into an EXCEL form, and the data were audited after the data entry was completed to establish a Chinese medicine prescription database for the treatment of influenza-like cases. It was necessary to standardize the names of Chinese medicines in the database due to the homonym phenomenon in the original prescriptions. The Chinese medicine names were unified with reference to the Pharmacopoeia of the People's Republic of China (2020 edition), the Chinese Materia Medica, and the Chinese Herbal Medicine.[36]

Data Mining Methods

Rule-Based Extraction of Diagnostic Features

We employed a rule-based method to transform unstructured inquiry text into structured diagnostic features.

(1) Keyword Dictionary

A dictionary was developed based on the inquiry form categories. Each category contained predefined keywords relevant to clinical descriptions.

(2) Keyword Matching

In Python, the inquiry text was segmented using jieba, and relevant terms were matched via regular expressions. The matched keywords were recorded as feature variables for each case. This process transformed unstructured narrative data into structured variables suitable for subsequent statistical analysis or machine learning applications. A partial example of the keyword extraction function is provided in Table 1.

Association Rule Mining with Permutation-Based Significance Testing

This section describes the use of the Apriori algorithm to identify statistically significant herb compatibility patterns in prescriptions for influenza-like illness (ILI). The analysis was performed using Python 3.12, incorporating both frequent pattern mining and permutation-based significance testing to reduce the risk of spurious associations. The following statistical indicators were computed for each rule: support, confidence, lift, lift z-score, effect size, and p-value. To assess robustness, a permutation test was applied by randomly shuffling feature–prescription pairs and recalculating metrics, thus estimating the probability of observing each rule by chance. Rules with $p < 0.05$ were considered statistically significant.

Table 1 Keyword Extraction Function

Keyword Extraction Function

```
import re
import jieba
def extract_keywords_with_nlp(questionnaire, question_keywords):
extracted_features = {}
for category, keywords in question_keywords.items():
matched_keywords = set()
keyword_pattern = “|”.join(map(re.escape, keywords))
for q in questionnaire:
words = jieba.lcut(q)
filtered_words = set(words)
found = re.findall(keyword_pattern, q)
matched_keywords.update([word for word in found if word in filtered_words])
if matched_keywords:
extracted_features[category] = list(matched_keywords)
return extracted_features
```

Complex Network

The complex network analysis method, increasingly applied in Chinese medicine research, reveals intricate relationships among various TCM elements. In this study, significant association rules were represented as a network in which nodes denoted diagnostic features or prescriptions, and edges indicated significant associations. In TCM, nodes may represent Chinese medicines, formulas, diseases, or symptoms, while edges capture the interactions between them. Such a network enables the identification of key information, including important nodes, sub-networks, and critical paths.

Results

Frequency of Use and Distribution of Drug Properties

A Microsoft Excel 2021 pivot table was used to calculate the frequency of occurrence of Chinese herbal medicines in prescriptions. Among the 457 prescriptions entered for the treatment of ILIs, a total of 105 Chinese herbal medicines were identified, with a cumulative frequency of 4043 occurrences. Among them, 21 appeared only once, 62 had frequencies of less than 10, 27 ranged from 10 to 100, and the remaining 16 exceeded 100, including 6 with frequencies over 200 (see Table 2).

For cold treatments, the most frequently used traditional Chinese medicine was Xingren (Semen Armeniacae Amarum), at 380 times. This was followed by Mahuang (Herba Ephedrae), Jiegeng (Radix Platycodonis), Gancao (Radix Glycyrrhizae), Shigao (Gypsum Fibrosum), and Houpo (Cortex Magnoliae Officinalis), with frequencies of occurrence of 375, 294, 290, 257, and 219 times, respectively.

The analysis of the Four Natures, Five Flavors, and Channel Tropism of the medicines revealed that cold-natured medicines, such as Zhishi (*Fructus Aurantii Immaturus*), Zhuru (Bamboo Shavings), Chaihu (*Radix Bupleuri)*, and Huangqin (*Radix Scutellariae*), as well as warm-natured medicines, such as Xinyi (*Flos Magnoliae*), Banxia (*Rhizoma Pinelliae*), Cangzhu (*Rhizoma Atractylodis*), Chenpi (*Pericarpium Citri Reticulatae*), Guizhi (*Ramulus Cinnamomi*) were used most frequently.

Among the Four Natures, cold-natured medicines appeared most frequently, followed by warm-natured medicines. The remaining natures, ranked by frequency, were slightly cold, neutral, slightly warm, extremely cold, and hot (Figure 1).

In the analysis of the Five Flavors, sweet, bitter, and acrid flavors were predominant (Figure 2). Regarding Channel Tropism, the most frequently targeted meridians were the Lung, Liver, Spleen, and Stomach meridians (Figure 3).

Table 2 Drug Frequencies (First 36)

Herb	Frequency	Percentage/%	Herb	Frequency	Percentage/%
Xingren	380	9.40	Caoguo	46	1.14
Mahuang	375	9.28	Baizhi	45	1.11
Jiegeng	294	7.27	Baizhu	38	0.94
Gancao	290	7.17	Niubangzi	38	0.94
Shigao	257	6.36	Digupi	37	0.92
Houpo	219	5.42	Shengjiang	34	0.84
Zhishi	183	4.53	Zhimu	34	0.84
Xinyi	154	3.81	Gegen	32	0.79
Banxia	137	3.39	Yiyiren	26	0.64
Cangzhu	132	3.26	Dazao	25	0.62
Chenpi	131	3.24	Shichangpu	23	0.57
Guizhi	124	3.07	Danpi	22	0.54
Zhuru	115	2.84	Lugen	22	0.54
Chaihu	114	2.82	Maidong	22	0.54
Fuling	110	2.72	Xuanshen	21	0.52
Huangqin	103	2.55	Maogen	20	0.49
Baishao	68	1.68	Dahuang	19	0.47
Binglang	57	1.41	Baixianpi	16	0.40

https://doi.org/10.2147/JMDH.S521388

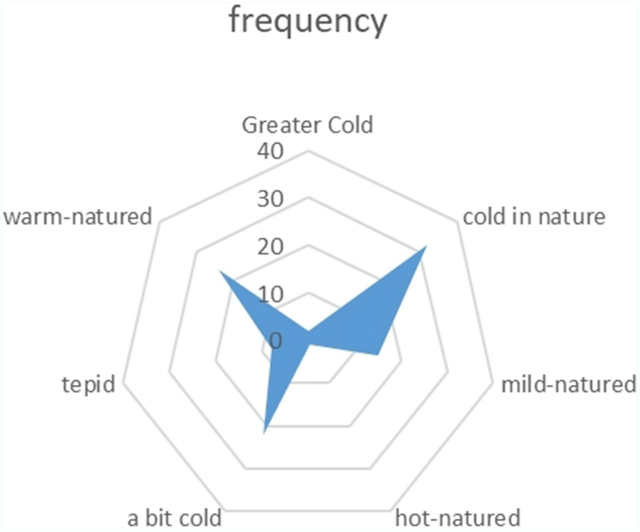

Figure 1 Analysis of the Four Natures of Medicines.

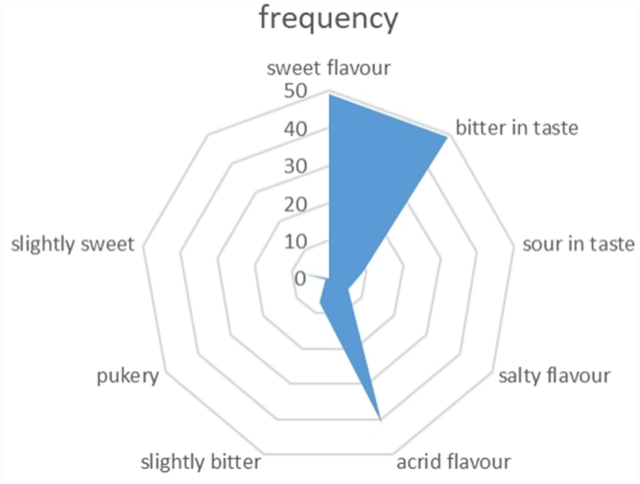

Figure 2 Analysis of the Five Flavors of Medicines.

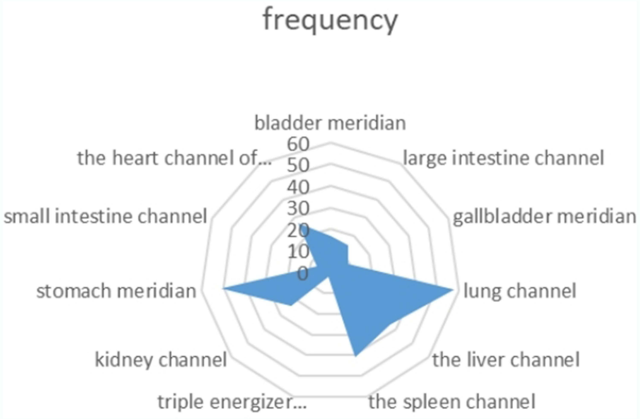

Figure 3 Categorization analysis of the drugs.

Association Rule Mining and Statistical Validation

Association rule mining using Apriori algorithm was performed using Python 3.12 on a dataset comprising 105 distinct herbs across 457 clinically validated prescriptions. Each prescription was transformed into a transaction and encoded into a Boolean matrix using the TransactionEncoder module from the mlxtend library.

To reduce spurious associations driven by high-frequency herbs, a permutation test with 500 iterations was implemented. In each iteration, herbs were randomly reassigned among prescriptions while preserving the number of herbs per prescription, thereby generating a randomized dataset. The Apriori algorithm was applied to each randomized dataset to establish the null distribution of lift values for all potential rules.

For each observed rule, the deviation of its lift from the null expectation was quantified using the Z-score, and statistical significance was assessed via the corresponding P-value. Additionally, effect sizes were calculated as the proportional increase of the observed lift relative to the null mean, and 95% confidence intervals (CIs) for lift under the null model were estimated. Only rules with $P< 0.05$ were retained as statistically significant, ensuring that the final set reflected both high-frequency and statistically robust herb combinations.

Frequent itemsets were identified via the Apriori algorithm with a minimum support of 0.03 and a confidence threshold of 0.50, which yielded 101 statistically significant rules ($P < 0.05$). When the minimum support was increased to 0.05 while keeping the confidence threshold constant, 77 statistically significant rules ($P < 0.05$) were obtained.

Significant rules were categorized by the number of drugs involved and the top 20 rules in two-drug and multi-drug groups were analyzed. This combined evaluation of support, confidence, lift, statistical significance, and effect size ensures that identified herb combinations are both frequent and statistically robust, providing reliable insights into traditional Chinese medicine compatibility.

In Table 3, several high-support rules revealed the central role of the Maxing Shigan Decoction core structure. For instance, the pair Xingren-Mahuang exhibited the highest co-occurrence (support=0.81, confidence=0.98, lift=1.19, p=0.002), indicating that nearly all prescriptions containing Xingren also included Mahuang. Similar robust associations were observed between Shigao and Xingren or Mahuang, as well as Gancao with the aforementioned herbs, thereby delineating the stable four-herb core of Maxing Shigan Decoction.

Table 3 Association Rules of the Drug Pair Combinations

Antecedents	Consequents	Support	Confidence	Lift	Lift_z_score	Effect_size	p_value
Xingren	Mahuang	0.81	0.98	1.19	6.68	0.24	0.002
Mahuang	Xingren	0.81	0.99	1.19	6.65	0.24	0.002
Jiegeng	Xingren	0.56	0.86	1.04	2.01	0.08	0.021
Jiegeng	Mahuang	0.54	0.84	1.03	1.81	0.07	0.038
Shigao	Xingren	0.51	0.91	1.10	3.37	0.14	0.002
Shigao	Mahuang	0.50	0.89	1.09	3.05	0.13	0.004
Xinyi	Mahuang	0.29	0.94	1.14	3.20	0.18	0.002
Zhishi	Jiegeng	0.29	0.72	1.12	1.96	0.09	0.026
Guizhi	Mahuang	0.26	0.98	1.19	3.38	0.23	0.005
Gancao	Xingren	0.26	0.90	1.09	2.02	0.12	0.032
Gancao	Mahuang	0.26	0.89	1.09	2.17	0.13	0.012
Guizhi	Xingren	0.26	0.95	1.14	2.80	0.18	0.009
Xinyi	Jiegeng	0.23	0.74	1.15	1.93	0.09	0.037
Gancao	Shigao	0.22	0.76	1.35	4.84	0.21	0.048
Banxia	Jiegeng	0.20	0.75	1.17	2.28	0.11	0.031
Zhuru	Jiegeng	0.19	0.76	1.18	2.49	0.12	0.029
Gancao	Mahuang	0.19	0.96	1.16	2.67	0.19	0.007
Caoguo	Houpo	0.10	0.98	2.04	7.63	0.54	0.037
Niu Bangzi	Jiegeng	0.08	1.00	1.55	4.54	0.38	0.006
Zhimu	Shigao	0.07	0.94	1.67	3.98	0.36	0.013

Beyond this core, extension modules were also detected. The combination of Guizhi with Mahuang (support=0.26, confidence=0.98, lift=1.19) suggests a frequent integration of Maxing Shigan Decoction with the principles of Guizhi Decoction to enhance exterior-releasing effects. Jiegeng emerged as a central hub herb, showing significant associations with Xingren, Mahuang, Xinyi, Niubangzi, Zhuru, and Banxia, consistent with its role in facilitating lung qi dispersal and directing herbs upward.

The pair Caoguo-Houpo demonstrated a remarkably high effect size (0.54) and lift (2.04), indicating a strong and specific co-usage pattern for resolving dampness. Likewise, the pair Zhimu- Shigao (support=0.07, confidence=0.94, lift=1.67, p=0.013) corresponds to the core herbal dyad of Baihu Tang (White Tiger Decoction), reinforcing the classical “heat-clearing and fire-purging” therapeutic principle. Furthermore, the pair Niu bangzi-Jiegeng showed a perfect confidence with high lift, which matches the core structure of Niubang Ganju Tang, suggesting an exclusive co-occurrence pattern in prescriptions involving Niu bangzi.

Collectively, these findings reveal both core herbal combinations and specific therapeutic modules, thereby providing quantitative evidence for the stability and extension strategies of classical Chinese medical prescriptions.

In the association rules from the Table 4 involving three or more herbs, the core structure of Maxing Shigan Decoction (Mahuang–Xingren–Shigao–Gancao) remained evident. For example, the combinations Jiegeng–Mahuang–Xingren (support=0.54, confidence=0.99, lift=1.20) and Xingren–Shigao–Mahuang (support=0.50, confidence=0.98, lift=1.20) demonstrated very high support and confidence, reflecting the central role of this classical prescription within the dataset.

Beyond this core structure, several extension combinations also showed strong significance. The rules Jiegeng–Houpo–Mahuang (support=0.26, confidence=1.0, lift=1.20) and Shigao–Houpo–Mahuang (support = 0.23, confidence = 1.0, lift = 1.20) suggest the collaborative role of Houpo, likely associated with its function in regulating qi, relieving fullness, and resolving dampness.

Furthermore, the combinations Zhishi–Mahuang–Xingren (support = 0.30, confidence = 1.0, lift = 1.20) and Xingren–Zhishi–Jiegeng (support = 0.24, confidence = 0.76, lift = 1.18) reveal a synergistic pattern between Zhishi and Jiegeng, both of which influence qi dynamics and strengthen the effects of dispersing the lung and relieving chest oppression.

Overall, the rules involving three or more herbs exhibit two main features: on the one hand, the classical structure of Maxing Shigan Decoction continues to dominate; on the other hand, extension modules represented by Jiegeng, Zhishi,

Table 4 Association Rules of the Three-Flavor Herbal Combinations

Antecedents	Consequents	Support	Confidence	Lift	Lift_z_score	Effect_size	p_value
Giegeng & Mahuang	Xingren	0.54	1.00	1.20	4.57	0.27	0.003
Xingren & Giegeng	Mahuang	0.54	0.97	1.19	4.19	0.25	0.005
Xingren & Shigao	Mahuang	0.50	0.98	1.20	4.59	0.27	0.003
Mahuang & Houpo	Xingren	0.37	1.00	1.20	4.20	0.26	0.003
Giegeng & Shigao &Mahuang	Xingren	0.31	1.00	1.20	3.09	0.25	0.016
Giegeng & Shigao	Mahuang	0.31	0.89	1.08	2.14	0.14	0.034
Zhishi & Mahuang	Xingren	0.30	1.00	1.20	3.80	0.26	0.003
Xingren & Xinyi	Mahuang	0.29	0.99	1.21	3.70	0.26	0.006
Xingren & Gancao	Mahuang	0.26	0.98	1.20	3.57	0.25	0.003
Guizhi &Xingren	Mahuang	0.26	1.00	1.22	3.09	0.26	0.003
Giegeng&Mahuang &Houpo	Xingren	0.26	1.00	1.20	2.64	0.23	0.028
Xingren & Zhishi	Giegeng	0.24	0.76	1.18	2.23	0.11	0.043
Cangzhu &Mahuang	Xingren	0.23	1.00	1.20	3.60	0.25	0.003
Zhishi &Mahuang	Giegeng	0.23	0.75	1.16	2.09	0.11	0.042
Shigao &Mahuang & Houpo	Xingren	0.23	1.00	1.20	2.36	0.19	0.038
Shigao &Houpo	Mahuang	0.23	0.90	1.10	2.21	0.15	0.038
Giegeng & Xinyi	Xingren	0.22	0.94	1.13	2.22	0.18	0.042
Xingren &Gancao	Shigao	0.21	0.82	1.45	4.64	0.27	0.045
Giegeng & Xinyi	Mahuang	0.21	0.93	1.14	2.35	0.18	0.018
Mahuang & Gancao	Shigao	0.21	0.82	1.45	5.05	0.25	0.034

and Houpo are frequently integrated into the core framework, forming functional clusters related to dispersing the lung, regulating qi, and resolving dampness. This reflects the flexible expansion of classical prescriptions within clinical applications.

Extraction of Key Features from Consultation Data

Based on the steps in Rule-Based Extraction of Diagnostic Features, TCM Diagnostic Features of ILI Cases are obtained, as shown in Table 5.

Complex Network Analysis

We organized all of the drugs into a co-occurrence matrix format and imported this into Gephi.0.10.1 software. The layout was set to Fruchterman Reingold, the zone was set to 10000, the gravity was set to 10, the speed was to one, and "run" was clicked in the filter window. The size of the nodes was decided according to the size of the degree value, the type of modularity was set to different node colors, the thickness of the edges was 1.5, and "refresh" was clicked to obtain the complex network graph. The data table window allowed us to see the status of each parameter of the complex network diagram. The software was used to calculate the relevant indicators.

There were 105 nodes with 1603 edges. The average degree was 15.267, the average weighted degree was 156.305, the network diameter was 4, and the network density was 0.147 (Figure 4). The K-core was greater than 26; and according to the graph, Mahuang, Xingren, Gancao, Shigao, Jiegeng, Houpo, Zhishi, Xinyi, Banxia, and Cangzhu were the nucleus of the drugs (Figure 5). The complex network diagram of the core drugs is shown in Figure 6.

Table 5 TCM Diagnostic Features of ILI

No.	Issues	Features
1	Personal Information	Gender, Age, Height, Weight, Place of birth, Current residence
2	Mental state	Mental alertness, Irritability, Temperament, Talkativeness
3	Facial complexion and tongue coating	Facial complexion (Sallow complexion, Pale complexion, Dark complexion, Flushed complexion, Bluish complexion), tongue coating
4	Fever	Fever, Degree of fever, Cold intolerance, Wind sensitivity
5	Pain	Headache, dizziness, body pain, Location of pain
6	Sweating condition	Sweating, Tendency to sweat, Time of excessive sweating, Aversion to heat or cold
7	Nasal discharge	Nasal discharge, Clear and watery nasal discharge, Yellow and thick nasal discharge
8	Cough	Cough, phlegm, Thin and clear or thick and sticky (phlegm), White or yellow (phlegm)
9	Bowel movement	Bowel movements in the past three days, Bowel movement frequency, Dry and hard or loose, soft, and sticky, Stool formation (sausage-shaped or not), Stool odor
10	Urination status	Urine color (clear, light yellow, or dark yellow), Urine volume (large or small), Urinary frequency
11	Thirst	Thirst, Water intake (high or low), Preference for warm or cold water
12	Halitosis	Halitosis, an abnormal oral odor
13	Nausea and vomiting	Nausea and vomiting, Dry heaving or actual vomiting, Vomited content
14	Dyspnea	Dyspnea, Inhalation difficulty or exhalation difficulty
15	Appetite	Appetite, Frequency of belching, Frequency of flatulence, Abdominal bloating or pain, vomiting or diarrhea
16	Dinner	Daily dinner time, Nighttime snacking, Pre-sleep milk consumption
17	Eating and drinking habits	Recent food and drink preferences, Frequently consumed food and drink, Food and drink consumed two days before illness
18	Extremities temperature	Hands and feet temperature (warm or cold), Abdominal temperature (cold or warm), Buttocks temperature (cold or warm)
19	Sleep	Sleep quality, Bedtime and wake-up time, Daytime sleepiness
20	Medical history	Previous illnesses, Names of recently taken medications, Duration of medication use
21	Main symptoms	Currently most uncomfortable symptoms, Symptoms in order of severity, from most to least severe
22	Discomfort status	Severity of discomfort (day or night)

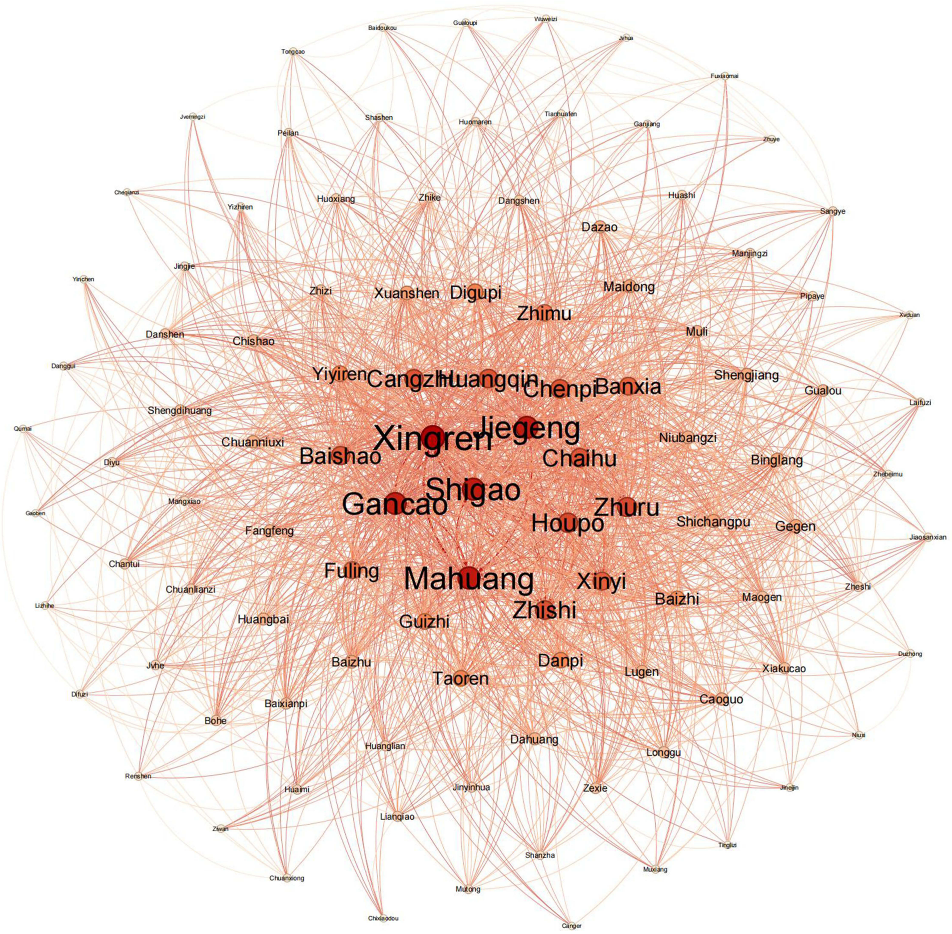

Figure 4 Overall complex network diagram.

Discussion

This study analyzes 457 electronic medical records of influenza treatment, not only extracting valuable experience in Traditional Chinese Medicine (TCM) for influenza but also revealing the medication patterns and principles used in the treatment process.

A rule-based diagnostic feature extraction method was employed to systematically extract structured data encompassing 22 aspects and over 70 specific diagnostic indicators from Traditional Chinese Medicine (TCM) inquiry texts of influenza-like illness (ILI) cases. By constructing a detailed keyword dictionary combined with jieba word segmentation and regular expression matching techniques, the method efficiently parsed and extracted information from unstructured texts. The extracted diagnostic features covered multidimensional content including personal information, mental state, facial complexion and tongue coating, fever, cough, and eating habits, fully reflecting the complexity and granularity of TCM diagnosis.This approach not only ensured the accuracy and interpretability of feature extraction but also provided a solid data foundation for subsequent data mining and machine learning modeling.

This study, based on association rule mining of 457 clinically validated prescriptions, revealed the core and extended compatibility features of traditional Chinese medicine (TCM) in the treatment of influenza-like illness (ILI). The reliability of the association rules was strengthened by the application of permutation testing, complemented by Z-score estimation and effect size analysis, which effectively minimized spurious associations arising from high-frequency herbs. The results demonstrated that the core structure of Maxing Shigan Decoction (Mahuang–Xingren–Shigao–Gancao) remained highly stable across two-herb, three-herb, and multi-herb combinations, with significant levels of support and confidence. These findings provide quantitative evidence for the stability and universality of this classical prescription in clinical practice, indicating that Maxing Shigan Decoction continues to serve as the most commonly employed foundational formula in the treatment of ILI.

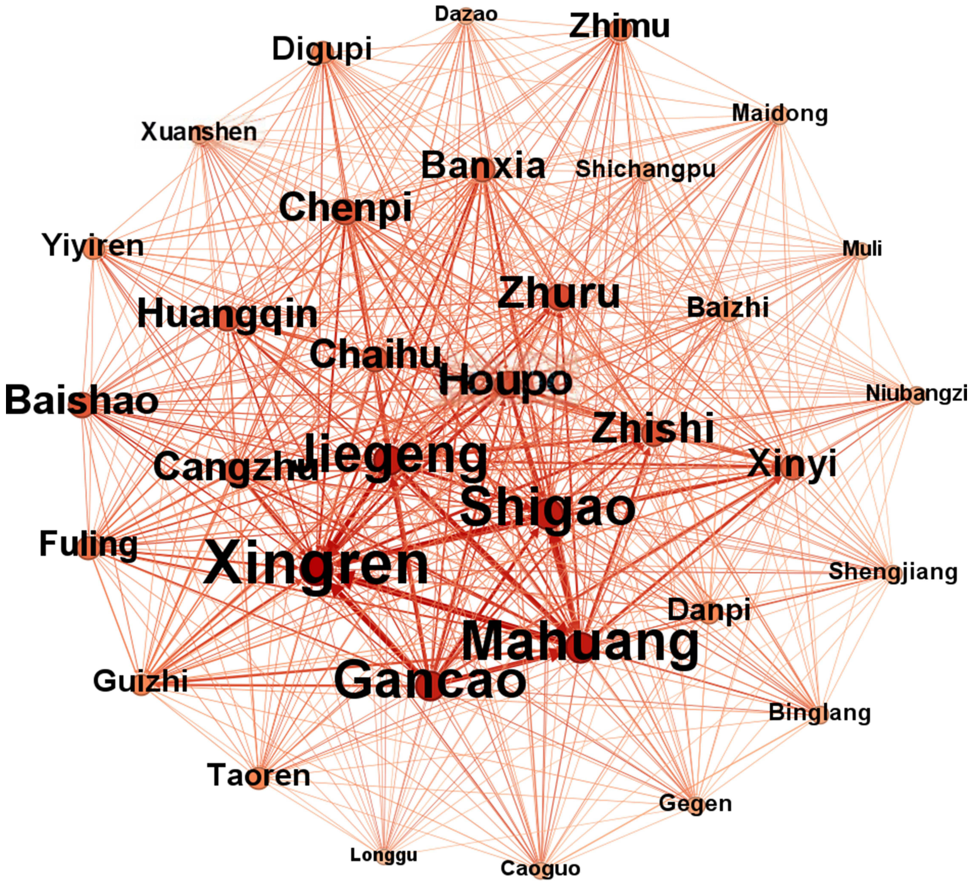

Figure 5 Complex network diagram with the K-core greater than 26.

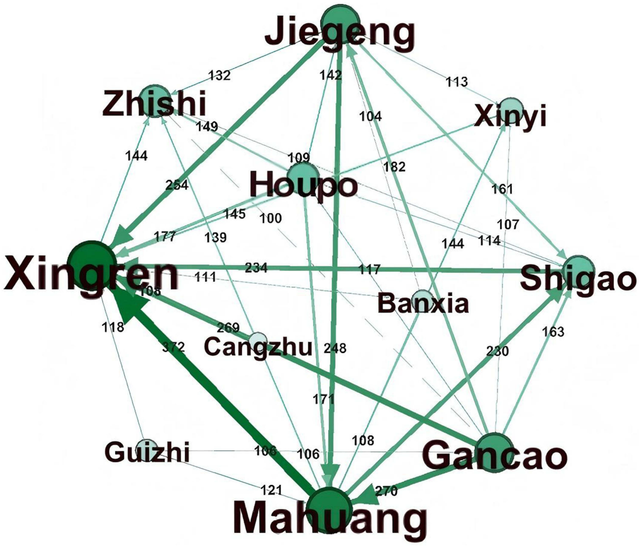

Figure 6 Complex network diagram of the core drugs.

Beyond the core structure, this study identified multiple extension modules that frequently co-occurred with the core herbs, reflecting the flexibility of clinicians in syndrome differentiation and treatment. Typical examples include the frequent incorporation of Jiegeng, Houpo, and Zhishi, which are often associated with concurrent symptoms such as sore

https://doi.org/10.2147/JMDH.S521388

throat, chest oppression, excessive phlegm, and impaired qi movement in ILI. This pattern of "targeted modification based on core formulas" embodies a fundamental principle of TCM compatibility, wherein the core prescription addresses the primary pathogenesis while extension herbs are employed to adjust for accompanying symptoms. The "core plus extension" model revealed by statistical analysis not only aligns with the clinical practice of syndrome-based treatment in TCM but also provides quantitative evidence for its flexibility in real-world applications.

Complex network analysis further elucidated the global structure of herb compatibility. In the Figure 4, Mahuang, Xingren, Gancao, Shigao, and Jiegeng occupied central positions with high degrees and strong edge weights, underscoring their pivotal role in the prescription system. In the Figure 5, although the number of edges decreased, the connections among core herbs became denser and stronger, further confirming the central role of Maxing Shigan Decoction and its modifications in ILI treatment. Moreover, herbs such as Houpo, Zhishi and Cangzhu remained connected to the core nodes, suggesting their functional roles in enhancing the original formula's effects and tailoring treatment to individual patient presentations.

In summary, this study integrated rule-based feature extraction, Apriori association rule mining with permutation testing, and complex network analysis to comprehensively investigate diagnostic features and herb compatibility in influenza-like illness (ILI). The methods effectively transformed unstructured clinical data, identified statistically significant herb combinations aligned with classical formulas, and revealed the overall prescription structure and variations.

Nonetheless, several limitations should be acknowledged. The rule-based feature extraction depended on a predefined keyword dictionary, which may limit the detection of implicit or novel expressions in clinical records. The association rule analysis was constrained by the sample size and representativeness of the prescription dataset, potentially affecting the generalizability of the results. Future studies should focus on integrating advanced natural language processing techniques, expanding and diversifying datasets, and incorporating pharmacological and clinical outcome data to enhance the robustness and clinical applicability of the findings.

Consent to Participate

This study only analyzed the medicines used in each case and focused on medicine analysis. As a data mining study, it did not include any individual patient details. Therefore, patient consent for reviewing medical records was not required by the Human Participants Ethics Committee of Shanghai Normal University, which approved the study. Patient data confidentiality was strictly maintained, and the study was conducted in compliance with the Declaration of Helsinki.

Funding

National Natural Science Foundation of China[72171170]; National Natural Science Foundation of China [82405422]; Guangdong Province Philosophy and Social Science Planning General Project [GD21CGL27]; Start-up funding from Guangdong Polytechnic Normal University [2021SDKYB019]; Heyuan Municipal Philosophy and Social Science Planning General Project [HYSK24P30].

Disclosure

The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Xiong Y, Li NX, Duan N, et al. Traditional Chinese medicine in treating influenza: from basic science to clinical applications. *Front Pharmacol*. 2020;11:575803. doi:10.3389/fphar.2020.575803
2. Wang Y, Liu M, Jafari M, et al. A critical assessment of traditional Chinese medicine databases as a source for drug discovery. *Front Pharmacol*. 2024;15:1303693. doi:10.3389/fphar.2024.1303693
3. Jiashuo W, Fangqing Z, Zhuang L, et al. Integration strategy of network pharmacology in traditional Chinese medicine: a narrative review. *J Traditional Chin Med*. 2022;42(3):479–486. doi:10.19852/j.cnki.jtcm.20220408.003
4. Zhou E, Shen Q, Hou Y. Integrating artificial intelligence into the modernization of traditional Chinese medicine industry: a review. *Front Pharmacol*. 2024;15:1181183. doi:10.3389/fphar.2024.1181183
5. Wu S, Su W, Fan Q, Shang H, Xiao W, Wang Y. Traditional Chinese medicine for the common cold: evidence and potential mechanisms. *Am J Chin Med*. 2023;51(03):487–515. doi:10.1142/S0192415X23500258

6. Gong L, Jiang J, Chen S, et al. A syndrome differentiation model of TCM based on multi-label deep forest using biomedical text mining. *Front Genetics*. 2023;14:1272016. doi:10.3389/fgene.2023.1272016
7. Li N, Yu J, Mao X, Zhao Y, Huang L. The research and development thinking on the status of artificial intelligence in traditional Chinese medicine. *Evid Based Complement Alternat Med*. 2022;2022:7644524. doi:10.1155/2022/7644524
8. Ma J, Wang Z, Guo H, et al. Mining syndrome differentiating principles from traditional Chinese medicine clinical data. *Comput Syst Sci Eng*. 2021;40(3):979–993. doi:10.32604/csse.2022.016759
9. Dou Z, Xia Y, Zhang J, et al. Syndrome differentiation and treatment regularity in traditional Chinese medicine for type 2 diabetes: a text mining analysis. *Front Endocrinol*. 2021;12:728032. doi:10.3389/fendo.2021.728032
10. Song Y, Zhao B, Jia J, et al. A review on different kinds of artificial intelligence solutions in TCM syndrome differentiation application. *Evid - Based Complementary Alternative Med*. 2021;2021:1741427X. doi:10.1155/2021/6654545
11. Guo Y, Wang T, Chen W, et al. Acceptability of traditional Chinese medicine in Chinese people based on 10-year's real world study with mutiple big data mining. *Front Public Health*. 2022;9:811730. doi:10.3389/fpubh.2021.811730
12. Qi F, Tang W. Traditional Chinese medicine for treatment of novel infectious diseases: current status and dilemma. *Biosci Trends*. 2021;15 (4):201–204. doi:10.5582/bst.2021.01263
13. Luo C, Yu H, Yang T, et al. Data mining and systematic pharmacology to reveal the mechanisms of traditional Chinese medicine in recurrent respiratory tract infections' treatment. *Evid Based Complement Alternat Med*. 2020;2020(1):8979713. doi:10.1155/2020/8979713
14. Xiong J, Wang J, Liu Y, et al. 41 Study on prescription medication mode and mechanism of traditional Chinese medicine in the treatment of noncritical COVID-19 based on data mining. *Trad Med Res*. 2023;8(6):36. doi:10.53388/TMR20221023001
15. Hao E, Su Z, Gong Y, et al. Analysis on application law of dampness-removing traditional Chinese medicines in treatment of coronavirus disease 2019. *Chinese Herbal Medicines*. 2021;13(4):518–524. doi:10.1016/j.chmed.2021.09.011
16. Wu D, Hou X, Xia Z, et al. Analysis on oral medication rules of traditional Chinese medicine prescriptions for prevention of COVID-19. *Chinese Herbal Medicines*. 2021;13(4):502–517. doi:10.1016/j.chmed.2021.10.007
17. Ma J, Wu HY, Chen YZ, Huang M, Zhang LS. Thoughts on traditional Chinese medicine treatment of novel coronavirus pneumonia based on two cases. *Chin J Integr Med*. 2021;27(5):375–378. doi:10.1007/s11655-020-3485-9
18. Yin LW, Gao YC, Li ZP, et al. Analysis of Chinese herbal formulae recommended for COVID-19 in different schemes in China: a data mining approach. *Comb Chem High Throughput Screening*. 2021;24(7):957–967.
19. Liu X, Sun P, Bao X, et al. Potential mechanisms of traditional Chinese medicine in treating insomnia: a network pharmacology, GEO validation, and molecular-docking study. *Medicine*. 2024;103(18):e38052. doi:10.1097/MD.0000000000038052
20. Zhang J, Zhang K, Chang X-W, et al. Exploring a novel traditional Chinese medicine prescription containing Chrysanthemi Flos for retinal diseases: an integrated strategy of data mining and network pharmacology. *Trad Med Res*. 2024;9(4):23. doi:10.53388/TMR20231023001
21. Hong M, Zhao YD, Zhong TL, et al. Out-of-set association analysis of lung cancer drugs and symptoms based on clinical case data mining. *Technol Health Care*. 2024;Preprint:1–11.
22. Hu L, Yang Y, Wang Z, et al. A real-world data analysis-based study of Chinese medicine treatment patterns after breast cancer surgery. *Medicine*. 2023;102(50):e36642. doi:10.1097/MD.0000000000036642
23. Weng J, Wu X, Shao P, et al. Medicine for chronic atrophic gastritis: a systematic review, meta-and network pharmacology analysis. *Ann Med*. 2023;55(2):2299352. doi:10.1080/07853890.2023.2299352
24. Zhou C, Wei J, Yu P, et al. Convergent application of traditional Chinese medicine and gut microbiota in ameliorate of cirrhosis: a data mining and Mendelian randomization study. *Front Cell Infect Microbiol*. 2023;13:1273031. doi:10.3389/fcimb.2023.1273031
25. Hong M, Sun WH, Lu M, et al. Analysis of the cluster efficacy and prescription characteristics of traditional Chinese medicine intervention for non-small cell lung cancer based on a clustering algorithm. *Technol Health Care*. 2023;31(5):1759–1770. doi:10.3233/THC-220644
26. Wang X, Zhang X, Li J, et al. Analysis of prescription medication rules of traditional Chinese medicine for bradyarrhythmia treatment based on data mining. *Medicine*. 2022;101(44):e31436. doi:10.1097/MD.0000000000031436
27. Xue D, Zhang Y, Song Z, et al. Integrated meta-analysis, data mining, and animal experiments to investigate the efficacy and potential pharmacological mechanism of a TCM tonic prescription, Jianpi Tongmai formula, in depression. *Phytomedicine*. 2022;105:154344. doi:10.1016/j.phymed.2022.154344
28. Zhang Y, Tong L, Chen G, et al. Analysis of medication rule of primary epilepsy based on Xiaocheng Yan's clinical experience collection of epilepsy. *Evid Based Complement Alternat Med*. 2022;2022:9539944. doi:10.1155/2022/9539944
29. Zhang C, Guo W, Yao X, et al. Database mining and animal experiment-based validation of the efficacy and mechanism of radix astragali (Huangqi) and rhizoma atractylodis macrocephalae (Baizhu) as core drugs of traditional Chinese medicine in cancer-related fatigue. *J Ethnopharmacol*. 2022;285:114892. doi:10.1016/j.jep.2021.114892
30. Bi S, Xu L, Chen S, Bu S, Xu Y. Detection of herbal combinations and pharmacological mechanisms of clinical prescriptions for coronary heart disease using data mining and network pharmacology. *Evid Based Complement Alternat Med*. 2021;2021:9234984. doi:10.1155/2021/9234984
31. Zhu X, Liu Y, Li Q, et al. Mining patterns of Chinese medicinal prescription for diabetes mellitus based on therapeutic effect. *Multimedia Tools Appl*. 2020;79(15–16):10519–10532. doi:10.1007/s11042-019-7226-z
32. Wu WT, Li YJ, Feng AZ, et al. Data mining in clinical big data: the frequently used databases, steps, and methodological models. *Military Med Res*. 2021;8(1):1–12. doi:10.1186/s40779-021-00338-z
33. Wang X, Pang F, Du XG. Analysis of traditional Chinese medicine symptoms in children with spastic cerebral palsy: a data mining study. *Multidiscip Healthc*. 2024;17:913–922. doi:10.2147/JMDH.S451768
34. Liu G, Yang H, Bai L, et al. Data mining-based analysis on medication rules of Chinese herbal medicine treating headache with blood stasis syndrome. *Heliyon*. 2023;9(4):1–12.
35. Dong N, Zhang X, Wu D, et al. Medication regularity of traditional Chinese medicine in the treatment of aplastic anemia based on data Mining. *Evid Based Complement Alternat Med*. 2022;2022:1605359. doi:10.1155/2022/1605359
36. Xia P, Gao K, Xie J, et al. Data mining-based analysis of Chinese medicinal herb formulae in chronic kidney disease treatment. *Evid Based Complement Alternat Med*. 2020;2020(1):9719872. doi:10.1155/2020/9719872