Journal of Pharmaceutical and Medical Research
Print ISSN: 2663-1954
Online ISSN: 2663-1962
DOI: https://doi.org/10.61784/jpmr3037

ANALYSIS OF PRESCRIPTION PATTERNS OF TRADITIONAL CHINESE MEDICINE FORMULAS FOR STROKE TREATMENT

XiaoMiao Gui[1,2], XinQi Hu[2], KeFan Yuan[3], RuiNa Wang[2*]
[1]Institute of Scientific and Technical Information of China, Beijing 100038, China.
[2]School of Public Health, Hubei University of Medicine, Shiyan 442000, Hubei, China.
[3]Xiantao Hospital of Traditional Chinese Medicine, Xiantao 433000, Hubei, China.
Corresponding Author: RuiNa Wang, Email: 2317517175@qq.com

Abstract: Stroke is a leading cause of death and disability worldwide. Despite the extensive use of traditional Chinese medicine (TCM) formulas in stroke treatment, systematic analyses of their multi-drug synergy and compatibility patterns remain scarce. This study leverages knowledge graph technology to integrate multi-source TCM formula data and constructs a knowledge graph for stroke-related TCM prescriptions. By combining frequency analysis, cluster analysis, and association rule mining, the study systematically uncovers medication patterns. A total of 1,403 validated formulas were included, identifying nine high-frequency herbs (frequency ≥ 200), such as Saposhnikoviae Radix and Glycyrrhizae Radix. These were categorized into four synergistic clusters. Five strong association rules were identified (e.g., "Ligustici Rhizoma and Almond → Ephedrae Herba" and "Scutellaria Baicalensis and Saposhnikoviae Radix → Glycyrrhizae Radix"). The utility of the knowledge graph in multidimensional retrieval and intelligent reasoning was validated. This study provides data support and a methodological paradigm for the standardized application and modernization of TCM in stroke treatment.
Keywords: Knowledge graph; Traditional Chinese medicine formulas; Prescription pattern; Stroke

1 INTRODUCTION

Stroke poses a major challenge to global public health due to its high disability rate and long-term rehabilitation needs[1], demanding more precise clinical interventions. Owing to their multi-target and multi-pathway synergistic effects, TCM formulas have unique advantages in stroke therapy[2]. However, the formulation principles often rely on empirical knowledge and lack systematic, data-driven validation[3]. Traditional studies based on single statistical methods have been inadequate in revealing hidden associations and synergistic modes among herbs[4], limiting the precision medicine potential of TCM[5].
Knowledge graphs, as an emerging tool for knowledge representation, integrate heterogeneous data through graph structures to visualize complex relationships among entities such as herbs, formulas, diseases, and symptoms[6]. They support intelligent reasoning and pattern discovery, offering new possibilities for the modernization of TCM[7]. In this study, we constructed a knowledge graph using Neo4j for stroke-related TCM prescriptions and applied frequency analysis, clustering, and association rule mining to systematically analyze medication patterns. The goal is to provide scientific evidence for clinical decision-making and new drug development.

2 MATERIALS AND METHODS

2.1 Data Sources and Preprocessing

Data were sourced from the *Encyclopedia of Chinese Medicine Prescriptions* and the TCM Formula Database (https://www.piccc.com/), covering classical texts and modern clinical prescriptions. A total of 1,692 raw records were collected through manual extraction and web scraping. The following preprocessing steps were applied:
Data cleaning: Duplicate entries and records lacking key information (e.g., formula name, composition) were removed, retaining 1,403 valid entries.
Standardization: Herb names were standardized based on the *Pharmacopoeia of the People's Republic of China (2020 edition)* (e.g., "Nanxing" standardized to "Tian Nanxing"), and traditional units (e.g., "liang," "qian") were converted to grams.
Structuring: The formula compositions were split into individual herbs using Python regular expressions to generate a structured CSV dataset.

2.2 Knowledge Graph Construction

Tools: Neo4j (v4.4.6) was used for graph data storage and management, and data import and querying were performed using the py2neo library.

Entity and relationship definitions: Four types of entities were defined: formulas, herbs, diseases, and symptoms. Formula attributes included name, usage, and source; herb attributes included dosage. Relationships included “COMPOSES” (herb → formula) and “TREATS” (formula → symptom/disease).
Construction process: Symptom entities were extracted via rule-based matching, assisted by a symptom dictionary based on *Chinese Symptomatology*. The final dataset was imported into Neo4j to complete the graph construction.

2.3 Analytical Methods for Medication Patterns

Frequency analysis: Herbs with occurrence frequency ≥ 200 were identified as high-frequency herbs.
Cluster analysis: Using SPSS 26.0, Euclidean distances between high-frequency herbs were calculated and hierarchical clustering was performed using the between-groups linkage method.
Association rule mining: The Apriori algorithm was applied with a minimum support of 50 and minimum confidence of 85% to discover frequent co-occurrence patterns among herbs.

3 RESULTS

3.1 Knowledge Graph Construction

The resulting stroke TCM knowledge graph comprised 7,451 nodes (diseases in blue, symptoms in red, formulas in orange, herbs in purple) and 9,976 relationships (COMPOSES and TREATS). For example, the “Ligustici Rhizoma Decoction” node clearly shows its herbal components (e.g., Ligustici Rhizoma, Saposhnikoviae Radix, Glycyrrhizae Radix) and target symptoms (e.g., hemiplegia, headache) (Figure 1).

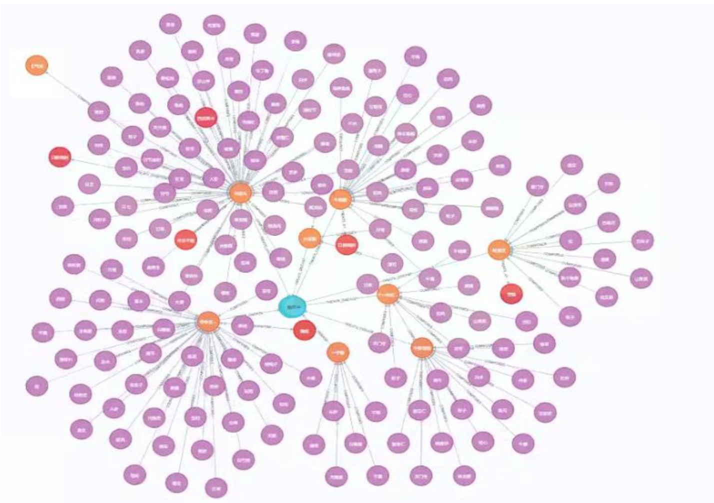

Figure 1 Knowledge Graph of Traditional Chinese Medicine Prescriptions for Stroke

3.2 Medication Pattern Analysis

3.2.1 High-frequency herbs

Nine herbs appeared ≥200 times in the dataset (Table 1): Saposhnikoviae Radix, Glycyrrhizae Radix, Ligustici Rhizoma, Ephedrae Herba, Angelicae Sinensis Radix, Ginseng Radix, Aconiti Lateralis Radix Praeparata, Angelicae Pubescentis Radix, and Notopterygii Rhizoma. These herbs primarily exhibit wind-dispelling, meridian-unblocking, and qi-invigorating effects, aligning with the stroke pathogenesis of “wind-phlegm and blood stasis with qi deficiency.” Saposhnikoviae Radix, ranked first, is traditionally used for wind-related conditions and has shown neuroprotective effects via anti-inflammatory and microcirculation improvement mechanisms. Glycyrrhizae Radix is frequently used for its harmonizing and detoxifying properties.

Table 1 Frequency of High-Frequency Drug Use (frequency ≥ 200 times)

Drug Name	Frequency (times)
Saposhnikoviae Radix	443
Glycyrrhizae Radix	357
Ligustici Rhizoma	316
Ephedrae Herba	290
Angelicae Sinensis Radix	288
Ginseng Radix	257
Aconiti Lateralis Radix Praeparata	240
Angelicae Pubescentis Radix	236
Notopterygii Rhizoma	207

3.2.2 Cluster analysis

Using SPSS, the nine high-frequency herbs were grouped into four clusters based on functional similarity:
Cluster A1: Saposhnikoviae Radix
Cluster A2: Glycyrrhizae Radix, Ligustici Rhizoma, Ephedrae Herba, Angelicae Sinensis Radix
Cluster A3: Dioscoreae Hypoglaucae Rhizoma
Cluster A4: Ginseng Radix, Aconiti Lateralis Radix Praeparata, Angelicae Pubescentis Radix, Notopterygii Rhizoma, Arisaema, Cinnamomi Core, Moschus, Gastrodiae Rhizoma, Atractylodes Macrocephala, Pinelliae Rhizoma, Asari Radix, Rhizoma Typhonii, Cinnamomi Cortex, Scutellaria Baicalensis, Ligustici Rhizoma

3.2.3 Association rule mining

Five strong association rules were identified (Table 2), such as:
These rules reflect therapeutic synergies such as "invigorating blood–promoting lung qi–dispelling pathogens," offering evidence for prescription decisions in complex stroke cases with external syndromes.

Table 2 Core Drug Groups and Association Rules

Antecedent Herbs	Consequent Herb	Support	Confidence (%)
Ligustici Rhizoma + Almond	Ephedrae Herba	67	90.54
Almond + Glycyrrhizae Radix	Ephedrae Herba	59	85.51
Scutellaria baicalensis + Saposhnikoviae Radix	Glycyrrhizae Radix	56	86.15
Ligustici Rhizoma + Almond + Glycyrrhizae Radix	Ephedrae Herba	50	92.73
Almond + Glycyrrhizae Radix + Ephedrae Herba	Ligustici Rhizoma	50	86.44

The combination of Ligustici Rhizoma (activating blood circulation and promoting qi movement) and Almond (relieving cough and asthma) results in a 90.54% occurrence probability of Ephedrae Herba (inducing sweating and relieving exterior syndrome). This rule reflects the synergistic effect of "activating blood circulation - ventilating lung - relieving exterior syndrome", which is applicable to the complicating syndrome of lung qi stagnation or external contraction of wind-cold after stroke. When Almond is combined with Glycyrrhizae Radix (harmonizing various medicines), the occurrence probability of Ephedrae Herba reaches 85.51%, indicating that Glycyrrhizae Radix may enhance the synergistic effect between Almond and Ephedrae Herba through alleviating spasm and detoxification. The combination of Scutellaria baicalensis (clearing heat and drying dampness) and Saposhnikoviae Radix (dispelling wind and relieving exterior syndrome) leads to an 86.15% occurrence probability of Glycyrrhizae Radix, reflecting the compatibility pattern of "clearing heat - dispelling wind - harmonizing". After the combination of Ligustici Rhizoma, Almond and Glycyrrhizae Radix, the occurrence probability of Ephedrae Herba is as high as 92.73%, which further verifies the multi-dimensional synergistic effect of "activating blood circulation - ventilating lung - harmonizing - relieving exterior syndrome". When Almond, Glycyrrhizae Radix and Ephedrae Herba are combined, the occurrence probability of Ligustici Rhizoma is 86.44%, embodying the therapeutic idea of strengthening blood circulation and qi promotion on the basis of "ventilating lung - harmonizing - relieving exterior syndrome".

As shown in Figure 2, it displays the correlation network among multiple medicines in stroke TCM prescriptions, visually presenting the compatibility relationships and their strengths between medicines. The nodes in the figure represent TCM medicines, and the edges represent the compositional relationships between medicines. The color, transparency, and label of nodes are displayed according to the degree value: the larger the degree value, the darker the node color, the smaller the transparency, and the larger the label font size, indicating that the corresponding TCM medicine is more important.

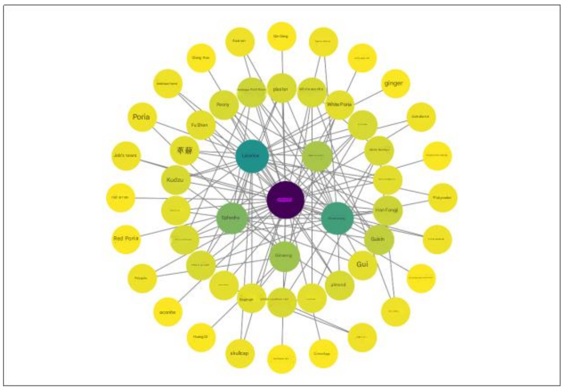

Figure 2 Drug Network Association

3.3 Validation of Knowledge Graph Application

Cypher, the query language used in Neo4j, enables efficient retrieval of formula components. For instance, a Cypher query can extract all herbs in the "Yizisan" formula, visually illustrating its structure and potential mechanism. Cypher also supports multi-tier relational queries, enhancing traceability and interpretability in TCM research.

4 DISCUSSION

4.1 Advantages and Challenges of Knowledge Graphs

The constructed knowledge graph breaks the isolation of traditional analyses by integrating multi-source data and visualizing complex herbal-formula-disease-symptom relationships. Its advantages include:
Explicit Relationships: The graph structure visualizes hierarchical links, such as "Saposhnikoviae Radix → Qi Deficiency Syndrome → Qi-Tonifying and Blood-Activating Therapy," offering new insights into multi-target mechanisms.
Intellignt Inference: Graph algorithms (e.g., link prediction, community detection) can suggest novel herbal combinations for formula development.
Challenges remain:
Data Standardization: TCM terms often have multiple variants (e.g., different processing methods for the same herb), requiring a comprehensive standard dictionary.
Natural Language Processing: Rule-based entity extraction struggles with complex symptom expressions, limiting extraction accuracy.

4.2 Scientific and Clinical Significance of Findings

High-frequency herbs such as Saposhnikoviae Radix and Glycyrrhizae Radix align with classical TCM theories like "treat wind by treating blood" and "harmonize prescriptions." Cluster A2 (Glycyrrhizae Radix, Ligustici Rhizoma, Ephedrae Herba, Angelicae Sinensis Radix) reflects the importance of qi and blood tonification in stroke treatment. The rule "Ligustici Rhizoma + Almond → Ephedrae Herba" illustrates how blood-invigorating and lung-relieving herbs enhance the efficacy of exterior-releasing herbs, guiding treatment for wind-cold syndromes in stroke.

5 CONCLUSION

This study applied knowledge graph technology to systematically explore TCM prescription patterns in stroke treatment. Nine high-frequency herbs were identified as core components with effects on wind expulsion, qi reinforcement, and blood circulation. Cluster and association rule analyses revealed synergistic groupings and compatibility principles. The knowledge graph offers a novel paradigm for the standardized and precise application of TCM in stroke and provides a foundation for future research integrating clinical data and advanced NLP techniques to modernize TCM practices[8-9].

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Parisa Fallahtafti, Amirhossein Habibzadeh, Negar Ghasemloo, et al. Stroke burden in North Africa and the Middle East, 1990-2021: an analysis based on the global burden of disease study. BMC neurology, 2025, 25 (1), 277-277.
[2] Yin Xinyi, Li Shutang, Wang Junwei, et al. Research progress of active compounds from traditional Chinese medicine in the treatment of stroke. European journal of medicinal chemistry, 2025, 291(5), 117599.
[3] Xu Min, Wu RuiXia, Li XiaoLi, et al. Traditional medicine in China for ischemic stroke: bioactive components, pharmacology, and mechanisms. Journal of integrative neuroscience, 2022, 21 (1), 26-26.
[4] Zhao Jianfeng, Zhang Wei, Yu Hai. Prospects and Analysis of Traditional Chinese Medicine Standards Through the Transition of Chinese Pharmacopeia. Journal of Pharmaceutical Innovation, 2022, 18 (2), 349-355.
[5] Zhang Peng, Zhang Dingfan, Zhou Wuai, et al. Network pharmacology: towards the artificial intelligence-based precision traditional Chinese medicine. Briefings in bioinformatics, 2023, 25 (1), 1-12.
[6] Zhang Zheng, Wu Hengyang, Wang Na. A Knowledge-Enhanced Disease Diagnosis Method Based on Prompt Learning and BERT Integration. Journal on Artificial Intelligence, 2025, 7 (1), 17-37.
[7] Gao Yanjun, Li Ruizhe, Emma Croxford, et al. Leveraging Medical Knowledge Graphs Into Large Language Models for Diagnosis Prediction: Design and Application Study. JMIR AI, 2025, 4(1), e58670: p1-p17.
[8] Alain García Olea, Ane G Domingo Aldama, Marcos Merino, et al. The Application of Deep Learning Tools on Medical Reports to Optimize the Input of an Atrial-Fibrillation-Recurrence Predictive Model. Journal of Clinical Medicine, 2025, 14 (7), 2297-2297.
[9] Priyanka Khalate, Shilpa Gite, Biswajeet Pradhan, et al. Advancements and gaps in natural language processing and machine learning applications in healthcare: a comprehensive review of electronic medical records and medical imaging. Frontiers in Physics, 2024, 12(12), 1445204-1445204.