



## OPEN ACCESS

## EDITED BY

Jorge E. Camargo,  
National University of Colombia, Colombia

## REVIEWED BY

Carlos Alberto Pereira De Oliveira,  
Rio de Janeiro State University, Brazil  
Xin Chen,  
Tongji University, China

## \*CORRESPONDENCE

HongPu Hu  
✉ hu.hongpu@imicams.ac.cn

RECEIVED 21 December 2024

ACCEPTED 11 July 2025

PUBLISHED 01 August 2025

## CITATION

Wang Y, Wan Y and Hu H (2025) Multi-task meta-attention network for traditional Chinese medicine diagnostic recommendation.  
*Front. Public Health* 13:1549679.  
doi: 10.3389/fpubh.2025.1549679

## COPYRIGHT

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

# Multi-task meta-attention network for traditional Chinese medicine diagnostic recommendation

YingShuai Wang<sup>1</sup>, YanLi Wan<sup>1</sup> and HongPu Hu<sup>1,2\*</sup>

<sup>1</sup>Institute of Medical Information/Library, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, <sup>2</sup>School of Marxism, School of Humanities and Social Sciences, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

**Background:** With the continuous growth of medical data and advancements in medical technology, there is an increasing need for personalized and accurate assisted diagnosis. However, implementing recommendation systems in healthcare presents numerous challenges, requiring further in-depth research.

**Objective:** This study explores the application of recommendation technology in smart healthcare. The primary goal is to design a deep learning model that effectively integrates medical knowledge for improved diagnostic support.

**Methods:** We first developed a feature engineering process tailored to the characteristics and requirements of medical data. This process involved data preparation, feature selection and transformation to extract informative features. Subsequently, a knowledge-matching deep learning model was designed to analyze and predict medical data. This model enhances evaluation metrics through its capabilities in nonlinear fitting and feature learning.

**Results:** Experimental results indicate that our proposed deep learning model achieves an average improvement of +2.7% in the core metrics Hits@10 compared to baseline models in the Traditional Chinese Medicine (TCM) clinical recommendation scenario. The model effectively processes medical data, providing accurate predictions and valuable insights to support clinical decision-making.

**Conclusion:** This research provides significant support for the advancement and application of smart medical technology. Our deep learning model demonstrates strong potential for medical data analysis and clinical decision-making. It can contribute to enhanced healthcare quality and efficiency, ultimately advancing the medical field.

## KEYWORDS

deep learning, feature engineering, recommendation technology, traditional Chinese medicine, clinical decision support, smart healthcare

## 1 Introduction

With the advancements in big data and artificial intelligence technologies, smart healthcare has gradually emerged as a new trend. Recommendation technology, with its strengths in personalized information delivery and decision support, offers new possibilities for the advancement of smart healthcare. By analyzing medical and health data, recommendation technology can provide personalized advice and services to both doctors

and patients, thereby enhancing the accuracy and efficiency of medical decision-making. Xiyu Shen (1) applied a deep learning-based hierarchical attention network to construct models for doctors and patients using consultation records. This approach strengthened the interaction between the doctor and patient vectors, assigned higher weights to doctors and patients with similar conditions, and calculated the doctor's recommendation value. Yinghua Wu (2) explored the application scenarios of ChatGPT and other next-generation AI technologies in smart healthcare, including areas such as medical management, medical technology for doctors, and patient healthcare. Quan Chen (3) investigated the application of ChatGPT and other next-generation AI technologies in smart healthcare. He developed a transfer relationship model between abnormal signs and drugs based on a graph neural network and integrated abnormal sign information to enable accurate drug recommendations. Recommendation systems in the medical and healthcare fields are continuously evolving, encompassing areas such as medication recommendations, prescription suggestions, assisted diagnosis, precision treatment, and health monitoring. Xiaojing Hu (4) designed a research questionnaire to assess patients' willingness to recommend, using the net recommendation value, and employed multivariate logistic regression analysis to explore the factors influencing patients' willingness to recommend. Fang Wang (5) analyzed the current state of health management services both domestically and internationally, and explored pathways for constructing intelligent prescriptions based on disease risk. Pinsky (6) analyzes the characteristics of standard datasets and pre-trained models in the medical field, integrates them with practical applications of artificial intelligence, and explores the requirements for large model development to promote the innovative advancement of medicine. Large Language Models (LLM) have archived initial success in medicine. Liu et al. used pre-training and fine-tuning to create a large language model-based diagnostic system, to assist physicians in formulating diagnosis and treatment plans (7). Li et al. developed a system that uses deep learning and language models to aid diabetes care and retinopathy screening, improving patient follow-up (8).

Although recommendation technology and machine learning demonstrate great potential in the field of smart healthcare, they still face numerous challenges. Firstly, the quality and scale of healthcare data are inconsistent, and the data often exhibit heterogeneity and incompleteness, which complicates model construction. Additionally, existing machine learning have limitations in medical applications, such as a lack of sufficient interpretability, difficulty in addressing the diversity of clinical scenarios, inadequate generalization ability, and the need for improved reliability and stability in actual clinical decision-making. The unique characteristics of medical field also pose numerous challenges for the development of large language models, including the issue of hallucinations generating inaccurate information, high deployment costs to ensure precision, lack of currency making it difficult to reflect the latest medical advancements, bias and toxicity potentially leading to unfair treatment recommendations, and privacy and security concerns when handling personal health information. The aim of this study is to conduct an in-depth exploration of the application of recommendation technology in smart healthcare, with the goal of providing theoretical guidance and technical support to advance the development of smart healthcare, and to further enhance the intelligence and humanization of healthcare services.

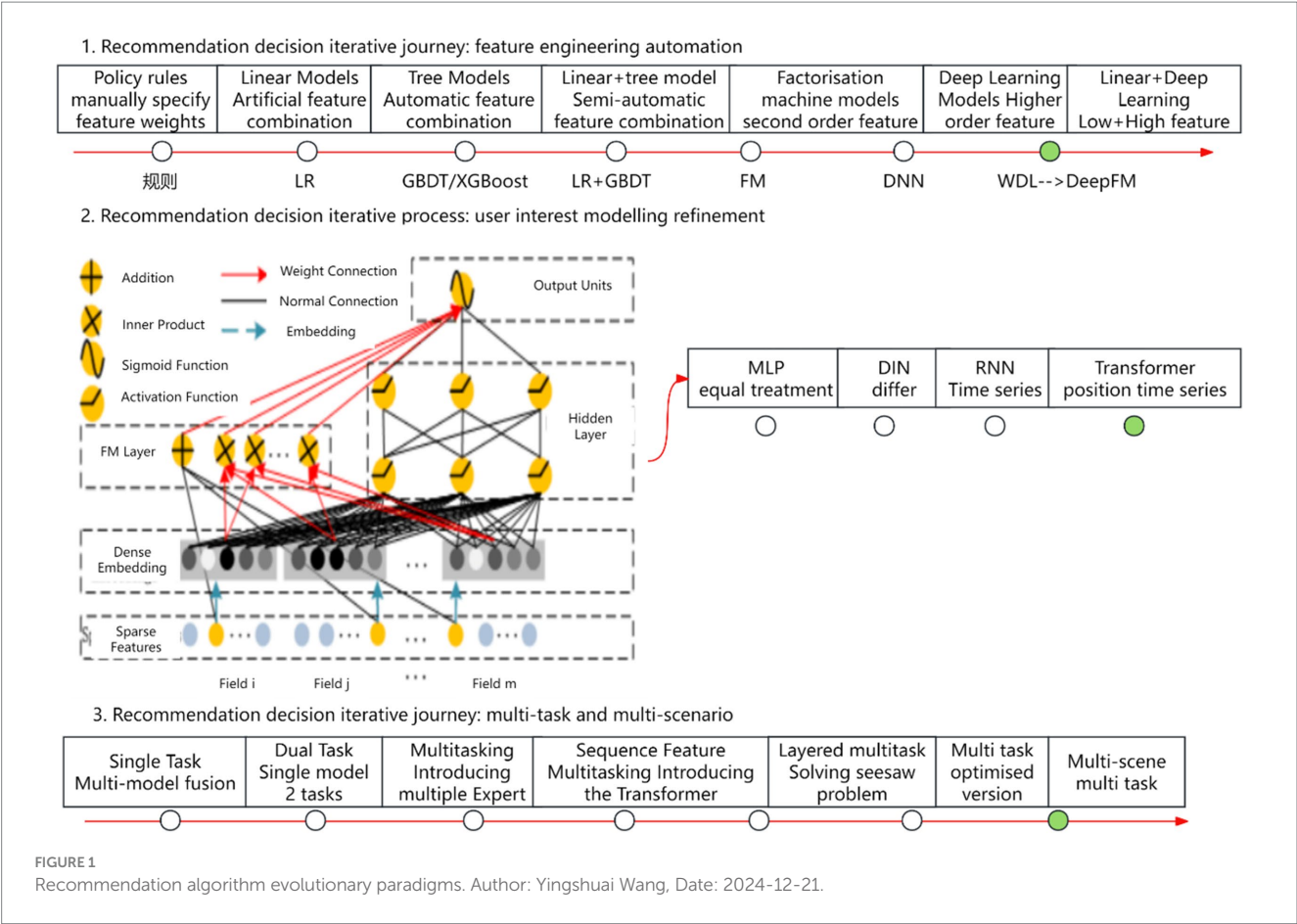
Recommendation algorithms have evolved along three primary paradigms, as illustrated in Figure 1. The first paradigm involves the automation of feature engineering, beginning with manually defined feature weights and strategy rules, progressing through logistic regression (LR) (9), gradient boosting decision trees (GBDT) (10, 11), and factorization machines (FM) (12). In recent years, deep neural networks (DNN) (13) have become dominant, owing to their remarkable expressiveness and flexibility. Models such as Wide & Deep (W&D) (14) exemplify this evolution by effectively combining low-order and high-order features. The second paradigm focuses on refining user interest modeling. It begins with the Multi-Layer Perceptron (MLP) (15), where user interests are treated uniformly. This is followed by the Deep Interest Network (DIN) (16), which differentiates user interests, and culminates in the Transformer model (17), incorporating both location and temporal sequences. The third paradigm centers on multi-task, multi-scenario modeling. It evolved from single-task, multi-model fusion to multi-task learning with the introduction of multiple Experts (18), then to sequence-based multi-tasking using the Transformer, and further to multi-layer multi-tasking. Ultimately, it leads to multi-scenario, cross-domain recommendation.

Personalized recommendation is a feasible solution for smart healthcare. However, existing recommendation technology is primarily applied in the commercial sector, whereas recommendation systems in the medical field differ significantly in terms of objectives, scenarios, and data. For instance, e-commerce recommendations are based on user purchase history to establish relationships between users and products, focusing on building user profiles and product profiles. In medical scenarios, however, patient evaluations or doctor's consultation information are typically used for modeling, and the data scale is relatively small. As a result, commodity recommendation models cannot be directly applied, presenting unique challenges and opportunities for exploration in modeling and design. Recommendation research in the medical field primarily relies on shallow deep learning techniques for model construction, with recommendation strategies largely based on statistical methods or machine learning mechanisms. These approaches typically use a single method for interactive vector calculations. Drawing from the modeling approach used in deep learning for commodity recommendations, and considering the unique features of recommendations in the medical field, the research framework is designed.

Most existing studies focus on data or knowledge integration, lacking deep collaboration among multi-source heterogeneous data, medical domain knowledge, and clinical needs. This limits model interpretability and clinical applicability. In complex clinical scenarios, flexible utilization of data and knowledge is essential. Inspired by the cognitive process of physicians considering patient profiles and guidelines, we propose a data-driven, knowledge-guided feature engineering to enhance diagnostic accuracy. Additionally, a multi-task learning algorithm based on meta-attention mechanisms is developed to prioritize relevant information, along with an automated recommendation framework mimicking cognitive management. These efforts aim to establish a dynamic data-knowledge-business model to improve disease diagnosis.

In summary, the contributions of this study are as follows:

- (1) Using real medical case data from Traditional Chinese Medicine, we develop a systematic feature engineering



framework. This includes the design of both high- and low-order features, knowledge representation features, and various feature interaction methods. These approaches help to fully explore the information contained in the original data.

- (2) We propose a deep matching neural network model that integrates TCM knowledge and modifies the loss function. Furthermore, an interaction mechanism between data and knowledge is introduced, enabling the model to better understand the underlying knowledge.
- (3) Feature engineering and model training are automated and applied to real-world scenarios. Compared to the baseline, both offline evaluation metrics and online results show significant improvements.

## 2 Methods

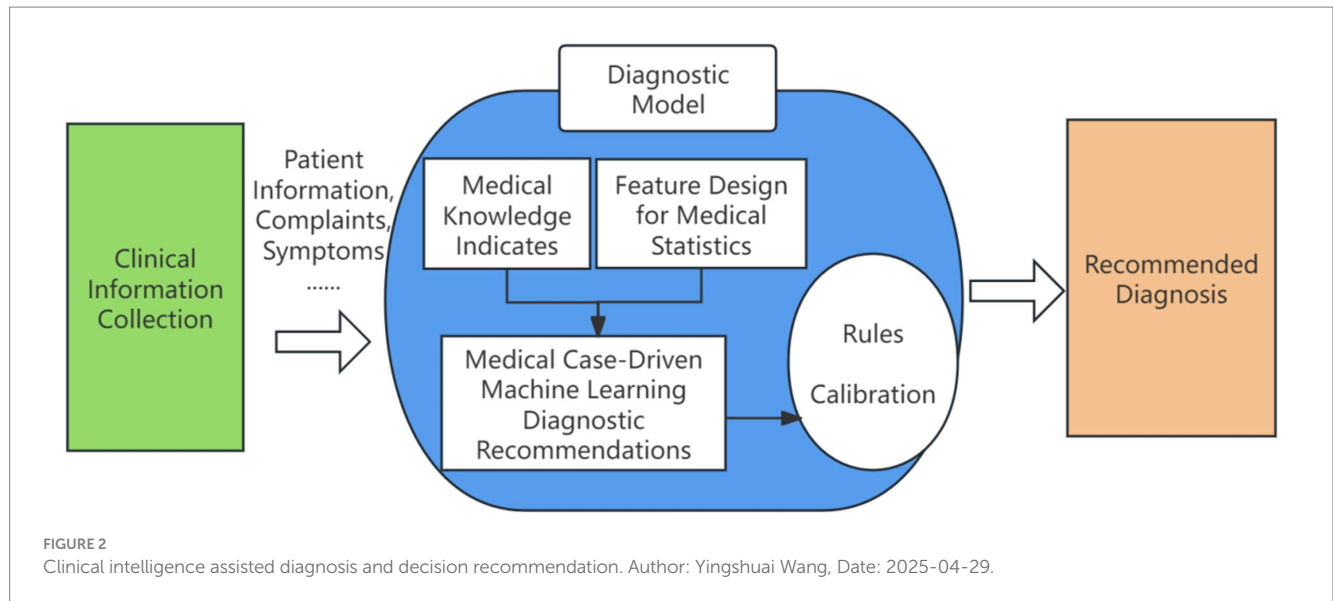
The recommendation framework incorporates automated feature engineering and model optimization, enabling its application across a range of tasks in the medical domain. These include medical case suggestions, prescription recommendations, drug recommendations, disease diagnosis, evidence element identification, and more. An overview of the intelligent clinical decision support and recommendation in Traditional Chinese Medicine is illustrated in Figure 2.

## 2.1 Feature engineering

### 2.1.1 Original features of Chinese medicine

This section outlines the key features derived from Traditional Chinese Medicine that serve as the foundation of the recommendation system. These include various types of information such as patient and physician details, symptoms, diagnostic findings, treatment plans, and medical case records, all of which help enhance to the accuracy and personalized recommendation.

- (1) Physician information: physician name, consultation department.
- (2) Patient information: patient name, gender, date of birth, age.
- (3) Objective environment: date of consultation, season of consultation.
- (4) Symptom information: chief complaint, history of present illness, carved symptoms, standard symptoms, source of symptoms, tongue, moss, pulse.
- (5) Diagnostic information: Chinese Medicine Diagnosis, Western Medicine Diagnosis, Chinese Medicine Evidence, source of evidence.
- (6) Treatment Information: Treatment principles and methods, formula name, formula composition, acupuncture and herbal therapy details, number of visits, and treatment results.
- (7) Medical case information: medical case type, medical case name, and medical case source.



## 2.1.2 Statistical features of herbs

### 2.1.2.1 Co-occurrence probability

The mathematical expression for co-occurrence probability is shown in Equation 1.

$$P(a|b) = \frac{\text{count}(a,b)}{\text{count}(a,B)} \quad (1)$$

$\text{count}(a,b)$  indicates the number of times a symptom and an evidence element appeared in the same medical case, while  $\text{count}(a,B)$  indicates the number of times a symptom and any evidence element appeared in the same medical case.

### 2.1.2.2 Confidence

The mathematical expression of confidence level is shown in Equation 2.

$$\text{confidence}(A \Rightarrow B) = P(B|A) = \frac{\text{count}(A \cap B)}{\text{count}(A)} \quad (2)$$

$\text{count}(a,b)$  indicates the number of times a symptom and an evidence element occurred in the same medical case in all medical cases, while  $\text{count}(a,B)$  indicates the number of times a symptom and any evidence element occurred in the same medical case in all medical cases.

### 2.1.2.3 Degree of support

The support of an association rule  $A \Rightarrow B$  indicates the ratio of the number of elements in the intersection of the item set  $A$  and the item set  $B$  to the number of elements in the total transaction set  $\text{count}(T)$ . The degree of support is used to evaluate the importance of association rules and indicates the universality of the current rule among all rules. The mathematical expression of support is shown in Equation 3.

$$\text{support}(A \Rightarrow B) = \frac{\text{count}(A \cap B)}{\text{count}(T)} \quad (3)$$

### 2.1.2.4 TFIDF

TFIDF (Term Frequency-Inverse Document Frequency) suggests that the importance of a word is positively correlated with its frequency in a document and negatively correlated with its frequency in the whole corpus. In the task of recommending 'evidence elements' in Traditional Chinese Medicine (TCM), it reflects the importance of 'evidence elements' to 'symptoms' as calculated by Equation 4.

$$TF = \frac{n_{ij}}{\sum_k n_{kj}} \quad (4)$$

Where  $n_{ij}$  indicates the number of occurrences of the current word in the document, and  $\sum_k n_{kj}$  means the sum of the occurrences of all words in the document. The mathematical expression of Inverse Document Frequency is shown in Equation 5.

$$IDF = \log \frac{N}{n_i + 1} \quad (5)$$

Where  $N$  denotes the total number of documents in the corpus and  $n_i$  denotes the number of documents that contain a particular word.

## 2.1.3 TCM domain knowledge features

TCM domain features are derived from the original features and include knowledge based on time, gender and age, Chinese and Western medicine diagnoses, and symptoms.

- (1) Knowledge derived based on time: year of birth, month of birth, Heavenly Stem and Earthly Branches, year calculates the five elements according to the innate weak organs, Heavenly Stem calculates the five elements of the year's fortune, plus excess or less than the year's fortune, the year and month of birth project the SITIAN in the spring, the main QI and



politeness, and the SITIAN in the spring, the main QI and politeness to get the innate constitution.

- (2) Based on the knowledge derived from gender age: after the female seven male eight points, the number of years combined with gender, divided by 8 arithmetic, 001 for female 0–7 years old, 002 for female 8–14 years old, 101 for male 0–8 years old, 102 for male 9–16 years old, and so on.
- (3) Based on knowledge derived from Chinese and Western medicine diagnosis: This includes site labeling (e.g., acute disease = 0, chronic disease = 1), a list of disease names from TCM, and the Western medicine disease classification system.
- (4) Symptom-based derived knowledge: common evidence elements, common evidence element frequency, common evidence element categories (a = disease site, b = essential substance, c = disease evil, d = pathological state, e = connecting word).

## 2.2 Model design

### 2.2.1 Meta-learning network for feature fusion

To better capture scene-specific information, a meta-learning fine-grained attention network is introduced. This network adapts to

the data distribution of various scenes via multi-objective learning, generating distinct scene representations while mitigating the impact of sample distribution variations on model performance. The structure of the meta-learning network is illustrated in Figure 3.

The features of each scenario are primarily categorized into three types: patient features, medicine features, and context features. Most of these features are sparse and high-dimensional in the CTR prediction task, and are typically transformed into low-dimensional dense vectors using embedding techniques. These vectors are randomly initialized, updated during model training, and concatenated with dense input features to form a comprehensive feature vector, as shown in Equation 6.

$$e_i = [e_u, e_i, e_s] \quad (6)$$

Where  $e_u$  means patient embedding,  $e_i$  indicates medicine embedding,  $e_s$  denotes scenario embedding.

The higher-order features of the patient are extracted using the self-attention mechanism with the Transformer decoder, where the decoding network comprises a Multi-head Self-Attention Network (MSA) and a Feed-Forward Network (FFN). The Self-Attention network (SAN) is implemented using Scaled Dot-Product Attention (SDPA) as defined in Equation 7.

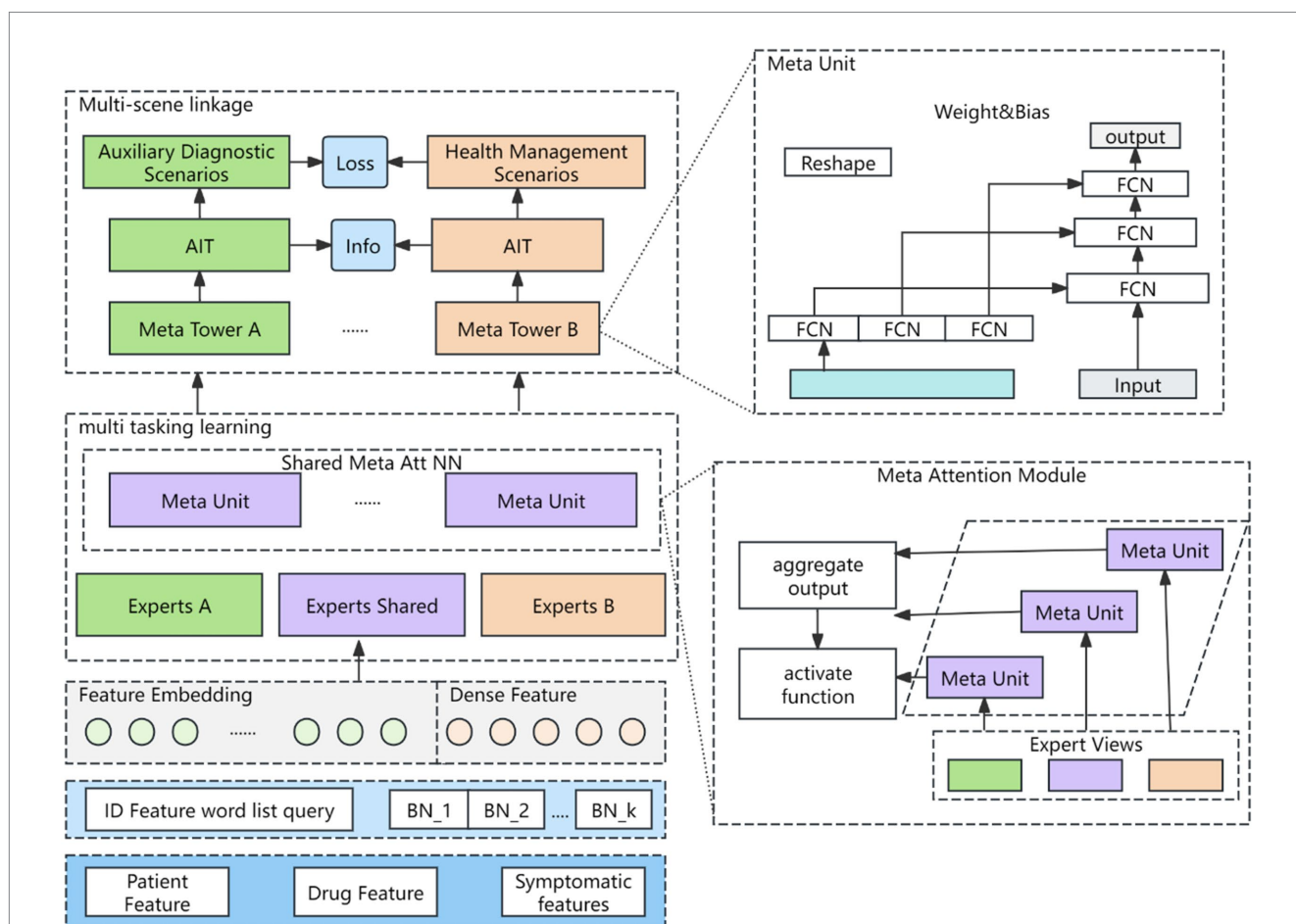


FIGURE 3  
Meta-learning network. Author: Yingshuai Wang, Date: 2024-12-21.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d}}\right)V \quad (7)$$

Where  $Q, K, V$  represent Queries, Keys and Values respectively,  $d$  denotes the dimension of Queries, Keys and Values. We use a multi-head attention mechanism to capture the relationship between the query matrix and the key matrix from different perspectives. The mathematical expression of the multi-head attention mechanism is shown in Equation 8.

$$S = \text{Multihead}(Q, K, V) = \text{Concatenate}(\text{head}_1, \text{head}_2, \dots, \text{head}_h)W^H \quad (8)$$

The attention expression for the  $i$ -th head is shown in Equation 9.

$$\text{head}_i = \text{Attention}(Q_i, K_i, V_i) \quad (9)$$

Where  $W^H \in R^{d \times d}$  is weight matrix,  $h$  is the number of attention heads. Furthermore, this study combines FFN and MSN to enhance the model's characterization. The key concepts are outlined in Equations 10, 11.

$$\text{feature}_{\text{high}} = \text{FFN}(S) \quad (10)$$

$$\text{FFN}(x) = \text{Relu}((xW_1 + b_1)W_2) + b_2 \quad (11)$$

Where  $x \in S, d_k = \frac{d}{h}, W_1 \in R^{d \times d_k}, W_2 \in R^{d_k \times d}, W_1$  and  $W_2$  are weight

matrices,  $b_1 \in R^{d_k}, b_2 \in R^d, b_1$  and  $b_2$  are the bias terms. After the features are characterized as described above, they are dynamically fused using a Squeeze-and-Excitation Network (SENET) (19). The SENET structure is shown as Figure 4.

The mathematical description of the network is shown in Equations 12, 13.

$$s = \sigma(\text{MLP}(z, W)) \quad (12)$$

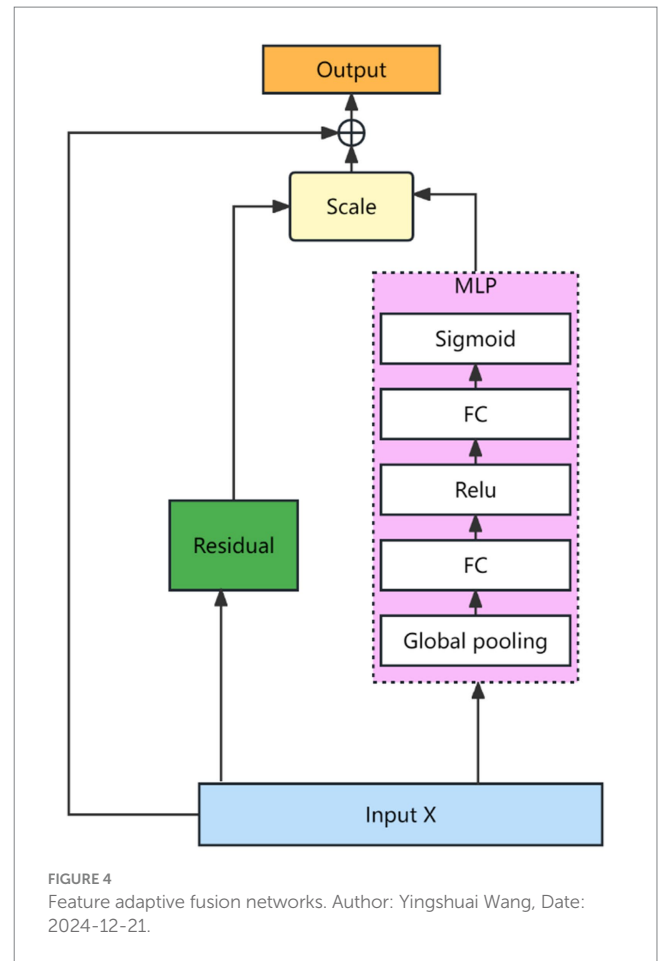
$$x_{l+1} = x_l + s \quad (13)$$

Where  $z$  denotes input features,  $W$  is the weight parameter,  $\text{MLP}$  denotes a multiple layer perceptual machine and  $\sigma$  is a nonlinear mapping operator. The Meta Unit module is calculated as in Equations 14, 15.

$$y = \text{Relu}(W_{\text{meta}}x + b_{\text{meta}}) \quad (14)$$

$$W_{\text{meta}} = \text{Reshape}(W * \text{embedding}_{\text{scene}} + b) \quad (15)$$

Where  $\text{embedding}_{\text{scene}}$  represents the embedding of the scene type, after MLP, we get the scene-specific representation, and generate the weight and bias by reshape to form the Meta Unit of the network. The Meta Unit parameters are different in different scenes. The input



is  $x$ , the output is  $y$ . The original fully connected network is  $y = \text{Relu}(Wx + b)$ , when the network training is completed,  $W$  and  $b$  are fixed, for all the samples,  $W$  and  $b$  are the same. The Meta Unit dynamically changes the  $W$  and  $b$  according to the samples.  $y = \text{Relu}(W_{\text{meta}}x + b_{\text{meta}})$ , where  $W$  and  $b$  are related to the type of scene. The computational logic of Meta Attention module is shown in Equation 16.

$$\alpha_{Ai} = V^T \text{Meta}_A([E_i]), 1 \leq i \leq S + K \quad (16)$$

Where  $\alpha_{Ai}$  denotes the attention scores of the experts for task  $A$ , the expert group consists of  $S$  shared experts and  $K$  unique experts.  $\text{Meta}_A$  represents the shared Meta Unit for task  $A$ . The parameters change dynamically with different scenarios, and  $V^T$  are network learning parameters.  $\alpha_{Ai}$  are scalar values, and then normalized by an activation function softmax for the  $S + K$  scores. Scene information is explicitly incorporated through the attention meta-network in the multi-task attention weights, and the information of different scenes is captured when calculating the attention scores. The computational logic of the Meta Tower module is as follows: each layer of the multi-task learning tower is implemented as a fully connected network. The meta tower is constructed by cascading multiple meta units, and the parameters adapt dynamically with different scenes. This architecture enhances the model's ability to characterize specific scenes through the layered structure of the tower.

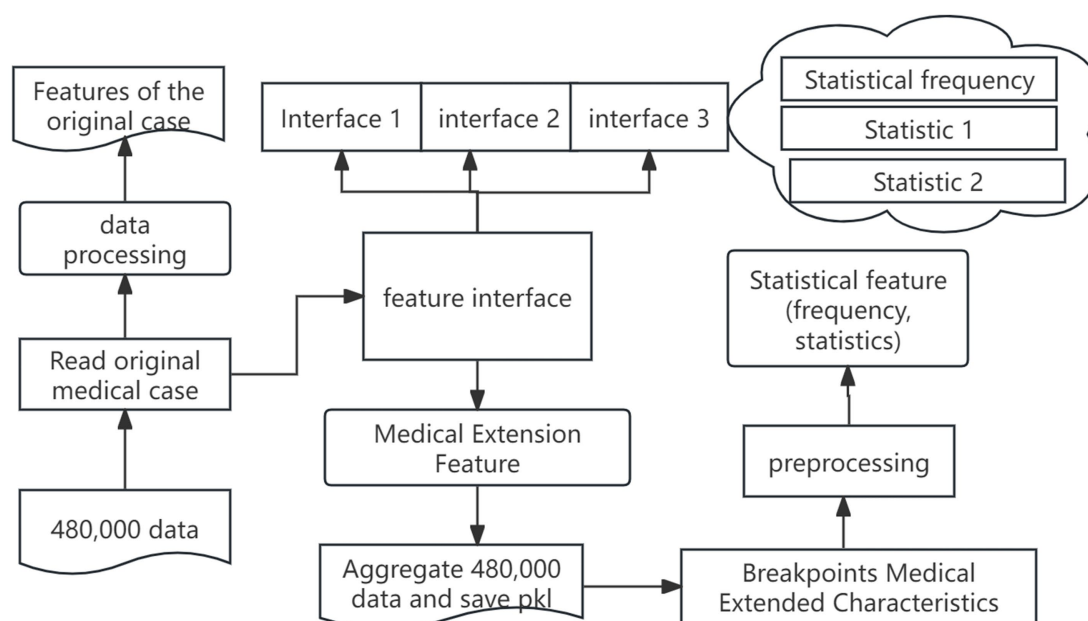


FIGURE 5  
Automated feature engineering. Author: Yingshuai Wang, Date: 2024-12-21.

### 2.2.2 Automated training framework

The automated training framework includes feature automation and recommendation model automation, and the feature automation architecture is shown in Figure 5.

The steps of feature automation are as follows: in the first stage, we perform raw-field processing by importing medical case from excel, handling missing values, and segmenting the data. Next, we extend the traditional Chinese medicine fields by transforming original symptom descriptions and integrating custom features via feature-interface. Finally, we unify all features into a standardized statistical input format, compute the necessary statistics, and serialize the extended feature set into PKL files, ensuring that the features stay continuously updated with the underlying data. The following outlines the key stages of the process, from data reading to model training, emphasizing the critical operations at each phase of the workflow.

#### 2.2.2.1 Data reading

Instantiate the data reading module and invoke the relevant functions to copy, shuffle, and sequentially retrieve the data.

#### 2.2.2.2 Model network

Firstly, abstract a general base network module. Secondly, carry out internal embedding, construct the network structure, design the loss function, and perform other related operations.

#### 2.2.2.3 Model training

Define the session graph, including the creation of the model and evaluation modules, the definition of the optimizer, and the input of score and loss into the evaluation module. Session graph computation involves tasks such as session initialization, executing `sess.run()` to calculate loss and evaluation metrics, and determining when to save the checkpoint.

## 3 Results

### 3.1 Experimental data

The data comes from the medical records of real patients. The following information is recorded in the case: the patient's personal information, information about the environment at the time of consultation, information about the description of the disease, and information about the doctor's analysis of the disease at the time of consultation and treatment. We screened 150,000 medical cases, selected 130,000 to generate the training set, and the remaining 20,000 to generate the test set. A sample example of the training data is shown in Table 1.

### 3.2 Comparison of methods

Text-CNN (20): Text-CNN is a multi-label framework that leverages convolution neural networks to construct model architectures.

MLP (15): Multiple layer Perceptron Machine is a baseline for deep learning and is widely used in recommendation systems.

MMOE (18): A model based on multi-task expert sharing, which can flexibly adjust the weights of experts for different tasks through gated networks.

Our model: The model proposed by us based on feature fusion of multi-task learning and meta-learning network, which can interact the information from different scenarios.

### 3.3 Evaluation metrics

The evaluation metrics used in this research are AUC, MR, MRR, Hits@10, and Unit Hit Rate.

TABLE 1 Examples of TCM samples.

Medicine Symptoms	Syndrome	Medical elements	Elements type	Disease name
Epigastric-pain, cracked-tongue, white-coating, stringy pulse	Deficiency of vital energy, congestion of heat and toxins.	zheng, qi, kui, xu, re, du, yong, sheng	b, b, d, d, c, c, d, d	Stomach ache
Nausea, vomiting, abdominal pain, vexation, fatigue, fat tongue, yellow moss, greasy moss, fine veins	Internal obstruction by wetness	Shi, xie, nei, zu	c, c, d	Vomiting
Generalized vomiting, excessive salivation, pale mouth, pale red tongue, white moss, sluggish pulse	Rebellious cold in the stomach	Wei, han, shang, ni	a, c, a, d	Vomiting
Weakness, Yellow urine, Stringy pulse	Stagnant heat in the liver meridian, blood stasis	gan jing, yu, re, yu	a, d, c, c	Liver accumulation
Tightness in chest, palpitation, shortness of breath, poor appetite, poor sleep, dry stools, dark red tongue, scanty coating	Double deficiency of the heart and spleen	xue, xin, xu, qi, pi	b, a, d, b, a	Tachycardia
Eye pain, itchy eyes, red tongue, thin moss, white moss, fine pulse, stringy pulse	Deficiency of liver and kidney	gan, xu, shen	a, d, a	Cataract

Author: Yingshuai Wang, Date: 2024-12-21.

- (1) AUC a metric used to measure the effectiveness of the ranking prediction model, the closer the value is to 1, the better the model is, mathematically defined as in Equation 17.

$$AUC = \frac{1}{|T^+||T^-|} \sum_{x^+ \in T^+} \sum_{x^- \in T^-} [g(x^+) > g(x^-)] \quad (17)$$

Where  $T^+$ ,  $T^-$  denote positive and negative sample sets,  $|T^+|$ ,  $|T^-|$  denote the number of positive sample and negative sample  $g(x)$  is the model predictive function,  $I(x)$  is an indicative function.

- (2) Hits@10 denotes the probability of hitting the real labels, which are ranked in the top 10 predicted objects by the model output, the higher the better, as defined in Equation 18.

$$Hits@10 = \frac{label \cap pred\_10}{numbers\ of\ label} \quad (18)$$

Where  $pred\_10$  denotes the top 10 ranked evidence elements predicted by the model.

- (3) MR (Mean Rank) is used to measure the likelihood of the model incorporating errors, the smaller the better, defined as in Equation 19.

$$MR = \frac{1}{|S|} \sum_{i=1}^{|S|} rank_i = \frac{1}{|S|} (rank_1 + rank_2 + \dots + rank_{|S|}) \quad (19)$$

- (4) MRR (Mean Reciprocal Ranking) reflects the generalization ability and robustness of the model, with higher values indicating better performance. The mathematical expression of MRR is shown in Equation 20.

$$MRR = \frac{1}{|S|} \sum_{i=1}^{|S|} \frac{1}{rank_i} = \frac{1}{|S|} \left( \frac{1}{rank_1} + \frac{1}{rank_2} + \dots + \frac{1}{rank_{|S|}} \right) \quad (20)$$

### 3.4 Experimental setup

The TCM recommendation model is implemented based on the TensorFlow framework, and one GPU (Tesla V100-PCIE-32GB) is used for training and testing. In order to be comparable between models, the data set and hyper-parameters are maintained the same. In order to prevent the interference of indicator fluctuation on the results, the models were trained five times for each experiment, and the evaluation indicators were averaged for each model.

### 3.5 Experimental results

The TCM evidence elements are divided into five categories: disease location, essence substance, disease evil, pathological state and association relationship, in order to enhance the learning ability of the model, the evidence elements are split according to the categories they belong to, and are viewed as five tasks for modelling. Table 2 shows the performance of each model in different datasets.

In the TCM evidence element recommendation task, there is a clear difference from the traditional e-commerce recommendation task. In e-commerce recommendation, there is no distinction between correct and incorrect products, while each item in a TCM medical case corresponds to an exact evidence element. This feature makes the evaluation of recommendation results more concerned with the accuracy of the first 10 evidence elements rather than the global comprehensive evaluation of recommendations. We introduce a custom metric called unit hit rate. By emphasizing the top-ranked correct evidence elements, this metric provides a more intuitive assessment of the model's recommendation performance in the TCM domain. The evaluation metric not only focuses on the accuracy of the recommendation results, but also on their importance and



TABLE 2 Multi-task meta-learning model effects.

Model	Hits@10	<i>p</i> value	MeanRank	<i>p</i> value	MRR	<i>p</i> value	AUC	<i>p</i> value
MLP	0.76644	base	3.00478	base	0.33280	base	0.90844	base
TextCNN	0.77080	0.1987	2.95720	0.4029	0.33812	0.5872	0.90991	0.2583
MMOE	0.76824	0.2018	2.98508	0.3964	0.33500	0.0973	0.90870	0.1476
Our model	<b>0.78713</b>	<b>0.0392</b>	<b>2.94366</b>	<b>0.0417</b>	<b>0.33959</b>	<b>0.0358</b>	<b>0.92974</b>	<b>0.0485</b>

Author: Yingshuai Wang, Date: 2025-04-27.  
The bold values in Table 2 represent the optimal performance for each evaluation metric in the model comparison experiments.

effectiveness in practical applications. The unit hit rate is calculated as Equation 21.

$$Hits^k = \frac{1}{N} \sum_{i=1}^N n_i \quad n_i = \begin{cases} 1 & \text{if } pred_i^k \text{ in } label_i \\ 0 & \text{otherwise} \end{cases} \quad (21)$$

Where *k* denotes the position of the predicted evidence elements after sorting, *N* represents the number of samples, *n<sub>i</sub>* denotes the correct evidence of the sample, and *pred<sub>i</sub><sup>k</sup>* denotes the predicted evidence at the *kth* position for the *ith* sample. This evaluation metric is calculated by dividing the number of correct evidence elements at the *kth* position in the prediction by the total number of samples, providing insight into the distribution of correct evidence elements within the prediction results. The metric was computed for the top 10 positions in the test results of MLP, TEXTCNN, MMOE, and the models presented in this study, as illustrated in Figure 6.

It can be seen that the models proposed in this research substantially outperform baselines in the first six positions, with an average number of correct evidence elements in the test samples of 4.489, and the distributions of correct evidence elements in the top 10 recommended results are all ahead of other models.

## 4 Discussion

### 4.1 Advanced characteristics of the model

Figure 7 illustrates our model performance compared to MLP, TextCNN, and MMOE across Hits@10, MeanRank, MRR, and AUC metrics, with standard deviation represented by error bars. Our approach consistently outperforms all baselines across these metrics while maintaining minimal variance. The results demonstrate our model superior precision and remarkable stability. This comprehensive evaluation provides compelling evidence of our method effectiveness and advantages over existing approaches.

### 4.2 Discussion of the ablation experiment

To more effectively compare the impacts of different enhancement points in the model, we conducted four ablation studies. These experiments focused on herbal statistical features (Test1), traditional Chinese medicine domain knowledge features (Test2), the meta-attention unit (Test3), and the feature adaptive fusion network (Test4). The results of the ablation studies are illustrated in Table 3.

### 4.3 Discussion of the unit hit rate behavior

Unit Hit Rate behavior is a novel metric introduced in this study, specifically designed for evaluating models in the medical domain. It emphasizes the precision of the top-ranked items in the model's recommendation list. Our model performs the best across all metrics, achieving 0.7176 in hits@1, 0.6371 in hits@2, 0.5054 in hits@3, and 0.6201 in Hits\_avg, which is the average of the first three positions, significantly outperforming other models. With a *p* value of 0.0279, the improvement is statistically significant compared to the baseline model, as shown in Table 4.

## 5 Conclusion

In this article, a knowledge-driven data scenario modelling approach is proposed. A representation based on association rules and domain knowledge is adopted, and the knowledge representation is transformed into mathematical vectors, which not only captures the relevance in the domain, but also provides more accurate information for the model. A data knowledge fusion neural network is proposed, which improves the model's understanding of knowledge by constructing auxiliary tasks and designing the interaction function between the main task and auxiliary tasks. A unit hit rate evaluation index is proposed to focus on the accuracy of the prediction of the forward position, which better measures the recommendation prediction effect of the model in the field of traditional Chinese medicine, and provides a more targeted direction for the iteration of the model. An automated framework for feature engineering and recommendation algorithm model training is designed and applied to a real TCM medical case evidence prediction task, demonstrating the model's effectiveness, which has been recognized by experts in the TCM field.

Although our approach has achieved significant results in TCM syndrome prediction, its association rules and knowledge representation depend on TCM specific theoretical systems, potentially limiting adaptability when transferred to Western medicine. While auxiliary task design enhances the model's perception of domain knowledge, the decision-making processes within deep neural networks still lack interpret ability, potentially undermining clinical expert trust in system outputs. Additionally, current evaluations primarily rely on historical case back-testing and offline metrics, without prospective clinical pilots or user experience studies, making it difficult to fully reflect the model's utility in real diagnostic scenarios. Future research will focus on developing generalized data-knowledge fusion algorithms, incorporating explainable techniques, advancing clinical application pilots, and establishing compliant data sharing or federated learning platforms to achieve greater practicality and universality.

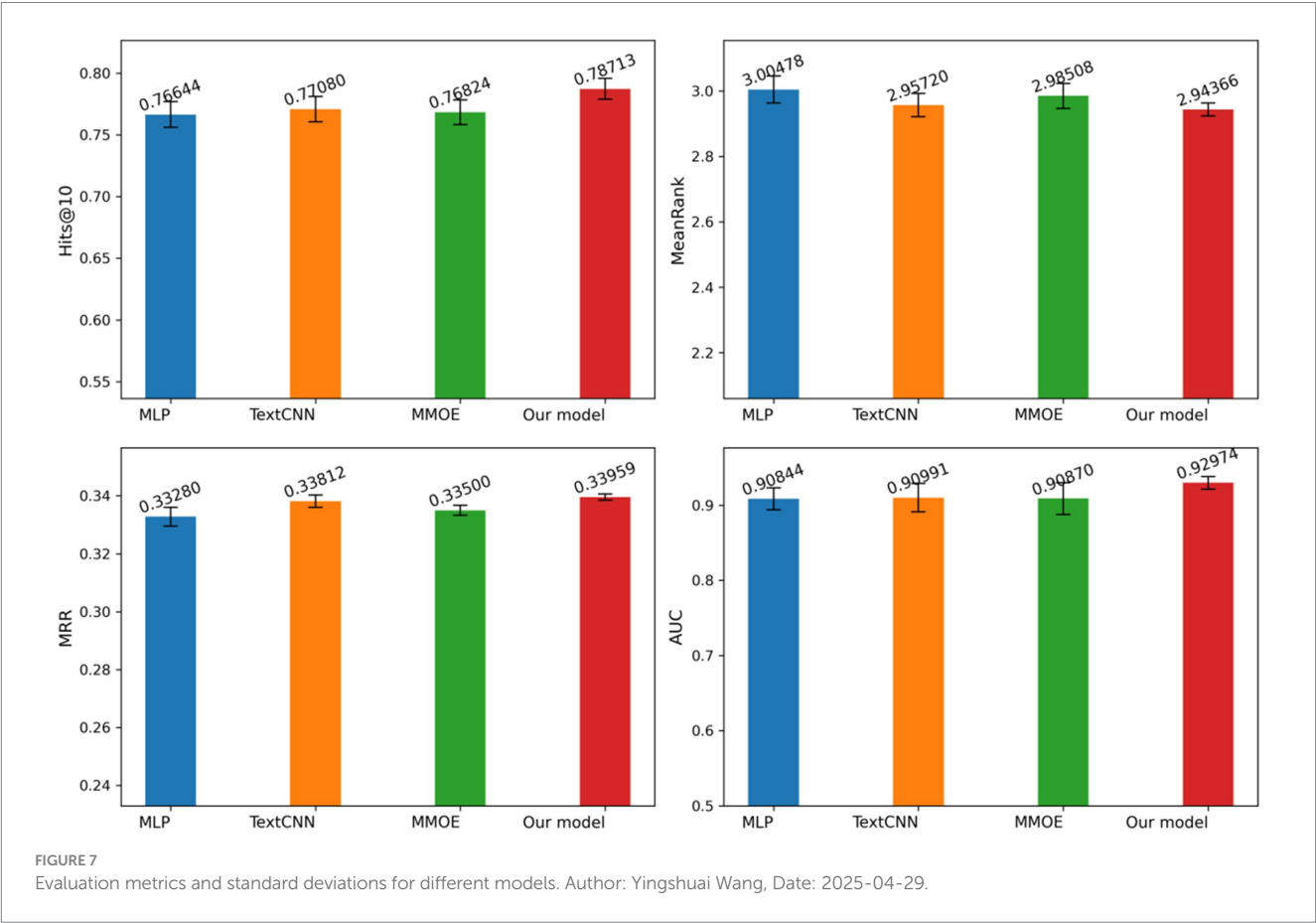
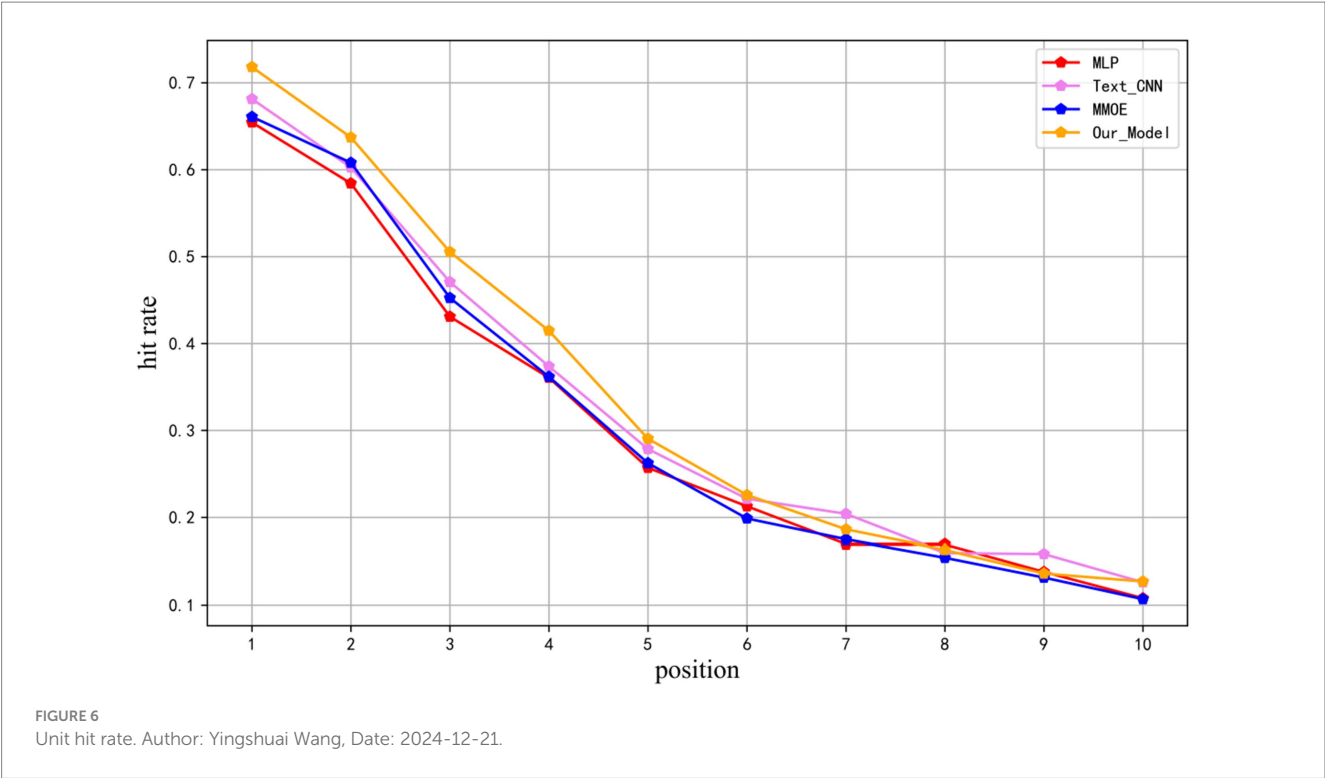


TABLE 3 Ablation study results.

AB test	Hits@10	<i>p</i> value	MeanRank	<i>p</i> value	MRR	<i>p</i> value	AUC	<i>p</i> value
Test1	0.76975	0.0478	2.95832	0.0492	0.33780	0.0563	0.91042	0.0573
Test2	0.77482	0.0498	2.96120	0.0531	0.33742	0.0481	0.91875	0.0493
Test3	0.77591	0.0378	2.95609	0.0467	0.33805	0.0479	0.90998	0.0462
Test4	<b>0.77307</b>	<b>0.0362</b>	<b>2.94366</b>	<b>0.0403</b>	<b>0.33959</b>	<b>0.0329</b>	<b>0.92974</b>	<b>0.0438</b>

Author: Yingshuai Wang, Date: 2025-04-27.  
The bold values in Table 3 indicate the best performance for each evaluation metric in the ablation experiments.

TABLE 4 Unit hit rate behavior.

Model	hits@1	hits@2	hits@3	Hits_avg	<i>p</i> value
MLP	0.6544	0.5842	0.4309	0.5565	base
Test_CNN	0.6814	0.6026	0.4708	0.5850	0.0973
MMOE	0.6609	0.6080	0.4525	0.5738	0.1856
Our model	0.7176	0.6371	0.5054	0.6201	0.0279

Author: Yingshuai Wang, Date: 2025-04-25.  
The bold values in Table 4 indicate the best performance for the 'unit hits behavior' evaluation metrics.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

This study was based on publicly available datasets. Ethical review and approval was not required for the study, in accordance with the legislation and institutional requirements.

Author contributions

YiW: Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Writing – original draft. YaW: Validation, Writing – review & editing, Resources. HH: Funding acquisition, Resources, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This study was

funded by the Chinese Academy of Medical Sciences Innovation Fund for Medical Sciences (CIFMS) [2022-I2M-1-019], National Social Science Fund of China (22&ZD141), The Non-profit Central Research Institute Fund of Chinese Academy of Medical Sciences (2024-ZHCH630-01), National Natural Science Foundation of China (61971446).

Acknowledgments

The authors thank all team members of the Health Information Management Section of the Institute of Medical Information, Chinese Academy of Medical Sciences, for their guidance on the methodology of this study.

Conflict of interest

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

Generative AI statement

The authors declare that no Gen AI was used in the creation of this manuscript.

Publisher’s note

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

References

1. Shen X, Cai X, Cao H. Research progress on recommender system incorporating medical knowledge graph. *Comput Eng Appl.* (2023) 59:40–51. doi: 10.3778/j.issn.1002-8331.2301-0006

2. Wu Y, Luo J, Lin C. Towards a new era of artificial intelligence: research and thinking on ChatGPT in smart medical application scenarios. *Data Commun.* (2023) 4:33–8. doi: 10.3969/j.issn.1002-5057.2023.04.009

3. Quan C, She D. A graph neural network drug recommendation method fusing patients' signs and medication data. *Data Anal Knowl Disc.* (2022) 6:113–24. doi: 10.11925/infotech.2096-3467.2021.1452

4. Xiaojing H, Wang P. Research on patients' willingness to recommend internet medical services and influencing factors. *Hosp Manag Forum.* (2023) 40:15–22. doi: 10.3969/j.issn.1671-9069.2023.02.004

5. Wang F, Hu H, Wan Y, et al. Research on community-wide health management model based on smart health prescription. *J Med Inform.* (2022) 43:62–6. doi: 10.3969/j.issn.1673-6036.2022.12.012

6. Pinsky MR, Bedoya A, Bihorac A, Celi L, Churpek M, Economou-Zavlanos NJ, et al. Use of artificial intelligence in critical care: opportunities and obstacles. *Crit Care.* (2024) 28:113. doi: 10.1186/s13054-024-04860-z

7. Liu X, Liu H, Yang G, Jiang Z, Cui S, Zhang Z, et al. A generalist medical language model for disease diagnosis assistance. *Nat Med.* (2025) 31:932–42. doi: 10.1038/s41591-024-03416-6
8. Li J, Guan Z, Wang J, Cheung CY, Zheng Y, Lim LL, et al. Integrated image-based deep learning and language models for primary diabetes care. *Nat Med.* (2024) 30:2886–96. doi: 10.1038/s41591-024-03139-8
9. Stoltzfus JC. Logistic regression: a brief primer. *Acad Emerg Med.* (2011) 18:1099–104. doi: 10.1111/j.1553-2712.2011.01185.x
10. Truong VH, Tangaramvong S, Papazafeiropoulos G. An efficient LightGBM-based differential evolution method for nonlinear inelastic truss optimization. *Expert Syst Appl.* (2024) 237:121530. doi: 10.1016/j.eswa.2023.121530
11. Chen T, Guestrin C. Xgboost: a scalable tree boosting system[C] Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. (2016): 785–794. Washington: University of Washington
12. Zaman N, Jana A. Automated recommendation model using ordinal probit regression factorization machines. *Int J Data Sci Anal.* (2024) 6:1–15. doi: 10.1007/s41060-024-00623-9
13. Sze V, Chen YH, Yang TJ, Emer JS. Efficient processing of deep neural networks: a tutorial and survey. *Proc IEEE.* (2017) 105:2295–329. doi: 10.1109/JPROC.2017.2761740
14. Cheng H T, Koc L, Harmsen J, Shaked Tal, Chandra Tushar, Aradhye Hrishi, et al. Wide & deep learning for recommender systems Proceedings of the 1st workshop on deep learning for recommender systems. (2016): 7–10. Association for Computing Machinery: New York
15. Wang Z, Ruan S, Huang T, Zhou H, Zhang S, Wang Y, et al. A lightweight multi-layer perceptron for efficient multivariate time series forecasting. *Knowl-Based Syst.* (2024) 288:111463. doi: 10.1016/j.knosys.2024.111463
16. Zhou G, Zhu X, Song C, Fan Ying, Zhu Han, Ma Xiao, et al. Deep interest network for click-through rate prediction Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining. (2018): 1059–1068. London: ACM
17. Vaswani A, Shazeer N, Parmar N, Uszkoreit Jakob, Jones Llion, Gomez Aidan N, et al. Attention is all you need. Proceedings of the 30th conference on neural information processing systems, (2017) 5998–6008. Long Beach, CA: NIPS
18. Ma J, Zhao Z, Yi X, Chen Jilin, Hong Lichan, Chi H., et al. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining. (2018): 1930–1939. London: ACM
19. Hu J, Shen L, Sun G. Squeeze-and-excitation networks Proceedings of the IEEE conference on computer vision and pattern recognition. (2018): 7132–7141. Salt Lake City, UT: IEEE
20. Kim Y, Convolutional neural networks for sentence classification. Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), (2014): 1746–1751. Qatar: Association for Computational Linguistics