ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

Adaptive optimization model of integrated traditional Chinese and Western medicine therapy based on reinforcement learning

Xiaozhong Liu[1], *Jingling Zuo*[2], *Heng Yin*[2], *Xiaoti Wu*[2] *and Fafeng Cheng*[1,*]

[1] School of Traditional Chinese Medicine, Beijing University of Chinese Medicine, Beijing, 100029 China
[2] Centre for Evidence-Based Chinese Medicine, Beijing University of Chinese Medicine, Beijing, China

SUMMARY: *An adaptive optimization model of integrated traditional Chinese and Western medicine based on reinforcement learning was proposed to solve the problems of dependence on experience in plan adjustment, insufficient description of individual difference response, and continuous optimization of combined intervention in the process of integrated traditional Chinese and Western medicine. In this model, western medicine clinical indicators, TCM syndrome characteristics, historical treatment trajectories and patient stage feedback are integrated into the state-action-reward framework, and the dynamic update of combined treatment strategy is realized by using deep Q network. Based on anonymized clinical records and time-series follow-up data, a continuous treatment decision sample was constructed, and the index improvement ability, syndrome adjustment effect, synergy gain, Q value convergence, patient satisfaction and treatment interruption of the model were comprehensively evaluated. The results showed that in the model group, the average fasting blood glucose decreased by 1.47 mmol/L, the average glycosylated hemoglobin decreased by 0.83%, the total syndrome score decreased by 9.2 points, the synergy gain index reached 0.34 in the sixth week, the average satisfaction score was 4.47, and the cumulative interruption rate in the sixth week was controlled at 14.4%. The research shows that reinforcement learning can provide computational support for continuous decision-making and joint strategy updating in integrated traditional Chinese and western medicine treatment, and provide new technical solutions for intelligent optimization of clinical intervention pathways.*

KEYWORDS: *Reinforcement learning; Integrated traditional Chinese and western medicine treatment; Adaptive optimization; Clinical decision support*

1 Introduction

In clinical scenarios such as chronic disease management, metabolic disease intervention, and complex symptom control, a single treatment pathway is often difficult to continuously adapt to the dynamic changes in patient status. Western medicine treatment emphasizes index monitoring, drug regulation and evidence-based intervention, which can quickly act on objective variables such as blood glucose, blood lipid, inflammatory factors and organ function. TCM treatment pays more attention to syndrome identification, overall regulation and individual differences, which has unique value in improving physical bias, relieving accompanying symptoms and improving long-term care effects. Therefore, integrated

*fafengcheng@gmail.com
https://doi.org/10.65102/is2026085
Published: 30 April, 2026

traditional Chinese and western medicine shows obvious complementary advantages, but the actual implementation process is not a simple superposition of two types of programs, but a continuous decision-making process with the evolution of the disease, the fluctuation of indicators and the transformation of syndromes. If we still rely on fixed schemes or empirical hierarchical treatment, it is easy to have problems such as lagging intervention rhythm, insufficient individual adaptation, and unstable collaborative efficiency.

With the digital accumulation of electronic medical records, laboratory examination data, medication records and four diagnostic information of traditional Chinese medicine, the clinical treatment process has a realistic basis for data-driven optimization. In recent years, machine learning and clinical decision support technology have continuously entered the medical application scenarios, making it possible to identify multivariate states, predict risk and recommend pathways. However, most of the existing intelligent medical models are still mainly based on static classification or single prediction, which are better at answering "what state is the current patient", but are difficult to deal with the sequential question of "how to jointly adjust the treatment in the next stage". For integrated traditional Chinese and Western medicine treatment, the patient state includes not only quantifiable physiological indicators of western medicine, but also the evolution of TCM syndromes, symptom combinations and physical signs. The state space is complex and the feedback cycles are different.

Reinforcement learning provides a new computational framework for this purpose. In this method, the treatment process can be abstracted as a closed-loop system of "state-action-reward", so that the agent can continuously learn a better intervention strategy according to the patient's current clinical indicators, syndrome characteristics and historical responses, and realize the iterative optimization of the scheme under the constraint of long-term benefits. Based on this idea, this paper focuses on the dynamic adaptation problem in integrated traditional Chinese and Western medicine treatment, and constructs an adaptive optimization model based on reinforcement learning. The improvement of western medicine indicators, TCM syndrome relief, treatment synergy, patient satisfaction and interruption risk are jointly incorporated into the reward design, and a joint decision-making mechanism for individual differences and stage evolution is formed. This study not only helps to improve the refinement and continuity of treatment plan generation, but also provides an interpretable modeling path for computer technology to participate in collaborative treatment of traditional Chinese and western medicine.

2 Review of related research

2.1 Research on decision-making model of integrated traditional Chinese and Western medicine treatment

The decision-making mode of integrated traditional Chinese and Western medicine is usually based on the parallel advancement of Western medicine "disease differentiation and classification" and traditional Chinese medicine "syndrome differentiation and treatment". Its core does not lie in the mechanical superposition of the two kinds of means, but in the dynamic combination of drugs, doses, intervention frequencies and conditioning methods according to the course of the disease, the severity of symptoms, the changes of laboratory indicators and the evolution trend of syndromes. Previous studies have shown that dynamic treatment regimes can better adapt to the changes of patient status in multi-stage processes, and are more consistent with real clinical decision-making logic than static regimes. Chakraborty and Murphy pointed out that clinical treatment is essentially a decision-making process of continuous updating, and the intervention at different time points should be consistent with the early response and subsequent goals [1]. Huang et al. further proposed that multi-stage treatment optimization

needs to incorporate the cumulative information into the decision-making basis, otherwise it is difficult to obtain stable long-term benefits [2]. Although this kind of research is mainly based on the modern medical statistical framework, it has important methodological implications for the phased adjustment of integrated traditional Chinese and Western medicine treatment.

In actual clinical practice, the decision-making of traditional integrated traditional Chinese and western medicine still relies more on the integration of physician experience. The western medicine part focuses on objective data such as blood glucose, blood lipid, inflammatory indicators and organ function, while the traditional Chinese medicine part discriminates syndromes based on the information of main symptoms, complications, tongue condition, pulse condition and constitution. The problem is that these two kinds of information are partial numerical and partial semantic, and their data structures are significantly different, which makes it difficult to form a unified computational expression for joint decision-making. Zhou et al. built TCM clinical data warehouse earlier and proved that TCM diagnosis and treatment information can enter the knowledge discovery and decision support process through standardization organizations [3]. The TCM intelligent auxiliary diagnosis system developed by Zhang et al further shows that TCM syndrome recognition can achieve high consistency with the help of artificial intelligence methods [4]. Pan et al., Tian et al. 's review also shows that machine learning has gradually entered the process of TCM diagnosis, syndrome differentiation analysis and auxiliary decision-making, but most of the research is still focused on diagnosis recognition or efficacy classification, and less in-depth into the sequential optimization level of combined TCM and Western medicine intervention [5, 6]. It can be seen that the decision-making mode of integrated traditional Chinese and western medicine has strong adaptability and practical basis in clinical practice, but its operation mechanism is still driven by experience, and its ability to integrate multi-source heterogeneous data in real-time is insufficient, and its utilization of long-term feedback in the treatment process is also limited. These limitations indicate that it is urgent to introduce reinforcement learning, clinical knowledge modeling and computer-aided decision making methods into integrated traditional Chinese and western medicine to improve the ability of multi-source information integration and continuous optimization.

2.2 Research on artificial intelligence-driven intelligent medical intervention methods

The intelligent medical intervention method driven by artificial intelligence is pushing the medical support system from "result judgment tool" to "process optimization tool". Early medical artificial intelligence mainly focuses on risk stratification, diagnosis recognition and prognosis prediction, and its computational goal is mainly to learn static mapping relationships from existing samples to classify or score the disease state at a certain point. Such methods have practical value in assisting abnormal identification and improving screening efficiency, but when the research object turns to integrated traditional Chinese and Western medicine treatment, only static prediction is obviously insufficient. The reason is that the integrated TCM and Western medicine intervention is not a single end point, but a continuous evolution of the treatment process: Western medicine indicators may fluctuate in a short period of time, and TCM syndromes are often accompanied by gradual transformation of symptom clusters. The feedback frequency and presentation of the two types of information are not consistent with the clinical interpretation logic, so it is more necessary to have a computational framework that can deal with sequential decision-making and delayed benefits.

In recent years, reinforcement learning has gradually become an important direction of intelligent medical intervention research. Jayaraman et al. pointed out that reinforcement learning can represent the clinical intervention process as a closed-loop system with the

interaction of state, action and reward, so that the model can not only stop at "identifying what state the patient is in", but further learn "what intervention should be taken in the current state" [7]. Frommeyer et al. conducted a systematic review of the research on precision medicine and dynamic treatment plan, and believed that reinforcement learning is particularly suitable for multi-stage treatment optimization, because it can continuously absorb historical feedback and adjust the next strategy during the continuous progress of treatment [8]. Liu et al. also emphasized in their review of clinical decision support in intensive care that, compared with conventional supervised learning, reinforcement learning is closer to the continuous decision-making mechanism in real medical scenarios [9].

From the perspective of specific applications, the existing research has made initial progress in the direction of infection control, diabetes management and multi-disease co-treatment. The artificial intelligence clinician model constructed by Komorowski et al. proves that deep reinforcement learning can learn sepsis intervention strategies from treatment records [10]. Wu et al. further integrated human expert knowledge into the value-based deep reinforcement learning model to improve the clinical acceptability of treatment strategies [11]. Wang et al. introduced a clinical knowledge guidance mechanism into the dosage recommendation of antibiotics, so that the reinforcement learning decision no longer completely relied on black-box trial and error [12]. In the metabolic disease scenario, Zheng et al. proposed an individualized management framework for multi-disease co-management of type 2 diabetes, indicating that electronic medical records can support sequential treatment learning [13]. Wang et al. proof-of-concept study showed that reinforcement learning had practical feasibility in optimal blood glucose control [14]. Javad et al. exploratory study on type 1 diabetes also showed that this method was helpful to deal with the individual differences under high-frequency monitoring [15]. Together, these results show that AI-driven medical intervention has gradually shifted from single-shot prediction to dynamic regulation.

However, as far as integrated traditional Chinese and western medicine treatment is concerned, there are still obvious gaps in existing methods. On the one hand, many models mainly focus on western medicine structured data, and rarely incorporate TCM syndromes, symptom semantics and follow-up descriptions into the unified state space. On the other hand, although some deep reinforcement learning studies can generate recommendation strategies, they do not consider the treatment synergy, patient compliance and interruption risk sufficiently [16, 17]. For this study, what really needs to be solved is not only "whether a certain drug is effective", but how to establish a sustainable and updated adaptive optimization mechanism of integrated traditional Chinese and western medicine treatment under the joint action of multiple time steps, multi-dimensional feedback and dual medical logic. The differences between the existing methods and the research focus of this paper are shown in Table 1.

Table 1: Comparison of characteristics of AI-driven medical intervention methods

Method Type	Main Data Basis	Computational Objective	Applicable Advantages	Limitations in Integrated Traditional Chinese and Western Medicine Treatment
Supervised Learning Prediction Model	Laboratory indicators, medical imaging, and structured medical records	State classification and efficacy prediction	Easy to train and intuitive in results	Difficult to handle sequential adjustment and unable to directly generate dynamic treatment strategies
Rule-Driven Decision System	Clinical guidelines, expert experience, and rule base	Fixed treatment plan matching	Strong interpretability	Slow response to individual differences and syndrome evolution
Deep Sequential Recommendation Model	Time-series medical records and medication history	Next-step intervention recommendation	Able to utilize historical trajectory information	Insufficient long-term reward modeling and limited expression of collaborative mechanisms
Reinforcement Learning Intervention Model	Multi-stage clinical states, behavioral feedback, and outcome rewards	Dynamic treatment strategy optimization	Suitable for continuous decision-making and delayed- reward scenarios	Integration of TCM semantic features and joint reward design still need further improvement

3 Proposed method: Adaptive optimization model of integrated traditional Chinese and Western medicine based on reinforcement learning

Integrated traditional Chinese and Western medicine treatment is not a passive implementation process after a one-time given plan, but a dynamic decision-making process that is continuously adjusted with the changes of symptoms, index fluctuations, syndrome transformation and treatment feedback. If the combined treatment plan is made only based on the judgment of a single outpatient visit or fixed follow-up rules, it is often difficult to identify the subtle deviation of the patient's state in time, and it is also difficult to deal with the reality that the improvement of western medicine indicators is not synchronized with the relief of TCM syndromes. Based on this, this paper models the treatment process of integrated traditional Chinese and Western medicine as a reinforcement learning environment, and constructs an adaptive optimization model of "patient state representation - joint intervention action generation - multi-objective reward feedback - strategy iterative update", so that the computer can learn a better collaborative intervention path of traditional Chinese and Western medicine in continuous treatment cycles. The overall structure of the model is shown in Figure 1.

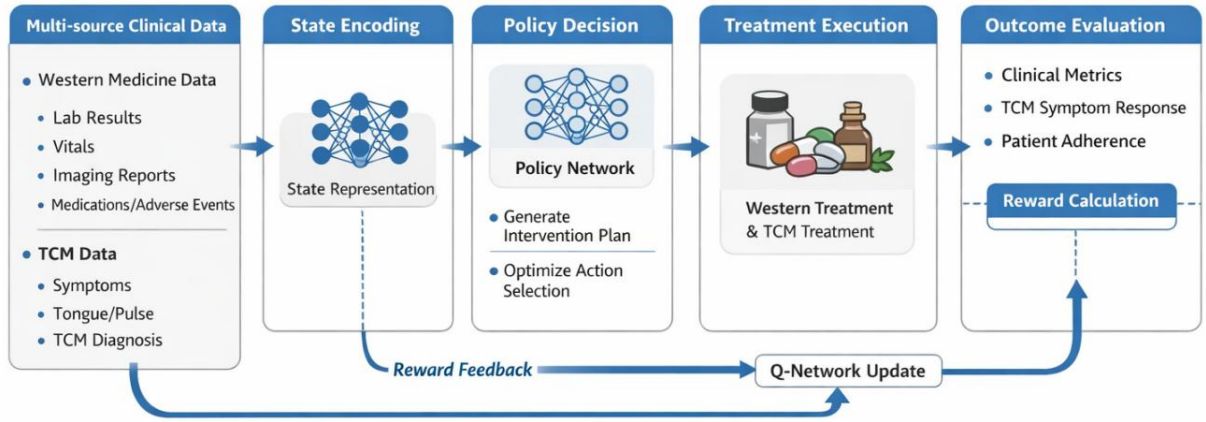

Figure 1: Overall architecture of adaptive optimization model of integrated Chinese and Western medicine therapy based on reinforcement learning

At the data input side, the model receives two types of information simultaneously. One is western medicine structured data, including laboratory test values, vital signs, imaging conclusion coding, previous medication records and adverse reactions information. The other category is TCM diagnosis and treatment information, including main symptoms, concurrent symptoms, tongue condition, pulse condition, syndrome label, constitutional tendency and follow-up text description. Considering the differences between the two types of data in expression form and time granularity, this paper first carries out unified preprocessing: The continuous indicators were interpolated with missing values and interval normalization, and the TCM symptom words and syndrome items were standardized. The follow-up texts were mapped into discrete semantic labels after word segmentation and medical entity extraction, and then the diagnosis and treatment records at different time points were reconstructed into time series samples according to the treatment cycle. After processing, the comprehensive state of the patient at time t can be expressed as follows.

$$s_t = \Phi(x_t^W, x_t^T, h_t) \tag{1}$$

where, x_t^W represents the western medicine index vector at time t, x_t^T represents the TCM syndrome and symptom feature vector, h_t is the historical treatment trajectory information, $\Phi(\cdot)$ is the multi-source feature fusion coding function. The significance of this formula is to map the originally scattered test data, symptom representation and previous response results into a unified state space, which provides a structured input for subsequent policy learning.

In terms of action design, this paper does not understand the treatment action as a single drug selection, but extends it to a combination of traditional Chinese and Western medicine intervention. Specifically, the action set A is composed of western medicine drug adjustment, TCM prescription addition and subtraction, intervention frequency setting, and follow-up intensity configuration. The reason for such treatment is that the optimization goal of integrated traditional Chinese and western medicine is not only to reduce a certain physiological index, but to pursue a balance between symptom relief, syndrome correction, compliance improvement and adverse risk control. Based on the current state s_t, the agent selects the joint treatment action a_t from the action space, and after executing it goes to the next state s_{t+1}.

In order to avoid the model only pursuing short-term index improvement and ignoring long-term synergy effects, this paper constructed a multi-objective reward function, which included the improvement of western medicine clinical indicators, the alleviation of TCM syndromes, treatment synergy and compliance into the feedback evaluation. The reward function is written as follows.

$$r_t = \lambda_1 \Delta C_t + \lambda_2 \Delta Z_t + \lambda_3 G_t - \lambda_4 D_t \quad (2)$$

Among them, ΔC_t represents the improvement range of the core clinical indicators of western medicine, ΔZ_t represents the decline of TCM syndrome score, G_t represents the synergistic benefit of the combined treatment of traditional Chinese medicine and western medicine, D_t represents the penalty term caused by adverse reactions, patient interruption or compliance decline, and λ_1-λ_4 is the weight coefficient set by verification. The key of this reward design is not to simply concatenate multiple indicators, but to transform the "parallel improvement" of traditional Chinese and western medicine treatment goals into a unified feedback signal that can be calculated and returned, so that reinforcement learning can truly serve the combined treatment scenario.

In the policy learning layer, this paper uses the deep Q network to construct the value function to approximate the long-term benefit of the state-action pair. Given the current parameter θ, the q-value update objective is as follows.

$$y_t = r_t + \gamma \max_{a'} Q(s_{t+1}, a'; \theta^-) \quad (3)$$

where, γ is the discount factor and θ^- is the target network parameter. The loss function can then be constructed as follows.

$$L(\theta) = \mathbb{E}[(y_t - Q(s_t, a_t; \theta))^2] \quad (4)$$

The model extracts small batch sequences from historical treatment samples through experience replay mechanism, uses Adam optimizer to iteratively update parameters, and periodically synchronizes the target network to reduce training oscillation. Considering the imbalance between high-yield samples and low-frequency complex samples in real clinical data, this paper retains key state transition segments in the experience pool to improve the recognition ability of the strategy for disease fluctuation nodes and syndrome turning nodes.

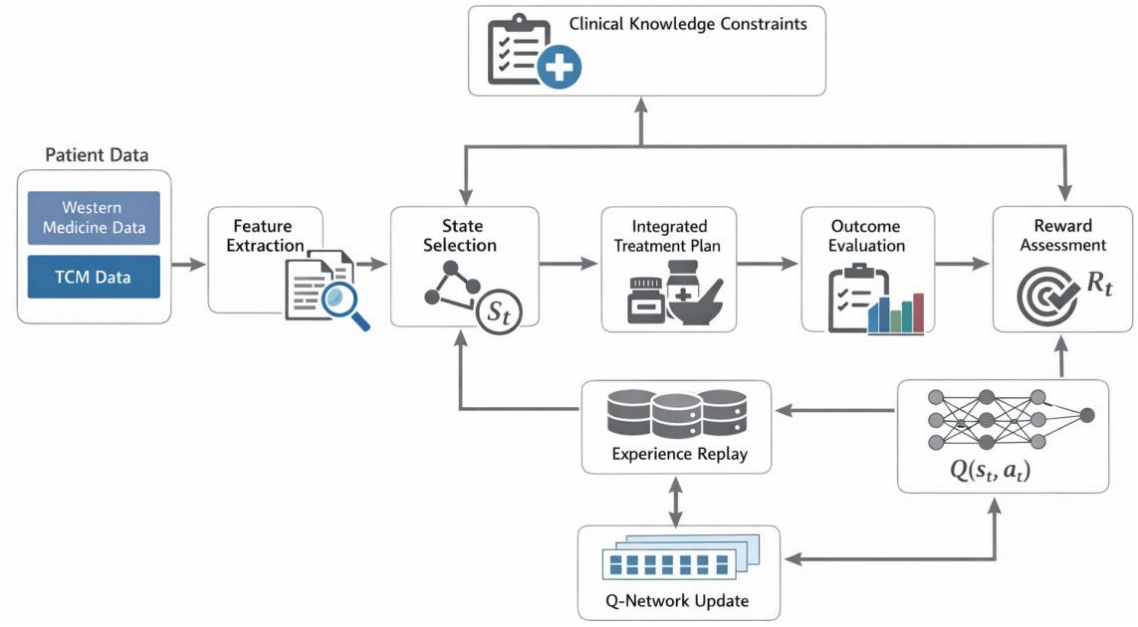

Figure 2: Modeling and reinforcement learning feedback update process of integrated Chinese and Western medicine treatment state

In order to enhance the interpretability of the model to the process of combined treatment, this paper also designs the state evolution and strategy feedback pathway, as shown in Figure 2. On the one hand, it reflects the main link from the patient's original information to the state coding, strategy selection, treatment implementation, and efficacy feedback. On the other hand, "expert knowledge correction branch" is added, that is, clinical guidelines constraints, TCM syndrome differentiation rules and taboo knowledge are embedded into the action screening stage to eliminate strategy combinations that are unreasonable in theory or have security risks. After the selection of knowledge constraints, the model limits the data-driven policy search to the set of clinically acceptable actions, so as to achieve a balance between policy optimization ability and medical safety.

4 Experimental design and performance evaluation

4.1 Setting of evaluation indexes and comparison methods

In order to systematically evaluate the adaptive optimization model of integrated traditional Chinese and western medicine treatment based on reinforcement learning constructed in this paper, the experimental part sets up evaluation indicators from four dimensions of clinical efficacy, syndrome improvement, strategy learning stability and patient behavior feedback, and introduces a variety of comparison methods for horizontal test. Considering that integrated traditional Chinese and Western medicine treatment is not a single endpoint optimization problem, it is often difficult to reflect the real value of combined treatment strategy in continuous intervention if the model is judged by only a certain kind of biochemical indicators or a single efficacy score. Therefore, in this paper, the improvement of western medicine clinical indicators, the change of TCM syndrome score, the synergy benefit of TCM and western medicine, the convergence of Q value, patient satisfaction and treatment interruption rate are incorporated into the performance evaluation system to ensure that the evaluation results can cover the core characteristics of the three levels of "efficacy, process and strategy".

The improvement effect of western medicine clinical indicators was expressed by the comprehensive improvement rate, which was used to measure the overall adjustment ability of the model output scheme to the key objective indicators. Let the MTH western medicine index of the patient before and after treatment be c_{im}^{pre} and c_{im}^{post} respectively, then the comprehensive improvement rate of western medicine can be expressed as follows.

$$E_c = \frac{1}{N}\sum_{i=1}^{N}\sum_{m=1}^{M}\omega_m \cdot \frac{c_{im}^{pre} - c_{im}^{post}}{c_{im}^{pre}} \tag{5}$$

Here, N is the number of patient samples, M is the number of indicators included in the evaluation, and ω_m is the weight of different clinical indicators. This index can comprehensively reflect the improvement range of blood glucose, inflammation level and metabolic parameters, and is suitable for investigating the direct treatment effect at the level of western medicine.

The TCM syndrome change is characterized by the syndrome integral decline rate. If z_i^{pre} and z_i^{post} denote the total syndrome scores of patients before and after treatment, the syndrome improvement index is defined as follows.

$$E_z = \frac{1}{N}\sum_{i=1}^{N}\frac{z_i^{pre} - z_i^{post}}{z_i^{pre}} \tag{6}$$

This index mainly reflects the changes of the main symptoms, concurrent symptoms and the overall syndrome burden. Different from the indicators in western medicine, the syndrome score emphasizes the combination of symptoms and the adjustment of the overall state, so it can be used as a key basis for observing the effectiveness of TCM intervention.

Considering that the research object of this paper is the combination of traditional Chinese and western medicine rather than the juxtaposition of the two sets of programs, the collaborative optimization gain index is further introduced in the experiment to determine whether the strategy generated by the model truly reflects the complementary effect of the combination therapy. Let R_i^{int} be the total benefit of the combined treatment, and R_i^w and R_i^t be the benefits of the Western medicine alone program and the traditional Chinese medicine alone program, respectively. Then the synergistic gain can be written as follows.

$$G_s = \frac{1}{N}\sum_{i=1}^{N}\frac{R_i^{int} - \max(R_i^w, R_i^t)}{\max(R_i^w, R_i^t)} \tag{7}$$

When G_s>0, it means that the joint scheme has extra gain compared with the single path. If the value is persistently low, it means that the model can generate composite actions, but it has not achieved true collaborative optimization.

In terms of reinforcement learning performance evaluation, this paper focuses on the stability and convergence of the policy learning process. To this end, the mean value of Q value fluctuation is introduced as the training stability index:

$$S_q = \frac{1}{T-1}\sum_{t=2}^{T}|Q_t - Q_{t-1}| \tag{8}$$

where T is the total number of training iterations and Q_t is the average Q value after the TTH update. The smaller the value is, the more stable the policy network becomes in the later training, and the less prone to violent oscillations in the learning process. At the same time, patient satisfaction was expressed by the mean of Likert scale, and the treatment interruption rate was calculated by the proportion of patients who did not complete the established intervention cycle. The two indicators were used to measure the actual performance of the model program in the level of acceptance and compliance, respectively.

In terms of comparison method Settings, this paper selects three representative baseline models. The first is the rule-driven combined treatment method, which performs scheme matching according to the preset diagnosis and treatment rules of traditional Chinese and western medicine, and does not have online learning ability. The second is supervised learning recommendation model, which uses the patient's historical characteristics to predict the next intervention action, and can complete static recommendation, but lacks long-term reward modeling. The third is the standard deep reinforcement learning model without collaborative reward, which retains the state-action learning framework, but does not explicitly introduce the synergy term of traditional Chinese and western medicine in the reward function, so as to test the actual contribution of the reward design in this paper. Through these three types of methods, the advantages of the proposed model can be verified from three levels of rule system, static

learning and ordinary reinforcement learning.

In order to ensure the robustness of the results, the unified sample division and cross validation strategy are used in the experiment. The training set, validation set and test set are relatively balanced in patient age, disease stage, baseline indicators and syndrome types, and the random influence caused by random initialization is weakened by multiple rounds of repeated experiments. The evaluation system can not only test the model's ability to improve clinical outcomes, but also reveal its comprehensive performance in strategy learning, joint optimization and patient feedback, which provides a more reliable quantitative basis for subsequent results analysis.

4.2 Source of data set and explanation of sample characteristics

In order to test the effectiveness of the adaptive optimization model of integrated traditional Chinese and western medicine based on reinforcement learning in continuous treatment decision-making, the experimental data are constructed by "real treatment records collation + time series sample reconstruction", instead of using pure simulated data directly. The reason is that integrated traditional Chinese and Western medicine treatment involves western medicine structured test information, TCM syndrome description, prescription adjustment records, and stage follow-up feedback. If only relying on rules to generate samples, although convenient for training, it is difficult to reflect the complex characteristics of symptom evolution, index fluctuation and treatment response in real clinical practice. Therefore, based on the anonymized electronic medical records and follow-up data of an integrated traditional Chinese and Western medicine clinic, this paper establishes a treatment sequence dataset for reinforcement learning training and testing.

The data content mainly includes four categories. The first is the basic information of the patient, such as age, gender, duration of disease and comorbidities. The second is clinical data of western medicine, including blood glucose, blood lipid, inflammatory indicators, liver and kidney function and related drug use records. The third is the information of TCM diagnosis and treatment, including the main symptoms, concurrent symptoms, tongue condition, pulse condition, syndrome classification, and the addition and subtraction of prescriptions. The fourth is the stage outcome information, such as the degree of symptom relief, return visit interval, satisfaction evaluation, and whether the treatment is interrupted. In order to meet the input requirements of the computational model, this paper standardized the continuous variables, encoded and mapped the text description of traditional Chinese medicine according to the terminology specification, and reorganized the original records into state transition samples for reinforcement learning according to the temporal logic of "initial diagnosis, follow-up and readjust".

After screening, a total of 612 samples were finally included, and 2874 treatment decision sequences were formed. The samples cover the common intervention scenarios of integrated traditional Chinese and western medicine, such as metabolic disorders, chronic inflammatory reactions, and syndromes accompanied by deficiency and reality, which can better reflect the characteristics of the coexistence of "short-cycle index changes" and "medium-cycle syndrome evolution" in combined treatment. From the sample characteristics, the age of patients was mainly distributed in 35-69 years old, the treatment cycle was concentrated in 6-12 weeks, and an average of 4.7 consecutive decision nodes were formed per patient. Western medicine data has a high dimension and relatively stable update frequency, while traditional Chinese medicine information shows the characteristics of strong semantics, high dispersion and obvious stage transformation. This heterogeneous feature structure provides a realistic data basis for multi-source state fusion and sequential decision learning of the model in this paper. At the same time, all the data used in this paper have been anonymized and do not involve information that can

identify individuals. Although the dataset is derived from real clinical practice, it still mainly reflects the diagnosis and treatment characteristics of a single center of integrated traditional Chinese and Western medicine clinic, so there are certain limitations in external generalization. However, in terms of state continuity, feedback observability, and treatment trajectory integrity required for training reinforcement learning models, this dataset is sufficient to support our experimental verification of adaptive optimization mechanisms.

5 Analysis and discussion of results

5.1 Discussion of Results

Integrating the results of western medicine clinical indicators, TCM syndrome scores, synergy gain, Q-value convergence, patient satisfaction and treatment interruption rate, it can be seen that the adaptive optimization model of integrated traditional Chinese and western medicine therapy based on reinforcement learning in this paper shows good comprehensive advantages in continuous treatment decision-making scenarios. The improvement is not only reflected in the local improvement of a single index, but in the more stable synchronous optimization between multi-dimensional objectives. On the one hand, the model can continuously adjust the intensity of western medicine and the addition and reduction strategy of traditional Chinese medicine according to the changes in the stage state of patients, so that the objective indicators such as blood glucose and inflammation fall faster, and the syndrome score is relieved more obviously. On the other hand, the joint reward mechanism makes the strategy learning not only pursue short-term numerical improvement, but also take into account the synergy benefit, compliance and treatment continuity. In other words, the model is closer to the dynamic intervention logic of "observation, adjustment and correction" in real clinical practice.

From the point of view of computer implementation, this result mainly benefits from three points. Firstly, the multi-source state coding unified the test data, symptom characteristics, syndrome labels and historical responses into the state space, which enhanced the recognition ability of the model for complex diseases. Secondly, the long-term reward update mechanism of reinforcement learning improves the sensitivity of the strategy to the subsequent results, so that the model output no longer stays in the static matching. Third, the introduction of synergy term and interruption penalty term in TCM and Western medicine makes the decision-making process more in line with the actual goal of combined treatment. It should be noted that the data in this paper mainly come from the clinical records of a single center, and the TCM syndrome score still has a certain color of empirical judgment. Therefore, the external generalization ability of the model still needs to be further verified on multi-center and multi-disease data. If more fine-grained follow-up data, knowledge graph constraints and online feedback mechanism from doctors can be combined in the follow-up, there is still room for further improvement in the stability and interpretability of the model in clinical decision-making.

5.2 Analysis of the improvement effect of Western medicine clinical indicators

In order to test the actual intervention effect of this model at the level of western medicine, four representative clinical indicators including fasting blood glucose, glycosylated hemoglobin, C-reactive protein and low-density lipoprotein cholesterol were selected for comparative analysis. These indicators correspond to the control of glucose metabolism, the stage of the overall blood glucose level, the degree of inflammatory response and the state of lipid metabolism, which can reflect the objective physiological regulation results of integrated traditional Chinese and western medicine treatment.

It can be seen from Table 2 that the improvement of the proposed model group in the four core indicators is higher than that of the other comparison methods. Among them, the average fasting blood glucose decreased by 1.47 mmol/L, which was 0.61 mmol/L, 0.44 mmol/L and 0.26 mmol/L higher than that of the rule-driven group, the supervised learning recommendation group and the standard deep reinforcement learning group, respectively. The average glycosylated hemoglobin decreased by 0.83%, which was 0.42, 0.27 and 0.15 percentage points higher than the other three groups, respectively. In terms of inflammation and lipid metabolism indicators, the average C-reactive protein and low-density lipoprotein cholesterol in the model group decreased by 2.79 mg/L and 0.58 mmol/L, respectively, maintaining the best results in the group. This indicates that the proposed model can not only improve metabolic related indicators, but also have a good continuous regulation ability to inflammatory load and lipid abnormalities.

Table 2: Comparison of the improvement effects of Western medicine clinical indicators under different methods

Method	Reduction in Fasting Blood Glucose (mmol/L)	Reduction in Glycated Hemoglobin (%)	Reduction in C-Reactive Protein (mg/L)	Reduction in Low-Density Lipoprotein Cholesterol (mmol/L)
Rule-Driven Group	0.86	0.41	1.72	0.29
Supervised Learning Recommendation Group	1.03	0.56	2.08	0.37
Standard Deep Reinforcement Learning Group	1.21	0.68	2.34	0.45
Proposed Model Group	1.47	0.83	2.79	0.58

From the perspective of the dynamic change process, the advantages of the model group in this paper are not limited to the end result, but continue to show throughout the treatment cycle. As shown in Figure 3, the baseline fasting blood glucose level of each group was similar, about 8.6 mmol/L. However, after the intervention, the model group showed a faster downward trend from the first week, and its value decreased from 8.60 mmol/L at baseline to 7.72 mmol/L at the second week. It further decreased to 7.16 mmol/L in the fourth week and stabilized at 6.83 mmol/L in the sixth week. In contrast, it was still 7.76 mmol/L in the rule-driven group at week 6, 7.42 mmol/L in the supervised learning recommendation group and 7.18 mmol/L in the standard deep reinforcement learning group. This result shows that the proposed model not only has an earlier effect, but also has a more stable control in the later stage, and there is no obvious rebound.

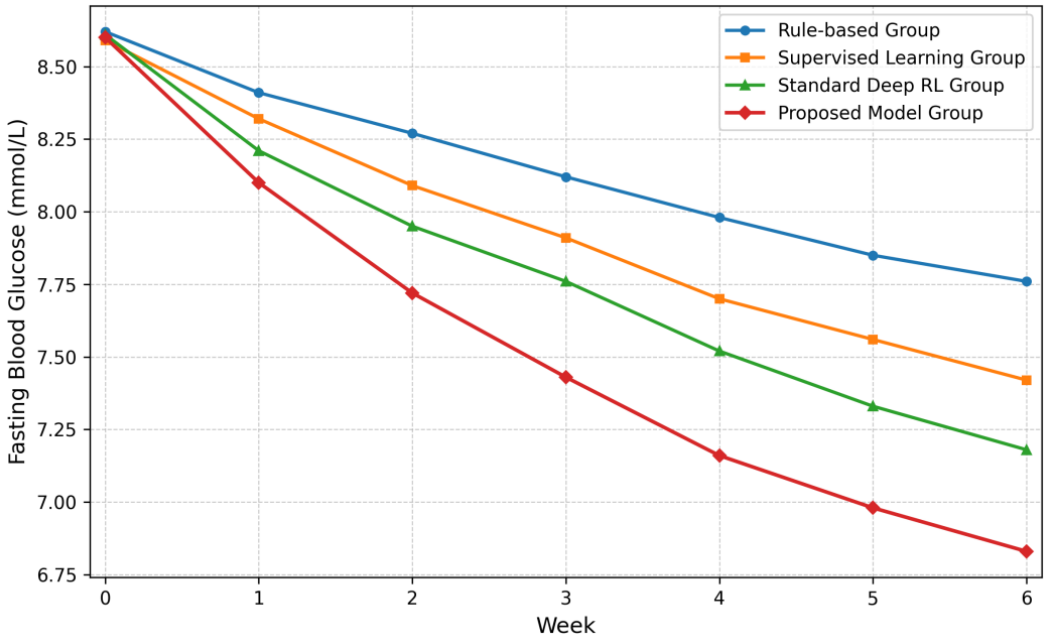

Figure 3: Weekly trends of fasting blood glucose in different methods

From the computational mechanism, this result is closely related to the long-term payoff modeling features of reinforcement learning. Different from the single-step recommendation method based only on the current state, the proposed model incorporates short-term indicator improvement and subsequent state stability in the reward function, so that the policy network is more inclined to output joint intervention actions that are "immediate and effective with low subsequent risk". Therefore, the western medicine treatment adjustment generated by the model does not show local radical suppression, but forms a smooth downward trajectory that is more in line with the continuous treatment law. In general, the model in this paper shows stronger continuous adjustment ability and better later stability in the improvement effect of western medicine clinical indicators, which provides more reliable computational support for the objective efficacy optimization in integrated traditional Chinese and Western medicine treatment.

5.3 Analysis on the effect of TCM syndrome changes

The changes of TCM syndromes can directly reflect the adjustment effect of the patient's overall state, and is also an important basis for testing whether the integrated traditional Chinese and Western medicine treatment truly realizes dynamic adaptation. Different from the indicators in western medicine that focus on physiological parameters, the syndrome score emphasizes the combination of symptoms, pathogenesis bias and the stage evolution of the body imbalance state, so it is more suitable for observing the differentiation regulation ability of reinforcement learning model in continuous intervention. To this end, this paper selects three indicators of total syndrome score, main disease remission rate and syndrome transformation stability to compare and analyze the improvement effects of traditional Chinese medicine under different methods.

It can be seen from Table 3 that the proposed model group performs best in terms of syndrome improvement. The average total score of symptoms was 18.6 points before treatment and 9.4 points after 6 weeks of treatment, with a total reduction of 9.2 points and a reduction rate of 49.5%. During the same period, the rule-driven group decreased from 18.4 points to 12.8 points, with a decrease of 5.6 points. The supervised learning recommendation group and the

standard deep reinforcement learning group decreased by 6.9 points and 8.0 points, respectively. If we further observe the remission rate of the main symptoms, the proposed model group reaches 71.3%, which is higher than 52.6% of the rule-driven group, 60.4% of the supervised learning recommendation group, and 66.8% of the standard deep reinforcement learning group, respectively. The results show that the proposed model not only promotes the relief of the main disease, but also improves the stability of the syndrome evolution process.

Table 3: Comparison of TCM syndrome improvement effects under different methods

Method	Total Syndrome Score Before Treatment	Total Syndrome Score After Treatment	Score Reduction	Reduction Rate / %	Main Symptom Relief Rate / %	Syndrome Transformation Stability
Rule-Driven Group	18.4	12.8	5.6	30.4	52.6	0.71
Supervised Learning Recommendation Group	18.5	11.6	6.9	37.3	60.4	0.78
Standard Deep Reinforcement Learning Group	18.7	10.7	8.0	42.8	66.8	0.84
Proposed Model Group	18.6	9.4	9.2	49.5	71.3	0.89

From the perspective of dynamic process, the syndrome score of the model group in this paper showed obvious continuity of decline. As shown in Figure 4, the initial difference of syndrome scores in each group was very small, all above 18 points. However, after the intervention began, the model group in this paper began to decline faster from the second week, from 18.6 points in the 0th week to 15.2 points in the second week. The score dropped further to 11.6 in the fourth week, and stabilized at 9.4 in the sixth week. In contrast, the rule-driven group still scored 12.8 until week 6, while the supervised learning recommendation group and the standard deep reinforcement learning group scored 11.6 and 10.7, respectively. More importantly, the curve of the model group in this paper tends to be stable in the later stage, indicating that the model does not frequently give drastic adjustment actions after the syndrome is relieved, but maintains good dialectical continuity.

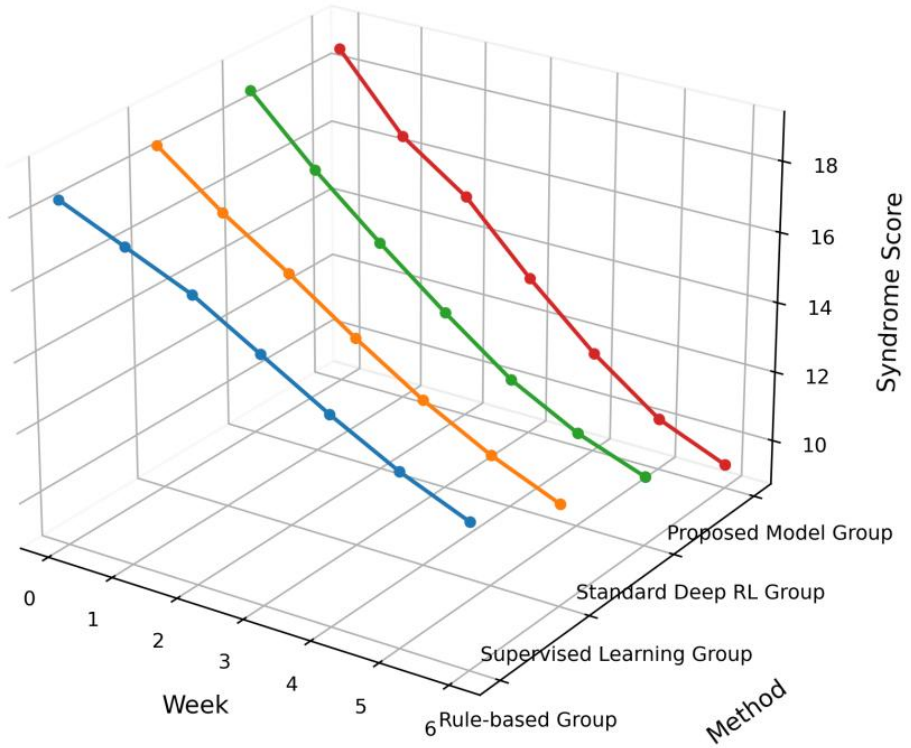

Figure 4: Variation trend of TCM syndrome total integral weekly degree under different methods

This result is closely related to the computational mechanism of the model. In this paper, the main disease, concurrent disease, tongue condition, pulse condition and previous treatment response are mapped into a unified state vector in the state coding, so that the reinforcement learning strategy can identify the complex situations that are common in TCM clinical practice, such as "symptoms are alleviated but not stable" and "symptoms are decreased but signs are not improved synchronously". At the same time, the syndrome score decline and collaborative income were included in the reward function to avoid the model only optimizing around a single western medicine value. Because of this, the treatment path output by the model in this paper is closer to the logic of "adding and subtracting with the syndrome and adjusting according to the mechanism" in the treatment of syndrome differentiation in traditional Chinese medicine, rather than mechanically repeating the existing plan.

5.4 Analysis of collaborative optimization effect of integrated traditional Chinese and Western medicine treatment

The key to the advantages of integrated traditional Chinese and Western medicine is not whether the two kinds of means are used at the same time, but whether they can form a synergistic regulatory relationship with consistent direction and complementary feedback in the continuous treatment process. To this end, the collaborative optimization gain index is used to compare different methods, which measures the additional benefit of combination therapy compared to a single path. As shown in Figure 5, the synergy gain index of the proposed model group was the highest overall, reaching 0.34 at the end of the 6-week intervention, which was significantly higher than 0.12 of the rule-driven group, 0.21 of the supervised learning recommendation group, and 0.28 of the standard deep reinforcement learning group. This shows that the proposed model does not only improve western medicine indicators and traditional Chinese medicine syndromes respectively, but also achieves stronger overall optimization ability at the

joint decision-making level.

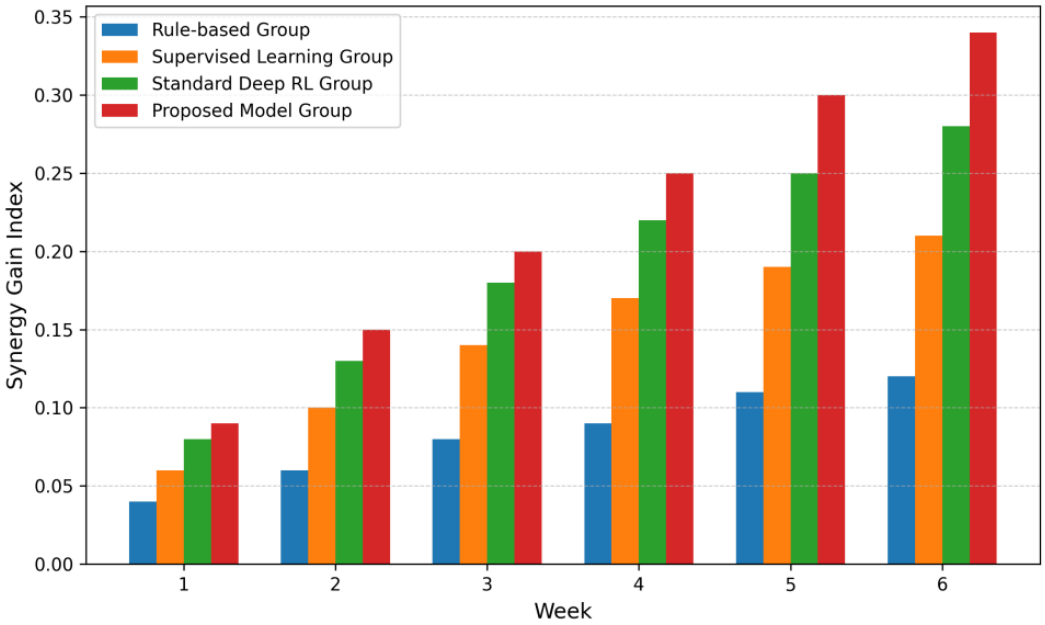

Figure 5: Variation trend of synergy gain index of integrated Chinese and western medicine treatment under different methods

From the perspective of the change process, the synergy gain of the model group in this paper continued to rise from the second week, from 0.09 in the first week to 0.25 in the fourth week, and further increased to 0.34 in the sixth week. In contrast, the increase of rule-driven group was small in the whole intervention phase, and it was always hovering at a low level. Although the supervised learning recommendation group improved compared with the rule method, it tended to be flat after the fourth week. The standard deep reinforcement learning group can form a certain joint benefit, but it is still weaker than the proposed model. This shows that it is difficult to continuously coordinate two types of treatment information simply by relying on fixed rules or static recommendations. However, with the help of the long-term benefit update mechanism of reinforcement learning, the model in this paper can make linkage corrections on the adjustment of western medicine and the addition and reduction scheme of traditional Chinese medicine according to the index response, syndrome changes and treatment feedback in the previous stage, so as to gradually amplify the synergistic effect.

This result shows that the joint decision-making method driven by reinforcement learning not only has the ability of "separate optimization", but also has the feature of "collaborative enhancement" in the treatment scenario of integrated traditional Chinese and Western medicine. Its computational value lies in the unified coding of TCM syndrome information and western medicine index changes, and the reward feed-back driven strategy network is used to learn a better combination path, so as to promote the transformation of combination therapy from experience splicing to data-driven dynamic coordination.

5.5 Analysis of Q value convergence, patient satisfaction and treatment interruption

In addition to the improvement of clinical indicators and syndromes, the effectiveness of reinforcement learning models in medical scenarios is also reflected in the stability of strategy learning, patient acceptance and treatment completion. Based on this, this paper further analyzes the performance of the proposed model from three aspects: Q value convergence,

patient satisfaction and treatment interruption rate. The results show that the proposed model not only shows better strategy convergence ability in the training phase, but also obtains higher patient recognition in the actual treatment sequence, and reduces the dropout phenomenon to a certain extent.

From the perspective of strategy learning process, the Q value of the proposed model rises faster in the early iteration stage, indicating that the model can identify the optimal treatment path from the historical state-action-reward samples at an early stage. After entering the middle and late stage, the fluctuation range of Q value gradually decreases, and basically becomes stable after the 70th round, indicating that the strategy network has transitioned from the exploration stage to the relatively stable utilization stage. In contrast, although the standard deep reinforcement learning model also shows a convergence trend, its later fluctuations are still obvious, indicating that the policy learning is more susceptible to local high reward sample disturbance when the collaborative reward and knowledge constraints of traditional Chinese and western medicine are not introduced. Table 4 shows that the average Q value of the proposed model is stable at 0.842 at the 100th round, and the change range of adjacent rounds is only 0.006, which is lower than 0.011 of the standard deep reinforcement learning group, indicating that the proposed model has better training stability in the combined treatment scenario.

Table 4: Comparison of Q-value convergence and patient satisfaction under different methods

Method	Average Q-Value at Round 20	Average Q-Value at Round 40	Average Q-Value at Round 60	Average Q-Value at Round 80	Average Q-Value at Round 100	Average Q-Value Fluctuation over the Last 20 Rounds	Average Satisfaction (5-Point Scale)
Rule-Driven Group	0.312	0.335	0.341	0.346	0.349	0.004	3.82
Supervised Learning Recommendation Group	0.458	0.536	0.588	0.612	0.627	0.013	4.06
Standard Deep Reinforcement Learning Group	0.521	0.641	0.734	0.789	0.821	0.011	4.29
Proposed Model Group	0.548	0.683	0.771	0.826	0.842	0.006	4.47

In terms of patient satisfaction, the model group maintained a high score after the second week of intervention, and the average satisfaction of the whole treatment cycle reached 4.47 points (5-point scale), which was higher than 3.82 points of the rule-driven group, 4.06 points of the supervised learning recommendation group, and 4.29 points of the standard deep reinforcement learning group. This result shows that the reinforcement learning driven CWM scheme not only produces improvement at the numerical level, but also shows stronger adaptation at the patient perception level. The reason is that the model can adjust the treatment action according to the stage feedback, reduce the discomfort caused by the mutation of the scheme, and make it easier for patients to form an understanding and trust of the treatment path.

Table 5: Variation of patient treatment interruption rate under different methods

Time Point	Rule-Driven Group / %	Supervised Learning Recommendation Group / %	Standard Deep Reinforcement Learning Group / %	Proposed Model Group / %
Cumulative Interruption Rate in Week 2	12.6	10.9	9.4	8.5
Cumulative Interruption Rate in Week 4	18.9	15.8	13.2	11.9
Cumulative Interruption Rate in Week 6	24.8	20.1	16.7	14.4

This was further validated by the treatment interruption situation. Table 5 shows that the cumulative interruption rate of the model group in the second week, the fourth week and the sixth week is 8.5%, 11.9% and 14.4%, respectively, which are the lowest among the four groups. The rule-driven group reached 24.8% at week 6, while the supervised learning recommendation group and the standard deep reinforcement learning group were 20.1% and 16.7%, respectively. From the perspective of the change trend, although the model group in this paper also dropped out to some extent with the progress of treatment, the growth slope was significantly lower than that of the other methods, indicating that it was easier to maintain patient retention in long-term continuous intervention. In general, the proposed model can not only learn a more stable strategy, but also perform better in terms of patient acceptance and continuous participation, which provides more direct support for its usability as an assistant decision-making tool for integrated traditional Chinese and Western medicine treatment.

6 Conclusions and future research directions

Focusing on the problems existing in the process of integrated traditional Chinese and Western medicine, such as insufficient utilization of stage feedback, lag of combined plan adjustment, and difficulty in continuously characterizing individual differences in response, this paper constructed an adaptive optimization model of integrated traditional Chinese and Western medicine based on reinforcement learning. In this study, western medicine clinical indicators, TCM syndrome characteristics, historical treatment trajectories and patient process feedback were integrated into the state-action-reward framework, and the continuous update of the combined treatment strategy was realized through the deep Q network. The results show that the proposed model is superior to the comparison methods in terms of the improvement of western medicine clinical indicators, the alleviation of TCM syndromes, the synergy gain of Chinese and western medicine, the stability of Q value convergence, patient satisfaction and treatment interruption control. It shows that reinforcement learning can not only support the optimization of treatment path, but also adapt to the complex decision-making characteristics of multi-objective parallel and asynchronous feedback arrival in the integrated traditional Chinese and Western medicine scenario.

The main contribution of this paper is that reinforcement learning is not directly applied to medical tasks, but the key links of the model are reconstructed according to the actual needs of integrated traditional Chinese and Western medicine: At the state level, the unified expression

of structural indicators and syndrome semantic information is realized. At the action level, the joint decision-making of western medicine regulation and Chinese medicine addition and subtraction is realized. This makes the model have a strong continuous learning ability, and also makes the computer-aided decision making and clinical combination treatment logic form a relatively close correspondence.

At the same time, there are still some limitations in this study. The data used were mainly from anonymous diagnosis and treatment records of a single center, and the sample coverage and disease diversity were still limited. Although TCM syndrome score has been standardized, its formation process still has certain experience judgment characteristics. The model validation mainly stays at the retrospective data level, and has not yet entered the prospective test in the real clinical process. Therefore, the current results are more suitable to be understood as a validation of the feasibility of the method rather than a definitive conclusion that can directly substitute for clinical decision making.

The follow-up research can be further carried out from three directions. Firstly, the data set of multi-center, multi-disease and longer follow-up period should be expanded to improve the external generalization ability of the model. Secondly, knowledge graph, clinical guideline constraints and human-computer collaborative feedback mechanism are introduced to enhance the interpretability and security of strategy generation. Thirdly, natural language processing and multimodal medical data acquisition technology are combined to further integrate TCM consultation texts, tongue images, pulse signals and western medicine time series monitoring information, so as to promote the integration of traditional Chinese and Western medicine treatment from experience collaboration to a higher level of data-driven collaboration. In general, the adaptive optimization method based on reinforcement learning provides a new path for integrated traditional Chinese and western medicine therapy with computational significance and application potential.

Funding

To study the biological basis of "Qi stagnation, phlegm dampness, internal fire" in NAFLD and the intervention effect of Jiefu Huatan prescription based on succinate-GPR91 non-energy metabolism signal pathway. (No. : CI2023C010LH)

References

[1] Chakraborty B, Murphy S A. Dynamic treatment regimes[J]. Annual review of statistics and its application, 2014, 1(1): 447-464.

[2] Huang X, Choi S, Wang L, et al. Optimization of multi-stage dynamic treatment regimes utilizing accumulated data[J]. Statistics in medicine, 2015, 34(26): 3424-3443.

[3] Zhou X, Chen S, Liu B, et al. Development of traditional Chinese medicine clinical data warehouse for medical knowledge discovery and decision support[J]. Artificial Intelligence in medicine, 2010, 48(2-3): 139-152.

[4] Zhang H, Ni W, Li J, et al. Artificial intelligence–based traditional Chinese medicine assistive diagnostic system: validation study[J]. JMIR medical informatics, 2020, 8(6): e17608.

[5] Pan D, Guo Y, Fan Y, et al. Development and application of traditional Chinese medicine

using AI machine learning and deep learning strategies[J]. The American journal of Chinese medicine, 2024, 52(03): 605-623.

[6] Tian D, Chen W, Xu D, et al. A review of traditional Chinese medicine diagnosis using machine learning: Inspection, auscultation-olfaction, inquiry, and palpation[J]. Computers in biology and medicine, 2024, 170: 108074.

[7] Jayaraman P, Desman J, Sabounchi M, et al. A primer on reinforcement learning in medicine for clinicians[J]. NPJ digital medicine, 2024, 7(1): 337.

[8] Frommeyer T C, Gilbert M M, Fursmidt R M, et al. Reinforcement learning and its clinical applications within healthcare: A systematic review of precision medicine and dynamic treatment regimes[C]//Healthcare. MDPI, 2025, 13(14): 1752.

[9] Liu S, See K C, Ngiam K Y, et al. Reinforcement learning for clinical decision support in critical care: comprehensive review[J]. Journal of medical Internet research, 2020, 22(7): e18477.

[10] Komorowski M, Celi L A, Badawi O, et al. The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care[J]. Nature medicine, 2018, 24(11): 1716-1720.

[11] Wu X D, Li R C, He Z, et al. A value-based deep reinforcement learning model with human expertise in optimal treatment of sepsis[J]. NPJ Digital Medicine, 2023, 6(1): 15.

[12] Wang Y, Liu A, Yang J, et al. Clinical knowledge-guided deep reinforcement learning for sepsis antibiotic dosing recommendations[J]. Artificial intelligence in medicine, 2024, 150: 102811.

[13] Zheng H, Ryzhov I O, Xie W, et al. Personalized Multimorbidity Management for Patients with Type 2 Diabetes Using Reinforcement Learning of Electronic Health Records: H. Zheng et al[J]. Drugs, 2021, 81(4): 471-482.

[14] Wang G, Liu X, Ying Z, et al. Optimized glycemic control of type 2 diabetes with reinforcement learning: a proof-of-concept trial[J]. Nature Medicine, 2023, 29(10): 2633-2642.

[15] Javad M O M, Agboola S O, Jethwani K, et al. A reinforcement learning–based method for management of type 1 diabetes: exploratory study[J]. JMIR diabetes, 2019, 4(3): e12905.

[16] Lyu L, Cheng Y, Wahed A S. Imputation-based Q-learning for optimizing dynamic treatment regimes with right-censored survival outcome[J]. Biometrics, 2023, 79(4): 3676-3689.

[17] Luckett D J, Laber E B, Kahkoska A R, et al. Estimating dynamic treatment regimes in mobile health using v-learning[J]. Journal of the american statistical association, 2020.

[18] Abebe S, Poli I, Jones R D, et al. Learning optimal dynamic treatment regime from observational clinical data through reinforcement learning[J]. Machine Learning and Knowledge Extraction, 2024, 6(3): 1798-1817.

[19] Wang L, Zhang W, He X, et al. Supervised reinforcement learning with recurrent neural network for dynamic treatment recommendation[C]//Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining. 2018: 2447-2456.

[20] Choi Y, Oh S, Huh J W, et al. Deep reinforcement learning extracts the optimal sepsis treatment policy from treatment records[J]. Communications medicine, 2024, 4(1): 245.

[21] Zhang T, Qu Y, Wang D, et al. Optimizing sepsis treatment strategies via a reinforcement learning model[J]. Biomedical Engineering Letters, 2024, 14(2): 279-289.

[22] Ng J Y, Wieland L S, Lee M S, et al. Open science practices in traditional, complementary, and integrative medicine research: A path to enhanced transparency and collaboration[J]. Integrative medicine research, 2024, 13(2): 101047.