Liu *et al.*
Journal of NeuroEngineering and Rehabilitation (2025) 22:107
https://doi.org/10.1186/s12984-025-01609-9

Journal of NeuroEngineering and Rehabilitation

RESEARCH

Open Access

Optimizing the dynamic treatment regime of outpatient rehabilitation in patients with knee osteoarthritis using reinforcement learning

Sijia Liu[1], Jiawei Luo[2] and Chengqi He[1*]

Abstract

Background Knee osteoarthritis (KOA) is a prevalent chronic disease worldwide, and traditional treatment methods lack personalized adjustment for individual patient differences and cannot meet the needs of personalized treatment.

Methods In this study, a dedicated knee osteoarthritis bank (KOADB) was constructed by collecting extensive clinical data from patients. Random forest was used to select the features that had the greatest impact on treatment decisions from 122 questionnaire items. The questionnaire design was optimized to reduce the burden on patients and ensure the validity of data collection. Then, based on the key features screened out, a dynamic treatment recommendation system was constructed by using deep reinforcement learning algorithms, including Deep Deterministic Policy Gradien(DDPG), Deep Q-Network(DQN) and Batch-Constrained Q-learning(BCQ). A large number of simulation experiments have verified the effectiveness of these algorithms in optimizing the treatment strategy of KOA. Finally, the applicability and accuracy of the model were evaluated by comparing the treatment behaviors with actual patients.

Results In the application of deep reinforcement learning algorithms to treatment optimization, the BCQ algorithm achieves the highest success rate (79.1%), outperforming both DQN (68.1%) and DDPG (76.2%). These algorithms significantly outperform the treatment strategies that patients actually receive, demonstrating their advantages in dealing with dynamic and complex decisions.

Conclusions In this study, a deep learning-based KOA treatment optimization model was developed, which was able to adjust the treatment plan in real time and respond to changes in patient status. By integrating feature selection and reinforcement learning techniques, this study proposes an innovative method for treatment optimization, which offers new possibilities for chronic disease management and demonstrates certain feasibility in the development of personalized medicine and precision treatment strategies.

Keywords Knee osteoarthritis, Feature selection, Reinforcement learning, Dynamic treatment

*Correspondence:
Chengqi He
hxkfhcq2015@126.com
[1] Institute of Orthopedics, West China Hospital, Sichuan University, Chengdu 610041, Sichuan, China
[2] West China Biomedical Big Data Center, West China Hospital, Sichuan University, Chengdu 610041, Sichuan, China

Introduction

KOA is a chronic degenerative disease characterized by irreversible structural changes [1]. There has been a significant increase in morbidity, number of patients, and disability-adjusted life years (DALYs) due to KOA [2]. From 1990 to 2017, the number of KOA patients in China increased from 26.1 million to 61.2 million, and the

incidence is also increasing [3]. DALYs per 100,000 people increased from 92.5 to 98.8. Globally, the prevalence of KOA in adults over 40 years of age is approximately 23%, and in adults over 45 years of age, 61% of patients show radiological evidence of KOA, with approximately half of these patients presenting with associated symptoms [4]. Female, obesity, and a history of knee injury are clear risk factors for KOA [5, 6]. The treatment of KOA includes non-drug therapy [7], medication [8], and surgical intervention [9]. With the main goals being pain relief and improvement of functional impairment [10–12]. However, due to the chronic and progressive nature of KOA, its management requires a patient-centered approach focusing on long-term care. This includes early intervention, comprehensive assessment, personalized treatment plans, and regular follow-ups.

In recent years, advancements in deep learning, natural language processing, and computer vision have led to significant progress in the application of artificial intelligence (AI) in medical imaging, personalized treatment, and disease prediction [13–16]. AI offers new possibilities for the diagnosis and treatment of KOA [17]. For example, Bayramoglu et al. [18] proposed a machine learning method for detecting early osteoarthritis. Leung et al. developed a deep learning prediction model using X-ray images and a ResNet34 architecture [19], which performed well in predicting the risk of knee replacement surgery. However, current AI methods face challenges, such as small dataset sizes, limited model generalizability, and a lack of standardization in imaging equipment across different medical institutions, which impacts the clinical applicability and generalization of these models [20, 21].

While AI shows promise in KOA diagnosis and treatment, the development of treatment recommendation systems has lagged [22]. Studies have shown that machine learning-based systems can improve the accuracy and efficiency of disease management [23]. For example, Li et al. developed a KOA management system [24]. Mustaqeem et al. [25] proposed an intelligent recommendation model for heart disease. However, there is still much room for exploration in the clinical validation and application of deep learning-based treatment recommendation systems.

Therefore, this study aims to develop an intelligent KOA rehabilitation treatment decision support system using deep reinforcement learning (DRL) based on patient follow-up data [26, 27]. This system will not only identify key features influencing treatment decisions but also simulate the effects of different treatment plans on patients, optimizing personalized treatment plans. Ultimately, the goal is to achieve precise long-term rehabilitation management and improve the quality of life for KOA patients.

Methods

Study design and data sources

A total of 2836 patients with knee osteoarthritis (KOA) who were admitted to the Rehabilitation Medicine Center of West China Hospital, Sichuan University from January 2012 to December 2023 were included in this retrospective and prospective mixed design. Data sources include:

(1) Historical data (2012–2022): A total of 2224 patients meeting the diagnostic criteria for KOA were screened from the hospital's electronic medical record (EMR) system. These patients underwent a second round of screening based on the inclusion and exclusion criteria (see Sect. "Inclusion and exclusion criteria"), and their contact information was traced using unique patient IDs for follow-up.
(2) Prospective data (2018–2023): An additional 612 patients were enrolled between January and December 2023. These patients were directly recruited by the research team from the outpatient clinics across multiple campuses of West China Hospital.

Inclusion and exclusion criteria

The inclusion criteria were: (1) age between 40 and 80 years old; (2) weight less than 160 kg and body mass index (BMI) below 40 kg/m^2; (3) diagnosis of unilateral or bilateral KOA based on the clinical and radiographic criteria of the American College of Rheumatology (ACR); (4) Kellgren & Lawrence grade II-III (mild to moderate OA) on anteroposterior and lateral X-rays of the knee joint [28].

Exclusion criteria included: (1) severe knee injury or surgery within 6 months prior to the first visit; (2) a history of hip or knee replacement (presence of prosthesis); (3) acute exacerbation of KOA or presence of joint effusion; (4) rheumatoid arthritis, infectious arthritis, or other systemic diseases affecting the knee joint; and (5) severe cardiovascular, respiratory, neurological conditions, vestibular dysfunction, or cognitive impairment. Cases of rheumatoid arthritis, traumatic arthritis, tuberculous arthritis, gouty arthritis, and other knee joint diseases were also excluded. A total of 2836 patients who met the inclusion and exclusion criteria were included in the study.

Data collection and pre-processing

We extracted demographic data (age, gender, height, weight/BMI) and knee-related clinical assessment items (function and structure assessments) from EMR, totaling 122 variables. These variables were collected weekly during the follow-up period by seven medical scales (Visual Analogue Scale, WOMAC OA Index, Activities of Daily Living, Berg Balance Scale, Quality of Life SF-36 Scale, Hamilton Anxiety Scale, Hamilton Depression Scale). Table A-1 of the Appendix lists all the questions selected for this study.

In this study, patients were followed for 52 weeks. If partial data entries were missing during a follow-up visit (e.g., pain score not recorded), two clinicians independently imputed the missing values based on the patient's historical data and clinical expertise, with cross-validation to ensure consistency. Cases with consecutive missing data spanning ≥ 2 weeks were considered lost to follow-up and excluded from the final analysis.

To vividly illustrate the joint distribution of these seven therapeutic actions, we adopted a binary vector representation, where each vector element is strictly limited to a value of either 0 or 1. If a patient receives a relevant treatment at a specific time point, the corresponding variable is assigned a value of 1; otherwise, it is assigned a value of 0. Figures 1 display the frequency of use of these seven treatment behaviors at different time points.

There are three treatment modalities for patients with KOA, covering a total of 7 specific treatment behaviors. This classification is based on the following considerations:

(1) Clinical Practice Basis: Through statistical analysis of KOA clinical data from the West China Biomedical Big Data Center, we found that some treatment modalities had a high use rate in patients and had similar treatment goals. Therefore, we categorized the treatment modalities based on the actual treatment trajectory.

(2) Data-driven classification approach: We first screened out key features from 122 treatment-related variables and applied a random forest model combined with hierarchical clustering to identify treatment patterns [29, 30]. Statistical results demonstrated that while 128 therapeutic combinations were theoretically possible, only 9 combinations were clinically applied (Table 1). To ensure computational efficiency and clinical applicability of the model, we ultimately selected three core treatment modalities:

1. Pharmacological Treatment: including glucosamine and non-steroidal anti-inflammatory drugs (NSAIDs).
2. Physical Therapy: including Power Bikes, Medium-frequency Pulsed Electrical Therapy, Ultra-shortwave Therapy and Ultrasound Therapy.
3. Injection Therapy: Intra-articular injections of platelet-rich plasma (PRP) and hyaluronic acid (HA).

In the reinforcement learning model, PRP and HA were classified under the "injection therapy" category, as both are intra-articular (I.A.) interventions primarily aimed at improving joint function and alleviating symptoms. Although PRP exhibits stronger tissue regeneration potential compared to HA's primary role in joint lubrication, clinical data analysis revealed that these two therapies are frequently used complementarily or

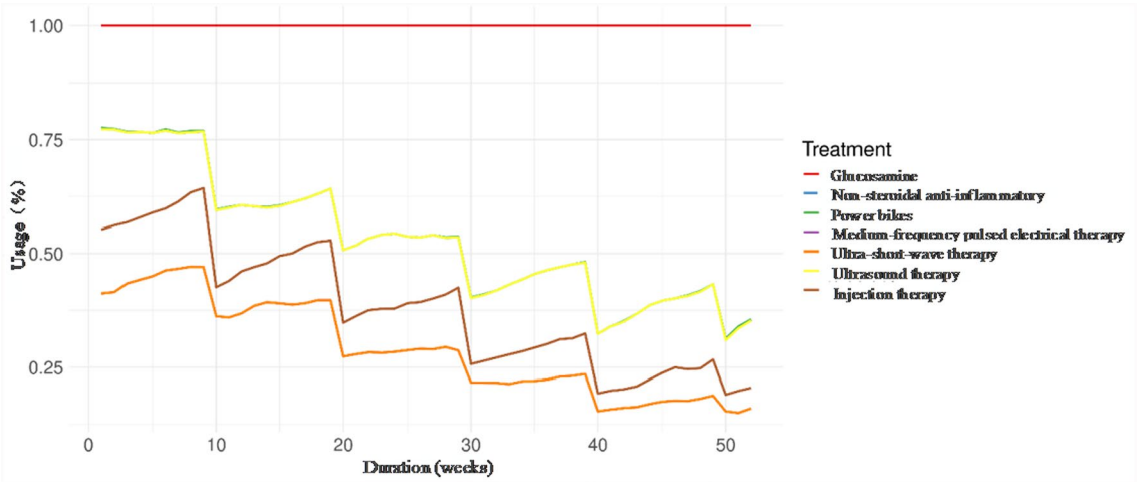

Fig. 1 Utilization of therapeutic behaviors at each time point

Table 1 Frequency of different combinations of behaviors

Number	Binary representation	Decimal representation	Frequency	Therapeutic behavioral combinations
1	1,000,000	64	52,215	Glucosamine
2	1,000,001	65	16,827	Glucosamine + Injection therapy
3	1,010,000	80	60	Glucosamine + Power bikes
4	1,010,001	81	44	Glucosamine + Power bikes + Injection therapy
5	1,110,000	112	12	Glucosamine + NSAIDs + Power bikes
6	1,110,010	114	22,662	Glucosamine + NSAIDs + Power bikes + Ultrasound therapy
7	1,110,011	115	12,871	Glucosamine + NSAIDs + Power bikes + Ultrasound therapy + Injection therapy
8	1,111,110	126	16,356	Glucosamine + NSAIDs + Power bikes + Medium-frequency pulsed electrical therapy + Ultra-short-wave therapy + Ultrasound therapy
9	1,111,111	127	26,425	Glucosamine + NSAIDs + Power bikes + Medium-frequency pulsed electrical therapy + Ultra-short-wave therapy + Ultrasound therapy + Injection therapy

NSAIDs Non-steroidal anti-inflammatory drugs

interchangeably in practice, justifying their unified classification. Glucosamine, despite not being universally recommended as a first-line treatment in some guidelines, was included as an independent therapeutic variable due to its widespread utilization in Chinese clinical practice, long-term safety profile, and potential symptomatic benefits. Importantly, the classification strategy prioritizes real-world treatment patterns over strict pharmacological distinctions, as the model's objective is to optimize global therapeutic pathways rather than dissect isolated biological mechanisms of individual interventions. This approach aligns with the reinforcement learning framework's emphasis on dynamic decision-making and practical applicability.

Feature selection

Among the clinical assessment variables, it was difficult for us to determine which features were crucial for optimizing treatment, resulting in the potential existence of some redundant features. When the number of features is vast, the algorithm is often susceptible to the "curse of dimensionality" [31]. Feature selection stands as a significant method for dimensionality reduction in high-dimensional data [29].

In this study, we first calculated and ranked the importance of all features using the Random Forest (RF) algorithm, an ensemble learning method that consists of multiple decision trees [30]. This method exhibits high tolerance for outliers and noise, and moreover, it is capable of providing the importance of each characteristic variable. We employed a bootstrap method with replacement to randomly select samples from the data, built multiple decision trees by randomly splitting each resampled dataset, and ultimately made predictions through a voting mechanism. We utilize the Out-Of-Bag (OOB) error [32] to quantify the importance G_i of each feature X_i, which provides an unbiased estimate of the model error rate:

$$G_i = \frac{1}{B}\sum_{j=1}^{B}\left|D_j - D_{ji}\right| \tag{1}$$

Here, B denotes the number of bootstrap samples, and i represents the i^{th} feature. D_j is the number of correct classifications by OOB, and D_{ji} is the number of correct classifications of samples after feature X_i is perturbed, indicating the feature's contribution to model accuracy.

After the random forest model preliminarily ranked the feature importance scores, the iterative feature elimination (IFE) method was used to gradually remove the features with low importance, and the model performance was evaluated, Fig. 2 shows the execution process of the iterative feature elimination algorithm, when the number of features is 20, the model achieves the highest accuracy of 0.9999. Finally, the 20 most important features that have a critical impact on treatment decisions for patients with knee osteoarthritis are retained.

After the random forest model ranked the features based on their importance scores, we applied iterative feature elimination (IFE) to iteratively remove low-importance features while monitoring model performance. Through this process, we identified candidate features that significantly influenced model performance. To further refine the feature set, we implemented a sequential backward selection (SBS) algorithm [33, 34], using pain scores as the dependent variable to assess classification accuracy. As shown in Fig. 1, the model achieves the highest accuracy of 0.9999 when the number of features is 20. In the end, we retained the 20 most critical features that had a significant impact on treatment decisions for patients with knee osteoarthritis (Fig. 3).

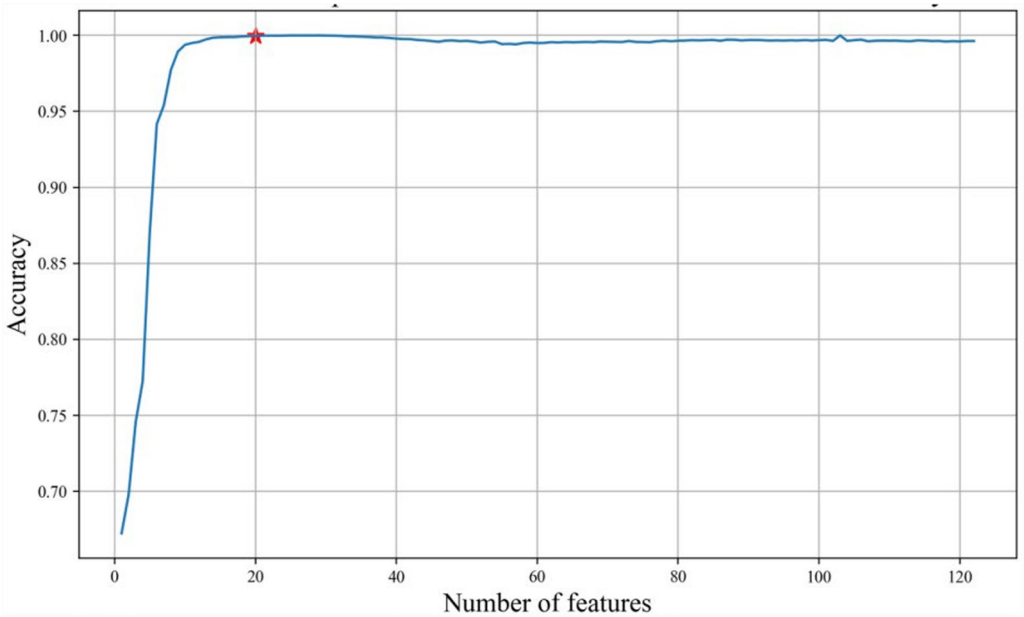

Fig. 2 The relationship between the number of features and the accuracy

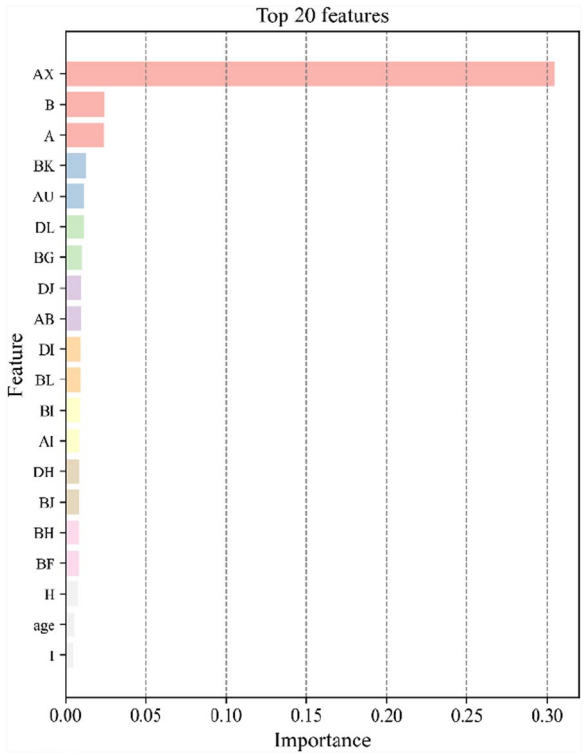

Fig. 3 Feature importance ranking

All patients were randomly divided into a training group (70%, 1985 patients), a test group (20%, 567 patients), and a validation group (10%, 284 patients).

Model building

Based on the selected key features, this paper uses reinforcement learning algorithms (including DDPG, DQN and BCQ) to construct a dynamic treatment recommendation system. Specifically, the construction of a dynamic treatment recommendation model for KOA patients is formalized as a Markov decision-making process with a finite time step [35, 36], as shown in Fig. 4. This decision-making process consists of a behavior space $\mathcal{A} \in \{0,1\}^{K}$, a state space $\mathcal{S}$, and a reward function $\mathcal{R} : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$. At each time step t, the physician observes the patient's current state $s_t \in \mathcal{S}$, selects $\widetilde{a}_t \in A$ from the behavioral space as the patient's current treatment based on the strategy function $\widetilde{\pi}(s_t)$, and receives immediate feedback r_t.

Reward function

This study designed a reward function f to evaluate the effectiveness of medical interventions for patients with KOA [37], by assessing the change in pain score of the score over time in patients with KOA. This function includes additional thresholds for more granular assessment of patient outcomes. The reward function f is a piecewise function that depends on the KOA pain score_t and score_{t+1}, and a set of coefficient and threshold parameters for two consecutive weeks. The definition is as follows:

$$r_t = f(\text{score}_t, \text{score}_{t+1}, \theta_1, \theta_2, \alpha, \beta, \gamma, \delta) \quad (2)$$

where score_t and score_{t+1} represent the KOA pain score at t week and $t+1$ week, respectively. θ_1 and θ_2 are the thresholds that define different intervals of the KOA pain

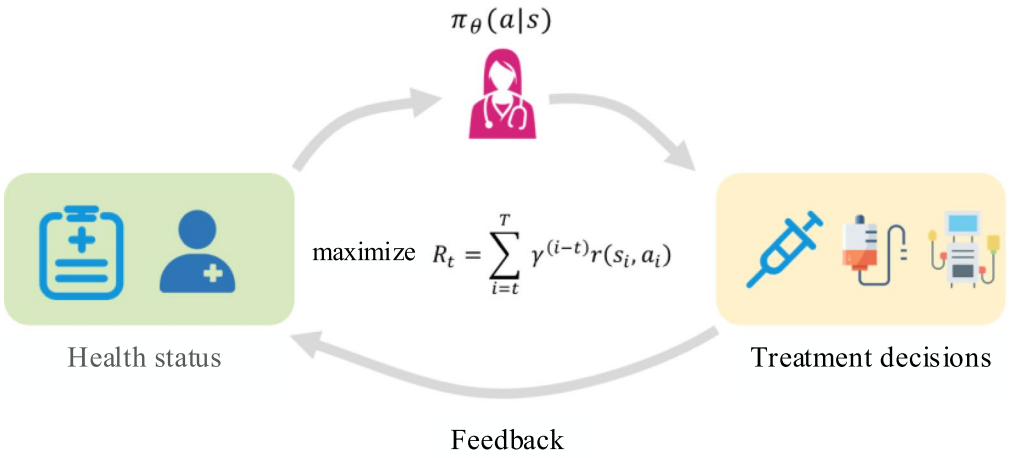

Fig. 4 Dynamic treatment optimization process for KOA patients

score, α,β,γ and δ are the coefficients for adjusting reward or punishment.

There are three types of function f:

1. Improvement (Score Reduction): If $\text{score}_{t+1} < \text{score}_t$, the function calculation reward is:

$$(1 + \alpha \times \text{score}_{t+1}) \times (\text{score}_{t+1} - \text{score}_t) \tag{3}$$

(2) Stable (No change in score): If $\text{score}_{t+1} = \text{score}_t$, different rules are applied according to the rating level:

a. If the score of the moment t is greater thanθ_2, return to $-\beta \times \text{score}_t$, penalties for high but stable ratings.

b. If the score of the moment t is less than or equal toθ_1, it will return γ, and the reward is very low and stable.

c. In other cases, a return of zero indicates that a medium stable score is neither rewarded nor punished.

(3) Deterioration (score increase): If $\text{score}_{t+1} > \text{score}_t$, the penalty for function calculation is:

$$-(1 + \delta \times \text{score}_t) \times (\text{score}_{t+1} - \text{score}_t) \tag{4}$$

where θ_1 and θ_2 are 0, 3, respectively.α and β set to 0.1,γ set to 0, δ set to 0.1.

BCQ algorithm

BCQ is an offline reinforcement learning algorithm that avoids accessing unseen state-action pairs by constraining action selection, thereby improving sample efficiency and strategy performance [38].

BCQ uses neural networks Q to estimate state-action value functions $Q(s,a)$. Network Q consists of two subnetworks, the first for estimating the value and the second for estimating the relative advantage function $A(s,a)$ for each action taken in a given state. The update of the Q network follows the rules of time-series differential learning, and its loss function is:

$$L(\theta) = \mathbb{E}_{(s,a,r,s')\sim D}\left[\left(r + \gamma \max_{a'} Q_{\overline{\theta}}(s',a') - Q_\theta(s,a)\right)^2\right] + \mathbb{E}_{(s,a)\sim D}\left[-\log \frac{\exp(A_\theta(s,a))}{\sum_{a'} \exp(A_\theta(s,a'))}\right] + \alpha \mathbb{E}_{s\sim D}\left[\frac{1}{|\mathcal{A}|}\sum_a A_\theta(s,a)^2\right] \tag{5}$$

where θ and $\overline{\theta}$ are the parameters of the Q with the first quartile and network and the target Q network, respectively. D represent the offline dataset, γ is the discount factor, and α is the regularization coefficients. The second and third terms of the loss function correspond to the maximization and regularization of the dominance function, respectively. In the action selection stage, BCQ first calculates the relative advantage of each action according to the dominance function, and then normalizes it and compares it with a threshold τ to obtain a binary action mask:

$$m(a|s) = 1\left(\frac{\exp(A_\theta(s,a))}{\max_{a'} \exp(A_\theta(s,a'))}\right) > \tau \tag{6}$$

Finally, the BCQ selects the action with the maximum Q value, and uses the action mask to constrain the action selection:

$$a = \arg\max_{a} \left(m(a|s) \cdot Q_\theta(s,a) - (1 - m(a|s)) \cdot C\right) \tag{7}$$

where C is a large negative constant that is used to rule out unseen actions. In this way, BCQ can learn a robust and efficient strategy from offline datasets without accessing the environment.

DDPQ algorithm

DDPG is a reinforcement learning algorithm for continuous action spaces [39]. It combines the ideas of DQN [40] and DPG (Deterministic Policy Gradient) [41].

The DDPG uses the Actor-Critic architecture and consists of two main components:

(1) Actor Network $\mu(s|\theta^\mu)$: It is a deterministic strategy that outputs deterministic actions a for a given state s.
(2) Critic network $Q(s,a|\theta^Q)$: It estimates the Q value of the action a taken in the state s.

DQN algorithm

DQN is a value-based reinforcement learning algorithm used to solve continuous control problems with high-dimensional state spaces [42, 43]. DQN uses a deep neural network to approximate the state-action value function $Q(s,a)$. The core idea of DQN is to learn the optimal Q function by minimizing the timing difference (TD) error [44]. TD error is defined as:

$$L(\theta) = \mathbb{E}_{(s,a,r,s')\sim D}\left[\left(r + \gamma \max_{a'} Q_{\theta^-}(s',a') - Q_\theta(s,a)\right)^2\right] \tag{8}$$

where θ is the parameter of the Q network, θ^- is the parameter of the target network, and D is the empirical replay buffer, γ is the discount factor.

Model training and validation

During model training, DDPG uses off-policy data and Bellman's equation to update the Critic network [38, 45]. The loss function of a Critic is defined as:

$$L(\theta^{\mathrm{Q}}) = \mathbb{E}_{(\mathrm{s,a,r,s'})\sim \mathrm{D}}\left(\mathrm{Q}\left(\mathrm{s,a}|\theta^{\mathrm{Q}}\right) - \mathrm{y}\right)^2 \tag{9}$$

where $y = r + \gamma Q'(s', \mu'(s'|\theta^{\mu'})|\theta^{Q'})$, D is the replay buffer, γ is the discount factor. The goal of the Actor network is to maximize the desired Q value. Therefore, the Actor's loss function is defined as:

$$\nabla_{\theta^\mu} J \approx \mathbb{E}_{s\sim D}\left[\nabla_a Q\left(s,a|\theta^Q\right)|a = \mu(s)\nabla_{\theta^\mu}\mu\left(s|\theta^\mu\right)\right] \tag{10}$$

To improve training stability, DDPG introduces two techniques:

(1) Target Networks: Create a slowly updating target network for both Actor and Critic.
(2) Exploration Noise [46]: Add noise to the Actor's output to encourage exploration. In each training iteration, a batch of transfers (s,a,r,s') is sampled from the replay buffer D. Then, use gradient descent to update the Critic and Actor networks. Finally, update the target network with a soft update:

$$\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'} \tag{11}$$

$$\theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'} \tag{12}$$

where $\tau \ll 1$ is the soft update rate. By combining the ideas of DQN and DPG, DDPG uses the Actor-Critic architecture and some training techniques to achieve effective policy learning on a continuous action space [47].

DQN samples a batch of transfer data (s,a,r,s') from the empirical replay buffer D and then uses gradient descent to minimize TD error and update the parameters θ of the Q network. The formula for updating the Q network is:

$$\theta \leftarrow \theta - \alpha\nabla_\theta L(\theta) \tag{13}$$

where α is the learning rate. To improve training stability, DQN introduces two important tips:

Experience Replay [48]: Store the transfer data in a buffer and randomly sample it to update the network parameters to break the correlation between the data.

Target Network: The TD target value is calculated using a separate target network, and the parameters θ^- of the target network are copied from the Q network at regular intervals to reduce the instability of training. When selecting actions, DQN uses the ϵ-greedy strategy to randomly select actions with the probability of ϵ, otherwise the action with the largest Q value is chosen:

$$a = \begin{cases} \arg\max_{\mathrm{a}} \mathrm{Q}_\theta(\mathrm{s,a}) & \text{with probability } 1 - \varepsilon \\ \text{random action} & \text{with probability } \varepsilon \end{cases} \tag{14}$$

Of these, ϵ is usually decayed gradually as training progresses to balance exploration and utilization. Through the above training process, DQN can learn an approximate optimal Q function and make decisions in the environment according to the learned strategy to achieve end-to-end reinforcement learning.

Through a large number of simulation experiments, the three reinforcement learning algorithms have

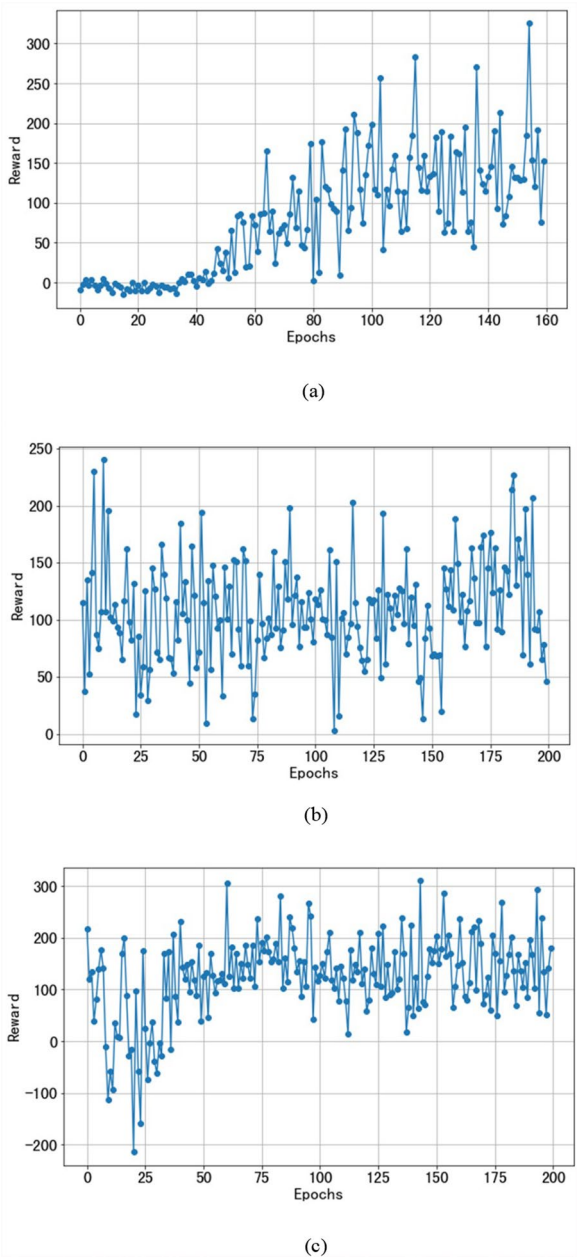

Fig. 5 The change trajectory of the average cumulative reward value during the training process of different models; **a**–**c** represent the training trajectories of BCQ, DDPG, and DQN, respectively

shown remarkable effects in optimizing treatment strategies. As shown in Fig. 5, the average reward values of the recommended behaviors of the DDPG, DQN, and BCQ algorithms on the training set are significantly better than those of the KOA patients in the original data. This suggests that our proposed model can be effectively learned and generalized to actual treatment decisions, which in turn may improve patient outcomes.

Table 2 Performance of the three reinforcement learning algorithms on the test set

Algorithm	R'	$R' - R$	Success rate
BCQ	32.6 (– 2.20, 228.05)	36.1 (– 1.30, 230.00)	0.791
DDPG	54(0.90, 190.15)	52.1(1.10, 186.60)	0.762
DQN	59.9(– 10.65, 276.80)	57.9(10.85, 278.75)	0.681

We use the saved model parameters to evaluate the actual performance of the strategy on the test set patient data. For each patient in the test set, we simulated the entire treatment process according to their historical treatment trajectory (state and original action sequence (s, a)). On this basis, we do not directly change the original action sequence, but use the trained strategy model π to infer the possible optimal action a'_t and the possible reward r'_t in each state. In this way, the state s_t of each step, the original action a_t, and the obtained counterfactual reward r'_t combine to form a counterfactual decision trajectory

$$s_1, a_1, r'_1, s_2, \ldots, s_T, a_T, r'_T. \tag{15}$$

where T represents the total number of steps taken throughout the treatment process. Define the cumulative counterfactual reward R' as:

$$\mathrm{R}' = \sum_{t=1}^{T} r'_t \tag{16}$$

At the same time, we also calculate the actual cumulative rewards R generated according to the original strategy:

$$R = \sum_{t=1}^{T} r_t \tag{17}$$

Finally, by comparing R' and R, calculate $\Delta R = \mathrm{R}' - \mathrm{R}$. We can calculate the proportion of patients whose cumulative reward exceeds the original strategy's cumulative reward when adopting the new strategy as a measure of the new strategy's effectiveness:

$$\text{Success Rate} = \frac{\sum_{i=1}^{N} 1(\Delta R_i > 0)}{\sum_{i=1}^{N} 1} \tag{18}$$

where N is the total number of test samples.

Results

Table 2 shows the performance of the three reinforcement learning algorithms on the test set in detail. Specifically, the median value of the inverse factual reward R' of the BCQ algorithm is 32.6, and the first and third quartiles are − 2.2 and 228.05, respectively. Compared to the actual reward, the median of the $R' - R$ was 36.1,

and the first and third quartiles were − 1.30 and 230.00, respectively, indicating a success rate for treatment optimization. The DDPG algorithm outperformed in this evaluation, The median of R' is 54, the first and third quartiles are 0.9 and 190.15, respectively. The median value of the $R' - R$ is 52.1, with the first quartile and the third quartile being 1.10 and 186.6, respectively. The success rate of treatment optimization reached 76%. Meanwhile, the optimized median R' value for the DQN algorithm is 59.9, with the first quartile and the third quartile being -10.65 and 276.8, respectively. The median value of $R' - R$ is 57.9, with the first quartile and the third quartile being 10.85 and 278.75, respectively. Demonstrating a 68.1% success rate of treatment optimization.

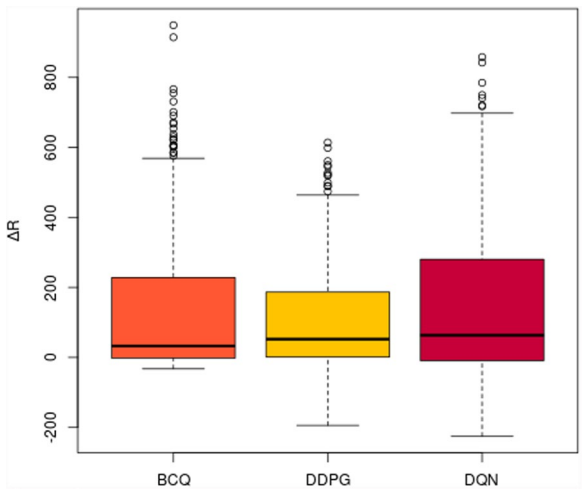

Fig. 6 Distribution of ΔR for three reinforcement learning algorithms

Overall, as evident from the data in Table 2, the BCQ algorithm performs best among these three algorithms, achieving the highest success rate of treatment optimization. Figure 6 further illustrates the detailed distribution of the $\Delta R = R' - R$ values.

By testing all time points for all patients with KOA in the test set using three trained deep reinforcement learning algorithms, we attempted to compare the distribution differences between the optimized combinations of treatment behavior choices and the actual treatment choices received by the patients. Figures 7, 8, 9 illustrate the distribution differences between the recommended and actual behaviors across all time points in the test set for the three algorithms.

The DQN algorithm exhibited a pronounced bias in its recommendations, particularly for the third therapeutic behavior combination (glucosamine + power bikes), which was recommended 26,373 times. In contrast, this combination was accepted only 14 times in real-world scenarios. This substantial discrepancy indicates that the DQN algorithm fails to adequately capture patients' actual behavioral preferences, potentially due to overfitting to specific features during training. Furthermore, while the DQN algorithm demonstrates relatively balanced recommendation frequencies across other treatment combinations, these frequencies remain markedly lower than real-world acceptance rates. For instance, the first treatment combination (glucosamine) was recommended merely 331 times, whereas its actual acceptance count reached 9718 instances. A detailed distribution comparison between the DQN-recommended behaviors and real-world behaviors is provided in Fig. 7.

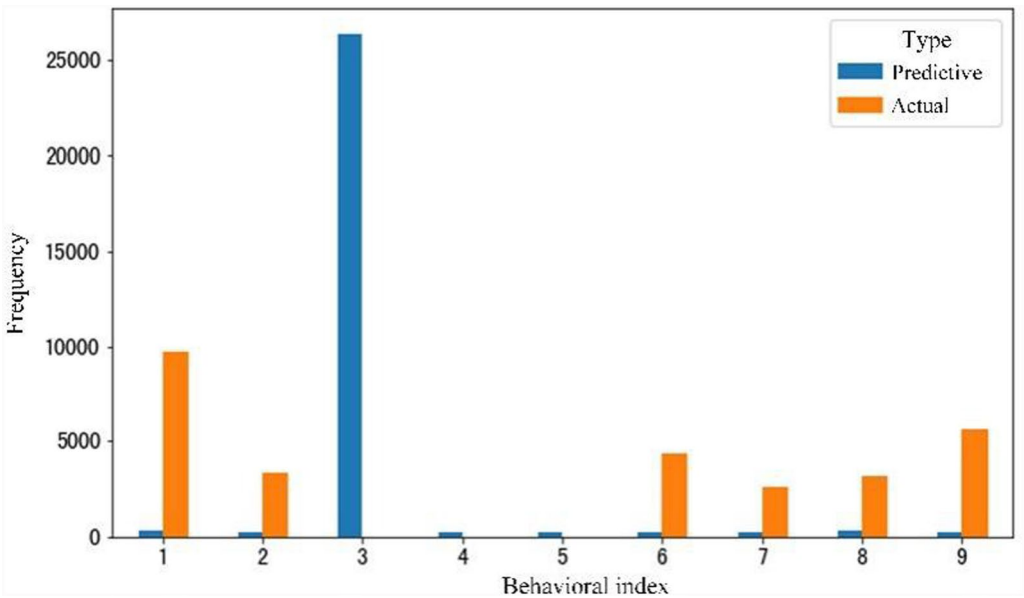

Fig. 7 DQN Recommendations vs. Actual Behavior Distribution

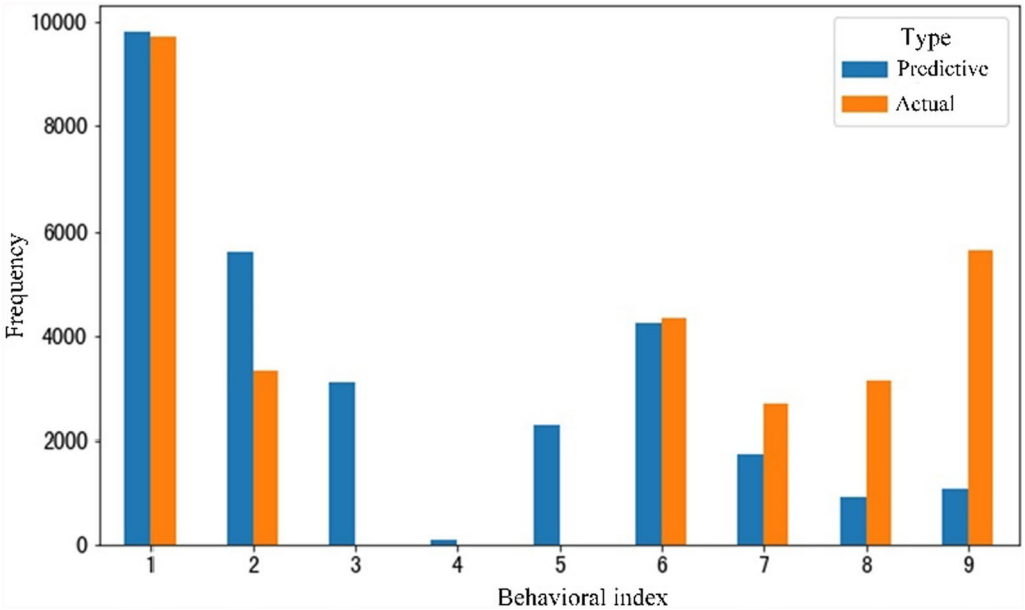

Fig. 8 BCQ Recommendations vs. Actual Behavior Distribution

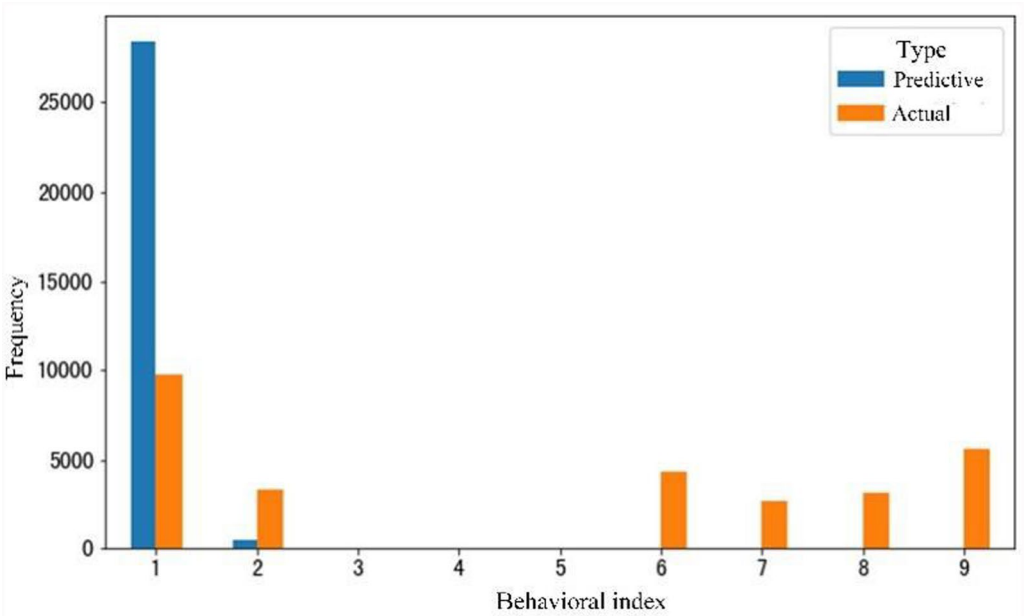

Fig. 9 DDPG Recommendations vs. Actual Behavior Distribution

Compared with the DQN algorithm, the BCQ algorithm demonstrates a more balanced recommendation distribution, with its outputs aligning more closely to patients' actual acceptance patterns. This improvement is particularly evident in the recommendations for the 5th therapeutic combination (glucosamine + NSAIDs + power bike) and the 7th combination (glucosamine + NSAIDs + power bike + ultrasound therapy + injection therapy). The BCQ algorithm recommended these combinations 4237 and 925 times, respectively. While their real-world acceptance counts reached 4343 and 3158 instances. These results suggest that BCQ better simulates clinicians' decision-making logic in treatment selection, likely attributable to its enhanced generalization capability and sensitivity to minority samples when processing heterogeneous data features. A

comparative analysis of the BCQ-recommended behaviors versus real-world behaviors is illustrated in Fig. 8.

The DDPG algorithm exhibited an overly conservative recommendation pattern, with an extreme bias toward the first treatment combination (glucosamine only). This combination was recommended 28,450 times, significantly exceeding its real-world acceptance count of 9718 instances. Notably, DDPG failed to recommend any other treatment combinations, suggesting potential limitations in its ability to generalize from training data or a skewed prioritization of specific features within the training set. A comparative analysis of the DDPG-recommended behaviors versus real-world behaviors is provided in Fig. 9.

Discussion

To identify the most influential features affecting the policy functions of different DRL algorithms, we calculated Shapley values for policy functions derived from three distinct DRL models. The Shapley value—a well-established concept in game theory and economics—quantifies a fair contribution of individual participants to collective outcomes in cooperative games [49]. Specifically, this Shapley value approach was employed to analyze feature importance in the algorithms' decision-making processes.

In the application of the Batch-Constrained deep Q-learning (BCQ) algorithm, the dominant features influencing strategy selection reflect direct correlations with knee osteoarthritis (KOA) patients' capacity to perform daily living activities. The characterizing Shapley values of the BCQ algorithm's policy function are presented in Fig. 10. In contrast, the Deep Deterministic Policy Gradient (DDPG) algorithm prioritizes fundamental physiological parameters. Age and weight emerge as the two most critical determinants, indicating that treatment strategies are strongly associated with patients' physiological status. Furthermore, ambulatory capacity on level surfaces, body height, and stair negotiation ability were also identified as significant features impacting the DDPG algorithm's decision-making process. These findings suggest the algorithm's reliance on comprehensive assessments of patients' global functional mobility. The corresponding Shapley values for the DDPG policy function are illustrated in Fig. 11.

In the analysis of the DQN algorithm, age, weight, and height are also determined to be the most significant influencing factors. This highlights the crucial importance of patients' basic physiological attributes in the selection of treatment strategies. At the same time, pain scores and the ability to walk on flat ground are identified as key features, reflecting the algorithm's focus on pain management and patients' daily activity capabilities. Figure 12 presents the Shapley values of the features for the DQN algorithm's policy function.

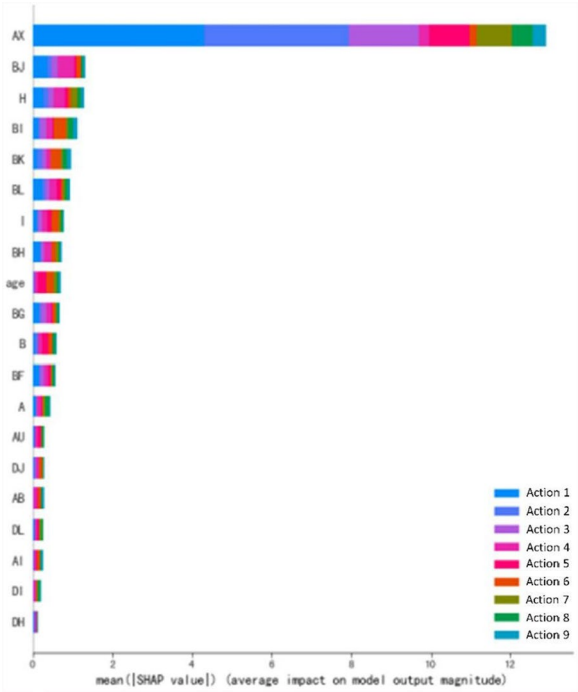

Fig. 10 BCQ policy function Shapley values

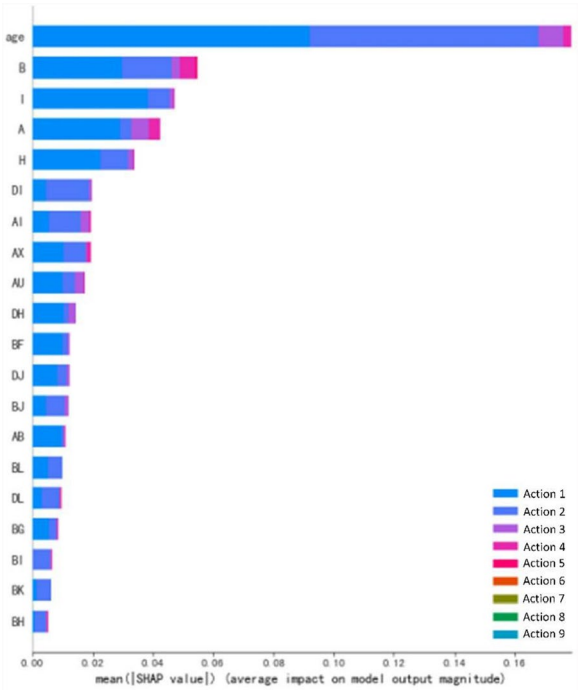

Fig. 11 DDPG policy function Shapley values

In summary, different algorithms exhibited significant differences in their treatment recommendations. Although the DQN algorithm performs well in terms of

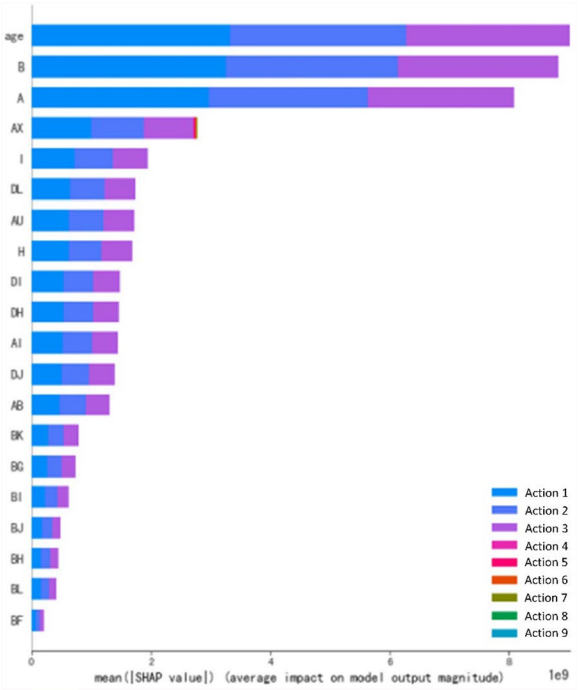

Fig. 12 DQN policy function Shapley values

diversity, it demonstrates an obvious bias in accurately matching the treatment selection of real patients. The BCQ algorithm has shown good adaptability and accuracy in simulating real treatment behaviors, suggesting it may be more suitable for integration into actual clinical decision support systems. The DDPG algorithm exhibits limitations in achieving policy diversity and complexity, and its policy learning mechanism requires further adjustment and optimization.

In this study, we categorized the treatment modalities for KOA and optimized the treatment strategy using reinforcement learning models. However, we acknowledge that there are some limitations to this classification method, which include the following:

1.Heterogeneity of injection treatments: PRP and HA have different biological mechanisms and clinical effects, they were grouped into the same category in the model because they both belong to the intra-articular injections and are often used interchangeably in clinical practice. However, this classification may ignore the actual efficacy differences between the two treatments in patients with KOA. Future studies could further refine the subcategories of injection treatments, (e.g., modeling PRP and HA separately) to improve the accuracy of individualized recommendations.

Clinical applicability of glucosamine: Our analysis found that glucosamine was frequently used among patients, so it was included in the treatment optimization model. However, it has a low recommendation level in international guidelines, and its clinical efficacy remains controversial according to some studies. Future research should incorporate data from larger-scale randomized controlled trials to validate its role in individualized treatment.

Dynamic adjustment of treatment modalities: Current models learn from historical treatment data but fail to fully account for disease progression in individual patients during different treatment stages. For example, some patients may prefer physical therapy early in the disease course but require medications or injections as the disease progresses. Therefore, time series modeling or causal inference methods could be introduced in the future to enhance the model's adaptability to dynamic treatment adjustments.

Surgical treatment options not included: This study focuses on non-surgical treatment only, whereas joint replacement remains the ultimate option for patients with severe KOA. Future research could be extended to surgical decision optimization and exploration of personalized strategies for both surgical and non-surgical treatments.

In summary, this study adopted a data-driven approach for classifying treatment modalities and implemented optimization through reinforcement learning. Nevertheless, more refined classification methods are needed to improve the clinical applicability of the model and the effectiveness of individualized recommendations.

Conclusion

In this study, we constructed a dynamic treatment recommendation system by integrating feature selection techniques with reinforcement learning algorithms, specifically DDPG, DQN, and BCQ. The efficacy of these algorithms in optimizing KOA treatment strategies was rigorously validated through extensive simulation experiments. Furthermore, to assess the suitability and precision of our model's recommendations, we compared them to the actual treatment behaviors received by patients. The results revealed that the BCQ algorithm exhibited the highest performance, achieving a treatment optimization success rate of 79.1%, while the DQN and DDPG algorithms followed closely with success rates of 68.1% and 76.2%, respectively. Notably, these algorithms significantly outperformed the actual treatment strategies employed by patients, demonstrating their prowess in navigating dynamic and complex decision-making landscapes. Our findings offer an innovative approach to optimizing KOA treatment, opening up new avenues for chronic disease management and showcasing the feasibility of personalized medicine and precision treatment strategies.

Appendix

See Table 3

Table 3 A list of questions from the questionnaire used in this study

Serial number	Question (variable name)	Variable name identifier	Description
1	Height (unit: cm)	A	Height (unit: cm)
2	Body weight (unit: kg)	B	Body weight (unit: kg)
3	The patient's primary diagnosis	C	Patient's primary diagnosis (knee osteoarthritis if not confirmed)
4	Stool control	D	Stool Control Ability: 0 = incontinence or coma 5 = occasional incontinence (< 1 per week); 10 = controllable
5	Urination control	E	Urine Control Ability: 0 = Incontinence or coma or need for human catheterization 5 = Occasional incontinence (1 < every 24 h, 1 > per week) 10 = controllable
6	Toilet	F	Toilet ability (refers to the process of going to the toilet by yourself: including going to the toilet, untying clothes, wiping, tidying up, flushing): 0 = dependence on others 5 = Partial help required 10 = Self-care
7	Embellish	G	Perform personal hygiene (e.g. washing your face, brushing your teeth, combing your hair, shaving, etc.): 0 = Need help 5 = Wash your face, comb your hair, brush your teeth, shave independently
8	Up and down stairs	H	Ability to go up and down stairs (up and down a flight of stairs is considered independent with a cane): 0 = No 5 = Need for assistance (physical or verbal instruction) 10 = Self-care
9	Walk on flat ground	I	Ability to walk on flat ground (primarily short walks, e.g. in and around hospital rooms, excluding long walks): 0 = immobile 5 = Walk independently in a wheelchair 10 = 1 person assisted to walk (physical or verbal instruction) 15 = Independent walking
10	Bed and chair transfer	J	Bed-to-chair and chair-to-bed transfer capabilities: 0 = Completely dependent on others, unable to sit 5 = Partial help required 10 = Full self-care
11	Dressing	K	Ability to dress (including putting on and taking off clothes, buttoning, zipping, putting on and taking off shoes and socks, tying shoes: 0 = Dependent 5 = Half help 10 = Self-care (buttoning, closing, zipping and putting on shoes)
12	Bathe	L	Ability to bathe alone: 0 = Dependent 5 = Self-care
13	Eating	M	Feeding ability (refers to the process of transporting food from the container to the mouth, chewing, swallowing, etc., e.g. sandwiching, serving, cutting bread with utensils): 0 = dependence on others 5 = Partial help required (rice picking, rice serving, bread cutting) 10 = Full self-care

Table 3 (continued)

Serial number	Question (variable name)	Variable name identifier	Description
14	From sitting to standing	N	Subject position: The patient is seated on a treatment table. Test command: Please stand up and try not to help with your hands 4 points: Able to stand up independently and maintain stability without hand support; 3 points: Able to stand up independently with the support of the hand; 2 points: Stand up on your own after a few attempts 1 point: Requires a small amount of help to stand up or stay stable Score 0: Requires moderate or significant assistance from others to be able to stand up or remain steady
15	Stand independently	O	Subject position: Standing position. Test command: Please try to stand as strong as possible 4 points: Able to stand safely for 2 min; 3 points: Able to stand under supervision for 2 min; 2 min: Able to stand independently for 30 s; 1 point: It takes several attempts to stand independently for 30 s; 0 points: Cannot stand for 30 s without assistance
16	Sit independently	P	Subject's position: Sit in a chair with feet flat on the floor and back off the chair. Test command: Please hold your upper limbs crossed in front of your chest and sit as firmly as possible 4 points: Able to safely remain seated for 2 min; 3 points: Able to sit under supervision for 2 min; 2 min: Able to sit for 30 s; 1 min: Able to sit for 10 s; 0 points: Can't sit for 10 s without backrest support
17	From standing to sitting	Q	Subject position: Standing position. Test Command: Please sit down and try not to help with your hands 4 points: Able to sit down safely with a little help from your hands; 3 points: need to use hand help to control the downward shift of the body's center of gravity; 2 points: You need to use the back of your legs against the chair to control the downward shift of your body's center of gravity; 1 point: Able to sit independently in a chair but unable to control the downward shift of the body's center of gravity; 0: Need help to sit down
18	Bed-chair transfer	R	Start by preparing a chair with armrests and a chair without armrests next to the treatment table. Subject's position: The patient is seated on a treatment table with their feet flat on the floor. Test Command: Please sit on a chair with armrests, then sit back on the bed; Then sit down in a chair without armrests and sit back on the bed 4 points: With a little help from your hands, you can transfer safely; 3 points: Hand assistance is required to be able to transfer safely; 2 points: Guardianship or verbal cues are required to complete the transfer; 1 point: One person is required to complete the transfer; 0 points: Two people are required to help or supervise the transfer to complete

Table 3 (continued)

Serial number	Question (variable name)	Variable name identifier	Description
19	Stand with your eyes closed	S	Subject position: Standing position. Test command: Close your eyes and try to stand as firm as possible 4 min: Able to stand safely for 10 s; 3 min: Able to stand under supervision for 10 s; 2 min: Able to stand for 3 s; 1 point: cannot stand for 3 s with eyes closed but can remain stable when standing with eyes open; Score 0: Need help to avoid falling
20	Stand with your feet together	T	Subject position: Standing position. Test command: Keep your feet together and stand as firm as possible 4 points: Able to stand independently with feet together and stand independently for 1 min; 3 points: Able to stand independently with feet together and under supervision for 1 min; 2 points: Able to stand with feet together independently but not stand for 30 s; 1 point: needs help to bring your feet together and be able to stand for 15 s; 0 points: Requires help to bring your feet together and cannot stand for 15 s after putting your feet together
21	Upper limb extension in a standing position	U	Subject position: Standing position. Test command: Raise your arms 90 degrees, straighten your fingers and stretch them forward as far as you can, taking care not to move your feet 4 points: Able to extend more than 25 cm; 3 points: Able to safely extend more than 12 cm; 2 points: Able to extend more than 5 cm; 1 point: Able to reach forward in the case of supervision; 0 points: Loss of balance when trying to reach forward
22	Pick up objects from the ground in a standing position	V	Subject position: Standing position. Test Command: Please pick up the slippers in front of your feet 4 points: Able to pick up slippers safely and easily; 3 points: Able to pick up slippers under supervision; 2 points: cannot be picked up but can reach a position 2-5 cm away from the slippers and maintain balance independently; 1 point: Unable to pick up and requiring supervision when trying to make an effort; 0: Can't try this activity or need help to avoid losing balance or falling
23	Turn around and look back	W	Subject position: Standing position. Test command: Turn to the left and look backwards with your feet still, then turn to the right and look back 4 points: Able to look backwards from both sides and shift the center of gravity well; 3 points: can only look backward from one side, the other side of the center of gravity shifts poorly; 2 points: can only turn to the side but can maintain balance; 1 point: Supervision is required when turning; 0 points: need help and avoid losing balance or falling
24	Turn around for a week	X	Subject position: Standing position. Test command: Please turn once, pause, and then turn again in the other direction 4 min: It takes only 4 s or less to make a safe turn in both directions; 3 min: can only safely make one turn in one direction in 4 s or less; 2 points: Able to make one safe turn, but it takes more than 4 s; 1 point: close monitoring or verbal cues are required when turning; 0 points: Need help when turning

Table 3 (continued)

Serial number	Question (variable name)	Variable name identifier	Description
25	Alternate steps with both feet	Y	Start by placing a step or a small stool in front of the subject about the height of the step. Subject position: Standing position. Test command: Place your left and right feet alternately on the steps/stool until each foot has stepped over the steps or stool 4 times 4 points: Able to stand independently and safely and complete 8 movements in 20 s; 3 points: Able to stand independently, but complete 8 movements for more than 20 s; 2 points: Able to complete 4 movements without help under supervision; 1 point: Able to complete 2 or more movements with minor assistance; Score 0: Need help to avoid a fall or can't try this activity
26	Stand with both feet in front of and back	Z	Subject position: Standing position. Test Command: (Demonstration to subject) Place one foot directly in front of the other and stand as steady as possible. If that doesn't work, place one as far in front of the other as possible so that the front heel is just in front of the back toe 4 points: Able to independently place one foot directly in front of the other foot for 30 s; 3 points: Able to independently place one foot in front of the other foot for 30 s; 2 points: Able to take a small step forward with one foot independently and hold it for 30 s; 1 point: needs help to move forward but can hold for 15 s; 0 points: Loss of balance when walking or standing
27	Stand on one leg	AA	Subject position: Standing position. Test command: Stand on one leg for as long as possible 4 points: Able to lift one leg independently and hold for more than 10 s; 3 points: Able to lift one leg independently and hold for 5–10 s; 2 points: Able to lift one leg independently and hold for 3–5 s; 1 point: Able to lift one leg for less than 3 s but able to maintain standing balance after effort; 0 points: Not able to attempt this activity or need help to avoid falling
28	Overall, how do you feel that your health:	AB	Compare health status one year ago: (1) Very good (2) Very good (3) Good (4) General (5) Poor
29	It is more health than you felt one year ago	AC	How it was a year ago: (1) Much better than 1 year ago (2) It's better than 1 year ago (3) It's similar to one year ago (4) It's worse than 1 year ago (5) It's much worse than 1 year ago
30	Only some of the planned activities can be completed	AD	Do you only do part of what you want to do: (1) Yes (2) No
31	The variety of activities is restricted	AE	Is there a restriction on the type of work or activity you want to do: (1) Yes (2) No
32	It takes more effort to complete the activity	AF	whether more effort is required to complete work or other activities" (1) Yes (2) No
33	Do things less carefully than usual	AG	Do things less carefully than usual: (1) Yes (2) No

Table 3 (continued)

Serial number	Question (variable name)	Variable name identifier	Description
34	Body aches and pains within the past four weeks	AH	Body aches and pains in the past four weeks: (1) No pain at all (2) There is very slight pain (3) There is slight pain (4) Moderate pain (5) Severe pain (6) Very severe pain
35	Pain interferes with work and household chores	AI	Does body aches and pains interfere with work and household chores: (1) No impact at all (2) There is a little impact (3) Moderate impact (4) The impact is great (5) The impact is very large
36	Heavy physical activity	AJ	The ability to perform heavy physical activity (e.g., running, weightlifting, strenuous exercise, etc.):(1) is very limited (2) somewhat limited (3) not limited at all
37	Carry-on daily necessities	AK	The ability to carry daily necessities (e.g., grocery shopping, shopping, etc.):(1) is very restrictive (2) somewhat limited (3) not limited at all
38	Be active in moderation	AL	The ability to perform moderate activities (e.g., moving tables, sweeping floors, doing tai chi, etc.):(1) is very restrictive (2) somewhat restrictive (3) not restrictive at all
39	Go up a few flights of stairs	AM	The ability to go up several flights of stairs: (1) very restrictive (2) somewhat limited (3) not restricted
40	Go up one staircase	AN	The ability to go up one staircase: (1) Very restrictive (2) Somewhat limited (3) No restrictive at all
41	Bend over, bend your knees, squat	AO	The ability to bend over, bend knees, and squat: (1) very limited (2) somewhat limited (3) not limited at all
42	100 m on foot	AP	Ability to walk 100 m: (1) Very restrictive (2) Somewhat restrictive (3) No restriction
43	800 m on foot	AQ	Ability to walk 800 m: (1) Very limited (2) Somewhat limited (3) No limit at all
44	Walk more than 1500 m	AR	Ability to walk more than 1500 m: (1) Very limited (2) Somewhat limited (3) No limit at all
45	Feeling overwhelmed at work or in everyday life	AS	You feel exhausted: (1) all the time (2) most of the time (3) more time (4) part of the time (5) a small part of the time (6) not feeling this way
46	Restrictions on work or activities	AT	Whether work or activities are restricted because you feel depressed or anxious (1) Reduced time spent at work or other activities: a. Yes b. No (2) What you want to do can only be partially done: a. Yes b. No (3) The type of work or activity you want to do is restricted:a. Yes b. No (4) Increased difficulty in completing work or other activities (e.g., requiring extra effort):a. Yes b. No
47	Everyday life impacts	AU	The extent to which physical pain affects daily life: (1) No impact at all (2) There is a little impact (3) Moderate impact (4) The impact is great (5) The impact is very large

Table 3 (continued)

Serial number	Question (variable name)	Variable name identifier	Description
48	Social activities are blocked	AV	Whether social activities are hindered by physical or emotional problems: (1) All the time (2) Most of the time (3) More time (4) Part of the time (5) A small part of the time (6) There is no such feeling
49	Restricted times for social activities	AW	To what extent has your health or emotional distress affected your normal social interactions with family, friends, neighbours or groups in the past 4 weeks? (1) No impact at all (2) There is a little impact (3) Moderate impact (4) The impact is great (5) The impact is very large
50	Physical pain sensations	AX	Physical pain sensation (0–10 points)
51	Frequency of joint pain	AY	Joint pain: When I walk on flat ground (0–10 min) When going up and down stairs (0–10 min) When you sleep in bed at night (0–10 min) When sitting or lying down (0–10 points) Standing (0–10 points)
52	Joint pain when walking	AZ	Degree of joint pain when walking: Walking on flat ground (0–10 points)
53	Joint pain when going up and down stairs	BA	Degree of arthralgia when going up and down stairs: up stairs (0–10 points); Down stairs (0–10 points)
54	When sleeping in bed at night	BB	How stiff your joints are when you wake up in the morning (0–10 points)
55	Joint pain when sitting or lying down	BC	Degree of joint pain when sitting or lying down: sitting (0–10 points), when going to bed and lying down (0–10 points)
56	Joint pain when standing upright	BD	Severity of arthralgia when standing: Standing (0–10 points)
57	Stiff joints when you wake up in the morning	BE	How stiff your joints are when you wake up in the morning: (0–10 points)
58	Stiff joints in the evening	BF	Joint stiffness during the day: How stiff your joints are during the day, after you sit, lie down or rest (0–10 points)
59	How difficult it is to go downstairs	BG	Difficulty when going downstairs: Descending stairs (0–10 points)
60	How difficult it is when going upstairs	BH	Difficulty when going upstairs: Climbing stairs (0–10 points)
61	How difficult it is to get up from a chair	BI	Difficulty getting up from a chair (0–10 points)
62	How difficult it is to stand	BJ	Difficulty in standing: (0–10 points)
63	How difficult it is when bending over	BK	Difficulty when bending over: (0–10 points)
64	How difficult it is to walk on a flat road	BL	Difficulty when walking on a flat road: (0–10 points)
65	How difficult it is to get on and off the bus	BM	Difficulty in getting on and off the bus: (0–10 points)
66	How difficult it is when shopping around	BN	Difficulty when shopping: (0–10 points)
67	How difficult it is when wearing socks	BO	Difficulty in wearing socks: (0–10 points)
68	How difficult it is to take off your socks	BP	Difficulty in taking off socks: (0–10 points)
69	How difficult it is to get out of bed	BQ	Difficulty getting out of bed: (0–10 points)
70	How difficult it is to lie down in bed	BR	Difficulty in getting into bed: (0–10 points)
71	How difficult it is to get in and out of the bathroom to take a shower	BS	Difficulty in getting in and out of the bathroom to take a shower: (0–10 points)
72	How difficult it is when sitting	BT	Difficulty when sitting: (0–10 points)
73	How difficult it is to go to the toilet	BU	Difficulty in going to the toilet: (0–10 points)
74	How difficult it is to do heavy chores	BV	Difficulty in doing heavy chores: (0–10 points)

Table 3 (continued)

Serial number	Question (variable name)	Variable name identifier	Description
75	How difficult it is to do simple household chores	BW	Difficulty in doing simple chores: (0–10 points)
76	Anxiety in the mood	BX	Anxious mood (worried, worried, feeling that something worst is about to happen, irritating easily):(0–4 points)
77	Nervous	BY	Nervousness (nervousness, fatigue, inability to relax, emotional reactions, crying, shaking, feeling uneasy):(0–4 points)
78	Afraid	BZ	Fear (fear of the dark, strangers, being alone, animals, riding in a car or traveling, and crowded situations):(0–4 points)
79	Insomnia	CA	Insomnia (difficulty falling asleep, waking up easily, not sleeping deeply, dreaming, night terrors, feeling tired after waking up):(0–4 points)
80	Cognitive function	CB	Cognitive function (memory, attention deficits, difficulty concentrating, poor memory):(0–4 points)
81	Depressive mood	CC	Depressive mood (loss of interest, lack of pleasure in past hobbies, depression, early awakening, light day and night):(0–4 points)
82	Somatic anxiety of the muscular system	CD	Somatic anxiety muscular system (muscle soreness, inflexibility, muscle tics, limb tics, teeth chattering, voice trembling):(0–4 points)
83	Somatic anxiety sensory system	CE	Somatic anxiety sensory system (blurred vision, chills and fever, weakness, tingling sensation):(0–4 points)
84	Cardiovascular symptoms	CF	Cardiovascular symptoms (tachycardia, palpitations, chest pain, pulsation, fainting, cardiac leakage):(0–4 points)
85	Respiratory symptoms	CG	Respiratory symptoms (chest tightness, suffocation, sighing, dyspnea):(0–4 points)
86	Gastrointestinal symptoms	CH	Gastrointestinal symptoms (dysphagia, belching, dyspepsia, intestinal movement, bowel sounds, diarrhea, weight loss, constipation):(0–4 points)
87	Genitourinary symptoms	CI	Genitourinary symptoms (urinary frequency, urgency, menopause, dreams, impotence, premature ejaculation):(0–4 points)
88	Autonomic nervous system symptoms	CJ	Autonomic symptoms (dry mouth, flushing, pallor, sweating easily, goosebumps, tension headache, hairs stand up):(0–4 points)
89	Behave during the meeting	CK	(1) General manifestations: nervousness, inability to relax, nervousness, finger biting, clenching fists, touching handkerchiefs, facial muscle twitching, non-stopping, hand trembling, frowning, stiff expression, high muscle tone, sighing breathing, paleness. (2) Physiological manifestations: swallowing, hiccups, fast heart rate at rest, rapid breathing (more than 20 beats/min), tendon hyperreflexia, tremor, dilated pupils, eyelid beating: (0–4 points)
90	It's hard to quiet yourself	CL	I find it hard to quiet myself:(0–4 points)
91	Thirsty	CM	I feel dry mouth: (0–4 points)
92	Frequency of physical activity	CN	Number of physical activities (e.g., running, playing, gymnastics, walking, etc.) in a week:(0–4 points)
93	It doesn't feel pleasant	CO	I don't seem to feel any pleasure or comfort at all: (0–4 points)
94	Dyspnea	CP	I feel breathless (e.g. wheezing or breathless):(0–4 points)
95	It's hard to start working on your own initiative	CQ	It's hard to get to work: (0–4 points)
96	Overreact	CR	Frequency of overreaction to things: (0–4 points)
97	I felt trembling	CS	Feeling trembling (e.g., shaking hands):(0–4 points)
98	A lot of energy is expended	CT	I feel like I'm expending a lot of energy; (0–4 points)
99	Worry about situations where you might panic or make a fool of yourself	CU	I'm worried about situations where I might panic or make a fool of myself: (0–4 points)

Table 3 (continued)

Serial number	Question (variable name)	Variable name identifier	Description
100	There is nothing to look forward to in the near future	CV	I don't think I have anything to look forward to in the near future:(0–4 points)
101	Feeling uneasy	CW	I'm nervous: (0–4 points)
102	It's hard to relax yourself	CX	I find it hard to relax:(0–4 points)
103	Feeling sad and depressed	CY	I feel depressed and depressed: (0–4 points)
104	Nothing that gets in the way of work is tolerated	CZ	I can't tolerate anything that prevents me from continuing to work:(0–4 points)
105	I felt like I was about to collapse	DA	I feel like I'm about to break down: (0–4 points)
106	You can't be passionate about anything	DB	I can't be enthusiastic about anything:(0–4 points)
107	I feel that I have no value as a human being	DC	I don't think I'm very worthy: (0–4 points)
108	It's easy to get irritated	DD	I find myself easily irritated: (0–4 points)
109	Even when there is no significant physical activity, the heart rhythm is not normal	DE	Feeling an irregular heart rhythm even when there is no significant physical activity: (0–4 points)
110	Scared for no reason	DF	Feeling scared for no reason: (0–4 points)
111	Feeling that life is meaningless	DG	Feeling that life is meaningless: (0–4 points)
112	Feel that life is full	DH	How you feel about your life in the past 1 month: (0–4 points)
113	Feel energized for a lot of time	DI	Time spent feeling energetic in the past four weeks: (0–4 points)
114	Feel at peace of mind	DJ	Time spent feeling calm in the last four weeks: (0–4 points)
115	A time when you feel in a bad mood, depressed, or depressed	DK	Time in the past four weeks when you felt bad, depressed or depressed: (0–4 points)
116	Feel happy time	DL	Time spent feeling happy in the past four weeks: (0–4 points)
117	General health	DM	Overall, what is your health: (0–4 points)
118	It seems to be more likely to get sick than others	DN	Feeling as if you are more likely to get sick than others: (0–4 points)
119	Be as healthy as everyone around you	DO	Feeling as healthy as everyone around you: (0–4 points)
120	I think my health is going bad	DP	I think my health is going bad: (0–4 points)
122	I am in very good health	DQ	My health is very good:(0–4 points)

Abbreviations

KOA	Knee Osteoarthritis
DDPG	Deep Deterministic Policy Gradient
DQN	Deep Q-Network
BCQ	Batch-Constrained Q-learning
DALYs	Disability-adjusted life years
DRL	Deep reinforcement learning
BMI	Body mass index
ACR	American College of Rheumatology
EMR	Electronic medical records
OOB	Out-Of-Bag
DPG	Deterministic Policy Gradient
TD	Timing difference
NSAIDs	Non-steroidal anti-inflammatory drugs

Acknowledgements
Thanks to Prof. Jiawei Luo and Prof. Quan Guo for their efforts in technical support. We would like to thank Xianghong Zhang for her contribution to data collection and collation.

Author contributions
Sijia Liu, Chengqi He developed the study protocol. Sijia Liu collected the data which was analysed by Jiawei Luo, Sijia Liu.Sijia Liu drafted and wrote the manuscript with input from Chengqi He.

Funding
This work was supported by the National Natural Science Foundation of China (No. 82102680) and Key R&D project of Sichuan Provincial Department of Science and Technology under grant number 2024YFFK0143.

Availability of data and materials
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
This study is a retrospective study and does not involve participant consent. This study has obtained approval from the Biomedical Ethics Review Committee of West China Hospital, Sichuan University.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 20 December 2024 Accepted: 17 March 2025
Published online: 08 May 2025

References

1. KU I. Liisa Kuhi a Contribution of Biomarker Collagen Type Ii Neoepitope C2c in Urine to the Diagnosis and Prognosis of Knee Osteoarthritis.
2. Li E, Tan J, Xu K, et al. Global Burden and Socioeconomic Impact of Knee Osteoarthritis: A Comprehensive Analysis. Front Med. 2024;11:1323091.
3. Ma Xl Hu, Yc WK, et al. Chinese Clinical Practice Guidelines in Treating Knee Osteoarthritis by Periarticular Knee Osteotomy. Orthop Surg. 2022;14:789–806.
4. Cui A, Li H, Wang D, et al. Global, Regional Prevalence, Incidence and Risk Factors of Knee Osteoarthritis in Population-Based Studies. EClinicalMedicine. 2020; 29:
5. Zeng Q-y, Zang C-h, Li X-f, et al. Associated Risk Factors of Knee Osteoarthritis: A Population Survey in Taiyuan. China Chinese medical journal. 2006;119:1522–7.
6. Dong Y, Yan Y, Zhou J, et al. Evidence on Risk Factors for Knee Osteoarthritis in Middle-Older Aged: A Systematic Review and Meta Analysis. J Orthop Surg Res. 2023;18:634.
7. Moseng T, Vlieland TPV, Battista S, et al. Eular Recommendations for the Non-Pharmacological Core Management of Hip and Knee Osteoarthritis: 2023 Update. Ann Rheum Dis. 2024;83:730–40.
8. Zhou Y, Wang Q, Chen L, et al. Daily Habits, Diseases, Drugs and Knee Osteoarthritis: A Two-Sample Mendelian Randomization Analysis. Front Genet. 2024;15:1418551.
9. Zeng C-Y, Zhang Z-R, Tang Z-M, et al. Benefits and Mechanisms of Exercise Training for Knee Osteoarthritis. Front Physiol. 2021;12: 794062.
10. Hunter DJ, March L, Chew M. Osteoarthritis in 2020 and Beyond: A Lancet Commission. The Lancet. 2020;396:1711–2.
11. Wellsandt E, Golightly Y. Exercise in the Management of Knee and Hip Osteoarthritis. Curr Opin Rheumatol. 2018;30:151–9.
12. Deyle GD, Allen CS, Allison SC, et al. Physical Therapy Versus Glucocorticoid Injection for Osteoarthritis of the Knee. N Engl J Med. 2020;382:1420–9.
13. Obuchowicz R, Strzelecki M, Piórkowski A. Clinical Applications of Artificial Intelligence in Medical Imaging and Image Processing—a Review. MDPI. 2024: 1870
14. Nishida N. Advancements in Artificial Intelligence-Enhanced Imaging Diagnostics for the Management of Liver Disease—Applications and Challenges in Personalized Care. Bioengineering. 2024;11:1243.
15. Petrovic K. Deep Learning in Personalized Medicine: Advancements and Applications. Journal of Advanced Analytics in Healthcare Management. 2023;7:34–50.
16. Yu Z, Wang K, Wan Z, et al. Popular Deep Learning Algorithms for Disease Prediction: A Review. Clust Comput. 2023;26:1231–51.
17. Yeoh PSQ, Lai KW, Goh SL, et al. Emergence of Deep Learning in Knee Osteoarthritis Diagnosis. Comput Intell Neurosci. 2021;2021:4931437.
18. Bayramoglu N, Englund M, Haugen IK, et al. Deep Learning for Predicting Progression of Patellofemoral Osteoarthritis Based on Lateral Knee Radiographs, Demographic Data, and Symptomatic Assessments. Methods of Information in Medicine. 2024;
19. Leung K, Zhang B, Tan J, et al. Prediction of Total Knee Replacement and Diagnosis of Osteoarthritis by Using Deep Learning on Knee Radiographs: Data from the Osteoarthritis Initiative. Radiology. 2020;296:584–93.
20. Willemink MJ, Koszek WA, Hardell C, et al. Preparing Medical Imaging Data for Machine Learning. Radiology. 2020;295:4–15.
21. Castiglioni I, Rundo L, Codari M, et al. Ai Applications to Medical Images: From Machine Learning to Deep Learning. Physica Med. 2021;83:9–24.
22. ul Abideen Z, Khan TA, Ali RH, et al. Docontap: Ai-Based Disease Diagnostic System and Recommendation System; proceedings of the 2022 17th International Conference on Emerging Technologies (ICET), 2022. IEEE,
23. Kute SS, Shreyas Madhav A, Kumari S, et al. Machine Learning–Based Disease Diagnosis and Prediction for E-Healthcare System. Advanced analytics and deep learning models. 2022; 127–147.
24. Li HHT, Chan LC, Chan P-K, et al. An Interpretable Knee Replacement Risk Assessment System for Osteoarthritis Patients. Osteoarthritis and Cartilage Open. 2024;6: 100440.
25. Mustaqeem A, Anwar SM, Khan AR, et al. A Statistical Analysis Based Recommender Model for Heart Disease Patients. Int J Med Informatics. 2017;108:134–45.
26. François-Lavet V, Henderson P, Islam R, et al. An Introduction to Deep Reinforcement Learning. Foundations and Trends® in Machine Learning. 2018; 11:219–354.
27. Mousavi SS, Schukat M, Howley E. Deep Reinforcement Learning: An Overview; proceedings of the Proceedings of SAI Intelligent Systems Conference (IntelliSys) 2016: Volume 2, 2018. Springer,
28. Schiphof D, Boers M, Bierma-Zeinstra SM. Differences in Descriptions of Kellgren and Lawrence Grades of Knee Osteoarthritis. Ann Rheum Dis. 2008;67:1034–6.
29. Anukrishna P, Paul V. A Review on Feature Selection for High Dimensional Data; proceedings of the 2017 International Conference on Inventive Systems and Control (ICISC), 2017. IEEE,
30. Parmar A, Katariya R, Patel V. A Review on Random Forest: An Ensemble Classifier; proceedings of the International conference on intelligent data communication technologies and internet of things (ICICI) 2018, 2019. Springer,
31. Liu K, Bellet A. Escaping the Curse of Dimensionality in Similarity Learning: Efficient Frank-Wolfe Algorithm and Generalization Bounds. Neurocomputing. 2019;333:185–99.
32. Salles T, Gonçalves M, Rodrigues V, et al. Broof: Exploiting out-of-Bag Errors, Boosting and Random Forests for Effective Automated Classification; proceedings of the Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval, 2015.
33. Bhadra T, Bandyopadhyay S. Supervised Feature Selection Using Integration of Densest Subgraph Finding with Floating Forward-Backward Search. Inf Sci. 2021;566:1–18.
34. Moradi P, Gholampour M. A Hybrid Particle Swarm Optimization for Feature Subset Selection by Integrating a Novel Local Search Strategy. Appl Soft Comput. 2016;43:117–30.
35. Hu Q, Yue W. Markov Decision Processes with Their Applications. Springer Science & Business Media, 2007
36. Zanini E. Markov Decision Processes. Citeseer. 2014
37. Eschmann J. Reward Function Design in Reinforcement Learning. Reinforcement Learning Algorithms: Analysis and Applications. 2021; 25–33.
38. Fujimoto S, Meger D, Precup D. Off-Policy Deep Reinforcement Learning without Exploration; proceedings of the International conference on machine learning, 2019. PMLR,
39. Tan H. Reinforcement Learning with Deep Deterministic Policy Gradient; proceedings of the 2021 International Conference on Artificial Intelligence, Big Data and Algorithms (CAIBDA), 2021. IEEE,
40. Mnih V, Kavukcuoglu K, Silver D, et al. Human-Level Control through Deep Reinforcement Learning. nature. 2015; 518:529–533.
41. Lillicrap T. Continuous Control with Deep Reinforcement Learning. arXiv preprint arXiv:150902971. 2015;
42. Byeon H. Advances in Value-Based, Policy-Based, and Deep Learning-Based Reinforcement Learning. International Journal of Advanced Computer Science and Applications. 2023; 14:
43. Alabdullah MH, Abido MA. Microgrid Energy Management Using Deep Q-Network Reinforcement Learning. Alex Eng J. 2022;61:9069–78.
44. Li SE. Deep Reinforcement Learning. Reinforcement Learning for Sequential Decision and Optimal Control. Springer. 2023: 365–402.
45. Sarkar T, Kalita S. A Weighted Critic Update Approach to Multi Agent Twin Delayed Deep Deterministic Algorithm; proceedings of the 2021 IEEE 18th India Council International Conference (INDICON), 2021. IEEE,
46. Nobakht H, Liu Y. Action Space Noise Optimization as Exploration in Deterministic Policy Gradient for Locomotion Tasks. Appl Intell. 2022;52:14218–32.
47. Wu M, Gao Y, Jung A, et al. The Actor-Dueling-Critic Method for Reinforcement Learning. Sensors. 2019;19:1547.
48. Zhang S, Sutton RS. A Deeper Look at Experience Replay. arXiv preprint arXiv:171201275. 2017;
49. The WE, Value S. Handbook of game theory with economic applications. 2002;3:2025–54.

Publisher's Note