# Universal decision-making system for life cycle maintenance of bridge based on deep reinforcement learning

Kehong Chen, Chengzhang Chai, Haijiang Li*
BIM for Smart Engineering, School of Engineering, Cardiff University, UK
ChenK43@cardiff.ac.uk

**Abstract.** Bridge maintenance is a complex system involving multiple stakeholders and multi-scale constraints. The traditionally used multi-criteria decision-making (MCDM) and heuristic algorithms have shortcomings such as subjectivity, weak scalability and long calculation time. Reinforcement learning is a data-driven method capable of learning decision-making strategies from large amounts of data. The purpose of this paper is to explore a universal decision-making framework to maximize the health performance of the bridge throughout its life cycle and optimize the cumulative maintenance economic cost. The U.S. National Bridge Inventory (NBI) data set was used to construct a bridge deterioration model. A decision-making model based on Deep Q-learning was trained, and its decision-making performance gradually increased with the number of training iterations. This article also simulated the trained model 100 times, and the results showed that 98% of the cumulative rewards were higher than the expected value.

## 1. Introduction

Bridges are significant elements of the modern transport system; however, with increasing traffic, the health of many bridges is deteriorating dramatically[1]. Reports indicate that a certain number of bridges are in poor health conditions in countries such as the United Kingdom, the United States, China, and Canada; moreover, the cost of bridge maintenance and renovation is expensive[2]. In the bridge maintenance process, in addition to the structural data of the bridge itself, the operating data and environmental data are usually taken into consideration. Recent research considers structural data, such as stress and strain, operating data like traffic volume and sufficiency rating, as well as environmental data, including temperature and humidity, for decision-making in bridge maintenance[3-5].

Bridge maintenance is a complex system. Traditionally, decisions based on the experience of engineers are often subjective and hysteretic[1]. Decision-making models provide a more sensible and objective way to help experts make decisions. Multi-criteria decision-making (MCDM), as a typical decision model, quantifies the priority of alternatives by evaluating the relationship between influencing factors, decision objectives and alternatives[6]. However, these kinds of methods are still based on expert knowledge, which is highly subjective. The heuristic algorithms obtain optimal maintenance policies by searching for the maximum or minimum value of the quantized equation[5], which requires the establishment of specific mathematical equations for a particular optimization problem, and the resulting decision model is challenging to apply to other vectors because it is not universally applicable.

The data-driven approach is effective in diminishing the impact of subjectivity on decision-making using statistical and machine learning techniques to identify hard-to-notice relationships and patterns from large amounts of data[7]. Supervised learning models have been developed to predict the deterioration of bridges[8]. However, the generality of supervised learning, where patterns and relationships are obtained through labelled samples, is limited by data labelling compared to methods that learn from unlabeled data. Unsupervised learning can learn relationships that are generalizable from a large amount of data, while its model training process has limited human involvement, which can largely avoid subjectivity. Nevertheless, there needs to be more cases of unsupervised learning algorithms being used in the field of bridge maintenance decision-making to validate the feasibility of using this approach for bridge maintenance decision-making systems. The agent in reinforcement learning can obtain rewards by interacting with the environment and accumulate experience in the process of accumulating rewards to choose a more wisdom action, which not only effectively avoids subjectivity in decision-making but also accumulates rewards in the process of iteration, which positive significance for the whole lifecycle bridge maintenance.

This study investigates the development of a deep reinforcement learning based universal framework for maintenance decision-making recommendations over the entire life cycle of bridges. The model offers decision recommendations on a 30-year cycle, which improves the health performance of bridges over the entire life cycle and reduces the cumulative maintenance economic cost. The proposed decision-making framework uses the U.S. National Bridge Inventory dataset, and the model agent learns decision-making strategies from historical health data and maintenance data of 1,426 bridges of different types to improve the generalizability of the model for different bridge types.

## 2. Background

Since 2018, the potential for the application of reinforcement learning in bridge maintenance has been progressively discovered, and reinforcement learning-based decision-making frameworks for bridge maintenance have begun to be progressively applied. Wei et al. proposed a deep reinforcement learning framework using deep neural networks to learn state-action Q-values and conducted a case study with a simple bridge deck with 7 components and a large-span cable-stayed bridge with 263 components, verifying that deep reinforcement learning can efficiently find optimal strategies for maintenance tasks of both simple and complex structures[9]. Lei et al. used a regional probabilistic model to simulate the stochastic process of bridge deterioration to maximize the life cycle maintenance cost-effectiveness of maintenance actions and maintain the health of regional bridges by developing a 100-year maintenance strategy for highway bridges that meets different budgetary constraints[10]. Ao and Alireza developed a parametric DQN model that synergistically integrates adaptive sequential decision support, life-cycle cost analysis, and probabilistic risk assessment for long-term bridge asset management systems[11]. Zhou et al. proposed a multi-intelligence reinforcement learning framework, which defines each bridge component as an intelligence interacting with the bridge environment to improve the ability to deal with complex maintenance decision-making problems[12]. Zachary and James-A proposed a hierarchical reinforcement learning framework, which naturally adapts to the structure of information and decision-making layers and improves the scalability of reinforcement-based learning decision-making framework[13].

According to the above review, the potential of the Reinforcement Learning framework to be applied in the decision-making domain of bridge maintenance has been effectively validated. However, the established decision models usually focus on maximizing the efficiency of whole-life maintenance or on improving the flexibility and extensibility of the decision models so that they can be adapted to different types of bridges or decision scenarios with different information structures. There is still a vacuum in the research of full-life-cycle bridge maintenance decision models with high generality.

## 3. Methodology

This study aims to develop a deep reinforcement learning based framework for maintenance decision-making recommendations over the entire life cycle of bridges. The primary focus is to establish a degradation model using probability distributions integrated into a Markov decision process framework, subsequently constructing the decision-making system of bridge maintenance based on Deep Q-learning. The framework of the Decision-Making System is shown in Figure 1.

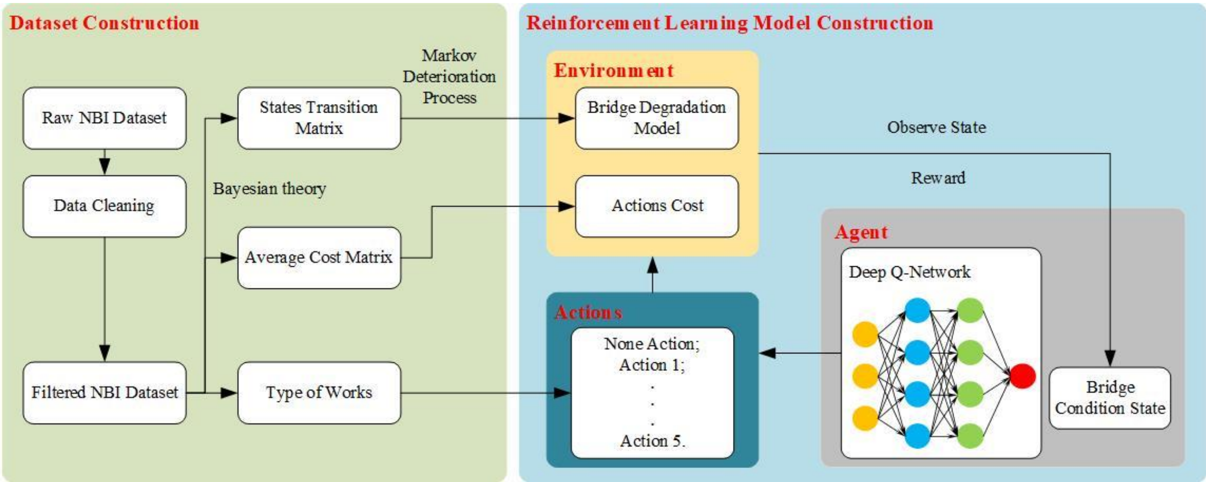


Figure 1: Framework of universal decision-making system for life cycle maintenance of bridge based on deep reinforcement learning

In the data cleaning stage, the data that was irrelevant to building a reinforcement learning model and the data with fewer observations were removed. The processed data was used to calculate the state transition matrix of bridge health and the average cost of each maintenance worker type based on Bayesian theory. Based on this, a bridge deterioration model was built, and an interactive environment for the reinforcement learning model was established. The actions that the agent of the reinforcement learning model can take are defined according to the type of maintenance work after data cleaning, and Deep Q-network is introduced to learn the decision-making strategy.

### 3.1 Markov Decision Process

Markov decision process (MDP) is a widely used theoretical basis for developing degradation models. The MDP provides a mathematical framework for quantifying sequential decision-making problems in stochastic environments [14]. The problem of developing maintenance and repair policies for the bridge can be modelled as a finite-state, discrete-time MDP[15]. Each state in the MDP is associated with a set of possible actions, and the transitions between states are probabilistic, influenced by the chosen actions and modelled using a probability distribution. The transition probabilities, $P(s_t, a_t, s_{t+1})$, represent the likelihood of moving from state $s_t$ to state $s_{t+1}$ under action $a_t$. These probabilities are derived from historical data and expert assessments concerning the impact of different maintenance actions on the state of the bridge.

In this study, Bayesian theory is used to calculate the state transfer probability distributions of bridges in the three initial states of Good, Fair and Poor. An MDP-compliant probabilistic model of bridge deterioration is built based on the matrix of state distributions of the bridge's health condition, which is the basis of the model presented in Chapter 4.

### 3.2 Deep Q-Learning

The basis of the reinforcement learning model learning to maximize the cumulative reward strategy is the interaction between the agent and the environment; the agent selects the action $a_t$ based on the current state $s_t$ of the environment, the reward $r_{t+1}$ and the state $s_{t+1}$ after taking the action $a_t$ will be fed back to the agent by the environment. Deep Q-Learning aims to learn the optimal policy by using a deep neural network to approximate the Q-function[13]. The Q-function $Q(s_t, a_t)$ is often updated using the temporal difference (TD) method during the iterative process [16]. The Bellman equation used in neural network training is as follows:

$$Qnew(s_t, a_t) \leftarrow (1-\alpha) \cdot Q(s_t, a_t) + \alpha \cdot (r_{t+1} + \gamma \cdot \max_a(s_{t+1}, a_t))$$

Where $\alpha$ is the learning rate, $\gamma$ is the discount factor, $r_{t+1}$ is the reward received after transitioning from state $s_t$ to $s_{t+1}$, and $\max_a(s_{t+1}, a_t)$ represents the maximum reward that can be obtained from the next state.

### 3.3 Simulation and Prediction

The trained Q function is used to simulate the state transition of the bridge over 30 years, and the maintenance actions selected for each year are listed. The simulation iterates through each time step, starting from the current state of the bridge. At each step, the model calculates the optimal maintenance action by selecting the action $a^*$ with the maximum Q-value for the current state $s_t$:

$$a^* = arg\ max(s_t, a_t)$$

These actions are recommended to the bridge maintenance team. The decision-making system considers the immediate and long-term implications of maintenance actions to ensure optimal resources allocation and prolongation of the bridge's lifespan.

In practice, if the agent always chooses the action with the highest value at every step, it could make exploring unused sequences difficult. To avoid being stuck in the so-called best choice and achieve a good balance between exploration and exploitation, the exploration-exploitation strategy, such as the epsilon-greedy strategy, is widely used[13].

## 4. Development

### 4.1 Data Preparation

The U.S. National Bridge Inventory (NBI), which is maintained by the FHWA Office of Bridge and Structures, provides the data support for this study. The inventory contains essential information, construction condition ratings and maintenance records for more than 620,000 highway bridges located in each state in the U.S. from 1992 to 2023. The evaluation specifications and data coding in the database are based on National Bridge Inspection Standards and the Recoding and Coding Guide for the Structure Inventory and Appraisal of the Nations Bridges[17].

In the NBI database, the condition rating of the bridge’s components was categorized into ten classes from 0 (Failed condition) to 9 (Excellent condition). The conditions of the bridges were classified according to the lowest condition rating among the components: Condition ratings from 7 to 9 are defined as good condition, condition ratings of 5 and 6 are fair condition, and condition ratings of less than 4 are poor condition. Additionally, the content of the type of work recorded in the inventory is shown in Table 1. The rating criteria mentioned above and Work

Type definitions provide the foundation for constructing a reinforcement learning environment and defining the actions of the agent in the following pages.

Table 1: Description of type of works

| Code | Description |
|---|---|
| 31 | Replacement of bridge or other structure because of substandard load carrying capacity or substandard bridge roadway geometry. |
| 32 | Replacement of bridge or other structure because of relocation of road. |
| 33 | Widening of existing bridge or other major structure without deck rehabilitation or replacement; includes culvert lengthening. |
| 34 | Widening of existing bridge with deck rehabilitation or replacement. |
| 35 | Bridge rehabilitation because of general structure deterioration or inadequate strength. |
| 36 | Bridge deck rehabilitation with only incidental widening. |
| 37 | Bridge deck replacement with only incidental widening. |
| 38 | Other structural work, including hydraulic replacements. |

### 4.2 State transition matrix construction

According to Bayesian theory, the data of this study is preprocessed. The frequency of bridge health condition transfer for the same bridge number under different work types from 2016 to 2021 was approximated as the probability of bridge health condition transfer. The data with low statistical significance, such as the work types 32, 36, and 37, were removed. The situation in which no work type was taken was defined as doing nothing, which represents the natural course of bridge deterioration in the absence of any maintenance measures. The calculated state transfer matrix for the bridge condition is shown in Table 2 to Table 7.

Table 2: Doing Nothing State transition probability

| State | Good | Fair | Poor |
|---|---|---|---|
| **Good** | 0.99 | 0.01 | 0 |
| **Fair** | 0 | 0.98 | 0.02 |
| **Poor** | 0 | 0 | 1 |

Table 3: Work 31 State transition probability

| State | Good | Fair | Poor |
|---|---|---|---|
| **Good** | 0.87 | 0.05 | 0.08 |
| **Fair** | 0.02 | 0.97 | 0.01 |
| **Poor** | 0.21 | 0.1 | 0.68 |

Table 4: Work 33 State transition probability

| State | Good | Fair | Poor |
|---|---|---|---|
| **Good** | 0.99 | 0.01 | 0 |
| **Fair** | 0 | 0.98 | 0.02 |
| **Poor** | 0 | 0 | 1 |

Table 5: Work 34 State transition probability

| State | Good | Fair | Poor |
|---|---|---|---|
| **Good** | 0.87 | 0.13 | 0 |
| **Fair** | 0.03 | 0.97 | 0 |
| **Poor** | 0 | 0 | 1 |

Table 6: Work 35 State transition probability

| State | Good | Fair | Poor |
|---|---|---|---|
| **Good** | 0.95 | 0.05 | 0 |
| **Fair** | 0.02 | 0.97 | 0.01 |
| **Poor** | 0.06 | 0.2 | 0.74 |

Table 7: Work 38 State transition probability

| State | Good | Fair | Poor |
|---|---|---|---|
| **Good** | 0.9 | 0.08 | 0.02 |
| **Fair** | 0.02 | 0.97 | 0.01 |
| **Poor** | 0 | 0.17 | 0.83 |

The NBI dataset also contains the cost incurred for each repair, and incorporating this type of data into the decision model can effectively improve its comprehensive ability to consider not only the health of the bridge but also the maintenance cost of the bridge over its entire lifecycle. The average cost for different types of work is shown in Table 8.

Table 8: Average cost of Actions

| Action | Cost/ $10^3$ \$ | Type of Work |
|---|---|---|
| 0 | 0 | None-action |
| 1 | 2279.69 | Work 31 |
| 2 | 1018.62 | Work 33 |
| 3 | 766.76 | Work 34 |
| 4 | 1005.71 | Work 35 |
| 5 | 3126.26 | Work 38 |

### 4.3 Reward setting

The quantification of rewards Deep-Q learning model is a critical aspect that directly influences the learning outcomes and efficacy of the decision recommendations. Rewards are assigned based on the maintenance outcomes associated with each action taken within the simulation environment. Specifically, rewards $R_t$ are computed by integrating several factors which are shows below:

$$R_t = R_{ht} + R_{ct}$$

Where $R_{ht}$ is the reward of health state transition, and $R_{ct}$ is the reward of cost efficiency.

Health state transition: Actions that enhance the health state of the bridge receive positive rewards. Conversely, actions resulting in health deterioration are penalized. The specific reward settings are shown in Table 9.

Table 9: Reward of health state transition

| State | Good | Fair | Poor |
|---|---|---|---|
| **Good** | 3 | -1 | -3 |
| **Fair** | 2 | 0 | -2 |
| **Poor** | 5 | 2 | -1 |

Cost efficiency: Maintenance actions receive negative reward based on average cost $C_a$, which is calculated as follows:

$$R_{ct} = \frac{1}{2} C_a \times 10^{-6}$$

## 4.4 Construction of Deep Q network

In this study, a bridge deterioration model is established based on the state transfer matrix of the bridge health condition calculated above, and the environment of the reinforcement learning model is established based on the bridge state transfer probability and the economic cost of maintenance work. The actions are defined as Action 0: None Action, Action 1: Work Type 31, Action 2: Work Type 33, Action 3: Work Type 34, Action 4: Work Type 35 and Action 5: Work Type 38. The states of bridge health condition are defined in three categories: Good, Fair and Poor. The model is trained with 30-time steps as an epoch, and the optimization objective is to obtain the highest total reward in 30 years under the combined consideration of bridge health and maintenance cost.

The Deep Q-Network initiates with an embedding layer translating discrete state inputs into a 10-dimensional vector, followed by three fully connected layers with 64, 128, and 64 neurons, concluding with an output layer mapping to the action space.

The hyperparameters are set as $\gamma$=0.95, lr=0.0015, the capacity of the experience buffer is 1000, and the training batch size is 32. The loss function is the mean square error (MSE) of the predicted and observed values of the total reward. The exploration probability $\varepsilon$ is initially set to 1, and exponentially decays to a minimum value of 0.1 along with the training episodes. Changes in the loss value during the iteration are shown in Figure 2, and the value matrix of actions is shown in Table 10.

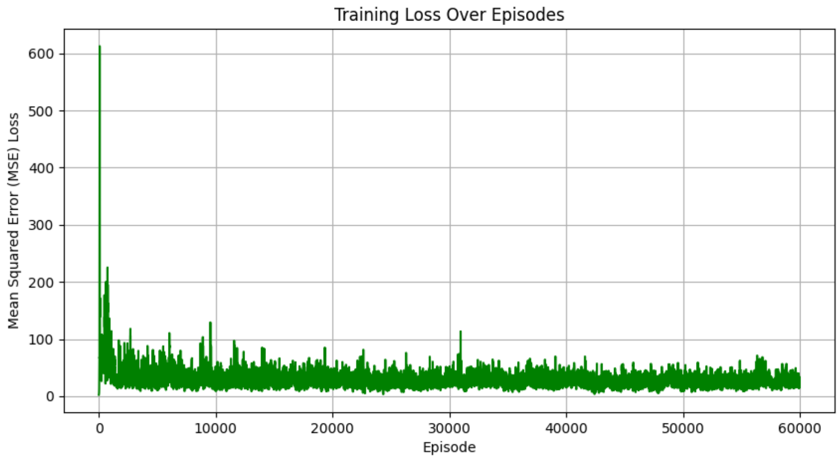


Figure 2: Mean Squared Error (MSE) Loss over episodes

Table 10: The value of the Action in each state

| | **Action 0** | **Action 1** | **Action 2** | **Action 3** | **Action 4** | **Action 5** |
|---|---|---|---|---|---|---|
| **Good** | 3.099 | 1.829 | 2.106 | 6.391 | 2.415 | 1.187 |
| **Fair** | 36.315 | 33.979 | 35.963 | 34.001 | 34.124 | 33.665 |
| **Poor** | 17.731 | 15.201 | 17.021 | 17.356 | 18.443 | 19.479 |

The reward settings of a well-constructed reinforcement learning model are usually based on expert knowledge. However, this study uses reward settings that lack the validation of expert knowledge to explore the feasibility, which could negatively affect the action value matrix.

### 4.5 Result

The blue dots in Figure 3 demonstrate the total reward value of this reinforcement learning model for each iteration during the 2000 epochs. The red line is the linear regression graph of the total reward with respect to the epochs, which represents the trend of the total reward. The green line denotes the expectation of the cumulative reward. According to Figure 3, the total reward has a higher value of the regression function value in comparison to the expectation value. The total reward is increased with the increase in the number of iterations of this model, which indicates that the performance of the agent is increased with the number of iterations.

The trained model was used to perform 100 simulations in the bridge deterioration model to simulate the change in the state of the bridge over 30 years under a random initial state and simulate the maintenance strategy chosen by the reinforcement learning agent. The total rewards of the results of these 100 simulations were compared with the expected cumulative rewards. According to Figure 4, 98 simulations resulted in higher total rewards than expected rewards, with a frequency of 98%.

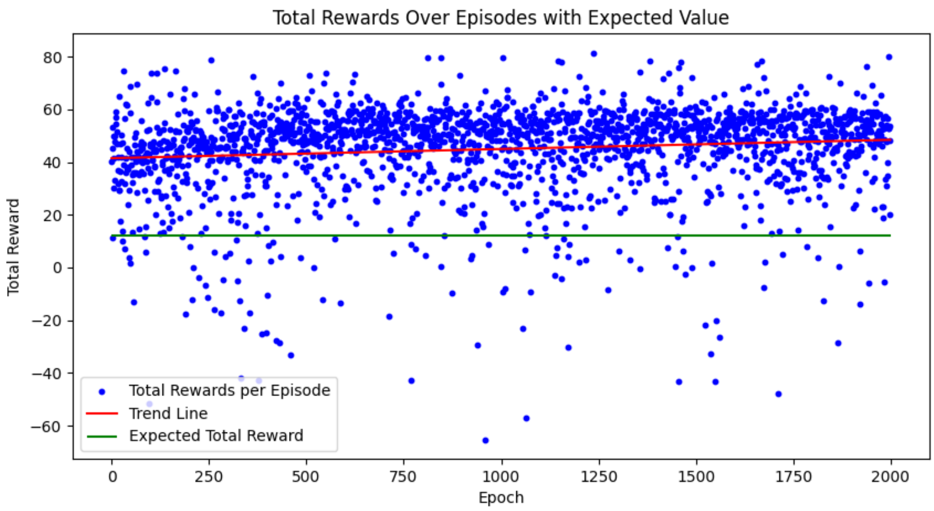


Figure 3: Total rewards trend over episodes with expected value

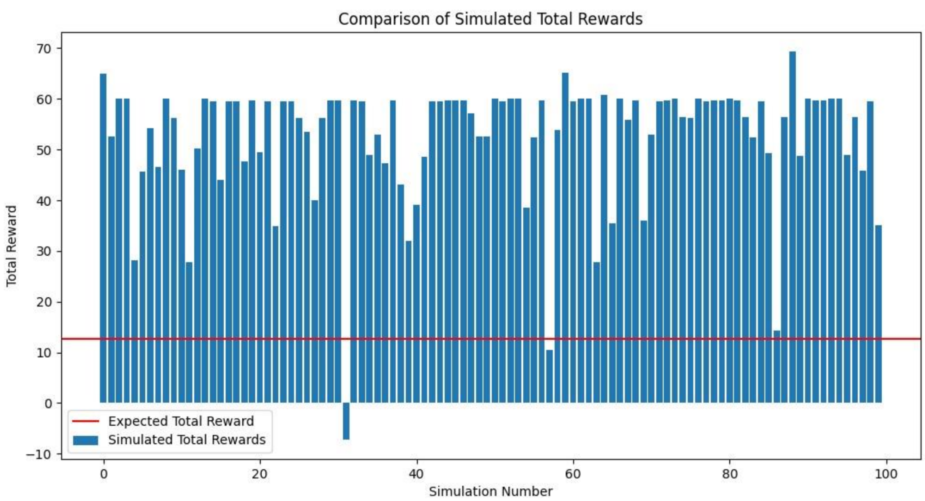


Figure 4: Comparison of simulated total rewards

## 5. Conclusion

In this study, a universal bridge maintenance decision-making system based on reinforcement learning is proposed for predicting the changes in bridge health, simulating the trends of bridge health over 30 years, and giving maintenance recommendations for each year, which is capable of effectively improving the full-life-cycle health performance of bridges and reducing the cumulative maintenance costs. In this study, a deterioration model is constructed based on state transfer matrix calculated from bridge data of different lengths and types from different states in the U.S NBI dataset. The training results show that the model performs well with a 98% frequency of cumulative gains above the expected gains when providing decision recommendations for bridge maintenance over 30 years.

The deterioration model constructed in this study contains limited state space. Additionally, only the economic cost of maintenance is included in the evaluation; the follow-up will be from the construction of more complex state space as well as the inclusion of the cost of maintenance time, the bridge life and other evaluation indexes to improve the decision-making ability of the model in complex environments.

## 6. Acknowledgements

This research is funded by the Cardiff University-China Scholarship Council joint program.

## References

[1] Wu, C., Wu, P., Wang, J., Jiang, R., Chen, M. and Wang, X., 2021. Critical review of data-driven decision-making in bridge operation and maintenance. Structure and infrastructure engineering, 18(1), pp.47-70.

[2] Jiang, Y., Yang, G., Li, H. and Zhang, T., 2023. Knowledge driven approach for smart bridge maintenance using big data mining. Automation in Construction, 146, p.104673.

[3] Gui, C., Zhang, J., Lei, J., Hou, Y., Zhang, Y. and Qian, Z., 2021. A comprehensive evaluation algorithm for project-level bridge maintenance decision-making. Journal of cleaner production, 289, p.125713.

[4] Contreras-Nieto, C., Shan, Y., Lewis, P. and Hartell, J.A., 2019. Bridge maintenance prioritization using analytic hierarchy process and fusion tables. Automation in Construction, 101, pp.99-110.

[5] Shen, Z., Liu, Y., Liu, J., Liu, Z., Han, S. and Lan, S., 2023. A Decision-Making Method for Bridge Network Maintenance Based on Disease Transmission and NSGA-II. Sustainability, 15(6), p.5007.

[6] Tan, T., Mills, G., Papadonikolaki, E. and Liu, Z., 2021. Combining multi-criteria decision making (MCDM) methods with building information modelling (BIM): A review. Automation in Construction, 121, p.103451.

[7] Cheng, M. and Frangopol, D.M., 2021. A decision-making framework for load rating planning of aging bridges using deep reinforcement learning. Journal of Computing in Civil Engineering, 35(6), p.04021024.

[8] Jaafaru, H. and Agbelie, B., 2022. Bridge maintenance planning framework using machine learning, multi-criteria decision analysis and evolutionary optimization models. Automation in Construction, 143, p.104585.

[9] Wei, S., Bao, Y. and Li, H., 2020. Optimal policy for structure maintenance: A deep reinforcement learning framework. Structural Safety, 83, p.101906.

[10] Lei, X., Xia, Y., Deng, L. and Sun, L., 2022. A deep reinforcement learning framework for life-cycle maintenance planning of regional deteriorating bridges using inspection data. Structural and Multidisciplinary Optimization, 65(5), p.149.

[11] Du, A. and Ghavidel, A., 2022. Parameterized deep reinforcement learning-enabled maintenance decision-support and life-cycle risk assessment for highway bridge portfolios. Structural Safety, 97, p.102221.

[12] Zhou, Q.N., Yuan, Y., Yang, D. and Zhang, J., 2022. An advanced multi-agent reinforcement learning framework of bridge maintenance policy formulation. Sustainability, 14(16), p.10050.

[13] Hamida, Z. and Goulet, J.A., 2023. Hierarchical reinforcement learning for transportation infrastructure maintenance planning. Reliability Engineering & System Safety, 235, p.109214.

[14] Lai, L., Dong, Y., Andriotis, C.P., Wang, A. and Lei, X., 2024. Synergetic-informed deep reinforcement learning for sustainable management of transportation networks with large action spaces. Automation in Construction, 160, p.105302.

[15] Durango, P.L. and Madanat, S.M., 2002. Optimal maintenance and repair policies in infrastructure management under uncertain facility deterioration rates: An adaptive control approach. Transportation Research Part A: Policy and Practice, 36(9), pp.763-778.

[16] Watkins, C.J.C.H., 1989. Learning from delayed rewards.

[17] Alogdianakis, F., Dimitriou, L. and Charmpis, D.C., 2022. Data-driven recognition and modelling of deterioration patterns in the US National Bridge Inventory: A genetic algorithm-artificial neural network framework. Advances in Engineering Software, 171, p.103148.