Frontiers in Emerging Multidisciplinary Sciences (Open Access)

Volume 03, Issue 08, August 2026,
Publish Date: 14-08-2026
DOI: https://doi.org/10.64917/fems/Volume03Issue08-05 **PageNo.56-61**

Deep Reinforcement Learning Approach for Optimal Automatic Driving of Urban Rail Trains

Dr. Luka Kranjc
Department of Artificial Intelligence Engineering Slovenian Center for Intelligent Systems Ljubljana, Slovenia

Dr. Maja Zupan
Center Laboratory of Machine Learning and Data Science Institute for Digital Innovation Maribor, Slovenia

ABSTRACT

Automatic train operation requires a control strategy capable of maintaining speed-tracking accuracy, passenger comfort, operational safety, and energy efficiency under continuously changing railway conditions. Conventional control approaches can provide effective regulation, but their performance may be constrained when train dynamics, disturbances, operational constraints, and nonlinear relationships become difficult to represent through fixed control rules. Reinforcement learning provides an alternative paradigm in which an intelligent controller learns operational decisions through interaction with an environment. This study develops a conceptual deep reinforcement learning approach for optimal automatic driving of urban rail trains, with particular emphasis on Deep Q-Network (DQN)-based decision-making. The proposed approach integrates train-state representation, discrete driving-action selection, reward-based optimization, and operational constraints into a unified automatic driving framework. The literature indicates that iterative learning control, self-anti-disturbance control, Q-learning, policy-gradient reinforcement learning, and comfort-evaluation methods provide important foundations for intelligent train control. However, these approaches also reveal a need for a more integrated framework capable of balancing speed tracking, energy consumption, and ride comfort. The proposed methodology therefore formulates automatic driving as a sequential decision-making problem and defines a multi-objective reward mechanism incorporating tracking error, acceleration variation, energy consumption, and operational constraints. The resulting framework provides a theoretical basis for adaptive and optimization-oriented automatic train driving and establishes directions for future experimental validation using real or simulated urban rail operating data.

KEYWORDS: Deep Reinforcement Learning, Deep Q-Network, Automatic Train Operation, Urban Rail Transit, Intelligent Train Control, Speed Tracking, Energy Optimization, Ride Comfort, Reinforcement Learning.

INTRODUCTION

1.1 Background

Urban rail transportation requires precise and reliable train operation because trains operate according to predefined schedules, station-to-station movement constraints, speed limits, and passenger-service requirements. Automatic driving systems must continuously regulate traction, coasting, and braking so that a train reaches the next station safely while maintaining an acceptable travel time, stopping accuracy, energy consumption, and level of passenger comfort. These requirements make automatic train driving a dynamic control problem rather than a simple speed-regulation task.

Traditional automatic train-driving strategies commonly rely on control laws and predefined operating rules. Iterative learning control has been investigated as a mechanism for improving automatic train-driving performance by learning from repeated operations (He, 2019). Similarly, non-parametric iterative learning control has been examined for automatic train-driving control, demonstrating the relevance of learning-based control mechanisms to railway applications (He and Xu, 2020). Self-anti-disturbance control has also been applied to train-speed regulation, particularly where external disturbances and uncertain operating conditions affect tracking performance (Lian et al., 2020; Yang et al., 2021).

The emergence of reinforcement learning provides a different theoretical perspective. Instead of determining a control action entirely from a predetermined mathematical relationship, reinforcement learning allows an agent to select actions according to the expected long-term reward associated with different operational decisions. Research on automatic train driving has already investigated reinforcement-learning-based approaches, including Q-learning and policy-gradient methods (Zhang, 2019; Zhang et al., 2020; Zhang, 2020). These studies provide a foundation for extending intelligent train control toward deep reinforcement learning.

1.2 Problem Statement

Automatic driving involves competing objectives. Aggressive traction may reduce travel time but increase energy consumption and passenger discomfort, whereas conservative control may improve comfort and energy efficiency while increasing travel time. Consequently, an automatic driving controller should not optimize a single variable independently. It should identify a suitable compromise among speed tracking, energy use, acceleration behavior, and operational constraints.

The central problem addressed in this study is therefore the formulation of a deep reinforcement learning framework capable of selecting appropriate automatic-driving actions while considering multiple operational objectives. A DQN-oriented approach is particularly relevant because the automatic-driving process can be formulated using a discrete set of driving actions, such as traction, coasting, and braking.

1.3 Research Relevance and Objectives

The research is relevant because intelligent train control can potentially improve the adaptability of automatic driving while reducing dependence on manually designed control rules. Existing research on reinforcement learning for train operation demonstrates that learning-based optimization is feasible, while studies of comfort and disturbance rejection emphasize the importance of considering operational quality beyond simple speed tracking (Lai et al., 2019; Wu and Zhang, 2017).

The objectives are to formulate a DQN-based automatic-driving framework, define an appropriate state-action-reward representation, integrate speed tracking with energy and comfort considerations, and critically examine how deep reinforcement learning can address limitations identified in existing intelligent train-control approaches.

1.4 Scope and Significance

The study focuses on the conceptual and methodological design of deep reinforcement learning for urban rail automatic driving. It does not claim experimental performance values that are not contained in the supplied references. Its significance lies in integrating the theoretical contributions of existing learning-based, disturbance-rejection, energy-optimization, and comfort-evaluation studies into a coherent deep reinforcement learning architecture.

LITERATURE REVIEW

Research into intelligent railway control has developed through several complementary directions. He (2019) investigated adaptive iterative learning control for automatic train-driving systems, emphasizing the ability of learning mechanisms to improve control behavior across repeated operations. He and Xu (2020) further examined a non-parametric iterative learning control algorithm for automatic train-driving control. These approaches demonstrate the value of using historical operating information to improve subsequent control actions. However, iterative learning is fundamentally associated with repeated task execution, whereas deep reinforcement learning can formulate operation as a sequential decision problem in which future rewards influence present decisions.

Disturbance rejection represents another important research direction. Lian et al. (2020) examined automatic speed control of high-speed trains based on self-anti-disturbance control. Yang et al. (2021) similarly investigated train-speed tracking using a self-anti-disturbance controller combined with an improved particle swarm algorithm. These studies indicate that robust speed tracking requires mechanisms capable of handling disturbances rather than relying solely on nominal operating conditions. For a reinforcement learning controller, this insight implies that disturbance-related state information and robustness-oriented reward design should be considered during policy development.

Reinforcement learning research provides a direct foundation for the proposed approach. Zhang et al. (2020) investigated intelligent train control using policy-gradient reinforcement learning, while Zhang (2020) examined automatic train-driving methods based on reinforcement learning. Zhang, Zhang, and Zhang (2019) specifically addressed energy-saving optimization of high-speed railway trains using the Q-learning algorithm. Collectively, these studies establish that reinforcement learning can be used to formulate train operation as an optimization problem involving sequential control decisions.

The transition from conventional Q-learning toward deep reinforcement learning is theoretically significant because DQN replaces a conventional action-value table with a neural network representation. This allows a larger and more complex state space to be represented without explicitly storing every state-action combination. For urban rail operation, the state may include speed error, current velocity, distance to a target station, acceleration, braking status, and operational constraints. The resulting state space can become considerably larger than that handled conveniently by tabular Q-learning.

Passenger comfort introduces another dimension to the optimization problem. Wu and Zhang (2017) evaluated urban rail train ride comfort using fuzzy reasoning, while Lai et al. (2019) proposed a comprehensive comfort evaluation system for subway passengers based on hierarchical analysis. These studies indicate that train-control quality should be assessed from a multidimensional perspective rather than exclusively through speed accuracy. Therefore, comfort-related variables can be incorporated into the reward function of an intelligent automatic-driving controller.

The literature reveals three major research gaps. First, existing control-oriented studies emphasize tracking and disturbance rejection, whereas reinforcement learning provides greater potential for sequential optimization. Second, Q-learning-based energy optimization and policy-gradient control demonstrate the applicability of reinforcement learning but leave room for deep state representation and more complex operating environments. Third, comfort research highlights passenger-oriented performance measures that should be integrated with speed and energy objectives rather than treated as an independent evaluation stage. The proposed DQN framework addresses these gaps by combining state-based deep decision-making with a multi-objective reward structure.

METHODOLOGY

3.1 Research Framework

The proposed methodology conceptualizes automatic train driving as a Markov decision process consisting of an environment, an intelligent agent, a state representation, an action set, and a reward function. The railway environment represents train dynamics, track conditions, speed restrictions, station locations, and operational constraints. The DQN agent observes the current state and selects a driving action intended to maximize cumulative future reward.

The overall architecture can be expressed as:

Train operating state → State preprocessing → DQN network → Driving-action selection → Train/environment response → Reward calculation → Network learning → Updated driving policy

This structure extends the reinforcement-learning direction investigated in automatic train driving (Zhang, 2020) while providing a deep function-approximation mechanism for representing complex operating states.

3.2 State Representation

The state vector should contain variables that sufficiently describe the operational situation of the train. A conceptual state vector can be expressed as:

[

S_t=[v_t,e_v,d_s,a_t,E_t,c_t,r_t]

]

where (v_t) represents current train speed, (e_v) represents speed-tracking error, (d_s) represents remaining distance to the target station, (a_t) represents acceleration, (E_t) represents cumulative energy consumption, (c_t) represents a comfort-related indicator, and (r_t) represents relevant operational restrictions.

Speed-tracking error is essential because the controller must determine whether the train is above or below its desired operating trajectory. Remaining station distance is particularly important during braking and stopping phases. Acceleration-related variables provide information about ride comfort, consistent with the importance of comprehensive comfort evaluation identified by Wu and Zhang (2017) and Lai et al. (2019).

3.3 Action-Space Design

A DQN requires a discrete action space. Accordingly, the train-control action can be represented using operational categories such as traction increase, moderate traction, coasting, moderate braking, and strong braking. The exact action levels would depend on the characteristics of the target train and traction-control system.

The principal advantage of this representation is that the network does not need to directly generate an unrestricted continuous control signal. Instead, it determines the most valuable operational action from a predefined set. This is conceptually consistent with Q-learning-based railway optimization investigated by Zhang, Zhang, and Zhang (2019).

3.4 Multi-Objective Reward Function

A major component of the proposed methodology is the reward function. Optimizing speed error alone may produce undesirable driving behavior, while optimizing energy alone may lead to excessive coasting or conservative operation. A composite reward is therefore proposed:

$$R_t=-\left(w_1E_v+w_2E_e+w_3E_c+w_4E_s+w_5E_o\right)$$

where E_v represents speed-tracking error, E_e represents energy consumption, E_c represents comfort-related penalties, E_s represents stopping or station-arrival error, and E_o represents operational-constraint violations. The coefficients $w_1,\ldots,w_5$ determine the relative importance of the objectives.

This formulation provides a direct mechanism for incorporating the energy-saving perspective explored by Zhang et al. (2019) and the comfort dimensions investigated by Wu and Zhang (2017) and Lai et al. (2019). It also allows disturbance-related performance to influence the learning objective, reflecting the concerns addressed by Lian et al. (2020) and Yang et al. (2021).

3.5 DQN Learning Mechanism

The DQN estimates the expected cumulative reward for each available action. For a state s, the network generates:

$$Q(s,a;\theta)$$

where θ represents the neural-network parameters. The selected action is determined by the action with the highest estimated value, subject to an exploration strategy during training.

The learning target can be represented as:

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

where r_t is the immediate reward, γ is the discount factor, and θ^- represents the parameters of a target network. The network is trained by minimizing the difference between predicted and target action values.

A replay mechanism can be used to store previous transition experiences and sample them during training. This reduces excessive dependence on consecutive observations and improves the diversity of training experiences. For railway applications, training scenarios should include normal operation as well as variations in speed profiles, disturbances, station distances, and braking requirements.

3.6 Operational Constraint Handling

Safety and operational restrictions cannot be treated merely as optional optimization objectives. A practical controller must prevent actions that violate mandatory speed or braking constraints. Therefore, constraint handling should operate at two levels: the learning reward should penalize undesirable decisions, while an external supervisory mechanism should prevent physically or operationally unsafe actions.

This distinction is important because reinforcement learning optimizes expected reward, whereas railway operation includes hard constraints that cannot be violated merely because a particular action has a high learned value. The disturbance-rejection studies of Lian et al. (2020) and Yang et al. (2021) reinforce the need for robust control behavior under non-ideal conditions.

3.7 Training and Evaluation Strategy

Training should use multiple simulated operational episodes representing different initial speeds, target station distances, speed restrictions, and disturbance conditions. The training objective should be convergence toward a stable policy rather than optimization of one isolated operating trajectory.

Evaluation should consider multiple indicators: speed-tracking accuracy, station stopping performance, energy consumption, acceleration variation, comfort, constraint violations, and computational stability. The proposed evaluation structure is consistent with the literature's emphasis on speed control, energy optimization, and passenger comfort (Lai et al., 2019; Zhang et al., 2019; Wu and Zhang, 2017).

RESULTS

The synthesis of the supplied literature and the proposed methodology indicates that deep reinforcement learning provides a coherent framework for integrating several previously separated objectives in automatic train driving. The first major finding is that train driving can be naturally formulated as a sequential decision problem. Existing reinforcement-learning studies demonstrate the applicability of Q-learning and policy-gradient mechanisms to intelligent train control (Zhang et al., 2020; Zhang, 2020). The proposed DQN formulation extends this direction by allowing the value function to be approximated through a neural network rather than a fixed state-action table. This is particularly relevant when train state descriptions contain multiple continuous operational variables.

A second finding concerns multi-objective optimization. Energy-efficient train operation has already been addressed through Q-learning (Zhang et al., 2019), while comfort has been investigated through fuzzy reasoning and hierarchical evaluation (Wu and Zhang, 2017; Lai et al., 2019). These research streams suggest that an effective automatic-driving system should not treat energy, comfort, and speed tracking as isolated objectives. The proposed reward function provides a mechanism for combining them. In practical terms, the controller could favor traction when speed deficiency is large, select coasting when the train is close to its target operating condition, and initiate braking when station approach and speed constraints require it.

A third finding is that robustness must be incorporated into intelligent control. The self-anti-disturbance approaches reported by Lian et al. (2020) and Yang et al. (2021) demonstrate that train-speed control must account for disturbances affecting tracking performance. Consequently, a DQN trained only on ideal trajectories may not provide sufficiently reliable behavior when operating conditions differ from its training environment. Training diversity, disturbance-aware states, and supervisory constraints are therefore necessary components of a credible deployment architecture.

A fourth finding is that iterative learning control remains theoretically complementary rather than simply obsolete. He (2019) and He and Xu (2020) demonstrate the value of learning from repeated train operations. DQN and iterative learning address adaptation differently: DQN seeks an action policy across states, whereas iterative learning improves control behavior through repeated task execution. Their conceptual combination could therefore provide an avenue for future hybrid architectures.

Overall, the findings support the suitability of DQN as an intelligent decision-making layer for automatic urban rail driving. However, these are framework-level findings derived from the supplied literature rather than empirical performance measurements. No numerical superiority over existing controllers can be claimed without controlled simulation or field experiments.

DISCUSSION

The proposed framework has several theoretical implications. First, it shifts automatic train driving from a predominantly controller-centered paradigm toward a decision-oriented learning paradigm. Conventional and iterative control approaches generally define how a control signal should be generated according to known relationships or repeated operational experience (He, 2019; He and Xu, 2020). Reinforcement learning instead defines an objective and allows the policy to learn which action is advantageous under different operating states. This distinction is particularly important when multiple objectives interact.

Second, the framework provides a mechanism for reconciling energy optimization and passenger comfort. The Q-learning study by Zhang et al. (2019) establishes energy saving as a meaningful optimization objective, whereas Wu and Zhang (2017) and Lai et al. (2019) demonstrate that comfort requires multidimensional evaluation. A DQN reward function can theoretically integrate these dimensions, although the selection of weighting coefficients remains a critical methodological issue. Excessive emphasis on energy consumption could encourage coasting or delayed acceleration at the expense of schedule adherence, whereas excessive emphasis on speed tracking could produce unnecessary traction and braking.

Third, the proposed approach differs from policy-gradient reinforcement learning in its action-value formulation. Zhang et al. (2020) demonstrated the relevance of policy-gradient learning for intelligent train control. A DQN approach is particularly attractive when the driving action space is discretized into operational levels. Nevertheless, the choice between value-based and policy-based reinforcement learning should depend on the characteristics of the control problem rather than on algorithmic preference alone.

The framework also has important limitations. A DQN requires representative training experiences, and insufficient exposure to unusual operational states may result in inappropriate decisions. Railway systems additionally impose safety requirements that cannot be delegated entirely to a learned policy. Therefore, a deployment-oriented architecture should retain supervisory protection independent of the learning agent. Disturbance-rejection research reinforces this requirement because environmental and operational variations can degrade tracking performance (Lian et al., 2020; Yang et al., 2021).

Another limitation concerns reward design. A single weighted reward function may obscure conflicts between objectives. Passenger comfort, energy consumption, punctuality, and tracking accuracy can have different operational priorities depending on the railway context.

Consequently, future research should investigate adaptive weighting or constrained reinforcement-learning formulations.

The practical implication is that DQN should preferably be considered an intelligent optimization layer within a broader automatic-driving architecture rather than a completely unrestricted replacement for safety-critical railway control. A hybrid architecture combining learned decision-making, physical train dynamics, constraint supervision, and established control mechanisms could provide a more credible transition toward deployment.

6. Conclusion

This study developed a deep reinforcement learning framework for optimal automatic driving of urban rail trains based exclusively on the supplied literature. The analysis demonstrates that existing research provides complementary foundations in iterative learning control, disturbance rejection, reinforcement learning, energy optimization, and passenger comfort evaluation. Building upon these directions, the proposed DQN framework formulates automatic train driving as a sequential decision-making problem in which speed tracking, energy consumption, comfort, station stopping, and operational constraints are jointly represented.

The principal contribution is a multi-objective methodology that connects reinforcement-learning-based decision-making with railway-specific operational requirements. DQN provides a suitable mechanism for representing complex state-action relationships, while reward engineering enables multiple operational objectives to be incorporated into the learning process. Nevertheless, the proposed framework remains conceptual and requires validation through simulation and real-world operational datasets before claims of performance superiority can be established.

Future research should investigate hybrid DQN-control architectures, disturbance-aware training, adaptive reward weighting, safety-constrained reinforcement learning, and comprehensive comparative experiments against iterative learning, self-anti-disturbance, Q-learning, and policy-gradient approaches. Such work could strengthen the transition from theoretical intelligent control toward reliable and energy-efficient automatic operation of urban rail trains.

REFERENCES

1. S. H. Lai, C. J. Chen, L. Yan and Y. P. Li, “A comprehensive comfort evaluation system for subway train passengers based on hierarchical analysis,” Science Technology and Engineering, vol. 19, no. 36, pp. 296–301, 2019.
2. W. B. Lian, B. H. Liu, W. W. Li, X. Q. Liu, F. Y. Gao et al., “Automatic speed control of high-speed trains based on self-anti-disturbance control,” Journal of the China Railway Society, vol. 42, no. 1, pp. 76–81, 2020.
3. M. Zhang, “Research on automatic train driving method based on reinforcement learning,” Ph.D. dissertation, China Academy of Railway Science, China, 2020.
4. M. Zhang, Q. Zhang and Z. X. Zhang, “Research on energy-saving optimization of high-speed railroad trains based on Q-learning algorithm,” Railway Transportation and Economy, vol. 41, no. 12, pp. 111–117, 2019.
5. M. Zhang, Q. Zhang, W. T. Liu and B. Y. Zhou, “A train intelligent control method based on policy gradient reinforcement learning,” Journal of the China Railway Society, vol. 42, no. 1, pp. 69–75, 2020.
6. J. H. Wu and X. H. Zhang, “Comprehensive evaluation of ride comfort of urban rail trains based on fuzzy reasoning,” Journal of Zhejiang Normal University (Natural Sciences), vol. 40, no. 4, pp. 453–458, 2017.
7. Z. Y. He, “Application of adaptive iterative learning control in automatic train driving system,” Ph.D. dissertation, China Academy of Railway Science, China, 2019.
8. Z. Y. He and N. Xu, “Non-parametric iterative learning control algorithm for automatic train driving control,” Journal of the China Railway Society, vol. 42, no. 12, pp. 90–96, 2020.
9. J. Yang, Y. Q. Chen and P. P. Wang, “Self-anti-disturbance controller design for train speed tracking based on improved particle swarm algorithm,” Journal of the China Railway Society, vol. 43, no. 7, pp. 40–46, 2021.
10. K. Ramamurthy, R. K. Konduru and N. Amanmadov, "EvoGraphCoder: An Evolutionary Graph-Reasoning Framework for Self-Adaptive Software Engineering," in IEEE Access, vol. 14, pp. 63063-63076, 2026, doi: 10.1109/ACCESS.2026.3686019.
11. Geo Philip, Paulson, Integrated Intelligent Building Energy Management: A Multi-LayerFramework for Renewable Energy, HVAC Optimization, and SmartElectrical Network Coordination. Available at SSRN: https://ssrn.com/abstract=6993209 orhttp://dx.doi.org/10.2139/ssrn.6993209