International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

DECENTRALIZED MULTI-AGENT REINFORCEMENT LEARNING ARCHITECTURE FOR RAILWAY TRACK DAMAGE DETECTION IN TRAIN-BASED MONITORING SYSTEMS

Aulia El Hakim[1*], Mohammad Erik Echsony [2], R. Gaguk Pratama Yudha [3], Aditya Diaz Prayogi [4], Annisa Salsa Bila Z [5]

[1,4,5] Computer Control Engineering, Department of Engineering, Madiun State of Polytechnic, Madiun, Indonesia – 63133

[*]Corresponding Author, e-mail: aim@pnm.ac.id

[2] Automation engineering technology, Department of Engineering, Madiun State of Polytechnic, Madiun, Indonesia

e-mail: erik_sony@pnm.ac.id

[3] Automation engineering technology, Department of Engineering, Madiun State of Polytechnic, Madiun, Indonesia

e-mail: gaguk@pnm.ac.id.

Abstract

Reliable monitoring of railway track conditions is essential to ensure operational safety and support predictive maintenance. This study proposes the first decentralized Multi Agent Deep Reinforcement Learning framework for real time rail damage detection in a Train Based Monitoring System. The architecture integrates three agents: DQN for front bogie vibration analysis, PPO for rear bogie vibration signals, and TD Learning for IMU based motion estimation. The models were trained using synthetic Root Mean Square vibration and IMU datasets within a Gymnasium based simulation environment representing eleven rail surface conditions. Individually, the TD Learning model achieved 49.0 percent accuracy, while the DQN and PPO models achieved 87.8 percent and 86.14 percent accuracy, respectively. A greedy ensemble fusion strategy was applied to combine the predictions of the three agents. The ensemble model achieved the best performance with 88.6 percent classification accuracy and a weighted F1 score of 0.887, demonstrating improved classification stability compared with standalone models. Several defects including corrugated rail, defect, and severe transverse crack were detected with perfect F1 scores. These results indicate that multi sensor reinforcement learning fusion improves robustness and provides a scalable solution for intelligent railway condition monitoring and automated defect detection.

AMS Subject Classification: 68T05, 93E35, 62M20, 62M45, 68T07, 93C85

Keywords: Decentralized Multi-Agent Reinforcement Learning; Train-Based Monitoring System; Railway Track Damage Detection; Ensemble Deep Reinforcement Learning; Vibration and IMU Sensor Fusion; Gymnasium Simulation Environment.

1. Introduction

The operational safety and service dependability of railway transport are strongly determined by the performance of rail condition monitoring systems that can identify defects at an early stage with high accuracy and on a continuous basis. In accordance with the Decree of the Board

Received: November 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

of Directors of PT Kereta Api Indonesia [1] concerning Regulation Dinas 10A on the maintenance of 1,067 mm gauge railway tracks, rail damage is generally categorized into several types: (1) longitudinal cracking, (2) longitudinal cracking occurring in rail transition zones, (3) star cracking patterns, (4) rail corrugation, (5) localized surface deterioration around the rail, (6) longitudinal cracking at the rail head, (7) wheel burn damage, (8) transverse cracking, (9) material defects, and (10) complete transverse fractures. Structural impairments including microcrack formation, progressive surface degradation, and deviations in rail geometry can induce irregular vibration signatures which, if left undetected, may escalate into service disruptions or catastrophic incidents [2], [3], [4]. Consequently, the development of a monitoring framework that functions in real time with high accuracy, while maintaining uninterrupted train operations, is critically required.

Traditional rail inspection approaches, including manual visual assessment and dedicated inspection trains, remain prevalent but exhibit notable constraints. Visual methods are labor intensive, time consuming, and inherently subjective, whereas inspection vehicles demand temporary service interruptions and involve substantial operational expenses [5]. To address these shortcomings, alternative strategies have been introduced, such as fixed wayside sensing systems [6], [7], [8] and aerial surveillance using Unmanned Aerial Vehicles (UAVs) [10]. Nevertheless, these approaches still fall short in delivering truly continuous monitoring with fine temporal resolution and adaptive responsiveness to changing track conditions [2], [4], [5], [9], [10]. Recent research indicates that track integrity can instead be inferred from the dynamic behavior of in service trains. By mounting accelerometers and Inertial Measurement Units on the bogie, vibration signatures and fluctuations in wheel–rail interaction forces can be captured, which closely reflect the level of structural deterioration [5], [11], [12].

The Train Based Monitoring System uses trains as mobile platforms for track inspection. However, many studies still rely on conventional machine learning with limited features and low adaptability to non stationary conditions [9], [13]. Multi Agent Reinforcement Learning enables agents to learn locally and collaborate, improving system performance. Although multi agent monitoring has been studied with wireless sensor networks [3], [14], most models remain centralized, limiting scalability and robustness. This study proposes a decentralized MARL framework for train based monitoring, where vibration and IMU sensor agents operate autonomously with minimal communication and adapt to dynamic conditions.

Using a Gymnasium simulation environment, the MADRL framework is trained to improve learning efficiency, speed convergence, and enable robust real time rail monitoring compared with centralized methods. This study develops a MADRL based Train Based Monitoring System integrating vibration and IMU sensors for real time defect classification. It also evaluates model optimization, communication strategies, and reinforcement learning parameters, and compares performance with conventional and visual inspection methods in terms of speed and accuracy.

Addressing the identified challenges requires a more adaptive and efficient rail monitoring model. This study proposes a decentralized multi agent framework that enables collaborative and autonomous monitoring through three agents: a front bogie vibration sensor for early pattern detection, a rear bogie vibration sensor for analyzing variation, and an IMU sensor for

Received: November 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

correlating vibration data with train dynamics. Each agent learns independently while exchanging limited information to improve overall performance. The model is trained within a Gymnasium based simulation environment that represents diverse rail defect scenarios, allowing pretraining to enhance convergence and learning efficiency prior to real world deployment. An ensemble deep reinforcement learning scheme is applied to optimize detection, enabling agents to adapt in real time and continuously refine defect detection and classification accuracy.

2. Literature Review

Railway infrastructure monitoring has been widely investigated [11], [12], [15], [16]. Cheng et al proposed a multi agent rail defect detection model using wireless sensor networks and soft multifunctional sensors, although the architecture remained centralized [3]. Alisjahbana developed a vibration-based rail maintenance detection method using the DR Train dataset, achieving up to 76 percent accuracy with onboard accelerometers and machine learning, but relying on a single sensor type [5]. Hakim applied a CRISP DM based machine learning approach to predict rail break locations in Indonesia using historical failure data, yet the model employed centralized supervised learning and focused mainly on fracture prediction [17], [18].

Extending prior work, this study introduces a Decentralized Multi Agent Reinforcement Learning architecture that combines vibration and IMU sensors for advanced rail monitoring. In contrast to centralized or single sensor models, the proposed framework supports autonomous and cooperative agent decisions, enabling real time adaptation to dynamic track conditions while improving scalability, robustness, and detection accuracy. Reinforcement Learning has also been widely explored [19]. Zhang et al presented a fully decentralized Multi Agent Reinforcement Learning model in which agents communicate and adapt without central control, relying on local observations and inter agent messaging, thereby strengthening system scalability and resilience [20].

This study offers several key innovations. It introduces a MARL based Train Based Monitoring model where onboard vibration and IMU sensors function as autonomous agents for real time defect detection and classification. The decentralized Multi Agent Reinforcement Learning framework improves accuracy, scalability, and adaptability under varying conditions. By combining deep reinforcement learning with multi sensor fusion in a decentralized structure, the system learns in dynamic environments, reduces dependence on central control, and enhances the robustness and efficiency of AI driven rail infrastructure monitoring.

3. Methodology

The proposed system employs a decentralized Multi Agent Reinforcement Learning architecture with three autonomous agents. The front bogie vibration sensor uses a Deep Q Network to detect early surface anomalies, the rear vibration sensor applies Proximal Policy Optimization to analyze response differences, and the IMU sensor utilizes Temporal Difference learning to relate vibrations to vehicle dynamics. Each agent operates independently within a Gymnasium environment, receiving observations, selecting classification actions, and obtaining rewards based on accuracy. Their decisions are combined through an ensemble layer

Received: November 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

to produce the final defect classification across ten PD 10A categories with severity levels Low, Medium, or High.

3.1 Simulation and Dataset Generation

The dataset was synthetically generated to represent bogie vibrations under various rail conditions. Continuous signals were segmented into 2 second windows, and the Root Mean Square amplitude was computed for each segment. Thus, every sample capture vibration energy within a defined interval, reflecting localized track irregularities. The dataset comprises three features: front bogie vibration vib_front, rear bogie vibration vib_back, and IMU angular motion imu_angle, expressed as follows.

$$s_t = [v_{f,\mathrm{RMS}}, v_{b,\mathrm{RMS}}, \theta_{\mathrm{IMU,RMS}}] \quad (1)$$

where $v_{f,\mathrm{RMS}}$and $v_{b,\mathrm{RMS}}$represent the RMS vibration amplitudes measured from the front and rear bogie frames, respectively, while $\theta_{\mathrm{IMU,RMS}}$denotes the RMS of the IMU's angular response within the same segment window. The RMS value for each signal $x(t)$in a time window T is computed as:

$$x_{\mathrm{RMS}} = \sqrt{\frac{1}{T}\int_0^T x(t)^2\, dt} \quad (2)$$

Each sample corresponds to one of eleven rail damage classes defined as follows.

$$C = \{c_0, c_1, c_2, c_3, c_4, c_5, c_6, c_7, c_8, c_9, c_{10}\} \quad (3)$$

The classification task in this study covers 11 categories of railway track conditions, including various crack patterns and surface anomalies as defined by PD 10A [1]. These are: (1) longitudinal crack, (2) transition crack, (3) star-shaped crack, (4) corrugated rail, (5) surface defect, (6) head crack, (7) wheel burn, (8) transverse crack, (9) general defect, (10) severe transverse crack, and (11) normal track condition.

3.2 Reinforcement Learning Formulation

Each agent interacts with environment $E$by observing state s_t, selecting action a_t, and receiving reward r_t, aiming to learn a policy $\pi(a_t \mid s_t)$that maximizes cumulative discounted reward. In this study, Agent 1 uses a Deep Q Network to classify rail defects from front bogie vibration data vib_front. The state $s_t$represents vibration amplitude features, while actions correspond to eleven PD10A defect classes. A positive reward is given for correct classification and negative otherwise. The DQN estimates the optimal action value $Q(s_t, a_t; \theta)$using a multilayer perceptron based on the Bellman optimality equation.

$$Q(s_t, a_t; \theta) \approx r_t + \gamma \max_{a'} Q(s_{t+1}, a'; \theta^-) \quad (4)$$

where $r_t$denotes the instantaneous reward, $\gamma$is the discount factor that balances immediate and future rewards, and θ^-represents the parameters of the target network used for stable learning. The agent iteratively updates the network weights θ to minimize the temporal difference (TD) error between the predicted and target Q-values. During the training process, the agent receives sequential vibration inputs from the simulated dataset (vib_front) and learns to maximize its cumulative reward by improving its classification accuracy across training batches. The

Received: November 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

implementation was realized using the Stable-Baselines3 DQN module with an experience replay buffer and ε-greedy policy for exploration-exploitation balance. The total reward per batch was recorded and visualized as the training progressed to observe the convergence behavior of the Q-function. The final DQN model $Q(s, a; \theta^*)$was stored and later integrated into the ensemble decision system. It provides the first classification hypothesis of rail defect types based solely on front bogie vibration characteristics.

After the front vibration signal was modeled using DQN, the second autonomous agent (Agent 2) was trained using the Proximal Policy Optimization (PPO) algorithm. This agent was designed to process vibration signals from the rear bogie sensor (vib_back), which experiences secondary vibrations that often differ in amplitude and phase from the front bogie due to rail discontinuities, wheel impacts, or propagation delays in the steel rail structure. For Agent 2 (PPO), the policy is optimized by maximizing a clipped surrogate objective:

$$L^{CLIP}(\theta) = \mathbb{E}_t[\min(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t)] \quad \ldots\ldots\ldots\ldots \quad (5)$$

where $r_t(\theta)$is the probability ratio between the new and old policy, and $\hat{A}_t$is the advantage estimate. The third agent (*Agent 3*) focuses on the inertial measurement unit (IMU) data, which captures angular motion of the bogie as the train traverses track irregularities. Unlike vibration sensors, IMU readings reflect the kinematic response of the vehicle body, such as pitch, roll, and yaw oscillations resulting from dynamic coupling between rail irregularities and train suspension. Therefore, a Temporal Difference (TD-Learning) method is employed, offering a lightweight yet effective learning mechanism for continuous adaptation of state–reward predictions. For Agent 3 (TD-Learning), the update rule follows:

$$V(s_t) \leftarrow V(s_t) + \alpha[r_t + \gamma V(s_{t+1}) - V(s_t)] \quad (6)$$

where $V(s_t)$ is the estimated value of the current state, $r_t$is the immediate reward, γ is the discount factor for future rewards, and, α is the learning rate controlling the update magnitude. This rule adjusts the current state value $V(s_t)$toward the temporal difference target $r_t + \gamma V(s_{t+1})$. In the context of IMU-based motion monitoring, each state $s_t$encodes the longitudinal angular motion (pitch angle) of the bogie as the train passes over various rail conditions. To achieve reliable and interpretable defect recognition, this study integrates three autonomous reinforcements learning agents DQN, PPO, and TD-Learning into an Ensemble Decision Layer. Each agent specializes in a specific sensing modality. The Ensemble Decision Layer integrates the normalized Q-values, policy probabilities, and value estimates from all agents using a weighted combination:

$$D_{ensemble} = \text{argmax}_i(w_1 Q_i + w_2 \pi_i + w_3 V_i) \quad (7)$$

Where, Q_i = normalized state–action value predicted by the DQN agent, π_i= categorical action probability output by the PPO agent, V_i= estimated state value from the TD-Learning agent, and w_1, w_2, w_3= weighting coefficients controlling each agent's contribution to the final decision. This equation defines a weighted voting mechanism that prioritizes agents with higher reliability in their sensing domain. The final predicted damage class D_{ensemble} corresponds to the maximum aggregated confidence score among the 10 predefined rail-defect categories.

Received: November 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

4. Experimental Evaluation

The experiments were conducted in a Gymnasium based simulation environment utilizing the RMS processed vibration and IMU dataset. This dataset reflects vibration energy derived from simulated train bogie dynamics, where each entry corresponds to the Root Mean Square amplitude calculated over a 1 second time window. With a total duration of 6 minutes, the data produce 360 sequential segments, and each segment functions as a discrete observation state in the reinforcement learning framework. Three sensor modalities were used in each segment: front bogie vibration ($v_{f,\mathrm{RMS}}$), rear bogie vibration ($v_{b,\mathrm{RMS}}$), and IMU angular response ($\theta_{\mathrm{IMU,RMS}}$). The extracted RMS features represent the effective vibration energy and angular displacement over consecutive rail sections, forming a time ordered depiction of the track condition along the route.

A. Deep Q-Network (DQN) Training

The DQN agent was trained to derive optimal classification policies from RMS features extracted from vibration and IMU signals in the simulated rail dataset. Each observation corresponds to a one second RMS segment of front and rear bogie vibrations together with IMU angular motion, representing localized track conditions along a six-minute trajectory consisting of 360 segments. the sensor features are combined into a state vector S_{t} and processed by a multilayer perceptron Q network that estimates $\mathrm{Q}(s,a)$ for eleven rail defect classes. The agent selects actions using an argmax policy, while experience tuples $(s_{\mathrm{t}}, a_{\mathrm{t}}, r_{\mathrm{t}}, s_{\mathrm{t+1}})$ are stored in a replay buffer. Mini batch sampling and temporal difference loss are used to update network parameters, enabling the model to learn the relationship between vibration patterns and rail defect categories.

The agent interacts with the environment (E) via the RailSingleAgentEnv, which supplies state observations as defined in Eq. 1 and returns a scalar reward r_t based on the classification accuracy for each segment. The DQN model was implemented using Stable Baselines3 with PyTorch, employing a Multilayer Perceptron policy to approximate the Q function. The main training parameters and environment configurations are presented in Table 1.

Table 1. DQN Training Configuration

Parameter	Description	Value / Setting
Framework	Reinforcement learning library	*Stable-Baselines3 (PyTorch)*
Environment	Custom environment for rail vibration data	RailSingleAgentEnv
Input Dataset	Source of vibration and IMU data	data/sintesis_dataset.csv
Policy Network	Type of neural network policy	MlpPolicy (Multilayer Perceptron)
Random Seed	Reproducibility parameter	42
Learning Rate	Step size for parameter update	5e-4

Received: November 06, 2025

Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Replay Buffer Size	Experience memory capacity	100000 transitions
Learning Starts	Warm-up timesteps before updates	2000
Total Training Steps	Number of timesteps trained	40.000 (10 batches × 4000)
Evaluation Interval	Reward monitoring after each batch	Every 2 000 steps
Reward Function	Cumulative accuracy-based reward per episode	Environment-defined
Logging & Plotting	Reward per batch visualization	Saved to models/reward_dqn_plot
Output Model	Trained policy checkpoint	models/dqn_agent1_front.zip

The DQN agent was trained with a learning rate of $5 \times 10^{-4}$and a discount factor $(\gamma)$of 0.98. The exploration rate $(\varepsilon)$was gradually reduced from 1.0 to 0.05 during the first 25% of training steps to balance exploration and exploitation. Cumulative rewards were evaluated every 2,000 timesteps, and a five-batch moving average was used to assess training stability. As illustrated in Figure 1, the reward curve increased rapidly during early training and stabilized around the eighth batch, indicating convergence of the Q network policy. The final evaluation produced a cumulative reward of 1,331.25 at batch 10 of 10, demonstrating that the agent successfully learned the relationship between front bogie RMS vibration signals and rail damage classes while maintaining stable decision performance, illustrated in Figure 1(a).

B. Proximal Policy Optimization (PPO) Training

The second model applies the Proximal Policy Optimization algorithm to evaluate vibration signals from the rear bogie sensor. PPO was chosen due to its ability to maintain policy stability while supporting exploration through a clipped surrogate objective. The loss formulation is given in Eq. 5. Training was conducted in the same RailSingleAgentEnv environment using synthetic RMS based vibration and IMU datasets. The agent takes rear bogie vibration amplitude as the main observation and predicts a discrete rail damage class ranging from 0 to 10. The training settings are summarized in Table 2.

Table 2. PPO Training Configuration

Parameter	**Value / Setting**
Framework	Stable-Baselines3 (PyTorch)
Environment	RailSingleAgentEnv using *dataset.csv*
Policy Type	MlpPolicy (Fully-connected MLP)
Random Seed	43

Received: November 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Total Timesteps	20 000
Learning Rate	3×10^{-4}
Rollout Steps per Update	2048
Batch Size	64
Entropy Coefficient	0.0
Clipping Parameter (ε)	0.2
TensorBoard Log	logs/ppo_agent2/
Checkpoint Interval	every 5000 steps

During training, the PPO agent was executed for twenty iterative batches, resulting in approximately 2.4 million environment interactions (20,000 timesteps). The average episode length (enplanement ≈ 3,950) indicates long and stable rollouts over segment-based vibration data, suggesting consistent exploration within the simulated rail environment. The adaptive reward shaping strategy gradually improved model performance, producing a steady upward reward trend as illustrated in Figure 2. Training stability was maintained by the clipped surrogate objective, where the approximate KL divergence (0.0556) and clip fraction (0.172) remained within a safe range, ensuring controlled policy updates. The entropy loss (−0.74) also indicates reduced randomness and convergence toward a more deterministic policy. In the final evaluation, the PPO agent achieved a cumulative reward of 13,240.8, demonstrating that the model successfully learned to map rear bogie RMS vibration features to the correct rail damage classes while maintaining stable learning behaviour throughout the training process.

The approximate KL divergence (0.0556) remained stable, indicating controlled PPO policy updates. The entropy loss (−0.74) and policy gradient loss (0.0022) suggest a balanced exploration–exploitation process and gradual convergence toward a deterministic policy. The value function loss (5.38×10^3) with zero explained variance indicates that the critic captured long horizon dependencies in the nonlinear rear bogie vibration data. After about 11,790 updates, the PPO model converged and reached a final cumulative reward of 13,240.8. As shown in Figure 1(b), the reward curve increases steadily and stabilizes at later timesteps, confirming reliable convergence and effective learning of vibration-based rail condition features.

C. Temporal Difference (TD-Learning) Training

The third agent employed a tabular Temporal Difference learning algorithm to classify rail conditions using IMU angular responses as the input feature. The continuous IMU signal was discretized into 20 state bins representing angular vibration levels, while the action space remained consistent with the other agents, covering 11 rail damage classes (0–10). Training was performed over 2,000 episodes using an ε greedy exploration strategy with learning rate $\alpha = 0.05$, discount factor $\gamma = 0.95$, and ε gradually reduced from 1.0 to 0.05. The detailed training configuration is presented in Table 3.

Received: November 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

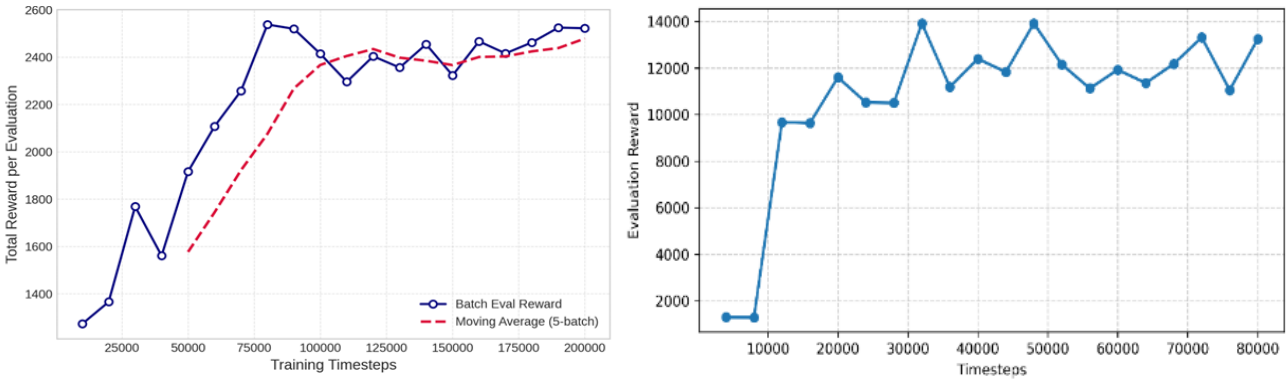
Table 3. TDL Training Configuration

Parameter	Description	Value / Setting
Input Dataset	Simulated IMU-angle data from vibration testbed	data/sintesis_dataset.csv
State Representation	Discretized IMU amplitude into N bins	N_BINS = 40
Action Space	Discrete (rail-defect classes 0–10)	N_ACTIONS = 11
Learning Rate (α)	Step size for state-value updates	0.05
Discount Factor (γ)	Future reward weighting	0.95
Exploration Policy	ε-greedy with exponential decay	ε = 1.0 → 0.05, decay = 0.999
Episodes	Number of training iterations	N_EPISODES = 2000
Max Steps per Episode	Uses environment episode length	None
Random Seed	Reproducibility parameter	123
Saved Parameters	Discretization edges, Q-table, and policy	models/td_agent3_bin_edges.npy models/td_agent3_qtable.npy models/td_agent3_policy.npy

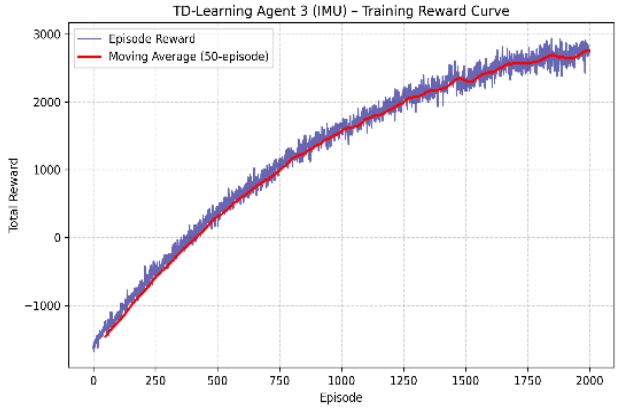
After 2000 episodes, the TD Learning agent showed stable convergence, as illustrated in Figure 1(c), which presents the episode reward and the moving average (50 episode) across the training process. The cumulative reward increased steadily and stabilized near the final episodes, indicating consistent learning progress. The average episode reward reached approximately 2,758. Under greedy policy evaluation, the learned Q table achieved a total reward of 2,336.65 over 3,959 steps, demonstrating that the agent successfully captured IMU angular motion patterns to distinguish different rail conditions. Overall, the results indicate that tabular TD Learning effectively learned a reliable state action mapping from discretized IMU data, providing an efficient baseline for real time rail monitoring in a multi sensor reinforcement learning framework.

Received: November 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)



(a) (b)



(c)

Figure 1.(a) Training Progress of DQN Agent-1, (b) PPO Agent-2, (c)TDL Agent 3

5. Discussion

The experimental results show that the decentralized MADRL framework achieves greater robustness, faster convergence, and higher classification accuracy than single agent or centralized reinforcement learning methods. Its decentralized design allows each agent to specialize in its sensor domain while coordinating through minimal information exchange, distributing computation and supporting scalable real time monitoring. The ensemble fusion layer further improves reliability by reducing noise and strengthening decision confidence. By combining vibration and IMU signals, the system identifies defect types and estimates severity levels, providing more informative diagnostics for maintenance planning.

A. Deep Q-Network (DQN) Testing

The normalized confusion matrix of the DQN model, illustrated in Figure 2(a), demonstrates a moderate yet stable classification capability across the eleven rail condition categories. The matrix shows clear diagonal dominance for several classes, indicating that the model was able to correctly identify multiple defect patterns based on the vibration signals. Nevertheless, the classification performance varies across categories, suggesting that some rail defects present similar vibration characteristics that make them more difficult to distinguish.

Several classes achieved very high recognition performance. In particular, Corrugated rail, Defect, and Severe transverse crack obtained perfect recall values of 1.000, with corresponding F1 scores of 0.974, 0.999, and 1.000 respectively, indicating that these vibration signatures

Received: November 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

were consistently detected by the model. The Transition crack category also demonstrated excellent performance with a recall of 0.994 and an F1 score of 0.989. Likewise, Surface defect and Star-shaped crack showed strong recognition capability with recall values of 0.950 and 0.958 and F1 scores of 0.955 and 0.932 respectively, confirming that their vibration patterns were effectively learned by the DQN agent.

In contrast, several categories exhibited weaker detection capability. The Head crack class showed the lowest recall at 0.381 with an F1 score of 0.522, indicating that many samples were misclassified, most likely due to overlapping vibration characteristics with other crack related defects. The Normal condition also showed relatively low precision of 0.536 despite achieving a high recall of 0.947, suggesting that a number of defective rail signals were incorrectly predicted as normal conditions. Additionally, Wheel burn and Transverse crack achieved moderate performance with recall values of 0.767 and 0.803 and F1 scores of 0.868 and 0.861 respectively, reflecting partial confusion among defects with similar vibration responses.

Overall, the model achieved an accuracy of 0.878, which corresponds to 87.8% classification accuracy across 3,960 evaluation samples. The macro average precision, recall, and F1 scores reached 0.902, 0.878, and 0.875, while the weighted average F1 score also reached 0.875. These results indicate that the DQN model is capable of identifying several rail defect patterns effectively, although the variation in class wise performance suggests that further improvement in feature discrimination is still required for certain defect categories. The detailed training configuration of the DQN model is summarized in Table 4.

Table 4. Performance Metrics of 3 Models for Rail Damage Classification

Damage Class	Precision			Recall			F1-Score			Support		
	DQN	PPO	TDL	DQN	PPO	TDL	DQN	PPO	TDL	DQN	PPO	TDL
Longitudinal crack	0.828	0.9076	0.818	0.856	0.8727	0.461	0.842	0.8898	0.590	360	18000	360
Transition crack	0.984	0.8815	0.000	0.994	0.9972	0.000	0.989	0.9358	0.000	360	18000	360
Star-shaped crack	0.908	0.9802	0.000	0.958	0.9750	0.000	0.932	0.9776	0.000	360	18000	360
Corrugated rail	0.950	0.9996	0.398	1.000	1.0000	0.981	0.974	0.9998	0.566	360	18000	360
Surface defect	0.961	0.9826	0.598	0.950	0.9667	0.407	0.955	0.9745	0.484	360	18000	361
Head crack	0.830	0.4482	0.846	0.381	1.0000	0.092	0.522	0.6190	0.165	360	18000	360
Wheel burn	1.000	1.0000	0.986	0.767	0.7667	0.381	0.868	0.8679	0.549	360	18000	360

Received: November 06, 2025

Transverse crack	0.929	0.9341	0.305	0.803	0.8667	0.639	0.861	0.8991	0.413	360	18000	360
Defect	0.997	1.0000	0.528	1.000	1.0000	0.628	0.999	1.0000	0.574	360	18000	360
Severe transverse crack	1.000	1.0000	0.596	1.000	1.0000	0.992	1.000	1.0000	0.745	360	18000	360
Normal	0.536	0.0000	0.441	0.947	0.0000	0.816	0.685	0.0000	0.573	360	17950	358
Overall Accuracy	-	-	-	-	-	-	0.878	0.8588	0.490	3960	19795	3959
Macro Average	0.902	0.8303	0.501	0.878	0.8586	0.490	0.875	0.8330	0.423	3960	19795	3959
Weighted Average	0.902	0.8305	0.501	0.878	0.8588	0.490	0.875	0.8333	0.423	3960	19795	3959

B. Proximal Policy Optimization (PPO) Testing

The confusion matrix of the PPO + Classifier model shown in Figure 2(b) indicates strong recognition for several rail damage categories using rear bogie vibration signals. The matrix presents clear diagonal dominance in many classes, suggesting that the model successfully learned distinctive vibration patterns for different rail conditions. In particular, Corrugated rail, Defect, and Severe transverse crack achieved perfect recall of 1.000 with F1 scores close to 1.000, indicating highly reliable detection. Transition crack also performed very well with precision of 0.9993, recall of 0.9972, and an F1 score of 0.9982. Similarly, Star shaped crack and Surface defect showed strong performance with F1 scores of 0.9764 and 0.9788.

Moderate performance was observed in several other classes. Longitudinal crack and Transverse crack achieved F1 scores of 0.8820 and 0.8849, indicating reasonably stable recognition despite some similarity in vibration signatures. Wheel burn obtained high precision of 0.9485 but a lower recall of 0.7667, resulting in an F1 score of 0.8479.

Some imbalance appears in certain categories. Head crack achieved very high recall of 0.9861 but low precision of 0.4504, producing an F1 score of 0.6184, indicating that many other defects were misclassified as this class. In contrast, the Normal class recorded zero precision, recall, and F1 score, suggesting that normal rail conditions were not correctly detected during evaluation.

Overall, the PPO model achieved 86.14% accuracy (0.8614) over 197950 samples. The macro average precision, recall, and F1 scores were 0.8290, 0.8612, and 0.8351, while the weighted average F1 score reached 0.8353. These results show that the model successfully learned several key vibration patterns for rail defect detection, although performance differences across some classes indicate that further improvement in training balance may still be required.

Received: November 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

C. Temporal Difference (TD-Learning) Testing

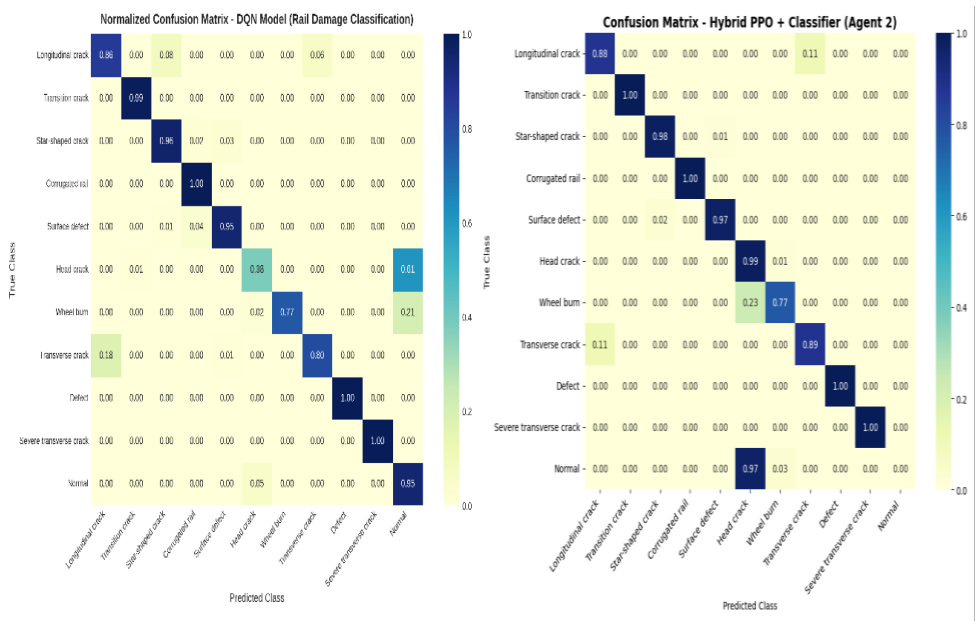
The TD Learning agent was evaluated after optimizing the discretization and exploration settings using quantile based IMU bins. The model achieved an accuracy of 0.490 (49.0%) over 3,959 samples, with a macro F1 score of 0.423, indicating limited but consistent ability to capture vibration patterns related to rail defects.

The normalized confusion matrix in Figure 2(c) shows that several classes were recognized with high recall. Severe transverse crack achieved the highest recall (0.992, F1 = 0.745), while Corrugated rail also showed strong recall (0.981) with an F1 score of 0.566. The Normal class reached a recall of 0.816 with an F1 score of 0.573, indicating reasonable detection of normal vibration signals.

Other classes showed moderate performance. Longitudinal crack achieved an F1 score of 0.590, while Defect reached 0.574. Wheel burn produced an F1 score of 0.549 with high precision (0.986) but lower recall (0.381). Meanwhile, Surface defect and Transverse crack obtained F1 scores of 0.484 and 0.413, reflecting confusion among similar vibration patterns.

However, several classes remained difficult to identify. Head crack showed very low recall (0.092) with an F1 score of 0.165, indicating frequent misclassification. More critically, Transition crack and Star shaped crack obtained zero precision and recall, meaning the agent failed to detect these defects. This suggests that their vibration signatures were either underrepresented or too similar within the discretized state space.

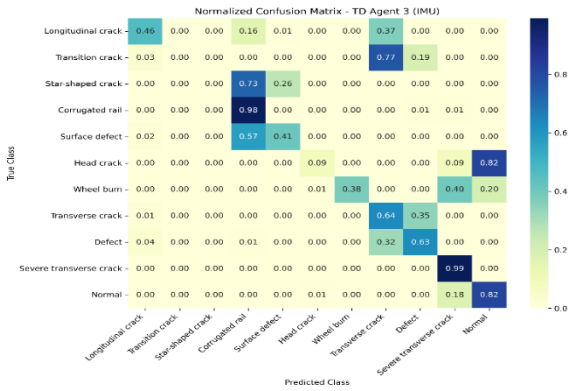
Overall, the TD Learning model achieved 49.0% classification accuracy, with macro and weighted average F1 scores of 0.423, indicating limited but measurable learning capability. Although the performance remains lower than that of more advanced reinforcement learning models, the results confirm that tabular TD Learning can still capture certain temporal vibration patterns and provide a lightweight baseline for rail defect classification. These findings highlight the potential of discretized reinforcement learning approaches as an initial framework for integrating vibration-based rail monitoring systems with more advanced learning architectures, the detailed training configuration is presented in Table 4.



Received: November 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

(a) b)



(c)

Figure 2. (a) Normalize Confusion matrix – DQN, (b) PPO Agent 2, (c) TDL Agent 3,

D. Ensemble Decision

The *Ensemble Normalized Confusion Heatmap* illustrated in Figure figure 3, visualizes the classification performance of the proposed multi-agent ensemble model across 11 categories of rail surface conditions. The horizontal axis represents the predicted labels, while the vertical axis corresponds to the true labels obtained from the testing dataset. The color intensity indicates the normalized classification accuracy for each class, with darker blue shades denoting higher accuracy values and lighter yellow regions reflecting lower prediction confidence or misclassification.

From the visualization, it can be observed that several damage classes such as Corrugated rail, Defect, and Severe transverse crack achieved notably high normalized accuracy values, reaching above 0.9. This indicates that the ensemble system consistently and accurately identifies these categories, which typically produce distinctive vibration and IMU signal patterns that are easily separable from other classes. In addition, the Wheel burn and Transverse crack categories also exhibit strong diagonal dominance (0.79 and 0.64 respectively), demonstrating that the ensemble framework maintains stable recognition of medium-severity defects with minimal cross-class confusion.

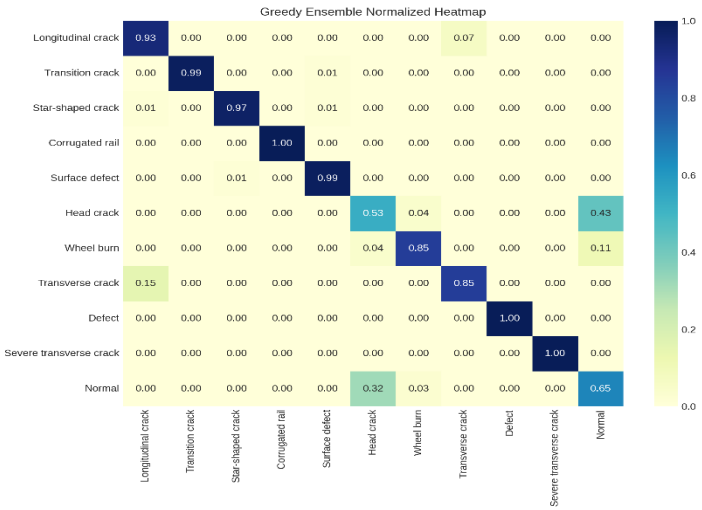


Figure 3. Ensemble Decision Reinforcement Learning

Received: November 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Conversely, the lower confidence regions observed in Longitudinal crack, Transition crack, and Star-shaped crack suggest that these fine-grained or early-stage damages have overlapping vibration characteristics. This overlap often results in partial misclassification with neighbouring classes such as *Surface defect* or *Normal condition*, where vibration amplitudes and frequency responses are less distinct.

Overall, the ensemble heatmap highlights the capability of the MADRL-based decision fusion mechanism to enhance discrimination across complex rail defect types. The ensemble effectively integrates the complementary strengths of the individual agents DQN (front vibration), PPO (rear vibration), and TD-Learning (IMU-based state estimation) through adaptive weighting. This results in robust and reliable classification boundaries, especially for severe and clearly defined damage types, validating the ensemble's suitability for real-time rail damage monitoring in inspection train systems and potential edge-computing deployment scenarios.

Table 5. Classification Performance of the Multi-Agent Ensemble Model Across 11 Rail Damage Classes

Class	Precision	Recall	F1-Score	Support
Longitudinal crack	0.848	0.931	0.887	72
Transition crack	1.000	0.986	0.993	72
Star-shaped crack	0.986	0.972	0.979	72
Corrugated rail	1.000	1.000	1.000	72
Surface defect	0.973	0.986	0.979	72
Head crack	0.594	0.528	0.559	72
Wheel burn	0.924	0.847	0.884	72
Transverse crack	0.924	0.847	0.884	72
Defect	1.000	1.000	1.000	72
Severe transverse crack	1.000	1.000	1.000	72
Normal	0.547	0.653	0.595	72
Acuracy			0.886	792
Macro Average	0.890	0.886	0.887	792
Weighted Average	0.890	0.886	0.887	792

Table 5 summarizes the classification performance of the proposed multi agent ensemble system, which combines three reinforcement learning models: DQN using front bogie vibration, PPO using rear bogie vibration, and TD Learning using IMU state estimation to identify eleven rail surface conditions. The evaluation uses precision, recall, and F1 score metrics. As illustrated in the confusion matrix in Figure 5, the ensemble model shows strong

Received: November 06, 2025

Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

diagonal dominance across most classes, indicating accurate recognition of the vibration patterns associated with different rail defects.

Several defect categories demonstrate very strong classification performance. In particular, Corrugated rail, Defect, and Severe transverse crack achieved perfect scores with precision, recall, and F1 score all equal to 1.000, indicating completely correct identification of these conditions. Other categories also exhibit high recognition capability, such as Transition crack with an F1 score of 0.993, Star shaped crack and Surface defect with F1 scores of 0.979, and Longitudinal crack with an F1 score of 0.887. These results suggest that the ensemble system successfully captures distinctive vibration signatures associated with both severe and moderate rail damage patterns.

Moderate performance appears in several categories that exhibit more complex or overlapping vibration characteristics. Wheel burn and Transverse crack both achieved F1 scores of 0.884, indicating reliable but not perfect detection. Meanwhile, Head crack produced a lower F1 score of 0.559, suggesting that this defect type still shares vibration similarities with other classes. The Normal rail condition also achieved a moderate F1 score of 0.595, indicating that distinguishing normal track behaviour from mild defects remains a challenge.

Overall, the ensemble model achieved an accuracy of 0.886 (88.6%) across 792 samples, with macro and weighted F1 scores of 0.887. The confusion matrix further confirms that most predictions concentrate along the diagonal, while only a few misclassifications appear, mainly involving Head crack, Normal, and Wheel burn categories.

When compared with the individual reinforcement learning models, the ensemble approach shows a clear improvement in performance. The TD Learning agent previously achieved only 49.0% accuracy, with significant difficulty detecting several crack categories. The DQN model improved performance to approximately 87.8% accuracy, but still showed misclassification in several defect types. The PPO classifier reached around 86.14% accuracy, yet it struggled to correctly detect the Normal condition and produced class imbalance in certain categories. By integrating the predictions of all three agents, the ensemble model achieved 88.6% accuracy, the highest among all evaluated approaches.

This improvement occurs because the ensemble mechanism leverages complementary information from different sensing perspectives. The DQN agent captures rapid vibration dynamics from the front bogie, the PPO agent models smoother temporal patterns from rear bogie signals, and the TD Learning agent contributes contextual state information derived from IMU measurements. By aggregating these heterogeneous predictions, the ensemble reduces prediction variance and improves overall robustness against noisy or ambiguous vibration signals, resulting in more stable rail condition classification.

6. Future Work

Future research will focus on extending the current simulation-based framework toward a hybrid digital twin environment integrated with a rail inspection simulator. The next stage will involve collecting real vibration and IMU signals from an inspection train prototype under controlled damage conditions. Frequency-domain features (FFT, spectral centroid, and band energy) will be extracted to enhance the discriminative power of each agent. Furthermore,

Received: November 06, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

transfer learning will be employed to bridge the gap between synthetic and real-world data, and the ensemble mechanism will be refined with an attention-based weighting strategy. Integration with ROS 2 and Gazebo will allow reinforcement learning agents to interact directly with the physical simulation for real-time adaptive policy updates.

7. Conclusion

This study proposed a Decentralized Multi Agent Deep Reinforcement Learning architecture for railway track damage detection within a Train Based Monitoring System. The framework integrates three reinforcement learning agents that process different sensing perspectives: DQN for front bogie vibration analysis, PPO for rear bogie vibration interpretation, and TD Learning for IMU based motion state estimation. The models were trained and evaluated using synthetic RMS vibration and IMU datasets in a Gymnasium based simulation environment representing eleven rail surface conditions.

Experimental evaluation shows different performance levels among the individual agents. The TD Learning agent achieved 49.0% accuracy with a macro F1 score of 0.423, indicating limited capability in distinguishing several defect categories. The DQN model significantly improved classification performance with 87.8% accuracy and a macro F1 score of 0.875, demonstrating strong recognition of several defect patterns such as corrugated rail, defect, and severe transverse crack. The PPO classifier achieved 86.14% accuracy with a weighted F1 score of 0.835, successfully capturing rear bogie vibration dynamics but still showing imbalance in several classes, particularly in normal rail condition detection.

To further enhance robustness, the outputs of the three agents were integrated using a greedy ensemble decision mechanism. The resulting ensemble model achieved the best overall performance with 88.6% classification accuracy and a weighted F1 score of 0.887 across 792 evaluation samples. Several defect categories including corrugated rail, defect, and severe transverse crack achieved perfect detection with F1 scores of 1.000, while transition crack, star shaped crack, and surface defect also demonstrated excellent performance with F1 scores above 0.97. Moderate performance was observed in wheel burn and transverse crack, whereas head crack and normal rail condition remained the most challenging classes due to overlapping vibration characteristics.

Overall, the results confirm that combining decentralized reinforcement learning agents with multi sensor vibration fusion improves classification stability and robustness compared with individual models. The ensemble approach successfully leverages complementary information from front bogie vibration, rear bogie vibration, and IMU based motion dynamics, resulting in more reliable rail condition identification. These findings demonstrate the potential of MADRL based TBMS frameworks as scalable and intelligent solutions for automated railway infrastructure monitoring and predictive maintenance.

8. Acknowledgement

The author wishes to express sincere gratitude to the Directorate of Research and Community Service, the Directorate General of Research and Development, and the Ministry of Higher Education, Science, and Technology of the Republic of Indonesia for the financial support provided for the implementation of this research. Appreciation is also extended to Politeknik

Received: November 06, 2025

Negeri Madiun for the facilities and technical assistance that made the successful completion of this study possible.

References

[1] PT Kereta Api Indonesia (Persero), "PD 10A (Perawatan Jalan Rel Dengan Lebar Jalan 1067 MM) | PDF." Accessed: Nov. 03, 2025. [Online]. Available: https://id.scribd.com/document/887347926/PD-10A-Perawatan-Jalan-Rel-Dengan-Lebar-Jalan-1067-Mm

[2] S. R. Nugroho, A. T. A. Salim, M. E. Echsony, S. N. Patrialova, and R. H. Setyawan, "Perangkat Pengukur Kondisi Kemiringan Rel Kereta Api Menggunakan Sensor Giroskop Berbasis GPS," J. Sci. Eng., vol. 7, no. 1, pp. 1–11, May 2024, doi: 10.33387/josae.v7i1.8009.

[3] X. Cheng, D. Yao, L. Yang, and W. Dong, "Collaborative Damage Detection Framework for Rail Structures Based on a Multi-Agent System Embedded with Soft Multi-Functional Sensors," Sensors, vol. 22, no. 20, p. 7795, Oct. 2022, doi: 10.3390/s22207795.

[4] T. A. Alvarenga, A. L. Carvalho, L. M. Honorio, A. S. Cerqueira, L. M. A. Filho, and R. A. Nobrega, "Detection and Classification System for Rail Surface Defects Based on Eddy Current," Sensors, vol. 21, no. 23, p. 7937, Nov. 2021, doi: 10.3390/s21237937.

[5] I. Alisjahbana, "Using In-Service Train Vibration for Detecting Railway Maintenance Needs," May 2024, Accessed: Nov. 03, 2025. [Online]. Available: https://arxiv.org/pdf/2405.09560

[6] R. Skrypnyk, U. Ossberger, B. A. Pålsson, M. Ekh, and J. C. O. Nielsen, "Long-term rail profile damage in a railway crossing: Field measurements and numerical simulations," Wear, vol. 472–473, p. 203331, May 2021, doi: 10.1016/j.wear.2020.203331.

[7] Y. Li, J. Liu, and Y. Wang, "Railway Wheel Flat Detection Based on Improved Empirical Mode Decomposition," Shock Vib., vol. 2016, pp. 1–14, 2016, doi: 10.1155/2016/4879283.

[8] C. Andrade and P. C. Neves, "Perceived Organizational Support, Coworkers' Conflict and Organizational Citizenship Behavior: The Mediation Role of Work-Family Conflict," Adm. *Sci.*, vol. 12, no. 1, p. 20, Jan. 2022, doi: 10.3390/admsci12010020.

[9] M. A. Indu and M. L. Smitha, "Enhancing the Efficiency of Damage Detection in Railway Track using Deep Learning," vol. 7, pp. 2421–2430, 2024.

[10] D.-Z. Dang, C.-C. Lai, Y.-Q. Ni, Q. Zhao, B. Su, and Q.-F. Zhou, "Image Classification-Based Defect Detection of Railway Tracks Using Fiber Bragg Grating Ultrasonic Sensors," Appl. Sci., vol. 13, no. 1, p. 384, Dec. 2022, doi: 10.3390/app13010384.

[11] J. Liu et al., "Dynamic responses, GPS positions and environmental conditions of two light rail vehicles in Pittsburgh," Sci. Data, vol. 6, no. 1, p. 146, Aug. 2019, doi: 10.1038/s41597-019-0148-9.

[12] L. Jing, K. Wang, and W. Zhai, "Impact vibration behavior of railway vehicles: a state-

Received: November 06, 2025

of-the-art overview," Acta Mech. Sin., vol. 37, no. 8, pp. 1193–1221, Aug. 2021, doi: 10.1007/s10409-021-01140-9.

[13] S. B. Setyawan, H. Arrosida, A. Elhakim, D. Shahab, and E. P. Nugroho, "Realtime road damage detection using transfer learning with Nvidia Jetson Nano," 2024, p. 020006. doi: 10.1063/5.0214103.

[14] W.-W. Wu, "Beyond business failure prediction," Expert Syst. Appl., vol. 37, no. 3, pp. 2371–2376, Mar. 2010, doi: 10.1016/j.eswa.2009.07.056.

[15] F. E. B. Setyawan, S. Supriyanto, F. Tunjungsari, W. O. N. Hanifaty, and R. Lestari, "Medical staff services quality to patients satisfaction based on SERVQUAL dimensions," Int. J. Public Heal. Sci., vol. 8, no. 1, pp. 51–57, 2019, doi: 10.11591/ijphs.v8i1.17066.

[16] W. A. Ciptaningrum A, J. I, B. RM, A. RAN, and Y. RGP, "Design of YOLOv5 Medium as Unmanned Rail Inspection in Braking Control System Based on Computer Vision," Int J Sci Eng Inf Technol, vol. 8, no. 2, [Online]. Available: https://journal.trunojoyo.ac.id/ijseit/article/view/26199/9861

[17] A. El Hakim, H. Hindersah, and E. Rijanto, "Application of reinforcement learning on self-tuning PID controller for soccer robot multi-agent system," Proc. 2013 Jt. Int. Conf. Rural Inf. Commun. Technol. Electr. Technol. rICT ICEV-T 2013, 2013, doi: 10.1109/RICT-ICEVT.2013.6741546.

[18] N. N. Hakim, "Implementasi Machine Learning pada Sistem Prediksi Kejadian dan Lokasi Patah Rel Kereta Api di Indonesia," J. Sist. Cerdas, vol. 3, no. 1, pp. 25–35, May 2020, doi: 10.37396/jsc.v3i1.58.

[19] Y. Wang, Y. Sun, W. Ji, and J. Xu, "Self-Tuning Parameters of a Maglev Control System Based on Q-Learning," Mechatronics Intell. Transp. Syst., vol. 3, no. 2, Jun. 2024, doi: 10.56578/mits030205.

[20] K. Zhang, Z. Yang, H. Liu, T. Zhang, and T. Başar, "Fully Decentralized Multi-Agent Reinforcement Learning with Networked Agents," 35th Int. Conf. Mach. Learn. ICML 2018, vol. 13, pp. 9340–9371, Feb. 2018, Accessed: Nov. 03, 2025. [Online]. Available: https://arxiv.org/pdf/1802.08757

Received: November 06, 2025