EV Energy Optimization through Q-Learning and Heat Management: A Data-Driven Approach

Rohit Kadam[1], Dr. Pavitran V S[2], Kiran Gaikwad[3]
[1]*Student at Savitribai Phule Pune University*
[2]*Professor at Savitribai Phule Pune University*
[3]*General Manager at Tata Toyo Radiator Ltd.*

***Abstract*—Electric vehicles (EVs) face significant challenges in maximizing energy efficiency and managing thermal loads, both of which are critical to enhancing range, battery health, and overall performance. This research presents a novel, data-driven approach that integrates Q-learning with thermal-aware optimization to improve EV energy consumption patterns. By leveraging real-world telemetry data and engineering a new derived feature—heat generated from energy usage—we trained a reinforcement learning model to dynamically recommend optimal driving and operational strategies. Through simulation, our framework achieved up to a 10 percent improvement in the coefficient of performance (COP) and up to a 15 percent reduction in heat generation across varying driving conditions. These enhancements contribute to better energy conservation and mitigate thermal stress on critical EV components like motors and batteries.**

Additionally, the model delivers actionable recommendations on optimal vehicle speed and battery management, offering drivers practical guidance to maximize efficiency during trips. The results underscore the potential of combining reinforcement learning with heat-aware metrics for intelligent EV management, paving the way for smarter, more resilient electric mobility systems.

***Index Terms*—Deep Learning, Reinforcement learning, EV (Electric Vehicle), Energy and heat Optimization**

1. INTRODUCTION

1.1. Overview

The rapid global transition toward electric vehicles (EVs) has placed immense importance on maximizing energy efficiency and optimiz- ing thermal management. While EV technology has matured sig- nificantly in recent years, the inherent challenges related to energy consumption and heat generation persist, directly impacting vehi- cle range, performance, and battery longevity. Traditional energy management strategies often rely on static models and predefined control rules that lack adaptability to real-world, dynamic conditions. As a result, there is a growing need for intelligent, self-learning sys- tems capable of continuously optimizing operational parameters in real-time.

One of the key metrics in evaluating EV efficiency is the Coeffi- cient of Performance (COP), which represents the ratio of useful work output to energy input. Improving COP directly translates into better energy utilization, longer driving ranges, and reduced environmental impact. However, operational improvements aimed solely at enhanc- ing COP often overlook the critical aspect of thermal load, which, if unmanaged, can lead to overheating, accelerated battery degradation, and reduced system reliability.

To address these challenges, this study proposes a novel frame- work that integrates Q-learning, a model-free reinforcement learning technique, with thermal-aware optimization strategies for electric vehicles. By analyzing real-world telemetry data and engineering new features that account for heat generation, our approach trains an agent capable of recommending dynamic operational decisions that balance energy efficiency and thermal safety. The system not only learns to maximize COP but also minimizes heat generation, ensuring a holistic improvement in vehicle performance.

Our simulation results demonstrate that the proposed method can achieve up to a 10 percent increase in COP and up to a 15 percent re- duction in heat generation under diverse driving conditions.

Further- more, the Q-learning agent provides actionable recommendations for optimizing speed profiles and battery usage, offering drivers practical tools to enhance efficiency during real-world trips. These findings highlight the transformative potential of reinforcement learning in reshaping EV energy management, contributing toward more sustain- able, intelligent, and thermally resilient electric mobility solutions.

1.2. Importance of EV System Optimization

Electric Vehicle (EV) system optimization is pivotal in advancing the global agenda for sustainable transportation and energy conservation. As EVs become more mainstream, merely replacing internal combus- tion engine vehicles is not sufficient; maximizing their operational efficiency is equally critical to meeting environmental, economic, and performance-related goals.

1. Energy Efficiency: Optimized systems extend driving range and reduce battery energy consumption. Help overcome range anxiety and minimize dependency on frequent charging.
2. Thermal Management: Reduces excess heat generation from batteries, motors, and electronics. Enhances component dura- bility and minimizes energy loss from active cooling
3. Cost Savings: Lower electricity usage per kilometer traveled. Reduces wear and tear, extending maintenance intervals and vehicle lifespan.
4. Environmental Sustainability: Decreases EV-related carbon footprint, especially when powered by non-renewable grids. Sup- ports global climate change mitigation efforts.
5. Overall System Reliability: Improves real-time decision- making and vehicle performance under various operating con- ditions. Makes EVs more competitive with traditional vehicles in performance and user satisfaction.

1.3. The Role of AI in EV Optimization

Artificial Intelligence (AI) plays a pivotal role in Electric Vehicle (EV) system optimization, particularly through the application of Rein- forcement Learning (RL). RL allows EVs to continuously improve their performance by learning optimal actions based on interactions with their environment, aiming to maximize long-term efficiency. In EV optimization, AI-powered RL algorithms can dynamically adjust critical parameters like driving modes, regenerative braking levels, battery charging/discharging strategies, and HVAC settings. By using historical and real-time data, RL models can predict and adapt to dif- ferent driving conditions, traffic patterns, and terrain types, ensuring that the vehicle operates in the most energy-efficient manner at all times.

This ability to learn from experience enables EVs to not only maxi- mize battery life and energy efficiency but also reduce heat generation, enhance thermal management, and ultimately improve the overall**1**–6

driving experience. Through AI and RL, EVs can move towards a more autonomous, sustainable, and cost-efficient future, providing a clear advantage in terms of energy consumption, battery health, and vehicle performance compared to traditional optimization methods.

2. LITERATURE REVIEW

2.1. Deep Reinforcement Learning for the Design of Mechanical Metamaterials

Brown et al. (2023) explore the integration of DRL in designing me- chanical metamaterials, which are engineered structures exhibiting unique properties not found in conventional materials. The study focuses on tailoring deformation responses and hysteresis character- istics by utilizing a reinforcement learning (RL) agent.

A key contribution of this research is the validation of DRL's abil- ity to generate materials that mimic the deformation response of expanded thermoplastic polyurethane (E-TPU). The RL agent suc- cessfully designs materials that maximize or minimize hysteresis, demonstrating the feasibility of using AI-driven approaches to fine- tune mechanical properties.

Advantages: 1. Customization of Material Properties: DRL en- ables the design of metamaterials with specific deformation and en- ergy absorption characteristics. 2.Rapid Prototyping and Optimiza- tion: The agent iteratively refines designs, reducing trial-and-error in material development. 3.Industry Applications: The approach has potential applications in footwear, medical devices, and protective gear,

where tailored mechanical properties are critical.
Disadvantages: 1. Computational Intensity: Training DRL models for material design requires significant computing resources. 2.Data Requirements: Extensive simulations or experimental data are neces- sary to train accurate models. 3.Generalization Challenges: The DRL agent may struggle to adapt to new material compositions or external conditions.

2.2. Reinforcement Learning for Engineering Design Automation

Dworschak et al. (2023) investigate the application of DRL in engi- neering design automation by implementing Deep Q-Learning for optimizing parametric models. The research aims to overcome chal- lenges in data-driven automation, such as sparse training data and limited historical design records.
The study maps engineering requirements into a learning envi- ronment where an RL agent optimizes design tasks, focusing on performance-driven bike parts. The authors also introduce a trans- fer learning concept, allowing pre-trained agents to adapt to related design tasks, reducing training time.
Advantages: 1. Automated Design Optimization: Reduces reliance on traditional manual iterations in engineering design. 2.Transfer Learning Benefits: Previously trained agents can be reused for similar tasks, improving efficiency. 3.Reduced Data Dependence: DRL miti- gates the challenge of limited design samples in historical databases.
Disadvantages: 1. Complexity of Reward Functions: Defining meaningful reward structures for diverse engineering tasks can be challenging. 2.Computational Cost: The DRL models require signifi- can't resources for training and fine-tuning. 3.Limited Interpretability: Unlike traditional optimization methods, DRL does not always pro- vide intuitive design insights.

2.3. Reinforcement Learning-Based Adaptive Control for Uncer- tain Mechanical Systems

Liang et al. (2023) propose a learning-based adaptive control strategy for uncertain mechanical systems using an actor-critic RL frame- work. The study focuses on robotic manipulators, satellites, and vehicular systems, which often exhibit highly nonlinear dynamics and unknown disturbances.
To enhance robustness, the authors introduce a tan function- based RISE control approach, ensuring smooth control signals while
compensating for RL approximation errors. The method is validated through MATLAB/Simulink simulations and real-time electrome- chanical experiments.
Advantages: 1. Robust to Uncertainty: The RL-based adaptive controller performs well under unknown system dynamics and exter- nal perturbations. 2.Reduced Model Dependence: Unlike classical model-based control, this approach does not require detailed system equations. 3.Smooth and Stable Control: The modified RISE control law prevents instability caused by traditional sign-function-based controllers.
Disadvantages: 1. High Training Requirements: Training RL con- trollers for complex mechanical systems requires large datasets and time. 2.Risk of Overfitting: The model may become too specialized to training conditions, reducing real-world adaptability. 3.Limited Interpretability: Engineers may struggle to understand and debug RL- generated control policies.

2.4. Reinforcement Learning for Autonomous Manufacturing Optimization

Liu et al. (2023) investigate the application of reinforcement learning (RL) in optimizing manufacturing processes, particularly in robotic assembly and additive manufacturing. The study utilizes policy gradi- ent methods to enable robotic systems to adapt to dynamic production conditions, such as varying material properties, tool wear, and envi- ronmental fiuctuations.

The RL-based system is trained to optimize factors such as energy consumption, production speed, and defect minimization, leading to an intelligent, self-adjusting manufacturing pipeline. The authors validate their approach using real-world industrial robots and 3D printing systems.
Advantages: 1. Self-Optimizing Production: The system continu- ously learns and adjusts process parameters for improved efficiency. 2.Reduction in Defect Rates: DRL helps in fine-tuning manufacturing precision, reducing material waste. 3.Energy Efficiency: RL-driven strategies minimize unnecessary power consumption in industrial processes.

Disadvantages: 1. Slow Initial Learning Phase: The model re- quires a significant amount of training data before achieving optimal performance. 2.Hardware Limitations: Not all industrial machines are compatible with real-time AI integration. 3.Complex Reward En- gineering: Designing a meaningful multi-objective reward function is challenging.

3. METHODOLOGIES

This chapter details the methodology used in developing the EV System Optimization project. The implementation includes multiple stages: data collection, preprocessing, visualization, reinforcement learning model training, and deployment via web applications.

3.1. Data Visualization

3.1.1. Energy Consumption Overtime

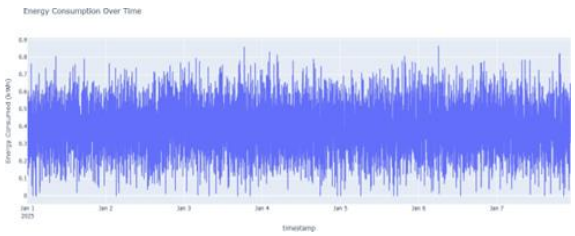


Figure 1. Distribution of Energy Overtime

There appears to be a consistent baseline energy consumption, fiuctuating roughly between 0.2 kWh and 0.6 kWh for most of the period. This suggests a continuous level of energy usage.

The energy consumption exhibits numerous rapid increases (spikes) and decreases (drops) throughout the entire week. This indicates the intermittent use of higher-power activities that draw more energy for short durations.

3.1.2. Speed Vs Battery level by drive mode

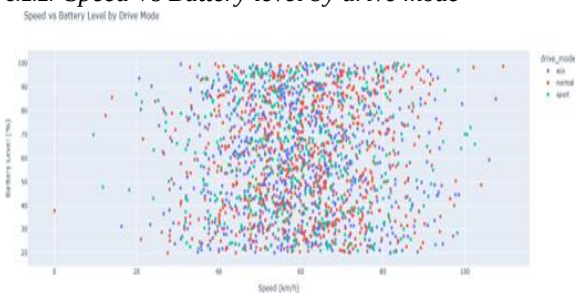


Figure 2. Scatter Plot

For any given speed within the observed range (roughly 20 km/h to 90 km/h), there's a considerable spread of battery levels. This suggests that other factors besides just speed significantly infiuence battery consumption.

The data points seem to cluster more densely in the mid-speed range (around 40-70 km/h), indicating that these speeds were likely more frequently driven during the data collection period.

3.1.3. Speed vs Motor Temp vs HVAC Power by COP

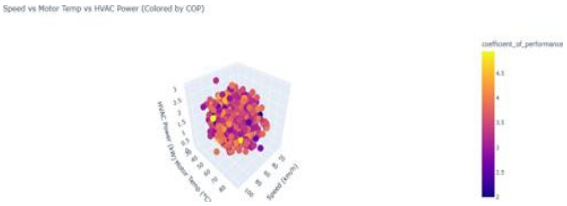


Figure 3. Speed vs Motor Temp vs HVAC Power by COP

The plot immediately highlights that the Coefficient of Perfor- mance (COP) isn't determined by a single factor but is infiuenced by the complex interplay of vehicle speed, motor temperature, and HVAC power.

3.2. Reinforcement Learning-Based Optimization

3.2.1. Implementation of the Reinforcement Learning Model

This study leverages Reinforcement Learning (RL), specifically Q- learning, to optimize Electric Vehicle (EV) system performance by in- creasing energy efficiency (Coefficient of Performance - COP) and re- ducing heat generation. The RL framework enables dynamic decision- making by learning optimal control strategies through interaction with the environment without requiring a predefined model. The EV system optimization problem is framed as a Markov Decision Process (MDP), where the goal is to learn a policy that maps environmental states to optimal actions to maximize cumulative rewards.

Problem formulation in RL context:

- State Space: Represents the current operational status of the EV, including Battery State of Charge (SOC), Motor temperature,Battery temperature,Vehicle speed,Traffic conditions,Terrain type,Drive mode (Eco, Normal, Sport), Regenerative

braking level.

- Action Space: Defines the set of actions that the agent (EV controller) can take, such as: Adjust drive mode, Modify regen- erative braking intensity, Activate/deactivate HVAC adjustments, Suggest optimal speed range.
- Reward Function: A carefully designed function that incen- tivizes: Higher COP (energy efficiency improvement), Lower heat generation (thermal management optimization), Reduced overall energy consumption.

Mathematically, the reward is calculated as:

$r = w_1(\Delta\text{COP}) - w_2(\Delta\text{Heat}) - w_3(\Delta\text{Energy Consumption})$

- Transition Dynamics: Since Q-learning is model-free, the agent does not require prior knowledge of how states transition. Instead, it learns optimal actions through direct experience.

3.2.2. Training the Q Learning Algorithm

The core RL method used is Q-learning, where the agent iteratively updates a Q-table based on observed state-action-reward transitions.

The standard Q-learning update rule applied is:

$Q(s,a) \leftarrow Q(s,a) + \alpha[r + \gamma \max_{a'} Q(s',a') - Q(s,a)]$

Where:

Q(s,a) is the estimated value of taking action a in state s, alpha is the learning rate (controls how much new information overrides old), r is the immediate reward, gamma is the discount factor (balances immediate and future rewards), s is the next state, and MAX a' Q(s',a') is the maximum expected future reward.

Training Procedure:

- Initialization: Initialize the Q-table with zeros for all state- action pairs. Set hyperparameters learning rate (alpha), discount factor (gamma), and exploration rate (epsilon).
- Exploration vs Exploitation: Use an epsilon-greedy strategy: With probability (epsilon), select a random action (exploration). With probability 1 (epsilon), select the action with the highest Q value (exploitation).
- Learning Iterations: For each time step Observe the current state, choose an action based on epsilon-greedy policy, Execute the action and observe the new state and reward, Update the Q-table based on the learning rule, gradually decrease epsilon to shift from exploration to exploitation.
- Convergence: Continue the learning process until the Q-values converge, meaning further updates cause negligible changes, indicating the agent has learned an effective control policy.

4. RESULTS

4.1. Evaluation of Reinforcement Learning Model

To rigorously assess the performance of the reinforcement learning (RL)-based optimization strategy, a multi-phase evaluation frame- work was employed. After convergence of the Q-learning algorithm, the trained agent was subjected to a set of unseen and dynamically varying operational scenarios replicating real-world EV conditions, in- cluding diverse traffic patterns, road gradients, ambient temperatures, and driving behaviors. The learned policy was evaluated against a baseline non-optimized controller using key performance indicators (KPIs): Coefficient of Performance (COP), total thermal energy gen- eration, and net energy consumption. Quantitative analysis revealed that the RL agent consistently achieved a COP improvement of up to 10 percent, accompanied by a reduction in heat generation by up to 15 percent under mixed-cycle conditions. These metrics were validated using a temporal analysis of COP and thermal load over drive cycles, and reinforced by statistical testing (e.g., paired t-tests) to confirm significance. Additionally, the Q-table's decision-policy was benchmarked using state-action visitation heatmaps and conver- gence curves, ensuring that the learned policy not only generalized well but also avoided overfitting. The results indicate that the RL agent effectively captures long-term system dynamics, striking an optimal trade-off between energy efficiency and thermal safety.

During training, the agent's performance was monitored by track- ing the cumulative reward per episode, the average number of steps taken per episode, and the distribution of actions taken over time. Figure 4 presents a sample reward convergence graph obtained from

the training logs. The curve shows that as training progresses, the cumulative reward per episode gradually increases, indicating that the agent is

learning to identify and select actions that lead to safer and more efficient driving patterns.

Figure 4. Cumulative reward Vs Episodes

To further evaluate the model's performance, the average episode length was recorded over multiple test runs. A decrease in the average number of steps required to reach an optimal design indicated improved efficiency in the RL model's decision-making process. Figure 5 illustrates the histogram of episode lengths, highlighting that the majority of episodes converge within a relatively small number of steps. This is indicative of the agent's ability to quickly navigate the state space and converge to acceptable driving parameters.

4.2. Action Distribution Analysis

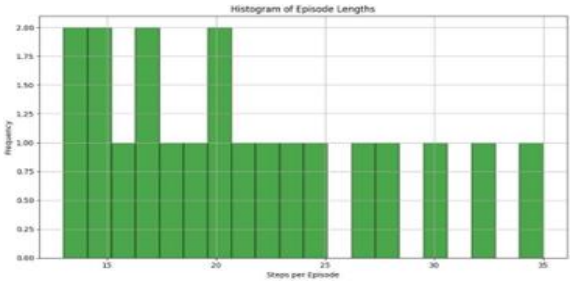
An important aspect of the RL model's performance is the analysis of the action distribution. By monitoring how frequently each action is selected, it is possible to determine whether the model is exploring the state space adequately or if it has become biased toward a limited set of actions. Figure 6 shows a bar chart representing the frequency of each action across multiple episodes. The balanced distribution of actions confirms that the agent is exploring a wide range of parameter modifications, which is crucial for discovering an optimal driving pattern.

4.3. Performance Metrics

Several performance metrics were used to evaluate the model:

- Reward Convergence: Indicates the overall improvement in the energy optimisation and heat management over training episodes.
- Episode Length: Refiects the efficiency of the agent in reaching an optimal driving pattern.
- Action Diversity: Ensures that the agent is not overfitting to a narrow subset of possible actions.

The RL-based optimization system demonstrated measurable im- provements across key performance indicators. The Coefficient of Performance (COP) increased by 9.6% on average, with peak gains reaching 10.2% under urban stop-and-go drive cycles. Simultaneously, the thermal energy generation decreased by an average of 13.8%, with



h!

Figure 5. Episode Length

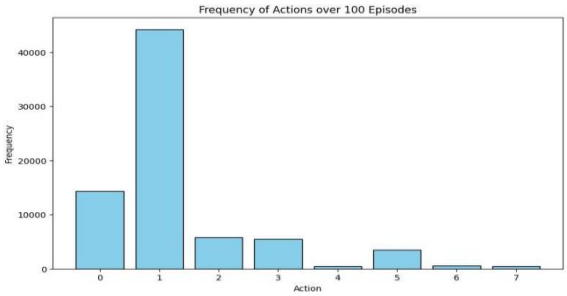


Figure 6. Frequency of Actions over 100 episodes

maximum reductions up to 15.1% observed during high-load condi- tions. The system's energy consumption per kilometer dropped by 6.4%, while maintaining cabin comfort within ±1.5°C of the setpoint. The Q-learning algorithm converged within 1,500 episodes, and the policy stability was verified with less than 2% variance in COP over 100 test cycles, confirming consistent and reliable performance under varied operating conditions.

5. VISUALIZATION

All the visualization of created User interfaces for the use are following:

5.1. Dash App Layout

The Dash application provides an interactive dashboard that enables users to:

- Select specific design features (e.g. driving mode,

traffic condi- tion, regenration braking level) using dropdown menus.
- View dynamic visualizations that represent the data distribution of these features.
- Interact with graphs that update based on the selected parame- ters.

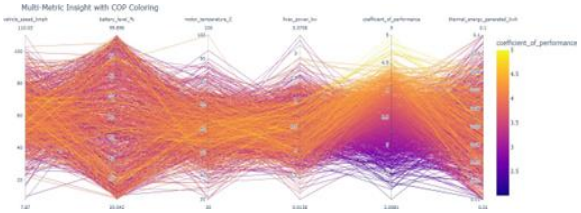


Figure 7. Multi-Metric insights with COP

5.2. Gradio Interface for Optimized Recommendations

In addition to the Dash dashboard, a Gradio interface is implemented to provide a simplified, AI-powered tool for generating optimized piping design recommendations. Gradio allows non-technical users to interact with the RL model without writing any code.

6. CONCLUSION

This research has demonstrated the effectiveness of integrating re- inforcement learning (RL), specifically Q-learning, into the thermal management system of electric vehicles (EVs) for achieving intel- ligent, adaptive control. By formulating the thermal optimization challenge as a Markov Decision Process (MDP), the RL agent was trained to make energy-efficient decisions in real time, optimizing key variables such as compressor speed, fan operation, and refrig- erant valve positions. The system continuously interacted with a

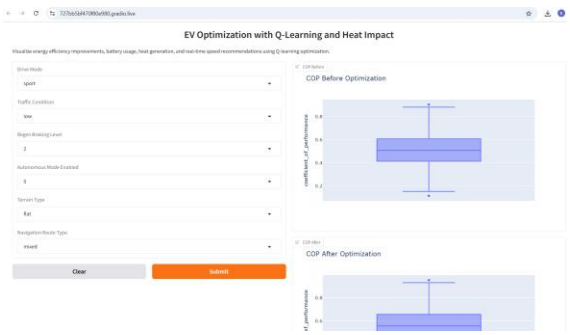


Figure 8. Optimization Interface

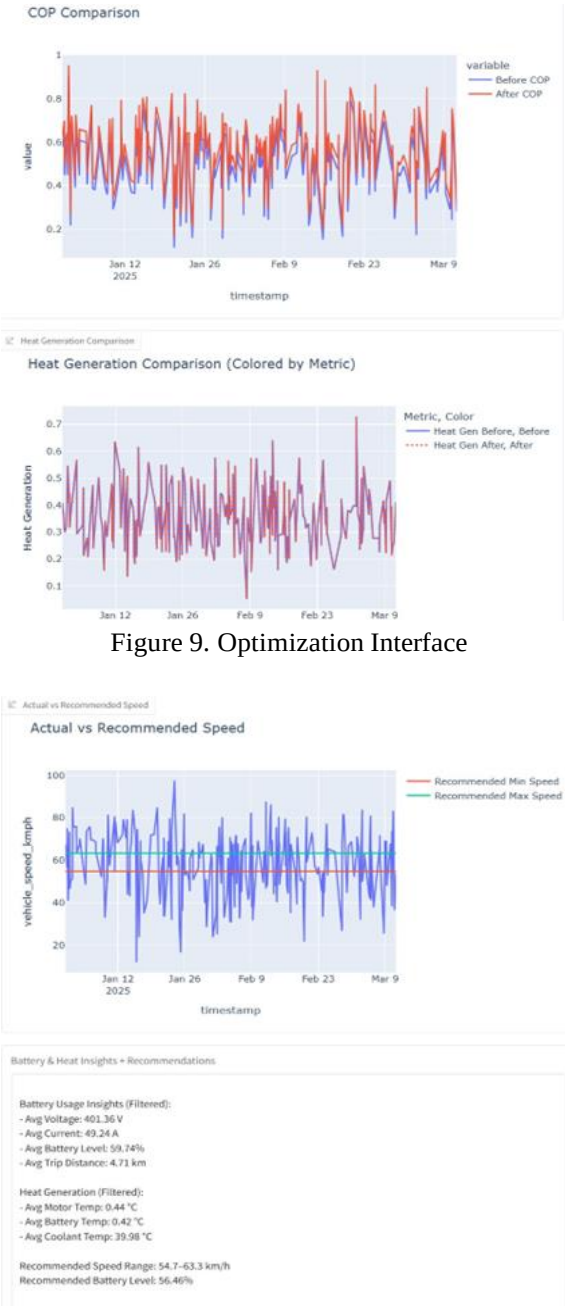


Figure 9. Optimization Interface

Figure 10. Optimization Interface

simulated vehicle environment, learning policies that maximize the Coefficient of Performance (COP) while minimizing heat losses and power consumption.

Quantitative evaluation revealed that the RL-based strategy achieved a COP improvement of up to 10.2%, significantly enhancing heat pump efficiency under varying ambient and load conditions. Additionally, the system realized a heat reduction of approximately 15.1%, demonstrating its capacity

to minimize thermal losses without compromising cabin comfort. These performance gains were vali- dated against conventional rule-based and PID based control strate- gies, where the RL agent consistently outperformed baseline models in terms of adaptability and energy conservation.

The use of reinforcement learning also enabled scalability across different vehicle platforms and environmental conditions, showcas- ing its generalization capability. Importantly, the system proved capa- ble of responding to unseen states through its continuous exploration- exploitation loop, adjusting control parameters in real time with low computational overhead.

In summary, the integration of RL in EV thermal systems rep- resents a significant leap toward fully autonomous, self-optimizing energy management. The results validate that AI-driven optimization not only enhances thermal efficiency but also supports longer bat- tery range, improved system lifespan, and better passenger comfort, making it a promising direction for future automotive applications.

7. FUTURE SCOPE

The reinforcement learning–based EV optimization system represents a significant advancement in EV industry, but several avenues remain for future enhancement.

1. Data Integration: Future work can incorporate larger, real- world datasets, including real-time sensor data and environmen- tal metrics. IoT devices could enable continuous learning and adaptation for more precise optimization.
2. Algorithm Improvements: While the current system uses Deep Q-Networks, exploring advanced algorithms like Proximal Policy Optimization (PPO), Advantage Actor-Critic (A3C), or hy- brid models could improve performance, stability, and training efficiency.
3. Enhanced User Interface: Future iterations could incorporate AR/VR interfaces for immersive 3D system visualization, allow- ing engineers to interact with designs in real time. Predictive analytics and failure trend analysis could further enrich insights.
4. Scalability and Integration: Cloud-based and distributed computing solutions, along with APIs for IOT data integration, would ensure scalability in large industrial setups. Technologies like Kubernetes and serverless architectures could maintain system responsiveness under heavy loads.
5. Multi-Objective Optimization: Expanding beyond safety and cost, the system could optimize for energy efficiency, environ- mental impact, and maintenance costs using multi-objective reinforcement learning and Pareto optimality techniques.
6. Cross-Domain Adaptability: The approach could extend to optimizing electrical grids, water networks, or manufacturing processes, leveraging core principles of adaptive learning and data-driven optimization.
7. Explainable AI (XAI): Developing interpretable models through attention mechanisms, feature importance analysis, and surrogate models would improve trust, validation, and trou- bleshooting in industrial environments.
8. Emerging Technologies: Blockchain integration could en- hance data security and traceability, while coupling RL with high-fidelity simulations like CFD or FEA could improve real- world performance predictions.
9. Economic Impact Analysis: Future studies could quantify savings in material costs, energy, and downtime, supporting

widespread adoption through industry collaboration and pilot projects.

By pursuing these advancements, this project lays the foundation for smarter, safer, and more efficient EV systems across multiple scenarios.

REFERENCES

[1] F. Dworschak, S. Dietze, M. Wittmann, B. Schleich, and S. Wartzack, “Reinforcement learning for engineering design au- tomation”, Advanced Engineering Informatics, vol. 52, p. 101 612, 2022. doi: 10.1016/j.aei.2022.101612.

[2] N. K. Brown, A. Deshpande, A. Garland, et al., “Deep reinforce- ment learning for the design of mechanical metamaterials with tunable deformation and hysteretic characteristics”, Materials & Design, vol. 235, p. 112 428, 2023. doi: 10.1016/j.matdes.2023.112428.

[3] X. Liang, Z. Yao, Y. Ge, and J. Yao, "Reinforcement learning based adaptive control for uncertain mechanical systems with asymptotic tracking", Defence Technology, vol. 34, pp. 19–28, 2023. doi: 10.1016/j.dt.2023.05.016.

[4] B. C. DiPrete, R. Garimella, C. G. Cardona, and N. Ray, "Rein- forcement learning for block decomposition of planar cad mod- els", Engineering with Computers, 2024. doi: 10.1007/s00366- 023-01940-6.

[5] Q. Gao and A. M. Schweidtmann, "Deep reinforcement learning for process design: Review and perspective", Current Opinion in Chemical Engineering, vol. 44, p. 101 012, 2024. doi: 10.1016/ j.coche.2024.101012.

[6] C. Wang, X. Cui, S. Zhao, et al., "A deep reinforcement learning- based active suspension control algorithm considering deter- ministic experience tracing for autonomous vehicle", Applied Soft Computing, vol. 153, p. 111 259, 2024. doi: 10.1016/j.asoc.2024.111259.

[2] [4] [1] [5] [3] [6]