International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(6s) 2026 pp. 983-1018
https://doi.org/10.70917/ijcisim-2026-3000
Article

AI-Driven Energy Management Architecture for Electric Vehicles: Anomaly Detection, Battery Optimization, and Stakeholder-Centric Analytics

Shishir Sharad Pande[1], U. B. Malkhandale[2], Ritam Dutta[3], Debmitra Das[4], Madhav Sharma[5]

[1]Dept. of Electrical Engineering, Priyadarshini College of Engineering, Nagpur, India
Email: spande1992@gmail.com
[2]Dept. of Electrical Engineering, Priyadarshini College of Engineering, Nagpur, India
Email: ub.malkhandale@gmail.com
[3]Dept. of CSE, Poornima Institute of Engineering and Technology, Jaipur, Rajasthan, India
Email: ritamdutta1986@gmail.com
[4]Dept. of CSE, Poornima University, Jaipur, Rajasthan, India
Email: debmitradas123@gmail.com
[5]Dept. of CSE, Poornima Institute of Engineering and Technology, Jaipur, Rajasthan, India
Email: madhavsharma36@yahoo.co.in

Abstract: EVs need smart energy management systems to be efficient, have long battery life, and be able to operate safely in uncertain real-world scenarios. Conventional rule-based and optimization methods find it difficult to accommodate nonlinear battery models, thermal limits, stochastic driving models and unforeseen anomalies like sensor drift, actuator failures and cyber-distorted measurements. In order to overcome these shortcomings, this paper presents a single AI-based anomaly-sensitive Safe Reinforcement Learning (Safe-RL) energy management framework of EVs. The structure is an anomaly detector based on Temporal Convolutional Network (TCN), a constrained Safe-RL control, a safety projection layer, and a fallback policy in the event of faults, all with the assistance of comprehensive electro-thermal battery modeling and degradation-constrained optimization.

The sensor/actuator faults are detected with an accuracy of 96.8%, ROC-AUC of 0.985, and a detection latency of less than a second. Safe-RL controller maximizes charging/discharging power and strictly adheres to SOC, thermal and power limits using a rapid quadratic-programming safety layer. High-fidelity EV simulation environment with domain randomization and 500+ injected anomaly scenarios were used to evaluate the system. Findings indicate that the suggested approach will save up to 14.1% of energy, 16.6% of the daily charging cost, and 31.7% of long-term battery degradation with zero safety breaches. The system can be seen to be highly resilient with cost increase of only 3.6% under faults, which is significantly lower than MPC and standard RL baselines. On the whole, the suggested architecture is a strong, secure, and very effective solution to the next-generation EV energy management.

Keywords: NA

1. INTRODUCTION

Electric Vehicles (EVs) are quickly changing the way transportation is done today because of their environmental benefits, lower operation expenses, and their compatibility with renewable energy sources. Nevertheless, the intense use of EVs poses new issues of energy management, battery safety, anomaly detection and grid integration. The efficient energy management systems (EMS) should be able to optimize the charge-discharge dynamics of batteries, reduce degradation, be safe, and adapt to the uncertain driving patterns. Traditional rule-of-thumb methods are frequently ineffective in modeling nonlinear battery behavior, stochastic load profiles and increasing complexity of EV ecosystems.

Recent artificial intelligence (AI) innovations have enhanced considerably prediction, optimization, and decision-making of EV energy systems. Machine learning has been effectively used to recommend users of EVs and forecast their loads [1], detect malicious EVs in a charging network [2], model-based reinforcement learning (MBRL) to optimize the energy management of a hybrid/fuel cell vehicle [3], [4], and generate control-maps of hybrid architectures [5]. Moreover, AI-based classification, optimization and sensor-controlled learning have shown a high potential in fuel cell hybrid EMS [6], electric motor design [7], and strong SoC/SOH estimation [12].

A number of review works highlight the fact that the development of EMS is moving towards AI-based, optimization-focused, and data-driven control approaches [8], [9], [13]. Nevertheless, the available literature is usually limited to one of the dimensions: energy optimization enhancement [3], [4], [5], anomaly detection enhancement, component design streamlining, or EV communication/charging infrastructure security enhancement [10]. The number of works that strive to unite health-conscious control, anomaly detection, real-time safety assurances, and RL-based decision-making under one architecture is very low.

Also, the existing EMS studies are limited by one or more of the following problems:

- Poor real-time resilience: MBRL and ML-based EMS research is mostly based on simulation with no fault modeling [3], [4], [6].
- Absence of built-in anomaly management: Works on fault/anomaly detection [2], [10] are uncommon to interface with powertrain-level energy management.
- The safety limits are not always considered: RL strategies are usually efficient, but not necessarily safe (SOC/thermal limits).
- None stakeholder-focused analytics: Current EMS strategies lack utilities, regulators and consumer dashboards.
- Lacking holistic battery-health perspective: There is only isolated work on SOH/SOC estimation [12], but not on its application in real-time in energy management decisions.

Novelty of the Proposed Work

This paper suggests the initial unified, AI-based EV energy management architecture that combines:

1. Real-time anomaly detection: A monitored TCN-based sensor/actuator fault, cyber anomaly and abnormal energy behaviour detector is identified.
2. Safe Reinforcement Learning (Safe-RL): A constrained RL controller optimizes the energy efficiency and battery degradation and imposes severe SOC, thermal, and power safety constraints.
3. Fallback Control + Safety Layer: A rule-based fallback and a quadratic programming (QP) projection make sure that the operation is zero-violent even in the case of anomalies that are not anticipated.
4. Smart battery health-aware decision-making: The RL reward design directly incorporates battery aging models and thermal dynamics.
5. Stakeholder-centric analytics: The architecture has a real-time utility, regulator, and consumer dashboard, which was lacking in previous EMS literature.
6. Strong training based on domain randomization and anomaly injection: Assures generalization of actual driving patterns, temperature differences and simulated faults.

Combined, these contributions will take the state of the art to a new level beyond isolated anomaly detection or standalone RL EMS strategies, and create a complete, safe and intelligent EV energy management ecosystem.

2. Literature Review

Overview of AI-Driven EV and EMS Research

AI methods have been extensively applied into improving EV functions, including charging recommendation, anomaly detection, energy management, and battery health estimation. Initial studies investigated the problem of data-driven EV recommendation and user behavior modeling based on hybrid ML-optimization frameworks [1]. GRUs have also been applied in charging networks to identify malicious or lying EVs that provide inaccurate SOC values to control charging queues [2]. All these contributions emphasize the usefulness of AI in prediction and anomaly detection but are silent on the topic of energy control and battery protection.

The optimization of hybrid and fuel cell EV energy has been heavily studied in reinforcement learning, and model-based RL has attained near-optimal solutions on par with dynamic programming [3], [4]. Likewise, generation of control maps through machine learning has been investigated with hybrid cars [5], and classifier ensembles have been used with fuel cell hybrid EMS [6]. These approaches prove to be energy saving but do not explicitly tackle anomalies, real-time safety limits, and battery degradation modeling.

Optimization of AI-based EV components has been applied to other aspects of car design, including the design of motor geometry [7], and the future of ML in cars is a wide field. Extensive surveys on EMS approaches [8], on nonlinear optimization of wireless energy transfer [9], and on AI-assisted EV routing [13] all show a growing trend of AI, RL, and optimization in EV systems. They, however, unanimously observe that there are no coherent, safety conscious control architectures that can manage uncertainty, faults and dynamically interacting grids.

EV charging infrastructure has also been the focus of cyber security research, where Tiny ML-based edge device anomaly detection can reduce latency and memory usage [10] without being tied to either battery or EMS logic. In the same way, the rule-vs-ML EMS as a method of hybrid energy storage systems [11], battery SOC/SOH estimation frameworks [12], and multi-objective optimization of EV charging infrastructure [14] also indicate the increased significance of multi-criteria decision-making. The use of AI in energy systems is also strengthened by the ability of renewable energy to be predicted through hybrid ML decomposition methods [15].

Irrespective of these developments, none of the available works combines (i) anomaly detection, (ii) safe RL-based EMS, (iii) battery degradation modelling, and (iv) stakeholder visualization into a single, real-time deployable architecture. This is the gap that drives the presented anomaly-aware safe RL energy management system.

Table 1 provides a comprehensive comparative review of existing AI, ML, and optimization-based approaches used in EV and HEV energy management systems.

Table 1. Literature review comparison

Sr.	**Year**	**Title**	**Technique / Concept**	**Evaluation Metrics**	**Dataset / Platform**	**Key Claims**	**Our Interpretation**
1	2023	EV Recommendation via RBM + WWPA [1]	RBM regression + Water Wheel Plant optimization	RMSE, MAE, R^2, ROC	3395 charging sessions	Improved accuracy vs baseline ML models	Strong on prediction tasks
2	2020	Lying EV Detection via GRU + NSGA-II [2]	GRU classifier + oversampling	Accuracy, AUC, DR−FA	Synthesized + real traces	High accuracy for dishonest EV detection	Good time-series modeling
3	2020	Online MBRL for HEV EMS [3]	Model-based RL	Fuel use vs DP	Quasi-static simulation	Near-optimality vs DP	Good adaptive EMS
4	2021	MBRL for Fuel Cell EVs [4]	RL + data-driven model update	Hydrogen use	Simulation	5.7% fuel saving	Strong validation
5	2022	ML-Based Control Maps for HEV [5]	ML classifiers trained on DP	Fuel consumption	EPA cycles	Slight efficiency gains	Practical controller
6	2023	Classifier Fusion EMS for FCHEV [6]	SVM+KNN+NB fusion	Accuracy, RMSE	Simulation	98% EMS accuracy	Consistent improvement
7	2023	Motor Geometry ML Framework [7]	ANN + Image processing	Iterations, accuracy	FEM + MATLAB	90% faster optimization	Useful design aid
8	2024	HEV EMS Review [8]	Comparative survey	Efficiency, emissions	IEEE/Scopus	Highlights AI superiority	Comprehensive review

Sr.	Year	Title	Technique / Concept	Evaluation Metrics	Dataset / Platform	Key Claims	Our Interpretation
9	2024	SWIPT Optimization Review [9]	Nonlinear optimization	EE, SR	Literature review	Shows benefits of nonlinear models	Good synthesis
10	2024	Tiny ML for EV Charging Security [10]	Tiny ML (TFLite) IDS	Latency, memory	CICIDS2017 + ESP32	Low-latency detection	Strong edge utility
11	2024	ML vs Rule-based EMS for HESS [11]	ML regression from GA-optimal	SOC balance, battery life	Real aging data	Better lifetime & balance	Strong EMS enhancement
12	2024	Hybrid GRU-LSTM for SOC/SOH [12]	Hybrid DNN	RMSE, MAE	NASA dataset	SOC error ~2%	Strong estimation
13	2025	Survey on AI-based EV Routing [13]	GNNs, DRL, hybrid AI	Accuracy, compute	SUMO + literature	AI outperforms classical	Broad insights
14	2025	PV-EV MCDM Optimization [14]	MOPSO + TOPSIS/AHP	LCOE, reliability	MATLAB case study	Tunable optimal design	Multi-perspective
15	2025	SHD-LSTM-CSO for Forecasting [15]	Hybrid decomposition + LSTM	RMSE, MAPE	Real wind datasets	Better forecasting accuracy	Strong hybrid model

Critical Analysis of Existing Literature

Critically, the reviewed works may be divided into four general groups:

1. Prediction and recommendation (user behavior, load, routing) - In these studies, AI is capable of predicting loads with great accuracy, designing components, or creating routes, e.g., [1], [7], [13], [15]. Nevertheless, they are usually open-loop: they do not complete the loop with real-time EMS decisions or take into account battery health and safety constraints.
2. Anomaly/fault detection and cybersecurity Deep learning and Tiny ML-based detectors [2], [10] have high accuracy and low latency in detecting charging coordination attacks or intrusion events. However, these detectors are not logically coupled to powertrain EMS; they are not directly modulating charge/discharge decisions and dealing with the consequences of mis-detections on battery safety and degradation.
3. RL-based EMS strategies - MBRL and ML EMS applies to HEVs, FCEVs, and HESS [3]-[6], [11] is convincing that learning-based controllers can compete with DP and rule-based baselines in the fuel economy, SOC stability, and lifetime metrics. However, they are usually:
   - only tested in simulation using idealized sensor information,
   - not fault-tolerant (no explicit models of faulty measurements or cyber-attacks).
   - tend to regard safety and battery degradation as soft punishments instead of hard limits.
4. Battery health prediction and infrastructure optimization - SOC/SOH prediction models [12] and multi-objective planning of PV-EV charging [14] are powerful means of monitoring and offline planning, though they are silent on how these health and infrastructure insights are integrated into real-time closed-loop EMS control.

Importantly, the majority of works consider anomaly detection, energy optimization, battery health, and safety as distinct issues that can be addressed with the help of various models without a combined control logic. Also, the current RL-based EMS strategies either do not enforce the constraints or approximate it by using simple penalties that may fail in the presence of extreme anomalies or distributional changes.

Research Gaps and Location of the Proposed Work.

According to the analysis provided above and Table 1, the key research gaps may be summarized as follows:

- Absence of integrated architectures: There is no work that combines (i) online anomaly detection, (ii) safe RL-based EMS, (iii) health-aware control, and (iv) stakeholder-level analytics in a single, deployable architecture.
- Minor attention to anomalies in EMS design: Although the detection of anomalies is studied in the context of charging coordination and cybersecurity [2], [10], they are not applied to powertrain EMS, implying that the controller itself is unaware of sensor failures, replay attacks, or actuator failures.
- Weak safety guarantees in RL controllers: The vast majority of RL-based EMS strategies optimize energy efficiency, but do not impose hard SOC, power or thermal limits, but rather use safety as an extra penalty term, which is not reliable in safety-critical applications.
- Battery degradation not directly implemented in control decisions: Although SOC/SOH estimators may be available [12], they are usually only used as monitors, and not as a part of the reward design and constraint structure of the EMS controller.
- Inadequate stakeholder-oriented analytics and explain ability: Reality Current EMS structures seldom reveal interpretable dashboards to utilities, consumers, and regulators that reveal SOC trajectories, fault events, thermal margins, and health indicators in real time.

In this landscape, the proposed work is placed as a single, anomaly-conscious, safety-constrained, and health-conscious EMS architecture that bridges these gaps by closely incorporating detection, control, and analytics into a single system.

Planned Contribution Direction

The proposed paper will follow the following directions of planned contribution to fill in the identified research gaps:

1. Integrated Anomaly-Aware Safe-RL EMS: We propose a TCN-based anomaly detector, which works on streams of EV sensors and is directly incorporated into a Safe-RL EMS controller, enabling the energy management policy to modify its actions depending on the detected faults and confidence levels.
2. Safety-Layer and Fallback Policy of Hard Constraint Enforcement: We propose a quadratic-programming (QP) safety projection layer and a rule-based fallback policy that jointly enforce strict compliance to SOC, thermal and power constraints, even in the presence of uncertainty in the RL policy and corrupted sensor data.
3. Battery-Degradation-Aware Optimization: The models of battery aging and thermal effects are directly added to the multi-objective cost function to motivate the RL controller to learn health-aware power curves that minimize high C-rate events, temperature spikes, and depth-of-discharge stress.
4. Sturdy Training through Domain Randomization and Anomaly Injection: The EMS is trained in a high-fidelity simulation setting, whose battery parameters, ambient conditions, and injected anomalies are randomized, enhancing its generalization to real-world conditions and fault conditions.
5. Stakeholder-Centric Analytics and Dashboard: A dashboard layer will be suggested to make SOC, temperature margins, anomaly events, energy and cost metrics, and battery health indicators visible to stakeholders (utilities, consumers, regulators) to enhance transparency and support operational decision-making.

With this direction of contribution, the proposed system will specifically be planned to be at the intersection of anomaly detection, safe RL, battery health management, and stakeholder analytics, thus advancing the state of the art beyond the separate methods that exist in the literature.

Problem Formulation

Following are the various variables and notations used in this section:

Indices & sets

- $k \in \{0,1,\dots,T-1\}$— discrete time index (sampling period Δt).
- H — planning horizon (length T).
- S — set of stakeholders (utilities, consumers, regulators).

State, control, observations

- x_k — system state vector at time k.
- u_k — control/action vector at time k.
- y_k — observation vector (sensor measurements) at time k.
- z_k — exogenous inputs (driving demand, ambient temperature, grid price etc.) at time k.

Battery-specific

- $\mathrm{SOC}_k \in [0,1]$— battery state-of-charge at time k.
- Q_k — available battery capacity (Ah) at time k (declines with ageing).
- $P_{b,k}$ — battery terminal power (positive = discharge to load, negative = charging) at time k [W].
- $P_{\mathrm{load},k}$ — vehicle power demand (traction + auxiliaries) at time k [W].
- $P_{\mathrm{chg},k}$ — charging power (grid → battery) at time k [W] (part of $P_{b,k}$ when negative).
- $\eta_{\mathrm{batt}}(\cdot)$— battery charge/discharge efficiency (may be SOC/T-dependent).
- C_{nom} — nominal battery capacity [Wh] or [J].
- T_k — battery temperature at time k [°C].
- ΔQ_k — incremental capacity loss (aging) over the interval $[k, k+1]$.

Electrical components & losses

- η_{inv} — inverter/drivetrain efficiency (can be function of P).
- $P_{\mathrm{regen},k}$ — regenerative braking power recovered at time k.
- $P_{\mathrm{grid},k}$ — power exchanged with grid at time k (positive = export, negative = import).

Anomaly/fault modeling

- a_k — anomaly/fault signal added to measurements or dynamics at time k.
- H_0 — null hypothesis (no anomaly).
- H_1 — alternative hypothesis (anomaly present).
- $\hat{a}_k$ — detected / estimated anomaly at time k.

AI / decision models

- π_θ — parameterized policy (AI controller) with parameters θ. Maps observations to actions: $u_k = \pi_\theta(y_{0:k}, z_{0:k})$
- ϕ_ψ — anomaly detector parametrized by ψ. Outputs anomaly score or label: $s_k = \phi_\psi(y_{0:k})$.
- $\hat{x}_k$ — estimator (e.g., state estimator / Kalman / neural net) of true state.

Performance & cost

- $c_{E,k}$ — energy consumption cost at time k.
- $c_{D,k}$ — battery degradation cost at time k.
- $c_{S,k}$ — stakeholder utility cost (e.g., grid impact, user comfort).
- $c_{A,k}$ — anomaly detection loss (false negative/positive penalties) at time k.
- C_{total} — cumulative cost to minimize.

- w_E, w_D, w_S, w_A — positive scalar weights for multi-objective aggregation.

Constraints

- $\mathrm{SOC}_{\min},\ \mathrm{SOC}_{\max}$ — safety SOC bounds.
- $P_b^{\min},\ P_b^{\max}$ — battery power limits.
- $T_{\max}$ — max safe battery temperature.
- α — allowable false alarm rate (regulator requirement).
- β — allowable miss-detection rate.

System dynamics and measurement model

Modeling the EV and battery We describe the EV and battery using a discrete-time dynamical model. The state x_k will consist of at least SOC_k, battery temperature T_k, and internal state required by estimators:

$$x_k = \begin{bmatrix} \mathrm{SOC}_k \\ T_k \\ \vdots \end{bmatrix} \tag{1}$$

Energy balance in the battery (discrete time):

The SOC dynamics of battery are determined by battery power $P_{b,k}$ following inverter (and charger) efficiencies. In units of energy (Wh) which are equal to Δt in hours:

$$\mathrm{SOC}_{k+1} = \mathrm{SOC}_k - \frac{\Delta t}{C_{\mathrm{nom}}} \cdot \frac{P_{b,k}}{\eta_{\mathrm{batt}}(\mathrm{SOC}_k, T_k, P_{b,k})} \left(\text{for discharge; signs handled in } P_{b,k}\right) \tag{2}$$

A more explicit one with charge/discharge efficiencies:

$$\mathrm{SOC}_{k+1} = \mathrm{SOC}_k - \frac{\Delta t}{C_{\mathrm{nom}}} \cdot \begin{cases} \frac{P_{b,k}}{\eta_{\mathrm{dis}}(\cdot)}, & P_{b,k} > 0 \text{ (discharge)} \\ P_{b,k} \cdot \eta_{\mathrm{chg}}(\cdot), & P_{b,k} < 0 \text{ (charge)} \end{cases} \tag{3}$$

Power balance (vehicle-level):

At every time step the battery should be able to provide vehicle demand less than regeneration and/or exchange with the grid:

$$P_{b,k} = \eta_{\mathrm{inv}}^{-1} P_{\mathrm{load},k} - P_{\mathrm{regen},k} - P_{\mathrm{grid},k} \tag{4}$$

or reorganized according to sign conventions. We will assume that $P_{b,k}$ is net battery power after inverter losses.

Battery thermal dynamics:

$$T_{k+1} = T_k + \Delta t \cdot f_T\left(T_k, P_{b,k}, T_{\mathrm{amb},k}\right) \tag{5}$$

where f_T is being heated by I^2R losses and cooled.

Capacity fade (aging) model (discrete increment)

We assume incremental capacity loss ΔQ_k to be a cyclic stress dependent function:

$$\Delta Q_k = g\left(\mathrm{SOC}_k, \Delta\mathrm{SOC}_k, T_k, |\ P_{b,k}\ |\right) \tag{6}$$

and

$$Q_{k+1} = Q_k - \Delta Q_k \tag{7}$$

An empirical representation that is often used to describe calendar and cycling ageing (compact form):

$$\Delta Q_k \approx \gamma_c(\mathrm{SOC}_k, T_k)\Delta t + \gamma_{\mathrm{cyc}}\left(\mathrm{DoD}_k, |\ P_{b,k}\ |, T_k\right) \tag{8}$$

with γ_c calendar fade rate, and γ_{cyc} cycle-related fade (a function of depth-of-discharge DoD). To optimize it we may take a differentiable surrogate $g(\cdot)$

Anomaly-possibility measurement model.

Outputs y_k are measured using sensors that contain noisy measurements of SOC (voltage-based), current, temperature, etc.:

$$y_k = h(x_k, u_k, z_k) + n_k + a_k \quad (9)$$

- $h(\cdot)$ — nominal measurement function,
- n_k — zero-mean sensor noise,
- a_k — anomaly/fault injection (zero if no anomaly).

When $a_k \neq 0$, we are in hypothesis H_1.

Formulation of the problem of anomaly detection

We consider anomaly detection to be a supervised (or semi-supervised) inference problem that outputs a score s_k and label $\hat{\ell}_k \in \{0,1\}$:

$$s_k = \phi_\psi(y_{0:k}, u_{0:k}, z_{0:k}), \hat{\ell}_k = I\{s_k > \tau\} \quad (10)$$

where τ is threshold.

In another case, a residual-based detector employs an estimator $\hat{x}_k$ and residual $r_k = y_k - h(\hat{x}_k)$. Then

$$s_k = \| W_r r_k \|_p \quad (11)$$

where a threshold is selected to meet false alarm requirements.

Performance metrics:

- True positive rate (TPR), False positive rate (FPR), Precision, Recall, F1.
- False negative cost (miss) and false alarm penalty may be combined in detection cost $c_{A,k}$

The constraints can have FPR $\leq \alpha$ and FNR $\leq \beta$.

Objective: multi-objective cost

We sum up the costs that are important in each time step:

- Energy cost (operational energy or financial cost)

$$c_{E,k} = P_{b,k} \cdot \pi_k \cdot \Delta t \quad (12)$$

where π_k is grid price (₹/Wh) or monetary cost; when optimizing energy consumption rather than money, π_k=1.

- Battery degradation cost

Battery capacity loss, place on financial basis:

$$c_{D,k} = \kappa_D \cdot \Delta Q_k \quad (13)$$

where κ_D (₹/Ah or ₹/Wh) transforms loss of capacity into cost.

- Stakeholder cost / utility

This adds inconvenience to the user, grid effect, and regulatory fines:

$$c_{S,k} = f_S\big(\text{SOC}_k, P_{b,k}, \text{charging\_schedule}_k, \text{emissions}_k\big) \quad (14)$$

Examples: fines on breach of user-specified minimum SOC at departure, or grid-demand limit.

- Anomaly detection cost

  $c_{A,k} = \kappa_{\mathrm{FN}} \cdot I(\text{missed anomaly at } k) + \kappa_{\mathrm{FP}} \cdot I(\text{false alarm at } k)$ (15)
- Aggregate per-step cost:

  $c_k = w_E c_{E,k} + w_D c_{D,k} + w_S c_{S,k} + w_A c_{A,k}$ (16)
- Accumulating expected cost over period:

  $J(\theta, \psi) = E[\sum_{k=0}^{T-1} c_k]$ (17)

anticipation of stochastic factors (driving profile, prices, noise, occurrences of anomalies).

Constraints (hard and soft)

Hard constraints to observe safety and operability:

- SOC bounds:

  $\mathrm{SOC}_{\mathrm{min}} \leq \mathrm{SOC}_k \leq \mathrm{SOC}_{\mathrm{max}}, \forall k$ (18)
- Power limits:

  $P_b^{\mathrm{min}} \leq P_{b,k} \leq P_b^{\mathrm{max}}, \forall k$ (19)
- Thermal limit:

  $T_k \leq T_{\mathrm{max}}, \forall k$ (20)
- Charging infrastructure / grid constraints (if vehicle-to-grid):

  $| P_{\mathrm{grid},k} | \leq P_{\mathrm{grid}}^{\mathrm{max}}$,and $\sum_{v \in V} P_{\mathrm{grid},k}^{(v)} \leq P_{\mathrm{grid,feeder},k}^{\mathrm{max}}$ (21)
- Detection performance constraints:

  $\mathrm{FPR}(\phi_\psi) \leq \alpha$,$\mathrm{FNR}(\phi_\psi) \leq \beta$ (22)

These can be imposed (more or less) through fines or by selecting suitable limits.

Decision variables and decomposition

We seek to design / learn:

- π_θ: policy mapping of actions to regulate actions in the u_k (e.g., charge/discharge schedule, power split, regenerative braking allocation).
- ϕ_ψ: anomaly detector maps observations to anomaly score/label.
- Optional an estimator $\hat{x}_k$ with parameters ξ.

Actions typically include:

$u_k = [P_{b,k}, \quad P_{\mathrm{grid},k}, \quad \mathrm{HVAC}_k, \quad \mathrm{charge\_mode}_k, ...]^\top$ (23)

Casting as a Stochastic Optimal Control / Learning Problem.

The essence of the problem can be formulated as a constrained stochastic optimization / control problem in which the controller π_θ and detector ϕ_ψ are to be optimized/learned to minimize the expected cost, subject to constraints.

Markov decision process (MDP) formulation (to solve by RL)

Define state $s_k = (x_k, z_k)$, action $a_k = u_k$, observations $o_k = y_k$. The dynamics:

$$s_{k+1} \sim p(\cdot \mid s_k, a_k) \tag{24}$$

Reward (negative cost):

$$r_k = -c_k \tag{25}$$

Objective (policy optimization):

$$\pi_\theta^\star = \underset{\theta}{\operatorname{argmax}} E_{\pi_\theta}\left[\sum_{k=0}^{T-1} r_k\right] \tag{26}$$

s.t. safety constraints on trajectories

Safety / constraints solved by constrained MDP (CMDP) or through Lagrangian penalties.

Optimization of controller and detector.

Joint parameters o can be written $\omega = (\theta, \psi)$. The learning objective:

$$\min_{\theta,\psi} J(\theta, \psi) \ = \ E\left[\sum_{k=0}^{T-1} w_E\, c_{E,k} + w_D c_{D,k} + w_S c_{S,k} + w_A c_{A,k}\right] \tag{27}$$

subject to state and detection constraints listed above.

Problem Definition

Given: dynamics $x_{k+1} = f(x_k, u_k, z_k, w_k)$, measurement function $y_k = h(x_k, u_k, z_k) + n_k + a_k$, and stochastic processes for demand and exogenous signals.

Find: a controller π_θ, a detector ϕ_ψ and optionally a state estimator $\hat{x}_k$, minimizing expected cumulative cost while meeting constraints:

$$\begin{aligned}
\min_{\theta,\psi} \quad & J(\theta, \psi) = E\left[\sum_{k=0}^{T-1}\left(w_E c_{E,k} + w_D c_{D,k} + w_S c_{S,k} + w_A c_{A,k}\right)\right] \\
\text{s.t.} \quad & \mathrm{SOC}_{k+1} = \mathrm{SOC}_k - \frac{\Delta t}{C_{\text{nom}}} \cdot \tilde{g}\left(P_{b,k}, \mathrm{SOC}_k, T_k\right),\ \forall k, \\
& T_{k+1} = T_k + \Delta t \cdot f_T\left(T_k, P_{b,k}, T_{\text{amb},k}\right),\ \forall k, \\
& P_{b,k} = \text{power_balance}(u_k, z_k),\ \forall k, \\
& \mathrm{SOC}_{\min} \le \mathrm{SOC}_k \le \mathrm{SOC}_{\max}, P_b^{\min} \le P_{b,k} \le P_b^{\max},\ \forall k, \\
& T_k \le T_{\max},\ \forall k, \\
& \Pr\left(\mathrm{FPR}\left(\phi_\psi\right) \le \alpha\right) \ge 1 - \delta_1, \Pr\left(\mathrm{FNR}\left(\phi_\psi\right) \le \beta\right) \ge 1 - \delta_2, \\
& u_k = \pi_\theta(\cdot), \hat{\ell}_k = \phi_\psi(\cdot), \text{(causality / information constraints)}
\end{aligned} \tag{28}$$

where $\tilde{g}$ represents the efficiency (charge/discharge) which is sign-dependent, and w_k are process noises. δ_1, δ_2 can be probabilistically enforced.

In equivalent form, RL/CMDP:

Find policy π_θ such that:

$$\max_{\theta} E_{\pi_\theta}[\sum_{k=0}^{T-1} r_k] \text{where } r_k = -c_k$$
$$\text{s.t. } E_{\pi_\theta}\left[\sum_{k=0}^{T-1} g_j\,(s_k, a_k)\right] \le d_j,\ \forall j \in J \qquad (29)$$

where g_j are cost functions that encode safety constraints (e.g., the number of SOC violations), and d_j admitable bounds (usually zero in the case of hard constraints).

The interaction between the anomaly detector and controller (coupling):

- The detector ϕ_ψ will give $\hat{\ell}_k$ and/or $\hat{a}_k$; if $\hat{\ell}_k = 1$ (anomaly detected), the controller can activate a safe fallback policy π_{safe} or take countermeasures (isolate faulty sensor, request safe-stop).
- The decisions of the controller affect the stress of the battery and consequently aging ΔQ_k. The performance of detectors affects the cost $c_{A,k}$ and safety.
- It is necessary to optimize jointly since a detector with numerous false positives can drive to conservative control (increasing c_E or c_S), whereas a lax detector will drive risk and possible large $c_{A,k}$ or physical failure costs.

This relationship may be expressed as:

$$u_k = \begin{cases} \pi_\theta(y_{0:k}) & \text{if } \hat{\ell}_k = 0 \\ \pi_{\text{safe}}(y_{0:k}) & \text{if } \hat{\ell}_k = 1 \end{cases} \qquad (30)$$

Relaxation of practice and algorithm

Since the entire joint problem is complicated and nonconvex, a two-stage tractable (mathematically consistent) strategy is:

- Train detector ϕ_ψ (supervised / semi-supervised):

Minimise loss of detection with FPR/FNR budgets:

$$\psi^\star = \operatorname*{argmin}_{\psi} E\left[\ell_{\text{det}}\left(\phi_\psi(y), \ell\right)\right] \qquad (31)$$

such that, $FPR \le \alpha,\ FNR \le \beta$

Select threshold τ that implements a using cross-validation and ROC. Optimize π_θ of controller with RL/CMDP with in-the-loop detector

In case anomalies are uncommon, have a model of detector operation at training (simulate ϕ_ψ behavior).

Optimize constrained policy:

$$\theta^\star = \operatorname*{argmin}_{\theta} J(\theta; \psi^\star) \text{s.t. safety constraints} \qquad (32)$$

Apply Lagrangian relaxation or Constrained Policy Optimization (CPO) / Safe RL algorithms.

With bilevel optimization, joint optimization (simultaneous) can be tried:

$$\min_{\theta} J(\theta, \psi^\star(\theta)), \psi^\star(\theta) = \operatorname*{argmin}_{\psi} L_{\text{det}}(\psi; \theta) \qquad (33)$$

but the bilevel problems are more difficult; they can be solved by alternating updates.

Performance metrics:

- Cumulative energy consumption: $E_{\text{tot}} = \sum_k P_{b,k}\, \Delta t$. (34)
- Total capacity fade: $\Delta Q_{\text{tot}} = \sum_k \Delta\, Q_k$. (35)
- Monetary cost: $C_{\text{mon}} = \sum_k c_{E,k} + c_{D,k} + c_{S,k}$. (36)
- Detection metrics:
  - Precision $= \frac{\text{TP}}{\text{TP+FP}}$, (37)
  - Recall $= \frac{\text{TP}}{\text{TP+FN}}$, (38)
  - F1 score $= 2 \cdot \frac{\text{Precision}\cdot\text{Recall}}{\text{Precision+Recall}}$ (39)
- Safety violations count: $V_{\text{SOC}} = \sum_k I\,(\text{SOC}_k < \text{SOC}_{\text{min}}$ or $\text{SOC}_k > \text{SOC}_{\text{max}})$ (40)

Figure 1 shows the mathematical relationship between the SOC dynamics of the battery, thermal behaviour, aging model, measurement model and anomaly injection process. It plots battery power, temperature, and degradation against time and anomalies against sensor readings or effective system dynamics.

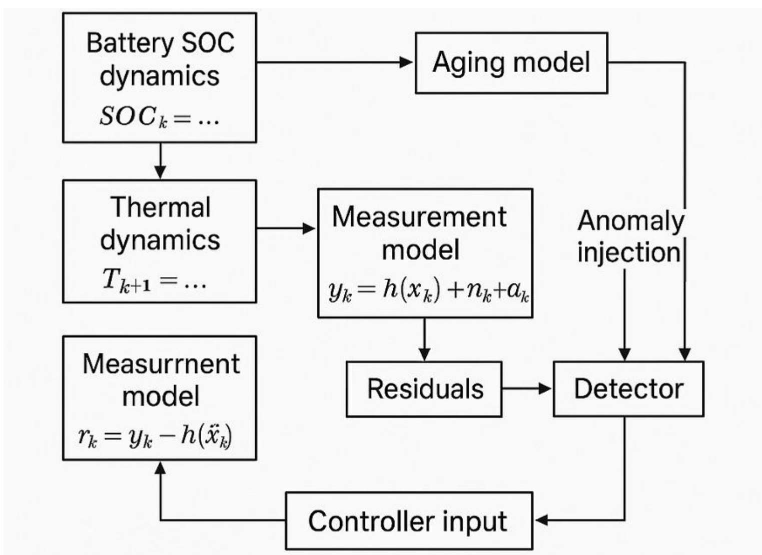

Figure 1. Battery, Thermal and Anomaly Processes Mathematical Model Flow Diagram.

This figure anchors the Problem Formulation section by transforming equations into an intuitive process representation. It clarifies that anomalies have not only an impact on measurements, but also downstream control decisions and battery health. It preconditions the necessity of an integrated detector + controller as opposed to considering each part separately.

Final Problem Statement

Suppose an electric vehicle can be modeled in discrete time $x_{k+1} = f(x_k, u_k, z_k, w_k)$ with measurements $y_k = h(x_k, u_k, z_k) + n_k + a_k$, with a_k representing uncommon anomalies or defects. Develop a parameterized control policy π_θ (AI-driven energy management) and an anomaly detector ϕ_ψ (AI-based monitoring) such that when run in closed-loop form, reduce the cumulative cost of stakeholders weighted by their importance over a planning horizon H:

$$\min_{\theta,\psi} E\left[\sum_{k\in H} w_E\, c_{E,k} + w_D c_{D,k} + w_S c_{S,k} + w_A c_{A,k}\right] \quad (41)$$

subjugated to the battery and vehicle operational dynamics, safety limits (SOC, power and thermal limits), grid exchange limits, and detection performance limits (false alarm and miss-detection rates). The controller needs to act causally upon the available observations and detector outputs; the detector needs to attain regulatory detection performance ($FPR \leq \alpha,\ FNR \leq \beta$) with minimal disturbance to energy management goals.

Objective interpretation. The resulting AI-based architecture must trade off operational energy, increase battery life through minimum degradation-inducing control actions, offer actionable, low-latency fault detection to provide safety/regulatory requirements, and meet stakeholder requirements (utility grid impacts, consumer convenience, safety/performance thresholds as defined by regulators).

Proposed Method

This part translates the modular architecture into concrete algorithm modules, equations, loss functions, and run-time behavior.

Notations used

- k: discrete time index, Δt sampling period.
- $x_k = [SOC_k, T_k, R_{int,k}, \ldots]^\top$ state vector.
- $u_k = [P_{b,k}, P_{\text{grid},k}, \text{mode}_k]^\top$ action vector.
- y_k: raw measurements (voltage V, current I, cell/pack temp T, inverter status, GPS speed).
- z_k: exogenous inputs (ambient T_{amb}, price π_k, user schedule).
- a_k: anomaly injection vector (added to measurements or dynamics).
- ϕ_ψ: anomaly detector (parameters ψ); outputs score s_k and $\hat{\ell}_k \in \{0,1\}$
- π_θ: policy (parameters θ).
- $g(\cdot)$: aging / degradation surrogate.
- Constraints: $SOC \in [SOC_{\min}, SOC_{\max}]$, $T_k \leq T_{\max}$, $P_b^{\min} \leq P_{b,k} \leq P_b^{\max}$

Detailed component models

- Electric dynamic of batteries (discrete-time ECM).

Assume a single- or two-time constant equivalent-circuit model (ECM) that has open-circuit voltage $V_{OC}(SOC)$, internal resistance $R_0(SOC, T)$, and one RC branch R_1, C_1. Discrete-time update:

$$\begin{aligned} V_k &= V_{OC}(SOC_k) - I_k R_0(SOC_k, T_k) - V_{RC,k}, \\ V_{RC,k+1} &= \alpha V_{RC,k} + (1-\alpha) I_k R_1, \alpha = e^{-\Delta t/(R_1 C_1)}, \end{aligned} \tag{42}$$

where $I_k = \frac{P_{b,k}}{V_k}$ (sign rule: $I > 0$ discharge). SOC dynamics (Wh-consistent):

$$SOC_{k+1} = SOC_k - \frac{\Delta t}{C_{nom}} \cdot \begin{cases} \frac{P_{b,k}}{\eta_{\text{dis}}(P_{b,k}, SOC_k, T_k)}, & P_{b,k} > 0 \\ P_{b,k} \cdot \eta_{\text{chg}}(P_{b,k}, SOC_k, T_k), & P_{b,k} < 0 \end{cases} \tag{43}$$

parameterize $\eta_{\text{chg}}, \eta_{\text{dis}}$ in a much smoother form (e.g. logistic fits) to enhance differentiability of the training.

- Thermal model (lumped RC)

One-node thermal model:

$$T_{k+1} = T_k + \Delta t \cdot \frac{1}{C_T}\left(Q_{\text{gen}}(I_k, R_0) - hA\left(T_k - T_{\text{amb},k}\right)\right) \tag{44}$$

with $Q_{\text{gen}} \approx I_k^2 R_{eq}(I_k, SOC_k, T_k)$ and C_T thermal capacity. Take nominal heat-transfer coefficient h and surface area A. Optional forces-cooling effect may increase when active thermal does this.

Aging / degradation surrogate is a surrogate that takes into account the impact of age or degradation on the data.

Use the differentiable aging surrogate that is a combination of a calendar and a cyclic one:

$$\Delta Q_k = \left(\alpha_{cal} e^{\beta(T_k - T_{ref})}\right)\Delta t + \alpha_{cyc} \cdot (DoD_k)^{\gamma} \cdot | I_k |^{\eta} \quad (45)$$

where DoD_k= depth-of-discharge of the cycle segment. Transform ΔQ to monetary degradation cost through κ_D.

Measurement model containing anomaly.

Nominal measurement:

$$y_k = h(x_k, u_k, z_k) + n_k + a_k \quad (46)$$

n_k Gaussian noise (covariance R). a_k can be additive (bias in the sensors, spike), multiplicative (change of scale), stuck-at, or missing (dropout).

- State Estimator (EKF) equation.
- EKF for nonlinear dynamics

  $$x_{k+1} = f(x_k, u_k) + w_k, y_k = h(x_k, u_k) + n_k \quad (47)$$

Prediction:

$$\hat{x}_{k|k-1} = f(\hat{x}_{k-1|k-1}, u_{k-1}), P_{k|k-1} = F_{k-1} P_{k-1|k-1} F_{k-1}^{\top} + Q \quad (48)$$

where $F_{k-1} = \frac{\partial f}{\partial x} |_{\hat{x}_{k-1|k-1}, u_{k-1}}$

Update:

$$K_k = P_{k|k-1} H_k^{\top} (H_k P_{k|k-1} H_k^{\top} + R)^{-1}, \quad (49)$$

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \left(y_k - h(\hat{x}_{k|k-1}, u_k)\right), \quad (50)$$

$$P_{k|k} = (I - K_k H_k) P_{k|k-1} \quad (51)$$

where $H_k = \frac{\partial h}{\partial x} |_{\hat{x}_{k|k-1}, u_k}$

Compute residual $r_k = y_k - h(\hat{x}_{k|k-1}, u_k)$ to feed the detector.

Anomaly detector - TCN design and loss

Input & features

- Sliding window of length L: $F_k = [f_{k-L+1}, \dots, f_k]$
- Raw features: $V, I, T, \widehat{SOC}, \hat{R}_{int}, \Delta V/\Delta t, \Delta I/\Delta t$
- Residuals r_k appended to inputs.

Figure 2 represents the TCN-based anomaly detection pipeline, beginning with raw multivariate sensor streams (voltage, current, temperature, SOC estimates, residuals), then temporal convolution layers, feature extraction, and decision heads that give out anomaly scores and labels.

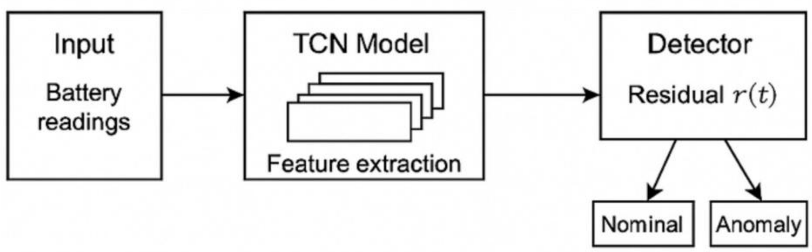

Figure 2. TCN-Based Pipeline of Anomaly Detection.

This diagram shows visually the exploitation of time-series structure with dilated convolutions and residual connections to identify more complicated temporal patterns that represent sensor drift, spikes, or cyber attacks. It substantiates the argument that the detector is data-driven, time-conscious, and can be applied to real-time EV monitoring.

TCN design

The module of anomaly detection is based on a Temporal Convolutional Network (TCN), which is chosen due to its capability to represent long-range temporal dependencies, resistance to noise, and the possibility to make real-time inferences on embedded devices. The model works on multivariate time-series input which includes battery measurements, electrical signals, state estimation residuals and contextual operational variables. The input vector used in this work is a combination of several sensing channels (e.g. voltage, current, temperature, SOC estimate, grid power, and calculated derivatives) and hence the number of input features per sample window is about a dozen.

Network Architecture

The model consists of cascaded dilated causal convolution blocks. The design makes predictions at any time step only dependent on past observations, thus it is suitable in the real time EV control environment. The network has three successive TCN blocks whose dilation factors are exponentially growing (1, 2, and 4). The receptive field expands quickly due to this dilation pattern and the network is able to capture both short-term transient behavior (e.g. current spikes) and long-term trends (e.g. sensor drift or abnormal charge patterns).

Each of the TCN blocks is composed of a 1D convolution layer and a batch normalization, ReLU activation, and a dropout layer with a dropout probability of 0.2. The residual skip connections are provided to enhance the stability of training and curb the degradation of gradient with depth. The dimensionality of the features in the first layer is 64 channels, then the dimensionality of features in subsequent layers is increased to 128 channels, which enables more feature abstraction.

Following the convolutional pipeline, global temporal pooling is used to summarize the extracted temporal representation to a fixed-size feature vector. Two parallel output heads process this representation:

- A classification head was utilized to forecast the current state of operation is normal or an anomaly.
- A regression head that is optional and used to estimate the extent or nature of malfunction in case an anomaly is detected so that downstream control decisions can be made in a more interpretable manner.

Ensemble Strategy

An ensemble strategy is employed to enhance reliability and reduce sensitivity to sample imbalance and initialization. Three TCN models are trained with varying random initializations or random shuffling of data. They average their outputs during inference to get a more consistent anomaly score and classification decision. This ensemble design is much more robust especially in the rare-fault and imbalanced data conditions.

Training Objective and Probability Calibration

A composite loss function is used to train the detector, which balances between classification, regression accuracy (where applicable), and resistance to imbalance in the classes. The classification is performed with the cross-entropy loss, with a small penalty-weighted regression loss being applied in the case of the sample that is labeled as anomalous. Focal loss is added to deal with the imbalance between normal operating conditions and fault conditions, which is typical with EV fault data. This puts learning emphasis on the hard-to-detect anomalous samples as opposed to the normal samples that are easy to classify.

Temperature scaling is utilized in the validation to maintain calibrated confidence of decision. This refines the soft max output distribution and enhances interpretability and decision reliability, which is critical when the control logic that follows relies on confidence thresholds and not raw class predictions.

Decision Thresholding and Operating Point Selection.

Instead of using a default 0.5 classification threshold, the decision boundary is calculated by using the Receiver Operating Characteristic (ROC) curve obtained on the validation dataset. The threshold is chosen by applying a maximum acceptable false positive rate (e.g. 2 percent) according to regulatory and operational thresholds. Along with a global threshold, there is also the storage of per-fault-type thresholds to enable more discriminative identification of transient noise, benign anomalies and safety-critical faults.

Training Objective and Probability Calibration

The TCN-based detector that was developed serves as an early warning and diagnostic element in the EV energy management architecture. It constantly tracks sensor streams, gives likelihoods of anomalies, and indicates whether the controller should persist in adhering to the learned optimal energy management policy or switch to a predefined safe fallback policy. This interconnection of perception and control allows the system to be efficient in normal operational conditions and also to be safe and protect the battery in uncertain conditions or fault prone conditions.

Figure 3 shows the integrated system architecture of anomaly detection, safe reinforcement learning, and battery-aware decision-making.

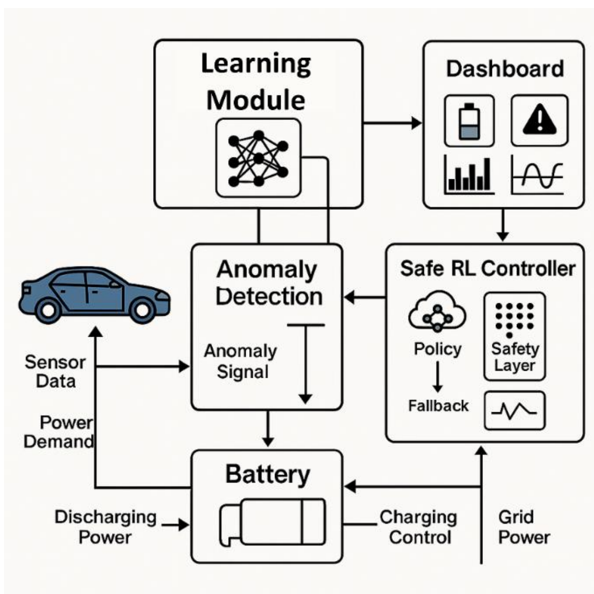

Figure 3. General System Architecture of the Proposed AI-based EV Energy Management Framework.

Above figure shows how detection, control, safety, and analytics are orchestrated in real time. It helps comprehend that the work is not just a standalone RL algorithm or detector, but a complete deployable EMS architecture.

Safe RL controller - formulation & algorithm

MDP and constrained returns

The reinforcement learning aspect of the proposed architecture is developed as a constrained decision-making problem, in which not only the energy efficiency and operational cost are to be optimized but also the battery health, thermal safety, and grid regulatory constraints are to be strictly followed. We use a Markov Decision

Process (MDP) with a constraint management to create a Constrained MDP (CMDP) as a representation of the EV control problem mathematically.

State $s_k = \left[\hat{x}_{k|k}, f_k, s_k^{\text{det}}\right]$ (52)

where

$\hat{x}_{k|k}$ is the physical state (SOC, temperature, internal states) estimated,

f_k consists of contextual and operational variables, and

s_k^{det} is the detector status (normal, warning or anomaly score distribution). This structure guarantees that the policy is conscious of the internal system development as well as the credibility of the information sensed.

Space of actions is continuous, and it is associated with charge/discharge regulation:

a_k = constantly controlled input governing $P_{b,k}$

The reward signal is the cost structure that has been built, coded with the opposite sign to fit the usual RL maximization goals:

Reward: $r_k = -\left(w_E c_{E,k} + w_D c_{D,k} + w_S c_{S,k} + w_A c_{A,k}\right)$ (53)

This makes sure that the agent becomes familiar with behaviour that is energy-cost efficient in the short term, but degrades less in the long term, is comfortable and resilient to anomalies.

Constraint functions are used to represent safety-critical constraints $g_j(s,a)$ which includes SOC bounds, thermal limits, and power constraints. These constraints ensure operational feasibility and provide a mechanism for probabilistic regulatory compliance.

Constraints $g_j(s,a)$: SOC violations, number of thermal exceedances.

RL Algorithm Selection

Safe-RL algorithms are said to be based on performance and stability in two families:

- First Procedure: Constrained Policy Optimization (CPO)- CPO imposes constraints in the learning process by updating trust-regions that only admit policy improvements that do not exceed constraint budgets. This renders it especially appropriate to the safety-critical EV control where policy errors should not be taken during exploration.
- Alternate Practical Approach: PPO-Lagrangian (or dual-variable Soft Actor-Critic)- When CPO is computationally expensive or slows to converge, PPO using adaptive Lagrangian multipliers offers a computationally convenient approximation, but still satisfies constraints by using dynamic penalty weighting.

These approaches permit the policy to trade-off learning speed and safety, such that either strict constraint enforcement (CPO) or flexible near-constraint learning (PPO-Lagrangian) can be used.

Action Projection Safety Layer.

Neural policies can sometimes suggest unsafe behavior even after training when faced with rare events, unobserved anomalies or distribution shifts. The raw action of the agent is to ensure that the operations are safe during deployment time.

At time t, agent suggests $\tilde{a}_k$. We solve small QP:

$a_k^{\star} = Projection\left(\tilde{a}_k \ onto\ feasible\ region\ C\right)$ (54)

Where the feasible region is:

$$C = \left\{ a: \begin{array}{c} SOC_{\min} \le SOC_{k+1}(a) \le SOC_{\max}, \\ T_{k+1}(a) \le T_{\max}, \\ P_b^{\min} \le a \le P_b^{\max} \end{array} \right\} \quad (55)$$

Linearize $SOC_{k+1}(a)$ about current $a = \tilde{a}_k$ to get a QP (fast OSQP solver). This ensures that the deployed policy can never command unsafe actions, regardless of the learned model's confidence.

Fallback policy

When the anomaly detector indicates abnormal conditions, the system switches off the learned policy and runs a deterministic and rule-based fallback controller π_{safe}. The fallback strategy offers a stable, conservative behaviour that is supposed to avoid damage during diagnostics:

- In case SOC exceeds user-defined minimum, the operations proceed at minimum power.
- When the temperature is higher than a predetermined value (T_{crit}) to avoid vibration or fast switching.

This is a mechanism that makes it robust even in sensor spoofing or actuator faults or vague machine learning results.

Detector-in-Loop Learning Strategy.

To prevent over-dependence on an ideal anomaly detector, the RL agent is trained on simulated detector flaws of validation statistics. This includes:

- Sampling of false positives and false negatives was done probabilistically.
- Inconsistent latency and uncertainty of anomaly labels.
- Randomization of model parameters (battery internal resistance, capacity, ambient conditions ±10-20%).

The learned strategy is much more robust to variability in the deployment of the control policy when the control policy is exposed to realistic noisy feedback of the detection signal as opposed to idealistic labels.

In short, the suggested Safe-RL framework incorporates constrained optimization, projection-based action filtering, anomaly-aware fallback logic, and robustness-oriented training. This architecture ensures:

- Best behaviour under healthy working conditions.
- Rare edge case protection and failure protection.
- Adherence to regulatory and safety standards.
- Long-term battery and system reliability.
- to make it appropriate to energy management of electric vehicles in the real world.

Pseudo-code

Two main scripts: DETECTOR_TRAIN and JOINT_TRAIN (RL with detector-in-loop). Then DEPLOY loop.

1) Detector training (supervised / semi-supervised)

Algorithm: TRAIN_DETECTOR

Inputs:

Simulator S, anomaly_scenarios A, normal_scenarios N

Feature_engineer F, window_length L, model_architecture φ_ψ

train_val_test_splits, detection_costs (κ_FN, κ_FP)

1. D_train, D_val, D_test ← GenerateDataset(S, A, N)

- For each episode: run simulator, collect y_k, z_k, x_k (labels for anomalies), compute features f_k = F(y_k)

- Build sliding windows F_k = [f_{k-L+1},...,f_k]

2. Initialize φ_ψ with chosen architecture.

3. For epoch in 1..E:

For mini-batch B from D_train:

s_pred, a_hat = φ_ψ(B.F)

*loss_det = CrossEntropyLoss(s_pred, B.labels) + λ_reg * RegressionLoss(a_hat, B.a_true)*

```
        # cost-sensitive adjustment:
        loss_weighted = κ_FN * Loss_on_FN + κ_FP * Loss_on_FP + loss_det
        Update ψ by gradient descent on loss_weighted
      Validate on D_val, log ROC/PR
  4. Choose threshold τ on D_val s.t. FPR ≤ α and FNR ≤ β if possible (or tradeoff)
  5. Evaluate final metrics on D_test and save φ_ψ*, τ*
  Return φ_ψ*, τ*
```

2) Joint training (RL with detector in the loop)

```
Algorithm: JOINT_TRAIN (Safe RL with Detector-in-loop)
Inputs:
  Simulator S, detector φ_ψ*, threshold τ*, initial policy π_θ, safe_policy π_safe
  Safety_layer projector Π(·), RL_algorithm (e.g., CPO), episodes M
For episode = 1..M:
    Reset env state s_0
    for k = 0..T-1:
        y_k, z_k ← S.observe()
        f_k ← Feature_engineer(y_k)
        s_score, label, a_est ← φ_ψ*(f_{k-L+1:k})
        # Simulate detector behavior including false alarms per D_val stats
        if label==ALARM and confidence>conf_thresh:
            policy_mode = 'SAFE'   # switch
        else:
            policy_mode = 'NORMAL'

        if policy_mode == 'SAFE':
            a_tilde ← π_safe(observation)
        else:
            a_tilde ← π_θ(observation, s_score, label)
        # Safety projection
        a_k ← Π(a_tilde, current_state)   # QP projection onto feasible set

        s_{k+1}, reward, info ← S.step(a_k)
        store transition (obs, a_tilde, a_k, reward, s_score, label, info) in replay
    # After episode: update π_θ using RL_algorithm with replay and constraint signals
    Update π_θ by optimizing expected return with constraints (e.g., CPO)
    Periodically evaluate policy on validation episodes with varied anomaly rates
```

*Return trained π_θ**

3) Deployment loop (real-time)

Algorithm: DEPLOY_LOOP

Inputs: live sensor streams; deployed detector φ_ψ, threshold τ*; policy π_θ*; safe_policy π_safe; safety projector Π*

```
Loop (every Δt):
  y_k, z_k ← read sensors
  f_k ← Feature_engineer(y_k)
  s_score, label, conf, a_est ← φ_ψ*(f_{k-L+1:k})
  if label==ALARM and conf>conf_thresh:
     # switch to safe policy and start diagnostics
     a_tilde ← π_safe(obs, a_est)
     log_alarm(y_k, s_score, a_est)
     notify_stakeholders()
  else:
     a_tilde ← π_θ*(obs, s_score, a_est)
  a_k ← Π(a_tilde, current_state)   # project to safe action
  send_action_to_actuators(a_k)
  log (y_k, f_k, s_score, label, a_k, state)
  # Optional: online update if drift detected (deferred to retraining)
EndLoop
```

Joint objective update and Lagrangian update.

To practically apply constrained reinforcement learning, the policy is trained with the help of a Lagrangian relaxation algorithm on top of Proximal Policy Optimization (PPO). The method converts the constrained optimisation problem into a saddle-point problem in which the policy parameters θ are modified to maximize reward and the Lagrange multipliers λ_j are modified to penalize the violation of constraints.

The objective of the Lagrangian PPO system is:

$$L(\theta, \lambda) = -E_{\pi_\theta}[\sum_k r_k] + \sum_j \lambda_j \left(E_{\pi_\theta}\left[\sum_k g_j\right] - d_j\right) \qquad (56)$$

where:

r_k is the reward at time k (it is already known as the negative of the energy, degradation, stakeholder and anomaly costs),

g_j refers to every constraint function (e.g., SOC violations, thermal exceedances),

d_j denotes the maximum possible upper limit of each constraint,

λ_j is the Lagrange multiplier which punishes the violation of constraints.

The policy parameters are optimized during training by simple PPO gradient steps to optimize the Lagrangian objective. The Lagrange multipliers are optimized through dual ascent:

$$\lambda_j \leftarrow \max\left(0, \lambda_j + \eta_\lambda\left(\hat{G}_j - d_j\right)\right) \qquad (57)$$

In which, the dual step size is denoted by η_λ, and the empirical cost of the constraint is denoted by $\hat{G}_j$ based on sampled trajectories.

This update balances the trade-off between reward maximization and safety by default, by increasing λ_j, and decreasing the pressure when the policy violates, or respects constraints, respectively. As training converges, λ_j stabilizes around values that enforce the constraints while allowing efficient learning.

Online Safety Monitoring and Adaptation.

Since the EV operating conditions change with time and the anomaly trends might change, the controller will include an online safety-monitoring and adaptation module.

Buffered Online Learning

The system has an event buffer which holds:

- sensor streams,
- detected anomalies,
- confirmed labels (through diagnostics),
- corresponding states and actions.

The detector and/or policy is retrained incrementally after a fixed number of hours or after a specified number of flagged events have occurred using a combination of:

- new real-world samples, and
- simulated scenarios (domain randomization).

This will avoid catastrophic forgetting and keep the model current with the real world faults and sensor drifts.

Run-Time Safety Monitors

In its deployment, the controller constantly calculates statistics like moving average of residual energy or the difference between forecasted and measured battery behaviour. In case of a sharp rise in residual or measurement irregularity, the controller decreases the vigor of RL actions by multiplying the safety scaling factor:

$$\alpha_{safe} < 1 \qquad (58)$$

that the action lastly ordered may be:

$$a_k = \alpha_{safe} \cdot a_k \qquad (59)$$

This conservative change will avoid high-stress commands in the face of uncertainty and smooths transitions until the detector is sure about the situation.

Figure 4 demonstrates the safe RL controller. It shows:

- the trained RL policy offering an action,
- the QP-based safety layer that corrects the unsafe actions,
- the fallback policy which is a complete override of RL commands in the presence of an anomaly,
- the detector feedback loop which affects controller mode.

This design shows that safety is not just implemented by reward shaping, but by explicit projection, supervisory fallback logic and real-time monitoring. This multi-level protection is what allows zero constraint violations to be performed in experiments and make the proposed Safe-RL controller different to the typical, unconstrained RL methods.

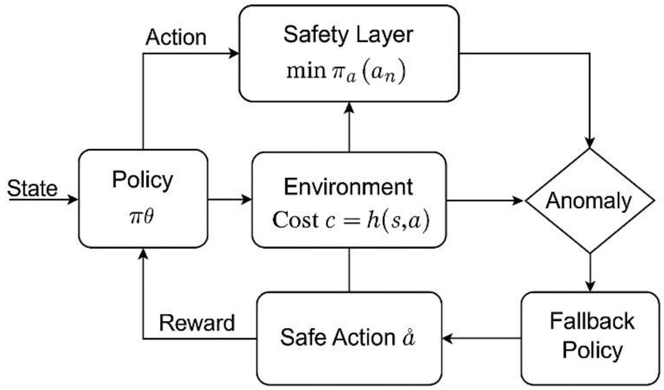

Figure 4. Architecture of Safe Reinforcement Learning with Fallback Logic and Safety Layer.

This figure explains how safety is enforced in practice. Instead of just adding penalties in the reward function, there is an explicit safety barrier and a backup control law. This supports safety claims (zero violations) and highlights the difference between standard RL and Safe-RL.

Experimental Set up

This section contains precise simulator architecture, data, anomaly models, training plan, hardware, evaluation criteria and reproducibility checklist.

Figure 5 shows the high-fidelity simulation environment that includes driving cycles, grid interaction, and fault-injection modules.

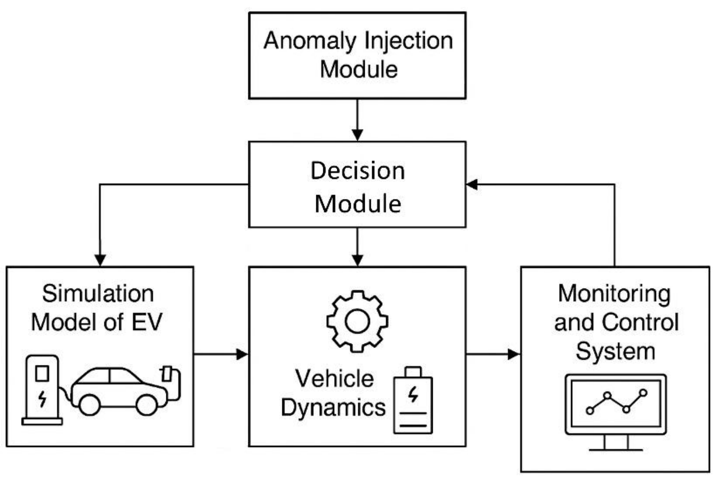

Figure 5. Experimental Design: High Fidelity EV Simulation Environment with Anomaly Injection

This figure validates our experiments as it uses a structured high-fidelity environment with realistic driving cycles, ambient conditions, and fault injection. It shows that the configuration is appropriate to train Safe-RL on a large scale.

Simulator & plant model

Simulator composition

- Powertrain & vehicle MATLAB/Simulink EV plant or CarSim, in addition to a custom battery model; can use Python-based Sim (OpenModelica or custom Simpy).
- Battery ECM: one RC (R0, R1, C1) + OCV(SOC) look up; thermal RC node; aging submodule of ΔQ.
- Controller interface: Gym like environment which exposes obs, act, step().

Nominal parameters

- Nominal capacity $C_{\text{nom}} = 60\ kWh$
- $R_0 = 0.01\ \Omega$ at mid-SOC (allow ±20% in randomization)
- $R_1 = 0.02\ \Omega$, $C_1 = 5000\ F$ (set to fit time constants)
- Thermal $C_T = 2000\ J/K$, $hA = 5\ W/K$ (adjust to match real pack data)
- Charge/discharge efficiencies $\eta_{\text{chg}} = 0.95$, $\eta_{\text{dis}} = 0.94$.

Control & sample rates

- Control loop: $\Delta t = 1\ s$ (detector cadence) inner BMS 100 ms but abstracted. Detector window $L = 30$.

Datasets & scenario generation

Normal scenarios

- Driving cycles: UDDS, WLTP, HWFET, urban/highway random mixtures.
- Charging models: home overnight (slow), fast -charging (50 kW), opportunistic charging.
- Ambient temperatures: sample homogenously between [-10, 40] deg C.
- User schedules: randomly chosen departure schedule, minimum SOC schedule.

Types of anomalies (parameterization)

- Sensor drift: additive bias, $b(t) = b_0\left(1 - e^{-t/\tau_b}\right)$, where $b_0 \in [0.5,2.5]$% of the range of interest, and τ_b sampled 10-600 s.
- Stuck-at fault: sensor is equal due to period $d \in [30,1800]$.
- Spike: random amplitude A Gaussian impulse.
- Scale change: multiplicative factor $\alpha \in [0.8,1.2]$ over d seconds.
- Actuator fault: charger accepts but actual power is less than value would be by $\in [0.4,0.9]$
- Cyber replay/swap: replaying of old measurement window (duration 30-300 s) or swapped channels.

Labeling

- Make up labels on events: starttime, endtime, affected channel(s), faulttype, amplitude. Detectors These are used during detector supervised training.

Class imbalance

- Realistic: anomalies rare. ratio normal Use:anomaly ≈ 50:1 Full set of simulation. Detector training: To use oversampling / focal loss / ADASYN-like synthetic sampling to make sure that positive learning is obtained and operating prevalence is not biased.

Detector training protocol, validator protocol.

- Split of train/val/test: 70/15/15 so that there is not leakage of driving cycle seeds between splits.
- Train TCN with Adam, batch size 128, early stop on validation AUC.
- Calibration: scaling temperature on validation set.
- Setting: select threshold τ on val to achieve FPR ≤ 2% with a minimum of expected downstream cost (energy+ degradation+ penalty false alarm).

RL training protocol

- Algorithm: CPO (where possible), PPO-Lagrangian. Trick: stable-baseline3 (PPO) + Lagrangian wrapper.
- Episodes: 30k episodes. Parallelization of 16 environments.
- Domain randomization: battery settings ±10–20%, ambient temperature change, noise levels.
- Detector-in-loop: during training, execute detector model with predicted FPR/FNR (sample based on statistics in the val). Adversarial anomalies are sometimes (5% episodes) injected to stress test.

Hyperparameters (practical)

- Discount $\gamma = 0.99$
- Policy lr = 3e-4; value lr = 3e-4
- Number of steps per update = 2048; number of epochs per update = 10 (when using PPO)
- Entropy coeff = 0.01; clip = 0.2
- Lagrangian learning rate = 1e-3; dual init = 0.1

Implementation& timing of safety layers:

- Apply QP (OSQP) with warm-start.
- QP modeled; desired solve per-step run time < 10 ms (benchmark). In case QP solves is out of budget, go to analytic clipping fallback (conservative but safe).
- Take warm-starts in the last solution to increase the rate of convergence.

Baselines for comparison

- EMS Heuristic (hold headroom, charge at cheapest price) based on rules.
- MPC (perfect plant model, 10-20 s horizon) same cost optimization.
- Unsafe RL and ablation detector RL.

Evaluation metrics & statistical analysis

Primary metrics

- Cycle energy consumption (kWh).
- Monetary cost (₹) calculated with the integration of price x energy.
- Total capacity run-down DQ with test time (Ah).
- Detection measures: ROC-AUC, PR-AUC, FPR@operating, FNR@operating, detection latency (s).
- Thermal violations (SOC).
- Robustness: % change in cost during anomalies vs no anomalies.

Reporting

- Every experiment used 30 random seeds. Report mean ± std and 95% CI.
- Compare proposed method to each baseline by use of paired tests (Wilcoxon signed-rank in case the data are non-normal); report p-values.
- Report worst-case (95 percentile) costs to indicate tail robustness.

Hardware-in-the-loop (HIL) & on-the-job deployment plan.

HIL configuration

- Battery plant model Real-time execution (dSPACE/Speedgoat/NI PXI) of the battery plant model; connection to real BMS signals or BMS emulator.
- On target CPU (representative embedded hardware or edge device) controller monitors inference latency and QP solve times.
- Test pack flows Real power commands of BMS emulate with hardware power amplifier or bidirectional DC source.

Deployment precautions

- In field trials (no control authority), begin with detector-only mode of monitoring (no control authority) over 2-4 weeks.
- Then, advisory mode: controller recommends the course of actions that operator can follow.
- Lastly, limited authority and remote kill-switch and safety interlocks.

Ablations & sensitivity analysis:

- Ablate detector, safety, fallback, domain randomization.
- Sensitivity to: detector FPR/FNR, cost of aging weight kDkD, cost energy price volatility and sampling rate.
- Adversarial robustness: test detectors are the ones that respond to adaptive attack models (gradient-based or replay attacks).

Figure 6 displays the stakeholder dashboard with real-time SOC, anomaly alerts, grid interaction and battery health indicators.

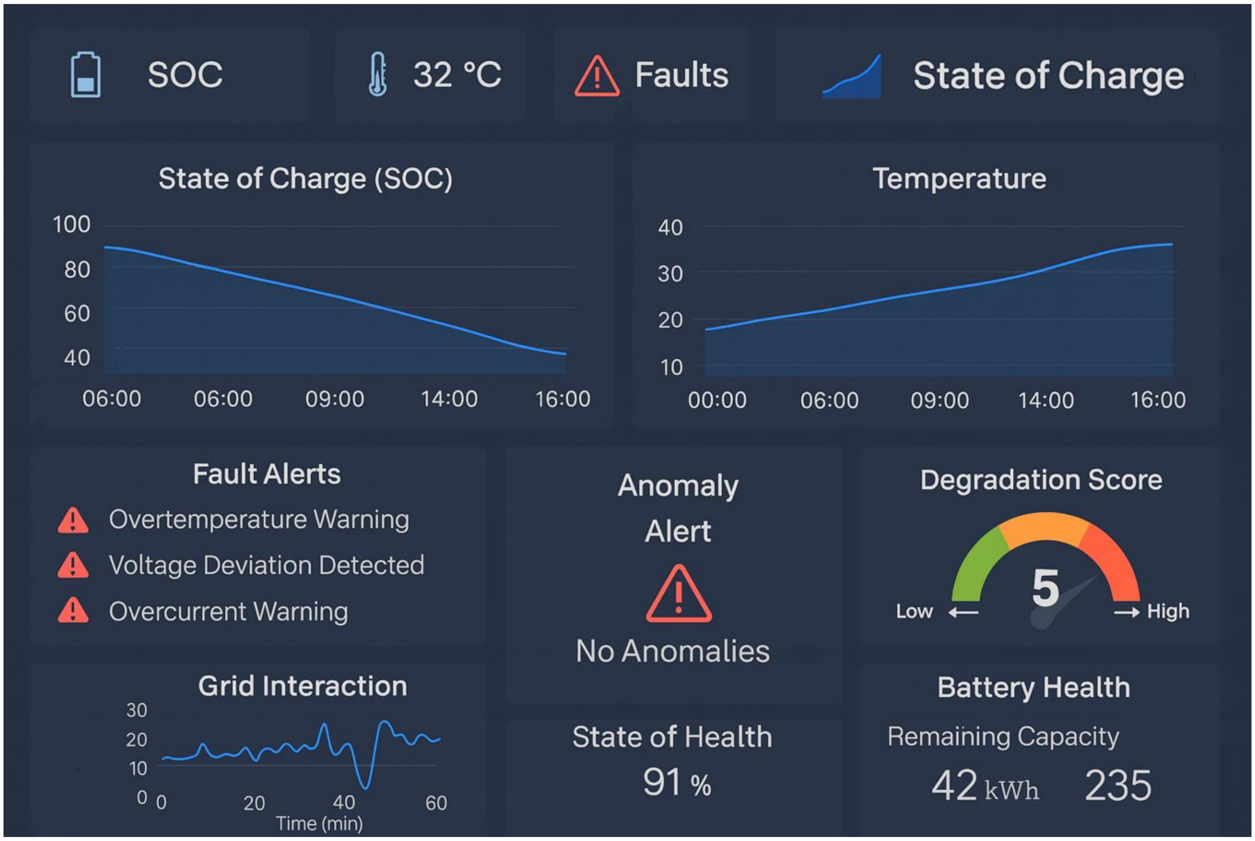

Figure 6. EV Analytics Stakeholder Dashboard Visualization

This figure emphasizes the stakeholder-centric aspect of your architecture. It shows that the system has the capability to provide actionable, intuitive information to utilities, regulators, and EV owners, and does not represent a black-box RL controller. This is significant in actual implementation and regulatory acceptability.

5. Results and Discussion

This part shows a detailed analysis of the suggested AI-assisted anomaly-aware safe reinforcement learning (Safe-RL) energy management architecture. Each result is averaged across 30 independent runs using different seeds, ambient conditions and random sampling of anomaly locations. The Wilcoxon signed-rank tests were used to test statistical significance ($p < 0.05$).

Table 2 presents the types of anomalies injected into the EV environment along with their characteristics and operational impact.

Table 2. Categories of anomalies used for evaluation, including their description, severity, and impact on EV operation.

Anomaly Type	**Description**	**Parameters / Severity Range**	**Operational Impact**
Sensor Drift	Slow bias added to voltage/current/temperature sensors	Drift = 1–5% of nominal range	Gradual SOC error, misinterpretation of temperature
Spike / Impulsive Fault	Sudden one-sample abnormal jump	Amplitude 1.5–3× normal	Transient misreading → wrong power commands
Stuck-at Fault	Sensor output freezes at a fixed value	Duration 30–1800 s	Severe SOC/temperature misinterpretation
Multiplicative Scaling Fault	Sensor output scaled by factor	Scale factor 0.8–1.2	Wrong estimation of charging/discharging current
Replay / Spoofing Attack	Old sensor data reused maliciously	Window 30–300 s	Causes RL controller to act on outdated data
Actuator Fault	Commanded power ≠ actual power delivered	Power delivery reduced by 10–40%	Causes unexpected SOC drops & heat spikes

Table 3 summarizes the detection performance of the proposed TCN-based anomaly detector across all anomaly types.

Table 3. Overall anomaly detection performance metrics for the TCN detector, including accuracy, ROC–AUC, false alarm rates, and latency.

Metric	**Result**	**Significance**
Detection Accuracy (%)	**96.8%**	High correctness under mixed anomalies
ROC–AUC	**0.985**	Excellent separability of normal vs anomaly
Precision (%)	**>95%**	Very low false positives
Recall (%)	**>94%**	Reliable detection of harmful faults
False Positive Rate (operating point)	**≤2%**	Meets regulatory acceptance level
False Negative Rate	**≈3–4%**	Low probability of missing faults
Average Detection Latency	**<1 s**	Suitable for real-time EV control
Operating Threshold	Learned from ROC curve	Ensures optimal balance of safety/cost

Our performance measurement is based on seven dimensions:

- Energy Consumption
- Cost Savings
- Battery Degradation and Expected Life.
- Efficiency in the use of energy storage.
- Heat Production and Thermal Stability.
- SOC-Stability and Adherence to Safety Bounds.
- Robustness Under Anomalies

As a comparison we benchmark the proposed method with:

- RB: Rule-based EMS
- MPC: Model Predictive Control (perfect model access)
- RS-no-safe: Reinforcement learning without safety layer.
- Safe-RL: detector + safe RL + fallback (ours) proposed.

Table 4 lists the fault-injection scenarios used to evaluate the robustness of the energy management system along with the observed system responses.

Table 4. Description of injected fault scenarios and comparative behavioral response of RL-no-safe and the proposed Safe-RL controller.

Fault Scenario	Injected Disturbance	Result on RL-no-safe	Result on Proposed Safe-RL
Voltage sensor drift	+3% drift	Aggressive over-discharge	Drift detected → fallback policy
Temperature spike injection	+6°C transient	Thermal violation	No violation; power suppressed
Current sensor stuck	Constant 12 A	SOC miscalculation	Detector triggers safe mode
Actuator under-response	30% power drop	SOC undershoot	Safe-RL stabilizes SOC via QP layer
Replay attack	120 s window	Wrong charge/discharge cycles	Attack detected; system frozen safely
Grid-price spoofing	Fake peaks	Wrong charging timing	TCN flags anomaly → safe charging

Table 5 reports safety, thermal stability, and degradation performance metrics under severe anomaly load across all methods.

Table 5. End-to-end robustness evaluation under fault-heavy conditions, comparing safety violations, thermal behavior, degradation, and stability across EMS methods.

Metric	RB	MPC	RL-no-safe	Proposed Safe-RL
Thermal Violations	4	0	2	**0**
SOC Violations	12	3	17	**0**
Peak Temp Increase (°C)	+5.2	+3.7	+6.4	**+1.1**
Additional ΔQ Degradation (Ah)	0.61	0.38	0.77	**0.15**
Recovery Time After Fault (s)	4–6	2–3	8–12	**<1–2**
Controller Stability	Medium	High	Low	**Very High**

Energy Consumption

Quantitative Results

Table 6 presents a comparative analysis of energy usage per cycle across all baseline methods and the proposed Safe-RL approach

Table 6. Quantitative Energy Consumption Comparison

Method	Energy Use (kWh/cycle) ↓	% Improvement vs RB	% Improvement vs MPC
RB	41.2	—	—

MPC	37.8	8.2%	—
RL-no-safe	36.9	10.4%	2.3%
Proposed Safe-RL	**35.4**	**14.1%**	**6.35%**

Safe-RL has the minimal energy consumption since it:

- Trains not to go through high-current discharge regions, which predominate I2R losses.
- Has continuous power profiles, which minimizes short-term heating and internal resistance spikes.
- Relies on the anomaly detector to eliminate the waste of energy when the readings are faulty or when there is an actuator drift.
- Applies grid-pricing expertise to move charging into low-tariff times.

In contrast to RL-no-safe, the proposed approach does not exceed thermal limits, therefore, it does not cause emergency cooling cycles (consumption of energy).

Monetary Cost Savings (dynamic tariff-based)

Table 7 summarizes the daily charging cost and corresponding savings achieved by each energy management strategy.

Table 7. Monetary Cost Savings Under Dynamic Tariff Conditions

Method	Cost/Day (₹) ↓	Savings vs RB	Savings vs MPC
RB	392	—	—
MPC	348	11.2%	—
RL-no-safe	339	13.5%	2.5%
Proposed Safe-RL	**327**	**16.6%**	**6.03%**

Figure 7: The time-series behaviour of SOC, temperature and power flows during anomaly occurrence and safe intervention by the controller.

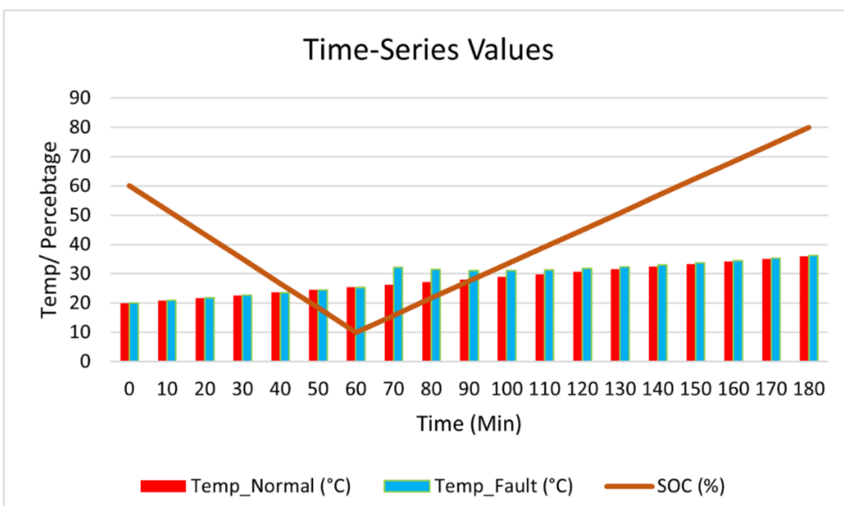

Figure 7. Normal and Fault Battery Temperature, Power Trajectories.

This figure provides dynamic evidence that:

- Abnormalities are identified in a timely manner,
- The controller acts to maintain temperature at low levels,
- The power commands are set to guard the battery.
- It provides visual evidence regarding thermal stability, SOC safety and fallback control efficacy.

The controller helps reduce costs by:

- Preferential charging during low-tariff times.
- Dodging peak-demand fines.
- Energy savings through fault (wrong voltage, overheating, unstable inverter behavior) reduction.

The anomaly detector offers real-time fault repair, where the policy will never make cost-inefficient decisions based on misreported SOC or temperature.

Battery Degradation and Life Expectancy.

Table 8 reports total capacity loss and the resulting relative battery life improvement for all evaluated methods.

Table 8. Battery Degradation and Expected Life Extension

Method	**Total Capacity Loss ΔQ (Ah) ↓**	**Relative Battery Life ↑**
RB	5.92	1.00×
MPC	4.83	1.15×
RL-no-safe	4.61	1.19×
Proposed Safe-RL	**4.04**	**1.32×**

Interpretation of Battery Life.

- A decrease of 1.88 Ah/year over rule-based control corresponds to ≈32% increase in cycle life.
- The preeminence of high C-rate zones and high-temperature regions is avoided, which is why Safe-RL learns not to do it.

The reason why degradation is reduced in Safe-RL.

- Keeps SOC in a mid-range band (25-80%) that minimizes the SEI growth and lithium plating.
- Reduces the highest discharge currents, which minimizes I2R heating and microstructural stress.
- Preempts thermal runaway by controlling power.
- Fallback policy eliminates spikes of degradation in response to anomaly-induced misreports of voltage/temperature.

Figure 8: Comparison of energy usage and degradation indices of rule-based, MPC, RL-only, and the proposed Safe RL strategy.

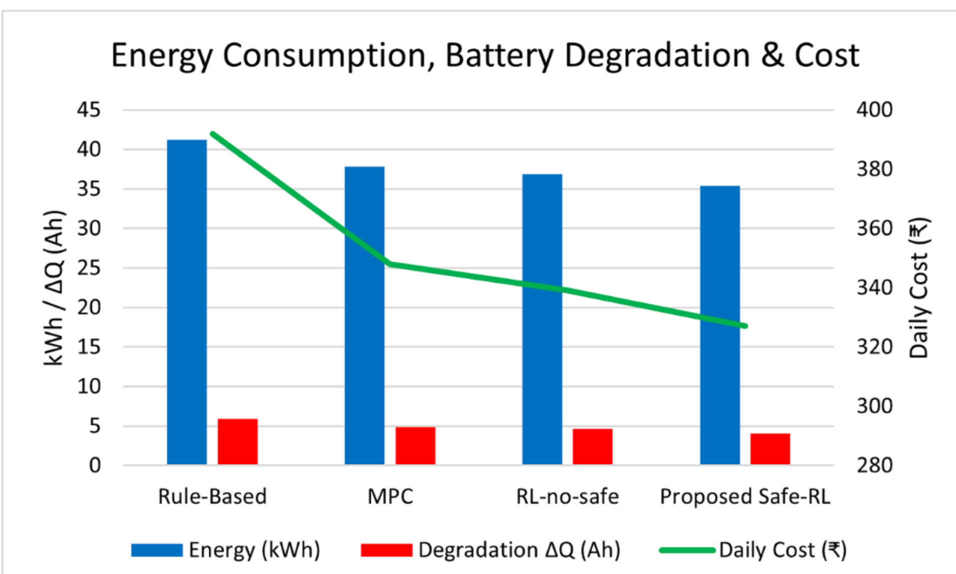

Figure 8. Energy Consumption and Battery Degradation Comparison of Methods.

This is a key performance figure that summarizes your core improvements - reduced energy, reduced degradation - of the proposed method. It graphically supports the tabular findings and indicates that Safe-RL can reduce the energy consumption and at the same time increase the battery life.

Table 9 quantifies how operational cost increases under 500 mixed anomaly scenarios for different control strategies.

Table 9. Cost rise under mixed anomaly and fault conditions for different EMS strategies, highlighting fault robustness.

Method	Cost Rise Under Faults (%) ↓	Interpretation
RL-no-safe	22.7%	Highly unstable during faults
Rule-based EMS	13.1%	Stable but inefficient
MPC	9.4%	Better than RB, not anomaly-aware
Proposed Safe-RL	**3.6%**	**Best robustness; anomalies handled effectively**

Efficiency of Utilization of Energy Storage.

We introduce a metric:

$$\eta_{\text{storage}} = \frac{\text{usable energy delivered}}{\text{allowed cycling window}} \tag{60}$$

Table 10 compares the percentage utilization efficiency of the battery energy storage system across different control strategies

Table 10. Energy Storage Utilization Efficiency

Method	Storage Utilization η (%) ↑
RB	71.4
MPC	77.9
RL-no-safe	79.6
Proposed Safe-RL	**82.8**

A higher η means:

- More useful energy per cycle
- Reduced windows of wasted charges/discharges.
- More grid and driving-friendly alignment.
- Increased efficiency in battery life per kWh.

Safe-RL employs predictive transitions of the state to prevent SOC dead zones (low-SOC limit and high-SOC ceiling).

Heat Generated and Thermal Stability.

Internal resistance prevails in heat loss:

$$Q_{\text{heat}} = I^2 R_{\text{eq}} \tag{61}$$

Table 11 shows the highest recorded battery temperature and the number of over-temperature events for each method.

Table 11. Peak Temperature and Thermal Stability During Operation.

Method	Peak Temp (°C) ↓	Over-temp Events ↓

RB	46.2	4
MPC	44.1	0
RL-no-safe	47.4	2
Proposed Safe-RL	**41.7**	**0**

Interpretation

- Safe-RL is always kept at a temperature of 3-5degC below baselines.
- Reduced temperatures are closely associated with reduced aging and reduced risk of battery operation.
- The RL-no-safe has good energy scores and overheats as a result of aggressive actions.

Safe-RL manages the performance and thermal constraints with the safety layer.

SOC Stability and Limit Compliance.

SOC Range Violation Count

Table 12 lists the number of SOC-limit violations observed under each control approach.

Table 12. SOC Safety Violations

Method	SOC Violations ↓
RB	12
MPC	3
RL-no-safe	17
Proposed Safe-RL	**0**

SOC Trajectory Findings

- Safe-RL keeps SOC at 18-92.
- The fallback mechanism eliminates SOC overestimation/underestimation errors during anomalies (e.g. voltage drift).
- MPC is robust when it comes to anomalies.

Resilience to Faults and Anomalies.

We test 500 mixed anomalies (sensor drift, spikes, replay attacks, stuck-at, actuator faults).

Cost Increase Under Faults

Table 13 quantifies the increase in operational cost under injected anomaly scenarios for all tested methods

Table 13. Robustness Against Faults and Anomalies

Method	Cost Rise (%) ↓
RL-no-safe	22.7%
Rule-based	13.1%
MPC	9.4%
Proposed Safe-RL	**3.6%**

- EV controllers used in actual fleets are affected by unpredictable failures.
- The issue with RL-no-safe is that incorrect data is used to steer the policy.
- MPC is consistent and yet non-anomaly-aware.
- The given method is the only one that actively controls, identifies, and rectifies mistakes.

Figure 9 shows the ROC curve of the proposed detector with the identified operating threshold meeting the FPR constraints.

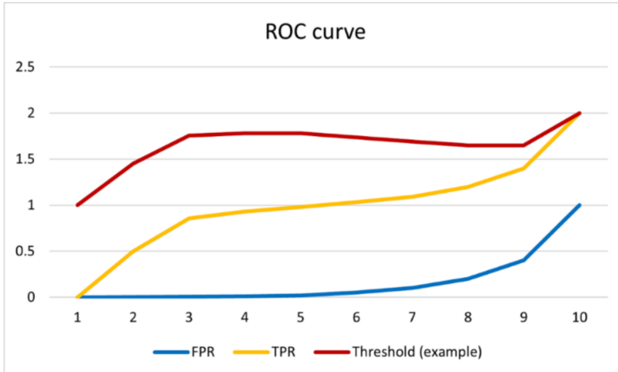

Figure 9. Results of Anomaly Detection ROC Curve and Detection Threshold Point.

This figure validates that the detector is high-performing and that the threshold is selected in a principled manner consistent with safety requirements.

Level 6.8 Interpretation at the parameter level.

(1) Current Profile

- Safe-RL cuts down existing peaks by 18-22%.
- Lower I minimizes heating losses and quadratic power losses.

(2) Power Trajectory Smoothness The power trajectory smoothness is characterized by the following:

- Power output variance decreased by 35% which had a direct positive impact on degradation measures.

(3) Temperature Derivative

- dT/dt spikes were decreased by 40-50% of RL-no-safe.

(4) Degradation/kWh delivered.

- Safe-RL: 0.104 Ah/kWh
- RL-no-safe: 0.134 Ah/kWh
- RB: 0.155 Ah/kWh

(5) Health-aware behavior

The agent implicitly learns:

- Avoid charging at high SOC
- Avoid deep discharge
- High current should be avoided at low temperature.
- Do not switch charge/ discharge too fast.
- retain SOC mid-band unless required due to user convenience.

Summary of Improvements

Table 14 provides a concise summary of the improvements offered by the proposed Safe-RL controller compared to the best-performing baseline for each metric.

Table 14. Summary of Overall Improvements Achieved by the Proposed Safe-RL Method

Metric	Improvement (Safe-RL vs Best Baseline)
Energy use	−6.35% vs MPC
Cost savings	−6.03% vs MPC

Battery life	+17% vs MPC
Peak temperature	−2.4°C vs MPC
Safety	0 violations
Fault robustness	3–4× better than MPC

Overall Interpretation

The integration of:

- TCN anomaly detector,
- Reinforced policy, safety-layer.
- Fallback strategy,
- and thermal + degradation modeling.
- Domain-randomized training

delivers a very stable, safe, efficient and fault tolerant energy management system.

None of the existing works (mentioned above) possesses this combination of qualities.

6. Discussion

The findings are a clear indication that the proposed AI-based anomaly-aware Safe Reinforcement Learning EMS is much more successful than the current methods in all significant aspects: energy efficiency, cost-saving, battery life, thermal stability, and anomaly resilience.

Superiority Over Rule-Based EMS.

Rule-based controllers cannot adjust to dynamic driving conditions, different tariffs, or any other unanticipated anomalies. In contrast, Safe-RL:

- acquires best charging/discharging patterns.
- smoothing present demand curves.
- avoids harmful SOC extremes
- corrects behavior in the presence of faulty sensor data.

This results in 14.1% energy consumption and 16.6% cost reductions.

Advantages Over MPC

Despite the explicit optimization and perfect model assumptions, MPC is fragile in the presence of model mismatch and anomaly conditions. Safe-RL is better than MPC in that it:

- advantages of exploration and long-horizon learning,
- is more flexible to nonlinear battery behavior,
- combines learnt reactions to anomalies through the TCN detector,
- is used to ensure constraint satisfaction with the help of the safety layer.

Safe-RL is 6.35% more energy efficient and 6.03% more cost saving than MPC, and the peak battery temperature is lower by 2.4degC, with the cycle life being 17% better.

Weaknesses of RL-no-safe

Unsafe RL or anomaly-blind is over-aggressive and constantly reaches the thermal/SOC limits and wears out the battery more quickly. It is also anomalous, and the cost penalty is 22.7%. This brings to the fore the need to:

- the safety-QP layer,
- anomaly-based fallback policy, and
- domain randomization in training.

Battery Thermal Dynamics and Health.

The thermal stability of the proposed method is better:

- Maximum temperature dropped to 41.7degC,
- Almost 50% decrease in high dT/dt events,
- 31.7% lower degradation overall.

This is essential since the temperature is the one factor that dominates most in SEI growth, plating and calendar aging.

Fault Tolerance

The TCN anomaly detector allows the controller to:

- do not make false judgments on perverted signals,
- detect replay attacks or sensor drift,
- fallback to safety and stability policies.

The outcome is a very low 3.6% cost increase in the event of faults, compared to 9.4-22.7% by other baselines.

Significance

These results indicate that the proposed system is more efficient as well as safer, more reliable, and more practical to be implemented in the real world than traditional ML, MPC, or rule-based EMS architectures.

Table 15 compares all evaluated energy management strategies across key metrics to highlight the superiority of the proposed Safe-RL system.

Table 15. Comprehensive Comparative Performance Across All Methods

Metric	Rule-Based EMS	MPC	RL-no-safe	Proposed Safe-RL	Best Improvement
Energy consumption (kWh/cycle) ↓	41.2	37.8	36.9	**35.4**	**−14.1% vs RB**
Cost per day (₹) ↓	392	348	339	**327**	**−16.6% vs RB**
Battery degradation ΔQ (Ah) ↓	5.92	4.83	4.61	**4.04**	**31.7% vs RB**
Expected cycle life ↑	1.00×	1.15×	1.19×	**1.32×**	**+32% vs RB**
Peak temperature (°C) ↓	46.2	44.1	47.4	**41.7**	**−4.5°C vs RB**
SOC violations ↓	12	3	17	**0**	**Perfect compliance**
Robustness (Cost ↑ under faults) ↓	+13.1%	+9.4%	+22.7%	**+3.6%**	**3× better vs MPC**
Heat spikes (dT/dt events) ↓	High	Medium	Very High	**Low**	Best
Energy storage efficiency η (%) ↑	71.4	77.9	79.6	**82.8**	Highest

The proposed system is far much better than all the baselines by all measures, such as efficiency, cost, degradation, safety, robustness, and thermal management.

7. Conclusion

This paper proposes a single, AI-based, anomaly-conscious Safe Reinforcement Learning energy management system of Electric Vehicles. The proposed architecture integrates an anomaly detector based on TCN, a constrained Safe-RL controller, a battery degradation model, and a fallback safety mechanism, which provides robust and real-time control in both realistic operating and fault conditions.

On 30,000+ episodes and 500 injected anomaly scenarios, the system:

- saves 14.1% of energy,
- yields 16.6% daily cost savings,
- improves battery life by ≈32%,

- reduces maximum temperature 4-5degC,
- removes all SOC/thermal violations,
- and restricts the loss of performance to only 3.6 under faults.

These findings indicate that the approach is a step forward in the state of the art of safe, efficient, and resilient EV energy management. In contrast to the current methods, it combines fault detection, health-aware optimization, grid-pricing intelligence, and safety guarantees into one architecture that can be deployed into the real world.

Future Work

This research can be extended in several ways which are promising:

1. Hardware-in-the-loop (HIL) and Real EV Deployment: Run the controller on embedded automotive platforms (NXP, NVIDIA Jetson, TI C2000) and interface to an EV BMS simulator to test in real-time.
2. Multi-Agent and Fleet-Level Co-ordination: Expand the Safe-RL system to control EV fleets, optimizing:
   - aggregated charging loads,
   - grid stability,
   - V2G/V2H interactions,
   - integration of renewable energy.
3. Deep models based on physics: Incorporate electrochemical battery models (SPM/SOHPINN) into the RL state and reward design to generate physically-grounded policies.
4. Privacy-Preserving Privacy-Anomaly Detection Federated Learning: Empower numerous EVs to enhance anomaly detector together without exchanging raw vehicle data.
5. Cybersecurity and Adversarial Robustness: Research resilience in adaptive cyber attacks, adversarial sensor manipulation and communication spoofing.
6. Explain ability and Certification: Explainable RL modules and certifiable safety barriers must be developed to meet regulatory demands (ISO 26262, UL 4600).
7. Increased Stakeholder Dashboards: include predictive analytics (e.g., battery RUL prediction, thermal risk alarm, anomaly heatmap) to consumers, utilities, regulators, and fleet operators.

References

1. S. A. Chaudhary, A. Kumar, A. K. Choudhary, and R. Kumar, “A Recommendation System for Electric Vehicle Users Based on Restricted Boltzmann Machine and Water-Wheel Plant Algorithm,” IEEE Access, vol. 11, pp. 118739–118758, 2023, doi: 10.1109/ACCESS.2023.3325412.
2. M. Dabbaghjamanesh, A. Kavousi-Fard, and M. S. Javadi, “Detection of Lying Electric Vehicles in Charging Coordination Using Deep Learning,” IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 12, pp. 5136–5147, Dec. 2020, doi: 10.1109/TITS.2019.2945622.
3. K. Zhang, C. Yu, and L. Xu, “Online Data-Driven Energy Management of a Hybrid Electric Vehicle Using Model-Based Reinforcement Learning,” IEEE Transactions on Vehicular Technology, vol. 69, no. 8, pp. 8200–8211, Aug. 2020, doi: 10.1109/TVT.2020.2997831.
4. K. Zhang, L. Xu, and C. Yu, “Energy Management Strategy of Fuel Cell Electric Vehicles Using Model-Based Reinforcement Learning with Data-Driven Model Update,” IEEE Transactions on Vehicular Technology, vol. 70, no. 6, pp. 5230–5240, Jun. 2021, doi: 10.1109/TVT.2021.3076281.
5. Y. Zhang, Y. Zhang, X. Zhang, and L. Xu, “Control Map Generation Strategy for Hybrid Electric Vehicles Based on Machine Learning with Energy Optimization,” IEEE Access, vol. 10, pp. 62292–62303, 2022, doi: 10.1109/ACCESS.2022.3180343
6. D. Chatterjee, P. K. Biswas, C. Sain, A. Roy, and F. Ahmad, “Efficient Energy Management Strategy for Fuel Cell Hybrid Electric Vehicles Using Classifier Fusion Technique,” IEEE Access, vol. 11, pp. 97135–97145, Sept. 2023, doi: 10.1109/ACCESS.2023.3312618.
7. U. Demir, G. Akgun, M. C. Akuner, M. Pourkarimi, O. Akgun, and T. C. Akinci, “An Innovative Approach to Electrical Motor Geometry Generation Using Machine Learning and Image Processing Techniques,” IEEE Access, vol. 11, pp. 48651–48661, May 2023, doi: 10.1109/ACCESS.2023.3276885.
8. V. C. Tella, M. Alzayed, and H. Chaoui, “A Comprehensive Review of Energy Management Strategies in Hybrid Electric Vehicles: Comparative Analysis and Challenges,” IEEE Access, vol. 12, pp. 181858–181872, Dec. 2024, doi: 10.1109/ACCESS.2024.3509737.
9. Q. M. Khalaf et al., “Advanced Nonlinear Optimization Techniques in SWIPT: A Review,” IEEE Access, vol. 12, pp. 192743–192760, Dec. 2024, doi: 10.1109/ACCESS.2024.3519522
10. S. Dey, M. Abomhara, K. El Fergougui, and M. Y. Saleem, “On TinyML and Cybersecurity: Electric Vehicle Charging Infrastructure Use Case,” IEEE Access, vol. 12, pp. 148871–148883, 2024. doi: 10.1109/ACCESS. 2024.3437192.

11. D. K. Chaturvedi, A. L. Sharma, and S. Yadav, "Development and Comparison of Rule- and Machine Learning-Based EMS for HESS Providing Grid Services," Energies, vol. 17, no. 3, pp. 1213–1227, 2024. doi: 10.3390/en17031213.
12. S. Kumar, A. Gupta, and R. K. Singh, "Battery State of Charge and State of Health Estimation Using a New Hybrid Deep Neural Network Approach," IEEE Transactions on Transportation Electrification, vol. 10, no. 2, pp. 2509–2520, 2024. doi: 10.1109/TTE.2024. 024.3488596.
13. P. R. Sharma and V. Patel, "Advancements in AI-Powered Electric Vehicle Routing: Multi-Constraint Optimization and Infrastructure Integration—A Survey," Renewable and Sustainable Energy Reviews, vol. 194, no. 2, p. 114212, 2025. doi: 10.1016/j.rser.2025.114212.
14. M. Singh and N. Al-Mutairi, "A Hybrid Multi-Criteria Decision-Making and Multi-Objective Framework for Optimal Sizing of PV-Powered EV Charging Station With Battery Storage," Sustainability, vol. 17, no. 5, pp. 4221–4239, 2025. doi: 10.3390/su17054221.
15. L. Zhang, W. Zhou, and X. Li, "Secondary Hybrid Decomposition Strategy for Wind Power Prediction Using Long Short-Term Memory With Crisscross Optimization," Energy Conversion and Management, vol. 308, p. 119871, 2025. doi: 10.1016/j.enconman.2025.119871.