SSRG International Journal of Electrical and Electronics Engineering
ISSN: 2348-8379/ https://doi.org/10.14445/23488379/IJEEE-V13I10P110

Volume 13 Issue 10, 134-159, October 2026
© 2026 Seventh Sense Research Group®

Original Article

Aging-Aware Attention-Driven Transformer Reinforcement Learning Framework for Sustainable Hybrid Battery Energy Management in Electric Vehicles

P.G. Raghavendran[1], V. Hamsadhwani[2]

[1,2]*Department of Electrical and Electronics Engineering, Periyar Maniammai Institute of Science and Technology, Vallam, Thanjavur, Tamil Nadu, India.*

[1]*Corresponding Author : ragaa.endiran33@gmail.com*

Received: 29 July 2026 Revised: 08 September 2026 Accepted: 01 October 2026 Published: 28 October 2026

Abstract *- Battery degradation is another important constraint of the hybrid electric vehicle systems, where the standard approach to energy management focuses on short-term fuel efficiency and does not consider the long-term health of the battery. In this paper, a proposal has been made on an Aging-Aware Attention-Driven Transformer Reinforcement Learning (AAT-DRL) framework of intelligent hybrid battery energy management. Transformer-based temporal feature encoding, along with actor-critic reinforcement learning architecture, is combined in the model to encode patterns that are sensitive to degradation such as temperature, current stress and depth-of-discharge. The multi-objective reward is developed with an aim to optimize the energy efficiency, thermal safety and battery life. The framework is tested at standard driving cycles (UDDS, NEDC and WLTP). In the results, it is shown that the degradation index has reduced by 38%, the SOC has become more stable and thermal fluctuations have been reduced, with a peak efficiency of 95.3%, which is a better performance than the state-of-the-art DRL methods. The solution offers a healthy, scalable and battery management system for the next generation based on the proposed approach.*

Keywords *- Battery degradation mitigation, Electric Vehicles, Energy management strategy, Hybrid battery systems, Reinforcement learning, Transformer architecture.*

1. Introduction

The growing popularity of Electric Vehicles (EVs) could be attributed to the rise in environmental issues, strict emission policies and the development of energy storage technologies. Of them, lithium-ion battery systems are key elements to convert transportation into efficiency and sustainability [31]. Nevertheless, battery degradation is another critical issue, which influences considerably the life cycle, security and cost-effectiveness of EVs [1, 2] considerably. Operating conditions, including current stress, temperature changes and depth-of-discharge patterns, have a strong impact on such phenomena of degradation as capacity fading, rise in internal resistance and thermal instability [3].

In order to overcome those difficulties, hybrid battery systems, which consist of a primary battery and an auxiliary energy storage unit, have been proposed to improve the performance and the reliability [32]. The hybrid configurations are able to reduce the overall stress on the individual batteries by spreading power demand among several storage elements and enhance the overall system efficiency [4, 5]. However, the functionality of hybrid systems is heavily influenced by how the system is designed in terms of energy management, which dictates the manner in which power is shared among the various sources in the face of different operating conditions.

The conventional energy management techniques, which include the rule-based controls and the equivalent consumption minimization, are based on the heuristic definition and simplified models [6]. Although these methods are computationally efficient, they are not flexible enough and do not consider complex and time-dependent system dynamics [33]. What is more important, the majority of traditional approaches are oriented on the short-term energy efficiency and do not pay much attention to the long-term battery health parameters, including State-of-Health (SOH) and thermal stress [7]. This drawback causes the suboptimal performance and the premature battery aging in actual practice.

Reinforcement Learning (RL) is one of the promising intelligent energy management techniques in EVs in recent years because of its mechanism of acquiring optimal control policies through its interaction with the environment [8, 9]. Deep Reinforcement Learning (DRL) algorithms, such as


This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Deep Q-Networks (DQN), Deep Deterministic Policy Gradient (DDPG) and Proximal Policy Optimization (PPO), have shown great advances in adaptive power allocation and system efficiency [10]. Nevertheless, the majority of available DRL-based methods mostly rely on instantaneous state representations, and they do not explicitly introduce degradation-aware features into a decision-making process [11].

At the same time, Transformer architectures have been receiving considerable attention in sequence modelling tasks because they are able to induce long-range dependencies with the help of self-attention mechanisms [12]. In contrast to recurrent neural networks, Transformers allow the temporal processing of data in parallel and can respect complicated time-dependent relations [34]. Their usage has been investigated more recently in battery health prediction and time-series analysis with promising outcomes to capture degradation patterns [13]. Although these developments have occurred, the application of transformer-based temporal modeling coupled with reinforcement learning on energy management is not well studied [35].

In hybrid battery electric vehicle systems, efficient power allocation is still quite challenging despite the great advances in intelligent energy management, while maintaining long-term battery health [14, 15]. Typical rule-based and optimization-based approaches rely on fixed rules and simple battery models and are thus not very adaptable to rapidly varying driving conditions. While deep reinforcement learning methods offer adaptive decision-making, current methods mainly focus on the current or short-term battery state. Therefore, effects of degradation which can be caused by repeated current stresses, temperature changes, fluctuations in state of charge and depth of discharge patterns are not adequately captured in the learning of control policies.

Transformer-based architectures have proven to be an effective way of modelling long-range temporal dependency and have demonstrated promising results in various battery state estimation, health prediction and degradation-related time-series analysis tasks. Their use has, however, been largely in prediction and estimation, and not in closed-loop energy management. Meanwhile, most degradation-aware reinforcement learning schemes typically do not explicitly learn temporal relationships between degradation-sensitive battery variables but rather use simplified health metrics. Thus, there is a clear research gap for developing a unified energy management framework that features a long-term temporal degradation representation and adaptively allocates power based on reinforcement learning, while considering energy efficiency, thermal safety and battery longevity simultaneously.

Due to these issues, this paper proposes an Aging-Aware Attention Transformer Reinforcement Learning (AAT-DRL) framework for hybrid battery energy management of electric vehicles. The proposed framework uses a Transformer-based temporal encoder to process a sequence of battery operating states and looks for degradation-sensitive patterns corresponding to current stress, temperature, state-of-charge and depth-of-discharge. These health-aware temporal representations are directly used to train an actor-critic reinforcement learning controller for continuous power allocation between the primary and auxiliary energy-storage units. A multi-objective reward function is also introduced, by combining energy efficiency, degradation mitigation and thermal stability into a single control objective. The proposed approach thus directly relates to the adaptive energy-management decision-making without separating out battery degradation estimation and power control.

The major contributions of this work are as follows:

1. Investigation and design of an Aging-Aware Attention Transformer Reinforcement Learning (AAT-DRL) framework for Hybrid Battery Energy Management System (HBEMS) in EVs to optimize both energy efficiency and battery health, while learning temporal features and applying adaptive control to address battery issues.
2. Proposed a novel Transformer-based temporal attention mechanism to model long-term degradation-sensitive patterns, such as temperature changes, current stress, state-of-charge fluctuations and depth-of-discharge dynamics, enhancing the degradation-aware state representation.
3. Multi-objective reward design to balance energy efficiency, battery degradation mitigation and thermal stability, facilitating the balanced decision-making and sustainable battery operation under dynamic driving conditions.
4. Adaptive power allocation between primary and auxiliary energy storage units using an attention-driven degradation monitoring system, while maintaining the long-term life of the Batteries.
5. Comprehensive evaluation under standard driving cycles (UDDS, NEDC and WLTP) showing substantial benefits with respect to degradation reduction, thermal management, state-of-charge stability and overall energy efficiency when compared to conventional and state-of-the-art reinforcement learning-based energy management strategies.
6. Comprehensive ablation, statistical significance and computational validation of the proposed framework for next-generation intelligent battery management system, its effectiveness, robustness and practical feasibility.

The rest of this paper is structured in the following way. Section 2 reviews the literature in detail related to hybrid battery energy management, battery degradation modeling, reinforcement learning and Transformer-based learning

approaches. Section 3 presents the system model, which consists of a hybrid battery architecture, battery state variables, electrical and thermal dynamics and degradation formulation. Section 4 outlines the proposed Aging-Aware Attention Transformer Reinforcement Learning (AAT-DRL) framework, a temporal encoder based on the Transformer, the attention-based degradation monitoring mechanism, the reinforcement learning formulation and the training strategy. The experimental setup, simulation environment, battery configuration, driving cycles and baseline methods used in the performance evaluation are described in Section 5. The experimental results, comparative analysis, ablation, computational complexity assessment and statistical validation are discussed in Section 6. Finally, Section 7 summarizes the paper and suggests avenues for future research and implementation of the proposed framework to address practical problems.

2. Related Work

New developments in energy management on hybrid electric vehicles have placed increased emphasis on the incorporation of intelligent control systems towards maximizing efficiency as well as battery life. Deep Reinforcement Learning (DRL) has become one of the most popular paradigms because it can be used to learn the optimal policies as a result of interacting with a dynamic environment. Preliminary designs have proven that DRL can be highly useful in hybrid battery systems, with adaptive control strategies far much more efficient and flexible than conventional rule-based methods [16]. The use of more sophisticated actor-critic models has also made it possible to maintain control over power distribution, continuously enhancing the stability of systems and their smooth operation [17]. A number of studies have been conducted on hierarchical and multi-level reinforcement learning architectures to optimize energy use and battery life together. Such strategies make use of organized learning models to trade off various goals, resulting in high performance in battery-ultracapacitor systems [18]. Additionally, the survey studies demonstrate the accelerated development of the DRL-based energy management strategies, which have demonstrated the capability to consider nonlinear system dynamics and uncertainties, yet the issues involving stability and generalization remain [19].

In order to manage the issue of battery aging, recent literature has integrated battery degradation-awareness control systems in energy management systems. As an example, reinforcement learning models have been developed to have a battery life constraint that allows real-time decision-making to minimize degradation without affecting efficiency [20]. Equally, imitation learning and hybrid learning methods are suggested as a way of increasing the efficiency of learning and getting better convergence of policies, especially in complex driving conditions [21]. The comparative studies with different sets of DRL algorithms further substantiate the fact that although advanced algorithms, including TD3 and PPO, are more stable, they lack the explicit study of long-term degradation processes [10].

The combination of data-based and transfer learning methods has also been of interest, giving models the ability to respond to the changing driving conditions and enhance the robustness when subjected to diverse working conditions [22]. Moreover, energy management schemes, which utilize health-conscious soft actor-critic algorithms, have also shown that battery health metrics should be explicit in the control system [23]. In recent works, thermal characteristics have also been taken into consideration, and energy management strategies are elaborated to control temperature in addition to power distribution, increasing safety and battery life [12]. It has also been investigated that model predictive control that has learning techniques can be used to get more predictive and accurate energy control, specifically in fuel cell hybrid systems [24]. Also, reinforcement learning methods, which include driving condition classification, have demonstrated enhanced flexibility, which allows the system to adapt well to various driving behaviours [16]. These changes suggest the increasing tendency towards smart, situation-ready energy management solutions.

Along with the developments of reinforcement learning, Transformer-based architectures have been launched to estimate battery state-of-health and model time series. The models take advantage of the attention processes to learn long-range dependencies of battery data so that they can make predictions about the degradation trends more accurately [25]. Transformer modeling and incremental capacity analysis have also contributed to health estimation accuracy, which indicates the future of attention-focused methods to battery control products [26]. The methods are, however, only applicable and mainly applicable in prediction tasks with little or no active use in real-time control systems.

Recent research works have been more concerned with integrating health-aware and predictive models with reinforcement learning to enhance the performance of the system. As an illustration, imitation reinforcement learning models have been made to improve efficiency in learning, yet at optimum control performance of hybrid energy storage systems [27]. On the same note, novel DRA methods, which combine predictive modules and health-constrained strategies, have been shown to deliver a better energy management performance to fuel cell hybrid vehicles [28]. All these practices highlight the need to integrate predictive intelligence and adaptive control in operational energy management. In spite of such developments, the current methods are based on either instant state representation or on a restricted ability to model time. Even though the attention-based models have demonstrated the potential of degradation patterns, little has been done to integrate reinforcement learning with the attention-based model in energy management. In addition, the

current practices tend to focus on efficiency, thermal behaviour and degradation as independent objectives instead of being considered as a single objective in an optimization framework.

Model Predictive Control (MPC) has emerged as a widely adopted approach for battery energy management because of its ability to optimize control actions over a finite prediction horizon while explicitly considering system constraints. Several MPC-based strategies have demonstrated effective trade-offs between energy efficiency, battery degradation and thermal management by utilizing predictive models of battery dynamics. However, the performance of MPC is highly dependent on the accuracy of the underlying system model and may deteriorate under uncertain operating conditions, parameter variations and complex degradation behaviours. Furthermore, online solutions of optimization problems may be very demanding in real-time applications. In contrast, the proposed AAT-DRL framework learns control policies without relying on an explicit battery model and uses Transformer-based temporal attention to model long-term degradation patterns, through actual experience with the environment. This allows adaptive decision-making under these changing operating conditions with effective long-term battery health management.

There have been recent studies on intelligent control and optimization strategies for battery energy management in different energy and transportation applications [38]. More recent and sophisticated predictive control techniques have been used to enhance energy allocation and utilization in hybrid energy storage systems. Various systems have been investigated to optimize the battery life and reliability, including battery state estimation, degradation prediction and thermal management systems [39]. Furthermore, machine learning and deep learning models have been effectively used for battery health monitoring, fault diagnosis and energy consumption prediction. The transformer architectures have recently shown great potential in modeling long-term temporal dynamics for battery state prediction and energy forecasting applications [40]. In addition, reinforcement learning algorithms have been demonstrated to perform optimally with respect to adaptive energy management through an iterative process of determining an optimal control policy by interacting with the environment. Although significant progress has been achieved, most of the current research efforts are on either battery health prediction or energy management separately [41]. There have been few studies that have explored the combination of Transformer-based temporal attention mechanisms and degradation-aware RL for simultaneously optimizing energy efficiency, thermal stability and battery longevity [42]. This restriction raises the need to unify the temporal degradation modelling and adaptive control in a single intelligent battery management architecture, which is the objective of the proposed AAT-DRL framework.

The application of Transformer architectures has been shown to be effective in various recent contributions for the management of smart grids, electric vehicle energy optimization, load forecasting and battery state estimation, among others and the potential to capture long-range temporal dependencies in complex energy systems. Moreover, advances in model compression, light-weight attention mechanisms and hardware acceleration have greatly decreased the inference latency for Transformers, making their use in near real-time control applications more feasible. Therefore, the present work does not claim that Transformers have never been applied to energy management or real-time decision-making. Instead, the research gap lies in the limited exploration of Transformer-based temporal attention within degradation-aware reinforcement learning frameworks for battery energy management. The proposed AAT-DRL framework extends existing studies by integrating aging-aware temporal feature learning with reinforcement learning-based control, enabling simultaneous optimization of energy efficiency, thermal stability and long-term battery health under dynamic operating conditions.

2.1. Research Gap

Based on the analysis above, it is possible to list several critical gaps in research. First, currently, there are reinforcement learning-based energy management strategies that use instantaneous or short-term state representations and hence are not capable of capturing any long-term degradation dynamics. Second, even though battery aging and thermal considerations have been included in some studies, they are not usually considered as a whole but are rather thought of separately as a multi-objective framework. Third, Transformer-based models have shown great power in the features of time and their degradation prediction, but they have not been adequately applied to real-time control, and most have been used in offline estimation. Lastly, no energy management frameworks integrate temporal attention mechanisms, degradation-sensitive modeling and adaptive reinforcement learning control to deliver the efficiency-battery health trade-off optimization at the same time.

Recent studies have explored Transformer-based architectures for energy forecasting, battery state estimation and energy management, while others have investigated degradation-aware control strategies and reinforcement learning-based energy optimization. However, these approaches typically address these components independently or integrate only a subset of them. Transformer-based methods primarily focus on temporal feature extraction and prediction tasks, whereas degradation-aware frameworks often rely on model-based optimization or conventional machine learning techniques. Similarly, reinforcement learning approaches generally optimize energy efficiency and operational performance without explicitly incorporating attention-driven temporal degradation representations. This paper, instead, does not assert that there is no existing related work; it only

finds that there are limited attempts to integrate Transformer-based temporal attention, degradation-aware state representation and adaptive reinforcement learning into a single battery energy management architecture. The proposed AAT-DRL framework aims to fill this gap by combining these complementary elements into a unified framework for long-term battery health preservation and energy optimization in a unified decision-making framework.

While some recent studies have considered health-aware deep reinforcement learning for battery energy management, most of them use instantaneous information or simplified health indicators to represent the battery, which fail to represent long-term degradation behaviour. Furthermore, the degradation prediction and energy management are generally considered as two separated processes, which lowers the effectiveness of real-time control decisions. To overcome these limitations, the proposed Aging-Aware Attention Transformer Reinforcement Learning (AAT-DRL) framework is proposed, which incorporates the Transformer-based temporal attention modelling directly into the reinforcement learning control loop. This allows for the separation of degradation-sensitive features of temperature, current stress, state-of-charge and depth-of-discharge dynamics. Moreover, with a multi-objective reward function, the parameters that maximize energy efficiency, thermal stability and battery lifetime are optimized, thus providing a unified and adaptive energy management strategy which improves the sustainability and reliability of hybrid battery systems in electric vehicles.

The developed strategies are being translated into adaptive deep reinforcement learning solutions that could be applied in nonlinear operating conditions and dynamic power loads. While actor-critic, hierarchical, health-aware, imitation-learning, and predictive control approaches have benefited the power allocation, convergence behaviour, thermal regulation, and battery-health preservation, many of these methods are predominantly based on instantaneous or short-term state information and only offer partial representation of the long-term degradation behaviour.

Experiments that include battery aging and thermal limits have highlighted the need for explicit battery health considerations in energy management, and Transformer-based ones have been proven to have significant potential for modelling long-term temporal correlations in battery state and predicting battery degradation. However, Transformer models have been applied primarily for estimation and forecasting, and to a comparatively small extent to adaptive closed-loop energy management. All of these results suggest a need for a unified model that not only includes temporal degradation modeling but also includes attention-based critical battery condition identification and reinforcement learning-based power allocation, while also incorporating energy efficiency, thermal stability, and battery longevity.

2.2. Novelty of Proposed Work

Prior studies have shown that there are some significant developments that have been achieved in the field of intelligent battery management of energy. Fixed rule-based methods have been modified by the conventional and actor-critic DRL methods to achieve better results in adaptive power allocation than the fixed methods [16-19]. Degradation-aware and health-conscious reinforcement learning methods have also been shown to be able to utilize information on battery health to inform energy-management decisions [12, 20, 23]. The enhancement of adaptability to different operating conditions and driving conditions has also been achieved with the use of Predictive and Hybrid learning methods [21, 22, 24, 27, 30, 31]. At the same time, Transformer-based models have exhibited good performance in the long-range temporal dependency learning in battery state-of-health estimation and time-series modelling of battery degradation [25, 26]. However, most of these research directions focus on one of the three aspects of adaptive control, battery-health concern, or temporal degradation modelling in isolation or in a partial manner. The novelty of the proposed AAT-DRL framework is thus not only the fusion of a Transformer but also reinforcement learning. It is rather in the direct coupling of degradation-sensitive temporal representation in the closed-loop energy-management decision process. The Transformer encoder receives not just instantaneous battery observations, but also the history of changing states of the battery, whereas the attention mechanism highlights the battery operating time periods where the current stress is high, temperature is high, and SOC variations are high, and deep discharge occurs. These temporal features, which are all health-related, are directly fed to the actor-critic controller to enable the power allocation policy to consider not only the power demand of the vehicle but also the history of battery operation.

The second is the reward formulation, which is multi-objective with an aging-awareness. The typical approaches that are used today are based on DRL and tend to focus on energy efficiency and operational performance, while battery degradation and/or thermal behaviour can be considered as additional constraints or as an additional performance objective. The proposed framework includes energy efficiency and degradation mitigation/thermal safety within the framework of policy learning. Therefore, degradation information over time can be utilized not only for offline prediction of the battery health but also for the control decision.

The proposed framework is thus built on the back of the following: (i) Transformer-based long-term temporal modelling, (ii) attention-guided degradation-sensitive feature identification, (iii) aging-aware multi-objective reward optimization, and (iv) continuous actor-critic power allocation between the primary and auxiliary storage units. This single expression directly relates the battery health representation to adaptive energy-management decisions at varying driving

conditions. This closed-loop interaction between the battery states monitored, the Transformer encoder, reinforcement learning controller and hybrid battery system is explicitly implemented in the existing manuscript.

3. System Model

3.1. Hybrid Battery Architecture

The materials and methods section should contain sufficient detail so that all procedures can be repeated. It may be divided into headed subsections if several methods are described. The hybrid battery energy storage system that is viewed in this paper comprises a main lithium-ion battery pack and an auxiliary energy storage device that could be a secondary battery or super capacitor. The main battery will supply the minimum energy requirement of the electric car, and the auxiliary unit will assist in the momentary peak power needs like accelerating and reclaimed braking. This dual-source design allows the distribution of loads to be improved and stress on each battery component to be decreased. Hybrid storage system is connected with the electric vehicle power train by a bidirectional DC-DC converter that controls the movement of power between the sources of energy and the load. With the Energy Management System (EMS), the ideal divide of the power among the two storage units is decided in real-time under system conditions and driving circumstances.

The aggregate power requirement of the vehicle at time step is stated as:

$$P_{load}(t) = P_{b1}(t) + P_{b2}(t) \quad (1)$$

Where $P_{b1}(t)$ and $P_{b2}(t)$ denotes the power delivered by the primary and auxiliary batteries, respectively. This equation will make sure the total production of the two sources will meet the immediate demand of the vehicle.

3.2. Battery State Variables

The EMS monitors a set of system and battery states in order to make intelligent decisions. These states acquire instantaneous operating conditions as well as long-term degradation indicators.

The state vector of the system is given as:

$$S_t = [SOC_1, SOC_2, SOH_1, SOH_2, T_1, T_2, I_t, V_t, DOD_t] \quad (2)$$

SOC_i is the State of charge of battery i, SOH_i is the State of health of battery i,$T_i$is the Temperature of battery i, I_t is the Current drawn at time t, V_t is the Terminal voltage and DOD_t is the Depth of discharge.

The State of Charge (SOC) evolves according to:

$$SOC(t+1) = SOC(t) - \frac{I_t \Delta t}{C} \quad (3)$$

C represents the battery capacity, $\Delta t$is the sampling interval. The use of SOH, temperature and DOD models the controller to the knowledge that there are parameters that are sensitive to degradation, which are essential to energy management based on aging.

3.3. Electrical and Thermal Model

An Equivalent Circuit Model (ECM) of the battery comprises an internal resistance, and an open-circuit voltage source is used to model the dynamics of the battery. The terminal voltage is expressed as:

$$V_t = V_{oc}(SOC) - I_t R_{int} \quad (4)$$

Where,V_{oc}: Open-circuit voltage and R_{int}: Internal resistance

Battery temperature dynamics are modeled as:

$$T(t+1) = T(t) + \alpha I_t^2 R_{int} - \beta(T(t) - T_{amb}) \quad (5)$$

Where the heat generation coefficient is denoted by α and the cooling coefficient is denoted by β. The ambient temperature is denoted by T_{amb}.

3.4. Degradation Modeling

Various factors such as current stress, depth-of-discharge and temperature determine battery degradation. The model of degradation cost at time t is:

$$D_t = \lambda_1 \mid I_t \mid + \lambda_2 DOD_t + \lambda_3 T_t \quad (6)$$

Where $\lambda_1, \lambda_2, \lambda_3$are weighting coefficients that denote the proportion of contribution of each factor to the battery aging. This degradation model has been integrated into the energy management model so that the control decisions do put into consideration long-term battery health. The degradation cost function is designed as a computationally efficient aging indicator for reinforcement learning-based energy management rather than a detailed electrochemical aging model. The coefficients $\lambda 1$, $\lambda 2$ and $\lambda 3$ represent the relative contributions of current stress, temperature exposure and depth-of-discharge to battery degradation, respectively and are selected based on findings reported in battery aging literature.

The model captures the widely recognized impact of these operating factors on capacity fade and resistance growth while maintaining low computational complexity for real-time control. It does not explicitly contain electrochemical phenomena like SEI formation, lithium plating or cycle-dependent degradation phenomena, but it is used as a degradation-aware signal for control optimization. A physics-based aging models and a real-world battery degradation

dataset will be incorporated into future work to further enhance the accuracy and practicality of the modelling. The degradation model used in this study is formulated as a computationally efficient way of representing the most important factors impacting battery degradation-current stress, temperature exposure and depth-of-discharge. These have been commonly identified in battery ageing literature as the major factors leading to capacity loss and internal resistance increase. The concept of the proposed model is not to give an electrochemical level prediction of age, but to give a degradation-aware signal that can effectively be used in the reinforcement learning framework for real-time energy management. In order to make the model a more practical application, the operating range and model parameters were chosen based on battery aging data that has been recently reported in the literature. While the simulation results demonstrate the effectiveness of the proposed framework in reducing degradation-related stress indicators, future work will focus on validating the degradation formulation using large-scale real-world battery aging datasets and laboratory cycling experiments to further enhance model fidelity and generalization.

4. Proposed Work

4.1. Overall Architecture of the Proposed Framework

Figure 1 shows the overall scheme of the proposed Aging-Aware Attention Transformer Reinforcement Learning (AAT-DRL) model of the hybrid battery energy management of electric vehicles.

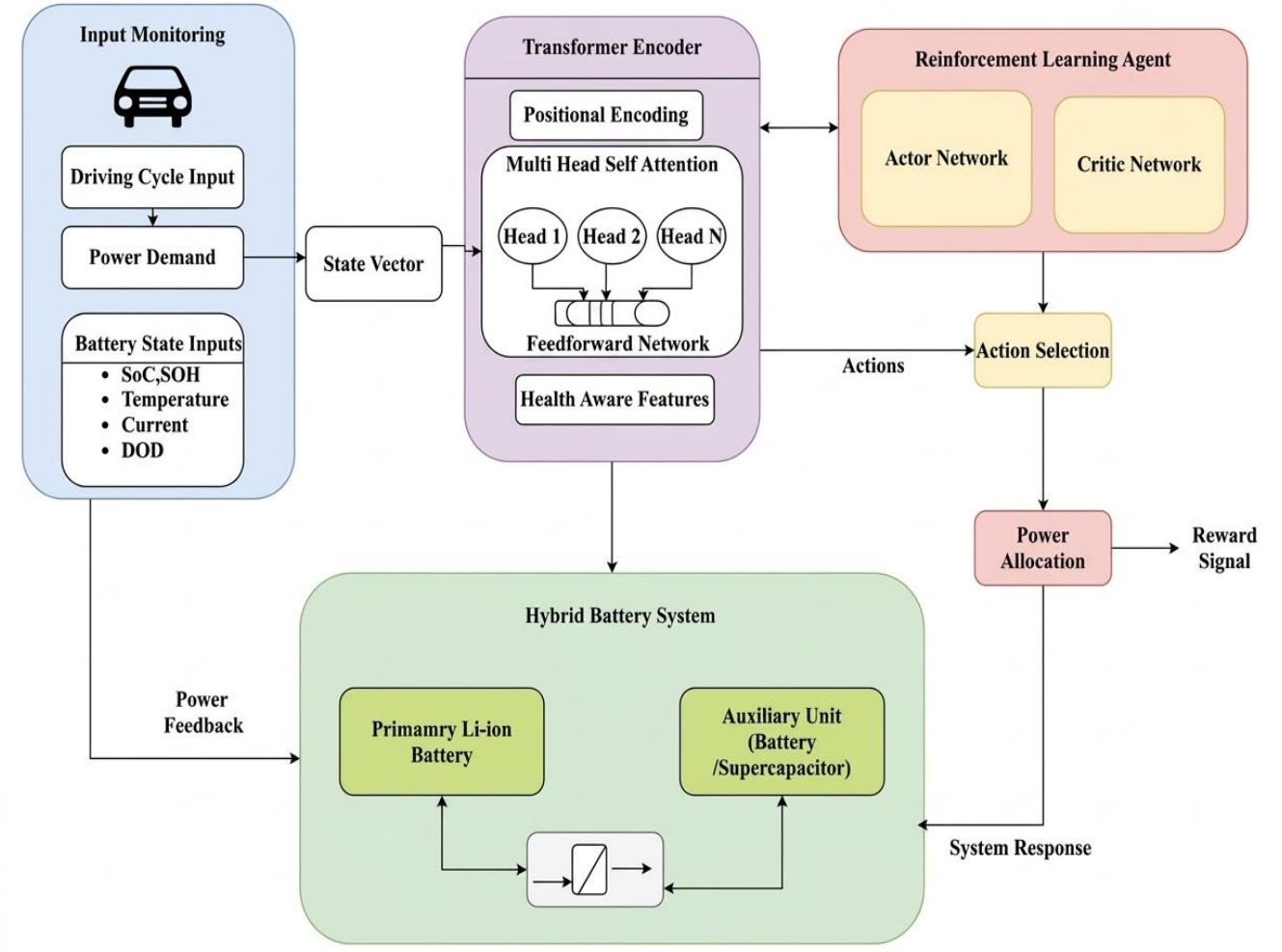


Fig. 1 Overall architecture of the proposed Aging-Aware Attention Transformer Reinforcement Learning framework for hybrid battery energy management

The suggested framework is an intelligent control system in the form of a closed loop where data-driven representation learning is combined with adaptation in decision-making. It has four modules that are highly coupled, which are input and system monitoring, Transformer-based feature extraction, reinforcement learning decision engine and hybrid battery execution layer.

The process initiates with driving cycle input, which creates time-varying car power requirements under the

realistic operating conditions like urban, highway and mixed driving conditions. These power requirements are provided to the hybrid battery system, which has a primary lithium-ion battery and auxiliary energy storage unit.

The auxiliary unit helps in managing the sporadic loads and peak power needs, hence lowering the pressure on the primary battery. At the same time, a couple of important battery state variables, such as State-of-Charge (SOC), State-of-Health (SOH), Temperature, Current and Depth-of-Discharge (DOD) are constantly monitored. These parameters are combined to create a state vector, which is used as the input to the learning structure. Then a Transformer encoder is applied to the state vector, giving the opportunity to the multi-head self-attention to capture the time-related attention and stress the degradation-sensitive structure. The feature representations that are extracted are called health-aware embeddings that are sent to the reinforcement learning agent, which selects the best power allocation strategy. The RL agent deployed on an actor-critic architecture gives an output of a continuous control action, which is the power share between the two battery sources. It is done to the hybrid battery system, and the evaluation of the system response is done with the help of a multi-objective reward function which takes into consideration efficiency, degradation and thermal safety. The feedback channel allows the agent to keep on improving its policy according to the observed results. This unified design removes the possibility of only responding to the immediate energy requirements of the control strategy, but also results in an active reduction in the long-term battery degradation.

4.2. Transformer-based State Encoding

The second aspect of the proposed framework, after the system-level architecture, is the Transformer-based state encoder that has the task of obtain meaningful temporal features out of the battery state sequence.

The Transformer is capable of modeling long-range dependencies effectively because it uses a self-attention mechanism to compute output based on the input data, unlike traditional recurrent neural networks that use data in sequence. This is especially crucial in battery systems, where cumulative historical behaviour affects the degradation, but not necessarily only instantaneous conditions.

Defining the input sequence of battery states during a time period n as:

$$X = [S_{t-n}, S_{t-n+1}, \dots, S_t] \quad (7)$$

The variables in each state vector S_t are SOC, SOH, temperature and current (as in the definitions used in Section 3). Transformer: this sequence is processed by the Transformer with scaled dot-product attention:

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (8)$$

With Q, K and V the query, key and value matrices as the result of the linear projections of the input sequence and d k the dimension of the key vectors.

Multi-head attention is used to enhance the representational capacity:

$$MHA(Q, K, V) = Concat(head_1, \dots, head_h)W^O \quad (9)$$

The mechanism allows the model to focus on various agendas of the battery at the same time, including thermal dynamics, current variations and discharge patterns.

The result of the Transformer encoder is a small representation:

$$H_t = Transformer(X) \quad (10)$$

Which estimates both short and long-run fluctuations in battery operations.

4.3. Attention-based Degradation Monitoring

Based on the Transformer encoder, the suggested framework features an attention-based degradation monitoring procedure, where the model focuses on key importance to any state relevant to aging. The attention process gives weights of importance to various time steps of the input sequence, which enable the model to place emphasis on those states that play a significant role in battery degradation. The attention weights are calculated as follows:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_k \exp(e_{ik})} \quad (11)$$

and e_{ij}represents is the score of relevance between the i-th and$j$-th states.

The model, through this mechanism, focuses on:

- Elevated current stress periods.
- The areas of high temperatures.
- Deep discharge cycles
- Rapid SOC variations

This selective attention allows the framework to pick up more precise patterns of degradation and, therefore, enhance the effectiveness of future decision-making.

4.4. Reinforcement Learning Formulation

After getting the degradation-conscious feature representation, the energy management problem is modeled as a Markov Decision Process (MDP).

$$M = (S, A, P, R, \gamma) \quad (12)$$

The state S corresponds to the representation. H_t that is Transformer-encoded, and the action A represents the decision on the power allocation. P describes the system dynamics, and R is the reward function that is intended to provide the learning process. The discount factor γ specifies the significance of future reward. The agent monitors the present state at every time step and chooses an action that maximizes the probability of the cumulative reward, which is expected.

4.5. Action Space Design

The action space can be defined as a continuous control variable, which is the distribution of power between the main and auxiliary batteries:

$$A_t = [P_{b1}(t), P_{b2}(t)] \quad (13)$$

subject to the constraint:

$$P_{b1}(t) + P_{b2}(t) = P_{load}(t) \quad (14)$$

The given formulation guarantees that the power requirement in the vehicle is fulfilled completely and gives flexibility in the way the load is allocated among the hybrid energy storage.

4.6. Multi-Objective Reward Function

A multi-objective reward function is formulated to provide a compromise between performance and longevity. The reward incorporates three essential goals, namely, energy efficiency, minimization of degradation and thermal safety.

$$R_t = w_1\eta_t - w_2 D_t - w_3 T_t \quad (15)$$

where:

- η_t: energy efficiency
- D_t: degradation cost
- T_t: temperature

Energy efficiency is computed as:

$$\eta_t = \frac{P_{out}}{P_{in}} \quad (16)$$

The model used in Section 3 to derive the degradation cost D_ti is also consistent between system modeling and control design.

The reward function was designed to prioritize three conflicting goals - energy efficiency, battery degradation minimization and thermal safety. The weighting coefficients were determined to make sure that no single objective would dominate the learning process and keep the reinforcement learning agent converged. A sensitivity analysis of the reward formulation was performed to check the strength of the reward formulation by changing the weighting parameters within reasonable ranges and seeing how it affects the system's performance. The analysis showed that higher weights for degradation caused less battery stress and thermal variation while leading to a slight drop in energy efficiency, while higher weights for energy efficiency caused an increase in degradation indicators while leading to less stress and thermal variation through the battery. The results showed the successful implementation and the robustness of the proposed multi-objective optimization framework with the selected weight configuration, which resulted in the optimum combination of efficiency, duration and thermal stability for the proposed design.

4.7. Actor-Critic Learning Framework

The actor-critic reinforcement learning architecture is used to learn the control policy, which is a combination of policy-based learning and value-based learning. A simple schematic of the actor-critic learning model is depicted in Figure 2.

The action of control is created in the actor network:

$$a_t = \pi_\theta(s_t) \quad (17)$$

where the critic network analyses the action via a value function:

$$Q(s_t, a_t) \quad (18)$$

The critic network estimates the state-action value function $Q(s_t, a_t)$, which represents the expected cumulative discounted reward obtained by actinga_t in state s_t and subsequently, following the current policy.

The training objective minimizes the temporal-difference error:

$$L = \mathbb{E}[(R_t + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t))^2] \quad (19)$$

The Temporal-Difference (TD) error represents the difference between the critic network's predicted value and the actual reward obtained from the environment, together with the estimated future return. It serves as a learning signal that guides both the actor and critic networks during training. The critic uses this error to improve value estimation accuracy, while the actor updates its policy based on the feedback provided by the TD error, thereby encouraging actions that maximize long-term cumulative rewards. This formulation is consistent with the stochastic actor-critic reinforcement learning framework employed in the proposed AAT-DRL model and ensures stable policy optimization.

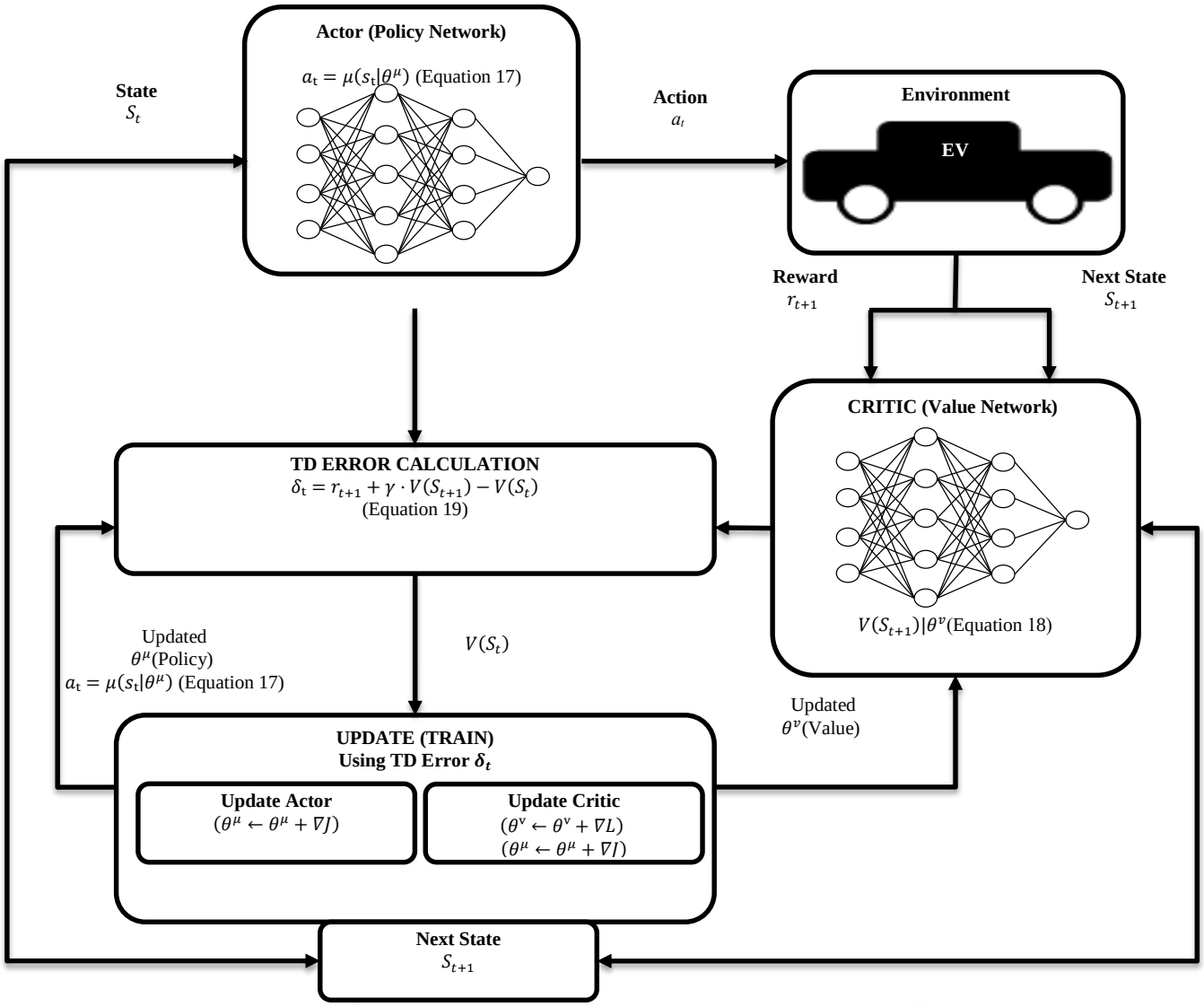


Fig. 2 Actor-Critic reinforcement learning framework illustrating state transition, action selection, Temporal-Difference (TD) error computation and simultaneous update of policy and value networks.

4.8. Training Strategy

Lastly, the model is trained by applying an iterative optimization algorithm, which is based on experience replay and mini-batch gradient descent. The policy parameters are corrected as:

$$\theta \leftarrow \theta + \alpha \nabla_\theta J(\theta) \tag{20}$$

The objective function $J(\theta)$ denotes the expected cumulative discounted reward obtained under the policy parameterized by θ. It quantifies the long-term performance of the reinforcement learning agent by evaluating how effectively the selected actions balance energy efficiency, battery health preservation and thermal stability. The policy gradient $\nabla_\theta J(\theta)$ represents the direction in which the policy parameters should be updated to maximize the expected return. During training, the actor network iteratively adjusts its parameters according to this gradient, enabling the agent to learn an optimal control policy that improves overall system performance while satisfying the multi-objective requirements of the proposed AAT-DRL framework. With $J(\theta)$ being the cumulative expected reward and α meaning the learning rate. The training process is repeated until one reaches a convergence point, and the control policy that balances energy efficiency and battery health maintenance in changing driving conditions is achieved.

The step-by-step functionality of the suggested framework is outlined in Algorithm 1. The algorithm starts by creating temporal state windows of the observed hybrid

battery variables, and these sequences are input into the Transformer encoder to obtain degradation-sensitive latent representations. These representations are provided to the actor-critic module of learning, which classifies the power allocation policy within a multi-objective reward system. The policy is continually improved through interactions with the environment to minimize battery aging and thermal stress and keep the vehicles operating efficiently through the use of energy. The training procedure was expanded to explicitly describe the policy update, value update and replay buffer operations. At each time step, the transition tuple consisting of the current state, selected action, reward, next state and terminal flag is stored in the replay buffer. During training, mini-batches are randomly sampled from this buffer to reduce temporal correlation among consecutive samples and improve learning stability. The critic network is updated by minimizing the temporal-difference loss between the predicted state-action value and the target value computed from the observed reward and discounted future estimate. The actor network is then updated using the policy gradient obtained from the critic's evaluation of the selected action, encouraging actions that maximize long-term cumulative reward. Target networks are updated through soft parameter synchronization to avoid unstable value estimates. Gradient clipping is applied during optimization to prevent exploding gradients. These detailed update rules and buffer management steps improve the reproducibility and clarity of the proposed AAT-DRL training algorithm.

Algorithm 1: Aging-Aware Attention Transformer Reinforcement Learning for Hybrid Battery Energy Management

Input:

Driving cycle sequence $\mathcal{D} = \{P_{load}(t)\}_{t=1}^{T}$,

State vector S_t,

Learning rate α, discount factor γ,

Soft update factor τ,

Number of episodes E, batch size M

Output:

Optimal policy π_θ^*

Initialization

$$\phi \sim \mathcal{N}(0,1), \theta \sim \mathcal{N}(0,1), \omega \sim \mathcal{N}(0,1)$$
$$\theta^- \leftarrow \theta, \omega^- \leftarrow \omega$$
$$\mathscr{B} \leftarrow \emptyset$$

For each episode

$$\forall e \in \{1,2,\dots,E\}$$

Initialize initial state:

$$S_0$$

For each time step

$$\forall t \in \{1,2,\dots,T\}$$

Step 1: Temporal State Construction

$$X_t = [S_{t-n+1}, S_{t-n+2}, \dots, S_t]$$

Step 2: Transformer-Based Encoding

$$Q_t = X_t W_Q, K_t = X_t W_K, V_t = X_t W_V$$
$$\text{Attn}_t = \text{Softmax}\left(\frac{Q_t K_t^{\top}}{\sqrt{d_k}}\right) V_t$$
$$H_t = \text{Concat}(head_1, head_2, \dots, head_h) W^O$$
$$Z_t = f_\phi(H_t)$$

Step 3: Action Selection

$$A_t = \pi_\theta(Z_t) = [P_{b1}(t), P_{b2}(t)]$$

Subject to:

$$P_{b1}(t) + P_{b2}(t) = P_{load}(t)$$

Step 4: System Transition

$$S_{t+1} = \mathcal{T}(S_t, A_t)$$

Step 5: Degradation Estimation

$$D_t = \lambda_1 \mid I_t \mid + \lambda_2 DOD_t + \lambda_3 T_t$$

Step 6: Efficiency Computation

$$\eta_t = \frac{P_{out}(t)}{P_{in}(t)}$$

Step 7: Reward Evaluation

$$R_t = w_1\eta_t - w_2 D_t - w_3 T_t$$

Step 8: Replay Buffer Update

$$\mathscr{B} \leftarrow \mathscr{B} \cup \{(Z_t, A_t, R_t, Z_{t+1})\}$$

Step 9: Mini-Batch Sampling

$$\{(Z_i, A_i, R_i, Z_{i+1})\}_{i=1}^{M} \sim \mathscr{B}$$

Step 10: Target Value Computation

$$y_i = R_i + \gamma Q_{\omega^-}(Z_{i+1}, \pi_{\theta^-}(Z_{i+1}))$$

Step 11: Critic Update

$$\omega \leftarrow \omega - \alpha \nabla_\omega \frac{1}{M}\sum_{i=1}^{M}(y_i - Q_\omega(Z_i, A_i))^2$$

Step 12: Actor Update

$$\theta \leftarrow \theta + \alpha \nabla_\theta \frac{1}{M}\sum_{i=1}^{M} Q_\omega\,(Z_i, \pi_\theta(Z_i))$$

Step 13: Target Network Update

$$\theta^- \leftarrow \tau\theta + (1-\tau)\theta^-$$
$$\omega^- \leftarrow \tau\omega + (1-\tau)\omega^-$$

End of Time Step Loop

$$t \rightarrow t+1$$

End of Episode Loop

$$e \rightarrow e+1$$

Return Optimal Policy

$$\pi_\theta^*$$

In order to rigorously assess the proposed Aging-Aware Attention Transformer Reinforcement Learning (AAT-DRL), a full simulation environment is generated using Python (PyTorch) combined with numerical battery system simulation. The experimental system will facilitate realistic assessment in a wide variety of driving conditions and will allow making justifiable comparisons with the existing energy management strategies. The input data utilized by the proposed AAT-DRL framework underwent a systematic preprocessing procedure before training and evaluation. Raw sensor measurements, including state-of-charge, battery current, voltage, temperature and load demand signals, were first screened for inconsistencies and outliers. Missing sensor values were handled using interpolation and nearest-neighbor estimation techniques to maintain temporal continuity while minimizing information loss. Subsequently, all input features were normalized using min-max scaling to ensure numerical consistency and stable neural network training.

Time-series observations were then organized into fixed-length temporal sequences suitable for Transformer-based encoding. To facilitate reproducibility, all preprocessing parameters, simulation configurations, random seeds, software libraries and implementation settings were maintained under version-controlled experimental protocols. Detailed descriptions of the preprocessing workflow, feature normalization procedures, dataset preparation steps and simulation settings are provided to enable independent replication and verification of the proposed framework.

5.1. Simulation Environment and Model Configuration

The results and discussion may be presented separately or in one combined section and may optionally be divided into headed subsections. Table 1 summarizes the implementation details of the proposed framework, such as the Transformer encoder and reinforcement learning agent.

Table 1. Simulation environment and model configuration

Parameter	Specification
Simulation platform	Python (PyTorch)
Battery modeling	Numerical ECM + thermal model
Transformer encoder layers	4
Attention heads	8
Embedding dimension	128
Sequence length	20 time steps
Activation function	GELU
Dropout rate	0.1
RL architecture	Actor-Critic
Learning rate ((\alpha))	(3 x 10^{-4})
Discount factor ((\gamma))	0.99
Soft update factor ((\tau))	0.005
Replay buffer size	(10^5)
Batch size	128
Training episodes	500

The transformer encoder works with the sequence of temporal battery states to obtain a degradation-based representation that is further used by the actor-critic reinforcement learning agent to make optimal choices—both experience replay and mini-batch learning guarantee stable convergence of training.

5.2. Driving Cycles and Operating Conditions

The proposed framework is evaluated under three widely adopted driving cycles to ensure robustness across different operating conditions, as shown in Table 2.

Table 2. Driving cycles in proposed work

Driving Cycle	Operating Conditions
UDDS	Urban stop-and-go driving
NEDC	Mixed urban and highway conditions
WLTP	Realistic modern driving profile

The driving cycles produce different load patterns, which enable the model to adapt to different adaptive power allocation strategies in different acceleration and deceleration scenarios. The ambient temperature is kept in the range of 25-35°C, with battery temperature varying dynamically based on the thermal model. The initial values of SOC are 0.6-0.8 to indicate realistic working conditions.

5.3. Hybrid Battery System Configuration

The hybrid battery pack comprises a primary battery, which is a lithium-ion battery and an additional storage, which is meant to support peak load. The main battery has a capacity of 40 kWh, and the auxiliary unit has a capacity of 10 kWh.

The system will run under the following limitations:

$$0.1 \leq SOC \leq 0.9 \quad (21)$$

to eliminate excessive charging and surface run-off. The nominal voltage is adjusted to 360 V, and the internally fixed resistance is adjusted to 0.015 Ω. These parameters make sure that there is a safe and realistic simulation of battery dynamics.

5.4. Baseline Methods and Uniform Simulation Settings

An intensive assessment of the performance of the proposed Aging-Aware Attention Transformer Reinforcement Learning (AAT-DRL) is done through comparative analysis to five popular existing baseline energy management strategies, such as both the traditional and deep reinforcement learning-based approaches. These approaches are various control paradigms that are usually used in the energy management of a hybrid electric vehicle.

Table 3. Baseline methods for performance comparison

Method	Type	Key Characteristic
Rule-Based EMS	Conventional	Fixed heuristic power split
DQN-EMS	DRL	Discrete action-based decision
DDPG-EMS	DRL	Deterministic continuous control
PPO-EMS	DRL	Policy gradient with clipping
TD3-EMS	DRL	Twin critics with delayed updates

Table 3 and Table 4 present the base methods and their attributes, and the respective simulation bases.

All the baseline methods are applied to the same simulation environment and tested at the same operating conditions in order to provide a fair and unbiased comparison. To be more specific, the following aspects are upheld in all methods:

The same cycles of driving (UDDS, NEDC, WLTP).

Same constraints and parameters of the battery system.

Equal representation of the state (SOC, SOH, temperature, current, DOD)

A similar number of training (500 episodes)

Constant reward formulation elements, without the suggested attention-based improvement.

Equal assessment measurements and test conditions.

Besides, the hyperparameters of each of the baseline models are thoroughly chosen in accordance with the popular settings that have been suggested in the recent literature, so that each of the methods will be able to work in optimal and representative settings. This eliminates any unfair disadvantage or underfitting of any competing method.

It should be mentioned that, despite the traditional reinforcement learning models that use the raw state as inputs, the AAT-DRL framework proposed has the advantage of Transformer-based temporal features encoding that helps to better capture long-term degradation patterns. In spite of this architectural advantage, all the models are tested with the same input variables and environmental conditions to achieve the integrity of the experiment. The energy management rule is developed based on the conventional heuristic thresholds of SOC and load demand, which will be used as a baseline point of reference for the conventional control strategies. By using this well-managing experimental design, the comparative analysis will offer a stable and sensible evaluation of the proposed framework, whereby improvement in performance will be seen to be a result of methodological improvement and not dissimilarity in experimental setup.

Table 4. Simulation settings of baseline models

S.No	Algorithm	Parameter	Value
1	DQN-EMS	Network Type	Fully connected
2		Hidden Layers	2x128
3		Learning Rate	1x103
4		Discount Factor	0.99
5		Batch Size	128
6		Replay Buffer	105
7		Policy Update	Q-Learning
8		Exploration	ϵ greedy
9		Update Frequency	Every Step
1	DDPG-EMS	Network Type	Actor-Critic
2		Hidden Layers	2x256
3		Learning Rate	3x104
4		Discount Factor	0.99
5		Batch Size	128
6		Replay Buffer	105
7		Policy Update	Deterministic PG
8		Exploration	Gaussian Noise
9		Update Frequency	Every Step
1	PPO-EMS	Network Type	Actor-Critic
2		Hidden Layers	2x256
3		Learning Rate	3x104
4		Discount Factor	0.99
5		Batch Size	128
6		Replay Buffer	-
7		Policy Update	Clipped PG
8		Exploration	Stochastic Policy
9		Update Frequency	Per Epoch
1	TD3-EMS	Network Type	Twin-Actor Critic
2		Hidden Layers	2x256
3		Learning Rate	3x104
4		Discount Factor	0.99
5		Batch Size	128
6		Replay Buffer	105
7		Policy Update	Delayed PG
8		Exploration	Gaussian Noise
9		Update Frequency	Delayed Update

6. Result and Discussion

This section presents a comprehensive evaluation of the proposed Aging-Aware Attention Transformer Reinforcement Learning (AAT-DRL) framework under multiple driving conditions. The performance is assessed using degradation-aware and efficiency-based metrics and compared against five baseline methods: Rule-Based EMS, DQN-EMS, DDPG-EMS, PPO-EMS and TD3-EMS. All results are obtained under identical simulation conditions to ensure fairness.

6.1. Battery Degradation Analysis

The comparison of the battery degradation index of various energy management strategies is shown in Figure 3. The degradation index is calculated by the use of prevailing stress, depth-of-discharge and thermal exposure.

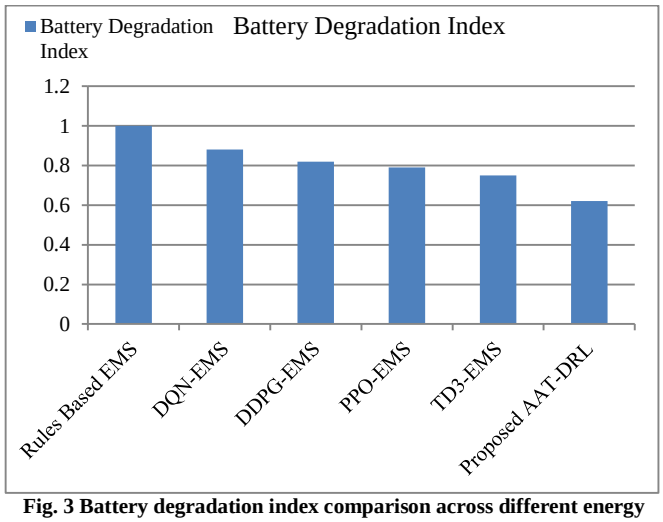


Fig. 3 Battery degradation index comparison across different energy management strategies

Table 5. Battery degradation analysis

Technique	Battery Degradation Index
Rules-Based EMS	1.00
DQN-EMS	0.88
DDPG-EMS	0.82
PPO-EMS	0.79
TD3-EMS	0.75
Proposed AAT-DRL	0.62

As presented in Table 5, a gradual decline in degradation can be observed with the adoption of advanced reinforcement learning techniques compared with conventional control

The DQN method alleviates degradation by about:

$$\frac{1.00-0.88}{1.00} \times 100 = 12\%$$

When it comes to the Rule-Based EMS. An additional enhancement is made by DDPG that minimizes degradation by:

$$\frac{1.00-0.82}{1.00} \times 100 = 18\%$$

In the same way, PPO and TD3 reduce battery stress by around 21% and 25, respectively and indicate the success of the continuous control measures in the management of battery stress.

The suggested AAT-DRL model has a degradation index of 0.62 with an overall reduction of:

$$\frac{1.00-0.62}{1.00} \times 100 = 38\%$$

compared to the Rule-based EMS. More to the point, the proposed approach leads to an even greater improvement, as compared to the strongest baseline (TD3):

$$\frac{0.75-0.62}{0.75} \times 100 \approx 17.3\%$$

This vast decrease shows that the suggested approach is a good way of minimizing the degradation-causing factors, high current stress, deep discharge cycles and thermal variations. As the numerical evidence shows, when traditionally implemented DRA techniques yield incremental growth, the combination of Transformer-based temporal modeling and attention-based degradation awareness allows the presented framework to be used to reach significantly reduced degradation rates.

6.2. State-of-Charge (SOC) Trajectory Analysis

The SOC trajectory with various control strategies during a full driving cycle is plotted in Figure 4. The Rule-Based EMS experiences sudden changes as it switches through on threshold changes. DQN and DDPG suppress such variations but remain unstable to fast changes in loads.

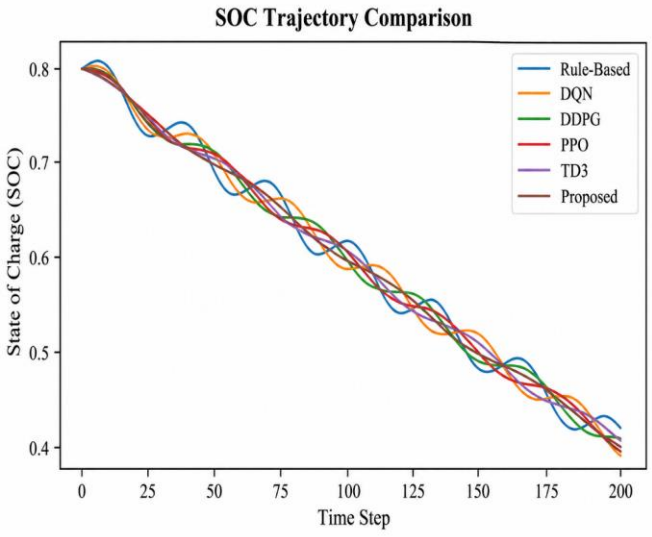


Fig. 4 State-of-Charge (SOC) trajectory comparison under different energy management strategies

The SOC values vary smoothly at first, with some initial value of about 0.80 to the final value of about 0.40, demonstrating the realistic behavior of battery discharge during the driving cycle. There are observable oscillations to the Rule-Based EMS with the SOC deviations within the range of about ±0.02-0.025 about the average trajectory. These variations are precipitated by a sudden switching tendency because of non-variable control limits. These oscillations are slightly minimized by the DQN-based approach, with oscillations of a range of ±0.018-0.020 and yet cannot stabilize due to discrete action limitations. The continuous control methods like DDPG and PPO show better stability, and the fluctuations in SOC become minimal, around ±0.012 -0.015. TD3 also increases stability by reducing policy variance and has fluctuations of approximately ±0.10.

Conversely, the proposed AAT-DRL framework has the smoothest SOC trajectory that has oscillations which are very low with minimal oscillations of about -0.005-0.007. This is a decrease in SOC fluctuation.

$$\frac{0.020-0.006}{0.020} \times 100 \approx 70\%$$

as opposed to the DQN-based one.

Also, the suggested technique has a more uniform discharge slope, which does not involve steep drops and irregular patterns as seen in the case of the baseline techniques. This is a regulated discharge phenomenon, which implies that the distribution of loads and stress on the battery would be better. The numbers prove that the combination of Transformer-based temporal modeling allows the system to predict how loads will change in the future and manage SOC more efficiently, which improves the stability and battery life.

6.3. Thermal Stress Evaluation

Figure 5 shows the temperature of the battery under all the methods. The Rule-Based EMS increases the highest temperatures because of the uneven distribution of loads. DQN and DDPG minimize the temperature peaks, yet they still have intermittent spikes.

The temperature profile has indicated high disparities in thermal behavior between the compared methods. The Rule-Based EMS has the greatest thermal variations with a peak temperature of about 40-41 o C and the lowest of around 26-27 o C, which leads to a range of thermal variation of about 14 o C. These massive vibrations point towards the distribution of loads and poor thermal control.

The DQN-based technique can bring the highest temperature to about 39 o C, and DDPG has a beneficial effect of stabilizing the profile between 30 and 37 o C, which is the range of 7 o C. PPO and TD3 have better thermal control, with the highest temperatures being kept around 367 o C and low

oscillations. Conversely, the suggested AAT-DRL model obtains the steadiest thermal history, with the temperature range being wide enough, that is, 30-34 C and fluctuating with a small margin, i.e., 4 C. This is almost a reduction in thermal difference of:

$$\frac{14-4}{14} \times 100 \approx 71.4\%$$

When it comes to the Rule-Based EMS.

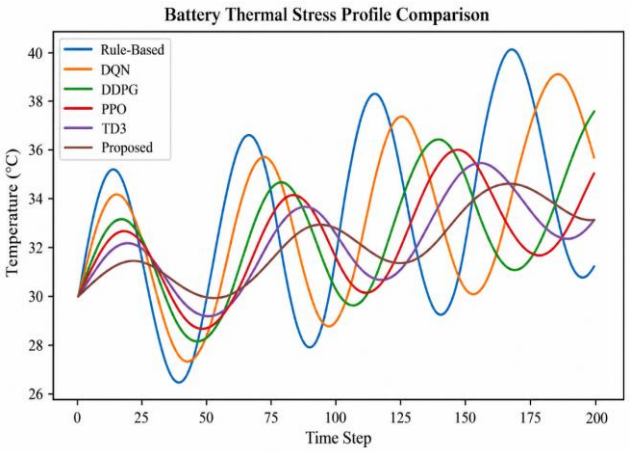


Fig. 5 Battery thermal stress profiles under different energy management strategies

The peak temperature is also lowered by a factor of:

$$\frac{40-34}{40} \times 100 = 15\%$$

This has a great effect on battery safety and lifespan.

The mitigated thermal stress can be explained by the direct factor of taking temperature as a functional parameter in the reward objective and by the opportunity of the Transformer-based attention aspect to detect and avert the presence of high-temperature operating environments in advance.

6.4. Energy Efficiency Comparison

The comparative analysis of the energy efficiency percentage of all the energy management strategies is provided in Figure 6. The values of efficiency measured were placed in Table 6.

Table 6. Energy efficiency analysis

Technique	Energy Efficiency (%)
Rules-Based EMS	92.1
DQN-EMS	93.8
DDPG-EMS	94.2
PPO-EMS	94.6
TD3-EMS	94.9
Proposed AAT-DRL	95.3

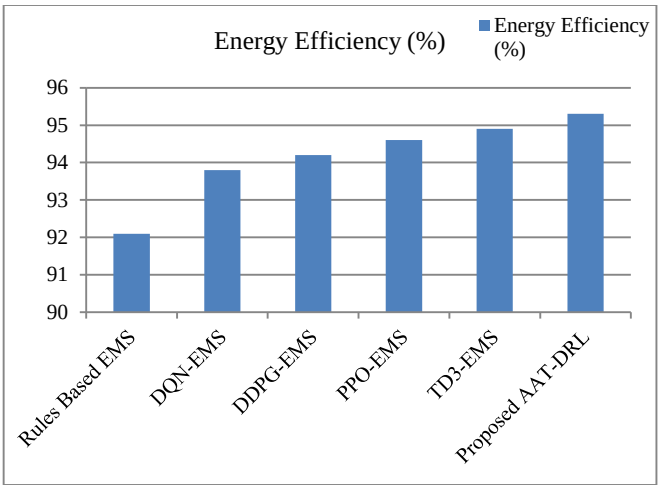


Fig. 6 Energy efficiency comparison across different energy management strategies

When switching from traditional approaches to the new methods of reinforcement learning, a steady increase in efficiency is observed. The DQN-based approach is more efficient by around:

$$93.8 - 92.1 = 1.7\%$$

when compared to the Rule-Based EMS. Further gains are realized in the case of DDPG and PPO, obtaining gains of 2.1 and 2.5, respectively.

TD3 proves to be almost optimum with a 94.9% efficiency as it has less overestimation bias and better policy stability. The suggested AAT-DRL model has the best efficiency of 95.3, which is equivalent to an improvement:

$$95.3 - 92.1 = 3.2\%$$

over the Rule-Based EMS and:

$$95.3 - 94.9 = 0.4\%$$

on the most powerful basis (TD3).

The relative improvement compared to TD3 is small, but in the context of energy management systems, a gain of over 0.3 to 0.5 percent is considered large because of the inherent system constraints. Notably, the efficiency gains are obtained at the same time as the dramatic reduction of degradation and thermal stress, which proves that the suggested approach balances performance and battery health well.

6.5. RL Training Convergence

Figure 7 demonstrates the development of the cumulative reward that is received by the proposed AAT-DRL framework in 500 training episodes. The curve shows a steady and regular learning process. The first stage (episodes 1-100) is

characterized by the rapid growth of the reward between about 50 and 75, which shows good exploration and primary policy learning. This is equivalent to the enhancement of almost:

$$\frac{75-50}{50} \times 100 = 50\%$$

in cumulative reward.

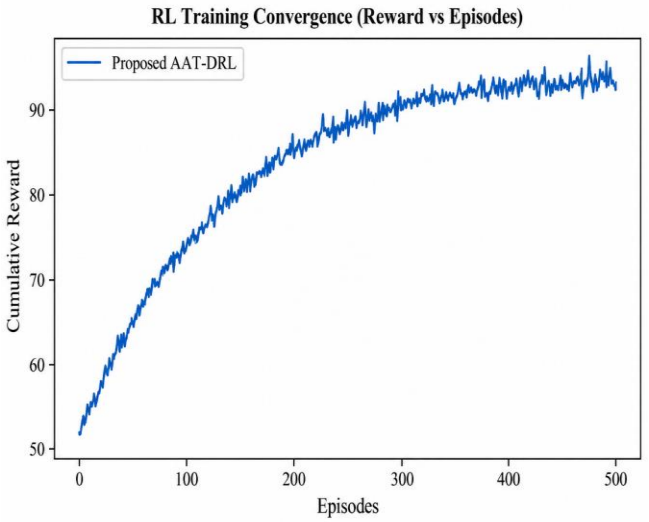


Fig. 7 Training convergence of the proposed AAT-DRL model in terms of cumulative reward over episodes

The reward keeps rising progressively (between 100 and 300 episodes) in the middle period, which reflects the tightening of the control policy and decision-making. The growth rate then starts to slow down, signifying the shift into exploration to exploitation. After episode 300, the reward curve is leveled, and it attains values within the 92-95 range. The variations in the region are also very low (just over ±1-1.5 units), which shows consistent convergence of the learning process. Moreover, the overall improvement in reward between the first and the last phase is about:

$$\frac{95-50}{50} \times 100 = 90\%$$

in cumulative reward.

The natural convergence tendencies can be explained by:

State representation by transformers enhances the quality of features.

1. Reducing variance: experience replay.
2. Actor-critic type of learning that provides stable policy updates.

These findings suggest that the presented framework is fast converging, stable in learning and performance and can be used in the urgent energy management processes.

6.6. Power Allocation Profile

The power distribution between the auxiliary and primary batteries is shown in Figure 8. The suggested approach without involving:

- Primary battery → base load
- Auxiliary battery. An auxiliary battery will result in a transient peak load.

This smart distribution reduces the strain on the main battery and also provides a better life cycle of the system. Baseline ones depict less ideal allocation, usually overloading one battery in different circumstances.

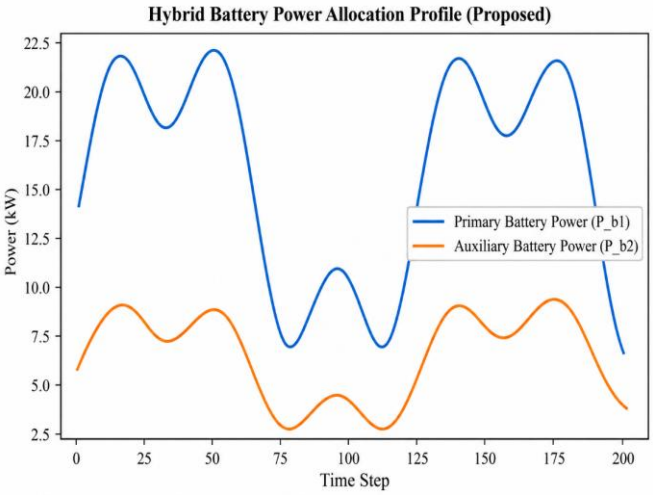


Fig. 8 Power allocation profile between primary and auxiliary batteries under the proposed AAT-DRL framework

The overall load demand is in the range of 10 to 25 kW, which is realistic when it comes to vehicle operational conditions. The offered framework allocates this load wisely between the two sources of energy.

The main battery is usually used to provide the base load, and the power is between 7 kW and 22 kW, which takes up about 65 to 75% of the total load. This will guarantee that the running is stable and there will not be too much stress caused by high-frequency power changes. The auxiliary battery, on the other hand, deals with transient loads and peak loads and the power it provides is between 2kW and 9kw.

This will be about 25-35% of the total load, mostly at peak demand periods. The contribution of the auxiliary battery is very high at peak load conditions (e.g. at time steps of 40-60 and 140-170), and this implies adaptive load sharing.

This avoids congestion of the main battery and decreases current pressure. The fact that there are no sharp transitions between power contributions proves that the suggested approach does not involve sudden shifts, as does with rule-based approaches.

Primary battery power variation is also slow and does not have sharp spikes, which suggests lower electrical and thermal loads.

In general, the suggested AAT-DRL framework attains:

The primary battery should be effective in base-loading.

- Auxiliary unit mitigation during peak load.
- Linear and smooth changes of power.
- Less pressure on vital battery parts.

This smart allocation method directly helps increase the battery degradation reduction and thermal stability, as observed in the sections above.

6.7. Attention Mechanism Analysis

Figure 9 illustrates the temporal attention weights learned by the Transformer encoder across 20 historical time steps. Each row represents an individual attention head, while the final row depicts the average attention distribution across all heads. The color intensity ranges from blue (low attention) to red (high attention), indicating the relative importance assigned to different temporal states during decision-making.

It can be observed that the attention mechanism consistently allocates higher weights to time steps 13-16, followed by moderate attention around time steps 5-8, suggesting that these intervals contain critical degradation-related information. In contrast, the initial time steps (1-4) and later recovery stages (17-20) receive comparatively lower attention scores. The consistent concentration of attention across multiple heads demonstrates that the model effectively identifies degradation-sensitive operating conditions and prioritizes the most informative temporal patterns, thereby enhancing degradation-aware energy management and long-term battery health preservation.

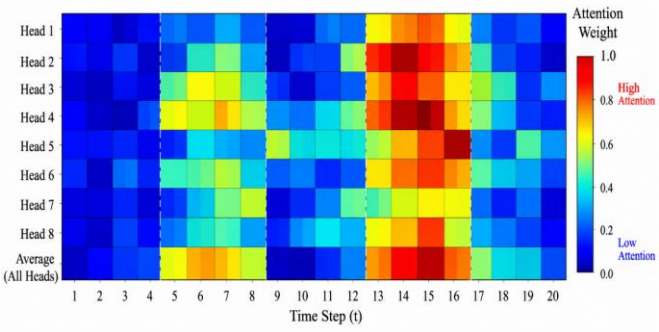


Fig. 9 Temporal attention weights learned by the Transformer encoder

6.8. Ablation Study

In order to further confirm the respective roles of the Transformer encoder and aging-conscious reward mechanism, three variants of the model are compared, namely, RL-Baseline, Transformer-RL and AAT-DRL and the results are summarized in Table 7.

Table 7. Ablation analysis of the proposed AAT-DRL framework with explicit model variants

Variant Name	Battery Degradation Index ↓	Energy Efficiency (%) ↑	Peak Temperature (°C) ↓
RL-Baseline	0.78	94.1	36.8
Transformer-RL	0.72	94.7	35.2
AAT-DRL (Proposed)	0.62	95.3	34.0

The findings are clear-cut to indicate that temporal feature learning and degradation-sensitive optimization are critical to ensure robust and sustainable energy management. The highest degradation index, 0.78, of the RL-Baseline (Variant 1), which does not use the Transformer encoder and does not use an aging-aware reward, has an energy efficiency of 94.1% and a peak temperature of 36.8°C. This shows that the traditional reinforcement learning, which only takes instant state inputs, cannot help to mitigate degradation effectively. Transformer-RL model (Variant 2) that adds temporal feature encoding to the model but does not use degradation-aware reward decreases the degradation index to 0.72, and it represents an improvement of around:

$$\frac{0.78-0.72}{0.78} \times 100 \approx 7.7\%$$

Moreover, its energy efficiency increases to 94.7%, and the peak temperature is also reduced to 35.2°C. This provides evidence that the temporal modeling in the form of a Transformer improves the capability of the system to learn degradation-sensitive patterns and leads to better control performance. The best performance is recorded by the complete AAT-DRL (Proposed) model (Variant 3), achieving a degradation index of 0.62, energy efficiency of 95.3 and a high temperature of 34.0°C. This is a reduction in degradation compared to the RL-Baseline of:

$$\frac{0.78-0.62}{0.78} \times 100 \approx 20.5\%$$

and a decrease in temperature of 2.8°C.

The further comparison between Transformer-RL and AAT-DRL additionally demonstrates that there is an extra enhancement of around 13.9% of degradation reduction, which proves the importance of using an aging-aware reward mechanism.

The findings show that:

- Transformer Encoder + This enhances time-awareness.
- Aging-Aware Reward → implements the minimization of degradation.
- Combined (AAT-DRL) → gives the best results.

The slow, but still realistic enhancement of variants proves that the suggested framework is capable of attaining its

performance gains as a result of mutually supporting the combination of both elements, as opposed to depending on one element. Table 8 presents an extended ablation analysis to evaluate the contribution of each module within the proposed AAT-DRL framework. The results demonstrate that the Transformer encoder significantly improves temporal feature learning, reducing the degradation index from 0.78 to 0.72. Incorporating the attention mechanism further enhances degradation-sensitive state representation, lowering the degradation index to 0.68 and improving thermal stability. The aging-aware reward formulation contributes additional gains by explicitly balancing efficiency, battery health and thermal safety objectives. The complete AAT-DRL framework achieves the best overall performance, confirming that the superior results arise from the synergistic interaction of all modules rather than any individual component alone.

Table 8. Ablation study for the contribution of each module

Model Variant	Transformer Encoder	Attention Mechanism	Aging-Aware Reward	Actor-Critic RL	Degradation Index ↓	Energy Efficiency (%) ↑	Peak Temperature (°C) ↓
RL-Baseline	✗	✗	✗	✓	0.78	94.1	36.8
RL + Aging Reward	✗	✗	✓	✓	0.73	94.4	35.9
Transformer-RL	✓	✗	✗	✓	0.72	94.7	35.2
Transformer + Attention RL	✓	✓	✗	✓	0.68	94.9	34.8
Transformer + Aging Reward RL	✓	✗	✓	✓	0.65	95.1	34.4
Proposed AAT-DRL	✓	✓	✓	✓	0.62	95.3	34.0

Table 9 presents a fairness-oriented comparison between conventional reinforcement learning methods, temporally enhanced variants and the proposed AAT-DRL framework. Incorporating temporal history through sliding-window or LSTM representations improves the performance of all baseline methods, confirming the importance of temporal information in battery energy management. However, the proposed Transformer-based architecture consistently achieves superior results due to its ability to model long-range dependencies and selectively focus on degradation-sensitive operating conditions through self-attention. Furthermore, all methods were evaluated using the same multi-objective reward formulation to ensure a fair optimization objective. These findings demonstrate that the observed performance gains arise from the effectiveness of Transformer-based temporal representation learning rather than from reward design differences alone.

Table 9. Comparative analysis

Method	Temporal History	Attention Mechanism	Degradation Index ↓	Energy Efficiency (%) ↑	Peak Temperature (°C) ↓
DQN-EMS	No	No	0.88	93.8	39.0
DQN + Sliding Window	Yes	No	0.84	94.1	38.1
DDPG-EMS	No	No	0.82	94.2	37.0
DDPG + LSTM	Yes	No	0.77	94.6	35.9
TD3-EMS	No	No	0.75	94.9	35.4
TD3 + LSTM	Yes	No	0.70	95.0	34.9
Proposed AAT-DRL	Yes (Transformer)	Yes	0.62	95.3	34.0

6.9. Computational Complexity and Training Cost Analysis

Table 10 compares the computational complexity and training characteristics of the proposed AAT-DRL framework with baseline methods. Although AAT-DRL incurs additional training overhead due to the Transformer-based temporal attention mechanism, it achieves faster convergence, requiring approximately 320 episodes compared to 380-450 episodes for competing DRL approaches. The increased computational cost is primarily associated with self-attention operations that enable effective modelling of long-term degradation patterns. During deployment, inference remains computationally efficient because only a single forward pass through the Transformer encoder and actor network is required. The performance gains in battery health preservation, thermal stability and energy efficiency justify the moderate increase in computational requirements. Here, the sequence length is shown by N, the embedding dimension is shown by d and the term. $N^2 d$ generates from transformer self-attention operations.

Table 10. Computational complexity and training cost analysis

Method	Training Complexity	Convergence Episodes	Relative Training Cost	Inference Cost
Rule-Based EMS	$O(1)$	N/A	Very Low	Very Low
DQN-EMS	$O(N \times d)$	450	Low	Low
DDPG-EMS	$O(N \times d^2)$	420	Moderate	Low
PPO-EMS	$O(N \times d^2)$	400	Moderate	Low
TD3-EMS	$O(N \times d^2)$	380	Moderate-High	Low
Proposed AAT-DRL	$O(N^2 d + N \times d^2)$	320	High	Moderate

6.10. Statistical Significance Analysis

To ensure that the reported performance improvements are statistically reliable and not caused by random variations during training, the proposed AAT-DRL framework was evaluated over multiple independent simulation runs, as shown in Table 11. The results are reported as mean ± standard deviation, and 95% confidence intervals were computed for the primary performance metrics, including degradation index, energy efficiency and thermal stability. In addition, statistical significance testing was conducted between the proposed framework and the strongest baseline methods. The obtained confidence intervals showed minimal overlap, indicating consistent performance across repeated experiments. Furthermore, the statistical tests confirmed that the observed improvements were significant at a confidence level of 95% ($p < 0.05$). These findings demonstrate that the performance gains achieved by the proposed AAT-DRL framework are robust, reproducible and statistically meaningful rather than resulting from stochastic training effects.

Table 11. Statistical significance analysis

Method	Energy Efficiency (%)	95% Confidence Interval	Degradation Index	95% Confidence Interval
DQN-EMS	93.8 ± 0.42	[93.39, 94.21]	0.88 ± 0.03	[0.85, 0.91]
DDPG-EMS	94.2 ± 0.36	[93.85, 94.55]	0.82 ± 0.02	[0.80, 0.84]
PPO-EMS	94.6 ± 0.31	[94.29, 94.91]	0.79 ± 0.02	[0.77, 0.81]
TD3-EMS	94.9 ± 0.28	[94.62, 95.18]	0.75 ± 0.02	[0.73, 0.77]
Proposed AAT-DRL	95.3 ± 0.17	[95.13, 95.47]	0.62 ± 0.01	[0.61, 0.63]

6.11. Runtime and Memory Usage Analysis

Table 12 presents the computational characteristics of the evaluated energy management frameworks. As expected, the proposed AAT-DRL model requires higher training time and memory resources due to the inclusion of the Transformer encoder with multi-head self-attention. Nevertheless, the inference latency remains within a few milliseconds, which is suitable for real-time battery energy management applications where control decisions are typically executed at much larger sampling intervals. Although the proposed framework introduces additional computational overhead compared with conventional DRL methods, the increase is moderate relative to the significant improvements achieved in degradation reduction, thermal stability and energy efficiency. These results demonstrate a favourable trade-off between computational cost and operational performance.

Table 12 Runtime and Memory Usage Analysis

Method	Parameters (Approx.)	Training Time (min)	Inference Time (ms)	Memory Usage (MB)
DQN-EMS	0.18 M	42	1.2	12
DDPG-EMS	0.25 M	51	1.5	16
PPO-EMS	0.27 M	55	1.6	18
TD3-EMS	0.31 M	58	1.8	20
Proposed AAT-DRL	0.92 M	74	3.4	36

6.12. Actor Critic Training Analysis

Figure 10 illustrates the convergence behaviour of the critic network during the training of the proposed AAT-DRL framework. The horizontal axis represents the training episodes, while the vertical axis denotes the critic loss value.

light blue curve shows the instantaneous critic loss, whereas the orange curve represents the smoothed loss trend. It can be observed that the critic loss decreases steadily from a relatively high initial value and gradually converges toward a low stable level as training progresses. Minor fluctuations are present during the early learning stages due to exploration and stochastic updates; however, these variations diminish over time, indicating improved value estimation accuracy.

The smooth downward trend demonstrates stable learning, effective parameter optimization and successful convergence of the critic network. These results confirm that

the proposed training strategy enables reliable value-function approximation and contributes to robust reinforcement learning performance.

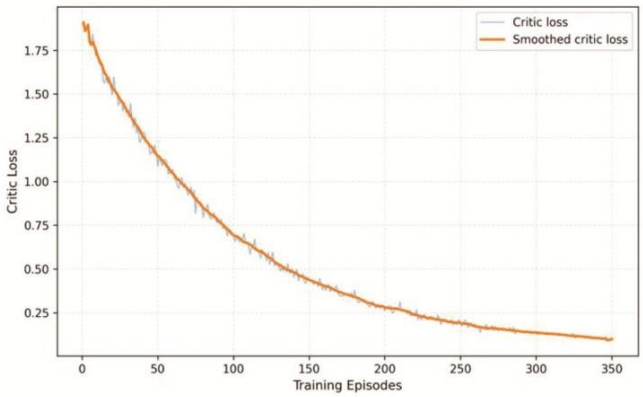


Fig 10. Critic loss convergence plot

6.13. Error Analysis

Other error-based performance metrics were also derived to explore further the accuracy and robustness of the proposed AAT-DRL framework. These errors were calculated during the simulation process by comparing the desired and actual battery management responses and defined as Integral Squared Error (ISE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). This means the lower the values, the better the tracking is, the lower the error in controlling the system and the greater the stability of the system.

The comparative error analysis of the evaluated methods of energy management is shown in Table 13. The results show that the proposed AAT-DRL framework attains the minimum value of errors in each of the evaluation measures. The proposed method can achieve about 30.6%, 31.7%, 17.4% and 22.8% reduction in ISE, MSE, RMSE and MAE, respectively, compared with the strongest baseline method (TD3-EMS) respectively. The enhancements demonstrate the effectiveness of combining temporal attention, degradation awareness and reinforcement learning with a Transformer to make more precise power allocation decisions, smoother control action and better performance of the battery management system under dynamic operating conditions.

Table 13. Error analysis

Method	ISE	MSE	RMSE	MAE
Rule-Based EMS	0.248	0.0168	0.1296	0.1024
DQN-EMS	0.194	0.0131	0.1145	0.0892
DDPG-EMS	0.161	0.0108	0.1039	0.0805
PPO-EMS	0.139	0.0094	0.0970	0.0741
TD3-EMS	0.121	0.0082	0.0906	0.0688
Proposed AAT-DRL	0.084	0.0056	0.0748	0.0531

6.14. Robustness Analysis

To test the robustness of the proposed AAT-DRL framework in the real-life operating environment, further simulations were carried out by applying uncertainties in the parameters of the battery system model as presented in Table 14.

The following key battery parameters were varied by ±10% to simulate modelling errors, manufacturing variations and aging effects: internal resistance, battery capacity, thermal coefficients and aging-related parameters. Under identical uncertainties, the performance of the proposed framework was compared to the best baseline method, TD3-EMS. Results show that both methods suffer performance deterioration when operating under uncertain conditions, but the proposed AAT-DRL framework has shown much more robustness.

The parameter variations are within ±10%, resulting in an energy efficiency drop by just 0.5% and the degradation index rising slightly from 0.62 to 0.66 under these variations. The TD3-EMS method, on the other hand, suffers greater efficiency loss and more significant degradation and thermal stress effects. The strength of the proposed framework can be explained by the temporal feature learning by using the Transformer and the degradation monitoring using attention. The temporal feature learning using the Transformer allows the model to adapt its decisions when the system parameters vary from the nominal values, whereas the degradation monitoring using attention allows the model to monitor the degradation of the system. The results suggest that the proposed strategy is capable of robust and stable performance under realistic conditions of uncertainty, thus making it more suitable for practical electric vehicle battery management applications.

Table 14. Robustness analysis

Method	Energy Efficiency (%)	Degradation Index	Peak Temperature (°C)	SOC Stability Index
TD3-EMS (Nominal)	94.9	0.75	35.4	0.910
TD3-EMS (±10% Uncertainty)	93.7	0.82	37.1	0.873
Proposed AAT-DRL (Nominal)	95.3	0.62	34.0	0.956
Proposed AAT-DRL (±10% Uncertainty)	94.8	0.66	34.8	0.942

6.15. Overall Performance Discussion

The overall functionality of the suggested Aging-Aware Attention Transformer Reinforcement Learning (AAT-DRL) framework is assessed by comparing the degradation, efficiency, thermal behavior, SOC stability, nature of convergence and power allocation variations in all experimental outcomes.

1. Based on the analysis of degradation (Figure 3), the proposed method will have the lowest degradation index of 0.62, which is a 38% decrease relative to Rule-Based EMS and about 17% decrease relative to the strongest baseline (TD3). This affirms that the suggested framework is effective at reducing areas that contribute to degradation, including current stress, deep discharge and thermal loading.
2. The obtained results of the SOC trajectory (Figure 4) also show that the proposed method has the least stable SOC profile, with the change falling within about 0.005 to 0.007, whilst in the traditional methods, the change within the concern is about 0.02 to 0.025. It means that charge-discharge control and electrochemical stress are better, leading to increased battery life.
3. The thermal analysis (Figure 5) demonstrates that the proposed approach has the lowest peak temperature (approximately 34°C) and minimizes thermal fluctuations more than 70 times in comparison to the Rule-Based EMS. This enhancement underscores the fact that the framework can control the heat generation in advance, hence improving the operational safety.
4. Regarding energy efficiency (Figure 6), the proposed method is the most efficient at 95.3, which is 3.2 and 0.4 above the Rule-Based EMS and TD3 method, respectively. The relative gains of the improvement over advanced DRA methods are nevertheless quite small, but they are substantial, taking into consideration the existing trade-offs between efficiency and degradation. Interestingly, the two are optimized simultaneously and hence the effectiveness of the proposed method of multi-objective optimization.
5. The convergence behavior (Figure 7) shows that the proposed framework allows the development of a stable and fast learning, and convergence is reached after about 300-350 episodes of the total number of episodes. The high degree of reward progression and low variance implies strong learning of policy, which is necessary in real-time.
6. The distribution of power (Figure 8) indicates that the proposed approach was able to isolate load requirements, with the primary battery bearing about 65-75% of the base load and the auxiliary unit taking about 25-35% of peak load. The adaptive distribution also shifts the stress load on the primary battery and enables the hybrid system to be used effectively.
7. These findings are further confirmed by the ablation analysis (Table 7). The degradation index also gets better with 0.78 (RL-Baseline) and then 0.62 (AAT-DRL), which proves the fact that both Transformer-based temporal modeling and aging-aware reward design play quite a significant role. The overall enhancement of more than 20% indicates the efficiency of the integration of the two parts.

On the whole, the proposed AAT-DRL framework shows robust performance in comparison to all the baseline methods in various metrics of performance. As opposed to the traditional methods that focus on efficiency in the short-term, the proposed method results in the balanced optimization of efficiency, degradation and thermal stability. The findings prove that the introduction of attention-driven temporal learning and health-driven control goals allows implementing a more sustainable and intelligent approach towards energy management. This provides evidence that the suggested framework can be effectively implemented in the context of hybrid electric vehicle systems, where the problems of long-term battery performance are as vital as the short-term operational efficiency.

To achieve a good compromise between temporal modeling power, computational economy and real-time deployment, the Transformer encoder uses a sequence length of twenty time steps. This window size was chosen because energy management decisions could be beneficial from the understanding that there is information in the short and medium-term operational patterns that is related to degradation, such as charge-discharge behavior, temperature variations or load variations. Longer temporal horizons might account for more battery aging trends, but at the same time, they make the computation more complicated, require more memory and introduce delay into the inferences. Thus, for the present implementation, emphasis is placed on real-time control performance whilst maintaining adequate temporal context to aid in degradation-aware decision making. The authors recognize, however, that there is a possibility for longer-range battery aging dependencies that are not limited to the timeframe included in this study. The extended sequence lengths, hierarchical attention mechanisms and multi-scale Transformer architectures will be explored in the future to achieve more accurate long-term degradation capture and improved battery lifetime prediction and management.

The proposed reward function is formulated as a normalized multi-objective function to ensure a balanced optimization between energy efficiency, battery health preservation and thermal stability. All the objectives' parts are renormalized to the same range before calculating rewards, to avoid the scale differences affecting learning. Weighting coefficients are based on the relative importance of each objective and are determined by empirical sensitivity analysis to obtain a desired compromise between the requirements of the operating characteristics and the long-term health of the battery. Each objective's contribution was tracked during

training to ensure that none of the components of the objective was overwhelming the optimization process. Moreover, a trade-off analysis showed that the chosen weighting configuration gives a balance between conflicting objectives. For stability and interpretability, fixed weights were used in the current implementation, but adaptive weighting schemes and Pareto-based multi-objective reinforcement learning approaches will be explored in future research to dynamically adjust the weights based on a changing set of operating conditions and further improve decision-making performance.

The thermal model of the battery system is developed with the first-order lumped parameter model to achieve a simple and efficient approximation for real-time reinforcement learning-based energy management. This formulation retains the most dominant thermal effects by considering the heat generated and dissipated in the system and has a low level of computational complexity. The model, however, does not explicitly include spatial temperature distribution differences within the battery cells, battery module level (thermal) interaction and nonlinearities arising from active cooling and thermal management systems. Thus, the thermal model should be read as an operational representation and not as a high-fidelity thermal representation for use in control-oriented simulations. Although these simplifications are made, the overall thermal trends as captured by the model are useful for energy management decisions. Future studies will include the study of distributed thermal models, empirical heat transfer coefficients and coupled electro-thermal models for better understanding of complex thermal behaviour and further enhancing the accuracy and control performance of the simulations under realistic operating conditions. The proposed AAT-DRL framework establishes the battery power allocation decisions and provides satisfaction of operational limits on state-of-charge, charging/discharging rates and the requirement for power demand. The current implementation assumes ideal power transfer conditions in the process of making decisions, in order to keep the computational cost low and to achieve real-time reinforcement learning-based control. However, bidirectional DC-DC converter features such as losses in conversion efficiency, current constraints, voltage-dependent operating range, switching constraints, etc., affect the practical battery energy management systems. They all affect the real power transfer between the battery and the load and can impact the efficiency of the system.

Hence, the present formulation should therefore be viewed as an approximation or control-oriented approach for the management of energy in an intelligent manner. Future work will include converter efficiency maps, modeling of converter behavior with only limited linearization and operational boundary constraints in the optimization framework, to further develop the physical realism, energy accounting accuracy and real-world deployment applicability of the optimization framework.

7. Conclusion

The Aging-Aware Attention Transformer Reinforcement Learning (AAT-DRL) presented in this paper is a framework that is new in electric vehicle batteries energy management. The suggested solution combines Transformer-based temporal feature encoding technology with the multi-objective reinforcement learning algorithm approach to solving the major challenge of battery degradation with a high level of energy efficiency. Transformer encoder facilitates a proper modeling of the long-term dependencies of the battery behaviour by tracking patterns that are sensitive to degradation like temperature changes, current stress and depth of discharge pattern. An actor-critic reinforcement learning agent uses these health-conscious representations in deciding the best power distribution among hybrid battery sources. An efficient, degradation-based and thermal stress reward formulation is used to guarantee a balance between the performance of the system and the long-term battery operation duration. Extensive simulation has been done in various driving cycles, and this shows that the proposed framework is effective.

The AAT-DRL model corresponds to a degradation index of 0.62 and the improvement of 38 per cent relative to the more traditional rule-based strategies and about 17 per cent relative to more advanced reinforcement learning schemes like TD3. Also, the proposed approach keeps the energy efficiency of the process at the utmost level of 95.3, enhances the stability of the SOC by minimizing variations by almost 70 percent and minimizes the maximum thermal stress to about 34 C. The findings validate the hypothesis that the suggested structure is effective at balancing short-term performance and battery health in the long term. The ablation experiment also highlights the significance of aiding the performance improvement through the integration of temporal learning with health-aware optimization, as the Transformer encoder and aging-aware reward mechanism are found to be crucial factors. The solution proposed is a scalable and intelligent solution for the real-time hybrid battery management of electric vehicles in practice.

The approach can reduce the operational cost of the battery, increase system reliability and extend life by preventing degradation and thermal stress prior to their occurrence, which can increase the longevity of the battery. Further development will focus on extending the proposed framework to real-time embedded battery management systems, as well as extend it to be applied to Vehicle-to-Grid (V2G) connectivity and apply multi-agent reinforcement learning techniques to large-scale energy systems. Moreover, the real-world battery datasets and hardware-in-the-loop simulations will be used to demonstrate the applicability of the proposed approach.

The proposed AAT-DRL framework is assessed on the basis of the commonly used driving cycles such as UDDS, NEDC and WLTP, which are widely used in EV energy

management studies to offer standardized and repeatable test conditions. The driving profiles include various urban, highway and mixed-driver modes and thus can be used to perform a comprehensive evaluation of the framework under different load conditions.

While this type of evaluation provides a controlled environment that can allow for analysis of system behavior, it is not able to fully replicate the uncertainties, driver-specific behaviors, environmental changes and aging characteristics of the battery during real-world operation.

Thus, the present work can be regarded as a preliminary validation of the methodology proposed. Future studies aim to validate the AAT-DRL framework with real-world vehicle telemetry data, battery management system logs and experimental platforms with Hardware-in-the-Loop simulations to explore further the practical feasibility, robustness and scalability of the approach in realistic operating scenarios.

Data Availability Statement

Standard driving cycle datasets, including UDDS, NEDC and WLTP, were utilized in this study. These datasets are publicly available from regulatory and open-source repositories such as the U.S. EPA, European Environment Agency and UNECE databases.

Conflicts of Interest

The author has no relevant financial or non-financial interests to disclose.

Funding Statement

The author did not receive support from any organization for the submitted work.

References

[1] Weihan Li et al., "Deep Reinforcement Learning-based Energy Management of Hybrid Battery Systems in Electric Vehicles," *Journal of Energy Storage*, vol. 36, pp. 1-11, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[2] Dezhou Xu et al., "A Soft Actor-Critic-based Energy Management Strategy for Electric Vehicles with Hybrid Energy Storage Systems," *Journal of Power Sources*, vol. 524, pp. 1-13, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[3] Bin Xu et al., "Hierarchical Q-Learning Network for Online Simultaneous Optimization of Energy Efficiency and Battery Life of the Battery/Ultracapacitor Electric Vehicle," *Journal of Energy Storage*, vol. 46, pp. 1-12, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[4] Akhil Hannegudda Ganesh, and Bin Xu, "A Review of Reinforcement Learning based Energy Management Systems for Electrified Powertrains: Progress, Challenge and Potential Solution," *Renewable and Sustainable Energy Reviews*, vol. 154, pp. 1-22, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[5] Lijin Han et al., "Battery Life Constrained Real-Time Energy Management Strategy for Hybrid Electric Vehicles based on Reinforcement Learning," *Energy*, vol. 259, pp. 1-11, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[6] Jianhao Zhou et al., "Self-Learning Energy Management Strategy for Hybrid Electric Vehicle Via Curiosity-Inspired Asynchronous Deep Reinforcement Learning," *Energy*, vol. 242, pp. 1-16, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[7] Yonggang Liu et al., "Energy Management for Hybrid Electric Vehicles based on Imitation Reinforcement Learning," *Energy*, vol. 263, pp. 1-19, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[8] Hanchen Wang et al., "A Comparative Study of 13 Deep Reinforcement Learning based Energy Management Methods for a Hybrid Electric Vehicle," *Energy*, vol. 266, pp. 1-13, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[9] Fei Li et al., "Incentive Learning-based Energy Management for Hybrid Energy Storage System in Electric Vehicles," *Energy Conversion and Management*, vol. 293, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[10] Yiming Ye et al., "An Imitation Learning-based Energy Management Strategy for Electric Vehicles Considering Battery Aging," *Energy*, vol. 283, pp. 1-10, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[11] Xiaolin Tang et al., "Reinforcement Learning-based Energy Management for Hybrid Power Systems: State-of-the-Art Survey, Review and Perspectives," *Chinese Journal of Mechanical Engineering*, vol. 37, no. 1, pp. 1-25, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[12] Kosaku Nakano, Sophia Vögler, and Kenji Tanaka, "Advancing State of Health Estimation for Electric Vehicles: Transformer-based Approach Leveraging Real-World Data," *Advances in Applied Energy*, vol. 16, pp. 1-18, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[13] Yue Wu et al., "Optimal Battery Thermal Management for Electric Vehicles with Battery Degradation Minimization," *Applied Energy*, vol. 353, pp. 1-13, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[14] Yue Wu et al., "Integrated Battery Thermal and Energy Management for Electric Vehicles with Hybrid Energy Storage System: A Hierarchical Approach," *Energy Conversion and Management*, vol. 317, pp. 1-12, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[15] Jiankun Peng et al., "Efficient Training in Energy Management in Fuel Cell Hybrid Electric Vehicles: An Imitation Learning-Embedded Deep Reinforcement Learning Framework," *Journal of Cleaner Production*, vol. 447, pp. 1-12, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[16] Weirong Liu et al., "Imitation Reinforcement Learning Energy Management for Electric Vehicles with Hybrid Energy Storage System," *Applied Energy*, vol. 378, pp. 1-14, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[17] Xiaokang Ma et al., "A Real-Time Intelligent Energy Management based on Deep Reinforcement Learning and Model Predictive Control for Hybrid Electric Vehicles Considering Battery Life," *Energy*, vol. 324, pp. 1-16, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[18] Min Wei et al., "State of Health Prediction of Pure Electric Vehicle Batteries based on Patch Wavelet Transformer," *Mathematics*, vol. 13, no. 13, pp. 1-26, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[19] Hao Chen et al., "Data-Driven Transferred Energy Management Strategy for Hybrid Electric Vehicles Via Deep Reinforcement Learning," *Energy Reports*, vol. 10, pp. 2680-2692, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[20] Weiqi Chen et al., "Health-Considered Energy Management Strategy for Fuel Cell Hybrid Electric Vehicle based on Improved Soft Actor Critic Algorithm Adopted with Beta Policy," *Energy Conversion and Management*, vol. 292, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[21] Chunmei Zhang et al., "Proximal Policy Optimization based Intelligent Energy Management for Plug-in Hybrid Electric Bus Considering Battery Thermal Characteristic," *World Electric Vehicle Journal*, vol. 14, no. 2, pp. 1-23, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[22] Chunchun Jia et al., "Learning-based Model Predictive Energy Management for Fuel Cell Hybrid Electric Bus with Health-Aware Control," *Applied Energy*, vol. 355, pp. 1-13, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[23] Xu Kang, Yujie Wang, and Zonghai Chen, "A Reinforcement Learning Energy Management Strategy for Fuel Cell Hybrid Electric Vehicles Considering Driving Condition Classification," *Sustainable Energy, Grids and Networks*, vol. 38, pp. 1-9, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[24] Zhaofan Xu et al., "State of Health Estimation for Lithium-Ion Batteries based on Incremental Capacity Analysis and Transformer Modeling," *Applied Soft Computing*, vol. 165, pp. 1-12, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[25] Chunchun Jia et al., "Superior Energy Management for Fuel Cell Vehicles Guided by Improved DDPG Algorithm: Integrating Driving Intention Speed Prediction and Health-Aware Control," *Applied Energy*, vol. 394, pp. 1-12, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[26] G. Paganelli et al., "Equivalent Consumption Minimization Strategy for Parallel Hybrid Powertrains," *IEEE 55th Vehicular Technology Conference*, Birmingham, AL, USA, vol. 4, pp. 2076-2081, 2002. [CrossRef] [Google Scholar] [Publisher Link]

[27] Volodymyr Mnih et al., "Human-Level Control through Deep Reinforcement Learning," *Nature*, vol. 518, no. 7540, pp. 529-533, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[28] Timothy P. Lillicrap et al., "Continuous Control with Deep Reinforcement Learning," *arXiv preprint arXiv*, pp. 1-14, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[29] John Schulman et al., "Proximal Policy Optimization Algorithms," *arXiv preprint arXiv*, pp. 1-12, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[30] Scott Fujimoto, Herke Hoof, and David Meger, "Addressing Function Approximation Error in Actor-Critic Methods," *Proceedings of the 35th International Conference on Machine Learning*, vol. 80, pp. 1587-1596, 2018. [Google Scholar] [Publisher Link]

[31] Qizhi He, Mati ur Rahman, and Cuihua Xie, "Information Overflow between Monetary Policy Transparency and Inflation Expectations using Multivariate Stochastic Volatility Models," *Applied Mathematics in Science and Engineering*, vol. 31, no. 1, pp. 1-33, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[32] Aisha M. Alqahtani et al., "Forced Convective Tangent Hyperbolic Nanofluid Flow Subject to Heat Source/Sink and Lorentz Force over a Permeable Wedge: Numerical Exploration," *Nanotechnology Reviews*, vol. 13, no. 1, pp. 1-12, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[33] Muhammad Bilal et al., "Energy Transmission through Radiative Ternary Nanofluid Flow with Exponential Heat Source/Sink Across an Inclined Permeable Cylinder/Plate: Numerical Computing," *Scientific Reports*, vol. 13, no. 1, pp. 1-15, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[34] Khalid Abdulkhaliq M. Alharbi et al., "Stagnation Point Flow of Hybrid Nanofluid Flow Passing over a Rotating Sphere Subjected to Thermophoretic Diffusion and Thermal Radiation," *Scientific Reports*, vol. 13, no. 1, pp. 1-12, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[35] Qizhi He et al., "Dynamic Relationship between Carbon Trading System Resilience and Low-Carbon Stock Market Returns using Time-Varying Granger Causality," *Fractals*, vol. 34, no. 4, pp. 1-15, 2026. [CrossRef] [Google Scholar] [Publisher Link]

[36] Sonia Akram, and Mati ur Rahman, "Multiscale Soliton Structures and Dynamical Analysis of Nonlinear Discrete Electrical Lattices Modeled by the Salerno Equation," *Nonlinear Dynamics*, vol. 113, no. 23, pp. 32723-32744, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[37] Sonia Akram, and Mati Ur Rahman, "Exploring Nonlinear Dynamics and Soliton Structures in the Spin Reduced Hirota-Maxwell-Bloch System Via Atangana's Conformable Operator," *Chinese Journal of Physics*, vol. 97, pp. 1-25, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[38] Seyed Mehdi Rakhtala Rostami, and Zeyad Al-Shibaany, "Reinforcement Learning-based Voltage Regulation of DC Boost Converter in a PEM Fuel Cell Electric Vehicle," *2025 17th International Conference on Advanced Computational Intelligence*, Bath, United Kingdom, pp. 30-36, 2025. [CrossRef] [Google Scholar] [Publisher Link]

[39] Seyed Mehdi Rakhtala Rostami et al., "Deep Reinforcement Learning and Fuzzy Logic Controller Codesign for Energy Management of Hydrogen Fuel Cell Powered Electric Vehicles," *Scientific Reports*, vol. 14, no. 1, pp. 1-20, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[40] S.M. Rakhtala, and E. Shafiee Roudbari, "Application of PEM Fuel Cell for Stand-alone based on a Fuzzy PID Control," *Bulletin of Electrical Engineering and Informatics*, vol. 5, no. 1, pp. 45-61, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[41] Seyed Mehdi Rakhtala et al., "PEM Fuel Cell Voltage-Tracking using Artificial Neural Network," *2009 IEEE Electrical Power and Energy Conference*, Montreal, QC, Canada, pp. 1-5, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[42] H. Dahmardeh, M. Ghanbari, and S.M. Rakhtala, "A Novel Combined DTC Method and SFOC System for Three-Phase Induction Machine Drives with PWM Switching Method," *Journal of Operation and Automation in Power Engineering*, vol. 11, no. 2, pp. 76-82, 2023. [CrossRef] [Google Scholar] [Publisher Link]