

Optimal Charging Management of Electric Vehicles in Distribution Networks Using Deep Reinforcement Learning

Hamed MirFathi¹, Hossein Askarian²,

¹Department of Electrical Engineering, AmirKabir University of Technology, Tehran, Iran
Hamed.mirfathi@aut.ac.ir

²Department of Electrical Engineering, AmirKabir University of Technology, Tehran, Iran
Askarian@aut.ac.ir

Abstract:

The rapid growth of electric vehicles (EVs) has created new opportunities for reducing greenhouse gas emissions and improving energy sustainability. However, the large-scale integration of EVs into power systems can introduce operational challenges for distribution networks if charging processes are not properly coordinated. Uncontrolled charging may lead to increased peak demand, higher network losses, voltage deviations, and potential overloading of distribution feeders. Therefore, developing intelligent charging management strategies for EVs has become an important research topic in modern power systems.

In this paper, a deep reinforcement learning (DRL)-based framework is proposed for optimal charging management of electric vehicles in distribution networks. The main objective of the proposed method is to determine appropriate charging and discharging actions for EVs in order to improve the operational performance of the network while maintaining acceptable battery energy levels. In the proposed approach, the EV charging problem is formulated as a sequential decision-making process in which a learning agent interacts with the network environment. The agent observes the system state, including the operating condition of the distribution network and the state of charge (SOC) of EV batteries, and then selects suitable actions such as charging, discharging, or remaining idle. A reward function is designed to guide the learning process by considering important network performance indices, particularly power losses and voltage profile.

The increasing penetration of EVs has motivated many researchers to investigate different approaches for EV charging scheduling. Conventional methods based on mathematical optimization or metaheuristic algorithms have been widely used in previous studies. Although these methods can provide effective solutions under certain assumptions, they often rely on deterministic information regarding load demand, EV arrival times, and charging requirements. In real-world environments, however, these parameters are uncertain and time-varying. Consequently, traditional optimization approaches may face limitations when dealing with dynamic and uncertain operating conditions. Reinforcement learning offers a promising alternative because it allows the control strategy to be learned directly through interaction with the environment without requiring precise mathematical modeling of uncertainties. In particular, deep reinforcement learning combines reinforcement learning with deep neural networks and enables the handling of complex and high-dimensional decision-making problems in modern power systems.

To evaluate the effectiveness of the proposed approach, the IEEE 33-bus distribution network is used as the test system. Power flow analysis is performed at each simulation step to accurately capture the impact of EV charging and discharging decisions on network operation. The learning agent continuously updates its control policy based on the received reward and gradually learns a charging strategy that improves network performance while maintaining a reasonable SOC level for EV users.

Simulation results demonstrate the effectiveness of the proposed DRL-based charging management strategy. The numerical results indicate that the proposed method significantly improves several key operational indices of the distribution network. Specifically, the total network energy losses decrease from 4.150 MWh to 3.420 MWh, corresponding to a reduction of approximately 17.59%. In addition, the average minimum voltage of the network increases from 0.9410 p.u. to 0.9635 p.u., representing an improvement of about 2.39% in the voltage profile. The proposed strategy also effectively reduces the network peak load from 4.850 MW to 4.120 MW, which indicates a peak reduction of approximately 15.05%.

Furthermore, the results show that the proposed method maintains an appropriate balance between improving grid performance and preserving EV battery energy. The average state of charge of vehicles changes from 0.820 to 0.765, indicating that the algorithm utilizes EV flexibility to support network operation while still maintaining a sufficient level of battery energy for vehicle usage. This demonstrates the capability of the reinforcement learning agent to learn an adaptive and balanced charging policy.

Overall, the obtained results confirm that deep reinforcement learning provides an effective and flexible approach for EV charging management in distribution networks. Unlike conventional optimization techniques that depend on fixed models and predefined scheduling rules, the proposed method can learn an adaptive control policy directly from interaction with the system environment. Therefore, the proposed framework can serve as a promising solution for enhancing the operational efficiency and reliability of distribution networks in the presence of high penetration levels of electric vehicles.

Keywords: Electric Vehicle Charging Management, Deep Reinforcement Learning, Smart Distribution Network, Operational Optimization, Voltage Profile

Date of sending the article: 2026/05/10

Acceptance date of the article: 2026/06/27

Name of the Corresponding Author: Hossein Askarian

Corresponding Author's Address: Department of Electrical Engineering, AmirKabir University of Technology, Tehran, Iran

مدیریت بهینه شارژ خودروهای الکتریکی در شبکه توزیع با استفاده از یادگیری تقویتی عمیق

نوع مطالعه: پژوهشی

میرحامد میرفتحی کمارعلیا، حسین عسکریان ابیانه، ۲

^۱ فارغ التحصیل کارشناسی ارشد مهندسی برق دانشگاه صنعتی امیرکبیر،

Hamed.mirfathi@aut.ac.ir

^۲ گروه مهندسی برق، دانشگاه صنعتی امیرکبیر، تهران، ایران،

Askarian@aut.ac.ir

چکیده:

افزایش نفوذ خودروهای الکتریکی در سال‌های اخیر، چالش‌های فنی قابل توجهی را در بهره‌برداری بهینه از شبکه‌های توزیع ایجاد کرده است؛ به گونه‌ای که شارژ کنترل نشده این وسایل نقلیه می‌تواند موجب افزایش پیک بار، تشدید تلفات توان و انحراف ولتاژ در باس‌ها شود. در این مقاله، یک چارچوب هوشمند مدیریت شارژ و دشارژ خودروهای الکتریکی مبتنی بر یادگیری تقویتی عمیق ارائه می‌شود. در رویکرد پیشنهادی، عامل یادگیری با در نظر گرفتن وضعیت لحظه‌ای شبکه و قیود دینامیکی باتری‌ها، سیاست کنترلی بهینه‌ای را برای زمان‌بندی توان تبادلی خودروها استخراج می‌کند. تابع پاداش به گونه‌ای طراحی شده است که کاهش تلفات شبکه و بهبود پروفیل ولتاژ را به صورت هم‌زمان مدنظر قرار دهد. ارزیابی عملکرد روش پیشنهادی بر روی شبکه استاندارد توزیع IEEE 33-bus و از طریق تحلیل پخش بار در هر گام زمانی انجام شده است. نتایج شبیه‌سازی بیانگر آن است که چارچوب ارائه‌شده می‌تواند ضمن مدیریت مؤثر بار ناشی از خودروهای الکتریکی، شاخص‌های بهره‌برداری شبکه را به طور معناداری بهبود بخشد.

کلمات کلیدی:

مدیریت شارژ خودروهای الکتریکی، یادگیری تقویتی عمیق، شبکه توزیع هوشمند، بهینه‌سازی بهره‌برداری، پروفیل ولتاژ

تاریخ ارسال مقاله: ۱۴۰۵/۰۲/۲۰

تاریخ پذیرش مقاله: ۱۴۰۵/۰۴/۰۶

نام نویسنده مسئول: حسین عسکریان ابیانه

آدرس نویسنده مسئول: گروه مهندسی برق، دانشگاه صنعتی امیرکبیر

۱. مقدمه

افزایش سریع نفوذ خودروهای الکتریکی در سال‌های اخیر، تغییرات قابل توجهی در الگوی مصرف انرژی و نحوه بهره‌برداری از شبکه‌های توزیع ایجاد کرده است. توسعه این فناوری به عنوان یکی از راهکارهای مهم برای کاهش انتشار گازهای گلخانه‌ای و وابستگی به سوخت‌های فسیلی شناخته می‌شود. با این حال، در صورت عدم مدیریت مناسب فرآیند شارژ، حضور تعداد زیادی از خودروهای الکتریکی می‌تواند چالش‌های فنی متعددی برای شبکه‌های توزیع ایجاد کند. شارژ همزمان و کنترل نشده این خودروها ممکن است منجر به افزایش قابل توجه پیک بار، تشدید تلفات توان، افت ولتاژ در باس‌های انتهایی شبکه و حتی اضافه بار در خطوط توزیع شود (Agah & Abbasi, 2012). از این رو، مدیریت هوشمند شارژ و دشارژ خودروهای الکتریکی به یکی از موضوعات مهم در حوزه بهره‌برداری از شبکه‌های توزیع و توسعه شبکه‌های هوشمند تبدیل شده است.

در سال‌های گذشته، پژوهشگران رویکردهای متعددی را برای زمان‌بندی و مدیریت شارژ خودروهای الکتریکی ارائه کرده‌اند. بخش قابل توجهی از این مطالعات مبتنی بر روش‌های بهینه‌سازی کلاسیک نظیر برنامه‌ریزی خطی، برنامه‌ریزی غیرخطی، برنامه‌ریزی عدد صحیح و الگوریتم‌های فراابتکاری بوده است (Sundstrom & Binding, 2011)، (He et al., 2012). در این روش‌ها معمولاً اهدافی نظیر کاهش هزینه انرژی، کاهش تلفات توان، بهبود پروفیل ولتاژ و یا هموارسازی منحنی بار دنبال می‌شود. با وجود مزایای این رویکردها، اغلب آن‌ها به مدل‌های دقیق ریاضی از سیستم و اطلاعات نسبتاً قطعی از پارامترهای محیطی مانند پروفیل بار، زمان ورود و خروج خودروها و سطح شارژ اولیه باتری نیاز دارند. در عمل، رفتار کاربران خودروهای الکتریکی و شرایط بهره‌برداری شبکه دارای عدم قطعیت‌های قابل توجهی است و این موضوع می‌تواند کارایی روش‌های بهینه‌سازی سنتی را کاهش دهد.

در سال‌های اخیر، استفاده از روش‌های مبتنی بر یادگیری ماشین در تحلیل و بهره‌برداری از شبکه‌های قدرت با رشد قابل توجهی همراه بوده است. این روش‌ها در مسائل مختلفی نظیر تشخیص اختلالات کیفیت توان، شناسایی رویدادهای شبکه و تحلیل داده‌های سیستم قدرت مورد استفاده قرار گرفته‌اند. به عنوان نمونه، در (Daliri et al., 2024) از ترکیب تحلیل موجک و الگوریتم‌های یادگیری ماشین برای تشخیص و طبقه‌بندی اغتشاشات کیفیت توان در ایستگاه‌های شارژ خودروهای الکتریکی استفاده شده است. نتایج این پژوهش نشان می‌دهد که استفاده از ویژگی‌های استخراج شده با تبدیل موجک گسسته در کنار الگوریتم Gradient Boost می‌تواند دقت بالایی در تشخیص اغتشاشات شبکه فراهم کند. همچنین در مطالعه (Daliri et al., 2025) یک چارچوب مبتنی بر مدل مارکوف پنهان (HMM) برای استخراج ویژگی از سیگنال‌های رویداد در شبکه قدرت ارائه شده و سپس با استفاده از الگوریتم XGBoost فرآیند تشخیص و طبقه‌بندی انواع خطاها و رویدادهای شبکه انجام شده است. نتایج این تحقیق نشان می‌دهد که ترکیب روش‌های استخراج ویژگی پیشرفته با الگوریتم‌های یادگیری ماشین می‌تواند عملکرد قابل توجهی در تشخیص سریع و دقیق رویدادهای شبکه ارائه دهد.

در ادامه، با گسترش کاربرد یادگیری ماشین در سیستم‌های قدرت، استفاده از یادگیری تقویتی (Reinforcement Learning - RL) برای مسائل تصمیم‌گیری و کنترل در شبکه‌های هوشمند مورد توجه قرار گرفته است. یادگیری تقویتی به دلیل ماهیت تعاملی و قابلیت یادگیری سیاست بهینه در محیط‌های پویا و نامعین، گزینه‌ای مناسب برای مسئله مدیریت شارژ خودروهای الکتریکی محسوب می‌شود (Wan et al., 2018). در این چارچوب، یک عامل یادگیرنده با مشاهده وضعیت محیط و دریافت پاداش ناشی از تصمیمات خود، به تدریج سیاست کنترلی مناسب را می‌آموزد. برخی پژوهش‌ها با استفاده از روش‌هایی نظیر Q-learning و نسخه‌های توسعه‌یافته آن به زمان‌بندی شارژ خودروهای الکتریکی پرداخته‌اند (Qian et al., 2023) و (Yin

(et al., 2022). نتایج این مطالعات نشان می‌دهد که روش‌های مبتنی بر یادگیری تقویتی قادرند بدون نیاز به مدل‌سازی دقیق عدم قطعیت‌ها، راهبردهای کنترلی مؤثری را برای مدیریت بار ناشی از خودروهای الکتریکی استخراج کنند.

با پیشرفت یادگیری عمیق و افزایش توان محاسباتی، ترکیب یادگیری تقویتی با شبکه‌های عصبی عمیق منجر به شکل‌گیری رویکرد یادگیری تقویتی عمیق (Deep Reinforcement Learning – DRL) شده است که در سال‌های اخیر کاربرد گسترده‌ای در مسائل پیچیده کنترل و تصمیم‌گیری پیدا کرده است. در حوزه سیستم‌های قدرت نیز مطالعات متعددی از این رویکرد برای مدیریت منابع انرژی و به‌ویژه زمان‌بندی شارژ خودروهای الکتریکی استفاده کرده‌اند. به عنوان نمونه، در (Liu et al., 2023) یک چارچوب مبتنی بر یادگیری تقویتی عمیق برای زمان‌بندی شارژ خودروهای الکتریکی با در نظر گرفتن پایداری ولتاژ در شبکه توزیع ارائه شده است. در مطالعه (Xiao et al., 2024) نیز یک راهبرد مبتنی بر DRL برای مدیریت هماهنگ شارژ و دشارژ خودروهای الکتریکی با هدف بهبود عملکرد بهره‌برداری شبکه پیشنهاد شده است. همچنین در (Jang & Oh, 2024) یک راهبرد بهره‌برداری مبتنی بر (V2G) با استفاده از یادگیری تقویتی عمیق برای افزایش انعطاف‌پذیری شبکه و مدیریت بار ارائه شده است.

علاوه بر این، برخی مطالعات اخیر تلاش کرده‌اند تا چارچوب‌های یادگیری عمیق را با سایر ابزارهای تحلیل شبکه ترکیب نمایند. برای مثال، در پژوهش (Rashid et al., 2025) از یادگیری تقویتی عمیق در کنار پیش‌بینی بار ناشی از شارژ خودروهای الکتریکی برای بهینه‌سازی توان راکتیو در شبکه توزیع استفاده شده است. همچنین در برخی پژوهش‌های جدید، استفاده از الگوریتم‌های پیشرفته‌تر یادگیری تقویتی نظیر Deep Q-Network (DQN)، Deep Deterministic Policy Gradient (DDPG) و

Optimization (PPO) برای مدیریت منابع انرژی در شبکه‌های توزیع مورد بررسی قرار گرفته است (Li et al., 2026) و (Sharif et al., 2026). این مطالعات نشان می‌دهد که روش‌های DRL قادرند در محیط‌های پیچیده و دارای عدم قطعیت، سیاست‌های کنترلی کارآمدی را برای مدیریت بار و بهبود شاخص‌های بهره‌برداری شبکه استخراج کنند.

با وجود پیشرفت‌های قابل توجه در این حوزه، همچنان چالش‌ها و خلأهایی در ادبیات موجود مشاهده می‌شود. نخست آنکه در بسیاری از مطالعات، تمرکز اصلی بر زمان‌بندی شارژ خودروهای الکتریکی بوده و امکان تبادل دوطرفه انرژی و نقش دشارژ خودروها در بهبود بهره‌برداری شبکه به‌طور کامل مورد بررسی قرار نگرفته است. دوم آنکه در برخی پژوهش‌ها تنها یک شاخص عملکردی نظیر هزینه یا تلفات توان مدنظر قرار گرفته و اثر همزمان تصمیمات کنترلی بر شاخص‌های مختلف شبکه از جمله تلفات توان و پروفیل ولتاژ به صورت جامع ارزیابی نشده است. علاوه بر این، در برخی مطالعات تحلیل عملکرد روش پیشنهادی بدون در نظر گرفتن محاسبات دقیق جریان توان در هر گام زمانی انجام شده است که می‌تواند منجر به برآوردهای غیرواقعی از عملکرد شبکه شود. همچنین در برخی موارد جزئیات مربوط به ساختار شبکه عصبی، فرآیند آموزش عامل یادگیری و معیارهای ارزیابی عملکرد به طور کامل گزارش نشده است که مقایسه نتایج را با سایر مطالعات دشوار می‌سازد.

در جدول ۱ تعدادی از مطالعات پیشین مورد بررسی قرار گرفته و مزایا و محدودیت‌های آن‌ها ارائه شده است. همان‌گونه که در جدول مرور ادبیات مشاهده می‌شود، پژوهشگران با استفاده از رویکردهای مختلفی از جمله مدل‌های ریاضی، روش‌های بهینه‌سازی کلاسیک، الگوریتم‌های فراابتکاری و یادگیری تقویتی به بررسی مسئله مدیریت شارژ خودروهای الکتریکی در شبکه‌های توزیع پرداخته‌اند. هر یک از این رویکردها دارای مزایا و قابلیت‌های خاص خود هستند، اما در عین حال با محدودیت‌هایی نظیر وابستگی به

- استفاده از چارچوب یادگیری تقویتی عمیق برای استخراج سیاست کنترلی کارا بدون نیاز به فرضیات محدودکننده در مورد رفتار قطعی بار و عدم قطعیت‌های مرتبط با حضور خودروهای الکتریکی.

در این مقاله، عامل یادگیری با مشاهده وضعیت لحظه‌ای شبکه توزیع و سطح شارژ باتری خودروهای الکتریکی، اقدام مناسب شامل شارژ، عدم شارژ یا دشارژ را انتخاب می‌کند تا شاخص‌های بهره‌برداری شبکه بهبود یابد. نتایج شبیه‌سازی نشان می‌دهد که چارچوب پیشنهادی می‌تواند با مدیریت هوشمند بار ناشی از خودروهای الکتریکی، به کاهش تلفات توان، بهبود پروفیل ولتاژ و کاهش پیک بار در شبکه توزیع کمک نماید.

ساختار ادامه مقاله به این صورت است: در بخش ۲، مدل‌سازی شبکه توزیع و پارکینگ خودروهای الکتریکی تشریح می‌شود. در بخش ۳، مسئله مدیریت شارژ به صورت ریاضی فرمول‌بندی شده و چارچوب یادگیری تقویتی عمیق معرفی می‌گردد. در بخش ۴، نتایج شبیه‌سازی و تحلیل عملکرد روش پیشنهادی ارائه می‌شود. در نهایت، بخش ۵ به جمع‌بندی و پیشنهادات برای تحقیقات آتی اختصاص دارد.

مدل‌های دقیق سیستم، نیاز به پیش‌بینی‌های قطعی از وضعیت بار و یا دشواری در تصمیم‌گیری‌های لحظه‌ای در شرایط عدم قطعیت مواجه‌اند. از این رو، توسعه روش‌هایی که بتوانند ضمن بهره‌گیری از مزایای رویکردهای موجود، محدودیت‌های آن‌ها را کاهش دهند، از اهمیت بالایی برخوردار است.

در این راستا، پژوهش حاضر یک چارچوب مبتنی بر یادگیری تقویتی عمیق برای مدیریت هوشمند شارژ و دشارژ خودروهای الکتریکی در شبکه‌های توزیع ارائه می‌دهد. نوآوری و برتری کار حاضر نسبت به مطالعات پیشین را می‌توان در چند محور خلاصه نمود:

- ارائه یک چارچوب تصمیم‌گیری تطبیقی که در آن فرآیند شارژ و دشارژ خودروهای الکتریکی به صورت گام‌به‌گام و بر اساس وضعیت لحظه‌ای شبکه تعیین می‌شود.
- در نظر گرفتن همزمان دو شاخص کلیدی بهره‌برداری شبکه شامل تلفات توان، انحراف ولتاژ و وضعیت شارژ باتری خودروها (SOC) در طراحی تابع پاداش عامل یادگیری.
- مدل‌سازی دقیق شبکه توزیع استاندارد IEEE 33-bus و انجام محاسبات جریان توان در هر گام زمانی به منظور ارزیابی واقعی اثر تصمیمات کنترلی بر عملکرد شبکه.

جدول ۱. جدول مرور ادبیات و مقایسه روش‌های مدیریت شارژ خودرو الکتریکی

محدودیت‌ها	مزایا	نوآوری	شبکه مورد مطالعه	هدف اصلی	رویکرد پژوهش	مرجع
نیاز به مدل دقیق شبکه و وابستگی شدید به پیش‌بینی قطعی	دقت محاسباتی بالا و تضمین بهینگی سراسری	استفاده از برنامه‌ریزی مخروط مرتبه دوم برای حل دقیق مسئله	شبکه توزیع استاندارد	کمینه‌سازی تلفات و تعادل بار	مدل‌سازی ریاضی (SOCP)	Qian et al., 2023
عدم توانایی مدل‌سازی رفتار غیرخطی و عدم قطعیت شدید	سرعت بالا و ساختار خطی ساده	مدل‌سازی اقتصادی دقیق با قیود فنی شبکه	پارکینگ متصل به شبکه توزیع	بیشینه‌سازی سود پارکینگ‌های هوشمند	بهینه‌سازی خطی (MILP)	Liu et al., 2023

Xiao et al., 2024	الگوریتم‌های فراابتکاری (ترکیب PSO و الگوریتم‌های نوین)	مدیریت شارژ و دشارژ با هدف کاهش هزینه	شبکه توزیع با نفوذ EV	ترکیب الگوریتم‌های تکاملی برای حل مسئله غیرمحدب	انعطاف‌پذیری بالا در مسائل پیچیده	زمان اجرای بالا و احتمال همگرایی به بهینه محلی
Jang & Oh, 2024	یادگیری تقویتی عمیق (DRL)	زمان‌بندی شارژ EV با تمرکز بر پایداری ولتاژ	شبکه توزیع شعاعی	استفاده از DRL برای کنترل تطبیقی ولتاژ	عملکرد مناسب در محیط پویا	نیاز به تنظیم دقیق پارامترهای یادگیری
Rashid et al., 2025	DRL برای شارژ و دشارژ هماهنگ EV	بهبود بهره‌برداری شبکه و مدیریت بار	شبکه توزیع با نفوذ بالای EV	تصمیم‌گیری تطبیقی مبتنی بر وضعیت لحظه‌ای شبکه	مدیریت مؤثر عدم قطعیت	طراحی پیچیده تابع پاداش
Li et al., 2026	DRL مبتنی بر V2G	افزایش انعطاف‌پذیری شبکه با تبادل دوطرفه انرژی	شبکه توزیع هوشمند	بهره‌گیری از ظرفیت باتری EV برای پشتیبانی شبکه	بهبود پایداری و مدیریت پیک	وابستگی به زیرساخت V2G
Sharif et al., 2026	DRL + پیش‌بینی بار EV	بهینه‌سازی توان راکتیو و پروفیل ولتاژ	شبکه توزیع استاندارد	ترکیب پیش‌بینی بار با کنترل مبتنی بر DRL	بهبود همزمان چند شاخص بهره‌برداری	افزایش پیچیدگی مدل و نیاز به داده پیش‌بینی
پژوهش حاضر	یادگیری تقویتی عمیق (DRL) با در نظر گرفتن محاسبه دقیق پخش بار	بهبود همزمان ولتاژ، کاهش تلفات و مدیریت پویا شارژ/دشارژ EV	شبکه استاندارد IEEE 33-bus	در نظر گرفتن همزمان چند شاخص عملکرد + محاسبه پخش بار در هر گام زمانی + تصمیم‌گیری آنی در شرایط عدم قطعیت	سازگاری بالا با شرایط پویا، عدم نیاز به مدل قطعی رفتار کاربران، عملکرد چندهدفه	نیاز به مرحله آموزش و هزینه محاسباتی در فاز یادگیری

۲. مدل‌سازی ریاضی مسئله

۲-۱. مدل شبکه توزیع

شبکه توزیع به صورت یک شبکه رادیال سه‌فاز معادل تک‌فاز متعادل مدل می‌شود. مجموعه باس‌ها و خطوط به ترتیب با (۱) و (۲) نشان داده می‌شوند.

$$N = \{1, 2, \dots, N\}, \quad (1)$$

$$L = \{1, 2, \dots, L\} \quad (2)$$

در روابط بالا، N بیانگر تعداد باس‌ها و L نشان‌دهنده خطوط می‌باشند.

باس ۱ به‌عنوان باس مرجع در نظر گرفته می‌شود و سایر باس‌ها به‌صورت رادیال به آن متصل هستند. برای هر خط $l \in L$ که بین باس‌های i و j قرار دارد، امپدانس معادل به صورت (۳) تعریف می‌شود.

$$Z_{ij} = R_{ij} + jX_{ij} \quad (3)$$

که Z_{ij} ، نشان‌دهنده امپدانس هر خط، R_{ij} و X_{ij} به ترتیب نمایانگر مقاومت و راکتانس خطوط می‌باشند.

توان اکتیو و راکتیو تزریقی در هر باس i در گام زمانی t به صورت (۴) تعریف می‌شود.

$$P_i(t) = P_{L,i}(t) + P_{EV,i}(t), \quad (4)$$

$$Q_i(t) = Q_{L,i}(t)$$

که در آن $P_{L,i}(t)$ و $Q_{L,i}(t)$ به ترتیب توان اکتیو و راکتیو بارهای پایه در باس i و $P_{EV,i}(t)$ توان خالص ناشی از شارژ/دشارژ خودروهای الکتریکی در باس i است (مثبت برای شارژ، منفی برای دشارژ). در این مقاله، بارهای خودروهای الکتریکی به‌صورت توان

$$\varepsilon = \{1, 2, \dots, N_{EV}\} \quad (8)$$

$$SOC_e(t) = \{SOC_e^{\min}, SOC_e^{\max}\} \quad (9)$$

دینامیک SOC برای هر خودروی الکتریکی به صورت گسسته و در افق زمانی T مدل می‌شود. نرخ شارژ/دشارژ در گام زمانی Δt برای خودروی e با $P_e(t)$ نمایش داده می‌شود که مثبت برای شارژ و منفی برای دشارژ است. با فرض طول هر گام زمانی Δt (بر حسب ساعت)، رابطه به‌روزرسانی SOC به صورت زیر است:

$$\begin{aligned} SOC_e(t+1) &= SOC_e(t) + \eta_c \frac{\max(P_e(t), 0) \Delta t}{E_e^{cap}} \\ &\quad - \frac{1}{\eta_d} \frac{\max(-P_d(t), 0) \Delta t}{E_e^{cap}} \end{aligned} \quad (10)$$

که در آن η_c راندمان شارژ و η_d راندمان دشارژ است. محدودیت‌های فیزیکی SOC به صورت زیر باید رعایت شود:

$$SOC_e^{\min} \leq SOC_e(t) \leq SOC_e^{\max} \quad (11)$$

همچنین توان تبدیلی خودرو با شبکه در هر لحظه محدود به توان نامی شارژر است:

$$-P_e^{max,dis} \leq P_e(t) \leq P_e^{max,ch} \quad (12)$$

۳-۲. تابع هدف و محدودیت‌های بهره‌برداری

هدف نهایی از مدیریت هوشمند، کمینه‌سازی یک تابع هزینه چندهدفه شامل تلفات توان و انحراف ولتاژ در کل دوره بهره‌برداری (T) است:

$$\min \sum_{t=1}^T \left(\alpha \cdot P_{loss}(t) + \beta \cdot \sum_{i=1}^N |V_i(t) - V_{ref}| \right) \quad (13)$$

اکتیو با ضریب توان واحد در نظر گرفته می‌شوند، لذا تغییری در بار راکتیو ایجاد نمی‌کنند.

معادلات توان برای هر باس i (غیر از باس مرجع) به صورت کلی از روابط توان جریان به دست می‌آید. در فرمول‌بندی عمومی جریان توان، تساوی توان اکتیو و راکتیو به صورت (۵) بیان می‌شود.

$$\begin{aligned} P_i(t) &= V_i(t) \sum_{j \in N} V_j(t) G_{ij} \cos \theta_{i,j}(t) \\ &\quad + B_{i,j} \sin \theta_{i,j}(t), \\ Q_i(t) &= V_i(t) \sum_{j \in N} V_j(t) (G_{ij} \sin \theta_{i,j}(t) \\ &\quad - B_{i,j} \cos \theta_{i,j}(t)) \end{aligned} \quad (5)$$

بیان می‌شود که در آن $V_i(t)$ اندازه ولتاژ، $\theta_{i,j}(t) = \delta_i(t) - \delta_j(t)$ اختلاف زاویه ولتاژ بین باس‌های i و j و G_{ij} و B_{ij} به ترتیب المان‌های ماتریس ادمیتانس شبکه هستند. حل این معادلات در هر گام زمانی، پروفیل ولتاژ و تلفات خطوط را تعیین می‌کند.

توان اکتیو تلفاتی در هر خط در گام زمانی t به صورت (۶) و تلفات کل شبکه بصورت (۷) تعریف می‌شود.

$$P_{ij}^{loss}(t) = R_{ij} \frac{P_{ij}^2(t) + Q_{ij}^2(t)}{V_i^2(t)} \quad (6)$$

$$P_{loss}(t) = \sum_{(i,j) \in L} P_{ij}^{loss}(t) \quad (7)$$

۲-۲. مدل خودروهای الکتریکی و دینامیک SOC

فرض می‌شود مجموعه‌ای از خودروهای الکتریکی ε در برخی باس‌های شبکه متصل هستند. هر خودروی الکتریکی $e \in \varepsilon$ دارای ظرفیت باتری نامی E_e^{cap} بر حسب (kWh) و حالت شارژ در گام زمانی t به صورت (۹) تعریف می‌شود.

به طوری که محدودیت‌های زیر برقرار باشد:

$$V^{min} \leq V_i(t) \leq V^{max} \quad (14)$$

$$I_{ij}(t) \leq I_{ij}^{max} \quad (15)$$

۳. متدولوژی: یادگیری تقویتی عمیق (DRL)

در این بخش، مسئله مدیریت شارژ خودروهای الکتریکی در قالب یک فرآیند تصمیم‌گیری مارکوف (MDP) فرمول‌بندی شده و الگوریتم شبکه Q عمیق (DQN) برای حل آن معرفی می‌گردد.

۳-۱. فرمول‌بندی MDP

یک مسئله یادگیری تقویتی شامل چهار مؤلفه اصلی است: وضعیت، اقدام، پاداش و نرخ تنزیل.

۱. فضای وضعیت ($\mathcal{S}$): وضعیت سیستم در هر گام زمانی t شامل اطلاعات محیطی و داخلی است:

$$s_t = \{t, P_{L,total}(t), SOC(t)\} \quad (16)$$

که در آن t زمان فعلی، $P_{L,total}(t)$ مجموع بار پایه شبکه و $SOC(t)$ بردار وضعیت شارژ تمامی خودروهای الکتریکی است.

۲. فضای اقدام ($\mathcal{A}$): در هر گام، عامل (Agent) باید برای هر خودرو یکی از سه تصمیم زیر را اتخاذ کند:

• $a = 0$: عدم شارژ (Idle)

• $a = 1$: شارژ با توان نامی (Charging)

• $a = 2$: دشارژ و تزریق توان به شبکه

(Discharging - V2G)

۳. تابع پاداش ($\mathcal{R}$): پاداش، بازخورد محیط به اقدام انجام شده است. برای هم‌راستایی با تابع هدف، پاداش به صورت معکوس با تلفات و انحراف ولتاژ تعریف می‌شود:

$$r_t = - \left(w_1 \cdot P_{loss}(t) + w_2 \cdot \sum_{i \in N} (V_i(t) - 1)^2 + w_3 \cdot Penalty_{SOC} \right) \quad (17)$$

در اینجا $Penalty_{SOC}$ جریمه‌ای برای جلوگیری از تخلیه بیش از حد باتری یا خارج شدن از محدوده مجاز است.

در تابع پاداش فوق، چندین شاخص عملکردی شامل تلفات شبکه، انحراف ولتاژ و وضعیت شارژ خودروهای الکتریکی در نظر گرفته شده است. برای ترکیب این شاخص‌ها، ضرایب وزنی به هر یک از مؤلفه‌ها اختصاص داده شد. انتخاب این ضرایب بر اساس تحلیل حساسیت و آزمون‌های تجربی انجام گرفت، به گونه‌ای که تعادل مناسبی میان کاهش تلفات شبکه، بهبود پروفیل ولتاژ و حفظ سطح مناسب شارژ خودروها برقرار شود. در فرآیند تنظیم پارامترها، مقادیر مختلفی برای وزن‌ها مورد بررسی قرار گرفت و در نهایت ترکیبی انتخاب شد که بهترین عملکرد کلی شبکه را از نظر شاخص‌های بهره‌برداری فراهم می‌کرد.

۳-۲. الگوریتم Deep Q-Network (DQN)

در محیط‌های با فضای وضعیت پیوسته و بزرگ، استفاده از جدول Q امکان‌پذیر نیست. DQN از یک شبکه عصبی عمیق با پارامترهای θ برای تخمین تابع ارزش $Q(s, a; \theta)$ استفاده می‌کند. هدف آموزش، کمینه‌سازی تابع زیان زیر است:

$$L(\theta) = E \left[(y_t - Q(s_t, a_t; \theta))^2 \right] \quad (18)$$

که در آن هدف (y_t) از رابطه بلمن (۱۹) به دست می‌آید:

$$y_t = r_t + \gamma \max_{a'} Q(s_{t+1}, a'; \theta^-) \quad (19)$$

در این رابطه، γ نرخ تنزیل و θ^- پارامترهای شبکه هدف هستند که برای پایداری آموزش به صورت دوره‌ای به‌روزرسانی می‌شوند. از مکانیزم "تجربه بازپخش" نیز برای از بین بردن همبستگی بین داده‌های آموزشی استفاده می‌گردد.

۳-۳. ساختار شبکه عصبی

برای تقریب تابع ارزش در الگوریتم DQN از یک شبکه عصبی پیشخور (Feedforward Neural Network) استفاده شده است. ورودی شبکه شامل بردار وضعیت تعریف شده در معادله (۱۶) بوده و خروجی شبکه مقدار تابع Q برای اقدامات مختلف عامل است. ساختار شبکه عصبی به صورت زیر در نظر گرفته شده است:

- لایه ورودی: شامل متغیرهای وضعیت سیستم
- لایه پنهان اول: ۱۲۸ نورون با تابع فعال سازی ReLU
- لایه پنهان دوم: ۶۴ نورون با تابع فعال سازی ReLU
- لایه خروجی: سه نورون متناظر با اقدامات ممکن عامل (شارژ، عدم شارژ، دشارژ)

برای بهروزرسانی وزن های شبکه از بهینه ساز Adam استفاده شده و تابع هزینه برابر با میانگین مربعات خطا (MSE) در نظر گرفته شده است.

۳-۴. فرآیند آموزش عامل یادگیری

در این پژوهش، عامل یادگیری از طریق تعامل مکرر با محیط شبیه سازی شبکه توزیع آموزش داده می شود. در هر اپیزود، عامل با مشاهده وضعیت سیستم اقدام مناسب را انتخاب کرده و پس از اجرای آن، وضعیت جدید شبکه و مقدار پاداش متناظر دریافت می شود. این فرآیند تا پایان افق زمانی بهره برداری ادامه می یابد.

برای ایجاد تعادل میان اکتشاف (Exploration) و بهره برداری (Exploitation) از سیاست ϵ -greedy استفاده شده است. در ابتدای فرآیند آموزش مقدار ϵ بزرگ انتخاب می شود تا عامل بتواند اقدامات مختلف را بررسی کند. سپس با پیشرفت فرآیند آموزش، مقدار ϵ به تدریج کاهش می یابد تا عامل بیشتر از سیاست آموخته شده استفاده کند.

در هر گام زمانی، تجربه حاصل به صورت یک چهارتایی شامل:

$$(s_t, a_t, r_t, s_{t+1})$$

در حافظه تجربه ذخیره شده و در مراحل بعدی برای آموزش شبکه مورد استفاده قرار می گیرد. در هر مرحله آموزش، مجموعه ای

تصادفی از این نمونه ها از حافظه انتخاب شده و برای بهروزرسانی پارامترهای شبکه مورد استفاده قرار می گیرد.

۳-۵. پارامترهای آموزش

پارامترهای اصلی مورد استفاده در فرآیند آموزش عامل یادگیری در این مطالعه در جدول ۲ ارائه شده اند.

جدول ۲. جدول پارامترهای اصلی در فرآیند آموزش عامل

یادگیری

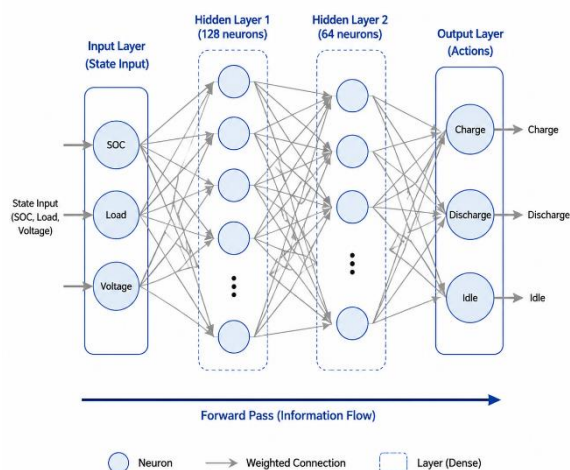
پارامتر	مقدار
Learning rate	۰.۰۰۱
Discount factor (γ)	۰.۹۵
Replay buffer size	۵۰۰۰
Batch size	۶۴
Target network update	هر ۲۰۰ گام
Initial	۱
Final	۰.۰۵
Number of training episodes	۱۰۰۰
Steps per episode	۲۴

این پارامترها بر اساس مطالعات پیشین در حوزه یادگیری تقویتی و همچنین آزمایش های تجربی تنظیم شده اند تا پایداری آموزش و همگرایی مناسب مدل تضمین گردد

ساختار شبکه عصبی پیشنهادی برای الگوریتم DQN در شکل ۱ نشان داده شده است.

شکل ۱. ساختار شبکه عصبی پیشنهادی

Deep Neural Network (DQN) Architecture



۳-۶. معیار همگرایی آموزش

فرآیند آموزش تا زمانی ادامه می‌یابد که عملکرد عامل به حالت پایدار برسد. برای ارزیابی همگرایی، میانگین پاداش دریافتی در چند اپیزود متوالی بررسی شده است. زمانی که تغییرات میانگین پاداش و شاخص‌های عملکرد شبکه شامل تلفات توان و انحراف ولتاژ کمتر از مقدار مشخصی باشد، فرآیند آموزش خاتمه یافته و سیاست آموخته‌شده به عنوان سیاست کنترلی نهایی در نظر گرفته می‌شود.

۴. نتایج شبیه‌سازی

۴-۱. سیستم مورد مطالعه و تنظیمات مدل

در این پژوهش، یک پارکینگ متمرکز خودروهای الکتریکی با ظرفیت ۲۰ خودرو در نظر گرفته شده است. این پارکینگ به عنوان یک موجودیت واحد (Aggregator) در باس ۳۳ شبکه IEEE 33-bus (دورترین باس از شینه اصلی) متصل شده است تا اثرات مدیریتی بر بهبود پروفیل ولتاژ در بحرانی‌ترین نقطه شبکه بهتر مشاهده شود.

پارامترهای فنی خودروهای مستقر در پارکینگ و تنظیمات سیستم در جدول ۳ ارائه شده است. فرض بر این است که اپراتور پارکینگ با استفاده از الگوریتم پیشنهادی، توان تبادلی کل پارکینگ با شبکه را در هر گام زمانی کنترل می‌کند.

جدول ۳. جدول پارامترهای اصلی در فرآیند آموزش عامل

یادگیری

واحد	مقدار	پارامتر
-	۲۰	تعداد خودروهای الکتریکی (N_{EV})
kWh	۴۰	ظرفیت باتری E_{batt}
kW	۷.۵	توان نامی شارژ/دشارژ P_{max}
-	۰.۹۵	راندمان شارژ/دشارژ
-	۹۰٪-۲۰٪	محدوده مجاز SOC

به منظور ارزیابی عملکرد راهبرد پیشنهادی مدیریت هوشمند شارژ خودروهای الکتریکی مبتنی بر یادگیری تقویتی DQ-Learning، شبیه‌سازی‌هایی در بازه زمانی ۲۴ ساعت انجام شده است.

در این آزمایش، وضعیت شارژ بدون مدیریت با شارژ هوشمند پیشنهادی مورد مقایسه قرار گرفته است. شاخص‌های ارزیابی شامل موارد زیر هستند:

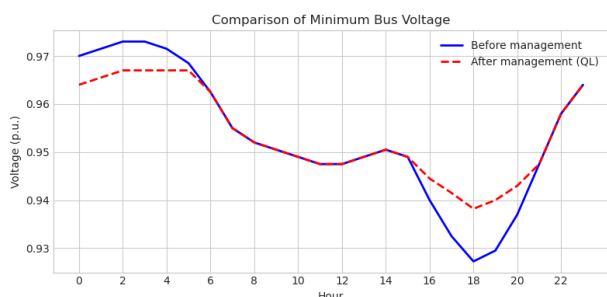
- تلفات کل شبکه توزیع
- پروفیل حداقل ولتاژ باس‌ها
- میانگین حالت شارژ (SOC) خودروهای الکتریکی
- بار اوج شبکه

الف) پروفیل ولتاژ

شکل ۲ مقایسه حداقل ولتاژ باس در طول شبانه‌روز را نشان می‌دهد.

در حالت بدون مدیریت، افت ولتاژ قابل توجهی در ساعات اوج مصرف (حدود ساعت ۱۷ تا ۱۹) مشاهده می‌شود، به‌طوری که مقدار حداقل ولتاژ به حدود ۰.۹۳ واحد پریونیت کاهش می‌یابد. این افت عمدتاً ناشی از شارژ همزمان خودروها در دوره‌های اوج بار شبکه است.

پس از بکارگیری الگوریتم یادگیری تقویتی، افت ولتاژ در ساعات بحرانی کاهش یافته و میانگین حداقل ولتاژ بهبود یافته است. (از ۰.۹۴۱۰ پریونیت به ۰.۹۶۳۵ پریونیت افزایش پیدا کرده است)، که نشان‌دهنده بهبود ۲.۳۹ درصدی در کیفیت ولتاژ است. این افزایش بیانگر توانایی الگوریتم در توزیع هوشمند توان شارژ و جلوگیری از افت ولتاژ در ساعات پرتراфик شبکه است.



شکل ۲. مقایسه پروفیل حداقل ولتاژ باس‌ها قبل و بعد از اعمال مدیریت هوشمند شارژ مبتنی بر الگوریتم DQ-Learning.

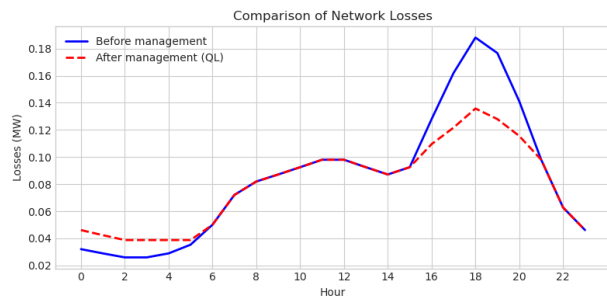
ب) تلفات شبکه

مقایسه تلفات شبکه در ساعات مختلف در شکل ۳ نشان داده شده است.

در حالت شارژ بدون مدیریت، تلفات شبکه در ساعات ۱۸ تا ۲۰ به طور محسوس افزایش یافته و مقدار کل انرژی تلف شده در شبانه روز معادل ۴.۱۵ مگاوات ساعت است.

پس از اعمال راهبرد پیشنهادی، شارژ خودروها به ساعات کم بار انتقال یافته و جریان خطوط متعادل تر شده است. در نتیجه، تلفات کل انرژی به ۳.۴۲ مگاوات ساعت کاهش یافته که معادل ۱۷.۵۹ درصد کاهش در تلفات شبکه است.

این کاهش قابل توجه نشان دهنده توانایی روش یادگیری تقویتی در بهبود بهره‌وری شبکه توزیع بدون نیاز به ارتقای سخت‌افزاری است.



شکل ۳. مقایسه تلفات ساعتی شبکه در حالت شارژ بدون مدیریت و شارژ هوشمند مبتنی بر یادگیری تقویتی.

ج) رفتار شارژ و انتقال بار

پروفیل توان شارژ و دشارژ خودروهای الکتریکی در شکل ۴ نشان داده شده است.

در حالت بدون مدیریت، شارژ عمدتاً در ساعات اوج بار انجام می‌شود که فشار زیادی بر شبکه وارد می‌سازد. در مقابل، با اجرای راهبرد هوشمند:

شارژ خودروها به ساعات بامداد و کم‌باری منتقل شده است.

در ساعات اوج مصرف، خودروها به صورت دشارژ (V2G) عمل کرده و به شبکه توان تزریق می‌کنند.

توان شارژ پیک به طور محسوس کاهش یافته است.

نتیجه این رفتار هوشمند در شکل ۴ قابل مشاهده است که نشان‌دهنده کاهش بار اوج شبکه می‌باشد.

با مدیریت پیشنهادی، پیک بار از مقدار ۴.۸۵ مگاوات به ۴.۱۲۰ مگاوات کاهش یافته که معادل ۱۵.۰۵ درصد کاهش بار اوج است. این نتیجه بیانگر عملکرد پیک‌زدایی مؤثر (Peak Shaving) توسط عامل یادگیری است که منجر به استفاده بهینه‌تر از ظرفیت شبکه شده است.

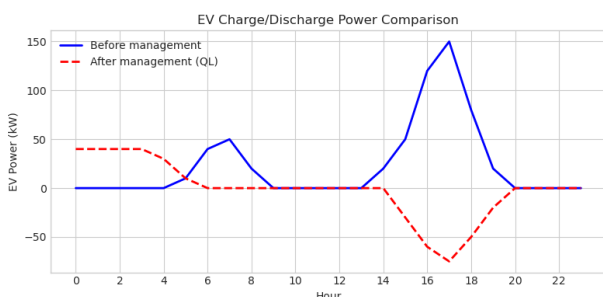
د) تحلیل وضعیت SOC خودروها

شکل ۵ تغییرات میانگین SOC خودروهای الکتریکی در طول شبانه‌روز را نشان می‌دهد.

در حالت بدون مدیریت، شارژ سریع و همزمان در ساعات اوج منجر به افزایش SOC تا حدود ۰.۹ واحد و متوسط روزانه ۰.۸۲ شده است.

در حالت هوشمند، الگوریتم با کنترل فرآیند شارژ و دشارژ، میانگین SOC را در مقدار ۰.۷۶۵ نگه داشته که بیانگر کاهش ۶.۷۰ درصدی نسبت به حالت بدون مدیریت است.

این کاهش جزئی بیانگر یک تبادل مثبت بین پایداری شبکه و سطح شارژ باتری است؛ به عبارتی، الگوریتم با فدا کردن مقدار کمی از انرژی ذخیره‌شده، عملکرد شبکه را از نظر پایداری و ولتاژ و کاهش بار بهبود می‌دهد.



شکل ۴. مقایسه پروفیل توان شارژ و دشارژ خودروهای الکتریکی قبل و بعد از مدیریت هوشمند.

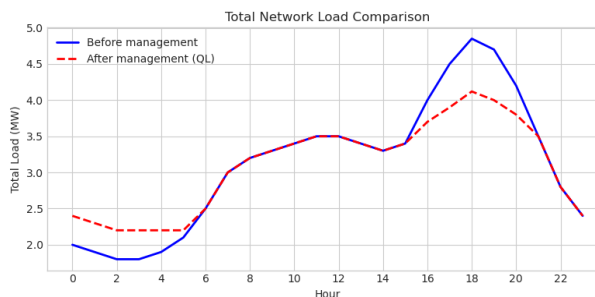
شامل الگوریتم ازدحام ذرات (PSO) و الگوریتم ژنتیک (GA) مقایسه شده است. همان‌طور که در جدول ۴ مشاهده می‌شود، در حالت بدون مدیریت تلفات کل شبکه برابر با ۴.۱۵۰ مگاوات ساعت است. استفاده از روش‌های PSO و GA این مقدار را به ترتیب به حدود ۳.۶۸۰ و ۳.۵۹۰ مگاوات ساعت کاهش داده‌اند. با این حال، روش پیشنهادی مبتنی بر یادگیری تقویتی عمیق توانسته است تلفات شبکه را تا مقدار ۳.۴۲۰ مگاوات ساعت کاهش دهد که نشان‌دهنده عملکرد بهتر این روش در کاهش تلفات شبکه توزیع است.

از نظر پروفیل ولتاژ نیز حداقل ولتاژ متوسط شبکه در حالت بدون مدیریت برابر با ۰.۹۴۱۰ پرینیت بوده، در حالی که این مقدار در روش‌های PSO و GA به ترتیب به ۰.۹۵۴۰ و ۰.۹۵۸۰ پرینیت افزایش یافته است. در روش پیشنهادی این شاخص به ۰.۹۶۳۵ پرینیت رسیده که بیانگر بهبود بیشتر وضعیت ولتاژ در باس‌های شبکه می‌باشد.

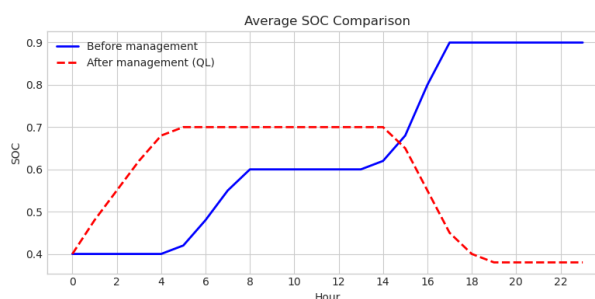
همچنین در شاخص بار اوج شبکه مشاهده می‌شود که مقدار آن از ۴.۸۵۰ مگاوات در حالت بدون مدیریت به حدود ۴.۳۶۰ و ۴.۲۵۰ مگاوات در روش‌های PSO و GA کاهش یافته است، در حالی که روش پیشنهادی توانسته است این مقدار را به ۴.۱۲۰ مگاوات برساند. این موضوع نشان‌دهنده توانایی بالاتر روش پیشنهادی در هموارسازی منحنی بار و مدیریت هوشمند فرآیند شارژ و دشارژ خودروهای الکتریکی است.

در مجموع، نتایج مقایسه‌ای نشان می‌دهد که رویکرد مبتنی بر یادگیری تقویتی عمیق قادر است عملکرد بهتری نسبت به روش‌های بهینه‌سازی متداول در بهبود شاخص‌های بهره‌برداری شبکه توزیع شامل تلفات، پروفیل ولتاژ و بار اوج ارائه دهد.

به‌طور کلی، عامل یادگیری توانسته است راهبرد شارژی را بیاموزد که تعادل مناسبی میان بهره‌وری شبکه و وضعیت ذخیره انرژی خودروها ایجاد می‌کند، بدون آنکه از قواعد ثابت یا زمان‌بندی از پیش تعیین‌شده استفاده شود.



شکل ۵. مقایسه پروفیل بار کل شبکه و کاهش بار اوج در اثر اجرای الگوریتم یادگیری تقویتی.



شکل ۶. مقایسه میانگین حالت شارژ (SOC) خودروهای الکتریکی در حالت بدون مدیریت و حالت هوشمند مبتنی بر یادگیری تقویتی.

۵) ارزیابی کلی عملکرد

جدول ۴ جمع‌بندی نتایج شبیه‌سازی را نشان می‌دهد:

جدول ۴. جمع‌بندی نتایج شبیه‌سازی

شاخص عملکرد	بدون مدیریت	PSO	GA	روش پیشنهادی (DQN)
تلفات کل شبکه (MWh)	۴.۱۵۰	۳.۶۸۰	۳.۵۹۰	۳.۴۲۰
میانگین حداقل ولتاژ (p.u)	۰.۹۴۱۰	۰.۹۵۴۰	۰.۹۵۸۰	۰.۹۶۳۵
میانگین SOC خودروها	۰.۸۲۰	۰.۷۹۵	۰.۷۸۲	۰.۷۶۵
بار اوج شبکه (MW)	۴.۸۵۰	۴.۳۶۰	۴.۲۵۰	۴.۱۲۰

به‌منظور ارزیابی بهتر عملکرد روش پیشنهادی، نتایج آن با حالت بدون مدیریت و همچنین دو روش متداول بهینه‌سازی فراترکاری

۵. نتیجه گیری

در این مقاله، یک چارچوب مدیریت هوشمند شارژ خودروهای الکتریکی مبتنی بر یادگیری تقویتی عمیق برای بهبود عملکرد شبکه‌های توزیع ارائه شد. با توجه به افزایش نفوذ خودروهای الکتریکی، شارژ همزمان این وسایل نقلیه می‌تواند منجر به افزایش تلفات شبکه و ایجاد انحراف ولتاژ در باس‌ها شود. در روش پیشنهادی، یک عامل یادگیری تقویتی با در نظر گرفتن وضعیت شبکه، سطح بار و وضعیت شارژ باتری خودروها آموزش داده می‌شود تا تصمیمات بهینه‌ای برای زمان‌بندی شارژ و دشارژ خودروها اتخاذ کند.

برای ارزیابی عملکرد روش پیشنهادی، شبیه‌سازی‌ها بر روی شبکه استاندارد توزیع ۳۳ باس در یک افق زمانی ۲۴ ساعته انجام شد. نتایج حاصل از شبیه‌سازی نشان داد که راهبرد پیشنهادی می‌تواند اثرات منفی ناشی از شارژ همزمان خودروهای الکتریکی را به طور قابل توجهی کاهش دهد. به طور مشخص، استفاده از روش ارائه‌شده منجر به کاهش تلفات شبکه، بهبود پروفیل ولتاژ و کاهش بار اوج شبکه شد. این نتایج نشان می‌دهد که رویکرد مبتنی بر یادگیری تقویتی قادر است با یادگیری رفتار بهینه در طول زمان، عملکرد شبکه توزیع را بدون نیاز به قوانین ثابت یا برنامه‌ریزی از پیش تعیین‌شده بهبود دهد.

به طور کلی، نتایج این پژوهش نشان می‌دهد که استفاده از روش‌های مبتنی بر یادگیری تقویتی می‌تواند راهکاری مؤثر برای مدیریت هوشمند شارژ خودروهای الکتریکی و افزایش بهره‌وری شبکه‌های توزیع باشد. در مطالعات آینده، می‌توان توسعه این چارچوب را با در نظر گرفتن عدم قطعیت در رفتار کاربران، قیمت‌گذاری پویا برق و ادغام منابع انرژی تجدیدپذیر بررسی کرد. همچنین پیاده‌سازی این روش در شبکه‌های واقعی و مقیاس بزرگ‌تر می‌تواند به ارزیابی دقیق‌تر کارایی آن در شرایط عملی کمک کند.

پیشنهادهایی برای کارهای آینده:

با توجه به نتایج به‌دست‌آمده، در مطالعات آینده می‌توان توسعه چارچوب پیشنهادی را با در نظر گرفتن عدم قطعیت در رفتار کاربران، قیمت‌گذاری پویا برق، ادغام منابع انرژی تجدیدپذیر، بررسی عملکرد روش پیشنهادی در شبکه‌های توزیع بزرگ‌تر و استفاده از ساختارهای پیشرفته‌تر یادگیری تقویتی را دنبال کرد.

مراجع

- Agah, S. M., & Abbasi, A. (2012). The impact of charging plug-in hybrid electric vehicles on residential distribution transformers. *Iranian Conference on Smart Grids*.
- Daliri, H. S., Dejahagah, M., Baghaee, H. R., Abyaneh, H. A., & Bakhshai, A. (2025). HMM-based feature extraction and machine learning methods for event detection and classification in microgrids. *IEEE Transactions on Smart grid*.
- Daliri, H. S., Dejahagah, M., Khodabakhshi-Javinani, N., Baghaee, H. R., Nafisi, H., & Abyaneh, H. A. (2024). Dwtgb-based power quality disturbance detection in electric vehicle charging station. *2024 14th Smart Grid Conference (SGC)*.
- He, Y., Venkatesh, B., & Guan, L. (2012). Optimal scheduling for charging and discharging of electric vehicles. *IEEE Transactions on Smart grid*, 3(3), 1095-1105.
- Jang, M.-J., & Oh, E. (2024). Deep-reinforcement-learning-based vehicle-to-grid operation strategies for managing solar power generation forecast errors. *Sustainability*, 16(9), 3851.
- Li, Y., Xiang, Y., Zheng, T., Wen, C., Wei, W., Tong, J., Hu, H., Sun, Z., Chen, T., & Zhang, Q. (2026). Deep Q-Network Based Optimal Charging Coordination of Electric Vehicles Considering Vehicle-to-Grid Technology. *Electricity*, 7(2), 44.
- Liu, D., Zeng, P., Cui, S., & Song, C. (2023). Deep reinforcement learning for charging scheduling of electric vehicles considering distribution network voltage stability. *Sensors*, 23(3), 1618.
- Qian, J., Jiang, Y., Liu, X., Wang, Q., Wang, T., Shi, Y., & Chen, W. (2023). Federated reinforcement learning for electric vehicles charging control on distribution networks. *IEEE Internet of Things Journal*, 11(3), 5511-5525.
- Rashid, T. S., Refat, K. A., Islam, M. N., Khan, R., Mostofa, E. R., Hossain, M. A., Kabir, H., & Ali, M. F. (2025). Machine learning-driven PV forecasting and coordinated PV-BESS dispatch for optimizing distribution grid performance under high EV penetration. *Energy Conversion and Management: X*, 101418.

- Sharif, M., Seker, H., & Javed, Y. (2026). Context-Aware Multi-Agent Coordination Framework for Intelligent Electric Vehicle Charging Optimization. *IEEE Access*.
- Sundstrom, O., & Binding, C. (2011). Flexible charging optimization for electric vehicles considering distribution grid constraints. *IEEE Transactions on Smart grid*, 3(1), 26-37.
- Wan, Z., Li, H., He, H., & Prokhorov, D. (2018). Model-free real-time EV charging scheduling based on deep reinforcement learning. *IEEE Transactions on Smart grid*, 10(5), 5246-5257.
- Xiao, Q., Zhang, R., Wang, Y., Shi, P., Wang, X., Chen, B., Fan, C., & Chen, G. (2024). A deep reinforcement learning based charging and discharging scheduling strategy for electric vehicles. *Energy Reports*, 12, 4854-4863.
- Yin, W., Qin, X., & Huang, Z. (2022). Optimal dispatching of large-scale electric vehicles into grid based on improved second-order cone. *Energy*, 254, 124346.