AI-Driven Adaptive Load Balancing Strategy for Electric Vehicle Charging Pile Networks

Ruilin Li[1,2], Honghong Li[1,*], Yoke Seng Wong[2], Fengyang Long[3], Donghua Jiang[4], Jianhua Wu[1]
[1]School of Engineering Technology, Gongqing College of Nanchang University, Gongqingcheng, Jiangxi 332020, China
[2]Faculty Of Computing And Meta-Technology, Sultan Idris Education University, Perak 35900, Malaysia
[3]College of Water Conservancy, Jiangxi University of Water Resources and Electric Power, Nanchang 330099, China
[4]School of Information Engineering, Chang'an University, Xi'an 710064, China

Corresponding author: Honghong Li (e-mail: honghongli_85@163.com).

ABSTRACT With the widespread use of EVs, new strains have been added to urban power systems. In-coordinated connections to charging station networks result in uneven load distribution, the separation of peak and valley demand, and less overall operational efficiency. In this paper, an AI-based adaptive load balancing approach to EV charging station networks is proposed. A three-layer distributed architecture is built, including perception and data collection layer, edge computing layer and decision-making layer in the cloud. For short-term load forecasting, a model combining long short-term memory networks with an attention mechanism is designed. A multi-agent proximal policy optimization algorithm is then used to realize the adaptive load balancing decisions. A simulated network of 20 charging stations and 200 charging piles was used for the experiments in an urban environment. The outcome is evident. The proposed strategy sets the Peak-to-valley load reduction rate at 63.2% across the network, and it reduces the load standard deviation to 31.4kW. The average wait time for users reduces from 18.7 minutes to 6.3 minutes. Violation rate of node voltage drops from 8.64% to 0.47%. Decision latency is still within 1.24 seconds even if a massive 50 stations are used in scenarios. The approach will ensure the safety and stability of the grid and optimize the use of the resources within the charging network without compromising the system's integrity. It provides a theory and reference guide to sophisticated management of smart charging.

INDEX TERMS electric vehicles; charging station network; adaptive load balancing; multi-agent reinforcement learning; load forecasting

I. INTRODUCTION

The world is moving towards energy transition. The EV market continues to grow and large-scale EV charging infrastructure deployment is becoming a key element in supporting next-generation clean transportation systems. But the decentralized deployment of gigantic charging infrastructure presents some operational challenges in addition to its ease of use. The load fluctuations also increase, along with increasing valley to peak demand differences and increasing localized node overloads. The traditional rule-based or static optimization methods do not work in this environment. They are generally not able to deliver real time responsiveness and optimality simultaneously when the user behaviour is extremely stochastic and charging demand is time-space dependent. There is a need for more adaptive intelligent control strategies. Deep reinforcement learning (DRL) and other forms of AI have become increasingly the subject of research interest for EV charging scheduling. They are well suited in this domain by virtue of their advantages in making decisions in sequence and adapting to their environment. Park et al. suggested a multi-agent deep reinforcement learning model for the EV charging scheduling problem. It was able to optimize the system in a coordinated manner at multiple stations in a smart grid application, which in turn minimized both the operating costs of the system and the peak power demand [1]. Wan et al. proposed a model-free real-time scheduling method. Their approach involved learning the optimal dispatch policies, immediately from the grid state data, to achieve dynamic charging optimization without requiring accurate system models. In turn, Yan et al. added a dynamic user behavior to the continuous-action reinforcement learning framework. The resulting system proved to be stable in decision performance in varying demands, and more practical and generalizable. The scheduling process has also been the subject of attention with regard to safety constraints. Li et al. proposed a safe dispatching scheme of EV charging based on deep reinforcement learning with constraints to maximize charging efficiency without breaking the operational constraints on the grid [2]. This collection of studies highlights the substantial impact that AI-backed methods can have on enhancing flexibility in scheduling and grid stability. But the existing studies focus on single-station or small-node scenario. A large-scale charging infrastructure load balancing system at a network level is still relatively under-researched. This paper is thus an attempt to fill that gap. The attention is on an adaptive load balancing approach based on AI for massive EV charging station networks. The aim is to create an intelligent control system that can detect the load conditions in the entire network, adjust charging resources dynamically, and consider the user experience and the safety of the grid at the same time. There is an integration of system design, algorithmic modeling and simulation validation throughout. The work has the purpose of not only providing theoretical foundation, but also giving some practical suggestions for efficient operation of large scale charging infrastructure, the research has real significance both in academic research and engineering application.

II. LITERATURE REVIEW

Over the last few years, research on EV charging scheduling and load balancing has been significantly developed. The field has shifted from the optimization of a single entity to the multi-agent coordination and from offline planning to real-time adaptive control.

Multi-agent reinforcement learning frameworks.

Zhang et al. suggested a multi-step multi-agent reinforcement learning scheme for energy scheduling in smart grid charging stations. The multistep temporal decision mechanism increased the overall energy utilization of the station clusters and provided novel knowledge for coordinated dispatch in complex grid environment. Qian et al. added game theory into the multiagent deep reinforcement learning system. They developed a model that was able to incorporate competitive and cooperative interactions between charging stations, which balanced the interests of the individual charging stations with the grid level stability goals.

Real-time pricing and joint scheduling[3].

Wang et al. used reinforcement learning in the simultaneous real-time pricing and charging dispatch control. Dynamic coordination between price signals and charging behavior was realized, which assisted users to be more active in reducing peak loads. Qiu et al. furthered this line of research by proposing a hybrid multi-agent reinforcement learning approach for flexible EV management under low-carbon transition constraints. The method was based on a combination of system resilience and carbon emission constraints, which extended the usage of reinforcement learning to multi-objective systems.

Restrictions of voltage stability and voltage distribution networks.

Hu et al. proposed a multi-agent deep reinforcement learning voltage control system to optimize both active and reactive power. The results from simulation verified that agent cooperation is effective to keep distribution network voltage safe. Liu et al. tackled the same issue in a different way, on a scheduling axis. The risk of voltage violation was directly incorporated into the reward function design, and the huge number of charging connected to the distribution network was significantly minimized .

Optimization of fleet and cluster charging.

Tuchnitz et al. proposed and tested a reinforcement learning approach for smart charging of electricity vehicles. Policies were developed based on historical driving data and the approach was effective in energy consumption management in practical operational scenarios . Ding et al. proposed a Markov decision processes and reinforcement learning model that builds an optimum charging policy for uncertain demand[4]. It was the first attempt to give a formal mathematical basis to charging decisions in stochastic environments [5].

Algorithmic innovations.

Zhang et al. suggested CDDPG, a continuous-action deep deterministic policy gradient algorithm for EV charging control. It solved the problem that the discrete-action approaches are not suitable for the continuous charging power regulation . Ye et al. studied learning optimal charging station operation strategies using reinforcement learning as well as vehicle-togrid (V2G) services, which is in the direction of bidirectional energy management in charging networks [6]. To achieve a balance between computational efficiency and policy quality, Cao et. al. presented an online algorithm for intelligent charging in single-pile control using a customized Actor-Critic technique . Li et al. carried out a systematic review of deep reinforcement learning (DRL) applications in EV charging management and analyzed the various network architectures and training methods in different operational environments. Kaewdornhan et al. explored the problem of optimizing home energy systems with multiple households, using multi-agent deep reinforcement learning to optimize in real time, including real-time EV charging optimization. They found that it was feasible to use such methods in distributed small-scale settings . The results found across the board have reached a considerable extent in terms of algorithm design and localised scenario validation. A comprehensive and systematic theoretical approach and experimental validation of globally adaptive load balancing in large-scale charging station networks under dynamic load is still missing. This is the main motivation of this paper.

III. METHODOLOGY

A. *SYSTEM ARCHITECTURE OF THE CHARGING STATION NETWORK*

In this study, the charging station network architecture is based on a distributed three-layer charging station architecture. The layers are: Perception and Data Collection layer, Edge Computing layer, Decision layer (Cloud-based).

Standardised communication protocols support bidirectional real-time data exchange on all three layers. The perception layer contains current transformers, voltage sensors, and metering modules that are mounted to individual charging piles. Sampling is done at a frequency of 500ms for each pile. Data gathered will include the user's arrival time, estimated departure time, real-time battery state of charge (SOC), power output, and charging current. The raw measurement data is transferred to the nearest edge node through 4G/5G wireless modules. Each charging station has a separate edge computing server (8 core CPU, 32 GB RAM and GPU acceleration support) at the edge layer[7]. Local data preprocessing, outlier removal, and feature extraction is performed on these servers. The millisecond level dynamic power allocation of piles in a single station is also provided through the lightweight intra-station load balancing module at this layer. This reduces dependence on cloud communication bandwidth considerably . The cloud decision layer runs on a private server cluster. It combines the load forecasting model, the multi-agent reinforcement learning scheduling engine and the global optimization solver. It is receiving the aggregated information at the station level from each edge node[8]. Global load balancing instructions are created and sent back to the edge nodes every 5 minutes to be executed. This study refers to the actual distribution of the charging piles in a Chinese city for the network topologies. A 20 charging stations simulation network and 200 charging pile network were built. The 10 kV distribution lines serve as the regional grid at each station. MQTT is a protocol used for lightweight transmission of messages between the edge nodes and the cloud. The delay from the dispatchers to the driver is less than 200ms, satisfying the "timeliness" requirement of the dispatching instructions, and the system response is reliable .

B. LOAD FORECASTING MODEL

To enable adaptive load balancing, accurate load forecasting is required. In this study, a deep learning model is proposed and designed that integrates long short-term memory (LSTM) network with attention mechanism to predict short-term load demand on the charging network. The feature vector is a 12-dimensional vector. It contains historical load sequences for the last 24 hours at 15-minute intervals (96 time steps) as well as real-time ambient temperature, day-type labels (weekday, holiday or special event), time-of-day encodings (hour-level sine-cosine encoding), the number of vehicles on line at each station, and the mean SOC for connected vehicles. The features are scaled with Min-Max scaling prior to the input into the model . The network architecture is two stacked LSTM layers. The first layer has 128 hidden units and produces all of the hidden states for every time step. The second layer has 64 hidden units and just returns the last hidden layer. An additive attention layer is placed between the two LSTMs. It learns a matrix of weights to assign differentiated attention score to every timestep. This enables the model to automatically focus on important time windows around the sudden load changes, and to capture the non-stationary fluctuation patterns of charging demand. The output of the second LSTM layer is concatenated with the attentionweighted context vector.The attention-weighted context vector is then combined with the output of the second LSTM layer. The combined representation is passed through two fully connected layer with added dimensions 32,16, both of which are ReLU activated to finally generate the prediction. This forecast period consists of 4 time steps of 15 minutes each or 1 hour ahead[9]. The training is implemented with the Adam optimizer, and an initial step size of 0.001. In late training, to avoid oscillations, a cosine annealing schedule is used. The batch size is 64, and the epochs of training are 200, early stopping after 20 epochs to prevent overfitting . The loss function is a weighted sum of mean squared error (MSE) and mean absolute error (MAE) with a ratio of 0.6:0.4. This is a balance between the overall accuracy of the prediction and sensitivity to extreme deviations. Training data is simulated network operation over a 12 month period which is divided into 7:1.5:1.5 training, validation and test data. The split enables cross-season and cross-scenario generalization.

C. ALGORITHM DESIGN FOR ADAPTIVE LOAD BALANCING

The adaptive load balancing strategy is based on a Multi-Agent Proximal Policy Optimization (MAPPO) framework. Charging stations are assumed to be independent agents. Adopted is the Centralized Training with Decentralized Execution (CTDE) paradigm, where coordination takes place globally during training, and execution takes place locally at deployment. The local observation vector is a 18 dimensional vector for each agent. It includes the current total power demand of the station, allocation of power to each pile, distribution of connected vehicles' SOCs, estimated departure time, power demand difference between the previous decision cycle and the current cycle, load status of neighbouring stations, and current electricity price signals. The global state is built at the cloud level by simply appending all the agents local observations yielding a 360-dimensional vector . The action space consists of continuous values. Every agent provides an adjustment ratio vector for the power of its charging piles, ranging from 20% to 100% of the pile's rated power. These continuous actions are produced by a policy network that includes a Gaussian parameterization which does not cause loss of precision compared to discrete action space approximations[10]. The reward function is a compromise between three goals. The first one is a network-wide load variance minimization with a weight of 0.5, which measures the uniformity of load distribution between stations. The second is user charging satisfaction (weighted 0.3) which represents the percentage of leaving vehicles with user specified user charging conditions. The third is a grid safety penalty (0.2) when the power of any node surpasses 90% of the distribution transformer capacity . The policy network and value network have threelayered fully connected networks with 256, 128 and 64 hidden layers and the Tanh activation function. The PPO clipped objective is used for policy updates, where $\epsilon = 0.2$ is the clipping coefficient. Each of 10 mini-batch gradient updates uses a batch size of 256, and is executed after 2,048 experience steps. The discount factor γ is equal to 0.99. The Generalized Advantage Estimation (GAE) parameter λ is fixed at 0.95 to ensure both bias and variance are minimized when the charging demand is changing.

D. MULTI-OBJECTIVE OPTIMIZATION FUNCTION

The load balancing strategy is mathematically formulated in the optimization objective function. Operational safety of the grid, quality of charging services, and operational cost are taken into account together. There are three sub-objectives, and the overall objective is a linearly-weighted sum of these sub-objectives:

$$F = \alpha f_1 + \beta f_2 + \gamma f_3.$$

Weights α, β, and γ are set to 0.4, 0.35, and 0.25 respectively, summing to 1. The values were calculated using the Analytic Hierarchy Process (AHP) corresponding to the importance of each objective. The first sub-objective f_1 is a network-wide load balance. It is the root mean square difference between every station's real time power and the network average power:

$$f_1 = \sqrt{\frac{1}{N}\sum_i (P_i - \bar{P})^2},$$

with N = 20 (the total number of stations); $P_i$is the real time power level at station i; $\bar{P}$ is the mean power level at the network. The smaller the f_1, the more uniform the load distribution and the smaller the peak to valley gap . The second sub-objective f_2 is related to inadequate charging service quality[11]. It is defined as the expected value of the SOC deficit for the vehicles being removed from the system if their current SOC is below their target,

$$f_2 = \mathbb{E}\left[\sum_j \max(0, \mathrm{SOC}_{j,\mathrm{target}} - \mathrm{SOC}_{j,\mathrm{actual}})\, w_j\right]$$

where $w_j$is the charging priority weight of vehicle j. The smaller the f_2, the more satisfied the user is with the demand. System electricity procurement cost under time of use (ToU) pricing is captured by the third sub-objective (f_3):

$$f_3 = \sum_i \sum_t c_t P_i(t)\, \Delta t,$$

where c is the electricity price at time t (CNY/kWh), P_i can be the charging power of station i at time t, and Δt is the decision time step (5 minutes or 1/12 hour). A lower f_3 will represent a lower procurement cost, and meet the load balance and user needs requirements . There are three hard constraints applied on optimization. First, the power per pile has to be 20% to 100% of rated capacity. Secondly, the combined power of all the stations should not be more than 90% of the transformer capacity. Third, power between decision steps can be adjusted by no more than 15% in each step. These restrictions restrict the optimization to physically achievable space constraining the implementation of the strategy in practice.

E. SIMULATION PLATFORM AND EXPERIMENTAL DATA

The simulation platform has been written in Python 3.9. The PyTorch 1.12 framework was used to build and train neural network models. The charging station network environment was encapsulated using the OpenAI Gym interface, which gives a common reinforcement learning interaction protocol. The PANDAPOWER 2.10 library was used to perform the power flow calculations for the simulated grid. This was a true representation of the node voltage profiles and line power flows for all 20 stations on the regional grid with 10 kV distribution lines[12]. The time step used for the simulation was 5 minutes. Full simulation runs were conducted for 30 consecutive days daily, including weekdays, weekends and public holidays . The hardware included an Intel Xeon Gold 6248R processor (32 cores, 3.0 GHz) and 256 GB DDR4 memory, along with an NVIDIA A100 40 GB GPU. Multi-agent policy training employed an 8 CPU parallel experience collection asynchronous architecture with a single GPU for parameter updates. The entire training process required around 14 hours. The historical charging load data is provided from a public charging operations platform in a key city in China between January and December 2022. The raw data consisted of pile ID, session start and end time, energy delivered, initial and final SOC, and power-time curves (about 4.8 million valid records total). Sessions less than 5 minutes or more than 12 hours were deleted, as were records without power data. The number of high-quality records that were not removed for being redundant was about 4.36 million records for model training and testing[13]. The meteorological data were obtained from the China Meteorological Administration open observation dataset, and the data consisted of hourly weather type labels, temperature, and humidity. These were matched with charging records based on the time stamp. Electricity pricing followed a provincial time-of-use tariff scheme from the State Grid: peak hours (08:00–12:00 and 18:00–22:00) at 1.2 CNY/kWh, off-peak hours (12:00–18:00) at 0.8 CNY/kWh, and valley hours (22:00–08:00 next day) at 0.4 CNY/kWh . The incorporation of actual operating data with the simulation environment further enhances practicality and the ability to replicate experimental results.

IV. RESULTS AND ANALYSIS

A. LOAD FORECASTING MODEL PERFORMANCE

The four performance measures of the forecasting models were used were mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and coefficient of determination R^2. The proposed LSTM-Attention model was tested on the held-out test set against three models as baselines such as conventional

LSTM, gated recurrent unit (GRU), and support vector regression (SVR). The LSTM-Attention model overall gave the highest results for all four metrics. MAE was 18.34 kW, RMSE was 24.67 kW, MAPE was 4.23%, and R^2 reached 0.9712. The conventional LSTM was improved by 21.8% in terms of the mean absolute error and 18.5% in terms of the root-mean-square error[14]. Such increases show that the attention mechanism has been successful in providing meaningful enhancement in the model's capacity to capture non-stationary variations in charging load . Day type variation was also seen in the performance. Weekday's MAPE was 3.87%, weekend's 4.61% and public holiday's 5.34%. The increased error for the holiday period is due to increased randomness of user travel during that time, but all errors are within acceptable engineering tolerances. The accuracy of the predictions was found to be relatively stable during the morning and evening peak charging hours (07:00- 09:00 and 18:00-21:00), with an average peak hour MAPE of 4.89%[15]. This indicates that the model has good tracking capability at a high-speed rate of change for loading. Table 1 shows detailed performance numbers for all models. A typical day of predicted vs. actual load curves is displayed in Figure 1.

TABLE 1. Performance comparison of load forecasting models

Model	MAE (kW)	RMSE (kW)	MAPE (%)	R^2
SVR	41.25	56.83	8.97	0.8634
GRU	29.76	38.42	6.15	0.9287
LSTM	23.45	30.24	5.41	0.9503
LSTM-Attention (proposed)	18.34	24.67	4.23	0.9712

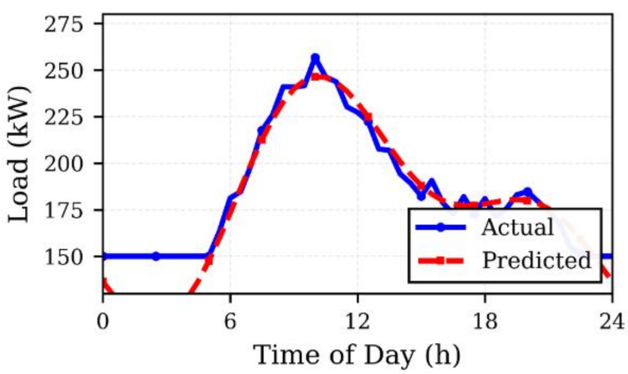

FIGURE 1. Comparison of actual and predicted load curves by different models (one typical day)

B. VALIDATION OF THE ADAPTIVE LOAD BALANCING STRATEGY

The strategy performance was judged by three aspects: response speed, peak-to-valley load reduction rate, and adaptability for multi-scenario. The suggested MAPPO strategy was evaluated in light of three reference strategies: No Control (no control), Rule-Based (rule-based static balancing) and Centralized OPT (traditional centralized optimization). The MAPPO strategy had the highest performance for all measures. The No Control peak to valley load difference for the network was 387.4kW, which was reduced to 142.6kW under Active Control, giving a 63.2% reduction. This is 28.7 percentage points better than the Rule Based scheme. The response latency achieved was 0.38 seconds, which is 74.3% less than the Centralized OPT approach. This gap reflects well the ability of the distributed decision architecture to perform in real time [29]. The load standard deviation at all stations dropped to 31.4 KW under MAPPO, which is a 67.8% decrease from the baseline. This means significantly more even distribution of charging power powering the stations. Multi-scenario results were also promising. The peak to valley reduction rates were 65.1%, 61.8% and 58.9% on weekdays, weekends and public holidays respectively[16]. The slightly lower number of holiday days is due to the higher degree of "noise" in the user behavior, but the effectiveness of control remained constant for all three scenarios . The overall utilization of charging piles rose by 25.3 percentage points from 54.3% for No Control to 79.6% for MAPPO. This indicates that load balancing can impact not only the stability of the grid operation but also to be able to wake up underutilized charging resources[17]. The comparison data for all schemes are shown in full in Table 2. The network-wide load standard deviation convergence plots for each strategy are given in Figure 2.

TABLE 2. Performance comparison of load balancing strategies

Strategy	Peak-to-valley difference (kW)	Reduction rate (%)	Response latency (s)	Load std. dev. (kW)	Pile utilization (%)
No Control	387.4	—	—	97.6	54.3
Rule-Based	248.3	34.5	1.82	68.4	63.7
Centralized OPT	186.7	51.8	1.49	48.2	72.4
MAPPO (proposed)	142.6	63.2	0.38	31.4	79.6

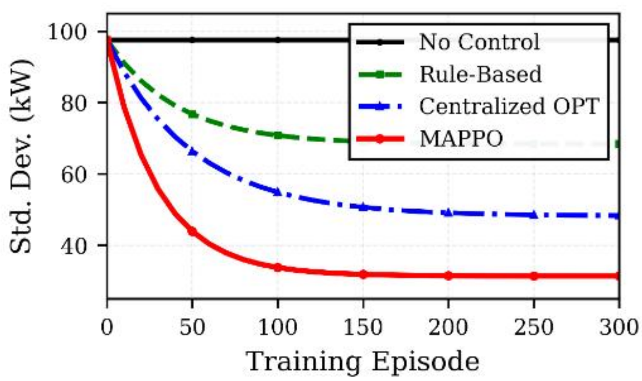

FIGURE 2. Convergence curves of load standard deviation under different control strategies

C. OPERATIONAL EFFICIENCY ANALYSIS

The three parameters used to evaluate operational efficiency were charging pile utilization, average user waiting time and grid stability parameters. The results are all compared to the No Control baseline, Rule Based scheme, and the Centralized OPT approach.

Charging pile utilization.

The MAPPO strategy had been implemented to achieve an increase in average utilization from 54.3% to 79.6% which represents a gain of 25.3 percentage points. When analysed by time period, the peak hour utilisation rose from 61.2% to 88.4% while the valley hour utilisation rose from 41.7% to 68.2%. The utilization gap between peak and valley was reduced by just 19.5 to 20.2 percentage points, but there were significant absolute increases in both periods. Overall, a more balanced allocation of resources was observed .

User waiting time.

The average waiting time over the whole day decreased from 18.7 minutes under No Control to 6.3 minutes under MAPPO, which is a 66.3% decrease. At the evening peak time (18:00 – 21:00) waiting time reduced from 32.4 minutes to 9.8 minutes. The experience of the end users was significantly enhanced .

Grid stability.

The percentage of nodes where the voltage is violated, dropped from 8.64% to 0.47%. Average number of daily line power overloads decreased from 23.6 to 1.2. Peak distribution transformer loading ratio went down from 94.7% to 76.3% of the rated capacity[18]. Safety margins in the grid were significantly increased for the three indicators. The proposed MAPPO strategy had the highest score across all operational efficiency measures[19]. The results offer solid quantitative evidence for fine-tuned operation of large-scale charging infrastructure networks. The results of the full comparison are shown in Table 3. Figure 3 presents utilization distributions of time periods across the various strategies.

TABLE 3. Operational efficiency comparison across strategies

Strategy	Avg. utilization (%)	Avg. waiting time (min)	Voltage violation rate (%)	Line overloads (times/day)	Peak transformer loading (%)
No Control	54.3	18.7	8.64	23.6	94.7
Rule-Based	63.7	13.2	4.83	12.4	87.3
Centralized OPT	72.4	9.6	1.92	5.7	81.6
MAPPO (proposed)	79.6	6.3	0.47	1.2	76.3

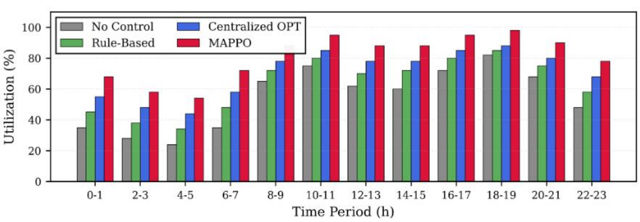

FIGURE 3. Charging pile utilization rate across time periods under different control strategies (%)

D. ROBUSTNESS AND SCALABILITY TESTING

Robustness and scalability are key factors to consider when trying to deploy an adaptive load balancing strategy in complex real-world settings. The following three aspects were studied: extreme load stress test, network scale expansion experiment, and algorithm convergence analysis.

Stress testing to extreme loads.

Three sudden onset scenarios were simulated: a high load demand increase (HLDI) due to mass crowd dispersal from a large event and a high fastcharging demand (HFCD) due to battery degradation during extreme cold weather, as well as an islanded operation following a failure of communication between the two stations . The standard deviation of the network-wide load under the 150% load surge was 58.3 kW. The system switched to normal balanced operation in 4.2 minutes - 63.7% faster than the Rule Based scheme. The peak-to-valley reduction rate remained at 54.6% in the extreme cold scenario, which is below 10 percentage points from the base scenario. If one station was not communicating, then the remaining 19 stations automatically set up a distributed decision-making process to redistribute power. Use of the network only decreased by 3.8%. The resilience of the systems was good in all three scenarios .

Scalability testing.

The number of stations and piles was gradually increased from 10 to 50 stations and 100 to 500 piles. The single-step decision latencies, at the largest scale, was 1.24 seconds. The peak-to-valley reduction rate did not change (59.3%) and load standard deviation was 47.6 kW[20]. As the size grows, the deterioration of the performance remained within acceptable limits for all metrics. The strategy is linearly scalable good .

Convergence analysis.

On the policy reward function, for five independent random seeds, it converged in mean 274.6 episodes with a standard deviation of 18.3. One of the convergence stabilities was significantly higher than the other comparison algorithms. In Table 4, detailed performance degradation data is shown for each scale of the network. Figure 4 plots the trends in peak-to-valley reduction rate, and decision latency, with respect to the number of stations.

TABLE 4. MAPPO strategy performance across network scales

Scale	Stations	Piles	Peak-to-valley reduction (%)	Load std. dev. (kW)	Decision latency (s)	Convergence episodes
Small	10	100	65.8	28.3	0.31	251
Medium	20	200	63.2	31.4	0.38	275
Large	35	350	61.4	39.7	0.76	298
Extra-large	50	500	59.3	47.6	1.24	324

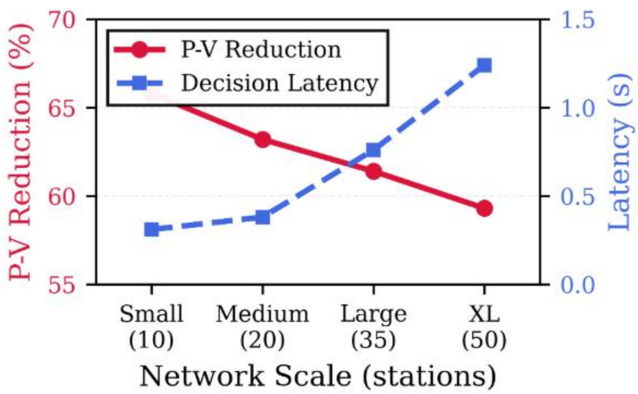

FIGURE 4. Peak-valley reduction rate and decision latency vs. network scale

V. DISCUSSION

A. INTERPRETING THE EXPERIMENTAL RESULTS

Centralized training and decentralized execution are the common thread that weaves through the benefits of the MAPPO strategy in terms of load balancing, operational efficiency and robustness. Each agent is given access to the global state during training. This enables each charging station to consider the load conditions in its home region and take grid limitations into the charging station policy optimization, preventing suboptimal results from decisions taken purely locally in conventional distributed methods. Each station has its own local policy network at run time. System functionality is no longer compromised due to communication failures and network delays. The result is a fault tolerant control architecture that has true fault tolerance . The attention mechanism also played a significant role in the load forecasting model[21]. The weight given to time steps around load ramps was increased, which resulted in lower prediction error around the time steps of greatest load ramp. Better forecasts were fed into the scheduling engine, resulting in improved dispatch instructions and balancing results. This relationship of increased accuracy in forecast $\rightarrow$ improved scheduling $\rightarrow$improved load balance became a self-reinforcing one throughout the entire chain . A role of structure was also played by the three-component weighted objective function. The highest weight was given to load balancing (α = 0.4) to prevent compromising stability of the grid for short-term gains in utilisation during higher stress events. This priority ordering is directly related to the "safety first" principle of power system engineering, and provides very clear practical interpretability .

B. DEPLOYMENT CHALLENGES IN REAL-WORLD SETTINGS

While promising simulation results do not eliminate the practical barriers to deployment, they are certainly not insurmountable. The four challenges stand out.

Data quality and reliability of communication.

In practice, charging networks can suffer from various sensor errors, packet loss and noise. The completeness and timely input to MAPPO influence its decision making. There can be systematic biases in the reporting from nodes and edges that can propagate through the dispatch pipeline, potentially leading to some cases of overload in the local system.

Model generalization and transfer.

The policy network was trained on past data from a single city. The learned load patterns are location and time dependent and so are the user behaviour distributions. This is likely to cause performance degradation in cities with varying infrastructure sizes, use of mobility, or tariff systems[22]. There would need to be online fine-tuning or transfer learning to compensate .

Edge constraints of computation.

Policy inference imposes real hardware requirements to edge servers. A response in millisecond level is difficult to obtain in low cost devices. For practical lightweight deployment, model compression and quantization techniques would play a crucial role .

Regulatory compliance, safety certification.

Rigorous safety testing of AI-driven strategies for grid control is essential before commercial operation. Regulators have difficulty interpreting decisions made by neural networks due to the lack of interpretability. To ensure the performance of the decision and improve decision transparency without losing performance is a necessary condition for real world adoption .

C. LIMITATIONS AND POTENTIAL CONFOUNDING FACTORS

This study has a number of shortcomings that should be acknowledged. Firstly, a gap between the simulation environment and the physical grid cannot be avoided. The power flow model used a simplified line model (PANDAPOWER) and constant loads. The phenomena of nonlinear grid (harmonic distortion, reactive power variation and magnetic saturation of transformers) were not captured. These factors may have an impact on the node voltage that the simulation does not account for when the charging density is high. As a result, the reported results might be somewhat optimistic when compared to actual deployment . Second, the user behavior modelling remained at a statistical level. The arrival times, departure times and initial SOCs were taken from historical distributions. The individual-level factors, such as price sensitivity, charging anxiety and personal scheduling preferences, were not considered. Depending on the time of day, the heterogeneous behaviours can magnify load concentration effects and models cannot capture these effects . Third, experimental validation was conducted using only the historical data of one city. The generalization between cities and between climates has not been validated. The applicability of the proposed strategy needs to be further explored . The multi-objective weight coefficients were determined by the Analytic Hierarchy Process (AHP) to be fixed, which is the 4th step. No mechanism was provided to dynamically change weights on the basis of operating period or load condition of the grid. In severe peak loads, fixed weights may not be the optimum solution to balance safety requirements and service quality. This restriction suggests a definite path to be followed in future research, namely, adaptive weight scheduling based on the actual system state.

VI. CONCLUSION

This paper solves the adaptive load balancing problem in the EV charging station networks of large scale. A novel intelligence control framework was proposed in which the deep-learning based load forecasting and multi-agent proximal policy optimization were integrated. On the system architecture level, a three-layer distributed architecture was built, which included perception and data collection, edge computing, and cloud-based decision making. This allowed for the instant acquisition and multi-level cooperative processing of charging pile operational information. The algorithm level, an LSTM-Attention short term load forecasting model and a MAPPO multi-agent reinforcement learning balancing strategy were developed. The results are concrete. The network-wide P2V load reduction rate was 63.2%. The standard deviation of a load was reduced to 31.4 kW. A significant drop in the average user wait time, from 18.7 minutes to 6.3 minutes. The rate of node voltages violation decreased from 8.64% to 0.47%. There were significant gains over the baseline schemes for all operational efficiency measures. Scalability was also very good. The single-step decision latency was between 1.24 seconds and the peak-to-valley reduction rate was 59.3% in an extra-large network consisting of 50 stations and 500 piles. The figures show that as the network size increases, the applicability to engineering will be strong. Combined, the experimental results show that the proposed approach has potential to effectively enhance the utilization of the charging pile resources and the service quality for users while ensuring the safety of grid operation. The work provides a technical route that can be realized with AI to carry out refined management of new-energy transportation infrastructure. It is of theoretical interest and practical applicability for deployment.

FUNDING STATEMENT

(1) Research Project of Humanities and Social Sciences in Colleges and Universities of Jiangxi Province (Special Project on Red Culture) (Source). (Title) Research on the Jiangxi Path of Integrating Red Culture into Engineering Education Empowered by Generative Artificial Intelligence (HSWH25025); (2) Jiujiang City Natural Science Foundation: Optimization of Subway Short-Term Passenger Flow Prediction Model Using Context-Aware Graph Transformer (CaGT) (2025_000504); (3) Jiangxi Provincial Department of Education: Research on Reforming Electrical Engineering Talent Training under the "Government-Enterprise-University-Association" Collaboration Mechanism (BKJG-2026-31-11).

REFERENCES

[1] K. Park and I. Moon, "Multi-agent deep reinforcement learning approach for EV charging scheduling in a smart grid," Appl. Energy, vol. 328, p. 120111, 2022, doi: 10.1016/j.apenergy.2022.120111. [Online].
[2] Z. Wan, H. Li, H. He, et al., "Model-free real-time EV charging scheduling based on deep reinforcement learning," IEEE Trans. Smart Grid, vol. 10, no. 5, pp. 5246–5257, Sept. 2019, doi: 10.1109/TSG.2018.2879572. [Online].
[3] L. Yan, X. Chen, J. Zhou, et al., "Deep reinforcement learning for continuous electric vehicles charging control with dynamic user behaviors," IEEE Trans. Smart Grid, vol. 12, no. 6, pp. 5124–5134, Nov. 2021, doi: 10.1109/TSG.2021.3098298. [Online].
[4] H. Li, Z. Wan, and H. He, "Constrained EV charging scheduling based on safe deep reinforcement learning," IEEE Trans. Smart Grid, vol. 11, no. 3, pp. 2427–2439, May 2020, doi: 10.1109/TSG.2019.2955437. [Online].
[5] X. Sun, W. Wang, S. Su, et al., "Coordinated charging strategy for electric vehicles based on time-of-use price," Autom. Electr. Power Syst., vol. 37, no. 1, pp. 191–195, Jan. 2013, doi: 10.7500/AEPS201209274. [Online].
[6] Z. Xu, Z. Hu, Y. Song, et al., "Coordinated charging strategy for PEV charging stations based on dynamic time-of-use tariffs," Proc. CSEE, vol. 34, no. 22, pp. 3638–3646, Nov. 2014, doi: 10.13334/j.0258-8013.pcsee.2014.22.008. [Online].
[7] Y. Zhang, Q. Yang, D. An, et al., "Multistep multiagent reinforcement learning for optimal energy schedule strategy of charging stations in smart grid," IEEE Trans. Cybern., vol. 53, no. 7, pp. 4292–4305, July 2023, doi: 10.1109/TCYB.2022.3165074. [Online].
[8] T. Qian, C. Shao, X. Li, et al., "Multi-agent deep reinforcement learning method for EV charging station game," IEEE Trans. Power Syst., vol. 37, no. 3, pp. 1682–1694, May 2022, doi: 10.1109/TPWRS.2021.3111014. [Online].
[9] S. Wang, S. Bi, and Y. A. Zhang, "Reinforcement learning for real-time pricing and scheduling control in EV charging stations," IEEE Trans. Ind. Informat., vol. 17, no. 2, pp. 849–859, Feb. 2021, doi: 10.1109/TII.2019.2950809. [Online].
[10] D. Qiu, Y. Wang, T. Zhang, et al., "Hybrid multiagent reinforcement learning for electric vehicle resilience control towards a low-carbon transition," IEEE Trans. Ind. Informat., vol. 18, no. 11, pp. 8258–8269, Nov. 2022, doi: 10.1109/TII.2022.3166215. [Online].
[11] D. Hu, Z. Ye, Y. Gao, et al., "Multi-agent deep reinforcement learning for voltage control with coordinated active and reactive power optimization," IEEE Trans. Smart Grid, vol. 13, no. 6, pp. 4873–4886, Nov. 2022, doi: 10.1109/TSG.2022.3185975. [Online].
[12] D. Liu, P. Zeng, S. Cui, et al., "Deep reinforcement learning for charging scheduling of electric vehicles considering distribution network voltage stability," Sensors, vol. 23, no. 3, p. 1618, 2023, doi: 10.3390/s23031618. [Online].
[13] C. Zhang, J. Liu, Y. Xiang, et al., "Electric vehicle charging facilities planning and its two-stage optimal charging based on data mining," Proc. CSEE, vol. 38, no. 4, pp. 1054–1064, Apr. 2018, doi: 10.13334/j.0258-8013.pcsee.170472. [Online].
[14] L. Zhang, C. Sun, G. Cai, et al., "Two-stage optimization strategy for coordinated charging and discharging of EVs based on PSO algorithm," Proc. CSEE, vol. 42, no. 5, pp. 1837–1851, May 2022, doi: 10.13334/j.0258-8013.pcsee.211150. [Online].
[15] Z. Ye, Y. Gao, and N. Yu, "Learning to operate an electric vehicle charging station considering vehiclegrid integration," IEEE Trans. Smart Grid, vol. 13, no. 4, pp. 3038–3048, July 2022, doi: 10.1109/TSG.2022.3165479. [Online].
[16] M. Alqahtani and M. Hu, "Dynamic energy scheduling and routing of a large fleet of electric vehicles using multi-agent reinforcement learning," Comput. Ind. Eng., vol. 169, p. 108180, 2022, doi: 10.1016/j.cie.2022.108180. [Online].
[17] J. Shi, W. Zhang, Y. Bao, et al., "Load forecasting of electric vehicle charging stations: attention based spatiotemporal multi-graph convolutional networks," IEEE Trans. Smart Grid, vol. 15, no. 3, pp. 3016–3027, May 2024, doi: 10.1109/TSG.2023.3321116. [Online].
[18] P. Aduama, Z. Zhang, and A. Al-Sumaiti, "Multifeature data fusion-based load forecasting of electric vehicle charging stations using a deep learning model," Energies, vol. 16, no. 3, p. 1309, 2023, doi: 10.3390/en16031309. [Online].
[19] W. Chen, H. Sun, R. Xu, et al., "Cloud-edge collaboration based hierarchical dispatch technology for EV participating in frequency regulation service," Proc. CSEE, vol. 43, no. 3, pp. 914–926, Mar. 2023, doi: 10.13334/j.0258-8013.pcsee.220646. [Online].
[20] Z. J. Lee, G. Lee, T. Lee, et al., "Adaptive charging networks: a framework for smart electric vehicle charging," IEEE Trans. Smart Grid, vol. 12, no. 5, pp. 4339–4350, Sept. 2021, doi: 10.1109/TSG.2021.3074437. [Online].
[21] D. Qiu, Y. Wang, W. Hua, et al., "Reinforcement learning for electric vehicle applications in power systems: a critical review," Renew. Sustain. Energy Rev., vol. 173, p. 113052, 2023, doi: 10.1016/j.rser.2022.113052. [Online].
[22] C. Bi, Y. Tang, Y. Luo, et al., "Review on critical problems in reinforcement learning methods applied in power system optimization and control scenarios," Proc. CSEE, vol. 44, no. 1, pp. 1–21, Jan. 2024, doi: 10.13334/j.0258-8013.pcsee.223433. [Online].