Energy Conversion and Management 292 (2023) 117340

Contents lists available at ScienceDirect

Energy Conversion and Management

journal homepage: www.elsevier.com/locate/enconman

Real-time energy scheduling for home energy management systems with an energy storage system and electric vehicle based on a supervised-learning-based strategy

Truong Hoang Bao Huy [a], Huy Truong Dinh [b], Dieu Ngoc Vo [c,d], Daehee Kim [a,*]

[a] Department of Future Convergence Technology, Soonchunhyang University, Asan-si, Chuncheongnam-do 31538, South Korea
[b] Faculty of Engineering, Vietnamese-German University (VGU), Binh Duong, Vietnam
[c] Department of Power Systems, Ho Chi Minh City University of Technology (HCMUT), 268 Ly Thuong Kiet Street, District 10, Ho Chi Minh City, Viet Nam
[d] Vietnam National University Ho Chi Minh City, Linh Trung Ward, Thu Duc City, Ho Chi Minh City, Viet Nam

ARTICLE INFO

Keywords:
Home energy management system
Supervised learning
Deep neural network
Energy storage system
Electric vehicle
Deep reinforcement learning

ABSTRACT

With rising energy costs and concerns about environmental sustainability, there is a growing need to deploy Home Energy Management Systems (HEMS) that can efficiently manage household energy consumption. This paper proposes a new supervised-learning-based strategy for optimal energy scheduling of an HEMS that considers the integration of energy storage systems (ESS) and electric vehicles (EVs). The proposed supervised-learning-based HEMS framework aims to optimize the energy costs of households by forecasting the energy demand and simultaneously scheduling the charging and discharging operations of ESS and EV. From the scenarios extracted from historical data, the HEMS optimization problem is solved using a mixed-integer linear programming (MILP) solver to collect the datasets on the optimal actions of the ESS and EV. Accordingly, a supervised learning method is used to learn the optimal actions of the MILP solver using deep neural networks (DNNs). Well-trained DNNs act as decision-making tools that are subsequently applied to predict near-optimal actions for ESS and EV based on real-time data. The effectiveness of the proposed method is demonstrated through simulation results and compared with deep reinforcement learning-based and forecasting-based methods. The results show that the proposed method can significantly reduce energy costs and improve the efficiency of ESS and EV operations. Overall, the proposed supervised-learning-based HEMS offers a practical and effective solution for residential energy management.

1. Introduction

1.1. Background

The energy landscape has undergone a significant transformation in recent years with the rise of renewable energy sources (RES) such as solar and wind power and the increasing demand for electricity by households and businesses. This has led to the development of smart grid technologies and home energy management systems (HEMS) designed to optimize energy usage, reduce carbon emissions, and lower energy costs [1]. Smart grids enable consumers to participate in demand response (DR) programs where they can adjust their energy usage in response to price signals or grid conditions [2]. An HEMS is a type of smart home technology that allows homeowners to monitor and control their energy usage through a centralized platform. An HEMS uses advanced optimization algorithms and machine learning techniques to predict energy consumption patterns and provide real-time recommendations for optimal energy usage. Moreover, the addition of solar photovoltaics (PV) and energy storage systems (ESS) to HEMS has become increasingly important in recent years, enabling households to generate their own energy and reduce their reliance on the grid. An ESS can store excess energy generated from RES and provide it during periods of high demand. Electric vehicles (EVs) are becoming increasingly popular, and many households are investing in them to reduce their carbon footprints. EVs can act as mobile storage systems that can be charged during periods of excess supply and discharge energy back to the grid or households during periods of high demand. Overall, the integration of solar PV systems, ESS, and EVs into HEMS can result in significant energy savings, reduced carbon emissions, and improved energy security

* Corresponding author.
E-mail addresses: trhbhuy@sch.ac.kr (T.H.B. Huy), truongdinhhuy@gmail.com (H. Truong Dinh), vndieu@hcmut.edu.vn (D. Ngoc Vo), daeheekim@sch.ac.kr (D. Kim).

https://doi.org/10.1016/j.enconman.2023.117340
Received 19 April 2023; Received in revised form 9 June 2023; Accepted 23 June 2023
Available online 6 July 2023

Nomenclature

Abbreviations

ACKTR	actor-critic using Kronecker-Factored Trust Region
DDPG	deep deterministic policy gradient
DDQN	double deep Q-learning
DNN	deep neural network
DQN	deep Q-network
DRL	deep reinforcement learning
ESS	energy storage system
EV	electric vehicle
G2H	grid-to-home
G2V	grid-to-vehicle
H2G	home-to-grid
H2V	home-to-vehicle
HEMS	home energy management system
MADDPG	multi-agent deep deterministic policy gradient
MDP	Markov decision process
MILP	mixed-integer linear programming
MPC	model predictive control
PD-DDPG	primal–dual deep deterministic policy gradient
PDF	probability distribution function
PPO	proximal policy optimization
PV	photovoltaic
RES	renewable energy source
RL	reinforcement learning
SL	supervised learning
SOC	state of charge
TRPO	trust region policy optimization
V2G	vehicle-to-grid
V2H	vehicle-to-home

Parameters and Variables

DOD^{ESS}	ESS depth of discharge
DOD^{EV}	EV depth of discharge
P^{EC}	energy consumption of home appliances
$P^{ESS,ch}$	ESS charging power (kW)
$\overline{P}^{ESS,ch}$	maximum ESS charging power (kW)
$P^{ESS,dch}$	ESS discharging power (kW)
$\overline{P}^{ESS,dch}$	maximum ESS discharging power (kW)
P^{G2H}	grid-to-home power (kW)
$\overline{P}^{G2H}$	maximum grid-to-home power (kW)
P^{H2G}	home-to-grid power (kW)
$\overline{P}^{H2G}$	maximum home-to-grid power (kW)
$P^{EV,ch}$	EV charging power (kW)
$\overline{P}^{EV,ch}$	maximum EV charging power (kW)
$P^{EV,dch}$	EV discharging power (kW)
$\overline{P}^{EV,dch}$	maximum EV discharging power (kW)
P^{PV}	solar PV power (kW)
$\overline{P}^{PV}$	maximum solar PV power (kW)
T	number of time intervals
$u^{ESS,ch}$	binary variable – 1 if ESS charging mode is activated; otherwise 0
$u^{ESS,dch}$	binary variable – 1 if ESS discharging mode is activated; otherwise 0
u^{G2H}	binary variable – 1 if grid-to-home mode is activated; otherwise 0
u^{H2G}	binary variable – 1 if home-to-grid mode is activated; otherwise 0
$u^{EV,ch}$	binary variable – 1 if EV charging mode is activated; otherwise 0
$u^{EV,dch}$	binary variable – 1 if EV discharging mode is activated; otherwise 0
$\Delta\tau$	time step (hours)
ε^{ESS}	ESS state of charge (kWh)
$\overline{\varepsilon}^{ESS}$	maximum capacity of the ESS (kWh)
ε^{EV}	EV state of charge (kWh)
$\overline{\varepsilon}^{EV}$	maximum capacity of the EV (kWh)
η^{PV}	conversion efficiency of the solar PV system
η^{ESS}	ESS charging/discharging efficiency
η^{EV}	EV charging/discharging efficiency
λ^{G2H}	buying electricity tariff (Cents/kWh)
λ^{H2G}	selling electricity tariff (Cents/kWh)
υ	solar irradiance (kW/m^2)

[3]. With the continued development of renewable energy technologies and advancements in artificial intelligence and machine learning, the potential to revolutionize energy management is significant.

1.2. Literature review

With the increasing demand for sustainable living and energy efficiency, HEMS has become an increasingly popular area of research and development in recent years. However, the HEMS optimization problem faces a major hurdle because of the unpredictable and variable nature of renewable generation, market prices, and electricity demand. Accurate forecasting of these uncertainty factors is often unrealistic. Addressing these challenges requires a range of approaches, including advanced uncertainty-aware optimization techniques such as stochastic programming, robust optimization, and model predictive control (MPC), to mitigate the impact of uncertainties. In [3], the authors proposed a scenario-based stochastic optimization approach for an HEMS that minimized the energy costs, discomfort index, and peak-to-average ratio. Dorahaki et al. [4] introduced a behavioral HEMS model wherein prospect theory was adapted to account for time discounting, and a scenario-based method was used to model uncertainties. The authors in [5] developed a stochastic solution framework for an HEMS that used scalarizing functions and lexicographic optimization, allowing the energy cost to be the primary objective, while other objectives are secondary. A two-stage stochastic programming was proposed in [6] to reduce the electricity procurement costs in HEMS by considering the battery degradation cost, uncertainties, and parameter sensitivity. The authors in [7] introduced an energy management model for smart homes with PV and ESS, wherein uncertain energy market prices were defined using robust optimization. In [8], an interval optimization method was proposed for an HEMS that combines DR and user tolerance degrees to minimize costs and emissions, while considering uncertainties in system parameters. Gazafroudi et al. [9] proposed an autonomous HEMS to manage energy generation, consumption, and trade using a hybrid interval-stochastic optimization approach. An HEMS scheduling system based on rolling horizon optimization was proposed in [10], which optimizes smart appliance scheduling and RES using a genetic algorithm while considering the uncertainties. The authors in [11] proposed an MPC scheme for optimizing energy usage in smart homes. Jin et al. [12] introduced Foresee, an HEMS based on a multiobjective MPC framework that optimizes energy efficiency and cost savings while meeting user needs. The authors in [13] presented an MPC to optimize electricity and gas consumption in smart homes equipped with hybrid heating, ESS, EV, and PV. A chance-constrained optimization model was proposed by Huang et al. [14] to optimize the energy use of an HEMS while considering the uncertainties in electricity prices and load forecasting. Despite the significant contributions of these studies to the field of HEMS, they rely heavily on modeling and

predicting uncertainties in the system. Owing to the difficulty of accurately predicting or knowing the distribution of uncertainties, the accuracy of scheduling results is affected by the precision of the mathematical or forecasting model [15].

With the development of artificial intelligence technology, reinforcement learning (RL) offers an efficient resolution to address stochastic decision-making problems and obtains a more dependable scheduling strategy by bypassing the requirement to model or predict uncertainties. RL based on the Markov decision process (MDP) theory has proven to be effective in making decisions when there is no prior knowledge of the environment. This approach has been utilized to address numerous problems such as energy management in residential and commercial buildings. RL involves an agent that learns how to make decisions by interacting with the environment. The agent learns through a trial-and-error approach by adjusting its policy based on the rewards it receives. In [16], a multi-agent Q-learning algorithm was proposed to minimize the energy bills and discomfort of an HEMS, wherein energy prices were predicted using an artificial neural network. Simulations showed that it significantly reduced electricity costs compared to a benchmark without DR. Xu et al. [17] suggested a data-driven HEMS approach with a Q-learning algorithm and feedforward neural network to minimize the power consumption of EVs and household appliances. In that study, an extreme learning machine was applied to predict electricity prices and solar generation trends, and a multi-agent Q-learning algorithm was employed for decision-making. Nevertheless, the power of EVs was not consistently adaptable, and the potential impacts of V2H were not considered. A Q-learning algorithm was proposed in [18] to reduce energy usage and bills by shifting the peak load demand to off-peak periods. This approach used a single agent with fewer states and actions and fuzzy reasoning as the reward function. This reference showed an 18.5 % reduction in electricity costs during peak periods while accounting for user preferences and feedback. The authors in [19] proposed a multi-agent RL approach for multicarrier energy management in residential areas. Their method modeled energy management as a nonlinear optimization problem and utilized Q-learning to solve it, outperforming conventional optimization-based programs. Although RL methods are widely used, they often encounter problems related to computational cost and data efficiency. Furthermore, most variables in the HEMS problem are continuous. RL has difficulty dealing with continuous states and actions because RL algorithms represent policy or value functions using simple functions or tables.

Deep reinforcement learning (DRL) algorithms use deep neural networks (DNNs) to approximate the value or policy function used to determine the optimal action in a given state. This allows DRL algorithms to handle high-dimensional inputs and learn complex policies in continuous spaces. Zhao et al. [15] suggested a proximal policy optimization (PPO) algorithm to optimize HEMS operation considering uncertainties. The approach scheduled household devices without relying on predictions and reduced the average energy costs by 4.59 % compared with the MPC method and 12.17 % compared with the deep Q-network (DQN) algorithm. In [20], a deep deterministic policy gradient (DDPG) algorithm was introduced to manage the HVAC and ESS in a smart home using real-world data, accounting for the temperature range and uncertainties. The authors in [21] explored the application of a DQN and double deep Q-learning (DDQN) for HEMS using ESS and real-world data. That study concluded that the DDQN algorithm outperformed the particle swarm optimization algorithm and that the DDQN was more effective and generalizable than the DQN algorithm. Ding et al. [22] proposed a primal–dual DDPG (PD-DDPG), which is a safe reinforcement learning approach for solving multi-energy HEMS problems by incorporating a long short-term memory (LSTM) neural network for electricity price forecasting. The PD-DDPG-based HEMS effectively minimizes the total energy cost and avoids constraint violations, outperforming the DDPG and DQN algorithms. A model-free approach called actor-critic using the Kronecker-factored trust region (ACKTR) was proposed in [23] for an HEMS with solar PV and ESS. The ACKTR method improved the sampling efficiency, handled uncertainties from customer behavior and fluctuating prices, and outperformed the MPC method. Case studies showed that the ACKTR method reduced costs by 25.37 % in the test scenario. The authors in [24] investigated the application of DDPG to manage different systems in a smart home. An MILP was transformed into a DRL model and achieved better results than a rule-based method, achieving 75 % self-sufficiency while minimizing comfort violations. Ye et al. [25] proposed a DDPG approach using a prioritized experience replay strategy for a real-time autonomous HEMS, where the prioritized experience replay approach enhanced the quality of the policy and accelerated the learning process. Their approach achieved lower daily energy costs and was more cost-effective and computationally efficient than the traditional methods. In [26], DQN and DDPG were developed for building energy management systems based on the Pecan Street Inc. dataset, which provided real-time feedback to consumers to improve their electricity usage efficiency. In [27], a DQN-based optimal management strategy was proposed for low-carbon HEMS to minimize user dissatisfaction, carbon trading, and energy consumption. This strategy outperformed day-ahead forecasting-based management and demonstrated good performance in stochastic environments with high stability and convergence ability. A trust region policy optimization (TRPO) was proposed in [28] to schedule smart appliances and optimize residential DR under uncertainty. The proposed method handled both discrete and continuous actions and was evaluated using real-world data, it was found to perform better than benchmark approaches. Although the application of DRL to HEMS problems has yielded promising results, DRL algorithms have certain disadvantages. They can be computationally expensive, particularly for training large neural networks in complex environments. Moreover, DRL agents may have difficulty generalizing new scenarios that differ from those in the training environment. An agent may overfit the training data and fail to perform well in new situations. This can limit the scalability to larger problems and real-world applications.

Recently, supervised learning (SL) methods have been developed to solve energy management problems. Such methods have been inspired by imitation learning from experts instead of learning from scratch-like RL algorithms. The general idea of this approach is to solve optimization problems based on historical data to obtain optimal actions and then use deep learning models to approximate the optimal actions with the input of historical states of the environment. In the context of energy management, supervised-learning-based approaches were applied to mimic and schedule HVAC operations in a residential home [29] and multizone commercial building [30]. In [31], imitation learning was proposed for the real-time power scheduling of several ESSs in a small microgrid,

Table 1
Summary of the reviewed references and the current study on artificial intelligence-based approaches for HEMS.

References	RES	ESS	Bidirectional EV	H2G	Method
[16]	x	x	x	x	multi-agent RL
[17]	x	x	x	✓	Q-learning
[18]	✓	✓	x	x	Q-learning
[19]	✓	x	✓	✓	Q-learning
[15]	✓	x	x	✓	PPO
[20]	✓	✓	x	✓	DDPG
[21]	✓	✓	x	✓	DQN and DDQN
[22]	✓	✓	x	x	PD-DDPG
[23]	✓	✓	x	✓	ACKTR
[24]	✓	✓	x	✓	DDPG
[25]	✓	✓	x	✓	PDDPG
[26]	✓	x	x	✓	DQN and DDPG
[27]	✓	✓	✓	✓	DQN
[28]	x	x	✓	✓	TRPO
[29]	x	x	x	✓	Supervised learning
[30]	x	x	x	✓	Supervised learning
[31]	✓	✓	x	✓	Supervised learning
[32]	✓	✓	x	✓	Supervised learning
Present study	✓	✓	✓	✓	Supervised learning

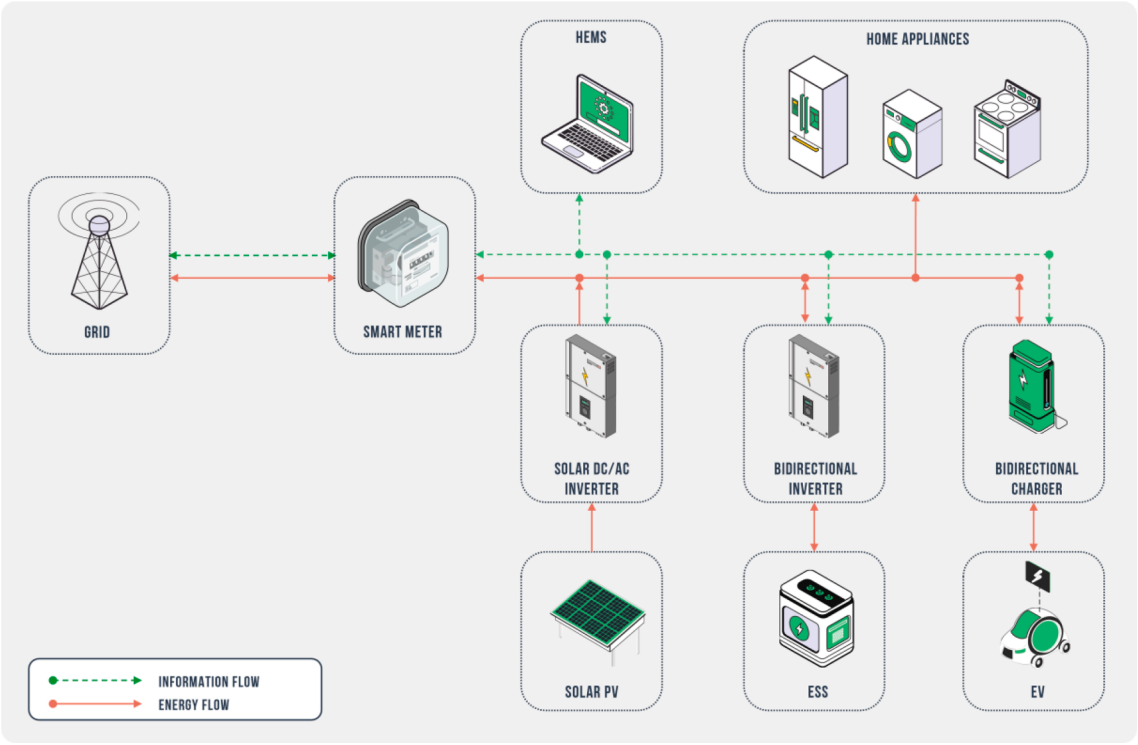

Fig. 1. Schematic diagram of the proposed HEMS.

which showed better performance than the MPC, PPO, and Q-learning methods. A similar approach was proposed in a previous study [32] to control ESS in different households.

1.3. Research gaps and motivation

A literature review shows that there has been much interest in studying HEMS problems. A comparison of the research studies on artificial intelligence-based methods in HEMS is provided in Table 1. The research gaps on this topic are listed as follows:

- In general, current solution approaches applied to HEMS have shortcomings. Because day-ahead strategies based on traditional methods [3–14] (e.g., stochastic optimization, robust optimization, etc.) rely on modeling and predicting uncertainties, inaccurate predictions or inaccurate distributions of uncertainties can result in suboptimal solutions [15]. Furthermore, they may not adapt to changing conditions in real time, such as changes in energy prices, weather conditions, or user behavior, making them less reliable. RL and DRL algorithms [15–28] usually start with no prior knowledge of the environment and must learn from scratch through trial and error, which can be time consuming. DRL algorithms may suffer from instability, non-convergence, or slow convergence. Therefore, it is necessary to develop a robust real-time energy scheduling strategy for the HEMS problem.
- The references focused on controlling and optimizing only devices that are available during the scheduling process, such as the ESS and HVAC. Several studies have not considered the energy scheduling of EVs in HEMS [15–18,20–26]. Today, EVs are becoming more mainstream and are increasingly seen as viable alternatives to traditional gasoline-powered vehicles. Unlike other devices, scheduling EVs can be challenging because of the unpredictable and intermittent nature of EV availability, including arrival time, departure time, and the remaining energy in the EV at arrival time. This can have a significant impact on the effectiveness of the HEMS. With the increasing popularity of EVs, the DR strategy should consider EV scheduling in HEMS, particularly for EVs with V2H capability.
- The application of supervised learning methods to HEMS problems remains limited. These studies only considered a simple HEMS model with a single device, namely HVAC [29,30] or ESS [31,32], and ignored EV scheduling. Scheduling multiple devices, especially EVs, may cause many difficulties in terms of solution quality and computation time owing to uncertainties. Thus, further research is required to address the scalability and applicability of HEMS using supervised learning methods with multiple devices in real-world scenarios.

1.4. Research contributions

Capturing research gaps from the literature review, this study develops an efficient supervised-learning-based HEMS framework to make real-time energy scheduling decisions for ESS and EV with the aim of minimizing daily energy costs. Because the HEMS problem is modeled as an MILP optimization problem in this study, the MILP solver can effectively achieve global optimal solutions when all the information is available. The information required for the HEMS problem includes the energy price, energy consumption, solar irradiation, and EV availability, which can be extracted from historical data or generated using scenario generation. In the supervised learning method, a dataset includes state-action pairs representing the optimal actions obtained by the MILP solver in response to different states. The optimal actions from the MILP solver include the charging/discharging power of the ESS and EV at specific time intervals, which are then approximated by training DNNs to map states to actions using the supervised learning method. The proposed method uses real-time data as inputs for the trained DNNs to predict the desired actions of the ESS and EV at the next time interval. To the best of our knowledge, this study is the first attempt to develop a supervised-learning-based strategy to optimally schedule the operation of a hybrid ESS and EV in an HEMS. The main contributions of this study are as follows:

- An HEMS is formulated as an MILP optimization problem with the integration of a solar PV system, ESS, and bidirectional EV, which aims to optimize the energy costs of households while maintaining their comfort and convenience. Furthermore, the HEMS interacts with the grid, enabling it to participate in DR programs and freely exchange energy with the grid. The developed HEMS considers both the energy demand and supply of households, the real-time electricity price, and the state of charge (SOC) of the ESS and EV to find the optimal charging and discharging operations for the ESS and EV using the MILP solver.
- A supervised learning method is integrated into the HEMS as a decision-making strategy. Based on various scenarios from historical data, the proposed method learns the optimal charge/discharge decisions of ESS and EV approximated by DNNs, thereby making more accurate predictions of real-time energy scheduling decisions. Specifically, well-trained DNNs are used to define near-optimal charge/discharge operations for the ESS and EV at each time interval.
- The performance of the supervised-learning-based HEMS framework is evaluated through simulations with real-world data, demonstrating its ability to effectively obtain an energy cost close to the ideal minimum while maintaining system constraints. A comparison of the proposed supervised-learning-based HEMS with the multi-agent deep deterministic policy gradient (MADDPG) algorithm and day-ahead forecasting-based method validates its superior performance and flexibility in adapting to different household input information.

1.5. Paper layout

The remainder of this paper is organized as follows. A description of the HEMS optimization problem is presented in Section 2. A detailed description of the proposed approach is presented in Section 3. The simulation results are presented in Section 4, and conclusions are presented in Section 5.

2. Problem formulation

Fig. 1 schematically illustrates an HEMS that integrates several components to provide an efficient and sustainable energy system for households. The main components of the HEMS are as follows:

- Home appliances: These are the devices and appliances in households that consume energy, such as lights, electronics, HVAC systems, and kitchen appliances. In this study, home appliances are freely used based on user preferences.
- Solar photovoltaic (PV) system: This system consists of solar panels that are used to convert solar energy into electricity. Since solar PV power output is dependent on solar irradiation, solar panels are installed on the roof or in a location that receives the maximum sunlight. Solar panels are connected to the solar DC/AC inverter which converts DC power to usable AC power.
- ESS: This system consists of a rechargeable battery and a bidirectional inverter. The bidirectional ESS inverter operates in a manner that allows both AC-DC and DC-AC conversions. ESS stores the excess energy generated by the solar panels or energy drawn by the grid during times of low electricity prices. The stored energy from ESS can be used flexibly during times of peak load demand or even deliver power to the grid.
- EV: Through a bidirectional charger, the EV can function as a storage system instead of being passively charged. The bidirectional EV charger operates similarly to the bidirectional ESS inverter, which converts AC to DC during charging and the reverse during discharging. When an EV is available at home, it can provide backup energy during intervals of high load demand and discharge energy back to the home when needed. Moreover, EV charging is scheduled at low electricity prices to reduce charging costs.
- HEMS: The HEMS controller is the system's brain. It collects real-time data and optimizes the operation schedule of the ESS and EV to minimize energy costs. Hence, residential users benefit from efficient energy management and a stable power supply.

An HEMS employs a communication network to transmit information between the user, utility grid, and HEMS. An HEMS gathers and controls data and communications from external sources, devices, and users. The data input to the HEMS include the energy usage of home appliances, state of charge (SOC) of the ESS and EV, solar radiation, and real-time electricity prices. The HEMS is integrated with the proposed approach to analyze the necessary information and provide optimal operation of the ESS and EV via control signals. Moreover, the proposed HEMS enables users to purchase and sell energy with the utility grid in a flexible manner. Communication between the utility grid and the HEMS is facilitated using a smart meter. This device is responsible for providing the HEMS with predetermined electricity rates from the utility grid and receiving information from the HEMS regarding the quantity of energy exchanged with the utility grid. Home energy demand can be satisfied by the utility grid, energy generated from solar PV panels, and energy stored by the ESS and EV. Another advantage of the DR strategy is that the excess energy can be fed back into the grid to generate revenue.

Several assumptions are considered for the HEMS paradigm in this study as follows:

- The effects of battery degradation of the ESS and EV due to charging/discharging cycles are ignored [1,33].
- For simplicity, the delay time of communication networks and power electronic inverters is negligible and is assumed to be neglected in this study [34–36].
- In the scenario of a sudden drop in the PV power, ESS and EV are assumed to maintain their charging/discharging power. In such unexpected scenarios, the HEMS controller adjusts the amount of energy exchanged with the grid to meet the household load demand.

The developed HEMS framework is expressed as a MILP formulation that spans 24 h in a day, with 24 intervals ($T = 24$) and a time step of 1 h ($\Delta\tau = 1$). The following subsections outline the mathematical formulation of each component of the HEMS and the optimized objective function.

2.1. HEMS modeling

2.1.1. Grid modeling

An HEMS operating in a prosumer environment can either purchase or sell energy to the utility grid. However, owing to the physical limitations of the distribution grid or contractual agreements, the quantity of electricity that can be exchanged between the home and the grid is subject to constraints imposed by utility companies. The grid model is formulated as follows [37,38]:

$$0 \le P_t^{G2H} \le u_t^{G2H} \cdot \overline{P}^{G2H}; \quad \forall t = 1, 2, \ldots, T \tag{1}$$

$$0 \le P_t^{H2G} \le u_t^{H2G} \cdot \overline{P}^{H2G}; \quad \forall t = 1, 2, \ldots, T \tag{2}$$

where P_t^{G2H} and P_t^{H2G} are the grid-to-home (G2H) and home-to-grid (H2G) powers at interval t, respectively; $\overline{P}^{G2H}$ and $\overline{P}^{H2G}$ are the limited powers that can be exchanged with the grid; and u_t^{G2H} and u_t^{H2G} are the binary variables indicating the G2H and H2G modes at interval t, respectively.

It is generally not feasible to carry out energy purchasing and selling procedures simultaneously, and this is enforced by the following constraints [37,38]:

$$0 \le u_t^{G2H} + u_t^{H2G} \le 1; \quad \forall t = 1, 2, \ldots, T \tag{3}$$

2.1.2. Solar PV system modeling

Typically, the expected solar PV output can be determined based on weather predictions, specifically solar irradiation. Thus, the expected power generation by a solar PV system during a given time interval t can be estimated using the following equation [5,38]:

$$P_t^{PV} = v_t \cdot \eta^{PV} \cdot \overline{P}^{PV} \cdot \Delta\tau; \quad \forall t = 1, 2, \ldots, T \tag{4}$$

where $\overline{P}^{PV}$ represents the maximum solar PV power, η^{PV} represents the conversion efficiency, and v_t represents the solar irradiance during the specified time interval t.

2.1.3. ESS modeling

Typically, ESSs are integrated into smart homes to provide both economic and technical advantages. The ESS can function as both a power source and energy consumer according to the operational mode (charging or discharging). In reality, the maximum amount of energy that can be charged or discharged by the ESS is restricted by its rated power, which is expressed as:

$$0 \le P_t^{ESS,ch} \le u_t^{ESS,ch} \cdot \overline{P}^{ESS,ch}; \quad \forall t = 1, 2, \ldots, T \tag{5}$$

$$0 \le P_t^{ESS,dch} \le u_t^{ESS,dch} \cdot \overline{P}^{ESS,dch}; \quad \forall t = 1, 2, \ldots, T \tag{6}$$

where $P_t^{ESS,ch}$ and $P_t^{ESS,dch}$ are the charging and discharging powers of the ESS at interval t, respectively; $\overline{P}^{ESS,ch}$ and $\overline{P}^{ESS,dch}$ are the maximum allowable charging and discharging powers of the ESS, respectively; and $u_t^{ESS,ch}$ and $u_t^{ESS,dch}$ are the binary variables denoting the charging and discharging modes of the ESS at interval t, respectively.

It must be ensured that the ESS charging and discharging functions are mutually exclusive:

$$0 \le u_t^{ESS,ch} + u_t^{ESS,dch} \le 1; \quad \forall t = 1, 2, \ldots, T \tag{7}$$

The amount of energy stored in the ESS can be modeled using Eq. (8) [39,40]:

$$\varepsilon_t^{ESS} = \varepsilon_{t-1}^{ESS} + \left(\eta^{ESS} \cdot P_t^{ESS,ch} - \frac{P_t^{ESS,dch}}{\eta^{ESS}} \right) \cdot \Delta\tau; \quad \forall t = 1, 2, \ldots, T \tag{8}$$

where ε_t^{ESS} is the amount of energy stored in the ESS at interval t and η^{ESS} is the charging/discharging efficiency of the ESS.

The overcharging/overdischarging of the ESS is prevented using the following equation:

$$(1 - DOD^{ESS}) \cdot \overline{\varepsilon}^{ESS} \le \varepsilon_t^{ESS} \le \overline{\varepsilon}^{ESS}; \quad \forall t = 1, 2, \ldots, T \tag{9}$$

where $\overline{\varepsilon}^{ESS}$ is the maximum capacity of the ESS; and DOD^{ESS} is the depth of discharge (DOD) of ESS.

The model assumes that the maximum capacity of the ESS is reached at the beginning and end of the scheduling period according to Eq. (10) [5].

$$\varepsilon_1^{ESS} = \varepsilon_T^{ESS} = \overline{\varepsilon}^{ESS} \tag{10}$$

2.1.4. EV modeling

This study explores the potential of EV by examining the vehicle-to-home (V2H) and home-to-vehicle (H2V) processes. A mathematical model is developed that treats an EV as a storage system composed of batteries. This approach allows the full utilization of EV capabilities. Eqs. (11) and (12) restrict the amount of power that an EV can exchange with the home. Furthermore, it stipulates that the EV cannot charge and discharge energy simultaneously, as shown in Eq. (13). These limits are expressed by the following equations [37]:

$$0 \le P_t^{EV,ch} \le u_t^{EV,ch} \cdot \overline{P}^{EV,ch}; \quad \forall t = 1, 2, \ldots, T \tag{11}$$

$$0 \le P_t^{EV,dch} \le u_t^{EV,dch} \cdot \overline{P}^{EV,dch}; \quad \forall t = 1, 2, \ldots, T \tag{12}$$

$$0 \le u_t^{EV,ch} + u_t^{EV,dch} \le 1; \quad \forall t = 1, 2, \ldots, T \tag{13}$$

where $P_t^{EV,ch}$ and $P_t^{EV,dch}$ are the charging and discharging powers of the EV at interval t, respectively; $\overline{P}^{EV,ch}$ and $\overline{P}^{EV,dch}$ are the maximum allowable charging and discharging powers of the EV, respectively; and $u_t^{EV,ch}$ and $u_t^{EV,dch}$ are the binary variables denoting the charging and discharging modes of the EV at interval t, respectively.

The SOC of the EV is modeled using Eq. (14), which is based on the SOC at interval $(t-1)$ and the energy charged/discharged from/to the home. The amount of energy stored in an EV is limited by its nominal capacity and *DOD* settings, as shown in Eq. (15). These constraints are expressed as follows [34,39]:

$$\varepsilon_t^{EV} = \varepsilon_{t-1}^{EV} + \left(\eta^{EV} \cdot P_t^{EV,ch} - \frac{P_t^{EV,dch}}{\eta^{EV}} \right) \cdot \Delta\tau; \quad \forall t = 1, 2, \ldots, T \tag{14}$$

$$\left(1 - DOD^{EV}\right) \cdot \overline{\varepsilon}^{EV} \le \varepsilon_t^{EV} \le \overline{\varepsilon}^{EV}; \quad \forall t = 1, 2, \ldots, T \tag{15}$$

where ε_t^{EV} is the amount of energy stored in the EV at interval t; $\overline{\varepsilon}^{EV}$ is the maximum capacity of the EV; η^{EV} is the charging/discharging efficiency of the EV; and DOD^{EV} is the DOD of the EV.

In contrast to an ESS, an EV is not available from home during certain intervals and is only engaged in the scheduling plan upon returning home. As stated in Eq. (16), the initial SOC of the EV is equivalent to the remaining energy at arrival time. At departure time, the EV must be completely charged, as shown in Eq. (17) [5,37].

$$\varepsilon^{EV,arrive} = \varepsilon^{EV,initial} \tag{16}$$

$$\varepsilon^{EV,depart} = \overline{\varepsilon}^{EV} \tag{17}$$

2.1.5. Energy balance

The proposed HEMS framework must guarantee that all the energy needs are met during the energy scheduling period, and that the energy balance in the home is maintained according to the following equation [1,3]:

$$P_t^{G2H} + P_t^{PV} + P_t^{ESS,dch} + P_t^{EV,dch} = P_t^{H2G} + P_t^{EC} + P_t^{ESS,ch} + P_t^{EV,ch}; \quad \forall t = 1, 2, \ldots, T \tag{18}$$

where P_t^{EC} is the energy consumption of home appliances at interval t.

2.2. Objective function

In this study, the HEMS optimization problem aims to optimally control the charging and discharging power of the ESS and EV to minimize the daily energy cost. The objective function is expressed as follows:

$$\mathbf{P1}: \quad \min \sum_{t=1}^{T} \left\{ \Delta\tau \cdot \left(\lambda_t^{G2H} \cdot P_t^{G2H} - \lambda_t^{H2G} \cdot P_t^{H2G} \right) \right\} \quad \text{s.t.}: (1) - (18) \tag{19}$$

where λ_t^{G2H} and λ_t^{H2G} are the buying and selling prices of electricity with the utility grid at interval t, respectively.

The vector of decision variables is given by:

$$\boldsymbol{x} = \left\{ \begin{matrix} P_t^{G2H}, P_t^{H2G}, P_t^{ESS,ch}, P_t^{ESS,dch}, P_t^{EV,ch}, P_t^{EV,dch}, \varepsilon_t^{ESS}, \varepsilon_t^{EV}, \\ u_t^{G2H}, u_t^{H2G}, u_t^{ESS,ch}, u_t^{ESS,dch}, u_t^{EV,ch}, u_t^{EV,dch} \end{matrix} \right\}; \quad \forall t = 1, 2, \ldots, T \tag{20}$$

To ensure maximum user convenience, this study assumes that home appliances operate flexibly according to user preferences. This study does not involve any load shifting or cutting in relation to the DR

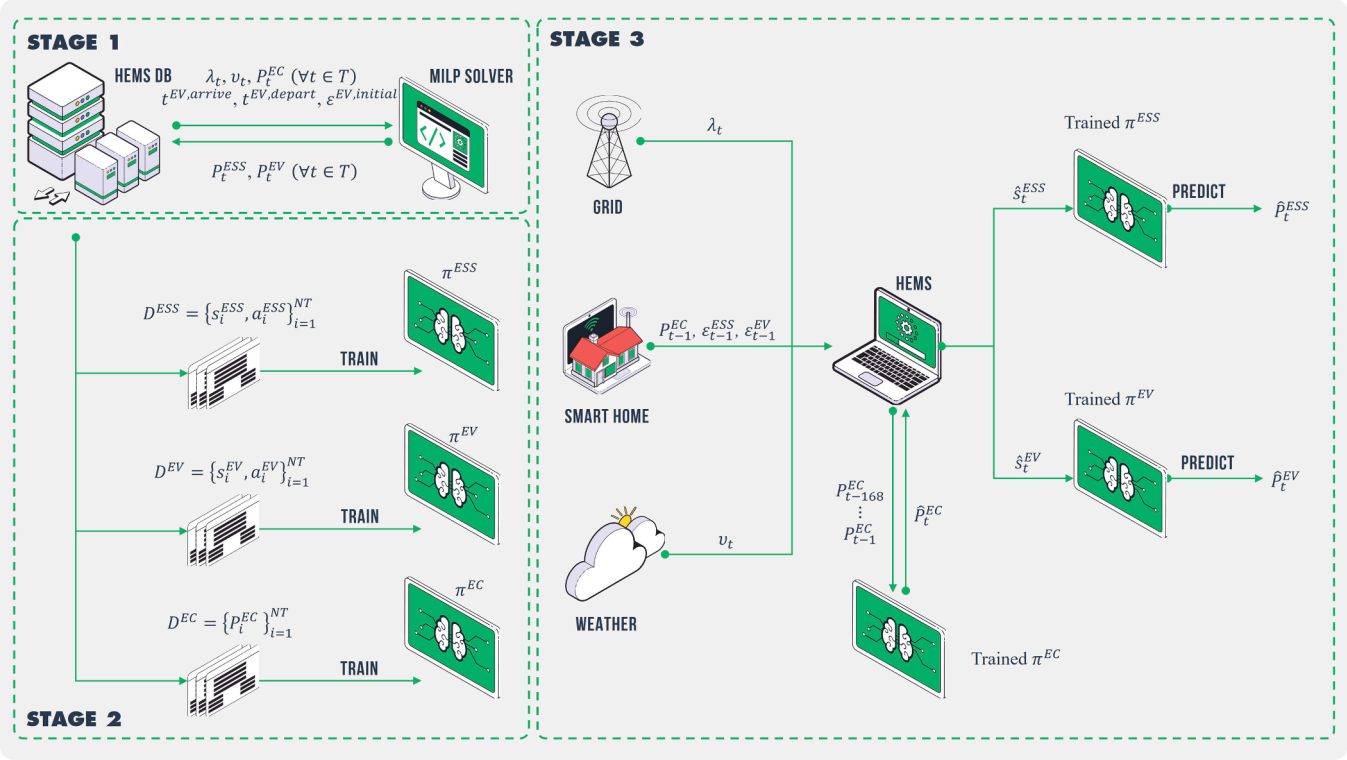

Fig. 2. Overall framework of the proposed supervised-learning-based HEMS strategy.

strategy.

3. Proposed methodology

The HEMS optimization problem can be considered a sequential decision problem. Solving problem **P1** is challenging because uncertain future data such as energy prices, solar PV generation, and EV availability are not accurately known in advance. Although problem **P1** can be solved with the predicted values, forecasting errors are unavoidable and can affect its optimal solution [31,41–43]. To overcome the above limitations, a real-time energy scheduling strategy, namely a supervised-learning-based HEMS framework, is proposed to schedule ESS and EV operations in real time. Two other artificial intelligence-based strategies, MADDPG-based and forecasting-based methods are also developed to compare and validate the performance of the proposed supervised learning method. Descriptions of the proposed method and the two comparable methods are presented in the following subsections.

3.1. Supervised learning method

In the proposed supervised learning method, decision variables, also known as actions, are defined online using real-time state variables. In the first stage, the MILP solver is used to solve the HEMS problem, which generates sets of state-action pairs (i.e., input–output pairs) for the ESS and EV. The proposed method then trains the DNNs to approximate the correct action for each state based on the optimal results of the MILP solver. The trained DNNs are then used to predict the desired actions of the ESS and EV in real time instead of a day-ahead prediction. By learning from the optimal actions of the MILP solver, the proposed method can achieve a higher level of performance than that achieved through trial and error. The overall framework of the proposed supervised-learning-based HEMS is depicted graphically in Fig. 2, which is also presented in Algorithm 1. The primary stages of the proposed framework are described in the following subsections.

3.1.1. Data preprocessing and aggregation

In practice, by gathering historical data or generating data using scenario generation, it is possible to collect numerous scenarios, in which each scenario corresponds to an energy scheduling cycle. For each scenario, historical data over the scheduling horizon of T intervals including energy price ($\lambda_1, \lambda_2, \ldots, \lambda_T$), energy consumption ($P_1^{EC}, P_2^{EC}, \ldots, P_T^{EC}$), solar irradiation ($\upsilon_1, \upsilon_2, \ldots, \upsilon_T$), arrival time ($t^{EV,arrive}$), departure time ($t^{EV,depart}$), and initial SOC of EV ($\varepsilon^{EV,initial}$) are selected as the input of the MILP optimization problem **P1** that is then solved by the MILP solver. Accordingly, the decision variables and other optimal values over the horizon of T intervals are also defined. Note that in this case, the optimal solution is the ideal solution (i.e., the best possible solution).

The optimal actions of the ESS and EV at each interval t are defined by combining their charging and discharging powers into a single variable as follows:

$$P_t^{ESS} = P_t^{ESS,ch} - P_t^{ESS,dch}; \quad \forall t = 1, 2, \ldots, T \tag{21}$$

$$P_t^{EV} = P_t^{EV,ch} - P_t^{EV,dch}; \quad \forall t = 1, 2, \ldots, T \tag{22}$$

Moreover, the net load between the energy consumption and the energy generated from the solar PV system at interval t is calculated using the following equation:

$$P_t^{net} = P_t^{EC} - P_t^{PV}; \quad \forall t = 1, 2, \ldots, T \tag{23}$$

Subsequently, all the obtained data are structured into a dataset of T state-action pairs $D^{ESS} = \{s_t^{ESS}, a_t^{ESS}\}_{t=1}^{T}$ with respect to the ESS decision, as follows:

$$s_t^{ESS} = \left[t, \lambda_t, P_t^{net}, \varepsilon_{t-1}^{ESS}\right] \tag{24}$$

$$a_t^{ESS} = \left[P_t^{ESS}\right] \tag{25}$$

where s_t^{ESS} is the state vector corresponding to the ESS, including the time interval (t), energy price at interval t (λ_t), net load at interval t (P_t^{net}), and the SOC of the ESS at the previous interval t (ε_{t-1}^{ESS}). The selling and

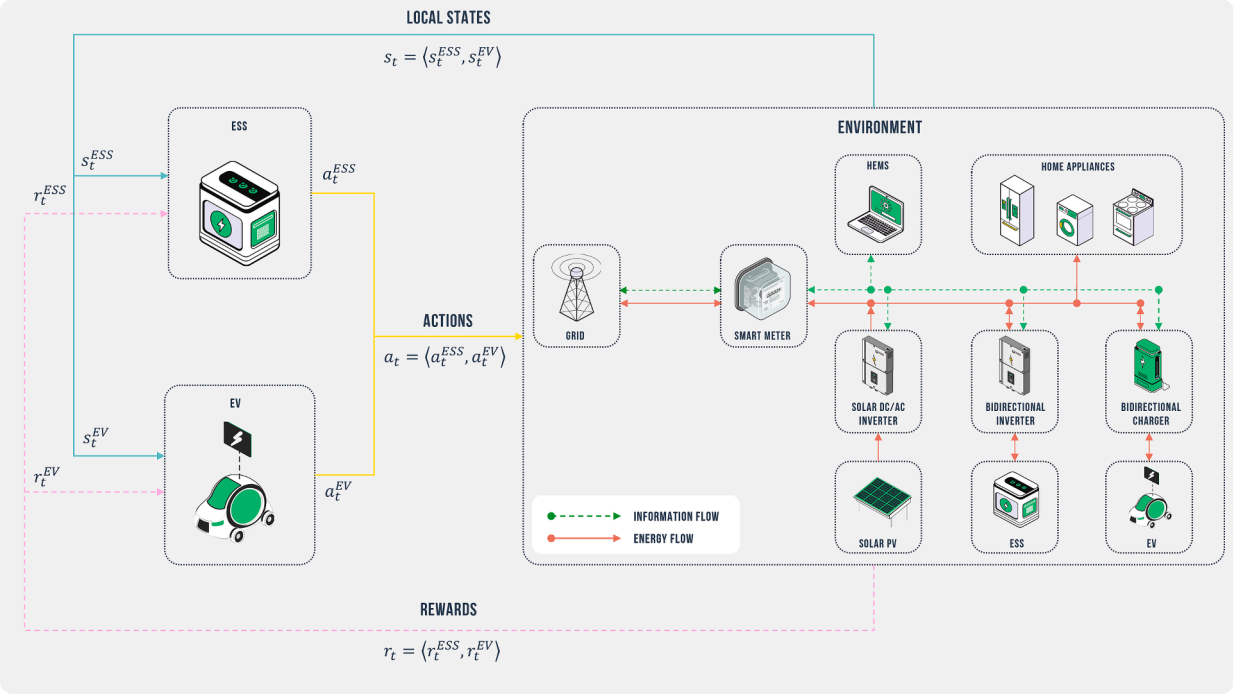

Fig. 3. The multi-agent environment structure of HEMS.

purchase prices of electricity are correlated, which is generally expressed through the energy price (λ_t) in the state vector. a_t^{ESS} refers to the corresponding optimal actions of the ESS at interval *t*.

Similarly, a dataset of *T* state-action pairs $D^{EV} = \{s_t^{EV}, a_t^{EV}\}_{t=1}^{T}$ with respect to EV decision can be formulated as follows:

$$s_t^{EV} = \left[t, \lambda_t, P_t^{net}, \varepsilon_{t-1}^{EV}\right] \tag{26}$$

$$a_t^{EV} = \left[P_t^{EV}\right] \tag{27}$$

where s_t^{EV} is the state vector corresponding to the EV, and a_t^{EV} refers to the corresponding optimal action of the EV at interval *t*.

This process is iteratively implemented for all considered scenarios, and the datasets of state-action pairs obtained from all considered scenarios are combined, as shown in lines 1–7 of Algorithm 1. Thus, solving problem **P1** with *N* scenarios (each scenario has *T* intervals) forms two datasets of *NT* state-action pairs: $D^{ESS} = \{s_i^{ESS}, a_i^{ESS}\}_{i=1}^{NT}$ and $D^{EV} = \{s_i^{EV}, a_i^{EV}\}_{i=1}^{NT}$.

3.1.2. Approximate mappings from states to actions using deep neural networks

Inspired by imitation learning, the supervised learning method learns and mimics the optimal actions obtained by the MILP solver. Assuming that all necessary information is available from historical data, the optimal actions of the ESS and EV can be easily achieved by solving problem **P1**. Accordingly, the MILP solver serves as a reliable expert and the optimal solutions achieved by solving problem **P1** are deemed perfect expert demonstrations. Based on expert demonstrations, the proposed supervised learning method uses a dataset (D) consisting of pairs of actions and states (s_i, a_i) to train a function approximator to map the states to the corresponding optimal actions. For a given state (s), a function approximator generates an expected action ($\hat{a} = \pi(s)$) to imitate the optimal decision of the MILP solver [44]. Accordingly, the original sequential decision-making problem is converted into a supervised-learning-based regression problem [31]. Owing to its widely recognized abilities in approximation and generalization, DNNs are chosen as the function approximator [31,41].

In the context of energy scheduling, two DNNs are trained on two different expert datasets for two mappings π^{ESS} and π^{EV} with respect to ESS and EV, respectively. This process is described in lines 8–11 of Algorithm 1. A lightweight deep feed-forward network is applied to handle a regression problem with continuous inputs and outputs. The networks are trained to minimize the loss function of the mean absolute error (MAE) using the Adam optimizer. Moreover, a gated recurrent unit-based recurrent neural network (RNN) is trained to predict the energy consumption for the next interval.

3.1.3. Real-time energy scheduling

The implementation process of the proposed supervised-learning-based HEMS for real-time energy scheduling is outlined in lines 12–22 of Algorithm 1. The generalizability of the trained DNNs is expected to lead to a real-time DR strategy that can perform well in future scenarios. In this study, it is assumed that home appliances are operated flexibly based on user preferences. No load shifting or cutting related to the DR strategy is performed. For these reasons, energy consumption in the future needs to be predicted using an RNN rather than accurately collected. At each interval *t*, the trained RNN utilizes the past 168 h of energy consumption data ($P_{t-168}^{EC},\ldots,P_{t-1}^{EC}$) as input to forecast the energy consumption data for the interval *t* ($\widehat{P}_t^{EC}$). Except for the energy consumption of home appliances, the remaining data are easily collected in real time. Utility grids consistently provide users with accurate electricity prices for 1–2 h in advance. Moreover, the forecast values of solar irradiation are also very accurate for the next hour through the meteorological center. Accordingly, the net load is calculated based on the forecasted energy consumption as $\widehat{P}_t^{net} = \widehat{P}_t^{EC} - P_t^{PV}$. At each interval *t*, the HEMS collects real-time data and defines two state vectors, $\hat{s}_t^{ESS}$ and $\hat{s}_t^{EV}$, as follows:

$$\hat{s}_t^{ESS} = \left[t, \lambda_t, \widehat{P}_t^{net}, \varepsilon_{t-1}^{ESS}\right] \tag{28}$$

$$\hat{s}_t^{EV} = \left[t, \lambda_t, \widehat{P}_t^{net}, \varepsilon_{t-1}^{EV}\right] \tag{29}$$

Because the EV is not available in some intervals of the energy-scheduling horizon, the SOC of the EV at this interval takes a value of

zero in Eq. (29) when the EV is not connected to the smart home. Upon obtaining the required information (i.e., state vectors), the proposed supervised-learning-based HEMS is adopted to iteratively make energy scheduling decisions for the ESS and EV charging/discharging over a scheduling cycle (i.e., $t = 1, 2,\ldots, T$), as shown in lines 12–22 of Algorithm 1. At each interval t, two trained DNNs π^{ESS} and π^{EV} take the inputs of the state vectors $\hat{s}_t^{ESS}$ and $\hat{s}_t^{EV}$ and then predict the expected actions $\hat{a}_t^{ESS}$ and $\hat{a}_t^{EV}$ for the ESS and EV, respectively. If the predictions deviate significantly from the expected actual values, the raw outputs (i.e., the predicted actions of the ESS and EV) of the trained DNNs are simply post-processed during energy scheduling at each interval to satisfy their related constraints, which are outlined in line 19 of Algorithm 1 and can be given as follows:

$$\hat{a}_t^{ESS} = \begin{cases} P_t^{ESS,ch}; & \text{if } \hat{a}_t^{ESS} > P_t^{ESS,ch} \\ -P_t^{ESS,dch}; & \text{if } \hat{a}_t^{ESS} < -P_t^{ESS,dch} \\ \hat{a}_t^{ESS}; & \text{otherwise} \end{cases} \tag{30}$$

$$\hat{a}_t^{EV} = \begin{cases} P_t^{EV,ch}; & \text{if } \hat{a}_t^{EV} > P_t^{EV,ch} \\ -P_t^{EV,dch}; & \text{if } \hat{a}_t^{EV} < -P_t^{EV,dch} \\ \hat{a}_t^{EV}; & \text{otherwise} \end{cases} \tag{31}$$

The predicted actions, $\hat{a}_t^{ESS}$ and $\hat{a}_t^{EV}$, are used to control the ESS and EV, thereby determining the SOCs of ESS and EV at interval t. This procedure is repeated until the end of the energy scheduling cycle ($t = T$).

3.2. Multi-Agent deep deterministic policy gradient (MADDPG)

3.2.1. Markov decision process (MDP) formulation

The HEMS optimization problem can be mathematically formulated as an MDP comprising several essential elements, including the environment, agent, state, action, and reward functions. At each time interval, the agent observes the current state of the environment and performs an action. The agent then receives an immediate reward from the environment, which transitions to a new state. Fig. 3 illustrates the iterative interaction between the agents and the environment in this MDP framework.

The MDP formulation for the HEMS problem can be given as follows:

Environment: This environment is represented by the proposed HEMS, as shown in Fig. 1.

Agent: The HEMS has two agents representing the ESS and EV. The agents can be trained to develop an optimal energy scheduling strategy based on the state of the environment.

State: The environmental state refers to the information that an agent can observe from the environment to make decisions. Similar to the proposed supervised learning method, environmental states contain real-time HEMS data. The state vectors are expressed as follows:

$$s_t^{ESS} = \left[t, \lambda_t, P_t^{net}, \epsilon_{t-1}^{ESS}\right] \tag{32}$$

$$s_t^{EV} = \left[t, \lambda_t, P_t^{net}, \epsilon_{t-1}^{EV}\right] \tag{33}$$

where s_t^{ESS} and s_t^{EV} denote the state vectors observed by the ESS and EV agents, respectively.

Action: The agents in an HEMS aim to make energy scheduling decisions according to the observed environmental state. The decision variables in problem **P1** include the charging and discharging power of the ESS and EV. Therefore, the action of each agent can be defined as:

Algorithm 1: Supervised learning-based HEMS framework

1 /* **Stage 1: Data preprocessing and aggregation** */
  Input: N training scenario $\mathbb{S}$
  Output: Dataset $D^{ESS} = \{s_i^{ESS}, a_i^{ESS}\}_{i=1}^{NT}$
  Dataset $D^{EV} = \{s_i^{EV}, a_i^{EV}\}_{i=1}^{NT}$
2 **for** each scenario S in $\mathbb{S}$ **do**
3 Solve problem **P1** to obtain $\boldsymbol{x}_t^*, \forall t = 1, ..., T$
4 Extract T state-action pairs for ESS by Eqs. (24) and (25)
5 Extract T state-action pairs for EV by Eqs. (26) and (27)
6 Append state-action pairs to D^{ESS} and D^{EV}
7 **end**
8 /* **Stage 2: Approximate mappings from states to actions using DNNs** */
9 Train a DNN π^{ESS} on D^{ESS} to map state (24) to action (25)
10 Train a DNN π^{EV} on D^{EV} to map state (26) to action (27)
11 Train a RNN π^{EC} to forecast the energy consumption data for the next hour
12 /* **Stage 3: Real-time energy scheduling for a specific scenario** */
13 **for** $t \leftarrow 1$ to T **do**
14 Forecast the energy consumption data for interval t ($\hat{P}_t^{EC}$)
15 Define the net load as $\hat{P}_t^{net} = \hat{P}_t^{EC} - P_t^{PV}$
16 Define states $\hat{s}_t^{ESS}$ and $\hat{s}_t^{EV}$ by Eqs. (28) and (29) with real-time data
17 Predict ESS action by $\hat{a}_t^{ESS} \leftarrow \pi^{ESS}(\hat{s}_t^{ESS})$
18 Predict EV action by $\hat{a}_t^{EV} \leftarrow \pi^{EV}(\hat{s}_t^{EV})$
19 Post-processing $\hat{a}_t^{ESS}$ and $\hat{a}_t^{EV}$ to meet all constraints of HEMS problem
20 Define charging/discharging power of ESS and EV accordingly
21 Implement energy scheduling operations to HEMS
22 **end**

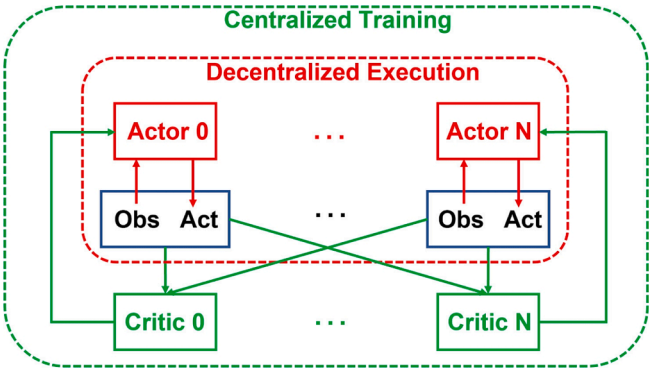

Fig. 4. MADDPG architecture [45,47].

$$a_t^{ESS} = \left[P_t^{ESS}\right] \tag{34}$$

$$a_t^{EV} = \left[P_t^{EV}\right] \tag{35}$$

Reward function: Two agents receive the same reward r_t after performing actions a_t^{ESS} and a_t^{EV}. As the objective of HEMS energy scheduling is to minimize the energy cost, the reward function at interval t is defined based on the energy cost at interval t. Furthermore, two penalty functions are imposed on the reward function to ensure that the constraints of the ESS and EV in the HEMS problem are fully satisfied. The reward function is defined as follows:

$$r_t = -\left[C_t^1 + C_t^2 + C_t^3\right] \tag{36}$$

where C_t^1 denotes the energy cost at interval t in Eq. (19).

In Eq. (36), C_t^2 is the penalty function for binding the SOC of the ESS to its maximum value in the final interval, which can be expressed as follows:

$$C_t^2 = \begin{cases} \omega^{ESS}\left(\bar{\varepsilon}^{ESS} - \varepsilon_t^{ESS}\right); & \forall t = T \\ 0; & \text{otherwise} \end{cases} \tag{37}$$

The term C_t^3 is a penalty function that forces the EV to be fully charged at departure time, which is defined as follows:

$$C_t^3 = \begin{cases} \omega^{EV}\left(\bar{\varepsilon}^{EV} - \varepsilon_t^{EV}\right); & \forall t = t^{EV,depart} \\ 0; & \text{otherwise} \end{cases} \tag{38}$$

where ω^{ESS} and ω^{EV} are weighting factors that are both set to 100 Cents/kWh in this study.

The objective of the MDP is to find the optimal energy scheduling policy π^* to maximize the expectation of discounted cumulative rewards over a horizon of T intervals, as follows:

$$\max_{\pi\in\Pi} J(\pi) = \mathrm{E}_\pi\left[\sum_{t=1}^{T}\gamma^{t-1} r_t\right] \tag{39}$$

where E_π is the expected value under policy π, Π is the set of all permissible policies, and $\gamma \in [0, 1]$ is the discount factor.

3.2.2. Learning algorithm

The MADDPG algorithm [45] is a variant of the DDPG algorithm [46] used for handling control problems in multi-agent environments and continuous action spaces. The key concept behind MADDPG is to allow each agent to learn its own individual policy while considering the actions of other agents in the environment. This is achieved by training each agent's policy using a centralized critic that receives the observations and actions of all the agents as inputs. This allows each agent access to a global view of the state of the environment, rather than just its own local observations. Fig. 4 depicts the MADDPG framework, in which each agent has an actor network and a critic network.

Before training, each agent p randomly initializes an original actor network μ_p and an original critic network Q_p that are parameterized by $\theta_p^{\mu_p}$ and $\theta_p^{Q_p}$. To improve the training stability, a target actor network $\mu_p^{'}$ and a target critic network $Q_p^{'}$ are also created, whereas their respective parameters $\theta_p^{\mu_p^{'}}$ and $\theta_p^{Q_p^{'}}$ are initialized to be identical to those of the original networks as $\theta_p^{\mu_p^{'}} \leftarrow \theta_p^{\mu_p}$ and $\theta_p^{Q_p^{'}} \leftarrow \theta_p^{Q_p}$.

For each agent, a replay buffer $\mathscr{D}$ is created to store a list of tuples $\langle \boldsymbol{s}_t, \boldsymbol{a}_t, \boldsymbol{r}_t, \boldsymbol{s}_{t+1}\rangle$ known as experiences, where $\boldsymbol{s} = \langle s_t^1, \ldots, s_t^P\rangle$,$\boldsymbol{a} = \langle a_t^1, \ldots, a_t^P\rangle$,$\boldsymbol{r} = \langle r_t^1, \ldots, r_t^P\rangle$, and $\boldsymbol{s}_{t+1} = \langle s_{t+1}^1, \ldots, s_{t+1}^P\rangle$. Training stability is also enhanced by the replay buffer, which allows agents to learn from mini-batches sampled from all experiences accumulated during training.

At the beginning of each training episode, an initial state is initialized and a random process is used to generate noise to accelerate the agent's exploration of the environment. Using the observed state s_t^p and noise N_t, each agent selects an action as follows:

$$a_t^p = \mu_p(s_t^p) + N_t \tag{40}$$

where $\mu_p(s_t^p)$ is the output (action) of the actor network.

At the end of the time interval, each agent calculates its reward r_t^p and observes a new state s_{t+1}^p. The experience $\langle \boldsymbol{s}_t, \boldsymbol{a}_t, \boldsymbol{r}_t, \boldsymbol{s}_{t+1}\rangle$ is stored in replay buffer $\mathscr{D}$, and the initial state is updated as $\boldsymbol{s}_t \leftarrow \boldsymbol{s}_{t+1}$. After every number of episodes, for each agent p, actor network μ_p and critic network Q_p are trained by randomly sampling B transitions from replay buffer $\mathscr{D}$. The transitions are used to update the network weights for both the original actor and critic networks and the target actor and critic networks. Let $\left\langle \boldsymbol{s}_t^j, \boldsymbol{a}_t^j, \boldsymbol{r}_t^j, \boldsymbol{s}_{t+1}^j \right\rangle$ represent the experience of each transition j. The weights of the original critic network ($\theta_p^{Q_p}$) associated with each agent p are updated through gradient descent algorithm using the mean-square Bellman error function as follows:

$$L\left(\theta_p^{Q_p}\right) = \frac{1}{B}\sum_{j=1}^{B}\left(y^{p,j} - Q_p(s_t^j, a_t^j)\right)^2 \tag{41}$$

where $Q_p(\boldsymbol{s}_t^j, \boldsymbol{a}_t^j)$ is the predicted output of the original critic network, and $y^{p,j}$ is its target value, which is given by.

$$y^{p,j} = r_t^{p,j} + \gamma Q_p^{'}\left(\boldsymbol{s}_{t+1}^j, a_{t+1}^1, \ldots, a_{t+1}^P\right)\Big|_{a_{t+1}^{p'}=\mu_p^{'}\left(s_{t+1}^p\right)} \tag{42}$$

where $a_{t+1}^p = \mu_p^{'}\left(s_{t+1}^p\right)$ is the action predicted by the target actor network, and $Q_p^{'}\left(\boldsymbol{s}_{t+1}^j, a_{t+1}^1, \ldots, a_{t+1}^p\right)$ is the value predicted by the target critic network.

Simultaneously, the weights of the original actor network ($\theta_p^{\mu_p}$) are updated based on the sampled policy gradient, as follows:

$$\nabla_{\theta_p^{\mu_p}} J(\theta_p^{\mu_p}) = \nabla_{\theta_p^{\mu_p}} \mu_p(s_t^p) \nabla_{a_t^p} Q_p(\boldsymbol{s}^j, \boldsymbol{a}) \tag{43}$$

where $\boldsymbol{a} = \left(\mu_1(s_t^1), \ldots, \mu_P(s_t^P)\right)$.

Once the weights of the original actor and critic networks are updated, the target actor and critic networks are updated as follows:

$$\begin{cases} \theta_p^{Q_p^{'}} \leftarrow \tau\theta_p^{Q_p} + (1-\tau)\theta_p^{Q_p^{'}} \\ \theta_p^{\mu_p^{'}} \leftarrow \tau\theta_p^{\mu_p} + (1-\tau)\theta_p^{\mu_p^{'}} \end{cases} \tag{44}$$

where τ refers to the learning rate.

In summary, MADDPG training is centralized, whereas MADDPG execution is decentralized, as illustrated in Fig. 4. In other words, the agent uses only the trained actor network μ_p^* and discards the trained critic network and replay buffer during execution. At each interval, agent p observes the local state s_t^p to perform the desired action as $\widehat{a}_t^p = \mu_p^*(s_t^p)$. Similar to supervised learning, the raw outputs of joint actions

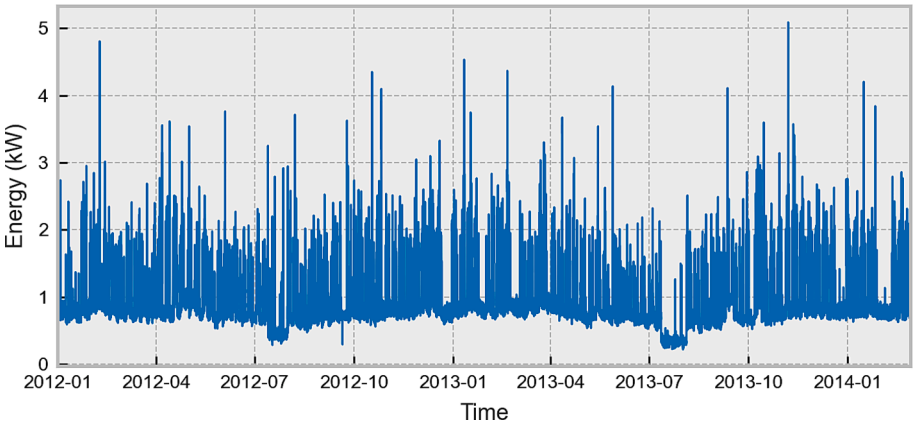

Fig. 5. Hourly energy consumption of smart home 1.

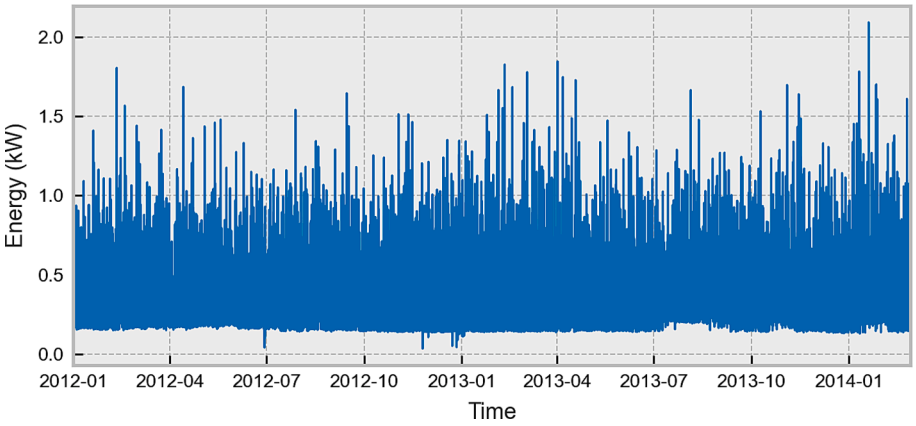

Fig. 6. Hourly energy consumption of smart home 2.

from all agents are post-processed to meet all the required constraints and then used to make real-time energy-scheduling decisions in the HEMS.

3.3. Forecasting-based method

The forecasting-based method is a day-ahead energy scheduling strategy that seeks to schedule ESS and EV operations in an HEMS over the next 24 h. Thus, it solves problem **P1** based on forecast data for the next 24 h. To this end, three RNN models are developed as time-series predictive models, which are trained on historical data, and are used to predict future observations. The inputs of the RNNs are the past 168 h of energy price data, energy consumption data, and solar irradiation data, and their outputs are forecasted future 24 h trends of energy consumption, solar irradiation, and electricity price. This method assumes that the EV data take the worst-case values for arrival time, departure time, and initial SOC. These predictions and hypothetical information are fed into the MILP solver to solve problem **P1** and define the ESS and EV energy schedules for the next scheduling cycle. Forecast-based values are used to control the ESS and EV during the day, despite unusual real-time fluctuations in energy price, energy consumption, and weather conditions.

4. Simulation results

In this section, the proposed framework is validated using real-world data, including energy price, energy consumption, solar irradiation, ESS configuration, and EV data for two different residential users. The simulations are performed over a 24 h time horizon with a time step of 60 min (12:00–11:00 h), which leads to 24 intervals in a daily energy scheduling cycle. The MILP formulation of the HEMS system is developed in Python and solved using the Gurobi optimizer. The simulations are performed on a 64-bit Intel (R) Core(TM) i7-7700 CPU at 3.6 GHz with 16 GB RAM.

Table 2
Data of solar PV system, ESS, and EV for two smart homes.

Device	Details	Smart home 1	Smart home 2
Solar PV	$\overline{P}^{PV}$	2 kW	1 kW
	η^{PV}	0.9	
ESS	$\overline{\varepsilon}^{ESS}$	5 kWh	2 kWh
	$\overline{P}^{ESS,ch}/\overline{P}^{ESS,dch}$	2 kW	0.5 kW
	η^{ESS}	0.98	
	DOD^{ESS}	0.8	
EV	$\overline{\varepsilon}^{EV}$	24 kWh	22 kWh
	$\overline{P}^{EV,ch}/\overline{P}^{EV,dch}$	3.3 kW	3 kW
	η^{EV}	0.98	
	DOD^{EV}	0.8	

4.1. Input data

In this study, two typical smart homes with different input data are used to evaluate the proposed model. The historical dataset for the hourly energy consumption associated with smart homes 1 and 2 are depicted in Figs. 5 and 6, respectively, which are extracted from the "Energy Consumption Data in London Households" dataset of the UK Power Network from Jan 2012 to Feb 2014 [48]. As shown in Figs. 5 and

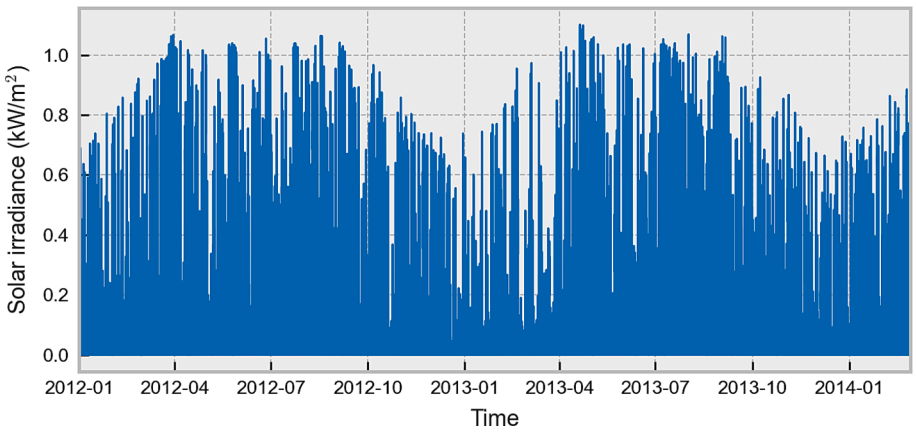

Fig. 7. Hourly global solar irradiance.

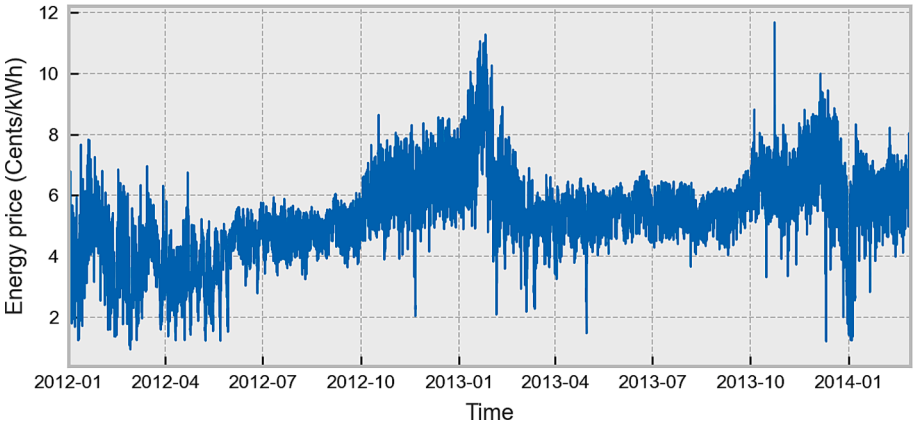

Fig. 8. Real-time energy price.

Table 3
Considered data for EV availability.

	Mean	Standard deviation	Min	Max
Arrival time (h)	16	3	14	19
Departure time (h)	8	3	5	10
Initial SOC (%)	50	25	30	95

6, smart homes 1 and 2 have two different types of load patterns with average energy consumption values of 22 kW/day and 9.53 kW/day, respectively. To increase the realism of the simulation, the configuration parameters of the solar PV system, ESS, and EV are selected based on the energy consumption of each smart home. Table 2 lists the data of the equipped devices for each corresponding smart home, which are obtained from [1,3,5].

In our simulations, solar irradiance is extracted from the European Commission in London (UK) from January 2012 to February 2014 [49], as depicted in Fig. 7. Because electricity tariffs in London are not available from open sources, real-time electricity prices from January 2016 to February 2018 taken from the Spanish Transmission Service Operator - Red Electric España [50,51] are considered in this study, as shown in Fig. 8. The selling price is assumed to be equal to the purchase price [3]. The maximum powers that could be purchased and sold between the home and utility grid are 10 and 6 kW, respectively. For the proposed strategies, data from January 2012 to January 2014 (i.e., 762 training scenarios) are used as training data, and data from February 2014 (i.e., 26 test scenarios) are used to test and evaluate the performance of the proposed method.

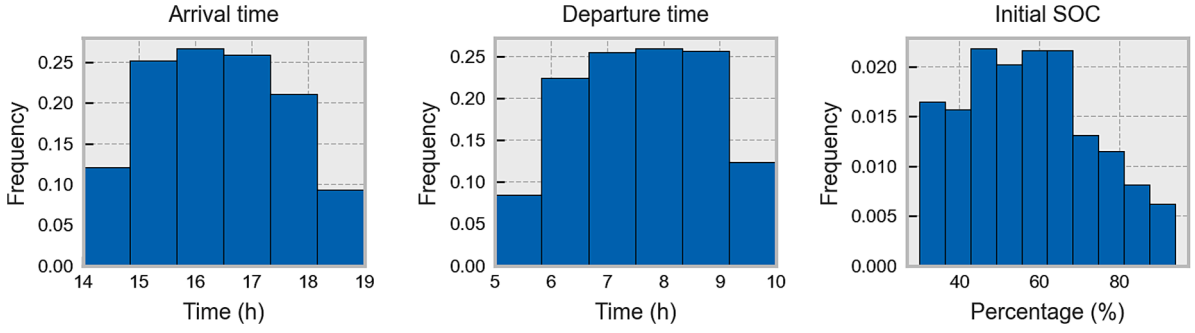

Fig. 9. Distributions of EV availability for smart home 1.

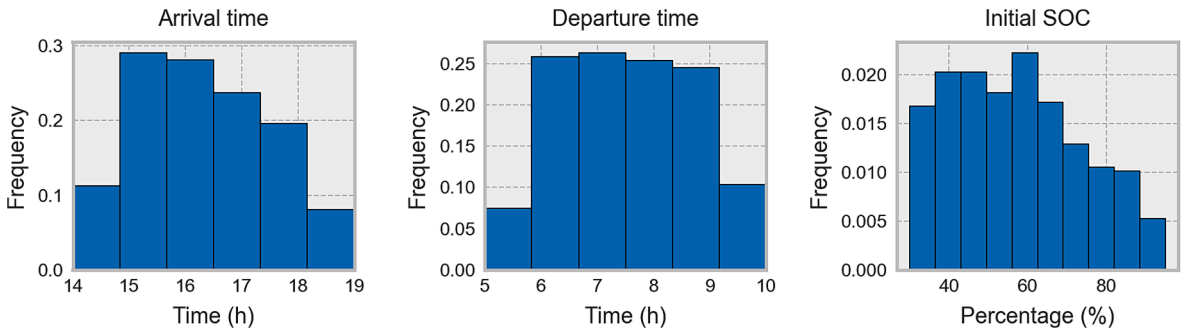

Fig. 10. Distributions of EV availability for smart home 2.

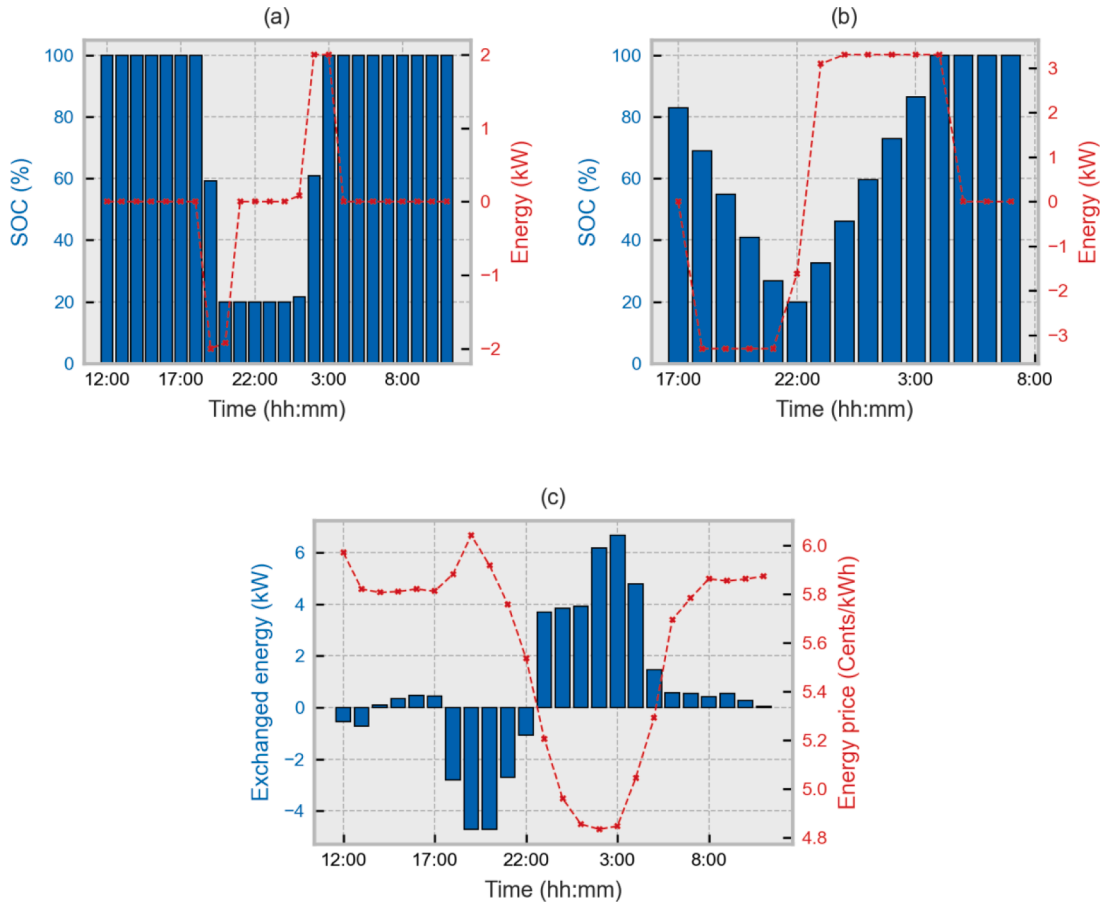

Fig. 11. Optimal results from the ideal theoretical method for smart home 1: (a) SOC and action of ESS, (b) SOC and action of EV, (c) energy exchange with grid.

The EV behavior, including arrival time, departure time, and initial SOC, is created via a scenario-based approach. Further details related to the scenario-based approach are available in [3]. To this end, a truncated Gaussian probability distribution function (PDF) is used to generate 788 scenarios of arrival time, departure time, and initial SOC with the data considered in Table 3, which are obtained from [1]. The distributions of all scenarios for EV availability in the two smart homes are shown in Figs. 9 and 10.

4.2. Ideal deterministic case

In this study, an ideal theoretical method is formulated as a deterministic MILP optimization model. In this method, all the required forecast data, including energy price, energy consumption, solar irradiation, and EV availability, are known precisely in advance, as shown in Figs. 5-10. The HEMS uses these data as inputs for the MILP optimization model. Accordingly, the MILP problem is solved using the Gurobi solver to determine the optimal actions of the ESS and EV to minimize the daily energy costs (as given in Section 2). The optimization results for a random day for smart home 1 are shown in Fig. 11.

As shown in Fig. 11, the charging mode of the ESS is activated to take advantage of the low-peak electricity tariffs. The ESS is discharged to supply energy when the power consumed by the home load reaches a peak. As EV often arrives at home at peak energy prices, EV charging is not immediately activated. Instead, the V2H of the EV is exploited to supply the load demand or to sell energy back to the grid. The EV is charged only during intervals of low electricity tariffs to minimize the charging costs. During energy scheduling, all constraints related to the ESS, EV, and grid are strictly satisfied, wherein the SOCs of the ESS and EV are always within their DOD and maximum capacities. To increase

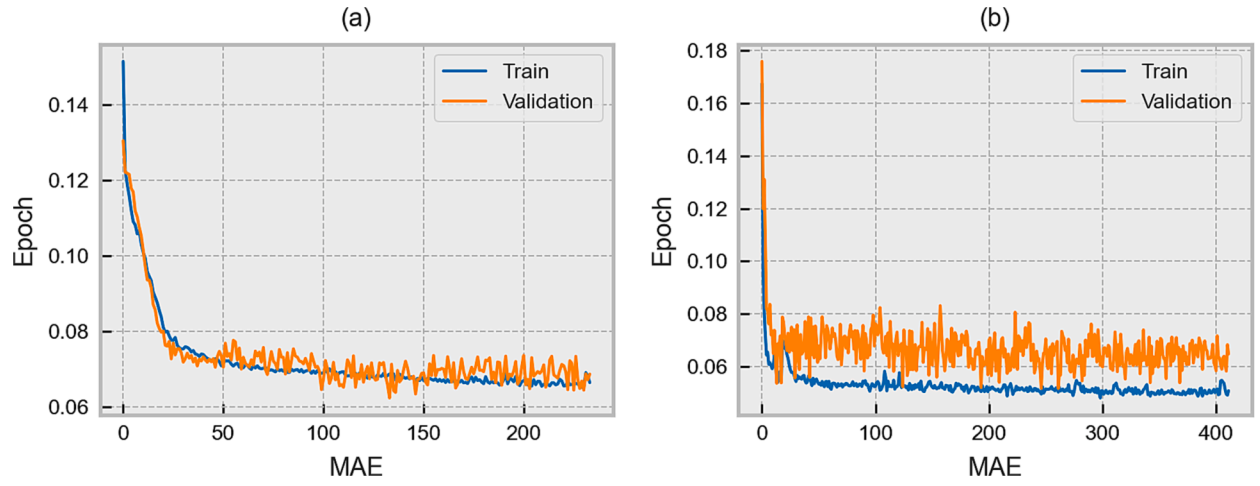

Fig. 12. Loss curves during DNNs training for smart home 1: (a) ESS and (b) EV.

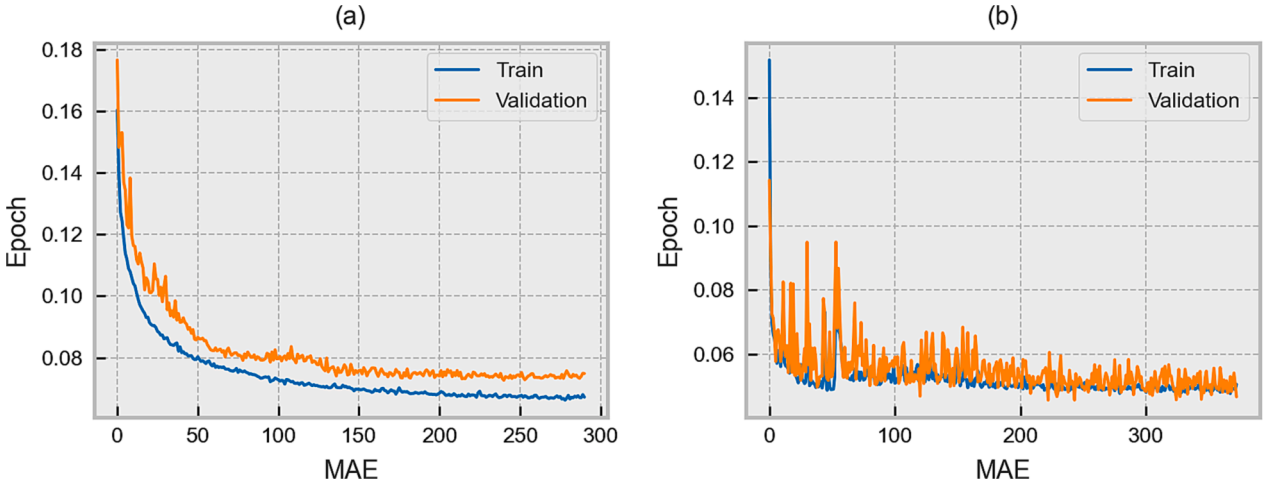

Fig. 13. Loss curves during DNNs training for smart home 2: (a) ESS and (b) EV.

consumer satisfaction, an EV is guaranteed to be fully charged at the time of its departure. Hence, the proposed HEMS model can be used to define intelligent charging and discharging decisions for ESS and EV. The energy cost via the MILP problem for this day is 71.85 Cents, which is the global optimal solution.

The results obtained from the ideal method confirm its effectiveness in the energy scheduling of ESS and EV. With perfect future information, the optimal actions obtained from the ideal theoretical method are considered perfect demonstrations and are used to train the DNNs, as discussed in Section 3.1. Moreover, the optimal energy costs achieved by the ideal method are essentially the minimum possible costs, which are the benchmark values for evaluating the performance of the proposed method.

4.3. Model training and convergence

In the supervised learning method, two DNN models are used to approximate the actions of the ESS and EV via supervised training. To accomplish this, each DNN is built with three hidden layers and one output layer, as suggested in [31]. The three hidden layers consist of 200, 100, and 50 neurons, respectively, and the popular rectified linear unit (ReLU) activation function is used for nonlinear transformation. The output layer produces a single output value and uses a linear activation function to solve the regression problem. The network is trained using the Adam optimizer with a loss function of MAE. Learning rate decay and early stopping mechanisms are applied to stabilize and expedite the training process, with an initial learning rate of 0.001.

From the historical data on energy price, energy consumption, solar irradiation, and EV availability in Section 4.1, the proposed supervised learning method considers data from 610 days to form 610 training scenarios with 14,640 training samples. Additionally, 152 holdout validation scenarios with 3,648 training samples are used to observe the training progress of the DNNs. Figs. 12 and 13 show the loss curves of the MAE for the ESS and EV training models for smart homes 1 and 2, respectively. The network training processes are completed between 250 and 400 epochs because of early stopping. The fact that there is a negligible difference between the training loss and validation loss suggests that there is little or no overfitting. This confirms that the DNNs are effectively trained. It is expected that well-trained DNNs can be used to predict near-optimal decisions for ESS and EV using real-time measurements.

In the MADDPG method, the actor and critic networks have two hidden layers that contain 256 and 128 neurons, respectively. The discount factor is set to 0.99, and the learning step size parameter is 0.001. The Adam optimizer is used to stabilize the direction of the gradient. In the MADDPG method, each episode corresponds to a scheduling scenario randomly selected from the 762 training scenarios. Because the training scenarios have different input data (e.g., energy price, energy consumption, solar irradiation, and EV availability), the episode rewards fluctuate considerably over a wide range. Thus, a gap metric is

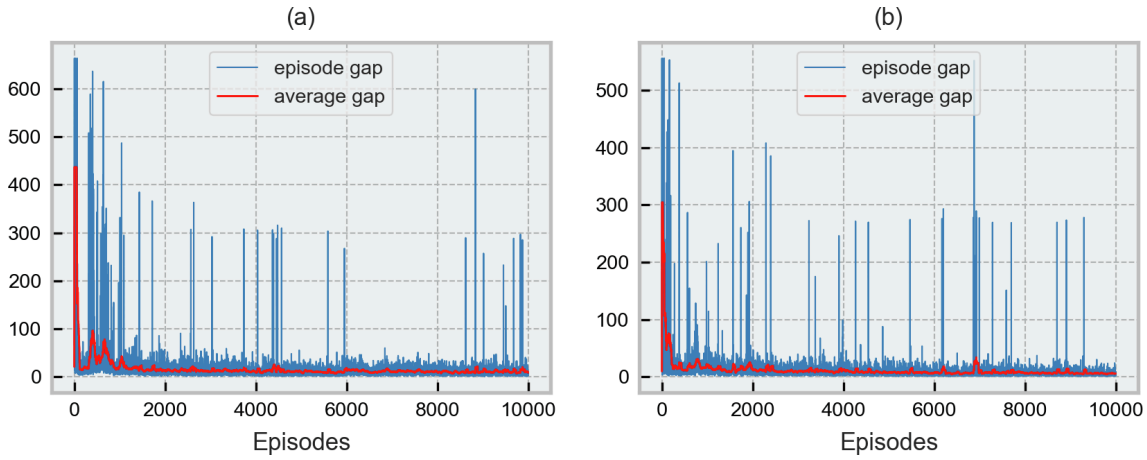

Fig. 14. Learning curves of the MADDPG algorithm during the training process: (a) Smart home 1 and (b) Smart home 2.

Table 4
Comparison of MAE and MAPE for all methods for 26 test scenarios.

Method	Smart home 1		Smart home 2	
	MAE	MAPE	MAE	MAPE
Base-load without DR	41.35	37.66 %	33.61	54.30 %
Forecasting method	15.59	15.21 %	11.54	22.76 %
MADDPG method	10.08	9.37 %	5.79	9.58 %
Supervised learning method	**2.50**	**2.08 %**	**1.38**	**2.18 %**

used to evaluate the convergence of the MADDPG method during the training process, which can be expressed as follows:

$$R_{gap} = R_n^{MADDPG} - J_n^{idea} \tag{45}$$

where R_n^{MADDPG} is the episode reward achieved by the MADDPG method for the n^{th} episode, and J_n^{idea} is the theoretical minimum energy cost achieved by the ideal method for the n^{th} episode.

Fig. 14 shows the learning curves of the MADDPG method over 10,000 episodes. The average values of the reward gap of the previous 50 episodes are also shown to observe the changing trend of the rewards more clearly. As shown in Fig. 14, the average gap is very high during the initial episodes. As the number of episodes increases, the average gap gradually decreases and is then stabilized. This indicates that the episode reward approaches the corresponding theoretical minimum value.

4.4. Numerical comparisons with other methods

In this section, the results obtained using supervised learning, MADDPG, and forecasting-based methods for HEMS are discussed and compared to evaluate the performance of the proposed solution method. To this end, the optimal result obtained by the ideal method is used as the reference point (or baseline). The ideal method assumes an impractical situation in which all forthcoming information is precisely known, leading to the attainment of the best possible solution (i.e., the lower bound of the objective cost). The 26 test scenarios associated with the 26 test days are used for performance comparison.

To simplify the comparison of energy cost, three metrics—mean absolute error (MAE), mean absolute percentage error (MAPE), and the performance gap of accumulated energy costs—are computed using the following formulas:

$$MAE = \frac{1}{N_{test}} \sum_{n=1}^{N_{test}} \left| J_n - J_n^{ideal} \right| \tag{46}$$

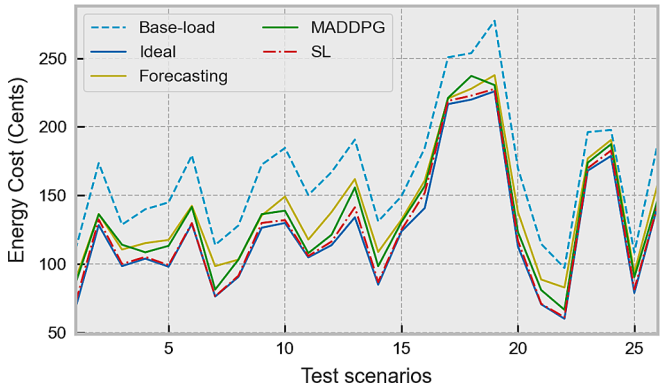

Fig. 15. Daily energy costs of all methods for smart home 1 for 26 test scenarios.

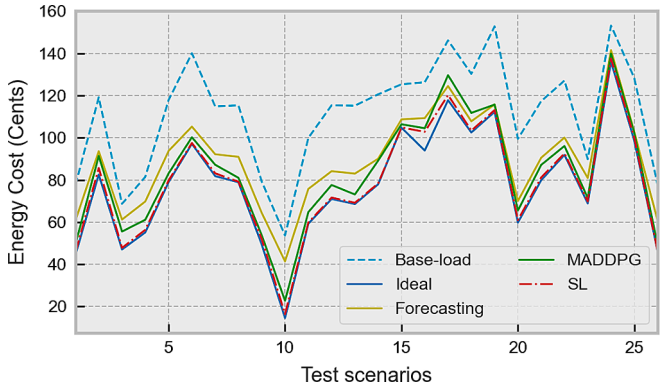

Fig. 16. Daily energy costs of all methods for smart home 2 for 26 test scenarios.

$$MAPE = \frac{100\%}{N_{test}} \sum_{n=1}^{N_{test}} \frac{\left| J_n - J_n^{ideal} \right|}{J_n^{ideal}} \tag{47}$$

$$J_{gap} = \frac{J_{acc} - J_{acc}^{ideal}}{J_{acc}^{ideal}} \tag{48}$$

where N_{test} is the total number of test scenarios; J_n^{ideal} and J_n are the energy costs achieved from the ideal method and the proposed method for a specific scenario, respectively; and J_{acc}^{ideal} and J_{acc} are the

Table 5
Comparison of accumulated energy costs and performance gaps for all methods in 26 test scenarios.

Method	Smart home 1		Smart home 2	
	Accumulated energy costs	Performance gap	Accumulated energy costs	Performance gap
Ideal method	3228.89	–	2018.99	–
Base-load without DR	4304.05	33.29 %	2892.97	43.28 %
Forecasting method	3634.38	12.56 %	2319.05	14.86 %
MADDPG method	3491.08	8.12 %	2169.79	7.49 %
Supervised learning method	**3293.90**	**2.01 %**	**2054.90**	**1.78 %**

accumulated energy costs achieved from the ideal method and the proposed method for all test scenarios, respectively.

Table 4 presents a comparison of the MAE and MAPE values for the different methods. For smart home 1, the MAE and MAPE achieved by the supervised learning method are approximately 2.50 and 2.08 %, respectively, which are much better than those of the MADDPG method (MAE of 10.08 and MAPE of 9.37 %) and forecasting-based methods (MAE of 15.59 and MAPE of 15.21 %). For smart home 2, the supervised learning method obtains an MAE of 1.38 and a MAPE of 2.18 %, which indicates that it also outperforms other methods in terms of MAE and MAPE indicators. There are significant differences between the MAE and MAPE values obtained by the supervised learning method and those obtained by the other methods for both smart homes, demonstrating the marked superiority of the proposed supervised learning method.

For a more intuitive assessment, Figs. 15 and 16 illustrate the daily energy costs achieved by the three methods related to smart homes 1 and 2, respectively, for the 26 test scenarios. The blue lines in Figs. 15 and 16 serve as reference points to represent the best possible solution achieved by the ideal method, which fully knows the required information in advance and eliminates forecasting inaccuracies. The cyan dotted lines illustrate the energy cost associated with the base load in the absence of DR. Figs. 15 and 16 indicate that the supervised learning strategy (red dotted line) is closer to the ideal results (blue line) than the MADDPG (green line) and forecasting-based methods (yellow line). Small gaps exist between the results obtained using the supervised learning method and the ideal method. By contrast, the forecasting-based method has the largest gap from the ideal results and exhibits the worst performance among the three methods. It can be inferred that the supervised learning method exhibits the best performance and strong generalizability for different scenarios.

Table 5 presents the performance gaps of the different methods for the 26 test scenarios. Referring to the results of smart home 1 in Table 5, the accumulated energy costs obtained from the supervised learning, MADDPG, and forecasting-based methods are 3293.90, 3491.08, and

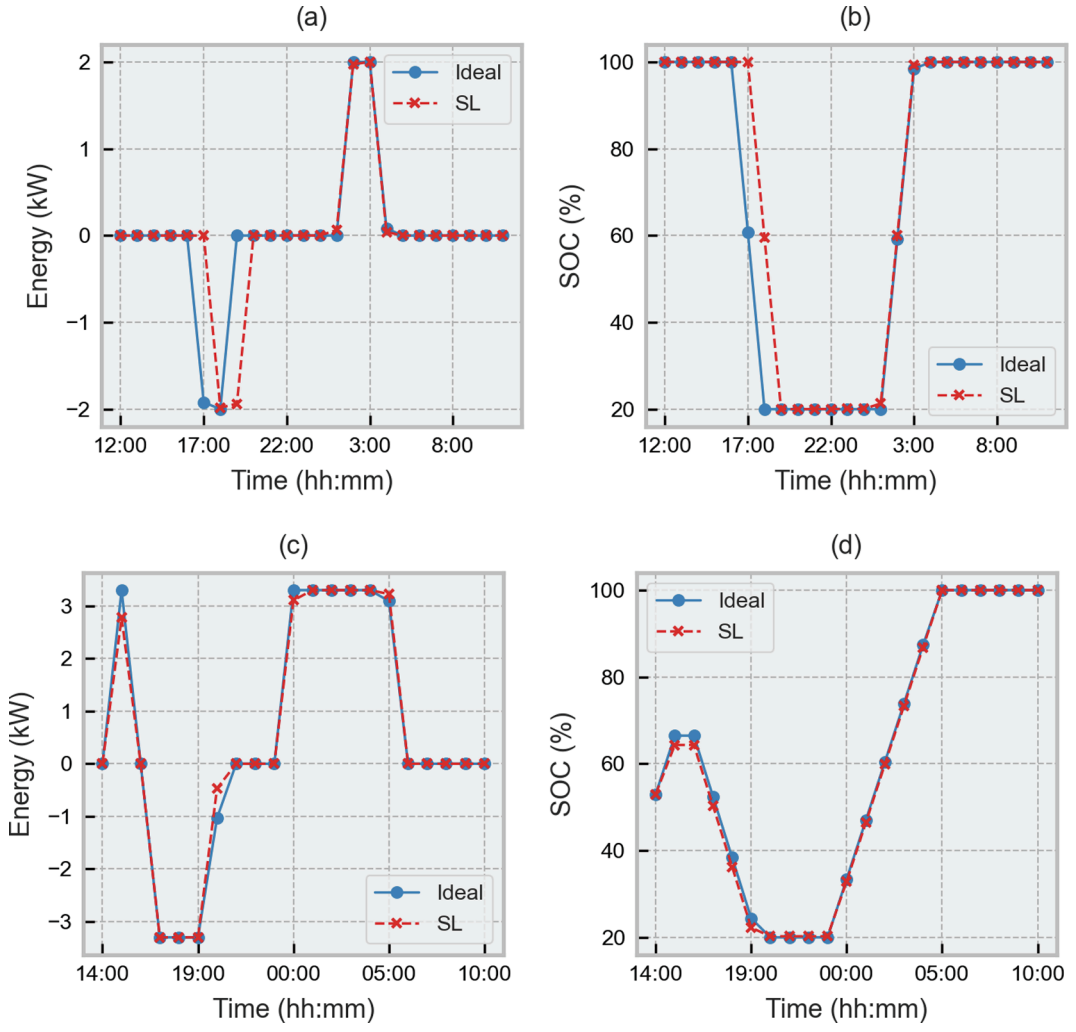

Fig. 17. Comparison of the ideal method and supervised learning method for smart home 1: (a) ESS action, (b) SOC of ESS, (c) EV action, (d) SOC of EV.

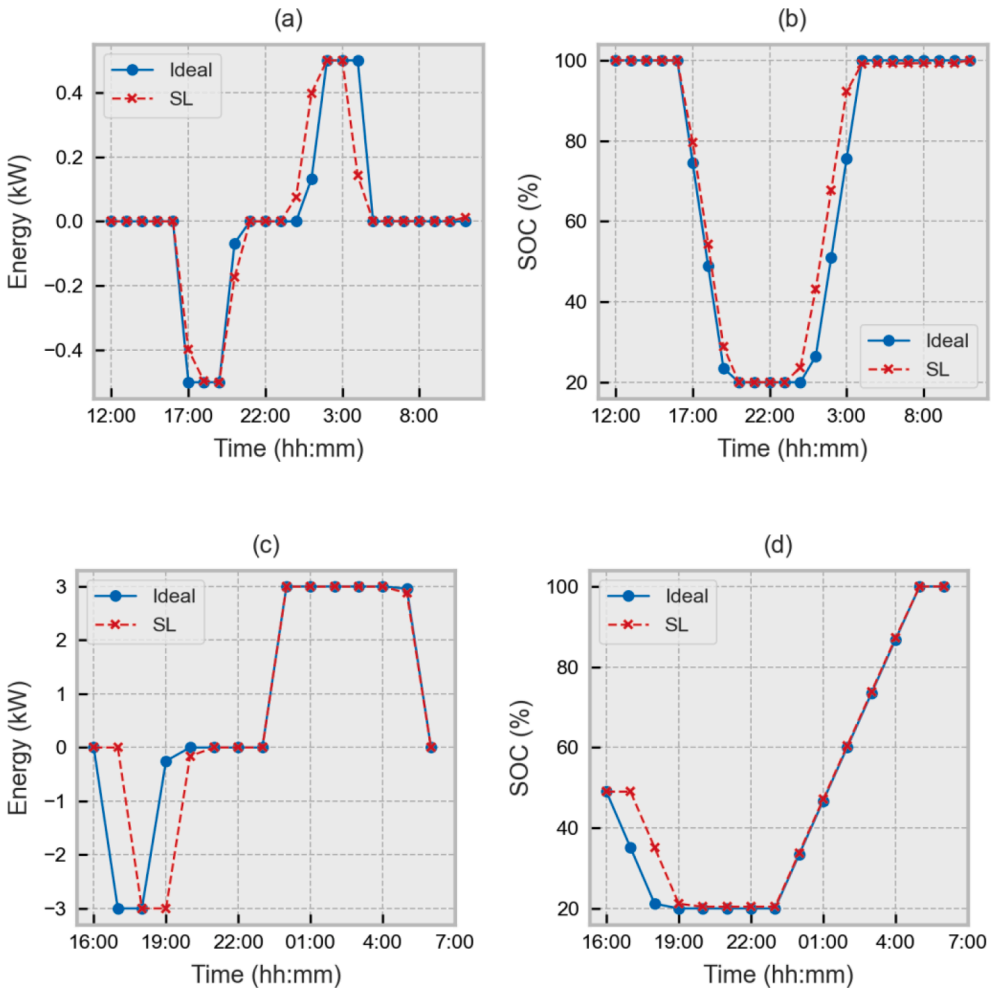

Fig. 18. Comparison of the ideal method and supervised learning (SL) method for smart home 2: (a) ESS action, (b) SOC of ESS, (c) EV action, (d) SOC of EV.

3634.38, corresponding to performance gaps of 2.01 %, 8.12 %, and 12.56 %, respectively. Similarly, the corresponding performance gaps related to these methods are 1.78 %, 7.49 %, and 14.86 % for smart home 2. The supervised learning method exhibits superior performance compared to the MADDPG and forecast-based methods by leveraging real-time measurements, resulting in a significant reduction in energy costs. The performance gaps of the supervised learning method for the two smart homes are relatively low (i.e., 2.01 % and 1.78 %, respectively), which are very close to those of the ideal method.

Of the three methods, the forecasting method is a day-ahead scheduling strategy that depends entirely on the forecast information. The poor performance of the forecasting method is due to inevitable prediction errors that significantly affect its optimal result when solving problem **P1**. The proposed supervised learning method takes advantage

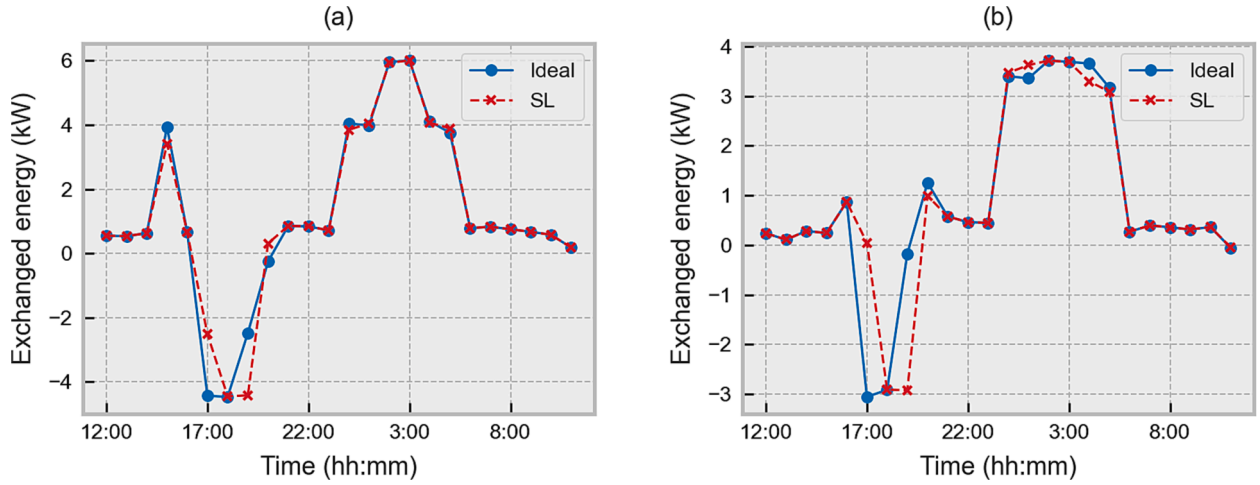

Fig. 19. Comparison of the ideal method and supervised learning method for energy exchange with grid: (a) Smart home 1 and (b) Smart home 2.

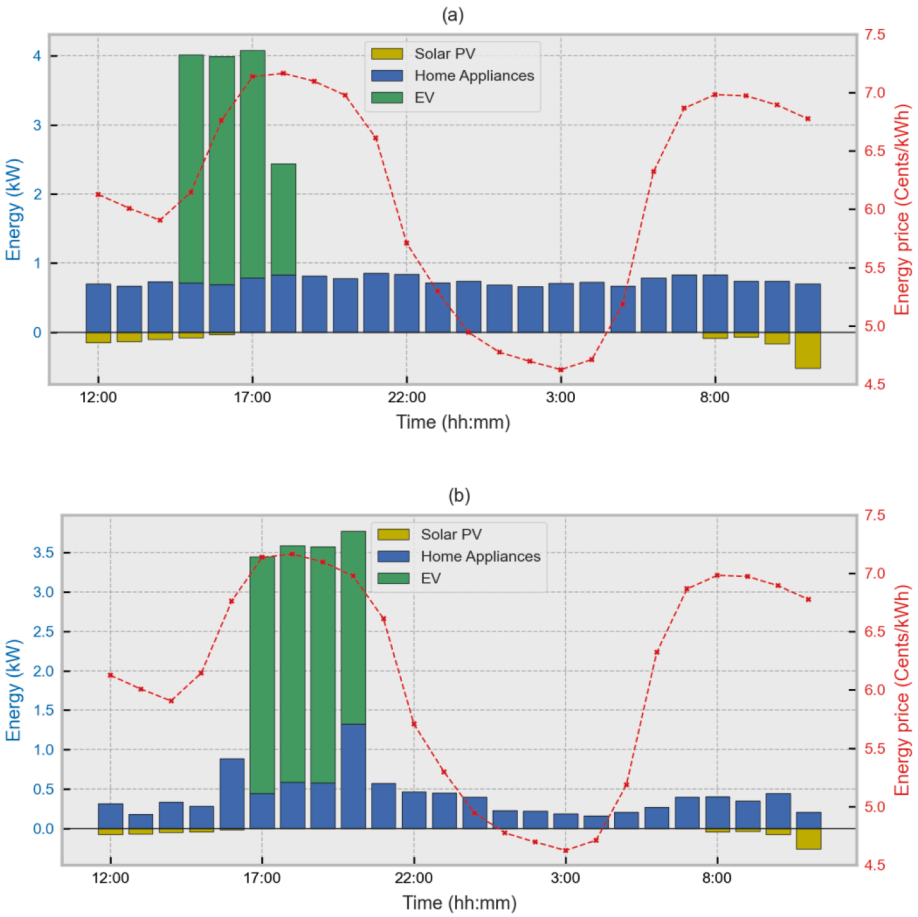

Fig. 20. Base-load pattern: (a) Home 1 and (b) Home 2.

of the generalization ability of DNNs to approximate the optimal actions from the MILP solver in regression problems. In other words, the idea of the proposed method is to learn from perfect demonstrations from a reliable expert instead of learning via trial and error (i.e., learning from scratch) as in the MADDPG method. From the comparative results, the proposed supervised learning method outperforms the other methods in terms of solution quality, robustness, and generalization. Hence, the proposed supervised learning method demonstrates superior performance in real-time energy scheduling.

4.5. Energy scheduling via supervised learning method for a specific scenario

A detailed analysis is conducted on a specific arbitrarily chosen scenario to further verify the decision-making capability of the proposed supervised-learning-based HEMS. The proposed supervised learning method assumes that no prior knowledge of the future is obtained, which is feasible in practice. To assess the performance of the proposed supervised learning method, an ideal method is used as the baseline. At each interval *t*, the actions of the ESS and EV obtained by the supervised learning method are compared with the optimal actions of the ideal method, as shown in Figs. 17(a), (c), 18(a), and (c). Accordingly, the SOCs of the ESS and EV at each interval *t* obtained using the proposed scheme are defined and presented in Figs. 17(b), (d), 18(b), and (d).

As shown in Figs. 17 and 18, the general trends of the actions provided by the two methods are relatively similar. The differences between the actions obtained by the two methods at certain intervals are not significant. Owing to the post-processing in the supervised learning method, all restrictions related to ESS and EV, such as permissible charging/discharging capacity and limited SOCs, are fulfilled during the energy scheduling cycle. Although EVs have different behaviors (i.e., arrival time, departure time, and initial SOC) in the considered scenario in the two smart homes, they are fully charged upon departure, which guarantees complete user satisfaction.

As shown in Fig. 19, the energy exchanges with the utility grid between the two methods are similar. The general strategy of the supervised learning method is to purchase energy and charge the ESS/EV at the off-peak of the electricity tariff (around early morning) and discharge the ESS/EV and sell energy at the on-peak of the electricity tariff (around evening), which is consistent with those of the ideal method. Moreover, the proposed method fully exploits the V2H capability of EVs, thereby contributing to optimal energy scheduling.

It should be emphasized that the supervised learning method makes decisions for real-time energy scheduling and requires only the current state (i.e., information for only one interval) as input to the trained DNNs. The cost of the ideal method is essentially a theoretical minimum and can only be obtained when the full information is available for the entire scheduling cycle (i.e., for all 24 intervals). For a quantitative comparison, the energy costs associated with the ideal and supervised learning methods for smart home 1 are 129.43 Cents and 129.95 Cents, respectively. The corresponding values for smart home 2 are 97.14 Cents and 97.53 Cents, respectively. The small variation between the two energy costs confirms the efficacy of the proposed supervised learning method for real-time energy scheduling for ESS and EV in HEMS.

Fig. 20 shows the load operation in both homes without the HEMS and DR strategy (i.e., base-load) for a given scenario. The base load model assumes that the ESS is not considered and that the EV is charged as soon as it arrives home until its battery is fully charged without considering the V2H mode. Because there is no HEMS, ESS, or DR strategy for the base load, the surplus energy generated by the solar PV

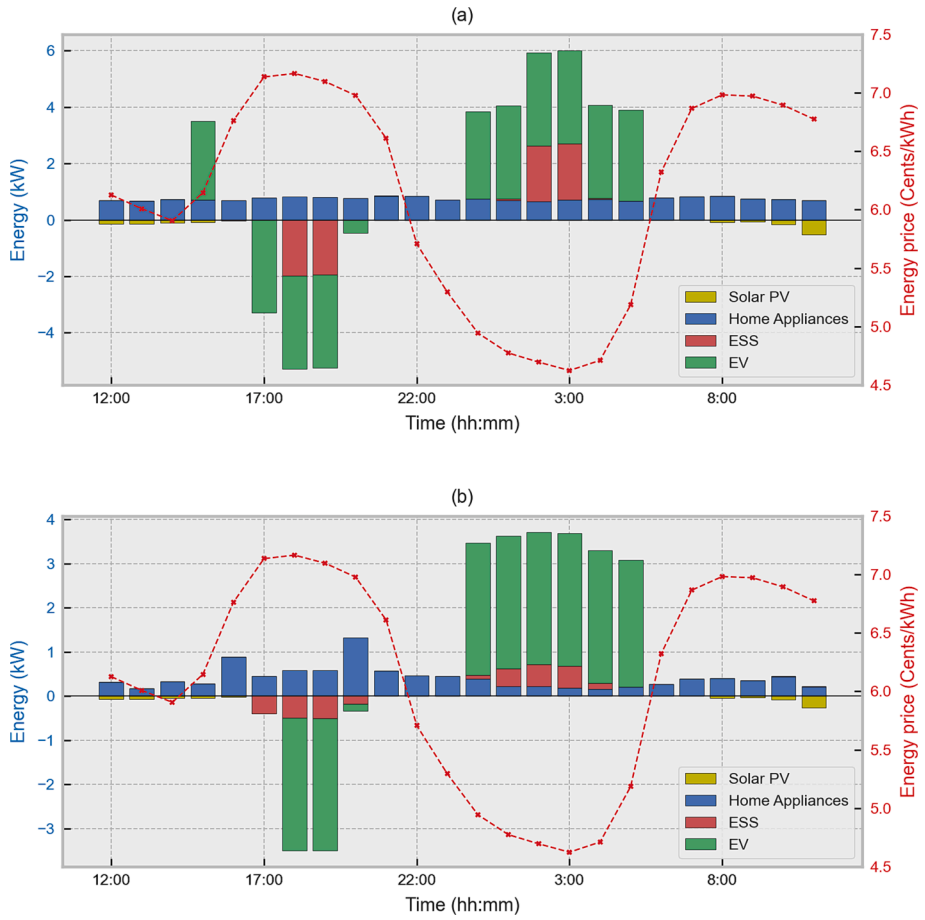

Fig. 21. Energy scheduling pattern using supervised learning method: (a) Smart home 1 and (b) Smart home 2.

panels cannot be stored and can lead to waste. The energy cost of the base case for the considered scenario is 179.28 Cents (for home 1) and 140.08 Cents (for home 2).

Fig. 21 shows the energy scheduling in both smart homes using the supervised-learning-based HEMS approach. The proposed method allows optimal decision-making to coordinate the energy flow control among the solar power, ESS, EV, and grid. Accordingly, the energy cost is reduced from 179.28 Cents to 129.95 Cents, which corresponds to a significant reduction of 27.52 % in the energy cost for smart home 1. The cost reduction for smart home 2 on this day is 30.38 %. The results show considerable reductions of 23.47 % and 28.96 % in the total cumulative energy costs for all test scenarios in both smart homes. It is important to note that this cost reduction is accomplished without compromising user comfort because home appliances are used arbitrarily according to user preferences.

Moreover, the execution time of the proposed method takes an average of 28 ms to perform decision-making for one single time-step scheduling, which can fulfill the timing requirements of the real-time execution on timescales of milliseconds [16,25,28,52]. Thus, the supervised-learning-based HEMS successfully provides an efficient real-

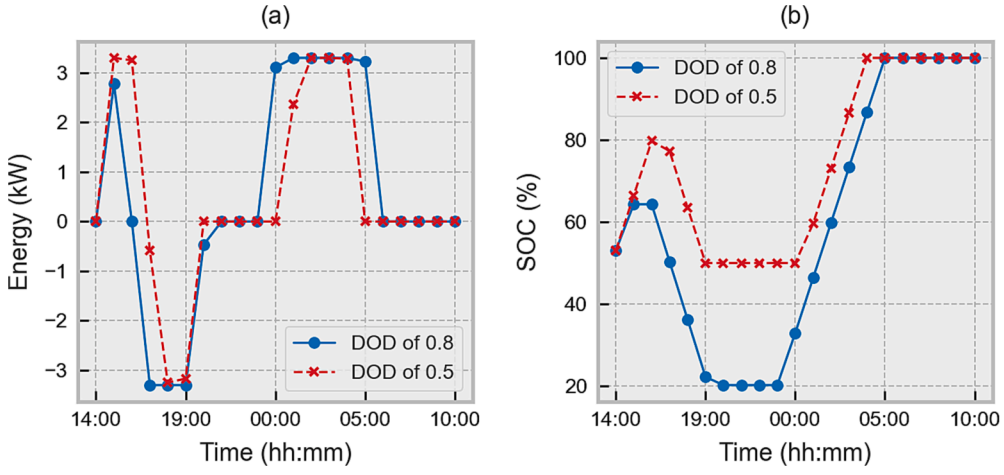

Fig. 22. Comparison of two DOD settings of EV for smart home 1: (a) EV action, (b) SOC of EV.

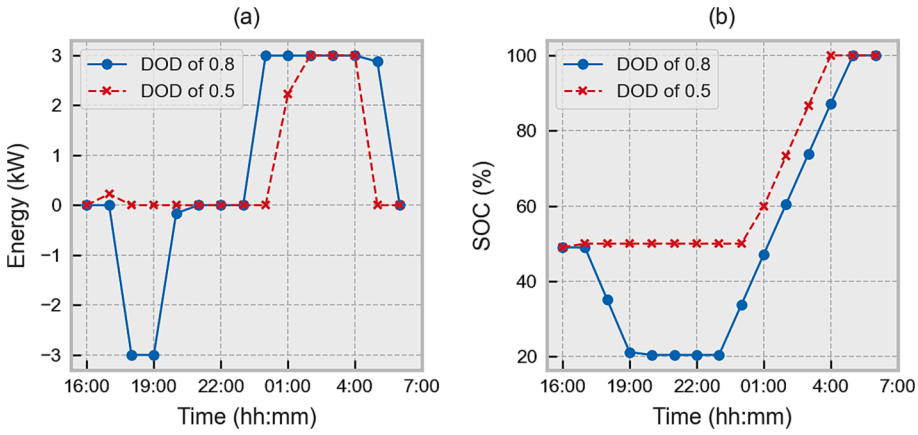

Fig. 23. Comparison of two DOD settings of EV for smart home 2: (a) EV action, (b) SOC of EV.

time energy scheduling method, supporting residential end users to participate in the DR program.

4.6. Analysis of the unexpected scenarios of EV scheduling

Based on the DR strategy, the user can decide whether EV participates in the energy scheduling scheme when EV arrives home. In case the EV does not participate in the energy scheduling scheme, it is scheduled to charge immediately until its battery is full, similar to the base-load without DR. In some unexpected scenarios, users may need EVs for a spontaneous period, and the minimum capacity of the EV may not satisfy user satisfaction. To this end, extensive research is performed to mitigate the impact of such unexpected scenarios. In the extensive research, the HEMS mathematical model is modified, wherein EV is charged immediately to reach 50 % capacity after it arrives home and connects to the HEMS. EV battery is maintained at a minimum of 50 % (i.e., DOD of 0.5) instead of 20 % (i.e., DOD of 0.8) during the energy scheduling. DNNs are trained to learn and predict EV actions for extensive scenarios.

Figs. 22 and 23 present comparisons of actions and SOC of EVs for two DOD settings for smart homes 1 and 2, respectively. As shown in these figures, EV actions are significantly different when changing the DOD from 0.8 to 0.5. It can be seen that the V2H possibility of the DOD of 0.5 is less exploited than that of the DOD of 0.8 due to the minimum capacity limit of EVs. As a result, the energy costs of the two smart homes are 142.51 Cents and 109.65 Cents, respectively, corresponding to increases of 9.67 % and 12.43 % compared to the DOD setting of 0.8. Total cumulative energy costs also increase by 6.78 % and 10.25 % for all test scenarios in smart homes 1 and 2, respectively. However, EVs always store a certain amount of energy to be used whenever needed, ensuring a certain level of user satisfaction in unexpected scenarios. The trade-off between energy costs and user satisfaction is inevitable in the energy scheduling process. Users can adjust the settings of the HEMS model based on their preferences. Accordingly, the proposed method can be easily updated and improved to effectively adapt to user requirements.

5. Conclusion

In this study, a supervised-learning-based HEMS framework was proposed as a real-time energy scheduling strategy to increase energy efficiency and reduce energy costs in smart homes. The developed HEMS model was defined by the penetration of solar PV, ESS, and EV, wherein the HEMS plays the role of an active prosumer in the electricity market. With the application of supervised learning, the strategy can learn the optimal actions of the ESS and EV (playing as expert demonstrations) from MILP solvers (playing as an expert) based on historical input data, which generates two mappings via supervised training using two DNNs. The performance of the supervised-learning-based HEMS framework was verified using two different smart homes with real-world data. From the scheduled energy patterns, the supervised-learning-based HEMS framework makes optimal decisions to effectively control the charging/discharging power of the ESS and EV based on real-time information. For the total accumulated energy costs, the performance gaps of the supervised learning method were 2.01 % and 1.78 %, respectively, which are very close to the ideal results for the two smart homes. The proposed method can also contribute to the full exploitation of the RES generation and storage capacities of ESS and EVs. Moreover, the comparative results indicate that the proposed method outperforms two other machine learning-based methods, MADDPG and forecasting-based methods, in terms of solution quality, robustness, and generalizability with a set of test scenarios. The supervised-learning-based approach can be adapted for future energy scheduling plans, leading to more personalized and efficient energy management. It can be concluded that the proposed supervised-learning-based HEMS is a cost-effective and uncertainty-aware solution for energy scheduling at the residential level.

Future works aim to extend and apply the proposed method in four directions to address the limitations of the present study. In the first future direction, the generalized performance of the proposed method needs to be enhanced in order to improve its robustness and adaptability to unexpected situations such as solar PV panel and ESS failures, grid outages, use of EVs during spontaneous intervals, etc. Secondly, the impact of the time delay of the communication networks and the power electronic inverters should be considered during the real-time energy scheduling process. The third future direction involves integrating the proposed method in the power hardware-in-the-loop simulation to verify its effectiveness and practicality. Finally, the proposed method is recommended for application to extended energy management problems in large-scale systems, such as microgrids, multi-energy systems, and EV charging stations in further studies.

CRediT authorship contribution statement

Truong Hoang Bao Huy: Conceptualization, Methodology, Software, Visualization, Writing – original draft. **Huy Truong Dinh:** Validation, Formal analysis, Writing – review & editing. **Dieu Ngoc Vo:** Supervision, Writing – review & editing. **Daehee Kim:** Supervision, Investigation, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1A4A2001810, 2020R1F1A1048664, 2022H1D8A3038040), by Institute for Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2022-0-01197, Convergence security core talent training business (SoonChunHyang University)), and by the MSIT (Ministry of Science, ICT), Korea, under the National Program for Excellence in SW, supervised by the IITP (Institute of Information & communications Technology Planning & Evaluation) in 2021 (2021-0-01399). This work is a result of a study on the "Leaders in INdustry-university Cooperation 3.0" Project, supported by the Ministry of Education and National Research Foundation of Korea (No. 1345356224) and this work was supported by the Soonchunhyang Research Fund.

References

[1] Shafie-Khah M, Siano P. A Stochastic Home Energy Management System Considering Satisfaction Cost and Response Fatigue. IEEE Trans Ind Inf 2018;14:629–38. https://doi.org/10.1109/TII.2017.2728803.
[2] Batchu R, Pindoriya NM. Residential Demand Response Algorithms: State-of-the-Art, Key Issues and Challenges. In: Pillai P, Hu YF, Otung I, Giambene G, editors. Wireless and Satellite Systems. Cham: Springer International Publishing; 2015. p. 18–32. https://doi.org/10.1007/978-3-319-25479-1_2.
[3] Huy THB, Dinh HT, Kim D. Multi-objective framework for a home energy management system with the integration of solar energy and an electric vehicle using an augmented ε-constraint method and lexicographic optimization. Sustain Cities Soc 2023;88:104289. https://doi.org/10.1016/j.scs.2022.104289.
[4] Dorahaki S, Rashidinejad M, Fatemi Ardestani SF, Abdollahi A, Salehizadeh MR. A home energy management model considering energy storage and smart flexible appliances: A modified time-driven prospect theory approach. J Storage Mater 2022;48:104049. https://doi.org/10.1016/j.est.2022.104049.
[5] Tostado-Véliz M, Gurung S, Jurado F. Efficient of many-objective Home Energy Management systems. Int J Electr Power Energy Syst 2022;136:107666. https://doi.org/10.1016/j.ijepes.2021.107666.
[6] Zeynali S, Rostami N, Ahmadian A, Elkamel A. Two-stage stochastic home energy management strategy considering electric vehicle and battery energy storage system: An ANN-based scenario generation methodology. Sustainable Energy Technol Assess 2020;39:100722. https://doi.org/10.1016/j.seta.2020.100722.
[7] Bahramara S. Robust Optimization of the Flexibility-constrained Energy Management Problem for a Smart Home with Rooftop Photovoltaic and an Energy Storage. J Storage Mater 2021;36:102358. https://doi.org/10.1016/j.est.2021.102358.
[8] Su Y, Zhou Y, Tan M. An interval optimization strategy of household multi-energy system considering tolerance degree and integrated demand response. Appl Energy 2020;260:114144. https://doi.org/10.1016/j.apenergy.2019.114144.
[9] Shokri Gazafroudi A, Soares J, Fotouhi Ghazvini MA, Pinto T, Vale Z, Corchado JM. Stochastic interval-based optimal offering model for residential energy management systems by household owners. Int J Electr Power Energy Syst 2019;105:201–19. https://doi.org/10.1016/j.ijepes.2018.08.019.
[10] Li S, Yang J, Song W, Chen A. A Real-Time Electricity Scheduling for Residential Home Energy Management. IEEE Internet Things J 2019;6:2602–11. https://doi.org/10.1109/JIOT.2018.2872463.
[11] Killian M, Zauner M, Kozek M. Comprehensive smart home energy management system using mixed-integer quadratic-programming. Appl Energy 2018;222:662–72. https://doi.org/10.1016/j.apenergy.2018.03.179.
[12] Jin X, Baker K, Christensen D, Isley S. Foresee: A user-centric home energy management system for energy efficiency and demand response. Appl Energy 2017;205:1583–95. https://doi.org/10.1016/j.apenergy.2017.08.166.
[13] Wang J, Liu J, Li C, Zhou Y, Wu J. Optimal scheduling of gas and electricity consumption in a smart home with a hybrid gas boiler and electric heating system. Energy 2020;204:117951. https://doi.org/10.1016/j.energy.2020.117951.
[14] Huang Y, Wang L, Guo W, Kang Q, Wu Q. Chance Constrained Optimization in a Home Energy Management System. IEEE Trans Smart Grid 2018;9:252–60. https://doi.org/10.1109/TSG.2016.2550031.
[15] Zhao L, Yang T, Li W, Zomaya AY. Deep reinforcement learning-based joint load scheduling for household multi-energy system. Appl Energy 2022;324:119346. https://doi.org/10.1016/j.apenergy.2022.119346.
[16] Lu R, Hong SH, Yu M. Demand Response for Home Energy Management Using Reinforcement Learning and Artificial Neural Network. IEEE Trans Smart Grid 2019;10:6629–39. https://doi.org/10.1109/TSG.2019.2909266.
[17] Xu X, Jia Y, Xu Y, Xu Z, Chai S, Lai CS. A Multi-Agent Reinforcement Learning-Based Data-Driven Method for Home Energy Management. IEEE Trans Smart Grid 2020;11:3201–11. https://doi.org/10.1109/TSG.2020.2971427.
[18] Alfaverh F, Denaï M, Sun Y. Demand Response Strategy Based on Reinforcement Learning and Fuzzy Reasoning for Home Energy Management. IEEE Access 2020;8:39310–21. https://doi.org/10.1109/ACCESS.2020.2974286.
[19] Ahrarinouri M, Rastegar M, Seifi AR. Multiagent Reinforcement Learning for Energy Management in Residential Buildings. IEEE Trans Ind Inf 2021;17:659–66. https://doi.org/10.1109/TII.2020.2977104.
[20] Yu L, Xie W, Xie D, Zou Y, Zhang D, Sun Z, et al. Deep Reinforcement Learning for Smart Home Energy Management. IEEE Internet Things J 2020;7:2751–62. https://doi.org/10.1109/JIOT.2019.2957289.
[21] Liu Y, Zhang D, Gooi HB. Optimization strategy based on deep reinforcement learning for home energy management. CSEE J Power Energy Syst 2020;6(572–82). https://doi.org/10.17775/CSEEJPES.2019.02890.
[22] Ding H, Xu Y, Chew Si Hao B, Li Q, Lentzakis A. A safe reinforcement learning approach for multi-energy management of smart home. Electr Pow Syst Res 2022;210:108120. https://doi.org/10.1016/j.epsr.2022.108120.
[23] Chu Y, Wei Z, Sun G, Zang H, Chen S, Zhou Y. Optimal home energy management strategy: A reinforcement learning method with actor-critic using Kronecker-factored trust region. Electr Pow Syst Res 2022;212:108617. https://doi.org/10.1016/j.epsr.2022.108617.
[24] Langer L, Volling T. A reinforcement learning approach to home energy management for modulating heat pumps and photovoltaic systems. Appl Energy 2022;327:120020. https://doi.org/10.1016/j.apenergy.2022.120020.
[25] Ye Y, Qiu D, Wu X, Strbac G, Ward J. Model-Free Real-Time Autonomous Control for a Residential Multi-Energy System Using Deep Reinforcement Learning. IEEE Trans Smart Grid 2020;11:3068–82. https://doi.org/10.1109/TSG.2020.2976771.
[26] Mocanu E, Mocanu DC, Nguyen PH, Liotta A, Webber ME, Gibescu M, et al. On-Line Building Energy Optimization Using Deep Reinforcement Learning. IEEE Trans Smart Grid 2019;10:3698–708. https://doi.org/10.1109/TSG.2018.2834219.
[27] Hou H, Ge X, Chen Y, Tang J, Hou T, Fang R. Model-free dynamic management strategy for low-carbon home energy based on deep reinforcement learning accommodating stochastic environments. Energ Buildings 2023;278:112594. https://doi.org/10.1016/j.enbuild.2022.112594.
[28] Li H, Wan Z, He H. Real-Time Residential Demand Response. IEEE Trans Smart Grid 2020;11:4144–54. https://doi.org/10.1109/TSG.2020.2978061.
[29] Dinh HT, Kim D. MILP-Based Imitation Learning for HVAC Control. IEEE Internet Things J 2022;9:6107–20. https://doi.org/10.1109/JIOT.2021.3111454.
[30] Kim Y-J. A Supervised-Learning-Based Strategy for Optimal Demand Response of an HVAC System in a Multi-Zone Office Building. IEEE Trans Smart Grid 2020;11:4212–26. https://doi.org/10.1109/TSG.2020.2986539.
[31] Gao S, Xiang C, Yu M, Tan KT, Lee TH. Online Optimal Power Scheduling of a Microgrid via Imitation Learning. IEEE Trans Smart Grid 2022;13:861–76. https://doi.org/10.1109/TSG.2021.3122570.
[32] Dinh HT, Lee K, Kim D. Supervised-learning-based hour-ahead demand response for a behavior-based home energy management system approximating MILP optimization. Appl Energy 2022;321:119382. https://doi.org/10.1016/j.apenergy.2022.119382.
[33] Paterakis NG, Erdinç O, Bakirtzis AG, Catalão JPS. Optimal Household Appliances Scheduling Under Day-Ahead Pricing and Load-Shaping Demand Response Strategies. IEEE Trans Ind Inf 2015;11:1509–19. https://doi.org/10.1109/TII.2015.2438534.
[34] Tostado-Véliz M, León-Japa RS, Jurado F. Optimal electrification of off-grid smart homes considering flexible demand and vehicle-to-home capabilities. Appl Energy 2021;298:117184. https://doi.org/10.1016/j.apenergy.2021.117184.
[35] Ustun TS, Hussain SMS. Standardized Communication Model for Home Energy Management System. IEEE Access 2020;8:180067–75. https://doi.org/10.1109/ACCESS.2020.3028108.
[36] Collotta M, Pau G. An Innovative Approach for Forecasting of Energy Requirements to Improve a Smart Home Management System Based on BLE. IEEE Trans Green Commun Netw 2017;1:112–20. https://doi.org/10.1109/TGCN.2017.2671407.
[37] Tostado-Véliz M, Arévalo P, Kamel S, Zawbaa HM, Jurado F. Home energy management system considering effective demand response strategies and uncertainties. Energy Rep 2022;8:5256–71. https://doi.org/10.1016/j.egyr.2022.04.006.
[38] Tostado-Véliz M, Icaza-Alvarez D, Jurado F. A novel methodology for optimal sizing photovoltaic-battery systems in smart homes considering grid outages and demand response. Renew Energy 2021;170:884–96. https://doi.org/10.1016/j.renene.2021.02.006.
[39] Alsaidan I, Khodaei A, Gao W. A Comprehensive Battery Energy Storage Optimal Sizing Model for Microgrid Applications. IEEE Trans Power Syst 2018;33:3968–80. https://doi.org/10.1109/TPWRS.2017.2769639.
[40] Arévalo P, Tostado-Véliz M, Jurado F. A novel methodology for comprehensive planning of battery storage systems. J Storage Mater 2021;37:102456. https://doi.org/10.1016/j.est.2021.102456.
[41] Bui V-H, Hussain A, Kim H-M. Double Deep Q -Learning-Based Distributed Operation of Battery Energy Storage System Considering Uncertainties. IEEE Trans Smart Grid 2020;11:457–69. https://doi.org/10.1109/TSG.2019.2924025.
[42] Shuai H, Fang J, Ai X, Tang Y, Wen J, He H. Stochastic Optimization of Economic Dispatch for Microgrid Based on Approximate Dynamic Programming. IEEE Trans Smart Grid 2019;10:2440–52. https://doi.org/10.1109/TSG.2018.2798039.
[43] Foruzan E, Soh L-K, Asgarpoor S. Reinforcement Learning Approach for Optimal Distributed Energy Management in a Microgrid. IEEE Trans Power Syst 2018;33:5749–58. https://doi.org/10.1109/TPWRS.2018.2823641.

[44] Osa T, Pajarinen J, Neumann G, Bagnell JA, Abbeel P, Peters J. An Algorithmic Perspective on Imitation Learning. FNT in Robotics 2018;7:1–179. https://doi.org/10.1561/2300000053.
[45] Lowe R, Wu Y, Tamar A, Harb J, Abbeel P, Mordatch I. Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments; 2020. 10.48550/arXiv.1706.02275.
[46] Lillicrap TP, Hunt JJ, Pritzel A, Heess N, Erez T, Tassa Y, et al. Continuous control with deep reinforcement learning; 2019. 10.48550/arXiv.1509.02971.
[47] Farag W. Multi-Agent Reinforcement Learning using the Deep Distributed Distributional Deterministic Policy Gradients Algorithm. In: 2020 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT); 2020. p. 1–6. 10.1109/3ICT51146.2020.9311945.
[48] Energy consumption data in London households n.d. https://data.london.gov.uk/dataset/ [accessed April 4, 2023].
[49] JRC Photovoltaic Geographical Information System (PVGIS) - European Commission n.d. https://re.jrc.ec.europa.eu/pvg_tools/en/tools.html [accessed June 7, 2022].
[50] Markets and prices | ESIOS electricity · data · transparency n.d. https://www.esios.ree.es/en/market-and-prices?date=03-04-2023 [accessed April 4, 2023].
[51] Jhana N. Hourly energy demand generation and weather n.d. https://www.kaggle.com/datasets/nicholasjhana/energy-consumption-generation-prices-and-weather [accessed April 4, 2023].
[52] Shuai H, He H. Online Scheduling of a Residential Microgrid via Monte-Carlo Tree Search and a Learned Model. IEEE Trans Smart Grid 2021;12:1073–87. https://doi.org/10.1109/TSG.2020.3035127.