JJEE
Volume 11 | Number 2 | June 2025 | Pages 264-288

Jordan Journal of Electrical Engineering

ISSN (print): 2409-9600, ISSN (online): 2409-9619

Homepage: jjee.ttu.edu.jo

Data-Driven Reinforcement Learning for Energy Management with Peer-to-Peer Trading and Demand-Side Management

Amany El-Zonkoly*

Department of Electrical and Control Engineering, College of Engineering and Technology, Arab Academy for Science, Technology and Maritime Transport, Alexandria, Egypt

Emails: amanyelz@yahoo.com, amanyelz@aast.edu

Received: Jan 15, 2025 *Revised: Feb 28, 2025* *Accepted: Mar 09, 2025* *Available online: Apr 6, 2025*

Abstract – This paper presents demand side management (DSM) for integrated energy systems of dairy farms - in order to lower the energy consumption - utilizing a variety of energy devices and load types, in addition to Peer-to-peer (P2P) energy trading. The decision tree regression model is used to anticipate the day-ahead PV power generation, utility grid energy pricing, and various loads consumption based on real-world data in order to implement DSM and P2P energy trading (P2P-ET). To handle the growing uncertainties - brought on by bidding actions, transaction volumes, and forecasted data on load profiles, the generation of renewable resources, and energy prices - a modified multi-agent reinforcement learning (MARL) is used for decision-making. To address such a situation, the DSM and P2P-ET problem is formulated as a finite Markov decision process. The mixed uncertainty is included as additional stochastically represented states and action scenarios in the modified MARL. It is found that incorporating additional stochastic states and action scenarios significantly enhances the RL agent's ability to adapt to diverse and dynamic environments, particularly when faced with unexpected variations in PV generation and energy price. These extra states and actions allow the agent to learn more nuanced strategies and respond effectively to probabilistic circumstances. The simulation result unveil that by using the proposed MARL algorithm to optimize the P2P-ET and DSM strategies, the average load can be reduced by 20.73%. Moreover,, the optimal P2P-ET results in buying 82.1% of the energy needed from the P2P community, while the remaining 17.9% is bought from the utility grid. Finally, by applying both the optimal DSM and P2P-ET, the average daily cost of energy can be reduced by 23.57%.

***Keywords*—** Peer-to-peer energy trading; Demand side management; Integrated energy system; Modified multi-agent reinforcement learning.

Nomenclature

P_t^{H2}	Power Content of the Electrolyzer's Produced Hydrogen Gas [kW]	η_{EL}	Electrolyser Efficiency
$P_t^{PV\text{-}EL}$	PV Electrical Power Supplied to Electrolyzer [kW]	η_c, η_e, η_h	Cold, Electric, and Heat Efficiencies of CCHP, respectively
$P_t^{c\text{-}CCHP}$, $P_t^{e\text{-}CCHP}$, $P_t^{h\text{-}CCHP}$	Cold, Electric, and Heat Power Output of CCHP, respectively [kW]	$\eta_{e\text{-}HCH}$	Electricity to Cold Efficiency of HCH
$P_t^{e\text{-}HCH}$, $P_t^{c\text{-}HCH}$	Input Electric Power, and Output Cold Power of HCH, respectively [kW]	P_{min}^{EL}, P_{max}^{EL}	Minimum and Maximum Electric Power Input to the Electrolyzer, respectively [kW]
p_t^{bn}	Energy Buying Price of Buyer N [$/kWh]	$P_{max}^{H2-CCHP}$, $P_{max}^{H2-CCHP}$	Minimum and Maximum Hydrogen Power Input to the CCHP, respectively [kW]
p_t^{sn}	Energy Selling Price of Seller N [$/kWh]	P_{min}^{h-HCH}, P_{max}^{h-HCH}	Minimum And Maximum Heat Power Input to the HCH, respectively [kW]

* Corresponding author

DOI: https://doi.org/10.5455/jjee.204-1736974366

MCP_t	Market Clearing Price [$/kWh]	$P_{min}^{e-HCH}, P_{max}^{e-HCH}$	Minimum and Maximum Electric Power Input to the HCH, respectively [kW]
c_t^e	Net Energy Trading Cost [$]	P_{max}^{e-CCHP}	Maximum Electric Power Output from the CCHP [kW]
c_t^D	Total Dissatisfaction Cost [$]	P_{max}^{h-CCHP}	Maximum Heating Power Output from the CCHP [kW]
$c_t^{D,n}$	Dissatisfaction Cost of Nth Load [$]	P_{max}^{c-CCHP}	Maximum Cold Power Output from the CCHP [kW]
p_t^S, p_t^B	Sell And Buy Prices of Energy of the UG, respectively [$/kWh]	P_{max}^{c-HCH}	Maximum Cold Power Output from The HCH [kW]
p_t^D	Dissatisfaction Price [$/kWh]	P_{max}^{e-load}	Maximum Electric Load [kW]
p_t^{P2P}	P2P Energy Trading Price [$/kWh]	P_{max}^{h-load}	Maximum Heating Load [kW]
$E_t^{Set,n}$	Rated Energy of Nth Load [kWh]	P_{max}^{c-load}	Maximum Cooling Load [kW]
E_t^n	Selected Energy of Nth Load [kWh]	P_{max}^{PV}	Maximum PV Generated Power [kW]
E_t^{P2P}	Energy Traded with P2P Community [kWh]	P_{max}^{P2P}	Maximum P2P Traded Power [kW]
E_t^{UG}	Energy Traded with UG [kWh]	P_{max}^{UG}	Maximum UG Traded Power [kW]
E_t^{PV}	PV Energy Generated [kWh]	P_{max}^{h-SD}	Maximum Charging/Discharging Power of Heat Storage Device [kW]
$E_t^{e\text{-}CCHP}$	Electric Energy Output of CCHP [kWh]	P_{max}^{c-SD}	Maximum Charging/Discharging Power of Cold Storage Device [kW]
E_t^{light}, E_t^{vent}, E_t^{feed}	Lighting, Ventilation, And Feeding Load Energy, respectively [kWh]	α_D	Dissatisfaction Coefficient
$E_t^{PV\text{-}EL}$	PV Energy Input to Electrolyzer [kWh]	τ_n	Tolerance Energy of Nth Load [kWh]
$E_t^{e\text{-}HCH}$	Electric Energy Input to HCH [kWh]	SOC_t^{SD}	State of Charge of SD [kWh]
$E_t^{h\text{-}SD}$, $E_t^{c\text{-}SD}$	Charging/Discharging Energy of Heat and Cold SD, respectively [kWh]	E_t^{heat}, E_t^{cool}	Heating And Cooling Load Energy, respectively [kWh]
$E_t^{h\text{-}CCHP}$	Heat Energy Output of CCHP [kWh]	$E_t^{c\text{-}CCHP}$, $E_t^{c\text{-}HCH}$	Cold Energy Output of CCHP and HCH, respectively [kWh]

Parameters

η_{EL}	Electrolyser Efficiency	P_{max}^{c-load}	Maximum Cooling Load [kW]
η_c, η_e, η_h	Cold, Electric, And Heat Efficiencies Of CCHP, respectively	P_{max}^{PV}	Maximum PV Generated Power [kW]
$\eta_{e\text{-}HCH}$	Electricity To Cold Efficiency Of HCH	P_{max}^{P2P}	Maximum P2P Traded Power [kW]
P_{min}^{EL}, P_{max}^{EL}	Minimum And Maximum Electric Power Input to the Electrolyzer, respectively [kW]	P_{max}^{UG}	Maximum UG Traded Power [kW]
$P_{max}^{H2-CCHP}$, $P_{max}^{H2-CCHP}$	Minimum And Maximum Hydrogen Power Input to the CCHP, respectively [kW]	P_{max}^{h-SD}	Maximum Charging/Discharging Power of Heat Storage Device [kW]
P_{min}^{h-HCH}, P_{max}^{h-HCH}	Minimum And Maximum Heat Power Input to the HCH, respectively [kW]	P_{max}^{c-SD}	Maximum Charging/Discharging Power of Cold Storage Device [kW]
P_{min}^{e-HCH}, P_{max}^{e-HCH}	Minimum And Maximum Electric Power Input to the HCH, respectively [kW]	α_D	Dissatisfaction Coefficient
P_{max}^{e-CCHP}	Maximum Electric Power Output from The CCHP [kW]	τ_n	Tolerance Energy of Nth Load [kWh]
P_{max}^{h-CCHP}	Maximum Heating Power Output from The CCHP [kW]	N	Number Of Loads
P_{max}^{c-CCHP}	Maximum Cold Power Output from The CCHP [kW]	E_{min} , E_{max}	Minimum And Maximum Load Energy, respectively [kWh]

P_{max}^{c-HCH}	Maximum Cold Power Output from The HCH [kW]	SOC_{min}^{SD}, SOC_{max}^{SD}	Minimum And Maximum State of Charge Of SD, respectively [kWh]
P_{max}^{e-load}	Maximum Electric Load [kW]	R^{SD}	Rate Of Charge/Discharge of SD [kWh]
P_{max}^{h-load}	Maximum Heating Load [kW]		

Sets

sE_t^{PV}	States Set of PV Generation	aE_t^{Cool}	Action Set of Cooling Loads
sp_t^B, sp_t^S	States Sets of UG Buying and Selling Energy Prices, respectively	aE_t^{Heat}	Action Set of Heating Loads
aE_t^{Light}	Action Set of Lighting Loads	ap_t^{P2P}	Action Set of P2P Energy Trading Price
aE_t^{Vent}	Action Set of Ventilation Loads	aE_t^{P2P}	Action Set of P2P Energy Traded

1. INTRODUCTION

For dairy farms, electricity is a significant and inevitable expense. Dairy farmers are under more and more strain due to rising energy prices and worries about energy security. Investigating ways to lower energy expenses is crucial to addressing these issues. Negotiating advantageous supplier contracts and prices while highlighting the significance of peer-to-peer energy trading (P2P-ET) is a typical strategy. Peer-to-peer (P2P) energy trading has several advantages, including establishing a competitive energy market that lowers electricity costs, lowering power outages, improving power systems' overall efficiency, and enhancing alternative energy sources. Another option is to use demand side management (DSM) programs to reduce on-peak power use.

In addition to reducing greenhouse gas emissions, the deployment of renewable energy sources (RES) on dairy farms can also reduce the cost of network-dependent energy consumption. Solar energy options provide dairy farmers with a more cost-effective and energy-efficient alternative. The location of a farm also affects how other renewable energy sources, such as hydrogen, are evaluated. Dairy farms are among the many types of farms located along Egypt's northwestern coastline. Consequently, the integrated energy system of such farms depends on the synthesis of green hydrogen from seawater, as proposed in [1], using PV-generated power.

Managing the farm's demand within reasonable comfort levels and utilizing the most efficient, sustainable, and cleanest energy generation technologies are important aspects of cost-effective energy use. Peer-to-peer energy markets have been shown to be a successful way to facilitate energy trading among prosumers, which can maximize profits for surplus or needed energy [2].

The P2P-ET platform facilitates energy sharing between consumers and prosumers through bidding processes. Through these bids, participants can negotiate prices and specify the amount of energy they wish to exchange. The market coordinator then coordinates the clearing and transaction processing of the market while taking into account the operating limits of the power system [3]. In [4], two different double-auction trading algorithms for energy trading were proposed. The research focused on the pricing issue in shared parking but did not address DSM or P2P-ET. In [5], the authors introduced a discriminating k-double auction approach. However, the research goal was to enhance the efficiency of P2P energy trading by applying double auction mechanisms and did not consider the issue of energy management.

The combination of DSM and P2P-ET to assist participants reduce their electricity expenditure has attracted the interest of numerous researchers. This combination was applied effectively to different sectors such as residential, commercial, and industrial sectors [6]. In [7], a P2P energy exchange model that considered user welfare was introduced. Different pricing mechanisms were compared and evaluated based on their influence on welfare distribution. A method that accounted for the unpredictability of wind-generated electricity was proposed in [8] to identify the optimal P2P trading strategies among a large number of microgrids. The proposed method adopted a transformed optimal power flow model for energy trading with optimal topology planning. The objective of [9] was to determine the best P2P-ET plan for prosumers who wanted to reduce their electricity costs while maximizing their utilization of renewable energy sources through power consumption control. However, authors in [7-9] did not take DSM into account.

Nevertheless, there are two significant issues with the integration of DSM and P2P-ET. First, this integration complicates the decision-making process because the choice factors are highly unpredictable. Furthermore, the large search space produced by the mixed DSM and P2P-ET decisions makes the calculation challenging. Better decision-making is demonstrated by reinforcement learning (RL). There are various benefits of using reinforcement learning in decision-making. First, reinforcement learning does not require prior information, which could be challenging to handle in practice, as it learns the best course of action through contact with its environment. Second, by employing both offline training and online implementation, reinforcement learning (RL) may be used flexibly for a variety of applications. Thirdly, compared to traditional optimization techniques, RL is simpler to apply in practical settings. Because of its high computational efficiency, RL can generate the best results in a look-up table.

The application of reinforcement learning (RL) to energy management issues is of great interest to many academics. To help energy suppliers adjust their selling prices to meet the demand from energy customers, the authors in [10] used the K-means methodology. In [11], a Q-learning method based on neural networks (NNs) was used to obtain the best DSM with predicted solar power generation and energy pricing. The ability of Q-learning to solve optimization problems in discrete action space has been shown. It must, however, discretize this space while handling P2P-ET and DSM situations that call for a continuous space of activities, which significantly affects the accuracy of the results. In [12, 13], data-driven DRL algorithms were introduced for the optimal DSM of residential participation with expected load behavior. Using a multi-agent DRL algorithm and considering each machine as an agent, [14] was able to assist an industrial customer in calculating their energy use. However, the P2P-ET was not addressed in [12-14]. In an attempt to streamline the energy trading process, a system for trading energy was presented more recently in [15] but without considering DSM or mixed uncertainty. A P2P-ET and DSM model for residential homes was developed in [16] using a multi-agent DRL technique to determine both the P2P energy traded and the energy consumption of various equipment. In order to reduce the cost of dissatisfaction and the electricity bill, the suggested multi-agent DRL was then applied to the competitive challenge that was designed. However, the states-actions pairs of each agent did not consider the demand uncertainty.

In [17], the authors used RL to create a double-auction based mechanism that maximizes P2P market participation. Additional challenges with P2P-ET data transfer include security and latency. Authors in [18] addressed these problems by proposing a Prosumer Recommender

System for P2P-ET using RL and blockchain, and authors in [19] suggested a Secure Energy Trading scheme for P2P-ET based on Smart Contracts that controls the energy load of EVs, businesses, and residential homes. A secure DSM solution for residential energy management using RL and Ethereum blockchain was introduced in [20] to address data security issues. The RL-based energy management research in [17-20] did not integrate all the factors taken into account in this paper regarding the P2P-ET, DSM, data-driven prediction, and various uncertainties in RES, energy prices, and various load types.

Dairy farms require three types of energy usage, which DSM programs are responsible for managing: power, heating, and cooling. In order to save energy, it is also recommended to use the low-grade waste heat from the power generation process for both heating and cooling instead of relying solely on the P2P community/utility grid (UG) to meet all of these demands. The "waste heat" can be used in two ways: trigeneration and cogeneration. Trigeneration, the method of producing useful electricity, cooling, and heating from the same energy conversion process, is also known as combined cooling, heating, and power (CCHP).

Natural gas, a limited fossil fuel, powers most of the trigeneration systems currently in use; more environmentally friendly fuels must be used in its place. Hydrogen has been identified as the option that produces the fewest pollutants and shows the greatest promise for long-term renewable energy sources. In [21], a study using hydrogen was carried out to investigate the performance of a diesel-based micro-scale trigeneration engine-generator combo. The findings demonstrated the potential of hydrogen as an energy vector, which is required to meet the upcoming, stricter greenhouse gas emission regulations. According to the calculations, hydrogen has a very high chance of beating traditional diesel fuel in terms of energy efficiency and nearly carbon emissions.

The integrated energy system (IES) for the farm is recommended to satisfy load demands. Unlike other types of integrated energy systems (IES), the unique characteristics of dairy farms influence the design and operation of their IES. They have precise and consistent energy needs for milking, cooling, ventilation, and other agricultural processes. Dairy farms employ a combined cooling, heating, and power (CCHP) system that can generate electricity and capture waste heat for a range of agricultural uses. Additionally, such an IES includes a variety of device types, including energy storage systems and power to gas (P2G) [22]. However, depending on the farm's location, different energy carriers can be available. There might not be as much natural gas accessible because dairy farms are usually located in rural areas. This makes it necessary to identify an appropriate fuel source for the CCHP. Seawater-produced hydrogen can be used as a CCHP fuel instead of diesel because the dairy farm being studied in this work is located near the seaside. Additionally, the location is perfect for PV unit installation, which creates the possibility of green hydrogen production. On a farm, such energy device considerations lead to a clean and sustainable IES.

A detailed comparison of the research mentioned above and this paper is given in Table 1 to enable a better understanding of the motivation and contributions of this study.

As indicated in Table 1, the author is aware of no prior use of data-driven based RL for decision support of the integrated P2P-ET and DSM problem with mixed uncertainty. RL algorithms were used by researchers in [10 - 20], but none of them integrated all the factors taken into account in this article concerning P2P-ET, DSM, data-driven prediction, and various uncertainties in RES, energy prices, and various load types. Furthermore, although some studies examined the use of heat and cold storage systems in conjunction with natural gas-

fueled CCHP, others focused on the use of green hydrogen as a fuel for CCHP without taking into account both types of storage systems. Additionally, a decision-making technique that had not previously been discussed in the literature—modified multi-agent reinforcement learning, or MARL—is used to resolve the uncertainties brought on by the uncertain forecast of PV generation and the UG's buying/selling energy prices.

Table 1. Comparison between this work and other - reported in literature – research works.

Ref.	P2P ET	DSM	Uncertainties			Data-driven prediction	RL
			RES	Energy price	loads		
[3]	√	√	•	•	•	•	•
[7]	√	•	•	•	•	•	•
[8]	√	•	√	•	•	•	•
[9]	√	•	•	•	•	•	•
[10]	•	√	•	•	•	•	√
[11]	•	√	√	√	•	√	√
[12]	•	√	•	•	√	√	√
[13]	•	√	•	•	√	√	√
[14]	•	√	•	•	•	•	√
[16]	√	√	•	•	•	•	√
[17]	√	•	•	•	√	•	√
[18]	√	√	•	•	•	•	√
[19]	√	√	•	•	•	•	√
[20]	•	√	•	•	•	•	√
This paper	√	√	√	√	√	√	√

Multi-agent reinforcement learning techniques previously used to solve optimal energy management problems were formulated with states representing the deterministic PV generation and UG energy prices. Based on these states, the actions were taken. However, due to the forecasting of these states, different values of each of these states result, which are stochastically represented with different probabilities. As a result, in a time slot, each of these states have more than one value. These values are also taken into consideration when deciding the corresponding actions of RL agents at that time slot and hence determining the corresponding rewards.

Therefore, in the proposed MARL, it is found that incorporating additional states and action scenarios significantly enhanced the RL agent's ability to adapt to diverse and dynamic environments, particularly when faced with unexpected variations in PV generation and energy price. These extra states and actions allow the agent to learn more nuanced strategies and respond effectively to probabilistic circumstances.

Furthermore, in addition to taking into account a P2P trading mechanism for additional electricity bill reduction, this paper explores DSM for lowering both the energy consumption cost and the discontent cost. Additionally, the MARL is used for decision making in order to address the increasing uncertainties brought about by bidding actions, transaction amounts, load profiles, and the generation of renewable resources. Therefore, the following is a summary of this paper's primary contributions:

- The DSM and P2P-ET problem is formulated as an finite Markov decision process (FMDP) using a data-driven framework, and decision support for various load types is provided by a modified multi-agent RL methodology that takes mixed uncertain situations into account.
- Utilizing real-world data, the Decision Tree Regression (DTR) model forecasts utility grid energy costs, PV power generation, and other load usage. To account for varying prediction accuracies, the improved RL incorporates extra stochastic scenarios.
- A double auction-based clearing energy price is suggested for a fair P2P-ET, taking into account the DSM and P2P-ET for a dairy farm with various loads.
- To evaluate the effectiveness of the suggested strategy for handling uncertain decision-making problems, the outcomes of the suggested MARL methodology are contrasted with those of alternative optimization techniques that incorporate probabilistic uncertainty functions.
- Consideration is given to a clean and sustainable integrated energy system that includes heat and cold storage devices as distributed resources, hydrogen-fueled CCHP, and PV-power to hydrogen (green hydrogen).

2. SYSTEM DESCRIPTION

The integrated energy system of the dairy farm under study consists of PV-supply device, in addition to energy storage and energy conversion devices, and loads as shown in Fig. 1.

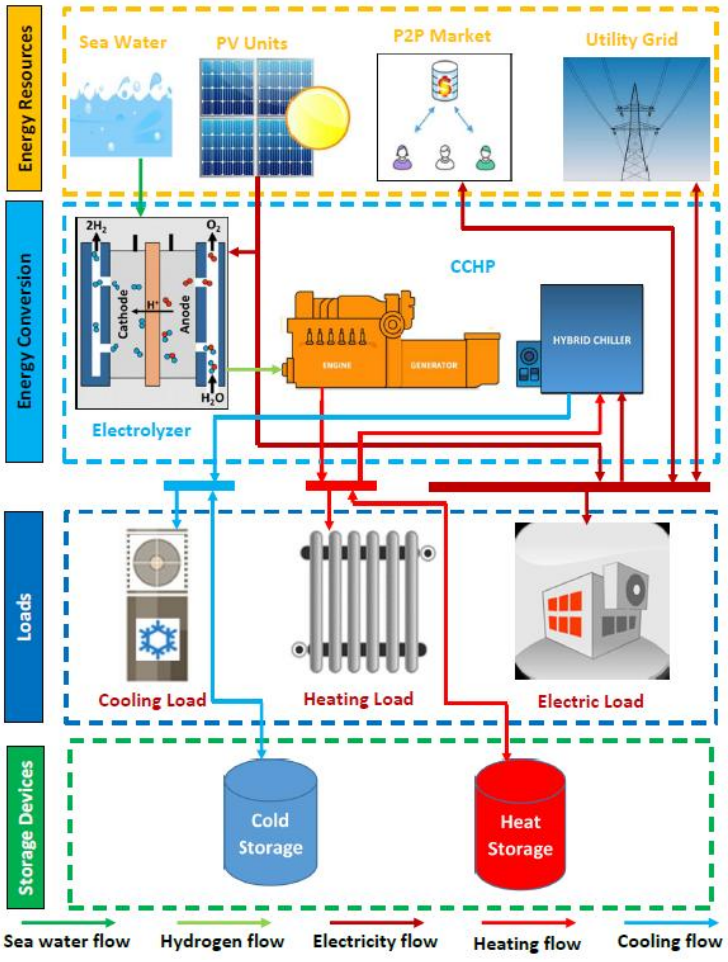

Fig. 1. The integrated energy system of the dairy farm.

2.1. The Electrolyzer

The efficiency of a hydrogen electrolyzer (EL) is typically measured by its electrical-to-hydrogen conversion efficiency. This efficiency represents the portion of the energy contained in the produced hydrogen gas to the electrical energy supplied to the EL. The PV-powered electrolyzer has the relationship given in Eq. (1).

$$\begin{cases} P_t^{H2} = \eta_{EL} * P_t^{PV-EL} \\ P_{min}^{EL} \leq P_t^{PV-EL} \leq P_{max}^{EL} \end{cases} \quad (1)$$

2.2. CCH Power

CCHP produces electricity from a hydrogen-fueled gas turbine in addition to waste heat. This heat is utilized for heating and cooling. For the cooling part, multiple hybrid electricity/heat cooling chillers (HCH) are used. The energy conversion relationships and operational constraints of the CCHP and HCH are given in Eqs. (2) and (3), respectively.

$$\begin{cases} P_t^{e-CCHP} = \eta_e * P_t^{H2} \\ P_t^{h-CCHP} = \eta_h * P_t^{H2} \\ P_t^{c-CCHP} = \eta_c * P_t^{H2} \\ P_t^{c-HCH} = \eta_{e-HCH} * P_t^{e-HCH} \end{cases} \quad (2)$$

$$\begin{cases} P_{min}^{H2-CCHP} \leq P_t^{H2} \leq P_{max}^{H2-CCHP} \\ P_{min}^{h-HCH} \leq P_t^{h-HCH} \leq P_{max}^{h-HCH} \\ P_{min}^{e-HCH} \leq P_t^{e-HCH} \leq P_{max}^{e-HCH} \end{cases} \quad (3)$$

2.3. Dairy Farm Loads

There are three types of loads in the dairy farm, namely electrical, heating and cooling loads. These loads represent the lighting, feeding, ventilation, heating and cooling loads, in addition to the electric power needed for the electrolyzer. Different energy resources should fulfill these loads within the constraints given in Eq. (4).

$$\begin{cases} P_{max}^{e-CCHP} + P_{max}^{PV} + P_{max}^{P2P} + P_{max}^{UG} \geq P_{max}^{e-load} \\ P_{max}^{h-CCHP} - P_{max}^{h-SD} \geq P_{max}^{h-load} \\ P_{max}^{c-CCHP} + P_{max}^{c-HCH} - P_{max}^{c-SD} \geq P_{max}^{c-load} \end{cases} \quad (4)$$

3. P2P ENERGY TRADING AND PROBLEM FORMULATION

3.1. Double Auction Mechanism for P2P Energy Trading

In this research, a double auction (DA) trading mechanism integrated with the MARL model is used to execute P2P energy trading. Participants submit bids and offers at the beginning of a trading period, indicating the quantity of energy to be traded as well as the required selling and buying prices. Next, the following formula is used to get the final market clearing price for this period (MCP_t) [17].

a) The selling and buying prices with the quantity are sorted as in Eq. (5) in ascending and descending order, respectively.

$$\begin{cases} p_t^{b1} > p_t^{b2} > \cdots > p_t^{bn} \\ p_t^{s1} < p_t^{s2} < \cdots < p_t^{sn} \end{cases} \quad (5)$$

b) The buying prices curve and selling prices curve will intersect as given in Eq. (6) at which

$$p_t^{bi} > p_t^{sj} > p_t^{b,i+1} \quad (6)$$

c) The final market clearing price is determined using the mid-price method given in Eq. (7) as

$$MCP_t = \frac{p_t^{bi} + p_t^{sj}}{2} \quad (7)$$

d) Once the MCPt is determined, participants who pass the auction trade through the P2P platform. On the other hand, participants who fail the auction need to trade with the UG for power balance.

Fig. 2 illustrates the determination of the MCP_t for 9 buyers and sellers. The prices are given as percentages of the buying/selling prices of energy of the utility grid.

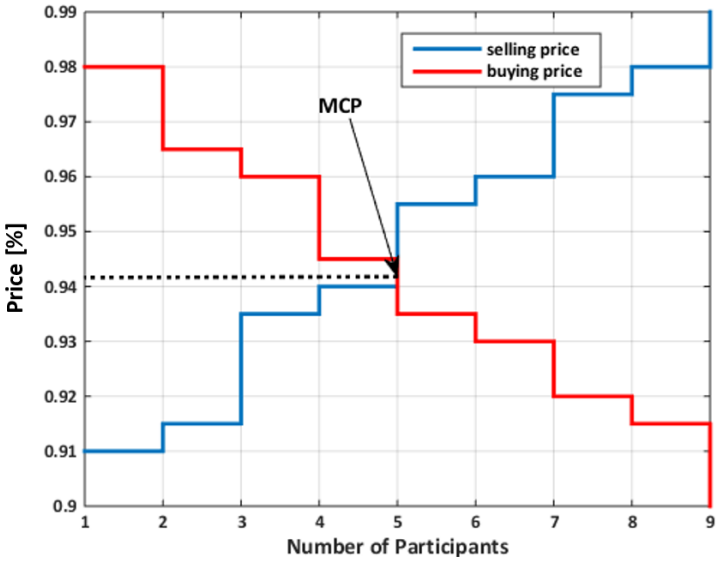

Fig. 2. MCP in a P2P market with N=9 buyers and sellers.

3.2. Problem Formulation

3.2.1. The Objective Function

The objective is to minimize the function given in Eq. (8)

$$C_t = \omega_1 c_t^E + \omega_2 c_t^D \quad (8)$$

in which, $\omega_1 + \omega_2 = 1$.

The resultant cost of traded energy consists of the cost of traded energy with the P2P community in addition to that traded with the UG. The trading process of the dairy farm has two cases, either buying or selling. These two cases are represented as follows in Eqs. (9) and (10).

In case of selling, i.e. E_t^{P2P} & $E_t^{UG} \leq 0$

$$c_t^E = \begin{cases} E_t^{P2P}\, MCP_t + E_t^{UG}\, p_t^S & if\ p_t^{P2P} \leq MCP_t \\ (E_t^{P2P} + E_t^{UG})\, p_t^S & if\ p_t^{P2P} > MCP_t \end{cases} \quad (9)$$

In case of buying, i.e. E_t^{P2P} & $E_t^{UG} > 0$

$$c_t^E = \begin{cases} E_t^{P2P}\, MCP_t + E_t^{UG}\, p_t^B & if\ p_t^{P2P} > MCP_t \\ (E_t^{P2P} + E_t^{UG})\, p_t^B & if\ p_t^{P2P} \leq MCP_t \end{cases} \quad (10)$$

The dissatisfaction cost given in Eq. (12) represents the deviation of the controllable loads' levels from their predefined set values taking into consideration a value tolerance of τ, as given in Eq. (11).

$$c_t^{D,n} = \begin{cases} 0 & if\ (E_t^{Set,n} - \tau_n) \le E_t^n \le (E_t^{Set,n} + \tau_n) \\ p_t^D \alpha_D \sqrt{\left(E_t^n - (E_t^{Set,n} - \tau_n)\right)\left(E_t^n - (E_t^{Set,n} + \tau_n)\right)} & else \end{cases} \quad (11)$$

$$c_t^D = \sum_{n=1}^{N} c_t^{Dn} \quad (12)$$

3.2.2. Operating Constraints

- Power balance constraints: electricity is needed in the dairy farm system to provide feeding loads, lighting, ventilation, and a portion of the cooling load by running the hybrid chiller. The PV system and CCHP provide the farm with its electricity, in addition to power exchanged with the utility grid and P2P platform. For electrical energy balance, the equality constraint given in Eq. (13) must be satisfied.

$$E_t^{PV} + E_t^{e-CCHP} + E_t^{P2P} + E_t^{UG} = E_t^{light} + E_t^{vent} + E_t^{feed} + E_t^{PV-EL} + E_t^{e-HCH} \quad (13)$$

As the farm cannot sell to one market and buy from another in the same period, the constraint in Eq. (14) must be satisfied for all time intervals.

$$E_t^{P2P} \times E_t^{UG} \ge 0 \qquad \forall\, t \in T \quad (14)$$

Regarding the cooling and heating loads, the energy balance is given in Eqs. (16) and (17). As the storage devices can be charging/discharging as given in Eq. (15).

$$E_t^{h-SD}\ \&\ E_t^{c-SD} \begin{cases} < 0 & if\ discharging \\ > 0 & if\ charging \end{cases} \quad (15)$$

$$E_t^{heat} = E_t^{h-CCHP} - E_t^{h-SD} \quad (16)$$

$$E_t^{cool} = E_t^{c-CCHP} + E_t^{c-HCH} - E_t^{c-SD} \quad (17)$$

- Load management constraints: when setting the loads' action sets for DSM, the maximum and minimum loads' levels must be taken into consideration as given in Eq. (18). As given in Eq. (18), all loads are controlled for flexible consumption levels between minimum and maximum values except the feeding load.

$$\begin{cases} E_{min}^{light} \le E_t^{light} \le E_{max}^{light} \\ E_{min}^{vent} \le E_t^{vent} \le E_{max}^{vent} \\ E_{min}^{cool} \le E_t^{cool} \le E_{max}^{cool} \\ E_{min}^{heat} \le E_t^{heat} \le E_{max}^{heat} \\ E_t^{feed} = E_{Set}^{feed} \end{cases} \quad (18)$$

- Storage devices constraints: the integrated power system of the dairy farm contains both heat and cold storage devices (SD). The operating constraints of such devices are given in Eq. (19), considering the charging/discharging condition in Eq. (15).

$$\begin{cases} SOC_t^{SD} = SOC_{t-1}^{SD} + E_t^{SD} \\ E_t^{SD} \le R^{SD} \\ SOC_{min}^{SD} \le SOC_t^{SD} \le SOC_{max}^{SD} \\ SOC_1^{SD} = SOC_T^{SD} \end{cases} \quad (19)$$

4. DATA-DRIVEN MULTI-AGENT RL FOR P2P-ET AND DSM

As previously mentioned, due to the large search space produced by the mixed DSM and P2P-ET decisions, the calculation challenge. Better decision-making is demonstrated by reinforcement learning (RL). There are various RL techniques such as deep Q-networks (DQN), actor-critic (AC) frameworks and multi-agent RL (MARL) frameworks. In this paper,

the MARL framework is applied as it offers superior stability and convergence compared to DQN and AC-based methods in multi-agent optimization problems. Regarding the stability issue, DQN relies on experience replay and target networks to stabilize Q-learning. However, in multi-agent settings, the environment becomes non-stationary as multiple agents simultaneously update their policies, making past experience obsolete.

Similarly, Actor-Critic (AC) methods use function approximation for both value estimation (critic) and policy updates (actor). In multi-agent settings, the policy gradients exhibit high variance, making learning unstable. On the other hand, the MARL approach uses a centralized critic that observes joint states, reducing non-stationarity while enabling decentralized execution for scalability. The Centralized Training with Decentralized Execution (CTDE) reduces instability by allowing centralized critics to learn a more structured representation of the environment while enabling decentralized agents to act independently [23].

When addressing the convergence issue, DQN-based multi-agent adaptations often fail in Markov games due to policy oscillation. On the other hand, Multi-Agent Actor-Critic (MAAC) attempts to mitigate this but still suffers from slow convergence in highly dynamic environments, while MARL stabilizes multi-agent training by allowing each agent to maintain its own policy while using a shared centralized critic for improved gradient estimation, leading to faster convergence [24].

In this paper, with data-driven framework, the P2P-ET and DSM problem is formulated as a FMDP to fit into a model-free multi-agent RL framework for decision support, as shown in Fig. 3, dealing with different types of loads, while considering mixed uncertain conditions.

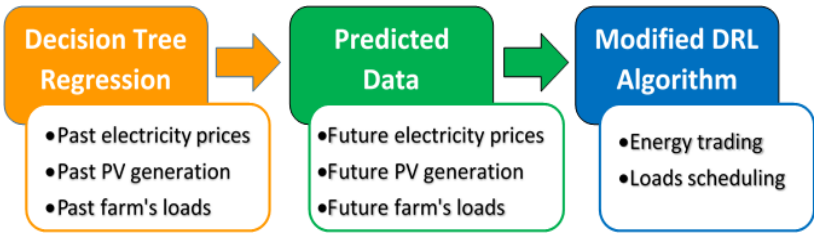

Fig. 3. Structure of the proposed data-driven based MARL solution.

The objective function to be minimized is described in Eq. (8) over a simulation period of T, which is described in Eq. (20)

$$\min \sum_{t=1}^{T} C_t \tag{20}$$

4.1. FMDP Formulation

This P2P-ET and DSM problem can be formulated as an FMDP, in which each agent takes an action corresponding to the states at a certain time step and is then granted a reward. The FMDP model contains three main elements—states, actions, and reward. This problem has six agents representing the four controllable loads (i.e. lighting, ventilation, heating and cooling loads), the P2P trading price and the P2P traded energy.

4.1.1. States

The states vector $\boldsymbol{s_t}$ at time t consists of a number of sub-vectors representing the PV generation, the buying energy prices and the selling energy prices offered by the UG and is given by:

$$s_t = [sE_t^{PV}, sp_t^B, sp_t^S] \quad (21)$$

4.1.2. Actions

The actions vector a_t at time t, which is described in (22), consists of a number of sub vectors representing the levels of controllable loads (E_t^{Load}), buying/selling P2P energy price (p_t^{P2P}), and the amount of energy traded through the P2P platform (E_t^{P2P}), as given in Eq. (23).

$$a_t = [aE_t^{Load}, ap_t^{P2P}, aE_t^{P2P}] \quad (22)$$

where,

$$\begin{cases} aE_t^{Light} = [E_t^{L1}, \dots, E_t^{Lm}, \dots] & Lm \in A_{Light} \\ aE_t^{Vent} = [E_t^{V1}, \dots, E_t^{Vm}, \dots] & Vm \in A_{Vent} \\ aE_t^{Cool} = [E_t^{C1}, \dots, E_t^{Cm}, \dots] & Cm \in A_{Cool} \\ aE_t^{Heat} = [E_t^{H1}, \dots, E_t^{Hm}, \dots] & Hm \in A_{Heat} \\ ap_t^{P2P} = [p_t^{P2P1}, \dots, p_t^{P2Pm}, \dots] & P2Pm \in A_{p-P2P} \\ aE_t^{P2P} = [E_t^{P2P1}, \dots, E_t^{P2Pn}, \dots] & P2Pn \in A_{E-P2P} \end{cases} \quad (23)$$

4.1.3. Reward

The reward function r_t represents the benefit obtained at time t corresponding to the (action a_t - state s_t) pair, which is the inverse of the cost function, and the total reward during the simulation period R, as given in Eqs. (24) and (25), respectively.

$$r_t = -C_t \quad (24)$$

$$R = \sum_{t=1}^{T} r_t \quad (25)$$

In order to calculate the reward value corresponding to certain state-actions combinations, while considering the operational constraints of the IES, the proposed algorithm is shown in Fig. 4 and proceeds as follows:

- The available PV power generation is known based on the state values at every time period. The electrolyzer will be powered by this PV power, which will also meet the operational constraint of Eq. (1).
- To meet the power balancing constraint of Eq. (13), any remaining PV electricity will either help supply the farm's electric needs or be traded.
- The electrolyzer's produced hydrogen will be utilized for fueling the CCHP unit in accordance with its operating constraints Eq. (3).
- The farm's electric loads will be supplied by the electricity produced by the CCHP unit. In order to satisfy the power balance constraint of Eq. (13) any remaining electric power will be traded.
- The CCHP unit's heat and cold powers will be utilized to meet the farm's corresponding loads while satisfying the power balance constraints outlined in Eqs. (16) and (17).
- In accordance with the operational constraints of the respective storage units, as stated in Eq. (21) the excess/shortage of heat and cold energy, if any, will be traded.
- The required energy to meet such loads is determined in accordance with the first four actions concerning the load levels to be taken into consideration. The quantity and direction of energy to be traded (i.e., bought or sold) will depend on whether the power balance constraint of Eq. (13) is satisfied. In order to meet the power balancing constraint of Eq. (13), the final action will decide what proportion of this energy is to be traded with the P2P community. Based on that percentage, the amount of energy to be traded with the utility grid will also be decided.

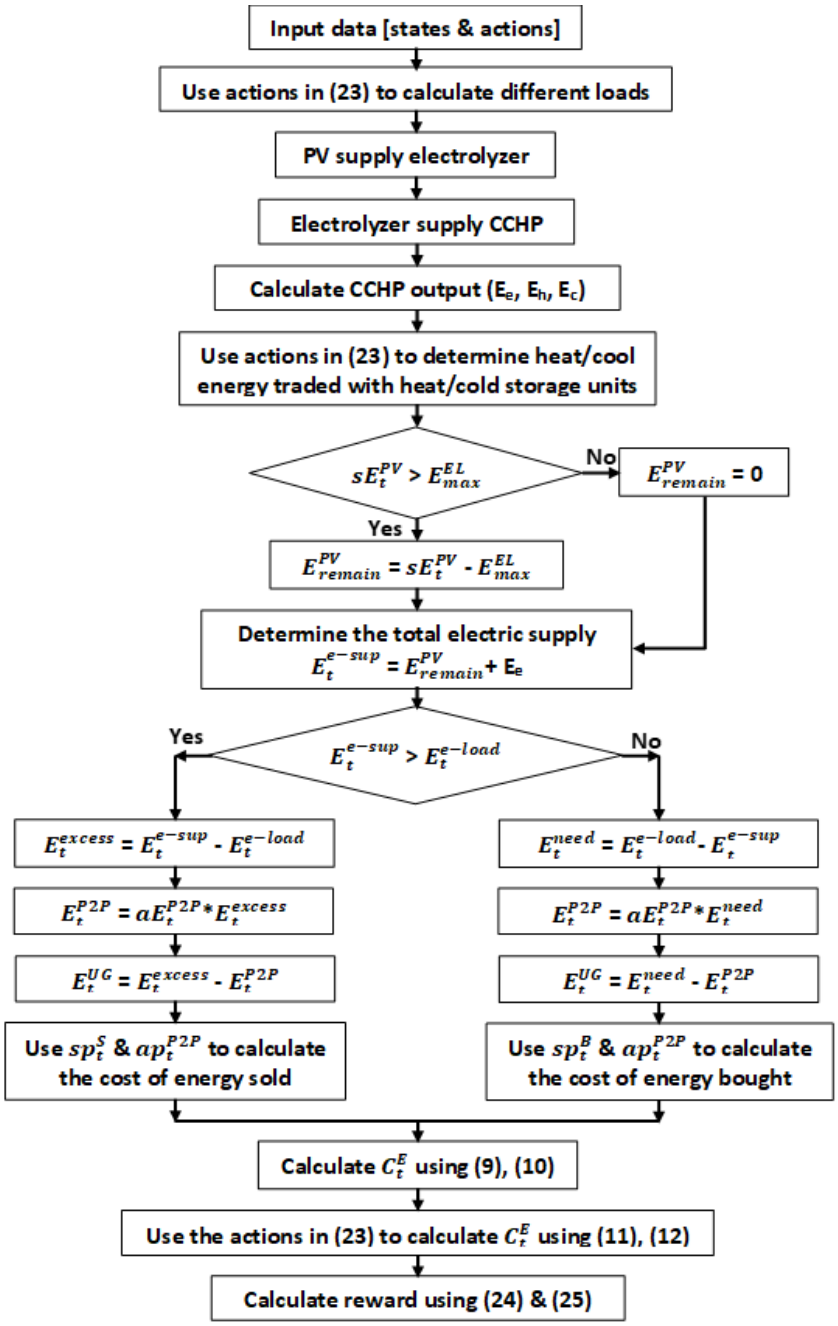

Fig. 4. The proposed algorithm for calculating the reward value corresponding to states-actions combinations.

4.2. Data-Driven Based Multi-Agent RL Algorithm

First, the predicted PV generation, UG energy prices and different loads consumption are acquired using the Decision Tree Regression (DTR) algorithm. For the MARL decision-making process, the Q-learning algorithm is applied to gain the expected rewards of each agent. The Bellman equation, given in Eq. (26), is used for the computation of Q-values for each state-action pair (s_t,), which provides an accurate approximation for rewards and updates.

$$Q(s_t, a_t) = r_t + \gamma \max[Q(s_{t+1}, a_{t+1})] \quad (26)$$

For learning, the previously constructed Q-table is updated in each training iteration. In this way, the optimal action with optimal Q-value in each state can be selected according to the respective reward [25].

4.3. Modified Multi-agent RL Algorithm with Uncertainties

While the DTR model provides accurate point predictions, it is important to recognize that real-world systems involve inherent uncertainties—such as load demand variability, weather fluctuations, and other stochastic factors. Deterministic predictions, even if accurate, do not capture these uncertainties. Probabilistic studies are essential to:

- Quantify the impact of uncertainty on system performance.
- Provide decision-makers with a range of possible outcomes for robust planning and risk assessment.
- Enhance the resilience of systems by preparing for extreme scenarios.

The need for probabilistic studies is therefore complementary to deterministic predictions, as they provide a more comprehensive evaluation.

As the data-driven prediction of RES, energy price, and load consumption can have different accuracies, the P2P energy trading and DSM problem have mixed uncertain conditions. The uncertainties arise due to the uncertain forecasts of PV generation and buying/selling energy prices of the UG (E_t^{PV}, p_t^B, p_t^S) affecting the problem states, in which additional states are included with accompanying probabilities. Additional uncertainties arise due to different loads forecasts are simulated as stochastic scenarios, which affects the reward function. In this paper, different scenarios are probabilistically modeled, in which a probability density function is assigned to each source of uncertainty. Different states scenarios are generated, and then the corresponding actions are selected. The reward corresponding to each state-action pair is then calculated using the modified reward function given in Eq. (27), which has two parts. The first part is deterministic (C_t), representing energy and dissatisfaction costs based on predicted values for PV power generated, loads levels, and UG energy prices. The second part is probabilistic C_t^{pr} representing the cost of each scenario (pr).

$$r_t = -[C_t + \sum_{pr=1}^{Spr} D_{pr}\, C_t^{pr}] \tag{27}$$

Where D_{pr} is the probability of each scenario [6].

In each iteration of the algorithm, each agent of the participating six agents observes the states s_t given in Eq. (21) and then chooses an action a_t of the corresponding set given in Eq. (23) using the exploration and exploitation mechanism. To realize the exploration and exploitation, the agent selects an action whose current Q-value is maximum. After taking an action, this action is used to evaluate the energy cost C_t using Eqs. (8)-(12) then the agent acquires an immediate reward r_t as given in Eq. (27). The agent observes the next state s_{t+1} and updates the Q-value (s_t, a_t) given in Eq. (26). This process is repeated until the state s_{t+1} is terminal. After each iteration, the agent checks the iterations' termination criterion, in this case, it is the number of iterations. If this termination criterion is not satisfied, the agent will move to the next iteration and repeat the above process.

5. SIMULATION RESULTS

5.1. Integrated Energy System

The integrated energy system of the dairy farm under study consists of a PV-supply device, energy conversion devices, energy storage devices, and loads. The operating

parameters of different system's components are given in Table 2. The predicted values of the different loads of the dairy farm are given in Fig. 5.

Table 2. Operating parameters of the integrated energy system's components.

Component	Operating parameter	Numerical value
PV system	Capacity [kW]	175
Electrolyzer	Capacity [kW]	100
	Efficiency, η_{EL}	0.61
CCHP	Capacity [kW]	100
	Electric Efficiency, η_{EL}	0.29
	Heat efficiency, η_{EL}	0.2
	Cold efficiency, η_{EL}	0.42
HCH	capacity [kW]	25
	efficiency η_{EL}	0.6
Heat storage device	Capacity [kW]	150
	Max. SOC [%]	90
	Min. SOC [%]	10
Cold storage device	Capacity [kW]	60
	Max. SOC [%]	90
	Min. SOC [%]	10

The energy-generating devices in this system are the PV-units, the electrolyzer, and the CCHP combined with an HCH. The heating and cooling loads of the farm are satisfied by the CCHP and HCH, in addition to the heat and cold storage units. Electric energy is utilized to operate the ventilation, lighting, and feeding loads in addition to part of the cooling load when the HCH is electrically operated. Therefore, the only energy that needs to be traded with the P2P platform and the utility grid is electric energy.

Consequently, the price used to calculate the electric energy cost is the electric energy price ($/kWh), which depends on the supplier, whether it is the UG or the P2P platform. In addition, the dissatisfaction price ($/kWh) is used to calculate the cost of dissatisfaction when any of the loads are deviate from their set value.

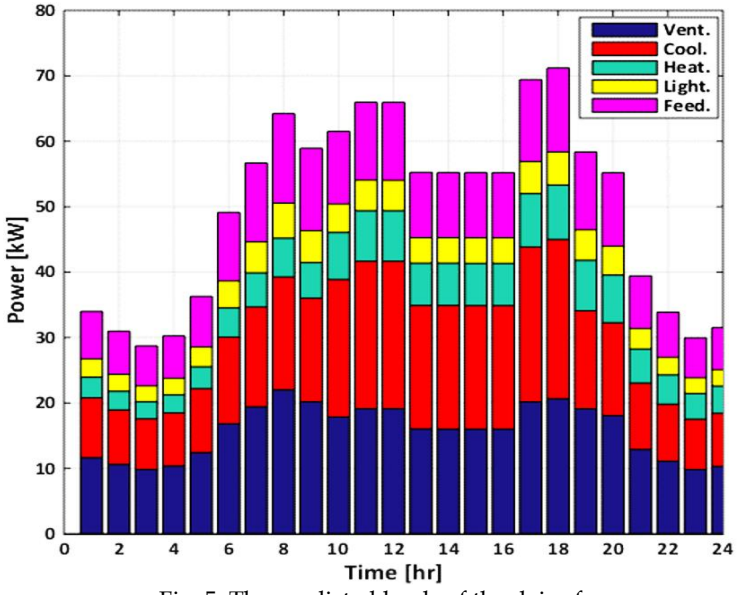

Fig. 5. The predicted loads of the dairy farm.

As for the energy prices of the utility grid in the case of buying or selling electric energy, the predicted prices are shown in Fig. 6.

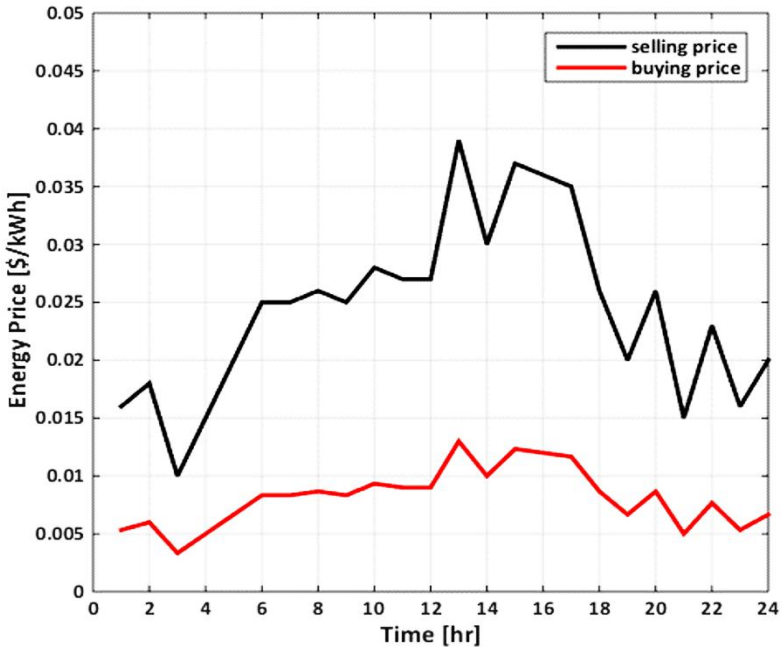

Fig. 6. The predicted selling/buying electric energy prices of the UG.

In this paper, there are 6 agents to be considered in the MARL algorithm as given in Eq. (23). To determine the action sets corresponding to the first 4 agents representing different load types, the acceptable levels of these loads are considered. Considering the ventilation load, it is recommended that the rate of ventilation during the summer be around 40 to 60 air changes per hour [26]. Therefore, it is set to vary between 0.7 and full load. Similarly, the recommended cooling, heating, and lighting levels in dairy farms are considered for the corresponding action sets of loads agents [27-29]. On the other hand, the feeding load is not considered for curtailment and hence, not considered for DSM. As for the other two agents, i.e. P2P trading price and P2P traded energy, their action sets are determined as a percentage of the energy price of the UG and a percentage of the total energy to be traded, respectively. The different action sets considered are given in Table 3.

Table 3. The action sets of MARL agents.

Agent ID	Action set
aE_t^{Light}	[0.75, 0.85, 0.95, 1] of full load
aE_t^{Vent}	[0.7, 0.8, 0.9, 1] of full load
aE_t^{Cool}	[0.7, 0.8, 0.9, 1] of full load
aE_t^{Heat}	[0.94, 0.96, 0.98, 1] of full load
ap_t^{P2P}	[0.98, 0.97, 0.96, 0.95, 0.94, 0.93, 0.92, 0.91, 0.9] of UG price when buying
	[0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98, 0.99] of UG price when selling
aE_t^{P2P}	[0.7, 0.8, 0.9, 1] of energy traded

5.2. Performance of the Decision Tree Regression Model and Considered Uncertainties

Real-world data is used in this paper for training the DTR model. The hourly data of electricity prices, PV generation, and different farm loads for one year, from January 1 to December 31, is utilized. As a sample result of the DTR model, Fig. 7 shows a comparison of

the predicted and actual ventilation load on January 1-3. The performance of the DTR model used to predict the required data was evaluated using the Mean-Square-Error (MSE) metric as shown in Table 4. As shown in Fig. 7 and Table 4, compared to real data, the DTR predicted datasets have low errors. In addition, consistency in the value of MSE across both training and test datasets confirms that the model is neither overfitted nor suffering from data leakage.

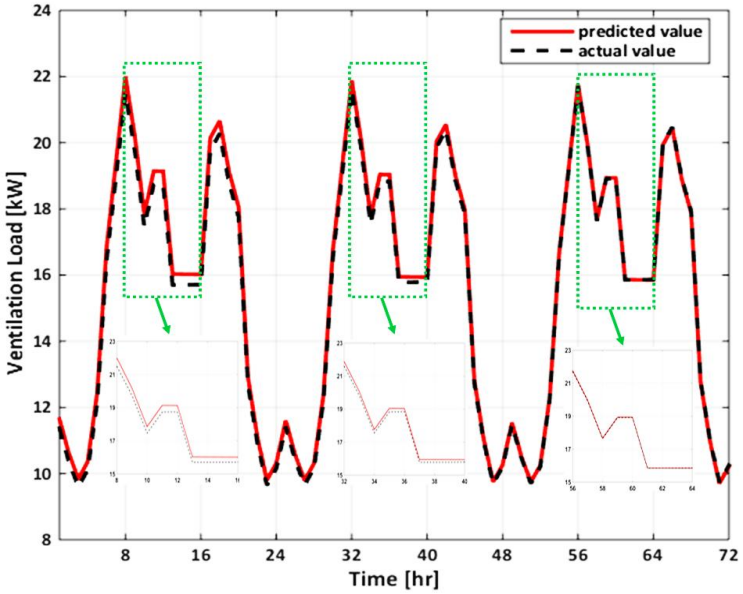

Fig. 7. Comparison between the predicted and the actual ventilation load on January 1-3.

Table 4. The MSE of different predicted datasets.

Dataset	Training-MSE	Test-MSE
Energy price	0.2081	0.2302
PV generation	0.0053	0.0135
Lighting load	0.0032	0.0101
Ventilation load	0.1809	0.2036
Heating load	0.0005285	0.0021
Cooling load	0.0672	0.0742
Feeding load	0.0098	0.0293

For considering the mixed uncertainties of this system, additional states are modeled representing different probabilistic scenarios regarding the PV generation and trading prices of the UG that arise from prediction accuracy.

In addition, the stochastic deviations of loads from their predicted values are also considered. In this paper, the stochastic variations in these parameters are assumed to follow normal distribution for uncertainty modeling [6]. The probability distribution functions are calculated with the predicted values of such parameters as their mean (μ) and with a standard deviation of 0.3. The probabilistic and deterministic values of the PV generation are shown in Fig. 8.

Fig. 9 (a) shows the convergence of the energy cost through the training process of the agents of the MARL algorithm. It can be seen in Fig. 9 (a) that the cost value converges to a minimum value as required. The reward convergence of the six agents through the training process is shown in Fig. 9 (b).

As shown in Fig. 9 (b), the reward value of each agent converged to a maximum value through the training iterations. In addition to converging to a minimum energy cost as required, the proposed MARL algorithm, with the additional probabilistic states-actions pairs,

succeeded to converge almost after 4000 iterations as shown in Fig. 9 (a). Furthermore, as shown in Fig. 9, the proposed MARL algorithm showed effectiveness in satisfying the objective minimum energy cost in spite of the highly mixed uncertainty of predicted data.

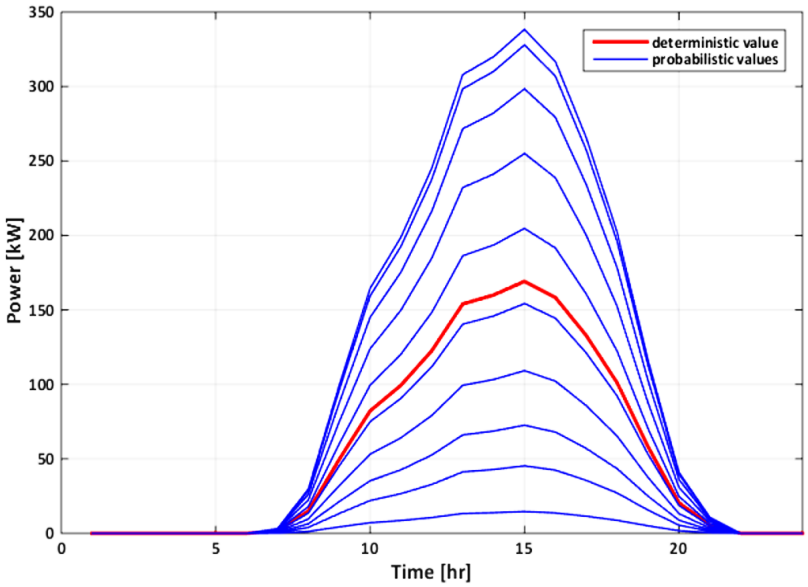

Fig. 8. The deterministic and probabilistic values of the PV generation.

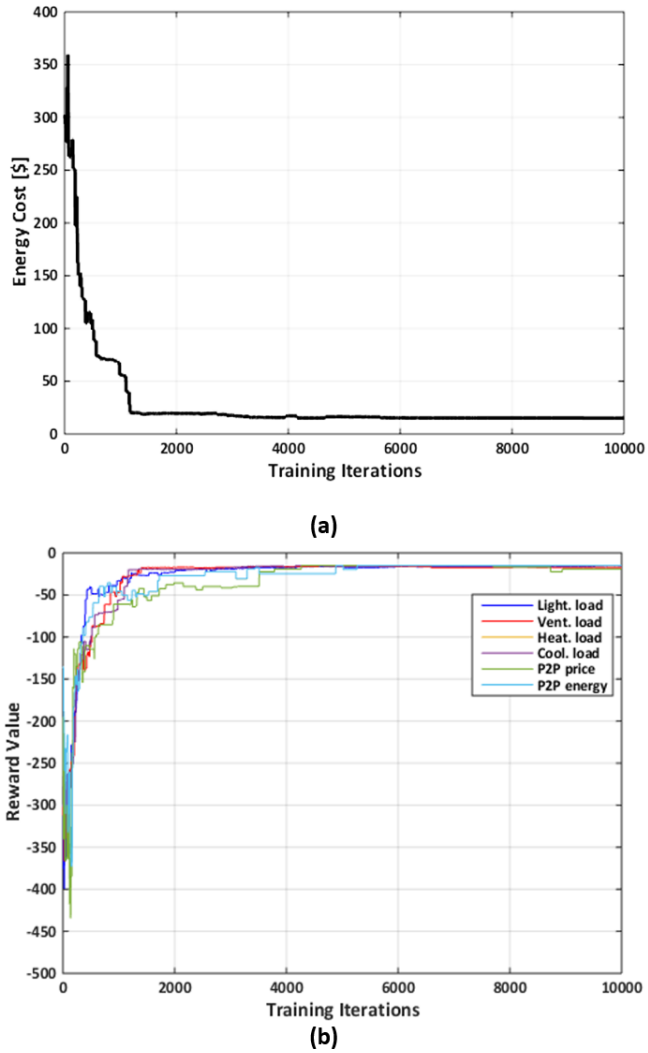

Fig. 9. The obtained through the training process of the MARL algorithm: a) the convergence of the energy cost; b) the reward convergence of the six agents.

5.3. Energy Trading Analysis

Table 5 presents the total day-ahead energy cost with and without considering MARL-based P2P energy trading or DSM or both. As shown in Table 5, applying DSM without P2P trading resulted in a reduction of 16.28% in the cost of purchased energy from the UG, while applying P2P trading without DSM resulted in a reduction of 14.21% in the overall cost of purchased energy. Finally, applying both P2P trading and DSM reduced the overall cost of purchased energy by 23.57%. This proves that the proposed MARL-based P2P-DSM solution can be more beneficial to the farm's financial system and is able to modify the energy consumption and trading behavior by managing and trading energy through the P2P market.

Table 5. The total day-ahead energy cost.

Case	DSM	P2P trading	Energy cost [$]
1	X	X	19.77
2	√	X	16.55
3	X	√	16.96
4	√	√	15.11

The intervals at which the farm system is buying and selling energy, and the corresponding energy cost are shown in Fig.10. Through the P2P trading algorithm 82.1% of the needed energy was bought from the P2P community, while the remaining 17.9% was bought from the utility grid. On the other hand, all of the excess energy was sold to the P2P community, which resulted in a considerable reduction in energy cost as shown in Fig. 10.

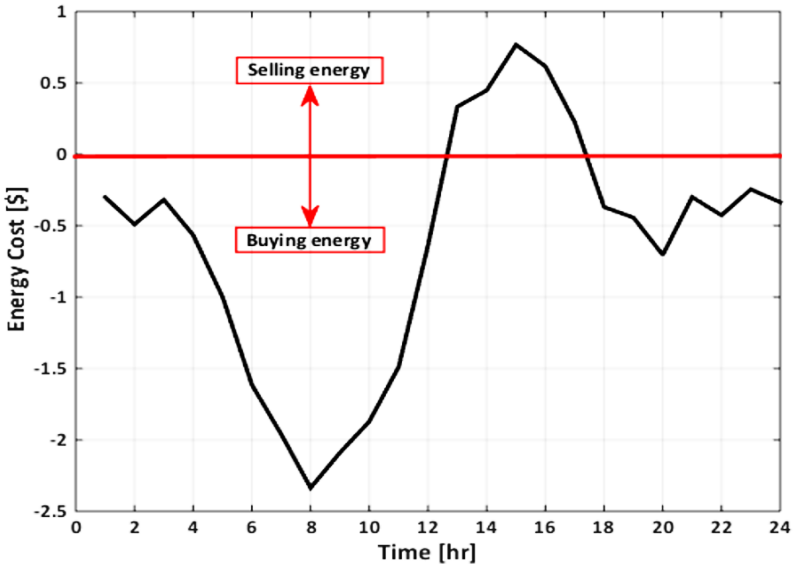

Fig. 10. The cost of energy trading for 24 hours.

5.4. Energy Management Analysis

As stated earlier, the integrated energy system of the dairy farm incorporates a hydrogen-fueled CCHP with a hybrid chiller in addition to the PV units. With the help of MARL-based DSM, the system can internally supply part, and sometimes all, of the required demand. Fig. 11(a) shows the PV power energizing the electrolyzer and the remaining PV power used/traded by the system, while Fig. 11(b) shows the total electric demand and the electric power supplied by the CCHP and PV units.

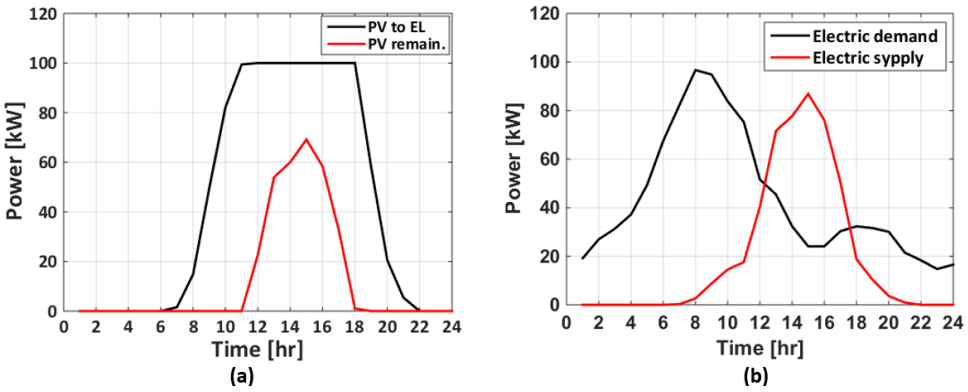

Fig. 11. a) The PV to electrolyzer power and the remaining PV power; b) the total electric demand and total electric power supplied by the CCHP and PV units.

As shown in Fig. 11(a), at the intervals in which the electrolyzer receives its rated power, there exists a surplus PV power that can be used/traded. The surplus PV energy represents about 22.41% of the total PV energy generated. On the other hand, as shown in Fig. 11(b), some intervals experience surplus electric energy when the generated power is greater than the demand, which is traded either with the P2P community or with the UG as shown in Fig. 10. Regarding the DSM, Fig. 12 shows the farm's loads with and without applying the MARL-based DSM. As shown in Fig. 12, reductions of 27.55%, 16.28%, 3.05% and 10.62% in the ventilation, lighting, heating and cooling loads, respectively, are achieved. Table 6 presents a comparison of different loads with and without applying MARL-based DSM in terms of load factor and energy savings. As shown in Table 6, the load factors of the lighting, heating and cooling loads encountered a reduction as the average loads were reduced while the peak loads were nearly the same as shown in Fig. 12. In the case of the ventilation load, the reduction of the peak load was greater than that of the average load, which resulted in a slight increase in the load factor. However, all of the farms' controllable loads experienced different degrees of energy savings as shown in Table 6.

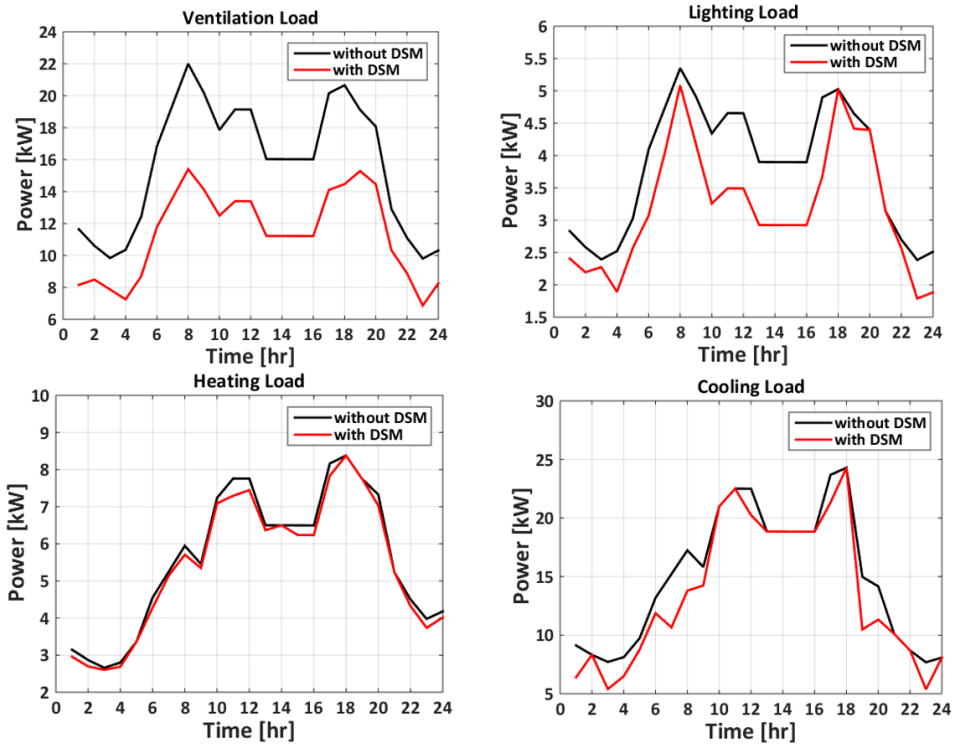

Fig. 12. The farm's loads with and without applying the MARL-based DSM.

Table 6. Comparison of different loads with and without applying MARL-based DSM.

Load type	Load factor		Energy savings	
	Without DSM	With DSM	Saving [kWh]	Percent [%]
Lighting load	3.8073	3.1874	14.8785	16.28
Ventilation load	0.7114	0.7363	103.5016	27.55
Heating load	0.668	0.6479	4.0266	3.05
Cooling load	0.6133	0.557	32.8144	10.62

Considering heating and cooling loads with respect to heat and cold power generation, Fig. 13 shows these powers. As shown in Fig. 13, at some intervals, the demands are lower than the available respective energies, which results in storing the surplus energies in the respective storage devices (SDs). At other time intervals when the generated energy is less than load demands, the stored energies are then used to compensate for such a supply shortage.

As shown in Fig. 13 (a), the heat energy generated by the CCHP along with the energy stored in the heat SD are adequate to supply the heat load, while keeping the SOC of the heat SD within a permissible range. On the other hand, as shown in Fig. 13 (b), at the interval between 2h and 10h, the SOC of the cold SD reached its minimum limit of 10%, while the cold energy generated by the CCHP is lower than the energy required to supply the cooling load. At this interval, the HCH is electrically operated to fulfill the required load.

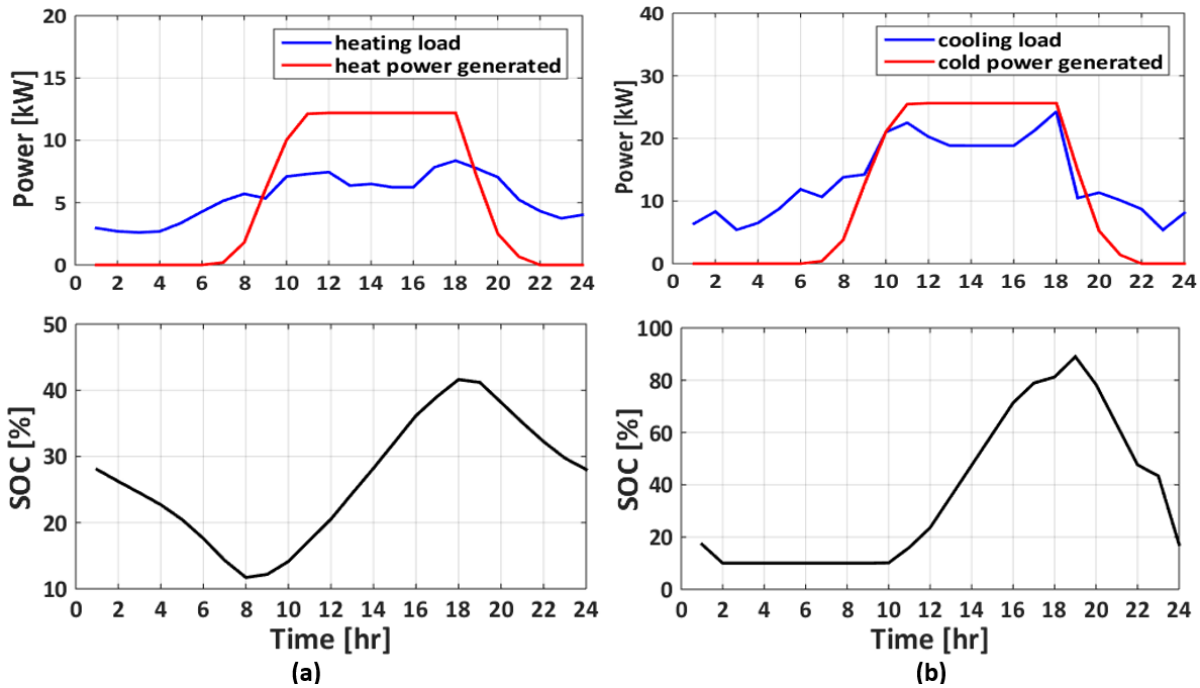

Fig. 13. a) Heating loads, heat power generation and SOC of heat SDs; b) cooling loads, cold power generation and SOC of cold SDs

5.5. Multi-Agent RL Algorithm Performance Compared to other Optimization Algorithms

In order to evaluate the proposed multi-agent RL algorithm, the P2P-DSM problem was solved using a multi-layer individual-based algorithm that was first proposed by the author in [30]. In the first layer, the DSM was addressed. Then, the P2P energy trading was addressed in the second layer. Mixed uncertainties were considered, and the algorithm was applied using

the firefly algorithm (FA), particle swarm optimization (PSO), and genetic algorithm (GA), for decision-making and optimization.

First, the FA is a nature-inspired metaheuristic optimization technique based on the flashing behavior of fireflies. It optimizes DSM and P2P energy trading by balancing energy demand and supply while minimizing costs. FA is particularly useful in handling non-linear, multi-objective optimization problems in energy management due to its ability to escape local optima. Second, the GA is an evolutionary algorithm that mimics natural selection, where candidate solutions evolve through selection, crossover, and mutation. In DSM and P2P energy trading, GA helps optimize energy scheduling and trading strategies by finding near-optimal solutions through an adaptive search process. It is widely used for solving complex optimization problems with multiple constraints.

Finally, the PSO is a swarm intelligence-based algorithm inspired by the collective movement of birds and fish. It optimizes DSM and P2P energy trading by iteratively improving potential solutions based on the experiences of individual particles (agents) and the swarm as a whole. PSO is well-suited for real-time energy management due to its fast convergence and robustness in dynamic environments.

These algorithms offer different strengths in optimizing energy cost reduction and improving system efficiency, making them valuable tools for managing uncertainties in DSM and P2P energy trading.

The simulation is run with 100 individuals and 10000 iterations on an Intel i7, 8 GB RAM laptop. The convergence of the cost function through iterations of each algorithm is shown in Fig. 14 and the computational times and minimum cost values are given in Table 7.

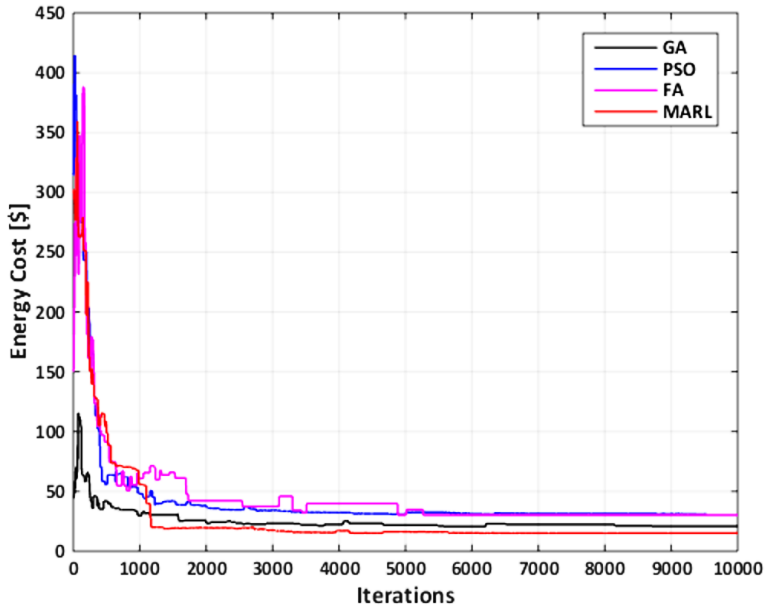

Fig. 14. Convergence of the cost function through iterations of each algorithms

Table 7. Computational times and minimum values of different algorithms.

Algorithm	Computational time	Minimum value [$]
MARL	27.9 s	15.11
GA	31.25 min	20.76
PSO	2.62 min	30.33
FA	1.53 min	30.24

As shown in Fig. 14, the proposed multi-agent RL algorithm succeeded in converging to the lowest value of energy cost, lower than any of the other three algorithms and its computational time was the lowest too as shown in Table 7. GA, PSO and FA also succeeded in converging to a minimum cost value but higher than that reached by multi-agent RL. As shown also in Fig. 14, the nearest algorithm to multi-agent RL was GA but with the highest computational time. As PSO and FA are almost near in nature, they both converged to nearly the same minimum value, but FA had a lower computational time.

6. CONCLUSIONS

This paper presented a clean and sustainable integrated energy system for a dairy farm, with a focus on DSM and P2P-ET for improved energy efficiency and cost savings. The investigation addressed the increased complexity of decision-making caused by the uncertainty of decision variables in DSM and P2P-ET, as well as the computational challenges arising from a large search space. To tackle these challenges, a modified MARL approach - formulating the DSM and P2P-ET problem as a FMDP - was developed and applied. The proposed MARL algorithm demonstrated - compared to other AI-based optimization methods - significant improvements, reducing the average daily energy cost by 23.57% and the average load by 20.73%. Moreover, it not only enhanced cost efficiency but also improved computational performance. These findings underscore the scientific contribution of integrating MARL into energy management systems, offering a scalable and adaptive solution for complex energy trading and DSM problems. The results demonstrate the practical applicability of the proposed approach in real-world energy systems, particularly for decentralized renewable energy management in agricultural and industrial sectors.

The proposed approach can be implemented using existing smart grid infrastructure and IoT-enabled energy management systems. Future research will focus on real-world deployment, testing the MARL-based framework in operational dairy farms or similar agricultural and industrial settings. Additionally, further investigation can address potential challenges such as data privacy, cybersecurity risks, and regulatory constraints in decentralized energy markets. By refining the model through pilot studies and integrating real-time energy pricing mechanisms, the approach can be fine-tuned to ensure robustness and adaptability in diverse real-world scenarios.

REFERENCES

[1] J. Guo, Y. Zheng, Z. Hu, C. Zheng, J. Mao, K. Du, M. Jaroniec, S. Qiao, T. Ling, "Direct seawater electrolysis by adjusting the local reaction environment of a catalyst," *Nature Energy*, vol. 8, pp. 264–272, 2023, doi: 10.1038/s41560-023-01195-x.

[2] P. Siano, G. Marco, A. Rolán, V. Loia, "A survey and evaluation of the potentials of distributed ledger technology for peer-to-peer transactive energy exchanges in local energy markets," *IEEE System Journal*, vol. 13, no. 3, pp. 3454–3466, 2019, doi: 10.1109/JSYST.2019.2903172.

[3] F. Alfaverh, M. Denai, Y. Sun, "A dynamic peer-to-peer electricity market model for a community microgrid with price-based demand response," *IEEE Transaction on Smart Grid*, vol. 14, no. 5, pp. 3976–3991, 2023, doi: 10.1109/TSG.2023.3246083.

[4] X. Haohan, X. Meng, Y. Hai, "Pricing strategies for shared parking management with double auction approach: Differential price vs. uniform price," *Transportation Research Part E: Logistics and Transportation Review*, vol. 136, p. 101899, 2020, doi: 10.1016/j.tre.2020.101899.

[5] S. Malik, S. Thakur, M. Duffy, J. Breslin, "Comparative double auction approach for peer-to-peer energy trading on multiple microgrids," *Smart Grids and Sustainable Energy*, vol. 8, no. 4, p. 21, 2023, doi: 10.1007/s40866-023-00178-x.

[6] A. El-Zonkoly, "Feasibility of blockchain-based energy trading within islanded microgrids in Alexandria, Egypt", *Journal of Energy Engineering*, vol. 147, no. 3, p. 04021009, 2021, doi: 10.1061/(ASCE)EY.1943-7897.0000754.

[7] M. Dynge, K. Berg, S. Bjarghov, Ü. Cali, "Local electricity market pricing mechanisms' impact on welfare distribution, privacy and transparency," *Applied Energy*, vol. 341, p. 121112, 2023, doi: 10.1016/j.apenergy.2023.121112.

[8] H. Hou, Z. Wang, B. Zhao, L. Zhang, Y. Shi, C. Xie, "Peer-to-peer energy trading among multiple microgrids considering risks over uncertainty and distribution network reconfiguration: A fully distributed optimization method," *International Journal of Electrical Power and Energy Systems*, vol. 153, p. 109316, 2023, doi: 10.1016/j.ijepes.2023.109316.

[9] H. Sahebi, M. Khodoomi, M. Seif, M. Pishvaee, T. Hanne, "The benefits of peer-to-peer renewable energy trading and battery storage backup for local grid," *Journal of Energy Storage*, vol. 63, p. 106970, 2023, doi: 10.1016/j.est.2023.106970.

[10] S. Ma, H. Liu, N. Wang, L. Huang, H. Goh, "Incentive-based demand response under incomplete information based on the deep deterministic policy gradient," *Applied Energy*, vol. 351, p. 121838, 2023, doi: 10.1016/j.apenergy.2023.121838.

[11] X. Xu, Y. Jia, Y. Xu, Z. Xu, S. Chai, C. Lai, "A multi-agent reinforcement learning based data-driven method for home energy management," *IEEE Transaction on Smart Grid*, vol. 11, no. 4, pp. 3201–3211, 2020, doi: 10.1109/TSG.2020.2971427.

[12] G. Xu, L. Liu, Y. Lu, Y. Zhao, L. Zhang, S. Song, "Perception and decision-making for demand response based on dynamic classification of consumers," *International Journal of Electrical Power and Energy Systems*, vol. 148, p. 108954, 2023, doi: 10.1016/j.ijepes.2023.108954.

[13] L. Xiong, Y. Tang, C. Liu, S. Mao, K. Meng, Z. Dong, F. Qian, "A home energy management approach using decoupling value and policy in reinforcement learning," *Frontiers of Information Technology & Electronic Engineering*, vol. 24, no. 9, pp. 1261–1272, 2023, doi: 10.1631/FITEE.2200667.

[14] R. Li, Y.C. Li, Y. Li, J. Jiang, Y. Ding, "Multi-agent deep reinforcement learning based demand response for discrete manufacturing systems energy management," *Applied Energy*, vol. 276, p. 115473, 2020, doi: 10.1016/j.apenergy.2020.115473.

[15] Q. Liu, D. Feng, Y. Zhou, H. Li, K. Zhang, S. Shi, "Equilibrium analysis for electricity market considering carbon emission trading based on multi-agent deep reinforcement learning," IEEE/IAS Industrial and Commercial Power System Asia, 2023, doi: 10.1109/ICPSAsia58343.2023.10294544.

[16] J. Wang, L. Li, J. Zhang, "Deep reinforcement learning for energy trading and load scheduling in residential peer-to-peer energy trading market," *International Journal of Electrical Power and Energy Systems*, vol. 147, p. 108885, 2023, doi: 10.1016/j.ijepes.2022.108885.

[17] H. Pereira, L. Gomes, Z. Vale, "Peer-to-peer energy trading optimization in energy communities using multi-agent deep reinforcement learning," *Energy Informatics*, vol. 5, p. 44, 2022, doi: 10.1186/s42162-022-00235-2.

[18] A. Kumari, R. Gupta, S. Tanwar, "PRS-P2P: a prosumer recommender system for secure P2P energy trading using Q-learning towards 6G," IEEE International Conference on Communications Workshops, 2021, doi: 10.1109/ICCWorkshops50388.2021.9473888.

[19] A. Kumari, A. Shukla, R. Gupta, S. Tanwar, S. Tyagi, N. Kumar, "ET-DeaL: A P2P smart contract-based secure energy trading scheme for smart grid systems," IEEE INFOCOM, 2020, doi: 10.1109/infocomwkshps50562.2020.9162989.

[20] A. Kumari, S. Tanwar, "A reinforcement-learning-based secure demand response scheme for smart grid system," *IEEE Internet of Things Journal*, vol. 9, pp. 2180–2191, 2022, doi: 10.1109/JIOT.2021.3090305.

[21] Y. Wang, Y. Huang, E. Chiremba, A. Roskilly, N. Hewitt, Y. Ding, D. Wu, H. Yu, X. Chen, Y. Li, J. Huang, R. Wang, J. Wu, Z. Xia, C. Tan, "An investigation of a household size trigeneration running with hydrogen," *Applied Energy*, vol. 88, no. 6, pp. 2176–2182, 2011, doi: 10.1016/j.apenergy.2011.01.004.

[22] F. Teng, Q. Zhang, T. Zou, J. Zhu, Y. Tu, Q. Feng, "Energy management strategy for seaport integrated energy system under polymorphic network," *Sustainability*, vol. 15, no. 53, pp. 1–22, 2023, doi: 10.3390/su15010053.

[23] X. Liu, B. Jin, "Information-theoretic multi-agent algorithm based on the CTDE framework," International Conference on Electronic Technology and Information Science, 2024, doi: 10.1109/ICETIS61828.2024.10593780.

[24] H. Sheikh, L. Bölöni, "Multi-agent reinforcement learning for problems with combined individual and team reward," International Joint Conference on Neural Networks, 2020, doi: 10.1109/IJCNN48605.2020.9206879.

[25] T. Ahammed, I. Khan, "Ensuring power quality and demand-side management through IoT-based smart meters in a developing country," *Energy*, vol. 250, p. 123747, 2022, doi: 10.1016/j.energy.2022.123747.

[26] N. Akdeniz, J. Mccarville, H. Schlesser, L. Seefeldt, R. Sterry, "Ventilation in dairy buildings," 2023, https://dairy.extension.wisc.edu/articles/ventilation-in-dairy-buildings.

[27] *Heat stress in dairy cows*, 2023, https://www.vostermans.com/ventilation/heat-stress-in-dairy-cows.

[28] *Optimal Lighting for Dairy Cows*, 2023, https://www.fas.scot/article/mmn-november-2023-optimal-lighting-for-dairy-cows.

[29]*Cleaning your milking system*, 2023, https://www.dairynz.co.nz/milking/milking-plant-maintenance/cleaning-your-milking-system/.

[30] A. El-Zonkoly, "Optimal energy management in smart grids including different types of aggregated flexible loads," *Journal of Energy Engineering*, vol. 145, no. 5, p. 04019015, 2019, doi: 10.1061/(ASCE)EY.1943-7897.0000613.