Optimizing Building Energy Management with Deep Reinforcement Learning for Smart and Sustainable Infrastructure

[1]Nabeel S. Alsharafa, [2]Suguna R, [3]Raguru Jaya Krishna, [4]Vijaya Krishna Sonthi, [5]Padmaja S M and [6]Mariaraja P
[1]Department of Information Technology, College of Science, University of Warith Al-Anbiyaa, Karbala, Iraq.
[2]Department of Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, Tamil Nadu, India.
[3]Department of Computer Science and Engineering, Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, Manipal, Karnataka, India.
[4]Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India.
[5]Department of Electrical and Electronics Engineering, Shri Vishnu Engineering College for Women Bhimavaram, Andhra Pradesh, India.
[6]Department of Electrical and Electronics Engineering, P. A. College of Engineering and Technology, Pollachi,Tamil Nadu, India.
[1]nabeel.alshreefy@uowa.edu.iq, [2]drsuguna@veltech.edu.in, [3]raguru.krishna@manipal.edu, [4]vijayakrishna1990@gmail.com, [5]padmaja_vvr@yahoo.com, [6]mariarajap@gmail.com

Correspondence should be addressed to Raguru Jaya Krishna : raguru.krishna@manipal.edu.

Article Info
Journal of Machine and Computing (http://anapub.co.ke/journals/jmc/jmc.html)
Doi: https://doi.org/10.53759/7669/jmc202404036
Received 11 September 2023; Revised from 02 January 2024; Accepted 09 February 2024.
Available online 05 April 2024.

Abstract – This study develops a new technique for optimising Energy Consumption (EC) and occupant satisfaction in business centres using Building Energy Management Systems (BEMS) that implement Deep Reinforcement Learning (DRL). Energy Management Models (EMM) are growing increasingly advanced and vital for intelligent power systems due to the growing demand for energy efficiency and the adoption of Renewable Energy Sources (RES), which are subject to variability. Flawed energy Consumption (EC) and problems are typical effects of traditional BEMS due to their unpredictability and failure to adapt to new environments. In this intended investigation, a DRL framework is demonstrated that may evolve its decision-making in real-time to control energy savings, electricity, and HVAC through input from the environment in which it operates. A pair of significant metrics, namely the cost of energy and room temperature stability, are employed to assess the effectiveness of the model compared to that provided by conventional rule-driven and predictive control systems. As investigated with different baseline models, the experimental findings proved that the DRL approach significantly reduced the cost of electricity while maintaining stable levels of comfort.

Keywords – Smart Grid, Deep Learning, Deep Reinforcement Learning, Renewable Energy, Energy Cost, Energy Storage Management.

I. INTRODUCTION

The introduction of advanced technology has made it achievable to develop "Smart Grids (SG)," which have the benefit of implementing electronics that are digital into traditional electric power systems. The introduction of this system was motivated by rising demands for a more predictable and trustworthy distribution of power, as well as by the objectives to boost the value of Energy Consumption (EC) by increasing the use of energy generated by Renewable Energy Sources (RES) [1]. The processing of data collected in real time by SG in order ensure demand-side management and indeterminate power fluctuations is done in order to set up a SG that is simultaneously flexible and dynamic. By setting up the implementation of freely accessible energy and controlling the random output of energy generated by RES, this builds a basis for the development of sustainable energy methods [2].

With the aid of Energy Management Systems (EMS) as part of SG, sustainable distribution and consumption of power are controlled. A segment of EMS, Building Energy Management Systems (BEMS), determines if the EC of a building is in line with the grid's infrastructure and the demands of the people who live there [3]. The primary methods by which BEMS enable the SG to perform with greater effectiveness is by regulating HVAC and electricity, two particularly energy-intensive building systems. The BEMS has been developed to enhance the well-being of occupants, fulfil the demands of sustainability, and reduce maintenance expenses and impact on the environment by regulating these systems [4].

This method uses the BEMS interface between SG owners and power optimization/Demand Response (DS) programs [5]. An array of EMS systems is available, from schedule-based rule-based systems to data-driven predictive models that predict the demand for energy in the future [6]. Whereas these models can provide an outline for EMS, they failed when it involves (i) integrated with fluctuating RE inputs, (ii) adapting energy distribution to specific owners' dynamic needs, and (iii) real-time adaptability.

Moreover, they are incapable of adapting on a personal basis to shifting energy conditions, which implies they cannot adapt. The key objectives of these technologies should be to maintain consumer safety while optimising EC—more intelligent, adaptive systems with autonomy for learning need to be developed to deal with these challenges. Deep Reinforcement Learning (DRL) networks have a tendency for success at this particular task.

The current investigation introduces a framework that combines the DRL and BEMS in an attempt to boost the business premise's livability and EC. The DRL approach accepts people's values, temperatures, and present-day EC as data and employs them to generate real-time decisions regarding controlling HVAC environments, lighting available, and Energy Storage Systems (ESS). The proposed model was experimented with in a five-floor commercial building of 10000 *sq. ft*. The data was collected for a duration of 90 days, and using two baseline models, (a) Traditional Rule-Based System (TRS) and (b) Predictive Control System (PCS), the proposed DRL model was compared. The simulations have shown that the proposed model has performed better than the other baseline models.

The paper is structured as follows: **Section 2** presents the literature study, **Section 3** presents the methodology, **Section 4** presents the experiment analysis, and **Section 5** concludes the work.

II. LITERATURE REVIEW

In [7] have presented a review article that analyzed 121 research works. The review analyzed the works in the current field and future scope related to intelligent control systems in smart buildings. The review paper listed key factors, such as comfort parameters, control systems, and occupant behaviour, as influential factors for constructing efficient EMS. [8] proposed an Artificial Intelligence Technique for Monitoring Systems in Smart Buildings (AIMS-SB) to achieve Energy Consumption (EC), production, and recycling using different factors. They developed an efficient prediction and monitoring system to manage RE production in smart cities.

The [9] came up with a Smart Building Energy Management System (SBEMS) that utilized factors like thermal and electrical power loops, RE sources, battery storage systems, and heat sharing and storage facilities. They developed a genetic algorithm-optimized learning model to analyse the effectiveness of charging scheduling of a bidirectional power network. Another effective model for energy scheduling in smart buildings was proposed by [10]. They utilized the Deep Reinforcement Learning (DRL) approach to classify the device demand and predict EC and demand for effective scheduling. They demonstrated their model using cloud infrastructure to reduce delay and cost in smart city energy distribution. [11-14] had developed a DRL-based Energy Management Model (EMM) for intelligent buildings. Their work focused on optimizing energy consumption using the DRL. They modelled distributed energy generation systems by establishing Q-learning-based EMM.

In [15-18] proposed a HVAC scheduling method for energy savings in buildings. They employed RL to collect, analyze, and infer EC data from a hotel testbed. Their attempt to design a purpose-oriented energy-saving methodology based on RL helped manage the HVAC systems efficiently. An online optimization of the BEMS system was proposed by [19-20]. They designed a DRL-based BEMS with Deep Q-learning and evaluated the model using the Pecan Street Inc. database. Their model had shown efficient EMM using energy scheduling strategies and consumer feedback.

III. METHODOLOGY

Description of the Building and Systems

In this study, the choice of building is a commercial office building located in Budapest, Hungary. The building has a covering area of 10,000 square meters over five floors. The building was built using energy-efficient materials such as low-emissivity (low-E) glass and high R-value insulation [21-24]. The occupancy rate in the building varies, with a typical daily peak of 800 occupants during standard office hours (9 AM to 5 PM) and decreasing to less than 100 occupants during off-hours and weekends. **Fig 1** shows the first-floor structure of the chosen building.

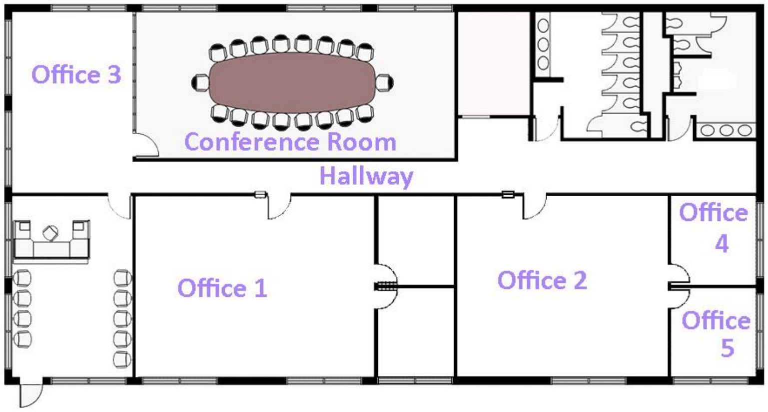


Fig 1. Floor Plan of the First Floor of The Selected Building

HVAC System: The HVAC system features a Variable Air Volume (VAV) setup with 20 high-efficiency air handling units (AHUs), each with a heat recovery system capable of achieving up to 75% thermal efficiency. The system is segmented into 50 zones to manage temperature and airflow control effectively. Temperature set points are adjusted between 20°C and 24°C based on occupancy rate and external temperature conditions. The system's economizer cycle is activated whenever the external temperatures are between 15°C and 20°C at such a temperature zone, and the cooling EC is reduced by utilizing outdoor air for cooling.

Electrical System: The power system includes dual feeders, automatic transfer switches for critical loads, and real-time control and monitoring EMS. The system included an Advanced Metering Infrastructure (AMI) with precision of ±1%. A 50 kW photovoltaic (PV) array is integrated into the building's electrical system through a grid-tie inverter that covers approximately 5% of its total electrical demand.

Lighting System: The building's lighting system includes LED fixtures with an average luminous efficacy of 100 lumens per watt. The system is designed to maintain average illuminance levels of 500 *lux* in workspaces, with the capability to adjust between 300 *lux* to 700 *lux* based on task requirements and natural daylight availability. Occupancy sensors in all significant spaces and daylight sensors near windows ensure optimal lighting conditions and minimize energy waste. The lighting control system is integrated with the BMS, allowing scene setting in meeting rooms and dynamic adjustments in response to occupancy patterns and daylight levels.

Problem Definition

The problem of optimizing BEMS with DRL is to dynamically control the building's energy systems, primarily HVAC and lighting, to minimize energy consumption and cost while maintaining or improving occupant comfort levels. This optimization must respect operational limits and indoor environmental quality standards. The challenge lies in the building's dynamic and complex nature, influenced by external (*e.g.*, weather, time of day) and internal (*e.g.*, occupancy, activities) conditions.

The objective function is designed to minimize the net EC of the building, accounting for energy efficiency, RE utilization, and occupant comfort. This can be expressed as EQU (1).

$$\min f(\mathbf{x}) = \alpha \cdot E_{\text{total}} - \beta \cdot E_{\text{PV}} + \gamma \cdot \Delta C \tag{1}$$

where, $f(\mathbf{x})$ is the objective function, E_{total} represents the total EC of the building, E_{PV} denotes the energy generated from photovoltaic sources, ΔC is a metric indicating deviation from optimal comfort levels, α, β, γ are weighting factors that reflect the relative importance of each term in the objective function, **x** symbolizes the set of controllable variables, such as HVAC settings, lighting intensity, and the operation of electrical devices.

Subject to Constraints:

Thermal Comfort Constraints: The system must maintain indoor thermal conditions within acceptable ranges, EQU (2) and EQU (3).

$$T_{\min} \leq T_{\text{indoor}} \leq T_{\max} \tag{2}$$

$$H_{\min} \leq H_{\text{indoor}} \leq H_{\max} \tag{3}$$

Energy Supply Constraints: The energy demand must not exceed the total supply from the grid and RE sources at any given time, EQU (4).

$$E_{\text{demand}}(t) \leq E_{\text{grid}}(t) + E_{\text{PV}}(t) \tag{4}$$

Operational Limits: Operational parameters for HVAC, lighting, and other systems must stay within predefined safety and efficiency ranges, EQU (5).

$$x_{\min,i} \leq x_i \leq x_{\max,i} \tag{5}$$

Building Energy Storage Model

The ESS is characterized by its storage capacity C_{ESS}, measured in kilowatt-hours (kWh), which determines the total amount of energy that can be stored. The system's charge and discharge efficiencies, η_{charge} and $\eta_{\text{discharge,}}$, respectively, influence the net amount of usable energy from the storage system. These efficiencies account for energy losses during conversion processes, such as inverting electrical energy from DC to AC.

The ESS is integrated into the BEMS to dynamically charge (store energy) during periods of low energy demand or high renewable production and discharge (release energy) during peak demand periods or when RE production is low. The decision to charge or discharge is governed by the DRL algorithm, which considers the building's current and predicted energy needs, the status of the grid, and the availability of RE.

The State-of-Charge (SOC) of the ESS at any time $t, SOC(t)$, is determined by the previous state of charge, the energy input (charging), and the energy output (discharging), adjusted for efficiencies, EQU (6).

$$SOC(t) = SOC(t-1) + \eta_{\text{charge}} \cdot E_{\text{in}}(t) - \frac{1}{\eta_{\text{discharge}}} \cdot E_{\text{out}}(t) \tag{6}$$

where, $E_{\text{in}}(t)$ is the Energy charged into the ESS at time t(kWh), $E_{\text{out}}(t)$ is the Energy discharged from the ESS at time t(kWh).

The ESS serves as a buffer to maximize the use of RE (*e.g.*, from the building's PV array) and minimize dependence on the grid, especially during peak tariff periods. It allows storing excess RE generated during peak production hours and its use during peak demand or low production periods.

The model optimizes the timing and quantity of energy to be stored or released, taking into account:

- Predicted RE generation based on weather forecasts and historical data.
- Predicted building energy demand based on occupancy patterns, scheduled activities, and historical consumption data.
- Grid energy prices and demand-response signals to minimize energy costs and participate in grid stabilization efforts.

Objective Function Modification

The introduction of the ESS into the BEMS necessitates a modification of the original objective function to include terms representing the cost or benefit of using the ESS, EQU (7)

$$\min f'(\mathbf{x}, SOC) = f(\mathbf{x}) + \delta \cdot C_{\text{grid-ESS}} \tag{7}$$

where, $f'(\mathbf{x}, SOC)$ is the modified objective function, including the ESS, δ represents the cost (or negative benefit) associated with charging from or discharging to the grid, $C_{\text{grid-ESS}}$ is the cost associated with grid interactions facilitated by the ESS.

HVAC Model

The HVAC system is critical for maintaining indoor comfort while minimizing EC. This model focuses on optimizing the HVAC operations through a balance of temperature control and energy efficiency. The HVAC system's operation is optimized by adjusting the temperature set points (T_{set}) and air flow rates (Q_{air}) based on occupancy levels (N_{occ}) and external weather conditions ($T_{\text{ext}}, H_{\text{ext}}$). The objective is to maintain indoor thermal comfort ($T_{\text{indoor}}, H_{\text{indoor}}$) within acceptable standards while reducing EC.

The energy consumption (E_{HVAC}) of the HVAC system can be modelled as a function of the airflow rate and the difference between the set point temperature and the external temperature, adjusted by the system's efficiency (η_{HVAC}), EQU (8).

$$E_{\text{HVAC}} = \eta_{\text{HVAC}} \cdot Q_{\text{air}} \cdot (T_{\text{set}} - T_{\text{ext}}) \tag{8}$$

where,

- E_{HVAC} = EC of the HVAC system (kWh).
- η_{HVAC} = Efficiency of the HVAC system.
- Q_{air} = Air flow rate (cubic meters per second, m^3/s).

- T_{set} = Temperature set point ($^\circ$C).
- T_{ext} = External temperature ($^\circ$C).

The objective function for the HVAC model within the DRL framework can be defined as minimizing the EC subject to maintaining indoor environmental conditions within predefined comfort ranges, EQU (9).
$\min E_{\text{HVAC}}(\mathbf{x})$, subject to:

$$\begin{gathered} T_{\min} \leq T_{\text{indoor}} \leq T_{\max} \\ H_{\min} \leq H_{\text{indoor}} \leq H_{\max} \\ CO2_{\min} \leq CO2_{\text{indoor}} \leq CO2_{\max} \end{gathered} \tag{9}$$

where $T_{\text{indoor}}, H_{\text{indoor}}$, and $CO2_{\text{indoor}}$ are the indoor temperature, humidity, and CO2 concentration, respectively. $T_{\min}, T_{\max}, H_{\min}$, and $H_{\max}$ are the minimum and maximum acceptable limits for temperature and humidity, and $CO2_{\min}$ and $CO2_{\max}$ define the acceptable indoor air quality levels.

Cost Model

The Cost Model aims to quantify the financial implications of operating the BEMS, primarily focusing on the HVAC and Energy Storage System (ESS) under an optimal EMM. This model calculates the total energy cost, considering varying electricity tariffs, operational costs, and potential savings from energy-efficient practices.

The cost of electricity is a significant factor in the total operational cost and varies according to the time of day, demand, and utility provider policies. Let $C_{\text{elec}}(t)$ denote the electricity cost at time t, which can be defined as EQU (10).

$$C_{\text{elec}}(t) = P(t) \cdot E_{\text{total}}(t) \tag{10}$$

where, $P(t)$ represents the price per kWh of electricity at time t, $E_{\text{total}}(t)$ is the total EC at time t, combining HVAC, lighting, and other systems; HVAC Operational Cost: The operational cost of the HVAC system, C_{HVAC}. It depends on its EC and the current electricity tariff. It can be expressed as EQU (11).

$$C_{\text{HVAC}} = \sum_t P(t) \cdot E_{\text{HVAC}}(t) \tag{11}$$

where, $E_{\text{HVAC}}(t)$ is the EC by the HVAC system at time t.
ESS Cost: The cost associated with the ESS, C_{ESS}, includes charging costs during low-demand periods and savings from discharging during peak tariff periods. It is calculated as EQU (12).

$$C_{\text{ESS}} = \sum_t [P(t) \cdot E_{\text{in}}(t) - P(t) \cdot E_{\text{out}}(t)] \tag{12}$$

where, $E_{\text{in}}(t)$ and $E_{\text{out}}(t)$ are the energy charged into and discharged from the ESS at time t, respectively.

The total cost, C_{total}, encompasses the cumulative costs of electricity consumption, HVAC operation, and ESS operation, EQU (13).

$$C_{\text{total}} = \sum_t [C_{\text{elec}}(t) + C_{\text{HVAC}} + C_{\text{ESS}}] \tag{13}$$

MDP Formulation

Fig 2 presents the proposed BEMS model and, for the model, defines a MDP formally as a five-tuple $M = (S, A, P, R, \gamma)$, where S represents the set of all possible environmental states, and A encompasses all possible actions. The transition probability function $P: S \times A \times S \rightarrow [0,1]$ captures the uncertainty in how states evolve based on actions undertaken by the agent. The reward function $R: S \times A \rightarrow \mathbb{R}$ alongside a discount factor $\gamma \in [0,1]$ completes this model. In this study, the *'agent'* refers to the BEMS controller observes the current state s_t, executes an action a_t, leading to the evolution of the environment to a new state s_{t+1}, and receives a reward R_{t+1}. In the following sections, this work will define the essential elements of the MDP framework, specifically focusing on the environmental states, actions, and reward function to formulate a comprehensive strategy for managing building energy efficiently.

State: The state (s_t) is represented as a set of variables such as indoor (T_{indoor}) and outdoor temperatures (T_{outdoor}), occupancy levels (N_{occ}), energy storage levels (E_{ESS}), and HVAC system status (P_{HVAC}) at time t that characterizes the current condition of the building and its immediate environment, EQU (14).

$$s_t = [T_{\text{indoor}}, T_{\text{outdoor}}, N_{\text{occ}}, E_{\text{ESS}}, P_{\text{HVAC}}] \tag{14}$$

Action: The action (a_t) at time t represents the set of decisions applied to the building's systems to influence its energy performance, EQU (15),

$$a_t = [A_{\text{HVAC}}, A_{\text{light}}, A_{\text{ESS}}] \tag{15}$$

where A_{HVAC} controls the HVAC operation, A_{light} adjusts lighting, and A_{ESS} decides ESS behavior. These actions are chosen based on the current state to optimize the building's EC while maintaining comfort and efficiency.

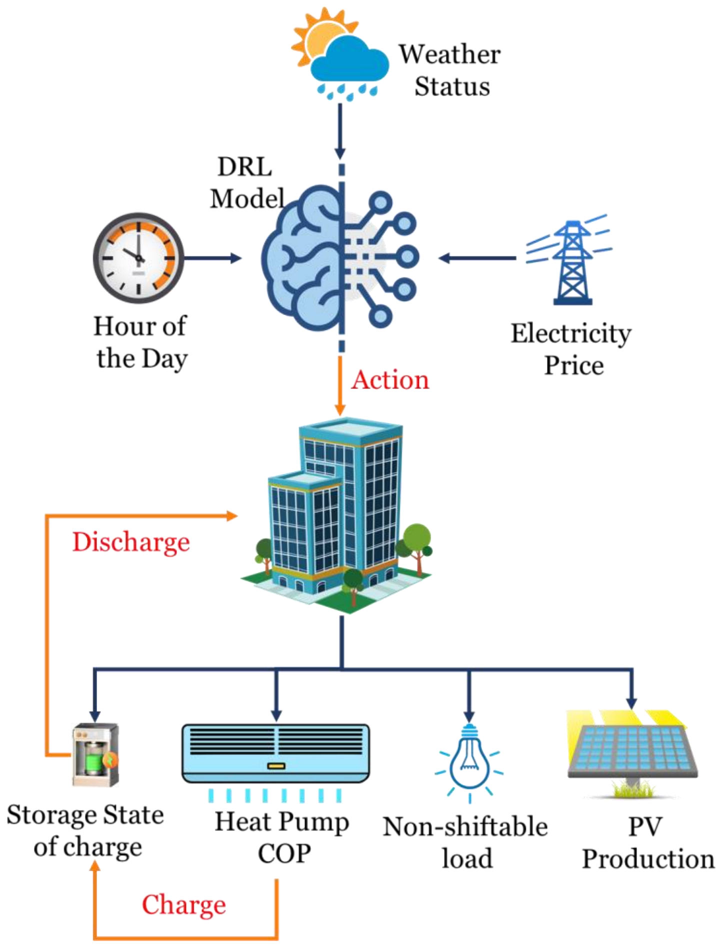


Fig 2. Proposed BEMS model

Reward: The reward function r_t at time t aims to direct the DRL agent toward minimizing EC and associated costs while upholding a comfortable indoor temperature range, EQU (16).

$$r_t = -\left(\alpha \cdot E_{\text{HVAC},t} + \beta \cdot D_{\text{ESS},t} + \gamma \cdot \left|\Delta T_{\text{comfort},t}\right|\right) \tag{16}$$

where, $E_{\text{HVAC},t}$ quantifies the HVAC system's energy consumption, $D_{\text{ESS},t}$ accounts for the ESS's depreciation, reflecting its usage and maintenance costs, $\left|\Delta T_{\text{comfort},t}\right|$ measures deviations from the ideal comfort temperature range. The factors α, β, and γ are weighting coefficients for HVAC efficiency, ESS preservation, and temperature comfort, respectively.

Action Value Function: The Action Value Function, denoted as $Q(s,a)$, is integral for calculating the expected return of choosing a specific action a in state s, and after adhering to the best policy after that. It quantifies the expected cumulative rewards, facilitating the identification of actions that yield optimal EC and occupant comfort. The action value function for BEMS can be expressed as EQU (17).

$$Q(s,a) = \mathbb{E}\left[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s, a_t = a\right] \tag{17}$$

where, s_t and a_t represent the current state and action, respectively, γ is the discount factor, r_{t+k+1} stands for the reward received after $k+1$ steps. The objective is to iteratively update $Q(s,a)$ values to mirror the real action values under an optimal policy π^* that maximizes the building's operational efficiency, EQU (18)

$$\pi^*(s) = \arg\max_a Q(s,a) \tag{18}$$

Crucially, the iterative updating of $Q(s,a)$ values based on actual experiences allows the DRL model to adapt to changing conditions and learn effective strategies for real-time EMM.

Transition Probability

Transition probabilities, denoted as $P(S_{t+1} \mid S_t, A_t)$, represent the likelihood of the system transitioning from a current state S_t to a new state S_{t+1} given an action A_t. Understanding and accurately modelling these probabilities is essential for predicting the system's behavior in response to various control actions, enabling the DRL agent to make informed decisions to optimise energy efficiency and occupant comfort.

The transition probability function can be expressed as EQU (19).

$$P(S_{t+1} = s' \mid S_t = s, A_t = a) \tag{19}$$

where, s' is the potential next state of the system, s represents the current state, a is the action taken in state s.

Q-Learning Algorithm for the BEMS

Q-learning is a model-free reinforcement learning technique that helps find the optimal action-selection policy for any finite Markov decision process. It works by learning an action-value function that ultimately provides the expected utility of taking a given action in a given state and following the optimal policy afterwards.

Algorithm for Q-Learning for BEMS Optimization

Inputs:

- s : Set of states
- A: Set of actions
- α (alpha): Learning rate
- γ (gamma): Discount factor
- ε (epsilon): Exploration rate
- $MAX_{episodes}$: Maximum number of episodes
- T_{STATES}: Set of terminal states

Procedure:

1 Initialize Q(s, a) for each state s in S and action a in A to zero. For terminal states in T_{STATES}, $Q(s,a) = 0$.

2 For episode $= 1$ to $MAX_{episodes}$ Do:
  - Initialize state s to an initial state of the BEMS.
  - While s is not in T_{STATES} Do:
    - o Choose action a from state s using ε-greedy policy:
      - ▪ With probability ε, choose a random action from A.
      - ▪ With probability $1-\epsilon$, choose $a = \text{argmax}_{a'} Q(s,a')$.

o Take action a, observe the next state s' and reward r.

  o Update the Q-value for the state-action pair (s,a) using EQU (20).

$$Q(s,a) \leftarrow Q(s,a) + \alpha \cdot \left(r + \gamma \cdot \max_{a'} Q(s',a') - Q(s,a)\right) \tag{20}$$

    o Update state s to s'.

  - End While

3 End For

4 Derive optimal policy π^* from Q-values:
  - For each state s in S do, EQU (21).

$$\pi^*(s) = \text{argmax}_a Q(s,a) \tag{21}$$

5 Return Q, π^*

IV. PERFORMANCE EVALUATION

In this study, simulations incorporate datasets reflecting actual conditions for solar energy production, constant energy demands, exterior temperature variations, and fluctuating electricity tariffs, all derived from the expansive Energy Information Administration (EIA) database. Given the emphasis on the cooling functionality of Building HVAC systems due to the extreme heat observed during summer, the test dataset spans from April to July 2023 for both the training and testing of the proposed model. More precisely, data from April to June are utilized for the neural network models' training phase, while July data is reserved for evaluating the model's effectiveness. To accurately represent the indoor environment's thermal dynamics within proposed BEMS optimization simulations, this work implements a simplified model for indoor temperature changes, structured as follows: EQU (22).

$$T_{\text{indoor}}(t+1) = T_{\text{indoor}}(t) + \Delta t \cdot \left(\frac{P_{\text{HVAC}}(t) - Q_{\text{loss}}(t) + Q_{\text{gain}}(t)}{C_{\text{air}}}\right) \tag{22}$$

where:

- $T_{\text{indoor}}(t)$ is the indoor temperature at time t,
- Δt is the time step between t and $t+1$,
- $P_{\text{HVAC}}(t)$ represents the power input from the HVAC system,
- $Q_{\text{loss}}(t)$ accounts for the heat loss,
- $Q_{\text{gain}}(t)$ includes heat gains from solar radiation,
- C_{air} is the specific heat capacity of air.
-

The simulation is configured with the system parameters as listed in **Table 1**.

Table 1: Parameters and Settings for the Model

Parameter	Value
Learning Rate (α)	0.1
Discount Factor (γ)	0.9
Exploration Rate (ϵ)	1 to 0.01
$\boldsymbol{\Delta t}$	1hour
Maximum $P_{\text{HVAC}}(t)$	10 kW
Average $Q_{\text{loss}}(t)$	2 kW
Average $Q_{\text{gain}}(t)$	3 kW
$\boldsymbol{C_{\text{air}}}$	$1.005\ \text{kJ/kg} \cdot \text{K}$

The proposed model is compared against the following two baseline models:

- Baseline 1: Traditional Rule-Based System (TRS)

  TRS utilizes static, predefined rules for controlling HVAC, lighting, and Energy Storage Systems (ESS). It operates on fixed schedules or setpoints, lacking responsiveness to real-time changes in environmental conditions or occupancy patterns.
- Baseline 2: Predictive Control System (PCS)

  PCS employs forecast data to make anticipatory adjustments in energy system controls. It aims to optimize operations based on expected weather conditions and occupancy but does not adapt in real time to unforeseen changes.

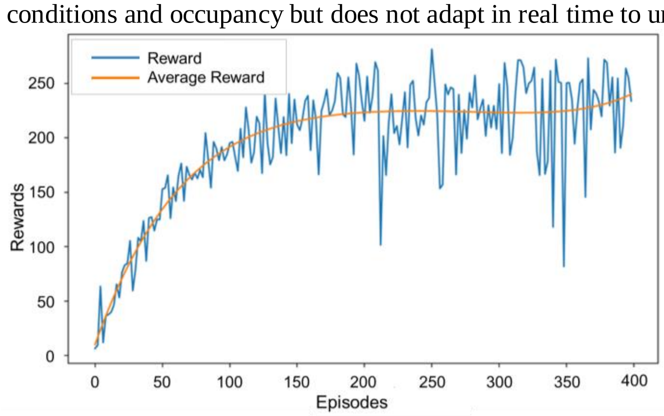
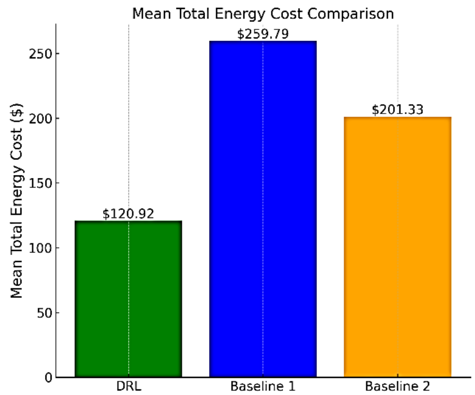
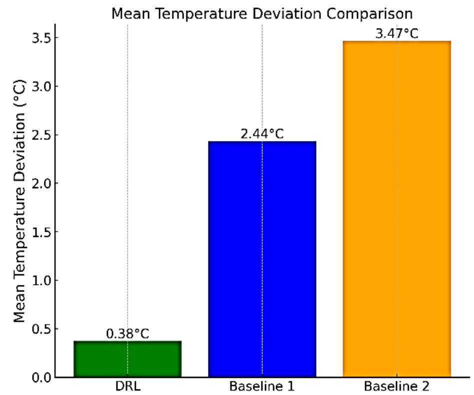


Fig 3. Convergence Analysis

Analyzing the convergence of the proposed DRL model for the BEMS, **Fig 3** indicates a convergence towards an optimal policy over the training episodes. Initially, there's a rapid improvement in rewards, typical of early learning, where significant gains are made as the model identifies better strategies. As training progresses, the average reward trendline shows a gradual plateauing, interspersed with diminishing fluctuations, signifying that the model's explorations yield less dramatic improvements and are settling into a stable policy. The smoothness of the average reward line towards the end of the training episodes suggests that the DRL model has largely converged, finding an equilibrium where the actions chosen consistently lead to higher rewards, which, in the context of BEMS, would correlate with optimized EC and maintained comfort levels. This plateauing trend is a hallmark of convergence in reinforcement learning, indicating that further training is unlikely to result in substantial performance gains, affirming the model's readiness for deployment in a real-world BEMS scenario.



(a)



(b)

Fig 4 (a). Mean total Energy cost, **Fig (b).** Mean temperature deviation

Fig 4 compares the performance of a DRL model with two baseline models in terms of mean total energy cost (**Fig 4 (a))** and mean temperature deviation **(Fig 4 (b)).** For the mean total energy cost, the DRL model is the most cost-effective, averaging $120.92, which is significantly lower than Baseline 1 and Baseline 2, which are $259.79 and $201.33, respectively. This suggests that the DRL model is the most efficient in managing energy consumption, thereby reducing costs. When looking at mean temperature deviation, which is a measure of comfort, the DRL model again outperforms the baselines with a minimal deviation of 0.38 *°C*. Baseline 1 has a higher deviation of 2.44 *°C*, and Baseline 2 is the least effective, with the most significant deviation of 3.47 *°C* from the desired temperature setpoint.

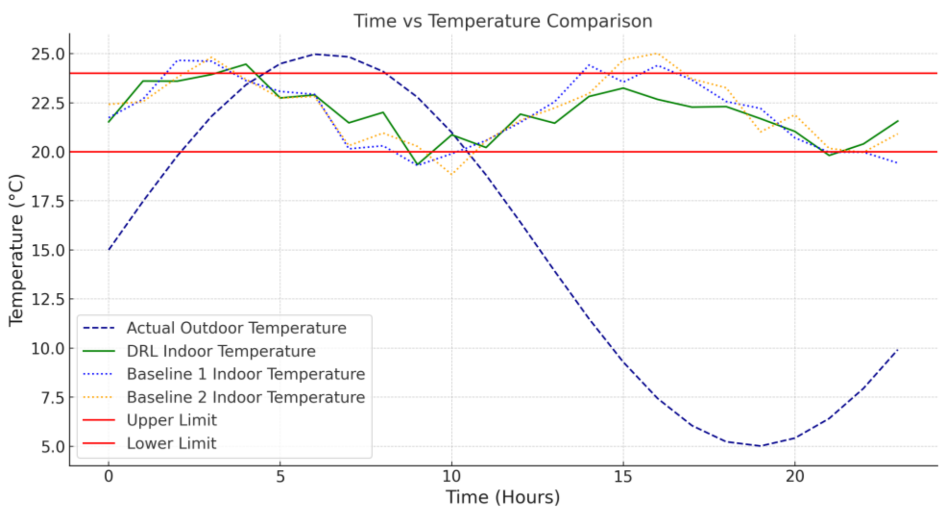


Fig 5. Time vs Temperature

Fig 5 presents the Temperature variation against time for the three compared models over a 24-hour period, where the DRL model adeptly maintains indoor temperature within the comfort zone, set between 20 *°C* and 25 *°C*, demonstrating its effectiveness in adapting to the actual outdoor temperature fluctuations. The DRL model consistently keeps the indoor temperature close to the optimal 22.5 *°C* mark, outperforming Baseline 1, which shows moderate temperature control with occasional peaks just touching the 25 *°C* upper limit, indicating slight deviations from ideal comfort. Baseline 2 exhibits the least effective control, with temperatures swinging beyond the upper and lower comfort thresholds, suggesting potential discomfort for occupants. This analysis underscores the DRL model's advanced capability for ensuring occupant comfort by dynamically responding to external temperature changes, as opposed to the more static nature of the baseline models.

The **Fig 6** shows the HVAC measurement for 24 Hrs. timeframe for the DRL model was compared against measurements from two baseline control systems over a 24-hour period of the day. Throughout the daylight hours and night, the DRL system's

EC boosted to 8 kW, rendering it probably the most practical EC. Baseline 1's energy levels were approximately 8.5 kW, which was slightly more than Baseline 2. However, with peaks hitting 9 kW, Baseline 2 demonstrated the maximum EC.

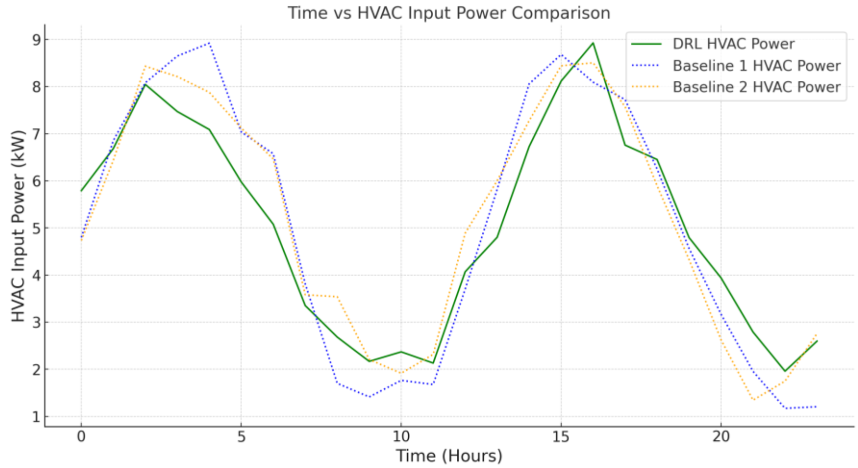


Fig 6. Time vs HVAC

V. CONCLUSION AND FUTURE WORK

The objective of the current study is to analyse the practicality and feasibility of the Deep Reinforcement Learning (DRL) framework within the framework of Building Energy Management Systems (BEMS). A 10,000-square-meter business premises in Hungary was utilised as a location for the research investigation. Energy consumption (EC) and sustaining a suitable temperature are two key variables the computer model considers when computing the BEMS tuning variables. Researchers tested the proposed approach to two standard baseline models: the Predictive Control System (PCS) and the Traditional Rule-Based System (TRS). Minimising energy costs and maintaining the temperature inside within the targeted level of comfort were two areas where the recommended design was performed. Improved and more balanced energy efficiency and comfort for consumers are the outcomes of higher efficiency made feasible by efficient investigation of real-time data and dynamic study. This is how the proposed framework addresses today's issues facing BEMS in an adaptable approach.

Implementing Renewable Energy Sources (RES) and increasing the number of building types will be the key objectives of future work.

Data Availability

No data was used to support this study.

Conflicts of Interests

The author(s) declare(s) that they have no conflicts of interest.

Funding

No funding agency is associated with this research.

Competing Interests

There are no competing interests.

References

[1] T. Kataray et al., "Integration of smart grid with renewable energy sources: Opportunities and challenges – A comprehensive review," Sustainable Energy Technologies and Assessments, vol. 58, p. 103363, Aug. 2023, doi: 10.1016/j.seta.2023.103363.

[2] A. H. Elsheikh et al., "Low-cost bilayered structure for improving the performance of solar stills: Performance/cost analysis and water yield prediction using machine learning," Sustainable Energy Technologies and Assessments, vol. 49, p. 101783, Feb. 2022, doi: 10.1016/j.seta.2021.101783.

[3] S. Durairaj and R. Sridhar, "MOM-VMP: multi-objective mayfly optimization algorithm for VM placement supported by principal component analysis (PCA) in cloud data center," Cluster Computing, vol. 27, no. 2, pp. 1733–1751, Jun. 2023, doi: 10.1007/s10586-023-04040-8.

[4] A. M. Gandhi et al., "SiO2/TiO2 nanolayer synergistically trigger thermal absorption inflammatory responses materials for performance improvement of stepped basin solar still natural distiller," Sustainable Energy Technologies and Assessments, vol. 52, p. 101974, Aug. 2022, doi: 10.1016/j.seta.2022.101974.

[5] P. Chithaluru, F. Al-Turjman, T. Stephan, M. Kumar, and L. Mostarda, "Energy-efficient blockchain implementation for Cognitive Wireless Communication Networks (CWCNs)," Energy Reports, vol. 7, pp. 8277–8286, Nov. 2021, doi: 10.1016/j.egyr.2021.07.136.

[6] “Grid Integration of Renewable Energy Sources using GA Technique for Improving Power Quality,” International Journal of Renewable Energy Research, no. v11i3, 2021, doi: 10.20508/ijrer.v11i3.12292.g8283.
[7] A. Gayathri et al., “Cooperative and feedback based authentic routing protocol for energy efficient <scp>IoT</scp> systems,” Concurrency and Computation: Practice and Experience, vol. 34, no. 11, Feb. 2022, doi: 10.1002/cpe.6886.
[8] R. C. Bheemana, A. Japa, S. S. Yellampalli, and R. Vaddi, “Negative capacitance FETs for energy efficient and hardware secure logic designs,” Microelectronics Journal, vol. 119, p. 105320, Jan. 2022, doi: 10.1016/j.mejo.2021.105320.
[9] P. Thamizharasu., “Revealing an OSELM based on traversal tree for higher energy adaptive control using an efficient solar box cooker,” Solar Energy, vol. 218, pp. 320–336, Apr. 2021, doi: 10.1016/j.solener.2021.02.043.
[10] T. Vino et al., “Multicluster Analysis and Design of Hybrid Wireless Sensor Networks Using Solar Energy,” International Journal of Photoenergy, vol. 2022, pp. 1–8, Oct. 2022, doi: 10.1155/2022/1164613.
[11] B. B. Sharma et al., “Designing and implementing a smart transplanting framework using programmable logic controller and photoelectric sensor,” Energy Reports, vol. 8, pp. 430–444, Nov. 2022, doi: 10.1016/j.egyr.2022.07.019.
[12] L. Sathish Kumar et al., “Modern Energy Optimization Approach for Efficient Data Communication in IoT-Based Wireless Sensor Networks,” Wireless Communications and Mobile Computing, vol. 2022, pp. 1–13, Apr. 2022, doi: 10.1155/2022/7901587.
[13] R. Thirumuru, K. Gurugubelli, and A. K. Vuppala, “Novel feature representation using single frequency filtering and nonlinear energy operator for speech emotion recognition,” Digital Signal Processing, vol. 120, p. 103293, Jan. 2022, doi: 10.1016/j.dsp.2021.103293.
[14] N. P. Singh., “Investigation on characteristics of Monte Carlo model of single electron transistor using Orthodox theory,” Sustainable Energy Technologies and Assessments, vol. 48, p. 101601, Dec. 2021, doi: 10.1016/j.seta.2021.101601.
[15] L. Maddisetti, R. K. Senapati, and R. JVR, “Accuracy evaluation of a trained neural network by energy efficient approximate 4:2 compressor,” Computers & Electrical Engineering, vol. 92, p. 107137, Jun. 2021, doi: 10.1016/j.compeleceng.2021.107137.
[16] B. M L., “Deep Learning-Based Smart Hybrid Solar Water Heater Erection Model to Extract Maximum Energy,” International Journal of Photoenergy, vol. 2022, pp. 1–8, Oct. 2022, doi: 10.1155/2022/2943386.
[17] V. Myilsamy, S. Sengan, R. Alroobaea, and M. Alsafyani, “State-of-Health Prediction for Li-ion Batteries for Efficient Battery Management System Using Hybrid Machine Learning Model,” Journal of Electrical Engineering & Technology, vol. 19, no. 1, pp. 585–600, Jun. 2023, doi: 10.1007/s42835-023-01564-2.
[18] V. Arumugham et al., “An Artificial-Intelligence-Based Renewable Energy Prediction Program for Demand-Side Management in Smart Grids,” Sustainability, vol. 15, no. 6, p. 5453, Mar. 2023, doi: 10.3390/su15065453.
[19] R. Abdulkader et al., “Soft Computing in Smart Grid with Decentralized Generation and Renewable Energy Storage System Planning,” Energies, vol. 16, no. 6, p. 2655, Mar. 2023, doi: 10.3390/en16062655.
[20] A. Lal Karn., “An Empirical Analysis of the Effects of Energy Price Shocks for Sustainable Energy on the Macro-Economy of South Asian Countries,” Energies, vol. 16, no. 1, p. 363, Dec. 2022, doi: 10.3390/en16010363.
[21] A. Lal Karn, P. Selvam Manickam, R. Saravanan, R. Alroobaea, J. Almotiri, and S. Sengan, “IoT Based Smart Framework Monitoring System for Power Station,” Computers, Materials & Continua, vol. 74, no. 3, pp. 6019–6037, 2023, doi: 10.32604/cmc.2023.032791.
[22] A. Rajagopalan et al., “Modernized Planning of Smart Grid Based on Distributed Power Generations and Energy Storage Systems Using Soft Computing Methods,” Energies, vol. 15, no. 23, p. 8889, Nov. 2022, doi: 10.3390/en15238889.
[23] S. Sengan, S. V, I. V, P. Velayutham, and L. Ravi, “Detection of false data cyber-attacks for the assessment of security in smart grid using deep learning,” Computers & Electrical Engineering, vol. 93, p. 107211, Jul. 2021, doi: 10.1016/j.compeleceng.2021.107211.
[24] K. S. Kumar, B. P. Esther, V. Indrgandhi, S. Sudhakar, L. Ravi, and V. Subramaniyaswamy, “Area Based Efficient And Flexible Demand Side Management To Reduce Power And Energy Using Evolutionary Algorithms,” Malaysian Journal of Computer Science, pp. 61–77, Nov. 2020, doi: 10.22452/mjcs.sp2020no1.5.