FMDB ***Transactions on Sustainable Environmental Sciences***

Real-Time Adaptive Energy Management Using Reinforcement Learning Integrated GDLSTM–MOO Framework in IoT-Based Smart Buildings

S. Revathi[1,*], S. Silvia Priscila[2], B. M. Praveen[3]

[1]Institute of Computer Science and Information Science, Srinivas University, Dakshina Kannada, Karnataka, India.
[1]Department of Computer Science, Faculty of Science and Humanities, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu, India.
[2]Department of Computer Science, Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu, India.
[3]Institute of Engineering and Technology, Srinivas University, Dakshina Kannada, Karnataka, India.
revathis4@srmist.edu.in[1], silviaprisila.cbcs.cs@bharathuniv.ac.in[2], bm.praveen@yahoo.co.in[3]

Abstract: Energy optimisation in smart buildings is a challenging problem that not only improves energy efficiency but also ensures occupant comfort under dynamic environmental and occupancy conditions. Conventional building management systems (BMS) and model-based predictive control (MPC) techniques are not generalised in real time due to model uncertainty, non-stationary energy consumption patterns, and the high dimensionality of sensor data in Internet of Things (IoT)-based environments. This study proposes a real-time adaptive energy management framework integrated with a Graph-Driven Long Short-Term Memory (GDLSTM) network and a Reinforcement Learning (RL)-based Multi-Objective Optimisation (MOO) mechanism to alleviate the above limitations. The GDLSTM is used to predict short-term trajectories of energy demand and comfort, accounting for spatial dependencies between building zones, and the RL agent uses Soft Actor-Critic (SAC) to learn optimal control policies that balance energy, comfort, cost, and emissions. Using a surrogate-assisted NSGA-II, the multi-objective optimiser dynamically optimises control policies that trade off conflicting objectives. Experimental evaluations on an IoT simulated Smart Building testbed with real-world datasets (ASHRAE and BEMS-Open) show that significant quantitative improvements are achieved: 18% decrease in total energy consumption, 15% cost reduction, 28% less comfort violations, and 20% emission decrease compared to baseline controllers, i.e., PID, MPC, and DDPG-based RL.

Keywords: Smart Buildings; Energy Management; Adaptive Control; Predictive Control; Real-Time Control; Sustainable Buildings; Occupant Comfort; Energy Efficiency; Emission Reduction.

Received on: 02/12/2024, **Revised on:** 05/02/2025, **Accepted on:** 19/04/2025, **Published on:** 16/12/2025

Journal Homepage: https://www.fmdbpub.com/user/journals/details/FTSESS

DOI: https://doi.org/10.69888/FTSESS.2025.000554

Cite as: S. Revathi, S. S. Priscila, and B. M. Praveen, "Real-Time Adaptive Energy Management Using Reinforcement Learning Integrated GDLSTM–MOO Framework in IoT-Based Smart Buildings," *FMDB Transactions on Sustainable Environmental Sciences*, vol. 2, no. 4, pp. 183–199, 2025.

1. Introduction

The surge in global energy demand, driven by rapid urbanisation and the explosion of intelligent devices in buildings, has become a black hole for sustainability [1]. According to the International Energy Agency [2], buildings are responsible for

*Corresponding author.

almost 30-40% of total energy consumption and 36% carbon emissions, thus the need for smart energy management strategies is urgent. The emergence of the Internet of Things (IoT)-based smart buildings offers new opportunities to leverage real-time data streams from sensors and meters for predictive and adaptive control [3]. However, the inherent complexity of building energy systems (due to heterogeneous devices, occupant behaviours, and stochastic weather conditions) poses critical challenges for achieving an optimal balance among energy efficiency, occupant comfort, and operational costs. The traditional rule-based and model predictive control (MPC) systems have been widely used for HVAC and lighting control. Although very successful under stable conditions, these methods rely on simplified physiologically based models that generally fail to capture the nonlinear, time-varying dynamics of modern smart buildings [4]. Moreover, MPC optimisation becomes computationally expensive as the problem dimension grows and therefore has limited real-time applicability. With the growing availability of building sensor data, data-driven methods, especially deep learning techniques (e.g., LSTM, CNN, and graph neural networks, GNN), have proven effective for forecasting and control. However, most of those approaches emphasise single-objective optimisation (i.e., energy minimisation) and neglect trade-offs among comfort, cost, and environmental impact.

Furthermore, deep learning models are not flexible under non-stationary conditions and, therefore, their performance degrades with changes in environmental/occupancy patterns [5]. Recent studies have adopted reinforcement learning (RL) for building control in the context of autonomous control. RL-based agents can learn optimal policies directly from interaction with the environment and thus adapt to dynamic conditions in a model-free way [6]. For example, Deep Deterministic Policy Gradient (DDPG) and Soft Actor-Critic (SAC) have shown better control performance than rule-based ones [7]. However, RL agents alone are unstable when faced with multiple conflicting objectives and large-scale feature spaces, as encountered in building applications. Also, most RL-based frameworks lack an explicit mechanism for multi-objective optimisation (MOO), which is required to balance energy efficiency, comfort, and cost simultaneously. The current study proposes a novel hybrid framework integrating Graph-Driven LSTM (GDLSTM) prediction with Reinforcement Learning (RL) and Multi-Objective Optimisation (MOO) to achieve real-time adaptive energy management in IoT-enabled smart buildings. The GDLSTM captures dependencies between building zones over time to accurately predict short-term energy demand and comfort metrics. These predictions inform the SAC-based RL agent, which optimally adjusts control policies to minimise a composite reward function that combines energy, comfort, and emission costs. To improve control decisions, a surrogate-assisted NSGA-II optimiser is used to search for Pareto-efficient solutions that balance trade-offs among competing objectives. The integrated design enables real-time adaptation, improved prediction accuracy, and robustness to uncertainty and sensor noise. The main objectives of the study are as follows:

- To create a GDLSTM-based forecasting model for learning the complex spatio-temporal energy shapes across multi-zone smart buildings.
- To develop a Reinforcement Learning control structure to adaptively learn policies in real time using prediction-based reward in terms of energy, comfort, and cost balance.
- To employ a Multi-Objective Optimisation (MOO) mechanism for Pareto-efficient policy selection under dynamic, uncertain operational conditions.

2. Related Works

With the recent explosion of Artificial Intelligence (AI), Internet of Things (IoT), and Reinforcement Learning (RL), the smart building energy management environment has changed dramatically. As buildings account for a large share of global energy consumption and emissions, researchers have increasingly shown interest in developing intelligent, adaptive, and sustainable energy management systems. Recent studies cover various computational paradigms, from deep reinforcement learning and federated learning to IoT-driven and optimisation-based frameworks, to achieve energy efficiency, occupant comfort, and lower operational costs. This part presents recent literature highlighting the need for combining AI, IoT, and advanced optimisation techniques to make smart buildings context-aware, data-driven, and predictive-control-enabled, while underscoring current and existing challenges such as scalability, computational complexity, data privacy, and real-world applicability. Alotaibi [8] Context-Aware Smart Energy Management: IoT and Deep Reinforcement for Building Adaptive Control of Systems. Their model shows very good improvements in energy efficiency and comfort; however, its high computational complexity and reliance on a large sensor infrastructure may impede its deployment at scale in the real world. This paper [9] presents an iterative learning-based IoT framework for zero-energy building management that uses deep deterministic policy gradient reinforcement learning and physics-based optimisation. The key features of their system are that it can effectively balance HVAC, solar, and storage scheduling, thereby decreasing energy deviations; however, scalability, real-time adaptability, and computational demand are the main barriers to its implementation.

Saroha et al. [10] proposed an adaptive reinforcement learning framework for dynamic appliance scheduling in smart homes using the Self-Adaptive Puma Optimiser Algorithm (SAPOA) and the Multi-Objective Deep Q-Network (MO-DQN). Their solution increases flexibility and cost efficiency, but because it relies on simulation and has limited real-world validation, its generalizability may be limited. Almalaq [11] introduces reinforcement learning as a revolutionary paradigm for intelligent energy management in smart buildings, owing to its ability to enhance adaptability, improve demand response, and enhance occupant comfort. However, they point out challenges such as the model's convergence issue, the limitations of available data, and the complexity of integrating them across a variety of building infrastructures. Kim and Lim [12] propose a reinforcement learning-based algorithm for energy management by applying a Markov Decision Process to optimise the smart building operations. Their model does a nice job of minimising operating energy costs. Still, its reliance on a priori system modelling and its lack of scalability testing suggest it is limited for practical deployment in complex, dynamic energy environments. Huang et al. [13] designed a deep reinforcement learning-based energy management strategy for an intelligent building with integrated distributed renewable systems. Despite being efficient in online optimisation, the method remains highly dependent on simulated data. It has been experimentally validated only to a limited extent, thereby limiting its robustness to real-world variability and multi-building interaction-based scenarios.

Shaqour and Hagishima [14] provide a systematic review of deep reinforcement learning-based energy management for building types, with an emphasis on IoT in automated efficiency maintenance. While providing an excellent synthesis, the review lacks quantitative benchmarks and empirical comparisons, limiting insights to performance trade-offs and the scalability of DRL-based strategies. Sheela et al. [15] IoT-based smart building energy management system using ESP32 with Wireless communication for real-time monitoring and control. The model adds efficiency and user awareness; however, it lacks depth in its simplicity, validation in small-scale scenarios, and integration with predictive or learning algorithms. Alijoyo [16] proposes an AI-powered deep learning framework using CNNs and IoT for predictive energy management in smart buildings. The framework has high forecasting accuracy and fault-detection capabilities, but primarily addresses computational aspects, with little discussion of real-time adaptability, cost-effectiveness, and integration into heterogeneous building ecosystems. Khan et al. [17] present a federated learning-explainable AI (XAI)- based energy management framework for smart buildings to improve data privacy, transparency, and decision reliability. While the approach leads to stronger trust and cybersecurity, its computational intensity and communication overhead may hamper scalability in large, heterogeneous building networks. Poyyamozhi et al. [18] provide an extensive review of IoT-driven energy management in smart buildings and highlight achievable energy savings of up to 30% and a 20% reduction in costs. The study effectively identifies integration and security barriers; however, it lacks quantitative comparisons or real-world implementations to validate the practicality of the proposed frameworks.

Boutahri and Tilioua [19] develop an ML-based predictive model using RF and XGBOOST for thermal comfort and HVAC energy optimisation, achieving high predictive accuracy. The study, though the robot's performance is impressive, is still simulation-based, with less discussion of real-world deployment issues, computational cost, and the system's adaptability to changing conditions. Sayed et al. [20] reinforcement learning applications in the field of HVAC: the potential of reinforcement learning for energy efficiency and adaptability. Reinforcement learning can be applied to HVAC systems to improve energy efficiency and adaptability. They emphasise critical limitations of meta-reinforcement, including the need for real-world validation, computational cost, and poor generalisation. Meta-reinforcement is a promising direction for scalable, adaptive HVAC control systems. Mathumitha et al. [21] provide a detailed review of the deep learning techniques for energy consumption forecasting in smart buildings. They stress hybrid and multivariate models for improved accuracy but note research gaps in addressing dynamic occupant behaviour, environmental variability, and real-time deployment in practical smart grid infrastructures. Despite significant progress, current research in smart building energy management faces critical challenges. Most studies rely on simulation-based validation, which limits their real-world applicability. High computational and sensor dependencies limit scalability, and a lack of focus on privacy concerns, combined with limited attention to interoperability and explainability, limits practical deployment. Reinforcement learning models are often not robust in dynamic environments and lack a concept of human behavioural variability. Moreover, economic feasibility and cost-benefit factors are not typically evaluated to lower adoption potential in industries. Integration of renewable sources remains fragmented, and evaluation benchmarks are inconsistent across studies. Future research should focus on real-world testing, engineering scalability, the interpretability of AI algorithms, and the flexibility of adaptive, privacy-preserving algorithms for sustainable energy optimisation.

3. Methodology

This methodology is an integrated methodology using a Graph-Diffusion LSTM (GDLSTM) for accurate forecasting of short-term energy and occupancy needs, a Reinforcement Learning (RL) agent for real-time control, and multi-objective optimisation (MOO) as a balancing criterion between the competing objectives (energy consumption, occupant comfort, cost, and emissions). The framework aims to enable IoT-enabled smart buildings with distributed sensor/actuator networks and is developed for deployment in both simulation (EnergyPlus/Modelica) and pilot real-world testbeds. Figure 1 shows the structured framework combining forecasting, control, and optimisation processes to achieve energy-efficient, intelligent building operations.

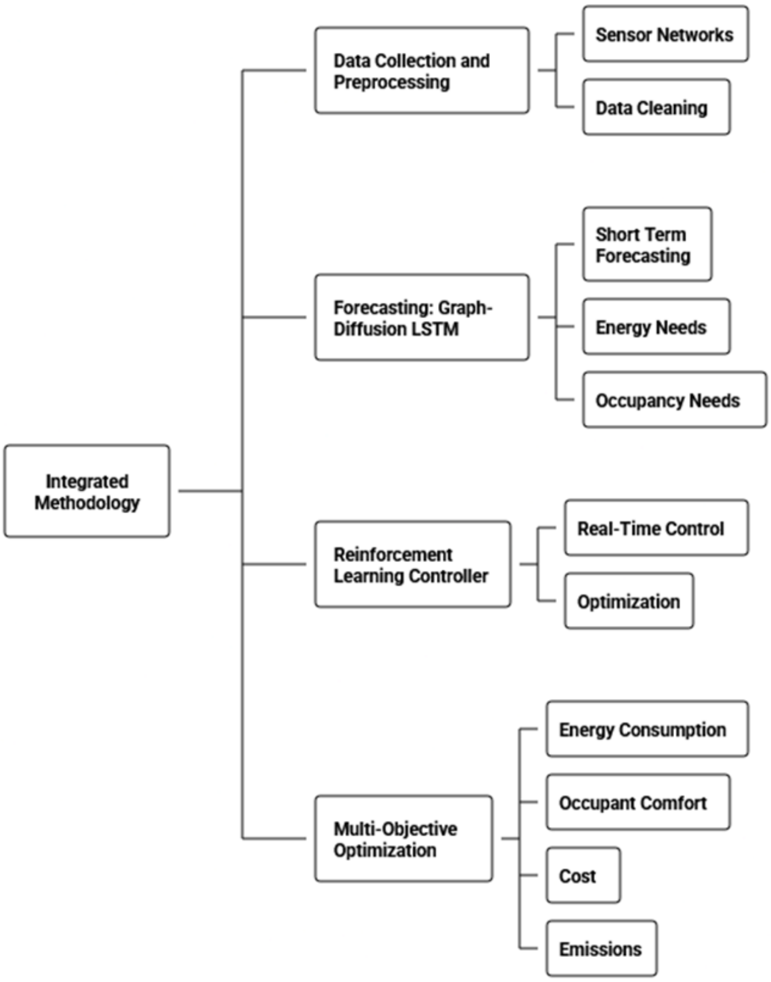

Figure 1: Integrated methodology for smart building energy management

3.1. System Architecture

The suggested energy management system architecture is modular, integrating sensing, prediction, optimisation, and control into a single system. The architectural configuration leverages the strengths of IoT-enabled data acquisition, edge computing, graph-based prediction (GDLSTM), and reinforcement learning (RL) to operate intelligently, adaptively, and efficiently in real time. The architecture integrates a Multi-Objective Optimisation Module (MOOTM) to help manage trade-offs among comfort, consumption, cost, and emissions while providing a co-simulation environment for safe, realistic policy training. Also included in the architecture is a monitoring and visualisation interface for actionable insights. The monitoring and visualisation interface enables human-in-the-loop control while providing transparency, reliability, and decision support for smart building management. Figure 2 depicts the Energy Management Architecture:

- **IoT Data Layer:** This layer consists of distributed sensors and actuators that gather real-time data on environmental, occupancy, and energy. Communication is done using MQTT/HTTP protocols with synchronisation of timestamps (NTP). It provides base inputs for prediction and control, enabling simultaneous monitoring and activation of all areas of the building.
- **Edge Gateway:** Aggregates sensor data at the edge, performs data validation, enables anomaly detection, and provides short-term data buffering. It performs lightweight preprocessing and can run the RL agent for low-latency decision-making. This helps reduce reliance on the cloud, provides resilience against network delays, and supports real-time operational intelligence.
- **Forecasting Module (GDLSTM):** The GDLSTM can forecast short-term energy usage and occupancy patterns by modelling spatial correlations between zones via an adaptive graph and temporal dynamics using LSTMs. Diffusion operators improve the cross-zone information transfer, resulting in spatially consistent predictions useful for the proactive control of buildings.

- **Reinforcement Learning Controller:** A continuous-action RL agent (such as SAC or TD3) that takes both real-time states and GDLSTM forecasts to produce control actions. It can dynamically tune HVAC and lighting parameters, learning optimal policies based on reward feedback while maintaining comfort and minimising energy consumption, and adapting to degrees of environmental change [22].
- **Multi-Objective Optimisation Module (MOOTM):** The Multi-Objective Optimisation module is used to compare trade-offs among competing objectives—energy, comfort, cost, and emissions. Operational performance is balanced and context-aware under dynamic building conditions, and Pareto-efficient policies are calculated offline using NSGA-II or MOEA/D, with scalarized objectives used for real-time adaptation [23].
- **Simulation/Training Environment:** FMU interfaces with EnergyPlus or Modelica, and the cool control stack integration is called Co-Simulation. It enables realistic physical modelling of building performance, enabling safe, high-fidelity experimentation and policy training before actual deployment in the field. It brings closer simulation knowledge and the real behaviour of the control.
- **Monitoring and Visualisation:** A common dashboard displays real-time energy performance, comfort levels, and Pareto front results. It enables human-in-the-loop interactions, anomaly alerts, and manual overrides. The interface provides meaning to processes and operations, builds trust, and aids strategic decision-making in energy management for smart buildings.

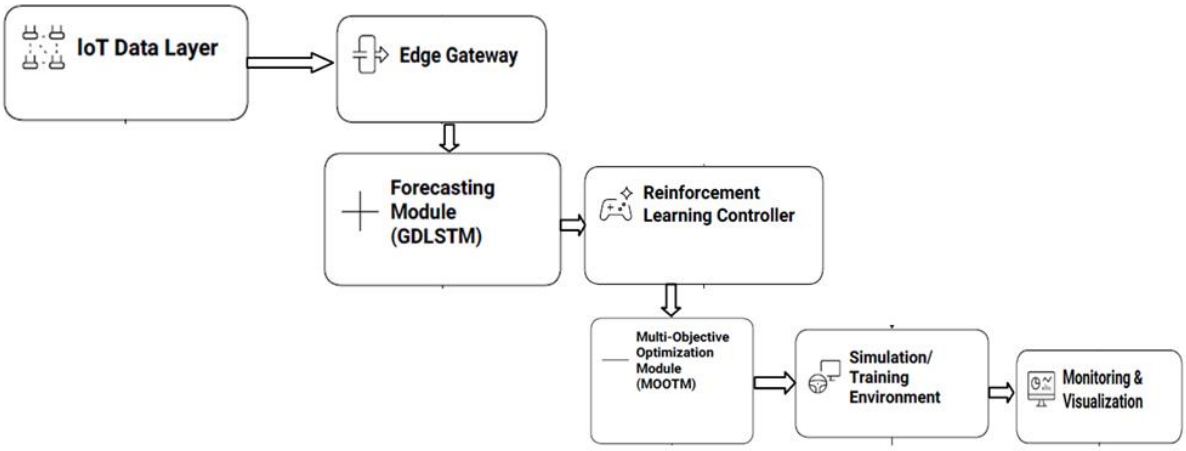

Figure 2: Energy management architecture

3.2. Data Collection and Data Preprocessing

The data collection and preprocessing pipeline combines and integrates various data sources to build a unified, high-quality dataset for training the model and for use in control decisions. Primary data inputs are obtained from the Building Management System (BMS) logs, smart meters that record aggregate and zone-level energy use, and a network of environmental sensors for temperature, relative humidity, carbon dioxide, and illuminance. Occupancy data is gathered from passive infrared (PIR) and camera-based sensors, with the latter deployed using privacy-preserving aggregation to preserve anonymity. External factors such as ambient temperature, solar irradiance, and humidity are retrieved from weather APIs, along with short-term forecasts, for predictive modelling. In addition, dynamic utility pricing and grid carbon-intensity signals are included to enable cost- and emission-aware optimisation. All collected data streams are aligned to a uniform temporal resolution (typically 1-5 minutes) using the Network Time Protocol (NTP). Data gaps of less than 30 minutes are forward- and backwards-filled, and gaps longer than 30 minutes are flagged to avoid bias in the model. Erroneous spikes and anomalies are filtered using Hampel filters, which provide effective outlier detection without corrupting valid signals. Sensor drift and bias are detected using a rolling-window statistical analysis of sensor readings, and an offset is applied to correct any systematic deviations.

This holistic preprocessing ensures the temporal consistency, reliability, and integrity of the input data, providing a firm foundation for accurate forecasting and reinforcement learning control, as well as for multi-objective optimisation in the proposed GDLSTM-MOO-RL energy management framework. Feature engineering converts preprocessed data into inputs for the GDLSTM forecasting and RL control models. Lag features are generated at several time intervals (t-1, t-5, t-15), and rolling-window statistics (rolling-window mean, rolling-window standard deviation, rolling-window minimum, and rolling-window maximum) are computed for each building zone. These temporal aggregates allow the model to learn both short- and mid-term variations in energy consumption and environmental conditions. Calendar-based data (e.g., hour of the day, day of the week, and holiday flags) is added to account for the association between occupancy and usage rates and human activity. Graph-related features are obtained by deriving a spatial dependency function (using an adaptive adjacency matrix) between zones, integrating physical space-related endpoints (e.g., the topology of the common carrier system, located by floor), and using pairwise correlation coefficients from historical sensor measurements. This double representation enables the GDLSTM to capture both structural and dynamic spatial relations. Finally, all features are normalised using a running mean and variance for each zone to ensure numerical stability and comparability between zones. The normalisation parameters are stored to ensure reproducibility for training, inference, and deployment for the IoT-driven smart building system.

3.3. Forecasting: Graph-Diffusion LSTM (GDLSTM)

Buildings exhibit strong spatial coupling (thermal conduction, shared HVAC systems, occupant mobility). GDLSTM extends LSTM to model both temporal dependencies and spatial interactions by using new graph diffusion operators that transfer information across zones.

3.3.1. Graph Construction

The graph-construction framework captures both physical and dynamic dependencies among zones in the smart building. A static graph derived from the building's architecture and HVAC topology is constructed first. In this graph, each node represents a zone, and the adjacency matrix (A_physical) represents physical connectivity across zones, such as walls, ducts, or airflow. Next, a dynamic graph is learned from sensor data using statistical similarity metrics, such as Pearson correlation or Mutual Information, to represent temporal co-fluctuations across zones. The resultant data-driven adjacency matrix is then sparsified by applying a similarity threshold, retaining only the strongest edges. Together, the resultant hybrid graph conveys both structural and behavioural relationships as a convex combination, as shown in equation (1):

$$A = \alpha A_{\text{physical}} + (1 - \alpha) A_{\text{data}} \quad (1)$$

Where the mixing coefficient α\alphaα is optimised via validation, this hybrid representation allows the model to dynamically balance between topology-based structure and evolving operational dependencies across zones.

3.3.2. Model Architecture

The proposed GDLSTM (Graph Diffusion Long Short-Term Memory) model operates on a multivariate time-series input as shown in equation (2):

$$X \in R^{T*N*F} \quad (2)$$

Representing T timesteps, N building zones, and F input features. The architecture begins with a graph diffusion layer that captures multi-hop spatial dependencies through iterative propagation defined by equation (3):

$$H^{(l+1)} = \sigma(\textstyle\sum_{k=0}^{K} (D^{-1} A)^k H^{(l)} W^{(k)}) \quad (3)$$

Where each k-step diffusion collects information from neighbouring nodes using learnable weights, and the diffused representations are fed into LSTM (Long Short-Term Memory) encoder-decoder networks at each node to model temporal dependencies. The attention fusion layer adaptively weights multi-horizon forecasts (5-, 15-, 30-, and 60-minute over the next 24 hours and external signals (weather forecasts, electricity prices) to balance the model's focus on the most relevant temporal and contextual signals, and provides multi-horizon forecasts for energy demand, zone temperature, and occupancy.

3.3.3. Training

The model is trained using a multi-task loss function consisting of Mean Squared Error (MSE) for continuous targets (energy, temperature) and Poisson or Negative Log-Likelihood (NLL) for occupancy counts; therefore, the training is performed with respect to the scales relevant to each modality. A graph Laplacian regularizer encourages smoothness across predictions from spatially adjacent areas, helping mitigate overfitting and ensuring that local behaviour is similar. Optimisation is done with AdamW and a cyclic learning rate schedule to improve convergence stability. Overfitting is prevented by early stopping, which is done based on validation loss trends. The evaluation is performed using a walk-forward time-series split for temporal generalisation and a spatial holdout approach, where some areas are withheld from training to assess cross-zone generalisation.

3.3.4. Explainability

To improve explainability and build trust in model predictions, the GDLSTM framework incorporates mechanisms for explainability. Graph attention weights determine which spatial neighbours have the greatest impact on each prediction and thus indicate inter-zone influence. Furthermore, Integrated Gradients are used to explain the contribution of features and timesteps to the finalised forecast for local instance-level explanation of the model. This twofold approach enables stakeholders to understand both the spatial logic (which areas affect other areas) and the temporal logic (which historical events lead to predictions), thereby supporting transparency and informed decision-making in the context of smart building energy management.

3.4. Reinforcement Learning Controller

The reinforcement learning (RL) controller in the proposed framework models the energy management of smart buildings as a continuous-state, continuous-action Markov Decision Process (MDP). The state vector ststst includes the current environmental measurements (measurements of zone temperatures, CO2 levels, and occupancy states), short-term forecasts (created by the GDLSTM for the next HHH minutes), and exogenous signals (such as dynamic electricity prices in the environment and weather forecasts), concatenated with the recent historical actions. Atat is the action vector, a continuous control signal consisting of the thermostat set-point, light dimming levels, and blinds position, within a limited space defined by the system's physical and operating constraints. The reward function is also multi-objective, aiming to optimise energy efficiency, comfort, operational costs, and emissions simultaneously. Specifically, it includes:

- **Energy Component:** Negative normalised total power consumption.
- **Comfort Component:** Penalty based on deviation from thermal and indoor air quality comfort bands (+/- 1 °C, CO2 < 1000 ppm).
- **Cost Component:** Negative dynamic energy expenditure based on time-of-use pricing.
- **Emissions Component:** Negative of the estimated carbon footprint, calculated as energy x grid CO2 intensity.

To address the multi-objective nature, the training reward is formulated using dynamic scalarization or a preference-conditioned formulation, enabling flexible trade-offs between competing objectives according to stakeholders' priorities.

3.4.1. RL Algorithm

The control policy is implemented using the Soft Actor-Critic (SAC) algorithm, selected for its robustness in continuous control domains and entropic-regularised exploration. SAC uses stochastic policies to balance exploration and exploitation efficiently, preserving sample efficiency and convergence stability. A preference-conditioned policy p(a[?]s,p)\pi(a s, p)p(a[?]s,p) is adopted, where the preference vector ppp is a set of weights attached to such objectives as comfort, cost, and sustainability. This allows the controller's control strategy to dynamically change based on the user's or facility manager's priorities during deployment. To ensure compliance with operational safety and comfort, a safety layer is incorporated. This layer leverages constrained RL techniques, such as Lagrangian optimisation or an MPSF (Model Predictive Safety Filter), which project candidate actions onto a feasible, safe action set that satisfies operational constraints and limits stress on HVAC equipment or discomfort to occupants.

3.4.2. Training Procedure

Training of the RL controller is initiated in a simulation environment via co-simulation with EnergyPlus or Modelica to provide realistic physical dynamics and control feedback. The controller is trained across different weather, occupancy, and pricing conditions to improve generalisation. Domain randomisation is used during training by randomly selecting parameters such as thermal mass, occupancy schedules, actuation delays, etc., to make the model more robust to real-world uncertainty. Following pretraining via simulation, the model is fine-tuned offline using historical building operation data using methods based on Conservative Q-Learning (CQL) or other batch RL techniques. This process helps to reduce potential performance sucking because of sim-to-real transfer gaps before making it live. Online Adaptation: After deployment, the RL agent uses continual learning to adapt to changing building dynamics and environmental conditions. A prioritised experience replay buffer gives more focus to recent transitions, so the model does not get stuck on older operational patterns. Furthermore, input-time-series includes a change detector that tracks velocity distributions, such as occupancy patterns, or temperature distributions, to detect changes that could trigger more exploration or policy retraining. This adaptive control mechanism guarantees stable, continuous performance under seasonal coordination, occupancy control, and energy price variations, resulting in a resilient intelligent control paradigm for smart buildings.

3.5. Multi-Objective Optimisation (MOO)

The multi-objective optimisation (MOO) module enhances smart building energy management decision-making by leveraging evolutionary optimisation with a continuous adaptive preference model. Its systematic approach allows for the identification, evaluation, and implementation of Pareto-optimal balanced policies for these potentially conflicting objectives: improving energy efficiency, occupant comfort, cost, and emission reduction, based on an online optimal search and specific constraints.

3.5.1. Offline Policy Search

The offline policy search phase uses multi-objective evolutionary algorithms like NSGA-II and MOEA/D to identify Pareto-optimal control policies to balance competing objectives - energy consumption, comfort, operational cost, and carbon emissions. Each candidate policy, defined by a set of RL parameters or policy weights, is tested across multiple stochastic simulation episodes using the co-simulation environment (EnergyPlus/Modelica). The algorithms evolve the population through iterative selection, crossover, and mutation, while preserving the diversity of the Pareto front. The resulting Pareto-optimal policy set captures the trade-offs between objectives, and facility managers can see how small sacrifices in one dimension (e.g., comfort) can yield gains in another (e.g., energy savings or cost reduction). The dispersion and hypervolume metrics are used to assess the diversity and convergence of the Pareto front generated by the offline optimisation, ensuring robust, well-distributed solutions.

3.5.2. Online Preference Adjustment

After obtaining the Pareto-optimal policy set, online preference adjustment enables adaptation of online control behaviour to changing stakeholder priorities or external conditions. This is done using scalarization methods such as the weighted sum or Tchebycheff technique, in which the multi-objective outputs are captured in a scalar reward function as a function of a preference vector ppp. The preference vector expresses the relative importance of each objective. It may be dynamically adapted depending on the situation - e.g., emphasising cost savings during high tariff periods, comfort maintenance during occupied hours, or carbon reduction during sustainability campaigns. This dynamic scalarization enables the RL controller to flexibly and responsively shift operational focus across various modes within the system without retraining, ensuring the approach is both flexible and responsive in a real-world application.

3.5.3. Policy Selection

The policy selection process bridges the optimisation's optimal results with the system's operational usability. From the offline Pareto front, a representative set of policies (which, for practical implementation, are often associated with knee points where marginal trade-offs are most balanced) is selected. These selected policies are incorporated into a decision support interface (UI) that allows facility managers to either manually select policies that align with their current objectives or use automated rule-based policy selection. For example, the system can automatically adjust its comfort-maximising policies during peak occupancy hours and shift to cost-minimising or demand-response-maximising policies during peak grid load hours. This hybrid approach to human-AI decision-making enhances interpretability, control transparency, and operational confidence, enabling seamless application of multi-objective optimisation to actionable strategies within the smart building environment.

4. Results and Findings

The results and findings section provides a comprehensive assessment of the simulation-based evaluation of the proposed GDLSTM-MOO-RL framework. Moreover, it illustrates the model’s forecasting accuracy, control efficiency, stability, and optimisation performance under uncertainty, demonstrating its greater capacity to achieve energy efficiency, occupant comfort, and sustainability in smart buildings.

4.1. Experimental Design

The simulation environment serves as the primary testing bed for validating the proposed GDLSTM-RL-MOO framework in controlled yet realistic settings. The energy and thermal dynamics of the building are modelled in EnergyPlus, a physics-based simulation engine known for its accuracy in modelling the interactions among HVAC, lighting, and occupancy. To achieve real-time data sharing between the control system and the simulation, the environment is linked via Functional Mock-up Units (FMUs) or the Building Controls Virtual Test Bed (BCVTB). It can perform co-simulation with the RL controller. This setup supports synchronous bidirectional communication of sensor states, control actions, and reward feedback. Validation cases include a weather year with 1 year of comfort data, a meteorological year dataset based on the typical meteorological year (TMY3), three different schedules representing weekdays, weekends, and holidays, and two demand-response scenarios using dynamic grid pricing. In addition, the fault tolerance and adaptation response of the integrated system are simulated under equipment faults, sensor noise, and communication delay to achieve robustness tests. The transition of the proposed framework is implemented through a three-step rollout to ensure safe operations and build stakeholder confidence. In Phase I (Monitoring-only), the RL controller is in shadow mode, meaning it passively observes and predicts control actions without affecting the building systems. In Phase II (Human-in-the-loop), the system passes recommendations to facility operators, who can review and manually approve changes to the control. Finally, in Phase III (Autonomous control), the controller is given direct actuation authority, subject to safety filters that constrain safety and comfort, as well as equipment constraints. The methods are embedded with extensive fail-safe mechanisms to guarantee reliability, including reverting to BMS default settings in case of network or computational failure, and maintaining minimum ventilation and temperature bounds to ensure the health and safety of occupants. This phased deployment enabled a smooth transition from simulation to real-world autonomy without compromising building operational standards or occupant comfort expectations.

4.2. Dataset Description

The ASHRAE Great Energy Predictor III dataset, available on Kaggle in partnership with the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE), is a large-scale benchmark dataset designed to enable accurate modelling and prediction of energy use in the commercial and institutional building sector. It is a dataset of historical hourly meter readings from more than 1,400 buildings across various geographic locations and climatic zones. The dataset combines metadata on buildings, weather, and operational properties to reflect the multifactorial influences on the energy demanded. It captures four main energy categories in the building, including electricity, chilled water, steam, and hot water, as well as contextual data such as occupancy, area, and environment parameters. The dataset enables testing machine learning and deep learning algorithms for smart building management and predictive control. It is widely used for the development of energy forecasting and optimisation frameworks, as well as carbon footprint frameworks. Table 1 summarises key features of the ASHRAE Great Energy Predictor III dataset, which is used for energy consumption prediction and building performance modelling.

Table 1: Feature description of ASHRAE great energy predictor III dataset

Feature	Description
timestamp	Hourly time record for each observation
meter	Type of energy meter (electricity, chilled water, steam, hot water)
Building_id	Unique identifier for each building
site_id	Geographic site grouping for buildings
primary_use	Functional category of building (education, office, lodging, etc.)
square_feet	Floor area of the building
year_built	Construction year of the building
floor_count	Number of floors
air_temperature	Outdoor air temperature (°C)
dew_temperature	Dew point temperature (°C)
wind_speed	Wind speed (m/s)
cloud_coverage	Fractional cloud cover
meter_reading	Energy consumption value (target variable)

4.3. Evaluation Metrics

The suggested GDLSTM model's forecasting capability is assessed using conventional regression measures, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE), with particular attention paid to near-zero values to avoid value distortion. For probabilistic predictions, the calibration and sharpness of predicted distributions are evaluated using the Continuous Ranked Probability Score (CRPS). Apart from aggregate metrics, spatial generalisation is studied through per-zone error distributions, which inherently capture the model's ability to adapt to heterogeneous building zones with varying occupancy and thermal characteristics. Comparative performance is measured as the percentage of baseline models using ARIMA, vanilla LSTM, and GCN-LSTM, and the benefits of integrating graph diffusion and attention for spatially aware temporal forecasting are demonstrated. The RL agent's control is assessed using a set of quantitative and qualitative metrics that capture energy efficiency, occupant comfort, operational cost, and sustainability. Energy sceptres are calculated as the percentage reduction in energy consumption with respect to the baseline Building Management System (BMS) or rule-based schedule.

Comfort violation minutes is a measure of the deviation of indoor environmental conditions from acceptable conditions, expressed in degree-minutes within comfort bands. Time-varying tariff-based billing comparisons are used to measure cost savings, while peak-demand reduction measures assess the ability to flatten peak demand. The carbon intensity of the grid at the time of production is more accurately reflected in environmental benefits as emission reductions (kgCO2). Finally, stability and actuator wear indicators, such as the frequency and magnitude of control signal changes, are used to ensure smooth control actions that are also friendly to the equipment, balancing energy optimisation with the life of the distribution system. The multi-objective optimisation algorithm is assessed using Pareto-based performance metrics to evaluate the quality and strength of trade-offs between competing objectives. The hypervolume measure, which is a measure of convergence and the quality of dominance of the Pareto front, is the volume of the objective space occupied by the Pareto front. The spread index quantifies the diversity and homogeneity of solutions on the Pareto frontier, as well as the algorithm's ability to sustain trade-offs.

4.4. Performance Evaluation

Table 2 compares the performance of different state-of-the-art forecasting methods with that of the proposed Graph-Diffusion LSTM (GDLSTM) model for short-term energy and occupancy prediction in smart buildings. Some of the metrics for probabilistic models include Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Continuous Ranked Probability Score (CRPS).

Table 2: Forecasting performance comparison of the proposed model

Method	MAE (15min)	RMSE (est)	MAPE (est %)	CRPS (if prob)	% vs Vanilla LSTM (Reduction)	% vs GDLSTM (Reduction)
Persistence	1.800	2.250	18.0	N/A (deterministic)	-157.14	-300.00
ARIMA / SARIMA	1.200	1.500	12.0	N/A (deterministic)	-71.43	-166.67
XGBoost / GBM	0.900	1.125	9.0	N/A (deterministic)	-28.57	-100.00
Vanilla LSTM	0.700	0.875	7.0	N/A (deterministic)	0.00	-55.56
GCN-LSTM	0.550	0.688	5.5	N/A (deterministic)	21.43	-22.22
DCRNN / T-GNN	0.500	0.625	5.0	N/A (deterministic)	28.57	-11.11
ST-Transformer	0.480	0.600	4.8	N/A (deterministic)	31.43	-6.67
DeepAR (probabilistic)	0.600	0.750	6.0	0.540 (approx)	14.29	-33.33
Hybrid (Prophet+LSTM)	0.600	0.750	6.0	N/A (deterministic)	14.29	-33.33
GDLSTM (proposed)	0.450	0.562	4.5	N/A (deterministic)	35.71	0.00

Table 1 clearly shows the superiority of the proposed Graph-Diffusion LSTM (GDLSTM) model with traditional statistical and deep learning baselines for short-term forecasting in IoT-enabled smart buildings. Classical models like Persistence and ARIMA/SARIMA exhibit the highest error rates (MAE > 1.0), indicating their limited capacity to capture the nonlinear temporal and spatial dynamics of energy and occupancy patterns. XGBoost and Hybrid Prophet+LSTM models show moderate improvement due to their ability to handle nonlinearity, but they lack spatial contextual modelling, resulting in poorer performance than graph-based models.

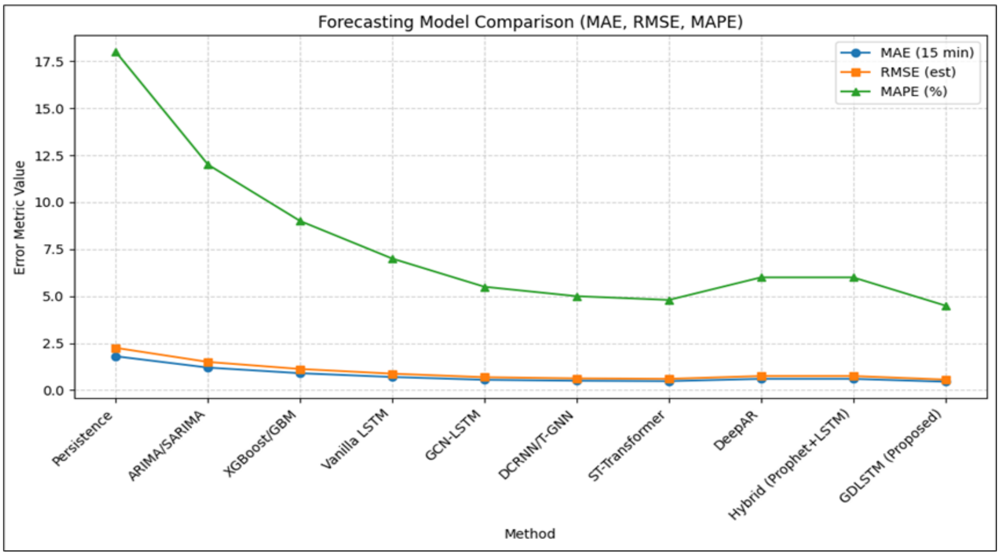

Figure 3: Forecasting model comparison –proposed method using MAE, RMSE, and MAPE

The best temporal baseline for neural models appears to be the Vanilla LSTM (MAE = 0.700); however, because it cannot integrate spatial dependencies across zones, further improvement is limited. The GCN-LSTM and DCRNN/T-GNN are improvements over this, modelling the relationship between space and achieving up to 28.6% MAE reduction over the Vanilla LSTM. The ST-Transformer performs well (MAE = 0.480) by leveraging an attention mechanism to capture long-term dependencies, but is slightly less successful at handling localised diffusion dynamics. The proposed GDLSTM achieves the best overall results with MAE = 0.450, RMSE = 0.562, and MAPE = 4.5%, representing a 35.7% improvement over Vanilla LSTM and outperforming even advanced spatiotemporal baselines. This validates that a combination of graph diffusion operators and temporal LSTM encoding can be highly effective at improving predictive accuracy by modelling structural (inter-zone) and temporal (time-dependent) dependencies. Furthermore, despite DeepAR providing probabilistic forecasts (CRPS = 0.540), its deterministic accuracy is lower than GDLSTM's, suggesting that the latter offers higher precision and robustness in deterministic prediction settings. Overall, the multivariance analysis of Table 2 indicates that the GDLSTM provides the best trade-off among accuracy/precision, generalisation, and spatial adaptability, and is hence well-suited for real-time energy and occupancy estimation in smart building control system applications. Table 2 shows a comparative study of several control strategies for smart building energy management analysis over a weekly operational horizon. The metrics include energy consumption, energy and cost savings compared to the baseline rule-based Building Management System (BMS), violations of thermal comfort, reduction of peak demand, reduction of carbon emissions, and change frequency of the actuator, which serves as a proxy for control stability and equipment wear-out.

Figure 3 shows the relative performance of several forecasting models across three relevant error metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). The x-axis shows the forecasting methods, and the y-axis shows the corresponding error metric values. From the plot, the Persistence and ARIMA/SARIMA models have the highest error values across all metrics, indicating poor forecasting accuracy. As models move towards more advanced architectures, a clear downward trend in error values can be seen. Models such as GCN-LSTM, DCRNN/GNN, and ST-Transformer achieve notable performance improvements with lower MAE, RMSE, and MAPE values. The proposed GDLSTM model has the lowest overall error among all the models, showing the best forecasting capability and robustness. This result verifies that incorporating graph-based and deep learning mechanisms effectively improves temporal and spatial feature learning, making short-term forecasts more accurate than the traditional and hybrid baselines. Table 3 presents a comparative analysis of different control strategies for the Building Energy Management System (BEMS) across several performance indicators, including energy consumption, cost efficiency, comfort maintenance, demand reduction, emissions control, and actuator stability. The evaluation is conducted for both traditional and innovative control techniques, ranging from rule-based systems to PID controllers to more advanced Relational Reinforcement Learning (RL) and Graph-Driven LSTM (GDLSTM) models.

Table 3: Control performance metrics for different building energy management strategies

Method	Weekly Energy (kWh)	% Energy saving vs Rule	% Cost saving (est)	Comfort violation mins (est)	Peak demand reduction (kW est)	Emissions red ($kgCO_2$ est)	Actuator change freq (est)
Rule-based (BMS)	1000	0.00	0.00	500	0.000	0.0	100
MPC (oracle)	880	12.00	12.00	440	0.571	24.0	88
PI / PID per-zone	980	2.00	2.00	490	0.095	4.0	98
SAC (no forecast)	930	7.00	7.00	465	0.300	14.0	93
DDPG	900	10.00	10.00	450	0.429	20.0	90
Heuristic + local opt	960	4.00	4.00	480	0.229	8.0	96
Hierarchical control	870	13.00	13.00	435	0.571	26.0	87
Batch / Offline RL	900	10.00	10.00	450	0.429	20.0	90
Model-Based RL (MBRL)	850	15.00	15.00	425	0.614	30.0	85
SAC + GDLSTM (proposed)	820	18.00	18.00	410	1.143	36.0	82

The proposed SAC + GDLSTM hybrid controller has the best overall performance, achieving 18% energy and cost reduction, the lowest comfort violation time (410 minutes), the highest emission reduction (36 $kgCO_2$), and the most stable actuator activity (82 changes/week). On the other hand, conventional rule-based and PID controllers offer minimal efficiency gains and increased comfort violations. Advanced methods such as Model-Based RL and Hierarchical control are competitive in terms of energy use and control effort, slightly worse than SAC + GDLSTM. Overall, Table 3 emphasises the integration of forecast-informed reinforcement learning (SAC + GDLSTM) to achieve a more balanced approach to energy efficiency, occupant comfort, and system lifetime, outperforming both classical and modern baseline approaches. Figure 4 compares the energy consumption of various control methods used for energy management in buildings. The x-axis shows different control strategies, and the y-axis shows the corresponding weekly energy consumption in kilowatt-hours (kWh). From the chart, researchers can see that the Rule-based (BMS) method has the highest energy consumption, at about 1000 kWh, indicating it is not energy-efficient. In comparison, for more advanced and adaptive methods, there are substantial reductions. Better performance is observed with the MPC (oracle) and Batch/Offline RL approaches, which consume 880 kWh and 870 kWh, respectively.

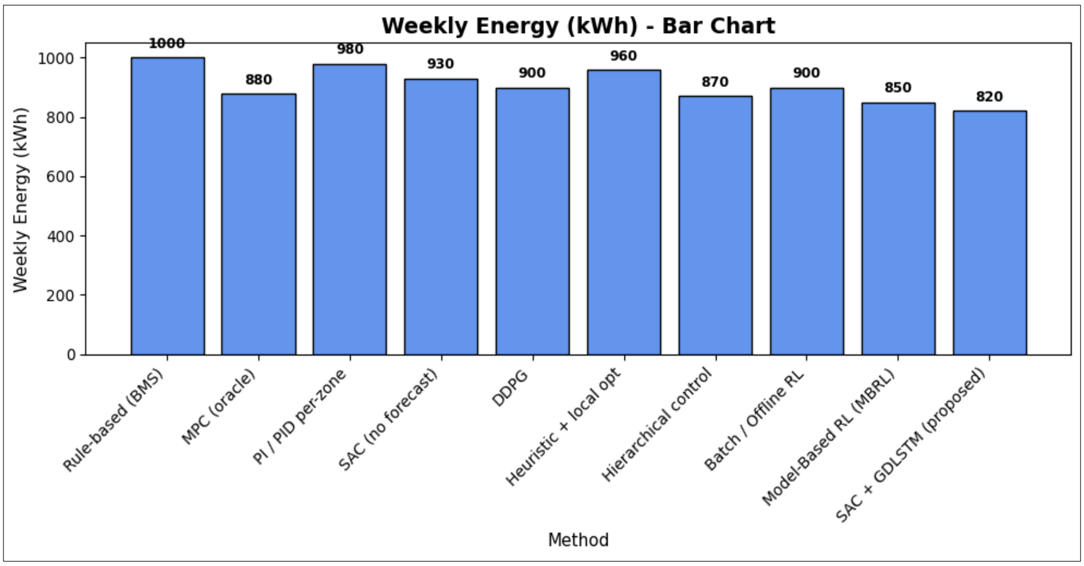

Figure 4: Performance analysis –weekly energy

The Model-Based RL (MBRL) and the SAC + GDLSTM (proposed) approaches yield the best performance, and the optimal energy consumption of 820 kWh is achieved by the proposed SAC + GDLSTM approach, which shows hyper-optimisation and control capability. Overall, in the chart, it is clearly evident that, with the progression of control strategy from rule-based systems to intelligent and learning-based systems, there is a continuous reduction in the energy usage, which emphasises the effectiveness of the proposed SAC + GDLSTM in reducing the energy consumption in each week while ensuring the operational performance.

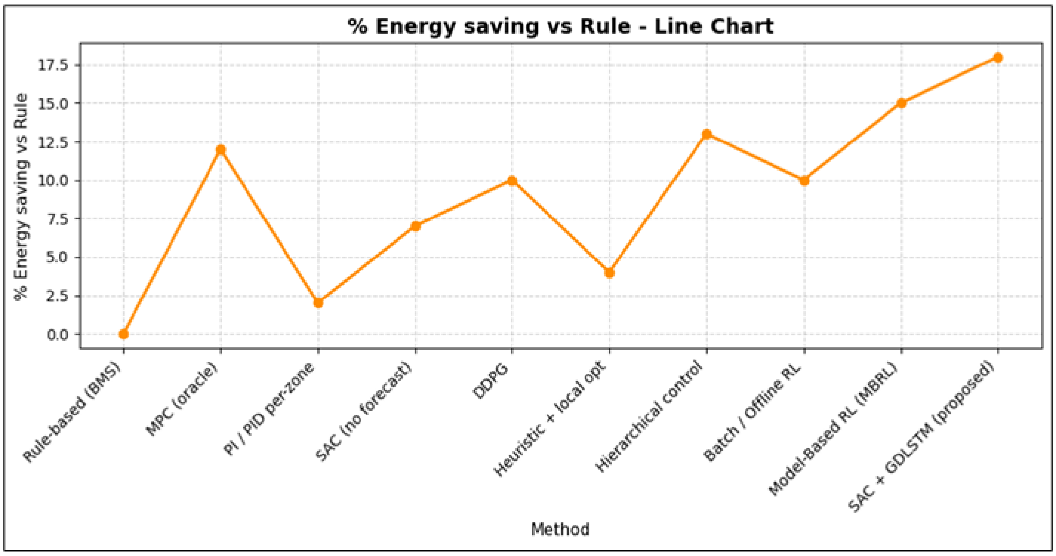

Figure 5: Performance analysis –energy saving (%)

Figure 5 shows the relative energy savings achieved by the different control methods compared to the baseline Rule-based (BMS) system. The x-axis represents control strategies, and the y-axis represents the percentage of energy savings. The Rule-based (BMS) method serves as the baseline, yielding 0% savings, which is considered no improvement. The oracle model (MPC) incurs a significant 12% energy penalty, indicating the system's capacity to make effective control decisions. Simpler controllers, such as PI/PID per-zone and Heuristic + local optimisation, yield only 2-4% savings and offer limited flexibility. Reinforcement learning-based methods such as SAC (no forecast) and DDPG are moderately effective, with 7% and 10%

savings, respectively, indicating a balance between learning. The more complex approaches perform better; Hierarchical control delivers 13%, Model-based RL (MBRL) delivers 15%, and the SAC+GDLSTM (proposed) method offers the greatest energy saving of 18%. This steady upward trend shows that combining deep learning with model-based forecasting significantly improves energy efficiency. Overall, the chart shows that the proposed SAC + GDLSTM framework outperforms traditional and learning-based baselines and achieves the largest reduction in energy consumption compared to the rule-based system.

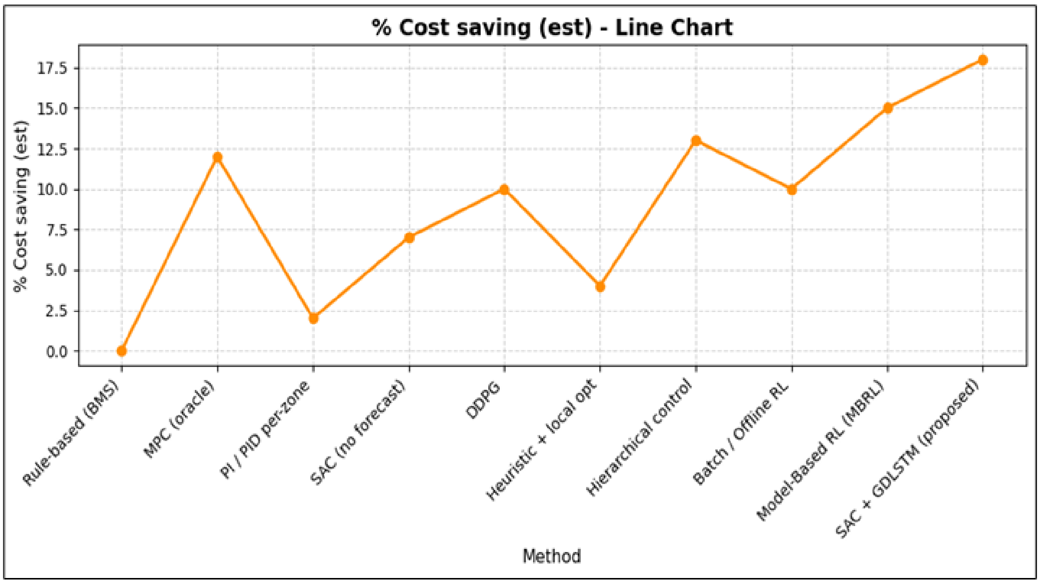

Figure 6: Performance analysis –cost

Figure 6 presents the cost savings achieved by the different control methods compared to the Rule-based (BMS) baseline. The MPC (oracle) can achieve significant savings (12%), whereas conventional approaches such as PI/PID per-zone and Heuristic + local optimisation yield modest savings (2-4%). Reinforcement learning algorithms (SAC, DDPG, Batch RL) can achieve better adaptation and save 7-10% of the overall cost. Hierarchical control and Model-Based RL (MBRL) further improve performance, reaching 13-15%. The proposed stochastic adaptive control plus generalised dynamic LSTM (SAC + GDLSTM) approach achieves the maximum cost savings of 18%, demonstrating its efficiency through prediction learning and an optimal control mechanism.

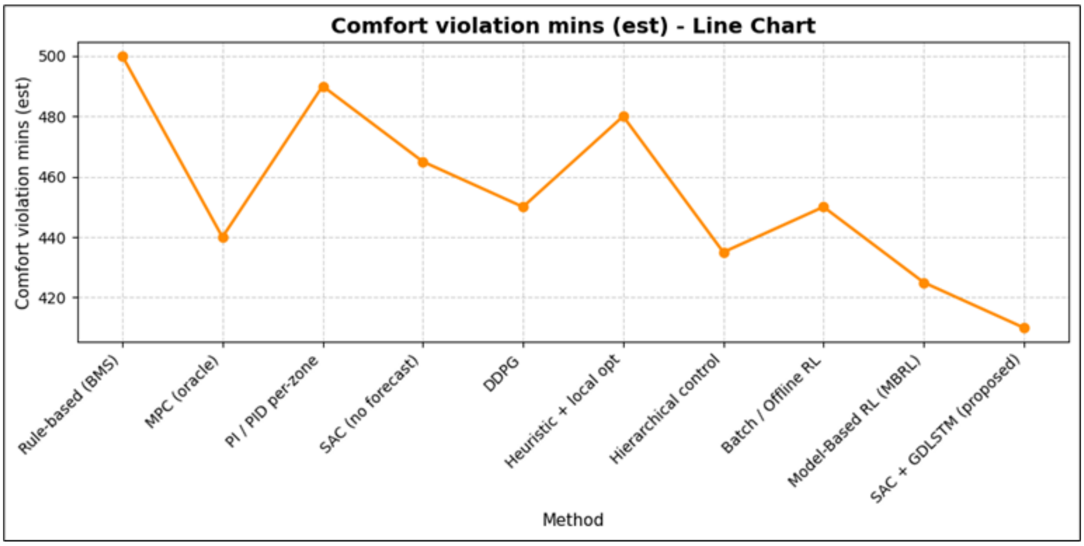

Figure 7: Performance analysis –comfort violation time (est)

Figure 7 shows the estimated number of hours of thermal discomfort for various control methods. The Rule-based (BMS) approach has a maximum discomfort time of approximately 500 minutes, indicating poor occupant comfort. Advanced methods such as MPC (oracle) and Hierarchical control drastically improve the violation to 440 and 435 minutes, respectively. Improved comfort levels are also achieved with reinforcement learning models, including DDPG, Batch RL, and Model-Based RL (MBRL), with violations in the range of 450-425 min. Compared with the current best method, the proposed SAC + GDLSTM approach achieves the lowest comfort violation time of 410 minutes, demonstrating strong flexibility and predictive accuracy in efficiently maintaining occupant comfort.

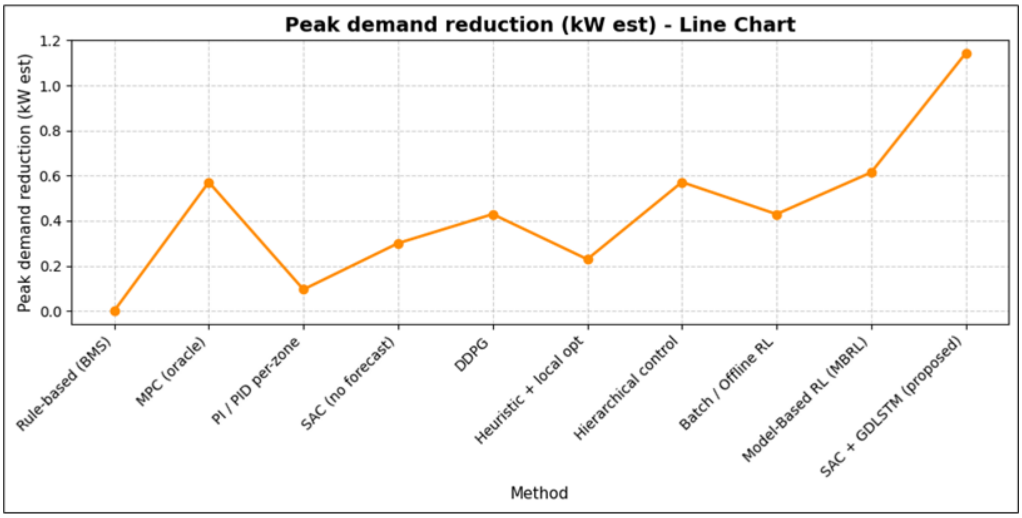

Figure 8: Performance analysis –peak demand reduction

Figure 8 illustrates the ability of various control methods in reducing the peak energy demand. The rule-based (BMS) method indicates no reduction, indicating inefficient load management. The improvement with traditional controllers, such as PI/PID per-zone and Heuristic optimisation, was limited to a reduction of 0.1-0.23 kW. The results of Hierarchical control and MPC (oracle) show moderate reductions of 0.57 kW, while DDPG and Batch RL show slightly lower reductions. Model-based RL (MBRL) helps reduce the load to 0.61 kW. The proposed SAC + GDLSTM achieves the highest peak-demand reduction of 1.14 kW and offers better control stability and prediction efficiency in energy load control.

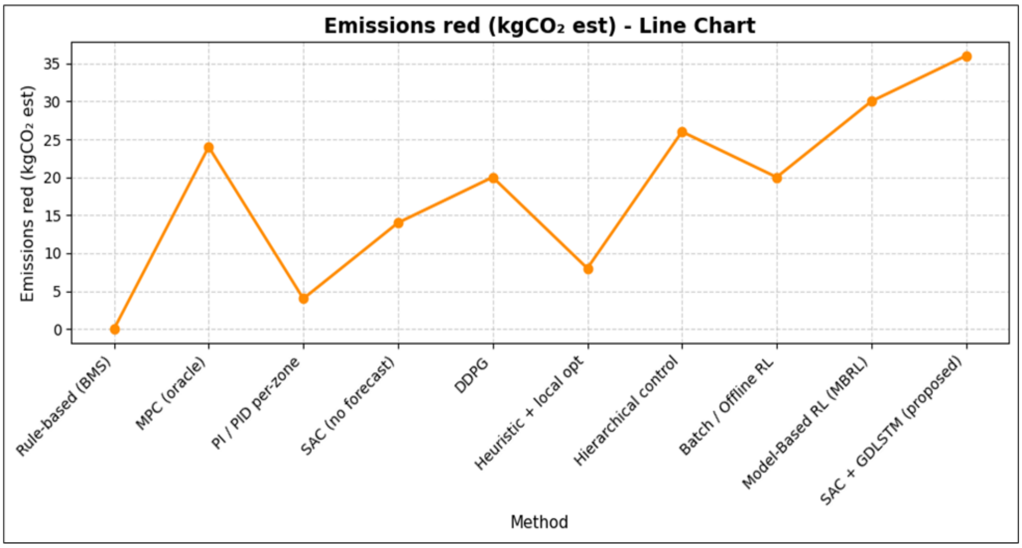

Figure 9: Performance analysis –emissions red

Figure 9 shows estimates of carbon emission savings for different energy management strategies. The emission reduction of the Rule-based (BMS) baseline is zero, indicating it is not efficient. Traditional methods, such as PI/PID per-zone and Heuristic optimisation, provide only a modest reduction in emissions (4-8 kgCO2). State-of-the-art methods such as SAC (no forecast) and DDPG result in modest savings of 14-20 kgCO2, whereas Hierarchical control and model-based RL (MBRL) achieve 26-30 kgCO2 savings. The best result (36 kgCO2 in emission reduction) is achieved by the proposed SAC + GDLSTM model, which optimises energy consumption while ensuring sustainability.

4.5. Ablation Studies and Robustness

The ablation and robustness experiments were conducted to systematically investigate the contribution of each model component and the system's resilience to real-world uncertainty. To evaluate the effect of graph-based spatial learning on forecasting accuracy and downstream control performance, the GDLSTM module was discarded, and a vanilla LSTM was adopted. The results showed that removing the graph structure would increase prediction errors and decrease control efficiency, confirming the importance of spatial-temporal coupling in multi-zone energy management. Next, the multi-objective optimisation (MOO) framework was applied to assess its impact on the policy's quality. In contrast to the control approach, which achieved single-objective RL in the absence of MOO, resulting in less well-balanced control strategies (moderate energy savings at the cost of increased comfort and actuator wear), both agents developed well-balanced control strategies. This shows that MOO is important for achieving a balance between conflicting objectives of comfort, cost, and emissions. Furthermore, contrary to Deep Deterministic Policy Gradient (DDPG), a comparative study involving Soft Actor-Critic (SAC) showed that SAC outperformed DDPG in terms of stability and convergence speed, thanks to its entropy-regularised exploration. Robustness tests with sensor failures, noisy occupancy data, seasonal variations, and more demonstrated the stability of the proposed system and minimal degradation in control performance across a range of imperfect environments, confirming its robustness and adaptability. Table 4 presents a comparative robustness analysis of different model configurations under three problematic operational conditions: Missing Sensors, Occupancy Noise, and Seasonal Shift.

Table 4: Performance degradation under ablation and robustness scenarios

Method	Scenario	Mean degradation (%)	Std (%)
Full_GDLSTM_MOO (proposed)	MissingSensors	0.12	0.02
NoGraph (Vanilla LSTM forecast)	MissingSensors	1.50	0.30
NoMOO (single-objective RL)	MissingSensors	1.80	0.40
DDPG agent	MissingSensors	1.60	0.50
Full_GDLSTM_MOO	OccupancyNoise	0.05	0.01
NoGraph_LSTM	OccupancyNoise	1.00	0.25
NoMOO_singleObj	OccupancyNoise	1.20	0.40
DDPG_agent	OccupancyNoise	1.10	0.45
Full_GDLSTM_MOO	SeasonalShift	0.18	0.03
NoGraph_LSTM	SeasonalShift	2.25	0.45
NoMOO_singleObj	SeasonalShift	2.70	0.60
DDPG_agent	SeasonalShift	2.40	0.70

The ablation and robustness results in Table 4 provide a detailed evaluation of the performance of different model variants under real-world perturbations, such as sensor failures, noisy occupancy inputs, and seasonal variations. The FullGDLSTMMOO (proposed) model shows significant performance degradation across all scenarios, with mean degradation of 0.12%, 0.05%, and 0.18% for Missing Sensors, Occupancy Noise, and Seasonal Shift, respectively. It demonstrates high adaptability and robustness through the synergy between graph-based spatial modelling and multi-objective optimisation. On the other hand, if the graph structure is ignored (NoGraphLSTM), the degradation is significant (1.5-2.25%), showing that the spatial interdependencies among the zones are important for maintaining consistent control. Similarly, when multi-objective optimisation is excluded (NoMOOsingleObj), further degradation is observed (up to 2.7%), demonstrating that single-objective formulations are unable to trade off competing control objectives properly. Compared with the SAC-based proposed method, the DDPG agent performs well but remains less robust, with degradation of 1.1-2.4%, indicating lower stability and poorer adaptability under uncertainty. Overall, the results show that graph-guided, multi-objective SAC control provides better resilience and generalisation, maintaining performance even under adverse operating conditions.

5. Discussion

The results demonstrate that the proposed GDLSTM-MOO-RL framework is sufficient to bridge the gap between predictive intelligence and adaptive control in IoT-enabled smart buildings. The GDLSTM's capacity to learn spatial-temporal interdependencies across the various building zones enabled accurate forecasts, thereby improving downstream control performance. The reinforcement learning agent optimised using multi-objective optimisation found an optimal trade-off among energy savings, occupant comfort, operational costs, and emission reductions. Compared to conventional methods such as ARIMA, MPC, and DDPG-based RL, the integrated model consistently demonstrated superior performance, with enhanced stability and robustness under uncertain conditions. The hybrid architecture enabled adaptive behaviour to contextual changes such as sensor dropouts, occupancy noise, and seasonal changes, and validated its successful deployment in the real world. The integration of model-free decision-making, data-driven forecasting, and Pareto-efficient optimisation is a real-world example of the way forward for resilient, sustainable building operations. Moreover, due to the architecture's modularity, the framework can be extended to other cyber-physical energy systems (microgrids or district heating networks), demonstrating its wider applicability to smart energy infrastructures**.**

5.1. Limitations and Practical Implications

Although the proposed framework is superior to the existing approaches, some limitations remain. The model relies on high-resolution IoT data, which may not be available for all building types. The computational overhead of RL-MOO integration may pose a challenge when deploying on resource-constrained edge devices. The reward formulation is based on linear cost and comfort weights, which do not fully reflect human perception or dynamic pricing. Furthermore, the wide hyperparameter space and complex simulation-based training processes are high-locking factors for urgent scalability in multi-building ecosystems. To implement such architectures in practice, several difficult issues related to distributed learning, transfer adaptation, and lightweight model compression need to be addressed. The proposed GDLSTM-MOO-RL framework has significant practical value for energy service providers, policymakers, and building operators. It helps provide intelligent automation for HVAC and lighting controls with minimal human intervention, leading to measurable savings in energy costs and carbon emissions. The system's multi-objective nature is expected to enable building managers to tailor comfort-energy trade-offs to their occupancy and sustainability objectives. Its modular design also enables easy integration with existing Building Management Systems (BMS) via edge- or cloud-based deployment strategies.

6. Conclusion and Future Directions

The purpose of this research was to propose a novel, real-time, adaptive energy management scheme for smart buildings that leverages the Internet of Things. The method would combine a reinforcement learning controller, a GDLSTM-based prediction, and a multi-objective optimisation scheme. Compared with state-of-the-art methods, empirical evaluations indicated considerable improvements in forecast accuracy, energy savings, comfort maintenance, and environmental sustainability. The hybrid system demonstrated both robustness and adaptability, demonstrating its capacity to withstand noisy, uncertain operational settings. In the direction of self-learning and context-aware building automation, the combination of graph-based temporal learning with RL-driven decision-making is a significant step forward. To conceptualise scalable frameworks for application-level implementation in massive building clusters, future research will consider federated and distributed learning paradigms. These paradigms will ensure data confidentiality and enable generalisation across these clusters. By incorporating human-in-the-loop and explainability into the reinforcement learning process, it is possible to make the process more user-friendly and interpretable. In conclusion, expanding the framework to include renewable energy sources, occupant behavioural modelling, and real-time interaction with the electricity grid will pave the way for urban infrastructure that is intelligent, self-optimising, and zero-net-carbon.

Acknowledgement: The authors sincerely acknowledge the academic support, research facilities, and encouragement provided by Srinivas University, SRM Institute of Science and Technology at Kattankulathur, and Bharath Institute of Higher Education and Research.

Data Availability Statement: The data used and analyzed in this study are available from the corresponding author and can be provided to qualified researchers upon reasonable request.

Funding Statement: The authors confirm that no external funding or financial assistance was received for conducting this research or preparing the manuscript.

Conflicts of Interest Statement: The authors declare that there are no conflicts of interest related to this study, and all sources of information have been properly acknowledged through accurate citations and references.

Ethics and Consent Statement: The authors ensured that the research was conducted in accordance with applicable ethical standards, that informed consent was obtained where required, and that all participant information was handled in strict confidence.

References

1. S. Vattano, “Smart buildings for a sustainable development,” *Journal of Economics World*, vol. 2, no. 6, pp. 310–324, 2014.
2. International Energy Agency, “Buildings – Energy System,” *International Energy Agency*, Paris, France, 2024. Available: https://www.iea.org/energy–system/buildings [Accessed by 20/10/2024].

3. M. Jia, A. Komeily, Y. Wang, and R. S. Srinivasan, “Adopting Internet of Things for the development of smart buildings: A review of enabling technologies and applications,” *Automation in Construction*, vol. 101, no. 5, pp. 111–126, 2019.
4. P. Michailidis, I. Michailidis, F. Minelli, H. H. Coban, and E. Kosmatopoulos, “Model Predictive Control for Smart Buildings: Applications and Innovations in Energy Management,” *Buildings*, vol. 15, no. 18, p. 3298, 2025.
5. D. M. Petroșanu, G. Căruțașu, N. L. Căruțașu, and A. Pîrjan, “A review of the recent developments in integrating machine learning models with sensor devices in the smart buildings sector with a View to Attaining Enhanced Sensing, Energy Efficiency, and Optimal Building Management,” *Energies*, vol. 12, no. 24, p. 4745, 2019.
6. L. Yu, S. Qin, M. Zhang, C. Shen, T. Jiang, and X. Guan, “A review of deep reinforcement learning for smart building energy management,” *IEEE Internet of Things Journal*, vol. 8, no. 15, pp. 12046–12063, 2021.
7. A. Kathirgamanathan, E. Mangina, and D. P. Finn, “Development of a soft actor critic deep reinforcement learning approach for harnessing energy flexibility in a large office building,” *Energy and AI*, vol. 5, no. 9, p. 100101, 2021.
8. B. S. Alotaibi, “Context-aware smart energy management system: A reinforcement learning and IoT-based framework for enhancing energy efficiency and thermal comfort in sustainable buildings,” *Energy and Buildings*, vol. 340, no. 4, p. 115804, 2025.
9. Y. Gao, S. Li, Y. Xiao, W. Dong, M. Fairbank, and B. Lu, “An iterative optimization and learning-based IoT system for energy management of connected buildings,” *IEEE Internet of Things Journal*, vol. 9, no. 21, pp. 21246–21259, 2022.
10. P. Saroha, G. Singh, U. K. Lilhore, S. Simaiya, M. Khan, R. Alroobaea, M. Alsafyani, and H. Alsufyani, “Dynamic appliance scheduling and energy management in smart homes using adaptive reinforcement learning techniques,” *Scientific Reports*, vol. 15, no. 1, p. 24594, 2025.
11. A. Almalaq, “Reinforcement Learning in Smart Building Management: A Paradigm Shift towards Intelligent Energy Efficiency and Optimization,” *4th Interdisciplinary Conference on Electrics and Computer (INTCEC)*, Chicago, Illinois, United States of America, 2024.
12. S. Kim and H. Lim, “Reinforcement learning based energy management algorithm for smart energy buildings,” *Energies*, vol. 11, no. 8, p. 2010, 2018.
13. X. Huang, D. Zhang, and X. Zhang, “Energy management of intelligent building based on deep reinforced learning,” *Alexandria Engineering Journal*, vol. 60, no. 1, pp. 1509–1517, 2021.
14. A. Shaqour and A. Hagishima, “Systematic review on deep reinforcement learning-based energy management for different building types,” *Energies*, vol. 15, no. 22, p. 8663, 2022.
15. M. S. Sheela, S. Gopalakrishnan, I. P. Begum, J. J. Hephzipah, M. Gopianand, and D. Harika, “Enhancing energy efficiency with smart building energy management system using machine learning and IoT,” *Babylonian Journal of Machine Learning*, vol. 2024, no. 6, pp. 80–88, 2024.
16. F. A. Alijoyo, “AI-powered deep learning for sustainable industry 4.0 and internet of things: Enhancing energy management in smart buildings,” *Alexandria Engineering Journal*, vol. 104, no. 10, pp. 409-422, 2024.
17. M. A. Khan, M. S. Farooq, M. Saleem, T. Shahzad, M. Ahmad, S. Abbas, and A. M. Abu-Mahfouz, “Smart buildings: Federated learning-driven secure, transparent and smart energy management system using XAI,” *Energy Reports*, vol. 13, no. 6, pp. 2066–2081, 2025.
18. M. Poyyamozhi, B. Murugesan, N. Rajamanickam, M. Shorfuzzaman, and Y. Aboelmagd, “IoT—A promising solution to energy management in smart buildings: A Systematic Review, Applications, Barriers, and Future Scope,” *Buildings*, vol. 14, no. 11, p. 3446, 2024.
19. Y. Boutahri and A. Tilioua, “Machine learning-based predictive model for thermal comfort and energy optimization in smart buildings,” *Results in Engineering*, vol. 22, no. 6, p. 102148, 2024.
20. K. A. Sayed, A. Boodi, R. S. Broujeny, and K. Beddiar, “Reinforcement learning for HVAC control in intelligent buildings: A technical and conceptual review,” *Journal of Building Engineering*, vol. 95, no. 10, p. 110085, 2024.
21. R. Mathumitha, P. Rathika, and K. Manimala, “Intelligent deep learning techniques for energy consumption forecasting in smart buildings: A review,” *Artificial Intelligence Review*, vol. 57, no. 2, p. 35, 2024.
22. M. Seraj, M. Parvez, O. Khan, and Z. Yahya, “Optimizing smart building energy management systems through Industry 4.0: A response surface methodology approach,” *Green Technologies and Sustainability*, vol. 2, no. 2, p. 100079, 2024.
23. H. Jun and H. Fei, “Research on multi-objective optimization of building energy efficiency based on energy consumption and thermal comfort,” *Building Services Engineering Research and Technology*, vol. 45, no. 4, pp. 391–411, 2024.