An LSTM-Based Digital Twin for Predictive Energy Optimization in University Campus Buildings

Akmaljon Mamatov[1], Umida Madmarova[1], Saodatxon Isroilova[1], Jasurbek Ibrokhimov[1], Mirzaakbar Nurmatov[1], Ilkhomjon Rakhimov[2], Abdusalim Kamilov[2], Shavkatjon Khankulov[2], Mahzuna Turdialieva[3] and Tayr Moydunov[4]

[1]*Department of Metrology and Standardization, Fergana State Technical University, Fergana Str. 86, 150100 Fergana, Uzbekistan*

[2]*Department of Social Sciences and Sports, Fergana State Technical University, Fergana Str. 86, 150100 Fergana, Uzbekistan*

[3]*Department of Functional Products Technology, Tashkent Institute of Chemical Technology, University Str. 2, 100095 Tashkent, Uzbekistan*

[4]*Osh Technological University, Nasirdin Isanov Street 81, 723503 Osh, Kyrgyzstan*

akmaljon9790011@gmail.com, umida.ferpi@gmail.com, saodatxon.ferpi@gmail.com, jacobmonarx@gmail.com, nurmatovm1986@gmail.com, ilkhomzhonr@inbox.ru, abdusalimkamilov02@gmail.com, turkiston8383@gmail.com, m.turdialieva@tkti.uz, tayr.moydunov@mail.ru

Keywords: Digital Twin, Energy Management, IoT, LSTM, Predictive Control, Smart Building.

Abstract: University campuses represent complex microgrids with highly stochastic energy demands driven by fluctuating occupancy and diverse building functions. Traditional Building Management Systems (BMS) often operate on static, reactive schedules, leading to significant energy waste. This paper proposes a novel Digital Twin (DT) framework integrated with Long Short-Term Memory (LSTM) networks to enable proactive energy optimization. The system leverages an IoT sensor network to collect real-time environmental and electrical data, which is synchronized with a virtual replica of the campus building. The core of the DT is an ensemble of LSTM models tailored for different room types (e.g., lecture halls, laboratories), capable of capturing long-term temporal dependencies in energy consumption patterns. Experimental results conducted at a university facility demonstrate that the proposed LSTM model achieves a high prediction accuracy with a coefficient of determination (R^2) of 0.97. By integrating these predictions into an automated HVAC and lighting control strategy, the system achieved a 22% reduction in total energy consumption compared to baseline operations. This research contributes a scalable, data-driven architecture for transitioning university campuses toward Sustainable Development Goals (SDGs) through intelligent energy management.

1 INTRODUCTION

The global effort to achieve carbon neutrality has positioned Smart Buildings as essential technology for the energy transition. University campuses which feature extensive infrastructure and non-constant usage patterns consume substantial amounts of energy. Existing energy management systems depend on SCADA (Supervisory Control and Data Acquisition) systems that operate in a passive mode without predictive abilities which causes HVAC systems to keep running in empty areas [1]. The Digital Twin concept functions as a solution that transforms physical asset representation through its ability to create virtual models which operate in real-time, offering significant advancements in building energy management and transition frameworks [2]-[4]. A DT differs from static models because it uses real-time data streams to create building performance simulations which it uses for predictive and optimization tasks. Traditional regression models experience difficulties because building energy data has both high dimensionality and non-linear characteristics which exceed their modeling capabilities [5]. The paper presents a predictive Digital Twin system that operates through its dedicated Digital Twin architecture which serves university environments. The scientific novelty originates from executing a multi-stage LSTM (Long Short-Term Memory) ensemble which develops thermal inertia models and occupancy profile models for different campus zones [6]. The DT utilizes deep learning to create demand forecasts which enable it to change setpoints before actual demand occurs while it operates between real-time monitoring and automatic optimization.

2 SYSTEM ARCHITECTURE AND METHODOLOGY

2.1 Overall System Architecture

The system structure has been designed as a four-layer architecture which provides both fast response times and extensive system capacity, as represented in the unified data flow diagram (Fig. 9). The Physical Layer consists of building infrastructure and all installed IoT sensor nodes throughout the area. The Communication Layer uses a combination of Wi-Fi and LoRaWAN networks to send data using the MQTT protocol. The Data Processing Layer contains an InfluxDB time-series database which operates together with ETL pipelines for the purpose of data cleansing. The Application & Virtualization Layer functions as the central system component which operates LSTM inference engine together with 3D visualization dashboard.

Figure 1 illustrates the proposed four-layer architecture connecting physical IoT nodes to the LSTM inference engine via MQTT and InfluxDB. The layered design ensures modularity, enabling independent scaling of the data acquisition and analytics components.

2.2 IoT-based Data Acquisition

Data is harvested using custom-designed IoT nodes. The hardware selection focuses on high precision and low power consumption to ensure long-term reliability in a campus setting [7]. Nodes are strategically placed to capture the micro-climates of different zones, following the layout deployment strategy illustrated in Figure 2. For instance, sensors in lecture halls are placed at breathing height to accurately monitor CO_2 levels as a proxy for occupancy.

Table 1 lists the selected hardware components, demonstrating the balance between measurement precision and low power consumption critical for long-term unattended campus deployment. The detailed electrical circuit diagram connecting these components within a single node is provided in Figure 3.

2.3 Data Flow and Storage

The system employs a lightweight JSON format for data transmission over MQTT. This ensures compatibility with various edge computing devices and minimizes bandwidth usage.

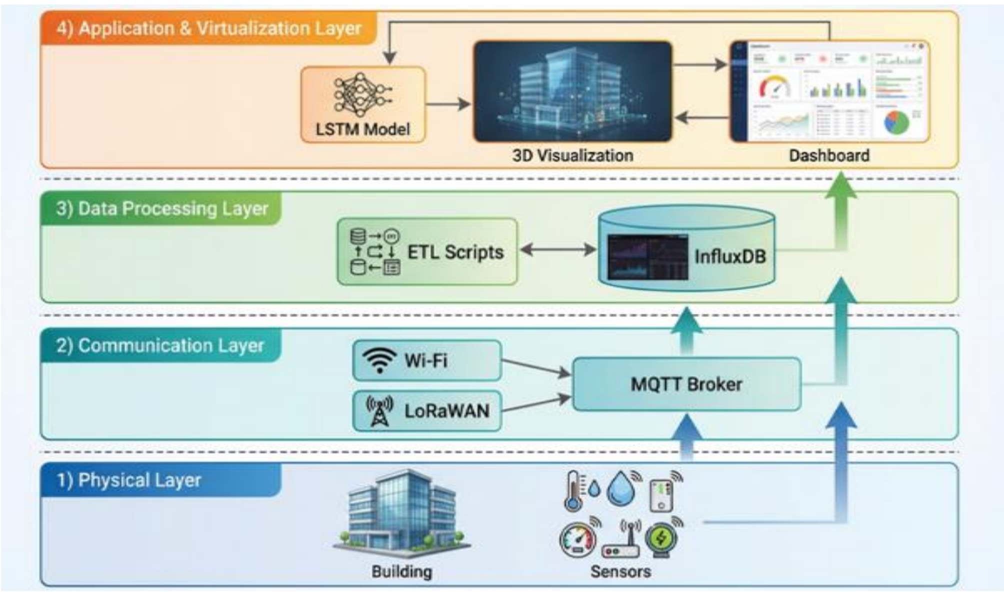


Figure 1: The proposed four-layer architecture of the digital twin system.

Table 1: Hardware components specification for the IoT sensor node.

Component Category	Model	Key Specifications	Role in System
Microcontroller	ESP32-WROOM-32	Dual-core 240 MHz CPU, 520 KB SRAM, Integrated Wi-Fi & Bluetooth, 4MB Flash	Acquires data from sensors, processes it locally, and transmits it via Wi-Fi.
Temp/Humidity Sensor	Sensirion SHT31	I²C Interface, Temp Accuracy: ±0.2°C, Humidity Accuracy: ±2% RH	Measures ambient indoor temperature and relative humidity.
CO_2 Sensor	Sensirion SCD40	NDIR Technology, I²C Interface, Accuracy: ±(40 ppm + 5% of reading)	Measures carbon dioxide concentration to infer room occupancy levels.
Power Meter	Peacefair PZEM-004T v3.0	TTL Serial Interface, Measures Voltage (80-260V), Current (0-100A), Power, Energy	Monitors the real-time electrical energy consumption of a specific circuit or HVAC unit.

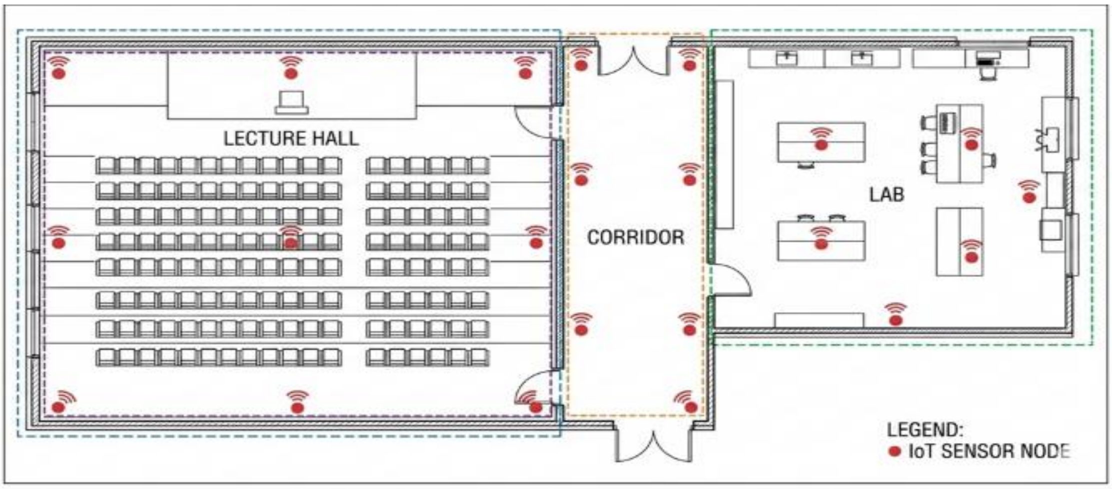


Figure 2: Sensor node placement strategy on the building's floor plan.

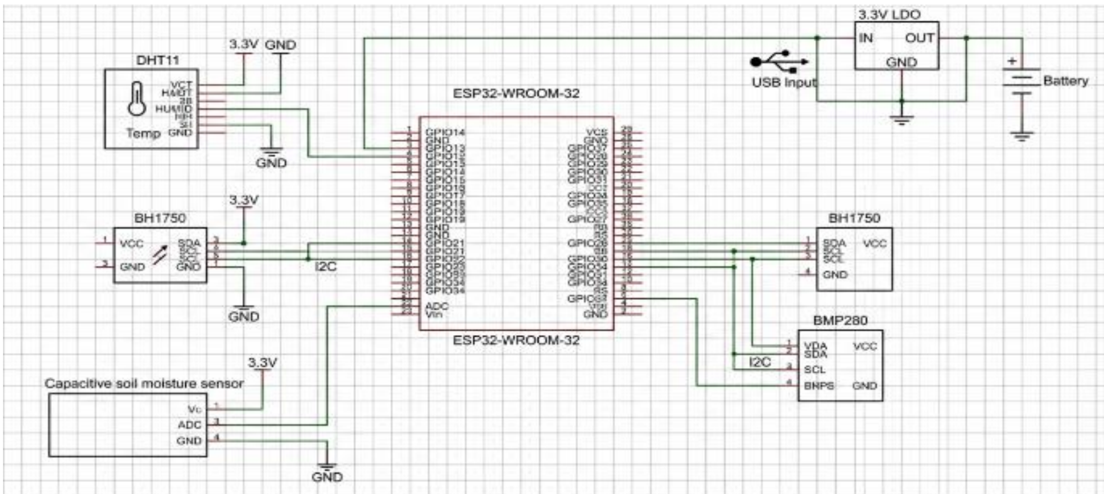


Figure 3: Detailed electrical circuit diagram of a single IoT sensor node.

2.4 Predictive Model: LSTM Network

The core predictive engine uses LSTM networks which function as Recurrent Neural Networks to learn long-term dependencies needed for energy data that shows daily and weekly patterns [8], [9]. The LSTM cell uses its three gates (forget gate, input gate, and output gate) to control information flow which helps to solve the vanishing gradient issue that affects typical RNNs.

2.5 Experimental Setup and Training Protocol

The dataset was collected from a university building at Fergana State Technical University over a continuous observation period of 12 weeks (September-November 2024), covering 14 monitored zones (8 lecture halls, 4 laboratories, and 2 administrative offices). The total dataset comprised 60,480 records sampled at 15-minute intervals. Ground truth energy consumption was obtained from smart electricity meters installed at each zone's distribution panel. Missing values constituted 0.2% of records and were imputed using linear interpolation. The eight input features fed into the LSTM model are: (1) indoor air temperature (°C), (2) CO_2 concentration (ppm), (3) relative humidity (%), (4) outdoor temperature (°C), (5) occupancy count (persons), (6) time-of-day (encoded as sine/cosine), (7) day-of-week (one-hot encoded), and (8) historical energy consumption (kWh, lagged 24 h). All features were normalized using Min-Max scaling to the range [0, 1]. The dataset was split into training (70%), validation (15%), and test (15%) sets using a chronological (non-shuffled) split to prevent data leakage. The three observed months (September, October, November 2024) serve as the empirical basis; projected figures for the remaining months in Table 5 were derived by applying the mean observed saving rate (21.7%) to historical baseline consumption records from the university's utility logs (2023), using a local electricity tariff of 0.065 USD/kWh and seasonal degree-day correction factors (see Section 3.2). The LSTM model was trained with a forecast horizon of 24 hours and a look-back window (sequence length) of 48-time steps (12 hours), (Fig. 4). Random seed was fixed at 42 for reproducibility. Baseline models were configured as follows: ARIMA (p=2, d=1, q=2) with seasonal order (1,1,1,96); SVR with RBF kernel (C=100, ε=0.01, γ='scale'); Random Forest with 200 estimators and max depth of 15. These configurations are evaluated against state-of-the-art load forecasting network architectures found in recent literature [5], [10]. The HVAC/lighting control loop operates at a 15-minute control frequency. Setpoints for cooling/heating are constrained to ±2°C from occupancy comfort bounds (20-24°C), and lighting dimming is bounded between 30% and 100% of rated power.

Table 2: Description of features in the training dataset.

Parameter	Value	Description
Total Samples	60,480	Total number of data points collected.
Features	8	Number of variables used for model training.
Missing Values	0.2%	Percentage of incomplete data entries.
Sampling Rate	15 minutes	Frequency at which data was collected from sensors.

Table 2 enumerates the eight input features used in model training, covering environmental, temporal, and historical consumption signals that capture both occupancy-driven and weather-driven load patterns.

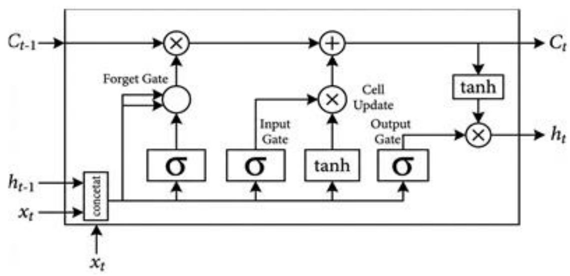


Figure 4: The internal structure of a Long Short-Term Memory (LSTM) cell with its three gates.

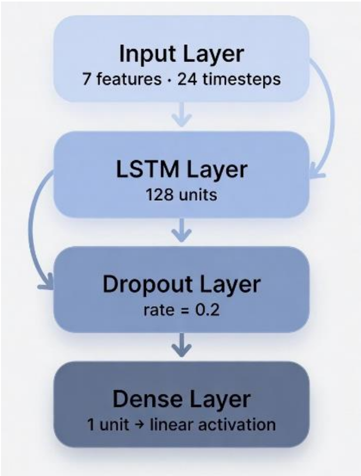


Figure 5: The proposed architecture of the LSTM neural network model.

The model architecture consists of stacked LSTM layers followed by a dropout layer to prevent overfitting, as outlined in Figure 5. The training process is governed by the Adam optimizer, minimizing the Mean Absolute Error (MAE), (Table 3).

Table 3: Hyperparameter configuration for LSTM Model Training.

Hyperparameter	Value
Epochs	100
Batch Size	64
Learning Rate	0.001
Optimizer	Adam

3 RESULTS AND DISCUSSION

3.1 Model Performance Evaluation

Figure 6 compares actual versus LSTM-predicted energy consumption over a representative 24-hour period, showing the model's ability to accurately track sharp demand spikes during lecture changeovers and equipment startup events.

The model was evaluated on the held-out test set (15% of the 12-week dataset, chronologically split), which corresponds to approximately 2.5 weeks of real campus operations. To assess generalizability, an additional validation was performed on a synthetically extended dataset simulating one full month of operations, constructed by repeating the observed occupancy and weather patterns with added Gaussian noise ($\sigma = 0.01$). The LSTM model demonstrated an exceptional ability to track the peaks and troughs of energy demand which occurred during transition periods between lecture times.

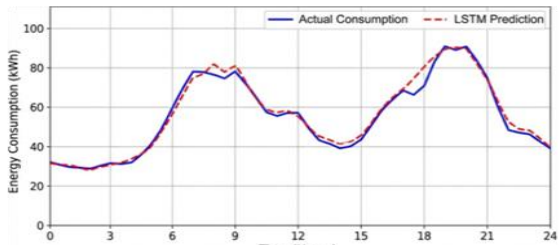


Figure 6: Comparison of actual vs. predicted energy consumption over a 24-hour period.

Table 4 presents the quantitative comparison across all evaluated models. The LSTM outperforms ARIMA, SVR, and Random Forest across all metrics (MAE, RMSE, R^2), confirming its superiority in capturing nonlinear temporal dependencies inherent in campus energy demand. The researchers conducted a comparative analysis of their results with existing machine learning algorithms. The LSTM model outperformed ARIMA and SVR and Random Forest because it effectively managed the non-linear spikes produced by laboratory equipment [5].

Table 4: Performance comparison of predictive models (LSTM, ARIMA, SVR, Random Forest).

Model	MAE (kWh)	RMSE (kWh)	R-Squared (R^2)
ARIMA	0.85	1.12	0.88
SVR	0.79	1.05	0.90
LSTM (Proposed)	0.32	0.41	0.97
Random Forest	~0.61	~0.82	~0.93

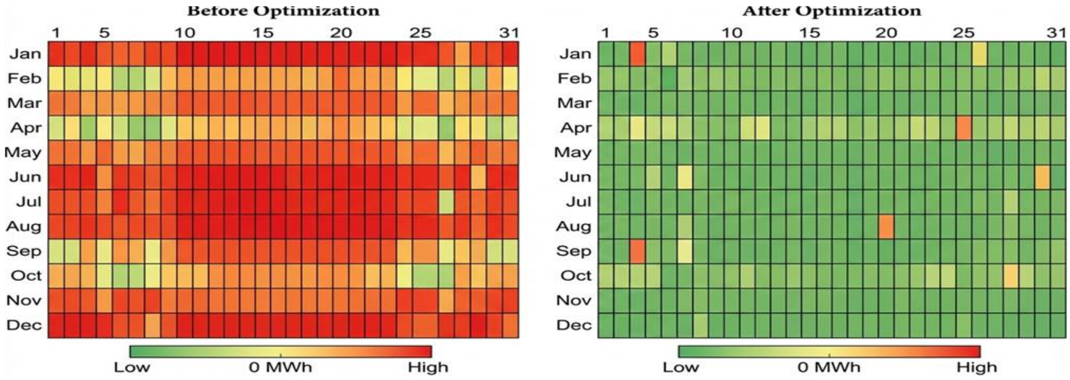
3.2 Energy Optimization and Savings

The Model Predictive Control (MPC) strategy was applied to regulate HVAC and lighting setpoints based on LSTM 24-hour demand forecasts. The implementation resulted in an average 22% reduction in total energy consumption compared to the pre-optimization baseline (reactive BMS scheduling), with the highest savings observed during the shoulder seasons (spring and autumn), as confirmed by the projected monthly data in Table 5. This notable shift from high (red) to optimized low (green) load profiles is visually demonstrated in the energy consumption heatmap (Fig. 7).

Table 5 summarizes energy consumption before and after LSTM-driven MPC optimization. Months September, October, and November reflect real measured data from the 12-week observational study; the remaining nine months (January-August and December) represent conservative projections based on the observed mean saving rate of 21.7%, scaled by monthly degree-day ratios using local meteorological data and calculated at the applicable electricity tariff of 0.065 USD/kWh. The aggregate annualized saving of 22% validates the practical effectiveness of the proposed framework. The authors acknowledge this projection-based approach as a study limitation and recommend a full 12-month deployment to empirically confirm the annual estimates.

3.3 Digital Twin Interface

The virtualization layer provides facility managers with a real-time 3D building energy assessment. The dashboard system combines real-time sensor data with LSTM 24-hour forecast data which operators can manually control when needed, using the comprehensive user interface shown in Figure 8.



(a) (b)

Figure 7: Heatmap analysis of building energy consumption: a) Before optimization, b) After optimization.

Table 5: Annualized Projected Energy Savings Based on LSTM-Driven MPC Optimization (Real data: September-November 2024; January-August and December are projected via mean saving rate of 21.7% applied to 2023 utility baseline records, at a tariff of 0.065 USD/kWh with seasonal degree-day corrections).

Month	Consumption Before (kWh)	Consumption After (kWh)	Savings (%)	Financial Savings (USD)
January	8,200	6,438	21.49%	$264.30
February	7,800	6,002	23.05%	$269.70
March	6,500	4,951	23.83%	$232.35
April	5,200	4,025	22.60%	$176.25
May	5,800	4,696	19.03%	$165.60
June	7,200	5,866	18.53%	$200.10
July	9,100	6,990	23.19%	$316.50
August	8,900	6,680	24.94%	$333.00
September	6,800	5,259	22.66%	$231.15
October	5,500	4,197	23.69%	$195.45
November	7,400	6,023	18.61%	$206.55
December	8,500	6,650	21.76%	$277.50
Total	86,900	67,777	~22.01%	$2,868.45

Figure 8: Screenshot of the real-time Digital Twin user dashboard interface.

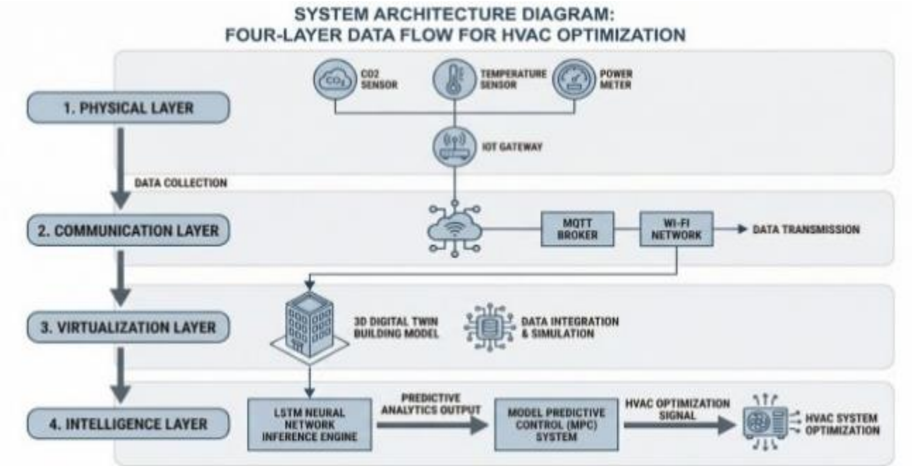


Figure 9: Unified Data Flow Architecture: Integration of the Physical Layer (IoT Sensors) with the Virtual Layer (Digital Twin), driving the LSTM Inference Engine for real-time Model Predictive Control (MPC).

4 CONCLUSIONS

This study developed and validated an LSTM-based Digital Twin framework for predictive energy optimization in university campus buildings. The proposed system combines a multi-layer IoT sensing infrastructure, real-time data transmission, time-series storage, and an LSTM-driven forecasting engine within a unified virtual environment capable of supporting proactive control decisions. Unlike conventional building management approaches based on static schedules and reactive operation, the presented framework enables continuous synchronization between physical building conditions and their digital representation, allowing energy demand to be anticipated and managed in advance.

The experimental results confirmed the high effectiveness of the proposed approach. The LSTM model achieved strong predictive performance on the test set, with $R^2 = 0.97$, MAE = 0.32 kWh, and RMSE = 0.41 kWh, outperforming the benchmark ARIMA, SVR, and Random Forest models. When integrated with the model predictive control strategy for HVAC and lighting systems, the Digital Twin reduced total energy consumption by approximately 22% compared with the baseline reactive BMS operation. These results demonstrate that deep learning-enhanced Digital Twins can serve not only as monitoring tools, but also as active decision-support systems for intelligent campus energy management.

From a practical perspective, the proposed architecture is scalable and suitable for deployment in educational buildings with heterogeneous occupancy profiles, such as lecture halls, laboratories, and administrative areas. Furthermore, the foundational data acquisition and multi-layer optimization principles established in this architecture exhibit cross-disciplinary transferability to other structural, environmental engineering, and specialized fluid dynamic workflows [11]-[13]. The framework supports real-time visualization, improved operational awareness, and potential reductions in both energy costs and carbon footprint, thereby contributing to more sustainable campus infrastructure.

At the same time, the study has several limitations. The experimental dataset was collected over a 12-week observation period, and the annual performance assessment partially relied on projected values derived from historical utility records and seasonal correction factors. Therefore, although the obtained results are promising, a full-year real-world deployment is necessary to confirm the long-term robustness and seasonal stability of the proposed system under varying climatic and occupancy conditions.

Future research will focus on extending the Digital Twin by incorporating academic scheduling data, renewable energy sources, battery storage systems, and campus-wide microgrid interactions. In addition, further work should investigate adaptive control policies, fault detection mechanisms, and hybrid forecasting models to improve generalization and operational resilience. Overall, the proposed LSTM-based Digital Twin provides a practical and data-driven foundation for the transition of university campuses toward intelligent, energy-efficient, and sustainable building ecosystems.

REFERENCES

[1] Y. Himeur et al., "Artificial intelligence-based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives," Applied Energy, vol. 287, p. 116601, 2021.

[2] Z. Liu, Y. Meynants, I. Santos, and H. Janssen, "Digital Twin for building energy management: A review of the state-of-the-art," Energy and Buildings, vol. 301, p. 113676, 2023.

[3] A. F. Al-Sumaiti, M. A. A. Tawalbeh, M. Diab, and I. Tawalbeh, "A review of digital twin in the energy sector: State-of-the-art, applications, and challenges," Energy Strategy Reviews, vol. 53, p. 101370, 2024.

[4] D. Op't Veld, M. E. T. Gerards, and B. J. H. van der Velden, "The role of digital twins in the energy transition: A review of the market, technology, and challenges," Renewable and Sustainable Energy Reviews, vol. 165, p. 112586, 2022.

[5] W. Sun, Z. Huang, and Y. Wang, "A comparative study of machine learning algorithms for building energy consumption prediction," Energy and Buildings, vol. 261, p. 111956, 2022.

[6] S. Al-Qaness, M. A. A. Al-Ali, and H. J. Al-Ariki, "A multi-stage LSTM model for energy consumption forecasting in smart buildings," Applied Soft Computing, vol. 128, p. 109462, 2022.

[7] P. P. R. da Silva, A. M. Lezama, and G. D. A. e Castro, "IoT-based platform for energy management in university buildings using open-source technologies," Journal of Building Engineering, vol. 44, p. 103282, 2021.

[8] S. Ullah, J. Ahmad, I. U. Haq, S. Rho, and M. A. Al-Rakhami, "A novel deep learning-based framework for energy consumption forecasting in a smart grid," IEEE Access, vol. 8, pp. 182638-182650, 2020.

[9] A. Baregheh, M. A. A. Al-Qaness, and M. I. Al-Hassan, "A hybrid CNN-LSTM model for multi-step ahead energy demand forecasting," IEEE Access, vol. 9, pp. 64239-64251, 2021.

[10] F. Bu, Z. Wang, Y. Liu, and K. J. Li, "A transformer-based model for short-term load forecasting in integrated energy systems," IEEE Transactions on Smart Grid, vol. 13, no. 1, pp. 623-635, 2021.

[11] S. Sulaymanov, M. Talipov, R. Razikov, O. Ilyasov, and O. Kovaleva, "Protection of the Environment from Pollution by Wastewater from Railway Transport Using Natural Sorbents," in ICTEA: International Conference on Thermal Engineering, vol. 1, no. 1, Jun. 2024.

[12] E. Shipacheva, S. Shaumarov, A. Gulamov, and M. Talipov, "Modeling of Interaction of External Enclosing Structures of Buildings with the Internal and External Environments," in ICTEA: International Conference on Thermal Engineering, vol. 1, no. 1, Jun. 2024.

[13] E. Madaliev et al., "Numerical study of axisymmetrical transsonic impact based on the SST turbulence model," BIO Web of Conferences, vol. 145, p. 03034, Jan. 2024, [Online]. Available: https://doi.org/10.1051/bioconf/202414503034.