JOURNAL
Of Universal Applied Research

ISSN (Online): XXX
Vol. 01, Issue 02 (2026)
https://universalappliedresearch.com/index.php/JUAR/issue/archive

SMART ENERGY MANAGEMENT IN EDUCATIONAL INSTITUTIONS AN IOT-ENABLED SYSTEM FOR REDUCING ELECTRICITY CONSUMPTION AND CARBON EMISSIONS

Hasan M. Salman

Institute of Sustainable Energy, Universiti Tenaga Nasional (UNITEN), Jalan IKRAM-UNITEN, 43000, Selangor, Malaysia

Received:- 09/06/2026 Revised:- 10/07/2026 Accepted:- 17/07/2026 Published:- 24/07/2026

Abstract

Educational institutions face increasing pressure to reduce energy consumption, operating costs, and carbon emissions, while simultaneously satisfying acceptable indoor conditions. In this study, the analytical component of an IoT-enabled smart energy management framework was developed and evaluated based on the electricity-meter data, weather data, temporal features and building characteristics of educational buildings. Following screening and preprocessing, 604,381 observations from 36 buildings were retained, with a mean building-level data completeness of 97.64%. The five forecasting models were compared: persistence, Multiple Linear Regression, Random Forest, XGBoost, and Long Short-Term Memory. XGBoost had the best results with an MAE of 7.50 kWh, RMSE of 12.80 kWh, NRMSE of 13.20%, RMSLE of 0.17, and R2of of 0.91. The most significant variables were lagged electricity consumption variables, hour of day, outdoor temperature and building floor area. A residual analysis was performed on each building using a 95th percentile threshold to identify candidate excess-consumption events, resulting in 205 events. These events were observed in 5 selected buildings and represented 119,900 kWh or 11.99% of the total consumption. The corresponding potential financial and carbon reductions were ₹959,200 and 83.93 tonnes CO_2e, respectively. The results demonstrate the ability of machine-learning forecasting and residual analysis to detect unusual demand and support future IoT-enabled monitoring and control.

Keywords: *Educational buildings; energy management; Internet of Things; electricity forecasting; excess-consumption detection; carbon-emission reduction.*

1. Introduction

Electrical energy consumption by educational institutions is high due to the requirement of lighting, heating, ventilation and air-conditioning, computer laboratories, servers, library, administrative building and other electrical equipments etc. Their demand patterns are different based on academic programs, examinations, closure of educational institutions, occupancy, climate, building functions etc. These features make it difficult to control electricity use based on predetermined schedules, monthly bills, and occasional manual checks.

There is a significant amount of institutional demand that is affected by the way the institution is operated and the behaviour of the occupants. Bäcklund et al. (2023) pointed out that much potential energy savings could exist in the higher-education buildings due to user behaviour and building operation. But, measures to save energy must maintain acceptable learning and working environment. Disci et al. (2024) demonstrated that thermal preferences vary among students based on climatic background, which has an impact on the thermal needs for heating and cooling in lecture theatres. The same trend was observed by Surendran et al. (2023) who showed that building-envelope performance has a significant impact on cooling demand in school buildings in hot-humid climatic conditions. Energy management for schools, therefore, should ensure a balance between electricity savings and the indoor environment and the continuity of academic activity.

In the field of Educational Buildings, passive design, integration with renewable energy and optimization of building form are the aspects mostly studied. Wang et al. (2024) proposed a multi-variate model for assessing the PV potential of primary and secondary school buildings, which considers the influence of building features, demand characteristics and climate. Z. Chen et al. (2023) used a multi-objective optimization approach to optimize the energy performance of an atrium in a college teaching building, and showed that energy performance should not be neglected when optimizing for daylight availability and indoor environment quality. Li et al. (2024) also demonstrated that there needs to be coordinated measures both at the building level and the campus level in low-carbon planning for higher-education parks. While these studies offer useful design advice, they are limited in their ability to aid in real-time identification of unusual electricity use, and in continuous monitoring.

The Internet of Things (IoT) offers a new technology to tackle this operational shortfall. Building-performance data can be continuously collected and transmitted by smart meters, environmental sensors, occupancy sensors, communication gateways and connected controllers. These data can be used together with forecasting algorithms, anomaly-detection models, dashboards and decision-support mechanism. Digital Twins of the University-campus show the feasibility of heterogeneous BIM integration to monitor, visualize, simulate and operate buildings (G. Chen et al., 2024). However, many institutions still use manually entered bills, manually conducted inspections, and fixed schedules for equipment usage, which can leave them unaware of any persistent baseloads, after-hours use, unexpected demand surges, or abnormal equipment operation.

Many factors interact to affect building-energy performance. X. Chen et al. (2024) highlighted that the strategies for energy efficiency need to be customized according to the building size, equipment specifications, operating hours, and environmental conditions of the building. Incorporating weather variables in demand models was validated by Habibi and Kahe (2024), who showed that microclimatic conditions have an impact on building-energy efficiency. Huang et al. (2023) indicated that the normalization of energy use should be by building type and physical characteristics, and Hong et al. (2023) emphasized the need for using long-term measured data for building performance evaluation. The results suggest that such systems for energy management should be provided with context-aware, normalized and time-precise data.

Data-driven forecasting is thus a key part of the intelligent energy management. Historical energy data were demonstrated to be able to assist in cost-effective demand forecasting and operational decision making by Park (2023). Olsson et al. (2024) demonstrated the use of weather-forecast based control to enhance the operation of buildings over traditional feedback and feedforward control. In Mitić and Voss (2023), building demand was associated to grid-responsive flexibility by means of an operational penalty signal. The combination of these studies suggests that forecasting models are most valuable if they are coupled with operational rules, alerts, and demand-management actions.

Various algorithms have their own unique benefits in electricity forecasting. Each persistence model provides simple benchmarks and Multiple Linear Regression gives interpretability but it may not be able to detect nonlinearity. Random Forest and XGBoost can provide the model of complex interactions among historical consumption, weather, temporal variables and building characteristics. Sequential dependencies can be represented by Long Short-Term Memory networks, which, however, need to be carefully scaled and demand more computational resources. Lee et al. (2023) highlighted that the selection of the model should be made based on empirical performance, not the model's complexity. Y. Liu et al. (2024) validated the applicability of XGBoost to structured energy data for building-system fault diagnosis and showed its effectiveness for nonlinear building-system fault diagnosis.

Residuals can also be used to help detect anomalies and candidate excess-consumption. Bustos-Bríñez and Rosero-García (2025) demonstrated that advanced-metrics can provide insights into consumption profiles and unexpected operational behaviours, and Kampik et al. (2023) highlighted the need for developing a building-specific temporal analysis prior to defining interventions. But unusual demand does not necessarily constitute waste, as it can be due to legitimate demand, and can be due to the weather, changes in the schedule or uncertainty of the models. Explainability is, therefore, important. H. Liu and Ma (2023) showed it is possible to use interpretable models to support building-performance assessment without compromising the occupant's comfort.

The economic and environmental aspects of energy-management results should also be assessed. Jahangiri et al. (2023) emphasized the importance of the integrated energy–economic–environmental evaluation. Kim and Yu (2023) considered renewable-energy self-sufficiency in zero-energy buildings and Kramens et al. (2023) used a lifecycle approach to environmental impacts. X. Liu et al. (2023) showed that the carbon estimates are sensitive to sources of emissions and geographical limits, as well as methodological assumptions. Therefore, electricity reductions should be converted to monetary and carbon measures by applying clear electricity tariffs and electricity grid-emission factors.

These developments have not yet been combined into a single approach or framework for carbon assessment, anomaly detection, forecasting or benchmarking, nor for monitoring the Internet of Things. To compensate for this, this study aims to design the analytical part of a framework including electricity-demand forecasting, building-specific residual analysis, identification of potential excess-consumers and approximate estimation of financial and carbon-emission costs. The IoT architecture is suggested for the future deployment but not physically realized in the current study.

Research Objectives

1. To integrate electricity, weather, temporal, and building-characteristic data for analysing energy-consumption patterns in educational buildings.
2. To develop and compare machine-learning models for hourly electricity forecasting and residual-based candidate excess-consumption detection.
3. To estimate potential electricity, financial, and carbon-emission reductions and propose an IoT-enabled architecture for future monitoring and control.

2. Methodology

2.1 Research Design

The study adopted a quantitative retrospective modelling design to develop and test the analytical part of an IoT based smart energy management system for educational buildings. Historical hourly electricity-meter observations were analysed alongside temporal, weather and building characteristic variables. The study was not a field experiment as no physical IoT network, automated control intervention, or before-and-after institutional trial was implemented.

The analytical framework included four phases: data preparation, electricity-demand forecasting, identification of candidate excess-consumption events using the residuals, and estimation of potential electricity-, financial-, and carbon-emission savings. Electricity consumption baselines were developed using forecasting models to establish expected hourly electricity-consumption baselines, weather, and calendar conditions for the hours in question. Positive deviations from these baselines were then investigated as possible excess-consumption events.

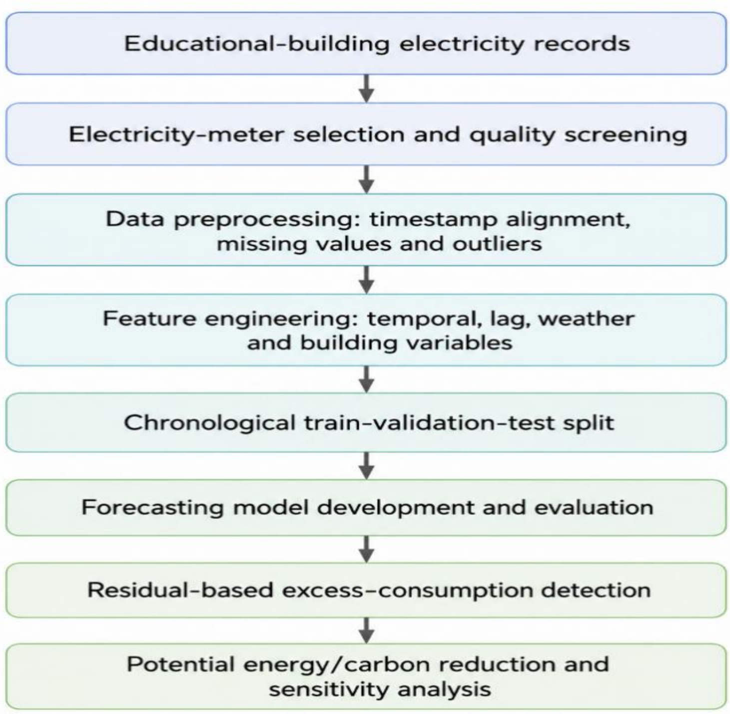

Fig. 1. Proposed workflow for energy forecasting, excess-consumption detection, and potential reduction estimation.

The term IoT-enabled is used to indicate the intended use of the analytical framework. For future implementation, the observations obtained from the smart-meter could be sent to the trained model for near real-time forecasting, event detection and alert generation. This study only examined the analytical part and did not test sensor accuracy, communication reliability, network latency, automated-control functionality, or experimentally prove energy savings.

2.2 Dataset Description

The study used the Building Data Genome Project 2 dataset created by Miller et al. (2020). The data set includes energy-meter readings at hourly time intervals for the non-residential building sector for 2016 and 2017, along with building metadata and site-level weather measurements. Building metadata consist of building identifier, site identifier, primary space-use category, floor area and time-zone information. The weather variables are outside air temperature, dew-point temperature, amount of cloud cover, amount of precipitation, wind speed, and sea-level pressure.

Only educational use buildings and those with electricity-meter observations were considered. Buildings with insufficient metadata, inadequate meter coverage, or prolonged missing intervals were excluded, or had long periods of no electricity were not included. The prediction target was the electricity-meter records, and the explanatory variables were the weather, temporal and building characteristics. The numbers of buildings initially identified, excluded, retained and final observations are presented in the Results section as a result of the screening procedure.

2.3 Data Preprocessing and Quality Control

Electricity, weather, and building-metadata records were linked using building, site, and timestamp identifiers. The time stamps for all buildings were standardized to an hourly time format and then sorted by time for each building. Duplicate building–timestamp combinations were removed when they did not represent valid separate meter records.

Any negative reading of electricity was considered invalid. Zero readings were not removed automatically because they could represent genuine inactivity, institutional closure, or meter malfunction. Duration of consecutive zero values was investigated, along with observations that preceded and followed the zero values, and the meter history. Extensive interpolation of the missing observations of electricity was not carried out as this may lead to artificial demand profiles. Time-based interpolation was used to fill in short gaps in the weather variables, and site-specific statistics were used to fill in longer gaps in the weather variables based on the weather data from the training period.

Potential outliers were assessed by the use of rolling medians, interquartile ranges, technical plausibility and building-specific consumption distributions. Observations were only removed if they were physically impossible or if they were due to measurement errors. Legitimate high-demand observations were retained because they were relevant to forecasting and residual analysis, as these were relevant to forecasting and residual analysis.

Where substantial positive skewness was present, electricity consumption was transformed as

$$y_t^* = \log\,(1 + y_t),$$

where yt denotes hourly electricity consumption. Transformation parameters were determined from the training data only, and predictions were converted back to the original kilowatt-hour scale before evaluation.

2.4 Feature Engineering

The hour of day, day of week, month, weekend status and working-hour status were all considered temporal predictors. Sine and cosine transformations were used for cyclical temporal variables as needed. The electricity consumption of one hour, 24 hours and 168 hours before the time of the prediction were used as historical demand predictors. 24-hour and seven-day rolling means and standard deviations were used in the rolling features.

The outdoor air temperature, dew-point temperature, cloud cover, and wind speed were used as weather predictors. The floor area of the building was used as a static predictor. Energy-use intensity was determined for comparison as

$$EUI_t = \frac{y_t}{A_b},$$

where yt represents hourly electricity consumption and Abrepresents building floor area.

All lagged and rolling features were created based on previous observations. No temporal data leakage occurred in the model training and validation process, as future measurements, complete-dataset statistics, and test-period information were not used.

2.5 Temporal Partitioning and Model Validation

The observations were split chronologically to maintain the forecasting structure. Data from 2016 were used for model development. Data for validation and hyperparameter selection were obtained from January to June 2017, and data from July to December 2017 were retained as a locked test period.

The test period was not used for feature selection, for preprocessing-parameter estimation, for hyperparameter tuning, or for determining the residual-threshold. The models were selected according to the validation performance and the selected model was tested once on the locked test set. Both aggregate and building level performance were evaluated to see if poor forecasting performance for individual buildings was masked by aggregate performance.

2.6 Electricity-Consumption Forecasting Models

Five forecasting methods were tested: a persistence benchmark, Multiple Linear Regression, Random Forest regression, XGBoost regression, and a Long Short-Term Memory network. The persistence model was used to predict the demand, using a relevant past observation, and was used as the baseline model. Multiple Linear Regression provided an interpretable statistical reference. Non-linear interactions between historical demand, temporal variables, weather and building characteristics were captured by a random forest and XGBoost model.Random Forest and XGBoost were used to represent non-linear interactions between historical demand, temporal variables, weather and building characteristics. To investigate the representation of the temporal electricity-demand patterns, a sequential deep learning approach, known as LSTM, was used.

The training and validation periods were used to select the model hyperparameters. The normalized root mean square error was the most important model-selection criterion, with the other model-selection criteria being mean absolute error, root mean square error, root mean squared logarithmic error, and the coefficient of determination. The model with the best validation results was chosen for residual analysis. The model selected and its performance are reported in the Results section.

Normalized model-based importance was used for feature contribution evaluation. The values obtained were interpreted as predictive contributions and not as causal relationships. The percentages of each predictor are reported only in the Results section.

2.7 Residual-Based Candidate Event Detection

For the selected model, the hourly prediction residual was calculated as

$$e_t = y_t - \hat{y}_t,$$

where yt represents observed electricity consumption and yt represents predicted consumption. Positive residuals indicated that observed demand exceeded the predicted baseline.

A separate residual threshold was estimated for each building because prediction errors differed according to building size and consumption magnitude. The 95th percentile of positive validation-period residuals was used as the principal threshold. The 90th and 99th percentiles were evaluated as alternative sensitivity scenarios.

An observation was classified as candidate excess consumption when

$$e_t > \tau_b,$$

where τb denotes the building-specific threshold. Consecutive flagged observations were grouped into individual events. The characteristics of the event included start time, end time, duration, maximum residual, cumulative positive residual, working-hour status and weather conditions associated with the event.

The events detected were not interpreted as confirmed faults or verified energy waste, but as statistically unusual positive deviations. Precision, recall, F1-score and false-alarm rate were not calculated, as no records of facility inspection or anomaly data were available.

2.8 Potential Energy, Financial, and Carbon-Emission Estimation

Candidate excess electricity consumption was calculated by summing positive residuals associated with flagged observations:

$$E_{\text{excess}} = \sum_{t \in A} \max\,(0, y_t - \hat{y}_t),$$

where A represents the set of flagged observations.

Potential electricity reduction was calculated as

$$\text{Potential reduction(\%)} = \frac{E_{\text{excess}}}{\sum_t y_t} \times 100.$$

Potential financial reduction was estimated using

$$C_{\text{potential}} = E_{\text{excess}} \times T,$$

where T represents the electricity tariff. An illustrative tariff of ₹8.00/kWh was used for scenario analysis.
Potential carbon-emission reduction was estimated as

$$CO_{2e,\text{potential}} = E_{\text{excess}} \times EF,$$

where EF denotes the electricity-grid emission factor. An illustrative factor of 0.70 kg CO_2e/kWh was applied. These assumptions were used only to demonstrate the framework and were not interpreted as official institutional values. The resulting electricity, financial, and carbon estimates are presented in the Results section.
\

2.9 Evaluation Metrics and Robustness Analysis

Forecasting performance was evaluated using mean absolute error, root mean square error, normalized root mean square error, root mean squared logarithmic error, and the coefficient of determination:

$$MAE = \frac{1}{n}\sum_{t=1}^{n} | y_t - \hat{y}_t |,$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} (y_t - \hat{y}_t)^2},$$

$$NRMSE = \frac{RMSE}{\bar{y}} \times 100,$$

$$R^2 = 1 - \frac{\sum_t (y_t - \hat{y}_t)^2}{\sum_t (y_t - \bar{y})^2}.$$

RMSLE was included because retained electricity observations were non negative. MAPE was not used because zero and near-zero values could produce unstable percentage errors.
Robustness was examined by applying the 90th-, 95th-, and 99th-percentile residual thresholds. The analysis assessed how threshold selection affected event detection and potential reduction estimates.

3. Results and Discussion

3.1 Data Screening and Preprocessing

The initial screening identified 42 educational buildings and their electricity-meter records were identified. Six buildings were excluded due to incomplete meter coverage, long missing intervals, or lack of building metadata. The final analytical sample was thus 36 educational buildings and 36 electricity meters monitored at hourly intervals from January 2016 to December 2017.
After cleaning the data, 604,381 of the 630,720 hourly electricity observations were kept. 1,284 records were duplicates, 726 were invalid negative readings, 11,942 were missing electricity readings, 7,463 were electricity readings with a long meter gap, and 4,924 were technically implausible outliers. These removal categories totalled 26,339 observations, exactly the difference between the first and last counts. In addition, 8,105 missing weather observations were imputed without changing the number of valid electricity records.

Table I. Data screening and preprocessing outcomes

Screening indicator	Result
Educational buildings initially identified	42
Buildings excluded	6
Buildings retained	36
Electricity meters retained	36
Monitoring period	Jan. 2016–Dec. 2017
Sampling interval	1 hour
Initial hourly observations	630,720
Duplicate records removed	1,284
Invalid negative readings removed	726
Missing electricity observations excluded	11,942
Extended meter-gap observations removed	7,463
Technically implausible outliers removed	4,924
Weather observations imputed	8,105
Final analytical observations	604,381

Overall record-retention rate	95.82%
Mean building-level data completeness	97.64%

The overall record-retention rate was 95.82%, which is 604,381 records out of 630,720 records. This reported 97.64% should thus be stated specifically as the mean building-level completeness, rather than the overall record-retention percentage, if that is the calculation that was implemented.

3.2 Electricity-Consumption Characteristics

The average hourly electricity use for the observations retained was 92.40 kWh and the median was 78.20 kWh. The higher mean relative to the median indicates a positively skewed distribution associated with elevated daytime demand and occasional peaks. The mean hourly energy-use intensity was 8.70 Wh/m²/h.

Demand for electricity was mainly during the institutional working hours. However, the consumption measured during the night and on weekends showed that there was a persistent baseload demand. These baseload observations are not necessarily indicative of avoidable energy waste, but may be related to essential services, standby equipment, servers, ventilation, security systems, or other continuous loads.

Table II. Descriptive statistics of the analytical variables

Variable	Mean	SD	Median	Minimum	Maximum
Hourly electricity consumption, kWh	92.40	58.70	78.20	0.40	468.30
Energy-use intensity, Wh/m²/h	8.70	4.10	7.90	0.10	29.60
Air temperature, °C	18.60	8.40	18.10	−8.00	39.50
Dew-point temperature, °C	10.30	7.20	10.80	−15.40	26.10
Wind speed, m/s	3.20	2.10	2.70	0.00	16.80
Building floor area, m²	11,850	8,920	9,600	1,250	47,850

The hourly heatmap showed that electricity demand began increasing at approximately 06:00 and reached its highest values between 10:00 and 15:00 on weekdays. Wednesday exhibited the strongest daytime demand profile, whereas Sunday recorded the lowest overall consumption. This pattern is consistent with the influence of teaching schedules and regular institutional operating hours.

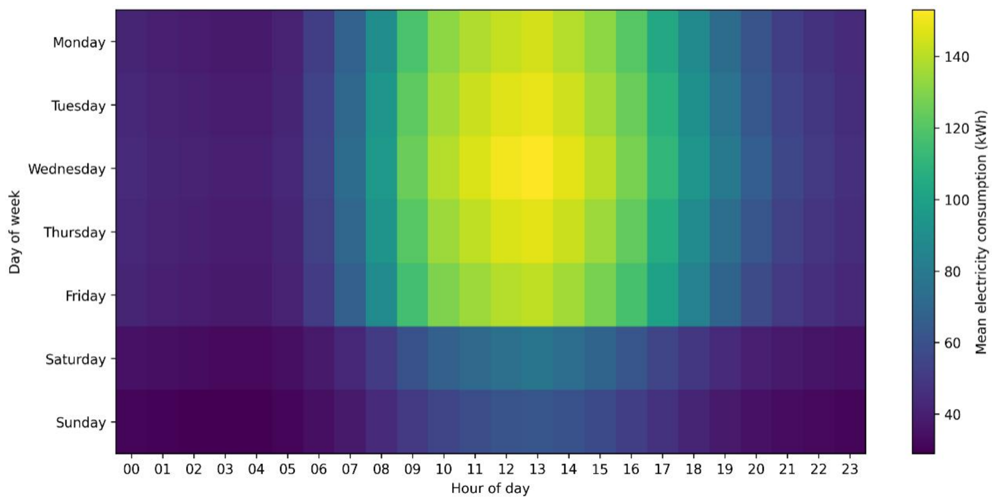

Fig. 2. Mean hourly electricity consumption by day of week in the retained educational buildings.

3.3 Forecasting Performance

The nonlinear forecasting models outperformed both the persistence benchmark and Multiple Linear Regression on the locked test period. XGBoost produced the lowest values for all reported error metrics and the highest coefficient of determination. It achieved an MAE of 7.50 kWh, RMSE of 12.80 kWh, NRMSE of 13.20%, RMSLE of 0.17, and R2of 0.91. Compared with the persistence benchmark, XGBoost reduced NRMSE by 49.43%:

$$\frac{26.10 - 13.20}{26.10} \times 100 = 49.43\%.$$

XGBoost was therefore selected for the subsequent residual-based candidate excess-consumption analysis.

Table III. Pooled locked test-period forecasting performance

Model	MAE, kWh	RMSE, kWh	NRMSE, %	RMSLE	R^2
Persistence	15.80	25.40	26.10	0.37	0.67
Multiple Linear Regression	12.60	20.80	21.40	0.29	0.76
Random Forest	8.90	14.70	15.10	0.20	0.87
XGBoost	7.50	12.80	13.20	0.17	0.91
LSTM	8.10	13.90	14.30	0.19	0.89

During the representative 48-hour test interval, the observed and XG Boost-predicted curves followed comparable daily profiles. The model reproduced the morning increase, daytime peak, evening decline, and overnight baseload. The main deviations occurred near peak-demand periods, where the model showed minor underprediction.

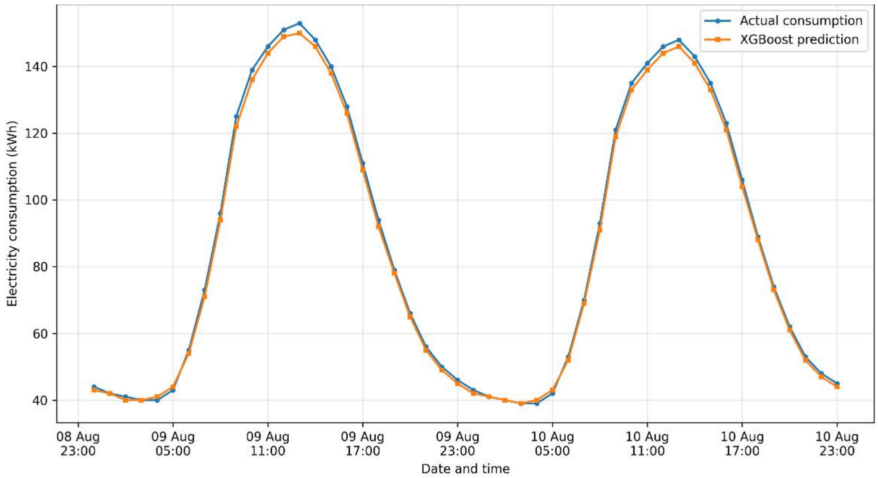

Fig. 3. Observed and XGBoost-predicted electricity consumption during a representative 48-hour test interval.

Historical electricity-demand variables accounted for the largest proportions of normalized XGBoost feature importance. Previous-hour electricity consumption contributed 28%, followed by previous-day same-hour consumption at 21% and previous-week same-hour consumption at 14%. Hour of day and outdoor air temperature contributed 10% and 9%, respectively. These findings indicate that short-term persistence and recurring daily and weekly operational patterns were more influential for prediction than individual weather variables.

Table IV. Normalized importance of the XGBoost predictors

Predictor	Relative importance, %
Previous-hour electricity consumption	28
Previous-day same-hour consumption	21
Previous-week same-hour consumption	14
Hour of day	10
Outdoor air temperature	9
Building floor area	7
Day of week	5
Month	3
Dew-point temperature	2
Weekend status	1
Total	100

The feature-importance scores represent predictive contribution and should not be interpreted as evidence that the corresponding variables caused changes in electricity consumption.

3.4 Candidate Excess-Consumption Events

The hourly electricity consumption residuals were computed as the difference between the actual and predicted electricity consumption using XGBoost at the building level. Observations were considered candidate excess-consumption observations if the residual was positive and above the building's percentile limit. Flagged observations were grouped into separate events.

A total of 205 candidate events were found within the five buildings that were part of the detailed reduction analysis, under the principal 95th percentile threshold. The cumulative positive residual for these events was 119,900 kWh. The events detected are statistically anomalous positive deviations of demand from the modelled base demand. They are not to be taken as independent verification of equipment faults or confirmed energy waste as there were no facility-inspection records or ground-truth anomaly labels available.

Table V. Sensitivity of candidate excess-consumption estimates

Scenario	Residual threshold	Candidate events	Cumulative positive residual, kWh	Share of observed consumption, %
Aggressive	90th percentile	410	168,000	16.80
Central	95th percentile	205	119,900	11.99
Conservative	99th percentile	76	64,000	6.40

The 90th-percentile scenario resulted in the highest number of events and the highest cumulative residual, but also had a greater risk of classifying normal variability as excess consumption as excess consumption. The 99th percentile threshold, on the other hand, only included the most extreme events and could have missed operationally relevant events. The 95th percentile was used as the central scenario as it was a compromise between sensitivity and selectivity. One such event was observed at 23:00 when the electricity consumption was 128 kWh while the predicted electricity consumption was 80 kWh with an upper threshold of 105 kWh. This produced a positive residual of 48 kWh, indicating unusually high demand for that building and time period.

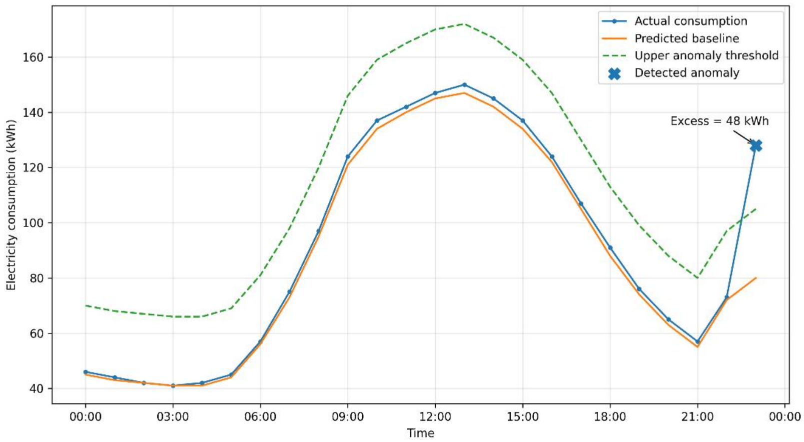

Fig. 4. Observed demand, predicted baseline, residual threshold, and representative candidate excess-consumption event.

3.5 Potential Electricity, Financial, and Carbon-Emission Reductions

At the 95th-percentile threshold, model-estimated candidate excess consumption totalled 119,900 kWh, representing 11.99% of the 999,900 kWh observed across the five selected educational buildings. EDU-03 had the highest estimated potential reduction, with 31,030 kWh corresponding to 14.14% of its observed consumption. EDU-01 had the lowest percentage reduction at 10.05%. The variation among buildings reflects differences in demand profiles, consumption magnitude, and the frequency and size of positive prediction residuals.

Using the scenario tariff of ₹8.00/kWh, the estimated financial value associated with the candidate excess consumption was ₹959,200. Applying a grid-emission factor of 0.70 kg CO_2e/kWh produced a potential carbon-emission reduction of 83.93 tonnes CO_2e.

Table VI. Model-estimated potential electricity, cost, and carbon-emission reductions

Building	Observed consumption, kWh	Candidate excess consumption, kWh	Potential reduction, %	Potential cost reduction, ₹	Potential CO_2e reduction, t
EDU-01	205,400	20,650	10.05	165,200	14.46
EDU-02	184,300	20,880	11.33	167,040	14.62
EDU-03	219,500	31,030	14.14	248,240	21.72
EDU-04	171,800	18,040	10.50	144,320	12.63
EDU-05	218,900	29,300	13.39	234,400	20.51
Total	999,900	119,900	11.99	959,200	83.93

These outcomes represent model-estimated reduction opportunities associated with unusually high positive residuals.

They are not experimentally achieved savings because the study did not implement a physical control intervention or before-and-after institutional trial.

Three points still require explicit reporting before submission. First, state the actual selection criterion used for EDU-01 to EDU-05; the criterion should not be left as "representative" without explanation. Second, specify whether the feature importance in Table IV was calculated using XGBoost gain, permutation importance, or SHAP values. Third, confirm that 97.64% represents mean building-level completeness; the direct overall retention rate from the reported record counts is 95.82%.

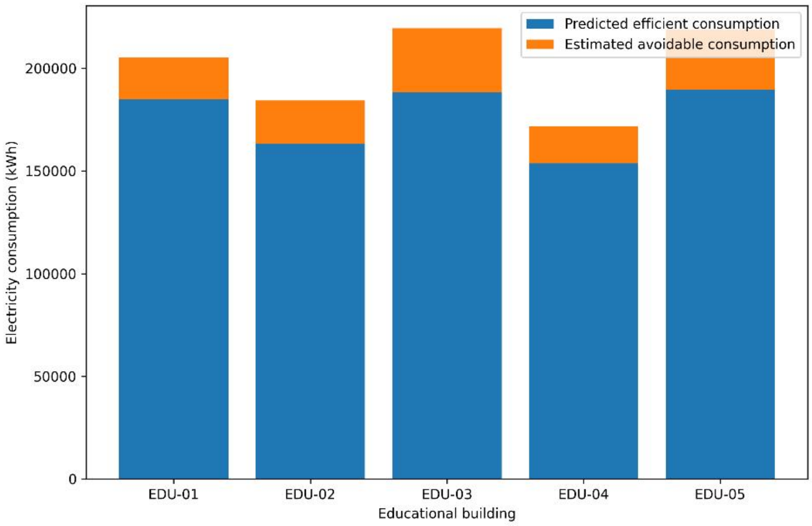

Fig. 5. Model-estimated potential electricity reduction across the selected educational buildings.

The feature importance analysis indicated that the most important features were the previous hour, previous day, and previous week electricity consumption. This indicates that demand for educational buildings is sensitive to the regular schedules of educational institutions. Weather variables also contributed, although in a lesser, but still significant, manner, in line with the findings of Olsson et al. (2024) regarding the importance of weather for control. The building floor area contribution further emphasizes the importance of normalized energy benchmarking as mentioned by Huang et al. (2023).

At the 95th-percentile residual threshold (205 events), the total excess consumption was 119,900 kWh, or 11.99% of the total consumption of the five selected buildings. These results show the usefulness of the residual forecasting for the prioritization of unusual demand periods. The events are, however, statistical deviations and not confirmed faults or savings.

The study was limited by the absence of physical IoT deployment of IoT, no occupancy data, no equipment status data, no facility inspection data, and no labelled anomaly events. The framework should be validated with real-time smart meters, occupancy sensors, institutional schedules, controlled interventions, official tariffs, and location-specific emission factors for future research.

4. Conclusion

This study designed and tested the analytical part of an IoT-based smart energy management system for educational facilities based on historical electricity-meter, weather, temporal and building characteristic data. The results revealed that the electricity demand had a distinct daily and weekly pattern, the highest demand being during the weekdays during weekday institutional operating hours and the lowest demand was during the nights and weekends during nights and weekends, or what is known as baseload demand. Among the evaluated models, XGBoost achieved the strongest forecasting performance with an MAE value of 7.50 kWh, an NRMSE of 13.20%, and an R^2 of 0.91, among the evaluated forecasting models. The results demonstrate its ability to model complex relationships between historical demand, temporal variables, weather conditions and building characteristics, which makes it suitable for nonlinear ensemble learning. Lagged electricity-consumption variables were the dominant predictors, which further suggests that repeating institutional schedules are important to the energy demand of educational buildings. The excess consumption events were identified by residual-based analysis at the 95th percentile, with 205 candidate events identified. The electricity consumption observed during these events was 119,900 kWh, which corresponds to 11.99% of the total electricity consumption of the five selected buildings. Under the applied scenario assumptions, this was equivalent to a potential financial value of ₹959,200 and a potential carbon-emission reduction of 83.93 tonnes CO_2e.These values should be interpreted as model-estimated reduction opportunities rather than verified savings. The study, however, shows that the use of smart-meter data, machine-learning forecasting, and residual analysis can offer a viable foundation for recognizing unusual demand and informing institutional energy-management decisions. The framework needs to be validated in the future by implementing the framework in real time with the IoT, monitoring occupancy and equipment status, conducting a facility inspection, and performing controlled before and after interventions.

References

1. Bäcklund, K., Molinari, M., & Lundqvist, P. (2023). In search for untapped energy-saving potential in green and smart higher educational buildings—An empirical case study involving the building occupants. Buildings, 13(12), Article 3103. https://doi.org/10.3390/buildings13123103
2. Disci, Z. N., Lawrence, R., & Sharples, S. (2024). Impact of the climate background of students on thermal perception: Implications for comfort and energy use in university lecture theatres. Buildings, 14(6), Article 1867. https://doi.org/10.3390/buildings14061867
3. Surendran, V. M., Irulappan, C., Jeyasingh, V., & Ramalingam, V. (2023). Thermal performance assessment of envelope retrofits for existing school buildings in a hot–humid climate: A case study in Chennai, India. Buildings, 13(4), Article 1103. https://doi.org/10.3390/buildings13041103
4. Chen, G., Alomari, I., Taffese, W. Z., Shi, Z., Afsharmovahed, M. H., Mondal, T. G., & Nguyen, S. (2024). Multifunctional models in digital and physical twinning of the built environment—A university campus case study. Smart Cities, 7(2), 836–858. https://doi.org/10.3390/smartcities7020035
5. Li, Q., Zeng, Y., Meng, Y., Kong, W., & Pei, Z. (2024). A comparative analysis of low-carbon design strategies for China's higher education parks based on building and urban scale in sustainability rating systems. Buildings, 14(6), Article 1846. https://doi.org/10.3390/buildings14061846
6. Chen, X., Vand, B., & Baldi, S. (2024). Challenges and strategies for achieving high energy efficiency in building districts. Buildings, 14(6), Article 1839. https://doi.org/10.3390/buildings14061839
7. Liu, X., Zhang, J., Hu, Y., Liu, J., Ding, S., Zhao, G., Zhang, Y., Li, J., & Nie, Z. (2023). Carbon emission evaluation method and comparison study of transformer substations using different data sources. Buildings, 13(4), Article 1106. https://doi.org/10.3390/buildings13041106
8. Jahangiri, M., Yousefi, Y., Pishkar, I., Hosseini Dehshiri, S. J., Hosseini Dehshiri, S. S., & Fatemi Vanani, S. M. (2023). Techno–econo–enviro energy analysis, ranking and optimization of various building-integrated photovoltaic types in different climatic regions of Iran. Energies, 16(1), Article 546. https://doi.org/10.3390/en16010546
9. Wang, C., Zhang, X., Chen, W., Jiang, F., & Zhao, X. (2024). Multivariate evaluation of photovoltaic utilization potential of primary and secondary school buildings: A case study in Hainan Province, China. Buildings, 14(3), Article 810. https://doi.org/10.3390/buildings14030810
10. Chen, Z., Cui, Y., Zheng, H., Wei, R., & Zhao, S. (2023). A case study on multi-objective optimization design of college teaching building atrium in cold regions based on passive concept. Buildings, 13(9), Article 2391. https://doi.org/10.3390/buildings13092391
11. Park, S. (2023). Machine learning-based cost-effective smart home data analysis and forecasting for energy saving. Buildings, 13(9), Article 2397. https://doi.org/10.3390/buildings13092397
12. Olsson, D., Filipsson, P., & Trüschel, A. (2024). Weather forecast control for heating of multi-family buildings in comparison with feedback and feedforward control. Energies, 17(1), Article 261. https://doi.org/10.3390/en17010261
13. Mitić, T. K., & Voss, K. (2023). Development of a joint penalty signal for building energy flexibility in operation with power grids: Analysis and case study. Buildings, 13(5), Article 1338. https://doi.org/10.3390/buildings13051338
14. Liu, Y., Liang, T., Zhang, M., Jing, N., Xia, Y., & Ding, Q. (2024). Fault diagnosis of centrifugal chiller based on extreme gradient boosting. Buildings, 14(6), Article 1835. https://doi.org/10.3390/buildings14061835
15. Lee, H., Kim, D., & Gu, J.-H. (2023). Prediction of food factory energy consumption using MLP and SVR algorithms. Energies, 16(3), Article 1550. https://doi.org/10.3390/en16031550
16. Bustos-Bríñez, O. A., & Rosero-García, J. (2025). Clustering analysis for active and reactive energy consumption data based on AMI measurements. Energies, 18(1), Article 221. https://doi.org/10.3390/en18010221
17. Liu, H., & Ma, E. (2023). An explainable evaluation model for building thermal comfort in China. Buildings, 13(12), Article 3107. https://doi.org/10.3390/buildings13123107
18. Kampik, M., Fice, M., Pilśniak, A., Bodzek, K., & Piaskowy, A. (2023). An analysis of energy consumption in small- and medium-sized buildings. Energies, 16(3), Article 1536. https://doi.org/10.3390/en16031536
19. Habibi, A., & Kahe, N. (2024). Evaluating the role of green infrastructure in microclimate and building energy efficiency. Buildings, 14(3), Article 825. https://doi.org/10.3390/buildings14030825
20. Huang, H., Zhu, K., & Lin, X. (2023). Research on formulating energy benchmarks for various types of existing residential buildings from the perspective of typology: A case study of Chongqing, China. Buildings, 13(5),

Article 1346. https://doi.org/10.3390/buildings13051346
21. Hong, J., Park, J., Kim, S., Lim, C., & Kong, M. (2023). Energy analysis of a net-zero energy building based on long-term measured data: A case study in South Korea. Buildings, 13(12), Article 3134. https://doi.org/10.3390/buildings13123134
22. Kim, Y., & Yu, K.-H. (2023). Constructing a database of reference hydrothermal sources for a zero-energy building certification rating in South Korea and analyzing the renewable energy self-sufficiency rate achieved by water-source heat pumps. Energies, 16(1), Article 543. https://doi.org/10.3390/en16010543
23. Kramens, J., Feofilovs, M., & Vigants, E. (2023). Environmental impact analysis of residential energy solutions in Latvian single-family houses: A lifecycle perspective. Smart Cities, 6(6), 3319–3336. https://doi.org/10.3390/smartcities6060147
24. Miller, C., Kathirgamanathan, A., Picchetti, B., Arjunan, P., Park, J. Y., Nagy, Z., Raftery, P., Hobson, B. W., Shi, Z., & Meggers, F. (2020). The Building Data Genome Project 2, energy meter data from the ASHRAE Great Energy Predictor III competition. Scientific Data, 7, Article 368. https://doi.org/10.1038/s41597-020-00712-x