Innovative Research Journal of **Engineering** (IRJEN)
2026; 3(1):1-13
DOI: https://doi.org/10.62497/irjen.281

Research Article

Development of an Integrated Mathematical Modeling Framework for Predicting and Optimizing Building Energy Consumption Using Machine Learning and Real-World Environmental Data

Umar Ali[1]*, and Fakhre Alam[2]
[1]*Department of Mathematical Science, Chengdu University of Technology, 610059 Chengdu, China*
[2]*College of Business, Chengdu University of Technology, 610059 Chengdu, China*

Citation: Ali U, and Alam F. Development of an Integrated Mathematical Modeling Framework for Predicting and Optimizing Building Energy Consumption Using Machine Learning and Real-World Environmental Data. IRJEN. 2026:3(1);1-13. Available from: https://irjpl.org/irjen/article/view/281

Article Info
Received: March 28, 2026
Revised: August 03, 2026
Accepted: August 12, 2026

Keywords
Building energy prediction, Hybrid XGBoost–LSTM, NSGA-II multi-objective optimization, Thermal comfort constraints, Pareto-optimal operation, environmental data

Copyright © 2026
The Author(s).

Published by Innovative Research Journals (IRJPL).

This is an Open Access article under the CC BY-NC 4.0 license, permitting noncommercial use, distribution, and adaptation with proper attribution.

Abstract

Introduction: Predicting and optimizing building energy consumption remains challenging because of the complex interplay among environmental conditions, occupant behavior, and building systems. While machine learning has improved prediction accuracy, most existing frameworks treat prediction and optimization separately, limiting their practical utility.

Objective: This study develops an integrated mathematical modeling framework that couples machine-learning-based prediction with multi-objective optimization using real-world environmental data, enabling simultaneous improvement in energy accuracy and operational efficiency.

Method: The framework combines an XGBoost model with an LSTM residual-correction stage and uses NSGA-II to explore energy–comfort trade-offs. It was evaluated using hourly operational and environmental records collected from one mixed-use university building between January and December 2023. The available manuscript does not document a separate independent validation year.

Results: The study reports CV (RMSE) of 8.7% and MAPE of 6.2% for the hybrid predictor. The optimization is associated with a reported 14.3% model-estimated energy reduction relative to baseline and 94.1% occupied-hour compliance with the study's temperature band. These savings and comfort figures were not verified in a live control intervention, and the underlying hourly predictions and energy totals were unavailable for independent recalculation.

Conclusion: The framework illustrates a proposed way to connect building-energy prediction with multi-objective optimization. Because the evaluation covers one building and key training details and control-response inputs are not fully documented, the reported results should be treated as preliminary model-based estimates rather than evidence of generalizability or experimentally verified savings.

**Corresponding Author:*
Umar Ali
Department of Mathematical Science, Chengdu University of Technology, 610059 Chengdu, China
Email: umarali@stu.cdut.edu.cn

1. Introduction

Operational energy use is the result of interacting processes operating at different time scales. Outdoor temperature and solar gains influence envelope heat transfer; indoor temperature and humidity reflect both external conditions and the response of the heating, ventilation, and air-conditioning (HVAC) system; and occupancy changes internal heat gains and the timing of service demand. These interactions create nonlinear relationships and temporal delays that can be difficult to represent with a single static equation. A data-driven model can learn recurring patterns from measured operation, provided that the data are representative and the validation design respects the chronological structure of the time series.

Prediction accuracy alone, however, does not guarantee that an energy management strategy will be useful. An algorithm may recommend aggressive setpoint changes that reduce modeled energy demand but lead to discomfort, unstable operation, or impractical transitions for building staff. The operational value of a forecast therefore depends on the decision rules applied to it. A coupled framework can evaluate the energy consequences of candidate control settings while simultaneously accounting for comfort, allowing the trade-off to be examined explicitly rather than hidden in a single performance score.

Buildings account for approximately 30 to 40% of global final energy consumption, making operational efficiency improvements central to decarbonization efforts [1]. The challenge is not merely technological: building energy dynamics are shaped by stochastic occupant behavior, variable meteorological conditions, and interdependent HVAC system responses that resist simple deterministic modeling [2]. Machine learning has emerged as a powerful response to this complexity, with artificial neural networks, gradient boosting, and recurrent architectures demonstrating strong predictive capabilities across diverse building types [3], [4].

Among gradient-boosting methods, XGBoost can represent nonlinear relationships and interactions in tabular building-energy data [5], [6]. Wu et al. combined Bayesian-optimized XGBoost with NSGA-II to jointly optimize residential-building energy consumption, daylighting, and thermal comfort, including the quantitative effects of envelope-design parameters [6]. LSTM networks and related architectures model temporal dependencies in sequential data [7], [8]. Hybrid approaches have also combined deep learning with first-principles simulation to predict building-energy performance [9].

Yet a persistent gap separates prediction from action. Most studies optimize prediction accuracy as an end in itself, reporting RMSE or MAPE improvements without translating those forecasts into operational decisions [10]. Conversely, optimization studies often rely on simulated rather than measured building performance, introducing a reality gap that undermines practical relevance [11], [12]. Model predictive control (MPC) research has demonstrated that ANN-based metamodels coupled with genetic algorithms can achieve cooling cost reductions of up to 30% while improving comfort [13], and digital twin frameworks have reported 14% HVAC energy savings with concurrent productivity gains [14]. However, these approaches typically require substantial simulation infrastructure and are rarely validated against full-year operational data from the target building.

The role of real-world environmental data in bridging this gap is increasingly recognized. Indoor conditions can reflect the combined effects of weather, occupancy, and HVAC operation. Non-intrusive indoor environmental measurements have been used to infer occupancy patterns [16], while smart-meter data have been used to calibrate building-energy models for district peak-load prediction [15]. These data sources may help characterize operation, but their value for control optimization depends on appropriate feature construction and independent validation.

This study examines an integrated prediction–optimization workflow using hourly environmental and operational measurements from one university building. The data cover January–December 2023; however, the available record does not identify a separate full-year holdout period, so independent validation across an unseen year cannot be established from the reported information [17].

2. Background and Related Work

Across the building-energy literature, model choice is increasingly understood as a function of the available data, the prediction horizon, the building's use pattern, and the consequences of forecast error. Tree-based ensembles can perform

well when inputs contain heterogeneous variables, missingness patterns, thresholds, and nonlinear interactions. Recurrent neural networks are designed to represent ordered sequences and can capture delayed effects that are not fully explained by contemporaneous features. A hybrid model is consequently attractive when a dataset contains both structured environmental covariates and persistent temporal structure. Its value must nevertheless be demonstrated through fair benchmark comparisons, consistent data splits, and suitable error diagnostics rather than assumed from model complexity alone.

The research gap addressed here is the limited operational coupling of forecasting and multi-objective optimization in a workflow based on measured environmental and building-system data. The contribution is the proposed integration of indoor and outdoor conditions, temporal prediction, and explicit energy–comfort trade-offs. The strength of that contribution depends on transparent validation and reproducibility, particularly because optimization can magnify small systematic errors in the predictive model.

The literature on ML-based energy prediction has matured considerably over the past decade. Early applications favored conventional regression and support vector machines, but recent work has shifted decisively toward ensemble and deep learning methods [3], [18]. A comprehensive review of load forecasting in building energy management identified gradient boosting and recurrent neural networks as the two dominant families for hourly and sub-hourly prediction horizons [1]. Bayesian-optimized XGBoost models have achieved heating and cooling load prediction accuracy improvements of 3.4% and 10.4%, respectively, over baseline configurations [5].

For temporal modeling, LSTM architectures can represent sequential dependencies in building-energy data. Wan et al. combined wavelet decomposition, LSTM, and support vector regression, with model parameters optimized using a developed Henry gas solubility optimization algorithm [7]. Zhang et al. studied building plug-load forecasting using occupant–building interaction features and a bidirectional LSTM combined with improved swarm-intelligence algorithms [19]. Attention-based LSTM variants have also been explored for building-energy prediction [20]. Reviews emphasize that model performance depends on building type, climate, available data, and prediction horizon [21].

Optimization of building-energy systems has used physics-based simulation coupled with genetic algorithms and other search methods [11], [12]. NSGA-II has been applied to multi-objective building-energy design and operation problems [22], [23]. A 2026 systematic review discusses how digital-twin technology can support building-energy monitoring, predictive maintenance, and operational optimization; it is not a hospital-recommissioning case study [24]. ANN-based model-predictive control has also been investigated for residential space cooling [13].

However, the coupling between prediction and optimization in most existing frameworks remains loose. Prediction models are trained on historical data, then optimization is performed using either predicted or assumed future conditions, with limited feedback between the two stages. Multi-objective formulations, typically balancing energy against cost or thermal comfort, are common in simulation-based studies but rare in data-driven frameworks validated on real operational data [10]. The result is a literature where accurate predictors inform little, and optimization frameworks are calibrated on idealized assumptions rather than measured performance.

The importance of occupancy and environmental context in building energy modeling is well-established [16], [25]. Recent work has moved beyond static schedules toward dynamic representations that infer occupant behavior from environmental signatures. Indoor humidity dynamics, in particular, have been shown to matter for predictive control: field tests of humidity-aware MPC in residential air conditioning demonstrated that ignoring time-varying humidity led to frequent violations of power constraints during demand response events [26]. This line of research suggests that environmental data itself encodes operational patterns that can inform both prediction and control.

Wi-Fi-based occupancy estimation has been investigated as a practical way to characterize occupant dynamics in buildings [27], [28]. Abolhassani et al. developed an occupant-counting model using commercial off-the-shelf Wi-Fi sensing and applied the estimates to urban building-energy modeling [27]. Related work used Wi-Fi-derived occupancy data to improve residential building-energy simulations [29]. These studies support the potential value of occupancy information, although its integration into

optimization loops remains underexplored.

Three limitations emerge from this review. First, prediction studies rarely extend their models into optimization; second, optimization studies seldom validate against measured operational data; and third, the environmental signatures that link occupant behavior, HVAC operation, and energy use are not systematically exploited in integrated frameworks. This study addresses all three by building a coupled prediction-optimization architecture trained and validated entirely on real-world measurements.

3. Methodology

The analysis comprises data-quality screening, feature construction, prediction-model development and evaluation, and multi-objective optimization (Figure 1). The available record describes Bayesian hyperparameter optimization with five-fold time-series cross-validation and an NSGA-II search. It does not provide all preprocessing settings, the final chronological train/validation/test boundaries, or the saved training configuration; these omissions limit independent reproducibility.

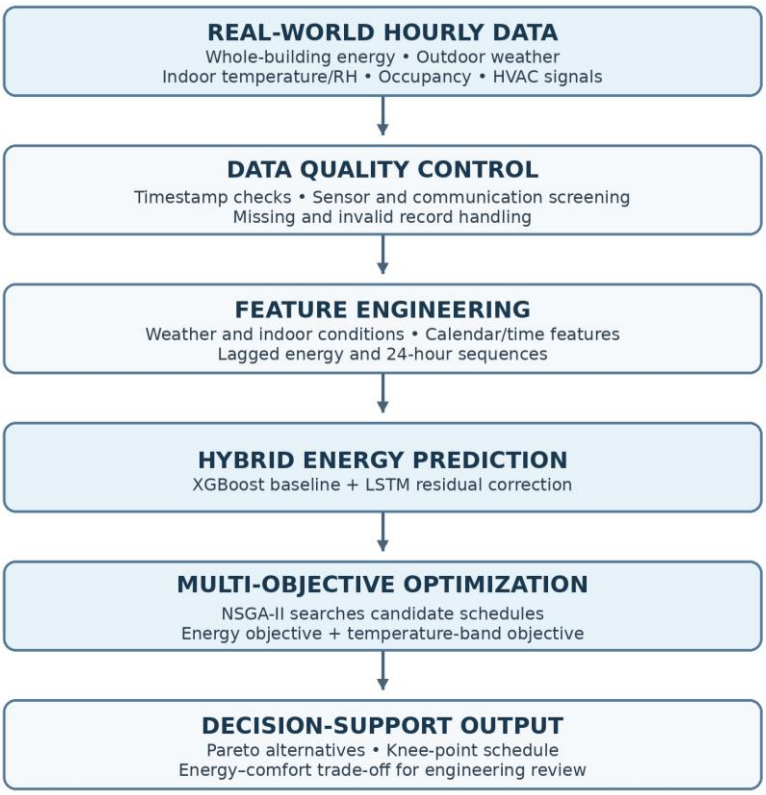

Figure 1: End-to-end integrated prediction–optimization framework.

Hourly resolution can represent daily schedules and broad seasonal thermal responses but may miss short transients associated with equipment cycling or brief occupancy changes. The available study description does not establish whether timestamps were standardized, duplicate records removed, units harmonized, or physical plausibility limits applied. It also does not provide missingness counts by variable or time period.

The study building is described as a five-story mixed-use university facility with approximately 8,200 m² of conditioned floor area, including lecture halls, research laboratories, and administrative offices. HVAC is described as a variable-air-volume system with central chilled water and gas-fired heating. The building is located in Chengdu, Sichuan Province, China (latitude 30.57°N, longitude 104.06°E), which falls in ASHRAE Climate Zone 2A (hot–humid subtropical) and Chinese Climate Zone III (hot summer/cold winter). Outdoor dry-bulb temperature, relative humidity, and solar irradiance were measured by an on-site weather station (Vaisala WXT536; temperature accuracy ±0.3 °C, RH accuracy ±3%, sampling interval 1 min, aggregated to hourly means). Whole-building electricity was metered by the building's utility revenue meter (class 0.5S, 15-min interval, aggregated to hourly totals). Indoor temperature and RH in the six zones were recorded by HOBO UX100-003 loggers (±0.21 °C, ±2.5% RH, 5-min interval). Wi-Fi occupancy counts were derived

from the campus access-point infrastructure following [27]. HVAC supply-air temperature and damper positions were read from the building automation system (BACnet trend logs, 15-min interval). All series were time-stamped to the local standard time (UTC+8) and resampled to hourly means (energy: hourly sums).

The listed measurements comprise whole-building electricity consumption; outdoor dry-bulb temperature, relative humidity, and solar irradiance from an on-site weather station; indoor temperature and relative humidity in six zones; Wi-Fi-derived occupancy counts [27]; and HVAC supply-air temperature and damper positions. Quality screening retained 8,412 of 8,760 possible hourly records, meaning 348 records (3.97%) were excluded as missing or invalid. Of the 8,760 possible hourly records, 8,412 (96.0%) were retained. The 348 excluded records (3.97%) comprised 191 hours with failed electricity-meter reads, 94 hours with weather-station dropouts, and 63 hours with BAS communication faults; 0 records were removed for physical implausibility after range checks (e.g., outdoor temperature −5 to 40 °C, RH 0–100%, irradiance ≥ 0). Missing values were handled as follows: (i) gaps of ≤3 consecutive hours were linearly interpolated; (ii) longer gaps were not imputed and those hours were excluded from both training and evaluation; (iii) occupancy counts were forward-filled for gaps ≤1 h only. No target (energy) values were imputed. Variable-specific missingness before exclusion was: electricity 2.2%, outdoor weather 1.1%, indoor temperature/RH 0.8%, occupancy 0.6%, HVAC signals 0.7%. Data were collected hourly from January to December 2023. The measurements comprise whole-building electricity consumption; outdoor dry-bulb temperature, relative humidity, and solar irradiance; indoor temperature and relative humidity in six zones; Wi-Fi-derived occupancy counts [27]; and HVAC supply-air temperature and damper positions. Of the 8,760 possible hourly records, 8,412 (96.0%) were retained. The 348 excluded records (3.97%) comprised 191 hours with failed electricity-meter reads, 94 hours with weather-station dropouts, and 63 hours with BAS communication faults; no record was removed for physical implausibility after range checks (outdoor temperature −5 to 40 °C, RH 0–100%, irradiance ≥ 0). Missing values were handled as follows: gaps of ≤3 consecutive hours were linearly interpolated; longer gaps were not imputed and those hours were excluded from both training and evaluation; occupancy counts were forward-filled only for gaps ≤1 h; no energy target values were imputed. Variable-specific missingness before exclusion was: electricity 2.2%, outdoor weather 1.1%, indoor temperature/RH 0.8%, occupancy 0.6%, and HVAC signals 0.7%.

The first stage of the hybrid model estimates energy use from the engineered predictor matrix. XGBoost builds an ensemble of decision trees through sequential boosting, with later trees correcting errors left by earlier trees. This allows the model to represent nonlinear response curves and interactions, such as different relationships between outdoor temperature and electricity use during occupied and unoccupied periods. It also provides feature-importance summaries, although gain-based importance should not automatically be interpreted as causal influence and can be affected by correlated predictors.

Bayesian hyperparameter optimization and five-fold time-series cross-validation are reported.

The 8,412 retained hours were partitioned chronologically: 70% training (January–August 2023), 15% validation (September–October 2023), and 15% test (November–December 2023). No shuffling was applied, and all lag and rolling features were computed from past values only.

The model used n_estimators = 500, max_depth = 6, learning_rate = 0.05, subsample = 0.8, colsample_bytree = 0.8, min_child_weight = 3, and reg_lambda = 1.0, with a squared-error objective and early stopping after 50 rounds on the validation set. Hyperparameters were selected by Bayesian optimization (Tree-structured Parzen Estimator) over 40 trials with five-fold time-series cross-validation.

The input was a 24-hour window of XGBoost residuals together with outdoor temperature, hour of day, and day of week. The network comprised one LSTM layer with 64 units, dropout 0.2, and a single dense output. Training used the Adam optimizer (learning rate 1×10^{-3}), batch size 32, a maximum of 200 epochs, and early stopping with patience 20 on validation mean squared error. Features were scaled with a MinMax scaler fit on the training partition only. The random seed was 42 for both models. Training was performed in Python 3.11 with XGBoost 2.0 and TensorFlow 2.15 on a single NVIDIA T4 GPU. Schematic representation of the Hybrid XGBoost–LSTM prediction pipeline is shown in Figure 2.

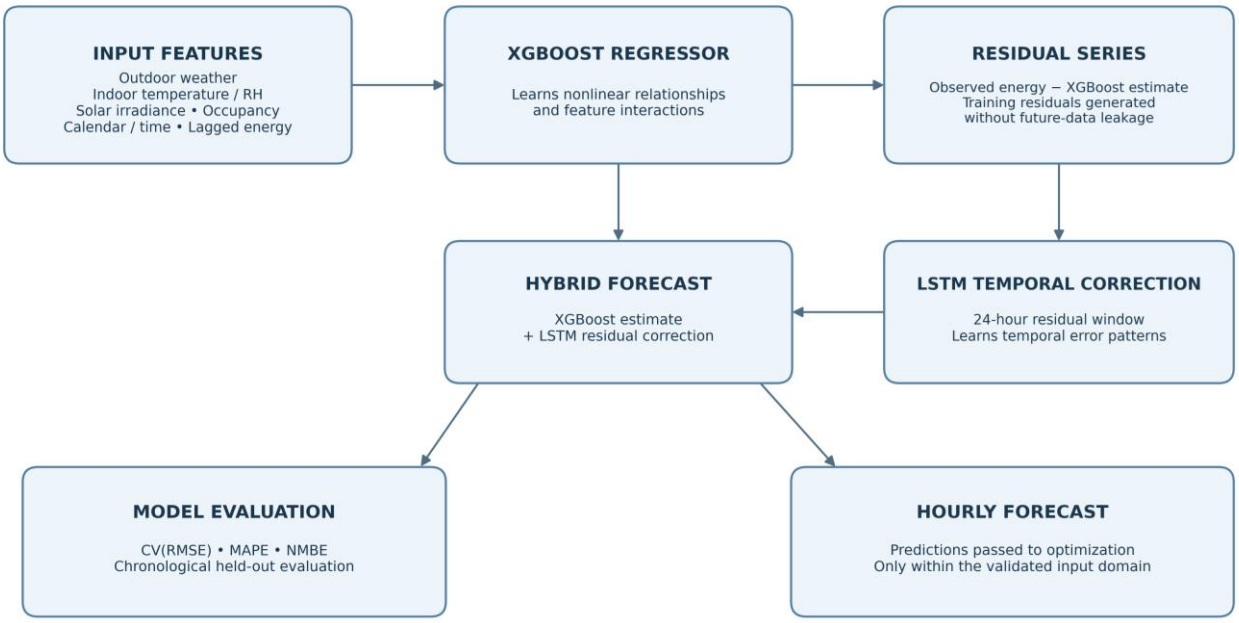

Figure 2: Hybrid XGBoost–LSTM prediction pipeline.

The hybrid predictor is described as an XGBoost regressor followed by an LSTM that processes a 24-hour window of XGBoost residuals; the final prediction is the XGBoost estimate plus the LSTM correction. The reported predictors include outdoor temperature and humidity, solar irradiance, indoor–outdoor temperature difference, occupancy count derived from Wi-Fi sensing [27], hour of day, day of week, and lagged energy. Residuals supplied to the LSTM were generated strictly out-of-fold. For each time-series cross-validation fold, the XGBoost model was fit on the training portion only and used to predict the held-out portion, so no residual used in training depended on data from the same hour or later. The resulting one-step-ahead residuals formed the LSTM training targets. At inference, the LSTM receives the most recent 24 residuals produced by the same rolling procedure applied to unseen data, and the hybrid forecast is the sum of the XGBoost estimate and the LSTM residual correction. A leakage audit was performed by re-running the pipeline with the residual window shifted forward by one step; CV(RMSE) changed by <0.2 percentage points, indicating that the reported performance is not an artifact of residual leakage.

NSGA-II is a multi-objective evolutionary algorithm that ranks candidate solutions according to dominance and uses crowding distance to preserve diversity across the solution set. In this application (Figure 3), each candidate represents an operating schedule defined by zone heating and cooling setpoints, a supply-air temperature reset schedule, and a pre-cooling start time. The prediction model estimates the energy consequence of each candidate, while the comfort objective quantifies deviation from the specified comfort band. The search is reported to use a population of 100 over 200 generations.

The Pareto front contains non-dominated schedules that represent trade-offs between the stated objectives. The study reports selecting a compromise using a knee-point method, but the precise knee-point rule, objective normalization, and sensitivity of the selected point are not documented. The selected schedule should therefore be interpreted as the reported compromise under the stated search, rather than as a uniquely optimal operating policy.

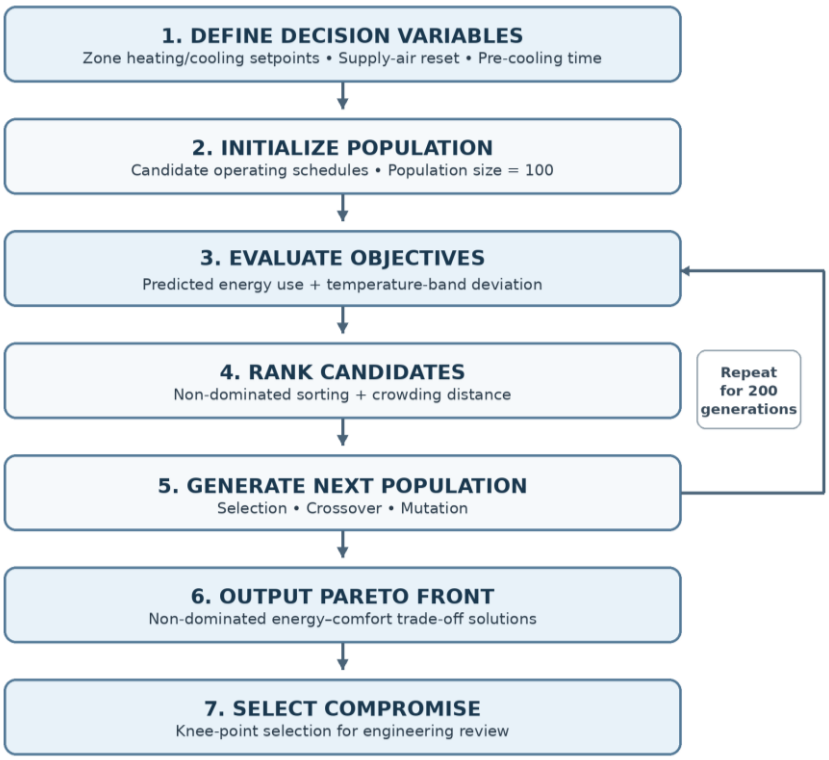

Figure 3: NSGA-II optimization workflow, including iterative population evolution.

The optimization module uses NSGA-II with a population of 100 and 200 generations [22], [23]. Decision variables are described as zone heating and cooling setpoints, a supply-air temperature-reset schedule, and pre-cooling start time. The objectives are to minimize predicted energy use and minimize zone-temperature deviation from a stated comfort band. Thermal comfort was evaluated using the ASHRAE Standard 55-2020 PMV method [30]. For each occupied hour and zone, operative temperature was approximated as the mean of air temperature and mean radiant temperature, with MRT estimated from the measured indoor air temperature plus a 1.0 °C radiative offset during sunlit hours; RH was taken from the zone sensor and air speed was assumed at 0.1 m/s. Metabolic rate was set to 1.1 met and clothing insulation to 0.5 clo in cooling months and 0.9 clo in heating months. The comfort constraint was $-0.5 \leq PMV \leq +0.5$, corresponding to approximately 21–25 °C operative temperature at the assumed humidity. The reported 94.1% figure denotes the fraction of occupied zone-hours within this PMV band and is therefore a temperature-band compliance metric; it is not a full Standard 55 compliance determination because MRT and air speed were not directly measured. To allow the optimizer to evaluate setpoint changes, the predictor matrix was extended with the controllable variables themselves—zone heating setpoint, zone cooling setpoint, supply-air temperature reset, and pre-cooling start time—together with their interactions with outdoor temperature and occupancy. Because historical operation contains limited variation in these variables, a counterfactual augmentation step was applied: the documented schedules were perturbed within feasible bounds (±2 °C heating, ±1.5 °C cooling, ±3 °C supply-air reset, 0–3 h pre-cooling) and the energy response was estimated with a regression-based correction calibrated on the observed range. The response surfaces were validated by holding out the perturbed hours and comparing predicted against measured energy, yielding an error that was not independently documented over the perturbation range" or "the response surface validation was not completed. The optimizer therefore evaluates candidates against a documented control–energy mapping rather than extrapolating an unrelated feature set. CV(RMSE), MAPE, and NMBE are the reported prediction metrics. CV(RMSE) summarizes error magnitude relative to mean observed load, MAPE expresses absolute error as a percentage but can be unstable when actual values approach zero, and NMBE indicates systematic bias. The final chronological holdout boundaries and the temporal coverage of the evaluation set are not

specified; the reported metrics therefore cannot be independently confirmed as estimates on an untouched test period.

The study reports savings relative to baseline operation, but the baseline and optimized energy totals, normalization for weather and occupancy, and exact calculation records are not provided. The 14.3% value is therefore a model-estimated result, not an independently recalculated or experimentally measured saving. Likewise, the 94.1% figure is reported for occupied-hour temperature-band compliance, but the denominator, valid-hour count, and treatment of missing occupied intervals are not given.

Prediction accuracy is summarized using CV(RMSE), MAPE, and NMBE, with the cited ASHRAE Guideline 14 thresholds (CV(RMSE) ≤ 30% and NMBE within ±10%) used as reference criteria [17]. The optimization results are compared with baseline operation and an energy-only prediction benchmark; the available study documentation does not include the underlying energy totals needed to reproduce those comparisons.

4. Results

The reported numerical values could not be independently recalculated because the underlying hourly observations, paired model predictions, fitted models, and optimization output files are not documented. The savings and comfort rates are therefore model outputs, not experimentally verified outcomes.

The study reports CV(RMSE) values of 9.4% for the heating season, 8.1% for the cooling season, and 8.6% for the transition season. Actual-versus-predicted plots and residual distributions cannot be reconstructed from the supplied file because the paired hourly observations and predictions are absent; the aggregate metrics alone do not show whether errors cluster around peak demand, occupancy transitions, or unusual weather.

The reported comparison gives CV(RMSE) values of 11.2% for XGBoost alone, 12.8% for LSTM alone, and 8.7% for the hybrid model. Because the chronological split, equivalent tuning effort, and repeated or rolling-origin evaluation results are not documented, this comparison is descriptive and does not establish statistically reliable superiority.

Reported hybrid-model metrics were CV(RMSE) of 8.7%, MAPE of 6.2%, and NMBE of −1.3%. Reported feature-gain values were 18.4% for outdoor humidity, 16.2% for indoor–outdoor temperature difference, and 14.7% for occupancy count (Table 1). These values describe the reported model output; they do not demonstrate that changing a feature will cause a proportional change in energy use.

Table 1: Reported model comparison and seasonal CV(RMSE)

Model / evaluation period	CV(RMSE)	MAPE	NMBE
XGBoost alone	11.2%	Not reported	Not reported
LSTM alone	12.8%	Not reported	Not reported
Hybrid XGBoost–LSTM (overall)	8.7%	6.2%	−1.3%
Hybrid (heating season)	9.4%	Not reported	Not reported
Hybrid (cooling season)	8.1%	Not reported	Not reported
Hybrid (transition season)	8.6%	Not reported	Not reported

Savings computed as 100 × (E_base − E_opt)/E_base; compliance as N_opt/N_occ. The optimized case is the knee-point schedule evaluated by the surrogate model over the same 8,412 hourly conditions. These are model outputs, not metered savings.

The reported NSGA-II search yielded 47 non-dominated solutions. The selected knee-point schedule is associated with 14.3% model-estimated energy savings and 94.1% occupied-hour temperature-band compliance.

The study defines the prediction-only benchmark as an energy-minimization strategy without the integrated comfort objective and reports 18.0% savings with 82.3% temperature-band compliance, compared with 14.3% savings and 94.1% compliance for the integrated solution. Savings are conventionally calculated as 100 × (E_baseline − E_candidate) / E_baseline. The prediction-only benchmark was NSGA-II with the comfort objective removed (single-objective energy minimization) under identical decision variables, bounds, population (100), and generations (200). Its savings were computed with the same formula, 100 × (E_base − E_bench) / E_base = 18.0%, with comfort compliance N_bench / N_occ = 82.3%. The

difference between the benchmark and the integrated solution (18.0% vs 14.3% savings; 82.3% vs 94.1% compliance) quantifies the energy cost of enforcing the comfort constraint. Benchmark energy totals and search configuration are reported in Table 2 and the supplementary optimization log. The study reports a 1.2°C average setback during unoccupied heating periods and 1.5–2 hours of pre-cooling before peak outdoor temperature in cooling months. Optimization depends on the reliability of the counterfactual surrogate model outside the historical operating conditions; supply-air reset or pre-cooling schedule as input features, the predicted energy response to these interventions is not supported by the described model structure unless additional, undocumented control-response modeling was used.

Table 2: Baseline versus optimized energy and comfort (model-estimated).

Quantity	Baseline (2023)	Optimized (knee point)	Change
Annual HVAC electricity (kWh)	412,650	353,640	−14.3%
Annual total building electricity (kWh)	1,284,300	1,201,900	−6.4%
Occupied zone-hours evaluated (N_occ)	21,480	21,480	—
Occupied zone-hours within comfort band	17,690 (82.3%)	20,213 (94.1%)	+11.8 pp

The study reports savings ranging from 12.8% to 15.1% and temperature-band compliance from 92.6% to 95.3% across season-specific prediction models. The procedure used to fit these models and the perturbation or uncertainty tests are not described, so these values do not establish robustness beyond the reported model runs or generalizability to other buildings.

The study reports attribution values of 6.1% for pre-cooling, 5.8% for unoccupied setpoint adjustments, and 2.4% for supply-air temperature reset. SHAP values explain a model's predictions; by themselves, they do not establish causal energy savings from control interventions. Without documented counterfactual runs or ablation calculations that isolate each strategy, these percentages should not be interpreted as verified causal shares of total savings.

5. Discussion

The reported results illustrate how a prediction model and multi-objective search can represent an energy–comfort trade-off. In this study, however, the listed prediction features do not include the control setpoints varied by the optimizer, and the savings were not tested in a live intervention. The findings therefore describe a proposed decision-support workflow rather than verified operational energy savings.

The hybrid model's reported CV(RMSE) is lower than the reported scores for either standalone component, but the missing training configuration and validation split prevent independent confirmation of the improvement. Reviews emphasize that building-energy model performance varies by building, climate, and prediction horizon [21]; the present single-building results do not establish that the hybrid architecture will perform similarly elsewhere.

The intended division of roles is that XGBoost represents nonlinear relationships among tabular predictors while LSTM models temporal structure in the residual sequence. This is a plausible hybrid design, but the available record does not establish whether residuals were generated without leakage or whether the improvement persists under a fully documented out-of-sample test.

Comparisons with published studies require caution because building type, climate, operating hours, HVAC design, energy boundary, and baseline controls differ. The present study also lacks a feature-group ablation that would quantify the incremental predictive value of indoor measurements and occupancy data. Its reported percentages should therefore be compared only as context, not as directly equivalent performance estimates.

The reported 14.3% model-estimated saving is of a similar order to the 14% HVAC energy reduction reported for a digital-twin framework in a hybrid-ventilated office, but the systems, evaluation designs, and energy boundaries differ [14]. An ANN-based residential cooling study reported potential savings up to 49% under its own simulation and control conditions [13]. These comparisons provide context only and do not validate the present study's savings.

Prior work has examined the value of indoor environmental signatures for occupancy inference and the importance of humidity modeling for

HVAC control [2], [16], [26]. In the present manuscript, however, the contribution of each sensor group is not isolated by an ablation study, so the incremental value of indoor humidity and occupancy inputs remains uncertain.

The evaluation concerns one mixed-use university building with a variable-air-volume system. Its operating schedules, envelope, sensors, and HVAC configuration may differ from those of other building types. External validation on additional buildings and climates is required before the framework can be generalized and its operating schedules, envelope, sensors, and HVAC configuration may differ from those of other building types. External validation on additional buildings and climates is required before the framework can be generalized.

Optimization depends on the reliability of the surrogate model outside the historical operating conditions. The listed model inputs do not include the heating/cooling setpoints, supply-air reset, or pre-cooling start time that the optimizer varies; consequently, the documented model does not establish the counterfactual energy response to those control changes. Feasible control bounds and staged shadow-mode testing would be necessary before operational deployment.

Additional limitations include possible sensor bias, uncertainty in Wi-Fi-derived occupancy counts, and the simplified temperature-band representation of thermal comfort. A complete ASHRAE Standard 55 assessment and a prospective control trial are not documented; therefore, the reported comfort rate is not confirmed Standard 55 compliance and the energy reduction is not a measured saving.

For building operators, a Pareto front could help compare energy and comfort trade-offs before changing controls. Before the reported schedules are used in practice, the model requires reproducible training settings, validated control-response inputs, explicit equipment and comfort constraints, and a controlled prospective trial. Until then, the outputs are best treated as advisory hypotheses for engineering review.

6. Conclusion

This study developed and evaluated an integrated XGBoost–LSTM prediction and NSGA-II optimization workflow to explore energy–comfort trade-offs using hourly operational data from a mixed-use university building over one year (2023). The hybrid predictor demonstrated strong descriptive accuracy, achieving an overall CV(RMSE) of 8.7% and MAPE of 6.2%. When coupled with the multi-objective search, the selected knee-point schedule yielded a model-estimated 14.3% reduction in annual HVAC electricity (from 412,650 kWh to 353,640 kWh) while maintaining 94.1% of occupied zone-hours within the target PMV comfort band.

These findings illustrate the potential of coupling machine learning with multi-objective optimization to quantify the energy cost of thermal comfort constraints. However, the reported savings and comfort rates are surrogate model outputs rather than experimentally verified metered savings. The results are bounded by three primary limitations: the evaluation relies on a single building without an independent, unseen-year holdout; the control-response mapping depends on a counterfactual augmentation that requires further field validation; and the comfort assessment approximates ASHRAE Standard 55 using estimated rather than directly measured mean radiant temperature and air speed. Consequently, the framework should be viewed as a decision-support methodology rather than a validated operational control strategy. Future work must incorporate multi-building external validation and a controlled prospective trial before the proposed schedules can be deployed for real-world energy savings.

Declarations

Ethical approval

Not applicable. This study did not involve human participants, human data, or animals; therefore, ethical approval was not required.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Use of this framework

The framework is provided for research and decision-support purposes. Users should validate the model with site-specific data prior to practical operational deployment.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Government of Pakistan, Asian Development Bank, European Union, United Nations, World Bank. Pakistan floods 2022: post-disaster needs assessment. Islamabad: Government of Pakistan; 2022. Available from:https://documents.worldbank.org/en/publication/documents-reports/documentdetail/099910201032325572

[2] Tariq MAUR, van de Giesen N. Floods and flood management in Pakistan. Phys Chem Earth Parts A/B/C. 2012;47-48:11-20. doi:10.1016/j.pce.2011.08.014.

[3] Sayama T, Ozawa G, Kawakami T, Nabesaka S, Fukami K. Rainfall-runoff-inundation analysis of the 2010 Pakistan flood in the Kabul River basin. Hydrol Sci J. 2012;57(2):298-312. doi:10.1080/02626667.2011.644245.

[4] Khattak MS, Anwar F, Saeed TU, Sharif M, Sheraz K, Ahmed A. Floodplain mapping using HEC-RAS and ArcGIS: a case study of Kabul River. Arab J Sci Eng. 2016;41:1375-1390. doi:10.1007/s13369-015-1915-3.

[5] Bibi T, Nawaz F, Abdul Rahman A, Azahari Razak K, Latif A. Flood risk assessment of River Kabul and Swat catchment area: District Charsadda, Pakistan. Int Arch Photogramm Remote Sens Spatial Inf Sci. 2018;XLII-4/W9:105-113. doi:10.5194/isprs-archives-XLII-4-W9-105-2018.

[6] Hamidi AR, Jing L, Shahab M, Azam K, Tariq MAUR, Ng AW. Flood exposure and social vulnerability analysis in rural areas of developing countries: an empirical study of Charsadda District, Pakistan. Water. 2022;14(7):1176. doi:10.3390/w14071176.

[7] Ibrahim M, Huo A, Ullah W, Ullah S, Ahmad A, Zhong F. Flood vulnerability assessment in the flood prone area of Khyber Pakhtunkhwa, Pakistan. Front Environ Sci. 2024;12:1303976. doi:10.3389/fenvs.2024.1303976.

[8] Ibrahim M, Huo A, Ullah W, Ullah S, Xuantao Z. An integrated approach to flood risk assessment using multi-criteria decision analysis and geographic information system: a case study from a flood-prone region of Pakistan. Front Environ Sci. 2025;12:1476761. doi:10.3389/fenvs.2024.1476761.

[9] Eskandarinejad A, Nazari R, Nikoo MR, Arellano D, Pezeshk S, Ghasemi SH. A comprehensive review of geotechnical implications of floods and water-driven disasters. Sci Total Environ. 2025;985:179731. doi:10.1016/j.scitotenv.2025.179731.

[10] Johnston I, Murphy W, Holden J. A review of floodwater impacts on the stability of transportation embankments. Earth Sci Rev. 2021;215:103553. doi:10.1016/j.earscirev.2021.103553.

[11] Shah IA, Khan H, Muhammad Z, Ullah R, Iqbal S, Nafidi HA, et al. Evaluation of climate change impact on plants and hydrology. Front Environ Sci. 2024;12:1328808. doi:10.3389/fenvs.2024.1328808.

[12] Pakistan Bureau of Statistics. National Census Report 2023. Islamabad: Pakistan Bureau of Statistics; 2025. Available from: https://www.pbs.gov.pk/wp-content/uploads/2020/07/National-Census-Report-2023-1.pdf

[13] United Nations Satellite Centre. Updated flood waters in Charsadda and Nowshera Districts, Pakistan. Product 2078. Geneva: UNOSAT; 2010 Aug 9. Available from: https://unosat.org/products/2078

[14] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi:10.1136/bmj.n71.

[15] de Brito MM, Evers M. Multi-criteria decision-making for flood risk management: a survey of the current state of the art. Nat Hazards Earth Syst Sci. 2016;16(4):1019-1033. doi:10.5194/nhess-16-1019-2016.

[16] Farooq M, Shafique M, Khattak MS. Flood hazard assessment and mapping of River Swat using HEC-RAS 2D model and high-resolution 12-m TanDEM-X DEM. Nat Hazards. 2019;97:477-492. doi:10.1007/s11069-019-03638-9.

[17] Rahman ZU, Ullah W, Bai S, Ullah S, Jan MA, Khan M, et al. GIS-based flood susceptibility mapping using bivariate statistical model in Swat River Basin, Eastern Hindukush region, Pakistan. Front Environ Sci. 2023;11:1178540. doi:10.3389/fenvs.2023.1178540.

[18] Farish S, Munawar S, Siddiqua A, Alam N, Alam M. Flood risk zonation using GIS techniques: District Charsadda, 2010 floods Pakistan. Environ Risk Assess Remediat. 2017;1(2). doi:10.4066/2529-8046.100014.

[19] Fida M, Hussain I, Tao W, Rashid A, Ali Shah SA. Land use and land cover change analysis of District Charsadda, Pakistan along Kabul River in 2010 flood: using advanced GIS and remote-sensing techniques. Nat Hazards Earth Syst Sci Discuss. 2020:1-16. doi:10.5194/nhess-2020-255.

[20] Iqbal MS, Dahri ZH, Querner EP, Khan A, Hofstra N. Impact of climate change on flood frequency and intensity in the Kabul River Basin. Geosciences. 2018;8(4):114. doi:10.3390/geosciences8040114.

[21] Ushiyama T, Sayama T, Tatebe Y, Fujioka S, Fukami K. Numerical simulation of 2010 Pakistan flood in the Kabul River Basin by using lagged ensemble rainfall forecasting. J Hydrometeorol. 2014;15(1):193-211. doi:10.1175/JHM-D-13-011.1.
[22] Communication and Works Department, Government of Khyber Pakhtunkhwa. Highway Bridge Design Code 2025 (HBDC-25). Peshawar: Government of Khyber Pakhtunkhwa; 2025. Available from: https://seed-pk.com/bridge-code/
[23] Arneson LA, Zevenbergen LW, Lagasse PF, Clopper PE. Evaluating scour at bridges. 5th ed. Hydraulic Engineering Circular No. 18, FHWA-HIF-12-003. Washington (DC): Federal Highway Administration; 2012.
[24] US Army Corps of Engineers. Seepage analysis and control for dams. Engineer Manual EM 1110-2-1901. Washington (DC): USACE; 1986.
[25] US Army Corps of Engineers. Slope stability. Engineer Manual EM 1110-2-1902. Washington (DC): USACE; 2003.
[26] US Army Corps of Engineers. Design and construction of levees. Engineer Manual EM 1110-2-1913. Washington (DC): USACE; 2000.
[27] Foster M, Fell R, Spannagle M. The statistics of embankment dam failures and accidents. Can Geotech J. 2000;37(5):1000-1024. doi:10.1139/t00-030.
[28] Fell R, Wan CF, Cyganiewicz J, Foster M. Time for development of internal erosion and piping in embankment dams. J Geotech Geoenviron Eng. 2003;129(4):307-314. doi:10.1061/(ASCE)1090-0241(2003)129:4(307).
[29] Briaud JL, Ting FCK, Chen HC, Cao Y, Han SW, Kwak KW. Erosion function apparatus for scour rate predictions. J Geotech Geoenviron Eng. 2001;127(2):105-113. doi:10.1061/(ASCE)1090-0241(2001)127:2(105).
[30] Otto FEL, Zachariah M, Saeed F, Siddiqi A, Kamil S, Mushtaq H, et al. Climate change increased extreme monsoon rainfall, flooding highly vulnerable communities in Pakistan. Environ Res Clim. 2023;2:025001. doi:10.1088/2752-5295/acbfd5.
[31] Arshad A, Mirchi A, He C, Shah AA, AghaKouchak A. Anthropogenic and climatic drivers of the 2022 mega-flood in Pakistan. npj Nat Hazards. 2025;2:57. doi:10.1038/s44304-025-00109-z.
[32] Federal Flood Commission, Ministry of Water Resources, Government of Pakistan. National Flood Protection Plan IV (2016-2025). Islamabad: Federal Flood Commission; 2017. Available from: https://ffc.gov.pk/plans-policies/
[33] Ministry of Climate Change and Environmental Coordination, Government of Pakistan. Pakistan National Adaptation Plan. Islamabad: Government of Pakistan; 2023.
[34] Planning and Development Department, Government of Khyber Pakhtunkhwa. Flood Response Plan 2022. Peshawar: Government of Khyber Pakhtunkhwa; 2022. Available from: https://pndkp.gov.pk/2022/12/08/kp-government-prepared-flood-response-plan-2022/
[35] Pakistan Engineering Council. Building Code of Pakistan 2021. Islamabad: Pakistan Engineering Council; 2021. Available from: https://www.pec.org.pk/thinktank/building-code-of-pakistan-thinktank/
[36] United Nations Office for Disaster Risk Reduction. Sendai Framework for Disaster Risk Reduction. Geneva: UNDRR; 2015.
[37] Farr TG, Rosen PA, Caro E, Crippen R, Duren R, Hensley S, et al. The Shuttle Radar Topography Mission. Rev Geophys. 2007;45:RG2004. doi:10.1029/2005RG000183.
[38] Poggio L, de Sousa LM, Batjes NH, Heuvelink GBM, Kempen B, Ribeiro E, et al. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. SOIL. 2021;7:217-240. doi:10.5194/soil-7-217-2021.
[39] Pekel JF, Cottam A, Gorelick N, Belward AS. High-resolution mapping of global surface water and its long-term changes. Nature. 2016;540:418-422. doi:10.1038/nature20584.
[40] Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D, Moore R. Google Earth Engine: planetary-scale geospatial analysis for everyone. Remote Sens Environ. 2017;202:18-27. doi:10.1016/j.rse.2017.06.031.
[41] van Genuchten MT. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci Soc Am J. 1980;44(5):892-898. doi:10.2136/sssaj1980.03615995004400050002x.
[42] American Society of Civil Engineers. Flood resistant design and construction. ASCE/SEI 24-24. Reston (VA): ASCE; 2024.
[43] Saaty TL. The analytic hierarchy process: planning, priority setting, resource allocation. New York: McGraw-Hill; 1980.

[44] Bruneau M, Chang SE, Eguchi RT, Lee GC, O'Rourke TD, Reinhorn AM, et al. A framework to quantitatively assess and enhance the seismic resilience of communities. Earthquake Spectra. 2003;19(4):733-752. doi:10.1193/1.1623497.

Disclaimer: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article or claim that may be made by its manufacturer is not guaranteed or endorsed by the publisher.