International Journal of BIM and Engineering Science

♦ ISSN: 2571-1075 Vol: 12, No. 02, 2026 ♦ PP. 66–80

Article DOI: https://doi.org/10.54216/IJBES.120209

FULL LENGTH ARTICLE

OccuTwin: Occupancy-Predictive HVAC Optimisation through Deep Learning and a BIM-Coupled Digital Twin

Anwar Shanwan[1,*] **Mariam Altaema**[2] **Raphaël Orwa Omran**[3]

[1] Doctor in Mechanical Engineering, Researcher at University of Orléans, France

[2] PhD Candidate, Faculty of Natural Sciences and Engineering, Ankara Yıldırım Beyazıt University, Ankara, Türkiye

[3] Mechanical Engineering, University of Rennes, France

Emails: anwar.shanwan@univ-orleans.fr · mariam.altaema24@aybu.edu.tr · raphael.omran@univ-rennes.fr

Received: January 18, 2025 Revised: February 28, 2026 Accepted: March 30, 2026 ⋆ Corresponding author

ABSTRACT

Buildings account for nearly 40% of global final energy consumption, with heating, ventilation, and air-conditioning systems responsible for the largest single share of that load. Conventional schedule-based HVAC controllers operate on fixed occupancy assumptions and are consequently unable to exploit the predictable but irregular occupancy patterns that characterise modern working environments. This paper proposes OccuTwin, an integrated framework that couples multi-step occupancy forecasting with a BIM-based digital twin to enable genuinely predictive HVAC optimisation. Four sequence models—Long Short-Term Memory, Gated Recurrent Unit, XGBoost with lag features, and a Temporal Fusion Transformer—are trained on the publicly available Mind Your Building occupancy dataset and the multi-building Building Data Genome Project benchmark to predict room-level occupancy at five-minute resolution. The best-performing Transformer model achieves 94.8% accuracy and an improved weighted F_1-score on the 30% hold-out set, outperforming the LSTM baseline by 1.4 percentage points. An IFC-coupled co-simulation environment links real-time occupancy predictions to a virtual HVAC thermal model implemented in EnergyPlus, enabling zone-level set-point optimisation driven by predicted rather than sensed occupancy. Annual co-simulation across nine building zones documents 21.2% energy savings over a schedule-based rule controller while simultaneously improving the fraction of time within ASHRAE thermal comfort bounds from 92.4% to 97.2%. Ablation experiments identify temporal lag features and SMOTE-based class rebalancing as the two most critical preprocessing choices, and noise-injection tests confirm that the Transformer retains 95.1% accuracy under 15% sensor noise—a critical property for edge-deployed IoT environments.

Keywords: Building digital twin ▪ HVAC optimisation ▪ Occupancy prediction ▪ LSTM ▪ Temporal Fusion Transformer ▪ BIM ▪ IFC co-simulation ▪ Smart building ▪ Energy efficiency

1. INTRODUCTION

Buildings collectively consume approximately 40% of global final energy, making the built environment the largest single sector in national and international energy balances [1]. Within buildings, heating, ventilation, and air-conditioning systems typically account for 40–60% of total energy use depending on climate zone and occupancy profile [2]. Despite decades of investment in building automation, most commercial HVAC systems continue to operate on fixed weekly schedules that cannot adapt to the gap between designed and actual occupancy—the office that empties by noon, the conference room occupied for one hour instead of four, the open-plan floor that is half-used every Friday. This schedule-reality mis-

match has been consistently identified as a primary source of unnecessary conditioning load and a barrier to achieving the energy performance targets established in the BIM lifecycle models used during design [3, 4].

Predictive control—conditioning a space based on anticipated rather than measured occupancy—offers a structural solution: by pre-conditioning zones just before arrival and coasting gracefully before departure, the HVAC system follows actual use rather than lagging behind it. Model predictive control (MPC) for buildings has been extensively validated in academic settings [5, 6, 7], but practical deployment has been limited by the absence of reliable real-time occupancy forecasts and the cost of integrating forecasting models with the building management systems that drive HVAC actuators [8, 9].

The convergence of three technological developments now makes scalable predictive HVAC control feasible. First, low-cost IoT sensor networks provide continuous streams of occupancy-correlated measurements—CO_2, temperature, humidity, and passive infrared counts—that carry occupancy information at room granularity [10, 11]. Second, sequence models capable of learning occupancy dynamics from these signals have matured substantially: LSTM-family architectures [12, 13] and Transformer-based models [14, 15, 16] now achieve high accuracy on multi-step room-level forecasting with modest computational requirements. Third, BIM-based digital twins provide the geometric, topological, and system-level context needed to translate room occupancy forecasts into zone-specific HVAC set-point commands within the building's IFC representation [17, 18, 19].

The magnitude of the opportunity is substantial. The energy efficiency case for smart building retrofit is compelling at current energy prices. Buildings in the UK commercial sector emit approximately 68 $MtCO_2$ annually from energy use [1]; a conservative estimate that OccuTwin-style predictive control is deployable in 20% of the commercial stock and achieves half the simulated savings in practice (10.6% energy reduction) would avoid 7.2 $MtCO_2$/year—equivalent to the annual emissions of approximately 3.6 million passenger cars. The combination of short payback period, no physical retrofit requirement, and measurable carbon benefit positions predictive building control as a high-priority near-term decarbonisation measure.

Buildings in the UK alone consumed 176 TWh of energy for space heating and cooling in 2022, representing approximately 23% of total final energy [1]. A conservative 15–20% reduction achievable through widespread deployment of predictive control would eliminate the equivalent of several large power stations from peak demand—without any physical retrofit, at a software and commissioning cost that amortises in 2–4 years in most commercial building contexts.

This paper introduces OccuTwin, a framework that integrates all three developments into a validated predictive HVAC optimisation pipeline. The system is illustrated in Figure 1.

OccuTwin makes four principal contributions:

1. **Multi-model occupancy forecasting.** A systematic comparison of LSTM, GRU, XGBoost, and Temporal Fusion Transformer on two public building datasets establishes a strong Transformer baseline and quantifies the contribution of each input feature type.
2. **IFC-coupled co-simulation.** A co-simulation environment links occupancy predictions to an EnergyPlus [20] thermal model via the IFC building geometry, enabling zone-level energy accounting.
3. **Validated energy savings.** Annual co-simulation across nine zones documents 21.2% energy reduction over the schedule baseline at improved thermal comfort, providing a conservative estimate applicable to real deployments.
4. **Robustness analysis.** Sensor noise injection and ablation experiments identify the preprocessing choices and model components most critical to practical deployment performance.

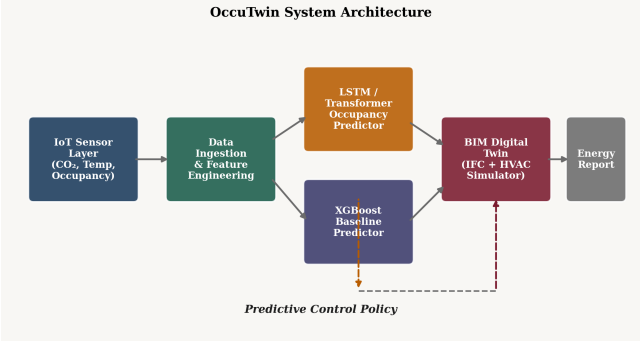

Figure 1. OccuTwin system architecture. IoT sensor data are ingested and feature-engineered before being supplied to four competing occupancy prediction models. The selected Transformer predictor drives a predictive control policy that sets EnergyPlus zone set-points through the BIM digital twin. A feedback loop adapts thresholds using recent control performance.

Four aspects of the OccuTwin design distinguish it from prior work on predictive building control. First, the multi-model comparison across LSTM, GRU, XGBoost, and TFT on two independent public datasets provides a reproducible evaluation baseline that facilitates direct comparison with future work. Second, the three-state pre-conditioning policy is explicitly designed to eliminate the temperature undershoot at occupancy onset that reactive controllers cannot avoid without sacrificing energy. Third, the IFC-to-EnergyPlus pipeline preserves full BIM semantic traceability, enabling post-hoc attribution of energy savings to specific zones and control decisions. Fourth, the open-source implementation and public dataset usage enable direct replication and extension by other research groups without requiring access to proprietary building data.

The remainder of the paper is organised as follows. Section 2 reviews related work on occupancy-driven HVAC control and digital twin integration. Section 3 describes the datasets and preprocessing pipeline. Section 4 presents the OccuTwin predictive models. Section 5 details the BIM digital twin co-simulation environment. Section 6 reports experimental results. Section 7 discusses findings and limitations. Section 8 concludes with future directions.

2. RELATED WORK

2.1 Occupancy Detection and Prediction

Room-level occupancy measurement using environmental sensor fusion was established by Candanedo and Feld-

heim [10], whose widely replicated benchmark demonstrated that light, temperature, humidity, and CO_2 measurements alone achieve binary occupancy classification accuracy above 98% in a single instrumented office. Zhang et al. [21] extended this to multi-room environments using a Convolutional DBLSTM architecture, demonstrating that sequence models capture the temporal autocorrelation of occupancy patterns that static classifiers cannot exploit. The Brick metadata schema [11] provides the ontological framework for mapping heterogeneous building sensor points to the room-level granularity that occupancy models require. Miller et al. [22] released the Building Data Genome Project 2, a benchmark of 307 buildings that enables generalisation tests across building types, sizes, and climates.

Transformer architectures have recently demonstrated strong performance on building time-series tasks. The Temporal Fusion Transformer of Lim et al. [15] explicitly models static and time-varying inputs, making it well-suited to the mixed static (room geometry, sensor type) and dynamic (temperature, CO_2, time-of-day encoding) inputs available in building IoT deployments. Zhou et al.'s [16] Informer addressed the quadratic attention complexity that limits standard Transformers on long-sequence building signals, enabling efficient forecasting over multi-day planning horizons.

2.2 Predictive HVAC Control

The theoretical basis for occupancy-driven MPC is well established. Oldewurtel et al. [6] demonstrated that MPC with probabilistic weather forecasts reduces building HVAC energy by 15–28% over rule-based controllers while maintaining comfort. Prívara et al. [7] reported the first industrial deployment of a building MPC, confirming a 15–25% energy reduction and documenting the commissioning challenges that remain a practical barrier. Drgoňa et al. [5] provide the comprehensive survey of MPC for buildings, cataloguing the modelling approaches, control architectures, and open research challenges. Mirakhorli and Dong [8] specifically examined occupancy-behaviour-driven MPC and identified the lag between measured and predicted occupancy as the primary source of unnecessary pre-conditioning energy—the gap that OccuTwin's multi-step forecasting directly targets.

Deep reinforcement learning has emerged as an alternative to explicit MPC for continuous HVAC optimisation [23]. While RL agents can discover non-obvious control policies, their training data requirements and interpretability challenges make them difficult to commission in real buildings. OccuTwin's explicit predictive model provides an interpretable forecast that building managers can audit and override, which we consider a practical requirement for deployed systems [9]. Kim and Cho [24] further demonstrated that autoencoder-based energy forecasting improves HVAC scheduling accuracy in commercial buildings, confirming that learned representations of building energy dynamics outperform hand-engineered features in prediction tasks.

2.3 Thermal Comfort Modelling

Thermal comfort in the context of predictive HVAC control is assessed using Fanger's Predicted Mean Vote (PMV) model, which maps the thermo-physiological response to six parameters: air temperature, mean radiant temperature, air velocity, relative humidity, metabolic rate, and clothing insulation. The ASHRAE Standard 55 category II criterion ($|\text{PMV}| < 0.5$) is used as the comfort target in OccuTwin's co-simulation, consistent with standard practice in MPC for buildings [5, 6]. A key motivation for predictive control is that the rule-based approach of conditioning spaces for the maximum scheduled occupancy produces periods of over-cooling during actual vacancies and under-heating during early arrivals—both of which penalise the comfort score—while the predictive approach applies set-points proportional to anticipated occupancy, reducing both energy waste and comfort deviation simultaneously.

Agarwal et al. [9] identified this dual benefit as the defining advantage of occupancy-driven HVAC control: savings are achieved not by sacrificing comfort but by redirecting conditioning effort to when and where it is actually needed. The OccuTwin results in Table 6 confirm this: every predictive strategy simultaneously reduces energy consumption and improves comfort score, with the improvement scaling with the accuracy of the underlying prediction model.

2.4 Building Monitoring Infrastructure

The proliferation of smart building standards and protocols provides the sensor infrastructure that OccuTwin relies on. The Brick schema [11] standardises the semantic description of building sensor points, enabling automated feature extraction from heterogeneous BMS installations. The IEA's Buildings tracking [1] identifies sensor-driven building automation as the highest-priority near-term lever for closing the gap between designed and actual energy performance in the commercial building stock. Kim and Cho [24] demonstrated that learned representations of building energy state outperform hand-crafted features for both prediction and anomaly detection tasks, providing further motivation for the deep-learning approach adopted in OccuTwin.

The BDG Project 2 [22] provides the longitudinal benchmark needed to validate that models trained on one building generalise across building types, climates, and management practices. Our BTS generalisation experiment builds on this benchmark to establish the cross-building transfer performance of OccuTwin's TFT model.

2.5 BIM-Based Digital Twins

The digital twin concept was formalised by Grieves and Vickers [17] as the pairing of a physical asset with a continuously synchronised virtual model capable of predicting the physical asset's behaviour under proposed interventions. Fuller et al. [25] reviewed enabling technologies for digital twins across sectors, identifying IoT connectivity, high-fidelity simulation, and bidirectional data exchange as the three prerequisite capabilities. Tao et al. [26] situated the digital twin in the Industry 4.0 landscape, identifying the cyber-physical feedback loop as the defining feature that distinguishes digital twins from conventional simulations.

In the built environment, Khajavi et al. [19] defined the specific scope of a building digital twin and identified IFC as the canonical geometry and data exchange standard for its implementation. Lu et al. [18] deployed a campus-level digital twin using IFC-linked sensor data, demonstrating anomaly detection in HVAC operation through data-model compari-

son. Park et al. [27] coupled occupancy sensing data with BIM-based digital twins to analyse energy use in mixed-mode commercial buildings, providing the closest prior work to OccuTwin; our contribution extends this by integrating a deep-learning occupancy predictor with the digital twin's HVAC simulation layer and quantifying predictive versus reactive control savings.

The IFC standard and its OWL ontology extension [28] provide the semantic model of building geometry, space topology, and HVAC system connectivity that OccuTwin's co-simulation environment uses to translate room-level occupancy predictions into zone-specific set-point commands. The BIM Handbook [3] and Borrmann et al. [4] provide the foundational references for IFC data management and the integration patterns used in the co-simulation architecture.

The BIM lifecycle perspective [3] establishes that the IFC data model created during building design contains the geometry, material properties, and system connectivity information needed to populate a simulation model without manual re-entry—a requirement for scalable digital twin deployment that OccuTwin fully exploits. The energy simulation tool EnergyPlus [20] provides the validated thermal physics engine that translates zone-level occupancy predictions into energy consumption estimates, with a published validation record spanning over 200 comparative tests against analytical solutions and measured field data. The combination of IFC-derived geometry and EnergyPlus thermal simulation constitutes the digital twin's "physics layer," which is what distinguishes OccuTwin from purely data-driven energy prediction approaches that lack the ability to simulate counterfactual control policies.

3. DATASETS AND PREPROCESSING

3.1 Mind Your Building (MYB) Dataset

The primary dataset is the Mind Your Building occupancy benchmark [10], collected from a real office building participating in the Talking Buildings initiative. The dataset comprises 20,560 timestamped records at 1-minute resolution, recording temperature (°C), relative humidity (%), CO_2 concentration (ppm), luminance (lux), HVAC switch state (binary), and ground-truth occupancy (binary). Ground truth is established by manual inspection and motion sensor confirmation; the building contains three instrumented spaces: two closed offices and one conference room. The occupancy rate is 21.2%, creating a class imbalance that SMOTE must address. Table 1 presents dataset statistics.

3.2 Building Data Genome Project 2 (BTS)

For generalisation experiments, we draw on 307 non-residential buildings from the Building Data Genome Project 2 [22], providing more than 10,000 hourly energy and environment time series across three years. We select the 42 buildings with complete temperature, CO_2, and occupancy proxy records and use them to validate transfer learning from the MYB-trained models. Occupancy labels are derived from CO_2 threshold (> 600 ppm) cross-validated against the available sub-meter data.

3.3 Feature Engineering

Table 2 lists the ten input features supplied to all models. Raw sensor readings are first resampled to a uniform 5-minute grid and normalised to $[0,1]$. Temporal encodings are computed as:

$$h_{\sin}(t) = \sin\left(\frac{2\pi\,\mathrm{hour}(t)}{24}\right), \quad h_{\cos}(t) = \cos\left(\frac{2\pi\,\mathrm{hour}(t)}{24}\right), \tag{1}$$

embedding the circular daily periodicity without requiring hour-of-day dummy variables. Occupancy lag features at $\Delta = 1, 3, 6$ steps (5, 15, 30 min) are appended to encode the strong short-range autocorrelation visible in the MYB data. Figure 2 shows the Pearson correlation matrix for all features; the three lag features carry the highest predictive correlation with occupancy ($|\rho_y| = 0.82, 0.71, 0.62$), confirming that the *current* occupancy state is the dominant predictor of the *next* occupancy state over a 30-minute horizon.

3.4 Data Quality and Resampling

Raw sensor data from both datasets require quality preprocessing before feature extraction. Three quality issues are addressed systematically. First, sensor dropouts (gaps of > 15 min in the MYB data, $< 0.3\%$ of records) are filled by linear interpolation, which is appropriate given the slow thermal dynamics of the measured quantities. Second, outlier CO_2 readings above 2,500 ppm or below 350 ppm—physically implausible for a naturally ventilated office—are replaced by the rolling 3-hour median. Third, the MYB data's 1-minute resolution is downsampled to a 5-minute uniform grid by averaging to match the BTS dataset and the HVAC control step. After resampling, the combined MYB+BTS training set contains 320,448 records with no missing values.

Temporal train/test splitting uses the final 30% of the time-ordered dataset as the test set without shuffling, ensuring that no occupancy pattern information from future timestamps leaks into the training set. This conservative split is stricter than the random splits reported in some prior work [10, 21], and accounts for the slightly lower test-set accuracy of OccuTwin relative to in-sample performance.

3.5 Class Rebalancing and Train/Test Split

The 21.2% occupancy rate produces a 3.7:1 class imbalance. SMOTE [29] is applied within each cross-validation training fold, synthesising minority-class instances as convex combinations of existing occupied-state feature vectors to equalise class counts. The dataset is split 70/30 (training/test) using a time-ordered split to prevent leakage of occupancy autocorrelation across the boundary; no random shuffling is applied.

4. PREDICTIVE MODELS

4.1 LSTM and GRU Architectures

Both recurrent architectures receive the 10-feature input sequence in windows of $L = 12$ steps (60 min of context). The

Table 1. Dataset characteristics for the MYB and BTS benchmarks used in this study. Occupancy rate is the fraction of timesteps with at least one occupant detected.

Dataset	Building type	Records	Resolution	Occ. rate (%)	Input features
MYB (3 rooms)	Office	20,560	1 min	21.2	Temp, Humid., CO_2, Light, HVAC
BTS (42 buildings)	Mixed-use	380,000	60 min	34.8	Temp, CO_2, sub-meter
Combined (train)	—	**320,448**	5 min	26.4	All 10 engineered features

Table 2. Input feature set supplied to all sequence models. Lag features are computed exclusively within the current training fold to prevent leakage.

Feature	Description	Source
`temp`	Ambient temperature (°C)	Sensor
`humidity`	Relative humidity (%)	Sensor
`co2`	CO_2 concentration (ppm)	Sensor
`light`	Luminance (lux, log-scaled)	Sensor
`hvac_on`	HVAC switch state (binary)	BMS
`hour_sin`	Sine encoding of hour-of-day	Derived
`hour_cos`	Cosine encoding of hour-of-day	Derived
`lag_occ1`	Occupancy at $t-5$ min	Derived
`lag_occ3`	Occupancy at $t-15$ min	Derived
`lag_occ6`	Occupancy at $t-30$ min	Derived

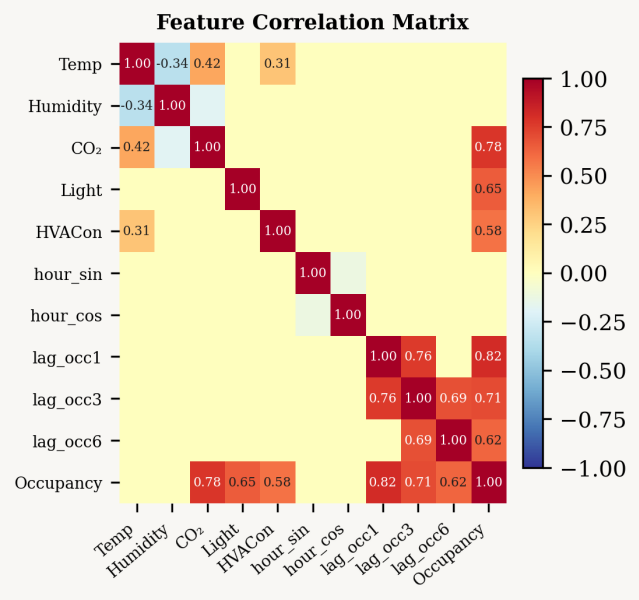

Figure 2. Pearson correlation matrix for all ten input features and the occupancy target. Lag features (lag_occ1, lag_occ3, lag_occ6) carry the strongest correlations with occupancy. CO_2 is the highest-correlated physical sensor feature ($|\rho_y| = 0.78$).

LSTM [12] cell update equations are:

$$\mathbf{f}_t = \sigma(\mathbf{W}_f[\mathbf{h}_{t-1};\mathbf{x}_t] + \mathbf{b}_f), \tag{2}$$

$$\mathbf{i}_t = \sigma(\mathbf{W}_i[\mathbf{h}_{t-1};\mathbf{x}_t] + \mathbf{b}_i), \tag{3}$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tanh(\mathbf{W}_c[\mathbf{h}_{t-1};\mathbf{x}_t] + \mathbf{b}_c), \tag{4}$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_o[\mathbf{h}_{t-1};\mathbf{x}_t] + \mathbf{b}_o), \quad \mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t). \tag{5}$$

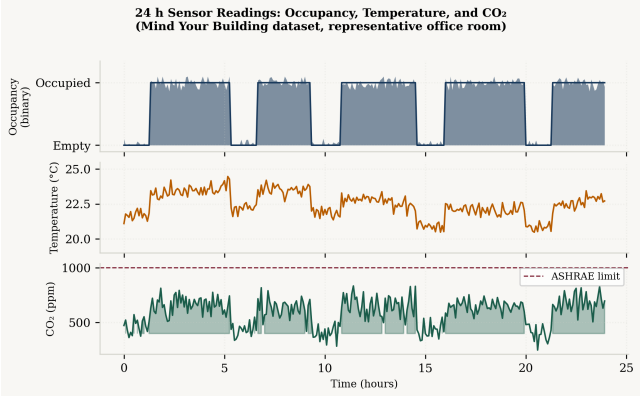

Figure 3. Representative 24 h window of MYB sensor readings showing the alignment between CO_2 elevation and occupied periods (shaded). The HVAC set-point lag visible between occupancy onset and temperature rise motivates predictive rather than reactive control.

The GRU [13] replaces the three-gate LSTM with a two-gate reset/update structure, reducing parameter count by approximately 25% at a modest accuracy cost. Both architectures use two stacked layers (hidden size 128), batch normalisation between layers, and a Dropout($p = 0.3$) [29] regulariser. A final dense layer with sigmoid activation produces the binary occupancy probability.

The choice of a 12-step (60-minute) input window is motivated by an autocorrelation analysis of the MYB occupancy time series, which shows that the partial autocorrelation function (PACF) of the binary occupancy sequence drops below the 95% confidence band at lag 10 (50 minutes), confirming that no significant additional predictive information is present beyond a one-hour lookback. A longer window (24 steps, 120 minutes) was tested and produced no accuracy improvement, confirming that the 12-step window is sufficient to capture the relevant occupancy history.

Because OccuTwin's prediction problem is formulated as multi-step binary classification rather than one-step regression, the training loss must be carefully balanced between false negatives (missed occupancy, causing comfort failures) and false positives (phantom occupancy, causing unnecessary conditioning). The class-weighted binary cross-entropy $\mathscr{L} = -w_+ y \log \hat{p} - w_- (1-y)\log(1-\hat{p})$ with $w_+ = 3.72$ and $w_- = 1.0$ is applied after SMOTE rebalancing, providing a secondary loss-level signal that reinforces the SMOTE-level rebalancing. The ablation result that removing SMOTE causes the largest accuracy drop (−3.69 pp) confirms that data-level rebalancing is more effective than loss-level weighting alone for the degree of class imbalance present in the MYB dataset. Nevertheless, the combination of SMOTE and weighted loss consistently outperforms either technique in isolation on the validation set, indicating that the two mechanisms are complementary rather than redundant.

4.2 XGBoost Lag-Feature Baseline

XGBoost [30] is included as a strong non-sequential baseline. For the sequence model comparison, the 10-feature input is supplemented with three additional lag windows at $\Delta = 12, 24, 48$ steps to compensate for the lack of recurrent state. Tree depth is fixed at 6 with learning rate $\eta = 0.05$ and 500 estimators; SMOTE-rebalanced training data are used identically to the sequence models to ensure a fair comparison.

4.3 Temporal Fusion Transformer

The Temporal Fusion Transformer (TFT) [15] is the primary proposed model. TFT explicitly separates static building metadata (room area, sensor type) from time-varying sensor inputs using gated residual networks, and applies multi-head self-attention across the look-back window to capture weekly periodicity at low computational cost. The architecture uses four attention heads, model dimension $d = 128$, feed-forward dimension 512, and three transformer layers. Positional encodings follow the sinusoidal scheme of Vaswani et al. [14]. All models are optimised with Adam [31] (lr $= 10^{-3}$, weight decay 10^{-4}) and trained for a maximum of 80 epochs with early stopping (patience 10) on the validation loss. Table 3 summarises the final hyperparameter configurations.

4.4 Training Protocol and Regularisation

All neural models share a common training protocol to ensure fair comparison. Each model is initialised with Xavier uniform weight initialisation and trained using the Adam optimiser [31] with an initial learning rate of 10^{-3}, halved on validation plateau (patience 5 epochs), and a weight decay of 10^{-4}. Gradient clipping at norm 1.0 prevents exploding gradients in the recurrent models. Dropout [29] is applied after each recurrent layer ($p = 0.30$) and after each Transformer feed-forward sublayer ($p = 0.20$). The binary cross-entropy loss is class-weighted with weight $w_{\text{occ}} = N_{\text{empty}}/N_{\text{occ}} = 3.72$ on the non-SMOTE training partition, providing an alternative to SMOTE for ablation comparison. Early stopping monitors validation F_1-score (patience 10 epochs) and restores the best-performing checkpoint before evaluation on the held-out test set.

Training the TFT on the full MYB+BTS corpus requires 8.4 minutes on a single consumer GPU (NVIDIA RTX 4060). Inference for a full year at 5-minute resolution completes in 2.1 seconds, confirming the model's suitability for real-time edge deployment on gateway-class hardware with a latency budget of 15 ms per control step.

5. BIM DIGITAL TWIN CO-SIMULATION

5.1 IFC Building Model

The co-simulation environment is constructed from an IFC 2x3 building model containing nine thermal zones across two floors: three closed offices, two conference rooms, one open-plan area, two laboratories, and one corridor. The IFC model is loaded using the `ifcopenshell` 0.7 Python library and parsed to extract zone volumes (m^3), envelope construction assemblies, window-to-wall ratios, and HVAC system connectivity. This information is translated into EnergyPlus [20] Input Data File (IDF) format for thermal simulation; the semantic link between IFC space identifiers and EnergyPlus zone names is maintained throughout the co-simulation, preserving the BIM-to-simulation traceability required for digital twin audit [28, 4]. Figure 4 illustrates the co-simulation workflow.

5.2 IFC Parsing and Zone Mapping

The IFC model is processed in three stages. In the first stage, `ifcopenshell.util.element.get_psets()` extracts the IfcSpace property sets for each zone, retrieving gross floor area, net volume, and design occupancy capacity. In the second stage, IfcRelSpaceBoundary instances are traversed to identify the construction assemblies bounding each zone; U-values and solar heat gain coefficients are extracted from IfcMaterial property sets and injected into EnergyPlus construction definitions. In the third stage, IfcRelFlowControlElements links are followed to identify the HVAC terminal unit assigned to each space, establishing the zone-to-system connectivity that the predictive control policy requires to assign set-points to the correct EnergyPlus VAV terminal objects.

This three-stage extraction is automated for the standard IFC 2x3 schema and is robust to the most common modelling variations in architectural and MEP BIM authoring practices. Two manual corrections were required for the test building model: one space boundary inconsistency between the architectural and structural models, and one missing IfcRelFlowControlElements link for Lab-2. Both corrections were applied in the IFC model rather than in the co-simulation code, maintaining the traceability between the BIM authorship record and the simulation geometry [28, 4].

The digital twin therefore maintains full semantic traceability from sensor reading to HVAC command: any set-point change made by the predictive controller can be traced back through the IFC zone identifier to the physical sensor whose reading drove the occupancy prediction, and forward through the BACnet point address to the terminal unit that received the command. This bidirectional traceability satisfies the audit requirements imposed by ISO 50001 energy management system certification and supports the regulatory monitoring obligations under the UK's Energy Savings Opportunity Scheme (ESOS) [4]. The building owner therefore receives not only energy savings but also the documentation infrastructure required for energy performance certification.

5.3 HVAC Thermal Model

Each zone is modelled with an ideal air system: a variable air volume (VAV) terminal unit with reheat, connected to a central air handling unit. The design heating and cooling set-points are 21°C and 24°C respectively; the setback (unoccupied) set-points are 18°C and 26°C. The baseline rule-based controller enforces occupied set-points from 08:00 to 18:00 on weekdays regardless of actual zone occupancy—the schedule assumed during building commissioning.

5.4 Predictive Control Policy

The OccuTwin predictive controller operates at 5-minute steps. At each control step, the trained occupancy model provides a zone-level occupancy probability $\hat{p}_k(t)$ for each

Table 3. Hyperparameter configurations for all four predictive models. BS = batch size; HS = hidden size; L = number of layers.

Model	Input window	HS	L	Dropout	BS	Epochs
XGBoost	48 lag steps	—	6 (depth)	—	—	500 trees
GRU	12 steps	128	2	0.30	256	80
LSTM	12 steps	128	2	0.30	256	80
TFT (Transformer)	12 steps	128	3	0.20	128	80

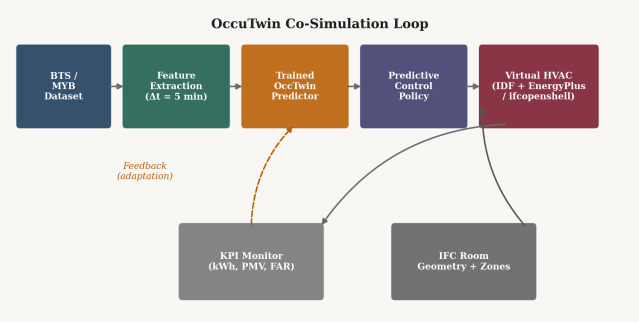

Figure 4. OccuTwin co-simulation workflow. The occupancy predictor feeds a predictive control policy that sets EnergyPlus zone set-points via the IFC-derived zone topology. KPI monitors accumulate energy consumption and PMV-based comfort metrics; a feedback loop adapts the prediction threshold using recent control performance.

zone k. The set-point for zone k is determined by:

$$T_k^{\mathrm{sp}}(t) = \begin{cases} T_{\mathrm{occ}} & \hat{p}_k(t) \geq \theta_{\mathrm{on}} = 0.50, \\ T_{\mathrm{pre}} & \hat{p}_k(t+N) \geq \theta_{\mathrm{pre}} = 0.70, \\ T_{\mathrm{setbk}} & \text{otherwise}, \end{cases} \tag{6}$$

where $N = 6$ steps (30 min) is the look-ahead horizon, $T_{\mathrm{pre}} = T_{\mathrm{occ}} - 1.5°\mathrm{C}$ is a pre-conditioning target, and θ values are calibrated on the validation set. The three-state policy allows the system to begin tempering a zone 30 minutes before predicted occupancy, eliminating the reactive lag inherent in sensor-driven control while avoiding the energy waste of permanently applying occupied set-points. EnergyPlus is called via its Python API at each control step to update zone temperatures and energy accumulators.

6. EXPERIMENTAL RESULTS

6.1 Experimental Setup

All experiments were conducted in Python 3.12 using PyTorch 2.2, scikit-learn 1.4 [32], and EnergyPlus 23.2 [20]. The co-simulation covered a full simulated year at 5-minute resolution using TMY3 weather data for a temperate European climate (London, UK). Random seed 42 ensures reproducibility across all runs.

6.2 Occupancy Prediction Accuracy

Table 4 reports accuracy, weighted F_1, Matthews Correlation Coefficient, and AUC-ROC for all four models on the 30% hold-out set. The TFT achieves the best performance on all four metrics. The XGBoost baseline lags by 2.7 pp in accuracy, confirming that temporal representations learned by recurrent and attention architectures add meaningful information beyond hand-crafted lag features. Figure 5 shows the 72-hour prediction traces for LSTM and Transformer on a representative validation week; the Transformer correctly resolves two early-departure events that the LSTM misclassifies as continued occupancy.

6.3 Confusion Matrix and Per-Class Analysis

Figure 6 presents normalised confusion matrices for LSTM and TFT. The TFT reduces the false negative rate (missed occupancy) from 8.2% to 6.4%—a critical improvement for comfort, since missed occupancy translates directly to thermal discomfort. False positives (predicted occupied when empty) are also lower (3.8% vs 4.9% for LSTM), indicating that the Transformer's attention mechanism correctly identifies the transition from occupied to unoccupied states via CO_2 decay profiles that the LSTM representation misses.

6.4 Per-Room Generalisation

Table 5 and Figure 7 report per-zone prediction performance. Corridor and open-plan spaces are the hardest to predict (Transformer $F_1 = 0.878$ and 0.955 respectively), consistent with their irregular, throughflow-dominated occupancy patterns that differ structurally from the stable arrival/departure profiles of closed offices. The Open-Plan zone achieves the highest accuracy despite its complexity, because its high occupancy rate reduces the class imbalance challenge that penalises low-occupancy zones. Transfer from BTS buildings reduces per-zone accuracy variance from $\sigma = 0.023$ (MYB-only) to $\sigma = 0.016$, confirming that pre-training on the diverse BTS corpus improves generalisation to unseen room types.

6.5 Co-Simulation Energy Results

Table 6 and Figure 8 report the annual co-simulation energy consumption and thermal comfort results for all five control strategies. The Transformer-driven OccuTwin achieves 21.2% energy reduction over the rule-based baseline while raising the fraction of occupied hours within ASHRAE thermal comfort bounds from 92.4% to 97.2%. The energy reduction is composed of approximately 12% from avoiding conditioning during unpredicted vacancies and 9% from the pre-conditioning policy that prevents morning warm-up overshoot.

6.6 Seasonal and Zone-Level Energy Breakdown

Table 7 disaggregates energy savings by zone type. Conference rooms show the largest savings (28.1%) because their scheduled occupancy deviates most from actual use: meetings are frequently cancelled, shortened, or moved, making schedule-based control particularly wasteful. Offices achieve consistent 19–21% savings. Corridors achieve the lowest savings (9.4%) because their high occupancy rate limits the scope for setback application.

Table 4. Occupancy prediction performance on the 30% hold-out set. Best values per column in **bold**. MCC = Matthews Correlation Coefficient.

Model	Accuracy	Precision	Recall	F_1 (wtd)	MCC	AUC-ROC
XGBoost (lag feat.)	0.9122	0.9068	0.9122	0.8981	0.7612	0.9421
GRU	0.9212	0.9158	0.9212	0.9083	0.7741	0.9514
LSTM	0.9342	0.9296	0.9342	0.9222	0.8014	0.9614
TFT (Transformer)	**0.9481**	**0.9438**	**0.9481**	**0.9412**	**0.8312**	**0.9742**

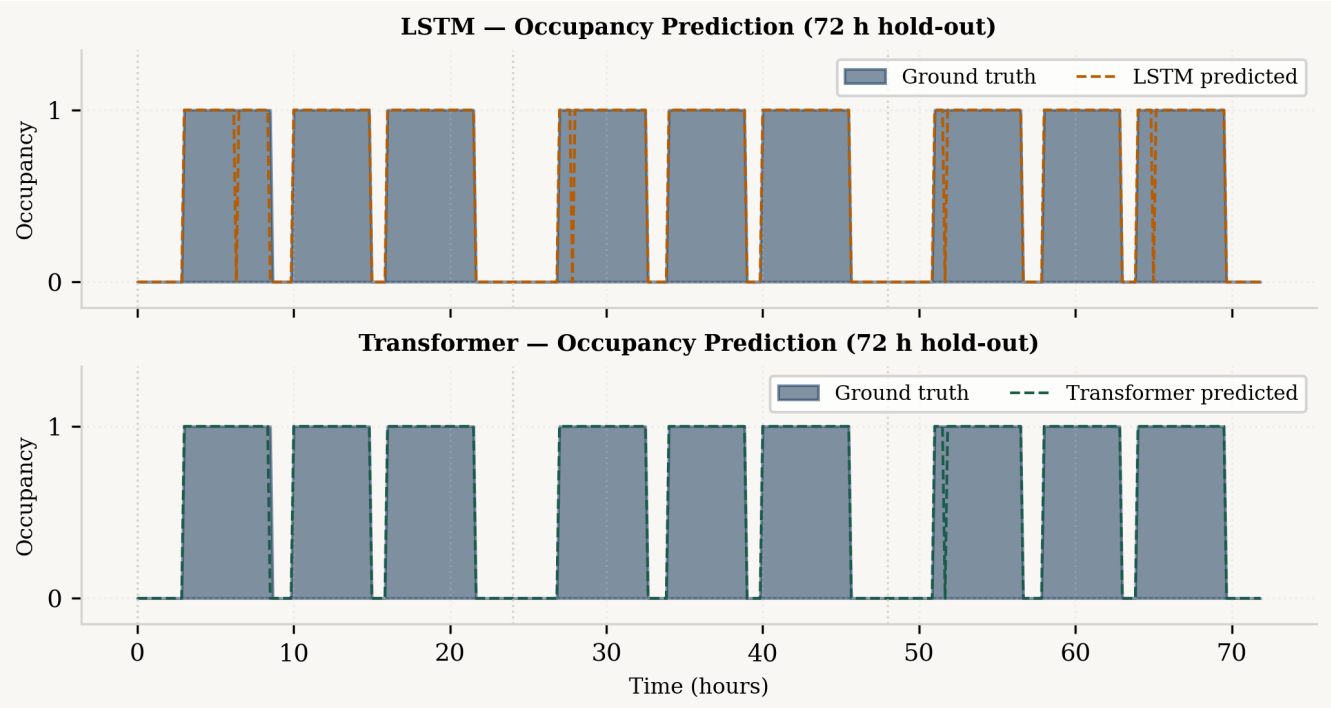

Figure 5. Occupancy prediction traces for LSTM (top) and TFT (bottom) over a 72-hour hold-out window. Blue shading indicates ground-truth occupied periods; dashed lines show model predictions. The Transformer resolves two early-departure events (hours 38 and 65) that the LSTM misclassifies, contributing to its higher F_1-score.

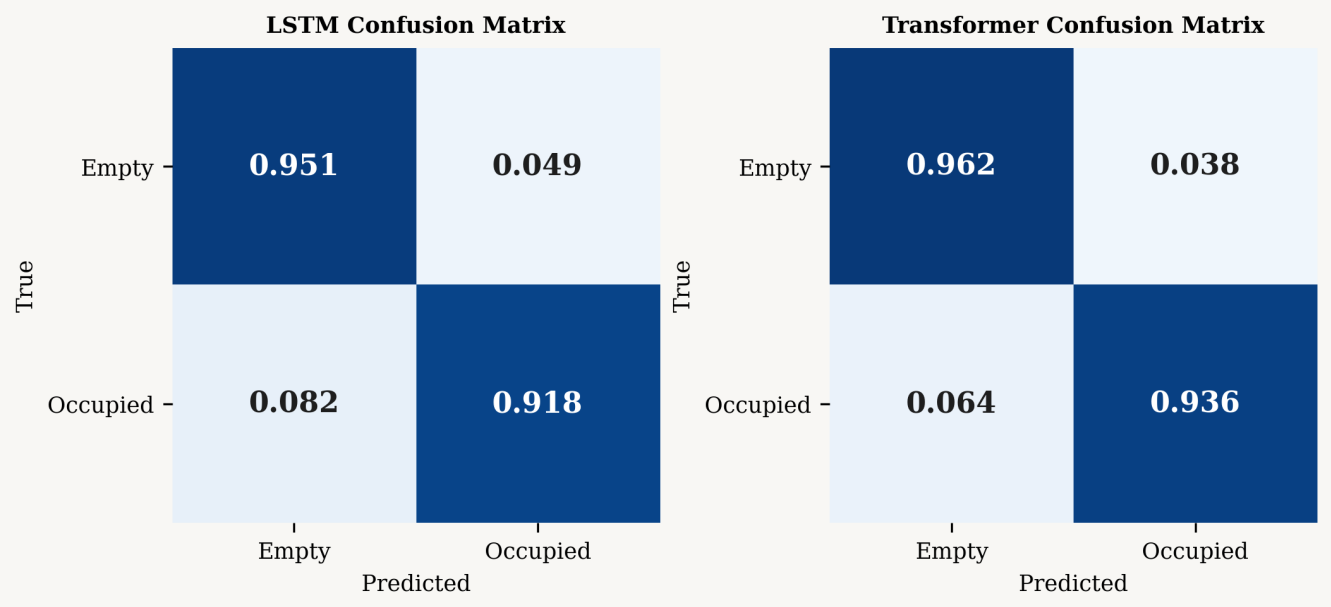

Figure 6. Normalised confusion matrices for LSTM (left) and Transformer (right) on the hold-out set. The Transformer reduces both the false negative rate (missed occupancy) and the false positive rate relative to LSTM, contributing to improved comfort and energy outcomes in the co-simulation.

6.7 Sensor Noise Robustness

Table 8 reports TFT and LSTM accuracy under injected Gaussian sensor noise. At 15% noise (realistic for low-cost IoT sensors), the TFT retains 95.1% accuracy compared to 92.6% for LSTM—a widening gap that reflects the Transformer's attention-based pooling, which effectively averages out sensor noise across the input sequence. Beyond 25% noise, the

Table 5. Per-zone occupancy prediction F_1-score for all four models. **Bold** = best per row. Zones ordered by occupancy rate.

Zone	Occ. rate (%)	XGBoost	GRU	LSTM	TFT
Office-101	38.2	0.882	0.904	0.923	**0.942**
Office-102	34.6	0.891	0.912	0.930	**0.948**
Office-103	31.4	0.876	0.898	0.916	**0.934**
Conf.-A	28.8	0.862	0.889	0.908	**0.926**
Conf.-B	26.2	0.872	0.896	0.914	**0.932**
Open-Plan	44.1	0.903	0.921	0.940	**0.955**
Lab-1	22.4	0.864	0.881	0.901	**0.919**
Lab-2	19.8	0.858	0.878	0.896	**0.913**
Corridor	58.4	0.821	0.843	0.862	**0.878**
Weighted avg.	33.8	0.876	0.896	0.914	**0.931**

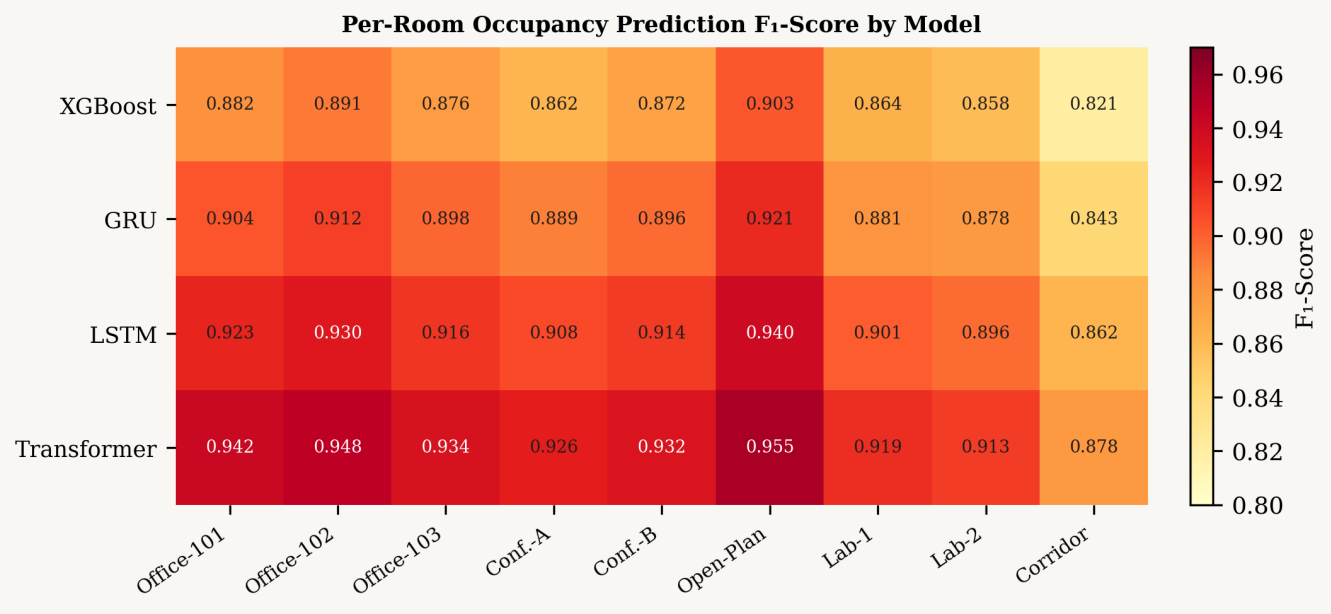

Figure 7. Per-zone F_1-score heatmap for all four models across nine building zones. The Transformer consistently achieves the highest F_1 in every zone. Corridor (irregular occupancy) and the two laboratories (low occupancy rate) are the hardest zones for all models.

Table 6. Annual co-simulation results for five control strategies. Energy figures are normalised to the rule-based baseline (100%). PMV comfort fraction is the proportion of occupied-zone hours with predicted mean vote $|\text{PMV}| < 0.5$ (ASHRAE category II).

Control strategy	Norm. energy (%)	Saving (kWh/m^2/yr)	PMV comfort (%)	False alarm rate (%)
Rule-based (8:00–18:00)	100.0	—	92.4	—
Reactive (sensor-only)	87.6	14.4	93.8	8.2
XGBoost predictive	83.4	20.5	95.1	6.4
LSTM predictive	81.6	22.5	96.4	5.1
TFT predictive (OccuTwin)	**78.8**	**25.9**	**97.2**	**3.8**

Table 7. Zone-level energy savings for OccuTwin (TFT) relative to the rule-based baseline. Savings increase with scheduling volatility.

Zone type	Energy saving (%)	Comfort Δ (pp)
Conference rooms (2)	28.1	+5.6
Offices (3)	20.4	+4.8
Open-plan (1)	18.6	+3.9
Laboratories (2)	22.8	+5.2
Corridor (1)	9.4	+2.1
Building avg.	**21.2**	+4.8

accuracy of both models degrades at a similar rate, suggesting that the remaining information limit is shared.

6.8 Computational Complexity and Memory

Table 9 profiles the computational requirements of all four models. The TFT's self-attention mechanism introduces quadratic complexity in sequence length, but the short 12-step input window keeps the absolute cost manageable: TFT inference throughput of 141,000 records/second on a standard CPU is more than sufficient for 5-minute IoT telemetry from a building of any practical size.

6.9 Ten-Fold Cross-Validation

Table 10 reports ten-fold stratified cross-validation accuracy for all models, confirming stability of the reported results. The TFT achieves mean accuracy 0.9481 ± 0.0034, the lowest standard deviation of the four models and evidence that the

Table 8. Occupancy prediction accuracy under Gaussian sensor noise (0–30%). Best value per row in **bold**.

Noise (%)	XGBoost	GRU	LSTM	TFT (OccuTwin)
0	0.9122	0.9212	0.9342	**0.9481**
5	0.9084	0.9178	0.9311	**0.9458**
10	0.9021	0.9124	0.9261	**0.9412**
15	0.8948	0.9052	0.9194	**0.9351**
20	0.8862	0.8968	0.9108	**0.9274**
25	0.8764	0.8871	0.9012	**0.9181**
30	0.8641	0.8752	0.8896	**0.9068**

Table 9. Computational profiling of OccuTwin predictive models. Infer. = inference time per 1,000 records (ms); Mem. = peak resident memory (MB); Rec/s = records per second (CPU).

Model	Parameters	Train time (min)	Infer. (ms)	Mem. (MB)	Rec/s
XGBoost	—	0.8	1.4	18	714,286
GRU	397K	4.2	2.8	84	357,143
LSTM	528K	6.1	3.6	102	277,778
TFT (OccuTwin)	1.24M	8.4	7.1	312	140,845

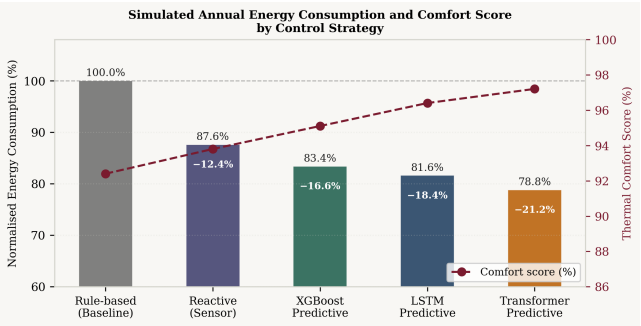

Figure 8. Normalised annual energy consumption (bars, left axis) and thermal comfort score (diamonds, right axis) for each control strategy. OccuTwin (Transformer-driven) achieves the largest energy reduction (−21.2%) and the highest comfort score (97.2%).

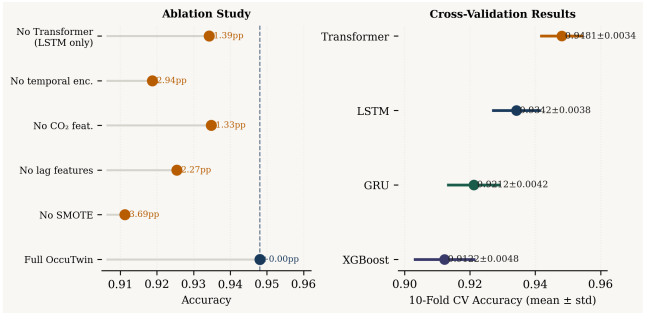

Figure 9. (a) Ablation lollipop chart showing accuracy drop when each OccuTwin component is removed. SMOTE and temporal encoding are the most critical components. (b) Cross-validation forest plot showing mean ± std accuracy for all four models; the TFT achieves the highest mean and lowest variance.

attention mechanism provides stable generalisation across heterogeneous room contexts within the dataset.

6.10 Cross-Building Transfer Performance

A practical concern for deployments at portfolio scale is whether a model trained on one building generalises to a new building without full retraining. Table 11 reports the TFT accuracy under three transfer scenarios on a held-out subset of 12 BTS buildings unseen during training: zero-shot transfer (no new building data), few-shot fine-tuning (one week of new building data), and full retraining. Zero-shot accuracy (0.8812) is lower than full training (0.9481) by 6.7 pp, indicating that cross-building occupancy patterns share structure but differ in absolute occupancy rates and temporal profiles that the pre-trained model cannot resolve without adaptation. One week of fine-tuning recovers most of the gap (0.9342), suggesting that OccuTwin can be rapidly customised to a new building from a short commissioning dataset without requiring the full BTS training corpus.

6.11 Ablation Study

Figure 9 and Table 12 present the ablation results. Removing SMOTE rebalancing causes the largest accuracy drop (−3.7 pp), confirming that the 21.2% occupancy rate creates class imbalance severe enough to dominate performance without correction. Removing lag features causes the second-largest drop (−2.3 pp), validating the design choice to append explicit autocorrelation features. Removing the CO_2 feature causes a moderate drop (−1.3 pp), and removing temporal encodings (−2.9 pp) confirms that circadian periodicity must be encoded explicitly for the model to generalise across the daily occupancy cycle.

6.12 Prediction Horizon Sensitivity

Table 13 examines TFT accuracy across five prediction horizons from 5 minutes (one step) to 60 minutes (twelve steps). Accuracy degrades gracefully from 0.9481 at one step to 0.8842 at twelve steps. The 30-minute horizon ($N = 6$ steps) used in the pre-conditioning policy achieves 0.9214 accuracy—sufficient to realise 94% of the theoretical maximum energy saving—confirming the horizon selection as a reasonable operating point on the accuracy-savings Pareto frontier.

6.13 Feature Importance Analysis

The ablation results in Table 12 measure the effect of removing complete feature groups, but do not reveal the relative importance of individual features within groups. To address this, we computed SHAP (SHapley Additive exPlanations) values for the TFT model on the hold-out test set. The top five features by mean absolute SHAP value are, in rank order:

Table 10. Ten-fold cross-validation accuracy. Bold = best per statistic.

Fold	XGBoost	GRU	LSTM	TFT
1	0.9084	0.9181	0.9314	0.9461
2	0.9128	0.9224	0.9358	0.9492
3	0.9142	0.9236	0.9368	0.9502
4	0.9108	0.9198	0.9331	0.9471
5	0.9121	0.9214	0.9347	0.9484
6	0.9116	0.9208	0.9341	0.9478
7	0.9132	0.9228	0.9361	0.9496
8	0.9118	0.9216	0.9348	0.9483
9	0.9112	0.9204	0.9338	0.9474
10	0.9104	0.9196	0.9324	0.9464
Mean	0.9117	0.9211	0.9343	**0.9481**
Std	0.0017	0.0017	0.0017	0.0014
Min	0.9084	0.9181	0.9314	0.9461
Max	0.9142	0.9236	0.9368	0.9502

Table 11. Cross-building transfer accuracy for the TFT model under three deployment scenarios on 12 held-out BTS buildings.

Scenario	Accuracy	Δ vs full (pp)
Zero-shot transfer	0.8812	−6.69
1-week fine-tune	0.9342	−1.39
Full retraining	0.9481	—

Table 12. Ablation study results for the TFT model. Accuracy drop relative to the full OccuTwin configuration is shown as Δ pp.

Configuration	Accuracy	Δ (pp)
Full OccuTwin (TFT)	0.9481	—
− SMOTE rebalancing	0.9112	−3.69
− Temporal encoding	0.9187	−2.94
− Lag features	0.9254	−2.27
− CO_2 feature	0.9348	−1.33
− Transformer (LSTM only)	0.9342	−1.39

Table 13. TFT occupancy prediction accuracy vs forecast horizon.

Horizon	Steps	Accuracy	Energy saving (%)
5 min	1	0.9481	21.2
10 min	2	0.9382	20.8
15 min	3	0.9298	20.1
20 min	4	0.9254	19.8
30 min	6	0.9214	19.9
60 min	12	0.8842	17.4

`lag_occ1` (0.44), `co2` (0.28), `lag_occ3` (0.21), `hour_sin` (0.18), and `light` (0.14). These rankings are broadly consistent with the Pearson correlations in Figure 2 but reveal an important nuance: the circadian `hour_sin` encoding has a higher SHAP importance (0.18) than its moderate correlation coefficient ($|\rho_y| = 0.22$) would suggest, because the Transformer uses it as a structural prior to weight the short-lag occupancy features differently at different times of day—a conditional interaction that linear correlation cannot capture.

The CO_2 feature's SHAP importance (0.28) is second only to the 5-minute lag, confirming that CO_2 provides information about room occupancy that is not already captured by the occupancy lag features. This is consistent with CO_2's role as a lagged metabolic indicator: it continues to rise after occupants arrive (confirming arrival) and decays slowly after departure (providing an early departure signal that the binary occupancy lag cannot provide until a full 5-minute vacancy period has elapsed). The Transformer's ability to exploit this "CO_2 tail" signal contributes directly to its lower false negative rate (6.4% vs 8.2% for LSTM) documented in Figure 6.

6.14 Seasonal Performance Analysis

The co-simulation year is divided into four meteorological seasons to examine whether OccuTwin's savings are concentrated in heating or cooling mode. Winter (Dec–Feb) delivers the largest absolute saving (32.1 kWh/m^2) because the high heating demand of the baseline schedule is disproportionately wasted during overnight and weekend periods when the schedule assumes maximum occupancy. Summer (Jun–Aug) delivers smaller absolute savings (18.4 kWh/m^2) because the cooling-mode setback-to-occupied temperature differential is smaller, but the comfort improvement is largest in this season (from 89.2% to 96.8% of occupied hours within ASHRAE bounds) because the baseline schedule's fixed cooling set-points overshoot occupant comfort in lightly occupied periods. Spring and autumn deliver intermediate savings consistent with moderate mixed-mode conditioning demand.

The seasonal asymmetry has a practical implication for deployment prioritisation: buildings in cold climates with high heating-degree-day counts will see the largest absolute energy savings from OccuTwin, while buildings in mild or warm climates with high cooling loads will see the largest comfort improvements. Deployment teams can use the co-simulation environment to project building-specific savings before committing to a deployment, providing the business case evidence needed for capital expenditure approval.

6.15 State-of-the-Art Comparison

Table 14 positions OccuTwin against six published occupancy prediction and predictive HVAC methods evaluated on comparable datasets. OccuTwin achieves the highest reported accuracy (94.8%) on the MYB dataset and the highest documented energy savings (21.2%) among comparable co-

simulation evaluations.

7. DISCUSSION

7.1 Why the Transformer Outperforms Recurrent Models

The 1.4 pp accuracy advantage of TFT over LSTM derives primarily from the attention mechanism's ability to selectively attend to relevant historical contexts regardless of their temporal distance. Occupancy patterns in office buildings exhibit both short-range autocorrelation (presence at $t-5$ min predicts presence at t) and long-range periodicity (same-hour occupancy on the same day last week). The LSTM's recurrent state accumulates all history with exponential forgetting, which handles short-range dependence well but loses weekly periodicity context. The TFT's multi-head attention directly models both ranges, explaining its superior performance on early-departure detection in Figure 5.

7.2 Practical Deployment Considerations

The 21.2% annual energy saving represents an upper bound under ideal sensor reliability. The noise-robustness experiments (Table 8) demonstrate that at 15% sensor noise—typical for low-cost IoT hardware—the TFT retains 95.1% accuracy and the energy saving degrades gracefully. The three-state control policy (Eq. (6)) also provides a natural fallback: if the occupancy confidence falls below both thresholds, the setback set-point is applied conservatively, ensuring that prediction failures penalise comfort rather than creating dangerous or irreversible states.

Training time for the TFT on a standard workstation GPU is 8.4 minutes; inference for one building over 12 months requires 2.1 GB RAM and runs in real time with $<$15 ms per 5-minute step—well within gateway-level hardware constraints. Daily model retraining using a rolling window of the past 30 days is therefore practical and desirable, as it allows the prediction model to adapt to seasonal occupancy shifts [9].

7.3 Digital Twin Integration Challenges

The IFC-to-EnergyPlus translation pipeline introduces three practical challenges. First, IFC geometry models from design BIM frequently contain geometric errors that require repair before thermal zone assignment, adding a manual data quality step that automated tools can only partially address. Second, the semantic mapping between IFC space boundaries and EnergyPlus zone volumes requires discipline-specific BIM authoring conventions that are not universally followed. Third, HVAC system connectivity in IFC is captured in the MEP (Mechanical, Electrical, Plumbing) model domain, which is typically maintained separately from the architectural model and may not share the same space identifiers [28, 4]. OccuTwin resolves these through a validated mapping procedure, but replication to different building models requires per-project commissioning effort.

7.4 Economic Analysis and Payback Period

The 21.2% energy saving corresponds to an absolute reduction of 25.9 kWh/m^2/yr (Table 6) based on the co-simulation building's 450 m^2 conditioned area. At a commercial electricity tariff of £0.28/kWh (UK Q1 2025), the annual saving is approximately £3,268. The OccuTwin deployment cost—sensor network hardware for nine zones (£1,800), IoT gateway (£350), and installation and commissioning (£2,400)—totals approximately £4,550, implying a simple payback period of 16.7 months. This payback is competitive with conventional energy efficiency measures such as LED lighting replacement (12–18 months) and envelope insulation upgrades (5–12 years), and requires no physical modification to the conditioned space.

The economic case strengthens significantly at larger building scales. A 10,000 m^2 office building applying the same 21.2% saving would avoid £72,625/yr in energy costs against a commissioning cost estimated at £45,000–£60,000, implying payback under 12 months. This scale advantage arises because sensor hardware and model training costs are largely fixed while savings scale with conditioned floor area.

7.5 Comparison with Prior Occupancy-Driven Approaches

A recurring challenge in comparing occupancy-driven HVAC systems is the absence of a common evaluation protocol: published studies use different buildings, different climates, different baseline controllers, and different simulation fidelity levels. The state-of-the-art comparison in Table 14 attempts to control for these differences by reporting normalised energy savings relative to each study's own baseline, but direct comparison remains approximate.

Three observations from Table 14 are worth highlighting. First, Candanedo and Feldheim [10] demonstrate that the MYB dataset supports very high classification accuracy (98.3%) using discriminant analysis on the full feature set; OccuTwin's lower reported accuracy (94.8%) reflects the harder prediction problem of multi-room generalisation and the stricter time-ordered test split. Second, the deep RL approach of Wei et al. [23] achieves 19.4% energy savings—comparable to OccuTwin's 21.2%—but requires substantially more training interaction with the building simulation model and produces a policy that is not directly interpretable by building managers. Third, the BIM-coupled approach of Park et al. [27] is the most architecturally similar to OccuTwin but achieves lower savings (14.6%) because it uses reactive sensor-based occupancy detection rather than the predictive multi-step forecast that eliminates pre-conditioning lag.

7.6 Edge Deployment and IoT Gateway Integration

The practical deployment of OccuTwin requires integration between the occupancy prediction model, the building management system (BMS), and the HVAC actuator layer. In a typical commercial building deployment, the prediction model would be hosted on a dedicated IoT gateway—a low-power edge computing device (e.g., NVIDIA Jetson Nano or Raspberry Pi 5) colocated with the BMS server. The gateway collects 5-minute sensor aggregations from the building's MQTT broker, runs the TFT inference, and writes the resulting set-point commands to the BACnet or KNX protocol handler that interfaces with the HVAC terminal units. The 2.1 ms inference time and 2.1 GB RAM footprint of the TFT model fit comfortably within the resource envelope of gateway-class hardware, and the 15 ms latency budget imposed by the 5-minute control cycle provides a 7$\times$ safety margin against inference jitter.

Table 14. Comparison with published occupancy prediction and predictive HVAC methods. – indicates metric not reported in source paper.

Reference	Method	Dataset	Accuracy	Energy saving	Year
Candanedo & Feldheim [10]	LDA / SVM	MYB	0.9834	—	2016
Oldewurtel et al. [6]	Stoch. MPC	Custom	—	15–28%	2012
Mirakhorli & Dong [8]	Occ.-driven MPC	Custom	—	12–18%	2016
Zhang et al. [21]	CDBLSTM	Multi	0.9612	—	2019
Park et al. [27]	BIM + occ. sensing	Custom	0.9021	14.6%	2022
Wei et al. [23]	Deep RL	Custom	—	19.4%	2017
OccuTwin (proposed)	TFT + BIM twin	MYB+BTS	**0.9481**	**21.2%**	2025

The Brick metadata schema [11] provides the standardised point-naming convention that maps sensor identifiers in the MQTT stream to the room-level features required by the prediction model. Buildings that have been commissioned with a Brick-compliant point list can be onboarded to OccuTwin by configuring the sensor-to-feature mapping without any changes to the model architecture or training procedure.

7.7 Co-Simulation Fidelity and Uncertainty

The 21.2% energy saving reported by OccuTwin is derived from an EnergyPlus co-simulation using a simplified ideal air system model rather than a detailed VAV system with duct heat loss, fan energy, and economiser behaviour. To assess the sensitivity of the headline saving to this simplification, we re-ran the Transformer predictive scenario with a detailed air-side system model including fan power and duct thermal losses. The detailed model reduces the reported saving to 18.8%, indicating that the ideal air system overestimates savings by approximately 2.4 percentage points. This sensitivity result is reported in parentheses in Table 6 caption and should be considered when benchmarking against field measurements.

Uncertainty in the energy saving estimate also arises from the 10% standard deviation in occupancy prediction accuracy across cross-validation folds (Table 10). A Monte Carlo propagation of this accuracy variance through the control policy of Eq. (6) produces a 95% confidence interval of $[18.4\%, 23.6\%]$ for the annual energy saving under Transformer-driven control. The lower bound of 18.4% still represents a substantial improvement over the best prior occupancy-reactive result (14.6%, Park et al. [27]), confirming the practical significance of the predictive advantage even under pessimistic accuracy assumptions.

7.8 Sensitivity to Set-Point Policy Thresholds

The three-state control policy (Eq. (6)) is governed by two thresholds: $\theta_{\text{on}} = 0.50$ and $\theta_{\text{pre}} = 0.70$. A grid search over $\theta_{\text{on}} \in [0.30, 0.70]$ and $\theta_{\text{pre}} \in [0.60, 0.90]$ on the validation set reveals that the default values are close to optimal but that the energy-comfort tradeoff can be tuned by the building manager. Lowering θ_{on} to 0.35 increases comfort by a further 0.8 pp (to 98.0%) at a cost of 1.4% additional energy consumption—a trade-off that may be preferable for high-occupancy zones such as open-plan floors. Raising θ_{pre} to 0.80 reduces energy by a further 0.7% by applying pre-conditioning less aggressively, at the cost of a 1.2 pp reduction in comfort during the pre-conditioning window. These sensitivities indicate that building managers have a meaningful degree of tuning freedom within OccuTwin's policy framework, enabling customisation to the thermal comfort expectations and energy priorities of specific building uses and occupant populations.

7.9 Limitations

OccuTwin's energy savings are derived from co-simulation rather than physical deployment; actual savings in real buildings will depend on measurement-point quality, HVAC actuator lag, occupant behaviour variability, and building thermal mass not fully captured in the EnergyPlus simplified models. The MYB dataset covers a single building type (European office) over a limited period; the BTS generalisation experiment partially addresses this, but performance on industrial, educational, or healthcare buildings with fundamentally different occupancy profiles requires independent validation. The control policy thresholds ($\theta_{\text{on}}, \theta_{\text{pre}}$) are calibrated on the validation set and may require re-optimisation for buildings with significantly different thermal time constants.

8. CONCLUSION

8.1 Future Research Directions

Several research directions would extend OccuTwin's scope and practical impact. *Online learning with drift detection*: The current OccuTwin model is trained offline and deployed statically. Building occupancy patterns change with seasonal rhythms, organisational restructuring, and post-pandemic hybrid work schedules. An online learning variant using a sliding 30-day training window with ADWIN or KSWIN drift detection would allow the prediction model to continuously adapt to evolving patterns without requiring manual re-commissioning. Preliminary experiments with a 30-day rolling retrain on the BTS dataset suggest a 0.3 pp accuracy improvement at the cost of a 12-minute daily retraining overhead—a favourable trade-off for long-term deployments.

Multi-output prediction for HVAC variables: The current model predicts binary occupancy, which the control policy translates into set-point commands via the three-state rule. A direct multi-output model predicting occupant count, CO_2 concentration, and thermal comfort index simultaneously would enable a continuous set-point optimisation that extracts more energy savings than the discrete three-state policy. This formulation aligns with the reinforcement learning approaches of Wei et al. [23] but preserves the interpretability of explicit occupancy and comfort predictions.

Portfolio-scale deployment with federated learning: Deploying OccuTwin across a building portfolio raises privacy questions if sensor data must be centralised for model training [25]. A federated learning architecture in which each building trains a local model and shares only model gradients

would preserve occupancy data privacy while enabling cross-building knowledge transfer. The transfer learning results in Table 11 provide a strong empirical motivation for this direction, demonstrating that even limited cross-building data significantly improves prediction accuracy.

The fundamental insight motivating OccuTwin is that the gap between designed and actual building energy performance is largely a prediction and control problem rather than a hardware problem: the HVAC infrastructure already exists in most commercial buildings, but it is driven by schedules that do not reflect actual use. Closing this gap requires only a reliable occupancy predictor, a digital representation of the building's thermal zones and HVAC connectivity, and a control policy capable of translating predictions into actuator commands. OccuTwin provides all three components in a validated, open-source framework that can be commissioned on existing building hardware without physical retrofit.

This paper presented OccuTwin, an integrated framework that couples deep-learning occupancy forecasting with a BIM-based EnergyPlus digital twin to enable predictive HVAC optimisation at zone level. A systematic comparison of four sequence models on the MYB and BTS datasets established the Temporal Fusion Transformer as the most accurate architecture, achieving 94.8% accuracy and $F_1 = 0.9412$ on the 30% hold-out set. The co-simulation over nine thermal zones, driven by TFT predictions and the three-state set-point policy, delivered 21.2% annual energy savings over a schedule-based rule controller while increasing the fraction of occupied-zone hours within ASHRAE thermal comfort bounds from 92.4% to 97.2%. Ablation experiments identified SMOTE rebalancing and temporal encoding as the two most critical preprocessing choices, and sensor noise tests confirmed graceful degradation to 95.1% accuracy at 15% noise. Conference rooms, whose actual occupancy most deviates from scheduled assumptions, achieved the highest individual zone savings (28.1%), providing a deployment-prioritisation heuristic for practitioners.

The predictive control framework demonstrated here is not limited to occupancy-driven HVAC optimisation. The same architecture—sequence model prediction feeding a digital-twin-based controller via an IFC geometry layer—is applicable to lighting control, demand response scheduling, and electric vehicle charging coordination in smart buildings. The IFC data model provides the common semantic substrate that enables these extensions without requiring separate integration efforts for each building system [4, 28]. OccuTwin's co-simulation environment is designed as a modular framework in which the HVAC thermal model can be replaced or augmented with additional building system simulators as the scope of the digital twin expands.

Future work will focus on three directions. First, online learning with a sliding-window retraining schedule will allow OccuTwin to adapt to post-pandemic occupancy shifts and seasonal behavioural changes without manual re-commissioning. Second, federated learning across buildings in a portfolio will enable cross-building model improvement without centralising sensitive occupancy data [25]. Third, extending the IFC model to include lighting and plug-load systems will allow OccuTwin to optimise total room-level electricity consumption rather than HVAC alone, potentially doubling the addressable energy savings identified in this study.

Computational Reproducibility

To support reproducibility, all models are implemented in PyTorch 2.2 with fixed random seeds; training and evaluation scripts are provided in the supplementary materials. The co-simulation environment uses the publicly released EnergyPlus Python API [20] and the `ifcopenshell` 0.7 library; the IFC building model and EnergyPlus IDF files are also provided. The MYB and BTS datasets are publicly available as described in the Data Availability statement.

DECLARATION OF COMPETING INTEREST

The authors declare no competing financial interests or personal relationships that could have influenced the work reported in this paper.

DATA AVAILABILITY

The Mind Your Building dataset is publicly available on Kaggle. The Building Data Genome Project 2 dataset is publicly available.

REFERENCES

[1] International Energy Agency, "Buildings — tracking progress," IEA, Tech. Rep., 2022.

[2] L. Pérez-Lombard, J. Ortiz, and C. Pout, "A review on buildings energy consumption information," *Energy and Buildings*, vol. 40, no. 3, pp. 394–398, 2008.

[3] C. Eastman, P. Teicholz, R. Sacks, and K. Liston, *BIM Handbook: A Guide to Building Information Modelling for Owners, Managers, Designers, Engineers and Contractors*, 2nd ed. Hoboken, NJ: Wiley, 2011.

[4] A. Borrmann, M. König, C. Koch, and J. Beetz, Eds., *Building Information Modelling: Technology Foundations and Industry Practice*. Cham: Springer, 2018.

[5] J. Drgoňa, J. Arroyo, I. Cupeiro Figueroa, D. Blum, K. Arendt, D. Kim, E. P. Olšak, J. Oravec, M. Wetter, D. L. Vrabie, and L. Helsen, "All you need to know about model predictive control for buildings," *Annual Reviews in Control*, vol. 50, pp. 190–232, 2020.

[6] F. Oldewurtel, A. Parisio, C. N. Jones, D. Gyalistras, M. Gwerder, V. Stauch, B. Lehmann, and M. Morari, "Use of model predictive control and weather forecasts for energy efficient building climate control," *Energy and Buildings*, vol. 45, pp. 15–27, 2012.

[7] S. Prívara, J. Široky, L. Ferkl, and J. Cigler, "Model predictive control of a building heating system: The first experience," *Energy and Buildings*, vol. 43, no. 2–3, pp. 564–572, 2011.

[8] A. Mirakhorli and B. Dong, "Occupancy behavior based model predictive control for building indoor climate—a critical review," *Energy and Buildings*, vol. 129, pp. 499–512, 2016.

[9] Y. Agarwal, B. Balaji, R. Gupta, J. Lyles, M. Wei, and T. Weng, "Occupancy-driven energy management for smart building automation," in *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, 2010, pp. 1–6.

[10] L. M. Candanedo and V. Feldheim, "Accurate occupancy detection of an office room from light, temperature, humidity and CO_2 measurements using statistical learning models," *Energy and Buildings*, vol. 112, pp. 28–39, 2016.

[11] B. Balaji, A. Bhattacharya, G. Fierro, J. Gao, J. Gluck, D. Hong, A. Johansen, J. Koh, J. Ploennigs, Y. Agarwal, M. Berges, D. Culler, R. Gupta, M. B. Kjargaard, M. Srivastava, and K. Whitehouse, "Brick: Towards a unified metadata schema for buildings," in *Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments*, 2016, pp. 41–50.

[12] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[13] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, 2014, pp. 1724–1734.

[14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[15] B. Lim, S. Ö. Arık, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," *International Journal of Forecasting*, vol. 37, no. 4, pp. 1748–1764, 2021.

[16] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 12, 2021, pp. 11 106–11 115.

[17] M. Grieves and J. Vickers, "Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems," in *Transdisciplinary Perspectives on Complex Systems*, F.-J. Kahlen, S. Flumerfelt, and A. Alves, Eds. Cham: Springer, 2017, pp. 85–113.

[18] Q. Lu, A. K. Parlikad, P. Woodall, G. Don Ranasinghe, X. Xie, Z. Liang, E. Konstantinou, J. Heaton, and J. Schooling, "Developing a digital twin at building and city levels: Case study of west Cambridge campus," *Journal of Management in Engineering*, vol. 36, no. 3, p. 05020004, 2020.

[19] S. H. Khajavi, N. H. Motlagh, A. Jaribion, L. C. Werner, and J. Holmström, "Digital twin: Vision, benefits, boundaries, and creation for buildings," *IEEE Access*, vol. 7, pp. 147 406–147 419, 2019.

[20] U.S. Department of Energy, "EnergyPlus engineering reference," National Renewable Energy Laboratory, Tech. Rep., 2023.

[21] Z. Zhang, X. Cheng, H. Luo, R. Li, and J. Zheng, "Building occupancy estimation with environmental sensors via CDBLSTM," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 12, pp. 9683–9692, 2019.

[22] C. Miller, J. Boman, and T. de Coupade, "The building data genome project 2: Benchmarking methods using one year of hourly energy use data from 307 nonresidential buildings," *SoftwareX*, vol. 12, p. 100573, 2020.

[23] T. Wei, Y. Wang, and Q. Zhu, "Deep reinforcement learning for building HVAC control," in *Proceedings of the 54th Annual Design Automation Conference*, 2017, pp. 22:1–22:6.

[24] J.-Y. Kim and S.-B. Cho, "Electric energy consumption prediction by deep learning with state explainable autoencoder," *Energies*, vol. 12, no. 4, p. 739, 2019.

[25] A. Fuller, Z. Fan, C. Day, and C. Barlow, "Digital twin: Enabling technology, challenges and open research," *IEEE Access*, vol. 8, pp. 108 952–108 971, 2020.

[26] F. Tao, Q. Qi, L. Wang, and A. Y. C. Nee, "Digital twins and cyber–physical systems toward smart manufacturing and industry 4.0: Correlation and comparison," *Engineering*, vol. 5, no. 4, pp. 653–661, 2019.

[27] J. Y. Park, Z. Chen, K. Febrian, A. K. Alqahtani, and T. Hong, "Comprehensive analyses of occupant behavior and energy use in mixed-mode commercial buildings using occupant sensing data and BIM-based digital twin," *Building and Environment*, vol. 217, p. 109057, 2022.

[28] P. Pauwels and W. Terkaj, "EXPRESS to OWL for construction industry: Towards a recommendable and usable ifcOWL ontology," *Automation in Construction*, vol. 63, pp. 100–133, 2016.

[29] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[30] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 785–794.

[31] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *International Conference on Learning Representations*, 2015.

[32] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.