UNIVERSITY OF WEST ATTICA
eISSN:3136-7658

Applied Manufacturing Intelligence Journal

Vol. 5, No. 3, July, 2026



A Few-Shot TimesNet-MoE Model for Occupancy Prediction in Smart Lighting

Octav Mănescu[1,*], Dumitru Zavoi[1]

[1] *Faculty of Automatics, Computer Science and Electronics, University of Craiova, Craiova, 200440, Romania*

Received 13 May 2026; Revised 26 August 2026; Accepted 31 August 2026; Published 03 September 2026

Abstract

Newly commissioned smart-lighting spaces often provide limited labelled occupancy histories, while room routines and sensing responses vary across a building. This study presents a few-shot TimesNet-MoE framework that combines period-aware temporal encoding, context-dependent expert weighting, and calibrated occupancy probabilities. Passive infrared, carbon dioxide, illuminance, switch-state, and calendar signals are organized into support-query episodes. Four experts emphasize motion, scheduling, environmental persistence, and occupancy transitions, with a shared prediction head linking their representations to a lighting-control interface. Evaluation covers five room categories and support budgets from four to 64 labelled context windows. At 16 support samples and a 10 min horizon, the framework achieved macro-F1 of 0.85, compared with 0.79 for plain TimesNet, while occupancy-probability mean absolute error decreased from 0.151 to 0.121. In the support-budget experiment, macro-F1 reached 0.900 and probability error reached 0.104 at 64 samples. Separate horizon and cross-room experiments yielded macro-F1 of 0.824 at 30 min and transfer F1 of 0.84. At 20% missing observations, F1 remained 0.82, compared with 0.72 for TimesNet. Inference required 46 ms and 0.53 J under the reported profiling configuration. Room-level normalized lighting energy ranged from 0.67 to 0.90. A separate controller-aggressiveness sweep paired 17.3% energy savings with 94.2% comfort compliance. The reported comparisons indicate that period-aware expert weighting can improve sparse-supervision occupancy prediction, with deployment decisions requiring explicit consideration of inference cost, calibration, and transition-level comfort.

Keywords: Few-Shot Occupancy Prediction, TimesNet-MoE, Smart Lighting Control, Multi-Modal Sensing, Expert Routing

1. Introduction

Room-level occupancy sensing provides the necessary information for coordinating lighting energy use with the presence and comfort of building occupants [1]. Wireless lighting networks can integrate motion detectors, environmental sensors, wall switches and local controllers to monitor the spatial distribution of room use [2]. Short-horizon occupancy forecasting extends the sensing function by anticipating occupant arrivals and vacancy transitions before changes in luminaire output [3]. Room-specific adaptation addresses the problem of limited labelled history during commissioning and different occupancy patterns in different areas of a building [4]. These requirements are particularly demanding in private offices, meeting rooms, laboratories, corridors, and quiet study areas, where the same sensor arrangement may observe different occupancy patterns. Motion events are intermittent during seated work, carbon dioxide responds with a delay, and changes in daylight can alter the measured illuminance independently of people. A newly installed controller therefore receives several imperfect sensing streams whose interpretation depends on recent history and room function. The engineering task is to learn this relationship from a small commissioning sample and retain the ability to respond to changes in use. A reliable predictor must distinguish sustained occupancy

* Corresponding author: oct.man@ace.ucv.ro

https://doi.org/10.68406/amij.2026.vol5iss3np6:79-93

➢ This is an open access article under the CC BY-NC-SA 4.0 license (https://creativecommons.org/licenses/by/4.0/).

© Authors retain all copyrights.

from brief passage, represent uncertainty around ambiguous transitions, and provide usable probabilities when individual sensing channels become unavailable. Therefore, few-shot learning and temporal specialization are directly related to the reliability of intelligent lighting deployment.

Occupancy prediction methods commonly include statistical models, recurrent neural networks, attention-based time-series architectures, and control-oriented heuristics. Statistical and heuristic methods offer interpretability, while nonlinear temporal dependencies associated with sensor delay and intermittent occupant motion can reduce their predictive performance [5]. Recurrent models represent sequential dependence, although their hidden states may overfit sparse room data and degrade under missing observations [6]. Transformer-based methods improve long-range dependency modelling, although their data requirements and limited periodic inductive bias can constrain small-scale applications [7]. TimesNet transforms one-dimensional time series into multi-period two-dimensional representations, enabling convolutional blocks to capture within-period and between-period variation. A single TimesNet backbone applies a shared temporal representation to office arrival peaks, meeting-room transitions, corridor pulses, and quiet library occupancy [8]. A shared representation may inadequately capture differences in anticipatory control requirements among rooms with similar average occupancy but distinct arrival, break, and vacancy patterns.

Few-shot learning can support rapid deployment when its formulation accounts for the physical and temporal structure of buildings. The support set in this case is a narrow window of sensor history with missingness, schedule bias and environmental lag. The model must identify transferable temporal features while preserving room-specific differences relevant to lighting control. Mixture-of-experts models route each segment to specialized temporal modules while retaining regime-specific representations. TimesNet-MoE combines period-aware feature extraction with an expert router that estimates the relevance of micro-motion, scheduled arrival, environmental accumulation, and departure dynamics to each segment. The resulting calibrated short-horizon probability provides an input to a dimming or switching policy for intelligent lighting.

This study addresses sparse commissioning data by integrating an established period-aware encoder with room-conditioned expert weighting and a calibrated lighting-control interface. The contribution lies in the task-specific integration of established temporal components and its evaluation across support budgets and room regimes. The framework addresses the preservation of recurring temporal structure through a shared encoder, adaptation of expert routing to local sensing conditions under limited supervision, and use of forecast probabilities to support gradual dimming around uncertain transitions. A compact expert bank is employed to model motion, scheduling, environment persistence and arrival/departure dynamics. The evaluation examines support-set size, transfer direction, forecast horizon, missing measurements, calibration, and the energy-comfort trade-off. The experiments examine these relationships through predictive-performance comparisons, robustness tests, and control-impact analysis. The remainder of the paper progresses from sensing requirements to model design and operational evaluation. Section 2 presents Occupancy Sensing, Few-shot Temporal Forecasting and Expert-based Learning, and introduces the research gap this paper aims to address. Section 3 introduces the occupancy-prediction task, episodic samples, period-aware encoder, and expert bank. Section 4 shows the adaptive routing, probability calibration and lighting-control interface. Section 5 presents the evaluation protocol and analyses of the predictive performance, transfer behavior, robustness, ablation and control results using six data figures and two tables. Section 6 presents the contribution, limitations, and directions for subsequent field validation.

2. Related Work

2.1. Occupancy Sensing for Lighting Control

Conventional occupancy detection for lighting control uses binary events from passive infrared or ultrasonic sensors to activate local switching rules. These systems offer low installation complexity and cost, with stationary-occupant detection remaining sensitive to the availability of motion evidence. Therefore, research on sensor fusion has combined motion events with carbon dioxide, illuminance, plug-load and switch data to infer a richer state of room use [9]. Sensor fusion is particularly relevant in offices and classrooms, where delayed environmental responses provide persistence cues during periods of limited motion. Sensor fusion introduces synchronization, calibration, and missing-data requirements that need explicit treatment in the sensing and control pipeline.

Lighting control imposes shorter response-time requirements than long-horizon building energy forecasting. A predictor must anticipate near-future occupancy while limiting false-vacancy decisions that can reduce user trust. Recent lighting-load forecasting research has used adaptive fusion to represent short-term changes in smart-building lighting demand [10]. Newly commissioned rooms often provide few labelled samples and have distinct commissioning histories. Transfer from previously instrumented zones can support adaptation, although differences in room-use patterns may introduce bias even when sensor sets are identical. Recent research has therefore focused on room-aware learning, metadata conditioning and calibration after deployment [11].

Occupancy-sensing research continues to examine the trade-offs among privacy, installation cost, and predictive accuracy. Privacy restrictions can limit camera-based counting in educational and corporate buildings despite its potential accuracy. Non-image sensors offer greater privacy compatibility, while their indirect observations remain sensitive to placement, ventilation, daylight, and device latency. Therefore, for smart lighting, temporal structure extraction can be performed on low-cost non-image streams. Few-shot formulation is also consistent with the commissioning practice of adapting models based on a small number of interval labels provided by facility staff.

2.2. Few-Shot Temporal Forecasting and Expert Models

Few-shot temporal forecasting extends limited-supervision learning to sequences characterized by temporal ordering, seasonality, and concept drift. Metric-learning methods compare new samples with learned prototypes, with occupancy prediction additionally requiring representations that preserve temporal relationships [12]. Vettoruzzo et al. [13] combined meta-learning with learnable time-series augmentation to support rapid adaptation to new subjects and activities from limited labelled samples. Task similarity varies across building applications with space function, occupant behavior, and sensor quality. These variations motivate an architecture that combines transferable temporal features with sensitivity to local context.

Recent self-supervised forecasting research has examined trend-error decomposition as a means of improving temporal representation under limited supervision [14]. TimesNet identifies dominant periods and reshapes sequences so that the 2D kernel can represent changes within and between periods [15]. Daily and subdaily rhythms in occupancy data motivate the use of period-aware representations. Application to small labelled support sets and heterogeneous room regimes needs additional adaptation and routing mechanisms. With scarce training data, a single temporal block may learn an aggregate representation with reduced sensitivity near occupancy transitions, where lighting decisions depend on timely state changes.

Mixture-of-experts models address heterogeneity by assigning different inputs to specialized subnetworks. Liu et al. [16] developed a sparse mixture-of-experts model with closed-loop routing for short-term building load forecasting and cross-building transfer. The present architecture uses continuous weighting of a compact expert bank for room-level occupancy prediction. For occupancy prediction, expert specialization has a physical interpretation: carbon dioxide accumulation, motion pulses, scheduled arrivals, and unscheduled passage require different temporal representations. The unresolved challenge is to integrate expert routing with period-aware temporal modeling while maintaining stable adaptation from a small support set. This paper positions TimesNet-MoE as a compact solution to that challenge for smart lighting.

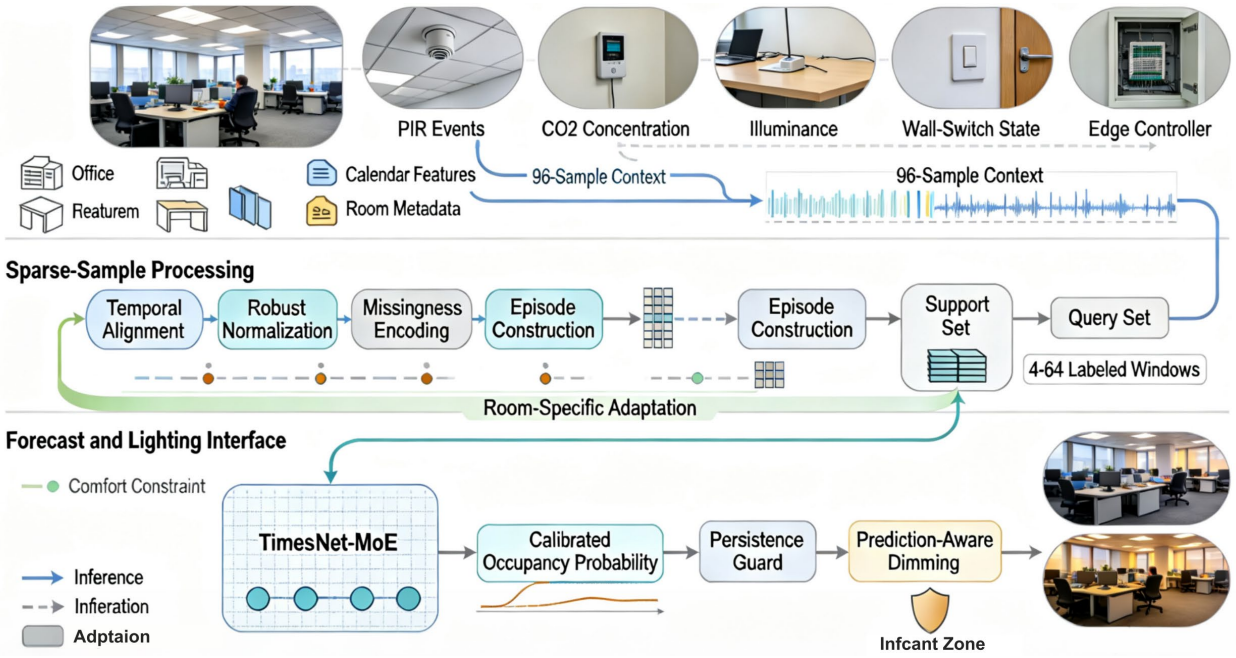


Figure 1. Few-shot smart lighting occupancy prediction pipeline

3. Few Shot TimesNet MoE Formulation

3.1. Problem Definition and Episodic Samples

A smart lighting zone is represented by a multivariate sequence sampled at a fixed interval. The input channels include passive infrared activation, carbon dioxide concentration, desk-level illuminance, wall-switch state, weekday

index, and minutes from the scheduled start of the main occupancy period. A newly commissioned room provides a limited labelled support set for adaptation. The learning objective is to estimate the probability of future occupancy by combining local observations with temporal information learned from other rooms. Figure 1 illustrates the prediction pipeline, linking sparse sensing streams, episodic adaptation, temporal encoding, and the lighting-control interface. A support sample is one labelled context window and its forecast target; a raw sensor reading and a labelled window represent different units of supervision.

The input window and binary prediction target are defined by the following equation:

$$X_{i,t} = \{x_{i,t-c+1}, \dots, x_{i,t}\}, y_{i,t}^{(h)} = \mathbf{1}(o_{i,t+h} > 0) \quad (1)$$

Here, $X_{i,t}$ denotes the context window for room i at time t, c is the historical length, h is the forecast horizon, and $o_{i,t+h}$ represents future occupancy. The model converts binary supervision into occupancy probabilities that preserve confidence information near the controller decision boundary. Each training episode contains support and query samples drawn from the designated temporal partitions. The support samples provide local adaptation evidence, while the query samples evaluate the resulting predictor. Future occupancy is a binary presence indicator. The event in the control-memory equation consequently represents room occupancy for any positive number of occupants.

This separation is formalized through an episodic equation:

$$\mathcal{D} = \{(\mathcal{S}_e, \mathcal{Q}_e)\}_e, \mathcal{S}_e = \{(X_j, y_j)\}_{j=1}^{K}, \mathcal{Q}_e = \{(X_q, y_q)\}_{q=1}^{M} \quad (2)$$

In the episodic equation, K specifies the number of labeled support examples and M specifies the query-set size. The shot count denotes the total budget of labelled context windows across the occupied and vacant classes. Support examples are balanced across occupied and vacant intervals where labels permit, while query examples retain the deployment distribution. Missing observations are accompanied by a binary mask and a time-since-observed channel. These indicators distinguish valid zero-valued measurements from unavailable sensor readings and preserve the duration of information gaps.

3.2. Period Aware Temporal Encoding

The encoder follows the TimesNet principle that temporal variation can be represented through several candidate periods. In smart lighting, relevant structures include short motion bursts, room-change intervals, meeting blocks, and daily routines. Input sequences are normalized using robust room statistics before frequency-based period discovery. The selected periods fold each sequence into two-dimensional tensors, with one axis representing positions within a period and the other representing repeated period instances. This arrangement exposes local variation and recurring structure to a common convolutional operation. Period-specific feature extraction follows the equation below:

$$Z_p = Reshape_p(Conv1D(X, B)), H_p = GELU\left(BN\left(Conv2D\left(Z_p\right)\right)\right) \quad (3)$$

The period-encoding equation maps the sequence X and missingness mask B to a feature representation H_p for candidate period p. The symbol B separates the sensor mask from the query-set size M. The convolutional block remains compact to limit adaptation complexity and controller memory requirements. Calendar features and room metadata enter after temporal encoding, allowing contextual information to influence routing while preserving sensor-derived temporal features. When a sequence length is incompatible with a selected period, padding positions retain an explicit validity mask. Features associated with padded positions are excluded from aggregation, preventing artificial boundary values from altering the representation. The resulting period-specific features form the shared input to the expert bank. Expert outputs require compatible feature dimensions for weighted fusion, while the input validity mask identifies unavailable or padded observations. Exact kernel sizes, feature widths, and period counts are needed to reproduce this schematic encoding block.

3.3. Expert Bank and Output Layer

The model uses four experts that share the period-aware input but differ in receptive-field emphasis and feature normalization. The micro-motion expert focuses on short bursts and small changes in passive infrared activity. The schedule expert emphasizes repeated daily structure and wall-switch patterns. The environment expert represents delayed carbon dioxide responses and changes in illuminance. The transition expert emphasizes arrival and departure, when occupancy changes rapidly. Their outputs are combined through adaptive routing and passed to a shared prediction head. Temperature scaling on validation episodes provides the final probability-calibration step.

The expert bank is organized around the temporal mechanisms observed by each sensing channel. Shorter kernels help the micro-motion expert preserve abrupt events whose interpretation depends on sensor placement and recent activity. The schedule expert uses a wider temporal field and incorporates weekday and time-offset embeddings.

Calendar-channel dropout during training limits excessive dependence on expected routines. Channel-wise normalization helps the environment expert accommodate differences in ventilation, daylight exposure, and sensor scale. The transition expert emphasizes first-order differences and short-term variability, providing a representation suited to changes between sustained occupancy and vacancy. These expert names denote architectural emphases. Demonstrating learned semantic specialization requires regime-conditioned routing records or targeted interventions in addition to aggregate usage shares.

The output head contains two projection layers. The first maps the routed expert representation to a compact state embedding, and the second produces a horizon-specific occupancy logit. Sharing the head across rooms reflects the common meaning of the prediction target and reduces the number of parameters adapted from sparse supervision. Room specificity is introduced through support-conditioned normalization and the gate. A missingness summary also enters the projection head so that the predictor can account for reduced observational support. Calibration near the control threshold remains particularly relevant because a confident vacancy estimate can trigger dimming during quiet occupied intervals.

4. Adaptive Expert Routing and Lighting Interface

4.1. Soft Gating Under Sparse Labels

The gating network receives temporal features, support-derived room descriptors, and missingness indicators. It assigns context-dependent weights to the mechanisms represented by the expert bank. Figure 2 illustrates the connection between period discovery, parallel experts, soft routing, and calibrated prediction. The room descriptor is computed from the permitted support interval and retained during query evaluation, while temporal features and missingness indicators change with the current input window.

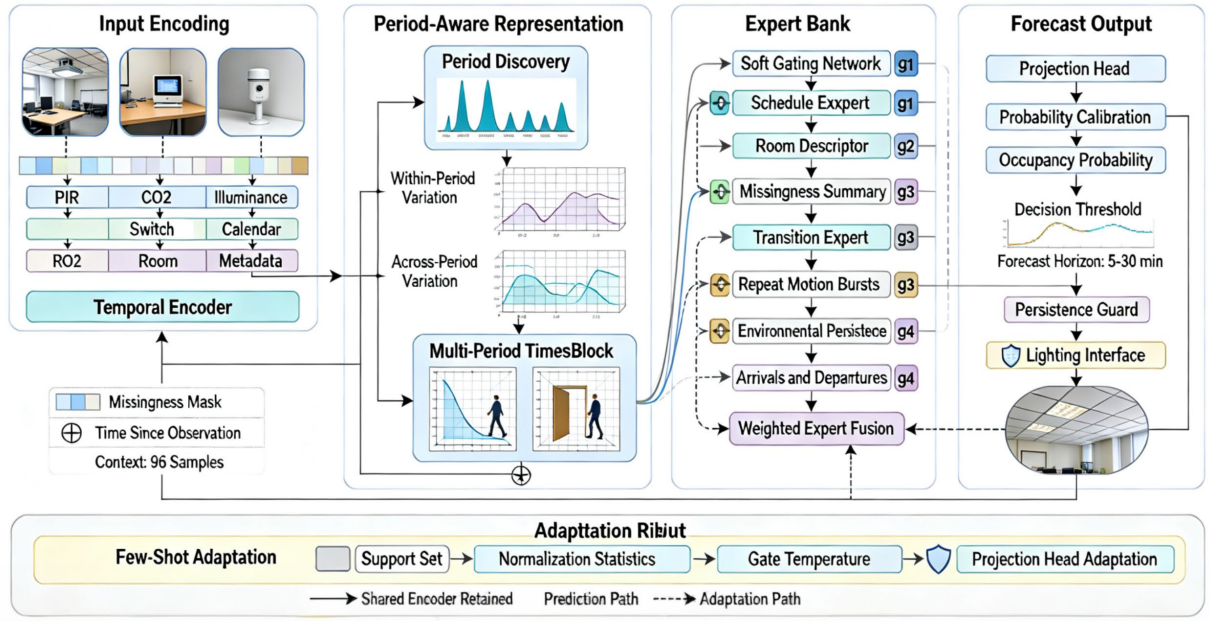


Figure 2. TimesNet-MoE architecture for few-shot occupancy forecasting

The routing equation assigns normalized expert weights and constructs the fused representation:

$$g_k = softmax\left(\frac{W_g[H;r;m]}{\tau}\right)_k, z = \sum_{k=1}^{E} g_k f_k(H) \quad (4)$$

Within this equation, H denotes the aggregated period-aware feature, r is the room descriptor, m summarizes missingness, and τ controls routing temperature. The function f_k represents expert k, and E is the expert-bank size. Soft routing accommodates segments in which several mechanisms remain plausible, such as scheduled arrivals accompanied by weak motion evidence. Entropy regularization discourages premature concentration on one expert during training. At inference, the weighted mixture evaluates the full expert bank and combines its outputs through continuous weights. This dense soft-routing formulation provides conditional weighting, with computational cost reflecting full-bank inference. The temperature parameter determines how sharply the gate distinguishes experts: a higher temperature spreads the weights, while a lower temperature concentrates them. Adaptation adjusts routing temperature using support

evidence. Output probability calibration uses a separate temperature acting on the prediction logit; reproducible implementation requires distinct records for the routing and calibration temperatures.

4.2. Calibration and Control Signal Generation

The predicted occupancy probability enters a lighting interface containing a decision threshold, persistence guard, and bounded output mapping. A zone becomes eligible for dimming when the forecast indicates vacancy throughout the required persistence interval. The guard preserves illumination during short interruptions in motion evidence and accommodates the delayed response of environmental sensors. Calibration is applied before threshold comparison so that confidence levels have a consistent interpretation across adapted rooms.

The control-memory equation combines the current occupancy decision with the preceding activation score:

$$u_t = \mathbf{1}(P(o_{t+h} = 1 \mid X_t) \geq \theta_c) + \alpha u_{t-1} \tag{5}$$

Here, u_t is an intermediate activation score, θ_c is the occupancy-probability threshold, and α determines the persistence of previous activation. The accumulated score represents an internal memory state that retains activation evidence for command generation. Its value is bounded before conversion to a physical dimming command; stability also depends on a nonnegative persistence coefficient below unity and an explicitly specified output mapping. This distinction preserves the operating range of the lighting interface when the current decision and retained activation both contribute to the score. The persistence guard is implemented separately through consecutive threshold checks, while the memory term smooths changes in the command.

Controller settings are selected using validation episodes and then retained during query evaluation. Threshold selection considers false-vacancy events, arrival delays, and manual overrides alongside aggregate prediction metrics. A user override takes precedence over the forecast-driven command and is recorded for subsequent inspection. Applying the same persistence duration and output mapping to every compared predictor allows the evaluation to attribute control differences to the forecast sequence.

4.3. Training Objective and Implementation

The training objective combines binary cross-entropy, a calibration penalty, and routing regularization. Cross-entropy supervises the occupancy estimate, the calibration term aligns predicted confidence with empirical frequency, and routing regularization discourages expert collapse. The architecture uses compact convolutional experts and shared period selection to constrain memory requirements. Section 5 reports latency and inference-energy measurements for the evaluated deployment configuration. Independent replication requires further specification of the exact loss weights, optimizer settings, and expert dimensions.

Training proceeds in two stages. Source-room episodes first train the period encoder and expert bank to represent transferable temporal mechanisms. Target-room adaptation then updates normalization statistics, routing temperature, and projection-head parameters using the designated support set. The pretrained encoder remains fixed during this stage, restricting the adaptation problem to parameters that can be estimated from limited supervision. This arrangement allows local sensor scale and room routine to influence predictions while retaining temporal features learned from source rooms.

Implementation choices address the processing and memory constraints of local lighting controllers. Continuous channels are clipped using robust percentiles and standardized with statistics estimated from the permitted support interval. Binary channels retain their native scale. Period selection is limited to a compact candidate set, and expert feature dimensions remain fixed across rooms. Episode-based mini-batches preserve support-query relationships during optimization. Early stopping requires a validation partition distinct from the final target-room query set. Verification of model-selection separation requires training logs identifying the stopping partition and documenting the reservation of final query labels for evaluation. Deployment profiling records input preparation, model inference, and probability calibration separately, allowing computation associated with sensing and control communication to be distinguished from the predictor itself.

5. Experimental Data and Analysis

5.1. Dataset Protocol and Baselines

The experimental evaluation covered five room categories: private office, laboratory, meeting room, corridor, and library. Predictive accuracy, probability calibration, inference latency, inference energy, and lighting-control performance were assessed under a common evaluation protocol. The input specification comprises passive infrared events, carbon dioxide concentration, illuminance, wall-switch state, weekday, and schedule offset. Occupancy is defined at future horizons of 5, 10, 15, 20, 25, and 30 min. A 96-sample history supplies temporal context, and labeled support sizes of 4, 8, 16, 32, and 64 intervals describe progressively less restrictive commissioning budgets. Temporally

separated support and query partitions are a recognized approach to evaluating adaptation without adjacent-window leakage [17]. Under the evaluation protocol, the initial two-day target-room interval supplies the support pool, and subsequent intervals supply the query pool. Source-room records remain separate from target-room query intervals. Overlapping windows require a boundary gap covering the context and forecast horizon to maintain separation between sensor events in the support and query partitions. Such separation is particularly relevant for quiet occupancy, where consecutive windows can remain nearly identical for extended periods. The present description establishes the room categories, forecast horizons, and support budgets. Independent reproduction additionally requires acquisition dates, sampling interval, individual room counts, ground-truth procedures, and usable window counts.

Table 1 records the configuration shared by the proposed model and its four comparison methods: LSTM, Informer, PatchTST, and plain TimesNet. The five methods receive the same sensor channels, context definition, support budgets, and target labels. Joint reporting of forecast quality and downstream operational indicators has been used in building-control evaluation to examine whether prediction improvements remain useful to a controller [18]. The selected measures are macro-F1, occupancy-probability mean absolute error, expected calibration error, inference latency, normalized lighting energy, and comfort compliance. Macro-F1 assigns equal importance to the occupied and vacant classes, whereas probability error preserves the query distribution faced by the controller. Calibration error describes agreement between predicted confidence and observed frequency and therefore addresses a different property from classification accuracy. The reference architecture contains four experts, with a default 10 min horizon for the principal model comparison. Each support example comprises a complete context window paired with its occupancy label. The adaptation protocol separates the designated support pool from the fixed query period. The reported aggregate values describe the available comparisons. The reported differences support descriptive performance comparisons; assessment of statistical significance requires repeat-level scores and confidence intervals. The configuration in Table 1 establishes the common operating conditions for the performance comparisons. Support-set size, forecast horizon, sensor channels, and metric definitions are retained consistently across methods. A reproducible metric specification additionally requires the probability-error target, calibration-bin construction, weighting convention, and operational definition of comfort compliance alongside the reported aggregates.

Table 1. Experimental configuration for few-shot smart lighting occupancy prediction

Item	Value	Item	Value	Purpose
Room types	5	Support sizes	4-64 samples	Few-shot evaluation
Context length	96 samples	Horizon	5-30 min	Short-term lighting control
Sensors	PIR, CO2, lux, switch	Experts	4	Heterogeneous occupancy regimes
Metrics	F1, MAE, ECE	Control	energy, comfort	Prediction and deployment quality

The five room categories represent distinct combinations of persistence, transition speed, and sensing ambiguity. An office may remain occupied during long intervals with few motion events, so short-term motion inactivity requires interpretation alongside slower environmental evidence. Laboratory use may extend beyond nominal schedules and involve larger illuminance changes caused by equipment or daylight. Meeting-room activity is organized around abrupt arrivals and departures, with occasional differences between booked and actual use. Corridors predominantly contain short passage events, making the controlled-zone definition essential when assigning occupancy labels. Libraries combine prolonged seated activity with quiet periods that can resemble vacancy in motion-only streams. These categories motivate testing transfer in both directions: a corridor-trained representation applied to an office can behave differently from an office-trained representation applied to a corridor. Room descriptors therefore remain available to the gate, while occupancy labels serve only as supervisory information. Source-target identities must be recorded for each transfer experiment. This allows changes associated with room function to be distinguished from those associated with sensor placement or measurement scale. The transfer-matrix labels identify room categories, while reproducible source-target assignments require individual room identities. Diagonal entries denote transfer between distinct rooms of the same category, with target query records reserved for evaluation. Per-room identifiers and category-level room counts are needed to reproduce the transfer assignments.

The preprocessing procedure was fixed before model comparison. Short gaps in continuous measurements are filled for at most three sample intervals, while longer gaps retain an explicit missingness mask and time-since-observed indicator. Room-specific robust scaling is estimated from the allowed support interval and then held fixed during query evaluation. Carbon dioxide detrending uses only past observations to preserve causal processing of the current input window. Illuminance clipping thresholds are likewise determined within the adaptation pool. Passive infrared events are encoded through binary activation and short-window counts, allowing a model to distinguish an isolated trigger from sustained activity. Wall-switch changes remain separate because they may describe a user intervention alongside occupancy evidence. The missing-sample comparison removes observations after the input specification has been fixed and retains the same query labels across methods. A controlled comparison requires identical dimming schedules,

persistence rules, comfort definitions, and override policies across predictors. Verification of consistency across controller settings requires configuration records alongside the aggregate results. Normalized daily energy refers to a common always-on reference within each room category; savings relative to a schedule baseline require a separate pairwise calculation. The energy-comfort sweep varies controller aggressiveness while retaining the forecast sequence. These definitions distinguish the effects of normalization and control policy from changes in predictive performance and support consistent comparisons with measured deployment records.

5.2. Predictive Accuracy and Transfer Behavior

Figure 3(a) specifies macro-F1 at five support sizes for each of the five methods. At four samples, the TimesNet-MoE series is 0.76, compared with 0.70 for TimesNet, 0.67 for PatchTST, 0.64 for Informer, and 0.61 for LSTM. At 16 samples, these values become 0.85, 0.79, 0.76, 0.73, and 0.70. The proposed series exceeds the 0.80 screening reference at eight samples, whereas the TimesNet series crosses that reference at 32 samples. The gap between the two remains 0.06 at both four and 64 samples, indicating a persistent performance advantage across the evaluated support budgets. Adaptation-budget sensitivity is a central consideration in few-shot temporal evaluation because support examples determine which local regimes are represented [19]. Figure 3(b) separately displays probability mean absolute error for the same support budgets. The proposed values decrease from 0.158 to 0.104, while TimesNet decreases from 0.191 to 0.132. At 16 samples, the difference is 0.030, corresponding to a 19.9% relative reduction against the TimesNet error. The paired changes indicate improved classification and lower probability error under the stated support-budget comparison. Attribution of the improvement to individual expert mechanisms requires matched component interventions. The improvement from 0.76 to 0.85 between four and 16 samples is larger than the 0.03 increase between 32 and 64 samples. Commissioning budgets should therefore be examined against marginal gains as well as absolute performance.

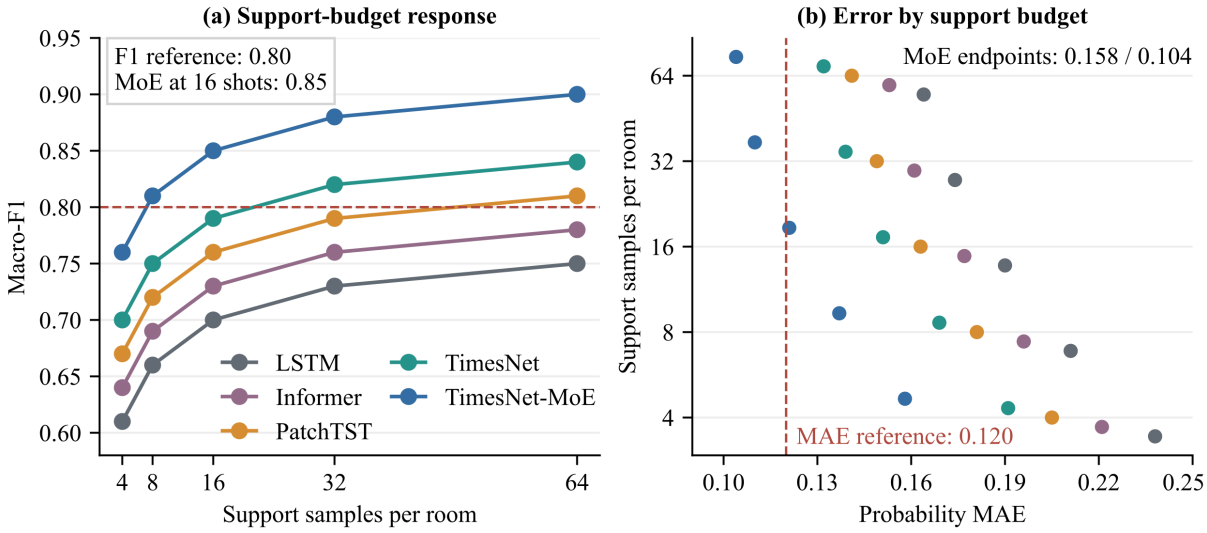


Figure 3. Few-shot accuracy and error trends: (a) macro-F1 under few-shot episodes; (b) occupancy probability MAE

Figure 4(a) gives cross-room F1 values of 0.68, 0.71, 0.74, 0.78, and 0.84 for the five methods in their stated order. The proposed value exceeds TimesNet by 0.06 and PatchTST by 0.10; interpretation alongside the within-room support-size series requires explicit consideration of the different transfer protocol.

Figure 4(b) resolves transfer into 25 source-target combinations. Office-to-office transfer is 0.86, office-to-library is 0.76, corridor-to-office is 0.75, and library-to-office is 0.73. These asymmetric entries indicate that an aggregate transfer score can obscure performance differences between specific source-target room pairs. Domain-shift analysis in building occupancy research treats transfer direction as relevant because the source and target may differ in activity persistence and sensing response [20]. The corridor diagonal is 0.86, while corridor-to-laboratory is 0.78, illustrating how a representation matched to brief passage can lose effectiveness in a space with irregular sustained activity. Figure 4(c) isolates routing under one to eight concurrently active zones. Adaptive routing decreases from 0.84 to 0.78, compared with 0.80 to 0.63 for static routing; the separation consequently grows from 0.04 to 0.15. Concurrent-zone count describes the evaluated workload. Interpretation of the accuracy trend concerns the evaluated workload, while assessment of controller scheduling capacity requires per-zone call frequency, communication delays, and hardware-load measurements. Figure 4(d) specifies aggregate usage shares of 34% for micro-motion, 28% for schedule, 22% for environment, and 16% for transition experts. These percentages describe utilization; assessing expert specialization requires per-regime assignments or interventions.

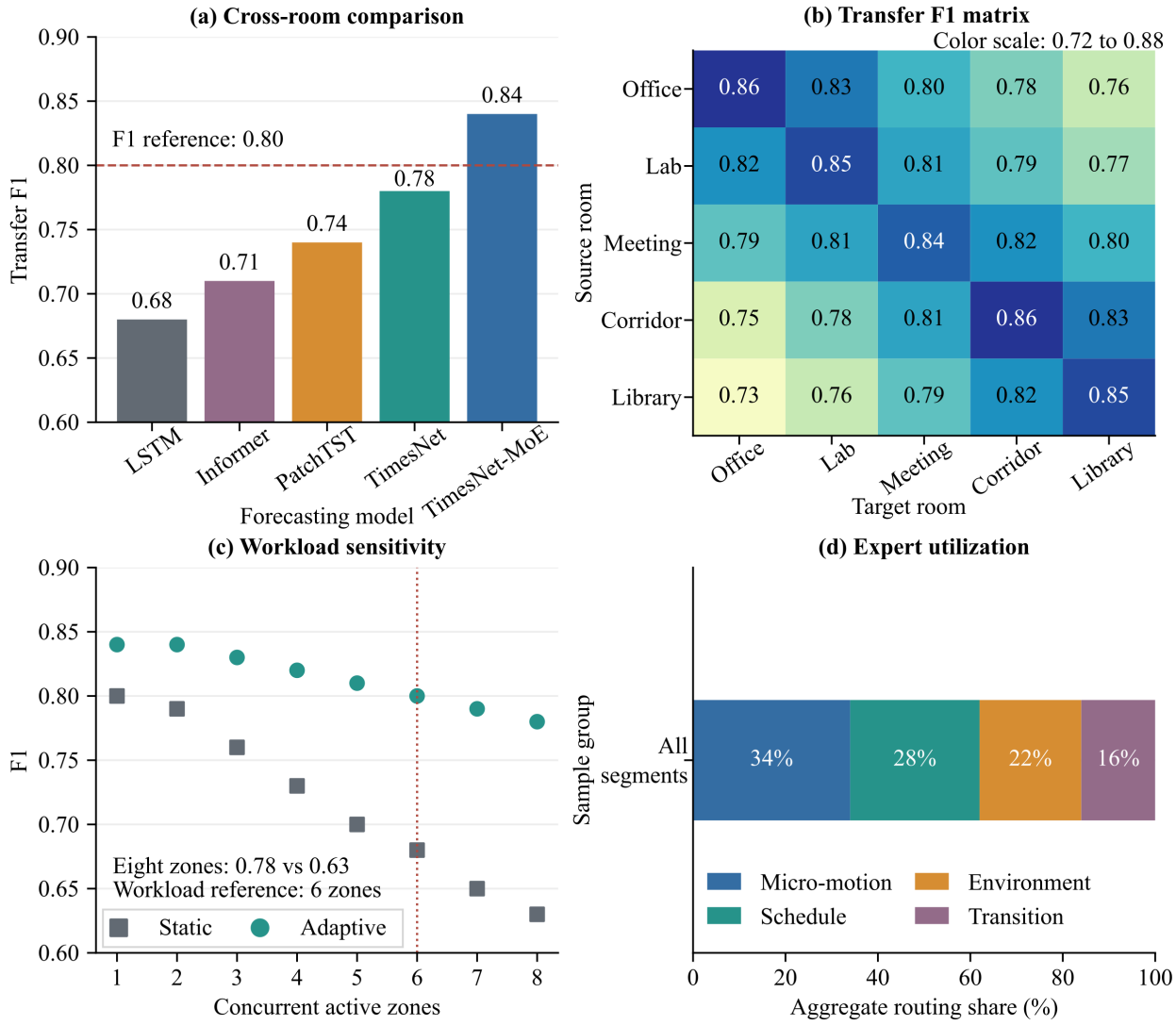


Figure 4. Transferability and routing behavior: (a) cross-room transfer; (b) source-target transfer matrix; (c) robustness to concurrent zones; (d) expert usage distribution

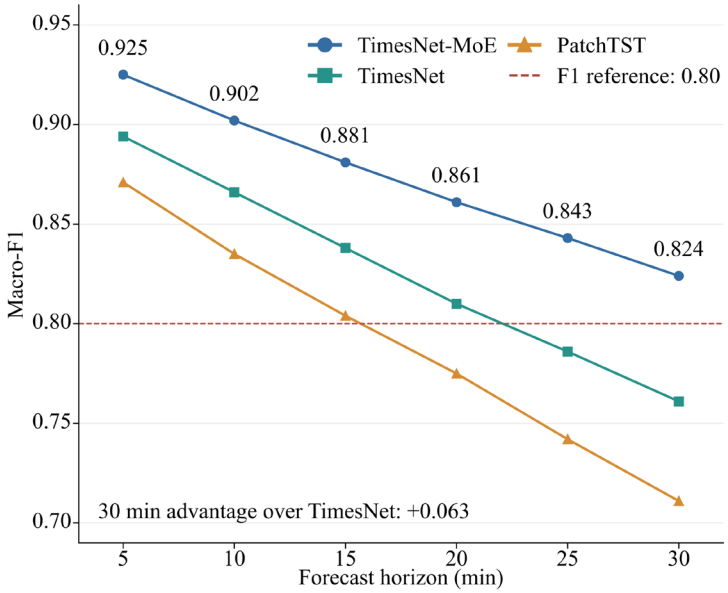


Figure 5. Horizon sensitivity in occupancy prediction

Figure 5 describes the relationship between forecast horizon and predictive accuracy over six horizons. TimesNet-MoE decreases from 0.925 at 5 min to 0.902, 0.881, 0.861, 0.843, and 0.824 at successive horizons. The corresponding

TimesNet sequence is 0.894, 0.866, 0.838, 0.810, 0.786, and 0.761; PatchTST decreases from 0.871 to 0.711. The F1 advantage over TimesNet increases from 0.031 at 5 min to 0.063 at 30 min, approximately doubling across the displayed horizons. Forecast-horizon selection has been examined in anticipatory lighting control because longer lead times can support gradual action while increasing occupancy uncertainty [21]. This pattern is compatible with a design that retains longer-duration regime information as immediate motion cues become less predictive, although confirming that mechanism requires controlled removal of the period encoder or gate. The 0.80 horizontal reference provides a descriptive F1 screening criterion for this comparison. Controller decisions use a separate occupancy-probability threshold, and operational acceptance additionally requires comfort-compliance evaluation. At 25 and 30 min, the proposed series remains above it, while the TimesNet series falls below it. The 15 and 20-min points provide intermediate operating conditions with proposed values of 0.881 and 0.861. The 15 and 20 min horizons provide candidate lead times for gradual dimming. Horizon selection requires transition-specific false-vacancy analysis, daylight conditions, and manual-override records alongside aggregate F1.

Table 2 fixes the principal comparison at 16 support samples and a 10 min horizon, separating that operating condition from the horizon sweep and transfer experiment. TimesNet-MoE achieved F1, probability error, and calibration error of 0.85, 0.121, and 0.057, compared with 0.79, 0.151, and 0.069 for TimesNet. The calibration difference is 0.012, or 17.4% relative to the TimesNet value. PatchTST records 0.76, 0.163, and 0.077, placing its classification and confidence measures between the simpler recurrent baseline and the two period-oriented configurations. Informer and LSTM have F1 values of 0.73 and 0.70, with probability errors of 0.177 and 0.190. These paired measures are relevant near the dimming threshold: two predictors can assign the same class while giving different probabilities that lead to different persistence times. The table also reports latency and energy-saving results, interpreted alongside the prediction metrics under the evaluation configuration. The four-expert reference architecture must remain distinct from the five-expert ablation point used later. The aggregate calibration error of 0.057 and the room-wise errors belong to separate reported comparison settings. Assessment of aggregation effects requires matched predictions, calibration-bin boundaries, and weighting rules across the comparison settings. Room-level reliability bins and class-specific confusion counts are required to assess quiet seated intervals, doorway motion, and unattended bookings. Such analysis would distinguish changes in false-vacancy detection, arrival detection, and majority-regime performance.

Table 2. Main comparison at 16 support samples and a 10 min prediction horizon

Model	F1	MAE	ECE	Latency *ms*	Energy saving *%*
LSTM	0.70	0.190	0.091	18	8.4
Informer	0.73	0.177	0.084	24	10.1
PatchTST	0.76	0.163	0.077	31	12.6
TimesNet	0.79	0.151	0.069	39	14.8
TimesNet-MoE	0.85	0.121	0.057	46	17.3

5.3. Robustness Ablation and Control Impact

Figure 6(a) specifies degradation under six dropped-sample rates from 0% to 25%. The TimesNet-MoE series decreases from 0.88 to 0.80, with a value of 0.82 at 20% missingness; TimesNet decreases from 0.82 to 0.69, and LSTM decreases from 0.79 to 0.62. At 20%, the F1 advantage over TimesNet is 0.10, compared with 0.06 under complete observations. The larger performance gap motivates targeted evaluation of masking and expert weighting under reduced motion evidence. Random removal represents intermittent packet loss; clustered outages and sensor-specific failures require separate trials because they may eliminate an entire temporal cue. Figure 6(b) specifies latency-energy pairs of 18 ms and 0.31 J for LSTM, 24 ms and 0.35 J for Informer, 31 ms and 0.40 J for PatchTST, 39 ms and 0.46 J for TimesNet, and 46 ms and 0.53 J for TimesNet-MoE. The proposed point exceeds the 40 ms reference by 6 ms and the 0.50 J reference by 0.03 J. Edge-inference evaluation commonly considers latency and energy alongside forecast performance when a model is assigned to a local building controller [22]. A 46 ms inference operation occupies approximately 0.077% of a one-minute scheduling interval. This ratio describes the timing allowance within the scheduling illustration. Sensor sampling rate and end-to-end controller utilization require separate measurements. Transfer of the latency and energy results requires the same hardware, numerical precision, batch size, and profiling boundary. The missing-sample sweep evaluates random removal. Generalization to channel-wide failures and temporally clustered outages requires additional failure-mode-specific evaluation.

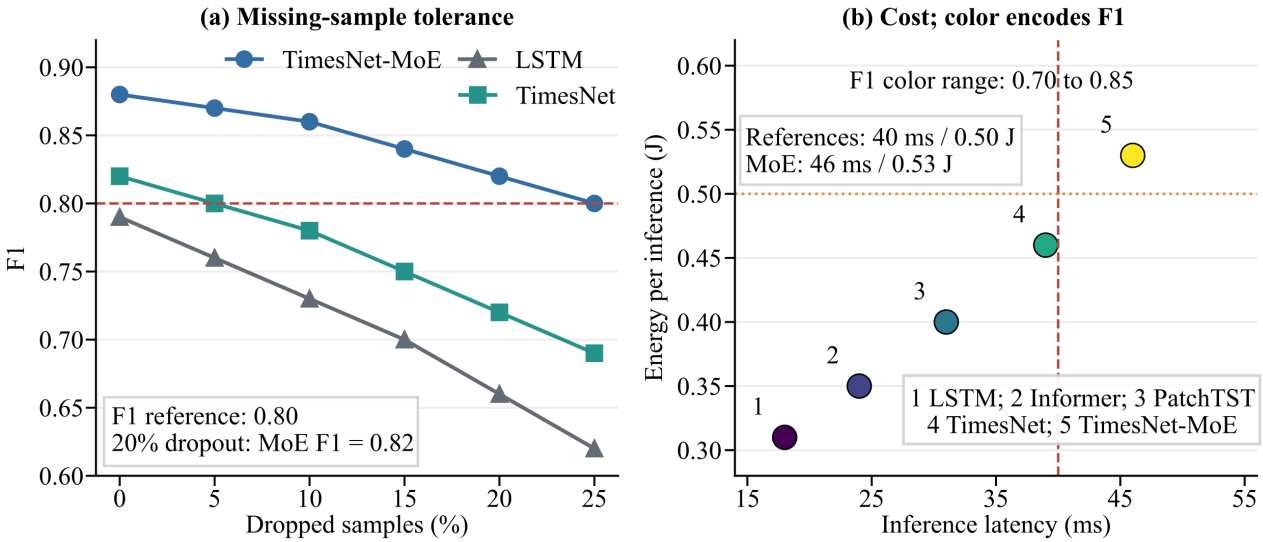


Figure 6. Robustness and deployment cost: (a) missing-sample tolerance; (b) edge cost versus accuracy

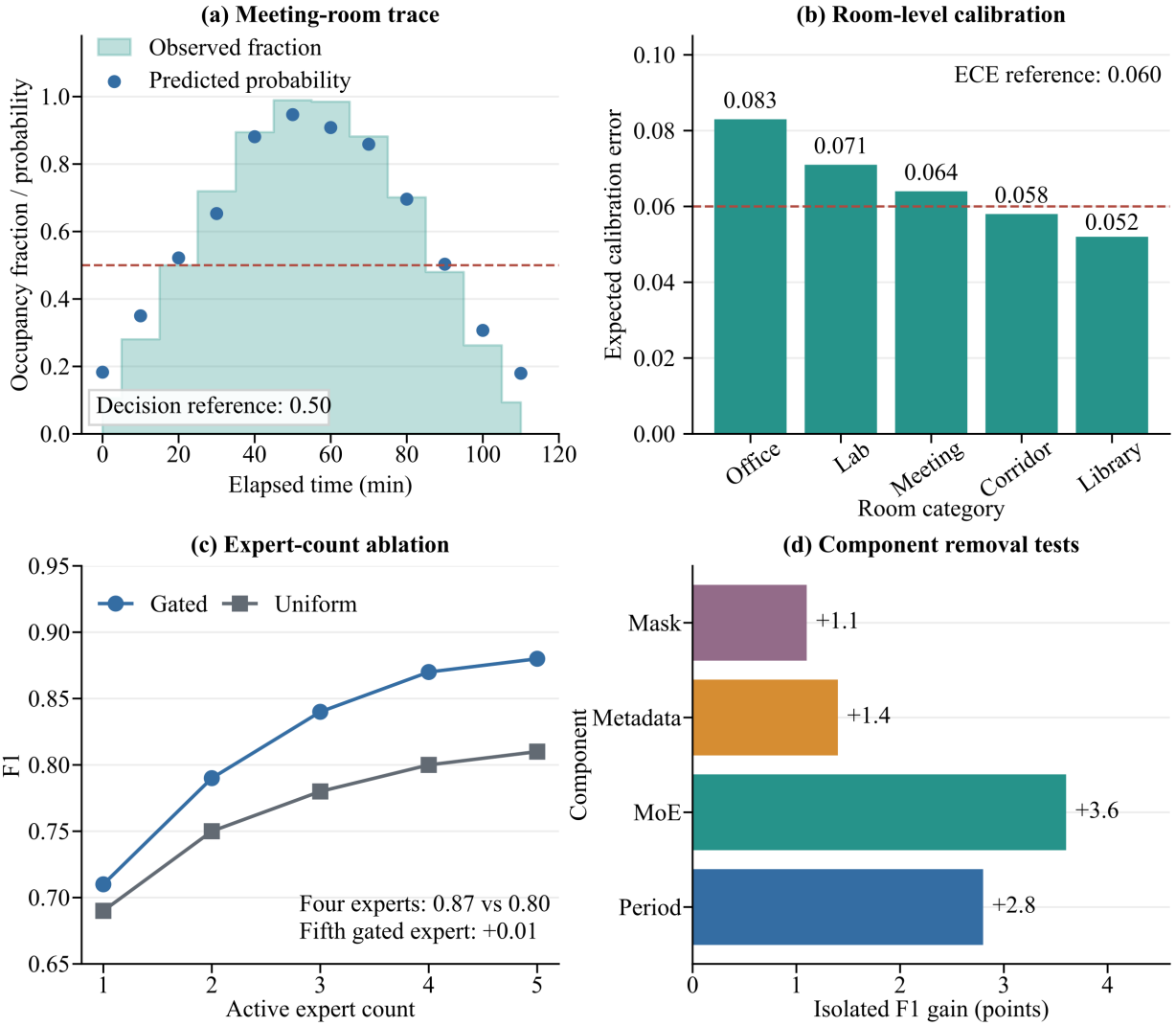


Figure 7. Temporal behavior and ablation evidence: (a) meeting-room occupancy fraction and probability; (b) room-wise calibration error; (c) expert-count ablation; (d) component contribution

Figure 7(a) presents the meeting-room occupancy-fraction trace and predicted probability sampled every 10 min. The observed sequence is an interval-level occupancy fraction used to illustrate temporal behaviour. The binary classification task uses a presence indicator, while the displayed curves support interval-level temporal interpretation. Binary prediction accuracy requires evaluation against the corresponding presence labels. At 20 min the observed fraction is

0.50, while the predicted probability is approximately 0.522; at 50 min the values are approximately 0.989 and 0.946. Near the trailing transition, the observed fraction decreases to approximately 0.263 at 100 min, while the prediction is approximately 0.307. These points describe transition lag and probability smoothing at the available sampling resolution. Figure 7(b) specifies room-level calibration errors of 0.083, 0.071, 0.064, 0.058, and 0.052 for office, laboratory, meeting room, corridor, and library. The first three exceed the 0.060 reference, indicating where room-specific calibration warrants closer inspection. Figure 7(c) compares one through five experts: gated F1 rises from 0.71 to 0.79, 0.84, 0.87, and 0.88, whereas uniform mixing rises from 0.69 to 0.75, 0.78, 0.80, and 0.81. Increasing the gated expert count from four to five adds 0.01 to F1, compared with a gain of 0.08 when increasing the count from one to two. Controlled component removal and matched training settings are established ways to attribute performance changes in temporal-model evaluation [23]. Figure 7(d) assigns component contributions of 2.8, 3.6, 1.4, and 1.1 F1 points to period encoding, mixture routing, metadata, and masking. These isolated ablations use a separate comparison setting and quantify individual component-removal effects in F1 percentage points. The contributions describe setting-dependent component-removal effects; assessment of their combined influence requires a joint ablation that accounts for interactions.

Figure 8(a) gives normalized daily schedule-baseline energy of 1.00, 0.94, 0.90, 0.86, and 0.83 across the five room categories, compared with prediction-aware values of 0.90, 0.82, 0.76, 0.71, and 0.67. Pairwise reductions against the schedule baseline are consequently 10.0%, 12.8%, 15.6%, 17.4%, and 19.3%. These relative reductions differ from the plotted normalized energy levels because each denominator is the corresponding schedule value. Lighting-design research examines simulation environments for assessing indoor illumination, providing context for joint consideration of energy use and occupant experience [24]. The larger reductions in corridors and libraries indicate greater scope for reducing scheduled illumination during vacancy; offices retain more lighting output during ambiguous quiet occupancy. Figure 8(b) separately specifies a controller-aggressiveness sweep with savings of 10.2%, 12.8%, 15.1%, 17.3%, and 19.4% paired with comfort compliance of 96.1%, 95.4%, 94.8%, 94.2%, and 93.5%. These five points represent different policy settings applied under common evaluation conditions. The 17.3% setting remains 0.2 percentage points above the 94% comfort floor, whereas the 19.4% setting falls 0.5 points below it. Increasing controller aggressiveness between these settings raises energy savings by 2.1 percentage points and reduces comfort compliance by 0.7 percentage points. Post-occupancy lighting evaluation has examined occupant light exposure and the interpretation of measurements in operational offices [25]. Within this policy sweep, the 17.3% savings setting is the highest-saving displayed setting that satisfies the specified 94% comfort floor. Its 0.2 percentage-point margin requires transition-level checks and an explicit comfort definition before operational acceptance.

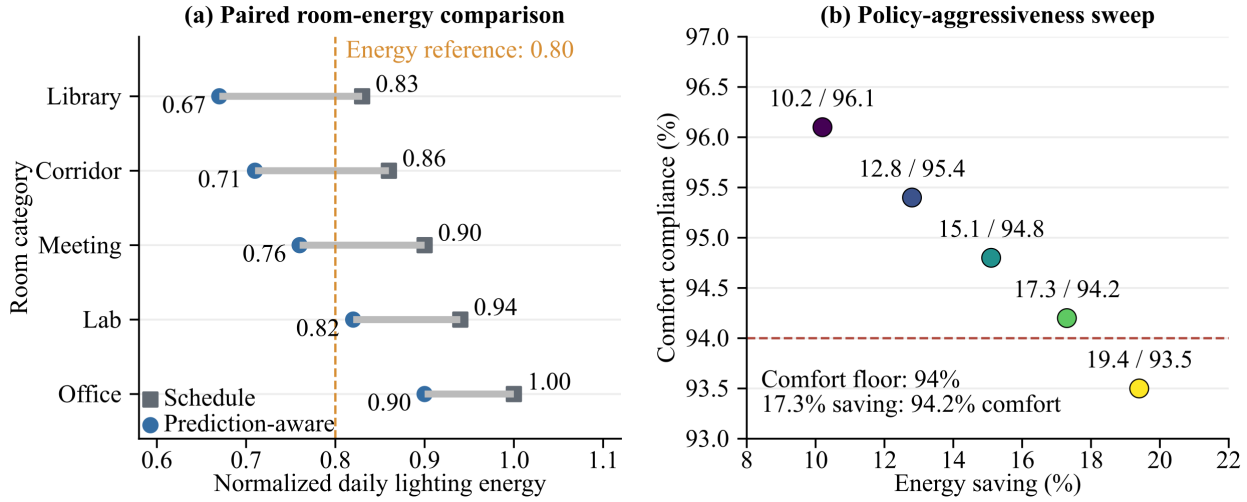


Figure 8. Smart-lighting control impact: (a) normalized lighting energy by room; (b) controller-aggressiveness sweep of energy saving and comfort compliance

Within the proposed architecture, a scheduled but unattended meeting can activate schedule-related features despite weak motion and environmental evidence. Quiet office work combines prolonged presence with limited movement, while strong ventilation can attenuate carbon dioxide accumulation. Doorway or corridor passage introduces spatial ambiguity because a brief sensor activation may originate from movement in a neighbouring zone. A difference between the booking schedule and actual stay has been identified as a cause of error in room-level energy management [26]. These conditions require separate records of false-vacancy events, false-occupancy events, and transition delays. Probabilistic forecasting research relates confidence calibration to the quality of downstream decisions [27]. The aggregate calibration error of 0.057 and the room-level range of 0.052-0.083 motivate examining reliability near the controller threshold, where small confidence errors may change a dimming decision. Maintenance and commissioning studies also consider sensor availability and labelling effort in the assessment of the actual cost of building-control

deployment [28]. Subsequent evaluation requires commissioning logs, timestamped packet-loss records, and manual lighting overrides aligned with predicted probabilities. Structured-outage testing first evaluates individual sensing-channel removal, followed by correlated motion and environmental losses. Reproducible hardware profiling requires the inference implementation, numerical precision, batch size, controller load, and energy-measurement boundary. Using consistent forecast and comfort definitions, the 46 ms latency and 0.53 J inference-energy results provide a basis for assessing controller resource requirements. Deployment assessment additionally requires installation costs, model-maintenance effort, and the consequences of incorrect vacancy interventions alongside potential energy reduction. These operational costs require separate measurement from inference energy.

6. Conclusion

This study developed a few-shot TimesNet-MoE framework for smart-lighting occupancy prediction, and the experimental comparisons showed improved predictive performance, calibration, and transfer behavior under limited labelling and heterogeneous room conditions.

1. Predictive performance under sparse supervision. At 16 support samples and a 10 min forecast horizon, the TimesNet-MoE configuration achieved a macro-F1 of 0.85, compared with 0.79 for plain TimesNet. Occupancy-probability mean absolute error decreased from 0.151 to 0.121, representing a relative reduction of 19.9%, while expected calibration error decreased from 0.069 to 0.057. Across the specified support budgets, macro-F1 increased from 0.76 at four samples to 0.90 at 64 samples, indicating diminishing gains as additional labels became available.

2. Transfer, robustness, and control trade-offs. Cross-room transfer F1 was 0.84 for TimesNet-MoE and 0.78 for TimesNet. With 20% of sensor samples removed, their respective F1 values were 0.82 and 0.72. The proposed configuration retained an F1 of 0.824 at a 30 min horizon, whereas TimesNet reached 0.761. These advantages accompanied higher inference costs: 46 ms and 0.53 J, compared with 39 ms and 0.46 J. In the separate controller-aggressiveness sweep, the 17.3% energy-saving setting maintained 94.2% comfort compliance, while the 19.4% setting reduced compliance below the specified comfort floor.

3. Contribution and applicability. The framework combines period-aware temporal representation with experts emphasizing motion, schedules, environmental persistence, and occupancy transitions. The observed performance gains are accompanied by higher inference latency and energy consumption, making controller resources an important deployment consideration. Applicability is presently bounded by the evaluated room categories and comparison settings. Wider generalization requires seasonal testing, repeat-level uncertainty estimates, complete acquisition metadata, and documented controller profiling.

Extended operational evaluation requires longitudinal records of routing stability, calibration drift, false-vacancy interventions, and manual overrides.

7. Declarations

7.1 Author Contributions

Octav Mănescu contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Dumitru Zavoi contributes to validation, analysis, investigation, data collection, draft preparation. All authors have read and agreed with the manuscript before its submission and publication.

7.2 Funding and Acknowledgments

This research received no specific financial support from any funding agency.

7.3 Institutional Review Board Statement

Not applicable.

7.4 Informed Consent Statement

Not applicable.

7.5 Date Availability

Not applicable.

7.6 Declaration of Competing 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.

7.7 Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors utilized Grammarly for linguistic polishing, logical organization and draft revision. The generative AI tool was not involved in research design, data collection, data analysis, or the equation of core conclusions. All outputs generated by AI were reviewed, revised and verified manually by the authors. The authors take full responsibility for all content of this manuscript.

8. References

[1] Cretu, G., Stamatescu, I., & Stamatescu, G. (2024). Modeling and prediction of occupancy in buildings based on sensor data using deep learning methods. IEEE Access, 12, 102994–103003. https://doi.org/10.1109/ACCESS.2024.3432584

[2] Parise, G., Zissis, G., & Martirano, L. (2025). Smart lighting systems, controls, and communication protocols: Introducing open communication protocols. IEEE Industry Applications Magazine, 31(2), 68–79. https://doi.org/10.1109/MIAS.2024.3482882

[3] Rafiq, I., Mahmood, A., Ahmed, U., Khan, A. R., Arshad, K., Assaleh, K., Ratyal, N. I., & Zoha, A. (2024). A hybrid approach for forecasting occupancy of building's multiple space types. IEEE Access, 12, 50202–50216. https://doi.org/10.1109/ACCESS.2024.3383918

[4] Vanhaeverbeke, J., Deprost, E., Verstockt, S., & van Hoecke, S. (2025). Cross-room CO_2-based presence detection for occupancy profiling. IEEE Access, 13, 120917–120930. https://doi.org/10.1109/ACCESS.2025.3588031

[5] Ao, X., Gong, Y., & He, A. (2025). A review of time series prediction models based on deep learning. IEEE Access, 13, 153696–153712. https://doi.org/10.1109/ACCESS.2025.3602791

[6] Banitaan, S., El Taher, T., Aldamasi, K., Hassoun, H., & Ahmed, S. (2026). Sensor-free occupancy forecasting for smart buildings: A Wi-Fi syslog approach with machine and deep learning. IEEE Access, 14, 10891–10909. https://doi.org/10.1109/ACCESS.2026.3654007

[7] Kong, X., Chen, Z., Liu, W., Ning, K., Zhang, L., Marier, S. M., Liu, Y., Chen, Y., & Xia, F. (2025). Deep learning for time series forecasting: A survey. International Journal of Machine Learning and Cybernetics, 16(7–8), 5079–5112. https://doi.org/10.1007/s13042-025-02560-w

[8] He, J., Yuan, K., Zhong, Z., & Sun, Y. (2024). Enhancing short-term power load forecasting with a TimesNet-Crossformer-LSTM approach. IEEE Access, 12, 56774–56788. https://doi.org/10.1109/ACCESS.2024.3383912

[9] Turgut, Z. (2025). XDOcc: An explainable artificial intelligence empowered deep framework for occupancy detection and occupant count estimation. IEEE Access, 13, 175386–175409. https://doi.org/10.1109/ACCESS.2025.3619449

[10] Chen, Z., Li, Q., & Huang, Z. (2026). Lighting-aware adaptive fusion for short-term lighting-load forecasting in smart-building lighting IoT. IEEE Access, 14, 145650–145660. https://doi.org/10.1109/ACCESS.2026.3735736

[11] Mena-Martinez, A., Davila Delgado, M., Alvarado-Uribe, J., & Ceballos, H. G. (2024). A real-life evaluation of supervised and semi-supervised machine learning approaches for indirect estimation of indoor occupancy. IEEE Access, 12, 118673–118693. https://doi.org/10.1109/ACCESS.2024.3449810

[12] Li, C., Denison, T., & Zhu, T. (2025). A survey of few-shot learning for biomedical time series. IEEE Reviews in Biomedical Engineering, 18, 192–210. https://doi.org/10.1109/RBME.2024.3492381

[13] Vettoruzzo, A., Bouguelia, M.-R., & Rögnvaldsson, T. (2025). Efficient few-shot human activity recognition via meta-learning and data augmentation. Neural Computing and Applications, 37(23), 19461–19477. https://doi.org/10.1007/s00521-025-11408-x

[14] Pederzoli, S., Del Buono, F., Vincini, M., & Guerra, F. (2026). Trend-error decomposition for self-supervised time series learning in multivariate forecasting task. IEEE Access, 14, 8618–8631. https://doi.org/10.1109/ACCESS.2026.3653488

[15] Cakiroglu, M. O., Kurban, H., Buxton, E., & Dalkilic, M. (2025). A novel discrete time series representation with de Bruijn graphs for enhanced forecasting using TimesNet. IEEE Access, 13, 123182–123198. https://doi.org/10.1109/ACCESS.2025.3588507

[16] Liu, X., Fu, Q., Chen, J., Liu, K., Liu, L., Wang, Y., & Lu, Y. (2026). Building-MoE: A closed-loop routing sparse mixture-of-experts time-series foundation model for building short-term load forecasting. Building Simulation, 19(4), 1007–1030. https://doi.org/10.1007/s12273-026-1403-6

[17] Altin, M., Cakir, A., & Toreyin, B. U. (2026). Time-series forecasting and evaluation on city sensing data: Pedestrian and microclimate benchmarks. IEEE Access, 14, 22618–22629. https://doi.org/10.1109/ACCESS.2026.3662469

[18] Papanashi, S., Hegde, N. V., Bhat, V. M., Bhagya, R., Reddy, B. R., & Patel, S. (2026). A lightweight, real-time human detection and tracking system for privacy-preserving occupancy monitoring using low-resolution thermal sensors. IEEE Access, 14, 16306–16325. https://doi.org/10.1109/ACCESS.2026.3658275

[19] Dintén, R., & Zorrilla, M. (2025). Using time series foundation models for few-shot remaining useful life prediction of aircraft engines. Computer Modeling in Engineering & Sciences, 144(1), 239–265. https://doi.org/10.32604/cmes.2025.065461

[20] Lin, C.-Y., Emon, S. H., & Tseng, T.-L. (2026). GeMeR-TS: Generative memory-based retrieval and mixture-of-experts for cross-domain time-series forecasting. IEEE Transactions on Big Data, 12(4), 1401–1419. https://doi.org/10.1109/TBDATA.2026.3689009

[21] Muslimin, N. A., Mansor, N. N., Mokhlis, H., Abdullah, M. N., & Ramli, S. P. (2025). Lighting system optimization, components, and simulation tools for building installations: A review. IEEE Access, 13, 160508–160535. https://doi.org/10.1109/ACCESS.2025.3578983

[22] Malka, M., Farhan, E., Morgenstern, H., & Shlezinger, N. (2025). Decentralized low-latency collaborative inference via ensembles on the edge. IEEE Transactions on Wireless Communications, 24(1), 598–614. https://doi.org/10.1109/TWC.2024.3497167

[23] Zhao, Z., Cao, L., & Wan, Y. (2025). Mixture of online and offline experts for non-stationary time series. Proceedings of the AAAI Conference on Artificial Intelligence, 39(21), 22867–22874. https://doi.org/10.1609/aaai.v39i21.34448

[24] Mohammadrezaei, E., Ghasemi, S., Dongre, P., Gračanin, D., & Zhang, H. (2024). Systematic review of extended reality for smart built environments lighting design simulations. IEEE Access, 12, 17058–17089. https://doi.org/10.1109/ACCESS.2024.3359167

[25] Hartmeyer, S. L., Davies, A., & Andersen, M. (2026). Post-occupancy evaluation of office lighting – What can we learn from light-dosimetry? LEUKOS, 22(1), 86–109. https://doi.org/10.1080/15502724.2024.2434976

[26] Goujili, N., Saumard, M., Cochard, N., & Jridi, M. (2025). Functional data analysis for prediction of employee presence. IEEE Access, 13, 114179–114198. https://doi.org/10.1109/ACCESS.2025.3580697

[27] Hu, K., & Su, M. (2026). Probabilistic forecasting of multi-energy loads in integrated energy systems via condition-aware dynamic coupling routing and conformal quantile calibration. IEEE Access, 14, 113437–113460. https://doi.org/10.1109/ACCESS.2026.3716060

[28] Kotbi, A. G. (2026). Integration gaps in smart indoor lighting: A systematic bibliometric review of human-centric control, IoT, and event-driven automation. Journal of Daylighting, 13(2), 575–597. https://doi.org/10.15627/jd.2026.32