An Intelligent Occupancy Detection Framework Using Multimodal Sensory Data in Smart Buildings

Giorgia Petrocelli
Department of Computer Science, Binghamton University, Binghamton, NY, USA.
contactgiorgia@binghamton.edu

Rozalia Utrata
Department of Computer Science, University of Central Florida, Orlando, FL, USA.
rozalia2004@ucf.edu

Abstract

Occupancy detection in smart buildings is a decision problem under uncertainty: single sensors fail under occlusion, drift, environmental lag, or changing schedules, while building control requires timely and calibrated presence estimates. This manuscript develops an intelligent occupancy detection framework that fuses heterogeneous sensory streams—passive infrared events, carbon dioxide, temperature, humidity, illuminance, acoustics, and plug or circuit power—through modality-specific encoders, reliability-weighted attention, temporal inference, and cost-sensitive decision making. The central contribution is a coherent probabilistic formulation that treats each sensor stream as an imperfect observation of a latent occupancy state and learns both discriminative representations and modality reliability. The framework separates representation learning, cross-modal fusion, and temporal smoothing so that missing or degraded modalities can be discounted rather than forcing a brittle fixed fusion rule. A hidden-state transition model and a conditional temporal objective enforce duration and switching regularities; calibration terms and an uncertainty decomposition support abstention, sensor maintenance, and risk-aware actuation. The paper specifies model assumptions, objective functions, inference steps, boundary conditions, and evaluation logic, and it analyzes failure modes such as carbon dioxide lag, passive infrared silence during stillness, acoustic contamination, and prior shift across zones. Because the framework is grounded in explicit probabilistic and decision-theoretic relationships, it can be instantiated with different neural encoders and label sources without altering the occupancy inference logic. The resulting design offers a technically defensible route from multimodal sensing to calibrated occupancy evidence for lighting, ventilation, and heating control.

Keywords

occupancy detection; multimodal sensing; data fusion; probabilistic inference; uncertainty calibration; smart buildings

1. Introduction

Reliable occupancy detection is a foundational function in smart buildings because lighting, ventilation, heating, and security policies depend on whether a zone is occupied, how many occupants are present, and how long the state persists. Environmental sensing has been used for this purpose because carbon dioxide, temperature, humidity, and illuminance respond to human presence and activity; statistical learning models applied to such measurements have demonstrated that office occupancy can be inferred from ordinary building sensors [1]. Earlier probabilistic work treated sensor evidence through belief networks, showing that

occupancy is better represented as a latent state than as a direct threshold crossing [2]. Reviews of occupancy sensing systems have since emphasized that no single sensing modality is uniformly reliable across room types, occupancy densities, and operational schedules [3].

The building-control literature provides a second motivation. User activity and occupancy are central to energy-intelligent buildings because schedules and static setpoints often misrepresent actual use [4]. Occupancy-driven automation can reduce wasted conditioning and lighting, but only if presence estimates are accurate enough for control decisions and timely enough for actuation [5]. WiFi and network-side occupancy sensing extend coverage where dedicated sensors are sparse, yet these signals are affected by device association behavior and privacy constraints [6]. Comparative reviews of occupancy estimation and detection reinforce that sensing choice, spatial resolution, temporal resolution, and ground-truth design jointly determine whether a method is appropriate [7].

Machine learning has shifted the field from rule-based thresholds to learned mappings from sensor histories to occupancy labels. Indoor environmental data have been used to train occupancy prediction models with several classical learners, demonstrating the value of feature engineering and temporal aggregation [8]. Data-analytics reviews further classify occupancy sensing approaches by sensing hardware, label acquisition, feature representation, and inference family, while noting persistent difficulties in transferability and evaluation consistency [9]. Demand-driven control studies show that occupancy measurement is not merely a classification problem; it is a measurement and decision problem in which false positives and false negatives have asymmetric operational costs [10].

The remaining design gap is not the absence of sensors or classifiers but the absence of a coherent framework that treats multimodal sensing as evidence rather than as a feature-concatenation exercise. Fixed fusion rules assume that all modalities are present and equally informative, which is rarely true in buildings. Passive infrared sensors can remain silent when occupants are still; carbon dioxide concentrations lag behind changes in occupancy; acoustic features are sensitive to machinery and media; power meters may aggregate multiple zones; and environmental variables drift with season and HVAC operation. A practical framework must therefore encode modality-specific structure, estimate reliability at inference time, model temporal persistence, calibrate probabilities, and connect occupancy evidence to cost-sensitive decisions.

This paper develops such a framework. Its central contribution is a probabilistic and decision-theoretic architecture for intelligent occupancy detection using multimodal sensory data in smart buildings. The framework represents each sensor stream through a modality-specific encoder, fuses the resulting evidence through reliability-weighted attention, imposes temporal coherence through a hidden-state or conditional sequence model, and produces calibrated occupancy probabilities that can be thresholded according to false-positive and false-negative costs. The design is intentionally modular: the same inference logic can accept different encoders, different label sources, and different building configurations without changing the statistical meaning of the occupancy estimate. The following sections define the observation model, fusion mechanism, temporal inference, training objective, calibration and decision rules, and the technical conditions under which the framework is expected to remain robust.

2. Related Work and Design Requirements

Occupancy detection methods can be grouped by the sensing physics they exploit. Environmental methods infer presence from human-induced changes in carbon dioxide, temperature, humidity, and illuminance [1], [8]. Event-based methods use passive infrared or door contacts, which provide high precision for transitions but poor sensitivity to stationary occupants [2], [3]. Network-based methods use WiFi association, probe requests, or traffic metadata to estimate presence over larger zones, but their coverage depends on device use and privacy policy [6]. Power-based methods detect loads associated with occupants, yet aggregate circuits and standby loads complicate inference [5]. Each modality therefore carries a distinct error profile, and the design question is how to combine them without assuming stationarity or complete observation.

Design studies of multimodal perception for smart environments emphasize co-located sensing, heterogeneous interfaces, calibration constraints, and the need to align data streams before inference [11]. Multi-sensor fusion research provides a broader vocabulary for this problem: fusion can occur at the signal level, feature level, or decision level, and the appropriate level depends on temporal synchronization, modality dimensionality, and the interpretability required by the application [12]. Multimodal analysis surveys similarly distinguish early, late, and hybrid fusion, noting that attention-based and reliability-aware schemes are useful when modalities are partially redundant and partially complementary [13]. For occupancy detection, feature-level fusion is attractive because low-level signals such as carbon dioxide slope, passive infrared event counts, and power variation are not individually sufficient but jointly informative. Decision-level fusion remains useful when subsystems have different update rates or when building operators require modular certification of sensors.

The requirements for a smart-building occupancy framework follow from these observations. First, the model must accept asynchronous and missing modalities. Sensor failures, maintenance, and packet loss are normal operating conditions, not exceptional cases. Second, the model must represent temporal persistence. Occupancy is a state with inertia: a single negative passive infrared sample during a period of desk work should not overturn strong carbon dioxide and power evidence, whereas a door event followed by a sustained drop in power and carbon dioxide should produce a transition. Third, the output must be probabilistic. Control policies need to distinguish confident absence from uncertain absence, and maintenance systems need to detect when a sensor modality has become uninformative. Fourth, the framework must support cost-sensitive decisions because a false positive that conditions an empty room wastes energy, while a false negative that leaves an occupied room unconditioned affects comfort and may create safety concerns. Fifth, evaluation must avoid leakage from future information and must report performance under realistic class imbalance, temporal drift, and zone-level transfer.

These requirements motivate a modular architecture rather than a single end-to-end classifier. A single classifier can learn useful interactions, but it obscures which modality supports a decision, makes missing-modality behavior difficult to reason about, and complicates calibration when the training distribution differs from deployment. The proposed framework therefore separates four functions: modality encoding, reliability-aware fusion, temporal occupancy inference, and decision making. The separation does not prohibit joint training; it defines the variables over which joint training and diagnostics operate.

3. Proposed Multimodal Occupancy Detection Framework

3.1 Observation Modeling and Modality Encoders

Let t index synchronized decision epochs, and let z_t ∈ {0,1} denote latent occupancy for a zone, where 1 indicates at least one occupant and 0 indicates absence. Each modality m ∈ {1,...,M} produces an observation x_t^m. The modalities may have different native sampling rates; before encoding, each stream is resampled to the decision epoch and accompanied by a mask q_t^m ∈ {0,1} that indicates whether the modality is available. For passive infrared, x_t^m may be an event count in a window; for carbon dioxide, a concentration and slope; for temperature and humidity, filtered values; for acoustics, a compact spectral or loudness representation; and for power, a circuit-level demand summary.

Each modality is mapped to a compact embedding by a modality-specific encoder:

Here φ_m is a function parameterized by θ_m, and e_t^m is the encoded evidence for modality m at epoch t. The encoder can be a multilayer perceptron for low-dimensional environmental features, a temporal convolutional network for carbon dioxide and power histories, or a recurrent network for event sequences. Long short-term memory networks are a natural choice when the encoder must retain information over irregular intervals because their gated state can attenuate irrelevant fluctuations and preserve evidence of persistent changes [14]. The use of separate encoders is deliberate: carbon dioxide has a slow, integrative response, while passive infrared and acoustic signals are event-like. Forcing these streams through a shared early encoder would obscure their distinct temporal statistics.

3.2 Reliability-Aware Fusion

After encoding, the framework fuses modality embeddings through reliability-weighted attention. A scalar reliability score is computed for each modality from its embedding and an availability mask, and the scores are normalized across available modalities. The fused representation is a convex combination of available embeddings:

$$e_t = \sum_\{m=1\}\^M \alpha_t\^m \, e_t\^m, \; \alpha_t\^m = softmax(\psi_m(e_t\^m) + \rho_m)$$

In this expression, e_t is the fused evidence vector, α_t^m is the attention weight assigned to modality m, ψ_m is a learned scoring function, and ρ_m is a modality-specific bias. The softmax is taken over available modalities; when q_t^m = 0, the corresponding weight is set to zero. The bias term allows the model to encode prior reliability learned from training, while the embedding-dependent score allows reliability to vary by context. For example, a passive infrared sensor may receive low weight when carbon dioxide and power indicate sustained occupancy but no recent motion, and a power meter may receive low weight when its circuit aggregates several zones and its variation is uninformative.

This fusion rule is more defensible than fixed averaging because it makes the evidence weighting explicit and inspectable. The weights can be logged during operation and compared with sensor health indicators. If a modality consistently receives low weight, the building operator can investigate calibration, placement, or failure. The rule also supports graceful degradation: if a modality is missing, the normalization redistributes weight over the remaining modalities rather than requiring imputation of the missing raw signal. Imputation can still be used inside an encoder, but the fusion layer does not depend on it.

3.3 Temporal Occupancy Inference

The fused evidence is converted into occupancy probabilities and then smoothed over time. A hidden Markov or conditional sequence formulation provides a transparent way to encode persistence and switching. In the hidden-state view, occupancy evolves according to a transition distribution and emits sensor evidence according to an emission distribution. For a sequence of epochs 1,...,T, the posterior over occupancy states can be written as:

$$\mathrm{P}(z_{1:T}|x_{1:T}) \propto \mathrm{P}(z_1)\prod_{t=2}^{T}\mathrm{P}(z_t|z_{t-1})\prod_{t=1}^{T}g(e_t|z_t)$$

Here P(z_1) is the initial state distribution, P(z_t|z_{t-1}) is the transition probability, and g(e_t|z_t) is the emission likelihood associated with the fused evidence. The transition matrix encodes expected persistence: the self-transition probability for occupied and unoccupied states should be high when decision epochs are short, but it should be calibrated to the zone's use pattern and to the decision epoch length. If a zone is a meeting room with frequent transitions, a single global transition matrix may be too smooth; in that case, the transition probability can be conditioned on time of day, door events, or calendar context. Hidden Markov models provide an established inference formalism for such latent-state problems [16].

An alternative is a conditional random field over the sequence, in which the posterior is modeled directly as a function of features rather than through a generative emission model. Conditional random fields can incorporate rich overlap features such as time since last passive infrared event, carbon dioxide slope, and power deviation without specifying their generative distributions [17]. The framework can use either formulation: the hidden-state model is preferable when interpretability and explicit transition probabilities are required, while the conditional model is preferable when labeled sequences are abundant and the feature set is heterogeneous. In both cases, the temporal layer outputs a calibrated probability p_t = P(z_t = 1 | x_{1:T}) or a causal filtering variant that uses only evidence up to t, depending on whether the application permits smoothing or requires real-time decisions.

3.4 Training Objective and Calibration

The model can be trained with supervised occupancy labels, weak labels from door events or schedules, or a combination. The objective separates classification accuracy from temporal consistency and calibration. A general form is:

$$L = L_cls + \lambda 1\, L_temp + \lambda 2\, L_cal + \lambda 3\, ||\theta||_2\text{^}2$$

In this expression, L_cls is a classification loss such as binary cross-entropy, L_temp is a temporal consistency or sequence likelihood term, L_cal is a calibration penalty, θ denotes all trainable parameters, and λ1, λ2, and λ3 are nonnegative weights. The temporal term discourages isolated state flips that are inconsistent with the transition model; it may be the negative log-likelihood of the sequence model or a penalty on changes that are not supported by a sustained evidence shift. The calibration term is necessary because a classifier trained only for discrimination can produce probabilities that are overconfident, especially under class imbalance. A simple calibration penalty is the Brier score over decision epochs:

$$L_cal = \left(\frac{1}{T}\right)\sum_{t=1}^{T}(p_t - z_t)^2$$

Here p_t is the predicted occupancy probability and z_t is the label. The Brier score is a proper scoring rule, so it rewards both discrimination and calibration. Training can also use

temperature scaling or isotonic regression after the main objective, but a calibration term inside the objective keeps the learned representation from becoming unnecessarily overconfident. Dropout-based approximate Bayesian inference offers another route to uncertainty estimates by treating dropout at inference time as an approximation to model uncertainty [18]. The optimizer for the resulting objective can be stochastic gradient descent with adaptive moment estimation, which is suitable for sparse and heterogeneous parameter updates [19].

3.5 Decision Rule and Uncertainty

Occupancy probabilities become control actions through a cost-sensitive rule. Let C_FP be the cost of a false positive and C_FN be the cost of a false negative. If only presence and absence decisions are allowed and the costs are linear in the decision, the Bayes decision rule is:

$$\hat{z}_t \;=\; 1\; if\; p_t\; C_FN \;>\; (1 \;-\; p_t)\; C_FP\; else\; 0$$

The rule states that occupancy should be declared when the expected cost of absence, p_t C_FN, exceeds the expected cost of presence, (1 - p_t) C_FP. In practice, the costs are policy parameters rather than fixed physical constants: a ventilation system may tolerate a higher false-positive rate than a security system, and a lighting zone may use a different threshold from a heating zone. The framework therefore exposes the threshold as a decision variable rather than hiding it inside a classifier.

Uncertainty can be decomposed to support abstention and maintenance. If an ensemble or Monte Carlo dropout produces K probability estimates p_t^(k), the total predictive variance can be approximated as:

$$u_t\text{^}2 \;=\; Var_k(p_t\text{^}(k)) \;+\; Mean_k[p_t\text{^}(k)(1 - p_t\text{^}(k))]$$

The first term captures disagreement among models or dropout samples, and the second term captures average Bernoulli variance. High total uncertainty can trigger a conservative fallback, a request for additional evidence, or a maintenance flag if it persists for a modality. This distinction is useful because epistemic disagreement may indicate distribution shift, while high Bernoulli variance may simply indicate proximity to a decision boundary.

4. Analytical Assessment and Technical Discussion

4.1 Missing-Modality Behavior and Sensor Degradation

The framework's reliability-weighted fusion is designed for missing and degraded modalities, but its behavior depends on whether the remaining modalities carry complementary information. If passive infrared is missing in a small office with a dedicated power meter and carbon dioxide sensor, occupancy inference may remain informative because the remaining streams respond to human presence. If the power meter is missing and the carbon dioxide sensor is poorly ventilated, the same office may become difficult to classify during short absences. The reliability weights do not create information; they reallocate trust. This distinction is important for deployment because a multimodal system should not be marketed as immune to sensor loss. The correct claim is that the fusion layer can degrade gracefully when redundancy exists and can expose low-reliability conditions when it does not.

Carbon dioxide sensing illustrates a specific failure mode. Carbon dioxide concentration responds to exhalation and ventilation, so its relationship to occupancy depends on room

volume, air exchange rate, and sensor placement [20]. A rise in carbon dioxide is strong evidence of recent occupancy, but a decline can lag behind departure. If the temporal model treats carbon dioxide as a direct occupancy indicator without accounting for the lag, it may overestimate occupancy after occupants leave. The transition model partially mitigates this by requiring sustained evidence for a state change, but the emission model should include history and slope, not just the current concentration. Similarly, passive infrared silence during still occupancy is not evidence of absence; it is evidence of no recent motion. The encoder should therefore represent time since the last event as a feature, and the fusion layer should learn that an old passive infrared event is weak negative evidence rather than a decisive absence signal.

4.2 Temporal Leakage, Label Noise, and Evaluation Logic

The framework's temporal component makes evaluation design consequential. Random train-test splits over decision epochs can leak information because adjacent epochs are highly correlated. A sequence-aware split that holds out whole days, weeks, or zones is more informative, although it is also more stringent. The evaluation should report performance at the decision epoch used by the controller, not only at a finer or coarser resolution that flatters the model. It should also report performance under class imbalance and under prior shift. If the training set contains mostly occupied periods and the deployment zone is mostly empty, calibration can drift even when ranking performance remains acceptable.

Label noise is another boundary condition. Occupancy labels may come from door counts, manual annotations, seat sensors, or schedule-based proxies, none of which is perfectly aligned with the latent state used by the controller. Door counts can miss occupants who enter together, manual labels can be sparse, and schedules can misrepresent actual use. The framework treats labels as observations of z_t rather than as ground truth, which is conceptually appropriate but does not remove the need for label-quality reporting. If labels are weak, the calibration term should be interpreted cautiously because it calibrates to the label process as well as to occupancy. A practical evaluation should therefore include a small high-quality validation set for calibration and a larger weakly labeled set for representation learning.

4.3 Computational and Deployment Trade-offs

The framework is modular, but modularity has costs. Separate encoders increase parameter count and require synchronization and masking logic. Attention-based fusion adds computation, although for a small number of modalities the cost is modest relative to the encoder. Temporal inference adds state and can complicate real-time deployment if the model uses bidirectional smoothing. A causal filter is preferable for control because it uses only past and present evidence, while a smoother is preferable for post-hoc analytics and fault diagnosis. The same trained model can support both if the temporal layer is implemented as a filter with an optional backward pass.

Privacy and governance are also technical constraints. Network-based occupancy sensing can reveal device presence patterns, and acoustic sensing can capture speech if features are not carefully restricted. The framework does not require raw audio retention; encoders can consume low-dimensional features such as loudness or spectral summaries and discard raw waveforms. Power sensing can reveal appliance usage, so circuit-level aggregation may be preferable to plug-level monitoring in shared spaces. These choices affect the evidence available to the model, which is why privacy-preserving feature design should be considered part of the sensing architecture rather than an afterthought.

The framework's practical significance lies in its separation of evidence, uncertainty, and action. Building operators can inspect modality weights, calibration curves, and uncertainty flags; control designers can set cost ratios without retraining the encoder; and maintenance teams can use persistent low reliability or high uncertainty as diagnostic signals. These properties are more valuable in real buildings than a marginal improvement in classification accuracy under a fixed benchmark condition, because real buildings experience sensor replacement, zone reconfiguration, and schedule change.

5. Conclusion

This manuscript presented an intelligent occupancy detection framework for smart buildings that treats multimodal sensory data as imperfect evidence about a latent occupancy state. The framework combines modality-specific encoders, reliability-weighted fusion, temporal occupancy inference, calibration-aware training, and cost-sensitive decision rules. Its central design choice is to separate representation, fusion, temporal smoothing, and decision making so that missing modalities, sensor degradation, and asymmetric control costs can be handled explicitly rather than absorbed into an opaque classifier.

The analytical assessment identified the conditions under which the framework is expected to be informative and the conditions under which it can fail. Passive infrared sensing is sensitive to motion but not stillness; carbon dioxide sensing is integrative but lagged; power and network sensing depend on infrastructure and occupancy-device coupling; acoustic sensing requires feature restrictions for privacy and robustness. Reliability weights can reallocate trust, but they cannot manufacture information that is absent from all modalities. Temporal models can enforce persistence, but they can also smooth over real transitions if the transition distribution is misspecified. Calibration can make probabilities more useful for control, but it remains sensitive to label quality and prior shift.

The framework is intended as a rigorous baseline for future empirical work rather than a claim of universal performance. Its equations define a trainable model, a sequence inference procedure, an uncertainty decomposition, and a decision rule that can be instantiated with different sensors and zones. The next technical steps are to evaluate the framework under sequence-aware splits, missing-modality perturbations, and cost-sensitive operating points, and to compare reliability-weighted fusion with fixed fusion and single-modality baselines. The broader implication is that occupancy detection in smart buildings should be evaluated as an evidence-to-decision pipeline: accuracy matters, but so do calibration, failure visibility, and the operational cost of being wrong.

References

1. Candanedo, L. M., & Feldheim, V. (2016). Accurate occupancy detection of an office room from light, temperature, humidity and CO2 measurements using statistical learning models. Energy and Buildings, 112, 28–39. https://doi.org/10.1016/j.enbuild.2015.11.071

2. Dodier, R. H., Henze, G. P., Tiller, D. K., & Guo, X. (2006). Building occupancy detection through sensor belief networks. Energy and Buildings, 38(9), 1033–1043. https://doi.org/10.1016/j.enbuild.2005.12.001

3. Yang, J., Santamouris, M., & Lee, S. E. (2016). Review of occupancy sensing systems and occupancy modeling methodologies for the application in institutional buildings. Energy and Buildings, 121, 344–355. https://doi.org/10.1016/j.enbuild.2015.12.018

4. Nguyen, T. A., & Aiello, M. (2013). Energy intelligent buildings based on user activity: A survey. Energy and Buildings, 56, 244–257. https://doi.org/10.1016/j.enbuild.2012.09.005

5. Agarwal, Y., Balaji, B., Gupta, R., Lyles, J., Wei, M., & Weng, T. (2010). Occupancy-driven energy management for smart building automation. In Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings (pp. 1–6). ACM. https://doi.org/10.1145/1878431.1878433

6. Balaji, B., Xu, J., Nwokafor, A., Gupta, R., & Agarwal, Y. (2013). Sentinel: Occupancy based HVAC actuation using existing WiFi infrastructure within commercial buildings. In Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems (pp. 1–14). ACM. https://doi.org/10.1145/2517351.2517370

7. Chen, Z., Jiang, C., & Xie, L. (2018). Building occupancy estimation and detection: A review. Energy and Buildings, 169, 157–170. https://doi.org/10.1016/j.enbuild.2018.03.031

8. Ryu, S. H., & Moon, H. J. (2016). Development of an occupancy prediction model using indoor environmental data based on machine learning techniques. Building and Environment, 107, 1–9. https://doi.org/10.1016/j.buildenv.2016.06.039

9. Saha, H., Florita, A. R., Henze, G. P., Sarkar, S., & Vaze, J. (2019). Occupancy sensing in buildings: A review of data analytics approaches. Energy and Buildings, 188–189, 263–278. https://doi.org/10.1016/j.enbuild.2019.02.020

10. Labeodan, T., Zeiler, W., Boxem, G., & Zhao, Y. (2015). Occupancy measurement in commercial office buildings for demand-driven control applications—A survey and detection system evaluation. Energy and Buildings, 93, 303–314. https://doi.org/10.1016/j.enbuild.2015.02.028

11. Liu, J., Luo, D., Fu, X., Lu, Q., & Kang, K. Y. (2022). Design strategy of multimodal perception system for smart environment. In Internet of Things for Smart Environments (pp. 93-115). Cham: Springer International Publishing.

12. Gravina, R., Alinia, P., Ghasemzadeh, H., & Fortino, G. (2017). Multi-sensor fusion in body sensor networks: State-of-the-art and research challenges. Information Fusion, 35, 68–80. https://doi.org/10.1016/j.inffus.2016.09.005

13. Atrey, P. K., Hossain, M. A., El Saddik, A., & Kankanhalli, M. S. (2010). Multimodal fusion for multimedia analysis: A survey. Multimedia Systems, 16(6), 345–379. https://doi.org/10.1007/s00530-010-0182-0

14. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

15. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems (Vol. 30, pp. 5998–6008).

16. Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2), 257–286. https://doi.org/10.1109/5.18626

17. Lafferty, J., McCallum, A., & Pereira, F. C. N. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In Proceedings of the 18th International Conference on Machine Learning (pp. 282–289). Morgan Kaufmann.

18. Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. In Proceedings of the 33rd International Conference on Machine Learning (pp. 1050–1059). PMLR.

19. Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In Proceedings of the 3rd International Conference on Learning Representations.

20. Calì, D., Matthes, P., Huchtemann, K., Streblow, R., & Müller, D. (2015). CO2 based occupancy detection algorithm: Experimental analysis and validation for office and residential buildings. Building and Environment, 86, 39–49. https://doi.org/10.1016/j.buildenv.2014.12.011