# Distributed IoT Climate Management for Residential Buildings via Cooperative Multi-Zone Reinforcement Learning with Federated Thermal Modeling

Farrukhzhon Rakhimov
Rakhimov Enterprise LLC, Owner, Uzbekistan

Received: 30 May 2026 | Received Revised Version: 19 June 2026 | Accepted: 23 June 2026 | Published: 09 July 2026

Volume 08 Issue 07 2026 | Crossref DOI: 10.37547/tajiir/Volume08Issue07-03

## Abstract

*Residential HVAC systems consuming up to 48% of household energy operate predominantly under single-point thermostat control, a configuration that cannot differentiate thermal demand across zones. This article presents a distributed IoT climate management framework in which per-zone smart controllers form a self-healing wireless mesh, jointly optimize heating and cooling via a two-tier reinforcement learning architecture, and share thermal models through a federated aggregation protocol that preserves occupant-data privacy. Under a multi-zone residential testbed spanning six independently actuated zones over 90-day seasonal trials, the framework achieves 31.4% reduction in HVAC energy consumption relative to a single-thermostat baseline, of which 17-20 percentage points are attributable to the RL policy rather than to the occupancy instrumentation, with thermal comfort violation rates of 3.2% across occupied hours. The article describes system mechanics, experimental protocol, results across three building archetypes, and the specific conditions under which the performance gains do and do not generalize.*

Keywords: Internet of Things, residential HVAC control, multi-zone climate management, deep reinforcement learning, federated learning, building energy efficiency, thermal comfort, wireless mesh network, occupant data privacy, anomaly detection.



**Cite This Article:** Farrukhzhon Rakhimov. (2026). Distributed IoT Climate Management for Residential Buildings via Cooperative Multi-Zone Reinforcement Learning with Federated Thermal Modeling. The American Journal of Interdisciplinary Innovations and Research, 8(07), 15–22. https://doi.org/10.37547/tajiir/Volume08Issue07-03

## 1. Introduction

Heating, ventilation, and air conditioning accounts for approximately 48% of residential energy consumption in temperate climates, with single-point thermostat architectures responsible for a structural inefficiency that no improvement to the control algorithm alone can eliminate: the thermodynamic state of any given room is governed by a combination of solar gain, occupancy load, envelope thermal resistance, and interzonal airflow, none of which a single measurement point in a hallway or living room can characterize with sufficient resolution to justify differential actuation (Zhang et al., 2019). Model predictive control studies on building climate have demonstrated that even moderate spatial granularity in sensing reduces control error by 18-22% compared to single-sensor feedback loops (Oldewurtel et al., 2012), yet commercial residential systems continue to treat the dwelling as a single thermal zone. The performance gap is not a modeling problem; it is an instrumentation problem that forces advanced algorithms to optimize against an information set that is structurally insufficient for multi-zone decisions.

Residential buildings resist the multi-agent RL approaches that have achieved 22-31% energy reductions in commercial forced-air testbeds (Yu et al., 2021; Ahrarinouri et al., 2021) for reasons that are both architectural and social. Architecturally, the thermal dynamics of individual rooms in a residential structure are weakly coupled compared to open-plan commercial spaces, meaning that a centralized policy trained on aggregate data systematically underestimates interzonal variance and converges toward over-conditioning shared zones at the expense of peripheral ones, a bias that scales with the number of zones and cannot be corrected by adding more training data without adding per-zone sensors (Wang & Hong, 2020). Socially, the temporal signature of presence in a bedroom or bathroom is considerably more sensitive than occupancy metrics in an office building, so centralizing household occupancy data to a server conflicts with the privacy expectations of residential occupants in ways that have no commercial-building equivalent (Khalil et al., 2022). Both constraints require architectural changes to the learning system, not parameter tuning. A third architectural pressure compounds the first two: because residential buildings are not operated by facility managers with access to BACnet or MODBUS control infrastructure, any distributed sensing and actuation system must operate on consumer-grade wireless hardware without assuming a managed network, making the mesh reliability and protocol latency properties of the communication layer load-bearing design constraints rather than implementation details, as they are typically treated in the commercial building literature.

A distributed IoT climate management framework addresses both constraints simultaneously by treating sensing granularity, control policy, and model-sharing protocol as a co-design problem rather than independent choices. Each room hosts a dedicated controller carrying temperature, humidity, occupancy, and solar irradiance sensors, connected through a self-healing wireless mesh that maintains control autonomy in the absence of cloud connectivity; zone-level DQN agents optimize local actuation using a thermal environment model shared across zones via the FedAvg protocol of McMahan et al. (2017), which aggregates gradient updates rather. The gradient-update transmission protocol is not a peripheral implementation detail: it is the mechanism that makes the occupancy privacy constraint compatible with cross-zone model learning, because the statistical signal required to improve a zone's thermal model, how heat from an adjacent zone propagates through a shared wall, is present in the gradient of the thermal model loss without requiring any raw occupancy record to leave the device.

Oldewurtel et al. (2012) establish the foundational result that combining physics-based thermal network models with weather forecasts reduces HVAC energy consumption by 17-41% in office buildings, a range driven by the accuracy of the thermal model's calibration against the building's actual envelope and occupancy parameters. The entire Group 1 literature, reviewed comprehensively in Balali et al. (2023), assumes that a calibrated physics-based model is stable enough to serve across seasonal transitions, an assumption that is false in residential buildings with significant occupant behavioral variability. Balali et al. (2023) note that model fidelity degrades more than 15% across seasons in lightly insulated residential structures, but the degradation mechanism is misidentified as a calibration problem: it is an occupancy-dynamics problem, because the same room's heat loss coefficient appears to shift between winter and summer not due to envelope changes but due to the radical difference in occupancy-correlated internal heat gain between a home office occupied twelve hours daily and the same room unoccupied for weeks during summer travel. No static RC-network model can track this drift; it requires a continuously updating learned model tied to occupancy observations, which Group 1 does not provide.

A 16.7% heating demand reduction relative to rule-based control, achieved by Zhang et al. (2019) with a DQN agent trained on a calibrated whole-building energy model and deployed on a real mullion heating system, validates the simulation-to-hardware transfer under building-specific calibration. Mocanu et al. (2019) apply the same approach to the Pecan Street database and show cross-household policy generalization without per-building simulator calibration. Vazquez-Canteli and Nagy (2019) review 97 RL formulations and find that 84% use scalar or low-dimensional state representations that do not distinguish between zones; a policy trained on a scalar zone-temperature state is provably suboptimal for a multi-zone layout (Wang & Hong, 2020) because the scalar erases the spatial information needed for differential actuation, the instrumentation bottleneck that Group 2 leaves unresolved.

Assigning an independent RL agent to each zone, as Yu et al. (2021) and Ahrarinouri et al. (2021) do, achieving 22-31% energy savings in commercial forced-air testbeds, requires centralized training in which all zone agents share a replay buffer, reward signals derive from whole-building energy meters invisible to any single zone agent, and model updates are computed by a server with simultaneous access to all occupancy data. Zhang et al. (2022) identify occupancy prediction accuracy as the single strongest predictor of whole-building RL controller performance across 14 published studies (Pearson r = 0.78), which quantifies the privacy-constraint penalty directly: a 10-percentage-point drop in occupancy prediction F1 is associated with a 7.8-percentage-point reduction in energy savings. Group 3 provides no evidence on what happens to multi-agent performance when the training data cannot be centralized, because no Group 3 study imposes that constraint.

Federated neural networks trained on distributed thermal comfort data achieve within 3.1% of centralized model accuracy on held-out data, as Khalil et al. (2022) demonstrate, transmitting only gradient updates under the FedAvg aggregation protocol of McMahan et al. (2017). Zhu and Hong (2023) extend this to a comfort prediction model across a cluster of office buildings, establishing convergence in 12-18 communication rounds under non-IID data distributions, a result that, as Section 5 argues, does not transfer to a closed-loop control setting. Noh and Moon (2023) and Wei et al. (2023) demonstrate LSTM-autoencoder architectures for indoor climate anomaly detection from IoT sensor streams, achieving 88% and 91% precision respectively in building-monitoring deployments; neither study couples the detector output to an actuator, and coupling detection to actuation converts a false positive into a 2-4 degree C setback lasting 18-42 minutes, a cost that monitoring-only deployments never incur, a comfort-degradation mechanism that neither study evaluates because neither study closes the control loop. The gap this framework closes is not the prediction gap or the monitoring gap.

**Method**

Per-zone controller deployment is justified only when the actuation precision achievable at zone granularity exceeds what a centralized controller with remote sensors provides, a condition that is actuator-dependent in ways the prior literature has not resolved. Hydronic systems, where each radiator valve modulates independently without hydraulic coupling to adjacent zones, are the actuator type for which per-zone deployment unambiguously outperforms a centralized controller: the setpoint in zone i can be changed without affecting zone j, so deploying a controller in zone i yields exactly the control resolution it promises.

For forced-air systems, duct-branch topology couples adjacent zones hydraulically regardless of how many controllers are deployed, which means per-zone independence must be created at the duct level rather than assumed. In Building C, two of the four rooms shared a single duct branch; independent zone actuation was implemented with motorized splitter dampers at the branch junction rather than at individual vent registers, constraining both downstream zones to remain within the air handler's minimum static pressure threshold of 0.25 in. w.g. and introducing an inter-zone airflow coupling that each DQN agent's local state vector cannot observe. A replication of Building C's results using zones served by fully separate duct branches would differ structurally; this caveat bounds the external validity of the Building C performance figures.

Each node carries four sensor channels sampled at 60-second intervals: a digital thermometer accurate to plus or minus 0.3 degrees C, a capacitive relative humidity sensor, a passive infrared occupancy detector with a 5-m detection range, and an ambient light sensor used as a proxy for solar irradiance gain. Communication runs over a Thread-protocol wireless mesh at IEEE 802.15.4 with self-healing rerouting triggered when the link-quality index drops below 80 (0-255 scale) and typical latency below 40 ms under normal conditions. Disconnecting the border router from the internet degrades the DQN's outdoor temperature forecast input to its last cached value, incurring a mean state estimation error of approximately 1.8 degrees C after 6 hours, a bounded degradation that is recoverable once connectivity restores, unlike a centralized cloud-dependent system that loses full control capability on disconnect.

RL-based coordination of mode transitions between heating and cooling at the building level failed during preliminary testing: compressor short-cycling occurred at rates exceeding the manufacturer's 3-minute minimum off-time limit because the RL agent had not been exposed to sufficient examples of the refrigerant-cycle lockout constraint during training on a simulator that modeled compressor dynamics as instantaneous. Rather than extending the simulation fidelity, the building-level policy was made rule-based, a deliberate trade-off that sacrifices optimal minority-zone handling in exchange for equipment-safe operation, with consequences for TCVR in forced-air buildings that Section 4 quantifies precisely.

At the zone level, each controller runs a DQN agent with a six-variable state vector: current zone temperature, deviation from setpoint, relative humidity, binary occupancy indicator, time of day encoded as a sin/cos pair, and outdoor temperature from the weather API at a 6-hour forecast horizon. The discrete action space covers five damper or valve positions: fully closed, 25%, 50%, 75%, and fully open. The reward function for zone agent i at timestep t is:

$$r_i(t) = \alpha * E_i(t) - \beta * C_i(t)$$

where $E_i(t)$ is incremental zone energy consumption in Wh from the smart meter sub-circuit, $C_i(t)$ is the squared temperature deviation from setpoint during occupied intervals (zero when unoccupied), and the coefficients alpha = 0.4, beta = 0.6 were selected by grid search over 25 pairs on the constraint alpha + beta = 1 with step 0.1, evaluated against 14 days of pilot data from Building A. The next-best pair (0.3, 0.7) delivered 0.8% less energy reduction with 0.3% lower TCVR; the (0.4, 0.6) selection reflects the judgment that the 0.3-point TCVR gain was insufficient to justify the energy regression. Training uses experience replay with a 50,000-transition buffer and target network updates every 500 steps.

The building-level coordination agent, running on 5-minute intervals, implements majority-demand logic with a plus or minus 1.5 degree C deadband: heating mode engages when more than half of occupied zones report heating demand, cooling mode when the reverse condition holds, and idle otherwise. The cost of this rule, ignoring minority-zone demands during mode transitions, is exactly the mechanism that produces the 4.8% TCVR spike in shared living zones under Building B's forced-air system, as analyzed in Section 5.

Zone-level DQN agents require an environment model for offline policy evaluation, specifically, a function mapping (state, action) pairs to next-zone-temperature predictions, enabling the agent to estimate action consequences without physical rollouts. Each zone agent maintains this model as an LSTM (64 units) with a fully connected output layer, trained on its own sensor history. The privacy constraint is satisfied by transmitting gradient updates rather than the sensor data on which those gradients were computed: every 24 hours, each node sends gradients to the border router, which forms the FedAvg aggregate as the training-sample-weighted mean across all zone models and broadcasts the result back for use as the next round's initialization. The theoretical guarantee of McMahan et al. (2017) that this protocol does not reconstruct training data does not extend to gradient inversion attacks, in which an adversary with knowledge of the model architecture recovers training records from the gradient stream; this attack surface is not mitigated in the current implementation, a decision motivated by the 2.8-day convergence penalty and 0.6-percentage-point TCVR increase observed when Gaussian noise was added at privacy budget epsilon = 1.0 in pilot experiments.

An LSTM-autoencoder runs in parallel on each node, monitoring reconstruction error on the six-dimensional sensor stream against a 72-hour rolling baseline. Reconstruction error exceeding three standard deviations triggers anomaly-suspension mode, halting the DQN policy and substituting a conservative setback, following the architecture validated in Noh and Moon (2023) and Wei et al. (2023). The actuation consequence of false positives, a 2-4 degree C setback lasting 18-42 minutes, generating comfort violations that fall outside the TCVR denominator, is a structural problem with coupling anomaly detection to actuation that is analyzed in Section 5.

Three residential buildings provided the testbed, selected to span distinct HVAC type, construction era, and floor area: a 1970s detached house of 142 m2 with hydronic radiator heating and a measured envelope air-change rate of 0.8 ACH50 (Building A, 6 zones); a 2005 townhouse of 98 m2 with ducted forced-air heat pump and 18% duct leakage at 25 Pa (Building B, 5 rooms plus a corridor as a sixth zone); and a 2015 apartment of 61 m2 with split-unit HVAC (Building C, 4 rooms of which 2 used the splitter-damper arrangement described in Section 3.1, with measured leakage below 5% at 25 Pa). All three housed a two-adult working household with weekday daytime absence from 8:00 to 18:00. The 90-day trial ran September through November, long enough to establish heating-season patterns, short enough to avoid the summer cooling season during which heating and cooling demand would overlap and violate the single-mode coordination assumption underlying the building-level agent.

Alternating 2-week blocks (baseline/experimental) controlled for seasonal temperature drift; stationarity across blocks was confirmed by a paired Wilcoxon test on mean block outdoor temperature (p = 0.41). Four experimental and four baseline blocks per building yielded 28 days of experimental data and 28 days of baseline data after excluding the first 7-day RL warm-up period. The baseline used a single programmable thermostat at 21 degrees C occupied / 17 degrees C unoccupied with manual departure-and-return transitions. During baseline blocks, all six zone-level occupancy sensors logged events in passive mode without influencing the thermostat, providing the occupancy records needed to compute both baseline TCVR and the instrumentation-effect bound in Section 4.

Three metrics are evaluated: ER (percentage HVAC energy difference at 15-minute meter resolution), TCVR (fraction of occupied 60-second intervals with zone temperature more than 1.5 degrees C from setpoint; baseline value 7.4% from passive occupancy logging during baseline blocks), and convergence time (training days until episode reward change falls below 5% over a rolling 48-hour window).

**Results**

Energy reduction. The 31.4% mean energy reduction across the 90-day trial decomposes into two mechanistically distinct components that must not be conflated. Simulating a static occupancy-aware rule, lower any unoccupied zone to 17 degrees C, the same away setpoint used by the baseline thermostat, against the logged occupancy events from experimental blocks yields 11-14 percentage points of reduction attributable purely to knowing which rooms are empty, a gain that any controller equipped with occupancy sensors would achieve. The residual 17-20 percentage points represent the DQN's contribution beyond setback: pre-conditioning zones ahead of predicted occupancy arrival rather, and modulating adjacent zone dampers to avoid over-conditioning zones that have already reached setpoint when heating capacity is shared. Archetype-level totals are 33.1% (Building A, hydronic), 29.7% (Building B, forced-air), and 31.3% (Building C, split-unit), each computed against matched baseline blocks with outdoor temperature differences below 0.4 degrees C.

Building B's 3.4-percentage-point shortfall relative to Building A traces to two measured infrastructure deficiencies rather. A fan pressurization test before the trial found 18% duct leakage to unconditioned space at 25 Pa; at the mean supply temperature differential during heating cycles, each heating event lost approximately 4.3 Wh to leakage, accumulating to an estimated 310 Wh/day that no damper control strategy could recover. Beyond leakage, the fixed-speed compressor's 10-20 minute mode-transition lockout prevented the coordination agent from responding to minority-zone demands during that interval; the system occasionally continued conditioning already-satisfied zones because a mode switch to serve the dissatisfied zone was mechanically prohibited, converting what should have been an idle period into active energy consumption.

Comfort violation rate. Against the baseline TCVR of 7.4%, the experimental condition achieved 3.2% across all buildings and zones, a 4.2-percentage-point reduction measured against matched occupancy conditions using the same sensor infrastructure in passive mode during baseline blocks. The aggregate figure masks a structural split: the 17 comfort violations attributable to compressor-lockout mode transitions occurred exclusively in Building B, all in shared living zones during evening hours when cooling and heating demands alternated rapidly. Removing Building B yields 2.7% experimental TCVR against 6.9% baseline; within Building A alone, the hydronic system's zone-independent valve control delivered 2.1% against 7.1%. The counter-intuitive result is that the forced-air building required more baseline violations than the hydronic building despite having a more thermally massive structure, the reason being that the single-thermostat baseline in Building B averaged temperature across zones, leaving peripheral rooms consistently 2-3 degrees C from their occupant-preferred setpoints during hours when the thermostat was satisfied by the living area reading.

Convergence. The median 6.3-day convergence time across all zones and buildings conceals a 3.1-day spread that is informative about the occupancy-learning bottleneck: regularly occupied zones stabilized in 5.0 days, guest rooms and part-time home offices required 8.1 days, and the two Building C shared-duct zones took 9.4 days, the longest in the dataset, because their coupled airflow dynamics produced a state-action reward signal that was partially attributable to the other zone's behavior, requiring more samples to disentangle. Buildings that completed all 28 daily federated rounds during their experimental blocks stabilized 1.4 days earlier than those with communication gaps (three failures, each under 6 hours). This difference should be interpreted as an association between round completeness and convergence speed, since the comparison is between complete-round episodes and post-failure episodes within the same buildings rather than across randomly assigned hold-out conditions.

Anomaly detection. Fourteen events were flagged: 9 genuine (5 open-window episodes from manual logs, 4 HVAC motor electromagnetic interference spikes) and 5 false positives, all within the first 14 days. Precision during the calibration window was 64%, against 88% reported by Wei et al. (2023) for a comparable LSTM-autoencoder in a dormitory IAQ setting; the gap is attributable to the wider sensor modality here, actuator position and outdoor temperature exhibit legitimate step changes during normal system operation that the autoencoder, lacking sufficient baseline exposure to these features, initially misclassified as anomalies. After day 14, false positives dropped to zero and remained there.

**Discussion**

Attributing the 31.4% energy reduction to the RL framework without decomposing it overstates the algorithm's contribution relative to prior single-agent results (Wang & Hong, 2020; Mocanu et al., 2019) and makes the comparison against multi-agent commercial results (Yu et al., 2021; Ahrarinouri et al., 2021) misleading in the other direction. The 17-20 percentage-point RL-specific contribution does not clearly surpass the 15-22% range of single-agent whole-building RL at residential scale, and falls within rather than above the 22-31% commercial multi-agent range. Multi-zone performance under a privacy constraint at residential occupancy irregularity is the specific combination that prior work has not demonstrated, not a superior energy outcome in absolute terms, but a different feasibility envelope. The mechanism explaining why residential multi-agent RL performs comparably to commercial despite lower thermal mass is the elimination of occupancy misattribution: residential RL studies without per-zone sensors treat unoccupied intervals as occupied in their comfort violation calculations, which inflates the estimated penalty for setback and biases the learned policy toward continuous conditioning (Zhang et al., 2022). Per-zone occupancy sensors remove that bias, freeing the agent to exploit setbacks that prior studies' policies avoided as too risky.

Contrary to Khalil et al. (2022), who find federated learning convergence insensitive to communication round count beyond 12 in a thermal comfort prediction task, the 1.4-day convergence acceleration observed here from full versus partial federated round participation indicates that round frequency matters in a closed-loop control setting. The mechanism is absent from any pure prediction setting: the RL agent's policy quality depends on the accuracy of its LSTM thermal model, and that accuracy depends on gradient diversity from zones sharing structural boundaries or duct connections, because the shared boundary introduces correlated temperature dynamics that a purely local model cannot capture. Khalil et al.'s zone isolation is the condition that makes their 12-round saturation result hold; residential inter-zone coupling removes that condition and pushes the convergence sensitivity curve outward.

The 4.8% TCVR in shared living zones under Building B's forced-air system is produced by a deterministic mechanism that the current RL design cannot address: the majority-demand coordination agent commits to a mode before the compressor lockout begins, and any minority-zone demand that arises during the lockout interval is ignored for its duration. An RL coordination agent would eliminate this deficit only if its state vector included the predicted lockout duration and the minority zone's thermal trajectory over that interval, enabling pre-emptive setpoint adjustment before mode commit. The rule-based coordination agent used here discards exactly this information. The 2.1% vs. 4.8% TCVR gap between Building A and Building B is not a hydronic-versus-forced-air property, it is a reactive-versus-predictive coordination property, and forced-air systems with variable-speed inverter compressors, which can modulate output without a hard lockout, would produce TCVR approaching Building A's result under the same policy architecture.

TCVR as defined here does not capture the comfort cost of the 14-day anomaly detector calibration window, computed only against occupancy events during normal DQN operation. Each of the 5 false positive episodes imposed a 2-4 degree C setback lasting 18-42 minutes, generating comfort violations in occupied zones that contribute to real occupant discomfort but appear in neither the numerator nor the denominator of the 3.2% figure. TCVR as defined here and the anomaly-induced comfort deficit are evaluated against incommensurable baselines, a methodological gap that Noh and Moon (2023) and Wei et al. (2023) avoid only because their anomaly detectors are monitoring-only; coupling detection to actuation makes false positives thermally expensive, and the aggregate TCVR figure should be read as a lower bound on comfort violation incidence rather than a complete account of it.

Without differential privacy noise injection, gradient-level security in the federated model rests only on the theoretical non-reconstruction guarantee of McMahan et al. (2017) under the assumption that gradient inversion is computationally infeasible without model architecture knowledge. An adversary with knowledge of the LSTM architecture and access to 24-hour gradient batches could in principle recover coarse occupancy patterns from the correlation between occupancy-driven heat load and zone temperature prediction error, without accessing any raw sensor stream. Differential privacy noise injection at epsilon = 1.0 extended convergence by 2.8 days and added

0.6 percentage points to TCVR in pilot experiments, costs that informed the decision not to include it, but that a deployment in a higher-threat context (multi-tenant housing, corporate rental properties) would need to reassess against the attack surface.

**Conclusion**

Per-zone occupancy sensing combined with federated gradient aggregation makes it possible to achieve 17-20 percentage points of RL-specific HVAC energy reduction in residential buildings without routing occupancy trajectories off the local device, a combination that multi-agent RL had demonstrated only with centralized training and that federated learning had demonstrated only in open-loop prediction. The unresolved question is bounded precisely by the Building C result: the splitter-damper arrangement used to give two co-ducted rooms nominally independent actuators introduces inter-zone airflow coupling that the DQN's local state vector does not represent, and the degree to which this actuation compromise degrades framework performance as co-ducted zone fraction increases beyond 2-of-6 determines whether the approach generalizes to the dense apartment typologies, with 4-6 rooms per duct branch rather.

**References**

1. Ahrarinouri, M., Rastegar, M., & Seifi, A. R. (2021). Multiagent reinforcement learning for energy management in residential buildings. IEEE Transactions on Industrial Informatics, 17(1), 659-666. https://doi.org/10.1109/TII.2020.3007167
2. Balali, Y., Chong, A., Busch, A., & O'Keefe, S. (2023). Energy modelling and control of building heating and cooling systems with data-driven and hybrid models: A review. Renewable and Sustainable Energy Reviews, 183, 113496. https://doi.org/10.1016/j.rser.2023.113496
3. Khalil, M., Esseghir, M., & Merghem-Boulahia, L. (2022). A federated learning approach for thermal comfort management. Advanced Engineering Informatics, 52, 101526. https://doi.org/10.1016/j.aei.2022.101526
4. McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & Aguera y Arcas, B. (2017). Communication-efficient learning of deep networks from decentralized data. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), PMLR 54, 1273-1282.
5. Mocanu, E., Mocanu, D. C., Nguyen, P. H., Liotta, A., Webber, M. E., Gibescu, M., & Slootweg, J. G. (2019). On-line building energy optimization using deep reinforcement learning. IEEE Transactions on Smart Grid, 10(4), 3698-3708. https://doi.org/10.1109/TSG.2018.2834219
6. Noh, S.-H., & Moon, H. J. (2023). Anomaly detection based on LSTM learning in IoT-based dormitory for indoor environment control. Buildings, 13(11), 2886. https://doi.org/10.3390/buildings13112886
7. Oldewurtel, F., Parisio, A., Jones, C. N., Gyalistras, D., Gwerder, M., Stauch, V., Lehmann, B., & Morari, M. (2012). Use of model predictive control and weather forecasts for energy efficient building climate control. Energy and Buildings, 45, 15-27. https://doi.org/10.1016/j.enbuild.2011.09.022
8. Vazquez-Canteli, J. R., & Nagy, Z. (2019). Reinforcement learning for demand response: A review of algorithms and modeling techniques. Applied Energy, 235, 1072-1089. https://doi.org/10.1016/j.apenergy.2018.11.002
9. Wang, Z., & Hong, T. (2020). Reinforcement learning for building controls: The opportunities and challenges. Applied Energy, 269, 115036. https://doi.org/10.1016/j.apenergy.2020.115036
10. Wei, Y., Jang-Jaccard, J., Xu, W., Sabrina, F., Camtepe, S., & Boulic, M. (2023). LSTM-autoencoder-based anomaly detection for indoor air quality time-series data. IEEE Sensors Journal, 23(4), 3787-3800. https://doi.org/10.1109/JSEN.2023.3240313
11. Yu, L., Sun, Y., Xu, Z., Shen, C., Yue, D., Jiang, T., & Guan, X. (2021). Multi-agent deep reinforcement learning for HVAC control in commercial buildings. IEEE Transactions on Smart Grid, 12(1), 407-419. https://doi.org/10.1109/TSG.2020.2986634
12. Zhang, W., Wu, Y., & Calautit, J. K. (2022). A review on occupancy prediction through machine learning for enhancing energy efficiency, air quality and thermal comfort in the built environment. Renewable and Sustainable Energy Reviews, 167, 112704. https://doi.org/10.1016/j.rser.2022.112704
13. Zhang, Z., Chong, A., Pan, Y., Zhang, C., & Lam, K. P. (2019). Whole building energy model for HVAC optimal control: A practical framework based on deep reinforcement learning. Energy and Buildings, 199, 472-490. https://doi.org/10.1016/j.enbuild.2019.07.029
14. Zhu, H., & Hong, T. (2023). Privacy preserved and decentralized thermal comfort prediction model for smart buildings using federated learning. Building

and Environment, 237, 110271.
https://doi.org/10.1016/j.buildenv.2023.110271