Journal of Science and Technology in Civil Engineering, HUCE, 2026, 20 (4): 21–34

DATA-DRIVEN DECISION SUPPORT FOR LOW-CARBON RETROFITS: BALANCING HEAT LOSS AND INDOOR ENVIRONMENTAL RISKS USING INTERPRETABLE MACHINE LEARNING

Nguyen Thi Thuy Hien[a,b,*], Ruoyu Jin[b]

[a]*Research group on Industry 4.0 in Transportation (I4T group), University of Transport Technology, 54 Trieu Khuc road, Thanh Liet ward, Hanoi, Vietnam*
[b]*Brunel University of London, Kingston Lane, Uxbridge, UB8 3PH, United Kingdom*

Article history:
Received 21/4/2026, Revised 04/6/2026, Accepted 22/6/2026, Available online 16/10/2026, Version of Record 25/12/2026

Abstract

Real estate decarbonization mandates necessitate improved building envelope thermal efficiency. These physical retrofits inadvertently restrict natural ventilation, increasing the risks of indoor dampness and poor air quality. Addressing this operational trade-off, this research proposes a quantitative framework integrating IoT sensor telemetry with static asset data. Formulated as a multi-label classification problem predicting three concurrent risks, the study evaluates a Random Forest model on a pilot portfolio of 81 residential units, achieving an average localized test Area Under the Curve (AUC) of 0.751. Spatial cross-validation yields a baseline AUC of 0.62, reflecting the physical complexity of generalizing predictive models to unseen architectural archetypes lacking historical telemetry. Applying the SHapley Additive exPlanations (SHAP) algorithm interprets model predictions, distinguishing between physical envelope degradation and occupant ventilation behaviors based on feature contributions. The framework maps these predictive probabilities into a prescriptive four-quadrant intervention matrix, designed to inform targeted capital and operational budget allocations. Longitudinal emission projections to 2040 quantify the environmental impact of delayed funding. Asset managers can use these outputs to balance energy efficiency targets with indoor environmental quality standards and justify accelerated green financing.

Keywords: low-carbon retrofit; IoT sensors; facility management; interpretable AI; trade-off optimization.

1. Introduction

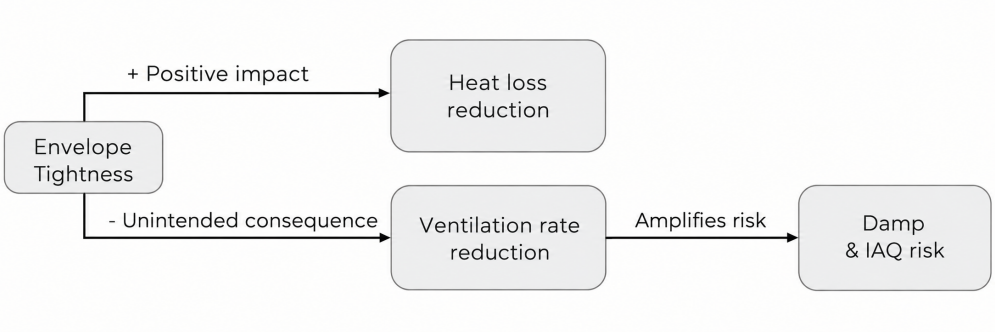
Existing building stocks generate 28% of global energy-related CO_2 emissions due to space heating, cooling, and ventilation requirements [1]. Reducing operational carbon in real estate portfolios requires proactive, data-driven retrofitting strategies [2]. Physical retrofits prioritize envelope thermal resistance and airtightness to lower space heating demands. Restricting natural air change rates establishes a direct trade-off with Indoor Environmental Quality (IEQ) [3] (Fig. 1(a)). Insufficient mechanical exhaust traps occupant-generated moisture and pollutants indoors [4]. Condensation on cold internal surfaces promotes structural dampness and mold growth [5]. Portfolio management requires a multi-objective approach to balance energy targets and health standards.

Multi-objective optimization (MOO) frameworks balance competing operational goals by evaluating energy consumption, thermal comfort, and capital costs [6]. These computational models assess

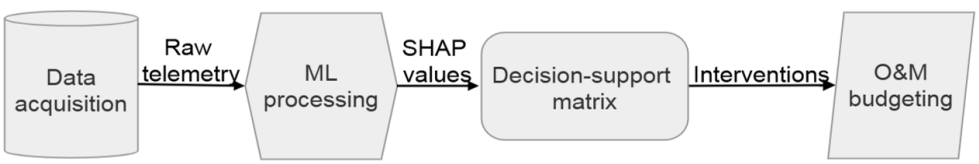
*Corresponding author. *E-mail address:* hienntt82@utt.edu.vn (Hien, N. T. T.)
https://doi.org/10.31814/stce.huce2026-20(4)-02 p-ISSN 1859-2996; e-ISSN 2734-9268
© 2026 The Author(s). Published by Hanoi University of Civil Engineering. Licensed under CC BY-NC-ND 4.0.

physical interventions ranging from insulation adjustments to HVAC upgrades [7]. Deterministic energy models assume static occupant behaviors and idealized material degradation rates. This reliance limits large-scale empirical validation [8]. Integrating continuous Internet of Things (IoT) telemetry with machine learning (ML) improves building asset management through data-driven energy forecasting and automated fault detection [9, 10]. Predictive ML frameworks isolate physical faults but rarely offer prescriptive engineering solutions [11]. The inability to translate raw probability scores into specific maintenance interventions restricts operational budget planning.

This research introduces a quantitative decision-support framework that advances building analytics from pure predictive fault detection into prescriptive facility management (Fig. 1(b)). The analytical pipeline processes raw environmental telemetry through machine learning algorithms to predict multiple risk targets simultaneously. The methodology applies SHapley Additive exPlanations (SHAP) values to extract localized feature importance, distinguishing physical degradation from behavioral patterns. These interpretable outputs directly inform a strategic decision matrix designed to guide physical interventions and prioritize operational budget allocations. The research scope encompasses the formulation of this predictive pipeline and its empirical validation using continuous sensor datasets from an existing residential portfolio, equipping asset managers with a pilot-scale screening tool to manage the trade-off.



(a) Decarbonization trade-off



(b) Decision-support framework

Figure 1. Research context and analytical framework

2. Materials and methodology

2.1. Data acquisition and preprocessing

The dataset integrates building metadata with high-frequency IoT telemetry (Table 1). The pilot portfolio comprises 81 social housing units located in a local authority in the UK. This region features a temperate maritime climate with high winter humidity, where large indoor-outdoor temperature gradients strictly dictate extended heating seasons and exacerbate condensation risks. Socioeconomically, these post-war estates (mean age: 58 years) are highly vulnerable to fuel poverty. Residents restrict heating and block natural ventilation pathways to reduce energy costs, amplifying

indoor moisture and CO_2 concentrations. This specific intersection of climate-driven dampness and constrained decarbonization budgets necessitates the proposed risk-balanced framework. Heat flux and condensation risks are calculated via indoor-outdoor temperature ΔT and moisture ΔHumidity gradients [12, 13]. Cubic spline interpolation reconstructs missing sensor readings to preserve thermodynamic dynamics [14]. This technique is restricted to signal gaps under four hours to maintain data consistency. Longer interruptions are excluded to prevent thermal inertia artifacts. A 7-day rolling window subsequently extracts maximum, minimum, and average values to capture weekly behavioral cycles [15]. Table 1 catalogs the raw telemetry variables. Feature selection isolated five core predictors, Building Age, Property Type, MaxCO_2, ΔT, and ΔHumidity, mitigating overfitting and computational latency during model training. All other rolling statistical indicators were excluded from the final dataset.

Table 1. Input feature catalog

Feature category	Feature name	Data type & Unit
Static metadata	Building Age	Continuous (Years)
	Property Type	Categorical (Nominal)
Thermodynamic drivers	Max/Min/Avg Temperature	Rolling 7D (°C)
	ΔT (Indoor-Outdoor gradient)	Time-series (°C)
Moisture indicators	Max/Min/Avg Humidity	Rolling 7D (%)
	ΔHumidity (Indoor-Outdoor gradient)	Time-series (%)
IAQ & ventilation	Max/Min/Avg CO_2	Rolling 7D (ppm)

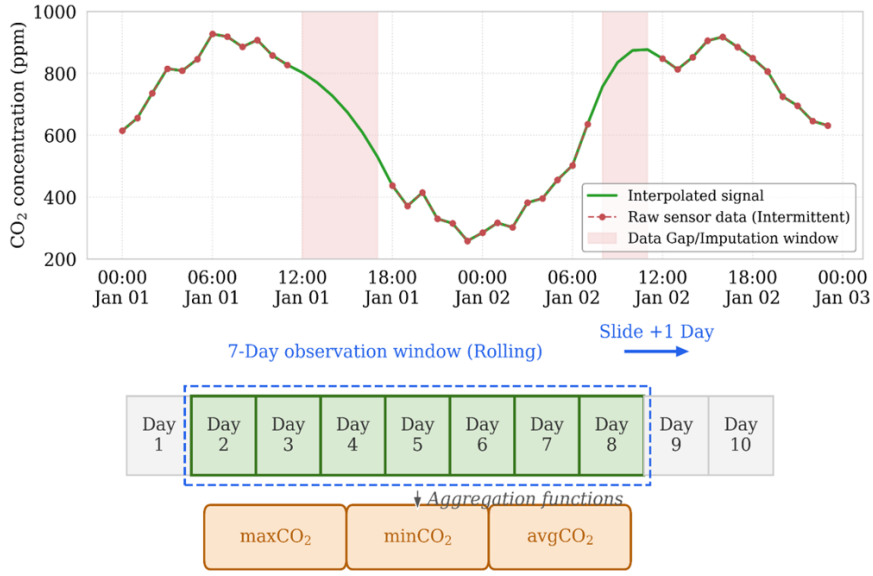


Figure 2. Data imputation and rolling aggregation pipeline

2.2. *Multi-label classification framework*

This study formulates predictive maintenance as a multi-label binary classification problem. For each building asset, an input feature vector $x_i \in \mathbb{R}^5$ represents the physical and environmental variables. The target output is a three-dimensional binary vector $y_i \in \{0, 1\}^3$, where components y_1, y_2,

and y_3 represent risks for Heat Loss, Dampness, and Indoor Air Quality (IAQ), respectively. Each risk component y_j is binarized using a portfolio-specific median threshold rather than absolute engineering, regulatory, or health standards (e.g., WHO indoor air quality guidelines). Applying absolute thresholds to aging post-war housing stocks classifies over 80% of assets as failing. Methodologically, this induces severe class imbalance, degrading the predictive performance and generalization of multi-label machine learning models. Practically, near-universal failure labeling obscures actionable insights for asset allocation. Establishing a relative benchmarking system ($y_j = 1$ for values exceeding the portfolio median, and $y_j = 0$ otherwise) mitigates this class imbalance while capturing relative operational vulnerability. This formulation enables asset managers to isolate the worst-performing 50% of the portfolio, prioritizing targeted capital and operational expenditure (CAPEX/OPEX) allocations under strict budget constraints. The model approximates the mapping function $f(x)$ using a Multi-Output Random Forest architecture (Fig. 3). Single-label methods isolate individual risks, whereas this multi-label approach captures correlations among simultaneous environmental failure modes [16]. The function $f(x)$ predicts the joint probability distribution of these risks within the defined input space X [17].

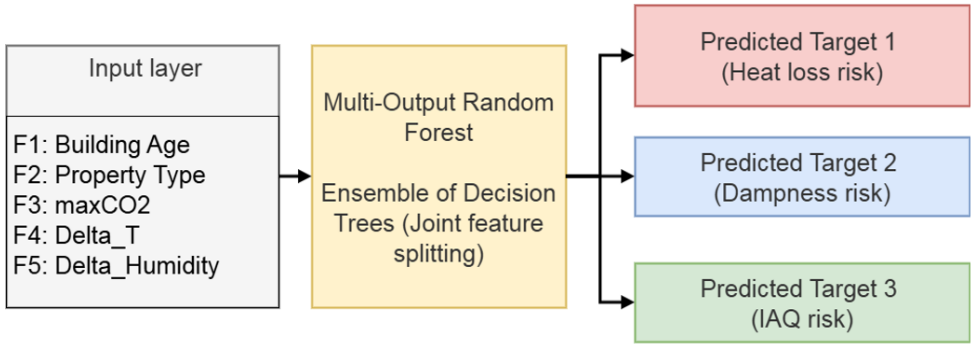


Figure 3. Proposed multi-task Machine Learning architecture

2.3. Model training and evaluation

Model evaluation requires rigorous partitioning to prevent data leakage. Building science datasets exhibit high spatial correlation within geographic clusters due to synchronized construction periods and shared microclimates. Random cross-validation introduces severe data leakage and overoptimistic performance estimates in panel-structured housing telemetry. This study employs a Group K-fold strategy ($K = 3$) based on project lot identifiers. Let L be the set of M unique lots; the partitioning ensures that any two folds L_a and L_b are disjoint ($L_a \cap L_b = \varnothing$). This spatial validation assesses generalizability by testing predictions on unseen housing developments [18]. Optimization maximizes the Area Under the Receiver Operating Characteristic Curve (ROC-AUC) via randomized search [19]. Performance assessment relies on Precision, Recall, F1-Score, and ROC-AUC derived from True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). Precision measures the accuracy of positive predictions:

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{3}$$

2.4. Model interpretability

Facility managers require interpretable results to justify maintenance interventions. This study applies SHAP to interpret the model predictions. Based on cooperative game theory, SHAP calculates the marginal contribution of each environmental feature to the predicted risk probability. For an original multi-label prediction model $f(x)$ and a local explanation model $g(z')$, the additive framework is formulated as:

$$g(z') = \phi_0 + \sum_{j=1}^{M} \phi_j z'_j \tag{4}$$

The binary coalition vector $z' \in \{0,1\}^M$ indicates the active status of M input features. The parameter ϕ_0 is the baseline expected model output, and ϕ_j is the Shapley value for feature j. The Shapley value ϕ_j evaluates the contribution of a feature across all possible combinations S, computed as:

$$\phi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} \left(f_x(S \cup \{j\}) - f_x(S)\right) \tag{5}$$

Feature contributions are quantified using the complete input set F, a subset $S \subseteq F \setminus \{j\}$, and the expected prediction $E[f(x)|x_S]$. This mechanism isolates anomaly triggers by decoupling structural degradation indicators (e.g., Building Age) from occupant behaviors (e.g., Max CO_2). This feature-level interpretability translates risk probabilities into targeted engineering interventions [20].

3. Results and discussion

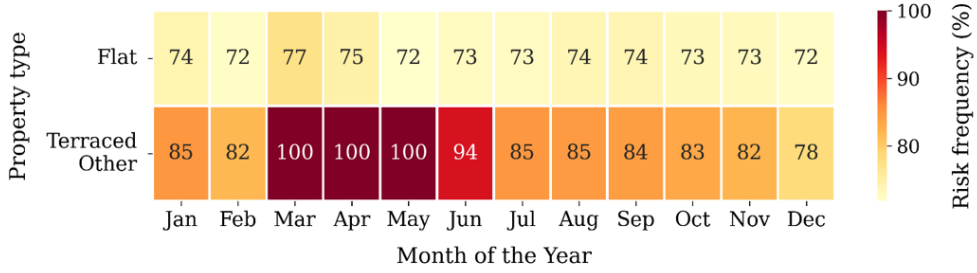
3.1. Portfolio characteristics and predictive performance

Continuous 7-day rolling window aggregation expanded the dataset from 81 physical assets to 808 temporal observations. Although this rolling extraction introduces temporal autocorrelation among sequential observations from the same unit, this dependence is controlled during model evaluation. As detailed in Section 2.3, Group K-Fold cross-validation, partitioned by project lot identifiers, ensures all temporal observations from a specific asset are assigned to a single fold. This structural separation prevents data leakage, ensuring accuracy metrics reflect the model's capacity to generalize to unseen assets rather than overfitting to dependent temporal sequences. Within the 81-unit pilot portfolio (Table 2), 'Terraced/Other' properties exhibit operational vulnerabilities. A mean temperature gradient of 4.8 ± 2.7 °C indicates thermal deficiencies and a dual-risk profile. The uniform building age (56.0 ± 0.0 years) reflects post-war batch-developed estates rather than a data anomaly. Despite the limited sample ($N = 24$), this subgroup provides representative evidence of structural patterns for that architectural era. Dynamic telemetry uncovers wide statistical dispersion in Max CO_2 concentrations, with an overall standard deviation of 1044 ppm (reaching 1133 ppm in flat properties). This variance confirms that occupant ventilation habits drive indoor conditions independent of envelope airtightness. Parallel to CO_2 variance, moisture telemetry reveals distinct condensation vulnerabilities across the portfolio. The dataset exhibits an overall average indoor-outdoor humidity gradient (ΔHumidity) of 31.4%, with maximum indoor humidity levels averaging 73.4%. These baseline metrics exceed the 60% threshold critical for mold proliferation. Notably, terraced properties demonstrate severe moisture retention (ΔHumidity: 40.2%; Max Humidity: 79.5%), aligning with their concentrated failure modes in Quadrant 1. This localized moisture accumulation directly substantiates the high dampness risk frequencies observed in Fig. 4 and justifies the necessity for active extraction interventions. Fig. 4(a) shows the temporal concentration of these vulnerabilities, showing sustained extreme risk frequencies across specific property types. Fig. 4(b) maps the spatial risk intensity across geographic

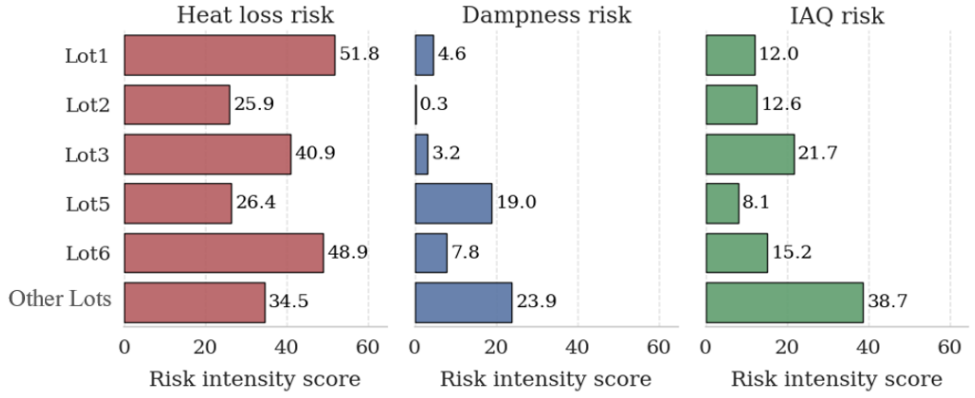
lots, confirming distinct localized failure modes. This pronounced spatiotemporal variance justifies the implementation of the spatial cross-validation approach.

Table 2. Descriptive statistics of the training portfolio

Property type	Sample size (N)	Building age (Mean ± SD)	Temperature gradient ΔT (Mean ± SD, °C)	Max CO_2 (Mean ± SD, ppm)	Heat loss risk (%)	IAQ risk (%)	ΔHumidity (Mean ± SD, %)	Max Humidity (Mean ± SD, %)
Flat	57	58.9 ± 17.0	5.7 ± 3.3	1151 ± 1133	51.6%	46.5%	30.3 ± 18.2	72.9 ± 18.8
Terraced/Other	24	56.0 ± 0.0	4.8 ± 2.7	1110 ± 815	49.7%	46.7%	40.2 ± 19.7	79.5 ± 18.3
Total/Average	**81**	**58.0 ± 14.9**	**5.4 ± 3.1**	**1139 ± 1044**	**51.0%**	**46.6%**	**31.4 ± 18.6**	**73.4 ± 19.2**



(a) Temporal risk frequency



(b) Spatial risk intensity

Figure 4. Spatio-temporal risk distributions

Table 3 details the spatial generalization performance. K-Nearest Neighbors (KNN) achieved a Spatial CV AUC of 0.6254, slightly higher than the Multi-Output Random Forest (RF) score of 0.6221. RF was selected as the primary architecture to ensure TreeSHAP compatibility for localized feature interpretation. Methodologically, decision tree ensembles process mixed data types, including nominal property classifications and continuous environmental telemetry, without requiring feature scaling. RF captures non-linear thermodynamic interactions, whereas distance-based algorithms like KNN are sensitive to scale variations in high-frequency sensor streams. The modest Spatial CV AUC (0.6221) reflects the physical complexity of generalizing predictive models to unseen architectural archetypes lacking historical telemetry. This metric differs from the localized test AUC (0.751)

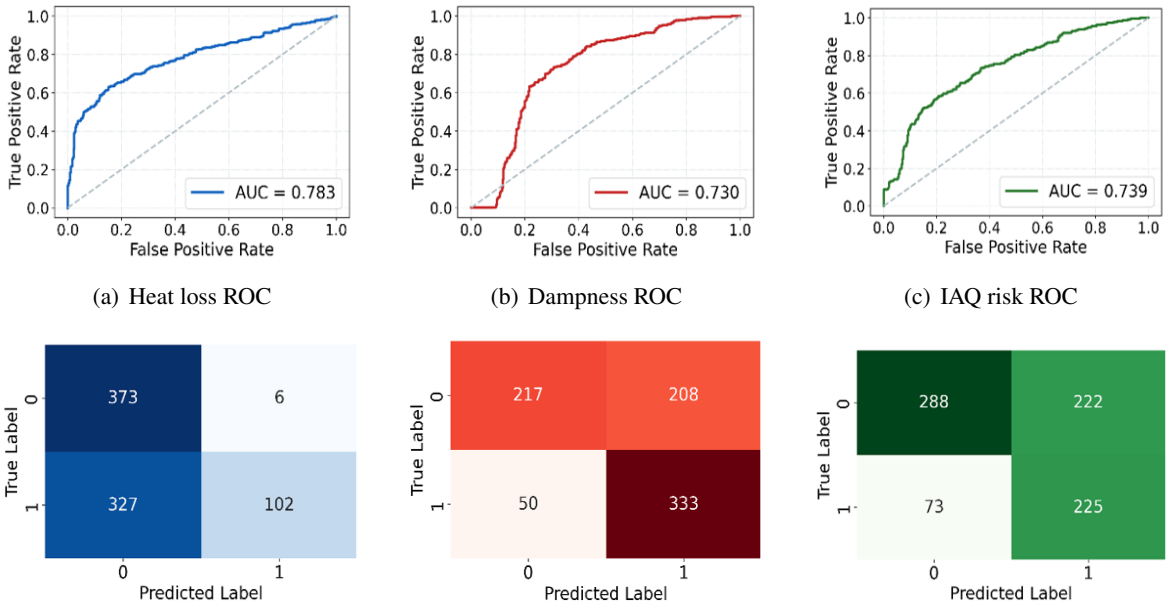
achieved under local calibration within known project lots. This framework functions as a pilot-scale screening tool to prioritize portfolio interventions rather than a definitive diagnostic instrument.

Table 3. Performance comparison and optimized hyperparameters of ML models

Machine Learning model	Spatial CV AUC	Optimized hyperparameters
K-Nearest Neighbors (KNN)	0.6254	weights: 'uniform', n_neighbors: 20
Random Forest (RF)	0.6221	n_estimators: 200, max_depth: 10, min_samples_leaf: 4, class_weight: 'balanced_subsample'
XGBoost	0.6155	n_estimators: 100, max_depth: 3, learning_rate: 0.05, colsample_bytree: 0.7
LightGBM (HistGBM)	0.6152	max_iter: 100, max_depth: 3, learning_rate: 0.01, l2_regularization: 0.0
Logistic Regression (Baseline)	N/A	Default parameters

Note: Spatial CV AUC represents the model's spatial generalization capability, evaluated using Group K-Fold Cross-Validation (split by project lots) to prevent data leakage between different building archetypes.

Random Forest (RF) serves as the primary model to enable TreeSHAP feature interpretation, a capability absent in distance-based algorithms. Local calibration on the pilot portfolio yields the final accuracy metrics (Fig. 5).



(a) Heat loss ROC (b) Dampness ROC (c) IAQ risk ROC

(d) Heat loss CM (e) Dampness CM (f) IAQ risk CM

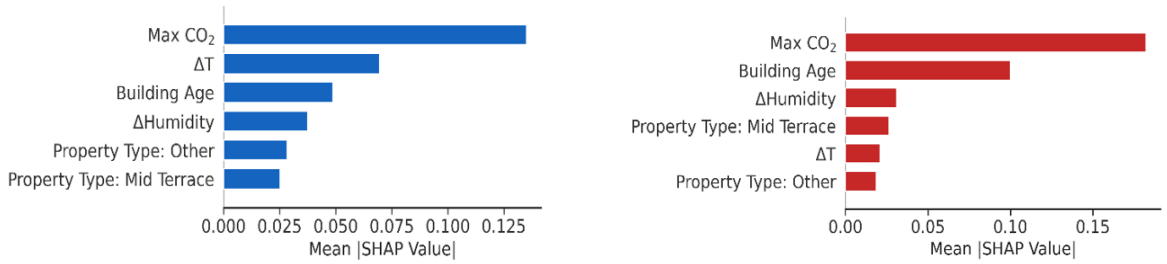
Figure 5. Performance metrics across risk domains

The Multi-Output Random Forest yields ROC-AUC scores of 0.783 for Heat Loss, 0.730 for Dampness, and 0.739 for IAQ risks. Deriving metrics from the confusion matrices, the Heat Loss classifier registers a 0.94 precision with only 6 false positive predictions (F1-Score: 0.37). The Dampness and IAQ models achieve F1-Scores of 0.72 and 0.60, respectively.

This performance distribution indicates an operational trade-off. Optimizing the Heat Loss classifier for precision limits the misallocation of capital expenditures (CAPEX) on unnecessary retrofits. The Dampness and IAQ models prioritize high recall rates, identifying 333 and 225 true positive cases, respectively. Capturing these instances supports the mitigation of respiratory health hazards.

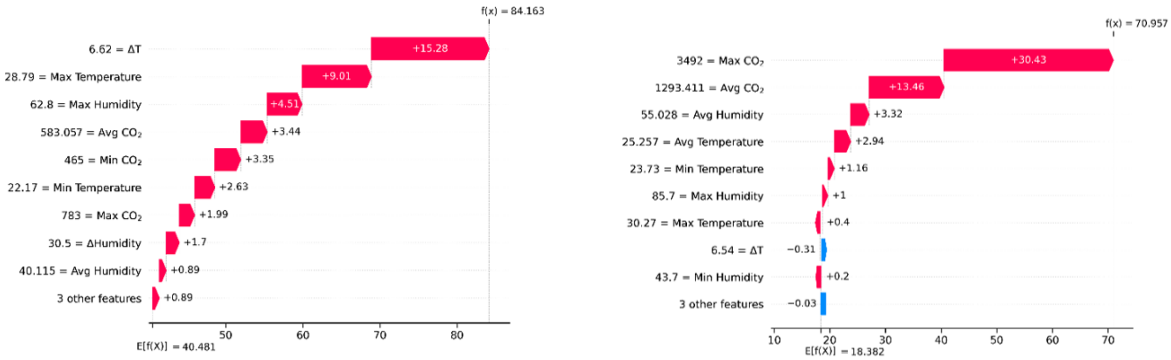
3.2. Feature importance and model interpretability

The SHAP algorithm facilitates both global and local model interpretability by quantifying individual feature contributions. At the macro level (Fig. 6), the dynamic behavioral metric Max CO_2 emerges as the leading predictive feature for both heat loss and IAQ risk classifications. The lower comparative importance of static structural parameters suggests that these environmental vulnerabilities are driven more by occupant habits than by material aging alone. Alongside these behavioral factors, the temperature gradient ΔT serves as the core physical driver for heat loss.



(a) Heat loss (b) IAQ

Figure 6. Global SHAP feature importance for risk prediction



(a) Heat loss (b) IAQ

Figure 7. Global SHAP feature importance for risk prediction

High CO_2 variance serves as a thermodynamic proxy for intermittent ventilation. Absent direct window-contact sensors or occupancy logs, rapid CO_2 decay events coupled with simultaneous internal temperature drops indicate occupant-driven heat loss. Fig. 7 isolates environmental drivers for specific units. In apartment B (Fig. 7(a)), feature contributions attribute an 84.16% heat loss probability primarily to the temperature gradient (+15.28) and indoor temperature (+9.01). Profiling apartment B, Fig. 7(b) links a 70.95% IAQ risk predominantly to CO_2 metrics (+30.43). This statistical attribution identifies suspected envelope vulnerabilities, such as localized thermal bridging, but does not constitute definitive physical proof. The dataset lacks physical validation records, such as infrared thermography. These local SHAP values therefore function strictly as diagnostic indicators.

The algorithmic outputs prioritize and direct targeted physical building audits, serving as a precursor to manual field inspections.

3.3. Decision-support framework and intervention strategy

Translating predictive probabilities into an operational strategy, the Energy-IEQ matrix quantifies the trade-off between thermal tightness and indoor environmental risks. A baseline 0.5 threshold segments the building portfolio into four intervention quadrants. Functioning as a tunable parameter rather than a rigid constraint, this boundary enables facility managers to dynamically scale intervention scopes. Raising specific thresholds narrows the targeted asset pool, aligning maintenance actions with constrained capital expenditure (CAPEX) budgets and prioritizing properties exhibiting advanced physical degradation.

Mapping algorithmic risk thresholds against localized feature contributions, the intervention strategy matrix (Table 4) provides a decision-support framework. For dual-risk profiles, Quadrant 1 indicates a need for combined External Wall Insulation (EWI) and Mechanical Ventilation with Heat Recovery (MVHR). From a psychrometric perspective, the regional winter climate exhibits high outdoor relative humidity alongside minimal absolute moisture. The MVHR heat exchanger draws and heats this cold outdoor air, substantially decreasing the relative humidity. The mechanism supplies continuous dry air with high moisture-absorption capacity, mitigating indoor condensation risks and preserving the thermal integrity of the retrofit. Sensible-only mechanical ventilation with heat recovery (MVHR) systems supersede enthalpy-based architectures within the analyzed maritime climate. Enthalpy cores transfer exhaust moisture into the supply stream, an exchange mechanism that amplifies condensation risks across high-occupancy residential retrofits. The sensible-only MVHR methodology extracts moisture-laden indoor air and applies sensible heating to the introduced low-absolute-moisture outdoor air, a thermal conditioning process that reduces supply relative humidity. Continuous low-humidity air supply purges internal latent loads in adherence to UK Building Regulations Part F [21] alongside CIBSE Guide A [22]. Quadrant 2 allocates non-structural operational expenditures (OPEX), such as active extraction ventilation, to highly sealed units. Addressing specific thermal deficiencies, Quadrant 4 focuses on isolated envelope upgrades, whereas Quadrant 3 establishes a safe operational baseline involving routine IoT monitoring. Within the resultant scatter distribution (Fig. 8), 'Terraced/Other' properties cluster in Q1, signaling a reliance on structural capital. 'Flats' align predominantly along the single-risk axes of Q2 and Q4.

Table 4. Translation from ML probability to technical interventions

Quadrant & resource	Diagnostic criteria	Engineering action
Q1: Deep retrofit (Priority 1, CAPEX)	High building age + High CO_2	Synchronized EWI and MVHR
Q2: Vent upgrade (Priority 2, OPEX)	High CO_2 + High humidity	decentralized Mechanical Extract Ventilation (dMEV) installation; behavioral nudges
Q4: Insulation focus (Priority 3, CAPEX)	High building age + Low temperature	Targeted CWI or Low-E glazing
Q3: Standard O&M (Priority 4)	Parameters within safety baselines	Routine IoT monitoring; no physical action

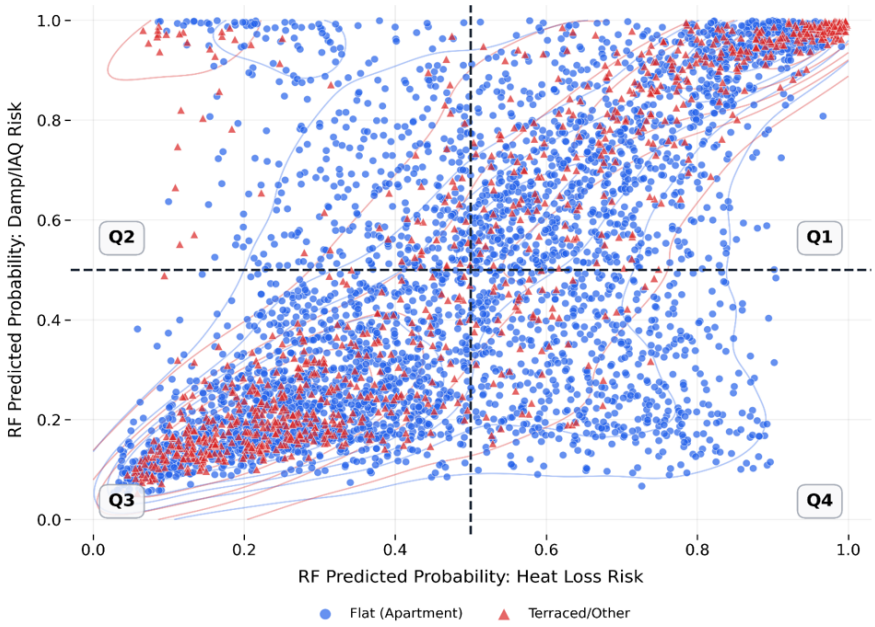


Figure 8. Intervention strategy matrix applying a 0.5 risk threshold. Each data point denotes a 7-day rolling-window observation (N = 808), distributing the portfolio across Q1 (High Heat Loss and IAQ Risk), Q2 (High IAQ Risk), Q3 (Safe Zone), and Q4 (High Heat Loss)

3.4. Financial and environmental impact estimation

Integrating the prescriptive matrix with financial planning quantifies portfolio resource allocation (Fig. 9). Quadrant 1 interventions for the 23 highest-risk units via whole-house retrofits incur an estimated £26,300 per unit [23], requiring a £604,900 baseline capital expenditure (CAPEX). Localized structural repairs in Quadrant 4 target defects at £8,000 per property [24, 25]. Regarding operational expenditure (OPEX), Quadrant 2 ventilation upgrades average £1,000 per unit. While Quadrant 3 assets require no physical modification, a 'Standard O&M' allocation of £100 per unit covers IoT recalibration, data hosting, and platform licensing [26]. Including this £3,500 maintenance provision, the cumulative portfolio intervention cost reaches £729,400. A ±20–40% market fluctuation band provides a contingency buffer against material inflation and supply chain disruptions. Linking currency allocations to algorithmically verified risks grounds the framework in applied asset management, establishing a quantitative basis for green financing.

Table 5 demonstrates that retrofitting 46 critical units decreases annual energy use by 108,500 kW h and drops emissions by 19,204.5 kgCO_2e. Under static pricing, these interventions carry lengthy payback periods of 53.7 years for deep retrofits and 28.6 years for envelope insulation. Incorporating energy inflation and carbon taxes into financial models can shorten these timelines. At present, baseline savings alone fail to cover upfront capital costs. Bridging this financial gap necessitates green financing or asset appreciation [27]. In Quadrant 2, ventilation upgrades add to the electrical load but remain essential for controlling respiratory hazards [28]. Delaying action incurs a measurable environmental cost: a 10-year phased rollout creates a 68-tonne carbon penalty compared to immediate execution (Fig. 10) [27, 29]. Quantifying this cumulative footprint provides direct evidence to justify accelerated capital funding.

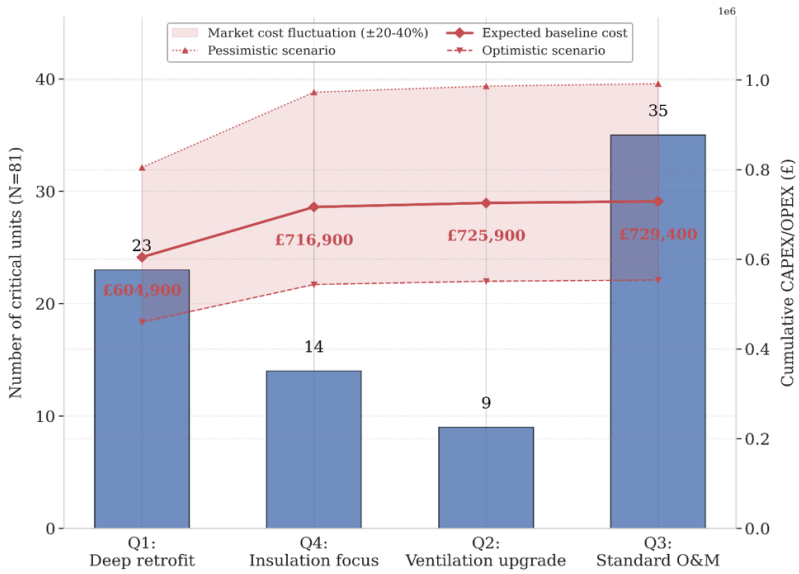
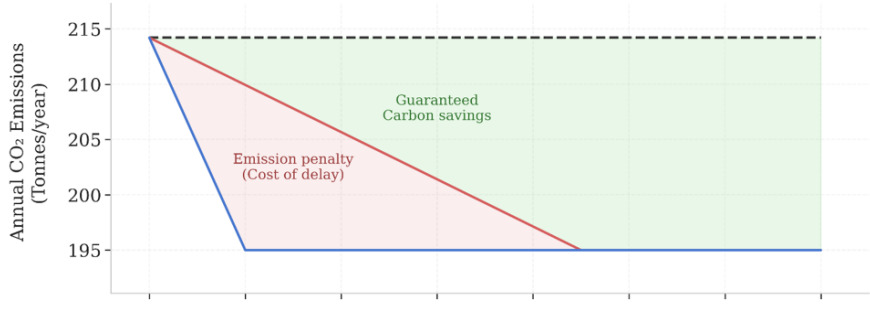


Figure 9. Resource allocation dashboard. Pareto distribution of critical intervention units (columns) relative to cumulative CAPEX/OPEX trajectories (lines) across varying market cost scenarios

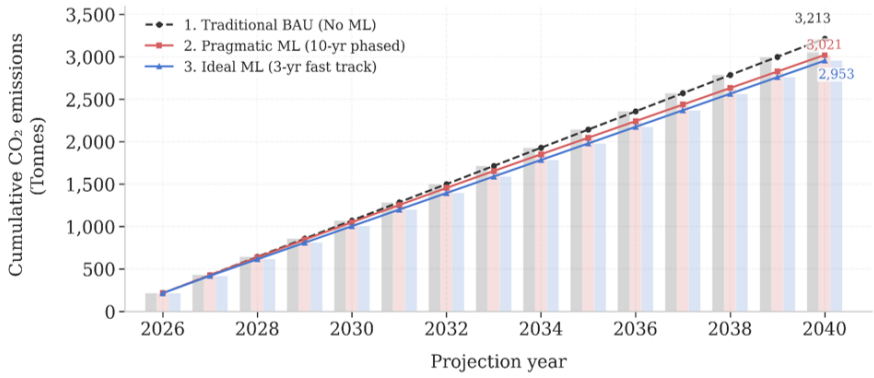
Table 5. Estimated carbon and O&M financial impacts

Quadrant	Intervention strategy	Target units	Estimated energy savings (kWh/yr)*	Estimated carbon mitigation ($kgCO_2e/yr$)*
Q1	Deep retrofit	23	80,500	14,248.5
Q2	Ventilation upgrade (dMEV)	9	N/A (Increased load)	N/A (Health focus)
Q3	Standard O&M	35	N/A (Baseline)	N/A (Baseline)
Q4	Envelope retrofit (Insulation)	14	28,000	4,956.0
Total	**Pilot target program**	**46**	**108,500**	**19,204.5**

*Notes: Calculations utilize steady-state heat transfer equations assuming standard exposed areas, a 3,200-hour heating season, and 85% boiler efficiency. U-value improvements reflect a 2.1 to 0.3 W/m^2K reduction. Carbon reductions apply the UK DESNZ factor of 0.177 $kgCO_2e/kWh$ for natural gas [30].



(a) Annual emissions and cost of delay



(b) Cumulative emissions

Figure 10. Projected CO_2 emission trajectories

4. Conclusions

Decarbonizing real estate necessitates a balance between thermal efficiency and indoor air quality. This study introduces a data-driven framework to navigate this physical trade-off, converting sensor telemetry and asset metadata into targeted maintenance strategies. By formulating the thermodynamic conflict as a multi-label classification problem, the Multi-Output Random Forest algorithm predicted concurrent environmental risks with ROC-AUC scores between 0.730 and 0.783. Spatial cross-validation verifies model generalizability across unseen housing developments [31]. SHAP value analysis successfully isolated structural envelope degradation from occupant ventilation behaviors, giving engineering teams direct evidence to locate the physical origins of building decay.

Practically, the research translates raw predictive probabilities into a structured intervention matrix. Mapping the inverse relationship between heat loss and condensation risks directs precise capital and operational budget allocations. This approach justifies financial investments primarily through respiratory hazard mitigation, shifting the focus away from energy-based metrics. Longitudinal carbon models also quantify the environmental cost of delayed funding, supplying asset managers with the quantitative basis needed to secure green financing.

A primary limitation of this study is the reliance on temperature gradients, rather than direct utility consumption, to evaluate thermal performance. Future iterations must integrate high-frequency smart meters to calibrate projected energy savings against absolute thermodynamic baselines. A secondary limitation is the reliance on CO_2 as a proxy for air exchange rates. The deployed IoT network lacks comprehensive IAQ metrics (e.g., TVOCs, particulate matter) and hardware telemetry for window operations or localized pollution sources. Future iterations should integrate window-contact sensors to empirically calibrate these behavioral proxies. Regarding dataset constraints, the pilot sample size of 81 units limits broad epidemiological extrapolation across diverse architectural styles. Applying spatial Group K-fold cross-validation counteracts this constraint by tightly controlling the risk of overfitting, ensuring the algorithm's predictive logic remains mathematically robust for these specific post-war estates. Expanding this empirical validation to larger portfolios will scale these insights, eventually enabling the integration of real-time air quality thresholds into automated heating and ventilation management systems.

Acknowledgment

This research utilized data provided by a social housing project in the UK. The author expresses gratitude to the project administration for granting access to the IoT sensor network and asset reg-

istries for this empirical analysis.

Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the author(s) used Gemini to improve the language or readability of the manuscript. The authors reviewed and edited the output and take full responsibility for the accuracy, originality, and integrity of the manuscript.

References

[1] International Energy Agency (2023). *Energy system: Buildings*.

[2] Mofidi, F., Akbari, H. (2020). Intelligent buildings: An overview. *Energy and Buildings*, 223:110192.

[3] Shrubsole, C., Macmillan, A., Davies, M., May, N. (2014). 100 unintended consequences of policies to improve the energy efficiency of the UK housing stock. *Indoor and Built Environment*, 23(3):340–352.

[4] Ortiz, M., Itard, L., Bluyssen, P. M. (2020). Indoor environmental quality related risk factors with energy-efficient retrofitting of housing: A literature review. *Energy and Buildings*, 221:110102.

[5] Sharpe, R. A., Thornton, C. R., Nikolaou, V., Osborne, N. J. (2015). Higher energy efficient homes are associated with increased risk of doctor diagnosed asthma in a UK subpopulation. *Environment International*, 75:234–244.

[6] Asadi, E., da Silva, M. G., Antunes, C. H., Dias, L. (2012). Multi-objective optimization for building retrofit strategies: A model and an application. *Energy and Buildings*, 44:81–87.

[7] Ascione, F., Bianco, N., De Stasio, C., Mauro, G. M., Vanoli, G. P. (2017). Artificial neural networks to predict energy performance and retrofit scenarios for any member of a building category: A novel approach. *Energy*, 118:999–1017.

[8] Ma, Z., Cooper, P., Daly, D., Ledo, L. (2012). Existing building retrofits: Methodology and state-of-the-art. *Energy and Buildings*, 55:889–902.

[9] Amasyali, K., El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews*, 81:1192–1205.

[10] Zhao, Y., Li, T., Zhang, X., Zhang, C. (2019). Artificial intelligence-based fault detection and diagnosis methods for building energy systems: Advantages, challenges and the future. *Renewable and Sustainable Energy Reviews*, 109:85–101.

[11] Hosamo, H. H., Svennevig, P. R., Svidt, K., Han, D., Kofoed Nielsen, H. K. (2022). A digital twin predictive maintenance framework of air handling units based on automatic fault detection and diagnostics. *Energy and Buildings*, 261:111988.

[12] Cengel, Y. A., Ghajar, A. J. (2014). *Heat and mass transfer: Fundamentals and applications*. 5th edition, McGraw-Hill Education.

[13] Altamirano-Medina, H., Davies, M., Ridley, I., Mumovic, D., Oreszczyn, T. (2009). Guidelines to avoid mould growth in buildings. *Advances in Building Energy Research*, 3(1):221–235.

[14] Moritz, S., Bartz-Beielstein, T. (2017). imputeTS: Time series missing value imputation in R. *The R Journal*, 9(1):207–218.

[15] Banos, O., Galvez, J.-M., Damas, M., Pomares, H., Rojas, I. (2014). Window size impact in human activity recognition. *Sensors*, 14(4):6474–6499.

[16] Boutell, M. R., Luo, J., Shen, X., Brown, C. M. (2004). Learning multi-label scene classification. *Pattern Recognition*, 37(9):1757–1771.

[17] Zhang, M.-L., Zhou, Z.-H. (2014). A review on multi-label learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, 26(8):1819–1837.

[18] Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8):913–929.

[19] Yang, L., Shami, A. (2020). On hyperparameter optimization of machine learning algorithms: Theory and practice. *Neurocomputing*, 415:295–316.

[20] Lundberg, S. M., Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, 30:4765–4774.

[21] HM Government (2021). *The Building Regulations 2010: Approved Document F – Ventilation, Volume 1: Dwellings*. Department for Levelling Up, Housing and Communities, UK Government.
[22] CIBSE (2015). *Guide A: Environmental design*. 8th ed., updated 2021. Chartered Institution of Building Services Engineers, London, UK.
[23] Climate Change Committee (2019). *UK housing: Fit for the future?*
[24] Energy Saving Trust. *Internal and external wall insulation for solid walls*. Retrieved April 7, 2026.
[25] Energy Saving Trust. *Cavity wall insulation*. Retrieved April 7, 2026.
[26] Chartered Institution of Building Services Engineers (2023). *Guide M: Maintenance engineering and management*. CIBSE.
[27] Climate Change Committee (2020). *Sixth carbon budget*.
[28] World Health Organization (2009). *WHO guidelines for indoor air quality: Dampness and mould*. WHO Regional Office for Europe.
[29] Department for Energy Security and Net Zero (2025). *National Energy Efficiency Data-Framework (NEED): Summary of analysis*. UK Government.
[30] Department for Energy Security and Net Zero (2025). *Greenhouse gas reporting: Conversion factors 2025*. UK Government.
[31] Nguyen, T.-H. T., Jin, R., Chen, W., Gan, L. (2026). Cross-dwelling validation of indoor environmental monitoring for operational risk screening in social housing portfolios. *Journal of Science and Technology in Civil Engineering (JSTCE) - HUCE*, 20(2S):44–54.