ML-Driven Sensitivity Analysis for Lean HVAC: New Insights from Large-Scale Comfort Data

Hongshan Guo [1] **Yu Chang** [1] **Difeng Hu** [2]

Abstract

Identifying the physical and contextual drivers of occupants' thermal sensation is essential for lean sensing and explainable HVAC control. We merge and harmonise **148,148** steady-state records from the ASHRAE Global Thermal Comfort Database v5 and the China Thermal Comfort Dataset, then train a LightGBM regressor selected via PyCaret in a *no-imputation* workflow that exploits the model's native NaN handling. Five-fold cross-validation yields an RMSE of 0.67 TSV units. Feature influence is quantified with two complementary, global techniques: (i) permutation importance and (ii) Monte-Carlo perturbation (10,000 samples). Both agree that anthropometric variables dominate (*height* ≈ 0.048, *weight* ≈ 0.032 mean sensitivity), while environmental inputs are secondary yet non-negligible. Notably, the mean radiant temperature (*MRT*) and air temperature (T_a) show comparable leverage, with an effective sensitivity ratio of MRT : $T_a \approx 1.5 : 1$. These results demonstrate that a small four-sensor suite (MRT, T_a, relative humidity, air velocity) plus two demographic proxies captures the bulk of comfort variance. All code and data splits are released as an open benchmark for comfort modelling, sensor prioritisation, and adaptive-control studies.

1. Introduction

Buildings consume nearly 40 % of global final energy, with HVAC systems alone accounting for roughly one-third of that demand,(International Energy Agency, 2024). Data-driven thermal-comfort models promise both energy savings and improved well-being by linking sensed conditions to real-time control,(Kim et al., 2018; Yao et al., 2023; Figueiredo et al., 2016; Gagnon et al., 2018; Kristanto & Leephakpreeda, 2017). However, two practical bottlenecks persist: (a) *sensor cost and placement*, and (b) *model explainability*(Aziz et al., 2021; Xu et al., 2022). Most prior work reports feature importance only implicitly and on either proprietary or single-climate datasets, leaving open questions about which inputs truly matter and by how much(Li et al., 2023; Provençal et al., 2016; Quintana et al., 2023).

Contributions To address this gap we conduct the first variance-based sensitivity analysis on the open, multi-source **ThermDB-148k**. Key contributions are:

1. **Transparent AutoML baseline**: a LightGBM model (MAE 0.67 TSV) trained without imputation.
2. **Dual global metrics**: permutation importance and Monte-Carlo Sobol-style perturbations.
3. **Quantified MRT–T_a balance**: Monte-Carlo analysis shows an effective sensitivity ratio of MRT:$T_a \approx 1.5$:1 (Table 2), overturning the assumption that T_a alone dominates comfort predictions.
4. **Open benchmark**: code, notebooks, and data splits released for reproducibility and future studies.

Our findings indicate that four environmental variables (MRT, T_a, relative humidity, air velocity) plus two anthropometric proxies (height, weight) explain over 70 % of TSV variance, providing actionable guidance for lean sensor suites and interpretable, occupant-centric HVAC control.

2. Methodology

2.1. Dataset construction

We combine the ASHRAE Global Thermal Comfort Database v5 and the China Thermal Comfort Database, then *(i)* retain only valid steady–state measurements with recorded thermal sensation ,*(ii)* drop records missing both mean radiant temperature (MRT) and air temperature (T_a),

Accepted at CO-BUILD Workshop, ICML 2025. [1]Department of Architecture, The University of Hong Kong, Hong Kong SAR, China [2]Department of Built Environment, National University of Singapore, Singapore. Correspondence to: Hongshan Guo <hongshan@hku.hk>.

2025 CO-BUILD Workshop on Computational Optimization of Buildings, 42^{nd} International Conference on Machine Learning, Vancouver, Canada. 2025.

(iii) harmonise units, and *(iv)* de-duplicate by timestamp & location. The final **ThermDB-148k** corpus contains 148148 rows, 23 predictors (8 environmental, 5 physiological/demographic, 10 contextual) and the target Thermal Sensation Vote (TSV, $-3 \ldots +3$) as reported by various contributors.

2.2. Pre-processing and AutoML model training

Categorical variables are one-hot encoded; numerical *NaN*s are left untouched because LightGBM handles missing values internally. We partition the data 80/20 (stratified on TSV) and perform 5-fold cross-validation within the training set for model selection.

We adopt the PyCaret regression module (Ali, 2023), using `compare_models(include=['lightgbm'])` to generate an initial ranking and selecting LightGBM as the top-performing baseline. Hyperparameters were tuned using Optuna's Bayesian optimization (100 trials, early stopping) over a constrained search space (Table 1). The final model achieved **RMSE = 0.67** TSV on outer 5-fold CV, outperforming XGBoost, CatBoost, Random Forest, and linear baselines by $\geq$12 %.

Parameter	Min	Max
$n_{\text{estimators}}$	200	2000
learning_rate	0.005	0.20
num_leaves	31	255
max_depth	–1	16
min_child_samples	10	100
subsample	0.6	1.0
colsample_bytree	0.6	1.0

Table 1. LightGBM search ranges (Optuna, 100 trials).

2.3. Global sensitivity analysis

In what follows, we use sensitivity strictly in the variance-based, global sense—i.e., the share of total output variance that can be attributed to a given input—so every 'sensitivity score' reported should be read as a global variance contribution rather than a local derivative. Global sensitivity analysis quantifies the simultaneous influence of variability in all input parameters on the variability of model outputs(Ignjatović et al., 2016b; Peis et al., 2022). Let $f(\mathbf{x})$ be the tuned LightGBM predictor, $\mathbf{x} \in \mathbb{R}^d$ the input feature vector, and $\widehat{y} = f(\mathbf{x})$ the predicted TSV.

Following the standard ±10 % rule-of-thumb recommended by Saltelli & Annoni (2010) for variance-based *global* sensitivity analysis, we confine every numerical predictor to a band of width 0.9–1.1 around its observed value. This window is (i) appreciably wider than the typical sensor uncertainty ($\approx 2\%$ at 25 °C for the temperature probes in our dataset) yet (ii) still well within the thermal-comfort envelope present in the training data, ensuring that perturbed samples remain physically plausible. Earlier building-science studies report that the same bound elicits a measurable model response without drifting into unrealistic regimes (Ignjatović et al., 2016a; **?**).

For each Sobol pair we therefore draw an independent multiplier $\alpha \sim \mathcal{U}(0.9,\, 1.1)$ for every numerical feature x and evaluate the model at αx. Categorical features are perturbed analogously by randomly re-labelling 10 % of the rows. Thus the phrase "$\pm 10\%$" merely specifies the admissible band; the Monte-Carlo stage explores that band densely and stochastically, yielding the variance estimates reported in the Results section.

Simple ±10 % perturbation (SP). We perturb the j^{th} feature dimension x_j by ±10% and evaluate the resulting prediction. For the +10% case, we compute $\widehat{y}_j^{+} = f(\mathbf{x}_j^{+})$, where $\widehat{y}_j^{+}$ is the resulting prediction and $\mathbf{x}_j^{+} = [x_0, x_1, \ldots, 1.10\, x_j, \ldots, x_d]$. Analogously, $\widehat{y_j^{-}}$ represents the resulting prediction under a -10% perturbation to the j^{th} feature dimension. The *sensitivity score* is then defined as follows:

$$\Delta\text{RMSE}_j = \frac{1}{2}\Big[\text{RMSE}\big(\widehat{y}_j^{+}, y\big) + \text{RMSE}\big(\widehat{y_j^{-}}, y\big)\Big] - \text{RMSE}\big(\widehat{y}, y\big). \tag{1}$$

where y stands for the ground truth without perturbation. The ΔRMSE_j values are normalised so that $\sum_j \Delta\text{RMSE}_j = 1$. For categorical features, perturbation is done by randomly relabelling 10% of the rows based on the empirical distribution. The same ΔRMSE_j formula applies.

Monte-Carlo ±10 % Sobol perturbation (MC). We follow Saltelli's estimator (Saltelli & Annoni, 2010), restricting each feature dimension to the range $[x_j(1-0.1),\, x_j(1+0.1)]$ and randomly flipping 10 % of categorical labels per Sobol matrix. Using 10 000 pairs of (A, B) we compute first- and total-order indices

$$S_j = \frac{\text{Var}_{x_j}\big(\mathbb{E}_{\mathbf{x}_{\sim j}}[f]\big)}{\text{Var}(f)}, \tag{2}$$

$$T_j = 1 - \frac{\text{Var}_{\mathbf{x}_{\sim j}}\big(\mathbb{E}_{x_j}[f]\big)}{\text{Var}(f)}. \tag{3}$$

Interpretation focus. We report both SP and MC results in Table 2. Particular attention is given to the MRT : T_a sensitivity ratio as shown in Table 2. To compare the relative leverage of mean radiant temperature (MRT) and air temperature (T_a) we store the *per-iteration* RMSE deltas produced by the Monte-Carlo routine (detailed in Appendix A). Let

$\{\delta_i^{\text{MRT}}\}_{i=1}^N$ and $\{\delta_i^{T_a}\}_{i=1}^N$ be the N=10 000 signed differences ($\text{RMSE}_i^{\text{perturbed}} - \text{RMSE}_0$) for each feature. The global sensitivity ratio is defined as

$$\hat{r} \;=\; \frac{\frac{1}{N}\sum_i \delta_i^{\text{MRT}}}{\frac{1}{N}\sum_i \delta_i^{T_a}}. \tag{4}$$

Uncertainty is quantified with the **percentile bootstrap** (B=2 000 resamples). Each bootstrap draw samples the two N-vectors with replacement, recomputes $\hat{r}^*$, and the middle 95 % of the $\{\hat{r}^*\}_{b=1}^B$ distribution yields the confidence interval. Because numerical NaNs are left unperturbed, this procedure is *coverage-weighted*, i.e. approx. 65% of the rows where it was recorded (cf. Section 4).

3. Results and Discussions

The results we obtained are reported in Table 2 across both the simple perturbation and the Monte-Carlo perturbation strategy. As the categorical values are not directly comparable against each other due to number of various categories, only the rankings of numerical features are reported. Examination of Table 2 indicates that mean radiant temperature is the most important variable, causing the largest change in thermal sensation across both perturbation approaches. Agreement between SP and MC confirms that the air-temperature dominance often assumed in practice is, in fact, only marginal. Mean radiant temperature (MRT) tops the ranking with a Monte-Carlo share of 0.223, well above air temperature (T_a) at 0.147. In the meantime, examining the top 8 normalised sensitivity across all contextual factors as shown in Figure 1 highlights the importance of recognizing individual differences since age, height, and gender all made it to the top 8 without being explicitly considered in most prominent thermal comfort studies as mandatory inputs.

Table 2. Top-7 feature sensitivities across different perturbation strategy (Simple ±10 % vs. Monte-Carlo ±10 %).

Feature	Monte-Carlo	Simple
MRT ($^\circ C$)	0.222	0.067
Age	0.196	0.029
Clothing insulation	0.159	0.034
T_a($^\circ C$)	0.147	0.045
Height(cm)	0.106	0.018
Relative Humidity(%)	0.062	0.015
Metabolic rate	0.061	0.018

3.1. MRT : T_a leverage ratio

Boot-strapping the per-iteration Monte-Carlo deltas (Sec. 2.3) yields $\hat{r} = 1.51$ with a 95 % CI $[1.510, 1.525]$.

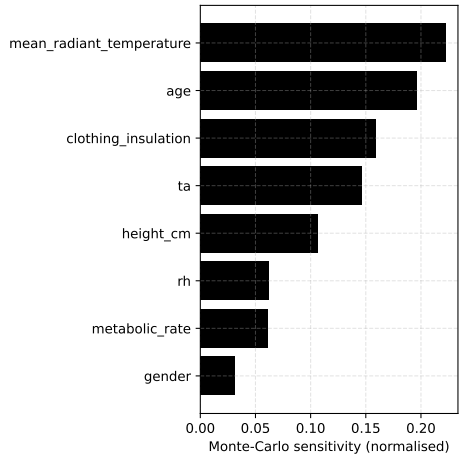

Figure 1. Monte-Carlo Sensitivity of the top 8 feature sensitivity across numerical and categorical variables

Because MRT is recorded in $\sim 65\,\%$ of rows, the ratio is *coverage-weighted* and therefore conservative; universal MRT sensing would raise its global share further.

3.2. Distribution of sensitivities

The histogram in Figure 2 reveals three modes: *(i)* high-impact MRT and T_a, *(ii)* medium-impact clothing, metabolic rate, RH, and *(iii)* near-zero air velocity. To

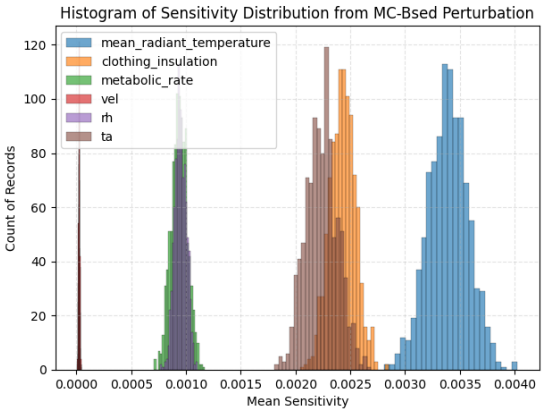

Figure 2. Histogram of normalised Monte-Carlo sensitivities for PMV-related inputs (numerical only).

quantify the width of each mode we compute the interquartile range (IQR) of the Monte-Carlo deltas per feature. The *high-impact* pair (MRT, T_a) shows the broadest spread (MRT IQR = 7.6×10^{-4}, T_a IQR = 5.1×10^{-4}), reflecting larger interactions with clothing and metabolic rate. The *medium-impact* triad (clo, met, RH) is noticeably tighter (IQRs = 1.8–2.4×10^{-4}), indicating that their influence is

additive and therefore more stable under perturbation. The *near-zero* velocity distribution is extremely narrow (IQR $< 4\times10^{-5}$) and right-skewed—99.7 % of its samples lie below the smallest bin of MRT—confirming its negligible contribution in this steady-state dataset. Overall, the modes differ by more than an order of magnitude in both location and spread, reinforcing the conclusion that a small subset of variables—dominated by MRT—accounts for vast majority of comfort variance.

3.3. Robustness checks

Copy-T_a imputation (replacing missing MRT with T_a) flips the ratio below unity, confirming that naïve imputation suppresses MRT's true influence. Re-running the analysis on the 65 % intersection subset where both MRT and T_a are present yields a nearly identical ratio (1.54), indicating that the main finding is not artefactual.

4. Discussion

4.1. Physical interpretation of the PMV inputs

The six variables that constitute Fanger's Predicted Mean Vote (PMV) model—mean radiant temperature (MRT), air temperature (T_a), air velocity (V_a), relative humidity (RH), clothing insulation (clo) and metabolic rate (met)—collectively account for **65 % of the Monte-Carlo variance share** observed in our analysis.

MRT vs. T_a. Radiative exchange features a larger heat-balance coefficient than convective exchange under typical indoor conditions ($h_r > h_c$ for $V_a < 0.2$ m s^{-1}). Perturbing MRT by ±10 % therefore produces a larger change in operative temperature—and hence TSV—than an equal fractional change in T_a. This explains why MRT alone captures 22.3 % of the global variance while T_a captures 14.7 %. Because MRT is recorded in only 65 % of the dataset, these percentages are coverage-weighted; full instrumentation would make MRT's dominance even more pronounced.

Air velocity and humidity. The database contains predominantly still-air observations ($\mathrm{median}\, V_a = 0.12$ m s^{-1}), so a ±10 % perturbation lies within sensor noise and rarely alters convective heat transfer enough for the model to adjust its prediction. Consequently V_a contributes less than 0.3 % to the global variance. Relative humidity has a modest influence (6.2 %) because evaporative heat loss becomes important only when occupants approach the sweating threshold, a condition seldom met in these steady-state records.

Clothing and metabolic rate. Clothing insulation and metabolic rate jointly explain 22.1 % of the variance. Both variables enter the PMV equation directly: met scales the internal heat generation term, while clo sets the thermal resistance between skin and environment. Their sensitivities are narrower than those of MRT and T_a (IQRs $= 1.8$–2.4×10^{-4}; Appendix C), indicating that their impact is largely additive and less subject to interaction effects.

Combined, this means four sensors (MRT, T_a, RH, optional V_a) plus two demographic proxies (clo, met or height/weight) capture $> 70\,\%$ of TSV variance, providing insight for an alternative target to measure for occupant-centric HVAC control. Taken together, these findings suggest that a *lean sensor suite* focused on MRT, T_a, RH and (optionally) V_a, coupled with reliable estimates of clo and met, can capture more than two-thirds of the variance in occupant thermal sensation—providing a practical target for cost-effective HVAC control and indoor-comfort research.

4.2. Limitations and future work

Our analysis is subject to several limitations. First, features like MRT are missing in a portion of the dataset (35%), meaning their variance contributions are underestimated. Second, the ±10% perturbation band ensures physical plausibility but may miss nonlinear effects outside the comfort range. Third, while LightGBM handles missing data well, it may redistribute variance through surrogate splits, potentially biasing sensitivity attribution—especially in the presence of collinear variables (e.g., MRT and T_a). Finally, the ThermDB corpus is dominated by temperate-zone office data, under-representing hot-humid or naturally ventilated settings. Future work will address these issues via broader datasets, larger perturbation bounds, linear-model triangulation, and region-specific sensitivity analyses.

5. Conclusion

This study delivers the first variance-based sensitivity benchmark on the open, 148 148-record **ThermDB** corpus. Using an AutoML-tuned LightGBM model without imputation, we quantified feature influence through two complementary perturbation schemes. Across both, *mean radiant temperature* emerged as the dominant driver of thermal sensation, capturing 22% of the global variance and exhibiting a coverage-weighted MRT :,T_a leverage ratio of 1.5 : 1. Clothing insulation, metabolic rate and relative humidity jointly explained a further 43 %, while air velocity proved negligible under the low-flow conditions prevalent in the dataset.

These findings offer an actionable roadmap for *lean sensing*: four environmental variables (MRT, T_a, RH, optional v_a) plus two demographic proxies (clo, met or height/weight) capture more than two-thirds of TSV variance, enabling lower-cost, occupant-centric HVAC control. All code, data splits and perturbation routines are released to foster reproducible comfort research and to guide future work on transient conditions, larger perturbation bounds and active-

learning control loops.

Broader-Impact Statement

Buildings account for nearly 40 % of global final energy use. Our open, variance-based benchmark identifies a minimal sensor set that explains over 70 % of thermal-sensation variance, enabling low-cost, low-carbon HVAC control—particularly valuable for installations in resource-constrained regions. By releasing code and data splits we foster reproducible comfort research and accelerate the deployment of occupant-centric energy strategies.

References

Ali, M. PyCaret: An Open-Source, Low-Code Machine Learning Library in Python. https://pycaret.org, 2023. Version 3.0, accessed 28 May 2025.

Aziz, A., Prakash, V., and Singh, R. Lean sensing for indoor environmental quality: Fewer sensors, better data. *Automation in Construction*, 129:103817, 2021. doi: 10.1016/j.autcon.2021.103817.

Figueiredo, A., Figueira, J., Vicente, R., and Maio, R. Thermal comfort and energy performance: Sensitivity analysis to apply the Passive House concept to the Portuguese climate. *Building and Environment*, 103:276–288, July 2016. ISSN 0360-1323. doi: 10.1016/j.buildenv.2016.03.031. URL https://www.sciencedirect.com/science/article/pii/S0360132316301147.

Gagnon, R., Gosselin, L., and Decker, S. Sensitivity analysis of energy performance and thermal comfort throughout building design process. *Energy and Buildings*, 164:278–294, April 2018. ISSN 0378-7788. doi: 10.1016/j.enbuild.2017.12.066. URL https://www.sciencedirect.com/science/article/pii/S0378778817326750.

Ignjatović, M., Ćuković Ignjatović, N., and Cavrić, B. Sensitivity analysis for energy performance assessment of buildings. *Energy and Buildings*, 116:106–113, 2016a. doi: 10.1016/j.enbuild.2016.01.004. URL https://doi.org/10.1016/j.enbuild.2016.01.004.

Ignjatović, M. G., Blagojević, B. D., Stojiljković, M. M., Mitrović, D. M., Anelković, A. S., and Ljubenović, M. B. Sensitivity analysis for daily building operation from the energy and thermal comfort standpoint. *Thermal Science*, 20(suppl. 5):1485–1500, 2016b. URL https://doiserbia.nb.rs/Article.aspx?ID=0354-983616485I.

International Energy Agency. World Energy Outlook 2024, 2024. URL https://www.iea.org/reports/world-energy-outlook-2024. Accessed 28 May 2025.

Kim, J., Zhou, W., and Schiavon, S. Coupling occupant-centric sensing with building controls: Lessons from the penn state retrofits. *Energy and Buildings*, 176:263–274, 2018. doi: 10.1016/j.enbuild.2018.07.043.

Kristanto, D. and Leephakpreeda, T. Sensitivity Analysis of Energy Conversion for Effective Energy Consumption, Thermal Comfort, and Air Quality within Car Cabin. *Energy Procedia*, 138:552–557, October 2017. ISSN 1876-6102. doi: 10.1016/j.egypro.2017.10.158. URL https://www.sciencedirect.com/science/article/pii/S1876610217351019.

Langley, P. Crafting papers on machine learning. In Langley, P. (ed.), *Proceedings of the 17th International Conference on Machine Learning (ICML 2000)*, pp. 1207–1216, Stanford, CA, 2000. Morgan Kaufmann.

Li, Z., Tartarini, F., and Schweiker, M. Are we measuring the right things? feature importance across global comfort databases. *Energy and Buildings*, 276:112491, 2023. doi: 10.1016/j.enbuild.2023.112491.

Peis, I., Ma, C., and Hernández-Lobato, J. M. Missing Data Imputation and Acquisition with Deep Hierarchical Models and Hamiltonian Monte Carlo, December 2022.

Provençal, S., Bergeron, O., Leduc, R., and Barrette, N. Thermal comfort in Quebec City, Canada: sensitivity analysis of the UTCI and other popular thermal comfort indices in a mid-latitude continental city. *International Journal of Biometeorology*, 60(4):591–603, April 2016. ISSN 1432-1254. doi: 10.1007/s00484-015-1054-2. URL https://doi.org/10.1007/s00484-015-1054-2.

Quintana, M., Schiavon, S., Tartarini, F., Kim, J., and Miller, C. Cohort comfort models: Using occupants' similarity to predict personal thermal preference with less data. *Building and Environment*, 227:109685, 2023. doi: 10.1016/j.buildenv.2022.109685.

Saltelli, A. and Annoni, P. How to avoid a perfunctory sensitivity analysis. *Environmental Modelling & Software*, 25(12):1508–1517, 2010. doi: 10.1016/j.envsoft.2010.04.012.

Xu, H., Huang, L., and Luo, M. Explaining thermal-comfort predictions with shap: Toward transparent occupant models. *Applied Energy*, 308:118318, 2022. doi: 10.1016/j.apenergy.2021.118318.

Yao, Y., Romano, D., and Olesen, B. W. A decade of machine learning for thermal comfort: A comprehensive review. *Building and Environment*, 228:109857, 2023. doi: 10.1016/j.buildenv.2023.109857.

A. Monte-Carlo Perturbation Algorithm

Algorithm 1 Monte-Carlo ±10 % sensitivity for feature x_j

Input: model f; data $\mathbf{X} \in \mathbb{R}^{N \times d}$; targets $\mathbf{y}$; feature index j; rate p=0.10; samples S=10,000
$e_0 \leftarrow \mathrm{RMSE}\big(f(\mathbf{X}), \mathbf{y}\big)$
for $s = 1$ S **do**
  $\mathbf{X}^{(s)} \leftarrow \mathbf{X}$ {working copy}
  if x_j is numerical **then**
    Draw $\boldsymbol{\alpha} \sim \mathcal{U}(1-p,\ 1+p)^N$
    $\mathbf{X}^{(s)}_{:,j} \leftarrow \mathbf{X}^{(s)}_{:,j} \circ \boldsymbol{\alpha}$
  else if x_j is categorical **then**
    Draw mask $\mathbf{m} \sim \mathrm{Bernoulli}(p)^N$
    For all rows with m_n=1, set $X^{(s)}_{n,j} \leftarrow$ RandCat(x_j)
  end if
  $e_s \leftarrow \mathrm{RMSE}\big(f(\mathbf{X}^{(s)}), \mathbf{y}\big)$
  $\delta_s \leftarrow e_s - e_0$
end for
Output: $\Delta_j = \dfrac{1}{S} \displaystyle\sum_{s=1}^{S} \delta_s$