



Feature Analysis and Predictive Modeling for Occupant-Centric Thermal Comfort in Naturally Ventilated University Cafeterias Using Advanced Machine Learning Models

Mim Mony¹, Nowshin Nowyer Oyshi², Lamia Jamal³, Furuja Tasneem⁴, Mahabuba Porosh Moni⁵, and Mohammad Nyme Uddin*⁶

Building Energy & Environmental Management-BEEM Enhanced by AI, Dhaka, Bangladesh
nymebd.uddin@connect.polyu.hk

Abstract. Thermal comfort is a critical aspect of outdoor environmental quality that significantly influences human health, well-being, and productivity, particularly in densely populated cities such as Dhaka, Bangladesh. Traditional predictive methods typically rely on simplified models and limited parameters, constraining their ability to capture the complex interactions between environmental factors and occupant behavior. Therefore, this study develops advanced machine learning models for feature analysis and predictive modeling to predict occupant-centric thermal comfort in naturally ventilated university cafeterias. A total of 500 samples were collected from two cafeterias during the summer season (June 2025 to August 2025). The dataset comprises 18 features covering environmental parameters, demographic data, human perception related, and architectural design characteristics. Random Forest (RF) and XGBoost were applied. Feature importance was evaluated using the Mean Decrease Impurity (MDI) approach, which estimates each feature's contribution based on the average reduction in impurity achieved across all splits within the ensemble models. Additionally, Permutation-based feature importance was employed to evaluate each variable's predictive contribution by measuring the increase in model error after randomly shuffling its values. Furthermore, SHapley Additive exPlanations (SHAP) analysis was used to interpret feature importance and increase model transparency. The findings consistently highlighted the importance of certain features in the analysis, such as Wind_Velocity, Humidity, Sweating_Effect etc. in predicting occupant comfort levels. This research provides critical insights into optimizing outdoor thermal comfort in naturally ventilated cafeterias, empowering evidence-based decisions in both architectural design and facility management. Future research should include broader datasets and multi-seasonal measurements to improve model generalization.

Keywords: Thermal Comfort, Naturally Ventilated University Cafeterias, Feature Analysis, Machine Learning Models.

1 Introduction

Predicting thermal comfort is crucial for optimizing built environments for human habitation, as it impacts health, productivity, and overall well-being [1]. Especially in densely occupied public spaces such as university cafeterias. In countries like Bangladesh with hot and humid climates, naturally ventilated cafeterias are widely used due to energy limitations and economic factors. But these environments often fail to maintain adequate thermal comfort because climatic variations and high occupant densities significantly change outdoor climates. In universities, cafeterias are not only dining areas but also social and study spaces where not only students but also teachers and staff spend long periods. Discomfort in such environments can negatively impact their health, learning mentality, social interaction, and overall campus experience. Despite increasing awareness and technological advancement, thermal comfort management in naturally ventilated cafeteria remains challenging due to the complex interaction between environmental factors, architectural design characteristics, and human perceptual responses.

Existing research on thermal comfort has relied mostly on traditional models, such as the Predictive Mean Vote (PMV), which depend on controlled air conditions and narrow climatic contexts. While valuable, these methods often oversimplify how people interact with their environments. They also fail to track real-time changes in naturally ventilated spaces, where occupants shape comfort through actions like adjusting clothing, airflow, or position. Machine Learning (ML) models have recently shown promise for complex environmental predictions and feature interactions. Several studies examine ML-based thermal comfort prediction in classrooms, offices, and homes, but few focus on university cafeterias. Outdoor comfort in these spaces is heavily influenced by climatic factors such as temperature, humidity, thermal radiation, and wind velocity [2].

This study develops advanced ML models for feature analysis and predictive modeling to predict occupant-centric thermal comfort in naturally ventilated university cafeterias. ML will transform how environment interact with both their occupants and environment, leading to improved efficiency, comfort, and sustainability [3].

2 Literature Review

Thermal comfort in naturally ventilated (NV) outdoor spaces has been extensively studied. Traditional models, such as PMV/PPD, have been widely applied to predict thermal comfort but often fail to capture the variability arising from occupant behavior, adaptive responses, and environmental dynamics typical of NV spaces [4]. While previous studies highlight the importance of environmental parameters like temperature, humidity, and air quality, they rarely integrate occupant-centric factors such as clothing type, activity level, health condition, and personal perception, which can significantly influence comfort outcomes [5]. Most research in NV educational spaces focuses on classrooms or lecture halls. Seasonal field studies have shown that neutral temperatures and hu-

midity levels in these spaces are influenced by adaptive behavior, such as window operation and clothing adjustment [7]. These findings indicate that occupants actively interact with the environment to maintain comfort, but previous models often fail to quantitatively capture these complex human-environment interactions. Ensemble algorithms, particularly Random Forest (RF) and XGBoost, are increasingly used due to their ability to handle non-linear relationships and heterogeneous datasets [8]. Feature importance analysis within these models, including permutation-based measures and mean decrease impurity, allows for the identification of the most influential environmental, architectural, and human-centric factors [6]. Explainable AI approaches such as SHAP further enhance model transparency, providing insights that can inform facility management and architectural design [9,10]. While prior research demonstrates the utility of ML in thermal comfort prediction, most studies are limited to indoor spaces with stable occupancy and controlled environments. Very few investigations focus on naturally ventilated cafeterias, where occupant behavior, clothing, activity patterns, and architectural design features create highly dynamic thermal conditions.

Table 1. Summary of recent review papers.

Paper	Year	No. of Re-viewed Papers	Years of Re-viewed Papers	Objectives	Key Findings / Gaps
[11]	2020	37	2005–2019	Explores predictive thermal-comfort models and their application.	Identified gaps include limited AI use.
[12]	2021	45	2005–2019	Reviews analytical thermal-comfort models using ANN- and RL.	Missing spatial features (e.g., room size, ceiling height).
[13]	2020	105	2010–2020	overview of occupant-centric comfort frameworks	Challenges include in sensing technologies.
[14]	2020	N/A	N/A	concepts of intelligent systems and evaluates current applications.	Future includes improved methods to adapt to dynamic comfort.
[15]	2019	33	1997–2018	Reviews the implementation of reinforcement learning.	Ignoring lighting and air quality, while excluding occupant behavior.

3 Methodology

This research used a mixed-method approach, integrating objective environmental measurements with subjective occupant surveys to thoroughly assess thermal comfort responses. The methodology follows a systematic process, as shown in figure 1:

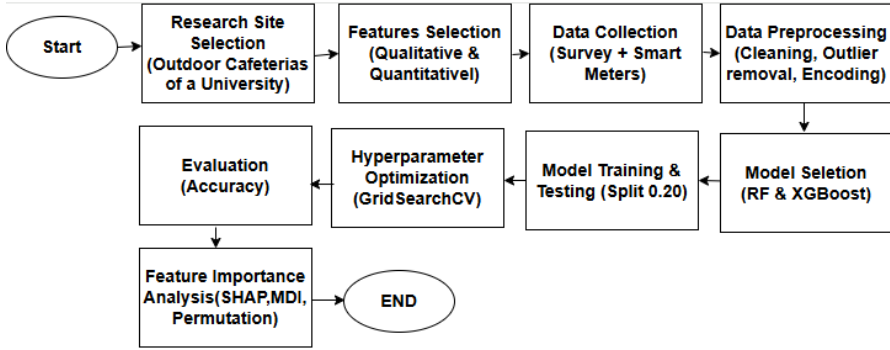


Fig. 1. Flowchart of Methodology

This study was conducted in two naturally ventilated university cafeterias located in Dhaka, Bangladesh. Both sites were chosen for their open layouts and high midday occupancy, where the thermal environment is influenced by heat from cooking activities, body warmth from large crowds, and the absence of mechanical cooling systems. A total of 500 samples were collected during the summer season (June 2025 to August 2025) to capture conditions under peak thermal stress. Each sample included 18 features spanning environmental, demographic, perceptual, and architectural parameters to capture both objective conditions and subjective experiences of comfort. All environmental parameters were measured using portable digital sensors to ensure consistency across both sites, while participant surveys captured subjective comfort ratings and demographic details.

Table 2. Characteristics of features

Category	Feature	Description	Unit
Environmental (7)	Air_Temperature_(°C)	Outdoor air temperature	°C
	Humidity_%	Outdoor relative humidity	%
	CO ₂ _(ppm)	CO ₂ concentration	Ppm
	Wind_Velocity_(m/s)	Air speed in the university area	m/s
	Light_Intensity_(lux)	Light intensity near the respondent	Lux
	TVOCs	Volatile organic compound s	mg/m ³
	Thermal_Comfort	Predicted Mean Vote – thermal comfort parameter	Categorical
Demographic (3)	Gender	Respondent's gender	Categorical
	Age	Respondent's age	Categorical
	Academic_Role	Role in academic setting	Categorical
Perceptual (6)	Health_Issue	Presence of environment-related health issue	Categorical

	Food_Temperature	Temperature perception of food	Categorical
	People_Density	Perceived crowding of the cafeteria	Categorical
	Clothing_Type	Type of clothing material worn	Categorical
	Sneaker_Effect	Perceived impact of wearing sneakers on thermal comfort (Yes, Neutral, No)	Categorical
	Sweating_Intensity	Self-reported sweating level while in the cafeteria (None, Light, Heavy)	Categorical
Architectural (2)	Cafeteria_Location	Location of cafeteria	Categorical
	Ceiling_Height_(mete)	Floor-to-ceiling distance	Meters

Data preprocessing involved analyzing feature distributions and variance, removing low-variance features and IQR-based outliers to enhance model robustness.

Random Forest (RF) and XGBoost were applied. The processed dataset was divided into 80% for training and 20% for testing. Model tuning was performed using GridSearchCV with 5-fold cross-validation. After training, feature importance was analyzed through multiple interpretability techniques. The Mean Decrease Impurity (MDI) metric was used first to estimate each variable’s contribution to reducing prediction uncertainty. MDI naturally handles multicollinearity and can highlight important features [16]. To confirm these results, permutation-based feature importance was applied. Finally, SHapley Additive exPlanations (SHAP) were used to interpret individual feature effects, providing a transparent understanding of how environmental, perceptual, and architectural variables influence comfort prediction.

4 Results and Discussion

XGBoost and RF models effectively classified thermal comfort states across six perceptual categories: cold, slightly cold, slightly cool, neutral, slightly warm, and hot. Each model achieved accuracies exceeding 70%. Feature importance derived from the MDI analysis revealed a subset of environmental and perceptual parameters as the key determinants. These findings were reinforced by permutation-based analysis, where randomizing these variables produced marked declines in model performance across both train and test sets. SHAP analysis provided additional clarity by revealing directional impacts, such as the tendency of elevated CO2 and humidity in the air shifted predictions towards a lower comfort level, whereas higher comfort levels were associated with moderate lighting and velocity of wind. A systematic comparison of the three importance techniques revealed a clear and consistent pattern. The MDI plots (Figs. 2&3) ranked Wind Velocity, Humidity, Light Intensity, Sweating Effect, Air Temperature, CO2, and Academic Role as the most influential predictors for both models. Permutation Importance (Figs. 4&5) produced a nearly identical hierarchy with Wind Velocity, Humidity, Sneakers Effect, Sweating Effect, and Air Temperature causing the most significant declines in model accuracy when permuted, followed by Light Intensity and Ceiling Height. The SHAP summary plots (Figs. 6&7) reinforced the same trend, positioning Wind Velocity, Humidity, Light Intensity, Air Temperature, CO2 and Sweating Effect as the strongest drivers across both models. From the three methods, Wind Velocity, Humidity, Light Intensity, Sweating Effect, Air Temperature, and

CO₂ level consistently appeared among the highest-ranked predictors, marking them as the most dominant factors. Wind Velocity improves comfort by enhancing heat loss through the movement of air, while Humidity reduces comfort as it slows down the evaporation of sweat. Light Intensity shapes how occupants perceive warmth, as extreme brightness or dullness increases discomfort. The Sweating Effect reflects the body's thermal strain, where higher sweating levels correspond to warmer sensations. Air Temperature directly affects whether conditions and influences whether occupants feel cold or warm, making it a key determinant of thermal comfort. Lastly, elevated CO₂ levels signal poor ventilation and crowded space, and such conditions typically lead to lower comfort. Other Variables, including Cafeteria Location, TVOC, Food Temperature, Clothing Type, and Sneakers Effect, showed a moderate but meaningful influence by subtly affecting occupants' thermal comfort. In contrast to earlier regression-based studies with limited feature sets that mainly identified temperature and humidity as key comfort drivers [17].

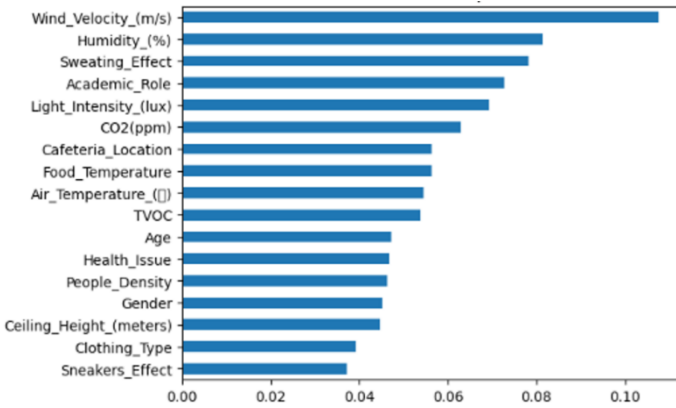


Fig. 2. MDI Feature Importance for XGBoost

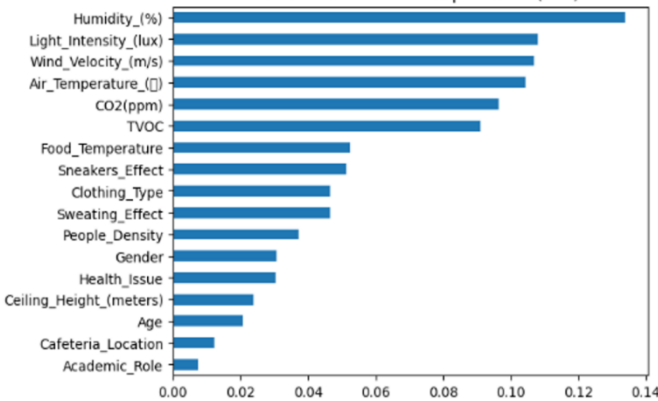


Fig. 3. MDI Feature Importance for RF

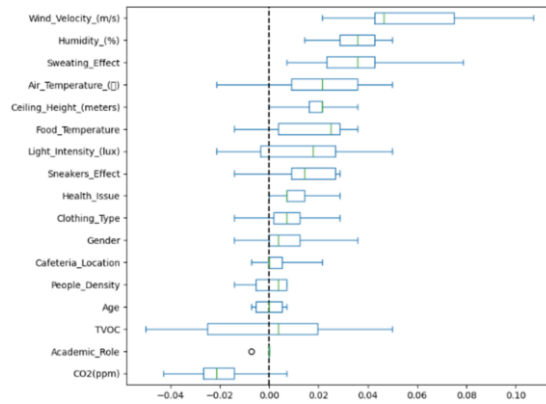


Fig. 4. Permutation Feature Importance for XGBoost

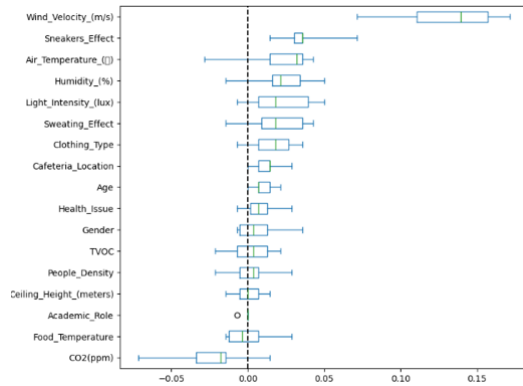


Fig. 5. Permutation Feature Importance for RF



Fig. 6. SHAP Summary Plot for XGBoost



Fig. 7. SHAP Summary Plot for RF

5 Conclusion

This study demonstrates that advanced machine learning techniques can effectively predict thermal comfort in naturally ventilated university cafeterias by capturing the complex interactions between environmental and perceptual factors in Dhaka, Bangladesh. Feature analysis and predictive modeling were performed using MDI, permutation-based, and SHAP methods, with Wind Velocity, Humidity, Light Intensity, Sweating Effect, Air Temperature, and CO₂ consistently emerging as the most influential variables. By pinpointing the key determinants of occupant comfort, the findings offer practical guidance for cafeteria management to enhance outdoor environmental quality. Strategies such as improving ventilation, regulating temperature and humidity, managing occupant density, and accounting for perceptual responses can significantly boost comfort, well-being, and productivity, contributing to healthier, more sustainable, and user-centered campus spaces.

The novelty of this research lies in its unique features and its mixed-mode approach, combining environmental, architectural, and human-centered factors for a comprehensive analysis of thermal comfort. Nevertheless, the study has some limitations, including seasonal weather variations, a narrow participant age range, challenges in classifying qualitative data, and limitations in measurement device accuracy. To strengthen the generalizability of the results, future research should incorporate data from multiple seasons and examine a wider variety of cafeteria layouts, including enclosed or mechanically ventilated spaces. Such efforts will support the development of adaptable,

occupant-focused strategies for promoting healthier and more comfortable environments.

Conflict of Interest

The authors declare that they have no conflict of interest.

Data Availability.

Data is available on request from the corresponding author.

References

1. Assymkhan, N., Kartbayev, A.: Advanced IoT-enabled indoor thermal comfort prediction using SVM and random forest models. *Int. J. Adv. Comput. Sci. Appl.* 15(8), 102–110 (2024). DOI: 10.14569/IJACSA.2024.01508102
2. Saha, R., et al. Monitoring Air Quality of Dhaka using IoT: Effects of COVID-19. in 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST). 2021.
3. Chen, T.-Y., Huang, C.-S., Sung, W.-P.: Improving summer outdoor comfort in metropolitan park: a data-driven approach using AI, experimental and design analysis. *Measurements in Engineering* 2, 466–483 (2025).
4. Boutabba, A., Albalkhy, W., Lafhaj, Z., Roussel, J., Yim, P., Danel, T.: Enhancing thermal comfort and energy efficiency in buildings using artificial intelligence: A systematic literature review. *Modular and Offsite Constr. Summit*, 61 (2025).
5. Luo, M., Xie, J., Yan, Y., Kee, Z., Yu, P., Wang, Z., Zhang, J.: Comparing machine learning algorithms in predicting thermal sensation using ASHRAE Comfort Database II. *Energy Build.* 210, 109776 (2020).
6. Jia, L.-R., Han, J., Chen, X., Li, Q.-Y., Lee, C.-C., Fung, Y.-H.: Interaction between thermal comfort, indoor air quality and ventilation energy consumption of educational buildings: A comprehensive review. *Buildings* 11(12), 591 (2021). <https://doi.org/10.3390/buildings11120591>
7. Qavidel Fard, Z., Zomorodian, Z.S., Korsavi, S.S.: Application of machine learning in thermal comfort studies: A review. *Energy Build.* 256, 111–129 (2022).
8. Wang, X., Yang, X., Gao, L., Zhao, S., Zhai, Y.: Thermal comfort in naturally ventilated university classrooms: A seasonal field study in Xi'an, China. *Energy Build.* 247, 111126 (2021). <https://doi.org/10.1016/j.enbuild.2021.111126>
9. Aparicio-Ruiz, P., Barbadilla-Martín, E., Guadix, J., Nevado, J.: Analysis of variables affecting indoor thermal comfort in Mediterranean climates using machine learning. *Buildings* 13(9), 2215 (2023).
10. Uddin, M.N.: Thermal and visual comforts of occupants for a naturally ventilated educational building in low-income economies: A machine learning approach. *Build. Eng.* 6, 101345 (2024). <https://doi.org/10.1016/j.job.2024.110015>
11. Lala, B., Manas Kala, S., Rastogi, A., Dahiya, K., Yamaguchi, H., Hagishima, A.: Multi-task learning for concurrent prediction of thermal comfort, sensation, and preference. *arXiv preprint arXiv:2204.04567* (2022)
12. Ngarambe, J., Yun, G.Y., Santamouris, M.: The use of artificial intelligence (AI) methods in the prediction of thermal comfort in buildings: Energy implications of AI-based thermal comfort controls. *Energy Build.* 211, 109807 (2020). <https://doi.org/10.1016/j.enbuild.2020.109807>

12. Ma, N., Aviv, D., Guo, H., Braham, W.W.: Measuring the right factors: A review of variables and models for thermal comfort and indoor air quality. *Renew. Sustain. Energy Rev.* 135, 110436 (2021). <https://doi.org/10.1016/j.rser.2020.110436>
13. Xie, J., Li, H., Li, C., Zhang, J., Luo, M.: Review on occupant-centric thermal comfort sensing, predicting, and controlling. *Energy Build.* 226, 110392 (2020). <https://doi.org/10.1016/j.enbuild.2020.110392>
14. Ghahramani, A., Galicia, P., Lehrer, D., Varghese, Z., Wang, Z., Pandit, Y.: Artificial intelligence for efficient thermal comfort systems: Requirements, current applications and future directions. *Front. Built Environ.* 6, 21 (2020). <https://doi.org/10.3389/fbuil.2020.00021>
15. Han, M., May, R., Zhang, X., Wang, X., Pan, S., Yan, D., Jin, Y., Xu, L.: A review of reinforcement learning methodologies for controlling occupant comfort in buildings. *Sustain. Cities Soc.* 51, 101748 (2019). <https://doi.org/10.1016/j.scs.2019.101748>
16. Bhadra, S., Sagan, V., Maimaitijiang, M., Maimaitiyiming, M., Newcomb, M., Shakoore, N., Mockler, T.C.: Quantifying leaf chlorophyll concentration of sorghum from hyperspectral data using derivative calculus and machine learning. *Remote Sens.* 12(13), 2105 (2020). <https://doi.org/10.3390/rs12132082>
17. Xue, J., Liu, W., Liu, K.: Influence of thermal environment on attendance and adaptive behaviors in outdoor spaces: A study in a cold-climate university campus. *Int. J. Environ. Res. Public Health* 18(11), 6139 (2021). <https://doi.org/10.3390/ijerph18116139>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

