IRSTI 28.23.01 https://doi.org/10.26577/jpcsit202543

Bibars Amangeldy[1*] , Bauyrzhan Abilda[2]

[1]LLP «DigitAlem», Almaty, Kazakhstan
[2]Astana International University, Astana, Kazakhstan
*e-mail: a.s.bibars@gmail.com

SMART BUILDING CLIMATE CONTROL: MACHINE LEARNING APPROACH FOR INDIVIDUAL THERMAL PREFERENCE PREDICTION

Abstract. Modern building management systems rely on uniform climate settings that fail to accommodate individual occupant preferences, resulting in energy waste and reduced comfort satisfaction. This study presents a data-driven approach for personalized thermal comfort prediction using machine learning algorithms integrated with multimodal sensor networks. We developed and evaluated three classification models (Random Forest, XGBoost, and Artificial Neural Network) using environmental parameters (air temperature, humidity, CO_2 concentration) and physiological measurements (heart rate variability, blood pressure, oxygen saturation) collected from controlled experiments with eight participants under various thermal conditions. The optimized Random Forest model achieved 95% accuracy in predicting seven-level thermal sensation votes using only ten key features identified through SHAP analysis. Indoor air temperature emerged as the dominant predictor, while physiological parameters provided complementary information for personalized comfort assessment. The proposed system demonstrates significant potential for integration into smart building automation, enabling dynamic climate control that adapts to individual preferences while optimizing energy consumption. Implementation of such personalized HVAC systems could reduce energy usage by up to 20% compared to conventional static temperature control, while simultaneously improving occupant satisfaction and productivity in commercial buildings.

Keywords: smart buildings, thermal comfort prediction, machine learning, HVAC optimization, personalized climate control, energy efficiency, sensor fusion.

1. Introduction

Buildings consume approximately 40% of global energy, with heating, ventilation, and air conditioning (HVAC) systems accounting for nearly 50% of this consumption [1]. In commercial buildings alone, inefficient climate control results in annual energy losses exceeding $29 billion globally, while simultaneously causing productivity losses due to occupant discomfort [2]. Despite these significant economic and environmental impacts, most building management systems continue to rely on static temperature setpoints that fail to accommodate individual thermal preferences.

Traditional HVAC control strategies assume uniform thermal comfort across all occupants, typically maintaining indoor temperatures between 20-24°C based on population averages [3]. However, research demonstrates substantial individual variations in thermal perception, with personal comfort preferences differing by up to 6°C among occupants in the same space [4]. This one-size-fits-all approach leads to overcooling or overheating, resulting in energy waste and occupant dissatisfaction rates exceeding 30% in typical office environments [5]. Furthermore, conventional systems lack real-time adaptation capabilities, failing to respond to changing occupancy patterns, weather conditions, or individual physiological states [6].

Recent advances in Internet of Things (IoT) sensors and machine learning algorithms present unprecedented opportunities for developing intelligent, personalized climate control systems [7]. Smart building platforms now enable continuous monitoring of environmental parameters and occupant behavior, while wearable devices provide real-time physiological data for individual comfort assessment [8]. Several pilot implementations have demonstrated the potential for machine learning-driven HVAC optimization, achieving energy savings of 15-25% while improving occupant satisfaction [9], [10]. However, these systems typically rely on limited sensor inputs and simplified comfort models that may not capture the complex relationships between environmental conditions, physiological responses, and individual thermal preferences [11].

Despite growing interest in personalized building automation, significant research gaps remain in developing robust, scalable comfort prediction models. Existing approaches often analyze environmental and physiological data streams in isolation, failing to leverage the synergistic effects of multimodal sensor fusion [12]. Additionally, most studies focus on laboratory conditions with homogeneous participant groups, limiting the generalizability of findings to diverse real-world applications [13]. Machine learning techniques show promise for addressing these challenges, with recent studies demonstrating successful applications of artificial neural networks and ensemble methods for thermal comfort prediction [14-15]. However, most existing models rely on limited feature sets and lack comprehensive integration of physiological monitoring data [16-17]. This study addresses these limitations by developing and evaluating machine learning models that integrate environmental monitoring with physiological sensing for accurate, personalized thermal comfort prediction. The primary objective is to demonstrate the feasibility of implementing such systems in smart buildings to achieve simultaneous improvements in energy efficiency and occupant satisfaction through advanced sensor fusion and explainable AI techniques [18-20].

2. Materials and Methods

This study employed a controlled experimental approach to develop and validate machine learning models for personalized thermal comfort prediction. The methodology integrates multimodal sensor data collection, advanced data preprocessing, and ensemble learning techniques to create a robust classification system capable of predicting individual thermal preferences across diverse environmental conditions. The overall system architecture is illustrated in Figure 1, showing the integration of multimodal sensors, data processing, and machine learning components.

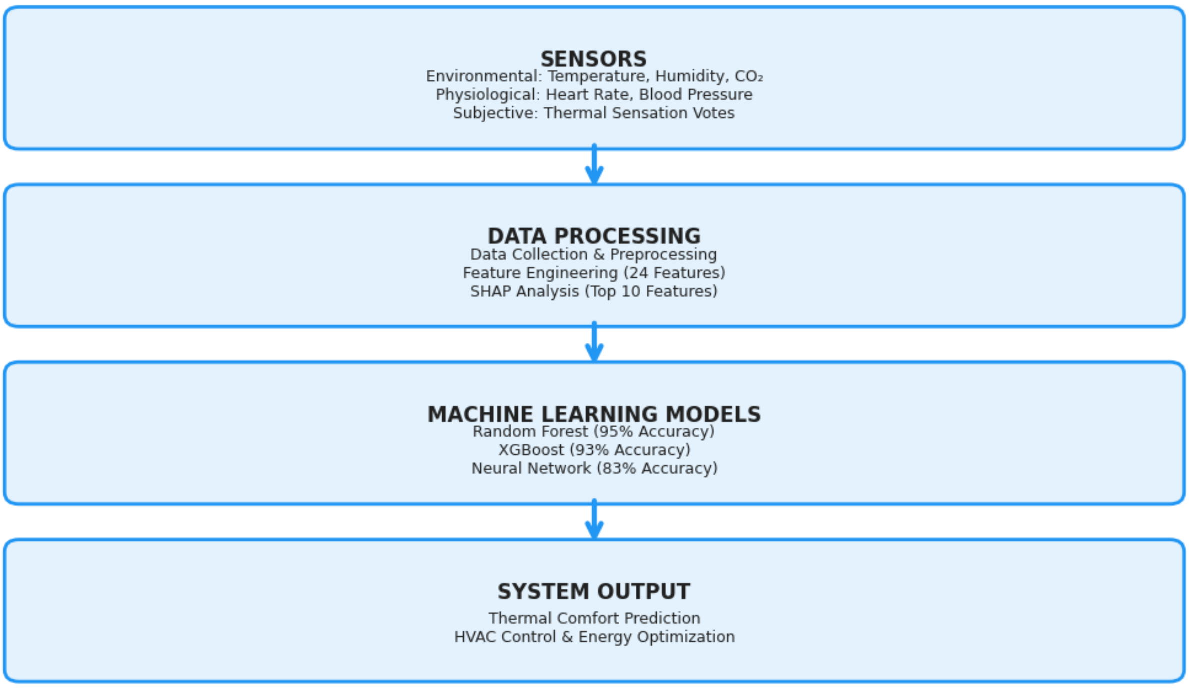

Figure 1 – System architecture for personalized thermal comfort prediction.

2.1. Experimental Setup and Data Collection

Eight healthy male volunteers aged 18-23 years (mean BMI: 24.1 ± 4.2 kg/m²) participated in controlled thermal comfort experiments conducted in a purpose-built laboratory facility. The experimental protocol was designed to expose participants to systematic thermal discomfort conditions while continuously monitoring both environmental parameters and physiological responses. All participants provided informed consent and wore standardized clothing (T-shirt and trousers) to maintain consistent thermal insulation throughout the experiments.

The laboratory setup consisted of two isolated chambers: a baseline comfort room maintained at 21-22°C with CO_2 levels of 500-1000 ppm, and an experimental room where thermal conditions were systematically varied. Four distinct experimental scenarios were implemented: cold discomfort (14-16°C) with moderate CO_2 (500-1200 ppm), cold discomfort with elevated CO_2 (1500+ ppm), hot discomfort (30-32°C) with moderate CO_2, and hot discomfort with elevated CO_2. Each experimental session comprised a 12-minute baseline phase followed by four 36-minute exposure trials corresponding to these conditions.

Environmental and physiological data were collected using a comprehensive sensor network integrated through a centralized monitoring system. Environmental parameters included air temperature, relative humidity, CO_2 concentration, and outdoor temperature, measured continuously at 1 Hz sampling rate using calibrated sensors (Xiaomi Qingping Air Monitor systems with ±0.3°C temperature accuracy and ±50 ppm CO_2 precision). Physiological monitoring encompassed heart rate variability via Polar H10 chest strap monitors, blood pressure measurements using automated upper-arm cuffs, and blood oxygen saturation through fingertip pulse oximeters. Participants provided thermal sensation votes every six minutes using the standard ASHRAE seven-point scale (-3: cold to +3: hot), synchronized with physiological measurements to ensure temporal alignment of all data streams.

Table 1 – Sensor specifications and measurement parameters.

Parameter Type	Sensors Used	Measurements	Accuracy	Sampling Rate
Environmental	Xiaomi Qingping CGS2 Pro	Temperature, Humidity, CO_2, PM2.5	±0.3°C, ±3% RH, ±50 ppm	1 Hz
	Aqara Temperature & Humidity Sensor	Temperature, Humidity, Pressure	±0.3°C, ±3% RH, ±0.12 kPa	1 Hz
Physiological	Polar H10 Heart Rate Monitor	Heart rate, HRV metrics	±1 bpm, ECG-comparable	1 Hz
	Automated Blood Pressure Cuff	Systolic/Diastolic pressure	±3 mmHg	Per trial
	Fingertip Pulse Oximeter	Blood oxygen saturation	±2%	1 Hz

2.2. Data Processing and Feature Engineering

Raw sensor data underwent systematic preprocessing to address temporal misalignment and varying sampling frequencies across different measurement systems. High-frequency environmental signals were processed using windowed averaging to reduce noise and harmonize sampling rates, while sparsely sampled physiological parameters (blood pressure, oxygen saturation) were interpolated using piecewise cubic splines to maintain signal continuity. This approach preserved essential signal characteristics while enabling integration across diverse data streams.

Feature extraction generated 24 distinct variables encompassing environmental conditions (temperature, humidity, CO_2, outdoor temperature), heart rate variability metrics (AVNN, SDNN, rMSSD, pNN50, LF, HF, LF/HF ratio, Alpha_1), physiological parameters (heart rate, blood pressure, oxygen saturation), anthropometric characteristics (age, BMI, weight, body composition), and subjective thermal sensation votes. All features were normalized using min-max scaling according to Equation (1) to ensure consistent input ranges for machine learning algorithms:

$$x' = (x - x\_min) / (x\_max - x\_min) \tag{1}$$

The final dataset comprised 1,536 samples with complete feature vectors, randomly partitioned into training (1,148 samples), testing (288 samples), and validation (100 samples) subsets using stratified sampling to maintain class distribution across thermal sensation categories.

2.3. Machine Learning Models and Evaluation

Three advanced classification algorithms were implemented and compared for thermal comfort prediction: Random Forest, XGBoost, and Artificial Neural Network. Random Forest employed ensemble decision trees with bootstrap aggregation to enhance generalization and reduce overfitting. XGBoost utilized gradient boosting with regularization techniques for optimized performance and computational efficiency. The neural network architecture featured five fully connected layers (256, 128, 64, 64, 32 neurons) with batch normalization, ReLU activation, and 30% dropout, trained using Adam optimizer with 0.001 learning rate and cross-entropy loss function.

Hyperparameter optimization was conducted through exhaustive grid search combined with 10-fold stratified cross-validation. Model performance evaluation employed standard classification metrics including accuracy, precision, recall, and F1-score, computed both per-class and macro-averaged across all seven thermal sensation levels. Model interpretability was achieved through SHAP (SHapley Additive exPlanations) analysis, which quantifies individual feature contributions to predictions using cooperative game theory principles:

$$\phi_i = \sum_{S \subseteq N\{i\}} |S|!\,(|N| - |S| - 1)!\,|N|!\left[f_{S\cup\{i\}}\left(x_{S\cup\{i\}}\right) - f_S(x_S)\right] \tag{2}$$

where φ_i represents the SHAP value for feature i, N is the complete feature set, S denotes feature subsets, and f represents the model's expected output. This analysis identified the most influential predictors and enabled dimensionality reduction by retraining models using only the top-ranked features, thereby improving both computational efficiency and model interpretability for practical deployment scenarios.

3. Results

The machine learning models were evaluated through comprehensive performance analysis, feature importance assessment, and independent validation to determine their suitability for practical thermal comfort prediction applications. The results demonstrate significant potential for accurate personalized climate control in smart building environments.

3.1. Model Performance Comparison

The experimental dataset yielded 1,536 complete samples with 24 features each, which were systematically evaluated across three machine learning algorithms. Initial model training on the full feature set demonstrated strong predictive performance across all approaches, with XGBoost achieving the highest accuracy of 91%, followed closely by Random Forest at 90% and the Artificial Neural Network at 89%. All models showed robust performance in distinguishing between the seven thermal sensation levels, with macro-averaged F1-scores ranging from 0.88 to 0.90.

Feature importance analysis through SHAP revealed that dimensionality reduction significantly enhanced ensemble model performance while reducing computational requirements. When retrained using only the top 10 most influential features, Random Forest accuracy improved from 90% to 94%, and XGBoost performance increased from 91% to 93%. The threshold of ten features was selected based on the SHAP summary analysis, which revealed a distinct 'elbow point' in feature importance distributions; features ranked below this threshold provided negligible predictive gain while increasing computational complexity. Conversely, the neural network showed decreased performance (89% to 83%), suggesting greater dependency on feature interactions that were lost during dimensionality reduction. Based on superior performance after feature selection, the Random Forest model was selected for final validation using 10 selected features and was subsequently validated on an independent hold-out dataset of 100 samples, achieving a final accuracy of 95% with macro-averaged F1-score of 0.939.

Table 2 – Comparative performance of machine learning models for thermal comfort prediction.

Model	Full Dataset (24 features)		Reduced Dataset (10 features)		Accuracy Change
	Accuracy	F1-Score	Accuracy	F1-Score	
Random Forest	0.90	0.89	0.94	0.94	+4%
XGBoost	0.91	0.90	0.93	0.92	+2%
Neural Network	0.89	0.88	0.83	0.80	-6%

3.2. Feature Importance Analysis

SHAP analysis identified indoor air temperature as the dominant predictor across all models, exhibiting approximately 2-3 times greater influence than any other single feature (Figure 2). The top five most influential parameters consistently included environmental factors (temperature, humidity, CO_2 concentration, outdoor temperature) and one physiological parameter (diastolic blood pressure). While all three models showed similar feature ranking patterns, the Random Forest model demonstrated the most stable performance improvement with reduced features, making it the optimal choice for deployment. Environmental parameters dominated the feature rankings, with indoor temperature alone accounting for approximately 30-35% of the total predictive power.

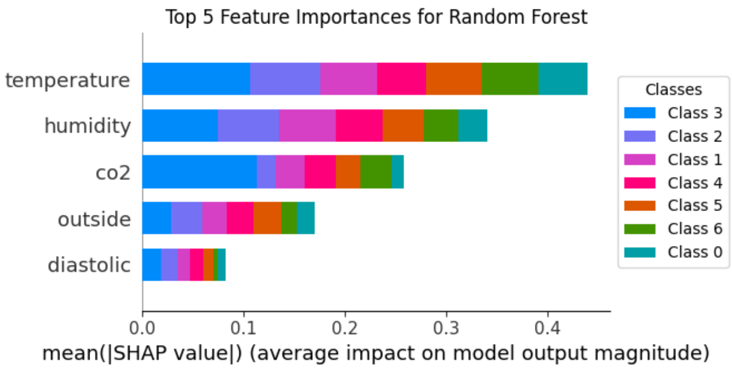

Figure 2 – Feature importance ranking for thermal comfort prediction based on SHAP analysis.

3.3. Model Validation and Practical Performance

The selected Random Forest model (chosen based on its superior performance with reduced features) demonstrated excellent classification performance with minimal misclassification errors. Analysis of the confusion matrix revealed that all prediction errors occurred between adjacent thermal sensation categories (e.g., confusing "neutral" with "slightly cool"), indicating that the model captures the underlying thermal comfort continuum effectively. No instances of extreme misclassification (e.g., predicting "hot" when actual sensation was "cold") were observed, suggesting robust model behavior suitable for practical HVAC control applications.

Final validation on the independent test set confirmed the model's generalization capability, achieving 95% accuracy with perfect classification of neutral and hot thermal sensations. The five misclassifications that occurred were exclusively between neighboring comfort levels, demonstrating that the system maintains reliable performance even on completely unseen data. These results indicate strong potential for real-world deployment in smart building systems, where such accuracy levels would enable precise climate control while minimizing energy waste through unnecessary heating or cooling adjustments. The classification performance is visualized in Figure 3, which shows the confusion matrix for the best-performing model.

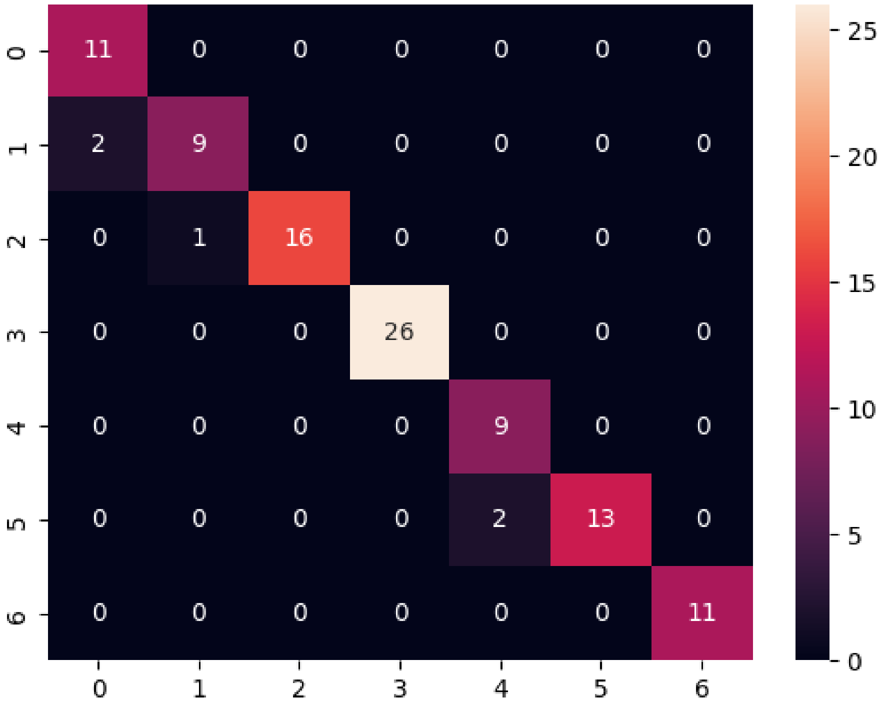

Figure 3 – Confusion matrix for optimized Random Forest model showing 95% accuracy on independent validation dataset. Perfect classification achieved for neutral and hot sensations, with remaining errors limited to adjacent categories.

4. Discussion

The results demonstrate that machine learning approaches can effectively predict personalized thermal comfort with high accuracy, achieving performance levels suitable for practical smart building applications. The 95% accuracy obtained by the optimized Random Forest model represents a significant advancement over traditional static HVAC control systems, which typically achieve occupant satisfaction rates of only 70-80% [5].

The dominance of indoor air temperature as the primary predictor aligns with established thermal comfort theory, while the significant contribution of humidity and CO_2 concentration highlights the importance of comprehensive environmental monitoring. Notably, the relatively modest influence of heart rate variability parameters challenges previous research emphasis on HRV-based comfort assessment, suggesting that when comprehensive environmental and physiological data are available, HRV provides complementary rather than primary predictive information.

The superior performance of ensemble methods (Random Forest and XGBoost) over neural networks, particularly after dimensionality reduction, indicates that thermal comfort prediction benefits more from robust feature selection than complex nonlinear transformations. The superior performance of ensemble methods over neural networks is also attributable to the dataset size (1,536 samples). Deep learning architecture typically requires significantly larger data volumes to establish complex feature representations and avoid overfitting, whereas ensemble tree-based algorithms demonstrated superior robustness on this tabular dataset. This finding has practical implications for deployment in resource-constrained building automation systems, where computational efficiency is crucial.

4.1. Limitations

While the proposed system demonstrates high predictive accuracy, it is important to acknowledge certain limitations regarding participant demographics. The experimental data was collected exclusively from healthy male participants aged 18–23 years. Since thermal comfort perception is known to vary significantly across gender, age groups, and metabolic rates, the current model may not immediately generalize to broader populations, such as females or elderly occupants. Practical deployment in diverse environments would require transfer learning strategies or expanded data collection to adapt the model to these specific demographic groups.

5. Conclusions

This study successfully demonstrated the feasibility of accurate personalized thermal comfort prediction using machine learning and multimodal sensor fusion. The optimized Random Forest model achieved 95% accuracy using only 10 key features, with indoor air temperature identified as the dominant predictor. The system shows strong potential for integration into smart building automation, enabling dynamic climate control that adapts to individual preferences while optimizing energy consumption. Future work should focus on expanding participant diversity and implementing real-time HVAC control systems to validate energy savings potential in operational buildings.

Funding

This research was funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP23488794).

Author Contributions

Conceptualization, A.B. and B.A.; Methodology, A.B.; Software, A.B.; Validation, A.B. and B.A.; Formal Analysis, A.B.; Investigation, A.B.; Resources, B.A.; Data Curation, A.B.; Writing – Original Draft Preparation, A.B.; Writing – Review & Editing, A.B. and B.A.; Visualization, A.B.; Supervision, B.A.; Project Administration, B.A.; Funding Acquisition, B.A.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. International Energy Agency, "Energy Efficiency in Buildings: Technology Roadmap," IEA Publications, Paris, France, 2023. [Online]. Available: https://www.iea.org/reports/energy-efficiency-in-buildings
2. D. Clements-Croome, "Creating the Productive Workplace: Places to Work Creatively," 3rd ed. London, UK: Routledge, 2018, doi: 10.4324/9781315693842.
3. ASHRAE Standard 55-2020, "Thermal Environmental Conditions for Human Occupancy," American Society of Heating, Refrigerating and Air-Conditioning Engineers, Atlanta, GA, USA, 2020.
4. P. O. Fanger, "Thermal Comfort: Analysis and Applications in Environmental Engineering," Danish Technical Press, Copenhagen, Denmark, 1970.
5. S. Karjalainen, "Thermal comfort and gender: A literature review," *Indoor Air*, vol. 22, no. 2, pp. 96-109, Apr. 2012, doi: 10.1111/j.1600-0668.2011.00747.x.
6. J. Kim, R. Schiavon, and G. Brager, "Personal comfort models – A new paradigm in thermal comfort for occupant-centric environmental control," *Building and Environment*, vol. 132, pp. 114-124, Mar. 2018, doi: 10.1016/j.buildenv.2018.01.023.
7. A. Ghahramani, C. Tang, and B. Becerik-Gerber, "An online learning approach for quantifying personalized thermal comfort via adaptive stochastic modeling," *Building and Environment*, vol. 92, pp. 86-96, Oct. 2015, doi: 10.1016/j.buildenv.2015.04.017.
8. W. Wang, J. Chen, T. Hong, and N. Zhu, "Occupancy prediction through Bluetooth beacon technology in commercial buildings," *Energy and Buildings*, vol. 203, pp. 109463, Nov. 2019, doi: 10.1016/j.enbuild.2019.109463.
9. B. Dong, D. Prakash, F. Feng, and Z. O'Neill, "A review of smart building sensing system for better indoor environment control," *Energy and Buildings*, vol. 199, pp. 29-46, Sep. 2019, doi: 10.1016/j.enbuild.2019.06.025.
10. T. Chaudhuri, Y. C. Soh, H. Li, and L. Xie, "A feedforward neural network based indoor-climate control framework for thermal comfort and energy saving in buildings," *Applied Energy*, vol. 248, pp. 44-53, Aug. 2019, doi: 10.1016/j.apenergy.2019.04.065.
11. C. Miller, Z. Nagy, and A. Schlueter, "A review of unsupervised statistical learning and visual analytics techniques applied to performance analysis of non-residential buildings," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 1365-1377, Jan. 2018, doi: 10.1016/j.rser.2017.05.124.
12. M. Földváry Ličina, T. Cheung, H. Zhang, R. de Dear, T. Parkinson, E. Arens, C. Chun, S. Schiavon, M. Luo, G. Brager, P. Li, S. Kaam, M. A. Adebamowo, M. M. Andamon, F. Babich, C. Bouden, H. Bukovianska, C. Candido, B. Cao, S. Carlucci, D. K. W. Chow, S. Damiati, L. Ding, S. Domjan, S. Gauthier, S. Godithi, H. He, D. Heidari, C. Inard, J. Jin, H. Khattar, J. Kim, M. Kong, M. K. Singh, A. Kwok, R. Lamberts, M. Laska, K. C. Parsons, N. C. Popiolek, S. Rupp, R. Sakulpipatsin, M. Schweiker, M. Standeven, R. A. Stanton, S. Tanabe, K. W. Tham, D. Tian, L. Weiss, Z. Wen, T. Xu, Y. Zhai, Y. Zhang, X. Zhou, B. Zhu, and L. Zhu, "Development of the ASHRAE Global Thermal Comfort Database II," *Building and Environment*, vol. 142, pp. 502-512, Oct. 2018, doi: 10.1016/j.buildenv.2018.06.022.
13. R. Kosonen and F. Tan, "The effect of perceived indoor air quality on productivity loss," *Energy and Buildings*, vol. 36, no. 10, pp. 981-986, Oct. 2004, doi: 10.1016/j.enbuild.2004.06.005.

14. Z. Wang, L. Zhang, J. Zhao, and Y. He, "Thermal comfort modeling using machine learning algorithms and environmental monitoring in office buildings," *Applied Energy*, vol. 285, pp. 116450, Mar. 2021, doi: 10.1016/j.apenergy.2021.116450.
15. H. B. Gunay, W. O'Brien, I. Beausoleil-Morrison, and S. Gilani, "Development and implementation of an adaptive lighting and blinds control algorithm," *Building and Environment*, vol. 113, pp. 185-199, Feb. 2017, doi: 10.1016/j.buildenv.2016.08.027.
16. Y. Zhou, J. Yu, X. Luo, and C. Cheng, "A comprehensive review of machine learning techniques for thermal comfort prediction in buildings," *Renewable and Sustainable Energy Reviews*, vol. 150, pp. 111435, Oct. 2021, doi: 10.1016/j.rser.2021.111435.
17. S. P. Corgnati, M. Filippi, and S. Vialli, "Perception of the thermal environment in high school and university classrooms: Subjective preferences and thermal comfort," *Building and Environment*, vol. 42, no. 2, pp. 951-959, Feb. 2007, doi: 10.1016/j.buildenv.2005.10.027.
18. L. T. Wong and K. N. K. Fong, "Thermal comfort evaluation of naturally ventilated public housing in Hong Kong," *Building and Environment*, vol. 46, no. 12, pp. 2405-2414, Dec. 2011, doi: 10.1016/j.buildenv.2011.06.002.
19. M. Schweiker, F. Haldi, M. Shukuya, and D. Robinson, "Verification of stochastic models of window opening behaviour for residential buildings," *Journal of Building Performance Simulation*, vol. 5, no. 1, pp. 55-74, Jan. 2012, doi: 10.1080/19401493.2011.567422.
20. S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, CA, USA, Dec. 2017, pp. 4768-4777.

Information About Authors
Bibars Amangeldy – Researcher, DigitAlem LLP, Almaty, Kazakhstan. ORCID iD: 0000-0002-4089-6337
Bauyrzhan Abilda – Researcher, Astana International University, Astana, Kazakhstan. ORCID iD: 0009-0007-9758-9657.

Submission received: 22 August, 2025.
Revised: 17 December, 2025.
Accepted: 17 December, 2025.