A Novel Feature Selection Method Based on an Improved Beluga Whale Optimization Algorithm for Thermal Comfort Prediction

C. Premila Rosy
Department of Computer Science, Idhaya College for Women, Affiliated to Bharathidasan University, Kumbakonam, India
premilarosy@yahoo.com

Saravanan K
Department of Computer Science, Periyar Maniammai Institute of Science and Technology, Vallam, Thanjavur, India
ks_tnj@yahoo.co.in

S. Josephine Theresa
Department of Computer Science, St.Joseph's College (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli, India
josephinetheresa_cs2@mail.sjctni.edu (corresponding author)

D. Meenakshi
Department of Computer Science, Periyar Maniammai Institute of Science and Technology, Vallam, Thanjavur, India
saravananmeena89@gmail.com

D. Geethamani
Department of Computer Science, Dr. N. G. P. College of Arts and Science, Coimbatore, India
geethumalar@gmail.com

Received: 14 January 2026 | Revised: 22 February 2026, 13 March 2026, 20 March 2026, 22 March 2026, and 24 March 2026 | Accepted: 26 March 2026

 DOI: https://doi.org/10.48084/etasr.17550

ABSTRACT

Heating, Ventilation, and Air Conditioning (HVAC) systems consume a significant portion of building energy and do not usually address individual thermal comfort requirements. To overcome this difficulty, the present study proposes a privacy-aware, deep learning-based thermal comfort prediction system using Federated Deep Learning (FDL). The preprocessing of the ASHRAE Global Thermal Comfort Database II was performed through outlier removal, Z-score normalization, categorical encoding, and class imbalance using the Synthetic Minority Oversampling Technique (SMOTE). The Cubic Transverse Mutation Beluga Whale Optimization (CTMBWO) algorithm was employed to effectively identify features without redundancy and enhance robustness. A federated deep neural network was trained on the chosen features without learning raw data. Chicken Swarm Optimization (CSO) was applied to optimize hyperparameters and maximize convergence and accuracy. Five-fold stratified cross-validation achieved a mean of 98.14 with low variance, outperforming baseline models. The proposed model facilitates energy-saving and privacy-conscious HVAC management in intelligent buildings.

Keywords-Heating, Hentilation, Air Conditioning (HVAC); Machine Learning (ML); Deep Neural Network (DNN); building types; Federated Deep Learning (FDL)

I. INTRODUCTION

Improving living standards and indoor comfort has become a priority. Heating, Ventilation, and Air Conditioning (HVAC) systems are used to control the indoor environment and consume considerable energy, typically in the range of 40–50%, while maintaining indoor thermal comfort [1]. It has been shown that 43% of occupants are not satisfied with their thermal comfort, exceeding the acceptable level of 20% outlined in ASHRAE Standard 55 [2]. The most likely source of dissatisfaction in these buildings is overcooling or overheating, causing energy wastage without meeting the comfort needs of individuals [3].

Research into personalized thermal comfort has gained considerable interest because these models can meet the specific comfort needs with effective energy consumption [4]. The specific models can predict the thermal comfort of occupants, thus integrating it with the indoor conditioning systems for sustainable development. Authors in [5] designed a controller to satisfy individual thermal preferences in real office environments. However, its performance and energy savings depend significantly on external weather, which introduces variability and limits prediction capabilities. Authors in [6] introduced a method to help households choose more efficient thermostat setpoints without sacrificing comfort. This method combines the adaptive comfort theory with predictive indoor temperature models to define adaptable setpoints. Accuracy depends on the reliability and granularity of smart thermostat data, which may vary across households.

Owing to the high subjectivity of thermal comfort, the perception of an individual can differ significantly from that of another [7]. This divergence adds difficulty to Machine Learning (ML) models, and outliers may emerge that can potentially skew the results of data-analyzing functions [8]. This may also affect the performance of the ML algorithms. Therefore, optimization becomes necessary to remove outliers and improve the performance of the models. Another issue is the daycare imbalance in the thermal comfort dataset. If occupant responses are not sufficiently represented for all classes in the Thermal Sensation Vote (TSV) category, this results in data imbalance, which may affect model performance [9].

As researchers develop more sophisticated techniques, they encounter additional challenges. One such challenge is the uneven data distribution, where some thermal comfort situations are well represented in the data, whereas others may have insufficient data to develop an accurate model, thus jeopardizing the model's performance. Therefore, it is significant to acquire high-quality data [10]. Two common models for thermal comfort assessment are the adaptive model and the PMV/PPD model [11]. The PMV model, based on thermal equilibrium between the human body and environment, uses chamber experiments in controlled indoor environments. It predicts thermal comfort by considering four environmental and two personal factors. Artificial Neural Networks (ANNs) define nonlinear relationships between inputs and outputs [12] to predict PMV index values for room thermal comfort. While ANN modeling requires multiple iterations without guaranteeing optimal solutions, the present study proposes Chicken Swarm Optimization (CSO) to achieve global Minimum Square Error (MSE) for optimal model parameters. CSO ensures minimum time slots within real-time limits. In addition, the study employs an improved Beluga Whale Optimization (BWO) algorithm [13] to assess software defects. Table I summarizes other related works.

TABLE I. SUMMARY OF RELATED WORKS

Study	Method	Advantages	Limitations
[14]	Transformer-based thermal comfort prediction	High accuracy using temporal attention; suitable for real-time monitoring	Requires large memory and GPU resources for training
[15]	Federated XGBoost for comfort estimation	Privacy-preserving model with reduced communication overhead	Slight drop in accuracy compared to centralized models
[16]	Hybrid feature selection with genetic algorithm and SVM	Enhanced feature relevance; interpretable model	Sensitive to initial population and mutation rate
[17]	LSTM-CNN ensemble for occupant comfort forecasting	Captures spatial-temporal patterns effectively	Performance depends heavily on parameter tuning
[18]	Reinforcement learning for HVAC optimization	Adaptive system control based on occupant feedback	Slow convergence in complex environments

Authors in [19] proposed an energy efficiency strategy for hotel buildings aimed at improving energy consumption while maintaining thermal comfort. The study primarily focused on facade modifications, such as sun shading and glass type. Other potential efficiency strategies, including HVAC system optimization, renewable energy integration, and occupant behavior, were not explored. Authors in [20] optimized the thermal comfort issues in the breathing layer at minimum energy cost. However, they did not focus on other important factors such as system design, integration with HVAC, and long-term performance evaluation. Authors in [21] proposed a thermal prediction model based on skin temperature and environmental factors. The proposed model was compared with a traditional ML algorithm to achieve high accuracy. However, the model did not focus on data imbalance and redundancy. To identify more insightful features, authors in [22] used a supervised forward sequential feature selection strategy. This approach prioritizes features that have a significant impact on predicting occupant thermal comfort.

II. PROPOSED METHODOLOGY

The proposed methodology begins with the ASHRAE II dataset [12], which contains the thermal comfort variables. Data preprocessing includes Z-score normalization, encoding categorical data, and applying the Synthetic Minority Oversampling Technique (SMOTE) to balance the classes. An improved Beluga optimization algorithm was used to select the most relevant features. These features were utilized as input for a Federated Deep Learning (FDL) model for privacy-preserving classification. FDL is capable of distributed training of models without any raw data sharing. CSO was used for performance improvement. The optimized model of the FDL accounts for feature and parameter optimization, thus providing an accurate prediction of thermal comfort. Figure 1 illustrates the workflow for the proposed methodology.

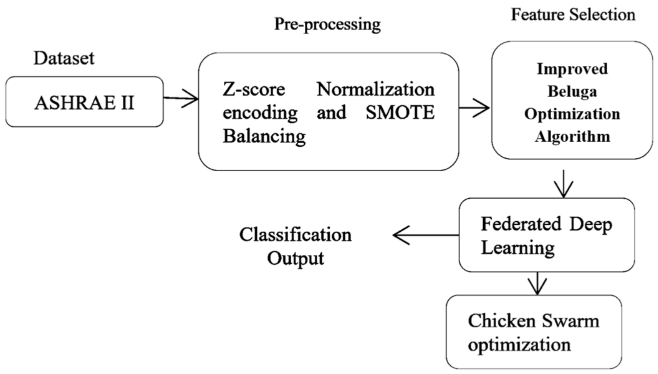

Fig. 1. Workflow of the proposed methodology.

A. *Dataset Preprocessing*

Raw data were obtained from the ASHRAE Global Thermal Comfort Database II, an open-access repository that collects the results of field studies from all over the world from 1995 to 2018. This database contains approximately 81,846 entries with corresponding records of subjective thermal comfort responses, along with objective environmental and physiological measurements. This dataset has been used to establish relationships between objective measurements and subjective thermal comfort. However, studies have been constrained by the size of the dataset, its relevance, and the practicality of implementation. Experimental data collected from real-world field studies suffered from noise, which requires preprocessing. The raw dataset contains outliers, missing values, or inconsistencies in the series.

Outliers were detected using a Z-score–based statistical method, suitable for continuous environmental and physiological variables in large-scale field datasets. For each numerical feature *x*, the Z-score was computed as:

$$Z = \frac{(x-\mu)}{\sigma}$$

where μ and σ represent the mean and standard deviation of the feature, respectively.

B. *Feature Selection*

With an increasing number of features, the BWO-based selection algorithm tends toward local optima and an uneven distribution of the population. While performing feature selection in a crowded area, there is a high probability of being trapped in a local optimum. In contrast, Cubic Transverse Mutation Beluga Whale Optimization (CTMBWO) presents a favorably distributed population such that all the populations are considered, thus providing a global optimum solution instead of a local one. Table II outlines the data availability of variables in the ASHRAE Global Thermal Comfort Database II in terms of percentage values calculated on the entire raw data. Table III provides the parameters used in the CTMBWO Algorithm.

TABLE II. AVAILABILITY AND MISSING-VALUE DISTRIBUTION OF VARIABLES IN THE ASHRAE II DATABASE

Category	Variable	Unit	Valid data (%)	Missing data (%)
Personal information	Weight	kg	30.1	69.9
	Height	cm	24.9	75.1
	Sex	–	82.0	18.0
	Age	years	53.3	46.7
Indoor environmental	Tair (air temperature)	°C	90.6	9.4
	Vair (air velocity)	m/s	81.1	18.9
	RH (relative humidity)	%	89.3	10.7
Thermal comfort factors	Met (metabolic rate)	W/m²	82.9	17.1
	Clo (clothing insulation)	m²·K/W	90.3	9.7
Outdoor environmental	Tout (outdoor temperature)	°C	72.9	27.1

TABLE III. PARAMETER USED IN CTMBWO ALGORITHM

Parameter	Typical range	Impact on optimization
Population size (N)	20–50	Larger values improve exploration but increase cost
Maximum iterations (T_{max})	50–200	Controls convergence depth
Step-size scaling factor (λ)	0.1–1.0	Affects exploration strength
Fitness weighting factor (α)	0.5–0.9	Balances accuracy vs. feature reduction
Random coefficients (r_1, r_2)	(0,1)	Introduces stochastic behavior
Mutation strength (δ)	0.01–0.1	Prevents local optimum trapping
Feature dimension (d)	Problem-dependent	Defines search space size

The step-by-step procedure of CTMBWO is:

- Initialization: An initial population of Beluga whales is generated using chaotic mapping to enhance diversity.
- Fitness evaluation: The fitness of each whale is computed using the proposed objective function.

- Exploration phase: Whale positions are updated using the exploration equation for a global search.
- Exploitation phase: Triangle wandering is applied to refine the solutions around promising regions.
- Diversity enhancement: Cauchy mutation and reverse learning are introduced to avoid premature convergence.
- Signal warning mechanism: Poorly performing individuals share directional cues to guide the population away from unfavorable regions.
- Termination: The same procedure is repeated until the maximum number of iterations is reached or the best feature subset is obtained.

C. Cauchy Mutation

Cauchy mutation and reverse-learning techniques are employed in BWO to help prevent entrapment in local optima. Target positions change after iterations with fitness recalculation. To avoid local optimum traps, Cauchy mutation and reverse learning randomly update positions with probability P. A mathematical model for the whale fall stage was provided in [13]. Position update is applied as:

$$X_i^{t+1} = X_i^t + \lambda \cdot r_1 \cdot (X_k^t - X_i^t) \quad (1)$$

where X_i^t is the position of the i^{th} whale at iteration t, X_k^t is a randomly selected whale, $r_1 \in (0,1)$ is a random number, and λ is the step-size scaling factor.

Adaptive probability for applying the hybrid strategy is defined as:

$$P_s = exp\,(1 - \frac{T}{T_{max}})^{10} + (t) \quad (2)$$

where T is the current iteration number, T_{max} is the maximum number of iterations, and ω is the adjustment factor (optimal when $\omega = 0.05$).

The reverse-learning strategy is formulated as:

Reverse-learning equation:

$$X_{best}^T = u_b + r(l_b - X_{best}^T) \quad (3)$$

where X_{new}^T is the best solution at iteration T, $X_{best}^{T^*}$ is a reverse solution, and $r \in [0,1]$ is the random number.

The new position of the candidate generation (hybrid strategy) is given by:

$$\begin{cases} X_{best}^{T+1} = b_3(X_{new}^T - X_{best}^{T^*} \\ Xx_{new}^{T+1} = Cauchy(0,1)X_{best}^{T+1} \end{cases} \quad (4)$$

The pseudo-information exchange coefficient is defined as:

$$\frac{\alpha \cdot exp(-\beta T)}{T_{max}}$$

D. Signal Warning in Sparrows

The alert sparrow mechanism conceptualizes a relevant behavior in which a dead whale induces failed individuals to convey helpful cues to other individuals, to avoid risky environments by imitating an alert behavior. This information encompasses the search space surrounding failed individuals, ensuring that other individuals are not trapped. The position update is defined as:

$$X_i^{T+1} = X_i^T + \alpha(X_{new}^{T+1} - X_i^T) + \beta R \quad (5)$$

where X_i^T is the position of the individual i at generation $T, X_i^{(T+1)}$ is the updated position of the individual i at generation $T+1, X_{new}^{T+1}$ is the target (or updated best) solution at generation $T+1, \alpha$ is the scaling factor controlling the movement direction toward the target solution, β is the scaling factor controlling the scope of the randomized search, and R is the random vector (usually drawn from a uniform or Gaussian distribution) used to introduce stochasticity and increase diversity in the search. This allows the algorithm to diffuse local knowledge from participants who perform poorly, encourage exploration by intentionally adding some randomization, and prevent premature convergence because of the diversity of search directions.

Algorithm 1: CTMBWO algorithm

```
Input: Objective function f(x), Beluga
whale population size N, Maximum number of
iterations Tmax
Output: The global optimal solution
Initialize the beluga whale population X =
{x1, x2, ..., xN}
Evaluate the fitness of each individual
using (5) and (6)
 Set iteration counter T = 0
while T < Tmax
           Calculate the beluga factor Bf
          if Bf > 0.5
       Enter the exploration phase and
update the position using (1)
         Else
       if Bf < Wf 0.5
                  Calculate parameters
P, Ps, CS, Xstep, Update position
                  using (2,4 and 5)
          if Ps > P
              Use the Cauchy mutation
strategy to calculate X_new^(T+1)
          Else
               Use a reverse learning
strategy to calculate X_new^(T+1)
               Assist other individuals in
optimizing their positions using (5)
               Increment iteration T = T + 1
    Else
         Increment iteration T = T + 1
End while
```

E. Classification

The substantial traction across various sectors over the past decade increased data availability and computing power, leading to more efficient deep-learning algorithms, enabling innovations such as smart buildings, autonomous vehicles, and intelligent systems. These algorithms require large data volumes that are typically stored on central servers, thereby creating challenges. Servers are vulnerable to attacks owing to the increased bandwidth and remote database systems. Processing vast amounts of data on single servers is time-consuming, and the connected devices may collect sensitive information that requires confidentiality under regulations.

Within a distributed framework, the FDL method considers M-permitted clients using m as the index for each client to analytically determine the fundamental loss function, as shown in:

$$\min_{\theta} l(\theta) = \sum_{m=1}^{M} \frac{n_m}{n} K_m(\theta) \quad (6)$$

where each client's local sample is n_m, and $K_m(\theta)$ can be written as:

$$K_m(\theta) = \frac{1}{n_m} \sum_{j \in R_k} L_q(\theta) \quad (7)$$

The dataset was organized using R_k with a size of $n \times m$. The FDL framework comprises a server and two clients in four stages. Clients first conducted basic training using the server-provided parameter, k. During model aggregation, the server uses parameters u and k from the clients for aggregation. The server then transmits the updated parameter t_l to the clients for retraining. Finally, clients revise the models using the combined parameters and assess their effectiveness.

III. RESULTS AND DISCUSSION

The study employed Python version 3.8 in conjunction with advanced deep learning frameworks, including PyTorch 1.10 and Tensor Flow 2.6, to effectively develop and optimize the models. Data processing and analysis were performed using Pandas 1.3.3, NumPy 1.21.2, and visual representations were created using Matplotlib 3.4.3 and Seaborn 0.11.2. The computational environment was bolstered by a high-performance workstation, featuring an NVIDIA RTX 3090 GPU with 24GB VRAM, 64GB DDR4 RAM, and an Intel Core i9-11900K processor tailored to handle intensive computations and large datasets. The parameters and values used in the process are listed in Table IV. This setup facilitated efficient model training, validation, and performance evaluation, thereby enhancing the research process.

TABLE IV. SIMULATION PARAMETER

Parameter	Value
Number of clients	10
Communication rounds	50
Local epochs	5
Optimizer	Adam
Learning rate	0.001
Batch size	32
Data sharing	No raw data sharing
Aggregation method	FedAvg

The proposed FDL system has a dataset spread among 10 federated clients, one client per building or zone. The data were divided according to a non-IID strategy, in which samples of a client have different temperature preference distributions and environmental conditions, representing the heterogeneity in real-world buildings. After cleaning, the dataset contained 15,162 samples: 28.60% (4,337) preferred warmer conditions, 56.10% (8,508) showed no change in preference, and 15.20% (2,317) preferred cooler conditions. The left panel shows the raw frequency distributions of variables such as temperature, humidity, and air velocity, which exhibit varying scales and skewness. After applying Z-score normalization, the right panel reveals a more uniform distribution centered around zero, indicating that the features have been rescaled to have comparable statistical properties. Table V presents comparisons among multiple feature selection strategies, including the proposed CTMBWO method and the traditional correlation-based methods.

TABLE V. PERFORMANCE COMPARISON OF FEATURE SELECTION TECHNIQUES IN THERMAL COMFORT PREDICTION

Feature selection	Model	Kappa	Accuracy	Package C
CTMBWO (proposed)	FDL	0.7823	0.9814	0.9617
	Random forest	0.6581	0.9275	0.8928
	XGBoost	0.6934	0.9402	0.9168
	SVM	0.6217	0.9013	0.8615
	BiLSTM	0.6419	0.9156	0.8787
Correlation-based method	FDL	0.7489	0.9513	0.9367
	Random forest	0.6342	0.9132	0.8737
	SVM	0.5984	0.8841	0.8412
	XGBoost	0.6483	0.9201	0.8842

Table VI compares the performance of the FDL-CSO model with that of traditional and ensemble classifiers. The FDL-CSO approach, which integrates FDL with CSO for hyperparameter tuning, outperformed all other models across all evaluation metrics. With an accuracy rate of 98%, a precision of 97%, a recall of 99%, and an F1-score of 98%, this system proved its strength in prediction and reliability. In contrast, Light GBM and gradient boosting showed strong performance, but fell short of the proposed method, particularly in recall. Basic models, such as logistic regression and decision trees, performed relatively poorly, emphasizing the effectiveness of the proposed hybrid optimization and federated framework for thermal comfort prediction.

TABLE VI. COMPARATIVE PERFORMANCE OF PROPOSED FDL-CSO WITH OTHER MODELS

Model	Accuracy	Precision	Recall	F1-score
FDL-CSO (proposed)	0.98	0.97	0.99	0.98
Light GBM	0.92	0.91	0.93	0.92
Gradient boosting	0.89	0.88	0.90	0.89
Ada Boost	0.84	0.85	0.83	0.84
Decision tree	0.81	0.79	0.83	0.81
Logistic regression	0.76	0.77	0.74	0.75

IV. CONCLUSION

This study presented a privacy-aware deep learning-based thermal comfort prediction system using the Federated Deep Learning (FDL) technique. The selection of parameters in the thermal comfort data largely influenced the efficiency of various Machine Learning (ML) techniques. Using Cubic Transverse Mutation Beluga Whale Optimization (CTMBWO), the model minimized redundancies and allowed the selection of the most informative features, thus significantly enhancing classification efficiency. This learning architecture allows for decentralized training while securing data privacy, with local clients training on their own data and sharing only model-parameter updates. Such a decentralized environment is an added advantage owing to its enhanced scalability for real-life smart building applications.

The hyperparameter tuning using Chicken Swarm Optimization (CSO) ensured near-optimal predictive performance in a deep learning model with the least manual input. Experimental validation over the ASHRAE II Global Thermal Comfort Database confirmed the success of the proposed approach. The results further proved that the CTMBWO-FDL-CSO combination had the best prediction accuracy and computational efficiency. Further studies could integrate real-time IoT sensor streams with adaptive federated learning to cater to dynamic occupant behavior. Cross-building transfer learning can improve the model generalization over various climatic zones.

DECLARATION OF COMPETING INTERESTS

The authors declare that they have no known competing financial or non-financial interests.

ACKNOWLEDGMENT

Not applicable to this work.

DATA AVAILABILITY

The ASHRAE II global thermal comfort dataset used in this study was obtained from [12].

REFERENCES

[1] H. Wang, D. Mai, Q. Li, and Z. Ding, "Evaluating Machine Learning Models for HVAC Demand Response: The Impact of Prediction Accuracy on Model Predictive Control Performance," *Buildings*, vol. 14, no. 7, Jul. 2024, Art. no. 2212, https://doi.org/10.3390/buildings14072212.

[2] J. Liu, Z. Zhai, Y. Zhang, Y. Wang, and Y. Ding, "Comparison of Energy Consumption Prediction Models for Air Conditioning at Different Time Scales for Large Public Buildings," *Journal of Building Engineering*, vol. 96, Nov. 2024, Art. no. 110423, https://doi.org/10.1016/j.jobe.2024.110423.

[3] A. Maboudi Reveshti, E. Khosravirad, A. K. Rouzbahani, S. K. Fariman, H. Najafi, and A. Peivandizadeh, "Energy Consumption Prediction in an Office Building by Examining Occupancy Rates and Weather Parameters Using the Moving Average Method and Artificial Neural Network," *Heliyon*, vol. 10, no. 4, Feb. 2024, Art. no. e25307, https://doi.org/10.1016/j.heliyon.2024.e25307.

[4] T. N. Dinh, G. S. Thirunavukkarasu, M. Seyedmahmoudian, S. Mekhilef, and A. Stojcevski, "Robust-mv-M-LSTM-CI: Robust Energy Consumption Forecasting in Commercial Buildings during the COVID-19 Pandemic," *Sustainability*, vol. 16, no. 15, Aug. 2024, Art. no. 6699, https://doi.org/10.3390/su16156699.

[5] H. Zhang, A. Tzempelikos, X. Liu, S. Lee, F. Cappelletti, and A. Gasparella, "The Impact of Personal Preference-Based Thermal Control on Energy Use and Thermal Comfort: Field Implementation," *Energy and Buildings*, vol. 284, Apr. 2023, Art. no. 112848, https://doi.org/10.1016/j.enbuild.2023.112848.

[6] R. Almeida, P. Georgieva, and N. Martins, "Heating Setpoint Recommendation Strategy for Thermal Comfort and Energy Consumption Optimization," *Energy and Buildings*, vol. 296, Oct. 2023, Art. no. 113406, https://doi.org/10.1016/j.enbuild.2023.113406.

[7] Q. Carton, J. K. Møller, M. Favero, D. Calì, J. Kolarik, and H. Breesch, "Predicting Individual Thermal Preferences in an Office: Assessing the Performance of Mixed-Effects Models," *Building and Environment*, vol. 261, Aug. 2024, Art. no. 111751, https://doi.org/10.1016/j.buildenv.2024.111751.

[8] Y. Wu, J. Fan, and B. Cao, "Comparison Among Different Modeling Approaches for Personalized Thermal Comfort Prediction When Using Personal Comfort Systems," *Energy and Buildings*, vol. 285, Apr. 2023, Art. no. 112873, https://doi.org/10.1016/j.enbuild.2023.112873.

[9] H. Zhang, S. Lee, and A. Tzempelikos, "Bayesian Meta-Learning for Personalized Thermal Comfort Modeling," *Building and Environment*, vol. 249, Feb. 2024, Art. no. 111129, https://doi.org/10.1016/j.buildenv.2023.111129.

[10] Y. Su, X. Zou, M. Tan, H. Peng, and J. Chen, "Integrating Few-Shot Personalized Thermal Comfort Model and Reinforcement Learning for HVAC Demand Response Optimization," *Journal of Building Engineering*, vol. 91, Aug. 2024, Art. no. 109509, https://doi.org/10.1016/j.jobe.2024.109509.

[11] C. Gu, J. J. Cao, and B. W. Wang, "Specific Emitter Identification Based on Hybrid Feature Selection," *Wireless Communication and Mobile Computing*, vol. 51, no. 5, 2024, Art. no. 267, https://doi.org/10.11896/jsjkx.230300216.

[12] S. Zhang, R. Yao, C. Du, E. Essah, and B. Li, "Analysis of Outlier Detection Rules Based on the ASHRAE Global Thermal Comfort Database," *Building and Environment*, vol. 234, Apr. 2023, Art. no. 110155, https://doi.org/10.1016/j.buildenv.2023.110155.

[13] S. Qiu, J. He, Y. Wang, and B. E, "A Feature Selection Method for Software Defect Prediction Based on Improved Beluga Whale Optimization Algorithm," *Computers, Materials & Continua*, vol. 83, no. 3, pp. 4879–4898, 2025, https://doi.org/10.32604/cmc.2025.061532.

[14] K. Chen and A. Ghahramani, "ComfortGPT: A Transformer-Based Architecture for Predicting Preferred Temperature Setpoints Leveraging Big Data," *Building and Environment*, vol. 248, Jan. 2024, Art. no. 111085, https://doi.org/10.1016/j.buildenv.2023.111085.

[15] I. L. Niza, I. M. Da Luz, and E. E. Broday, "Thermal Comfort Assessment in University Classrooms: A Discriminant Analysis for Categorizing Individuals According to Gender and Thermal Preferences," *Atmosphere*, vol. 14, no. 9, Aug. 2023, Art. no. 1325, https://doi.org/10.3390/atmos14091325.

[16] Y. Bai, Z. Dong, and L. Liu, "Hybrid Feature Selection-Based Machine Learning Methods for Thermal Preference Prediction in Diverse Seasons and Building Environments," *Building and Environment*, vol. 269, Feb. 2025, Art. no. 112450, https://doi.org/10.1016/j.buildenv.2024.112450.

[17] Z. Ma, J. Wang, S. Ye, R. Wang, F. Dong, and Y. Feng, "Real-Time Indoor Thermal Comfort Prediction in Campus Buildings Driven by Deep Learning Algorithms," *Journal of Building Engineering*, vol. 78, Nov. 2023, Art. no. 107603, https://doi.org/10.1016/j.jobe.2023.107603.

[18] Y. Boutahri and A. Tilioua, "Reinforcement Learning for HVAC Control and Energy Efficiency in Residential Buildings with BOPTEST Simulations and Real-case Validation," *Discover Computing*, vol. 28, no. 1, Apr. 2025, Art. no. 45, https://doi.org/10.1007/s10791-025-09544-y.

[19] N. Nasrullah and M. A. Hamdy, "Air Conditioning Energy Efficiency and Thermal Comfort in Hotel Buildings in Hot and Humid Tropical Climates: The Case Study of Makassar City Coastal Area Hotel, South Sulawesi, Indonesia," *Engineering, Technology & Applied Science Research*, vol. 14, no. 6, pp. 18290–18299, Dec. 2024, https://doi.org/10.48084/etasr.8463.

[20] M. Kanaan, S. Amine, and E. Gazo-Hanna, "Computational Simulation and Analysis of Local Thermal Comfort and Indoor Air Quality in Space with Displacement Ventilation," *Engineering, Technology & Applied Science Research*, vol. 14, no. 5, pp. 16383–16388, Oct. 2024, https://doi.org/10.48084/etasr.7948.

[21] B. Yang *et al.*, "Comparison of Models for Predicting Winter Individual Thermal Comfort Based on Machine Learning Algorithms," *Building*

and Environment, vol. 215, May 2022, Art. no. 108970, https://doi.org/10.1016/j.buildenv.2022.108970.

[22] R. Zhu, X. Zhang, L. Yang, Y. Liu, Y. Cong, and W. Gao, "Correlation Analysis of Thermal Comfort and Physiological Responses Under Different Microclimates of Urban Park," *Case Studies in Thermal Engineering*, vol. 34, Jun. 2022, Art. no. 102044, https://doi.org/10.1016/j.csite.2022.102044.