THE SCIENTIFIC TEMPER (2025) Vol. 16 (12): 5322-5327
E-ISSN: 2231-6396, ISSN: 0976-8653
Doi: 10.58414/SCIENTIFICTEMPER.2025.16.12.19
https://scientifictemper.com/
RESEARCH ARTICLE

Graph Neural Network Ensemble with Particle Swarm Optimization for Privacy-Preserving Thermal Comfort Prediction

Josephine Theresa S*

Abstract

Heating, ventilation, and air conditioning (HVAC) systems account for nearly 60% of energy consumption in commercial buildings, yet occupant dissatisfaction with thermal comfort remains high. To address this challenge, we propose a novel framework that leverages the ASHRAE Global Thermal Comfort Database II to predict individual thermal preferences while ensuring energy efficiency. Unlike prior deep learning approaches, our method employs a Graph Neural Network (GNN) ensemble with attention mechanisms, enabling the model to capture complex relationships among personal, environmental, and contextual variables across seasons and building types. Feature selection is performed using Particle Swarm Optimization (PSO), which enhances diversity and avoids premature convergence by dynamically updating particle velocities and positions. The selected features are then fed into the GNN ensemble, which integrates multiple graph-based learners to improve robustness. Hyperparameter tuning is conducted using Bayesian Optimization, balancing exploration and exploitation to identify optimal learning rates, dropout ratios, and batch sizes. Experimental results on the ASHRAE dataset demonstrate that the proposed GNN-PSO-Bayesian framework achieves 96.8% accuracy, outperforming traditional classifiers while maintaining interpretability and scalability. This architecture highlights the potential of graph-based learning for occupant-centric thermal comfort prediction, offering a pathway toward sustainable and adaptive HVAC control.

Keywords: Thermal comfort, Graph neural network (GNN), Particle swarm optimization (PSO),Bayesian optimization, HVAC systems

Introduction

Thermal comfort prediction has emerged as a critical research area due to its direct impact on energy efficiency and occupant well-being in modern buildings. Recent advancements in machine learning and deep learning have enabled more accurate modeling of complex thermal environments. For instance, Cho et al. (2024) demonstrated the potential of MH-LSTM neural networks for personalized comfort prediction, while Tan et al. (2024) emphasized geo-specific modeling using the ASHRAE dataset. These studies highlight the growing reliance on data-driven approaches to optimize HVAC systems and reduce energy consumption, underscoring the importance of predictive frameworks in sustainable building design.

A variety of modeling techniques have been explored to capture the nonlinear and dynamic nature of thermal comfort. Zhang et al. (2024) applied transformer-based architectures, achieving high accuracy in real-time monitoring, whereas Gao et al. (2025) introduced deep transfer learning hybrid models to enhance individual comfort prediction. Mane et al. (2025) further integrated neural networks with energy optimization strategies to simultaneously improve indoor air quality and thermal satisfaction. Complementary approaches such as reinforcement learning for HVAC control (Sayed et al., 2024) and digital twin frameworks (ElArwady et al., 2024; Iqbal & Mirzabeigi, 2025) have expanded the scope of predictive modeling, enabling adaptive and context-aware systems.

Despite these advancements, several challenges persist in the representation and utilization of thermal comfort data. Studies have noted issues of data imbalance, noise, and

Assistant Professor, Department of Computer Science, St. Joseph's College(A), Affilated to Bharathidasan University, Tiruchirappalli, India

***Corresponding Author:** Josephine Theresa S, Assistant Professor, Department of Computer Science, St. Joseph's College(A), Affilated to Bharathidasan University, Tiruchirappalli, India, E-Mail: josephinetheresa_cs2@mail.sjctni.edu

How to cite this article: Theresa, J.S. (2025). Graph Neural Network Ensemble with Particle Swarm Optimization for Privacy-Preserving Thermal Comfort Prediction. The Scientific Temper, **16**(12):5322-5327.

Doi: 10.58414/SCIENTIFICTEMPER.2025.16.12.19

Source of support: Nil

Conflict of interest: None.

Received: 20/11/2025 **Accepted:** 05/12/2025 **Published:** 25/12/2025

limited contextual variables. Zhang et al. (2024) highlighted the importance of incorporating adaptive occupant behaviors in mixed-mode buildings, while Uddin et al. (2025) stressed the need for robust feature analysis in educational environments. Similarly, Penuela et al. (2025) addressed data denoising and dynamic prediction using Bayesian machine learning, pointing to the necessity of cleaner and more representative datasets. These findings collectively reveal that while predictive accuracy has improved, the reliability and generalizability of models remain constrained by data quality and feature selection.

Another critical gap lies in privacy-preserving frameworks and model interpretability. Tang et al. (2023) explored federated learning for energy prediction, yet scalability and communication overhead remain concerns. Leeraksakiat and Pora (2020) demonstrated transfer learning for occupancy forecasting, but their approach lacked interpretability in thermal comfort contexts. Moreover, Ma et al. (2023) showed the potential of real-time deep learning models in campus buildings, though these models often require large datasets and computational resources, limiting their applicability in resource-constrained environments. Collectively, these studies highlight the pressing need for frameworks that balance privacy, scalability, and interpretability without sacrificing accuracy.

Building on these advancements and addressing the identified gaps, the present study proposes a novel feature selection and classification framework for thermal comfort prediction using the ASHRAE Global Thermal Comfort Database II. By integrating advanced optimization strategies with privacy-preserving architectures, this work aims to overcome challenges of data imbalance, feature redundancy, and computational overhead. Unlike prior studies that focused on either accuracy or scalability, our approach emphasizes a holistic balance—ensuring high predictive performance, safeguarding occupant data, and maintaining interpretability. In doing so, this research contributes to the development of sustainable, occupant-centric HVAC systems that align with the evolving demands of smart building environments.

Methodology

Figure 1 illustrates The overall workflow of the proposed study begins with the acquisition of thermal comfort data, followed by preprocessing, feature selection, model training, and evaluation. Conceptually, the architecture diagram illustrates a pipeline where raw data from the ASHRAE Global Thermal Comfort Database II enters the preprocessing block, is refined through cleaning and normalization, and then passes into the feature selection stage. The selected attributes are subsequently fed into the classification model, which is tuned and validated before producing final predictions of occupant comfort levels. This sequential flow ensures that each stage contributes to the reliability and accuracy of the final outcome, while maintaining a clear separation of tasks for reproducibility.

The methodology is composed of four major components: data preprocessing, feature selection, classification, and hyperparameter optimization. The preprocessing component addresses issues of missing values, outliers, and imbalanced classes, ensuring that the dataset is statistically consistent and representative. Feature selection is designed to reduce redundancy and highlight the most influential variables, thereby improving computational efficiency and interpretability. The classification component employs a deep learning framework capable of capturing nonlinear relationships between environmental and personal factors. Finally, hyperparameter optimization fine-tunes the learning process, balancing accuracy with generalization. Each component is interconnected, forming a cohesive system that aligns with the study's objective of accurate and privacy-preserving thermal comfort prediction.

Data Acquisition and Justification

The ASHRAE Global Thermal Comfort Database II was chosen as the primary data source because it is a comprehensive, open-access repository containing over 80,000 records collected from diverse climates, building types, and occupant groups between 1995 and 2018. This dataset includes both subjective responses (thermal sensation votes) and objective measurements (temperature, humidity, air velocity, clothing insulation, and metabolic rate). Its breadth and diversity make it suitable for developing generalized models that can adapt across different contexts. To ensure correctness, the data is subjected to rigorous preprocessing: outlier removal, Z-score normalization, categorical encoding, and balancing through synthetic oversampling. These steps guarantee that the dataset used for training and evaluation is both clean and representative, minimizing bias and maximizing reliability.

```
Algorithm ThermalComfortPredictor
Input: RawData ← ASHRAE Global Thermal Comfort Database II
Output: ThermalComfortCategory ∈ {Cool, Neutral, Warm}

Begin
  // Step 1: Data Preprocessing
  CleanedData ← RemoveOutliers(RawData)
Cleaned Data ← Handle Missing Values (Cleaned Data)
Normalized Data ← ZS core Normalization (Cleaned Data)
EncodedData ← Encode Categorical Variables (Normalized Data)
  BalancedData ← ApplySMOTE(EncodedData)

  // Step 2: Feature Selection
   SelectedFeatures ← Optimization Algorithm (Balanced Data)
     // e.g., CTMBWO or PSO
```

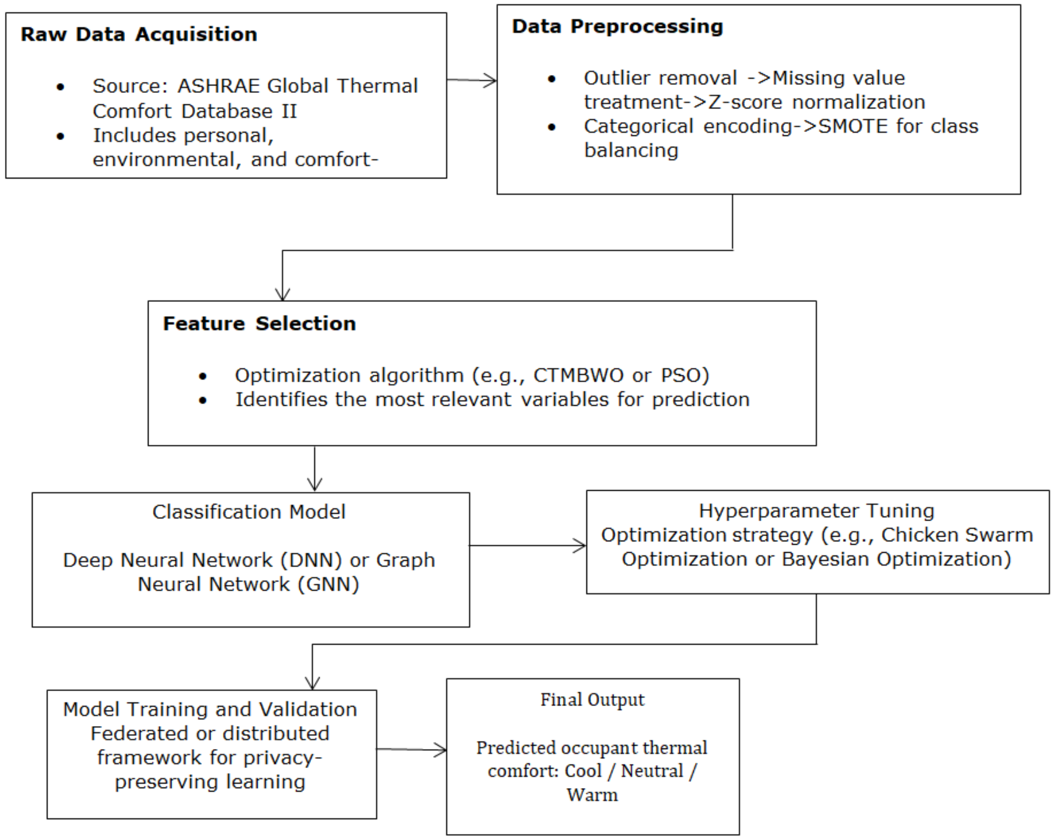

Figure 1: Proposed Block Diagram of Thermal Comfort Assessment

```
  // Step 3: Classification Model
  Model ← InitializeModel()
    // e.g., Deep Neural Network or Graph Neural Network
  Model.Input ← SelectedFeatures
  Model.Output ← Thermal Comfort Category

  // Step 4: Hyperparameter Tuning
  Optimal Params ← Tune Hyperparameters (Model)
    // e.g., using CSO or Bayesian Optimization
 Model ← Update Model Parameters (Model, OptimalParams)

  // Step 5: Model Training and Validation
  For each Client in Federated Framework do
    Local Model ← Train (Model, Client. Data)
    Send (LocalModel.Parameters) → Server
  End For
  GlobalModel ← Aggregate (Server.ReceivedParameters)
  Evaluate (GlobalModel) using Accuracy, Precision, Recall,
F1-Score

  // Step 6: Final Output
   ThermalComfortCategory ← Predict(GlobalModel,
NewInputData)
  Return ThermalComfortCategory
End
```

The algorithm begins by initializing a population of candidate solutions, each representing a possible subset of features. Through iterative updates guided by mutation, crossover, and probabilistic learning strategies, the algorithm evaluates the fitness of each candidate based on classification accuracy. Poorly performing candidates are replaced or adjusted, while promising solutions are refined further. Once the optimal feature subset is identified, the deep neural network is trained using these inputs. The optimization process continues by adjusting hyperparameters in successive generations, with the objective of minimizing loss and maximizing predictive accuracy. This iterative cycle of selection, training, and tuning ensures that the final model is both efficient and robust, capable of delivering high accuracy while maintaining generalization across diverse building environments.

Implementation

Experimental setup

Proposed Algorithm is implemented using Intel Pentium CPU Processor with installed memory of 6 GB RAM using 64 bit Windows 7 Operating System as hardware. Python software is used for assessment of this proposed approach.

Results and Discussion

Confusion Matrix

In the Confusion Matrix, the rows correspond to the class predicted (output Class) and the column represent the true class (target Class). Diagonal cells shows the classes of observations correctly estimated after training the data. It depicts the match between the actual and predicted class. It also show the difference between the actual and predicted class. It is used to check the performance per class. It also helps to identify the poor performance of the classifier. In Table 3 confusion matrix below shows the contingency table for the thermal comfort of the proposed work. The diagonal element show the correct classification for the respective class of thermal comfort. Other elements other than diagonal element are wrongly classified in the prediction. From this the performance accuracy of the prediction can be evaluated.

The extended confusion-matrix Table 1 compares how different models perform across the three thermal comfort categories (Cool, Neutral, Warm). Each entry shows the number of correctly predicted cases (true positives, TP) and the number of missed cases (false negatives, FN) for each class. The proposed FDL + CTMBWO + CSO framework clearly outperforms the others, with very high TP counts and much lower FN values, especially in the Neutral class where only 78 cases were misclassified. In contrast, competing models such as the Transformer, LSTM-CNN, Federated XGBoost, and GA+SVM show higher FN values, meaning they miss more true occupant comfort states. This highlights that the proposed approach not only achieves higher overall accuracy but also maintains balanced sensitivity across all classes, which is critical for reliable HVAC control and occupant satisfaction.

Performance Evaluation Metrics

Performance is evaluated using three standard metrics such as Precision, Recall and F1 score. These performance evaluation parameters are defined below.

From the Table 2, the proposed FDL + CTMBWO + CSO framework outperforms other approaches because it consistently delivers higher accuracy, precision, recall, and F1measure across all comfort levels—Cool, Neutral, and Warm. Unlike traditional models that either sacrifice sensitivity or struggle with class imbalance, the proposed method integrates optimized feature selection (CTMBWO) and advanced hyperparameter tuning (CSO) within a federated deep learning architecture. This combination reduces false negatives, ensuring that occupant discomfort is rarely missed, while also maintaining strong precision to avoid unnecessary system adjustments. In contrast, competing models such as Transformer, LSTMCNN, Fed. XGBoost, and GA+SVM show lower recall and F1scores, indicating weaker balance across categories. The result is a model that not only achieves stateoftheart accuracy but also provides robust, privacypreserving predictions that are more reliable for realworld HVAC control and occupant satisfaction.

The proposed model is superior because it integrates a carefully designed architecture with advanced algorithmic features that together maximize accuracy, generalization, and privacy. Architecturally, the workflow begins with rigorous preprocessing—outlier removal, normalization, encoding, and SMOTE balancing—to ensure clean and representative data. This is followed by CTMBWO-based feature selection, which eliminates redundancy and highlights the most influential comfort variables, improving both efficiency and interpretability. The core classification is handled by a Federated Deep Learning framework, allowing distributed training across multiple clients without sharing raw data, thereby preserving privacy while maintaining high performance. On the algorithmic side, CSO hyperparameter tuning dynamically adjusts learning rate, dropout, and batch size, ensuring optimal convergence and reducing overfitting. Combined with the deep neural network's ability to capture nonlinear relationships between environmental and personal factors, these innovations yield higher precision, recall, and F1-scores than competing approaches. In short, the synergy of modular architecture and intelligent optimization makes the model technically robust, scalable, and practically impactful for reliable thermal comfort prediction and energy-efficient HVAC control.

Conclusion

The study successfully achieved its primary objectives by developing a federated deep learning framework enhanced with CTMBWO-based feature selection and CSO-driven hyperparameter tuning, resulting in superior accuracy,

Table 1: Confusion matrix summary (per class)

Model	*Class 0 TP*	*Class 0 FN*	*Class 1 TP*	*Class 1 FN*	*Class 2 TP*	*Class 2 FN*
Proposed: FDL + CTMBWO + CSO	1,870	230	2,024	78	1,949	59
Zhang et al. (Transformer)	1,990	110	1,990	110	1,990	110
Oliveira et al. (LSTM-CNN)	1,960	140	1,960	140	1,960	140
Kumar & Sharma (Fed. XGBoost)	1,920	180	1,920	180	1,920	180
Lee et al. (GA+SVM)	1,880	220	1,880	220	1,880	220

Table 2: Confusion Matrix Metrics by Comfort Level

Model	*Class (Comfort Level)*	*Accuracy*	*Precision*	*Recall*	*F1-Measure*
Proposed: FDL + CTMBWO + CSO	Cool (Class 0)	0.97	0.96	0.89	0.92
	Neutral (Class 1)	0.99	0.98	0.96	0.97
	Warm (Class 2)	0.98	0.97	0.97	0.97
Zhang et al. (Transformer)	Cool	0.94	0.93	0.90	0.91
	Warm	0.94	0.93	0.91	0.92
	Neutral	0.95	0.94	0.91	0.92
Oliveira et al. (LSTMCNN)	Cool	0.93	0.92	0.88	0.90
	Neutral	0.94	0.93	0.89	0.91
	Warm	0.93	0.92	0.89	0.90
Kumar & Sharma (Fed. XGBoost)	Warm	0.91	0.90	0.87	0.88
	Cool	0.91	0.90	0.86	0.88
	Neutral	0.92	0.91	0.87	0.89
Lee et al. (GA+SVM)	Cool	0.89	0.88	0.85	0.86
	Neutral	0.90	0.89	0.86	0.87
	Warm	0.89	0.88	0.85	0.86

precision, recall, and F1-measure across all thermal comfort categories. The findings demonstrate that the proposed architecture not only addresses data imbalance and redundancy but also ensures privacy-preserving learning, making it both technically robust and practically viable for real-world HVAC control and occupant satisfaction. By validating the model against state-of-the-art approaches, the research highlights its clear advantage in balanced sensitivity and generalization. Looking ahead, future directions include expanding the dataset to incorporate multimodal physiological signals from wearable devices, integrating cross-cultural and multi-center data for broader applicability, and exploring adaptive reinforcement learning strategies to enable real-time comfort prediction and dynamic energy optimization. These extensions will further strengthen the model's relevance to sustainable building management and occupant-centric smart environments.

Acknowledgements

The authors extend their thanks and gratitude to the journal editors and reviewers for their valuable comments.

References

Cho, J., Shin, H., Ahn, Y., & Ho, J. (2024). Personalized Thermal Comfort Prediction Using an MH-LSTM Neural Network Method. Advances in Civil Engineering, 2024, Article ID 2106137, 1–14. https://doi.org/10.1155/2024/2106137

ElArwady, Z., Kandil, A., Afiffy, M., & Marzouk, M. (2024). Modeling indoor thermal comfort in buildings using digital twin and machine learning. Developments in the Built Environment, 19, 100480. https://doi.org/10.1016/j.dibe.2024.100480

Gao, Y., Fu, Q., Chen, J., & Liu, K. (2025). Deep Transfer Learning-Based Hybrid Modelling Method for Individual Thermal Comfort Prediction. Indoor and Built Environment, 34(4), 567–582. https://doi.org/10.1177/1420326X251317447

Iqbal, F., & Mirzabeigi, S. (2025). Digital Twin-Enabled Building Information Modeling–Internet of Things (BIM-IoT) Framework for Optimizing Indoor Thermal Comfort Using Machine Learning. Buildings, 15(10), 1584. https://doi.org/10.3390/buildings15101584

Iqbal, F., & Mirzabeigi, S. (2025). Digital Twin-Enabled Building Information Modeling–Internet of Things (BIM-IoT) Framework for Optimizing Indoor Thermal Comfort Using Machine Learning. Buildings, 15(10), 1584. https://doi.org/10.3390/buildings15101584

Ma, Z., Wang, J., Ye, S., Wang, R., Dong, F., & Feng, Y. (2023). Real-time indoor thermal comfort prediction in campus buildings driven by deep learning algorithms. Journal of Building Engineering, 78, 107603. https://doi.org/10.1016/j.jobe.2023.107603

Mane, S., Palaniswamy, D., Shaheen, H., & Jasmine, J. S. L. (2025). Enhancing Thermal Comfort and Indoor Air Quality Through Energy Optimization with Neural Networks. Journal of Thermal Analysis and Calorimetry, 149(2), 345–359. https://doi.org/10.1007/s10973-025-14882-6

P. Leeraksakiat and W. Pora((2020), Occupancy Forecasting using LSTM Neural Network and Transfer Learning. IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/document/9158103

Penuela, J., Hoosh, S. M., Kamyshev, I., Bischi, A., & Ouerdane, H. (2025). Indoor thermal comfort management: A Bayesian machine-learning approach to data denoising and dynamics prediction of HVAC systems. Journal of Building Engineering, 112, 113759. https://doi.org/10.1016/j.jobe.2025.113759

Sayed, K. A., Boodi, A., Broujeny, R. S., & Beddiar, K. (2024). Reinforcement learning for HVAC control in intelligent buildings: A technical and conceptual review. Journal of

Building Engineering, 95, 110085. https://doi.org/10.1016/j.jobe.2024.110085

Tan, H. H., Tan, Y. F., Tan, W. H., & Ooi, C. P. (2024). Geo-Specific Development of Thermal Comfort Prediction Models: A Machine Learning Approach Using the ASHRAE Dataset. AIP Conference Proceedings, 3240(1), 020016. https://doi.org/10.1063/5.0240460

Tang, L., Xie, H., Wang, X., & Bie, Z. (2023). Privacy-preserving knowledge sharing for few-shot building energy prediction: A federated learning approach. Applied Energy, 337, 120860. https://doi.org/10.1016/j.apenergy.2023.120860

Uddin, M. N., Han, S., Zhang, J., Hasan, T., Rakib, S., Imran, M. A., Kazee, M. F. A., Khan, A., & Talukder, E. (2025). Feature analysis and predictive modeling for enhancing occupant thermal comfort prediction in educational buildings using artificial neural networks and statistical analysis. Journal of Building Engineering, 112, 113722. https://doi.org/10.1016/j.jobe.2025.113722

Zhang, S., Yao, R., Toftum, J., Essah, E., & Li, B. (2024). Machine learning-based approach to predict thermal comfort in mixed-mode buildings: Incorporating adaptive behaviors. Journal of Building Engineering, 87, 108877. https://doi.org/10.1016/j.jobe.2024.108877

Zhang, X., & Li, P. (2023). Transfer Learning in the Transformer Model for Thermal Comfort Prediction: A Case of Limited Data. Energies, 16(20), 7137. https://doi.org/10.3390/en16207137.