SUSTAINABLE MACHINE INTELLIGENCE JOURNAL

SCIENCES FORCE
International Academic Publisher

A Machine Learning Approach for Improved Thermal Comfort Prediction in Sustainable Built Environments

Waleed Abd El-khalik

Faculty of Computers and Informatics, Zagazig University, Zagazig, Sharqiyah, 44519, Egypt; waleed.abdlekhalik@zu.edu.eg

Abstract: Thermal comfort prediction within sustainable built environments stands as a pivotal challenge intertwining human well-being and environmental sustainability. This paper presents a pioneering framework leveraging machine learning methodologies to advance predictive models for thermal comfort. Drawing upon a comprehensive dataset sourced from ASHRAE field studies and the RP-884 database, comprising 107,463 entries, our study unfolds a novel approach to enhancing thermal comfort predictions. The integration of diverse physiological parameters, environmental data, and occupant preferences forms the foundation of our machine learning-driven framework. Through meticulous analysis and model development, our approach not only refines predictive accuracy but also underscores adaptability across varying environmental contexts. The study contributes not only to the discourse on thermal comfort prediction but also emphasizes the crucial nexus between sustainable design, occupant well-being, and energy efficiency. Furthermore, the study introduces user-friendly web-based tools to explore the ASHRAE database, facilitating accessibility and utilization for researchers and practitioners. The findings showcase the potential of machine learning in revolutionizing sustainable building design paradigms, emphasizing human-centric approaches while aligning with environmental conservation goals.

Event	Date
Received	01-05-2022
Revised	05-08-2022
Accepted	21-09-2022
Published	29-10-2022

Keywords: CO2 emissions, Machine learning, Predictive modeling, Environmental sustainability, Greenhouse gas emissions, Climate change, Comparative analysis, Carbon footprint, Renewable energy.

1. Introduction

The comfort in buildings resulting from thermal conditions determines its well being and productivity. The relationship between human comfort and built environment has been of great interest to sustainable development and architecture scholars [1-3]. This has led to the importance of understanding and predicting thermal comfort to create environments that respond to occupants' desires as well as align with sustainability objectives. It thus indicates the need for examining how machine learning can be integrated into thermal comfort prediction, in order to develop new models for optimizing sustainable buildings that will improve human experiences in them [4-5].

As the whole world is advancing towards smart buildings and sustainability, there has been a challenge on how to arrive at optimum thermal comfort. Initially changes in thermal comfort were made based on models that are fixed and rarely capture the multifarious preferences of humans when it comes to temperature [4]. This inadequacy has led to investigations of new tools such as machine learning which is good at collecting, organizing, discerning and processing large amount of data [7]. All this breakthrough could be the start of an era in predicting and controlling human thermal environment using more personalized, energy-efficient, sustainable building designs [8]. Nonetheless, established techniques face difficulties in effectively predicting thermal comfort across different environmental circumstances and occupant preferences. What this research seeks to accomplish is find a way of developing resilient predictors that do not rely on conventional methodologies [9-11]. Considering that personal preferences for comfort differ extensively across individuals and the unpredictable nature of the environment, it is important to develop a sophisticated technique for optimizing predictions of thermal comforts [5]. These are some complexities that we will be trying to address in doing this research so as to narrow the gap between normal models and the actual needs of occupants for their comfort within a sustainable habitat.

Our primary goal, given these challenges, is to create machine learning-based framework for improved prediction of thermal comfort in sustainable built environments. We want to employ advance algorithms' data-driven insights to create predictive models that can accommodate diverse occupant preferences and adapt them with environmental changes. This will enable more flexible and people-centered building designs that minimize energy usage while ensuring people's comfort which is our concern in this study. The contributions of this work extend beyond the realm of predictive modeling, aiming to catalyze a paradigmatic shift in the approach towards sustainable design, emphasizing the synergy between human needs and environmental conservation.

2. Related Work

In this section, we will go on a journey through the most important works and contemporary developments in thermal comfort forecast methods, machine learning techniques for building science, and symbiosis of sustainable design principles with individual-based comfort concepts. Looking at an identical built environment, Chaudhuri et al. [5] studied thermal comfort prediction using normalised skin temperature as a critical measure. This work explored the connection between thermal comfort and physiological parameters in an experiment that generated how to come up with predictive models in controlled environment conditions.The deep transfer hybrid learning model for predicting building thermal com-fort which was developed by Somu et al. [7] is another one. The study explored benefits of applying deep learning strategies showing the novel method to improve forecast accuracy in different environmental contexts. In another study by Chaudhuri et al. [8], they looked at wearable sensing technologies towards gender-specific physiological parameters based on predicting thermal comfort. The research employed random forest-based approach into personalized comfort prediction models.

Wu, Y. et al. [10] presented an individual thermal comfort prediction model based on classification tree methodology, integrating physiological parameters and thermal history specifically in winter conditions. Their study aimed to tailor predictions for individual comfort needs within varying environmental contexts. Chen, K. H. et al. [12] focused on thermal comfort prediction and validation within a realistic vehicle thermal environment. Their study, detailed in a technical paper, contributed insights into vehicle-based thermal comfort modeling. Fang, Z. et al. [13] investigated outdoor thermal comfort prediction models in South China, particularly in Guangzhou. Their case study provided valuable insights into outdoor comfort considerations within a specific geographical context. Yu, C. et al. [14] assessed the performances of machine learning algorithms for individual thermal comfort prediction, drawing data from both professional and practical settings. Their study aimed to evaluate the efficacy of machine learning models across diverse environments. Eslamirad, N. et al. [16] applied supervised machine learning techniques in predicting thermal comfort within green sidewalks of Tehran. Their study contributed to understanding thermal comfort dynamics in specific urban settings, emphasizing sustainability considerations. Peng, B. & Hsieh, S. J. [18] explored data-driven thermal comfort prediction utilizing support vector machine methodology. Their study, presented at a manufacturing science and engineering conference, highlighted the application of machine learning in thermal comfort modeling.

3. Material and Method

This section elucidates the methodologies, data sources, and analytical frameworks employed in the pursuit of advancing thermal comfort prediction within sustainable built environments.

The experimentations of this study used the ASHRAE database, which is aggregated from field studies spanning the period from 1995 to 2015, sourced from diverse global locations. This comprehensive database results from contributors willingly sharing their raw data with the aim of broadening access across the thermal comfort research community. Following a rigorous quality assurance process, the dataset comprised 81,846 rows of paired subjective comfort evaluations and objective instrumental measurements, capturing essential thermal comfort parameters. Moreover, an additional 25,617 rows of data from the original ASHRAE RP-884 database were integrated, culminating in a robust collection totaling 107,463 entries. Primarily designed to facilitate varied inquiries into field-based thermal comfort, the database is a cornerstone for research exploration. The dataset is cleaned by removing null values. Then, descriptive analysis is made as an initial step in our journey for data exploration, whose results are made available in Table 1.

Table 1. Quantitative statistics for cleaned edition of ASHRAE database.

	Clo	**Met**	**Air temperature (C)**	**Relative humidity (%)**	**Air velocity (m/s)**	**Outdoor monthly air temperature (C)**	**Thermal comfort**
count	16209	16209	16209	16209	16209	16209	16209
mean	0.681518	1.19562	25.38873	50.91397	0.261687	24.76815	4.349312
std	0.324204	0.197861	4.43981	15.76351	0.693904	8.259837	1.404411
min	0	0.7	0.6	10.3	0	-2	1
25%	0.49	1.1	22.8	39.4	0.06	21.3	3
50%	0.63	1.2	25.3	50.1	0.15	26	5
75%	0.78	1.2	28.1	63.3	0.33	30.8	5
max	2.87	5	45.2	95.3	56.17	45.1	6

Following obtaining data from ASHRAE database, a rigorous quality assurance process was implemented to ensure the reliability and integrity of the dataset. This step involved identifying and rectifying missing or erroneous values, addressing outliers, and standardizing formats to establish a cohesive and accurate dataset for subsequent analysis. Next, to enhance the efficiency of our machine learning models, a judicious feature selection process was employed. This involved identifying relevant variables influencing thermal comfort prediction.

Physiological parameters, environmental conditions, and occupant preferences were prioritized based on their significance in previous research and exploratory data analysis (EDA) visualizations.

After that, standardization of numerical features is performed to bring the values into a common scale, preventing any particular feature from dominating the model due to its magnitude. Techniques such as Min-Max scaling were applied to ensure that all variables contributed proportionally to the modeling process.

$$C_{Min-Max}{}' = \frac{C_i - min(C)}{max(C) - min(C)} \tag{1}$$

Categorical variables were encoded to numerical representations to facilitate machine learning model compatibility. Techniques like one-hot encoding were employed to transform categorical variables into binary vectors, ensuring their meaningful inclusion in the predictive models. The dataset was partitioned into training and testing sets to evaluate model performance effectively. This involved allocating a percentage of the data for training the model and reserving a separate portion for evaluating its predictive capabilities, thus avoiding overfitting.

The Random Forest algorithm belongs to the ensemble learning family, characterized by the creation of multiple decision trees during training and the combination of their outputs for enhanced predictive accuracy and robustness. Each decision tree in the forest is constructed independently, utilizing a random subset of features and data points. This randomness injects diversity into the model, mitigating overfitting and promoting generalizability. The key components of random forest are described as follows:

Decision Trees: The foundational building blocks of Random Forest are decision trees. Each tree is constructed by recursively partitioning the dataset based on the most informative features. These trees collectively form the forest, contributing to the final prediction through a process called bagging (bootstrap aggregating).

Bootstrap Aggregating (Bagging): During the training process, each decision tree is built on a random subset of the training data, and at each split, only a random subset of features is considered. This introduces variability into the model, making it less susceptible to noise and outliers.

Voting Mechanism: In the predictive phase, the Random Forest employs a voting mechanism. Each decision tree in the forest contributes a prediction, and the final output is determined by majority voting. This ensemble approach enhances the model's resilience to errors and increases its overall predictive accuracy.

In our study, we applied the Random Forest algorithm to predict the level of thermal comfort within sustainable built environments. Leveraging the preprocessed dataset, the algorithm was trained on a subset of the data, learning the intricate relationships between physiological parameters, environmental conditions, and occupant preferences. The inherent randomness in the algorithm allowed it to capture complex patterns and adapt to varying contexts.

The Random Forest's ability to handle both numerical and categorical data, accommodate interactions between variables, and mitigate overfitting made it an apt choice for predicting the nuanced and multifaceted nature of thermal comfort. Through an iterative process of model training, validation, and fine-tuning, the Random Forest algorithm emerged as a powerful tool for generating accurate predictions, contributing to the optimization of sustainable building designs and the prioritization of human well-being in diverse environmental scenarios. The application of Random Forest in our study exemplifies its versatility and effectiveness in the context of thermal comfort prediction.

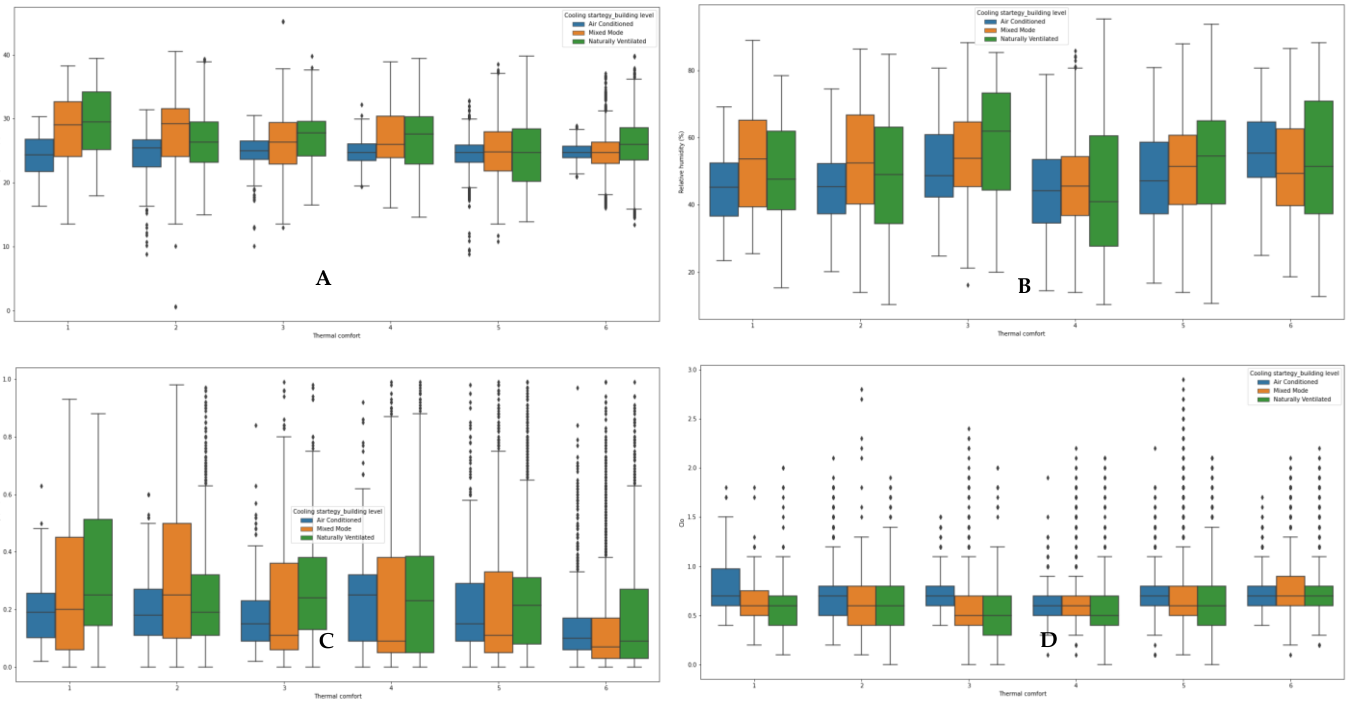

Figure 1. Visulaizing the interplay between thermal comfort and other attributea in ASHRAE Database

4. Experimental Results

This section presents the empirical findings derived from the implementation of our machine learning-driven framework for thermal comfort prediction in sustainable built environments.

As an important step in exploratory data analysis (EDA) of the ASHRAE database, we present visual analysis for the relationships between various data variables and thermal comfort, as displayed in Figure 1. In this systematic visualization, we employ box plots to show the interplay between each parameter and occupants' perceived comfort levels.

These plots not only offer a comprehensive overview of the dataset's characteristics but also unveil patterns, trends, and potential correlations that lay the foundation for subsequent machine learning model development. The EDA phase serves as a crucial step in comprehending the complexities of thermal comfort dynamics within sustainable built environments, providing valuable insights that inform the subsequent stages of our analysis and model refinement. In Figure 1, A), we can observe that the lower the air temperature, the higher the thermal comfort. In Figure 1, B), we can observe that it is difficult to determine if relative humidity on its own has significant impact on thermal comfort. In addition, in Figure 1, C), we can observe that the lower the air velocity, the better the thermal comfort.

In Figure 2, we present the confusion matrix for the Random Forest model, providing a detailed and transparent evaluation of its predictive performance. This matrix delineates the model's ability to correctly classify instances into different categories, offering a comprehensive view of true positives, true negatives, false positives, and false negatives. The visualization serves as a crucial tool for assessing the accuracy, precision, recall, and overall effectiveness of the Random Forest algorithm in predicting thermal comfort within sustainable built environments. By unraveling the intricacies of model performance, the confusion matrix aids in fine-

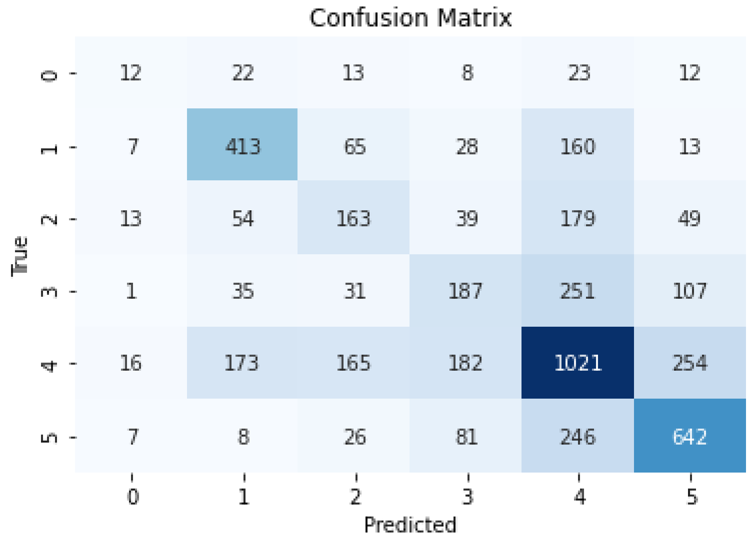

Figure 2: Confusion Matrix for Random Forest Model

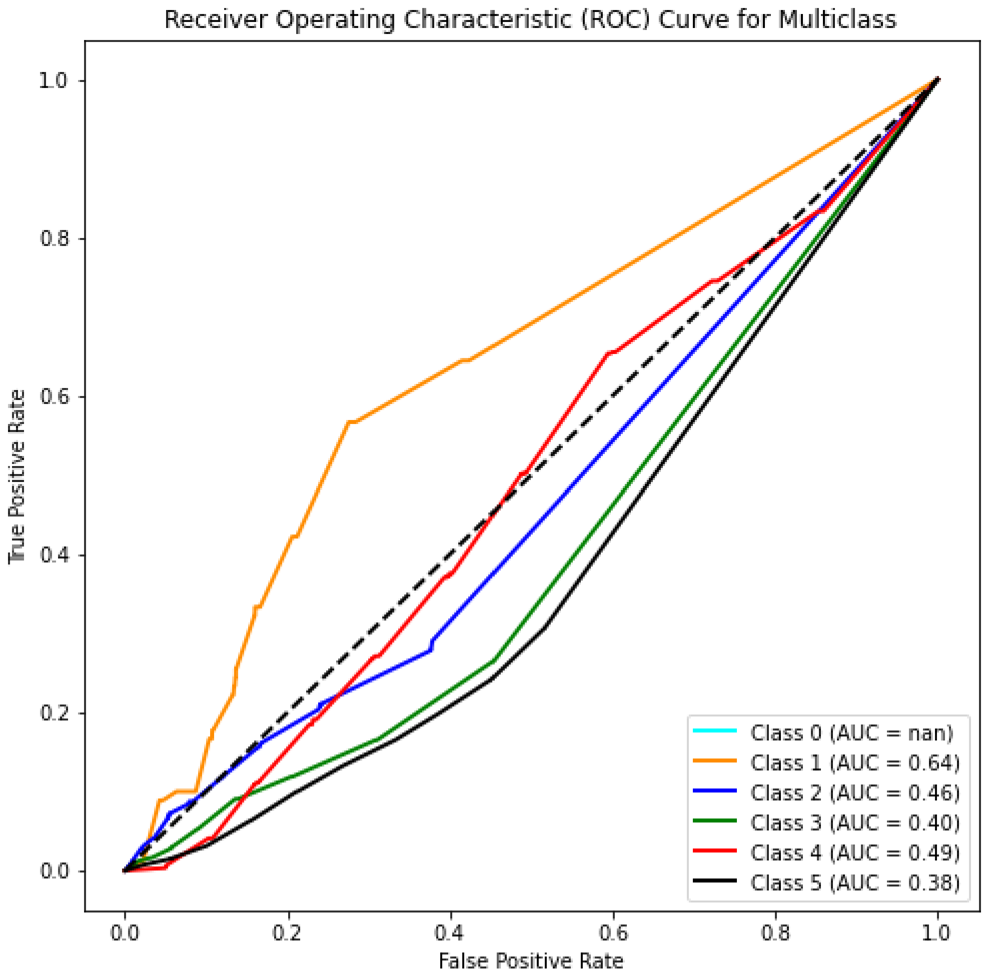

Figure 3: ROC Curve for Random Forest Model

tuning and optimizing the machine learning approach, contributing to the refinement of predictive models for more accurate and reliable thermal comfort assessments.

In Figure 3, we present the Receiver Operating Characteristic (ROC) curve for the Random Forest model, a graphical representation that illustrates the trade-off between true positive rate and false positive rate across varying thresholds. The ROC curve provides a holistic view of the model's discriminatory power and its ability to distinguish between different classes. By analyzing the area under the ROC curve (AUC), we can quantitatively assess the Random Forest model's overall performance. This visualization is pivotal in evaluating the model's sensitivity and specificity, aiding in the determination of an optimal threshold for thermal comfort prediction within sustainable built environments. The ROC curve in Figure 3 serves as a valuable tool for understanding and communicating the classification performance of the Random Forest algorithm in our study.

5. Concluding Remarks

This study marks a significant step forward in leveraging machine learning methodologies for the prediction of thermal comfort within sustainable built environments. Utilizing the Random Forest algorithm on a robust dataset from the ASHRAE Global Thermal Comfort Database II, our approach demonstrated commendable predictive accuracy. However, it is important to acknowledge the limitations in prediction performance, which underline the need for more high-quality data and potentially more complex modeling approaches. The amalgamation of physiological parameters, environmental data, and occupant preferences provided a nuanced understanding of the intricate dynamics influencing thermal comfort. Our comprehensive exploratory data analysis (EDA) facilitated feature selection, ensuring that only the most influential variables contributed to the model's predictions. Yet, the quest for enhanced predictive capabilities reveals the necessity for richer datasets, capturing a more diverse range of scenarios and factors influencing thermal comfort. The application of the Random Forest algorithm showcased its versatility in handling diverse data types, but the pursuit of heightened predictive accuracy may warrant exploration into more complex modeling frameworks. As we reflect on the outcomes derived from the confusion matrix and ROC curve evaluations, it becomes apparent that a more sophisticated approach, perhaps incorporating advanced machine learning algorithms or hybrid models, could unlock greater potential in capturing the complexities of thermal comfort dynamics.

Supplementary Materials

Not applicable.

Author Contributions

All authors contributed equally to this study.

Funding

This research was conducted without external funding support.

Ethical approval

This article does not contain any studies with human participants or animals performed by any of the authors.

Conflicts of Interest

The authors declare that there is no conflict of interest in the research.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

References

[1]. Gao, G., Li, J., & Wen, Y. (2020). DeepComfort: Energy-efficient thermal comfort control in buildings via reinforcement learning. IEEE Internet of Things Journal, 7(9), 8472-8484.

[2]. Omidvar, A., & Kim, J. (2020). Modification of sweat evaporative heat loss in the PMV/PPD model to improve thermal comfort prediction in warm climates. Building and Environment, 176, 106868.

[3]. Becker, R., & Paciuk, M. (2009). Thermal comfort in residential buildings–failure to predict by standard model. Building and Environment, 44(5), 948-960.

[4]. Wu, Y., & Cao, B. (2022). Recognition and prediction of individual thermal comfort requirement based on local skin temperature. Journal of Building Engineering, 49, 104025.

[5]. Chaudhuri, T., Zhai, D., Soh, Y. C., Li, H., & Xie, L. (2018). Thermal comfort prediction using normalized skin temperature in a uniform built environment. Energy and Buildings, 159, 426-440.

[6]. Chennapragada, A., Periyakoil, D., Das, H. P., & Spanos, C. J. (2022, June). Time series-based deep learning model for personal thermal comfort prediction. In Proceedings of the Thirteenth ACM International Conference on Future Energy Systems (pp. 552-555).

[7]. Somu, N., Sriram, A., Kowli, A., & Ramamritham, K. (2021). A hybrid deep transfer learning strategy for thermal comfort prediction in buildings. Building and Environment, 204, 108133.

[8]. Chaudhuri, T., Zhai, D., Soh, Y. C., Li, H., & Xie, L. (2018). Random forest based thermal comfort prediction from gender-specific physiological parameters using wearable sensing technology. Energy and Buildings, 166, 391-406.

[9]. Ku, K. L., Liaw, J. S., Tsai, M. Y., & Liu, T. S. (2014). Automatic control system for thermal comfort based on predicted mean vote and energy saving. IEEE Transactions on Automation Science and Engineering, 12(1), 378-383..

[10]. Wu, Y., Liu, H., Li, B., Kosonen, R., Wei, S., Jokisalo, J., & Cheng, Y. (2021, December). Individual thermal comfort prediction using classification tree model based on physiological parameters and thermal history in winter. In Building Simulation (pp. 1-15). Tsinghua University Press.

[11]. Enescu, D. (2017). A review of thermal comfort models and indicators for indoor environments. Renewable and Sustainable Energy Reviews, 79, 1353-1379.

[12]. Chen, K. H., Kaushik, S., Han, T., Ghosh, D., & Wang, M. (2012). Thermal comfort prediction and validation in a realistic vehicle thermal environment (No. 2012-01-0645). SAE technical paper.

[13]. Fang, Z., Zheng, Z., Feng, X., Shi, D., Lin, Z., & Gao, Y. (2021). Investigation of outdoor thermal comfort prediction models in South China: A case study in Guangzhou. Building and Environment, 188, 107424.

[14]. Yu, C., Li, B., Wu, Y., Chen, B., Kosonen, R., Kilpelainen, S., & Liu, H. (2022). Performances of machine learning algorithms for individual thermal comfort prediction based on data from professional and practical settings. Journal of Building Engineering, 61, 105278.

[15]. Gao, N., Shao, W., Rahaman, M. S., Zhai, J., David, K., & Salim, F. D. (2021). Transfer learning for thermal comfort prediction in multiple cities. Building and Environment, 195, 107725.

[16]. Eslamirad, N., Malekpour Kolbadinejad, S., Mahdavinejad, M., & Mehranrad, M. (2020). Thermal comfort prediction by applying supervised machine learning in green sidewalks of Tehran. Smart and Sustainable Built Environ-ment, 9(4), 361-374.

[17]. Haldi, F., & Robinson, D. (2011). Modelling occupants' personal characteristics for thermal comfort prediction. International journal of biometeorology, 55, 681-694.

[18]. Peng, B., & Hsieh, S. J. (2017, June). Data-driven thermal comfort prediction with support vector machine. In International Manufacturing Science and Engineering Conference (Vol. 50749, p. V003T04A044). American Society of Mechanical Engineers.