Advanced and Sustainable Technologies
ASET
E-ISSN 2976-2294

Volume 5, No 2, September 2026 [160-171]

Integrating Heart Rate Data into Machine Learning Models of Indoor Thermal Comfort Prediction

Ismail Ishaq Ibrahim [1]*, Shazmin Aniza Abdul Shukor [1] and Muhajir Ab Rahim[1]

[1]Faculty of Technology & Electrical Engineering, Universiti Malaysia Perlis (UniMAP), Kampus Alam Pauh Putra, 02600, Arau, Perlis, Malaysia.

Received 7 April 2026, Revised 6 May 2026, Accepted 5 June 2026

ABSTRACT

Maintaining occupant comfort necessitates accurate, real-time thermal comfort models. The Predicted Mean Vote (PMV) is a cornerstone index, but its calculation requires parameters that are often impractical to measure directly. Integrating heart rate will facilitate personalized comfort models. Further, it could anticipate comfort needs and make proactive adjustments when needed. Machine learning can enhance predictive accuracy of the models, and it enables continuous learning adaptation to changing conditions. This study leverages the computational environment of MATLAB to develop and compare two machine learning-based models, Support Vector Machine (SVM) and Random Forest (RF), for predicting PMV using heart rate (HR), air temperature (Ta), and relative humidity (RH). Both models were trained and evaluated based on Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R^2). This research concludes that RF provides a robust and accurate data-driven framework for real-time PMV prediction, paving the way for enhanced, physiology-responsive systems that can be used to achieve significant gains in energy resiliency and systematic sustainability from excessive carbon expenditure while simultaneously enhancing human bioclimatic comfort. This research also exemplifies the Smart Structure (SDG 11) mandate. It upgrades standard building hardware with IoT-driven intelligence, moving towards a resilient, sustainable and safe ecosystem.

Keywords: Heart Rate, Machine Learning, Physiological Sensing, Predicted Mean Vote (PMV), Thermal Comfort.

1. INTRODUCTION

Thermal comfort prediction plays a critical role in the design and operation of energy-efficient buildings, as it directly influences occupant well-being, productivity, and satisfaction. The Predicted Mean Vote (PMV) model, introduced by Fanger, has been widely adopted as the standard analytical approach for assessing thermal sensation under steady-state conditions [1]. However, PMV relies on simplified assumptions about human thermoregulation and environmental homogeneity [2], which limit its applicability in real-world indoor environments characterized by occupant diversity [3], physiological variability, and dynamic conditions [4], [5].

Recent advances in machine learning (ML) have enabled data-driven approaches to thermal comfort modeling by learning complex, nonlinear relationships directly from measured data [6], [7]. Supervised learning algorithms such as Support Vector Machines (SVM) and Random Forests (RF) have gained increasing attention due to their strong generalization performance and robustness in regression tasks involving heterogeneous and noisy outputs [8]. In particular,

*Corresponding author: ismailishaq@unimap.edu.my

integrating physiological indicators, such as the LH/HF ratio of heart rate variability, with environmental variables has been shown to improve the personalisation and accuracy of thermal comfort prediction [9][10].

Despite the growing body of literature on ML-based thermal comfort models, there remains a need for a systematic comparison of different algorithms under consistent experimental conditions, especially when physiological signals are incorporated [11][12]. Moreover, many studies focus primarily on predictive accuracy without sufficient analysis of model interpretability and underlying behavioural trends.

Therefore, this study aims to: (i) develop ML-based PMV prediction models utilizing SVM and RF from heart rate, air temperature, and relative humidity as data; (ii) compare the predictive performance of SVM and RF regression models using rigorous cross-validation; and (iii) interpret the learned relationships through feature importance analysis and partial dependence plots. The results provide insight into the suitability of ensemble learning approaches for personalized thermal comfort assessment and align with SDG 11, "Smart Structure".

2. METHODS

2.1 Experimental Design and Participants

The experimental study was designed to investigate the relationships among physiological responses, environmental conditions, and perceived thermal comfort under controlled indoor conditions. A total of 20 university students (male and female, aged 20-30 years) voluntarily participated in the experiment. All participants were healthy, with no known cardiovascular or thermoregulatory disorder and provided informed consent prior to participation.

Each participant underwent a single three-hour experimental session, during which they simulated typical sedentary office work activities, such as reading, typing, and other computer-based tasks. To minimize confounding effects, participants were instructed to avoid vigorous physical activity, caffeine intake, and heavy meals at least 2 hours before the experiment. The study protocol was designed to reflect a realistic office working environment while allowing controlled variation of indoor thermal conditions [13][14][15].

2.2 Experimental Environment

The experiment was conducted in a closed indoor room equipped with a split-type air-conditioning system. The room layout, furniture arrangement and lighting conditions were kept constant throughout all sessions to ensure experimental consistency.

The air-conditioning system temperature setpoint was systematically adjusted every one hour, resulting in three distinct thermal exposure periods during each experimental session. This design allowed participants to experience multiple thermal conditions while maintaining a continuous working posture [13]. Relative humidity was not actively controlled but was allowed to vary naturally in response to the air-conditioning operation. No external ventilation or additional heat sources were introduced during the experiment.

Figure 1 is the experimental setup for the multi-sensor IoT device. The 3D-printed prototype integrates atmospheric sensors (BMP280, DHT11) and a physiological sensor (MAX30102) for real-time telemetry. Commercial-grade reference instruments (a thermometer and a pulse oximeter) are used for empirical validation and sensor drift calibration.

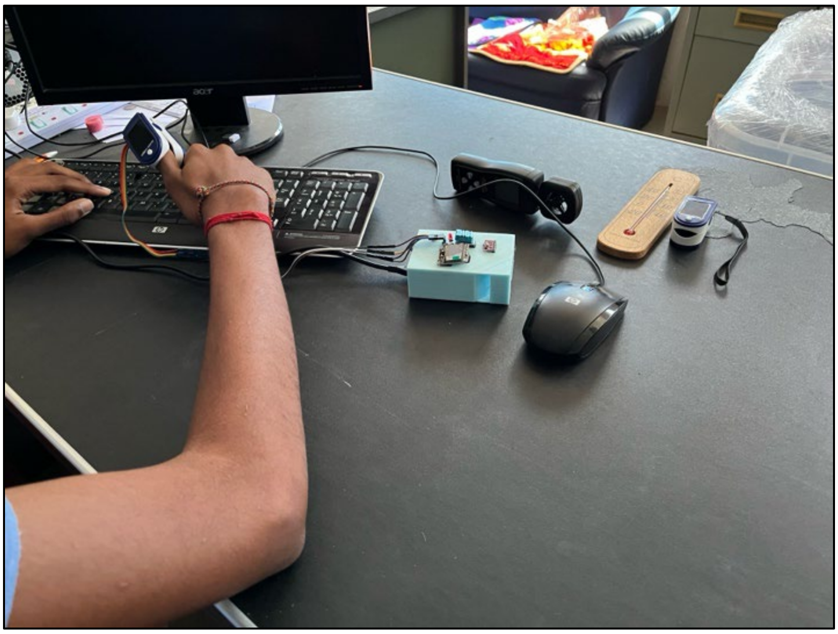

Figure 1: The setup for the data acquisition experiment.

2.3 Instrumentation and Data Acquisition

High-resolution sensors were used to capture both environmental and physiological variables at high temporal frequency. Sensor selection prioritized accuracy, stability and compatibility with automated data logging.

Air temperature was continuously measured using a BMP280 environmental sensor, which provides high-resolution and stable readings suitable for indoor environmental monitoring [16]. The sensor was placed near the participant's working area to capture representative ambient conditions.

Relative humidity was monitored using a DHT11 digital sensor. It was selected for its reliability in relatively humidity-rich indoor settings. The DHT11 sensor provides stable and calibrated humidity readings suitable for human comfort modeling.

Figure 2 represents a multi-sensor environmental and physiological data acquisition system based on the ESP8266 microcontroller. It utilizes a shared I2C bus architecture to integrate various digital sensors and a display unit, making it an efficient edge-computing device for IoT-based data acquisition.

Heart rate data were collected using the MAX30102 optical heart rate sensor, which measures photoplethysmographic (PPG) signals to estimate real-time heart rate. The sensor was securely attached to the participant's finger to ensure stable signal acquisition while minimizing motion artifacts [17]. Heart rate was selected as a physiological indicator due to its sensitivity to both metabolic activity and thermal stress, making it a relevant proxy for individual thermophysiological response.

All sensors were interfaced with the ESP8266 microcontroller, which served as the central unit for data acquisition and communication. The ESP8266 was selected due to its integrated Wi-Fi capability, low power consumption, sufficient processing capacity for multi-sensor sampling, and widespread use in Internet of Things (IoT)-based environmental monitoring system.

The ESP8266 coordinated sensor polling, timestamping, preprocessing and wireless transmission of data to a cloud platform. Its onboard Wi-Fi module enables real-time data upload without the need for additional communication hardware, making it suitable for long-duration indoor monitoring experiments [18]. Figure 3 depicts the custom-designed multi-sensor platform employed for the synchronized acquisition of environmental parameters and physiological signals.

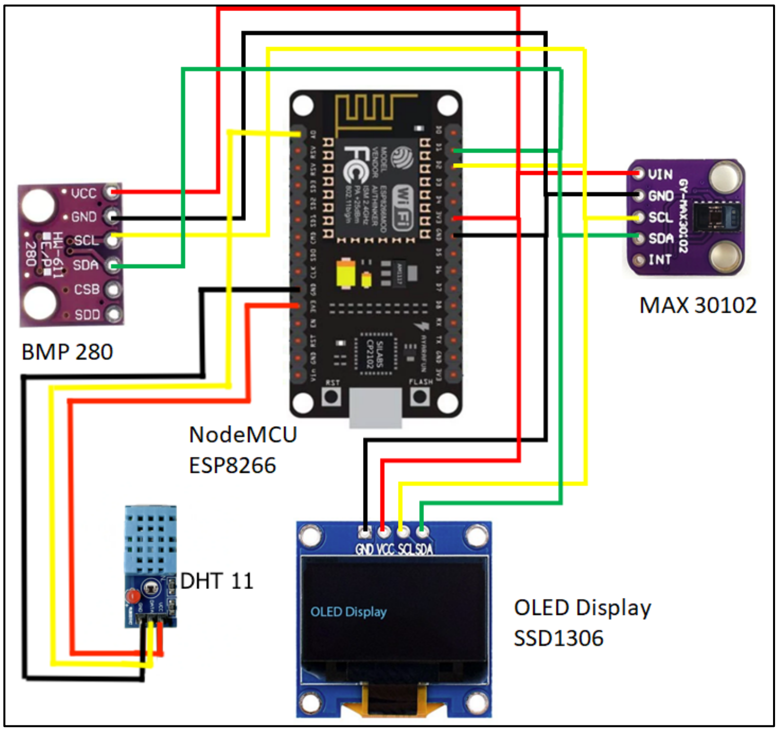

Figure 2: Data acquisition device schematic diagram.

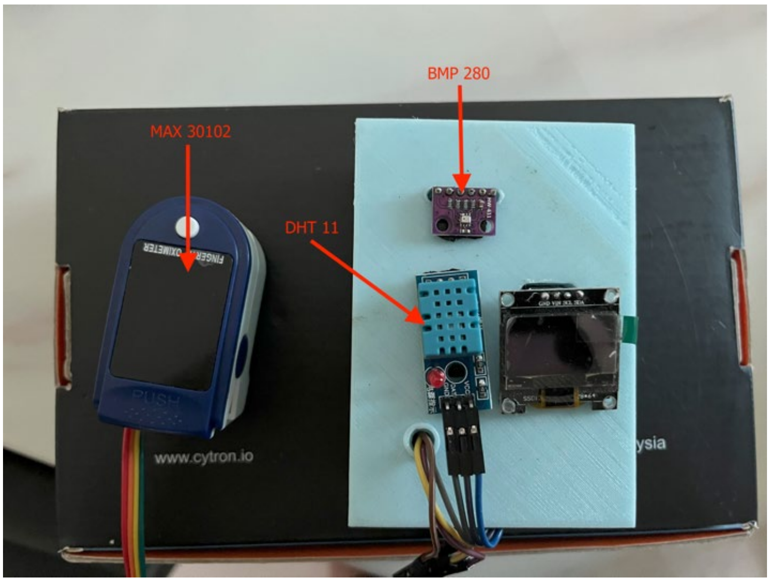

Figure 3: Data acquisition device.

2.4 Data Logging and Transmission

All sensor data, including air temperature, relative humidity and heart rate, were sampled and transmitted at intervals ranging from 5 to 10 seconds, depending on network stability and sensor response time. Data were automatically uploaded to a cloud-based Google Sheets platform in real time, enabling continuous and synchronized data storage throughout the experiment [19][20].

This approach ensured minimal data loss, accurate temporal alignment between environmental and physiological measurements, and scalability for long-duration monitoring.

2.5 Subjective Thermal Comfort Assessment

To capture perceived thermal comfort, participants will be asked to complete a subjective thermal comfort questionnaire every 30 minutes during the experiment. The questionnaire was based on the thermal sensation scale commonly used in PMV-related studies, allowing participants to report their instantaneous thermal perception.

The 30-minute interval was selected to balance sufficient temporal resolution to capture perceptual changes and to avoid excessive participant burden or distraction [9][13]. Each reported thermal comfort vote was timestamped and later synchronized with the corresponding physiological and environmental data collected during the same period.

2.6 Data Processing and Synchronization

Following data collection, all recorded data were aggregated and preprocessed prior to model development. The processing steps included removing invalid or missing sensor readings, synchronizing sensor data with subjective comfort votes based on timestamps, aggregating high-frequency sensor data to match the temporal resolution of subjective comfort votes, and normalizing input variables to reduce scale bias during machine learning training [21]. Each valid data instance consisted of heart rate, air temperature, relative humidity and corresponding subjective thermal comfort vote. Because sensor readings were captured every few seconds while subjective votes were captured every 30 minutes, a windowing technique was applied. This involved taking the mean of the environmental physiological data over a 10-minute window immediately preceding each subjective vote, ensuring that the model inputs accurately represented the physiological state that led to the participant's subjective sensation.

2.7 Machine Learning Framework

The processed dataset was used to develop machine learning regression models for thermal comfort prediction. SVM regression and RF regression were selected due to their established performance in nonlinear regression tasks and prior applications in thermal comfort research.

Model performance was evaluated using 10-fold cross-validation, ensuring robust estimates of generalization performance while minimizing bias and variance. Evaluation metrics included Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and the coefficient of determination (R^2) [22]. To enhance interpretability, feature importance analysis and partial dependence plots were employed for the RF model.

3. RESULTS AND DISCUSSION

The predictive performance of SVM and RF models was evaluated using ten-fold cross-validation. Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and coefficient of determination (R^2) were used as evaluation metrics. The averaged results across all folds are summarized in Table 1.

The RF model consistently outperformed the SVM model across all error metrics, achieving lower MAE and RMSE values and a higher R^2. Although the absolute R^2 values are modest, this outcome is not unexpected given the inherent subjectivity and noise associated with the thermal comfort perception and physiological signals. The random forest model effectively isolated the underlying trends from noisy physiological data using its ensemble logic. Importantly, the RF model demonstrated improved stability and generalization compared to SVM, as evidenced by reduced variance across folds.

Table 1: Model performance of all models.

Models	Performance		
	MAE	RMSE	R^2
PMV	2.774	2.936	-12.71
SVM	0.449	0.718	0.181
RF	0.461	0.679	0.267

The R^2 result for PMV is negative, and this is actually a common and highly revealing outcome in thermal comfort research, as the PMV model is evaluated against real-world data. This means that the PMV model is predicting the actual thermal sensation worse than if it had simply guessed the flat average of the real occupant's votes for every single data point.

Figure 4 provides a visual validation of the model's robustness and generalizability via a 10-fold cross-validation analysis. Each boxplot represents the interquartile range (IQR) of the MAE across 10 folds. The PMV model shows a tightly clustered distribution but an exceptionally high error rate (centred around ~2.75). The low variance here suggests that the model is consistently inaccurate. The model's rigid mathematical structure cannot capture the complexities of the real-world dataset collected by the sensors. Both machine learning models show a dramatic reduction in MAE. The SVM has the lowest median MAE and a very narrow IQR, indicating high stability across different data subsets. RF shows a similarly low MAE with a slightly larger vertical spread compared to SVM. This suggests that the RF is highly accurate on average; its performance is a bit more sensitive to the specific stochastic nature of the decision tree within certain folds of data.

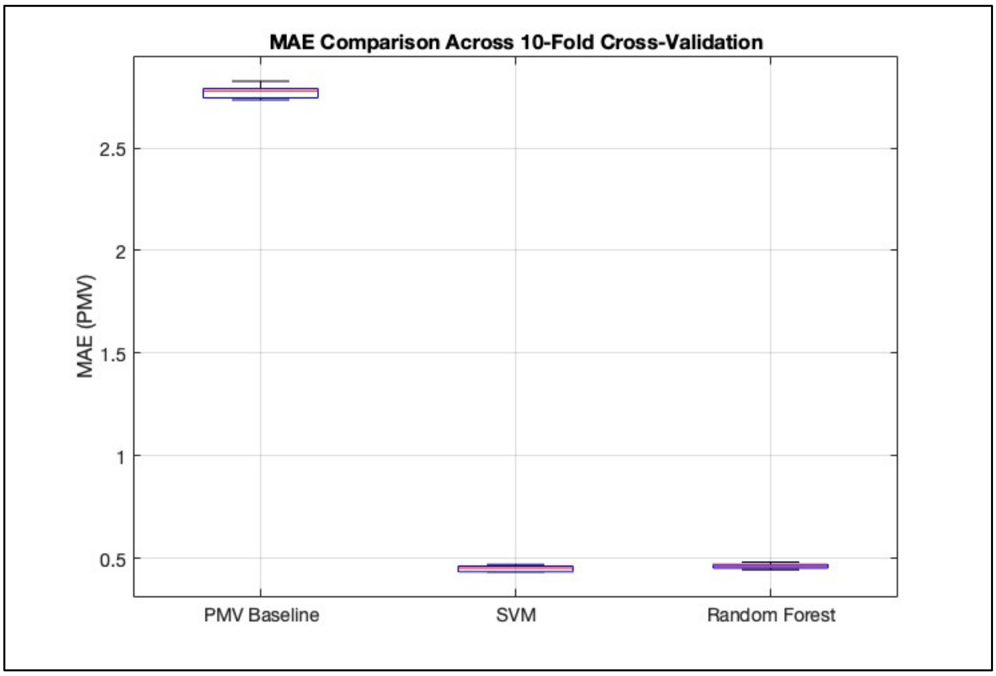

Figure 4: MAE comparison across 10-fold cross-validation.

Figure 5 illustrates the distribution of the Root Mean Square Error (RMSE) across a 10-fold cross-validation procedure. The PMV baseline exhibits an RMSE clustered around 2.9 to 3.0. The RMSE is slightly higher than the previous MAE (~2.7), indicating that the PMV model not only has a high average error but also suffers from occasional large outliers or "misses". The narrow boxplot at such a high error level confirms that the model consistently fails to capture the dataset's underlying variance, making it unsuitable for high-precision environmental-physiological monitoring.

The machine learning models exhibit a dramatic shift toward the lower bound of the y-axis, signifying a successful capture of the target variable's distribution. SVM distribution is extremely tight, centered around 0.7. This suggests high reliability: the model produces a consistent error magnitude across data folds. RF median RMSE is marginally lower than SVM (closer to 0.65); there is also a red outlier cross below the bottom whisker. This indicates that in at least one specific fold, the RF model achieved exceptionally high accuracy.

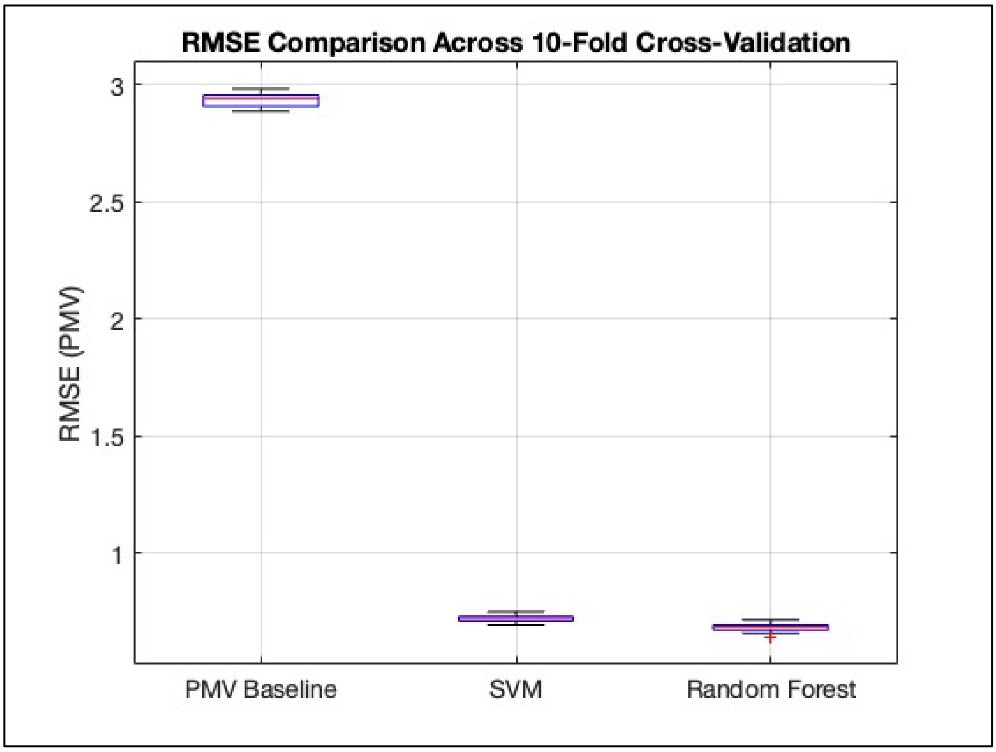

Figure 5: RMSE comparison across 10-fold cross-validation.

The relatively weak performance of the SVM model can be attributed to its reliance on a global kernel function, which enforces smoothness across the entire feature space. While this property is advantageous in low-noise scenarios, it can limit the model's ability to capture localized nonlinear interactions between physiological and environmental variables. In contrast, the RF model achieved the highest R^2 (0.267) and the lowest RMSE (0.679). This indicates that RF leverages an ensemble of decision trees to adaptively partition the input space, enabling it to model complex, heterogeneous relationships more effectively.

The lower RMSE suggests RF is more robust against large errors. In a heart rate monitoring context, preventing large "misses" is more critical than a slightly better average error.

Figure 6 presents the scatter plot of predicted versus measured PMV values for the RF model. The predictions exhibit a clear positive correlation with the measured PMV, with most data points distributed near the identity line. Deviations from the ideal line reflect the stochastic nature of thermal sensation and the limitations of using a limited set of predictors to represent human thermal perception. The RF model shows a slightly higher dispersion at the "Neutral" (0) and "Slightly Warm" (+1) points. In the measured -3 and -2 columns, the RF predictions are more symmetrically distributed around the identity line. This suggests the RF model is better at capturing the "extremes" of thermal discomfort.

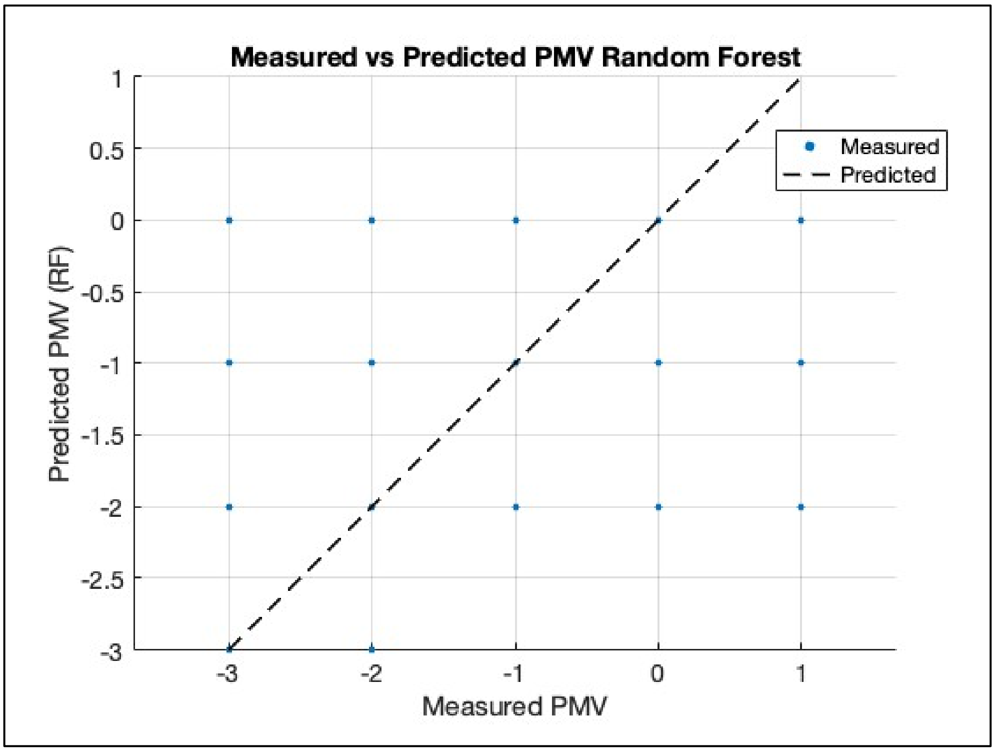

Figure 6: Measured vs Predicted PMV using RF.

Figure 7 illustrates the performance of a kernel-based regressor. The points in the SVM plot appear slightly more “constrained” toward the center of the y-axis range. This is characteristic of the RBF kernel, which tends to “smooth” the prediction manifold. For measured values at -3, many SVM predictions are clustered higher (closer to -2.5). This suggests a slight conservative bias. The SVM is less likely to predict extreme discomfort compared to RF. The cluster around the -1 and 0 measured points is tighter in the SVM than in the RF. It is highly accurate in normal conditions but less reactive to “extreme” conditions.

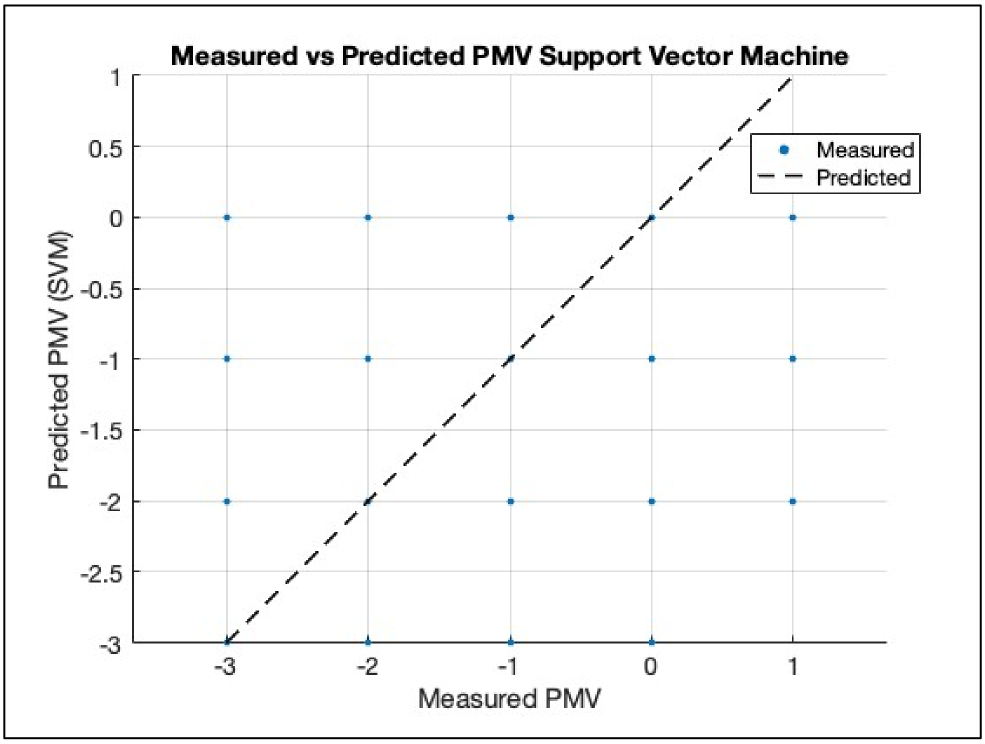

Figure 7: Measured vs Predicted PMV using SVM

Figure 8 shows the RF feature importance, which provides a hierarchy of thermal drivers. The temperature is confirmed as the primary driver, and the heart rate is identified as the biomarker of variance. While temperature tells us the environmental state, heart rate tells the model how the individual is coping with it. Although ranked lowest, humidity is vital for identifying latent heat stress, which heart rate often reflects through increased cardiovascular load.

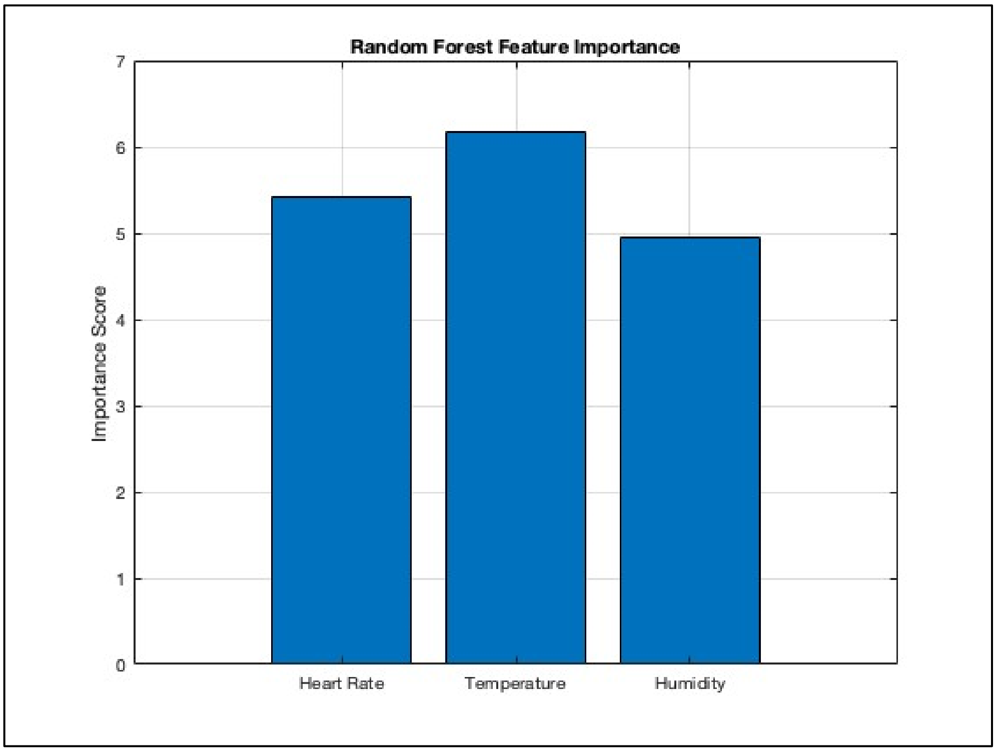

Figure 8: Random Forest feature importance.

Figure 9 shows the partial dependence plot. It allows the visualization of the “Tipping Point” of human comfort. The heart rate curve shows that the predicted discomfort remains relatively stable until a specific “biological trigger point” is reached, after which it climbs sharply. This marks the point at which the body’s autonomic nervous system shifts from passive to active thermoregulation. The linear-to-exponential rise in the temperature graph shows that the human perception of heat is not 1:1. As it moves away from the neutral zone, the discomfort accelerates. This phenomenon was captured perfectly by the RF model.

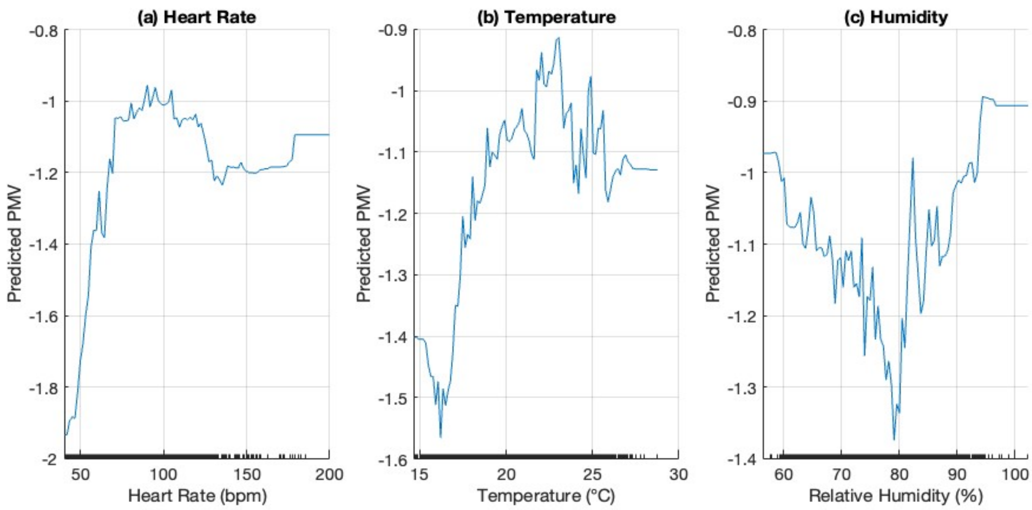

Figure 9: Random Forest partial dependence plot.

4. CONCLUSION

This study presented a comparative analysis of SVM and RF regression models for predicting thermal comfort (PMV) using physiological and environmental inputs. The result demonstrates that the RF model consistently outperforms the SVM model in terms of predictive accuracy and robustness, despite the inherent variability and subjectivity of thermal sensation data.

The superior performance of the RF model is attributed to its ensemble-based, non-parametric structure, which enables effective modelling of nonlinear interactions and localized patterns in the data. In contrast, the SVM model's globally smooth kernel representation limits its flexibility to noisy physiological measurements.

Future work will focus on incorporating additional physiological signals, such as skin temperature, skin humidity, perspiration rate, or shivering rate; expanding the dataset across diverse climatic conditions; and integrating adaptive control strategies to enable real-time, occupant-centric thermal comfort optimization.

ACKNOWLEDGEMENTS

This research is partially supported by the German Academic Exchange Service or *Deutscher Akademischer Austauschdienst* (DAAD).

REFERENCES

[1] Shukor, S. A. A., Kohlhof, K., Jamal, Z. A. Z. Development of a PMV-based thermal comfort modelling. Proceedings of the IASTED International Conference on Modelling and Simulation, issue June 2019 (2007) pp.670-675.

[2] Bogatu, D. I., et al. Human physiology for personal thermal comfort-based HVAC control - A review. Elsevier Ltd., (2023). https://doi.org/10.1016/j.buildenv.2023.110418

[3] Avelino, A. D., da Silva, L. B. PMV as a thermal evaluation method for air-conditioned spaces in hot climates: A systematic review. Ciência e Natura, vol 42 (2020) pp.e29. https://doi.org/10.5902/2179460X41375

[4] Liu, S., Schiavon, S., Das, H. P., Jin, M., Spanos, C. J. Personal thermal comfort models with wearable sensors. Build. Environ., vol 162, issue July (2019). https://doi.org/10.1016/j.buildenv.2019.106281

[5] Jung, D., Cheon, S., Kim, D., Yoon, J., Member, B. K. Short time remote heart rate measurement based on mmWave FMCW radar frame structure. vol 14, issue 8 (2023). https://doi.org/10.1109/LAWP.2023.3241153

[6] Lala, B., Kala, S. M., Rastogi, A., Dahiya, K., Hagishima, A. Are you comfortable now: Deep learning the temporal variation in thermal comfort in winters. Conf. Proc. IEEE Int. Conf. Syst. Man Cybern., vol 2022-Octob (2022) pp.1848-1855. https://doi.org/10.1109/SMC53654.2022.9945533

[7] Chen, Y., Liang, E., Berardi, U., Yang, J., Jiang, X., Cui, G. A model-based predictive control method of the ground-source heat pump system for maintaining thermal comfort in low-energy buildings. 2022 IEEE 7th International Conference on Power and Renewable Energy, ICPRE 2022, issue 52108100 (2022) pp.1080-1085. https://doi.org/10.1109/ICPRE55555.2022.9960567

[8] Huang, K., Lu, S., Li, X., Feng, K., Chen, W., Xia, Y. Predicted mean vote of subway car environment based on machine learning. Big Data Mining and Analytics, vol 6, issue 1 (2023) pp.92-105. https://doi.org/10.26599/BDMA.2022.9020028

[9] Liu, W., Lian, Z., Liu, Y. Heart rate variability at different thermal comfort levels. Eur. J. Appl. Physiol., vol 103, issue 3 (2008) pp.361-366. https://doi.org/10.1007/s00421-008-0718-6

[10] Fayyaz, M., Farhan, A. A., Javed, A. R. Thermal comfort model for HVAC buildings using machine learning. Arab. J. Sci. Eng., vol 47, issue 2 (2022) pp.2045-2060. https://doi.org/10.1007/s13369-021-06156-8
[11] Pao, S. L., et al. A physiological-signal-based thermal sensation model for indoor environment thermal comfort evaluation. Int. J. Environ. Res. Public Health, vol 19, issue 12 (2022) pp.1-16. https://doi.org/10.3390/ijerph19127292
[12] Ihianle, I. K., MacHado, P., Owa, K., Adama, D. A. Towards achieving thermal comfort through physiologically controlled HVAC system. International Conference Automatics and Informatics, ICAI 2022 - Proceedings, (2022) pp.245-249. https://doi.org/10.1109/ICAI55857.2022.9960018
[13] Zhan, J., He, W. Evaluation and prediction of elderly thermal comfort at varying ambient temperatures based on electroencephalogram signals and machine learning. Proceedings - 2022 15th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, CISP-BMEI 2022, issue 52078281 (2022) pp.0-4. https://doi.org/10.1109/CISP-BMEI56279.2022.9980300
[14] Nkurikiyeyezu, K., Shoji, K., Yokokubo, A., Lopez, G. Thermal comfort and stress recognition in office environment. HEALTHINF 2019 - 12th International Conference on Health Informatics, Proceedings https://doi.org/10.5220/0007368802560263
[15] Part of 12th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2019, issue Biostec (2019) pp.256-263.
[16] Nkurikiyeyezu, K. N., Suzuki, Y., Lopez, G. F. Heart rate variability as a predictive biomarker of thermal comfort. J. Ambient Intell. Humaniz. Comput., vol 9, issue 5 (2018) pp.1465-1477. https://doi.org/10.1007/s12652-017-0567-4
[17] Bosch. Bosch BMP280 datasheet. (2021).
[18] Patil, A., Hulwan, S. Heart-rate variability estimation using photoplethysmography signal. vol 9, issue 2 (2020) pp.67-74.
[19] Sung, W. T., Shih, J. A. Indoor thermal comfort environment monitoring system based on architecture of IoT. Proceedings - 2018 International Symposium on Computer, Consumer and Control, IS3C 2018, (2019) pp.165-168. https://doi.org/10.1109/IS3C.2018.00049
[20] Hormazábal, N., Franco, P., Urtubia, D., Ahmed, M. A. Data-driven smart avatar for thermal comfort evaluation in Chile. Buildings, vol 13, issue 8 (2023) pp.1-24. https://doi.org/10.3390/buildings13081953
[21] Ma, Y., Chen, C., Zeng, S., Guan, X., Lu, C. Data-driven edge offloading for wireless control. (2023) pp.1-16.
[22] Morresi, N., et al. Sensing physiological and environmental quantities to measure human thermal comfort through machine learning techniques. IEEE Sens. J., vol 21, issue 10 (2021) pp.12322-12337. https://doi.org/10.1109/JSEN.2021.3064707
[23] Zhou, Z., Qiu, C., Zhang, Y. A comparative analysis of linear regression, neural networks and random forest regression for predicting air ozone employing soft sensor models. Sci. Rep., vol 13, issue 1 (2023). https://doi.org/10.1038/s41598-023-49899-0

Conflict of interest statement: The authors declare no conflict of interest.

Author contributions statement: Conceptualization, Ismail I. Ibrahim & S. A. Abdul Shukor; Methodology, Ismail I. Ibrahim; Software, Ismail I. Ibrahim; Formal Analysis, Ismail I. Ibrahim, S. A. Abdul Shukor & M. A. Rahim; Investigation, Ismail I. Ibrahim; Writing & Editing, Ismail I. Ibrahim & S. A. Abdul Shukor; Supervision, S. A. Abdul Shukor & M. A. Rahim; Project Administration, S. A. Abdul Shukor; Funding Acquisition, Ismail I. Ibrahim & S. A. Abdul Shukor.

Declaration of Generative AI Use: During the preparation of this work, the authors used an AI tool to improve readability and language. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication. The authors also employed AI tool to generate preliminary code snippets for data analysis. All code was reviewed, tested, and verified by the authors, who take full responsibility for the methodology and results presented.

Data Availability Statement: All data generated or analyzed during this study are included in this published article.