IOP Conference Series:
Earth and
Environmental Science

Evaluating machine learning models in EEG-based thermal comfort studies: a comparative approach

To cite this article: Y Zhou *et al* 2025 *IOP Conf. Ser.: Earth Environ. Sci.* **1500** 012058

Evaluating machine learning models in EEG-based thermal comfort studies: a comparative approach

Y Zhou[1], T Zhou[1*], C Wang[1], I Lun[1], J Darkwa[2], D Du[1] and R Zhang[1]

[1] Department of Architecture and Built Environment, University of Nottingham Ningbo China, Ningbo 315100, China
[2] Department of Architecture and Built Environment, University of Nottingham, University Park, Nottingham, NG7 2RD, United Kingdom

*E-mail: Tongyu.Zhou@nottingham.edu.cn

Abstract. Thermal comfort is a critical determinant of human health, productivity, and well-being in indoor environments. While numerous studies have utilised electroencephalography (EEG) to explore human physiological responses to varying thermal conditions, comprehensive analyses that synthesise the effectiveness of various machine learning (ML) approaches for interpreting EEG data remain limited. To address this gap, this study compares various EEG feature sets and ML algorithms using a single EEG dataset. The dataset consists of EEG signals collected from 40 participants exposed to two distinct thermal conditions: a baseline comfortable state and an overheating state induced by wearing heavy clothing. To this end, our objective is to investigate the most pertinent EEG signal features, such as mean power density, power spectral densities, and so on, and evaluate the performance of popular machine learning models for predicting thermal comfort. We examine classifiers including Support Vector Machines (SVM), Random Forests (RF), and various neural network configurations, comparing their efficacy in analysing EEG data. The results indicate that the LDA classifier demonstrates high accuracy when using mean power density features in each 1 Hz frequency range. The SVM classifier, utilizing power density ratios of EEG frequency bands, exhibits robustness in recall and F1 scores. Additionally, the CNN classifier effectively captures complex patterns in the EEG data, showcasing the potential of deep learning methods. These findings contribute to the optimization of indoor environmental controls and advance the field of environmental engineering by providing insights into the neurophysiological impacts of thermal conditions.

Keywords: Thermal comfort, Electroencephalography (EEG), Machine learning, Neurophysiological responses, Indoor environment

1. Introduction

Morden lifestyles have significantly increased the time people spend indoors, estimated to account for more than 90% [1]. This tendency has led researchers to increasingly focus on indoor environmental quality parameters, with thermal comfort emerging as a key factor [2–4]. Thermal comfort reflects individuals' psychological satisfaction with their thermal environment [5,6], and is often assessed using indexes such as predicted mean vote (PMV) [7], standard effective temperature (SET) [8], universal thermal climate index (UTCI) [9] and so on. Despite their widespread use, these indexes often fail to accommodate individual preferences, as they operate as "one-size-fits-all" models [10]. This limitation has led to growing interest in using physiological data, such as electroencephalogram (EEG) signals, to develop personalised thermal comfort models. Especially in this recent decade, machine learning (ML) techniques have been under fast development and shown promise in analysing complex physiological data. Many existing studies have explored ML methods for classifying and predicting occupants' states in indoor environments.

Recently, a number of studies have explored the use of EEG signals in thermal comfort research. Shan et al. [11] investigated the correlations between EEG signals and subjective perceptions, revealing that frontal asymmetrical EEG activity effectively correlates with subjective assessments and task performance under varying thermal conditions. This also suggests that EEG can serve as a more objective metric than traditional questionnaire-based methods. Additionally, Shan et al. [12] lately employed machine learning techniques to classify occupants' real-time thermal comfort states using passive EEG measurements. The findings indicated that classifiers, particularly LDA, achieved high classification accuracy. Ren et al. [4] contributed to this area by measuring physiological parameters (i.e. EEG and skin temperature) to develop predictive models for thermal comfort using various machine-learning algorithms. Their study emphasised the importance of a comprehensive dataset, which improved predictive accuracy significantly when multiple physiological signals were combined. Maruyama's investigation [13] into the pleasantness of wind through EEG signals further underscores that physiological measurements can provide insights into thermal comfort and help enhance indoor environmental quality. Studies by Zhan et. al. [10] and Wu et. al. [14] have demonstrated the efficacy of using EEG to assess individual thermal comfort under different ambient temperatures and environments. Their research indicates that machine learning models, particularly the random forest (RF) algorithm, can accurately predict thermal comfort states based on EEG features, and reinforce the viability of EEG in practical applications for improving indoor climate control systems. Finally, Lee et al. [15] proposed an EEG-based circumplex model of affect, which highlights the potential for identifying individual differences in thermal comfort and emotional states. This growing body of work suggests that EEG can be instrumental in real-time monitoring and adaptation of thermal comfort conditions.

Despite these advances, a lack of comparative analysis across algorithms and features persists, leaving readers unclear about the most effective approaches for EEG-based thermal comfort prediction. This study addresses this gap by evaluating the performance of various machine learning algorithms and EEG feature sets using a common dataset. The dataset consists of EEG recordings from 40 participants exposed to two thermal conditions. By comparing model performance across four evaluation metrics, this research aims to identify the most effective combinations of features and algorithms, providing valuable insights for future applications in personalised thermal comfort and building management.

2. Methodology

This study begins by conducting a comprehensive literature review to identify existing research on EEG-based thermal comfort prediction. Articles were sourced from various databases, focusing on the machine learning algorithms and EEG features employed in these studies. We have collected EEG data on campus from 40 participants under controlled thermal conditions, including baseline comfort and overheating induced by heavy clothing. This dataset was used to reproduce and adapt methodologies identified in the literature. We applied a ten-fold cross-validation technique to ensure reliability. Four performance metrics were used to assess model accuracy and reliability.

2.1 Previous methodologies

To identify relevant studies, we conducted a systematic search across two databases: Google Scholar and Web of Science. Using the keywords ("EEG" OR "electroencephalogram") AND ("thermal environment" OR "thermal comfort") AND ("ML" OR "machine learning"), we retrieved articles published within the past ten years. Based on the titles, abstracts and content, we conducted a preliminary scan of the search results to identify machine learning studies on classifying thermal comfort states with physiological data. A screening of the literature follows to exclude studies that could not be replicated or validated in our experiments. After applying these criteria, five key research articles were identified, as summarized in table 1.

Table 1. Qualified methodologies on using ML to classify and predict EEG data under thermal comfort and discomfort states.

Algorithm	Used Feature(s)	Reference
LDA [16]	The mean power density in each 1 Hz frequency range and three different sets of power density ratio feature	Shan et.al. [11,12] and Wu et. al. [14]
Naïve Bayes (NB) [17]	The mean power density in each 1 Hz frequency range	Shan et. al. [12]
K Nearest Neighbours (KNN) [18]	The mean power density in each 1 Hz frequency range	Shan et.al. [12]
7 different KNNs [18]	Delta wave and theta wave of O2 channel; beta wave of FC5 channel	Pratama et al [19]
SVM [20]	Three different sets of power density ratio feature	Wu et. al. [14]
RF [21]	Delta, theta, alpha and beta waves obtained by Butterworth bandpass filter [22]	Ren et. al. [23]
Back Propagation Neural Network (BPNN) [24]	Delta, theta, alpha and beta waves obtained by Butterworth bandpass filter [22]	Ren et.al. [23]
Convolutional Neural Network (CNN) [25]	Delta, theta, alpha and beta waves obtained by Butterworth bandpass filter [22]	Ren et.al. [23]

IOP Conf. Series: Earth and Environmental Science **1500** (2025) 012058 doi:10.1088/1755-1315/1500/1/012058

2.2 EEG data

We utilised an EEG dataset collected from an experiment involving 40 university students. The experiment was conducted in a classroom with controlled indoor environmental conditions. The air conditioning system maintained a consistent temperature of 26°C (±0.5°C) and relative humidity of around 50% (±5%), which are widely recognized as conditions that typically produce neutral thermal comfort for lightly clothed individuals. Participants initially wore clothing they deemed comfortable and remained in this condition for 30 minutes. Afterward, participants put on a heavy garment designed to deliberately induce thermal discomfort and remained in this condition for an additional 30 minutes. EEG data was collected continuously during both thermal conditions, by using the EPOC X (EPOC+, Emotiv Inc., USA), a portable wireless EEG device equipped with 14 electrodes (F3, F4, FC5, FC6, F7, F8, T7, T8, P8, O1, O2, AF3, AF4). The electrodes follow the international 10-20 system for placement, as depicted in figure 1.

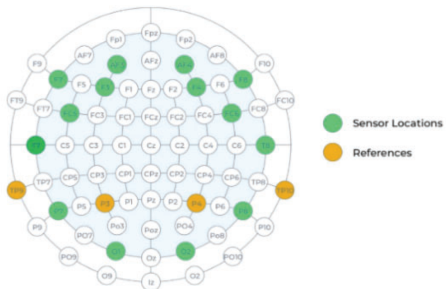

Figure 1. EPOC X electrodes mapping.

The EEG signal data collected from the experiments requires preprocessing to minimise interference and improve the accuracy of the final results. We used the EEGLAB toolbox in MATLAB 2023a for this purpose [26]. The preprocessing steps include:

1. Importing the raw data into EEGLAB and assigning each electrode its channel based on the international 10-20 system.
2. Filtering out data with frequencies lower than 0.5 Hz, higher than 30 Hz, and at 50 Hz.
3. Selecting multiple one-minute segments of stable and high-quality EEG signals from each 30-minute EEG file.
4. Applying the ICA algorithm to identify and remove eye movement and electrocardiogram (ECG) interference.

After preprocessing, we obtained 536 sets of EEG signal in total, with 238 sets of comfort data and discomfort data each.

2.3 Selection of EEG features and algorithms

This study employed five feature extraction methods for processing EEG signals, as summarised in table 1. Equation (1) calculates the first features-the mean power density in each 1 Hz frequency range, where $P(f)$ is the average power density at frequency; $f, X(f, t)$ is the Fourier transform of the signal at time t, and T is the total duration.

IOP Conf. Series: Earth and Environmental Science **1500** (2025) 012058 doi:10.1088/1755-1315/1500/1/012058

$$P(f) = \frac{1}{T}\int_0^T |X(f,t)|^2\, dt \tag{1}$$

The second feature is power spectral density (PSD) of delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), and beta (13-30 Hz) bands. These four bands were isolated by butterworth bandpass filter [22]. Wu et. al. [14], utilised equation (2) to calculate PSD values. Then unit total sum normalisation (UTSN) is applied to standardise the values and normalise results of delta (0.5-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), and beta (13-30 Hz) bands.

$$f^e = \frac{1}{\omega_2 - \omega_1}\int_{\omega}^{\omega} 10\log(|P(\omega)|^2)\, d\,\omega \tag{2}$$

Where $P(\omega)$ is the Fourier transform of the length of the EEG epochs. ω_1 and ω_2 are the upper and the lower frequency bounds of a specific band.

Previous research has indicated the power density ratios between low-frequency and high frequency are more sensitive for assessing human states [27,28]. Therefore, they obtained three sets of features:

$$x^{e1} = \left(f_\delta^e, f_\theta^e, f_a^e, f_\beta^e\right)^T \tag{3}$$

$$x^{e2} = \left(\frac{f_\delta^e}{f_\alpha^e}, \frac{f_\theta^e}{f_\alpha^e}, \frac{f_\delta^e}{f_\beta^e}, \frac{f_\theta^e}{f_\beta^e}\right)^T \tag{4}$$

$$x^{e3} = \left(\frac{f_\delta^e}{f_\alpha^e + f_\beta^e}, \frac{f_\theta^e}{f_\alpha^e + f_\beta^e}\right)^T \tag{5}$$

Where the $()^T$ represents the transformation of vectors.

Pratama et al. [19] used the last feature extraction method, which calculates PSD values of delta and theta frequency ribbons in the O2 channel and beta frequency ribbons in the FC5 channel, respectively, by using the welch function in MATLAB. The features are then normalised to reduce the number of outliers. Various machine learning algorithms are employed by these five studies in table1, covering linear models, ensemble learning, and even deep learning models. Specifically, they include LDA [16], SVM [20], NB [17], different KNN [18], RF [21], BPNN [24] and CNN [25].

2.4 Validation and evaluation methods

This study employs 10-fold cross-validation, a widely-used method, to assess the model's accuracy. The original dataset was randomly divided into 10 non-overlapping subsets, or folds. In each iteration, one-fold is selected as the test set, while the remaining nine folds are used as the training set. The model was trained on the training data and validated on the test data, repeating the process 10 times to ensure each fold was used as the test set once. The average performance across all folds was calculated to provide a reliable estimate of model performance. Four evaluation metrics, accuracy [29], precision [29], recall [29], and F1-score [29], were used to measure the model's classification performance.

IOP Conf. Series: Earth and Environmental Science **1500** (2025) 012058 doi:10.1088/1755-1315/1500/1/012058

3. Results

Subjective questionnaires revealed two distinct thermal states. In the baseline state, the mean thermal sensation was –0.2, indicating a neutral condition. In the overheating state, the score increased to 1.24, corresponding to the “warm” to “hot” range on the PMV scale. The results of comparative analysis of EEG signal classification methods under 2 thermal comfort states are presented in figures 2-5.

Figure 2 illustrates the classification outcomes when the classifiers are trained on the feature of mean power density in each 1 Hz frequency range. The LDA classifier, as depicted in plot (a), demonstrated a median accuracy of 76.25% with a range from 83.40% to 74.10%. The NB classifier, shown plot (b), exhibited median accuracy of 77.56%, with a broader range indicating higher variability. The KNN classifier, presented in plot (c), showed a median accuracy of 76.56%, with precision, and F1 score distributions that suggest a balance between these metrics.

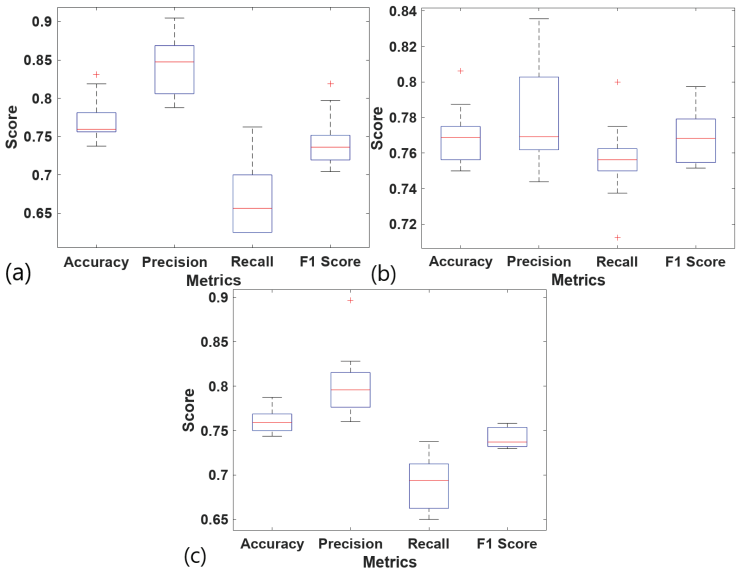

Figure 2. Four metrics in box plot for LDA, NB and KNN classifiers trained by feature - mean power density in each 1 Hz frequency range.

Figure 3 extends the analysis to classifiers trained on features derived from Delta, theta, alpha, and beta waves. The BPNN classifier, as shown in plot (a), achieved a median accuracy of 60.27%. The CNN classifier, displayed plot (b), showed a median accuracy of 77.90%, highlighting its ability to capture complex patterns within the EEG data. The RF classifier, presented in plot (c), achieved a median accuracy of 78.20%.

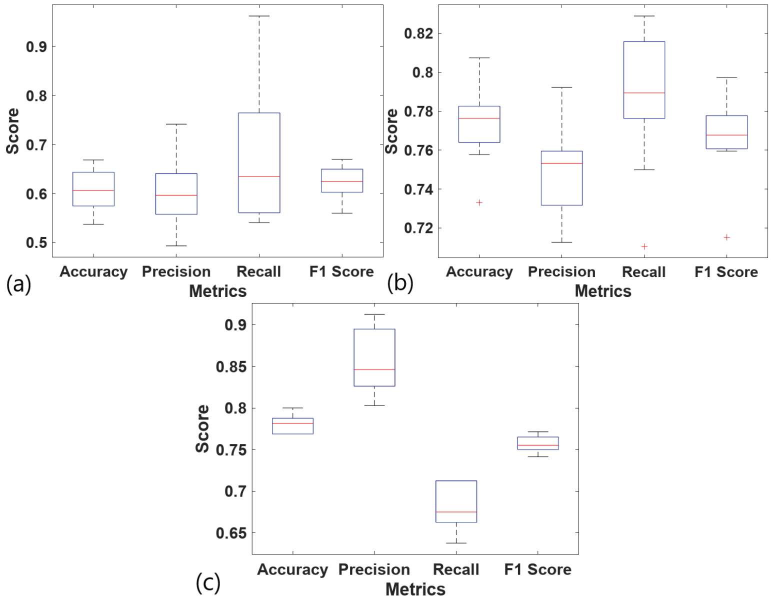

Figure 3. BPNN, CNN and RF classifiers are trained using feature - Delta, theta, alpha and beta waves, and their accuracy and performance are evaluated through 10-fold cross-validation.

In figure 4, box plot (a) and box plot (b) demonstrate the 4 metrics of LDA and SVM classifiers with 3 features. As for LDA classifier, it achieved a median accuracy of 85.05% by using the second feature. It also performed relatively better in other three metrics with the smallest boxes, representing integrated data. SVM classifier achieved the highest median of accuracy, precision, recall and F1 score by using the second feature.

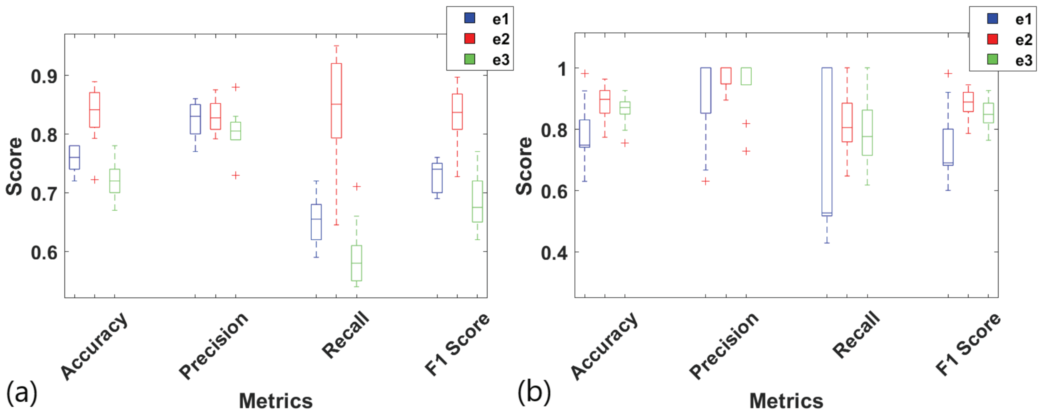

Figure 4. LDA classifier and SVM classifier are trained using 3 feature sets (i.e. explained in equation 3, 4 and 5). In the legend, e1, e2 and e3 represents the 1st, 2nd and 3rd feature sets.

Figure 5 provides a detailed comparison of the KNN classifier with different distance metrics. The six variants—fine-KNN in plot (a), medium-KNN in plot (b), cubic-KNN in plot (c), cosine-KNN in plot (d), coarse-KNN in plot (e), and weighted-KNN in plot (f)—are trained using specific EEG channel features. The box plots reveal that the fine-KNN variant, as shown in the first plot of the first row, achieved a high median accuracy.

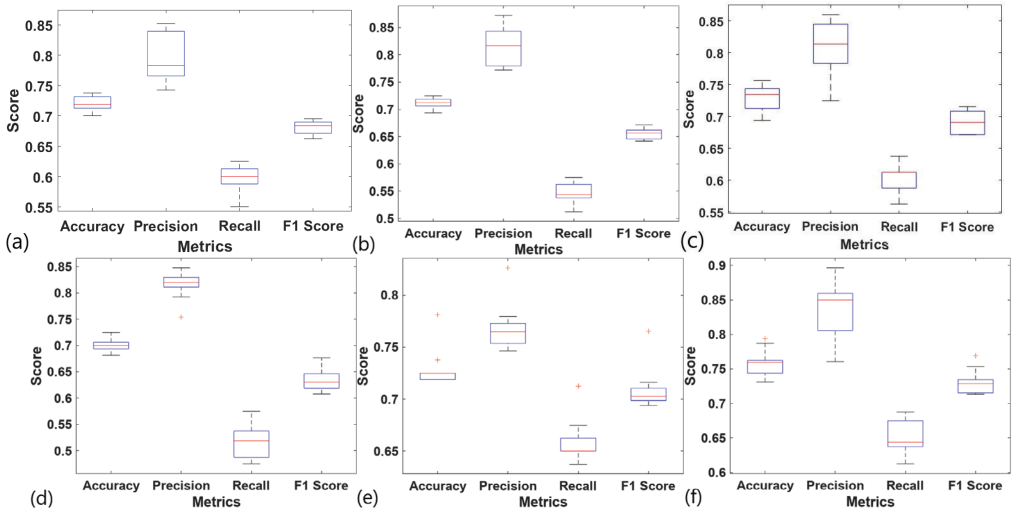

Figure 5. 6 KNN classifiers are trained using feature-Delta wave and theta wave of O2 channel; beta wave of FC5 channel. The first row of plots from left to right are fine-KNN, medium-KNN and cubic-KNN. The second row are cosine-KNN, coarse-KNN and weighted-KNN.

In summary, the results indicate that the choice of classifier and feature set significantly influences the predication performance of thermal comfort states. In the reproduction results, the LDA classifier using feature-mean power density in each 1 Hz frequency range, is in line with expectations and the conclusions of the authors, and its performance is the best among the LDA, NB and KNN classifiers. Although the RF classifier using PSD values of delta, theta, alpha and beta waves as features performs best in terms of accuracy and precision, it sacrifices the recall value. When LDA and SVM classifiers use three different features, the performance of feature-$\left(\frac{f_\delta^e}{f_\alpha^e}, \frac{f_\theta^e}{f_\alpha^e}, \frac{f_\delta^e}{f_\beta^e}, \frac{f_\theta^e}{f_\beta^e}\right)^T$ is better. Finally, when reproducing multiple KNN variants using the same feature, weighted KNN performs best.

4. Discussion and conclusion

This study compares methodologies integrating machine learning techniques with EEG signal analysis to classify and predict thermal comfort states. Using features such as mean power density and power density ratios, several classifiers were evaluated, including LDA, SVM, CNN, and KNN variants. The results indicate that feature selection significantly impacts model performance. The LDA classifier, utilising mean power density in each 1 Hz frequency range, achieved a mean accuracy of 77.06%, demonstrating high accuracy in minimising false predictions. The SVM classifier, which used power density ratios of the delta, theta, alpha, and beta waves, attained a recall of 83.80%, excelling in identifying thermally comfortable states. The balance between precision and recall, as reflected in the SVM's favourable F1 score, highlights its overall robustness.

Additionally, the CNN classifier demonstrated the potential of deep learning methods, yielding high F1 scores and effectively capturing complex patterns in EEG data. Consistent with previous studies, these findings align with prior studies by Shan et al. [12] and Wu et al. [14], corroborating the efficacy of EEG signals for thermal comfort prediction. Future research should explore more extensive datasets, diverse populations, and additional environmental conditions to validate these

IOP Conf. Series: Earth and Environmental Science **1500** (2025) 012058 doi:10.1088/1755-1315/1500/1/012058

findings further. In conclusion, this study contributes to optimising indoor environmental controls by systematically comparing machine learning methods and EEG features for thermal comfort prediction. The integration of neurophysiological data with machine learning presents a promising approach for creating adaptive and personalised indoor environments, paving the way for more effective and efficient environmental controls.

5. Reference

[1] Wargocki P 2013 The Effects of Ventilation in Homes on Health *Int. J. Vent.* **12** 101–118
[2] Costantino A, Ferrara M, Arnesano M and Fabrizio E 2023 Off-the-shelf wearable sensing devices for personalized thermal comfort models: A systematic review on their use in scientific research *J. Build. Eng.* **70** 106379
[3] Pirani M, Best N, Blangiardo M, Liverani S, Atkinson R W and Fuller G W 2015 Analysing the health effects of simultaneous exposure to physical and chemical properties of airborne particles *Environ. Int.* **79** 56–64
[4] Ren J, Tang M and Novoselac A 2022 Experimental study to quantify airborne particle deposition onto and resuspension from clothing using a fluorescent-tracking method *Build. Environ.* **209** 108580
[5] ASHRAE 2017 *Thermal Environmental Conditions for Human Occupancy* (Atlanta, GA: American Society of Heating, Refrigerating and Air-Conditioning Engineers)
[6] de Dear R J and Brager G S 2002 Thermal comfort in naturally ventilated buildings: revisions to ASHRAE Standard 55 *Spec. Issue Therm. Comf. Stand.* **34** 549–561
[7] Fanger P O 1972 *Thermal Comfort: Analysis and Applications in Environmental Engineering* (New York: McGraw-Hill)
[8] Gagge A 1973 Standard effective temperature-A single temperature index of temperature sensation and thermal discomfort *Proc. of The CIB Commission W 45 (Humen Requirements) Symposium, Thermal Comfort and Moderate Heat Stress*, Building Research Sta (HMSO) pp 229–250
[9] Błażejczyk K, Jendritzky G, Bröde P, Fiala D, Havenith G, Epstein Y, Psikuta A and Kampmann B 2013 An introduction to the universal thermal climate index (UTCI) *Geogr. Pol.* **86** 5–10
[10] Zhan J and He W 2022 *15th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)* (Beijing, China: IEEE) pp 1–5
[11] Shan X, Yang E-H, Zhou J and Chang V W-C 2018 Human-building interaction under various indoor temperatures through neural-signal electroencephalogram (EEG) methods *Build. Environ.* **129** 46–53
[12] Shan X and Yang E-H 2020 Supervised machine learning of thermal comfort under different indoor temperatures using EEG measurements *Energy Build.* **225** 110305
[13] Maruyama Y, Nakamura R, Tsuji S, Xuan Y, Mizutani K, Okaze T and Yoshimura N 2024 Classification of pleasantness of wind by electroencephalography *PLOS ONE* **19** e0299036
[14] Wu M, Li H and Qi H 2020 Using electroencephalogram to continuously discriminate feelings of personal thermal comfort between uncomfortably hot and comfortable environments *Indoor Air* **30** 534–543
[15] Lee J, Lee J, Kim T W, and Koo C 2022 EEG-Based Circumplex Model of Affect for Identifying Interindividual Differences in Thermal Comfort *J. Manag. Eng.* **38** 04022034
[16] Xanthopoulos P, Pardalos P M and Trafalis T B 2013 Linear Discriminant Analysis *Robust Data Mining* ed P Xanthopoulos, P M Pardalos and T B Trafalis (New York, NY: Springer New York) pp 27–33
[17] Murphy K P 2006 Naive bayes classifiers *Univ. Br. Columbia* **18** 1–8
[18] Peterson L E 2009 K-nearest neighbor *Scholarpedia* **4** 1883
[19] Pratama R, Achmad B, Avoressi D D, Waruwu M M and Luckyarno Y F 2021 Personal thermal comfort prediction based on EEG signal *J. Eng. Sci. Technol.* **16** 4582–4599
[20] Suthaharan S 2016 Support Vector Machine *Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning* ed S Suthaharan (Boston, MA: Springer US) pp 207–235
[21] Breiman L 2001 Random Forests *Mach. Learn.* **45** 5–32
[22] Zheng H, Pan L and Li T 2023 Research on indoor thermal sensation variation and cross-subject recognition based on electroencephalogram signals *J. Build. Eng.* **76** 107305

[23] Ren J, Zhang R, Cao X and Kong X 2024 Experimental study on the physiological parameters of occupants under different temperatures and prediction of their thermal comfort using machine learning algorithms *J. Build. Eng.* **84** 108676

[24] Li J, Cheng J, Shi J and Huang F 2012 Brief Introduction of Back Propagation (BP) Neural Network Algorithm and Its Improvement *Advances in Computer Science and Information Engineering* ed D Jin and S Lin (Berlin: Springer) pp 553–558

[25] Yamashita R, Nishio M, Do R K G and Togashi K 2018 Convolutional neural networks: an overview and application in radiology *Insights Imaging* **9** 611–629

[26] Delorme A and Makeig S 2004 EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis *J. Neurosci. Methods* **134** 9–21

[27] Chang P, Arendt-Nielsen L and Chen A 2002 Dynamic changes and spatial correlation of EEG activities during cold pressor test in man *Brain Res. Bull.* **57** 667–675

[28] Yao Y, Lian Z, Liu W and Shen Q 2008 Experimental study on physiological responses and thermal comfort under various ambient temperatures *Physiol. Behav.* **93** 310–321

[29] Grandini M, Bagli E and Visani G 2020 Metrics for Multi-Class Classification: an Overview *ArXiv Prepr. ArXiv200805756*