Machine learning-based approach to predict thermal comfort in mixed-mode buildings: incorporating adaptive behaviors

Article

Accepted Version

Zhang, S., Yao, R. ORCID: https://orcid.org/0000-0003-4269-7224, Toftum, J., Essah, E. ORCID: https://orcid.org/0000-0002-1349-5167 and Li, B. (2024) Machine learning-based approach to predict thermal comfort in mixed-mode buildings: incorporating adaptive behaviors. Journal of Building Engineering, 87. 108877. ISSN 2352-7102 doi: 10.1016/j.jobe.2024.108877 Available at https://centaur.reading.ac.uk/115742/

To link to this article DOI: http://dx.doi.org/10.1016/j.jobe.2024.108877

Publisher: Elsevier

Machine learning-based approach to predict thermal comfort in mixed-mode buildings: Incorporating adaptive behaviors

Shaoxing Zhang [a, b], Runming Yao [a, b, *], Jørn Toftum [c], Emmanuel Essah [b], Baizhan Li [a]

[a] Joint International Research Laboratory of Green Buildings and Built Environments (Ministry of Education), Chongqing University, Chongqing, 400045, China

[b] School of the Built Environment, University of Reading, UK

[c] Section for Indoor Environment, Department of Civil Engineering, Technical University of Denmark, Denmark

Abstract

Mixed-mode (MM) buildings are designed to provide mechanical air conditioning and natural passive cooling as regulated by occupants. This would enable the potential of shifting the narrow comfort range in HVAC (heating, ventilation and air conditioning) buildings to a wider range similar to NV (naturally ventilated) buildings. Recent studies have provided evidence of higher degrees of thermal adaptation among occupants in MM buildings. However, limited attention has been given to understanding the linkages between these expanded ranges and the specific adaptive behaviors or contextual factors that influence them. This paper aims to investigate the influence of occupants' adaptive behaviors on thermal comfort in MM buildings. A one-year field study in two MM office buildings with 5,096 valid questionnaires was conducted in Chongqing, China, under hot summer and cold winter climatic characteristics by developing machine learning algorithms compared with classic thermal comfort models. Results show that incorporating adaptive behaviors as input variables enhances the performance of machine learning algorithms, leading to improved overall model performance, while the classic thermal comfort index PMV (predictive mean vote) presents the limited accuracy but the best recall in most cases. This paper also demonstrates that some

energy-inefficient thermal adaptations were found in MM buildings during the HVAC mode, such as using air conditioning in mild spring and autumn, and frequent window openings during cooling periods of summer. It is therefore valuable for future research to further focus on how MM buildings both incorporate positive features and reduce negative features during the HVAC and NV modes.

Keywords: Adaptive thermal comfort, PMV, Adaptive model, Adaptive behaviors, Machine learning

Abbreviations

AC	Air conditioning
Clo	Clothing level
DT	Decision Tree
FN	False negative
FP	False positive
HVAC	Heating ventilation and air conditioning
KNN	K- Nearest Neighbor
Met	Metabolic rate
MM	Mixed mode
NB	Naive Bayes
NV	Naturally ventilated
PMV	Predicted Mean Vote
RH	Relative humidity
RH_{out}	Outdoor relative humidity
SVM	Support Vector Machine
T_a	Air temperature
T_r	Radiant temperature
T_o	Operative temperature
T_{out}	Outdoor air temperature
TN	True negative
TP	True positive
Vel	Air velocity

1. Introduction

Providing thermally comfortable conditions in buildings not only benefits occupants' health, satisfaction, productivity, and well-being [1] [2] [3], but also directly influences building energy usage from HVAC (heating, ventilation and air conditioning) systems [4]. It is suggested that increasing cooling setpoints by 2.8°C (22.2 to 25°C) and decreasing heating setpoints by 1.1°C (21.1 to 20°C) can contribute to total HVAC energy savings of 27% and 34%, respectively [5]. To better understand the essential *cause-and-effect* variables behind thermal comfort, many researchers have focused on developing empirical experiments to study human thermal perception in both well-controlled climate chambers and real-world buildings [6]. Relevant research findings were later employed to develop solid theories for mathematically describing thermal comfort, which have been then successfully incorporated into international and national standards, such as the PMV index in ISO 7730 [7] and ASHRAE 55 [8], adaptive models in EN 16798 [9] and UK's CIBSE Guide A [10], aPMV model in China's GB/T 50785 [11], etc. These thermal comfort models are stipulated by the standards to determine acceptable thermal environments for occupants and to calculate heating/cooling loads for equipment sizing in NV (naturally ventilated) and HVAC buildings.

However, for mixed-mode (MM) buildings with a hybrid strategy of using both natural ventilation and mechanical devices, operable evaluation approaches are still in progress. According to EN 15251 [12], MM buildings should be evaluated using both PMV and adaptive approaches depending on operation modes, whereas ASHRAE 55 updated the applicability of using the adaptive approach from "*must be no mechanical cooling system for the space*" (ASHRAE 55-2004 [13]) to "*no mechanical cooling system or heating system in operation*" (ASHRAE 55-2020 [8]). Although the statement "*in operation*" loses the strict requirements for adopting adaptive comfort theory in buildings, these "*black or white*" binary distinctions could increase the difficulty of applying appropriate thermal comfort models in practice because turning on/off the heating or cooling systems may occur frequently in MM buildings, making constant switch of evaluation methods less feasible from an engineering perspective.

1.1 Adaptive thermal comfort in MM buildings

Many field studies have already been conducted around the world to investigate the "*hybrid philosophy*" in relation to occupants' thermal comfort in MM buildings, as shown in Table 1. Most of the findings suggested that the adaptive model with a wider temperature range is more suitable for MM buildings because MM buildings provide more control opportunities for thermal environments, such as operating windows [14] and adjusting clothes [15], which can positively affect occupants' subjective satisfactions and thermal expectations. One common criticism about using the PMV index in HVAC buildings is that it encourages buildings to operate in a very narrow temperature range [16] and it isolates humans from the natural rhythms of the outdoors [17], resulting in an addiction to narrow artificial environments maintained by avoidable energy consumptions, but this narrow range has the potential to be optimized in MM buildings. Leaman and Bordass [18] compared post-occupancy evaluation (PoE) results from 21 MM and 64 HVAC buildings in the UK and discovered that occupants in MM buildings were more tolerant of changing indoor environments than occupants in HVAC buildings. Thermal comfort surveys from Indonesia [19] and India [20] also demonstrated that the upper limit and general range of comfort temperatures in MM buildings can be extended by 2-3°C and 5°C, respectively, when compared to HVAC buildings. Therefore, adaptive models with a more flexible and wider temperature range are often recommended for assessing thermal environments in MM buildings, particularly during their HVAC mode compared with fully HVAC operated buildings.

Table 1. Thermal comfort research in MM budlings.

Reference	Year	Location	Climate	Building type	Outdoor temperature limits (°C)	Sample size	Key finding
Barbadilla -Martín et al. [21]	2014	Seville, Spain	Temperate	Office	8, 33	5,000 responses from 54 subjects	The neutral temperatures calculated by the MM adaptive model are lower than the NV

							requirements in ASHRAE-55 and EN 15251.
Luo et al. [22]	2015	Shenzhen, China	Subtropical	Office	16, 31	834 responses from 50 subjects	PMV can't account for adaptive behaviors and deviated from actual thermal response in AC mode.
Kim et al. [23]	2017	Sydney, Australia	Subtropical	Residential	9, 27	1,525 responses from 42 homes	Residents in Sydney preferred to use AC and fans rather than open windows.
Rupp et al. [24]	2018	Florianópolis, Brazil	Subtropical	Office	17, 27	5,400 responses from three buildings	Customized adaptive models were developed for both HVAC and NV modes in Brazil context.
Khoshbakht et al. [25]	2019	Brisbane, and Gold Coast, Australia	Mediterranean subtropical	Office	10, 30	1,001 responses from three buildings	Control strategy can significantly affect thermal perceptions.
Kim et al. [14]	2019	Sydney, Australia	Subtropical	Office	15, 22	877 responses from 31 subjects	Occupants adapted more during NV mode compared to HVAC mode.
Ming et al. [15]	2020	Chongqing, China	Humid subtropical	Office	15, 38	827 responses from 29 subjects	Occupants actively adjusted clothes in four seasons.
Jia et al. [26]	2020	Tianjin, China	Humid continental	Office	-12, 38	583 responses from one building	Occupants adapted more in NV mode compared to AC mode. Adaptive models are more accurate than PMV.
Gaffoor et al. [27]	2021	Hybrid	Temperate oceanic	Hybrid	-4,35	1,121 responses from public datasets	Indoor operative temperature is the most influential factor.
Khadka et al. [28]	2022	Tokyo, Yokohama, and	Humid subtropical	Office	7, 34	3,000 responses from 17	The comfortable temperature range in MM mode is wider

Odawara, Japan		building	than in FR (free-running) mode.

On the contrary, the classic PMV index received less positive attention in MM buildings due to its limitation to maximize the potential of energy saving [29] and low predictive accuracy in real buildings [30]. Several field studies reported that occupants' neutral or comfort temperatures in MM buildings during HVAC mode are generally higher (1.5°C in [26], 2°C in [20], and 2.1°C in [29]) than PMV predicts. But a few studies have shown that PMV can accurately predict occupants' actual thermal sensation during the HVAC period of MM buildings [31]. Although current standards recommend clear guidance for evaluating thermal environments in HVAC and NV buildings (PMV [8] for HVAC, adaptive model [10] or aPMV [11] for NV buildings), an effective and distinct approach for MM buildings is lacking and has yet to be developed because a simple combination of stipulations from NV and HVAC buildings may not fully represent the actual thermal adaptations and energy-related behaviors of occupants in MM buildings, resulting in inappropriate or misleading outcomes.

1.2 Machine learning algorithms in thermal comfort research

One of the major trends in thermal comfort research is the shift towards the implementation of personalized models trained by machine learning algorithms [6]. Unlike traditional physical or empirical thermal comfort models that can't be updated or modified with extra input variables in MM buildings (age, gender, climate, season, building type, time of day, etc.), machine learning algorithms are highly data-driven and have the advantage of being able to accommodate new variables and patterns with high accuracy and efficiency [32]. Many studies already employed machine learning algorithms to investigate how specific contextual factors affect indoor thermal comfort, such as wrist temperature [33], facial temperature [34], heartrate variability [35], occupied time [36], gender [37], age [38], etc.

What's more, the emergence of Internet of Things (IoT) technology or digital transformation enables building management to learn about occupants' thermal comfort directly from real-world data generated through new and enormous data sources empowered by smart sensors and devices, such as smartphone [39], wristband [33], thermal camera [34], RGB camera [40], EEG (electroencephalogram) measurements [41], automatic sphygmomanometer [42], etc. These cutting-edge methods have the potential to revolutionize thermal comfort research by analyzing affluent input features and large-volume datasets, allowing complex relationships between variables to be identified.

Previous studies have shown that incorporating additional input dimensions from contextual factors or IoT data collected in climate chambers or field studies can significantly enhance the predictive performance of machine learning-based models. For instance, most studies have reported achieving remarkable accuracy of over 60% [43], which outperforms the predictive accuracy of classic thermal comfort models, such as PMV with 30-40% accuracy [30]. However, merely using metric "*accuracy*" to evaluate a model could be misleading due to randomness when the dataset is imbalanced [43]. What's more, machine learning in thermal comfort research is often employed to enhance the predictive performance of models by embedding additional inputs, but such inputs are typically utilized in a data-driven manner to improve model accuracy rather than to elucidate the underlying mechanisms of thermal comfort. Therefore, although machine learning can improve prediction, it may not necessarily contribute to a deeper understanding of the theoretical foundations of thermal comfort. While current studies have demonstrated that occupants in MM buildings have a wider acceptable or neutral temperature range than in HVAC buildings due to more flexibly adaptive approaches, there is still a need for a quantitative proof of the inner interactions of these extra factors and how they influence occupants' actual thermal comfort. This will provide a more comprehensive basis for designing and operating MM buildings.

1.3 Objective of this study

To address the gaps identified in the previous sections, this research aims to investigate the influence of occupants' adaptive behaviors on thermal comfort in mixed-mode buildings. It explores the specific impacts of adaptive strategies on occupants' comfort experiences and evaluates the effectiveness of machine learning algorithms in capturing and predicting these dynamics compared with classic thermal comfort models. The schematic overview of this research as presented in this paper is shown in Fig. 1, illustrating how each aspect of our methodology contributes to achieving specific objectives. Firstly, a one-year "*right-here-right-now*" field study in MM buildings with both air conditioning and natural ventilation modes was conducted, and relevant adaptive behaviors were recorded as extra model inputs. Secondly, the applicability of two popular machine learning algorithms, namely naive bayes (NB) and random forest (RF), was investigated with and without feature selection process. Their performances were assessed and compared with classic thermal comfort models using four evaluation metrics: accuracy, precision, recall, and F1-score. Finally, adaptive behaviors that can provide a positive impact during the establishment of machine learning models were selected and then examined using the Mann-Whitney test to see if there is statistical significance of thermal environments before and after any specific adaptation was implemented.

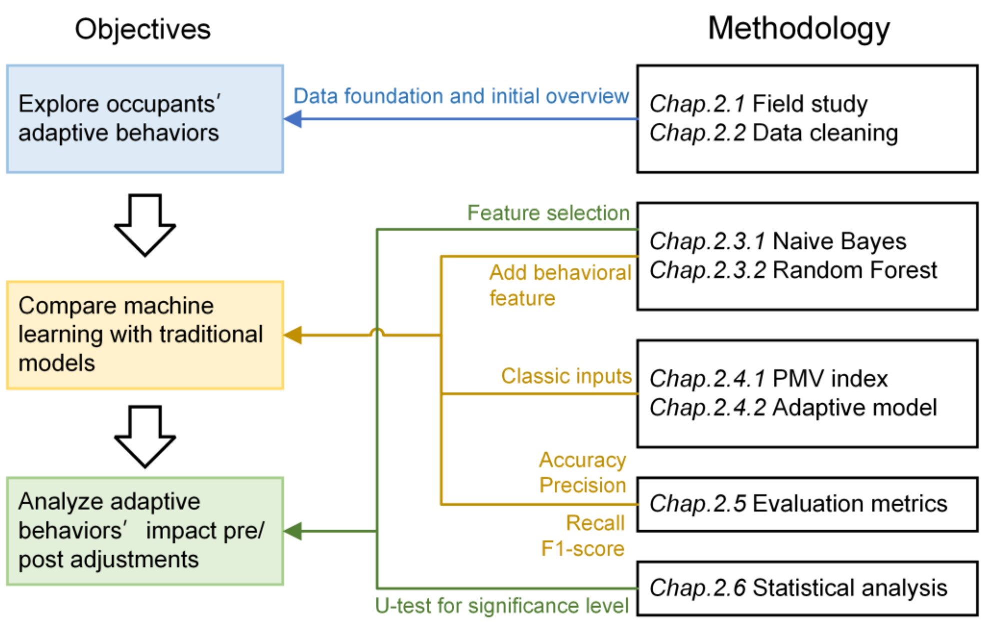

Fig. 1 Schematic overview of research methodology in relation to objectives

2. Methodology

In this study, a one-year investigation was conducted in five offices from two MM buildings at Chongqing University, China. The Chongqing area is situated at 29.56°N 106.55°E and is classified under the Cfa Köppen climate type (humid subtropical), ASHRAE climate zone 3A (warm-humid), and HSCW zone (hot summer and cold winter) in China's Thermal Design Code GB 50176 [44]. Chongqing is recognized as a "*furnace city*" due to its high outdoor temperatures and humidity levels. The daily average temperature exceeds 20 °C for seven months of the year, and the average humidity values consistently remain above 80% [45]. During the winter months, the daily mean air temperature in Chongqing typically ranges from 5 to 10 °C. The city also experiences a low level of sunshine rate in winter, with a rate of only 13% calculated as the length of time with sunshine over the total length of daytime [46].

The *right-here-right-now* field studies have been carried out in five office rooms from two academic buildings at Chongqing University. A total of twelve subjects participated voluntarily with payments and were asked to conduct their typical daily work activities. Monitoring devices were positioned near each subject to gather real-time environmental data. Subjects were required to complete and submit questionnaires hourly at least eight times a day and nine days per month over the entire year through the online survey platform *Wenjuanxing* (https://www.wjx.cn), and over 5,000 valid questionnaires were collected.

2.1 Experimental settings

2.1.1 Environmental monitoring

According to ASHRAE Standard 55-2020 [8], the assessment of acceptable thermal comfort in a steady state requires consideration of six significant factors, namely metabolic rate, clothing insulation, air temperature, radiant temperature, air speed, and humidity. Additionally, in occupant-controlled naturally conditioned spaces, the outdoor air temperature is regarded as a crucial factor in defining acceptable thermal conditions. This study did not include data on radiant temperature and air velocity because: 1) the absence of any significant radiant sources in the investigated offices led to the assumption that the radiant temperature was similar to the air temperature [47]; and 2) it was inappropriate to place an anemometer for each subject to frequently measure air velocity and measurement of air velocity was not feasible within the frame of the project. Given the fact that no ceiling fans or desk fans were used, the air velocity was assumed to be at 0.1 m/s in the steady state.

The air temperature and relative humidity of both indoor and outdoor environments were measured utilizing HOBO UX100-011 data loggers, with ±0.21℃ uncertainty for air temperature and ±3.5% for relative humidity. The data collection was configured with a frequency of 10 seconds. Indoor data loggers located close to each subject while maintaining a safe distance from potential heat sources such as computers and lamps. Outdoor data loggers were placed exterior to each office and shielded from direct sunshine by the implementing an anti-radiation louver ventilation hood.

2.1.2 Field Survey

This study conducted a one-year survey of thermal comfort in office environments from December 2020 to November 2021, involving twelve master and PhD students recruited voluntarily from five different offices in Chongqing, China (Figs. 2 and 3). Students primarily worked in the office but also attended lectures in other classrooms or conducted experiments outside the office. In order to gather sufficient data while minimizing disruptions to the subjects' daily work, they were permitted to select three days within the first, middle, and last ten days of each month to take the online survey during the study period. Subjects were carrying out typical office tasks from 9:00 a.m. to 6:00 p.m., such as talking, reading, writing, and computer operation. Brief departures from the office were allowed, such as having lunch or using the toilet. Subjects can adjust their comfort levels by turning on/off the air conditioner, adjusting the AC set point, changing clothing, drink hot/cold water, and opening/closing windows as needed. After performing these adaptive behaviors, subjects were instructed to document them in the subsequent online questionnaire, which was to be filled once per hour. The detailed statistics of subjects are presented in Table 2.

Table 2. Statistics information of 12 investigated subjects

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
Height (cm)	155	165	180	180	171	173	165	182	160	177	153	166
Weight (kg)	55	60	71	85	67	68	52	73	54	75	50	67
BMI	22.89	22.04	21.91	26.23	22.91	22.72	19.10	22.04	21.09	23.94	21.36	24.31
Gender	F	M	M	M	M	M	F	M	M	M	F	M
Clothing level (clo)	0.70±0.30	0.74±0.36	0.75±0.43	0.65±0.17	0.66±0.32	0.62±0.30	0.96±0.45	0.76±0.26	0.82±0.29	0.51±0.31	1.01±0.22	0.91±0.27
Metabolic rate (met)	1.22±0.24	1.22±0.22	1.10±0.10	1.11±0.13	1.11±0.11	1.21±0.26	1.26±0.27	1.21±0.26	1.13±0.12	1.20±0.18	1.22±0.25	1.40±0.13

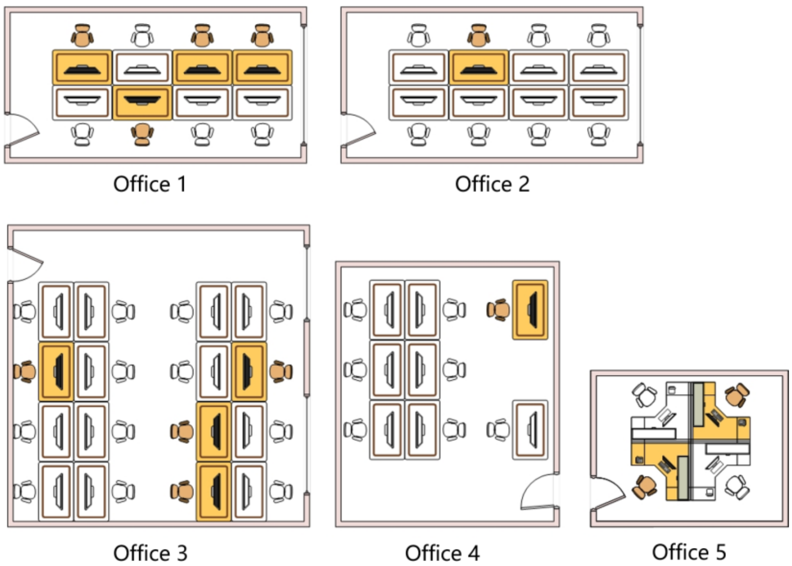

Fig. 2 Layout of investigated offices (surveyed participants are marked in yellow)

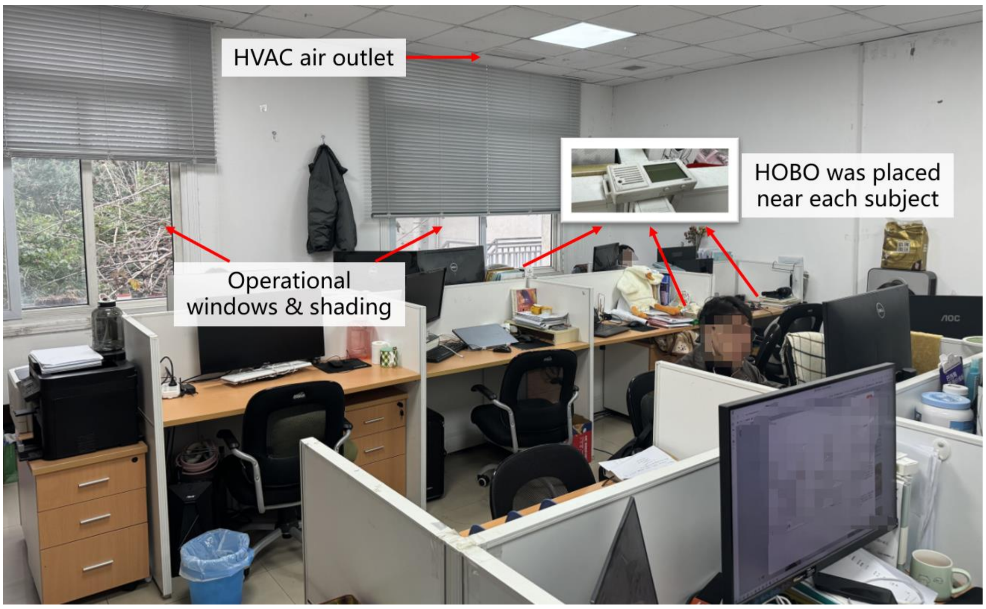

Fig. 3 Field-deployed instruments and occupants of Office 3

To minimize intervening with subjects, two thermal feedbacks were asked: 1) thermal sensation (*continuous ASHRAE scale from -3 cold to 3 hot* [8]); 2) thermal preference (*McIntyre scale of cooler, no change, and warmer*). Additionally, metabolic rate and clothing level were also required to be provided in each questionnaire according to the descriptions in ASHRAE 55-2017 [8] Informative Appendix L '*Measurements, surveys, and evaluation of comfort in existing spaces: parts 1 and 2*'. Thermal preference was chosen as the primary predictive response in this study because it directly indicates thermal discomfort and expresses subjects' preferences in practice, thereby guiding the HVAC system to take appropriate action.

2.2 Data preprocessing

2.2.1 Data cleaning

During the data collection stage, missing values and outliers could occur due to the improper placement, insufficient power and storage space of the data logger. To maintain the quality of the dataset, any data points that had missing values for air temperature, relative humidity, metabolic rate, clothing value, or thermal feedback were removed from the dataset. Outliers, which are typically distinguished by their extremely high or low values, can emerge as a result of dubious responses or incorrect data coding and have a significant impact on regression, model training/building, and overall statistical analysis. As a result, it is critical to accurately identify and effectively address outliers in order to avoid biases or inaccuracies in the results.

In thermal comfort studies, stochastic-based methods, such as the 3-Sigma rule or the Boxplot rule, are commonly used for outlier detection. These methods consider a new observation to be an outlier if its probability density is low relative to the statistical distribution established from prior data. However, it has been found that the 3-Sigma rule may fail to detect outliers, as the presence of extreme values can alter the overall pattern of the data distribution and impact the detection process due to its reliance on mean-based calculations [48]. In contrast, median-based approaches like the Boxplot rule are less susceptible to this influence and tend to produce more reliable results. The

Boxplot rule was employed in this paper for outlier elimination. After removing all the null values and outliers from the original dataset, the remaining data were utilized for further analysis.

2.2.2 Data matching and pre-coding

The data in this study were collected from data loggers and online questionnaires. Each questionnaire represented a single data point, which included several columns of features for further examination. To match the air temperature and relative humidity from the data logger with the questionnaire, the submission time of each questionnaire was located within the timestamps of the data logger near to the subject. The average values of environmental measurements collected one minute before and one minute after the questionnaire's submission time were used to match each questionnaire. For example, if a questionnaire was submitted at 10:31:04, its environmental data would be matched with the average measurements taken between 10:30:04 and 10:32:04.

Label encoding method was used to convert text data into numeric data during the training processes of machine learning algorithms because it was found to achieve higher accuracy than another commonly used method *one-hot encoding* on predicting thermal comfort data [49]. To reduce the imbalanced effects of data units, the input variables were standardized using the function *StandardScaler* in the Python package *Scikit-Learn* package [50]. The detailed descriptions of investigated adaptive behaviors are shown in Table 3

Table 3. Investigated adaptive behaviors and their relation to thermal comfort

Adaptive behavior	Filled value in survey	Relation to thermal comfort
Change AC setpoint	Off or setpoint value	Impact indoor temperature, affect body temperature
Open Window	Four opening levels	Impact indoor air circulation and temperature distribution, provide fresh air
Drink hot water	Y or N	Raise body temperature, improve comfort in cool environments
Drink cold water	Y or N	Lower body temperature, improve comfort in warm environments
Add clothe	Y or N	Maintain body temperature, improve comfort in cool environments
Reduce clothe	Y or N	Lower body temperature, improve comfort in warm environments
Leave office	Y or N	Alter individual heat load and metabolic rate
Have dinner	Y or N	Provide energy and heat, affect metabolism and body temperature regulation
Go upstairs	Y or N	Increase metabolic rate, affect heat generation and dissipation

2.3 Machine learning algorithms

This study compared the predictive outcomes of two machine learning algorithms, naive bayes (NB) and random forest (RF), with traditional thermal comfort models. The NB algorithm is well-known for its quick completion and solid theoretical foundations [51], whereas the ensemble tree method RF has recently gained popularity in the thermal comfort community due to its favourable performance.

2.3.1 Naive Bayes

The Naive Bayes (NB) algorithm is a well-established and straightforward probabilistic method employed in classification problems. This method is rooted in Bayes theorem and assumes that the features involved in the classification task are independent of one another. Despite the fact that this assumption is rarely met in real-world applications, NB algorithm efficient and even outperforms sophisticated rules in terms of efficiency, simplicity, and robustness to missing data and noise [51].

The Bayes theorem states that the probability of a class, given certain features, is proportional to the prior probability of that class multiplied by the probability of the features given that class. In a particular prediction problem, the predictor features $x_1, x_2, \ldots, x_1$ are utilized to estimate the probability of the response category C_k $(k = 1, \ldots, K$ possible categories). This relationship can be formulated as follows [52]:

$$C_k = f(x_1, x_2, \ldots, x_m) \tag{1}$$

The NB algorithm employs the concept of conditional probability as:

$$P(C_k|x_1, x_2, \ldots, x_m), \qquad \forall\, k = 1, \ldots, K \tag{2}$$

This posterior can be represented through the Bayes Theorem as:

$$P(C_k|x_1, x_2, \ldots, x_m) = \frac{P(x_1, x_2, \ldots, x_m|C_k)P(C_k)}{P(x_1, x_2, \ldots, x_m)} \tag{3}$$

where $P(x_1, x_2, \ldots, x_m|C_k)$ is the likelihood of the features given the class, $P(C_k)$ is the prior probability of the class, and $P(x_1, x_2, \ldots, x_m)$ is the evidence.

The fundamental simplification in the NB algorithm is the assumption of independence between features, meaning that the likelihood of the features given a class can be determined as the product of the individual likelihood probabilities:

$$P(x_1, x_2, \ldots, x_m | C_k) = P(x_1|C_k)P(x_2|C_k) \ldots P(x_m|C_k) \quad (4)$$

Then, equation (3) can be transformed by incorporating equation (4):

$$P(C_k|x_1, x_2, \ldots, x_m) = \frac{P(C_k)}{P(x_1, x_2, \ldots, x_m)} \prod_i^m P(x_m|C_k) \quad (5)$$

Since the evidence term $P(x_1, x_2, \ldots, x_m)$ is the same across all classes, the NB classifier selects the class with the highest posterior probability as the final prediction, given the observed features, as expressed by the following equation:

$$\hat{y} = \underbrace{argmax}_{k\in\{1,2,\ldots,K\}} P(C_k) \prod_i^m P(x_m|C_k) \quad (6)$$

The NB algorithm has been widely applied in various domains related to energy and buildings, including battery health management [53], investment in renewable energy [54], and smart grid systems [55]. Despite criticisms of the NB approach for its overly simplistic assumption of independence between features [56], there are several reasons why it often exhibits good performance, including [51]: 1) this intrinsic simplicity often results in low variance in the probability estimate; 2) the potential biased assumption may not influence the outcome directly as long as the rank order is correct; and 3) simple extension to original NB structure can further improve its performance.

To ensure reproducibility in this research, the training and testing data were split in an 8:2 proportion for each $NB_{default}$ model training under the *random state* of 42 by *Scikit-Learn* package (the fixed seed value of the *random state* ensures that the sample splitting results are consistent [50]). Rather than using all features for the NB model, default inputs from the classic PMV index were used to train the initial $NB_{default}$ model to better grasp the knowledge from prior thermal comfort research. Subsequently, one environmental or behavioral feature from questionnaires was added to the initial inputs to train an updated model NB_{update}, and the evaluation metrics were observed to see if they improved. If the performance was improved with the addition of a specific feature, that feature was marked as *selected* and included in the final training model NB_{select} (Fig. 4). The above process was carried out separately for data collected in each season.

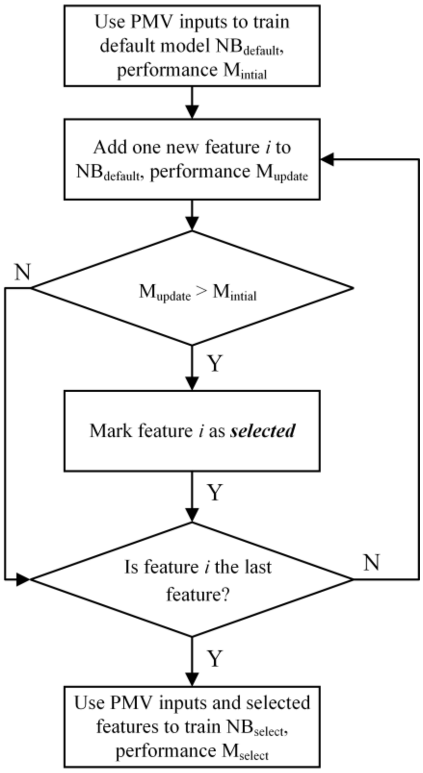

Fig. 4 Diagram of training NB models for one season

2.3.2 Random Forest

Random Forest (RF) is a stochastic-based machine learning algorithm that is known for its robustness and ability to avoid overfitting of a single tree. Traditional tree-based models suffer from the loss of generalization accuracy on unseen data, and the RF algorithm can overcome this by combining and training multiple decision trees on random subsets of data, then aggregating the results from each tree with selected variables to produce a final prediction [57].

One of central hyperparameters in the RF setting is *mtry*, which is defined as the number of randomly selected candidate variables from which each split is chosen, and is often set to the square root of the input number in classification problems [58].

$$mtry = \sqrt{p} \tag{7}$$

where p is the number of inputs.

The splitting criterions of the RF algorithm include two main kinds: "*Gini*" and "*entropy*". Gini (or Gini impurity) index measures how often a randomly selected element from the set would be labelled incorrectly if it were labelled randomly according to the distribution of labels in the subset [59]. It is calculated through:

$$Gini = 1 - \sum_{i=1}^{k} {p_j}^2 \tag{8}$$

where k is the number of classes, and p is the probability of class j.

The entropy index explains a set's disorder or randomness, with lower entropy indicating greater order or structure [59]. It is calculated through Shannon entropy [60]:

$$Entropy = -\sum_{i=1}^{k} p_j log_2(p_j) \tag{9}$$

where k is the number of classes, and p is the probability of class j.

Both splitting criteria are effective in many applications [46], but Gini impurity is faster to compute and preferable in large datasets because it is a linear measure as opposed to Entropy's logarithmic measure. The entropy is more sensitive to changes in the distribution of classes and is preferred for maximizing information gain at each split [61].

The remaining tuning hyperparameters in the RF algorithm include *number of trees, maximum depth,* and *node size*: **1) Number of trees** (or estimator) determines the total number of trees in the forest. A larger number of trees usually results in a more robust model, but at the expense of increasing computation time and becoming too complex, and it can begin to fit the noise in the dataset rather than the underlying patterns; **2) Maximum depth** represents the maximum depth of each tree, and small max depth could lead to a shallow tree and poor performance on training data, whereas a large max depth will result in overfitting and poor generalization to unseen data; and **3) Node size** (or minimum sample leaf) calculates the minimum number of samples required to arrive at a leaf node, which serves as the final decision point, and generates a prediction at a tree for a given input sample [62].

Above hyperparameters in the RF algorithm control the splitting principles and the randomness of aggregating trees, thus influencing the model performance. However, Mantovani et al. [63] discovered that the effects of tuning hyperparameters in the RF algorithm are much smaller than other machine learning algorithms, such as support vector machines (SVM). In building engineering, the RF algorithm has been extensively utilized and achieved appropriate results [64][65][66], and thermal comfort related studies are as listed in Table 4. Most of these studies reported acceptable predictive accuracy levels ranging from 70% to 90%. In comparative performance studies, the RF algorithm consistently demonstrated outstanding performance, often with the highest accuracy or only a minor difference from the top-performing algorithm.

Table 4. Random forest algorithms used in thermal comfort studies

Publication	**Matrix**	**RF**	**SVM**	**AdaBoost**	**DT**	**KNN**	**NN**	**NB**	**Logistic**	**PMV**	**Specific focus**
Luo et al. [49]	Accuracy	**65.1%**	57.7%	61.6%	57.9%	61.7%	60.4%	52.3%	59.9%	43.1%	Algorithm comparison
Gao et al. [67]	Accuracy	**51.41%**	37.93%	42.94%	43.33%	41.43%	50.35%	40.43%		33.35%	Climate difference
	F1-Score	**52.93%**	40.91	42.41%	43.34%	41.93%	50.67%	39.40%			
Zhang et al. [68]	Accuracy	**70.6%**	47.1%		49.6%		54.6%				Office building
Yu et al. [69]	Accuracy	79.4%		77.8%		**80.2%**					Feature selection
Aryal and Becerik-Gerber [70]	Accuracy	78%	77%		76%	**80%**					Personalized comfort systems
Li et al. [71]	Accuracy	**72.0%**	57.4%				67.7%				Individual prediction
	Kappa	**50.5%**	0				40.4%				
Farhan et al. [72]	Accuracy	52.1%	**56.7%**	51.1%						33.2%	Individual prediction
Vellei et al. [73]	Error	**0.2**			0.26				0.42		Relative humidity influence
Li et al. [74]	Accuracy	85%									Facial infrared thermography

Chaudhuri et al. [75]	Accuracy (Male)	92.86%	Gender difference
	Accuracy (Female)	94.29%	
Shetty et al. [76]	Accuracy	97.73%	Desk fan usage
Liu et al. [77]	Accuracy Kappa	77% 43%	Prediction using wrist temperature
Wu et al. [78]	Accuracy	70.4%	Prediction using local skin temperature

In this study, 20% of the data was divided into test samples with a *random state* of 42, and the grid search method under 5-fold cross-validation was used to evaluate RF predictive performance by identifying the best parameters among the above four hyperparameters within the range of [10, 25, 50, 100, 200] for a number of estimators, ["*Gini*", "*entropy*"] for criterion, [3, 5, 10, 15, 20, 25] for max depth, and [1, 2, 5, 10] for min sample leaf. The RF models have been fed with two rounds of input features: round one uses all features and round two uses the features selected from the NB_{select} modes.

2.4 Classic thermal comfort models in standards

2.4.1 PMV index

Developed by Fanger during 1970s, PMV (Predictive Mean Vote) index was based on American and European experiments in well-controlled climate chambers, which views thermal comfort as a physiological phenomenon of the human body and considers thermal sensation to be the result of heat transfer between the human body and its surroundings [8]. The PMV index has been widely adopted for evaluating the HVAC operation in thermal comfort standards (ISO 7730 [7], ASHRAE 55 [8], EN 16798 [9], CIBSE Guide A [10], and GB/T 50785 [11]). Despite its widespread acceptance, the PMV index also faces challenges in its application to built environments [6]: 1) its two environmental inputs, "radiant temperature" and "air velocity", require expensive instruments and human assistance for accurate measurement, and two personal inputs, "clothing level" and "metabolic rate", are often simplified or assumed due to the difficulty in automated collection; 2) The index was designed to predict average comfort of a large population, but usually performs poorly at an individual level; 3) It provides limited opportunity for further adaptations or updates, as all relationships between inputs and outcome are clearly stated.

Several thermal comfort studies have revealed that the PMV index only provides approximately half the accuracy compared with fine-tuned machine learning algorithms, as demonstrated in Table 1. To further assess the performance of this index, this study uses the boundary of ±0.5 to represent the hot and cold limits for evaluating thermal comfort, which is classified as Category II in ISO and EN standards, and Category I in American, UK, and Chinese standards. The PMV values were calculated using Python package *pythermalcomfort* developed by Tartarini and Schiavon [79].

2.4.2 Adaptive model

The "*one size fits all*" feature of the PMV index could not only lead to poor satisfaction at the individual level but also risk expending a significant amount of energy in maintaining a uniform indoor environment in buildings year-round, regardless of outdoor climates and individual preferences or adaptations. To address these limitations, many field studies have been conducted in actual buildings to validate the adaptive comfort concept in various climatic zones. In 1976, Humphreys performed a meta-analysis of over 30 field surveys conducted between 1930 and 1975, utilizing over 200,000 records to further promote the adaptive principle [80]. He discovered that the relationship between indoor preferred temperature and monthly mean outdoor temperature was linear in naturally ventilated buildings, whereas curvilinear in heated or cooled buildings. This linear relationship of the adaptive model in the most recent ASHRAE 55-2020 standard was stimulated for determining acceptable indoor temperature in naturally conditioned spaces as [8]:

$$Upper\ 80\%\ acceptability\ limit = 0.31\overline{t_{pma(out)}} + 21.3 \tag{10}$$

$$Lower\ 80\%\ acceptability\ limit = 0.31\overline{t_{pma(out)}} + 14.3 \tag{11}$$

where $\overline{t_{pma(out)}}$ is the prevailing mean outdoor air temperature, which is a weighted-mean value from the previous days (from 7 to 30 days [8]) with decreasing weighing factors.

Although this linear regression-based model includes the adaptive concept that people's neutral temperature can vary with changes in outdoor thermal conditions, it is primarily based on statistical analysis and does not provide detailed explanations of how the human body is affected by its surroundings, as the PMV index does [81]. To better examine its predictive performance, the comfort boundaries of the adaptive model in ASHRAE 55 were also used in this paper as comparative benchmarks for the analysis in the spring and autumn seasons.

2.5 Evaluation metrics

The performance of thermal comfort models developed using NB and RF machine learning algorithms, as well as classic PMV/adaptive models, was evaluated using a confusion matrix: accuracy, precision, recall, and F1-Score:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{12}$$

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

$$Recall = \frac{TP}{TP + FN} \tag{14}$$

$$F1 - score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} \tag{15}$$

where TP is true positive, FP is false positive, TN is true negative, and FN is false negative.

Accuracy is the most commonly used evaluation metric in thermal comfort studies (Table 1) because it indicates the fraction of successful predictions among all records and provides a clear understanding of how frequently the established model is correct. However, using only one index to evaluate model performance may produce misleading results. The metric of accuracy can be misleading when the class distribution is imbalanced, such as in a binary labeled dataset where only 1% of the data are labeled as "+". If the model predicts all examples as "-", it will still have an accuracy of 99% [77]. To address this issue, precision can be used, which is defined as the ratio of true positive predictions to all positive labels. In the 1% "+" example mentioned above, the lack of true positive predictions would result in a precision of 0, as the numerator in equation (9) would be 0. In other words, precision indicates how much the model can be trusted when it predicts an individual as positive. Therefore, precision is frequently prioritized in situations where *false positives* are costly, such as a stock trading system in which predictive buying signals are extremely important despite missing a few opportunities [82]. Conversely, recall measures the model's ability to identify all positive units in the dataset [43]. If *false negatives* are important, as in medical diagnosis, where identifying all potential cases of disease is important even if it means finding healthy people incorrectly [83], recall becomes the priority. To account for both *false positives* and *false negatives*, the F1 score, which provides a balanced average of precision and recall, is recommended. Therefore, indicators such as precision, recall, and F1-score, which provide more detailed performance information, are critical in evaluating the outcomes of machine learning algorithms [84].

As thermal preference was used as a predictive response in three classes: "*warmer*", "*no change*", and "*cooler*" (Fig. 5, rows and columns represent actual and predicted classes, respectively), the corresponding precision, recall, and F1 score for 3-class problems were examined across all classes by using the unweighted mean of the metric score.

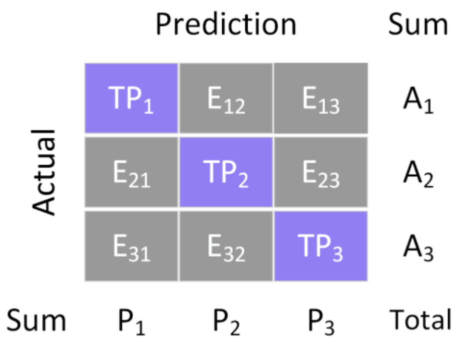

Fig. 5 Confusion matrix for 3-class classification

Then the overall confusion matrix can be updated as [85]:

$$Accuracy = \frac{\sum_{i=1}^{3} TP_i}{Total} \tag{16}$$

$$Precision\ (macro) = \frac{1}{3}\sum_{i=1}^{3}\frac{TP_i}{P_i} \tag{17}$$

$$Recall\ (macro) = \frac{1}{3}\sum_{i=1}^{3}\frac{TP_i}{A_i} \tag{18}$$

$$F1 - socre\ (macro) = \frac{1}{3}\sum_{i=1}^{3}\frac{2}{\frac{1}{\frac{TP_i}{P_i}} + \frac{1}{\frac{TP_i}{A_i}}} \tag{19}$$

where TP_i is the number of correct predictions in class-i, E_{ij} is the number of wrong prediction *j* in class-i, A_i is total sum of actual samples in class-i, P_j is the total sum of prediction *j*, and *Total* is the overall number of samples. Equations (13) to (15) denote that these macro indices place particular emphasis on *false positives*, *false negatives*, and their combined effects in 3-class classification problems.

In this research, the relevant calculations were also achieved by the functions *accuracy_score*, *precision_score, recall_score*, and *f1_score* in Python's *Scikit-Learn* package [50]. In particular, the parameter "*average*" in each *score* function was set to "*macro*" to evaluate the precision, recall, and F1-score of multi-class classification problems by giving equal weight to each class.

2.6 Statistical analysis

To investigate the statistical significance of adaptive behaviors on thermal preference, several statistical analyses were performed based on the Mann-Whitney test. The Mann-Whitney test (also known as the U test) is a non-parametric test that compares the medians of two independent samples to determine whether data from two populations are significantly diverse [86]. It performs well on small datasets and does not require the normality assumption in the data distribution as the parametric t-test does [87]. The Mann-Whitney test has already been used in several thermal comfort studies to determine whether there was a significant difference between data from two groups, for example, between PMV and TSV [88], between comfort state and discomfort state [89], as well as the attitudes of different bill payers toward programmable thermostats [90]. The Mann-Whitney test carried out in this paper was implemented using function *add_stat_annotation* in Python package *statannot* [91].

3. Results

3.1 Data overview

After removing null values and outliers from the original dataset (n: 5,135), a total of 5,096 data points remained available for analysis, as shown in Fig. 6. Due to participants being in their daily work, occasional forgetfulness may result in missed questionnaire submissions. While efforts were made to remind them, we did not force subjects to fill out the questionnaire. The final valid response rate for collected questionnaires is approximately 50%. Our study implemented several measures to mitigate participant fatigue and ensure the reliability of responses. These included a concise questionnaire design, a web-based data collection platform, limited experimentation days, and staged monetary compensation. Analysis of abnormal responses in Fig. 7, particularly in relation to deviations from optimal air temperature ranges in ISO 7730, reveals a low incidence of anomalous behavior. Further examination of individual responses, such as those from subjects 2 and 12, highlights a correlation between elevated metabolic rates and preferences for cooler ambient conditions. These findings collectively underscore the conscientious approach of participants towards questionnaire completion, reinforcing the validity of our study's data collection process.

Table 5 summarizes the statistical information of the dataset over this one-year experiment in four seasons. Because of the Spring Festival vacation and students' busy lectures or other experiment schedules, the data sums in winter and autumn are relatively low compared with spring and summer. In general, the indoor and outdoor temperatures are highest in the summer, lowest in the winter, and moderate in the spring and autumn. The average values of outdoor air temperatures during autumn (14.78°C) and winter (7.12°C) are lower compared to spring (19.28°C) and summer (27.82°C). The five offices studied are located in MM buildings, allowing occupants to adjust their thermal comfort through various means such as turning on/off air conditioning, opening/closing windows, changing clothing, drinking hot/cold water, etc. Fig. 8 depicts the relationship between variables as presented in Table 5, with the matrix diagonal showing the density plot of the three thermal preferences ("*no change*", "*warmer*", and "*cooler*") through different colors. It is evident that the "*No change*" preference was prevalent, while the "*Warmer*" preference was more commonly reported at low indoor/outdoor air temperatures, high indoor humidity levels, and low clothing levels. As the subjects carried out their daily office work regardless of the thermal environment, their metabolic rates appear to remain relatively constant.

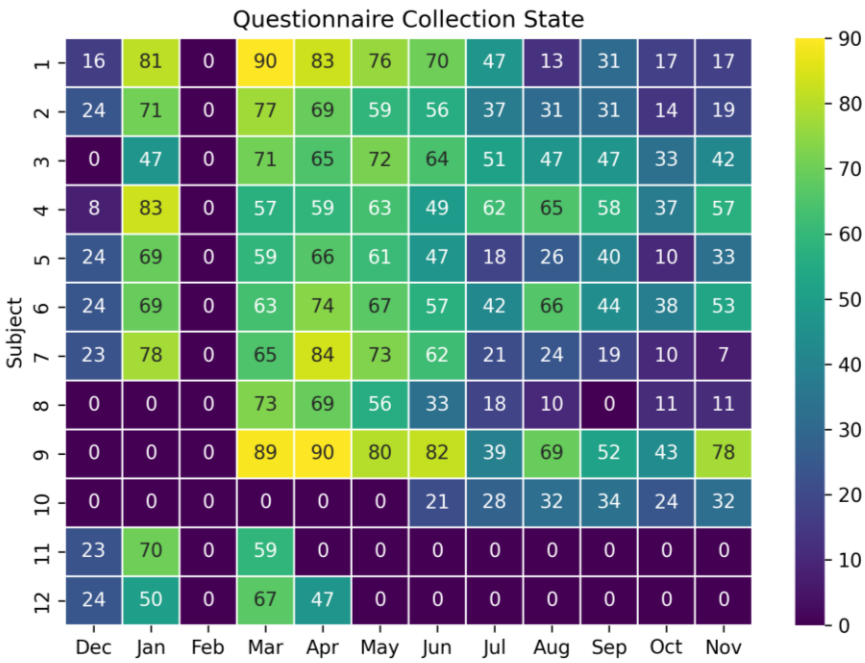

Fig. 6 Questionnaire collection state of the one-year experiment

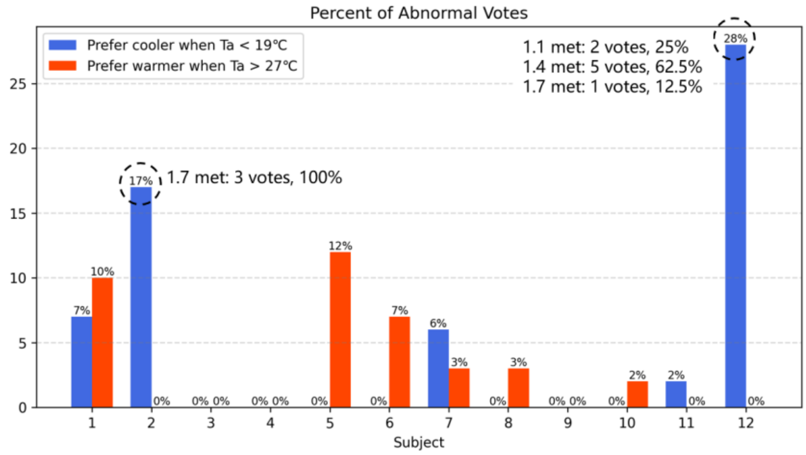

Fig. 7 Percent of abnormal votes when air temperature is below 19℃ or beyond 27℃

Table 5. Results of investigated environmental and subjective parameters

Parameter	Winter	Spring	Summer	Autumn
Indoor Air temperature (℃)	21.55 ± 3.02	23.33 ± 2.41	26.40 ± 1.26	22.13 ± 2.07
Outdoor Air temperature (℃)	7.12 ± 1.56	19.28 ± 3.47	27.82 ± 3.47	14.78 ± 3.58
AC setpoint (℃)	23.53 ± 2.55	23.88 ± 1.90	25.03 ± 1.38	25.04 ± 3.05
Indoor humidity (%)	35.78 ± 8.24	58.91 ± 8.32	58.61 ± 8.94	54.91 ± 10.38
Clothing level (Clo)	1.17 ± 0.31	0.77 ± 0.24	0.43 ± 0.13	0.92 ± 0.30
Metabolic rate (Met)	1.24 ± 0.24	1.19 ± 0.21	1.16 ± 0.20	1.15 ± 0.16
PMV	0.12 ± 0.71	0.08 ± 0.66	0.38 ± 0.60	-0.11 ± 0.67
Data sum	784	2083	1643	586

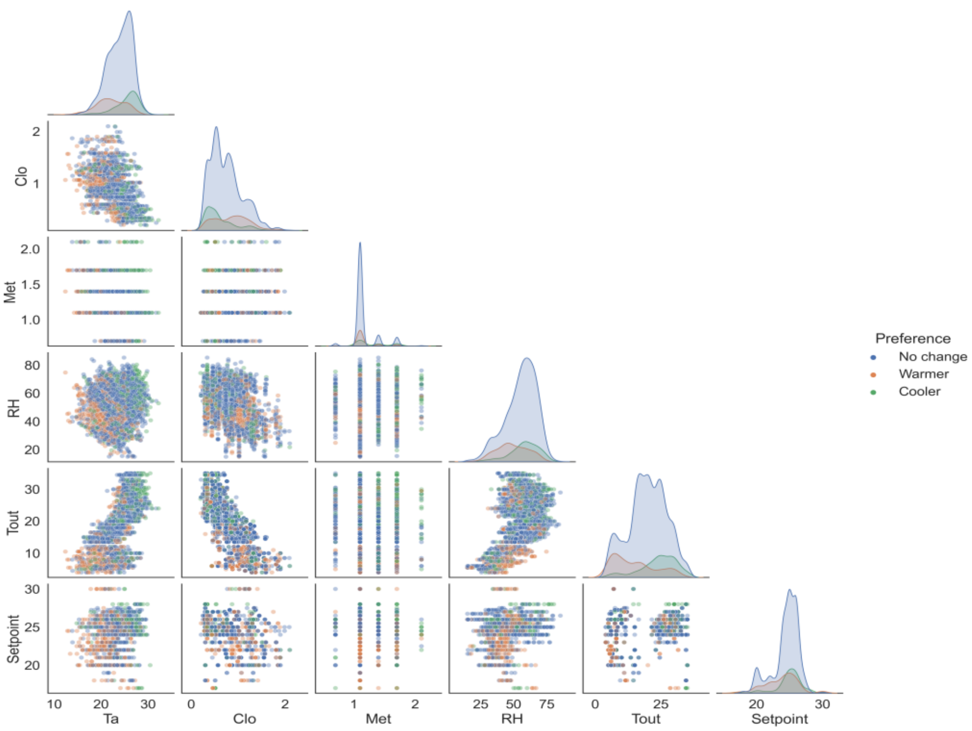

Fig. 8 Relationship between different variables

Fig. 9 shows the adaptive behaviors and thermal preferences among questionnaires. As Chongqing is located in the hot summer and cold winter (HSCW) climate zone in China, it shows typical needs of cooling in summer and heating in winter. In these mix-mode offices, investigated subjects turned to use air conditioning frequently in summer (73%) and winter (96%) compared with spring and (13%) autumn (13%). Despite the fact that the buildings were air conditioned 96% of the time in winter, the status of windows being completely closed was only 19%. Subjects also drank water for roughly half of the time in order to actively maintain thermal comfort, with the intake of cold water being highest in summer (28%) and hot water highest in winter (42%). For general thermal preference in each season, subjects in spring expressed a high level of satisfaction of 82%, which is higher than standard requirements that "*a thermal environment that a substantial majority (more than 80%) of the occupants find thermally acceptable*" [8]. This percentage, however, is not met in the other three seasons, with obvious cooling needs in summer (23%) and heating needs in autumn and winter (35%) even under the operation of HVAC systems.

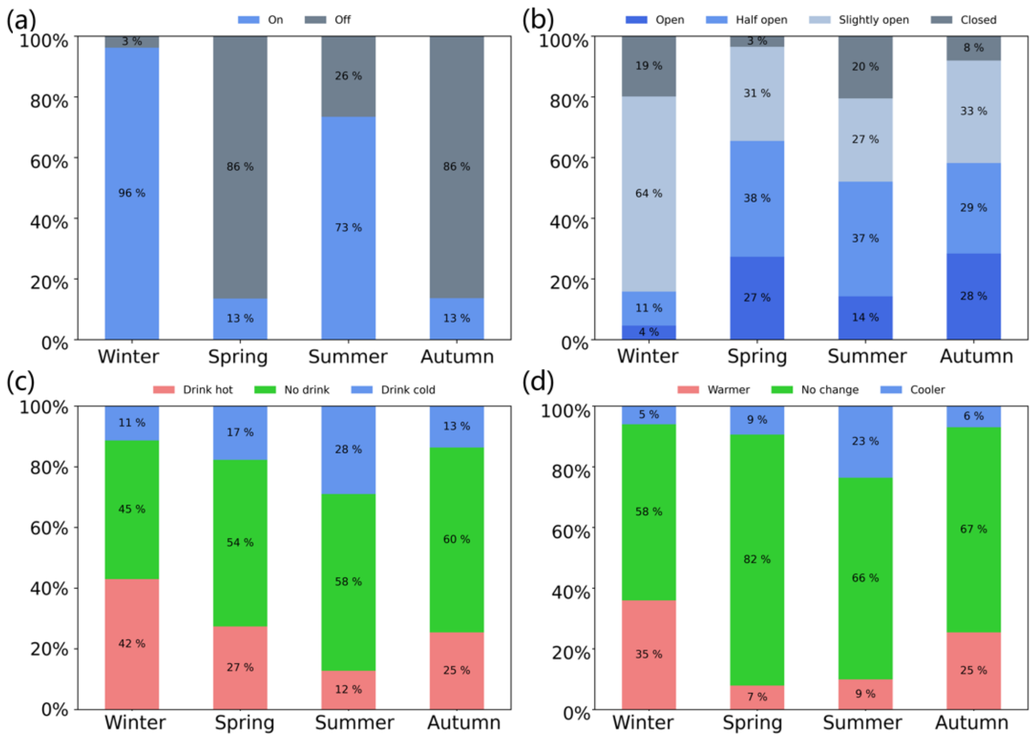

Fig. 9 Distribution of adaptive behaviors and thermal preference: (a) AC on/off; (b) windows open/close; (c) drink water; (d) thermal preference

The classic PMV index, developed by Fanger, incorporates two subjective factors, namely "*metabolic rate*" and "*clothing level*" into its numerical calculation of thermal comfort to represent the heat generated within the body and the heat retention/loss due to the insulation and the air permeability of clothing [92]. Since the participants in this study performed daily office work such as typing, reading, or talking, their metabolic rates remained relatively constant across different seasons or thermal environments. Therefore, the adaptation of clothing level is depicted in Fig. 10 at binned intervals of one degree of outdoor air temperature. In Table 5, the mean values of clothing level are 1.17, 0.77, 0.43, and 0.92 for winter, spring, summer, and autumn, respectively. Fig. 10 illustrates the decreasing linear relationship between outdoor air temperature and clothing level in spring (orange color) and autumn (red color). However, during summer and winter, the clothing level remains consistently low (green color) and high (blue color) with limited opportunities for people to adjust.

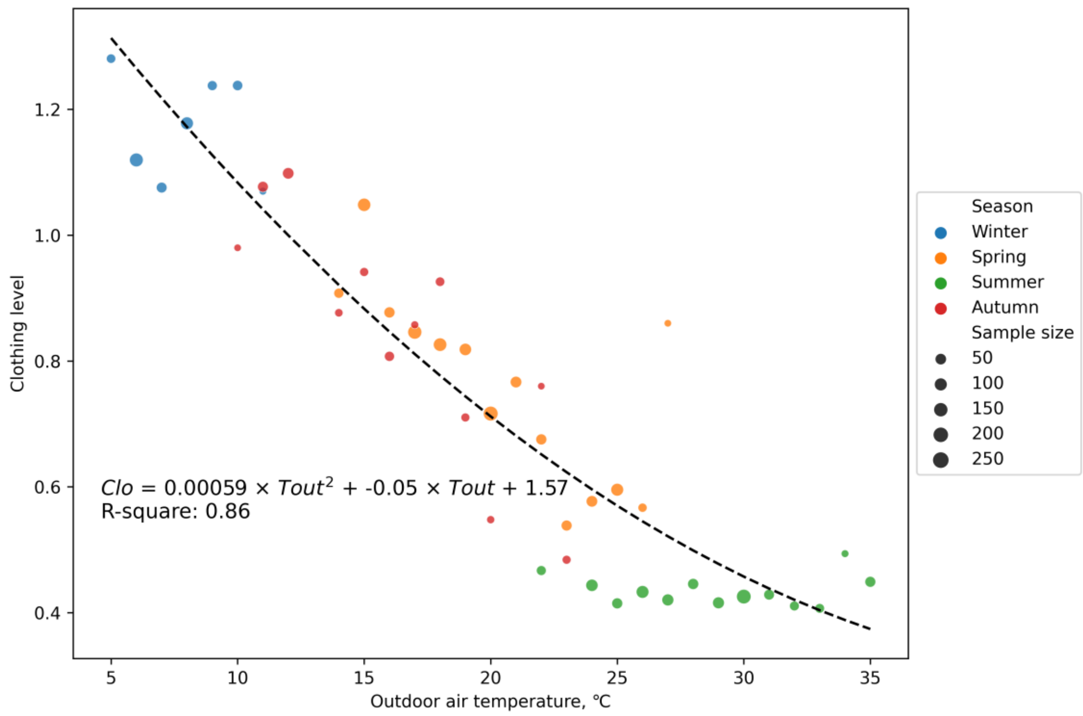

Fig. 10 Variations of clothing insulation in four seasons based on outdoor air temperature

3.2 Predictive performance of machine learning algorithms

3.2.1 Naive Bayes

Instead of randomly adding or throwing all features into the NB model, four PMV index inputs, namely T_a, Clo, Met, and RH, were used to train the initial $NB_{default}$ model, as T_r and Vel were not collected during the experiment. Table 6 shows the predictive performance of each $NB_{default}$ model over the four seasons. Each cell in Table 6 with a grey background indicates that the $NB_{default}$ model predicts data from that season, whereas cells without a grey background indicate that the $NB_{default}$ model predicts data from other seasons. The bold number represents the best evaluation metrics of each $NB_{default}$ model across four seasons. The winter $NB_{default}$ model shows the least generalizable potential, with only 12% accuracy in predicting spring data and 10% F1-score in predicting summer data. However, this poor generalization did not appear significantly among other $NB_{default}$ models in spring, summer, and autumn, with accuracies ranging from 53% to 81% and F1-scores ranging from 31% to 44%. Because spring data is the most imbalanced, with 82% "*no change*" votes (Fig. 9d), the $NB_{default}$ models trained from spring, summer, and autumn data all performed best accuracy on spring test data (81%, 75%, and 78%), posing the risk of "*accuracy cheating*", in which the model turns to generate more "*no change*" outcomes to improve accuracy. As precision and recall are concerned with predictive positives and actual positives, respectively, a trade-off between these two metrics is usually unavoidable under certain conditions [93]. This is also supported by Table 6, which shows that the best precisions and recalls do not appear simultaneously on test data from the same season. The high precisions in summer test data (58%) indicate that the summer $NB_{defualt}$ models are more likely to ignore tags than add incorrect tags [94].

Table 6. The prediction power of $NB_{default}$ model for each season

	Winter test data				Spring test data				Summer test data				Autumn test data			
	A	P	R	F1	A	P	R	F1	A	P	R	F1	A	P	R	F1
Winter $NB_{default}$	**62%**	**43%**	**41%**	**40%**	12%	30%	34%	8%	13%	25%	30%	10%	34%	32%	34%	23%
Spring	53%	**53%**	39%	31%	**81%**	35%	**40%**	37%	69%	45%	39%	**37%**	66%	32%	38%	34%

NBdefault																
Summer NBdefault	62%	47%	40%	40%	**75%**	42%	41%	42%	70%	**58%**	41%	41%	66%	50%	**42%**	**42%**
Autumn NBdefault	54%	45%	42%	38%	**78%**	49%	**44%**	44%	67%	40%	43%	42%	64%	**55%**	**44%**	**44%**

Note: A: accuracy; P: precision; R: recall; F1: F1-score. In binary classification problems, the F1-score is always between precision and recall. However, in multi-class classification, the final F1-score may be lower than both precision and recall due to Simpson's Paradox [95], which is caused by an imbalanced representation of subgroups when attempting to interpret the overall performance of subgroups. The values presented in the shadowed background correspond to model performance trained and predicted using the current season's data, while those in the non-shadowed background correspond to model performance trained and predicted using data from other seasons.

Fig. 11 depicts the Pearson correlation between four $NB_{default}$ model inputs. Winter data has the highest sum of dependence degree (1.57 in Fig. 11a), which could threaten the NB algorithm's fundamental assumption that each input is independent of each other, resulting in poor model generalization of winter $NB_{default}$ in Table 6. During the winter, the air conditioner was turned on 96% of the time to provide heating (Fig. 9a). It will raise the indoor air temperature while significantly lowering the RH level and cause a strong negative correlation (-0.66) between T_a and RH in Fig. 11a. The high negative correlation values between T_a and Clo in spring (-0.58 in Fig. 9b) and autumn (-0.37 in Fig. 9d) also indicate people's active adaptive behaviors of adding and reducing the number of clothes during cold and hot conditions.

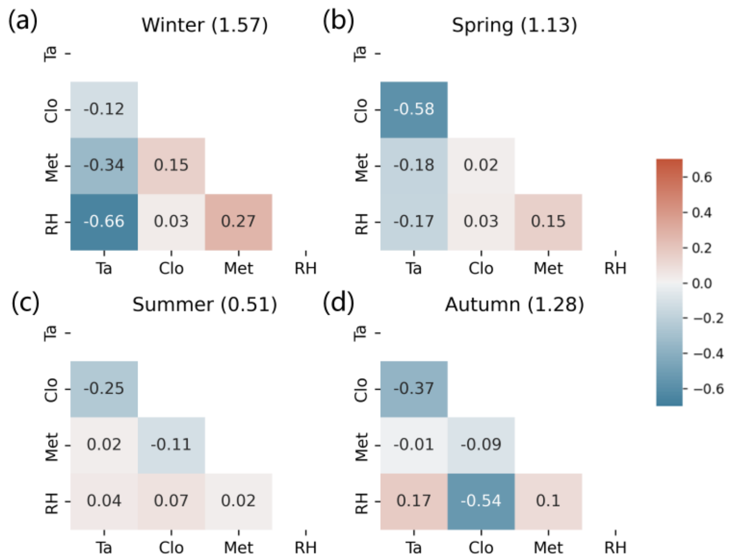

Fig. 11 The Pearson correlation coefficient between four $NB_{default}$ inputs (absolute sums shown after season's name)

The changes in predictive power after adding one feature at a time to the original four inputs of the $NB_{default}$ model are depicted in Fig. 12, with red and green colors representing a decrease or increase in performance and circle size indicating the change rate. During the spring season, several features failed to influence the performance of the NB_{update} model (no circles plotted), and two features "*max Tout*" and "*min Tout*" contributed positively to all evaluation metrics, namely improving accuracy, precision, recall, and F1-score simultaneously. Therefore, these two features were chosen for the training of final spring NB_{select} model. Similarly, the added features for winter, summer, and autumn are **1)** "*drink cold water*" (sum: 1); **2)**"*max Tout*", "*average Tout*", "*window open*", "*weather*", "*go out of office*", "*outdoor RH*", "*drink hot water*", "*stay time in office*", "*min Tout*", "*go upstairs to enter office*" (sum: 10); and **3)** "*weather*", "*go out of office*", "*drink hot water*", "*have dinner*" (sum: 4).

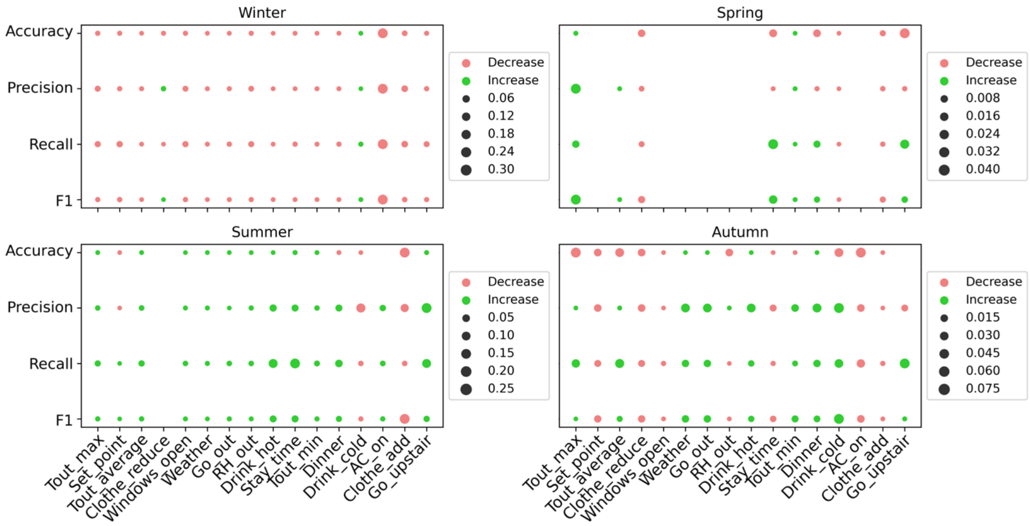

Fig. 12 Prediction power changes of the NB_{update} model compared with the $NB_{default}$ model

Table 7 shows the final predictive performance of the NB_{select} models after all the selected features have been added to the NB models. The accuracy of the spring NB model, which added two new features "*max* T_{out}" and "*min* T_{out}", remains at 81%, but precision, recall, and F1-score increased by 9%, 3%, and 6%, respectively. These metrics increases for winter and autumn remain in the 1% to 2% range. However, for the summer data, 10 new features were added to the training process, and accuracy, precision, recall, and F1-score increased by 3%, 12%, 10%, and 14%, respectively.

Table 7. Prediction power of NB_{select} model for each season

	Winter test data				Spring test data				Summer test data				Autumn test data			
	A	P	R	F1	A	P	R	F1	A	P	R	F1	A	P	R	F1
Winter NB_{select}	**63%**	**43%**	**42%**	**41%**	12%	33%	35%	9%	11%	19%	30%	7%	33%	31%	33%	22%
Spring NB_{select}	57%	46%	41%	41%	**81%**	44%	**43%**	43%	70%	46%	40%	39%	63%	**54%**	45%	**46%**
Summer NB_{select}	58%	58%	34%	27%	**75%**	41%	43%	41%	73%	**70%**	**51%**	**55%**	67%	51%	40%	37%
Autumn NB_{select}	57%	48%	44%	41%	**78%**	50%	**45%**	45%	65%	39%	43%	41%	64%	**54%**	44%	**45%**

3.2.2 Random Forest

Table 8 and Fig. 13 show the best pairs of hyperparameters of the RF_{all} models and RF_{select} models discovered by grid search, as well as their predictive performance. Table 8 shows that criteria entropy is preferred during AC-conditioned seasons (winter and summer), whereas criteria Gini is preferred during non-AC-conditioned seasons (spring and autumn), with the exception of season autumn from the RF_{all} model. This preference for criteria entropy during AC-conditioned seasons can be attributed to the increase in adaptive AC-related behaviors, such as frequent adjustments in temperature setpoints and usage patterns, particularly prevalent during winter and summer. Criteria entropy, being more sensitive to changes in class probabilities, effectively captures the variability introduced by these adaptive behaviors. Conversely, criteria Gini, which focuses on the majority class (or "*in favor of variables with high category frequencies*"[96]), are more suitable for non-AC-conditioned seasons where adaptive behaviors are less pronounced. The observed discrepancies between AC and non-AC conditions underscore the importance of considering contextual factors and adaptive behaviors in model development to improve model efficiency and predictive accuracy.

The complexity of the RF_{select} models is generally lower than the complexity of the RF_{all} models because fewer input features were utilized and the grid search results of the number of estimators, max depth, and min sample leaf were all reduced at significant levels in Table 8. It is noteworthy that the observed reduction in complexity of the RFselect models compared to the RF_{all} models is reflected not only in the utilization of fewer input features but also in the significant adjustments identified through grid search. For instance, the grid search results in Table 8 demonstrate notable reductions in the number of estimators and max depth for RF_{select} models across various seasons, indicating a more streamlined and efficient model architecture. Specifically, the reduction in the number of estimators from 100 to 25 in summer and the decrease in max depth from 25 to 5 in summer highlight the simplification of the RFselect models, aligning with the principle of Occam's razor in favoring simpler models when achieving comparable predictive performance.

Fig. 13 depicts the feature importance rankings from trained RF models. It is clear that some classic PMV inputs (T_a, RH, Clo) contribute the most to model establishment and remain at the top, followed by outdoor conditions (T_{out}, Weather), and adaptive behaviors (window opening, AC adjustments, water drinking, etc.). One interesting outcome is that the adaptative behaviors of adding and removing clothes have no contribution to the RF models, because clothing levels during the summer are already kept at a relatively low level and have very limited potential for further adjustments, so the RF training process in summer ignores these two features.

Table 8. Hyperparameter optimization of the RF models using all features and selected features during two rounds

Parameter	Search space	RF_{all}				RF_{select}			
		Winter	Spring	Summer	Autumn	Winter	Spring	Summer	Autumn
Number of estimators	[10, 25, 50, 100, 200]	200	25	100	100	10	25	25	100
Criterion	["Gini", "entropy"]	Entropy	Gini	Entropy	Entropy	Entropy	Gini	Entropy	Gini
Max depth	[3, 5, 10, 15, 20, 25]	25	20	25	25	5	25	20	20
Min sample leaf	[1, 2, 5, 10]	1	2	2	1	1	2	1	1

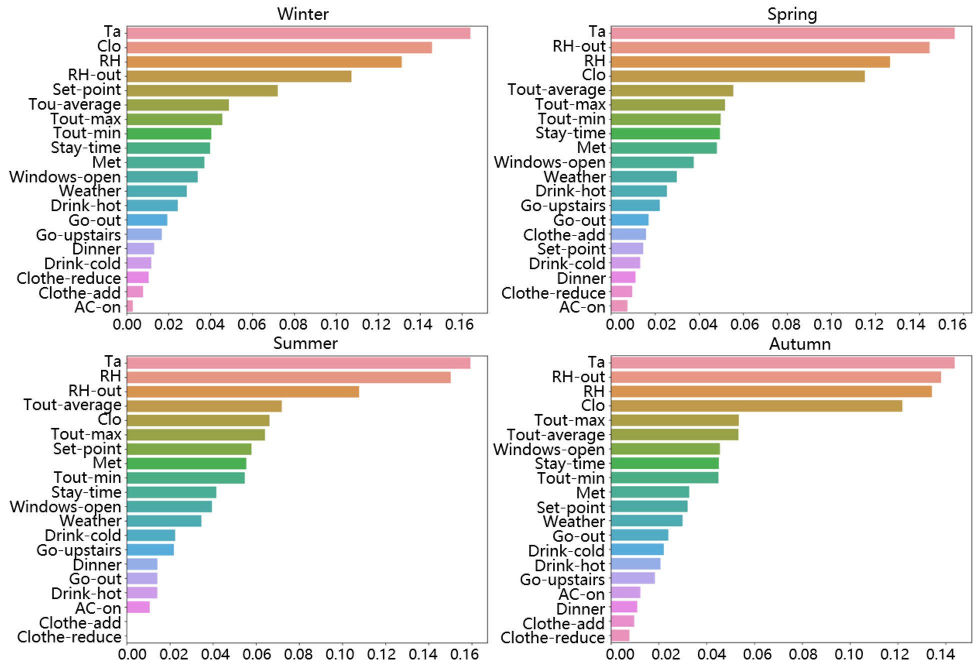

Fig. 13 Feature importance of RF models in four seasons

3.2.3 Performance comparison between machine learning models and PMV index

Fig. 14 summarizes the predictive results of the RF models combined with NB models in section 3.2.1 and PMV predictions. In general, the RF models produce the best results (green colors). Within RF models, the RF_{all} and RF_{select} models perform similarly in summer and autumn, but in winter, the RF_{all} models (light green) slightly outperform RF_{select} models (dark green), and this trend reverses in spring. The NB models with all features (light blue) have very poor predictive accuracy in all four seasons, as irrelevant features can significantly affect their performance. This is consistent with Luo et al's [49] finding that the NB models have the lowest accuracy among all the machine learning algorithms when evaluating 10,618 samples drawn from twelve features of the ASHRAE global database [97]. However, after only using selected features from the framework in this paper, the NB models exhibit competitive performance (dark blue) similar to the RF models. The RF models, on the other hand, are more robust to the interference from unfavorable features because each tree in the forest can choose to give low weights to the less contributed features during training.

Classic PMV models in Fig. 14 have the lowest predictive accuracy in these 3-class classification problems from winter to autumn (pink, 46% to 62%) when compared to

NB_{select} and RF models (62% to 88%). This poor performance in practice has also been frequently criticized in thermal comfort studies as being sometimes equivalent to random "guessing", such as around 50% accuracy in binary classification [98], 34% accuracy in seven-class classification [30], and 6% Cohen's kappa coefficient in 3-class classification [77]. The precision of the PMV models in this study is also very low in the spring, summer, and autumn, as the PMV index is population-based and this study only includes 12 surveyed subjects, resulting in bias due to individual sensitivity. However, the recall of PMV in winter, spring, and autumn shows superior results, which may lead to the final F1-score competitive (40% to 47%) to the NB_{select} and RF models (40% to 45%). This low precision and high recall prediction will suffer from incorrect label returns but will benefit from a high chance of detecting all actual positives, namely at the cost of introducing irrelevant results to avoid missing relevant results. Therefore, the PMV index could have many undiscovered potentials when *false negatives* are critical and *false positives* are less important. For example, in infant care [99] or hospital settings [100], where missing positive detections of thermal comfort can have serious consequences, the PMV index could be used as a supplement to ensure a higher level of recall rate for final decision making.

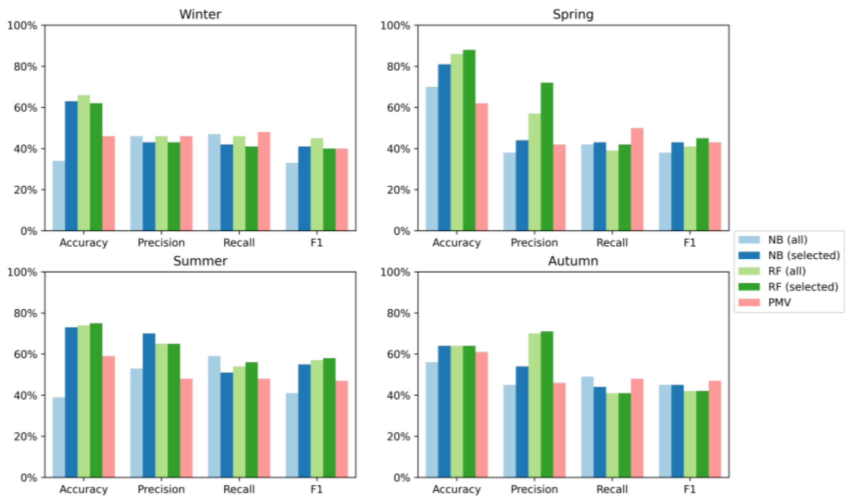

Fig. 14 Evaluation metrics of the NB, RF, and PMV models using all features and selected features

Fig. 15 depicts the training speed of the NB, RF, and PMV models in seconds on a log scale. The RF models require the longest time to train, ranging from 3.4 to 5.7 minutes. This may be due to the search grid method used to find the best hyperparameters. On the contrary, the NB models only take 0.04 to 0.06 seconds to train, resulting in a nearly real-time response. This is because NB models are straightforward and based on the mathematically well-grounded Bayes theorem with few hyperparameters to tune. The classic PMV index, which embeds several iterations and heat balance equations to calculate intermediate variables (such as determining clothing surface temperature), presents calculation times ranging from 0.6 to 1.2 seconds. After reducing feature sums from the original NB and RF training processes, the speeds of training NB_{select} and RF_{select} models (columns 4 and 2 in Fig. 15) are slightly slower than training NB_{all} and RF_{all} models (columns 3 and 1 in Fig. 15).

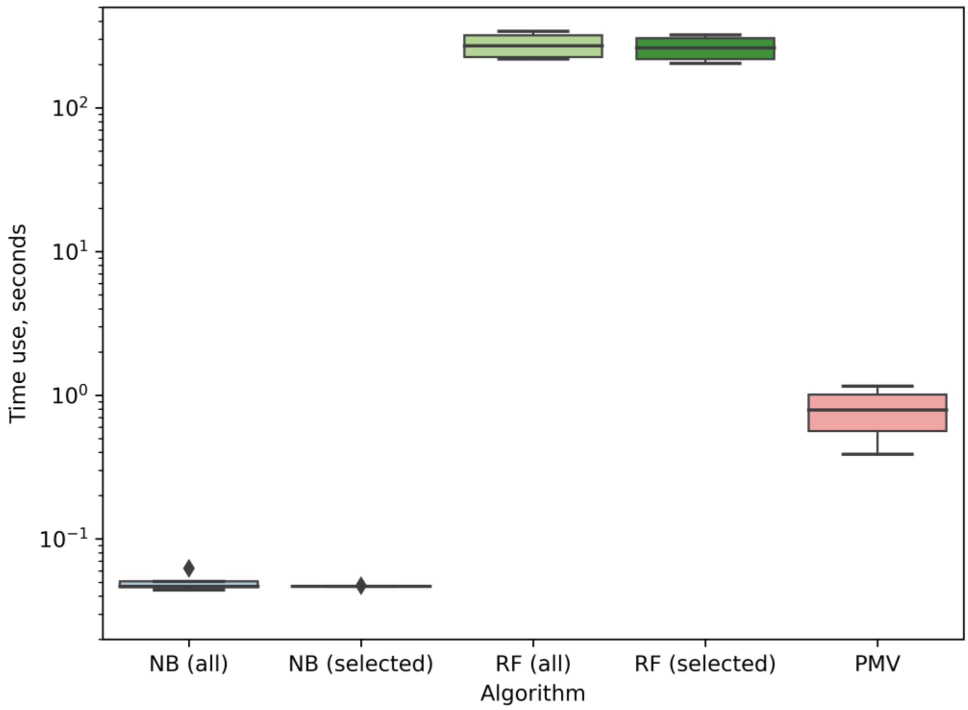

Fig. 15 Model training speed of NB, RF, and PMV models with log scale in y-axis (Results obtained using a laptop equipped with an Intel i7-8750H CPU, 2.20GHz)

3.3 Performance of the classic adaptive model in spring and autumn

The ASHRAE 55-2020 [8] standard suggests using a graphic method (adaptive model) to evaluate thermal comfort in occupant-controlled naturally conditioned spaces with no mechanical cooling or heating system operations with the valid prevailing mean outdoor temperature ranging from 10°C to 33.5°C. This graphic-based adaptive model's x-axis was changed from monthly mean outdoor air temperature to prevailing mean outdoor air temperature in the 2013 version of ASHRAE 55, while monthly mean outdoor air temperature is still allowed when prevailing mean outdoor air temperature is unavailable. The study in this paper was carried out in the hot summer and cold winter (HSCW) zone in Chongqing, China, where the climate is moderate in spring and autumn under high NV potentials. Therefore, Fig. 16 depicts the relationship between the indoor operate temperature and the monthly mean outdoor air temperature during the transition seasons of spring and autumn. Due to the high cost of collecting T_r and the lack of obvious radiance sources discovered during the investigation, T_r was assumed to be equal to Ta, resulting in the operative temperature of the adaptive model being the same as T_a ($T_o = (T_a + T_r)/2$ in ASHRAE 55 [8]).

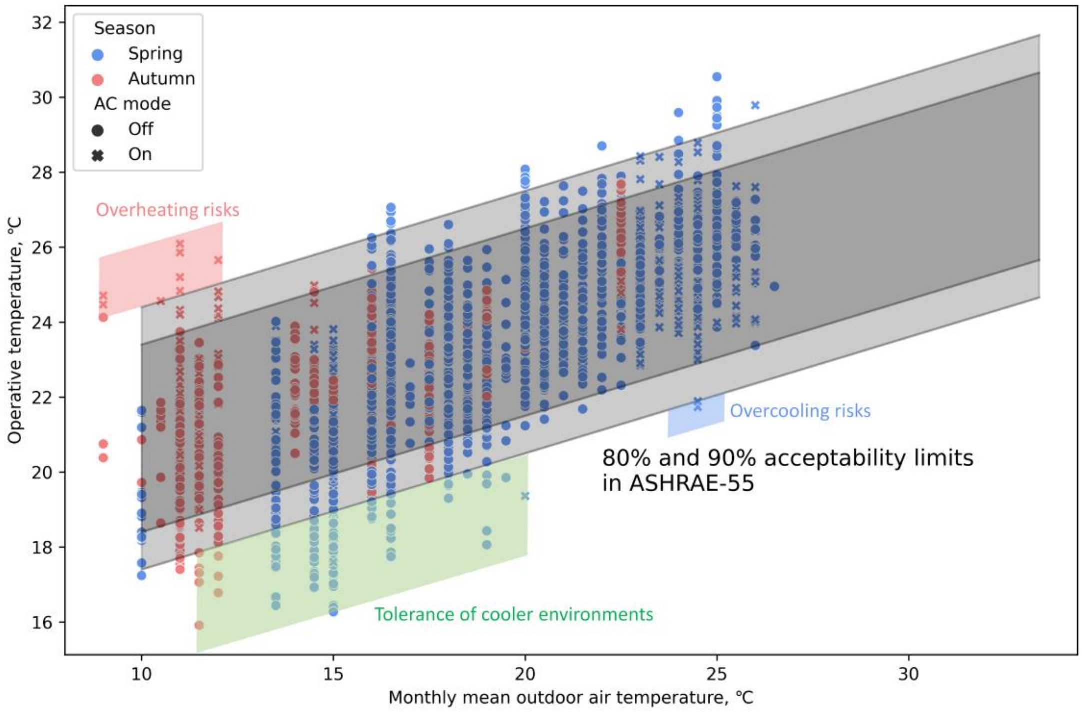

Fig. 16 Indoor temperature plotted against the 80% and 90% acceptability limits in ASHRAE55 the monthly means were used to interpolate the relationships in the adaptive model, as suggested by ASHRAE 55-2020 section 5.4.2.1.3 [8].)

In Fig. 16, spring and autumn seasons clearly show some cooling and heating demands

(× mark), as the values of outdoor air temperature in spring (blue background) are generally higher than the values in autumn (red background). However, a few overheating and overcooling risks emerge at the top left and bottom right when the AC was turned on (× mark) and the indoor temperature has been raised or decreased outside of the comfort zones. Some votes in the bottom left area indicate that people are more tolerant of cold thermal environments in offices (○ mark in the green background) and chose not to turn on the AC even when the conditions exceeded the ASHRAE 55 limits.

To better quantify the predictive performance of the adaptive model in transition seasons, Fig. 17 displays the evaluation metrics of the adaptive models, PMV, and the machine learning algorithms under selected features when the AC was turned off, in accordance with the ASHRAE 55 stipulation that no cooling or heating systems should be in operation when referring to the adaptive model. The NB and RF models (green colors) have the best overall performance, while the PMV (pink) still presents low accuracy/precision but a high recall rate. The adaptive models (blue colors) show relatively high accuracy/precision but low recall, which is the opposite of the PMV trend. The balanced metrics F1-score of the adaptive model is also lower (32% to 37%) than other models (41% to 46%).

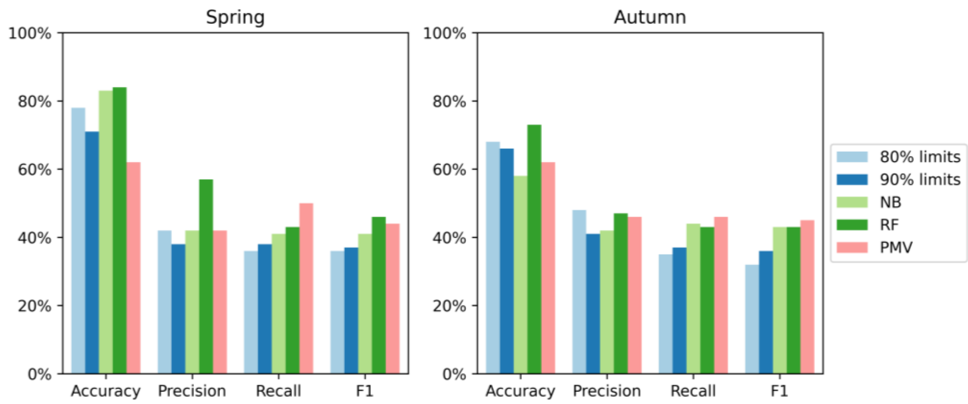

Fig. 17 Evaluation metrics of NB, RF, PMV, and adaptive models in spring and autumn when the AC was turned off

3.4 Adaptive behaviors in winter and summer

As discussed in the introduction, while machine learning models have proven effective in predicting thermal comfort in MM buildings, their inherent lack of interpretability raises concerns regarding the understanding of the underlying relationships between input variables and predictions. In order to address this interpretability aspect, the relationships between various input variables through statistical analysis are explored in this section. The training process of the NB model demonstrates that outdoor conditions and several personal behavioral variables during the winter and summer improve the overall predictive performance of machine learning algorithms. To better illustrate their specific impacts, the features "*drink water*" in winter and "*window status*" in summer were chosen for statistical analysis using the Mann-Whitney test when the AC was turned on, as both features contributed positively to the model training process.

3.4.1 Drinking cold and hot water in winter

Fig. 18 displays the results of the Mann-Whitney test on thermal preference in four seasons. Within one season, the significance levels of preference are all significant (from $p<=0.01$ to $p<=0.0001$), with one exception of warmer and cooler preferences in winter (ns), namely indoor environments of warmer preference and cooler preference in winter are quite similar but people present two opposite thermal states. In Fig. 19, the Mann-Whitney test on adaptive behaviors from winter data reveals that occupants' comfort temperature (green color, no change) when drinking cold water (room temperature water) is significantly higher ($p<=0.0001$) than when not drinking cold water. However, when drinking hot water, the comfort temperature is significantly lower $p<=0.001$ than when not drinking hot water. Therefore, the habit or adaptive behaviors of drinking hot water may contribute to lower comfort temperatures when compared to drinking cold water, which is beneficial for saving HVAC energy during the winter.

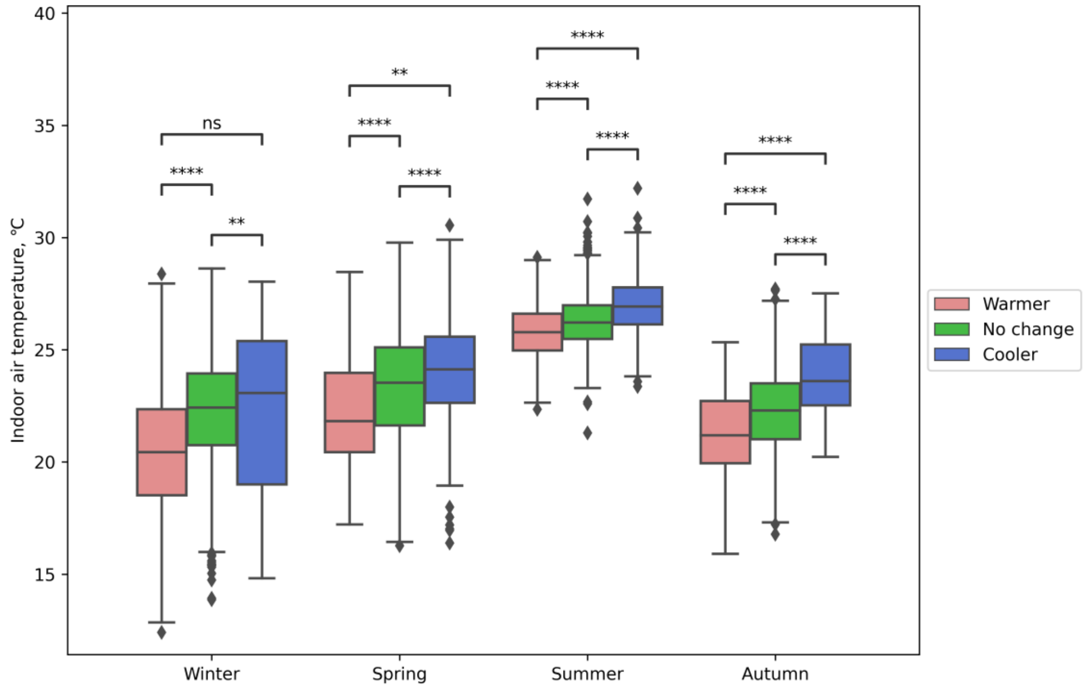

Fig. 18 Mann-Whitney test of thermal preference in four seasons (****: p <= 0.0001; ***: 0.0001 < p <= 0.001; **: 0.001 < p <= 0.01; *: 0.01 < p <= 0.05; ns: 0.05 < p <= 1)

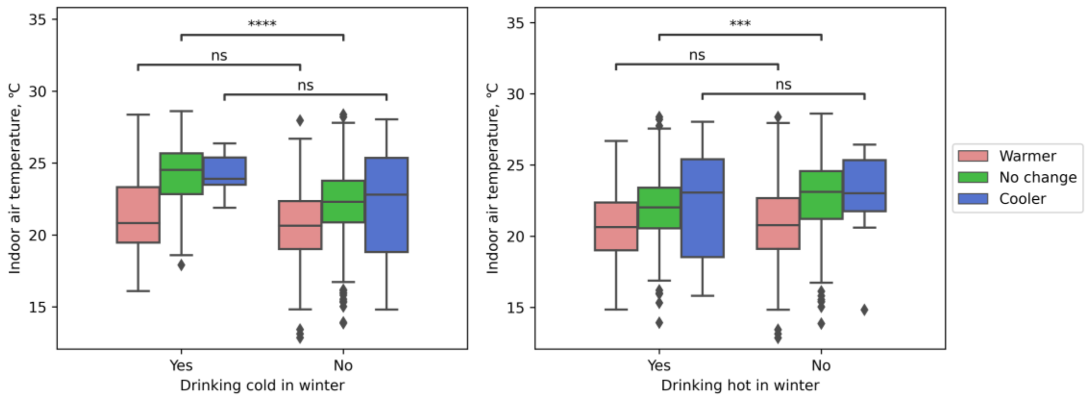

Fig. 19 Mann-Whitney test of drinking cold and hot water in winter

3.4.2 Windows status in summer

The Mann-Whitney test results on window status in Fig. 20 show that the differences in indoor air temperatures between prefer warmer (red color) and prefer cooler (blue color) during the windows "*totally open*" period are not significant (ns), whereas the results are all significant (p<=0.0001) for three non "*totally open*" periods. The differences in comfort temperatures between period “totally open” and the other three periods are increasing: from not significant (ns) on period “half open”, to 0.01 significant level (p<=0.01) on period “slightly open”, and finally to 0.0001 significant level (p<=0.0001) on period “closed”. Meanwhile, the general values of comfort temperatures (green colors) are reduced while the windows are closed more thoroughly.

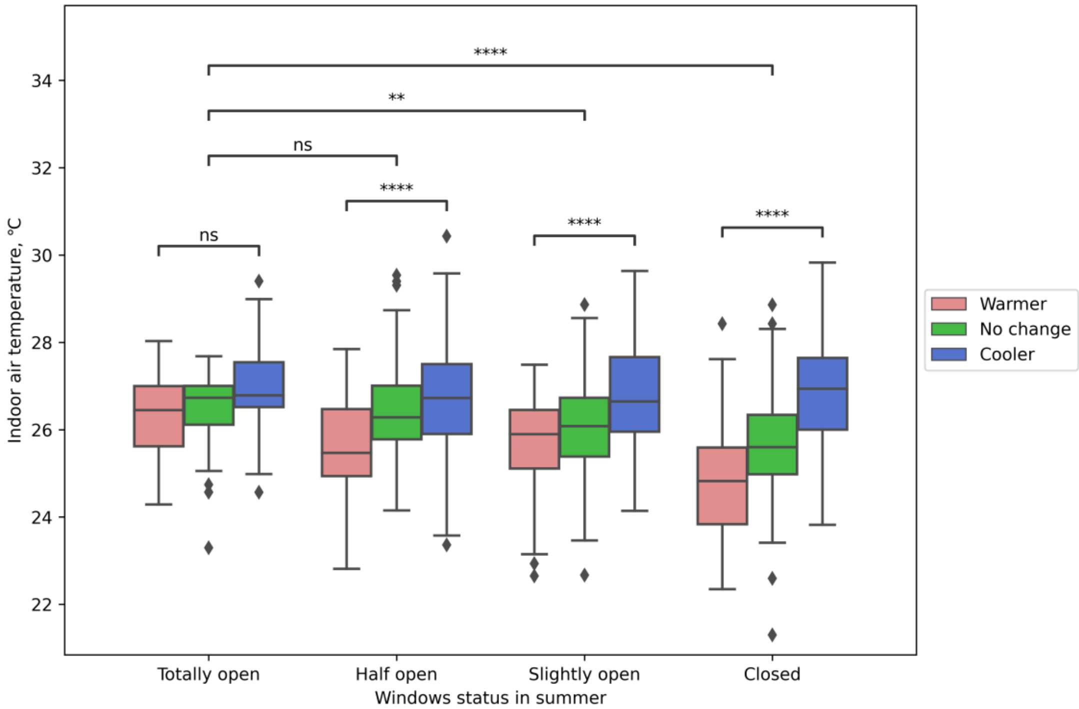

Fig. 20 Mann-Whitney test of windows status in summer

To better illustrate how windows behaviors influence indoor conditions, Fig. 21 plots the linear regressions of outdoor and indoor air temperatures binned at one-degree intervals at four window states. As the windows are closed more tightly, the overall indoor temperature drops. Furthermore, as long as the windows are not completely closed (periods “slightly open”, “half open”, and “totally open”), the indoor thermal environments change more obviously with outdoor temperatures, with linear gradients ranging from 0.08 to 0.12 (blue, orange, and green colors). When the windows are closed, the indoor conditions remain much more stable, with a linear gradient of 0.02 (red color).

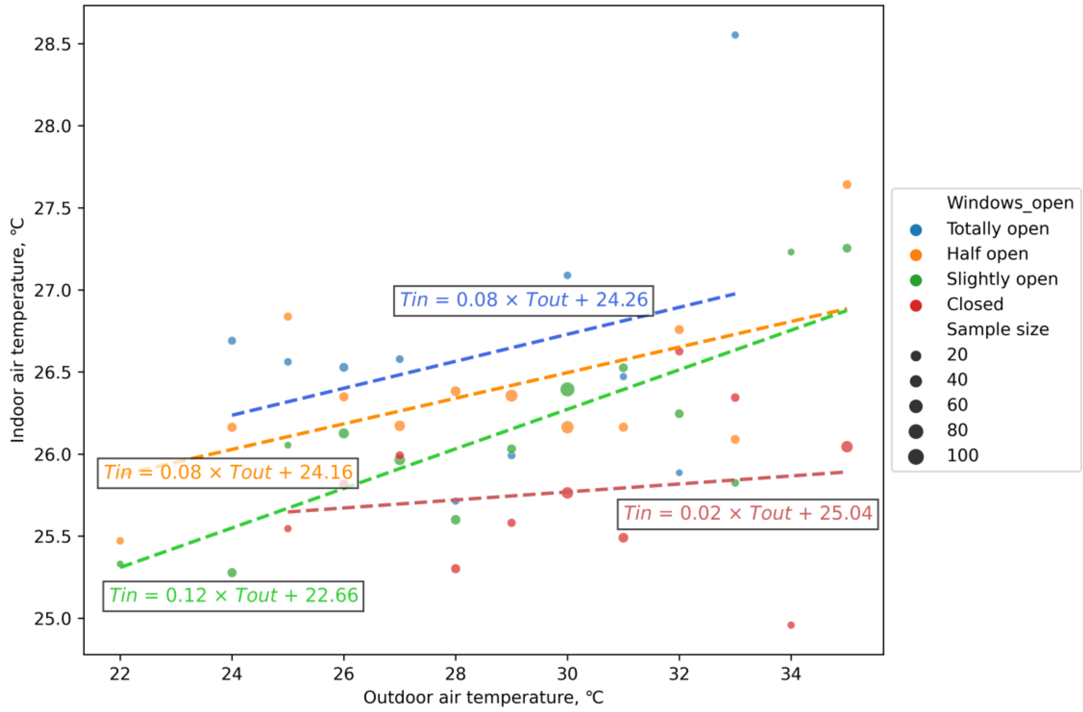

Fig. 21 Linear regressions of windows status in summer

3.5 Impact of outdoor air velocity on adaptive behaviors and thermal comfort

We also extracted outdoor wind speed data from the meteorological station in Chongqing and analyzed its influence on adaptive behavior and thermal comfort in mixed-mode buildings. Fig. 22 illustrates the distribution of wind force levels over 365 days, with force levels 1 and 2 comprising the majority, force level 3 being less common, and force levels 0 and 4 being even rarer. Fig. 23 displays the proportion of different degrees of windows opening under various wind force levels. As wind force levels increase, the proportion of “Open” and “Half open” significantly rises, while the proportion of “Slightly open” and “Closed” decreases. This suggests that occupants in mixed-mode buildings are more inclined to ventilate indoor spaces as outdoor wind speed increases, enhancing their own comfort.

Fig. 24 illustrates the distribution of indoor comfort air temperatures when subjects voted for "No change" under different wind force levels. For wind force levels 0-2, there is minimal difference in indoor comfort temperatures across various windows opening degrees. However, at wind force level 3, the indoor temperature when windows are closed is notably higher than when they are open, indicating that in such conditions, opening windows significantly lowers indoor comfort temperatures. Considering the current lack of specific and rigid recommendations for outdoor wind speed in thermal comfort standards, we suggest that future research could dive deeper into this aspect to supplement and enrich the comprehensiveness of existing thermal comfort knowledge.

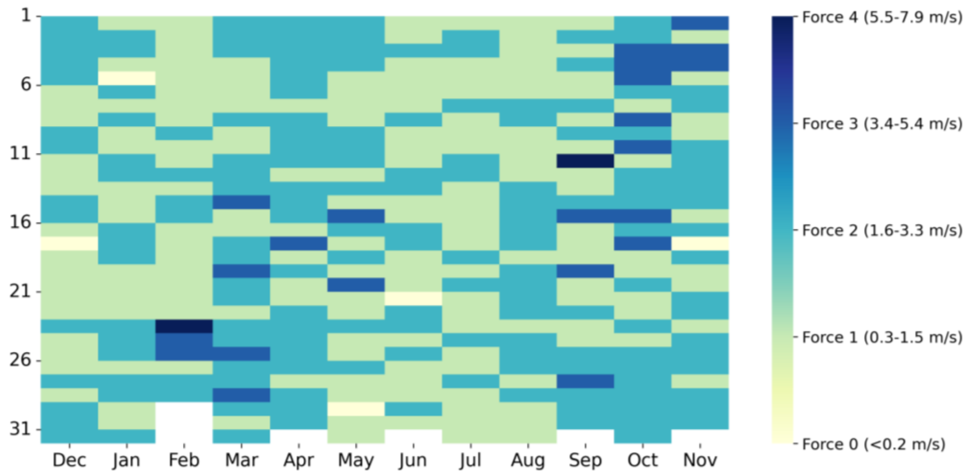

Fig. 22 Distribution of outdoor wind force levels during the one-year experiment

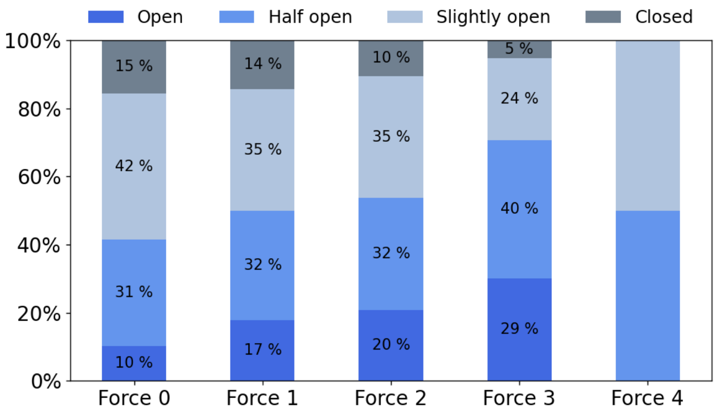

Fig. 23 Proportion of windows opening degrees under different wind force levels

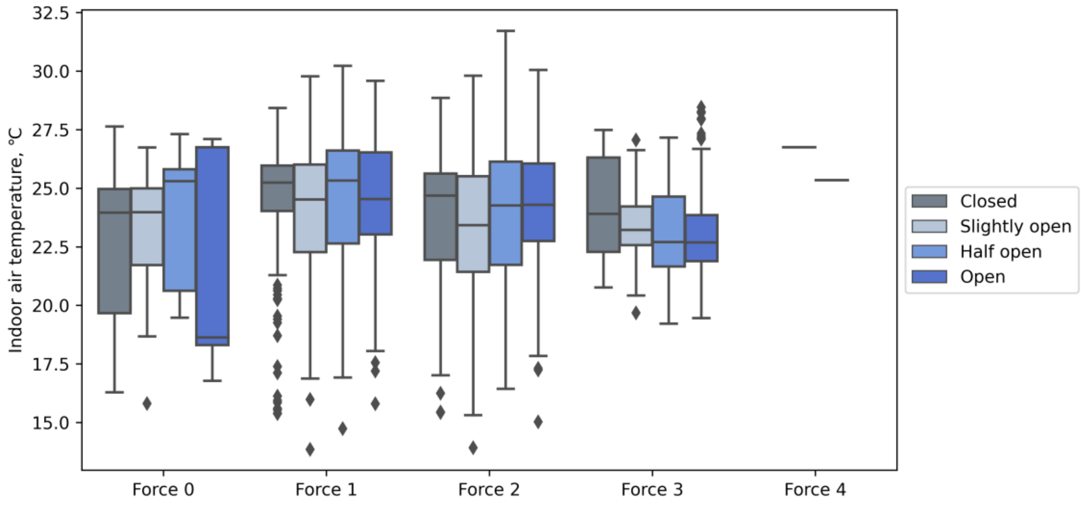

Fig. 24 Distribution of indoor comfort air temperatures under preference vote "No change"

4. Discussion

4.1 Predictive performance of machine learning algorithms and classic thermal comfort models

As many thermal comfort studies have already shown [6][49][69], machine learning models (NB and RF) generally outperform traditional thermal comfort models (PMV and adaptive) in terms of accuracy (Figs. 14 and 17). However, this paper compared three other evaluation metrics: precision, recall, and their combined effects F1-score. It was found that the PMV index shows highest predictive recalls in winter, spring, and autumn even beyond well-tuned RF models that further makes its F1-score comparative to machine learning algorithms, with only one exception from summer case in Fig. 14. The unexpectedly high recalls observed in the PMV index can be attributed to several factors: **1) Conservative predictions:** the PMV index tends to make conservative predictions by prioritizing the detection of actual true labels (preferences) even at the expense of potentially predicting some wrong labels (low precisions), which may lead to higher recall rates as the model turns to minimize false negatives (missed positive labels); **2) Sensitive to actual votes:** the PMV index may be more sensitive to variations in actual thermal states experienced by subjects, capturing a wider range of thermal preferences or discomfort levels across different seasons, which could contribute to its ability to detect a larger proportion of actual true labels, resulting in higher recall rates; **3) Inherent model characteristics:** the PMV index is well-known for its solid theoretical deduction process, which may inherently possess certain characteristics that facilitate higher recall rates, aiming to capture a broad spectrum of thermal states experienced by subjects; **4) Contextual factors:** the specific environmental conditions, occupants' adaptive behaviors, and building characteristics together may together influence the performance of the PMV index, which means that other study may or may not find PMV with high recall.

Overall, the unexpected high recalls observed in the PMV index across different seasons highlight its ability to effectively capture a substantial portion of actual thermal preferences or discomfort levels among subjects. However, it's essential to acknowledge the potential limitations of the PMV index, such as its lower precision and the risk of predicting wrong labels, especially in situations where false positives are costly. Consequently, combining the PMV index with customized machine learning approaches can offer a more comprehensive evaluation of occupants' real thermal states, leveraging the strengths of both traditional thermal comfort models and advanced machine learning techniques.

Although random forest models generally achieve optimal performance in most scenarios, it is important to note that they also have inherent limitations that should be considered in practical applications [58]: **1) Interpretability:** its structure is typically complex, consisting of numerous sub-decision trees, which can be challenging to interpret directly compared to a single decision tree or linear regression; **2) Potential overfitting:** when it becomes overly complex or when insufficient parameter tuning techniques are employed during model training, the model turns to capture noise or random fluctuations in the training data rather than the underlying patterns or relationships; **3) High computational resources:** its computational demands and memory requirements can become challenging as data volume or dimension increases, especially when resources are limited.

Adaptive models in ASHRAE-55 provide acceptability limits of 80% and 90% comfort zones indicated by graphic methods. The 80% limits have a wider range, which means that as the indoor temperature range extends, fewer people are satisfied. When predicting thermal preferences with 80% and 90% limits, as shown in Fig. 17, the 80% limits have higher accuracy and precision but lower recall and F1-score. From an accuracy perspective, the boundaries of 80% limits are more reasonable, but for picking more actual positives, 90% limits could have better performance.

4.2 Generalization of machine learning algorithms

The ability to generalize results from training samples to unseen data is widely recognized as an important capability of any model. However, the findings of a meta-analysis [101] that gathered models from several scientific literatures and cross-validated their performance on various public datasets show that one of the most modest models performs the best on all other datasets, while one of the most robust models performs nearly the worst. For the machine learning algorithm, even though it has many techniques to avoid overfitting [102], its data-driven nature faces the lack of systematic modeling procedures, physical fundamentals, and interactions with real-world scenarios [6], thus creating barriers to generalization.

In this study, the NB models were trained individually using data specific to each season. However, to testify their generalization during the prediction phase, the discussion encompasses the scenario where each model is utilized to predict data for all four seasons (spring, summer, autumn, and winter). Table 6 shows that the winter $NB_{default}$ model has 62% accuracy and 40% F1-score on winter data but only 12% accuracy and 9% F1-score on spring data. What's more, the models trained from spring, summer, and autumn data all achieve highest accuracy on spring data (81%, 75%, and 78%). This could be caused by the imbalanced data distribution in spring data that over 82% vote are "*no change*".

Imbalances in the dataset can lead to biased predictions, particularly affecting minority classes and overall model performance metrics. As discussed before, the trained Naive

Bayes models exhibited varying degrees of performance across seasons, with notable disparities observed in predictive accuracy and F1-scores when applied to different seasons. For instance, while the winter $NB_{default}$ model demonstrated relatively high accuracy and F1-score on winter data, its performance significantly decreased when predicting spring data. This disparity in performance across seasons highlights the impact of imbalanced data distribution, particularly evident in spring where over 82% of votes indicated "no change." Consequently, solely relying on accuracy as an evaluation metric may incentivize machine learning algorithms to prioritize the majority class, potentially compromising the model's ability to generalize. By incorporating additional metrics such as precision, recall, and F1-score, we provide a more comprehensive assessment of model performance, mitigating the influence of data imbalance and ensuring a more robust evaluation framework.

We have chosen to disclose this issue and trained another kind of advanced machine learning algorithm for performance compensation, "*Random Forest"*, which is known to be more robust to imbalanced data. In addition to model training process, resampling techniques, such as oversampling or undersampling, can partially address the issue of data imbalance. However, it's crucial to recognize that these techniques may introduce the problem of overfitting or underfitting [103], exacerbating the imbalance or reducing the representativeness of the data, respectively. Therefore, careful consideration of the trade-offs associated with resampling techniques is necessary to effectively manage the challenges posed by imbalanced data.

One of the central problems in machine learning is to identify relevant information subset or *feature-selection* for making accurate predictions. Fig. 14 depicts that when all features are used for NB training, the predictive accuracies in winter and summer (34% and 39%) are even lower than PMV accuracies (46% and 62%), with only half the accuracies of using selected features for NB training (63% and 73%). The precision of the RF model in spring drops from 72% to 57% when all features are used without *feature-selection*. These findings emphasize the importance of including the appropriate variables in the training process of machine learning algorithms.

Another important consideration during the generalization is the response time of model prediction. Fig. 15 shows that the RF models will take more than three minutes to train and establish, whereas the NB models and the PMV index only take around one second. These time differences are tolerable in situations where there will be no serious consequences, such as young adults working in offices who feel slightly cool or warm. However, some delays in predicting thermal state can result in life-threatening conditions, such as hypothermia and hyperthermia to the elderly who live alone at home [72]. Unfortunately, the classic PMV index or adaptive model are population-based that often perform less well on individuals [6]. The machine learning algorithms thus can provide new opportunities from adding specific features to the model establishment with improved predictive power and acceptable training speed.

4.3 Marked features in MM buildings

4.3.1 Classic thermal comfort models for MM buildings

Generally, the PMV index and graphic-based adaptive model are recommended for evaluating thermal environments in HVAC and NV buildings in current international and national thermal comfort standards [7] [8] [9] [10] [11]. This is because people in HVAC buildings are assumed to rarely adapt themselves and have few opportunities to control their own thermal environments, whereas people in NV buildings have no heating or cooling devices to operate. People in the study, however, actively adapted themselves during heating and cooling periods in winter and summer by drinking water and opening windows. This creates challenges when applying the PMV index directly to the HVAC mode of MM buildings.

On the other hand, people in this study during spring and autumn chose to turn on/off the AC whenever they felt thermally uncomfortable. Fig. 16 demonstrates that some data points within the ASHRAE 55 adaptive models' 80%/90% comfort zones were under AC operation, implying that the MM buildings may not provide acceptable thermal environments simply relying on the natural ventilation in Chongqing's spring and autumn. Furthermore, a significant proportion of comfort points in cooler environments were identified as uncomfortable (green background) by ASHRAE 55 adaptive models, and a few overheating (red background) and overcooling (blue background) risks emerged during the autumn and spring, respectively. The data in this paper proves that this adaptation appears in Chongqing and office buildings as well.

4.3.2 Water drinking

Several studies have already shown that people will actively consume hot/cold water to actively adapt to thermal environments [22][104]. This study further discovered that during the heating season, occupants' comfort temperatures were significantly higher if cold water was consumed ($p<=0.0001$), or lower if hot water was consumed ($p<=0.001$), compared to no drinking adaptations (Fig. 19). From the energy perspective, drinking cold water here will be unfavourable because higher comfort temperatures may result in higher heating demands from HVAC systems. Our investigated offices are all equipped with water dispensers that can conveniently provide cold or hot liquids, allowing occupants to choose according to their own preferences. Yu et al. [105] conducted field studies on residents' thermal comfort in Tibet, China, where temperatures and humidity are extremely low, with annual average air temperature ranging from 5.93℃ to 9℃ in four investigated cities. They discovered that people there have unique ways of protecting themselves from the cold, such as frequently using stoves in kitchens and drinking hot butter-sweet tea. These findings highlight the fact that the adaptive opportunities provided by buildings, as well as people's customs/habits, will result in varying HVAC energy consumption outcomes.

4.3.3 Windows opening

Previous publications already indicated a high frequency of open windows in non-heating/cooling periods [47]. This study further discovers that in MM buildings, the practice of windows openings occurred extensively (over 80% of the time) even during heating and cooling seasons. These behaviors could lead to increased air change rate and heat transfer through building facades, causing indoor environments to significantly fluctuate with outdoor conditions (Fig. 21), resulting in extra HVAC energy consumption. However, as the windows are opened wider, occupants' comfort temperatures will rise significantly (Fig. 20). On the other hand, opening windows can provide people with more enjoyable views and engagements with outdoor nature, as well as adequate fresh air and indoor air movements. Several ASHRAE-sponsored field studies [106] (6148 responses from 53 buildings) discovered that when the thermal sensation range was -0.7 to 1.5, larger percentages of people (47% to 84%) preferred more air movements, while smaller percentages (3% to 7%) preferred less. Therefore, cooling or heating thermal environments with similar characteristics from NV can not only improve occupants' subjective satisfaction but also save the required energy of HVAC systems by raising HVAC setpoints higher during cooling seasons. To validate this, Chen et al. [107] proposed a CFIAC (Ceiling-fan-integrated air-conditioning) framework offering non-uniform distributions of indoor air-speed and temperature capable of compensating for 1.2 to 1.5 PMV scale units in the 26℃ to 28℃ temperature range.

To address the challenges of optimizing energy efficiency during summer months when occupants tend to open windows with the air conditioner on, the following potential solutions and recommendations are proposed:

- **Implement smart windows technology:** install smart windows equipped with sensors and actuators that automatically adjust opacity or ventilation based on outdoor conditions and occupant preferences.
- **Optimize the ventilation system:** one reason for users opening windows is to improve indoor air quality. Therefore, with the installation or update of mechanical systems that provide adequate fresh air, the need for window opening behavior may be significantly reduced.
- **Enhance occupants' awareness of energy conservation:** provide occupants with information and guidelines on energy-efficient behaviors, including the appropriate use of windows, thermostats, and HVAC systems; encourage occupants to utilize natural ventilation during cooler times of the day and minimize reliance on air

conditioning.

4.4 Limitation and further work

Although we tried to comprehensively evaluate both classic thermal comfort models and machine learning algorithms for evaluating adaptive behaviors and predicting thermal comfort in MM buildings, some limitations and future research recommendations still remain:

1. It’s important to recognize that our study sample primarily consisted of young and middle-aged adults working in office settings. Therefore, the generalizability of our findings to other demographic groups, such as elderly individuals or individuals in non-office environments, may be limited. Future research should aim to include a more diverse range of participants to better understand how different demographic groups adapt to microclimate modifications in various settings.

2. This paper focused specifically on the climate conditions and building practices prevalent in Chongqing, which exhibit distinct seasonal demands for cooling in summer and heating in winter. While the insights gained from our research are valuable for this specific context, caution should be applied when extrapolating these findings to regions with different climate patterns and building practices. It is recommended that future studies consider conducting similar investigations in different geographical locations with varying climate conditions to assess the transferability of our models and observations. It is important to note that regions with diverse climate patterns may have unique thermal comfort requirements, and thus the applicability of our conclusions may vary.

3. The grid search method employed in hyperparameter tuning process of machine learning models can be computationally intensive and may not guarantee finding the absolute best hyperparameter combinations. Due to the vast search space and time constraints, it is possible that alternative hyperparameter configurations with potentially superior performance were not explored.

4. This study highlights the significance of both the number of input features and the

identification of dominant factors in achieving accurate models for understanding adaptive thermal comfort phenomena in MM buildings. However, the widespread use of smart devices in the future may enable data collection at any time rather than just the moment of a questionnaire, posing challenges in intuitively understanding the data and developing robust models for practice.

5. Future studies could address the following research questions to advance our understanding of thermal comfort in mixed-mode buildings:

- How do specific adaptive behaviors, such as adjusting air conditioning settings or opening windows, interact with each other in response to varying climatic conditions?
- What are the energy implications of these adaptive behaviors in terms of heating and cooling demand, and how do they contribute to overall building energy consumption?
- Are there differences in the effectiveness of adaptive strategies between different geographical regions with distinct climate patterns?
- How do occupant preferences and habits influence their adaptive behaviors, and how can building design and operation be optimized to align with these preferences while minimizing energy consumption?
- What are the trade-offs between occupant comfort and energy efficiency in mix-mode buildings, and how can these be balanced through design interventions or operational strategies?

5. Conclusions

The mixed mode of building is thought to be capable of positively extending occupants' comfort temperature to a wider range compared to fully air-conditioned buildings. Our study employs machine learning algorithms and classic thermal comfort models to investigate thermal preference during a one-year field study in Chongqing, China. The novelty of this research lies in integrating adaptive behaviors, which are difficult to include and quantify in conventional thermal comfort models, into machine learning algorithms to achieve a more comprehensive thermal comfort assessment and analysis of energy-related adaptive behaviors. This contributes to providing potential criteria and recommendations for identifying energy-inefficient behaviors in mixed-mode buildings, thereby enhancing building energy efficiency related to occupants' adaptations, and highlighting the potential trade-offs associated with building operation modes. The following findings are noteworthy:

(1) The spring, summer, and autumn naive bayes models all achieve best accuracy on spring data, with 82% voting "*no change*". However, relying solely on accuracy as the evaluation method can result in misleading results because it can be heavily influenced by the data distribution pattern. The high accuracy of machine learning models could be attributed to "*cheating*" on imbalanced data sources by giving more weights to majority votes instead of creating useful knowledge. To avoid falling into the trap of "*accuracy cheating*" during model training and evaluations, it is essential to include additional evaluation indices such as precision, recall, and F1-score, which will provide a more comprehensive understanding of model performance.

(2) In general, random forest models outperform naive bayes and classic thermal comfort models in terms of accuracy, precision, and F1-score. They are also resistant to irrelevant feature disruption, but at the cost of longer training times. However, the naive bayes models can provide training speed in real-time and are better suited for scenarios with time constraints. The classic PMV index with a transparent explanation between the human body and the surrounding physical environment has limited accuracies in most cases but unexpectedly high recalls, indicating that the PMV index has the potential to be a useful supplement for a more comprehensive evaluation.

(3) The concept of mixed-mode buildings can embody both positive and negative attributes of naturally ventilated and fully air-conditioned buildings. Our study discovers that occupants in mixed-mode buildings adopt energy-inefficient behaviors by using air-conditioning in moderate spring and autumn, with winter occupants requiring higher temperatures due to non-hot drinking habits, and summer occupants constantly opening windows while using air conditioning. Therefore, the operation of mixed-mode buildings should aim to minimize the usage of mechanical devices when outdoor conditions are moderate, while resorting to normative heating or cooling operations during adverse weather conditions. This will maximize the advantages of mixed-mode buildings by leveraging the strengths of naturally ventilated and fully air-conditioned buildings, rather than amplifying their shortcomings.

Acknowledgements

We appreciates the grants support from the National Natural Science Foundation of China (Grant No. 52278090), the Ministry of Science and Technology of the People's Republic of China (Grant No. 2022YFC3801504), and the Natural Science Foundation of Chongqing, and China (Grant No. cstc2021ycjh-bgzxm0156) and SuDBE International Research Centre. Mr. Shaoxing Zhang acknowledges the financial support from the program of the China Scholarship Council (No. 202006050214).

References

[1] L. Arakawa, V. Soebarto, and T. Williamson, "Performance evaluation of personal thermal comfort models for older people based on skin temperature, health perception, behavioural and environmental variables," *J. Build. Eng.*, vol. 51, p. 104357, 2022.

[2] S. Vosoughkhosravi, L. Dixon-grasso, and A. Jafari, "The impact of LEED certification on energy performance and occupant satisfaction: A case study of residential college buildings," *J. Build. Eng.*, vol. 59, p. 105097, 2022.

[3] E. Kükrer and N. Eskin, "Effect of design and operational strategies on thermal comfort and productivity in a multipurpose school building," *J. Build. Eng.*, vol. 44, p. 102697, 2021.

[4] W. Li *et al.*, "Indoor temperature preference setting control method for thermal comfort and energy saving based on reinforcement learning," *J. Build. Eng.*, vol. 73, p. 106805, 2023.

[5] T. Hoyt, E. Arens, and H. Zhang, "Extending air temperature setpoints: Simulated energy savings and design considerations for new and retrofit buildings," *Build. Environ.*, vol. 88, pp. 89–96, 2015.

[6] J. Kim, S. Schiavon, and G. Brager, "Personal comfort models – A new paradigm in thermal comfort for occupant-centric environmental control," *Build. Environ.*, vol. 132, pp. 114–124, 2018.

[7] International Standard Organization, "ISO 7730 Ergonomics of the Thermal

Environment—Analytical Determination and Interpretation of Thermal Comfort Using Calculation of the PMV and PPD Indices and Local Thermal Comfort Criteria.” 2005.

[8] ASHRAE, “Thermal Environmental Conditions for Human Occupancy, ANSI/ASHRAE Standard 55-2020.” Atlanta, 2020.

[9] E. UNI, “EN 16798-1:2019 Energy Performance of Buildings-Ventilation for Buildings-Part 1: Indoor Environmental Input Parameters for Design and Assessment of Energy Performance of Buildings Addressing Indoor Air Quality.” Thermal Environment, Lighting and Acoustics 16798.1, Brussels, Belgium, 2019.

[10] CIBSE, “CIBSE Guide A: Environmental design. 8th edition.” London. http://www.cibse.org/getattachment/Knowledge/CIBSE-Guide/CIBSE-Guide-A-Environmental-Design-NEW-2015/Guide-A-presentation.pdf.aspx Accessed 3 November 2019, 2015.

[11] MOHURD, *Evaluation standard for indoor thermal environment in civil buildings (GB/T 50785-2012)*. Ministry of Housing and Urban-Rural Development (MOHURD), Beijing, China, 2012.

[12] European Committee for Standardization, “CEN EN 15251,” *Indoor Environmental Input Parameters for Design and Assessment of Energy Performance of Buildings Addressing Indoor Air Quality, Thermal Environment, Lighting and Acoustics*. Brussels, Belgium, 2007.

[13] ASHRAE, “Thermal Environmental Conditions for Human Occupancy, ANSI/ASHRAE Standard 55-2004.” Atlanta, 2004.

[14] J. Kim, F. Tartarini, T. Parkinson, P. Cooper, and R. de Dear, “Thermal comfort in a mixed-mode building: Are occupants more adaptive?,” *Energy Build.*, vol. 203, p. 109436, 2019.

[15] R. Ming *et al.*, “Assessing energy saving potentials of office buildings based on

adaptive thermal comfort using a tracking-based method," *Energy Build.*, vol. 208, p. 109611, 2020.

[16] E. Arens, M. A. Humphreys, R. De Dear, and H. Zhang, "Are 'class A'temperature requirements realistic or desirable?," *Build. Environ.*, vol. 45, no. 1, pp. 4–10, 2010.

[17] G. Brager, "Mixed-Mode Cooling," *UC Berkeley Cent. Built Environ.*, 2006.

[18] A. Leaman and B. Bordass, "Are users more tolerant of 'green'buildings?," *Build. Res. Inf.*, vol. 35, no. 6, pp. 662–673, 2007.

[19] S. Aisyah, S. Ahmad, H. Bahadur, and S. Wonorahardjo, "Field study on adaptive thermal comfort in office buildings in Malaysia, Indonesia, Singapore, and Japan during hot and humid season," *Build. Environ.*, vol. 109, pp. 208–223, 2016.

[20] S. Manu, Y. Shukla, R. Rawal, L. E. Thomas, and R. De Dear, "Field studies of thermal comfort across multiple climate zones for the subcontinent: India Model for Adaptive Comfort (IMAC)," *Build. Environ.*, vol. 98, pp. 55–70, 2016.

[21] E. Barbadilla-Martín, J. M. Salmerón Lissén, J. Guadix Martín, P. Aparicio-Ruiz, and L. Brotas, "Field study on adaptive thermal comfort in mixed mode office buildings in southwestern area of Spain," *Build. Environ.*, vol. 123, pp. 163–175, 2017.

[22] M. Luo, B. Cao, J. Damiens, B. Lin, and Y. Zhu, "Evaluating thermal comfort in mixed-mode buildings: A field study in a subtropical climate," *Build. Environ.*, vol. 88, pp. 46–54, 2015.

[23] J. Kim, R. de Dear, T. Parkinson, and C. Candido, "Understanding patterns of adaptive comfort behaviour in the Sydney mixed-mode residential context," *Energy Build.*, vol. 141, pp. 274–283, 2017.

[24] R. F. Rupp, R. de Dear, and E. Ghisi, “Field study of mixed-mode office buildings in Southern Brazil using an adaptive thermal comfort framework,” *Energy Build.*, vol. 158, pp. 1475–1486, 2018.

[25] M. Khoshbakht, Z. Gou, and F. Zhang, “A pilot study of thermal comfort in subtropical mixed-mode higher education office buildings with different change-over control strategies,” *Energy Build.*, vol. 196, pp. 194–205, 2019.

[26] X. Jia, B. Cao, Y. Zhu, and B. Liu, “Thermal comfort in mixed-mode buildings: A field study in Tianjin, China,” *Build. Environ.*, vol. 185, p. 107244, 2020.

[27] M. A. Gaffoor, M. Eftekhari, and X. Luo, “Evaluation of thermal comfort in mixed-mode buildings in temperate oceanic climates using American Society of Heating, Refrigeration, and Air Conditioning Engineers Comfort Database II,” *Build. Serv. Eng. Res. Technol.*, vol. 43, no. 3, pp. 379–401, 2022.

[28] S. Khadka, M. Shrestha, and H. B. Rijal, “Investigation of the thermal comfort and productivity in Japanese mixed-mode office buildings,” *J. Eng. Res. [TJER]*, vol. 19, no. 1, pp. 63–72, 2022.

[29] M. P. Deuble and R. J. de Dear, “Mixed-mode buildings: A double standard in occupants’ comfort expectations,” *Build. Environ.*, vol. 54, pp. 53–60, 2012.

[30] T. Cheung, S. Schiavon, T. Parkinson, P. Li, and G. Brager, “Analysis of the accuracy on PMV – PPD model using the ASHRAE Global Thermal Comfort Database II,” *Build. Environ.*, vol. 153, pp. 205–217, 2019.

[31] S. Natarajan, J. Rodriguez, and M. Vellei, “A field study of indoor thermal comfort in the subtropical highland climate of Bogota, Colombia,” *J. Build. Eng.*, vol. 4, pp. 237–246, 2015.

[32] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[33] A. Costantino, M. Ferrara, M. Arnesano, and E. Fabrizio, “Off-the-shelf wearable sensing devices for personalized thermal comfort models: A systematic review on their use in scientific research,” *J. Build. Eng.*, vol. 70, p. 106379, 2023.

[34] D. Li, C. C. Menassa, and V. R. Kamat, “Robust non-intrusive interpretation of occupant thermal comfort in built environments with low-cost networked thermal cameras,” *Appl. Energy*, vol. 251, p. 113336, 2019.

[35] K. N. Nkurikiyeyezu, Y. Suzuki, and G. F. Lopez, “Heart rate variability as a predictive biomarker of thermal comfort,” *J. Ambient Intell. Humaniz. Comput.*, vol. 9, no. 5, pp. 1465–1477, 2018.

[36] B. Peng and S. Hsieh, “Data-driven thermal comfort prediction with support vector machine,” *Int. Manuf. Sci. Eng. Conf.*, vol. 50749, p. V003T04A044, 2017.

[37] P. Li, P. Dai, D. Cao, B. Liu, and Y. Lu, “Non-intrusive comfort sensing: Detecting age and gender from infrared images for personal thermal comfort,” *Build. Environ.*, vol. 219, p. 109256, 2022.

[38] Z. Wu *et al.*, “Using an ensemble machine learning methodology-Bagging to predict occupants’ thermal comfort in buildings,” *Energy Build.*, vol. 173, pp. 117–127, 2018.

[39] B. G. Heusinkveld, G. Sterenborg, G. J. Steeneveld, J. J. Attema, R. J. Ronda, and A. A. Holtslag, “Smartphone app brings human thermal comfort forecast in your hands,” *Bull. Am. Meteorol. Soc.*, vol. 98, no. 12, pp. 2533–2541, 2017.

[40] F. Jazizadeh and W. Jung, “Personalized thermal comfort inference using RGB video images for distributed HVAC control,” *Appl. Energy*, vol. 220, no. March, pp. 829–841, 2018.

[41] X. Shan and E. Yang, “Supervised machine learning of thermal comfort under different indoor temperatures using EEG measurements,” *Energy Build.*, vol.

225, p. 110305, 2020.

[42] H. Kim, D. Jung, H. Choi, and T. Hong, “Advanced prediction model for individual thermal comfort considering blood glucose and salivary cortisol,” *Build. Environ.*, vol. 224, p. 109551, 2022.

[43] M. Grandini, E. Bagli, and G. Visani, “Metrics for multi-class classification: an overview,” *arXiv Prepr. arXiv*, p. 2008.05756, 2020.

[44] MOHURD, *Thermal design Code for Civil Building (GB/T 50176-2016)*. China Architecture and Building Press, Beijing, China, 2016.

[45] R. Yao, V. Costanzo, X. Li, Q. Zhang, and B. Li, “The effect of passive measures on thermal comfort and energy conservation. A case study of the hot summer and cold winter climate in the Yangtze River region,” *J. Build. Eng.*, vol. 15, pp. 298–310, 2018.

[46] R. Yao, B. Li, and J. Liu, “A theoretical adaptive model of thermal comfort – Adaptive Predicted Mean Vote (aPMV),” *Build. Environ.*, vol. 44, no. 10, pp. 2089–2096, 2009.

[47] H. Liu, Y. Wu, B. Li, Y. Cheng, and R. Yao, “Seasonal variation of thermal sensations in residential buildings in the Hot Summer and Cold Winter zone of China,” *Energy Build.*, vol. 140, pp. 9–18, 2017.

[48] S. Zhang, R. Yao, C. Du, E. Essah, and B. Li, “Analysis of outlier detection rules based on the ASHRAE global thermal comfort database,” *Build. Environ.*, vol. 234, p. 110155, 2023.

[49] M. Luo *et al.*, “Comparing machine learning algorithms in predicting thermal sensation using ASHRAE Comfort Database II,” *Energy Build.*, vol. 210, p. 109776, 2020.

[50] P. F *et al.*, “Scikit-learn: Machine learning in Python,” *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[51] D. J. Hand and K. Yu, “Idiot’s Bayes—not so stupid after all?,” *Int. Stat. Rev.*, vol. 69, no. 3, pp. 385–398, 2001.

[52] M. E. Tipping, “Sparse Bayesian learning and the relevance vector machine,” *J. Mach. Learn. Res.*, pp. 211–244, 2001.

[53] S. S. Y. Ng, Y. Xing, and K. L. Tsui, “A naive Bayes model for robust remaining useful life prediction of lithium-ion battery,” *Appl. Energy*, vol. 118, pp. 114–123, 2014.

[54] M. Izanloo, A. Aslani, and R. Zahedi, “Development of a Machine learning assessment method for renewable energy investment decision making,” *Appl. Energy*, vol. 327, p. 120096, 2022.

[55] M. S. Ibrahim, W. Dong, and Q. Yang, “Machine learning driven smart electric power systems : Current trends and new perspectives,” *Appl. Energy*, vol. 272, p. 115237, 2020.

[56] J. Rennie, L. Shih, J. Teevan, and D. Karger, “Tackling the poor assumptions of naive bayes text classifiers,” *Proc. 20th Int. Conf. Mach. Learn.*, pp. 616–623, 2003.

[57] T. K. Ho, “Random decision forests,” *Proc. 3rd Int. Conf. Doc. Anal. Recognit.*, vol. 1, pp. 278–282, 1995.

[58] P. Probst, M. N. Wright, and A.-L. Boulesteix, “Hyperparameters and tuning strategies for random forest,” *Wiley Interdiscip. Rev. data Min. Knowl. Discov.*, vol. 9, no. 3, p. e1301, 2019.

[59] B. Sovdat, “Updating formulas and algorithms for computing entropy and gini index from time-changing data streams,” *arXiv Prepr. arXiv*, p. 1403.6348, 2014.

[60] C. E. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.

[61] W.-Y. Loh, “Classification and regression trees,” *Wiley Interdiscip. Rev. data Min. Knowl. Discov.*, vol. 1, no. 1, pp. 14–23, 2011.

[62] A. Géron, “Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow,” *O’Reilly Media, Inc*, 2022.

[63] R. G. Mantovani, A. L. D. Rossi, J. Vanschoren, B. Bischl, and A. C. P. L. F. Carvalho, “To tune or not to tune: recommending when to adjust SVM hyper-parameters via meta-learning,” *2015 Int. Jt. Conf. neural networks*, pp. 1–8, 2015.

[64] A. Hussien, W. Khan, A. Hussain, and P. Liatsis, “Predicting energy performances of buildings’ envelope wall materials via the random forest algorithm,” *J. Build. Eng.*, vol. 69, p. 106263, 2023.

[65] X. Zhou, J. Ren, J. An, D. Yan, X. Shi, and X. Jin, “Predicting open-plan office window operating behavior using the random forest algorithm,” *J. Build. Eng.*, vol. 42, p. 102514, 2021.

[66] Y. Wu and B. Cao, “Recognition and prediction of individual thermal comfort requirement based on local skin temperature,” *J. Build. Eng.*, vol. 49, p. 104025, 2022.

[67] N. Gao, W. Shao, M. Saiedur, J. Zhai, K. David, and F. D. Salim, “Transfer learning for thermal comfort prediction in multiple cities,” *Build. Environ.*, vol. 195, p. 107725, 2021.

[68] H. Zhang, X. Yang, R. Tu, J. Huang, and Y. Li, “Thermal Comfort Modeling of Office Buildings Based on Improved Random Forest Algorithm,” *2022 IEEE 11th Data Driven Control Learn. Syst. Conf.*, pp. 1369–1376, 2022.

[69] C. Yu, B. Li, Y. Wu, B. Chen, and R. Kosonen, “Performances of machine learning algorithms for individual thermal comfort prediction based on data from professional and practical settings,” *J. Build. Eng.*, vol. 61, p. 105278, 2022.

[70] A. Aryal and B. Becerik-gerber, “Thermal comfort modeling when personalized comfort systems are in use: Comparison of sensing and learning methods,” *Build. Environ.*, vol. 185, p. 107316, 2020.

[71] Q. Y. Li, J. Han, and L. Lu, “A random forest classification algorithm based personal thermal sensation model for personalized conditioning system in office buildings,” *Comput. J.*, vol. 64, no. 3, pp. 500–508, 2021.

[72] A. A. Farhan, K. Pattipati, B. Wang, and P. Luh, “Predicting individual thermal comfort using machine learning algorithms,” *2015 IEEE Int. Conf. Autom. Sci. Eng.*, pp. 708–713, 2015.

[73] M. Vellei, M. Herrera, D. Fosas, and S. Natarajan, “The influence of relative humidity on adaptive thermal comfort,” *Build. Environ.*, vol. 124, pp. 171–185, 2017.

[74] D. Li, C. C. Menassa, and V. R. Kamat, “Non-intrusive interpretation of human thermal comfort through analysis of facial infrared thermography,” *Energy Build.*, vol. 176, pp. 246–261, 2018.

[75] T. Chaudhuri, D. Zhai, Y. C. Soh, H. Li, and L. Xie, “Random forest based thermal comfort prediction from gender-specific physiological parameters using wearable sensing technology,” *Energy Build.*, vol. 166, pp. 391–406, 2018.

[76] S. S. Shetty, D. Chinh, M. Gupta, and S. K. Panda, “Learning desk fan usage preferences for personalised thermal comfort in shared offices using tree-based methods,” *Build. Environ.*, vol. 149, pp. 546–560, 2019.

[77] S. Liu, S. Schiavon, H. P. Das, M. Jin, and C. J. Spanos, “Personal thermal comfort models with wearable sensors,” *Build. Environ.*, vol. 162, p. 106281, 2019.

[78] Y. Wu, Z. Zhang, H. Liu, H. Cui, and Y. Cheng, “Optimal local skin temperatures for mean skin temperature estimation and thermal comfort

prediction of seated person in thermally stratified environments," *J. Therm. Biol.*, vol. 111, p. 103389, 2023.

[79] F. Tartarini and S. Schiavon, "SoftwareX pythermalcomfort: A Python package for thermal comfort research," *SoftwareX*, vol. 12, p. 100578, 2020.

[80] M. Humphreys, "Field Studies of Thermal Comfort Compared and Applied," *Build. Serv. Eng.*, vol. 44, no. 27, pp. 5–23, 1976.

[81] R. Yao *et al.*, "Evolution and performance analysis of adaptive thermal comfort models–a comprehensive literature review," *Build. Environ.*, vol. 217, p. 109020, 2022.

[82] J. Singh and M. Khushi, "Feature learning for stock price prediction shows a significant role of analyst rating," *Appl. Syst. Innov.*, vol. 4, no. 1, p. 17, 2021.

[83] B. R. Jackson, "The dangers of false-positive and false-negative test results: false-positive results as a function of pretest probability," *Clin. Lab. Med.*, vol. 28, no. 2, pp. 305–319, 2008.

[84] W. Jung and F. Jazizadeh, "Human-in-the-loop HVAC operations: A quantitative review on occupancy, comfort, and energy-efficiency dimensions," *Appl. Energy*, vol. 239, pp. 1471–1508, 2019.

[85] D. Ballabio, F. Grisoni, and R. Todeschini, "Multivariate comparison of classification performance measures," *Chemom. Intell. Lab. Syst.*, vol. 174, pp. 33–44, 2018.

[86] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *Ann. Math. Stat.*, pp. 50–60, 1947.

[87] F. McElduff, M. Cortina-Borja, S.-K. Chan, and A. Wade, "When t-tests or Wilcoxon-Mann-Whitney tests won't do," *Adv. Physiol. Educ.*, vol. 34, no. 3, pp. 128–133, 2010.

[88] R. Maiti, “PMV model is insufficient to capture subjective thermal response from Indians,” *Int. J. Ind. Ergon.*, vol. 44, no. 3, pp. 349–361, 2014.

[89] T. Chaudhuri, Y. C. Soh, H. Li, and L. Xie, “Machine learning driven personal comfort prediction by wearable sensing of pulse rate and skin temperature,” *Build. Environ.*, vol. 170, p. 106615, 2020.

[90] J. Langevin, P. L. Gurian, and J. Wen, “Reducing energy consumption in low income public housing: Interviewing residents about energy behaviors,” *Appl. Energy*, vol. 102, pp. 1358–1370, 2013.

[91] P. Virtanen *et al.*, “SciPy 1.0: fundamental algorithms for scientific computing in Python,” *Nat. Methods*, vol. 17, no. 3, pp. 261–272, 2020.

[92] P. Fanger, *Thermal comfort, Analysis and Applications in Environmental Engineering*. Copenhagen: Danish Technical Press, 1970.

[93] M. Buckland and F. Gey, “The relationship between recall and precision,” *J. Am. Soc. Inf. Sci.*, vol. 45, no. 1, pp. 12–19, 1994.

[94] D. Waterworth, S. Sethuvenkatraman, and Q. Z. Sheng, “Advancing smart building readiness: automated metadata extraction using neural language processing methods,” *Adv. Appl. Energy*, vol. 3, p. 100041, 2021.

[95] P. Taylor, C. R. Blyth, and C. R. Blyth, “On Simpson’s paradox and the sure-thing principle,” *J. Am. Stat. Assoc.*, vol. 67, no. 338, pp. 364–366, 1972.

[96] S. Nembrini, I. R. König, and M. N. Wright, “The revival of the Gini importance?,” *Bioinformatics*, vol. 34, no. 21, pp. 3711–3718, 2018.

[97] V. Földváry Ličina *et al.*, “Development of the ASHRAE Global Thermal Comfort Database II,” *Build. Environ.*, vol. 142, pp. 502–512, 2018.

[98] D. Fay, L. O. Toole, and K. N. Brown, “Gaussian Process models for ubiquitous user comfort preference sampling; global priors, active sampling and outlier rejection,” *Pervasive Mob. Comput.*, vol. 39, pp. 135–158, 2017.

[99] Z. Ostrowski, M. Rojczyk, and I. Szczygiel, “Dry heat loses of newborn baby in infant care bed: use of a thermal manikin,” *J. Phys. Conf. Ser.*, vol. 745, no. 3, p. 032087, 2016.

[100] F. Yuan *et al.*, “Thermal comfort in hospital buildings–A literature review,” *J. Build. Eng.*, vol. 45, p. 103463, 2022.

[101] L. Waldron *et al.*, “Comparative meta-analysis of prognostic gene signatures for late-stage ovarian cancer,” *JNCI J. Natl. Cancer Inst.*, vol. 106, no. 5, p. dju049, 2014.

[102] X. Ying, “An overview of overfitting and its solutions,” *J. Phys. Conf. Ser.*, vol. 1168, p. 022022, 2019.

[103] G. J. Babu, “Resampling methods for model fitting and model selection,” *J. Biopharm. Stat.*, vol. 21, no. 6, pp. 1177–1186, 2011.

[104] R. Yao, J. Liu, and B. Li, “Occupants’ adaptive responses and perception of thermal environment in naturally conditioned university classrooms,” *Appl. Energy*, vol. 87, no. 3, pp. 1015–1022, 2010.

[105] W. Yu, B. Li, R. Yao, D. Wang, and K. Li, “A study of thermal comfort in residential buildings on the Tibetan Plateau, China,” *Build. Environ.*, vol. 119, pp. 71–86, 2017.

[106] E. Arens, S. Turner, H. Zhang, and G. Paliaga, “Moving air for comfort,” *ASHRAE J*, vol. 51, pp. 18–28, 2009.

[107] W. Chen *et al.*, “Ceiling-fan-integrated air conditioning: Airflow and temperature characteristics of a sidewall-supply jet interacting with a ceiling fan,” *Build. Environ.*, vol. 171, p. 106660, 2020.