Energy & Buildings 284 (2023) 112856

Contents lists available at ScienceDirect

Energy & Buildings

journal homepage: www.elsevier.com/locate/enb

How do people set air conditioning temperature setpoint in urban domestic–Behavior model in Chinese three climate zones based on historical usage data

Junmeng Lyu [a,b], Jinbo Li [c], Zisheng Zhao [c], Xiongwei Miao [c], Heng Du [a], Dayi Lai [a], Yuxin Yang [a], Zhiwei Lian [a,b,*]

[a] School of Design, Shanghai Jiao Tong University, Shanghai 200240, China
[b] China Institute of Urban Governance, Shanghai Jiao Tong University, Shanghai 200030, China
[c] Guangdong Midea Air-Conditioning Equipment Co., Ltd., Guangdong, China

ARTICLE INFO

Article history:
Received 23 November 2022
Revised 14 January 2023
Accepted 2 February 2023
Available online 4 February 2023

Keywords:
Air-conditioning control model
Air-conditioning usage behaviors
China climate zone
Urban governance
Machine learning

ABSTRACT

Predicting user behavior in setting air conditioning (A/C) can help develop automatic A/C control techniques, but it is challenging. Existing studies have largely focused on A/C switching behavior models, with fewer studies exploring temperature-setting behavior models in various climate zones. This study analyzed differences in household A/C temperature setpoint behaviors in three climate zones, using usage data collected from embedded sensors in A/C during the cooling season. Further, five machine learning (ML) algorithms were applied to model the A/C temperature setpoint behavior for the three climate zones. The results showed that the behavior is related to time, indoor air temperature (T_{in}), outdoor air temperature (T_{out}), temperature setpoint before settings ($T_{set'}$), and cumulative time the A/C is turned on (t_c). A/C usage behavior varied significantly among users in different climate zones. Users in hot summer and warm winter (HSWW) have a higher acceptable T_{in} and the opposite in Cold zones. In HSWW zones, users were able to find their preferred setting temperature more quickly. By applying the random forest algorithm and proper data pre-processing, the model can predict the "Warmer", "Keep setpoint unchanged" and "Cooler" settings with 0.700–0.810 macro F1 score and 0.936–0.961 accuracy. This shows the excellent performance of ML algorithms in predicting user behavior with large data sizes. The T_{in}, $T_{set'}$, and t_c are important input features. This work aims to model A/C temperature setpoint behavior through ML algorithms to help develop intelligent A/C control methods for different climate zones in the future.

1. Introduction

As living standards improve, air conditioning (A/C) users are increasingly demanding intelligent, energy-saving systems [1,2]. A growing number of A/C engineers are exploring the possibility of automatically changing A/C temperature setpoints [3,4]. This "smarter" A/C can automatically increase or decrease the room temperature as needed, avoiding the energy waste that can result from maintaining constant air supply parameters for extended periods [5–7]. At the same time, it can improve the quality of life for users. To develop such intelligent systems, it is necessary to have a thorough understanding of how users interact with their A/Cs [8,9].

A/C temperature setpoint behavior is an important part of user interaction with A/C [10]. In the past, users had to manually change the temperature through the controller when they were too hot or too cold. However, recent advancements in research have allowed A/C engineers to create automatic control technology that takes into account thermal comfort [11]. There are two main approaches to this. The first is to incorporate thermal comfort models into the automatic control system [12–15], but these models often require input such as operative temperature and clothing insulation, which can be difficult to obtain through embedded sensors in A/Cs [16]. The second approach is to monitor physiological parameters that reflect a person's thermal sensation in real time using wearable devices or non-contact measurements and use them as inputs for the automatic control system [17–20]. However, this approach also

* Corresponding author at: School of Design, Shanghai Jiao Tong University, Shanghai 200240, China.

E-mail address: zwlian@sjtu.edu.cn (Z. Lian).

https://doi.org/10.1016/j.enbuild.2023.112856
0378-7788/

Nomenclature

T_{in}	Indoor air temperature, °C
T_{out}	Outdoor air temperature, °C
T_{set}	Air conditioning setpoint temperature after setting behavior, °C
$T_{set'}$	Air conditioning setpoint temperature before setting behavior, °C
$T_{set'\text{-}in}$	Air conditioning setpoint and indoor temperature difference, °C
ΔT_{set}	Air conditioning setpoint temperature change, °C
t_c	Cumulative time, s
$T_{out\text{-}in}$	Indoor and outdoor temperature difference, °C
TP_i	Number of samples in category "i" that are correctly predicted
FN_i	Number of samples that actually are category "i", but are predicted to be other categories.
FP_i	Number of samples predicted to be category "i", but actually other categories
RH	Relative humidity, %

Abbreviations

A/C	Air-conditioning
ANN	Artificial Neural Network
D	Day
Gaussian NB	Gaussian Naive Bayes
H	Hour
HSCW	Hot summer and cold winter
HSWW	Hot summer and warm winter
KNN	K-Neighbors
LR	Logistic regression
M	Month
rbf	Radial basis function
RF	Random forest
SD	Standard deviation
SMOTE	Synthetic minority over-sampling Technique
SVC	Support vector classification

has limitations as it does not consider other factors such as the acoustic and visual environments that can affect a person's physiological responses [21,22]. Additionally, non-contact measurement devices like thermal cameras and millimeter wave radar have limitations in terms of installation angle, measurement distance, and the imbalance between instrument precision and cost [23].

Many researchers focused on developing occupant behavior models for automatic A/C control using a large sample of user behavior data. These models allow for a better understanding of the interaction between occupants and A/Cs. These models do not require any additional physiological measurements, making them more cost-effective. Previous studies have explored the factors that influence occupants' A/C setting behavior, as summarized in Table 1. Many studies have shown that a person's daily schedule, indoor air temperature (T_{in}), and outdoor air temperature (T_{out}) are closely connected to their use of A/Cs. Further, some researchers have developed models based on data samples that take into account various important influencing factors to predict A/C usage. Xia et al. [24] used statistical and cluster analysis on 102 samples of bedroom A/Cs in Guangzhou, China to create a model that predicts when the A/C will be turned on or off using time, date, and room type as inputs. Yao et al. [25] employed statistical methods to create a model for predicting the schedule for turning an A/C on and off, as well as the desired temperature setpoint, in an apartment located in Ningbo, China. Similarly, Song et al. [26] utilized statistical analysis to develop a model for predicting the trigger

Table 1
Previous studies on A/C usage behavior.

Ref.	Location	Climate zone	Cooling/ Heating	Sample size	Occupant behavior	Factors
[27]	New Jersey, America	HSCW climate	Cooling	13 A/Cs	Turn on/off A/Cs	T_{out} and users' daily schedule
[28]	Guangzhou, China	HSWW climate	Cooling	10 A/Cs	Turn on/off A/Cs; A/C setpoint temperature setting	T_{in} and users' daily schedule
[29]	Kuala Lumpur, Malaysia	Tropical climate	Cooling	38 A/Cs	Turn on/off A/Cs	Users' daily schedule
[30]	Fukuoka, Japan	Moderate climate	Cooling	20 A/Cs	Turn on/off A/Cs	T_{out} and users' daily schedule
[31]	Eight cities, China	Cold climate, HSCW climate, HSWW climate	Cooling	34 A/Cs	Turn on/off A/Cs	T_{out}, users' daily schedule, and cumulative time of A/C on (t_c)
[32]	Tokyo, Japan	Moderate climate	Cooling & Heating	39 A/Cs	Turn on/off A/Cs	T_{out}, the preference and background of the individual subject
[33]	Fukuoka, Japan	Moderate climate	Cooling	8 A/Cs	Turn on/off A/Cs	T_{in}
[34]	Hongkong, China	HSWW climate	Cooling & Heating	554 questionnaires	Turn on/off A/Cs; A/C setpoint temperature setting	T_{in}, bedding insulation, and familiarity with A/C operation
[35]	Seven cities, China	Cold climate, HSCW climate, HSWW climate	Cooling & Heating	400 A/Cs	Turn on/off A/Cs	T_{out} and t_c
[26]	Tianjin China	Cold climate	Cooling & Heating	43 A/Cs	Turn on/off A/Cs; A/C setpoint temperature setting	T_{in}, T_{out}, climate zone, education level, and household income
[36]	Sydney and Wollongong, Australian	HSCW climate	Cooling & Heating	42 A/Cs	Turn on/off A/Cs; A/C setpoint temperature setting	T_{in} and T_{out}
[37]	Chongqing, China	HSCW climate	Cooling	1287 A/Cs	A/C setpoint temperature setting	T_{in}, T_{out}, room type, and individual thermal preferences

temperature when the A/C is turned on for 43 devices in Tianjin, China using T_{in}, occupant thermal sensation, clothing insulation, and personal characteristics as input variables. However, most of the models proposed in these studies are based on user behavior for turning the A/C on and off, and time-series models for the A/C operation state. Few studies have focused on temperature setpoint setting behavior. This is largely due to the difficulty in obtaining the temperature setpoints of A/Cs. In addition, most models require additional sensors beyond those installed in the A/C in order to gather the necessary input variables for self-control. This leads to higher costs and makes implementation difficult.

Luo et al. [9] recently conducted a study that used data from A/C sensors in Chinese hot-summer and warm-winter (HSWW) climate zones to model A/C temperature setting behavior for the first time. The model can predict "increase setting temperature" and "decrease setting temperature" settings with an accuracy of 72.1–87.3 %. The input variables for the model only required data from the A/C sensors, without the need for any additional complicated variables. However, the study has the limitation that the proposed model is only applicable to the specific climate in China. It may not be useful in other climate zones. Although the basic mechanism of human-thermal environment relationship does not vary with climate, the long-term impact of climate can affect how people in different zones perceive and tolerate indoor temperatures. As shown in Table 1, climate-related factors have a significant impact on A/C usage by users. In China, which has a diverse range of climate types [38], the Cold zone, hot summer and cold winter (HSCW) zone, and HSWW zone have high A/C usage rates. Few studies conducted in China and other areas with similar climates to analyze the A/C temperature setpoint behaviors based solely on data collected from A/C sensors. This suggests that it is meaningful to analyze the various factors that impact the A/C temperature setpoint behaviors in different climate zones in China and create a model that can be applied to various climate zones.

The research conducted on the intersection between the machine learning (ML) field and the indoor artificial environment field is currently developing rapidly. ML algorithms were utilized to build highly accurate models for predicting user thermal sensation [39,40], natural ventilation and window opening behavior [41], and more. Many of the existing models for predicting A/C operation behavior were based on statistical methods. As the era of big data comes, ML algorithms have shown excellent performance in handling classification tasks with large sample size datasets. But it is yet to be determined if ML algorithms can create high-performing models for predicting A/C temperature setpoint behavior in various climate zones.

This study will analyze datasets of summer household A/C usage in Cold, HSCW, and HSWW zones in China, provided by an A/C manufacturer. The research will aim to answer the following questions: 1. What factors influence the behavior of users setting their A/C temperature? 2. Are there any differences in A/C temperature setting behavior in different climate zones? 3. How can ML algorithms be used to create a model for A/C temperature setpoint behavior during the summer based on data collected by A/C sensors in various climate zones?

2. Methods

2.1. Data set

The A/C device samples in this study were 1,006 units, all installed in domestic buildings in three climate zones in China. The devices were located in most of the areas where A/C was commonly used during the summer, distributed throughout southern and northern China. The locations of the A/C devices can be seen in Fig. 1. The bedroom area of these subject households was 12–20 m^2 and the living room area was about 30–40 m^2. The average room window size was 2–5 m^2. The subject households were 2–5 persons per family. The data used in this study were obtained from embedded sensors in A/Cs. This dataset included information on the status of the A/C, T_{in}, T_{out}, and user interaction behaviors. The dataset consisted of 1,753,579 historical usage data samples from all devices over the cooling months (from June to September) in 2019 and 2020. Detailed information can be found in Table 2.

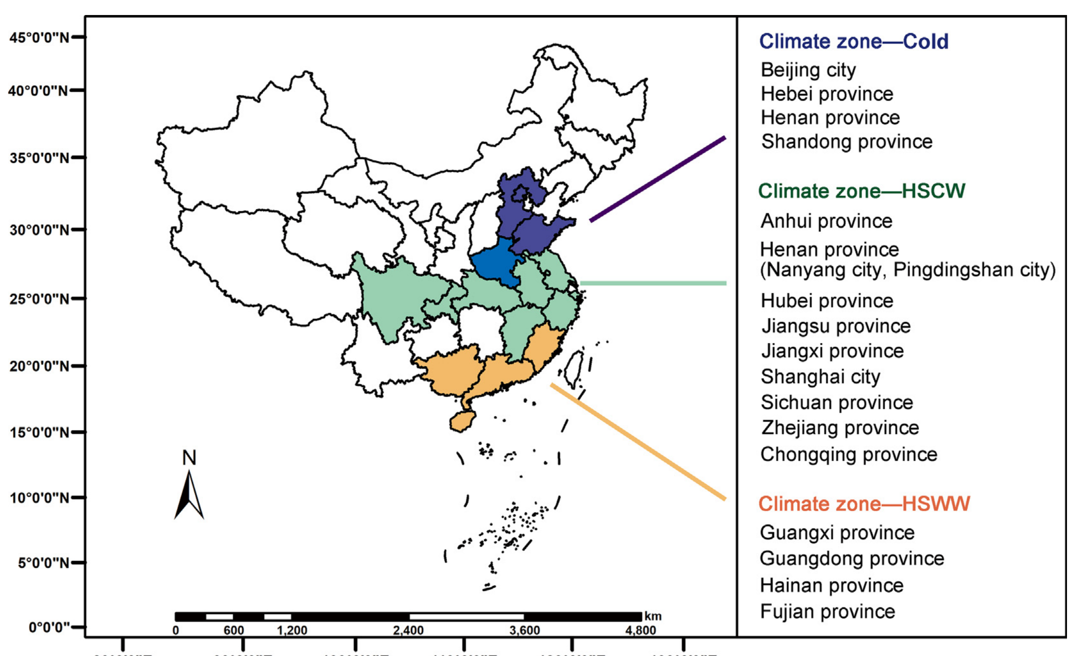


Fig. 1. Location of the sample devices.

Table 2
The information of the raw data set.

Climate Zone	A/C Devices number	Samples number	Sampling time	Original parameters
Cold zone	369	685,216	June to September 2019 and June to September 2020	Device ID, Month (M), Day (D), Hour (H), reflector, temperature setpoint (T_{set}), on/off, T_{in}, T_{out}, and so on
HSCW zone	321	498,019		
HSWW zone	316	570,344		

2.2. Development of A/C control model

The study investigated the behavior of A/C temperature setpoints, including increasing the temperature (Warmer), keeping it unchanged (NC), and decreasing it (Cooler). The model development process is depicted in Fig. 2, which consists of six modules. In the "Data Processing" module, samples with abnormal or null data were cleaned and sorted by climate zone. The processed data was then used to generate derived data, such as the difference between outdoor and indoor temperature (T_{out-in}) and the cumulative time the A/C is on (t_c). The "Data Analysis" module utilized analysis of variance (ANOVA) and Pearson correlation analysis to analyze differences in A/C temperature setpoint behavior among the climate zones and identified variables associated with the behavior. However, the sample size for each A/C temperature setpoint behavior was unbalanced, so the "Data Balancing" module applied the syn-

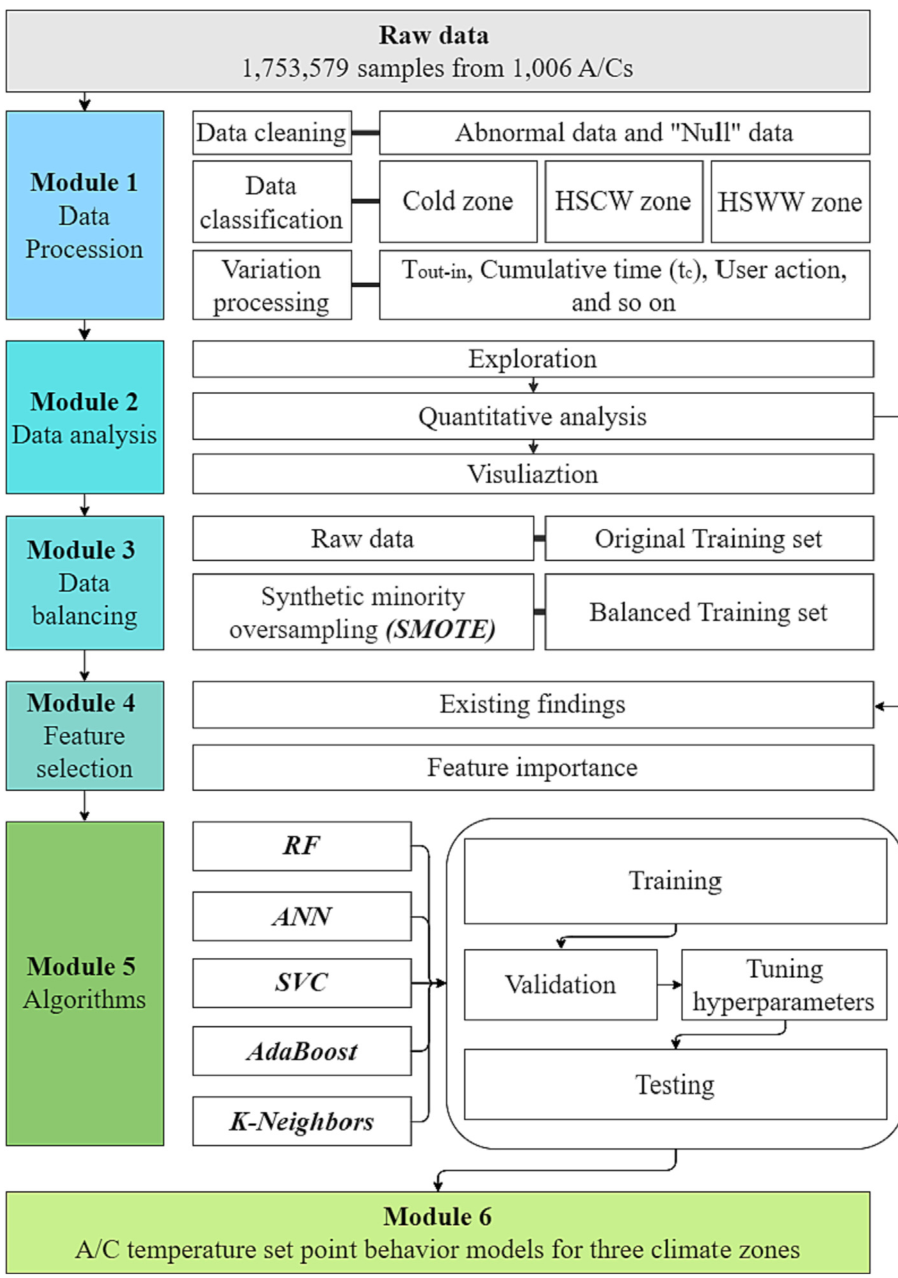


Fig. 2. Diagram of modeling of A/C temperature setpoint behavior.

thetic minority oversampling technique (SMOTE) to balance the training set. In the "Feature Selection" module, variables related to behaviors in each climate zone were chosen as model features based on the results from the "Data Analysis" module. The "Algorithms" module employed five algorithms to model the behavior and compared their performance on the test set. Finally, the A/C temperature setpoint behavior model for the three climate zones was developed using the optimal algorithm. Details will be stated in the following sections. This study used the Python 3.10.6 compiler language to develop the model on the "Jupyter notebook". The compiled model can be found in the supplementary materials and can be directly implemented.

2.2.1. User interaction data pre-processing

In order to build the model successfully, the raw data set was pre-processed [9]. To ensure the validity of the sample data, those with missing or non-reasonable values were removed. For instance, samples with T_{set} above 30 °C and T_{out} above 100 °C were deleted. The remaining samples were then sorted by ascending time series and assigned sequential numbers from "1″ to "n". The T_{set} of the sample numbered "n-1″ was added to the sample numbered "n", creating a new data column referred to as the "A/C temperature setpoint before the temperature setting behavior ($T_{set'}$)". The difference between T_{set} and $T_{set'}$ (ΔT_{set}) was used to determine the A/C setting behavior: If ΔT_{set} > 0, the user behavior is "Warmer"; if ΔT_{set} = 0, the user behavior is "NC"; if ΔT_{set} < 0, the user behavior is "Cooler". The difference between $T_{set'}$ and T_{in} ($T_{set'\text{-}in}$), $T_{out\text{-}in}$, and t_c were also calculated and included in the processed dataset. Detailed information of final dataset can be found in Table 3. The final dataset consisted of 1,642,083 historical usage data samples, as exemplified in Table 4.

2.2.2. Data exploration

Before developing the model, it is imperative to determine if there is a statistically significant difference in A/C temperature setpoint behavior among different climate zones. Thus, ANOVA and Pearson correlation analysis were employed to determine whether there were significant differences between climate zones in the data gathered by the A/C sensors. The Pearson coefficient between the two variables were strongly correlated at greater than 0.6, moderately correlated at 0.4–0.6, and weakly correlated at 0.2–0.4. The symbol "*" indicates significant difference ($p < 0.05$), "**" indicates significant difference ($p < 0.01$). Due to the large sample size of this data, the effect size Cohen's d was calculated to avoid giving subtle but not practically significant differences. The Cohen's d was 0.15–0.4 for a small effect size, 0.4–0.75 for a medium effect size, and greater than 0.75 for a high effect size. All data were statistically analyzed using the SPSS software (version 26.0; SPSS, Inc., Chicago, IL, USA).

Table 3
The sample size of the processed data set.

Climate Zone	A/C Devices number	Sample size	
Cold zone	369	640,389	Warmer: 23,735 NC: 594,546 Cooler: 22,108
HSCW zone	321	465,859	Warmer: 18,522 NC: 604,966 Cooler: 16,901
HSWW zone	316	535,835	Warmer: 15,243 NC: 506,488 Cooler: 14,104

2.2.3. Data balancing

The total number of user temperature setting behaviors in the processed dataset was 1,642,083, with 3.5 % classified as "Warmer" and 3.3 % classified as "Cooler". The "NC" category had about 25 times the number of samples compared to the other two categories. To address the issue of data imbalance, SMOTE was applied to the training set to balance the data [42]. The SMOTE works by randomly selecting a minority class sample ("Cooler" or "Warmer") and generates synthetic samples that are similar to the selected sample but differ in some feature values. These synthetic samples are added to the dataset to increase the size of the minority class to match that of the majority class ("NC"). Further details on the use of SMOTE can be found in study [42]. In addition to the three-classification model, the binary classification model, which distinguishes between "Warmer" and "Cooler" categories, was also built in this study (see Section 3.2). As the number of samples for the two classifications was relatively balanced, the SMOTE oversampling method was not utilized in the binary classification model.

2.2.4. Algorithm comparison

Various ML algorithms demonstrate varying levels of effectiveness when faced with different classification tasks, which can be influenced by the number of features, sample size, and data distribution characteristics. Researchers have identified seven commonly utilized ML classifiers in indoor environment studies: Random Forest (RF), Artificial Neural Network (ANN), Support Vector Classification (SVC), Logistic Regression (LR), AdaBoost, K-Neighbors (KNN), and Gaussian Naive Bayes (Gaussian NB) [43–45]. The characteristics of these algorithms are summarized in Table 5. Of these, LR and the Gaussian NB are not applicable to this study due to the non-linear separability of the samples and the interdependence of features. Therefore, this study compares the performance of the remaining five algorithms in modeling A/C temperature setpoint behavior models, with the aim of selecting the most suitable algorithm. The Scikit-learn ML library developed in Python was used to invoke algorithms.

2.2.5. Training, validation, testing, and model performance

In order to develop the model, the "train_test_split" function was utilized to split the data set into a training set (60 % of the data), a validation set (10 % of the data), and a test set (30 % of the data).

The accuracy, as calculated by Eq. (1), is an indicator that reflects the performance of the model. It is determined by dividing the number of correctly predicted data points by the total number of data points. However, this indicator is reliable when the sample

Table 4
Example of processed data samples.

No.	M	D	H	T_{set} (°C)	$T_{set'}$ (°C)	Time interval (s)	t_c (s)	T_{in} (°C)	T_{out} (°C)	ON-OFF	ΔT_{set} (°C)	T_{set} setting behavior	$T_{out\text{-}in}$ (°C)	$T_{set'\text{-}in}$ (°C)
11	7	14	23	26.5	26.5	0	0	25.4	25	on	0.5	Warmer	−0.4	1.1
12	7	15	18	26.5	26.5	67,544	67,544	26.1	27.2	on	0	NC	1.1	0.4
13	7	15	18	26	26.5	6	67,550	26.1	27.2	on	−0.5	Cooler	1.1	0.4
14	7	15	18	26	26	32	67,582	26.3	27.2	on	0	NC	0.9	−0.3

Table 5
Characteristics of seven basic ML classification algorithms.

Algorithm	Characteristics		Applicable to this study?
	Advantages	Disadvantages	
RF [46]	(1) Fast training speed; (2) Strong model generalization ability; (3) Not easily over-fitted; (4) The features can be ranked in order of importance.	(1) Unexplainable (black box model); (2) Not suitable for small sample size data sets.	√
AdaBoost [47]	(1) Cascading of multiple weak classifiers to improve model accuracy; (2) Can handle both continuous and discrete values; (3) Adaptively focus on training for misclassified samples to improve accuracy.	(1) Not great for imbalanced data; (2) Slow training speed.	√
ANN [48]	(1) Great for handling complex nonlinear problems; (2) With the function of associative memory; (3) Strong distribution storage and learning ability.	(1) Unexplainable (black box model); (2) Slow training speed.	√
LR [49]	(1) Fast training speed; (2) Explainable; (3) Suitable for binary classification problem.	(1) Requires linearly separable samples; (2) Sensitive to multicollinearity data.	×
SVC [50]	(1) Suitable for small sample size data sets; (2) The solution is globally optimal.	(1) If the amount of data is large, the training time is long; (2) Not great for multiple classification.	√
KNN [51]	(1) No training and no parameter estimation required; (2) Insensitive to outliers; (3) Suitable for multi-classification problems.	(1) Not great for imbalanced data; (2) Complexity of computation	√
Gaussian NB [52]	(1) Simple to Implement; (2) Suitable for small sample size data sets; (3) Can handle multiple classification tasks.	(1) Not great for imbalanced data; (2) It assumes that all the features are independent. However, conditional independence assumption does not always hold.	×

size of data from different classes is balanced. In this study, the dataset is unbalanced, so the macro F1 score, calculated by Eq. (2) [53], is introduced as an additional evaluation indicator. The macro F1 score accounts for unbalanced datasets and helps to prevent misevaluation. Therefore, both the accuracy and macro F1 score are utilized in this study to comprehensively evaluate the performance of the multiclassification models.

$$Accuracy = \frac{\sum_{i=1}^{q}(TP_i)}{\sum_{i=1}^{q}(TP_i + FP_i + FN_i)} \tag{1}$$

$$Macro\ F1 = \frac{1}{q}\sum_{i=1}^{q}\frac{2TP_i}{2TP_i + FP_i + FN_i} \tag{2}$$

In Eqs. (1) and (2), q is the number of classes in the model. In this study, the binary classification model ("Warmer", "Cooler") and the three-classification model ("Warmer", "Cooler", and "NC") of temperature setpoint behavior will be developed separately, q is 2 and 3, respectively; TP_i is the number of samples in class "I" that are predicted to be correct; FP_i is the number of samples that are predicted to be class "I", but are actually other classes; FN_i is the number of samples that are actually class "I", but are predicted to be other classes.

3. Results & discussion

3.1. Historical user interactions data exploration

3.1.1. Descriptive statistics for the study sample

Fig. 3 shows the proportion of A/C temperature setpoints in different climate zones. The A/C temperature setpoints formed a normal distribution in all three climate zones, with a peak of 26 °C. More than 85 % of the samples in the Cold zone and the HSCW zone had A/C temperature setpoints within a moderate range of 24 °C–28 °C. In the HSWW zone, 15 % of the sample had temperature setpoints below 24 °C, which was approximately-three times higher than the other two zones.

The relationship between T_{out}, T_{in}, and $T_{set'}$ was analyzed in the study of three climate zones, as depicted in Fig. 4. There were significant differences observed among the zones. The mean T_{in} for households in the Cold zone was 26.5 °C, which was lower than those in the HSCW and HSWW zones by 0.5 °C and 1.1 °C, respectively. The T_{in} in the Cold zone showed a moderate linear correlation with the $T_{set'}$ (r = 0.34, P < 0.01). In contrast, the T_{in} in the HSCW and HSWW zones demonstrated a weaker linear correlation with the $T_{set'}$ (r = 0.27, P < 0.01 and r = 0.24, P < 0.01). However, the correlation between T_{in} and T_{out} was higher in these two summer climate zones (r = 0.41, P < 0.01 and r = 0.44, P < 0.01) compared to the Cold zone. In addition, no correlation was found between T_{out} and $T_{set'}$ in any of the three climate zones (r < 0.1). In Cold zones, A/C users were able to adjust T_{in} below 28 °C by decreasing the A/C setpoint. However, in HSCW and HSWW zones, particularly the latter, the T_{in} may still be higher than the comfort temperature range, even if the A/C setpoint was low. Typically, there is a strong linear relationship between the T_{in} and the A/C setpoint. However, the findings of this study did not reflect this relationship, particularly in HSWW zones. This deviation from the expected correlation may be attributed to the varying indoor cooling load in different climate zones. Specifically, the indoor cooling load in Cold zones tends to be lower compared to HSCW and HSWW zones, allowing A/C systems in these areas to more easily meet the cooling needs of the indoor environment. Conversely, the high indoor cooling load in HSWW zones increases the likelihood that the A/C system will be unable to fully satisfy the cooling requirements.

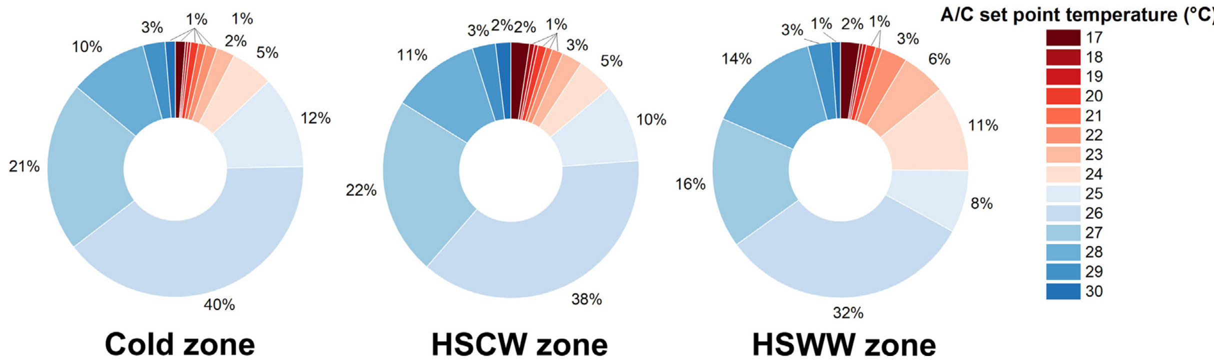


Fig. 3. Percentage of each A/C temperature setpoint.

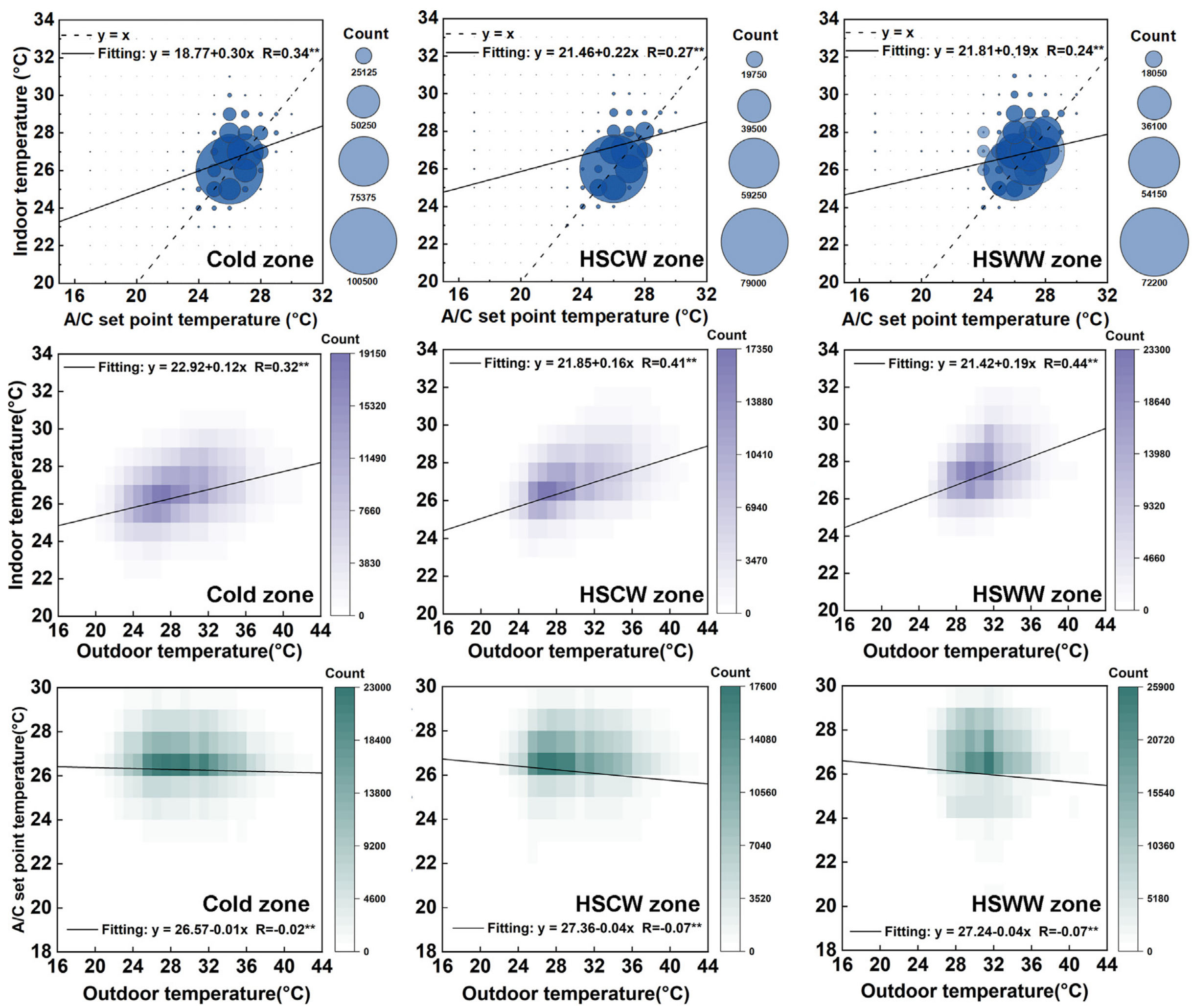


Fig. 4. Relationship between indoor temperature, outdoor temperature, and A/C temperature setpoint.

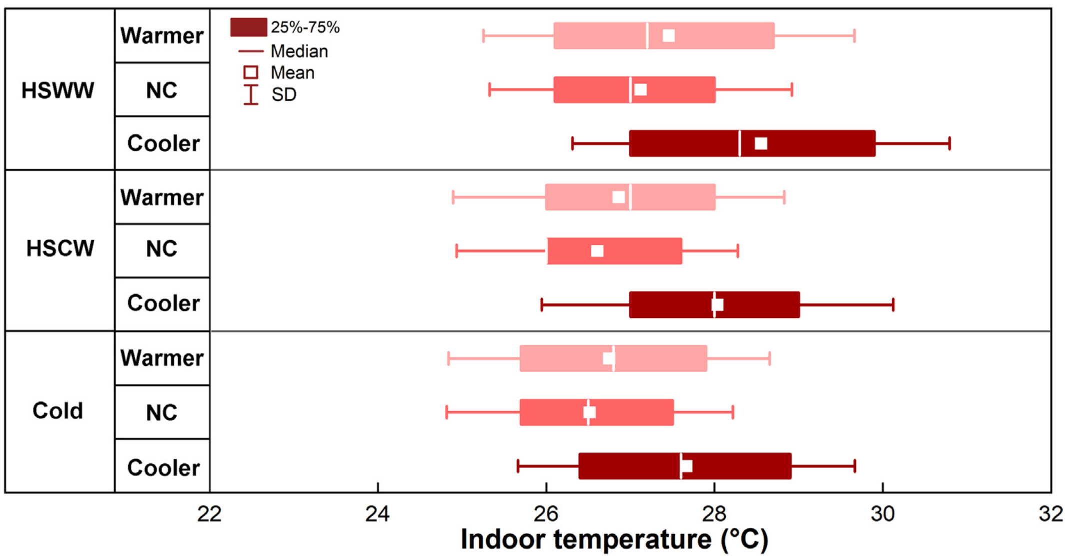
Table 6 presents the T_{out}, T_{in}, and $T_{set'}$ at various periods of the day. The period from 23:00 to 7:00 the following day is defined as the nighttime sleep time [54]; the period from 7:00 to 19:00 is considered daytime, and the period from 19:00 to 23:00 is referred to as the pre-bedtime of the night. Across all three climate zones, the T_{out} and T_{in} were lowest during the nighttime sleep time, followed by the pre-bedtime of the night and the daytime. The mean $T_{set'}$ in the cold zone remained relatively constant throughout the day, while the mean $T_{set'}$ in the other two climate zones increased at night.

3.1.2. Relation between temperature setting behavior and indoor temperature

The mean T_{in} was calculated for each climate zone when the user performed the "Warmer", "Cooler" and "NC" settings, and the results are presented in Fig. 5. The HSWW zones had signifi-

Table 6
Indoor temperature, outdoor temperature, and A/C temperature setpoint at different times of the day.

Climate zone		Times of the day		
		Daytime (7:00–19:00)	Pre-bedtime (19:00–23:00)	Nighttime sleep time (23:00–7:00)
Cold zone	T_{in} (°C)	26.9 ± 1.8	26.7 ± 1.8	26.2 ± 1.6
	T_{out} (°C)	31.7 ± 4.3	29.3 ± 4.0	27.0 ± 3.6
	$T_{set'}$ (°C)	25.8 ± 2.0	25.6 ± 2.1	25.9 ± 1.9
HSCW zone	T_{in} (°C)	27.3 ± 1.9	27.1 ± 1.9	26.3 ± 1.5
	T_{out} (°C)	32.4 ± 4.6	30.7 ± 4.2	28.6 ± 3.7
	$T_{set'}$ (°C)	25.6 ± 2.5	25.4 ± 2.7	26.0 ± 1.9
HSWW zone	T_{in} (°C)	27.8 ± 2.1	27.8 ± 2.1	26.8 ± 1.5
	T_{out} (°C)	32.8 ± 3.5	31.2 ± 3.3	29.8 ± 2.8
	$T_{set'}$ (°C)	25.8 ± 2.4	25.5 ± 2.7	26.2 ± 1.9



		Cold Cooler	Cold NC	Cold Warmer	HSCW Cooler	HSCW NC	HSCW Warmer	HSWW Cooler	HSWW NC	HSWW Warmer
Cold	Cooler							* P<0.05		
	NC	** d=0.64						** P<0.01		
	Warmer	** d=0.50	** d=0.18					N/A: Not applied		
HSCW	Cooler	** d=0.16	N/A	N/A						
	NC	N/A	N/A	N/A	** d=0.78					
	Warmer	N/A	N/A	N/A	** d=0.61	N/A				
HSWW	Cooler	** d=0.44	N/A	N/A	** d=0.28	N/A	N/A			
	NC	N/A	** d=0.41	N/A	N/A	** d=0.33	N/A	** d=0.16		
	Warmer	N/A	N/A	** d=0.40	N/A	N/A	** d=0.31	** d=0.56	** d=0.42	

Fig. 5. Indoor air temperature corresponding to the setting behaviors of users in different climate zones.

cantly higher T_{in} compared to the Cold zone, with medium effect sizes ($p < 0.01$, $d > 0.4$) for all three groups. In addition, there was a small effect size significant difference between the HSCW and HSWW zones ($p < 0.01$, $0.4 > d > 0.15$). A/C users in HSWW zones had a higher acceptable T_{in} range in summer, indicating better tolerance for high temperatures and worse tolerance for low temperatures. Conversely, A/C users in Cold zones exhibited the opposite trend. These findings are consistent with the research of Ji et al. [55], which shows that individuals accustomed to colder climates prefer lower temperatures than those accustomed to hot climates, resulting in enhanced tolerance for cold environments.

In daily experience, individuals tend to increase the A/C temperature setpoint when the T_{in} is low and decrease the setpoint when the T_{in} is high. The results of this study indicate that the mean T_{in} when individuals perform the "Cooler" setting was significantly higher than the other two settings in three different climate zones. However, the mean T_{in} when the "Warmer" setting was performed was slightly higher than the mean T_{in} when the setpoint was left unchanged ($p < 0.001$, $d > 0.15$), which was an unexpected finding. This may be connected to the daily habits of the user. The percentage of time periods in which A/C users perform different temperature setpoint settings is depicted in Fig. 6.

During the hours of 23:00–7:00, there was a decrease in the frequency with which users change the A/C temperature setpoint compared to other times of day. This may be due to the fact that the A/C cannot be artificially changed during the user's nighttime sleep period. During the hours of 19:00–23:00 in the evening, users tended to have a higher T_{in}. When compared to the sleeping hours, the buildings had a higher indoor load during the pre-bedtime period, resulting in a slightly higher mean T_{in} for "Warmer" setting compared to "NC" setting. This result indicates that the A/C temperature setpoint behavior of users is influenced by their daily rest-

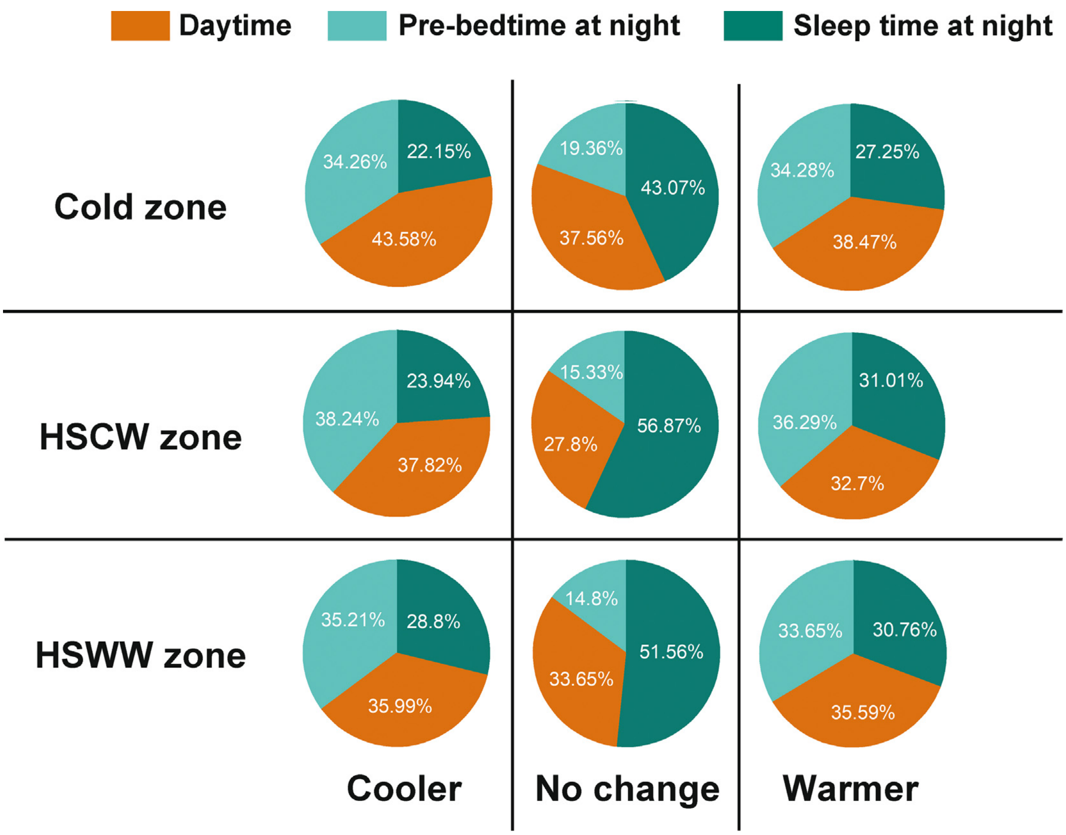
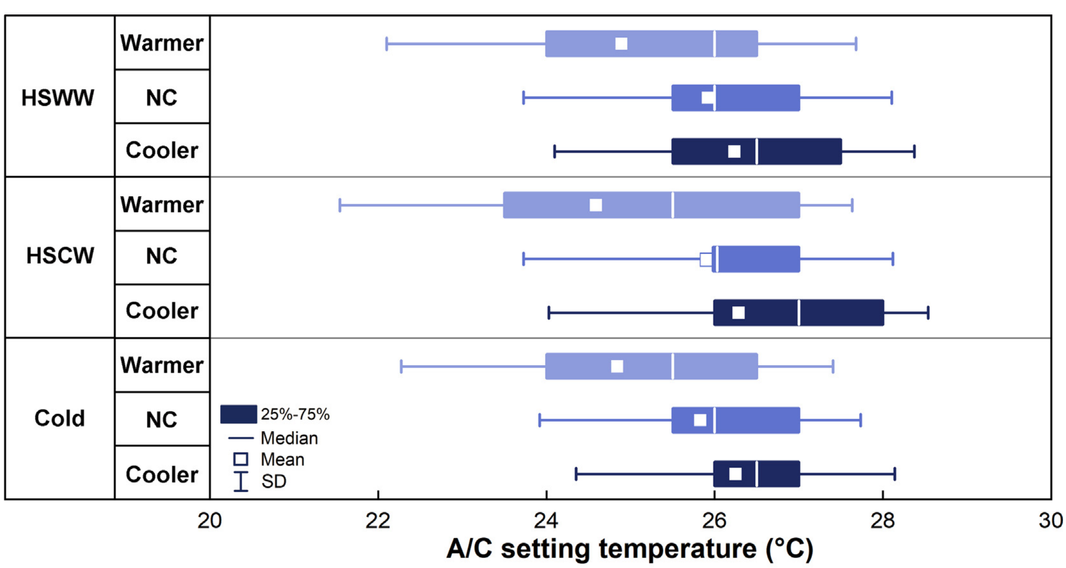


Fig. 6. Periods of the day when users perform different A/C temperature setpoint settings behaviors.



		Cold			HSCW			HSWW		
		Cooler	NC	Warmer	Cooler	NC	Warmer	Cooler	NC	Warmer
Cold	Cooler									
	NC	** d=0.22								
	Warmer	** d=0.62	** d=0.43							
HSCW	Cooler	N/A	N/A	N/A						
	NC	N/A	N/A	N/A	** d=0.16					
	Warmer	N/A	N/A	N/A	** d=0.65	** d=0.52				
HSWW	Cooler	N/A	N/A	N/A	N/A	N/A	N/A			
	NC	N/A	N/A	N/A	N/A	N/A	N/A	** d=0.16		
	Warmer	N/A	N/A	N/A	N/A	N/A	N/A	** d=0.59	**d=0.43	

* P<0.05
** P<0.01
N/A: Not applied

Fig. 7. The A/C temperature setpoint corresponding to the setting behaviors of users in different climate zones.

ing time. These findings are consistent with the results of previous studies [27,29,30].

3.1.3. Relation between temperature setting behavior and A/C temperature setpoint before settings

The mean A/C temperature setpoint was calculated for each climate zone before the "Warmer" and "Cooler" settings were performed and when the setpoint remained unchanged, as depicted in Fig. 7. For three climate zones, the mean A/C temperature setpoint before the "Cooler" setting was significantly higher than that of the "NC" setting ($P < 0.01$, $0.15 < d < 0.4$). The mean A/C temperature setpoint before the "Warmer" setting was significantly lower than that of the other two settings ($P < 0.01$, $0.4 < d < 0.8$). No significant differences in mean A/C temperature setpoint between climate zones were observed when the same A/C setting behavior was performed ($p < 0.01$, $d < 0.15$).

3.1.4. Relation between temperature setting behavior and cumulative time of A/C on

The variations in ΔT_{set} over time for all A/C samples after turning on in different climate zones were analyzed. The results, shown in Fig. 8, indicate that the temperature set behavior is correlated with the t_c. Within the first 30 min of turning on the A/C, the setpoint temperature may vary by up to 10 °C. In the Cold zone, 0.3 % of the total sample had a setpoint temperature variation greater than 5 °C, with 70 % of these occurring within the first 60 min of turning on the A/C. The setpoint temperature then gradually stabilizes as the turn-on time increases. In the HSCW climate zone, 0.5 % of the total sample had a setpoint temperature variation greater than 5 °C, with 70 % occurring within the first 70 min of turning on the A/C. Furthermore, users in the HSWW zone appeared to be able to find a preferred setpoint temperature more quickly, with only 0.1 % of the sample in this zone experiencing a setpoint temperature change greater than 5 °C, with 70 % occurring within the first 50 min of turning on the A/C. This may be due to the weak linear relationship between A/C setpoint temperature and T_{in} in the HSWW zone, leading users in this climate to be more inclined to adjust the setpoint directly to a lower temperature when turning on the A/C, thus reducing the decision time for the A/C temperature setpoint.

3.2. The performance of data-driven models

Based on the results of the previous data exploration, the date, time, T_{in}, T_{out}, t_c, $T_{set'}$, $T_{out\text{-}in}$, and $T_{set'\text{-}in}$ were utilized as features to build urban household A/C temperature setpoint behavior models in three climate zones in this section. The models are divided into binary classification model ("Warmer" and "Cooler") and three-classification model ("Warmer", "Cooler", and "NC").

3.2.1. The performance of models

The hyperparameters involved in the optimization of each algorithm are shown in Table 7. The grid search method was employed to find the optimal hyperparameters of different algorithms. The training set was trained with various combinations of hyperparameters, and the resulting models were validated on the validation

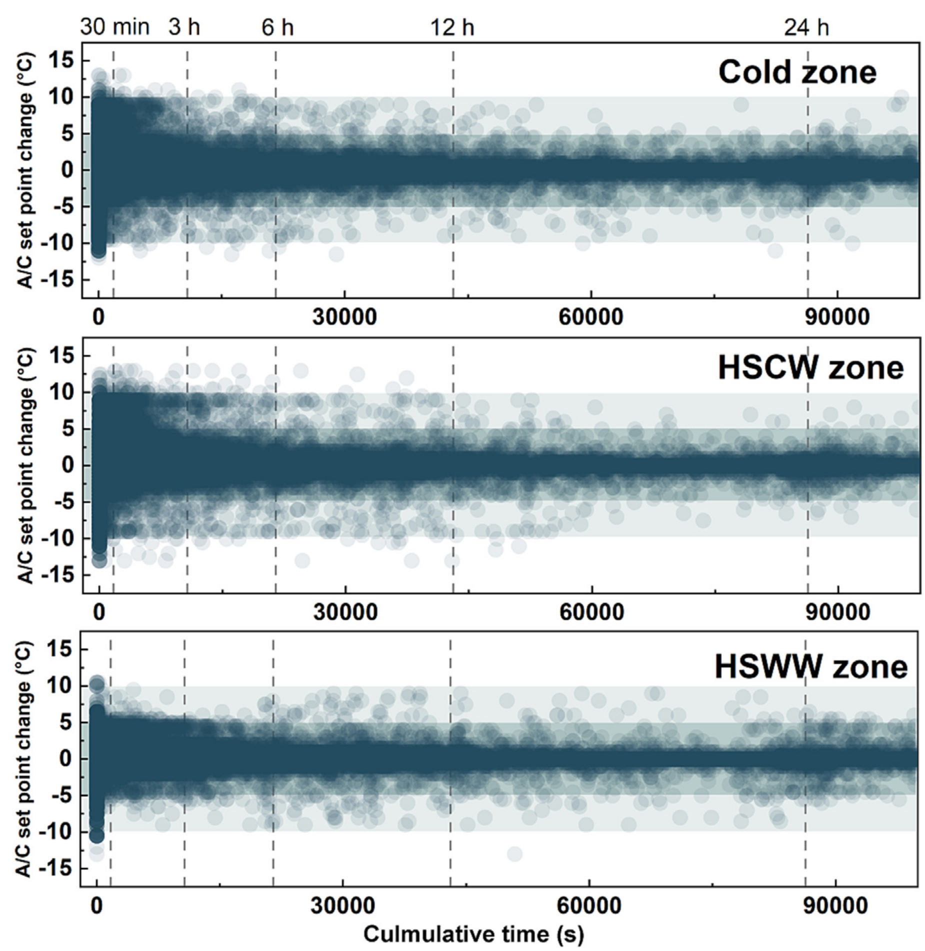


Fig. 8. A/C temperature setpoint settings with the cumulative time of the device on.

Table 7
The hyperparameters involved in the optimization.

Algorithm	Hyperparameters	Range
RF	n_estimaters (numbers of trees in the forest) max_depth (the maximum depth of the tree)	1 to 101, with an interval of 5 1 to 20, with an interval of 1
ANN	hidden layer (number of hidden layer) number of neurons in each hidden layer	2, 3 3 to 20, with an interval of 1
SVC	C Gamma Kernel	0.001, 0.01, 0.1, 0.5, 1.0 0.01, 0.1, 1, 10, 100 "poly", "rbf", "sigmoid"
AdaBoost	learning_rate n_estimaters algorithm	0.1 to 1.0, with an interval of 0.1 1 to 101, with an interval of 5 "SAMME", "SAMME.R"
K-Neighbors	algorithm	"auto", "ball_tree", "kd_tree", "brute"

set. The combination with the highest macro F1 score was selected as the best hyperparameter for the algorithm [56]. As an example, Fig. 9 shows the macro F1 scores of the validation set for different hyperparameter combinations when using both RF and a four-layer ANN to model the three-classification behavior of users in HSCW zone. The maximum macro F1 score for the validation set

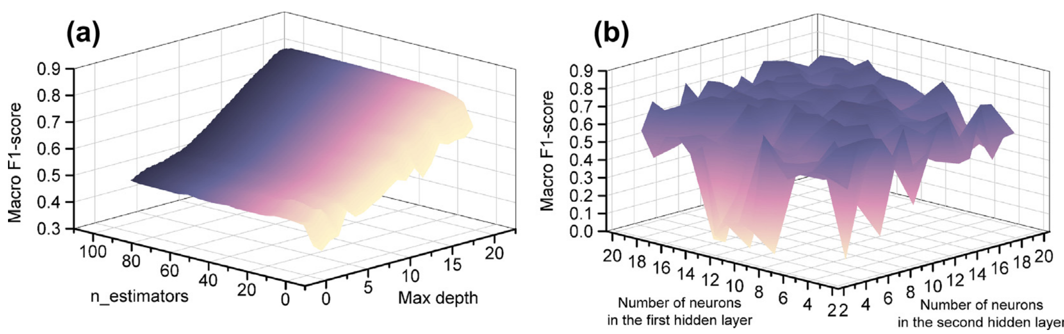


Fig. 9. Grid search for tuning parameters.

Table 8
The hyperparameters involved in the optimization.

Algorithm	Climate zone	Binary classification model	Three-classification model
RF	Cold	n_estimaters = 80 max_depth = 18	n_estimaters = 95 max_depth = 20
	HSCW	n_estimaters = 65 max_depth = 20	n_estimaters = 95 max_depth = 20
	HSWW	n_estimaters = 80 max_depth = 20	n_estimaters = 100 max_depth = 20
ANN	Cold	hidden layer sizes = (12,17,18)	hidden layer sizes = (13,16,13)
	HSCW	hidden layer sizes = (18,19,11)	hidden layer sizes = (14,12,15)
	HSWW	hidden layer sizes = (11,18,17)	hidden layer sizes = (17,11,16)
SVC	Cold	C = 1; Gamma = 0.01; Kernel = "rbf"	C = 1; Gamma = 0.01; Kernel = "rbf"
	HSCW	C = 1; Gamma = 0.01; Kernel = "rbf"	C = 1; Gamma = 0.01; Kernel = "rbf"
	HSWW	C = 1; Gamma = 0.01; Kernel= "rbf"	C = 1; Gamma = 0.01; Kernel = "rbf"

AdaBoost	Cold	learning_rate = 1;
		n_estimaters = 40;
		algorithm: SAMME.R
			learning_rate = 1;
			n_estimaters = 100;
			algorithm: SAMME.R

	HSCW	learning_rate = 1;
		n_estimaters = 100;
		algorithm: SAMME.R
			learning_rate = 1;
			n_estimaters = 100;
			algorithm: SAMME.R

	HSWW	learning_rate = 1
		n_estimaters = 100
		algorithm: SAMME.R
			learning_rate = 1;
			n_estimaters = 100;
			algorithm: SAMME.R

K-Neighbors	Cold	Algorithm: "ball_tree"	Algorithm: "brute"
	HSCW	Algorithm: "auto"	Algorithm: "auto"
	HSWW	Algorithm: "auto"	Algorithm: "auto"

using RF was obtained when the "n_estimaters" was set to 96 and the "max_depth" was set to 19. For the four-layer ANN, the maximum macro F1 score was achieved when the "Number of neurons in first hidden layer" was 9 and the "Number of neurons in second hidden layer" was 10. The remaining algorithms had their hyperparameters tuned according to the same process. The optimal hyperparameters of different algorithms are summarized in Table 8.

The results of the behavior model performance for each algorithm after tuning are presented in Table 9. Upon comparison of the models obtained from various algorithms, it was found that the A/C temperature setpoint behavior models in three climate zones using RF demonstrated the best performance. The confusion matrixes of the optimal models are depicted in Figs. 10 and 11. The accuracy for the three-classification model ranged from 0.936 to 0.961, while the macro F1 score ranged from 0.700 to 0.810. For the binary classification model, the accuracy ranged from 0.794 to 0.909 and the macro F1 score ranged from 0.793 to 0.909. Overall, the three-classification model demonstrated higher accuracy and lower macro F1 score compared to the binary classification model.

The three-classification model demonstrated great performance in the "NC" class but was less effective in the "Warmer" and "Cooler" classes. In contrast, the binary model performed well in both

Table 9
Performance of binary classification and three-classification models for A/C temperature setpoint behavior.

Climate zone	Algorithms	Binary classification models		Three-classification models	
		Accuracy	Macro F1 score	Accuracy	Macro F1 score
Cold zone	RF	***0.794***	***0.793***	***0.936***	***0.700***
	ANN	0.623	0.622	0.899	0.664
	AdaBoost	0.694	0.694	0.762	0.441
	K-Neighbors	0.641	0.642	0.684	0.392
	SVC	0.603	0.602	0.772	0.524
HSCW Zone	RF	***0.909***	***0.909***	***0.961***	***0.810***
	ANN	0.764	0.764	0.902	0.794
	AdaBoost	0.893	0.892	0.822	0.524
	K-Neighbors	0.674	0.674	0.702	0.401
	SVC	0.723	0.722	0.842	0.664
HSWW Zone	RF	***0.861***	***0.861***	***0.952***	***0.713***
	ANN	0.703	0.702	0.911	0.681
	AdaBoost	0.831	0.831	0.810	0.443
	K-Neighbors	0.632	0.632	0.711	0.364
	SVC	0.674	0.674	0.822	0.658

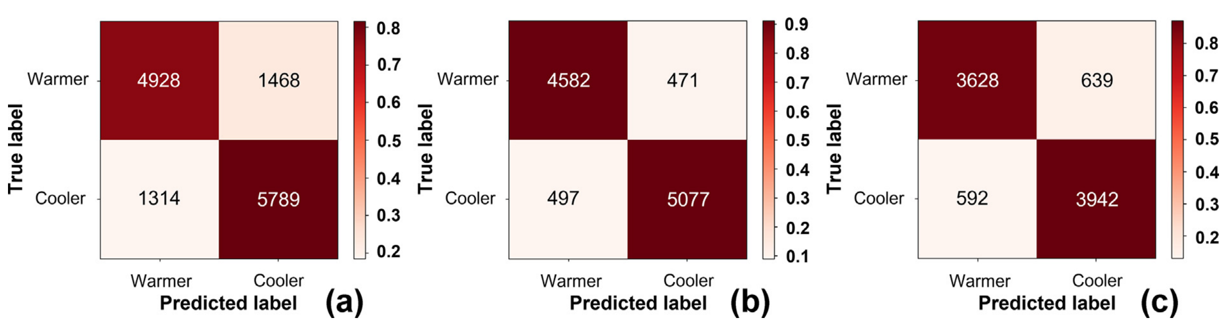


Fig. 10. Confusion matrix of output results for the best binary classification model (a) Cold zone; (b) HSCW zone; (c) HSWW zone.

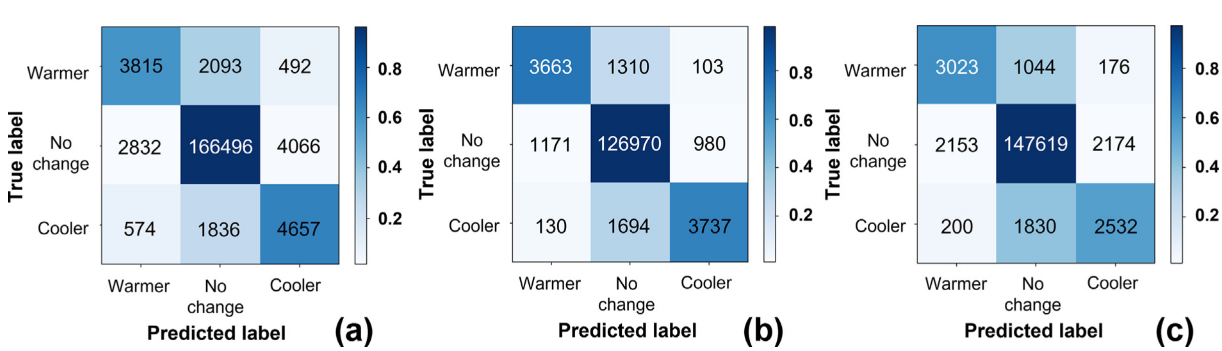


Fig. 11. Confusion matrix of output results for the best three-classification model (a) Cold zone; (b) HSCW zone; (c) HSWW zone.

classes. As a result, the binary model had a higher macro F1 score. However, the three-classification model had a higher accuracy due to the larger number of "NC" samples in the test set. Accuracy assesses the overall performance of the model but ignores the performance of the model on different classes. The macro F1 score specifically evaluates the performance of the model on individual classes. It is worth noting that in real-world use, individuals tend to keep A/C setpoints constant for longer periods of time. Therefore, relying solely on the macro F1 score without considering accuracy may not fully capture the model's practical long-term performance. Utilizing multiple evaluation indicators, including both accuracy and macro F1 score, offers a more comprehensive evaluation of the model's classification performance on different classes as well as its practical application.

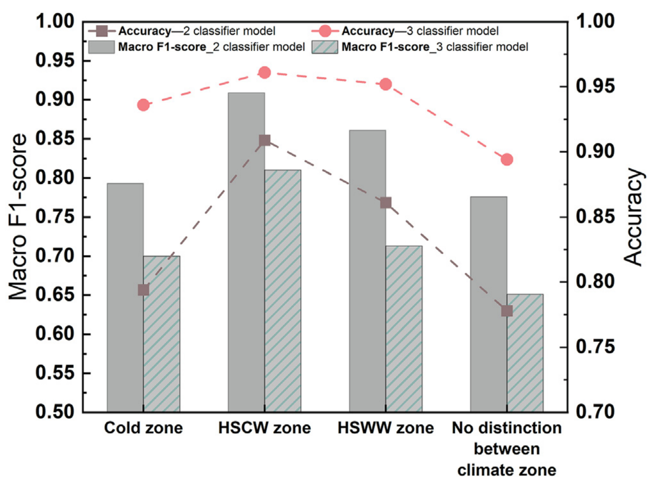


Fig. 12. Performance of the model with and without climate zone distinction.

The RF algorithm was utilized to build binary and three-classification models of A/C temperature setpoint behavior without considering distinctions between climate zones, and the performance of this model was compared to the climate-sensitive model proposed in this study. The result is depicted in Fig. 12. The climate-sensitive model for predicting A/C user behavior demonstrates better performance. This finding aligns with the results from the previous section, which indicate that users in different climate zones exhibit distinct habits for A/C control. Therefore, climate zone differences should be taken into consideration when building models of user behavior.

3.2.2. Feature importance

Fig. 13 illustrates the importance of the features for binary and three-classification models proposed in this study. The T_{in}, $T_{set'}$, and t_c are important input features. Among them, t_c is frequently overlooked, but this study found that it plays a crucial role in distinguishing "Warmer" and "Cooler" setting behaviors. As this study is limited to summer cooling conditions, the mean outdoor air temperature from June to September is relatively similar, thus the feature importance of month is lower compared to the other features.

3.3. Practical applications and future perspective

By applying the A/C temperature setpoint behavior model to A/Cs in each of the three climate zones (please find attached models), a new intelligent A/C automatic control strategy applicable to the three climate zones can be established, as depicted in Fig. 14.

The A/C applying this automatic control strategy monitors the T_{in}, T_{out}, t_c, $T_{set'}$, and the time information in real time through sensors, and adjusts the temperature setpoint accordingly. However, this model is a human behavior model at present and does not further realize the optimization of environment. In the future, it is possible to combine this model with other thermal comfort models to optimize the A/C system for both comfort and energy efficiency. As an example, the proposed model could be used to automatically lower

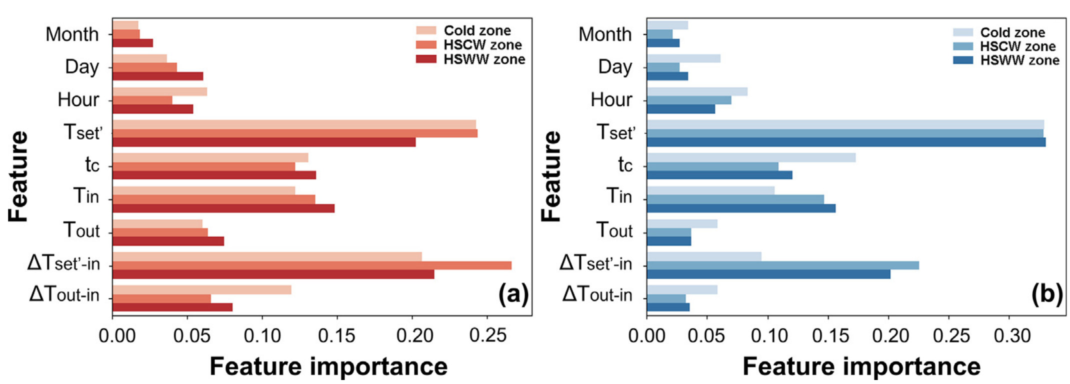


Fig. 13. Importance of model features. (a) Binary classification model (b) Three-classification model.

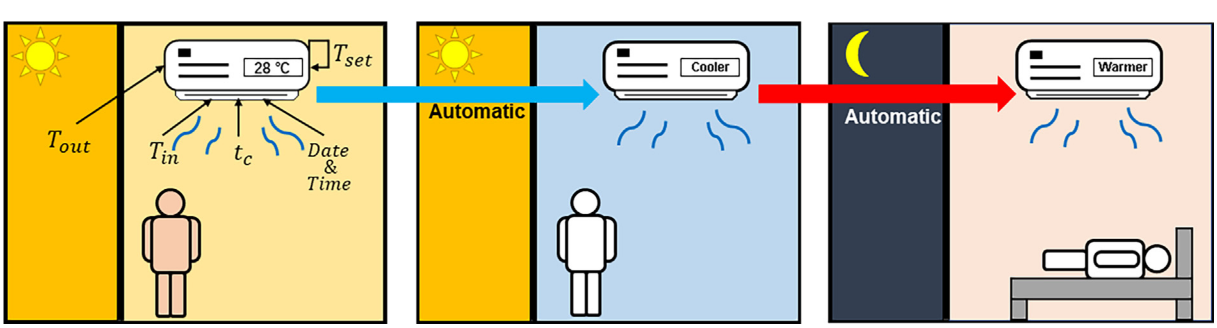


Fig. 14. An intelligent A/C control strategy based on temperature setpoint behavior model.

the daytime temperature setpoint and increase the nighttime setpoint in HSCW zone, as previous research has shown this to be a reasonable strategy for improving sleep quality and thermal comfort in HSCW zone [20,57,58]. This study did not include subjective feedback from occupants and did not collect other behavioral information (e.g., opening and closing windows). If environmental parameters, A/C usage, subjective feedback from users, and natural ventilation rate are all recorded, a more comprehensive analysis of A/C user temperature setpoint behavior can be performed. In addition, this study identified the most effective algorithm from a selection of five basic ML classifiers. Several studies have successfully modeled human behavior using reinforcement learning or long short-term memory algorithms, and it is possible that unsupervised algorithms could be used to model A/C temperature setpoint behavior in the future. As individuals often have different thermal preferences, it may be beneficial to develop an A/C temperature setpoint behavior model specific to individual users based on data collected from a single unit. Such a model could create a customized indoor environment that meets the thermal comfort needs of the user while minimizing energy consumption.

4. Conclusions

Based on a large sample size of the A/C usage dataset, this study analyzed the A/C temperature setpoint behavior of users and built the A/C temperature setpoint behavior model for the three climate zones by ML algorithms. The study found that the urban household A/C temperature setpoint behavior is related to the time, T_{in}, T_{out}, t_c, and $T_{set'}$. In the HSWW zone, 15 % of the samples had setpoints below 24 °C, which was about three times higher than the other two zones. In the HSCW and HSWW zones, T_{in} showed a weaker linear correlation with the $T_{set'}$, but a higher correlation with T_{out} compared to the Cold zone. Users in HSWW zones have a higher range of acceptable T_{in} during the summer months, while the opposite is observed in Cold zones. Notably, users in HSWW zones can find the preferred temperature setpoint more quickly after turning on A/C. Based on these analysis results, five ML algorithms were used to develop A/C temperature setpoint models for three climate zones, and the performance of the different algorithms was compared. By applying the RF algorithm and proper data preprocessing, the binary classification model can predict the "Warmer" and "Cooler" settings with 0.793–0.909 macro F1 score and 0.794–0.909 accuracy. The three-classification model can predict the "Warmer", "NC" and "Cooler" settings with 0.700–0.810 macro F1 score and 0.936–0.961 accuracy. The T_{in}, t_c, and $T_{set'}$ are important input features. The climate-sensitive model proposed in this study has better prediction performance than the model without distinguishing climate zones. Finally, the practical applications of the models proposed in this study were presented. In the future, the model proposed in this study can be combined with other thermal comfort models to optimize the A/C system for both thermal comfort and energy efficiency.

CRediT authorship contribution statement

Junmeng Lyu: Visualization, Formal analysis, Writing – original draft. **Jinbo Li:** Data curation. **Zisheng Zhao:** Data curation. **Xiongwei Miao:** Data curation. **Heng Du:** Methodology, Writing – review & editing. **Dayi Lai:** Methodology, Writing – review & editing. **Zhiwei Lian:** Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Data availability

Data will be made available on request.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the National Key R&D Program of China (2022YFC3803201).

Appendix A. Supplementary data

Supplementary data (the trained models of the urban domestic A/C temperature setpoint behaviors of users for each of the three climate zones are attached in the supplementary materials, please find the attached "ReadMe.txt" for usage details) to this article can be found online at https://doi.org/10.1016/j.enbuild.2023.112856.

References

[1] Z. Zhang, R. Gao, Y. Liu, M. Liu, Y. Wang, W. Zhu, L. Zhou, A. Li, Smart air supply terminal for floor-standing room air conditioners based on the identification of human positions, Build. Environ. 202 (2021), https://doi.org/10.1016/j.buildenv.2021.108041.
[2] D. Li, C.C. Menassa, V.R. Kamat, Personalized human comfort in indoor building environments under diverse conditioning modes, Build. Environ. 126 (2017) 304–317, https://doi.org/10.1016/j.buildenv.2017.10.004.
[3] M. Aftab, C. Chen, C.-K. Chau, T. Rahwan, Automatic HVAC control with real-time occupancy recognition and simulation-guided model predictive control in low-cost embedded system, Energy Build. 154 (2017) 141–156, https://doi.org/10.1016/j.enbuild.2017.07.077.
[4] A.P. Batista, M.E.A. Freitas, F.G. Jota, Evaluation and improvement of the energy performance of a building's equipment and subsystems through continuous monitoring, Energy Build. 75 (2014) 368–381, https://doi.org/10.1016/j.enbuild.2014.02.029.
[5] T. Parkinson, R. de Dear, G. Brager, Nudging the adaptive thermal comfort model, Energy Build. 206 (2020), https://doi.org/10.1016/j.enbuild.2019.109559.
[6] T. Hoyt, E. Arens, H. Zhang, Extending air temperature setpoints: Simulated energy savings and design considerations for new and retrofit buildings, Build. Environ. 88 (2015) 89–96, https://doi.org/10.1016/j.buildenv.2014.09.010.
[7] J. Lyu, X. Feng, Y. Cheng, C. Liao, Experimental and numerical analysis of air temperature uniformity in occupied zone under stratum ventilation for heating mode, J. Build. Eng. 43 (2021), https://doi.org/10.1016/j.jobe.2021.103016.
[8] Z. Wang, R. de Dear, M. Luo, B. Lin, Y. He, A. Ghahramani, Y. Zhu, Individual difference in thermal comfort: A literature review, Build. Environ. 138 (2018) 181–193, https://doi.org/10.1016/j.buildenv.2018.04.040.
[9] M. Luo, K. Jiang, J. Wang, W. Feng, L. Ma, X. Shi, X. Zhou, Data-driven thermal preference prediction model with embedded air-conditioning sensors and historical usage behaviors, Build. Environ. 220 (2022), https://doi.org/10.1016/j.buildenv.2022.109269.
[10] X. Zhuang, C. Wu, Saving energy when using air conditioners in offices—Behavioral pattern and design indications, Energy Build. 76 (2014) 661–668, https://doi.org/10.1016/j.enbuild.2013.11.042.
[11] T. Leephakpreeda, Implementation of adaptive indoor comfort temperature control via embedded system for air-conditioning unit, J. Mech. Sci. Technol. 26 (2012) 259–268, https://doi.org/10.1007/s12206-011-0920-z.
[12] J. Wu, X. Li, Y. Lin, Y. Yan, J. Tu, A PMV-based HVAC control strategy for office rooms subjected to solar radiation, Build. Environ. 177 (2020), https://doi.org/10.1016/j.buildenv.2020.106863.
[13] L. Zampetti, M. Arnesano, G.M. Revel, Experimental testing of a system for the energy-efficient sub-zonal heating management in indoor environments based on PMV, Energy Build. 166 (2018) 229–238, https://doi.org/10.1016/j.enbuild.2018.02.019.
[14] G.Y. Yun, H.J. Kong, J.T. Kim, The effect of seasons and prevailing environments on adaptive comfort temperatures in open plan offices, Indoor Built Environ. 21 (2012) 41–47, https://doi.org/10.1177/1420326X11419929.
[15] G.Y. Yun, J.H. Lee, K. Steemers, Extending the applicability of the adaptive comfort model to the control of air-conditioning systems, Build. Environ. 105 (2016) 13–23, https://doi.org/10.1016/j.buildenv.2016.05.027.
[16] H. Du, Z. Lian, D. Lai, L. Duanmu, Y. Zhai, B. Cao, Y. Zhang, X. Zhou, Z. Wang, X. Zhang, Z. Hou, Evaluation of the accuracy of PMV and its several revised models using the Chinese thermal comfort Database, Energy Build. 271 (2022), https://doi.org/10.1016/j.enbuild.2022.112334.
[17] W. Liu, Z. Lian, Q. Deng, Y. Liu, Evaluation of calculation methods of mean skin temperature for use in thermal comfort study, Build. Environ. 46 (2011) 478–488, https://doi.org/10.1016/j.buildenv.2010.08.011.

[18] L. Lan, Z. Lian, W. Liu, Y. Liu, Investigation of gender difference in thermal comfort for Chinese people, Eur. J. Appl. Physiol. 102 (2008) 471–480, https://doi.org/10.1007/s00421-007-0609-2.

[19] X. Tian, Z. Fang, W. Liu, Decreased humidity improves cognitive performance at extreme high indoor temperature, Indoor Air 31 (2021) 608–627. https://search.ebscohost.com/login.aspx?direct=true&db=eih&AN=149927517&site=ehost-live&scope=site.

[20] T. Cao, Z. Lian, J. Zhu, X. Xu, H. Du, Q. Zhao, Parametric study on the sleep thermal environment, Build. Simul. 15 (2022) 885–898, https://doi.org/10.1007/s12273-021-0840-5.

[21] T. el Aarbaoui, J. Méline, R. Brondeel, B. Chaix, Short-term association between personal exposure to noise and heart rate variability: The RECORD MultiSensor Study, Environ. Pollut. 231 (2017) 703–711, https://doi.org/10.1016/j.envpol.2017.08.031.

[22] X. Zhou, Y. Liu, M. Luo, S. Zheng, R. Yang, X. Zhang, Overall and thermal comfort under different temperature, noise, and vibration exposures, Indoor Air. 32 (2022) e12915, https://doi.org/10.1111/ina.12915.

[23] A. Čulić, S. Nižetić, P. Šolić, T. Perković, V. Čongradac, Smart monitoring technologies for personal thermal comfort: A review, J. Clean. Prod. 312 (2021), https://doi.org/10.1016/j.jclepro.2021.127685.

[24] D. Xia, S. Lou, Y. Huang, Y. Zhao, D.H.W. Li, X. Zhou, A study on occupant behaviour related to air-conditioning usage in residential buildings, Energy Build. 203 (2019), https://doi.org/10.1016/j.enbuild.2019.109446.

[25] J. Yao, Modelling and simulating occupant behaviour on air conditioning in residential buildings, Energy Build. 175 (2018) 1–10, https://doi.org/10.1016/j.enbuild.2018.07.013.

[26] Y. Song, Y. Sun, S. Luo, Z. Tian, J. Hou, J. Kim, T. Parkinson, R. de Dear, Residential adaptive comfort in a humid continental climate – Tianjin China, Energy Build. 170 (2018) 115–121, https://doi.org/10.1016/j.enbuild.2018.03.083.

[27] W. Kempton, D. Feuermann, A.E. McGarity, "I always turn it on super": user decisions about when and how to operate room air conditioners, Energy Build. 18 (1992) 177–191, https://doi.org/10.1016/0378-7788(92)90012-6.

[28] Y. Zhang, H. Chen, Q. Meng, Thermal comfort in buildings with split air-conditioners in hot-humid area of China, Build.Environ. 64 (2013) 213–224, https://doi.org/10.1016/j.buildenv.2012.09.009.

[29] S.A. Zaki, A. Hagishima, R. Fukami, N. Fadhilah, Development of a model for generating air-conditioner operation schedules in Malaysia, Build. Environ. 122 (2017) 354–362, https://doi.org/10.1016/j.buildenv.2017.06.023.

[30] A. Hagishima, R. Fukami, sheikh ahmad zaki shaikh salim, N. Ikegaya, N. Hanip, J. Tanimoto, Cross-cultural comparison of occupants' behaviour towards cooling of residences between Japan and Malaysia, in: CLIMA 2016 - Proceedings of the 12th REHVA World Congress, 2016.

[31] X. Ren, D. Yan, C. Wang, Air-conditioning usage conditional probability model for residential buildings, Build. Environ. 81 (2014) 172–182, https://doi.org/10.1016/j.buildenv.2014.06.022.

[32] M. Schweiker, M. Shukuya, Comparison of theoretical and statistical models of air-conditioning-unit usage behaviour in a residential setting under Japanese climatic conditions, Build. Environ. 44 (2009) 2137–2149, https://doi.org/10.1016/j.buildenv.2009.03.004.

[33] J. Tanimoto, A. Hagishima, State transition probability for the Markov Model dealing with on/off cooling schedule in dwellings, Energy Build. 37 (2005) 181–187, https://doi.org/10.1016/j.enbuild.2004.02.002.

[34] Z. Lin, S. Deng, A questionnaire survey on sleeping thermal environment and bedroom air conditioning in high-rise residences in Hong Kong, Energy Build. 38 (2006) 1302–1307, https://doi.org/10.1016/j.enbuild.2006.04.004.

[35] J. Wu, C. Liu, H. Li, D. Ouyang, J. Cheng, Y. Wang, S. You, Residential air-conditioner usage in China and efficiency standardization, Energy. 119 (2017) 1036–1046, https://doi.org/10.1016/j.energy.2016.11.038.

[36] J. Kim, R. de Dear, T. Parkinson, C. Candido, Air conditioning usage and environmental control behaviour in residential contexts, in: 2017. https://doi.org/10.26868/25222708.2017.389.

[37] L. Yan, M. Liu, K. Xue, Z. Zhang, A study on temperature-setting behavior for room air conditioners based on big data, J. Build. Eng. 30 (2020), https://doi.org/10.1016/j.jobe.2020.101197.

[38] Department of Construction of the PRC, Thermal Design Code for Civil Buildings, GB/T 50176, (2016).

[39] T. Chaudhuri, Y.C. Soh, H. Li, L. Xie, A feedforward neural network based indoor-climate control framework for thermal comfort and energy saving in buildings, Appl. Energy 248 (2019) 44–53, https://doi.org/10.1016/j.apenergy.2019.04.065.

[40] Z. Wang, H. Yu, M. Luo, Z. Wang, H. Zhang, Y. Jiao, Predicting older people's thermal sensation in building environment through a machine learning approach: Modelling, interpretation, and application, Build. Environ. 161 (2019), https://doi.org/10.1016/j.buildenv.2019.106231.

[41] Y. Chen, L.K. Norford, H.W. Samuelson, A. Malkawi, Optimal control of HVAC and window systems for natural ventilation through reinforcement learning, Energy Build. 169 (2018) 195–205, https://doi.org/10.1016/j.enbuild.2018.03.051.

[42] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, Synthetic minority over-sampling technique, J. Artif. Intell. Res. 16 (2002) 321–357.

[43] M. Luo, J. Xie, Y. Yan, Z. Ke, P. Yu, Z. Wang, J. Zhang, Comparing machine learning algorithms in predicting thermal sensation using ASHRAE Comfort Database II, Energy Build. 210 (2020), https://doi.org/10.1016/j.enbuild.2020.109776.

[44] W. Zhang, Y. Wu, J.K. Calautit, A review on occupancy prediction through machine learning for enhancing energy efficiency, air quality and thermal comfort in the built environment, Renew. Sustain. Energy Rev. 167 (2022), https://doi.org/10.1016/j.rser.2022.112704.

[45] Z. Qavidel Fard, Z.S. Zomorodian, S.S. Korsavi, Application of machine learning in thermal comfort studies: A review of methods, performance and challenges, Energy Build. 256 (2022) 111771. https://doi.org/https://doi.org/10.1016/j.enbuild.2021.111771.

[46] M. Belgiu, L. Drăguţ, Random forest in remote sensing: A review of applications and future directions, ISPRS J. Photogramm. Remote Sens. 114 (2016) 24–31, https://doi.org/10.1016/j.isprsjprs.2016.01.011.

[47] R.E. Schapire, Explaining AdaBoost, in: B. Schölkopf, Z. Luo, V. Vovk (Eds.), Empirical iNfereNce: Festschrift iN HoNor of Vladimir N. VapNik, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013, pp. 37–52.

[48] A.K. Jain, J. Mao, K.M. Mohiuddin, Artificial neural networks: a tutorial, Computer (Long Beach Calif). 29 (1996) 31–44, https://doi.org/10.1109/2.485891.

[49] R.E. Wright, Logistic regression, in: Reading and Understanding Multivariate Statistics, American Psychological Association, Washington, DC, US, 1995, pp. 217–244.

[50] W.S. Noble, What is a support vector machine?, Nat Biotechnol 24 (2006) 1565–1567, https://doi.org/10.1038/nbt1206-1565.

[51] J. Laaksonen, E. Oja, Classification with learning k-nearest neighbors, in: Proceedings of International Conference on Neural Networks (ICNN'96), 1996: pp. 1480–1483 vol.3. https://doi.org/10.1109/ICNN.1996.549118.

[52] A.H. Jahromi, M. Taheri, A non-parametric mixture of Gaussian naive Bayes classifiers based on local independent features, in: 2017 Artificial Intelligence and Signal Processing Conference (AISP), 2017, pp. 209–212, https://doi.org/10.1109/AISP.2017.8324083.

[53] K. Takahashi, K. Yamamoto, A. Kuchiba, T. Koyama, Confidence interval for micro-averaged F1 and macro-averaged F1 scores, Appl. Intell. 52 (2022) 4961–4972, https://doi.org/10.1007/s10489-021-02635-5.

[54] X. Xu, J. Zhu, C. Chen, X. Zhang, Z. Lian, Z. Hou, Application potential of skin temperature for sleep-wake classification, Energy Build. 266 (2022), https://doi.org/10.1016/j.enbuild.2022.112137.

[55] W. Ji, R. de Dear, J. Kim, Y. Zhu, B. Cao, S. Liu, Study on the influence of climatic thermal exposure environment changed from cold to hot on human thermal preference, Build Environ. 207 (2022), https://doi.org/10.1016/j.buildenv.2021.108430.

[56] J. Zhang, T. Zhao, X. Zhou, J. Wang, X. Zhang, C. Qin, M. Luo, Room zonal location and activity intensity recognition model for residential occupant using passive-infrared sensors and machine learning, Build. Simul. 15 (2022) 1133–1144, https://doi.org/10.1007/s12273-021-0870-z.

[57] T. Cao, Z. Lian, S. Ma, J. Bao, Thermal comfort and sleep quality under temperature, relative humidity and illuminance in sleep environment, J. Build. Eng. 43 (2021), https://doi.org/10.1016/j.jobe.2021.102575.

[58] Y. Sun, H. Zhang, Y. Yan, L. Lan, T. Cao, Z. Lian, X. Fan, P. Wargocki, J. Zhu, X. Xu, Comparison of wrist skin temperature with mean skin temperature calculated with Hardy and Dubois's seven-point method while sleeping, Energy Build. 259 (2022), https://doi.org/10.1016/j.enbuild.2022.111894.