

RESEARCH

Open Access



# Evaluation of thermal comfort in air-conditioned rooms based on structure/control-related parameters and data-mining method

Shunan Zhao<sup>1</sup>, Lin He<sup>1</sup>, Xin Wu<sup>2</sup>, Guowen Xu<sup>2</sup>, Junlong Xie<sup>2</sup> and Shanshan Cai<sup>2\*</sup>

## Abstract

Evaluating the thermal environment and thermal comfort in an air-conditioned room is an essential for estimating the performance of air-conditioning systems. However, multiple component structures and control-related parameters often lead to a long test cycle and large number of tests, significantly affecting the testing efficiency and speed. To address these problems, in this study, a data-mining method was proposed to predict and evaluate the thermal environment of an air-conditioned room. Owing to the limited amount of experimental data, the sample data were expanded by the simulation data of a collaborative platform between the air-conditioning system and air-conditioned room. Data-mining models, including the support vector regression (SVR), backpropagation (BP), and multiple linear regression (MLR) models, were developed and achieved good accuracy in evaluating the thermal environment by considering air-conditioning systems with various structures and control parameters. In the multiple-input single-output evaluation method, the prediction accuracy of the SVR model was higher than those of the BP and MLR models with respect to the vertical air temperature difference, temperature uniformity, temperature drop rate, and draft rate, while the result was the opposite in terms of the predicted mean vote indices. In the multiple-input multiple-output evaluation method, there was a decline in prediction accuracy and an increase in efficiency prediction compared with multiple-input single-output evaluation.

**Keywords** Data mining, Thermal environment evaluation, SVR model, BP model, MLR model

## 1 Introduction

Evaluating the thermal environment and thermal comfort in an air-conditioned room is vital for estimating the performance of air-conditioning systems. At present, the thermal environment testing of a room with an air conditioner follows GB/T 33658 [1] in China. However,

multiple structure- and control-related parameters affect the performance of air-conditioning systems as well as the thermal environments of air-conditioned rooms. To improve the testing efficiency and speed, it is important to identify an efficient method for predicting and evaluating the indoor thermal environments of air-conditioned rooms.

### 1.1 Classical methods for thermal comfort evaluation

Thermal comfort evaluation indexes are used to quantify thermal sensation of human beings in an indoor environment. The indicators are closely related to the variations and distributions of the room temperature, humidity, air speed, and some human related factors, such as the

\*Correspondence:

Shanshan Cai  
shanshc@hust.edu.cn

<sup>1</sup> State Key Laboratory of Air-conditioning Equipment and System Energy Conservation & GREE Electric Appliances, Inc. of Zhuhai, Zhuhai, China

<sup>2</sup> School of Energy and Power Engineering, Huazhong University of Science and Technology, Wuhan, China

activity level and clothing insulation. Commonly used indexes include thermal sensation vote (TSV), which reflects a person's description of how hot or cold the surrounding environment is; thermal comfort vote (TCV), which indicates a state of consciousness of a person to express satisfaction with the thermal environment; predicted mean vote (PMV), which reflects the feelings of the vast majority of people in the same environment; draft rate (or named as draught rate, DR), which describes the percentage of dissatisfaction of the measured population with the sense of the blowing air; and standard effective temperature, which reflects the heat exchange pattern between the human body and the external environment on the basis of simulating the physiological regulation of the human body. In addition, thermal comfort related indicators, such as the cooling rate (or heating rate), temperature deviation, temperature uniformity, temperature fluctuation, vertical air temperature difference, percent dissatisfied (PD), and predicted percentage of dissatisfied (PPD) are normally considered in the thermal comfort test of air-conditioning devices and standards [2, 3].

Indoor thermal comfort and its related indicators can be evaluated by conventional methods, such as onsite measurements or field surveys, laboratory tests with human beings or with a thermal manikin, and simulations. Through 1632 questionnaires containing thermal comfort indicators such as TSV and TCV, Yin et al. [4] concluded that TSV and TCV are highly correlated in high-density blocks in Harbin; the correlation between transitional seasons and winter is stronger than that in summer. Using manikins, Gao et al. [5] experimentally demonstrated the clothing insulation of eight garment combinations at different air speeds and wind directions; it is concluded that the airflow may lead to a decrease on the clothing insulation and lowered the values of the PMV and new standard effective temperature. Yang et al. [6] studied the effects of air velocity, air temperature, and average radiation temperature on PMV through 4 months of experiments in a 2000 m<sup>3</sup> test building and concluded that it is more energy efficient to achieve the same thermal comfort by adjusting the air speed than by adjusting the temperature. Wu et al. [7] studied the vertical temperature differences, PMV, and DR of different hybrid systems in the occupied zone and reported that the supply air temperature had a slight impact on the PMV-PPD for the studied hybrid systems. Tawackolian et al. [8] experimentally studied the effect of wind speed and intermittent ventilation cycles on DR and proposed a method to reduce DR in a neutral environment. CFD tools and PMV models are often used in the prediction of thermal indoor comfort. Embaye et al. [9] simulated radiators at different flow rates and evaluated the indoor thermal environment by analyzing the indexes

DR, percentage experience draught (PED), and PD which are used to evaluate indoor draught. Using a PMV model considering radiation, San et al. [10] analyzed the temperature, velocity, and PMV fields by CFD and performed a comprehensive assessment of the thermal comfort of the tested cooling system. Furthermore, energy consumption is often considered when using CFD methods to study thermal comfort. Aryal et al. [11] analyzed the PMV and the cooling load of supplementary air effected by partitions in an air-conditioned building by CFD and concluded that the installation of partitions resulted in a decrease in thermal comfort and an increase in energy consumption and proposed corresponding improvement measures. Shan et al. [12] coupled CFD with the building energy model to analyze the PMV field and energy consumption of cooling; it is found that the coupled method is more accurate than the separated calculations by comparing with experimental data.

## 1.2 Data mining methods applied for thermal comfort evaluation

With the development of interdisciplinary subjects, many studies on thermal comfort evaluation have been conducted by applying data-mining methods. These methods have a unique advantage in the processing of nonlinear and complex data [13, 14]. At current stage, they are widely applied to predict thermal sensations based on indoor thermal environment parameters or used for system optimization control based on thermal comfort indicators. Among the data-mining methods, the multiple linear regression (MLR) model has a simple structure and has obtained satisfactory results in the prediction of PMV [15]. Based on the PMV predictive model of MLR, Hang et al. [16] proposed an enhanced predictive control practical system, which is able to optimize thermal comfort conditions according to the season. Comparing to some other methods, MLR has the benefit of reducing computing time and improving prediction accuracy. Brik et al. [17] used MLR algorithm to build a prediction model for thermal comfort indicator PMV of occupants throughout entire year and conducted that MLR has faster calculations and more accurate results compared to polynomial regression, random forest regression, and multi-layer perceptron. Broday et al. [18] used the conventional method and MLR method to calculate the PMV and compare the two methods with TSV from experiments to obtain that the PMV calculated by MLR is more accurate. In addition to the MLR model, a backpropagation (BP) neural network has been built with dry bulb temperature, wind speed, relative humidity, and average radiation temperature as the input variables and PMV indices as the output variables. The prediction results are consistent with the experimental data of a human thermal comfort

study. Liu et al. [19] built a neural network evaluation model which connected personal thermal comfort with air-conditioning control based on BP method. Yao et al. [20] established a BP neural network model with air temperature, relative humidity, mean radiant temperature, air velocity, metabolic rate, and clothing index as inputs and PMV as output; compared to the experiment data, the prediction accuracy is within 95%, which indicates high reliability and accuracy. Wan et al. [21] integrated BP neural network algorithm in an optimization model for ventilation system; by optimizing thermal comfort indexes of PMV and DR, the system control strategy is improved to enhance the ventilation performance. The support vector regression (SVR) model can also be applied to the prediction of PMV indices, and the results indicate good agreement between the SVR predicted values and those obtained from the conventional thermal comfort evaluation. Chaudhuri et al. [22] proposed a predicted thermal state model that used the peripheral skin temperature and its gradient characteristics of a single body position to assess thermal state and concluded that support vector regression has higher prediction accuracy than extreme learning machines. Viani et al. [23] utilized SVR algorithm to study the indoor thermal comfort of an intelligent building to predict indoor temperature and adjust PMV value in advance; the input parameters are outdoor air temperature and outdoor air humidity, and the output is indoor air temperature; the results indicate that the error after 48 h is within 1 °C, which has shown good performance. Based on the collected environmental parameters, Qin et al. [24] applied a SVR model to realize online optimization of control strategies in air-conditioning unit to maintain the thermal comfort of the indoor environment. In addition, there are many other data mining methods used to study thermal comfort, and the accuracy of different methods may vary. Luo et al. [25] compared the values of PMV calculated by nine machine learning methods and found that the PMV calculated by the algorithms dealing data with high dimensions such as random forest, artificial neural network, and gradient boosting machine behave better accuracy than other studied machine learning methods. Mustafaraj et al. [26] used a linear parametric autoregressive model with external inputs (ARX) and a nonlinear autoregressive model with external inputs (NNARX) to predict dry-bulb temperatures and relative humidity in offices using climate data from interior and exterior regions of the building; it is concluded that NNARX had better accuracy than ARX by comparing the prediction results with the measured data. Wu et al. [27] proposed an intelligent ensemble machine learning method for predicting thermal sensation and concluded that the method is more accurate by comparing the calculated thermal comfort indexes such

as PMV and SET with artificial neural network and SVR. Data mining methods are also used to simplify the calculation of thermal comfort indexes. Buratti et al. [28] used artificial neural networks to develop a new algorithm for PMV calculation that only uses the monitored outdoor and indoor air temperature and relative humidity as input parameters for training and concluded that the result derived from the new algorithm is closer to the PMV value of the questionnaire. By using wet-bulb temperature and global temperature instead of relative humidity and mean radiation temperature in the original model, Athajariyakul et al. [29] proposed a neural-PMV model based on a feedforward neural network to calculate real-time PMV. Castilla et al. [30] used artificial neural network and 7-order polynomial to calculate PMV and concluded that the artificial neural network has improved accuracy comparing with the original PMV model.

### 1.3 Can the thermal comfort indices directly correlate with structure/control-related parameters?

Although data-mining methods have been widely applied in the prediction of PMV indices and system control strategies, they have not been used to investigate the mapping relationship between both structure and control related parameters of an air-conditioning unit with indoor thermal environment evaluation indices of an air-conditioned room. The structures of the main components in the air-conditioning devices may highly affect the supply air status and further affect the indoor thermal comfort. To solve this problem, in this study, data-mining methods were used to explore the possibility of correlating structural parameters (such as evaporator tube diameter, fin spacing, and tube spacing) and control parameters (such as set temperature and air supply mode) of the air-conditioning system with the thermal environment evaluation indices in the room, as shown in Fig. 1. This approach provides an alternate method for simplifying the procedures of the thermal comfort test of air conditioners and improving the overall test efficiency.

## 2 Data acquisition

The dataset is important for data mining. The data sources in this study were a thermal environmental comfort test and a collaborative simulation platform between the air-conditioning system and air-conditioned room. Both experimental data and simulation data compose the data set used for data-mining techniques. The experimental data were derived from the thermal comfort test and four types of air conditionings systems with different structure parameters were tested under multiple conditions. The simulation data were obtained based on a collaborative simulation method which help expand the amount of sample data.

### 2.1 Thermal comfort test

The thermal comfort test was conducted in a thermal environment testing laboratory of an air conditioner, as illustrated in Fig. 2. The laboratory included an inner room, which was the main thermal test laboratory, a controlled outer environment, and a control system. The indoor unit of the air conditioner was placed in the

inner room, and the outdoor unit was placed in the outer chamber. To simulate various outdoor conditions, environmental control units were also installed in the outer chamber. Based on the requirements of the standard, 147 measuring points were arranged in the test room, as depicted in Fig. 2. The working conditions of the outer chamber are presented in Table 1. Under the refrigeration

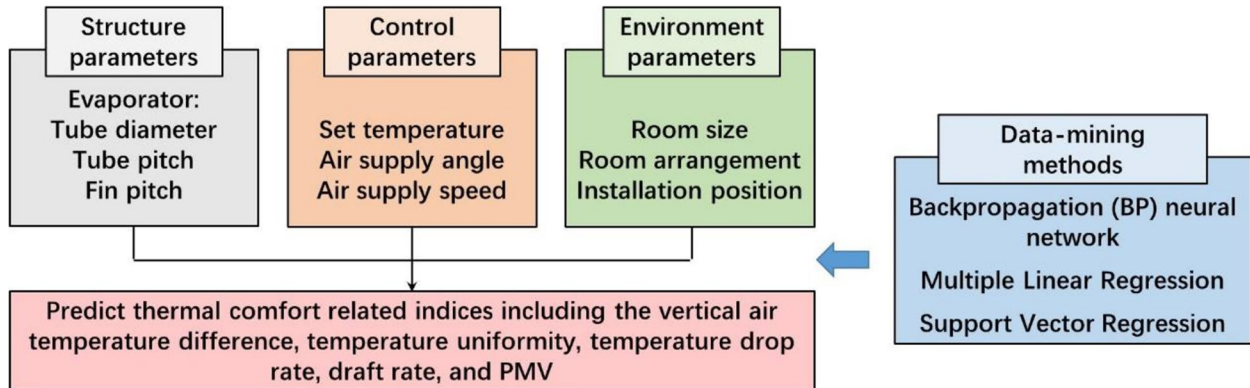


Fig. 1 Schematic of the objective in this study

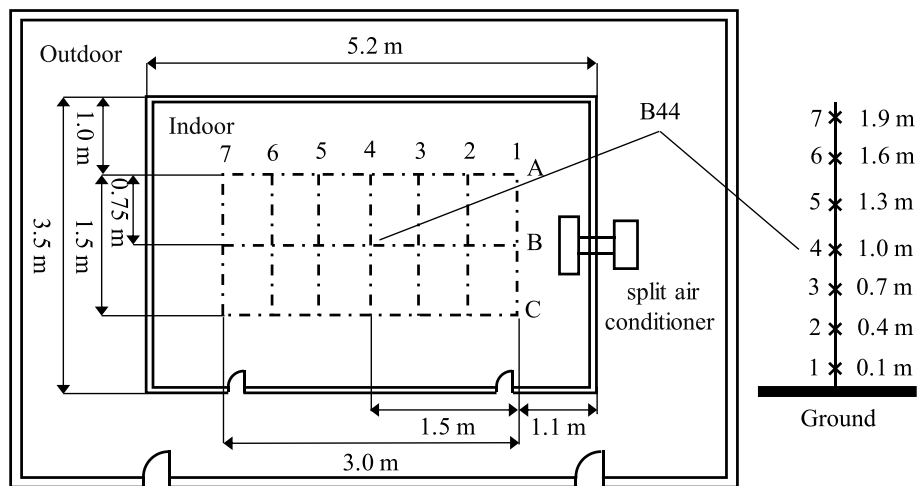


Fig. 2 Schematic of air conditioner testing laboratory

Table 1 Working conditions of outer chamber

Item	Environmental parameters	Setting values	
		Working condition of refrigeration	Working condition of heating
1	External chamber dry bulb temperature /°C	35 ± 0.5	7 ± 0.5
2	Outer chamber wet bulb temperature /°C	24 ± 0.5	6 ± 0.5
3	Heat load /W	70% of the rated capacity	

condition, the dry bulb temperature of the outdoor air is  $35 \pm 0.5$  °C, the wet bulb temperature of the outdoor air is  $24 \pm 0.5$  °C, and the setting temperature of the air conditioner is 23, 24, 25, 26, and 27 °C, respectively. Under the heating condition, the corresponding environmental parameters are  $7 \pm 0.5$  °C,  $6 \pm 0.5$  °C, and 20 °C, respectively. However, in this study, only the data derived in the cooling condition, including the temperature and velocity fields, were added in the sample data set used for data mining. The thermal load of the laboratory is 70% of its rated capacity. A total of 147 temperature measurement points was arranged in the room space. During the testing periods, the ambient temperature and humidity of the outer chamber were maintained at the required conditions. Data collection was initiated after the tested air conditioner began operating, and the interval of data collection was 1 min with a total collection time of approximately 3 h.

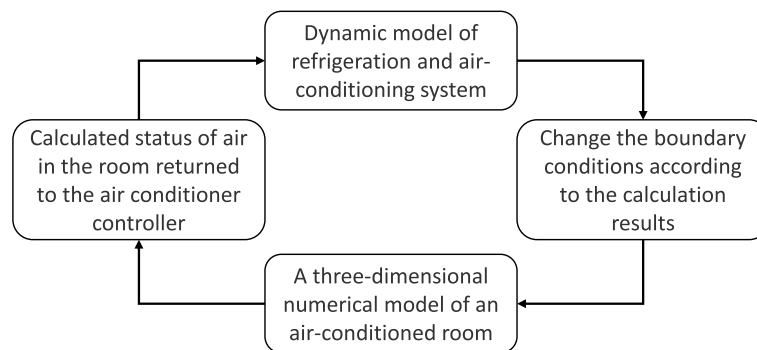
**2.2 Collaborative simulation platform**

Data derived from the collaborative simulation platform between the air-conditioning system and air-conditioned room helped expand the dataset used for the data mining. The collaborative simulation platform comprised an air-conditioning system model and a three-dimensional numerical model of an air-conditioned room [31]. The working principle of the collaborative simulation is illustrated in Fig. 3. The air-conditioning system model passes the air supply state data to the inlet boundary of the three-dimensional numerical room model, and the room thermal environment model feeds back the corresponding calculation results of the room temperature field as an input to the air-conditioning system model, which is expected to occur in real time. Therefore, the supply air parameters of the indoor unit can be provided by the air-conditioning system model, and the operation mode of the air-conditioning system can be adjusted according to the return air parameters. The three-dimensional

numerical model can consider the temperature fluctuations in the outdoor chamber and the heat transfer through the building envelope. Through a collaborative calculation platform, 150 groups of thermal environmental evaluation indices under different parameters were obtained. In this study, the structural parameters, i.e., tube spacing, tube outer diameter, and fin spacing, were mainly concerned with the evaporator, which represents the indoor unit of the air conditioner during the cooling stage. The control parameters of air conditioners included the set temperature, air supply method, and air supply speed.

**3 Model development**

Based on both experimental and simulated results, the set of data was provided by considering variations in structure and control parameters of the air-conditioning system, as shown in Fig. 4. The process of constructing the data-mining model is illustrated in Fig. 5. By testing different types of air conditioners, the thermal comfort tests mainly output air temperature distributions (few of them output velocity variations) in the tested room. These data are first used to validate the results predicted by the Simulink/Fluent collaborative simulation model. With the validated platform, more types of air conditioners with different structural parameters are simulated to predict the temperature and velocity distributions in the same tested room. Thermal comfort indices are computed based on the distribution profiles. Both the test data and simulated data derived on multiple types of air conditioners are collected to form the data set which will be used for data learning. With the dataset used for the data-mining study preprocessed, the input variables were determined based on feature selection. They included the tube spacing of the evaporator, tube outer diameter of the evaporator, fin spacing of the evaporator, set temperature, air supply angle, and air supply speed. The dataset was divided into a training set and a test set in a ratio



**Fig. 3** Working principle of collaborative simulation



Structure parameters			Air supply and control parameters	Number of data
Tube diameter (mm)	Fin pitch (mm)	Tube pitch (mm)		
5.00	1.19	19.05	Air supply speed: 3.0 m/s Set temperature: 27°C Air supply method: Vertical air supply	15
5.00	1.26	19.05		
5.00	1.33	19.05		
5.00	1.40	19.05		
5.00	1.47	19.05		
4.75	1.40	19.05		
5.25	1.40	19.05		
5.50	1.40	19.05		
5.75	1.40	19.05		
6.00	1.40	19.05		
5.00	1.40	18.10		
5.00	1.40	20.00		
5.00	1.40	20.96		
5.00	1.40	21.91		
5.00	1.40	22.86		
Air supply speed (m/s)	2.5, 3.5			30
Set temperature (°C)	23, 24, 25, 26			60
Air supply method	Upper air supply / Lower air supply / Left air supply			45

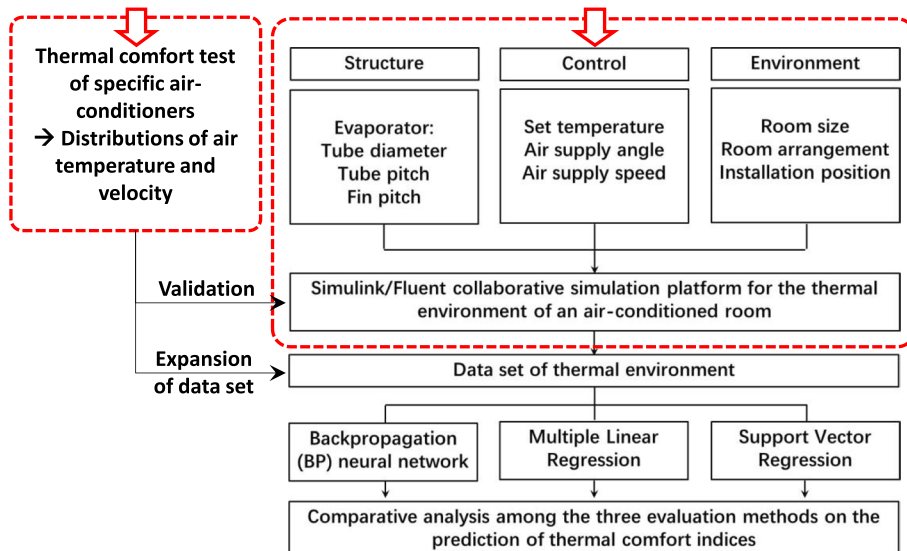


Fig. 4 Schematic of the methodology

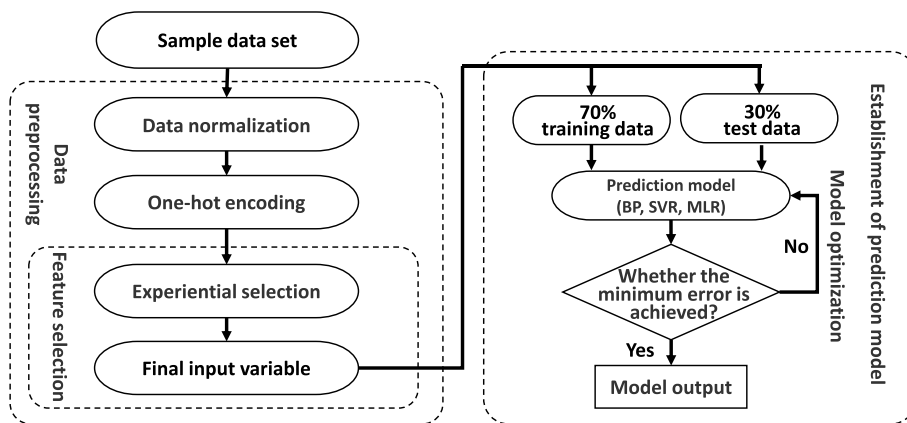


Fig. 5 Process of constructing data-mining model

of 7:3, and the optimal output model was obtained after adjusting the parameters.

### 3.1 Data-mining method

Based on the literature review, three types of data-mining methods, namely the BP neural network model, MLR model, and SVR model, were selected.

#### 3.1.1 BP neural network model

The BP neural network model [32] represents a multi-layer feedforward network. It has the classification ability of arbitrary complex systems and the excellent mapping ability of multivariate functions. The basic component units of the BP neural network are the neurons. Neurons constitute the input layer, hidden layer, and output layer of the BP neural network, and neurons between layers are directly connected by weights and thresholds. The values of the layers are calculated using Equations (1) and (2).

$$v_b^B = f \left( \sum_{m=1}^M \omega_{mb} x_{km} - \theta_{mb} \right) \tag{1}$$

$$v_c^C = f \left( \sum_{b=1}^B \omega_{bc} v_b^B - \theta_{bc} \right) \tag{2}$$

where  $v_b^B$  denotes the output value of the  $b^{th}$  neuron in the hidden layer,  $v_c^C$  denotes the output value of the  $c^{th}$  neuron in the output layer,  $f$  is the activation function,  $\omega_{mb}$  and  $\omega_{bc}$  denote the values of the weights, and  $\theta_{mb}$  and  $\theta_{bc}$  denote the values of the thresholds. The activation functions adopted by the BP neural network are the rectified linear unit and sigmoid function in the hidden and output layers, respectively. The values of the weights and thresholds are calculated using the gradient descent method.

#### 3.1.2 MLR model

The MLR model is used to describe the relationship between multiple input variables and dependent variables. This model is widely used in regression prediction problems owing to its simple structure. In general, it is calculated using Eq. (3):

$$y_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \dots + b_k x_{ki} \tag{3}$$

Here,  $y_i$  denotes the  $i$ th value of the dependent variable  $y$  ( $i = 1, 2, 3, \dots, n$ ),  $n$  denotes the number of dependent variables,  $k$  denotes the number of independent variables,  $x_{ki}$  denotes the  $i$ th sample value of the  $k$ th independent variable,  $b_0$  is a constant, and  $b_k$  represents the regression coefficient of each independent variable.

#### 3.1.3 SVR model

The SVR model [33] is used for small samples, high dimensions, nonlinear regression, and classification problems. The SVR method is described in Eqs. (4)–(8). Considering a sample set  $\{(x_i, y_i)\}_{i=1}^N \in R^{N \times D}$ , through the nonlinear mapping function  $\phi(x_i)$ , the input data  $x_i$  in the original sample are mapped to the high-dimensional space, and the linear regression equation is constructed in the high-dimensional feature space. Its expression can be transformed into Eq. (4), and the coefficients  $\omega$  and  $b$  are estimated by minimizing Eq. (5).

$$y = \sum_{i=1}^n \omega \phi(x_i) + b \tag{4}$$

$$\min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^N (\theta_i + \theta_i^*) \tag{5}$$

where  $C$  is the regularization constant, and  $\theta_i$  and  $\theta_i^*$  denote the positive and negative relaxation constants, respectively. The constraint conditions for Eq. (5) are given in Eq. (6).

$$\begin{cases} y_i - [\omega \phi(x_i) + b] \leq \varepsilon + \theta_i \\ [\omega \phi(x_i) + b] - y_i \leq \varepsilon + \theta_i^* \end{cases} \tag{6}$$

The final nonlinear regression equation can be obtained as shown in Eq. (7), for which the constraint conditions are shown in Eq. (8).

$$f(x) = \sum_{i=1}^N (\alpha_i - \alpha_i^*) K(x_i, x) + b \tag{7}$$

$$\begin{cases} \sum_{i=1}^N (\alpha_i - \alpha_i^*) = 0 \\ 0 \leq \alpha_i, \alpha_i^* \leq C \end{cases} \tag{8}$$

where  $\alpha_i$  and  $\alpha_i^*$  are Lagrange multipliers, and  $K(x_i, x)$  denotes the kernel function. Among the various kernel functions, the Gaussian radial basis kernel function is commonly used in the analysis of nonlinear data. Its expression is given by Eq. (9).

$$K(x_i, x) = \exp(-g|x_i - x|^2) \tag{9}$$

where  $x_i$  denotes the  $i$ th sample point,  $x$  is an independent variable, and  $g$  represents the tolerance coefficient.

## 3.2 Data preprocessing

### 3.2.1 Data normalization

In the process of model training, the prediction accuracy is reduced because of the dimensionality of variables. The purpose of normalization is to eliminate the dimensional effects in the sample data. Normalization is used to scale

the values of all variables in the sample data between 0 and 1 to reduce the amount of calculation and improve the prediction accuracy of the model. The data normalization method adopted was the maximum–minimum normalization method; the calculation formula is shown in Eq. (10).

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{10}$$

where  $x$  denotes the original value of a variable;  $x_{min}$  and  $x_{max}$  denote the minimum and maximum values of the variable, respectively; and  $x_{norm}$  denotes the value of  $x$  after normalization.

### 3.2.2 One-hot encoding

One-hot encoding is used for nonordered variables, which convert categorical data attributes into numerical data attributes. A list is used to present the categorical attributes. The list assigns a value of 1 in cases matching the attribute to be represented; otherwise, it assigns a value of 0. The air supply mode in this study was a

**Table 2** One-hot coding of air supply method

Air supply method	State 1	State 2	State 3	State 4
Upper air supply	1	0	0	0
Lower air supply	0	1	0	0
Vertical air supply	0	0	1	0
Left air supply	0	0	0	1

**Table 3** Experimental settings of thermal environment evaluation model

Evaluation method	Prediction data	Model	Hyperparameters			
			Iteration	Number of hidden-layer neurons	C	Gamma
Multiple-input single-output evaluation	Vertical temperature difference	SVR	/	/	100	0.6
		BP	20	400	/	/
		MLR	200	/	/	/
	Temperature uniformity	SVR	/	/	100	0.1
		BP	15	300	/	/
		MLR	100	/	/	/
	Temperature drop rate	SVR	/	/	5	0.1
		BP	15	250	/	/
		MLR	300	/	/	/
	Draft rate	SVR	/	/	280	0.25
		BP	15	200	/	/
		MLR	300	/	/	/
PMV	SVR	/	/	380	0.25	
	BP	15	300	/	/	
	MLR	350	/	/	/	
Multiple-input multiple-output evaluation			/	/	100	0.06

nonordered discrete-type feature, and the representation status of each air supply method after one-hot encoding is presented in Table 2.

### 3.3 Experimental settings

A greedy strategy was used to adjust the simulation settings, which refers to the process of adjusting the hyperparameters. The result obtained by the greedy strategy may not be the global optimal solution, but it can simplify the calculation procedure and save training time. The optimal combination of the hyperparameters of the thermal environment evaluation model is presented in Table 3.

## 4 Results and discussion

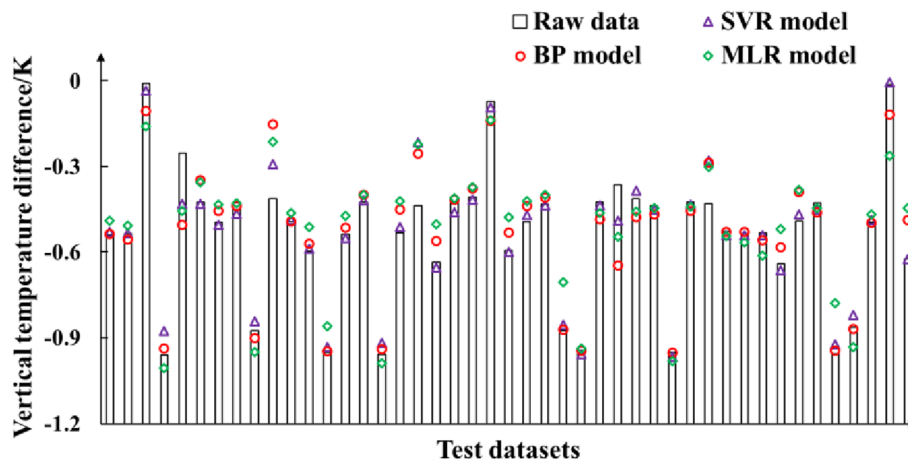
The original values of the 45 sample points and the predicted values obtained using the three data-mining methods in the test datasets were plotted and compared in the multiple-input single-output models. Then, the best data-mining method was selected to construct the multiple-input multiple-output models. In this section, the results are presented and discussed.

### 4.1 Multiple-input single-output evaluation

#### 4.1.1 Vertical temperature difference

The vertical temperature difference was used to evaluate the thermal discomfort of the human body. The temperature values of the measuring points at the head (the height was assumed to be 1.6 m) and ankle (the height was assumed to be 0.1 m) in the same vertical





**Fig. 6** Prediction results of vertical temperature difference

**Table 4** Test dataset evaluation indices of vertical air temperature difference

Model	Evaluation indices		
	RMSE	MAE	$R^2$
SVR model	0.059	0.034	0.936
BP model	0.090	0.058	0.853
MLR model	0.105	0.082	0.799

line direction were obtained. Figure 6 depicts that the predicted values of some sample points differ significantly from the raw values, such as sample points 10, 18, and 34. At these points, the predicted values derived from the three models were all greater than the raw values. The predicted values obtained by the three models for sample point 29 were lesser than the raw values. For the other sample points, the predicted values obtained by the SVR model were closer to the raw values, and its prediction accuracy was the best. The predicted values obtained by the MLR model were far from the raw values, and its prediction accuracy was the worst. The prediction accuracy of the BP neural network model ranked second among the three models. The evaluation indices of the test dataset are listed in Table 4. The root mean square error (RMSE) and mean absolute error (MAE) of the SVR model were 0.059 and 0.034, respectively, while the RMSE and MAE for the BP model were 0.09 and 0.058, respectively, and those for the MLR model were 0.105 and 0.082, respectively. The  $R^2$  of the SVR model was 0.936, which was considerably higher than those of the BP and MLR models. The SVR model was the best in predicting the vertical temperature difference.

In Table 4, the RMSE represents the square root of the squared deviation between the real and predicted data

and is used to describe the degree of deviation between the predicted and real data, which is calculated using Eq. (11). The MAE represents the average value of the deviation between the real and predicted data, which is calculated using Eq. (12). The goodness of fit ( $R^2$ ) indicates the degree of fit with the true value, and the calculation formula is shown in Eq. (13).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y_i')^2} \tag{11}$$

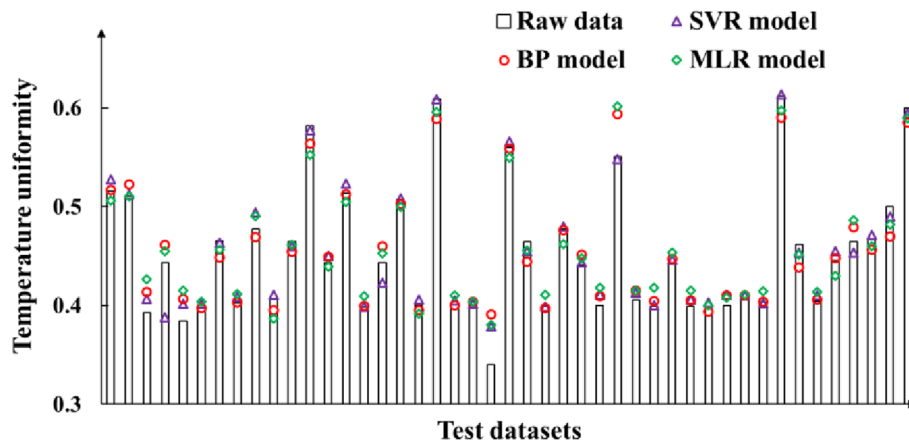
$$MAE = \frac{1}{n} \sum_{i=1}^n |(y_i - y_i')| \tag{12}$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i' - y_i)^2}{\sum_{i=1}^n (\bar{y}_i - y_i)^2} \tag{13}$$

where  $y_i$  represents the real data,  $y_i'$  represents the predicted data,  $n$  represents the total number of predicted data, and  $\bar{y}_i$  represents the average value of the predicted data in Eqs. (11)–(13).

#### 4.1.2 Temperature uniformity

Temperature uniformity was utilized to evaluate the difference in air temperature among all the measuring points at the same time. The instantaneous temperature uniformity at a certain moment was represented by the standard deviation of the instantaneous air temperature of all 147 measuring points at this time. A scatter plot of the predicted temperature uniformity is depicted in Fig. 7. The prediction results of the three models were close to the raw values. The individual differences were larger; for instance, the predicted value obtained by the



**Fig. 7** Prediction results of temperature uniformity

**Table 5** Test dataset evaluation indices of temperature uniformity

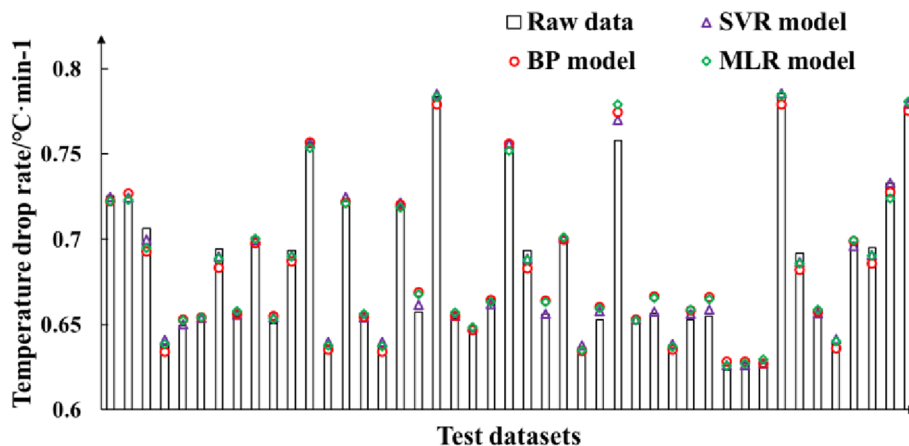
Model	Evaluation indices		
	RMSE	MAE	$R^2$
SVR model	0.013	0.008	0.962
BP model	0.015	0.011	0.943
MLR model	0.016	0.013	0.936

SVR model was smaller than the raw value at sample point 4, and the relative error was close to 10%. The predicted values of the BP and MLR models at sample point 29 were greater than the raw values, and the relative error was close to 15%. The evaluation indices of the test dataset are listed in Table 5. The RMSE and MAE of the SVR model were 0.013 and 0.008, respectively, which were the smallest among the three models. The  $R^2$  of the SVR model was 0.962, which was considerably higher than

those of the BP and MLR models. In general, the SVR model yielded the best prediction accuracy, followed by the BP and MLR models.

#### 4.1.3 Temperature drop rate

The temperature drop rate is defined as the rate at which the thermal environment of the room reaches a steady state. As illustrated in Fig. 8, the predicted results of the temperature drop rate derived from the proposed three drop rate models were all close to the raw values. The maximum relative error of the sample points was approximately 2%. In comparison with the previous prediction of the vertical air temperature difference and temperature uniformity, the errors between the raw values and predicted results were all minor. Therefore, the prediction accuracies of the three models for the temperature drop rate were relatively high. The evaluation indices of the three data-mining models on the output of the temperature drop are presented in Table 6. The RMSE and MAE of the three models were all less than 0.01, and the  $R^2$  values



**Fig. 8** Prediction results of temperature drop rate

**Table 6** Test dataset evaluation indices of temperature drop rate

Model	Evaluation indices		
	RMSE	MAE	R <sup>2</sup>
SVR model	0.0048	0.0028	0.995
BP model	0.0062	0.0048	0.981
MLR model	0.0054	0.0035	0.986

were greater than 0.98, indicating that all the three models exhibited high predictive precision for the temperature drop rate, while the SVR model was slightly better than the other two models.

**4.1.4 Draft rate**

The draft rate represents the percentage of occupants who are dissatisfied because of the loss of heat from the human body due to air flow. The magnitude of such an index depends on the indoor temperature, air speed, turbulence intensity, physical activity level, and clothing of the person, among other factors. The local draft rate indices of measuring point *i* during the collection time can be calculated using Eq. (14), where the draft rate in the air-conditioned room is the average of all the measured points.

$$DR_i = (34 - t_a)(u_a - 0.05)^{0.62}(0.37u_a I_T + 3.14) \tag{14}$$

where  $u_a$  denotes the local wind speed and  $I_T$  denotes the local turbulence intensity, which is 40%.

The simulation results for the draft rate are depicted in Fig. 9. The relative error of the three models at sample point 3 with the largest error was approximately 10%. The predicted value of the SVR model was closer to the raw value, and the prediction effect was the best. Table 7 indicates that the R<sup>2</sup> of the SVR model was slightly higher

**Table 7** Test dataset evaluation indices of draft rate

Model	Evaluation indices		
	RMSE	MAE	R <sup>2</sup>
SVR model	0.230	0.111	0.970
BP model	0.282	0.200	0.955
MLR model	0.266	0.148	0.960

than those of the BP and MLR models (approximately 0.970), while the RMSE and MAE were lower, with values of 0.230 and 0.111, respectively. In general, the predicted value of the SVR model was the closest to the raw value, and the prediction accuracy was higher than those of the BP and MLR models.

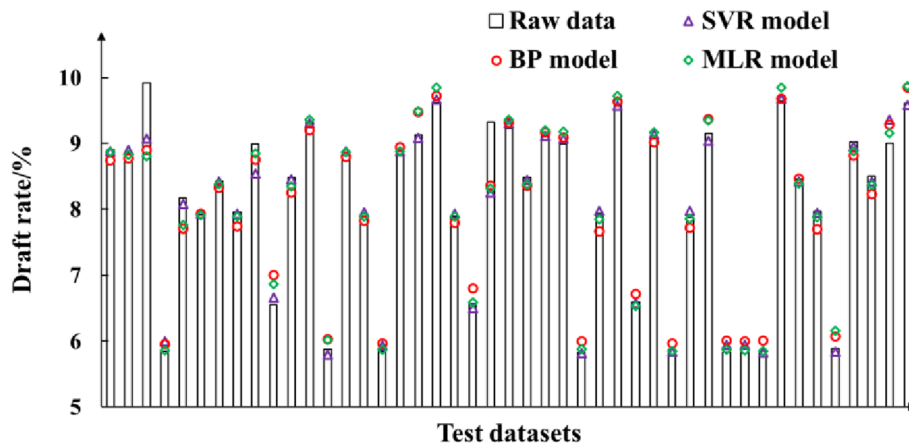
**4.1.5 PMV**

PMV is an internationally recognized comprehensive evaluation index that considers many factors related to human thermal comfort. In the specification, the PMV considers the seven-level thermal sensation evaluation standard by a large sample of people as the thermal comfort indices. The PMV evaluation and human thermal load calculation models are given by Eqs. (15) and (16).

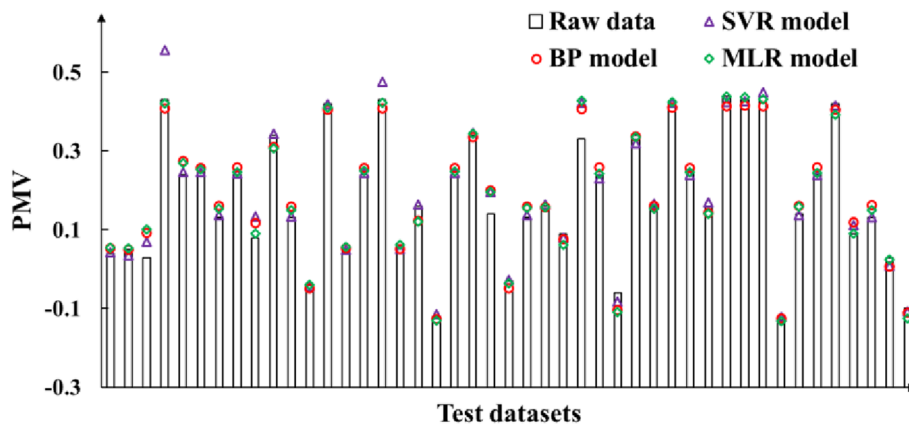
$$PMV = \left(0.303e^{-0.036M} + 0.028\right)TL \tag{15}$$

$$TL = (M - W) - 3.05[5.733 - 0.007(M - W) - P_a] - 0.42 \times (M - W - 58.15) - 1.73 \times 10^{-3}M(5.867 - P_a) - 1.4 \times 10^{-3} \times M(34 - t_a) - 3.96 \times 10^{-8}f_{cl}[(t_{cl} + 273)^4 - (t_r + 273)^4] - f_{cl}h_c(t_{cl} - t_a) \tag{16}$$

where  $M$  denotes the metabolic rate of the human body;  $W$  denotes the heat consumed by external work;  $P_a$  denotes the partial pressure of water vapor;  $t_a$  denotes the



**Fig. 9** Prediction results of draft rate



**Fig. 10** Prediction results of PMV

**Table 8** Test dataset evaluation indices of PMV

Model	Evaluation indices		
	RMSE	MAE	$R^2$
SVR model	0.030	0.0250	0.968
BP model	0.026	0.0209	0.974
MLR model	0.025	0.0166	0.975

air temperature;  $f_{cl}$  denotes the clothing area coefficient, which is related to the thermal resistance of clothing;  $t_{cl}$  denotes the surface temperature of clothing;  $t_r$  denotes the average radiation temperature; and  $h_c$  denotes the convective heat transfer coefficient.

The plot of the PMV index prediction is depicted in Fig. 10. The predicted values of the BP and MLR models were close to the raw values, while the predicted value of the SVR model was different from the raw value to some extent. Therefore, the BP and MLR models were better than the SVR model in predicting the PMV indices. The evaluation indices of the PMV prediction model test set are listed in Table 8. The RMSE and MAE of the BP and MLR models were both less than 0.26, and the  $R^2$  values of these two models reached 0.974 and 0.975, respectively, which were slightly higher than that of the SVR model. The relative error of the optimal PMV index prediction model was maintained within 20%, which was larger than that of the other models. This is mainly because the values of the PMV indices were distributed around 0; in other words, the denominator was close to 0 when calculating the relative error of some sample points, thus causing the overall relative error to be larger.

**4.2 Multiple-input multiple-output evaluation**

All the models built were multiple-input single-output thermal environment assessments for a single thermal

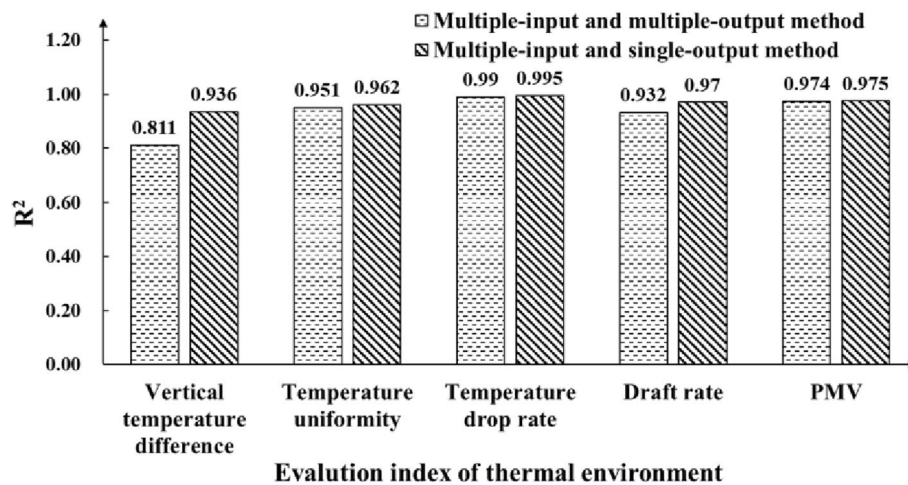
**Table 9** Test dataset evaluation indices of the multiple-input multiple-output evaluation

Prediction data	Evaluation indices		
	RMSE	MAE	$R^2$
Vertical temperature difference	0.084	0.043	0.811
Temperature uniformity	0.016	0.009	0.951
Temperature drop rate	0.003	0.002	0.990
Draft rate	0.341	0.155	0.932
PMV	0.027	0.015	0.974

environment evaluation index. According to the analysis, among the three data-mining methods, the SVR model performed well in prediction, with a relatively small error and a relatively high goodness of fit. To save calculation time and improve the prediction efficiency of the model, a multiple-input multiple-output thermal environment evaluation model was built using the SVR model, through which five thermal environment evaluation indices could be predicted simultaneously. The results of the test set of the multiple-input multiple-output SVR models are listed in Table 9. Except for the vertical air temperature difference, the  $R^2$  values for predicting other evaluation indices were all greater than 0.93, which are acceptable to a certain extent. However, as illustrated in Fig. 11, the prediction accuracy of the multiple-input multiple-output evaluation methods was reduced.

**5 Conclusions**

In this study, three data-mining methods—the MLR model, BP neural network model, and SVR model—were utilized to build the thermal environment evaluation model. The evaluation indices of the thermal environment, including the vertical air temperature difference,



**Fig. 11** Comparison of goodness of fit between the multiple-input multiple-output evaluation and the multiple-input single-output evaluation

temperature uniformity, temperature drop rate, draft rate, and PMV indices, were evaluated.

The simulation results indicated that the prediction accuracy of the SVR model was higher than those of the BP and MLR models with respect to the vertical air temperature difference, temperature uniformity, temperature drop rate, and draft rate in the multiple-input single-output evaluation method. In terms of PMV index prediction, the prediction accuracies of the BP and MLR models were slightly higher than that of the SVR model, and the relative error could be maintained at approximately 20%. The performance of the SVR model in predicting a single indoor thermal environment evaluation index was good, and it was thus selected for the multiple-input multiple-output evaluation method. The  $R^2$  of the prediction results was greater than 0.93, except for the prediction of the vertical air temperature difference, which was 0.81. In comparison with the multiple-input single-output evaluation method, the prediction accuracies of the corresponding indices were reduced; however, replacing five multiple-input single-output evaluation models with one multiple-input multiple-output evaluation model could simplify the complexity and save computation time. Therefore, the multiple-input single-output evaluation method is recommended for high-precision prediction, while the multiple-input multiple-output evaluation method is recommended for high-efficiency prediction.

**Acknowledgements**

This work was funded and supported by the State Key Laboratory of Air-conditioning Equipment and System Energy Conservation (Grant No. ACSKL2019KT11).

**Authors' contributions**

Conceptualization, S.Z., L.H. and S.C.; methodology, X.W. and S.C.; validation, G.X.; formal analysis, X.W.; investigation, S.Z.; resources, S.Z.; writing—original draft preparation, X.W.; writing—review and editing, S.C. and J.X.; supervision,

J.X.; project administration, L.H. and S.C.; funding acquisition, L.H. The author(s) read and approved the final manuscript.

**Availability of data and materials**

The data that support the findings of this study are available from the corresponding author upon reasonable request.

**Declarations**

**Competing interests**

The authors declare that they have no competing interests.

Received: 10 June 2022 Accepted: 20 January 2023

Published online: 08 February 2023

**References**

- General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China, Standardization Administration of the People's Republic of China. (2017). Thermal comfort requirements and evaluation for indoor environment. In *GB/T 33658-2017*. Standards Press of China.
- 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.
- ANSI/ASHRAE, ANSI/ASHRAE Standard 55. (2020). *Thermal environmental conditions for human occupancy*. American Society of Heating, Refrigerating and Air-Conditioning Engineers.
- Yin, Q., Cao, Y., & Sun, C. (2021). Research on outdoor thermal comfort of high-density urban center in severe cold area. *Building and Environment*, 200, 107938.
- Gao, S., Ooka, R., & Oh, W. (2021). Experimental investigation of the effect of clothing insulation on thermal comfort indices. *Building and Environment*, 187, 107393.
- Yang, K. H., & Su, C. H. (1997). An approach to building energy savings using the PMV index. *Building and Environment*, 32(1), 25–30.
- Wu, X., Liu, Y., Liu, G., Wang, F., & Wang, Z. (2017). Effect of supply air temperature on indoor thermal comfort in a room with radiant heating and mechanical ventilation. *Energy Procedia*, 121, 206–213.
- Tawackolian, K., Lichtner, E., & Kriegel, M. (2020). Draught perception in intermittent ventilation at neutral room temperature. *Energy and Buildings*, 224, 110268.

9. Embaye, M., AL-Dadah, R. K., & Mahmoud, S. (2016). Numerical evaluation of indoor thermal comfort and energy saving by operating the heating panel radiator at different flow strategies. *Energy and Buildings*, *121*, 298–308.
10. Shan, X., Xu, W., Lee, Y., & Lu, W. (2019). Evaluation of thermal environment by coupling CFD analysis and wireless-sensor measurements of a full-scale room with cooling system. *Sustainable Cities and Society*, *45*, 395–405.
11. Aryal, P., & Leephakpreeda, T. (2015). CFD analysis on thermal comfort and energy consumption effected by partitions in air-conditioned building. *Energy Procedia*, *79*, 183–188.
12. Xiaofang, S., Na, L., Kaiyu, S., Tianzhen, H., Yi-Kuen, L., & Wei-Zhen, L. (2020). Coupling CFD and building energy modelling to optimize the operation of a large open office space for occupant comfort. *Sustainable Cities and Society*, *60*, 338–348.
13. Xing, L., Li, L., Gong, J., Ren, C., Liu, J., & Chen, H. (2018). Daily soil temperatures predictions for various climates in United States using data-driven model. *Energy*, *160*, 430–440.
14. Zhao, H., & Magoulès, F. (2012). A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews*, *16*(6), 3586–3592.
15. Zhang, W., Liu, F., & Fan, R. (2018). Improved thermal comfort modeling for smart buildings: A data analytics study. *International Journal of Electrical Power*, *103*, 634–643.
16. Hang, L., & Kim, D. (2018). Enhanced model-based predictive control system based on fuzzy logic for maintaining thermal comfort in IoT smart space. *Applied Sciences*, *8*, (7), 1031.
17. Brik, B., Esseghir, M., Merghem-Boulaïhia, L., & Snoussi, H. (2019). *ThermCont: A machine learning enabled thermal comfort control tool in a real time* (pp. 294–300). 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC).
18. Broday, E. E., Moreto, J. A., Xavier, A. A. D. P., & de Oliveira, R. (2019). The approximation between thermal sensation votes (TSV) and predicted mean vote (PMV): A comparative analysis. *International Journal of Industrial Ergonomics*, *69*, 1–8.
19. Liu, W., Lian, Z., & Zhao, B. (2007). A neural network evaluation model for individual thermal comfort. *Energy and Buildings*, *39*(10), 1115–1122.
20. Yao, J., & Xu, J. (2010). Research on the BPNN in the prediction of PMV. *Advanced Materials Research*, *29-32*(3), 2804–2808.
21. Wan, T., Bai, Y., Wang, T., & Wei, Z. (2022). BPNN-based optimal strategy for dynamic energy optimization with providing proper thermal comfort under the different outdoor air temperatures. *Appl Energy*, *313*, 118899.
22. Chaudhuri, T., Zhai, D., Soh, Y. C., Li, H., & Xie, L. (2018). Thermal comfort prediction using normalized skin temperature in a uniform built environment. *Energy and Buildings*, *159*, 426–440.
23. Viani, F., & Polo, A. (2017). A forecasting strategy based on wireless sensing for thermal comfort optimization in smart buildings. *Microwave and Optical Technology Letters*, *59*(11), 2913–2917.
24. Qin, H., & Wang, X. (2022). A multi-discipline predictive intelligent control method for maintaining the thermal comfort on indoor environment. *Applied Soft Computing*, *116*, 108299.
25. Luo, M., Xie, J., Yan, Y., Ke, Z., Yu, P., Wang, Z., & Zhang, J. (2020). Comparing machine learning algorithms in predicting thermal sensation using ASHRAE comfort database II. *Energy and Buildings*, *210*, 109776.
26. Mustafaraj, G., Lowry, G., & Chen, J. (2011). Prediction of room temperature and relative humidity by autoregressive linear and nonlinear neural network models for an open office. *Energy and Buildings*, *43*(6), 1452–1460.
27. Wu, Z., Li, N., Peng, J., Cui, H., Liu, P., Li, H., & Li, X. (2018). Using an ensemble machine learning methodology-bagging to predict occupants' thermal comfort in buildings. *Energy and Buildings*, *173*, 117–127.
28. Buratti, C., Vergoni, M., & Palladino, D. (2015). Thermal comfort evaluation within non-residential environments: Development of artificial neural network by using the adaptive approach data. *Energy Procedia*, *78*, 2875–2880.
29. Atthajariyakul, S., & Leephakpreeda, T. (2005). Neural computing thermal comfort index for HVAC systems. *Energy Conversion and Management*, *46*(15-16), 2553–2565.
30. Castilla, M., Alvarez, J. D., Ortega, M. G., & Arahall, M. R. (2013). Neural network and polynomial approximated thermal comfort models for HVAC systems. *Building and Environment*, *59*, 107–115.
31. He, L., Zhao, S., Xu, G., Wu, X., Xie, J., & Cai, S. (2021). Prediction and evaluation of dynamic variations of the thermal environment in an air-conditioned room using collaborative simulation method. *Energies*, *14*, 5378.
32. Shi, Y., Zhao, X., Zhang, Y., & Ren, N. (2009). Back propagation neural network (BPNN) prediction model and control strategies of methanogen phase reactor treating traditional Chinese medicine wastewater (TCMW). *Journal of Biotechnology*, *144*(1), 70–74.
33. Bansal, S., Roy, S., & Larachi, F. (2012). Support vector regression models for trickle bed reactors. *Chemical Engineering Journal*, *207-208*, 822–831.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.