Build Simul (2022) 15: 2111–2125
https://doi.org/10.1007/s12273-022-0911-2

Development of data-driven thermal sensation prediction model using quality-controlled databases

Xiang Zhou[1], Ling Xu[1], Jingsi Zhang[2,1], Lie Ma[2], Mingzheng Zhang[2], Maohui Luo[1] (✉)

1. School of Mechanical Engineering, Tongji University, Shanghai 200092, China
2. Guangdong Midea Air-Conditioning Equipment Co., LTD, Guangdong 528311, China

Abstract

Predicting the thermal sensations of building occupants is challenging, but useful for indoor environment conditioning. In this study, a data-driven thermal sensation prediction model was developed using three quality-controlled thermal comfort databases. Different machine-learning algorithms were compared in terms of prediction accuracy and rationality. The model was further improved by adding categorical inputs, and building submodels and general models for different contexts. A comprehensive data-driven thermal sensation prediction model was established. The results indicate that the multilayer perceptron (MLP) algorithm achieves higher prediction accuracy and more rational results than the other four algorithms in this specific case. Labeling AC and NV scenarios, climate zones, and cooling and heating seasons can improve model performance. Establishing submodels for specific scenarios can result in better thermal sensation vote (TSV) predictions than using general models with or without labels. With 11 submodels corresponding to 11 scenarios, and three general models without labels, the final TSV prediction model achieved higher prediction accuracy, with 64.7%–90.7% fewer prediction errors (reducing SSE by 3.2–4.9) than the predicted mean vote (PMV). Possible applications of the new model are discussed. The findings of this study can help in development of simple, accurate, and rational thermal sensation prediction tools.

Keywords

thermal comfort;
machine learning;
adaptive thermal comfort;
multilayer perceptron

Article History

Received: 14 March 2022
Revised: 26 May 2022
Accepted: 27 May 2022

1 Introduction

With continuous improvement in living standards, the thermal comfort of building occupants has received increased attention, especially in the field of indoor environment conditioning. Methods and tools to predict and improve thermal comfort can help increase energy-efficiency in indoor environments (O'Brien et al. 2020; Zhang et al. 2022), with a greater satisfaction rate (Wagner et al. 2007; Frontczak en al. 2012) and working performance (Elnaklah et al. 2020; Tang et al. 2021).

It is known that the thermal comfort of building occupants is affected by many factors (Fanger 1970) aside from air temperature (Ta), radiant temperature (Tr) (Zhou et al. 2019), air speed (Va) (Zhu et al. 2015), relative humidity (RH), metabolic rate (Met) (Luo et al. 2018b), and clothing insulation (Clo), including age, thermal experiences (Lin et al. 2021), and outdoor climate conditions (Nicol and Humphreys 1973; De Dear and Schiller Brager 2001). Many models have been developed to quantify how different factors affect thermal sensation and thermal comfort. Many indexes can be found in the Berkeley comfort tool (Schiavon et al. 2014), the "Pythermalcomfort" package in Python (Tartarini and Schiavon 2020), and the "COMF" package in R programming (Schweiker 2016). The predicted mean vote (PMV) and adaptive comfort model are two classic indexes that have been incorporated in standards such as the ASHRAE standard 55 (ASHRAE 2017) and EN 15251 (Nicol and Humphreys 2010).

However, conventional thermal sensation vote (TSV) and thermal comfort prediction approaches have shortcomings. The prediction accuracy of the PMV model was low in field studies, particularly in NV buildings. Cheung et al. (2019) reported that the accuracy of PMV in predicting thermal

E-mail: luomaohui@tongji.edu.cn

Nomenclature

Algorithm abbreviations

GPR	Gaussian process regression
KNN	K nearest neighbors
ML	machine learning
MLP	multilayer perception
PI	permutation importance
RF	random forest
SVM	support vector machine

Thermal comfort-related abbreviations

AC	air-conditioned
Clo	clothing insulation (clo)
HVAC	heating, ventilation, and air-conditioning
Met	metabolic rate (met)
NV	naturally ventilated
PMV	predicted mean vote
RH	relative humidity (%)
TSV	thermal sensation vote
Ta	indoor air temperature (°C)
Tout	outdoor air temperature (°C)
Tr	mean radiant temperature (°C)
Va	air speed (m/s)

Other symbols

RMSE	root mean squared error
MAE	mean absolute error
R^2	correlation coefficients (r) squared
SSE	sum of squares for residuals

sensation was only 34%. The PMV and adaptive models lack self-learning capacity (Zhou et al. 2020), and are unable to adapt themselves to new contexts. To increase adaptability to new contexts, researchers have proposed correction coefficients such as the ePMV model (Fanger and Toftum 2002), aPMV model (Yao et al. 2009), and demand factor (Luo et al. 2018a). These coefficients can extend the scope of model application to some extent; however, once the coefficients are determined, the model is only suitable in certain contexts.

With advances in data-driven methods in recent years, many researchers have attempted to apply machine-learning (ML) algorithms in the field of thermal comfort to handle complex problems. The ML approach can achieve 5.3%–42% higher accuracy in thermal sensation prediction than conventional PMV and adaptive models (Kim et al. 2018; Wu et al. 2018; Cosma and Simha 2019; Li et al. 2019; Wu et al. 2021). Many studies (Lu et al. 2019; Luo et al. 2020; Zhou et al. 2020) have presented ML algorithms for TSV prediction using ASHRAE thermal comfort databases, with high prediction accuracy. More information on these studies can be found in a recent review article (Xie et al. 2020). Although data-driven thermal comfort prediction seems promising, there remain issues that require further investigation. First, with the limited number of data samples, the scope of model application is limited. Second, current data-driven thermal comfort studies focus mainly on improvement of model prediction accuracy without considering the rationality of the model. ML algorithms may be able to achieve high accuracy, but the output results may violate thermal comfort principles. Third, simple thermal sensation prediction tools have not been developed.

Inspired by previous research, a comprehensive thermal sensation prediction model with high accuracy, reasonable output, and wide applicability was developed in this study. Model development details are presented in Sections 2, 3, and 4. The model performance and application are presented in Section 5.

2 Methodology

2.1 Dataset

Three thermal comfort databases including 131,698 data samples collected from field studies in 89 cities and 23 countries worldwide were used in this study. The ASHRAE RP-884 database (De Dear 1998) provided 25,369 samples; the ASHRAE thermal comfort database II (Földváry Ličina et al. 2018) provided 81,965 samples, and the Chinese thermal comfort database developed with the support of the China National Key R&D Program from 2015–2020 provided 24,364 samples. The Chinese thermal comfort database is not publicly available, but will be published once the project is completed. These databases are the most quality-controlled and wide-ranging thermal comfort databases, allowing development of data-driven thermal sensation prediction models. The databases include information such as subjective thermal sensation and thermal comfort votes, measured thermal environment parameters, time and geographic information, personal factors such as clothing and activity level, and HVAC (heating, ventilation, and air-conditioning) operation modes.

2.2 Inputs and output

To select the proper input variables, six parameters in the PMV model were considered, including Ta, Tr, Met, Clo, Va, and RH. Three labels indicating the HVAC operation modes, climate zones, and HVAC operation season were selected to distinguish between indoor scenarios. HVAC operation mode was used to distinguish between AC and NV modes. Five climate labels (A/B/C/D/E) were used to distinguish between outdoor climate conditions based on the Koppen climate classification (Peel et al. 2007). The HVAC operation season label was used to distinguish between cooling and heating seasons in the AC scenario. The daily average outdoor temperature (Tout) was also selected as an input variable to consider the impact of the outdoor environment. The output variable of the machine-learning model was a continuous thermal sensation voting scale ranging from −3 to 3. Table 1 presents the input and output variables.

2.3 Algorithms

Weka (Waikato Environment for Knowledge Analysis, version 3.8.3) (Witten et al. 2000), an open-source machine-learning and data-mining platform based on Java programming, was used to develop the machine-learning model. Based on the findings of our previous study (Zhou et al. 2020), the following five algorithms were selected and compared.

Table 1 Descriptions of inputs and output

Parameter	Unit	Range/category	Type	Note
Ta	°C	8–38	Input	Suitable for both AC and NV scenarios
Tr	°C	4–40		
Met	met	0.7–3		
Clo	clo	0–2.0		
Va	m/s	0–2.0		
RH	%	0–100		
Tout	°C	3.9–38.5		
HVAC operation mode		AC/NV		
Climate zone		A (tropical) B (dry) C (humid subtropical) D (temperate continental) E (polar)		E was not included in the database
HVAC operation season		Cooling/heating		Suitable for AC scenarios
TSV		Scale from −3 to 3	Output	

A random forest (RF) is an ensemble algorithm that makes a mean prediction based on decision trees. The "RandomForest" package in Weka was used. The number of trees was set to 1000, and the size of each bag was set to 20% of the training set size.

A support vector machine (SVM) is a supervised learning method based on the principle of the Vapnik–Chervonenkis theory and structural risk minimization of statistical learning theory. The "LibSVM" package in Weka was used. Epsilon–SVR (regression) and "RBF" (Gaussian) kernels were used. The "gamma" and "cost" were tuned from 0–32 and 0–4, respectively.

Multilayer perception (MLP) is a backpropagation neural network method. The "MultilayerPerception" package in Weka was used. The "hidden layers", "learningRate", and "momentum" were tuned from 2–32, 0–1, and 0–1, respectively.

The K-nearest neighbors (KNN) algorithm is based on the distance theory and searches for nearest neighbors. The "IBk" package in Weka was used. The "KNN" algorithm was tuned from 1–64.

Gaussian process regression (GPR) is a method that does not require hyperparameter tuning and is based on probability theory and stochastic processes. The "GaussianProcesses" package in Weka was used. The "RBF" kernel was used; the "gamma" was tuned from 0–32.

2.4 Model performance evaluation

The root mean squared error (RMSE) and mean absolute error (MAE) were used to evaluate the model performance. Smaller RMSE and MAE values indicate better model performance. These indices are defined in Eqs. (1) and (2).

$$\mathrm{RMSE} = \sqrt{\frac{(p_1 - a_1)^2 + \cdots + (p_n - a_n)^2}{n}} \tag{1}$$

$$\mathrm{MAE} = \frac{|p_1 - a_1| + \cdots + |p_n - a_n|}{n} \tag{2}$$

$$\text{with} \qquad a_0 = \frac{a_1 + \ldots + a_n}{n}, \quad p_0 = \frac{p_1 + \ldots + p_n}{n}$$

where a_i is the target value, and p_i is the predicted value.

The ten-folds cross-validation method was used for model evaluation. The dataset was randomly split into ten folds by average, with nine folds for training, and one fold for testing. The averages of ten training and testing sessions were considered as the final results.

2.5 Overall model development process

Figure 1 shows a flowchart of the model development. Five machine-learning algorithms were compared in terms of

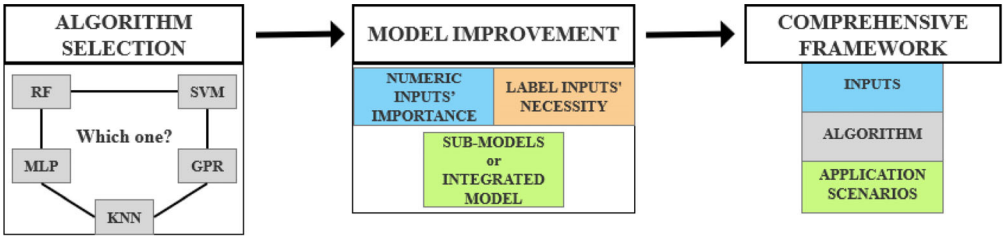


Fig. 1 Overall model development procedure

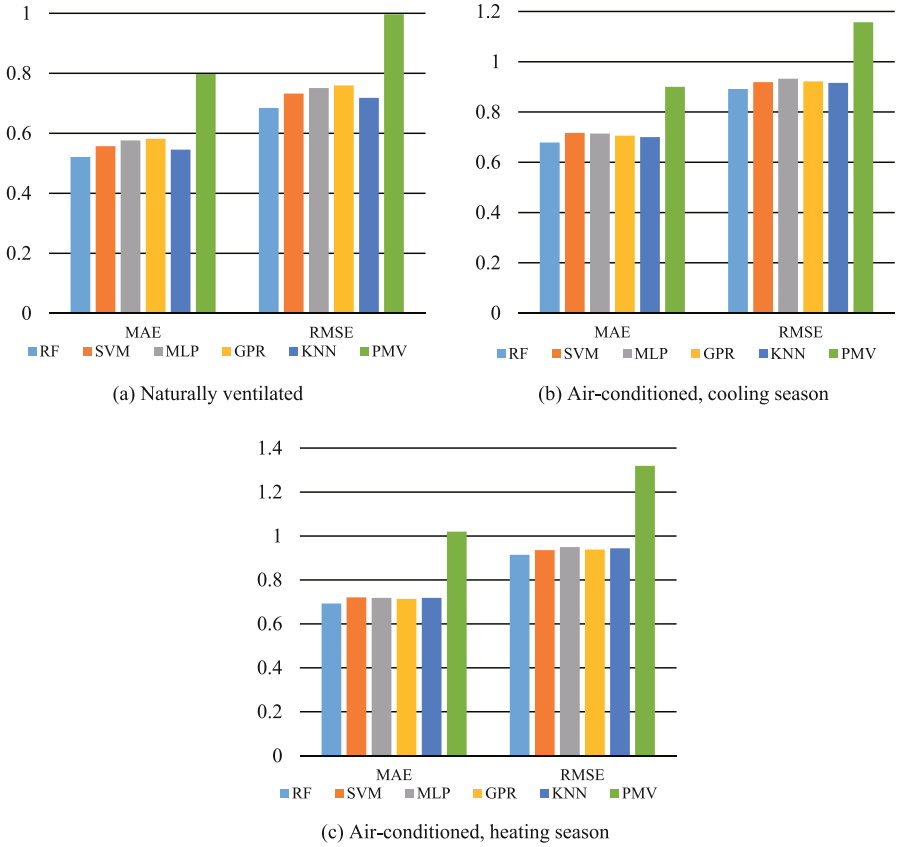
prediction accuracy and result analysis. The model was improved by considering the importance of each numeric input, the necessity of each label input, and the submodels and general model for different indoor scenarios. A comprehensive framework was established from input feature selection, algorithm determination, and application scenario options. Model development details are presented in the following sections.

3 Algorithm selection

Selecting an appropriate machine-learning algorithm for development of the final model is critical. Five algorithms were compared in terms of model prediction accuracy, predicted TSV in different temperature bins, and output analysis.

3.1 Model prediction accuracy

Figure 2 shows the MAE and RMSE for TSV prediction with different algorithms and the PMV model. In Figure 2(a), for naturally ventilated scenarios, the MAE and RMSE of the five algorithms decreased by 35.9% and 31.3%, respectively, compared with the PMV model. In Figure 2(b), for air-conditioned scenarios in the cooling season, the MAE and RMSE of the five algorithms decreased by 22.5% and 21.2%, respectively, compared with the PMV model. In Figure 2(c), for air-conditioned scenarios in the heating



(a) Naturally ventilated

(b) Air-conditioned, cooling season

(c) Air-conditioned, heating season

Fig. 2 Comparison of model prediction accuracy: (a) NV scenarios; (b) AC scenarios in cooling season; (c) AC scenarios in heating season

season, the MAE and RMSE of the five algorithms decreased by 30.7% and 29.4%, respectively, compared with the PMV model. For all three indoor scenarios shown in Figure 2, the five algorithms produced much smaller MAE and RMSE values than the PMV model. The difference between the five algorithms was 10.5% for the MAE and 9.8% for the RMSE.

Figure 3 compares the actual TSV, predicted TSV, and PMV for different air temperature bins. The sum of squares of residuals (SSE) and R^2 were used to describe the approximation degree between actual and predicted TSVs. The SSE indicates the accumulated squared error between predicted values and target values; R^2 describes the fitting degree from 0 to 1. Generally, a smaller SSE and an R^2 closer to 1 indicate a better model. SSE and R^2 are defined in Eqs. (3) and (4), where y is the target value (actual TSV), and y' is the corresponding predicted value (predicted TSV).

$$\mathrm{SSE} = \sum (y - y')^2 \tag{3}$$

$$R^2 = 1 - \frac{\mathrm{SSE}}{\sum (y_i - \bar{y})^2} \tag{4}$$

where $\bar{y} = \dfrac{y_1 + y_2 + \ldots + y_n}{n}$

In Figure 3, as the air temperature increases, the predicted TSVs are consistent with the actual TSV; the PMV is significantly different from the actual TSV. Comparing the bars in Figure 3, the five algorithms produced much smaller SSE and larger R^2 values than the PMV model, a 93.2% decrease in SSE and a 29.7% increase in R^2 on average for NV scenarios, a 90.7% decrease in SSE and a 150.3% increase in R^2 on average for AC scenarios in the cooling season, and a 66.3% decrease in SSE and a 49.1% increase in R^2 on average for AC scenarios in the heating season. Thus, the machine-learning algorithms demonstrated better TSV prediction performance than the PMV model.

3.2 Model rationality analysis

To further analyze the rationality of the model, the effects on the model output (the predicted TSV) were analyzed with changes in the six classic inputs. Figure 4 shows how the model output varies with the input parameters. Figure 4(a)

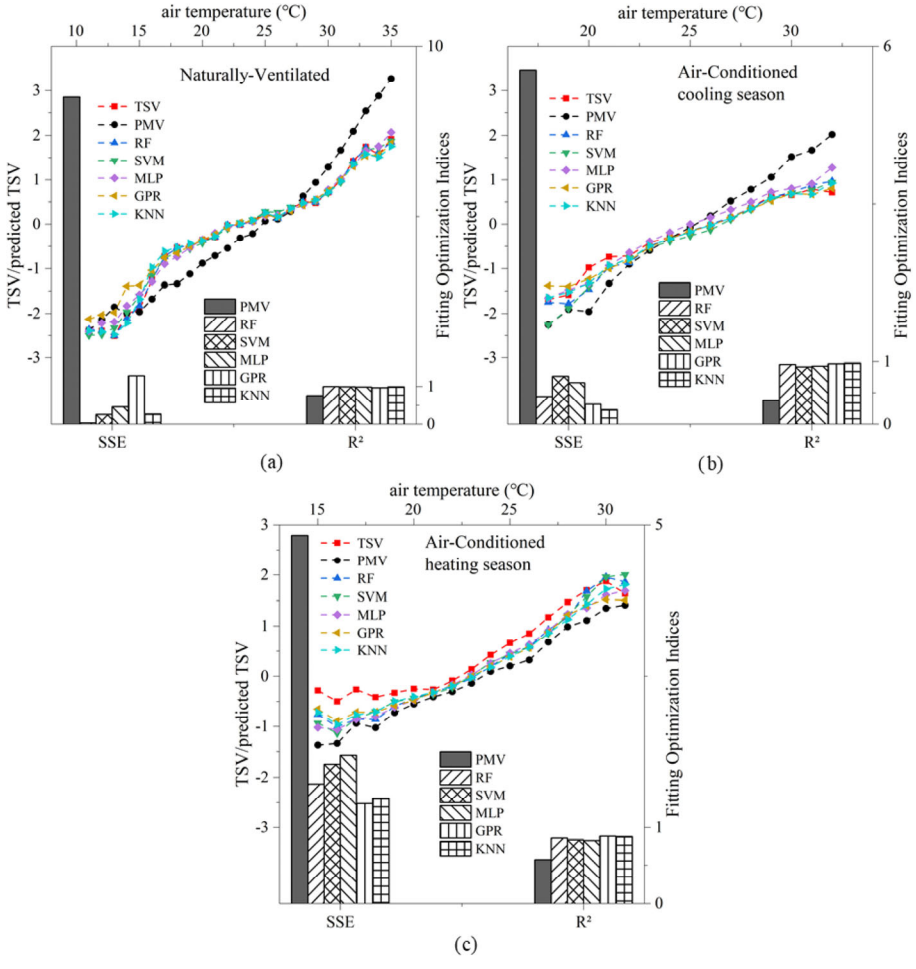


Fig. 3 Comparison of actual TSV, predicted TSV, and PMV as air temperature increases: (a) NV scenarios; (b) AC scenarios in cooling season; (c) AC scenarios in heating season

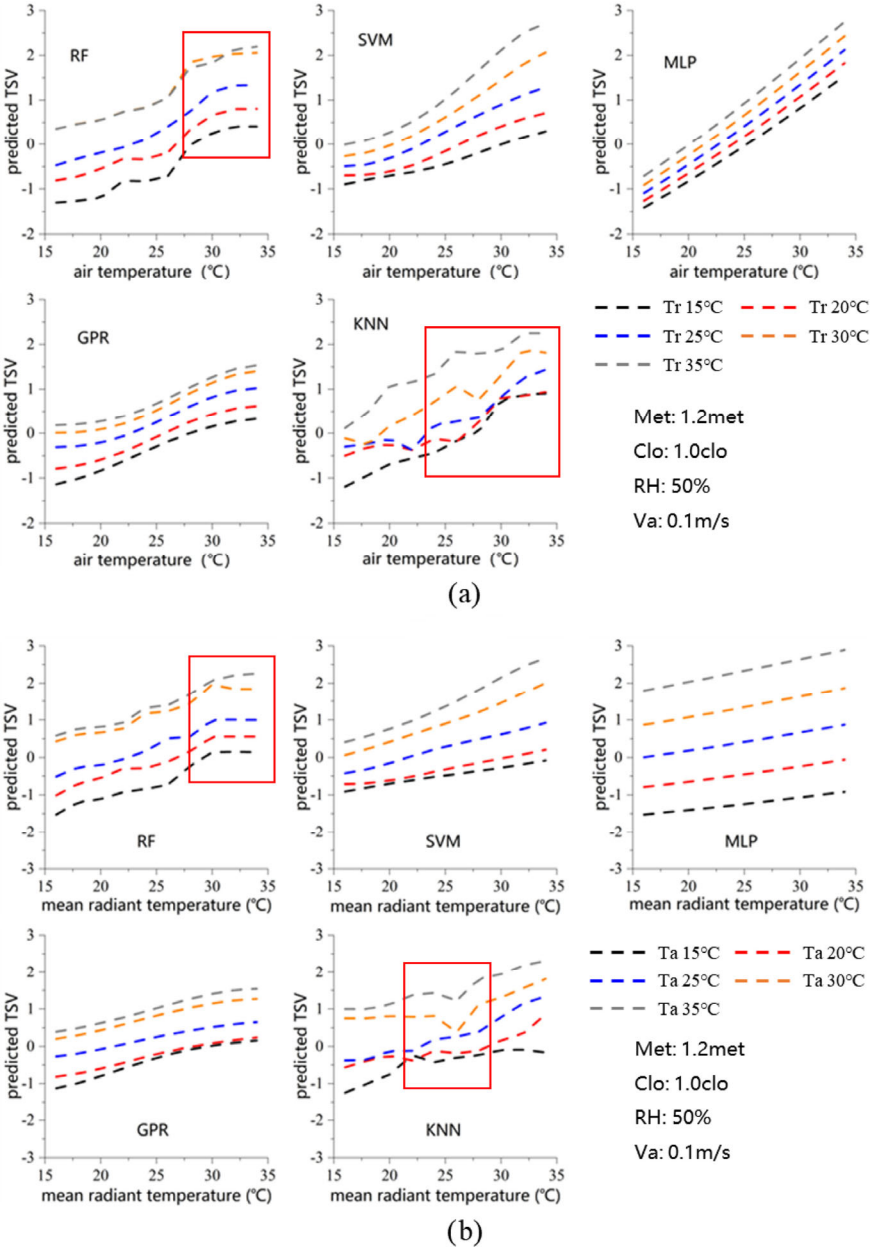


Fig. 4 Model output analysis: (a) varying Ta; (b) varying Tr (Note: the red brackets indicate areas with unreasonable results. Similar analysis for other parameters can be found in Appendix A in the ESM of the online version of this paper)

compares the predicted TSV with different algorithms with an increase in air temperature and the other five parameters (Tr, Met, Clo, Va, RH) set as constants. For models using the MLP, SVM, and GPR algorithms, the predicted TSV increased gradually with an increase in air temperature. However, for models using the RF or KNN algorithms, the predicted TSV exhibited a flat trend or a negative trend (see marked boxes) with an increase in air temperature, which was not consistent with higher air temperatures leading to warmer thermal sensations. In Figure 4(b), with an increase in mean radiant temperature, the predicted TSVs from machine-learning algorithms were compared, with the other five parameters (Ta, Met, Clo, Va, RH) set as constants. The RF and KNN algorithms produced some irrational results, indicated in the boxes; the predicted TSV exhibited negative growth with an increase in the mean radiant temperature. Similar analyses were conducted for Met, Clo, Va, and RH. More information is presented in Appendix A, Figure A1, which is available in the Electronic Supplementary Material (ESM) from the online version of this paper.

From the rationality analysis, we found that not all machine-learning algorithms produce rational TSV predictions. The circled areas in Figure 4 and in figures in

the Appendix indicate that some ML model outputs violate basic principles of thermal comfort. For example, it is widely known that a higher Ta and Tr, produce a warmer thermal sensation, assuming other parameters are the same. The circled areas in Figure 4 do not follow this knowledge. The irrational output may be caused by the data distribution and learning strategies of the algorithms. As the data samples were collected from field studies, the environmental parameters could not be controlled, unlike studies conducted in a climate chamber; the data distribution may be non-uniform. In addition, different algorithms use different learning strategies; some algorithms cannot adapt to discontinuities in the data distribution.

Our analysis showed that MLP outputs were consistent with thermal comfort principles; the RF, SVM, GPR, and KNN algorithms all produced an unreasonable output in some cases. MLP exhibits powerful learning for nonlinear regression, and the outputs are continuous with matching linear output layers. The learning strategy of MLP can better adapt to a discontinuous data distribution of thermal comfort data samples. Thus, we chose the MLP algorithm for the final model.

4 Modeling improvement

Some modeling improvements were made to the MLP algorithm. First, the importance of each numerical input parameter was calculated and ranked. Second, the necessity of each label input was investigated. Third, the differences between the submodels (one model for each scenario) and general models (a single model for multiple scenarios) were compared. A comprehensive modeling framework was established.

4.1 Importance of numeric inputs

The PI (permutation importance) method was used to evaluate the importance of each numeric input. The process is described as follows. First, we prepared two identical test datasets (set X and set Y). The target parameter sequence was randomly sorted in set Y; the other parameters were left unsorted. We tested the model with sets X and Y and compared the errors of the two test sets. The PI value for a specific parameter can be regarded as a decrease in the model error from adding the parameter to the input feature combination. A larger PI value indicates greater parameter importance. Ranking the importance of each parameter can help future users efficiently collect thermal comfort data.

In this study, the RMSE was used to measure the model error, and the PI value was determined using Eq. (5). To ensure the reliability and stability of the results, the data for each parameter were randomly sorted five times. The average PI value was considered as the final permutation importance of the target parameter.

$$\mathrm{PI} = \mathrm{RMSE}_{\text{disturbed test set}} - \mathrm{RMSE}_{\text{undisturbed test set}} \tag{5}$$

Figure 5 shows the permutation importance of each numeric input. For the AC scenarios, the importance ranks in order from greatest to least are Ta, Tr, Clo, RH, Va, and Met, as shown in Figure 5(a). For the NV scenarios, the ranks from greatest to least are Ta, Tout, Tr, Clo, RH, Va, and Met, as shown in Figure 5(b). Tout was the second most important parameter in the NV scenarios. The parameter importance determined using current databases may change with addition of data samples.

The prediction accuracy for different input combinations was explored based on the importance of the inputs. Table 2 presents different input combinations for AC and NV scenarios. For the AC scenarios, the inputs included combination a (Ta) to combination f (Ta, Tr, Clo, RH, Va, Met). For the NV scenarios, the inputs included combination a (Ta) to combination g (Ta, Tout, Tr, Clo, RH, Va, Met).

Figure 6 shows the decrease in the MAE and RMSE of the TSV prediction with added numeric input parameters.

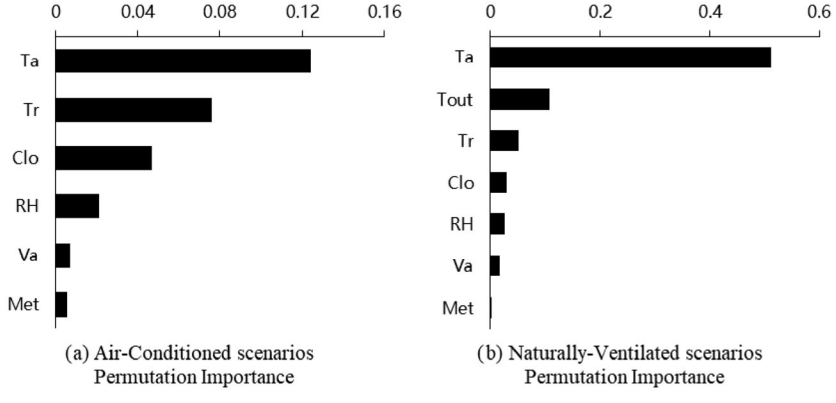


Fig. 5 Numeric input importance ranking: (a) AC scenarios; (b) NV scenarios

Table 2 Numeric input combinations for AC scenarios

Scenario	Sequence number	Input combinations
AC	a	Ta
	b	Ta, Tr
	c	Ta, Tr, Clo
	d	Ta, Tr, Clo, RH
	e	Ta, Tr, Clo, RH, Va
	f	Ta, Tr, Clo, RH, Va, Met
NV	a	Ta
	b	Ta, Tout
	c	Ta, Tout, Tr
	d	Ta, Tout, Tr, Clo
	e	Ta, Tout, Tr, Clo, RH
	f	Ta, Tout, Tr, Clo, RH, Va
	g	Ta, Tout, Tr, Clo, RH, Va, Met

Both the MAE and RMSE decreased gradually with addition of input parameters; the decreases did not exceed 6.38%. Adding input parameters can improve model accuracy; parameters with a higher importance rank contribute more to accuracy improvement.

4.2 Necessity of categorical inputs

In addition to numeric inputs, three categorical inputs (HVAC operation mode, HVAC operation season, and climate zone) were added to distinguish different indoor scenarios. To examine the necessity of these label inputs, the following tests were conducted.

4.2.1 Labeling AC and NV scenarios

To test model applicability in different scenarios, we used the NV model with NV data to predict AC scenarios, and then used the AC model with AC data to predict NV scenarios. Figure 7(a) shows the cross-testing process. The AC-A and NV-A models were derived from the AC-A scenario (air-conditioned, climate zone A, cooling season) and NV-A scenario (naturally ventilated, climate zone A), respectively. These models were tested using AC-A and NV-A data.

Figures 7(b) and 7(c) compare the AC-A and NV-A model performance in different scenarios. The predicted TSVs were compared with the actual TSV in each temperature bin. For the AC-A scenario, the AC-A model produced a much smaller SSE and much larger R^2 than the NV-A model, as shown in Figure 7(b). For the NV-A scenario, the NV-A model produced a much smaller SSE and much larger R^2 than the AC-A model, as shown in Figure 7(c).

Cross-tests between the AC and NV models were also conducted for climate zones B, C, and D. Table 3 summarizes the SSE and R^2 results. In predicting the TSV for AC scenarios, the AC models produced a much smaller SSE and a much larger R^2. A similar phenomenon was observed for TSV prediction in NV scenarios, indicating that labeling AC and NV scenarios is necessary in TSV prediction models.

4.2.2 Labeling climate zones

Cross-tests between different climate zones were conducted to evaluate the necessity of labeling them. For NV scenarios, NV-A, NV-B, NV-C, and NV-D models were developed using data samples from the NV-A, NV-B, NV-C, and NV-D scenarios, respectively. For AC scenarios, the cooling-A, cooling-B, cooling-C, cooling-D, heating-B, heating-C, and heating-D models were built using data from cooling-A (AC scenarios in climate zone A in cooling season), cooling-B, cooling-C, cooling-D, heating-B, heating-C, and heating-D scenarios, respectively.

Figure 8 compares the TSV prediction performance in NV scenarios and four climate zones. The NV-A model produced a smaller SSE and larger R^2 than the other three

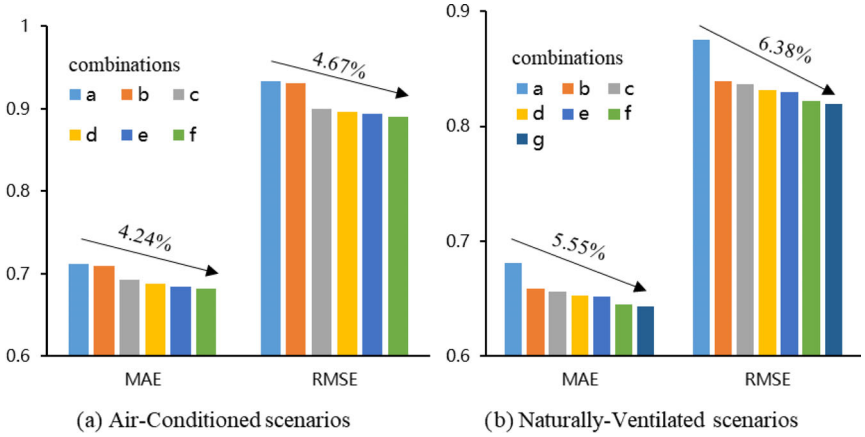


Fig. 6 Improving model by adding numeric inputs: (a) AC scenarios; (b) NV scenarios

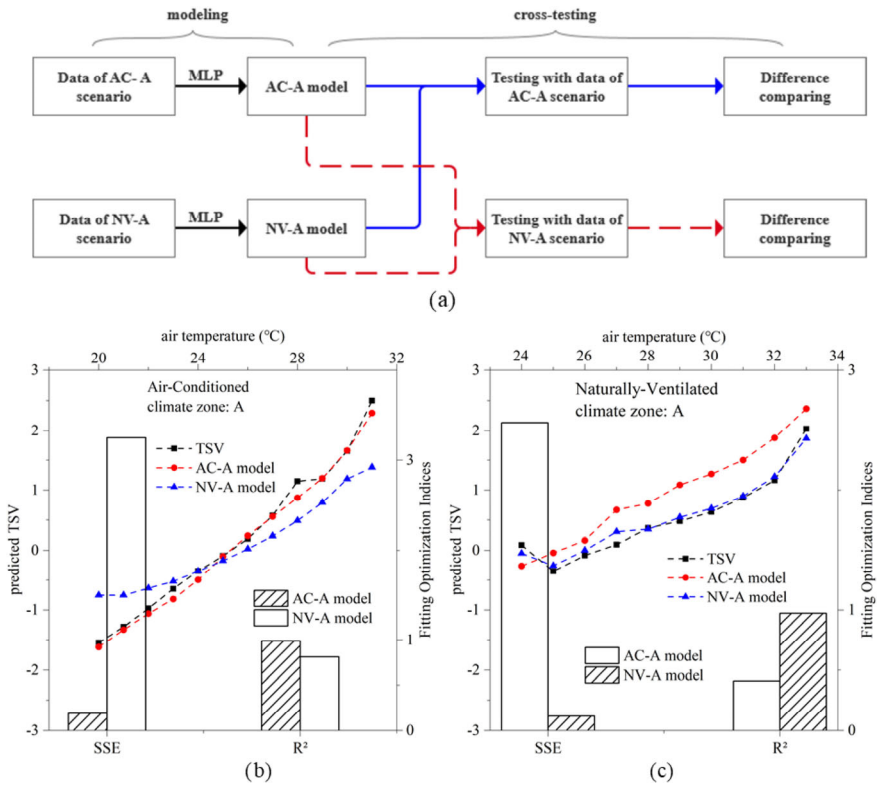


Fig. 7 Cross-testing on AC and NV labels: (a) cross-testing process; (b) TSV prediction performance in AC-A scenario; (c) TSV prediction performance in NV-A scenario

Table 3 SSE and R^2 for cross-tests between AC and NV models

Testing scenario	AC-A		AC-B		AC-C		AC-D	
Model	AC-A	NV-A	AC-B	NV-B	AC-B	NV-B	AC-D	NV-D
SSE	**0.189**	3.252	**0.272**	7.583	**0.259**	6.069	**0.354**	4.103
R^2	**0.989**	0.811	**0.991**	0.738	**0.989**	0.752	**0.976**	0.725
Testing scenario	NV-A		NV-B		NV-C		NV-D	
Model	AC-A	NV-A	AC-B	NV-B	AC-C	NV-C	AC-D	NV-D
SSE	2.567	**0.118**	17.93	**0.478**	13.381	**0.427**	5.935	**0.760**
R^2	0.410	**0.973**	0.273	**0.966**	0.469	**0.953**	0.779	**0.972**

models for climate zone A, as shown in Figure 8(a). The NV-B model produced a smaller SSE and larger R^2 than the other three models for climate zone B, as shown in Figure 8(b). Similar results are observed in Figure 8(c) and Figure 8(d). Thus, labeling climate zones can improve TSV prediction performance for NV scenarios. Table 4 summarizes the SSE and R^2 results from cross-tests between climate zones during the cooling and heating seasons for AC scenarios. For each climate zone, the corresponding model produced a smaller SSE and a larger R^2 than the other models, suggesting that labeling climate zones is also necessary for AC scenarios.

4.2.3 Labeling cooling and heating season

Cross-tests were conducted to compare differences between the cooling and heating seasons. Cooling-B, cooling-C, cooling-D, heating-B, heating-C, and heating-D models were built using the corresponding data samples. Only a cooling model was built for climate A because it represents a tropical climate (Table 1) and only includes cooling season data. The cooling season models were used to predict the heating season, and the heating season models were used to predict the cooling season. The prediction errors were compared and evaluated.

Figure 9 compares the TSV performance of the cooling-C and heating-C models in predicting cooling-C and heating-C scenarios. For the cooling-C scenario, the TSV predicted by the cooling-C model was more consistent with the actual TSV than the TSV predicted by the heating-C model, as

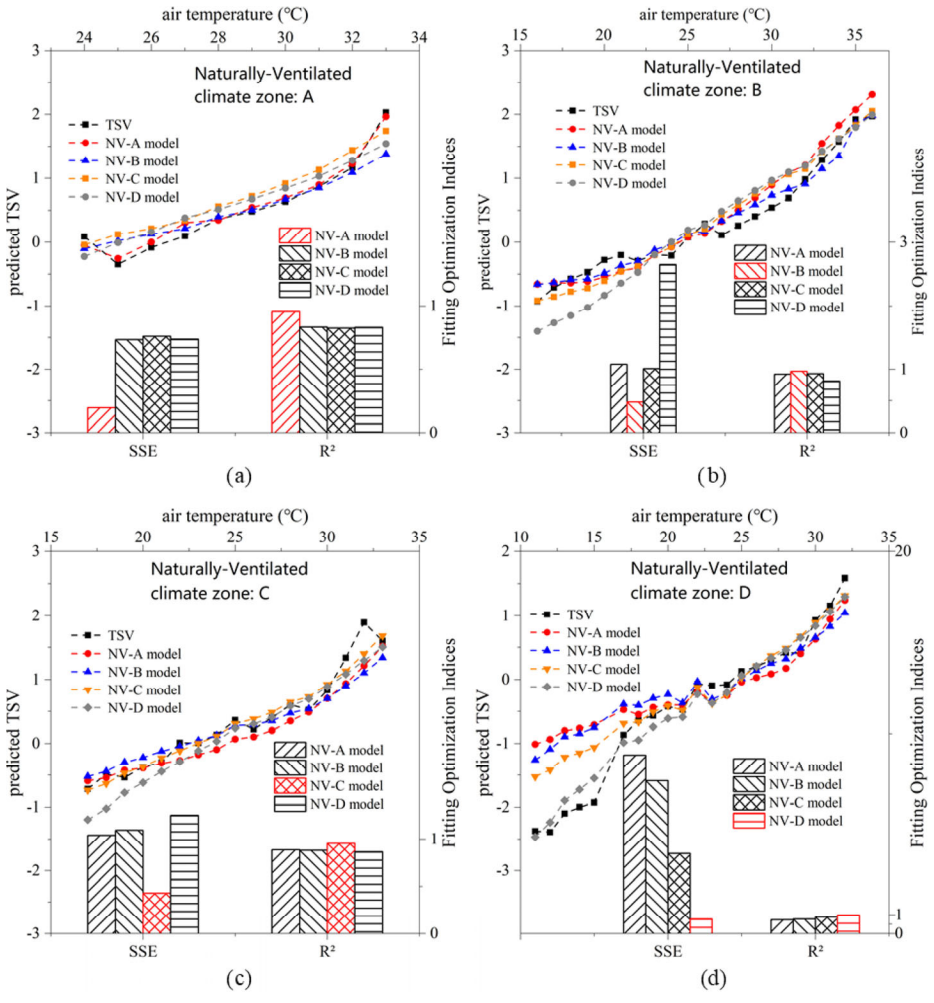


Fig. 8 Cross-testing of A, B, C, and D climate zones for NV scenarios: (a) predicting NV-A scenario; (b) predicting NV-B scenario; (c) predicting NV-C scenario; (d) predicting NV-A scenario

Table 4 Climate zone cross-tests in cooling and heating seasons

	Index	SSE	R^2	SSE	R^2	SSE	R^2	SSE	R^2
	Testing scenario	Cooling-A scenario		Cooling-B scenario		Cooling-C scenario		Cooling-D scenario	
Cooling season	Cooling A model	**0.189**	**0.989**	0.517	0.982	1.957	0.920	0.531	0.961
	Cooling B model	0.644	0.962	**0.272**	**0.992**	1.824	0.925	0.585	0.957
	Cooling C model	0.343	0.980	0.504	0.983	**0.259**	**0.989**	0.645	0.952
	Cooling D model	0.557	0.966	0.936	0.968	2.076	0.915	**0.187**	**0.986**
	Testing scenario	—		Heating-B scenario		Heating-C scenario		Heating-D scenario	
Heating season	Heating-B model			**0.091**	**0.950**	3.720	0.682	2.514	0.841
	Heating-C model	—		1.313	0.279	**0.142**	**0.988**	0.527	0.967
	Heating-D model			0.856	0.530	0.343	0.971	**0.281**	**0.982**

shown in Figure 9(a). The SSE of the cooling-C model was 96.5% smaller than that of the heating-C model. The R^2 of the cooling-C model reached 0.98, much larger than that of the heating-C model (0.70). For the heating-C scenario, the heating-C model also produced a much smaller SSE and larger R^2 than the cooling-C model, as shown in Figure 9(b), suggesting much better performance of the cooling-C model than the heating-C model in TSV prediction of the cooling-C scenario, and much better performance of the heating-C model than the cooling-C model in TSV prediction of the heating-C scenario. Similar results were also observed in climate zones B and D (Table 5).

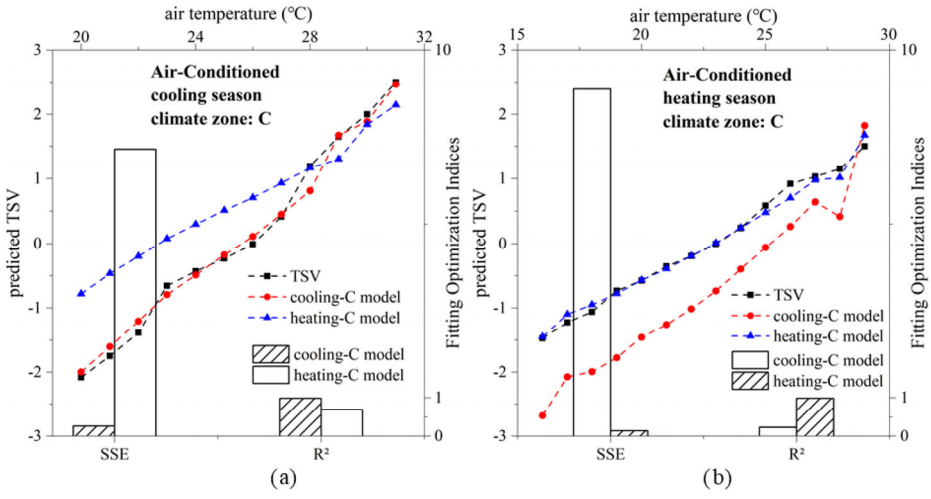


Fig. 9 Cross-testing of cooling and heating seasons for AC scenarios in climate zone C: (a) predicting cooling-C; (b) predicting heating-C

Table 5 SSE and R^2 for climate zone cross-tests in cooling season

Testing scenario	Cooling-B scenario		Heating-B scenario		Cooling-D scenario		Heating-D scenario	
Index	SSE	R^2	SSE	R^2	SSE	R^2	SSE	R^2
Cooling-B model	**0.272**	**0.991**	11.509	0.323				
Heating-B model	18.526	0.359	**0.091**	**0.950**		—		
Cooling-D model					**0.187**	**0.986**	5.041	0.682
Heating-D model			—		5.733	0.575	**0.281**	**0.982**

4.2.4 *Ranking necessity of different categories*

Through the cross-tests, we found that labeling AC and NV scenarios, different climate zones, and cooling and heating seasons can help improve the TSV prediction performance of the model. To address their relative importance, the absolute difference in SSE for each cross-test was compared between categories. For example, in Figure 7(b), the absolute difference between the SSE of the AC-A model and that of the NV-A model represents the difference between the AC-A and NV-A scenarios. In Figure 8(a), the absolute difference between the SSE of the NV-A model and the SSE of the other three NV models represents the difference between the different climate zones.

Figure 10 compares the differences between categories. The mean difference between AC and NV scenarios was 0.45, between the cooling season and heating season was 0.83, and between different climate zones was 0.08 and 0.11 for AC and NV scenarios, respectively. The difference between climate zones was larger for the NV scenarios than for the AC scenarios.

4.3 Submodels and general models

Labeling HVAC operation modes and climate zones can improve TSV prediction performance. There are usually

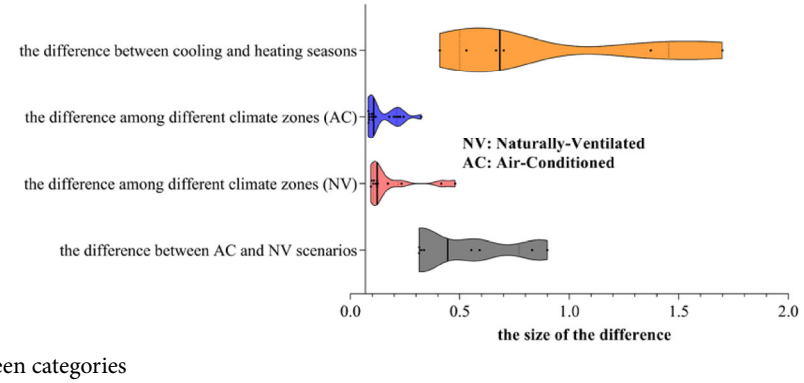


Fig. 10 Differences between categories

two approaches for establishing a comprehensive model. One is to build a general model containing all categories; the other is to build submodels for each scenario based on the corresponding data samples. To determine the best approach, the TSV prediction performance was analyzed for the submodels and the general model.

Figure 11 shows a comparison between the cooling-A model and the general model for the cooling-A scenario. Figure 11(a) shows the comparison process. The cooling-A model is based on the cooling-A scenario data; the general AC model with labels is based on all AC scenario data plus two categorical parameters (climate zone and HVAC operation season). An general AC model without labels was also developed for comparison.

Figure 11(b) shows a comparison of the TSV prediction performance of the cooling-A model and the two general AC models. The cooling-A model produces a smaller SSE and larger R^2 than the two general AC models. The general AC model with labels produces a smaller SSE and larger R^2 than the general model without labels. Similar comparisons were conducted for the other scenarios; the results are summarized in Tables 6 and 7.

For TSV prediction in AC scenarios, the SSE produced by submodels were 28.8%–79.2% lower than that of the general AC model with labels, and 82.9–96.4% lower than that of the general AC model without labels. For TSV prediction in NV scenarios, the SSE produced by submodels were 11.6%–53.9% lower than that of the general NV model with labels, and 39.1–63.9% lower than that of the general NV model without labels. These results suggest that submodels for specific scenarios achieve better TSV prediction performance than the general models.

4.4 Modeling framework

From the analysis, a comprehensive data-driven TSV prediction framework was established considering AC and NV scenarios, different climate zones, and cooling and heating seasons. Figure 12 shows the modeling framework with a series of submodels.

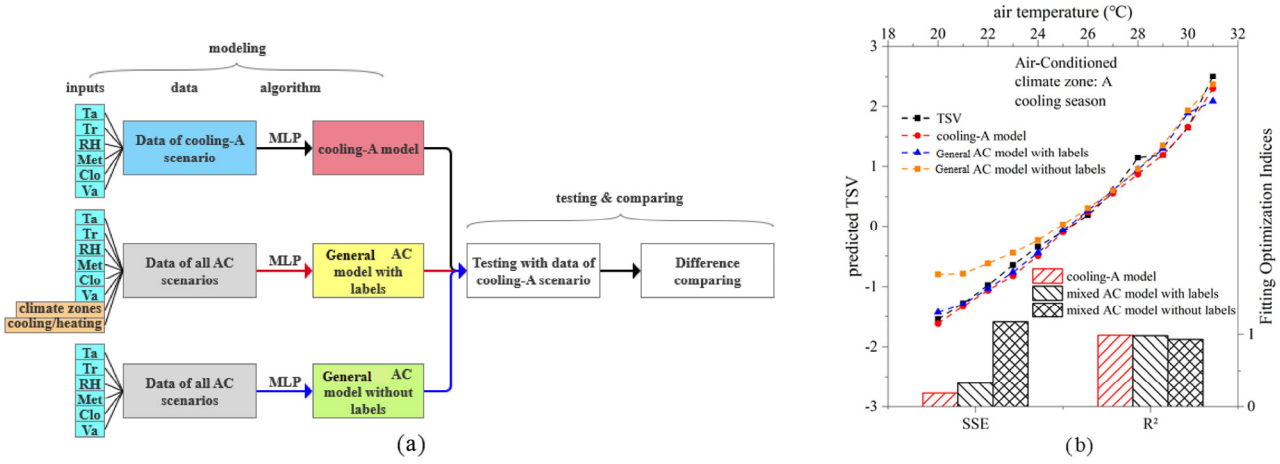


Fig. 11 Comparing submodels and general models: (a) comparison process; (b) TSV prediction performance for cooling-A scenario

Table 6 Comparison of SSE for AC sub-models and general AC models

AC submodel	Cooling-A	Cooling-B	Cooling-C	Cooling-D	Heating-B	Heating-C	Heating-D
Testing scenario	Cooling-A	Cooling-B	Cooling-C	Cooling-D	Heating-B	Heating-C	Heating-D
Relative improvement compared with general AC model with labels	43.0 ↓	28.8% ↓	79.2% ↓	48.6% ↓	56.1% ↓	69.4% ↓	78.1% ↓
Relative improvement compared with mixed AC model without labels	84.0% ↓	93.1% ↓	91.1% ↓	86.4% ↓	96.4% ↓	82.9% ↓	91.8% ↓

Table 7 Comparison of SSE for NV submodels and general NV models

NV submodel	NV-A	NV-B	NV-C	NV-D
Testing scenario	NV-A	NV-B	NV-C	NV-D
Relative improvement compared with general NV model with labels	53.9% ↓	11.6% ↓	15.2% ↓	21.5% ↓
Relative improvement compared with general NV model without labels	63.9% ↓	44.9% ↓	39.1% ↓	59.6% ↓

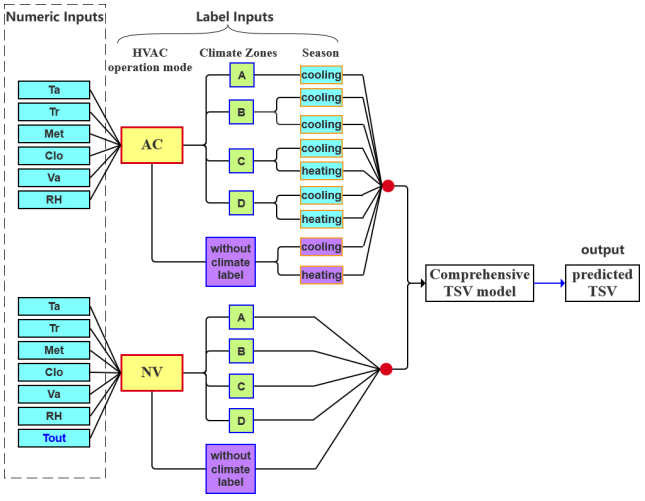


Fig. 12 TSV modeling framework for different sub-scenarios

For AC scenarios, the numeric input options included Ta, Tr, Met, Clo, Va, and RH. For NV scenarios, the numeric input options included Ta, Tr, Met, Clo, Va, RH, and Tout. Three options corresponding to the three label inputs (AC and NV scenarios, climate zones (A, B, C, or D), and the heating and cooling seasons) were also included. Given the relatively low importance of labeling climate zones (Figure 10), an option without climate labels was set as an alternative for cases with unknown climate zones. For AC scenarios, climate zone A only includes a cooling season; zones B, C, and D include both cooling and heating seasons. For NV scenarios, no cooling or heating season labels were used.

Eleven submodels corresponding to 11 scenarios, and three general models without climate zone labels were packaged to form the final comprehensive data-driven TSV prediction model. The final model can output a continuous TSV ranging from −3 to 3, and can adapt to different sub-scenarios. The comprehensive model has a 64.7%–90.7% lower prediction error than the conventional PMV model (reducing SSE by 3.2–4.9). The model in Figure 12 had an 11.6%–63.9% lower prediction error than the conventional data-driven model using all numeric and label inputs (reducing SSE by 0.58–3.2).

5 Model application

The data-driven TSV prediction model can be useful for evaluating thermal comfort in a building environment. One such application is to determine indoor thermal comfort zones for different climates to guide the design and operation of building environment conditioning systems.

For NV scenarios, the acceptable thermal temperature range with different outdoor air temperatures can be determined by the model without climate labels (Figure 12), with target TSV [−1, +1]. Considering Shanghai as an example, assuming that the other five parameters are fixed (Tr = Ta, RH = 60%, Met = 1.1met, Clo, and Va are related to Ta), the thermally neutral indoor temperature (TSV = 0) can be obtained (white line in Figure 13). Similarly, the upper and lower indoor temperature limits can be obtained (red and blue lines in Figure 13). With the monthly mean outdoor air temperatures for April, July, and October, the corresponding temperature limits can be obtained (dots in Figure 13). The indoor temperature limits for NV buildings can be determined in other cities in different climate zones using the same procedure (Table 8).

Based on the model framework in Figure 12, a prototype online tool was developed and will be continuously updated. To use the tool, it is not necessary to provide all numeric and input features shown in Figure 12. The tool outputs thermal sensation predictions based on whatever parameters are provided by the user (within the scope of inputs shown in Figure 12).

To test if the data-driven TSV model is extendible and generalized enough to predict partially unseen situations, we tried to develop models with international datasets while testing the model with the Chinese dataset. The results were shown in Figure A2 in the Appendix, which is in the ESM of the online version. For AC-cooling scenario, the model developed with two international datasets worked well on the Chinese dataset prediction. For NV scenario, models developed with international datasets showed large discrepancy when predicting the Chinese dataset. Adding half Chinese data into the training data could largely reduce the discrepancy. These might because the Chinese dataset has different climate classification criteria with the two international datasets. More analysis can be conducted to investigate this issue.

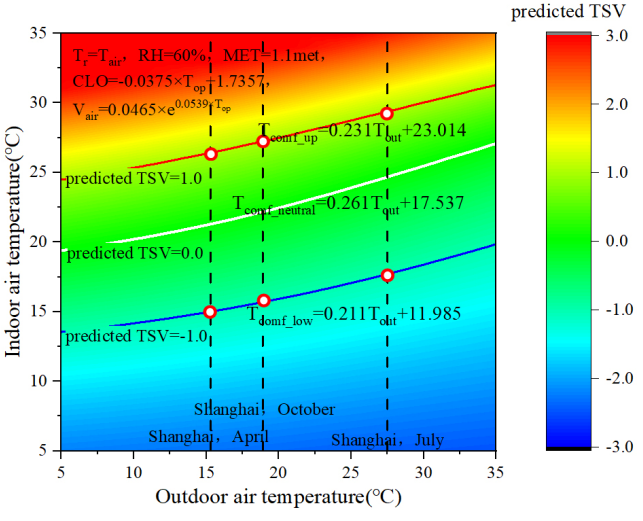


Fig. 13 Acceptable indoor temperature range in NV scenario for Shanghai

Table 8 Acceptable indoor temperature range for NV buildings in different cities

City	Temperature (°C)	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Beijing	Monthly mean outdoor air temperature			7.7	14.4	19.4	24.5	26.4	25.6	20.4	13	5.4	
	Upper thermal acceptable temperature			24.8	26.3	27.5	28.7	29.1	28.9	27.7	26.0	24.3	
	Lower thermal acceptable temperature			13.6	15.0	16.1	17.1	17.5	17.4	16.3	14.7	13.1	
Shanghai	monthly mean outdoor air temperature	4.5	6.3	9.9	15.3	20.6	24.3	27.5	27	24.4	18.9	13.6	7.4
	Upper thermal acceptable temperature	24.1	24.5	25.3	26.6	27.8	28.6	29.4	29.3	28.7	27.4	26.2	24.7
	Lower thermal acceptable temperature	12.9	13.3	14.1	15.2	16.3	17.1	17.8	17.7	17.1	16.0	14.8	13.5
Guangzhou	monthly mean outdoor air temperature	13.9	14.2	18.3	22.4	26.1	27.2	28.8	28	27.4	24.4	20.2	15.5
	Upper thermal acceptable temperature	26.2	26.3	27.2	28.2	29.1	29.3	29.7	29.5	29.4	28.7	27.7	26.6
	Lower thermal acceptable temperature	14.9	15.0	15.8	16.7	17.5	17.7	18.1	17.9	17.8	17.1	16.2	15.2
Nanjing	monthly mean outdoor air temperature			8.9	15.7	20.6	24.8	28.6	27.7	23.5	16.9	10.6	5.0
	Upper thermal acceptable temperature			25.1	26.6	27.8	28.7	29.6	29.4	28.5	26.9	25.5	24.2
	Lower thermal acceptable temperature			13.9	15.3	16.3	17.2	18.0	17.8	16.9	15.5	14.2	13.0

Note: when the mean outdoor temperature is less than 5 °C, the NV acceptable temperature range is no longer applicable where space heating may be needed.

6 Conclusions

In this study, a comprehensive data-driven thermal sensation prediction model was established using an MLP algorithm and three quality-controlled thermal comfort databases. The following observations were made during model development, improvement, and application.

With higher TSV prediction accuracy and more reasonable output results, the MLP algorithm was more suitable for TSV prediction than the other four tested algorithms (RF, SVM, KNN, GPR) in this specific case.

Labeling AC and NV scenarios, climate zones, and cooling and heating seasons can improve model performance. Submodels for specific scenarios can produce higher TSV prediction accuracy than general models with or without labels.

A data-driven TSV modeling framework was established using 11 submodels corresponding to 11 sub-scenarios, and three general models without climate zone labels. The new model can distinguish AC and NV scenarios, climate zones, and cooling and heating seasons, and reduce prediction errors by 64.7%–90.7% (reducing SSE by 3.2–4.9) compared with the conventional PMV model.

Electronic Supplementary Material (ESM): The Appendix of this paper is available in the online version at https://doi.org/10.1007/s12273-022-0911-2.

Acknowledgements

This study was supported by the National Natural Science Foundation of China (No. 52178087), the China National Key R&D Program during the 13th Five-year Plan Period (No. 2018YFC0704500), and the Fundamental Research Funds for the Central Universities (No. 22120210537). The authors would like to thank Guangdong Midea Air-Conditioning Equipment Co., Ltd. for their support.

References

ASHRAE (2017). ASHRAE Standard 55. Thermal Environmental Conditions for Human Occupancy. Atlata, GA, USA: American Society of Heating, Refrigerating and Air-Conditioning Engineers.

Cheung T, Schiavon S, Parkinson T, et al. (2019). Analysis of the accuracy on PMV–PPD model using the ASHRAE Global Thermal Comfort Database II. *Building and Environment*, 153: 205–217.

Cosma AC, Simha R (2019). Machine learning method for real-time non-invasive prediction of individual thermal preference in transient conditions. *Building and Environment*, 148: 372–383.

De Dear R (1998). A global database of thermal comfort field experiments. *ASHRAE Transaction*, 104: 1141–1152.

De Dear R, Schiller Brager G (2001). The adaptive model of thermal comfort and energy conservation in the built environment. *International Journal of Biometeorology*, 45: 100–108.

Elnaklah R, Fosas D, Natarajan S (2020). Indoor environment quality and work performance in "green" office buildings in the Middle East. *Building Simulation*, 13: 1043–1062.

Fanger PO (1970). Thermal Comfort. Copenhagen: Danish Technical Press.

Fanger PO, Toftum J (2002). Extension of the PMV model to non-air-conditioned buildings in warm climates. *Energy and Buildings*, 34: 533–536.

Földváry Ličina V, Cheung T, Zhang H, et al. (2018). Development of the ASHRAE global thermal comfort database II. *Building and Environment*, 142: 502–512.

Frontczak M, Schiavon S, Goins J, et al. (2012). Quantitative relationships between occupant satisfaction and satisfaction

aspects of indoor environmental quality and building design. *Indoor Air*, 22: 119–131.

Kim J, Zhou Y, Schiavon S, et al. (2018). Personal comfort models: Predicting individuals' thermal preference using occupant heating and cooling behavior and machine learning. *Building and Environment*, 129: 96–106.

Li W, Zhang J, Zhao T, et al. (2019). Experimental study of human thermal sensation estimation model in built environment based on the Takagi–Sugeno fuzzy model. *Building Simulation*, 12: 365–377.

Lin Y, Yang L, Luo M (2021). Physiological and subjective thermal responses to heat exposure in northern and southern Chinese people. *Building Simulation*, 14: 1619–1631.

Lu S, Wang W, Lin C, et al. (2019). Data-driven simulation of a thermal comfort-based temperature set-point control with ASHRAE RP884. *Building and Environment*, 156: 137–146.

Luo M, Wang Z, Brager G, et al. (2018a). Indoor climate experience, migration, and thermal comfort expectation in buildings. *Building and Environment*, 141: 262–272.

Luo M, Wang Z, Ke K, et al. (2018b). Human metabolic rate and thermal comfort in buildings: The problem and challenge. *Building and Environment*, 131: 44–52.

Luo M, Xie J, Yan Y, et al. (2020). Comparing machine learning algorithms in predicting thermal sensation using ASHRAE Comfort Database II. *Energy and Buildings*, 210: 109776.

Nicol JF, Humphreys MA (1973). Thermal comfort as part of a self-regulating system. *Building Research and Practice*, 1: 174–179.

Nicol F, Humphreys M (2010). Derivation of the adaptive equations for thermal comfort in free-running buildings in European standard EN15251. *Building and Environment*, 45: 11–17.

O'Brien W, Wagner A, Schweiker M, et al. (2020). Introducing IEA EBC Annex 79: Key challenges and opportunities in the field of occupant-centric building design and operation. *Building and Environment*, 178: 106738.

Peel MC, Finlayson BL, McMahon TA (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, 11: 1633–1644.

Schiavon S, Hoyt T, Piccioli A (2014). Web application for thermal comfort visualization and calculation according to ASHRAE Standard 55. *Building Simulation*, 7: 321–334.

Schweiker M (2016). Comf: An R package for thermal comfort studies. *The R Journal*, 8: 341.

Tang J, Liu Y, Du H, et al. (2021). The effects of portable cooling systems on thermal comfort and work performance in a hot environment. *Building Simulation*, 14: 1667–1683.

Tartarini F, Schiavon S (2020). Pythermalcomfort: A Python package for thermal comfort research. *SoftwareX*, 12: 100578.

Wagner A, Gossauer E, Moosmann C, et al. (2007). Thermal comfort and workplace occupant satisfaction—Results of field studies in German low energy office buildings. *Energy and Buildings*, 39: 758–769.

Witten IH, Frank E(2000). Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. San Francisco, CA, USA: Morgan Kaufmann Publishers.

Wu Z, Li N, Peng J, et al. (2018). Using an ensemble machine learning methodology-Bagging to predict occupants' thermal comfort in buildings. *Energy and Buildings*, 173: 117–127.

Wu Y, Liu H, Li B, et al. (2021). Individual thermal comfort prediction using classification tree model based on physiological parameters and thermal history in winter. *Building Simulation*, 14: 1651–1665.

Xie J, Li H, Li C, et al. (2020). Review on occupant-centric thermal comfort sensing, predicting, and controlling. *Energy and Buildings*, 226: 110392.

Yao R, Li B, Liu J (2009). A theoretical adaptive model of thermal comfort—Adaptive Predicted Mean Vote (aPMV). *Building and Environment*, 44: 2089–2096.

Zhang J, Zhou X, Lei S, et al. (2022). Energy and comfort performance of occupant-centric air conditioning strategy in office buildings with personal comfort devices. *Building Simulation*, 15: 899–911.

Zhou X, Liu Y, Luo M, et al. (2019). Thermal comfort under radiant asymmetries of floor cooling system in 2 h and 8 h exposure durations. *Energy and Buildings*, 188: 98–110.

Zhou X, Xu L, Zhang J, et al. (2020). Data-driven thermal comfort model via support vector machine algorithms: Insights from ASHRAE RP-884 database. *Energy and Buildings*, 211: 109795.

Zhu Y, Luo M, Ouyang Q, et al. (2015). Dynamic characteristics and comfort assessment of airflows in indoor environments: A review. *Building and Environment*, 91: 5–14.