Building and Environment 242 (2023) 110569

Contents lists available at ScienceDirect

Building and Environment

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

Towards inclusive underground public transportation: Gender differences on thermal comfort

Renlu Qiao [a,b], Zhiqaing Wu [a,b], Shuo Gao [c], Qingrui Jiang [a,b,*], Xiaochang Liu [a], Chenyu Huang [a], Li Xia [d], Mingze Chen [e]

[a] *Shanghai Research Institute for Intelligent Autonomous Systems, Tongji University, 1239 Siping Road, Shanghai, PR China*
[b] *College of Architecture and Urban Planning, Tongji University, 1239 Siping Road, Shanghai, PR China*
[c] *University of Oxford, 11a Mansfield Road, Oxford OX1 3SZ, UK*
[d] *School of Management, University of Science and Technology of China, Hefei, China*
[e] *The University of British Columbia, Department of Forest and Resources Management, 2424 Main Mall, Vancouver, BC V6T 1Z4, Canada*

A R T I C L E I N F O

Keywords:
Thermal comfort
Underground transportation space
Gender difference
Data augmentation
Gradient boosting tree
Nonlinear explanation

A B S T R A C T

As urbanization and societal needs grow, the demand for inclusive, efficient, and sustainable underground public transportation has increased, necessitating a greater focus on enhancing passengers' thermal comfort. However, limited research has addressed the subway station environment, and restricted datasets are challenging to build an accurate nonlinear model. Therefore, this study proposes a thermal comfort data generation algorithm based on 5161 measured data via the Variational Autoencoder (VAE) framework to break data collection constraints and increase thermal comfort model accuracy. By employing a gradient boosting algorithm on the extensively generated dataset, this study investigated factors influencing thermal comfort in subway stations, emphasizing individual gender differences and dynamic environmental variables. The VAE model enhanced the thermal comfort dataset, resulting in a 29% and 35% increase in male and female test set accuracy, respectively. The study suggests that males with a BMI over 27 were more sensitive to temperature. As for the gender-friendly environmental features, people feel most comfortable at temperatures between 24 and 29 °C and relative humidity levels between 20% and 70%. Additionally, the study reveals that a stable environment promotes enhanced thermal comfort for passengers. These insights contribute to developing sustainable, equitable, and user-friendly underground public transportation systems.

1. Introduction

1.1. Background

With the remarkable economic growth of cities, urban habitat is inevitably threatened by several issues, such as urban land scarcity and overpopulation. Significantly, urbanization is accelerating, and urban land resources are limited and facing immense challenges [1,2]. To alleviate the problem of urban land scarcity, governments have established comprehensive UUS development regulations and provided additional UUS over the past decades. There are various functions of underground space, such as storage, underground parking, underground public transportation, underground commercial area, etc. [3]. Among them, the subway system is the main kind of underground public transportation and urban rail transit system [4]. Until December 31, 2020, a total of 44 cities (excluding Hong Kong, Macao, and Taiwan) opened 233 urban rail transit lines, 4660 stations, 7545.5 km of operation, 25.28 million actual train trips, 17.59 billion passenger trips, 10.91 billion inbound trips [5].

However, these systems often pose significant challenges to passengers' thermal comfort due to the unique environmental conditions of underground spaces. Such challenges can negatively impact passengers' health, behavior, and quality of life. Therefore, addressing thermal comfort in underground spaces is an essential issue for sustainable cities and societies. This paper aims to investigate the influencing mechanism of the underground transportation system environment on passenger thermal comfort from individual gender differences, sustainable development, and machine learning perspectives. The paper presented a literature review of the current state-of-the-art research on the topic and proposed potential solutions to improve gender-friendly thermal comfort. The study's findings can inform policymakers and transportation

* Corresponding author. Qingrui Jiang, Shanghai Research Institute for Intelligent Autonomous Systems, Tongji University, 1239 Siping Road, Shanghai, PR China.
E-mail address: ucbqqj1@ucl.ac.uk (Q. Jiang).

https://doi.org/10.1016/j.buildenv.2023.110569
Received 28 April 2023; Received in revised form 23 June 2023; Accepted 26 June 2023
Available online 5 July 2023
0360-1323/

Abbreviations

UUS:	urban underground space
PMV	predicted mean vote
TEMP	temperature
RH	relative humidity
WV	wind velocity
CI	clothing insulation
BMI	body mass index
WVD	wind velocity difference
TempD	temperature difference
RHD	relative humidity difference
AI	artificial intelligence
MLP	multi-layer perceptron
SVM	support vector machine
VAE	variational autoencoder
AE	auto-encoder
ASHRAE	American Society of Heating, Refrigerating, and Air-Conditioning Engineers
GBDT	gradient boosting decision tree
GOSS	gradient-based one-Side sampling
EFB	exclusive feature bundling
BO	bayesian optimization
SHAP	Shapley additive explanation
MLM	machine learning model
GAN	generative adversarial network
ISO	International Organization for Standardization
LightGBM	light gradient boosting machine
ADASYN	Adaptive Synthetic
SMOTE	Synthetic Minority Over-sampling Technique
ELM	Extreme learning machine
SCN	Self-organizing competitive network
RF	Random Forest

system designers to create inclusive and sustainable transportation systems and infrastructure.

1.2. Related work

Research on thermal comfort has predominantly focused on laboratory experiments and above-ground buildings, such as offices, houses, and non-residential buildings [6]. However, with the global shortage of land resources and the proliferation of underground spaces, it has become increasingly crucial to conduct research on thermal comfort in this environment. In particular, subway spaces in underground areas require more attention, given that passengers' behavior is often dynamic, and they constantly interact with the surrounding physical environment [7]. The variability in spatial conditions can influence passengers' perception of the thermal environment, necessitating a more nuanced understanding of thermal comfort in these spaces. While the widely used PMV model only considers six parameters, including air TEMP, RH, WV, average radiation TEMP, CI, and metabolic rate [8], it lacks information about the dynamic changes in the environment, such as differences in air TEMP, WV, and humidity. Therefore, previous models are inadequate in accurately describing the thermal environment of underground spaces. It is necessary to incorporate more dynamic information about the environment to establish an accurate model of human thermal comfort in underground spaces and to promote the development of more comfortable and energy-efficient underground transportation to meet the needs of different stages.

The thermal comfort of underground spaces has become an increasingly important research topic in recent years due to the unique environmental characteristics of these spaces and the growing number of people using them [9,10]. A multitude of studies have been conducted to explore thermal comfort and the influencing factors in underground spaces. Among the key factors, air TEMP stands out as one of the most significant contributors to thermal comfort [11,12]. Extensive research has shown that the air TEMP in underground spaces is affected by various factors, including external air TEMP, ventilation rate, and forced air ventilation cabins [13]. Other key factors, such as RH and WV, also play an essential role in determining thermal comfort, as highlighted in previous research [14–16]. For instance, Li et al. conducted a study on the thermal comfort of underground space in China, concluding that air TEMP, RH, and WV significantly impacted people's perceived thermal comfort [17].

The thermal comfort of underground spaces is affected not only by environmental factors but also by human factors, which have been increasingly investigated in recent years. Various individual physiological and psychological factors, as well as behavioral factors, have been identified as having a considerable impact on the thermal comfort of underground spaces. For instance, age, gender, BMI, and health status are among the individual physiological factors that have been found to influence thermal comfort [18–20]. In addition, psychological factors like anxiety, stress, and mood, as well as behavioral factors such as CI, metabolic rate, and activity level, have been shown to affect thermal comfort [21,22]. For example, wearing heavy clothing or engaging in physical activity can increase metabolic rate, resulting in higher heat production and increased thermal discomfort [23,24]. Gender has been identified as a critical personal factor that significantly affects thermal comfort. Studies have demonstrated notable gender differences in thermal comfort, with females generally experiencing lower comfort than males in the same underground environment [25,26]. For instance, a study conducted in Qatar, India, and Japan found that females were more likely to feel uncomfortable in air-conditioned offices due to higher humidity and air TEMP [27]. Another study investigated gender differences in thermal comfort in underground shopping malls and found that females had lower levels of perceived thermal comfort than men, which the authors attributed to physiological differences, such as body size and composition, and differences in clothing choices between genders [9]. These studies underscore the importance of considering gender differences when evaluating thermal comfort in underground spaces. Designers and engineers must account for these gender differences when designing underground spaces to ensure optimal thermal comfort for all users.

Several studies have demonstrated the potential of AI in thermal comfort research, with promising results. For instance, Asma et al. employed a SVM algorithm to predict thermal sensations by utilizing a variety of environmental factors such as air TEMP, RH, and Age [28]. Similarly, Badr et al. used neural networks to model the effects of indoor environmental factors on thermal comfort in office buildings [29]. In addition, the use of AI has enabled the development of personalized thermal comfort models, which consider individual factors such as age, gender, and metabolic rate, leading to more accurate predictions [30,31]. Despite these achievements, some challenges exist in the use of AI in thermal comfort studies. One of the critical challenges is the need for large amounts of data to train and validate AI models effectively [32]. Previous data for thermal comfort was only available in limited quantities, ranging from hundreds to thousands of data, which makes it challenging to train high-precision predictive models. This limited data adversely affects the accuracy and generalization ability of the models. However, recent research has addressed this issue from a data enhancement perspective through VAE and GANs framework [33,34]. In Qiao et al.'s study, a novel variational autoencoder-based model for subsurface space data generation was proposed. This model maps thermal comfort data from the subsurface space to a highly compressed potential layer space and generates data in an unsupervised manner. Through this model, the prediction accuracy was improved by 41.34%–45.31%, demonstrating the potential of data enhancement methods to improve the accuracy of AI models.

1.3. Challenges

Despite extensive research on thermal comfort in underground spaces, the field remains riddled with challenges that require further attention. Initially, the thermal comfort datasets employed in previous studies were sparse, consisting of only hundreds or thousands of data, rendering them insufficient to support the training of high-precision prediction models. To address this inadequacy, we utilized the VAE method to augment the data and create a more robust dataset for subsequent model development, thereby complementing the existing data deficit. Moreover, linear regression models are commonly used in previous research, which exhibit limited accuracy and model capability as they are constrained to linear functions and fail to consider local variations across different value domains. To address these two limitations of traditional linear regression, our study employs an ensemble algorithm for nonlinear fitting and effectively improves the model's accuracy. Then, we use an interpretable algorithm to explore the intricate relationship among variables across various value domains. Thirdly, the conventional static metrics used to describe thermal comfort fail to provide accurate insights, as thermal comfort is highly variable and is dependent on environmental conditions and human behavior. To capture the dynamic changes in thermal comfort perception, we have incorporated dynamic environmental indicators such as WVD, TempD, and RHD in our study. Finally, traditional PMV models are limited, as they do not consider individual differences in physiological and behavioral factors that significantly affect thermal comfort. To address this, our study highlights the significance of physiological differences between males and females, paving the way for more inclusive underground environments for public transportation. Therefore, it is imperative to explore and address the above-mentioned challenges to advance our knowledge and understanding of thermal comfort in underground spaces and develop more effective and efficient solutions to enhance passenger comfort and improve the operation of sustainable and inclusive transportation systems.

2. Data

2.1. Study area

We carried out two-period field surveys from June to July 2016 and from June to July 2017 to conduct a comprehensive investigation into the thermal environment of Beijing Metro. Beijing, the capital city of China, has a typical north-temperate semi-humid continental monsoon climate, with average air TEMPs ranging from 26.3 °C during the summer months. To ensure the representativeness of the study, we selected the Nanluoguxiang and Beitucheng stations of Beijing Metro Line 8, taking into account the station structure and pedestrian flow. Beijing Metro Line 8 is known as the busiest bus line in Beijing's metro network, with a daily ridership of over 900,000 passengers, making it one of the busiest metro lines in the world. The line, which runs from the northwest to the southeast of the city, is also the longest fully automated subway line in the world, stretching 62.5 km and having 18 stations [35]. The inspection locations included the outside area, exit gate, ticket counter, security checkpoint, platform, and subway carriage of Nanluoguxiang station, as well as the security checkpoint, ticket counter, and outside area of Beitucheng Station. This study considers the stations' characteristics, aiming to provide a comprehensive analysis of the thermal environment and the thermal comfort of passengers in Beijing Metro.

2.2. Original thermal comfort data collection

This study employed a combination of objective measurements and subjective surveys to comprehensively investigate the thermal comfort experience of participants throughout their entire journey within the subway system. To minimize the potential bias caused by participants becoming overly familiar or fatigued due to repeated testing, the experimental sessions were spaced one week apart, with each session lasting approximately 30 min. The questionnaire administered during the study covered various aspects including gender, age, height, weight, clothing, and thermal comfort perceptions [33].

2.3. Individual characteristics measurement

This study focused on young commuters aged between 18 and 30 years, as this age group constitutes a significant portion, accounting for approximately 40% of the total number of Beijing subway passengers [7]. Therefore, the age distribution of the population selected for this study is representative. The physical attributes of the participants, such as their height, weight, and body mass index, were measured and recorded for further analysis. Additionally, to facilitate the determination of their thermal resistance, the participants were instructed to don weather-appropriate clothing suited to the current outdoor weather conditions. The thermal resistance of the clothing was gauged according to the guidelines of the ASHRA [36].

2.4. Thermal environmental characteristics measurement

In order to gain a precise understanding of the environmental characteristics inherent to underground spaces, this study utilized advanced instrumentation. This includes the MASTECH MS6508 digital temperature and humidity meter, with a precision of ± 0.1 °C (within the range of 0–45 °C) and $\pm 3\%$, and the BENETECH GM8903 anemometer, boasting a precision of ± 0.1 m/s. These instruments were employed to measure physical characteristics such as air temperature, RH, and WV within the vicinity of the underground space of Line 8. As riding the subway is a dynamic process, where the physical environment is influenced by the behavior of passengers, this study took into account the entire process of a passenger's subway ride from Nanluoguxiang station to Beitucheng station as a round of experimental sampling. In contrast to other environments characterized by static activities, such as office settings that typically involve seated postures, the experience of commuting on a subway entails a dynamic process. Recognizing this dynamic process, we have endeavored to provide a comprehensive description of the underground environment by collecting data on three key dimensions TempD, RHD, and WVD. Furthermore, this study also investigated passengers' perceptions of the thermal environment during the subway ride to evaluate the thermal comfort of the underground space. Since this study attempts to discover a nonlinear relationship between the underground space environment and passenger thermal comfort, we abstract it into an empirical regression model. To avoid confounding the results, we used a five-point continuous scale conforming to ISO Standard 10,551 [37] to evaluate thermal comfort (Table 1).

2.5. Data summary

Thermal comfort is crucial in enhancing environmental quality, reducing energy consumption, and promoting productivity.

Table 1
Evaluation of thermal comfort.

Comfortable	Slightly comfortable	Uncomfortable	Extremely uncomfortable	unbearable
0	1	2	3	4

Nevertheless, the limitations of objective conditions make it challenging to record dynamic environmental parameters accurately and extensively collect passengers' physiological and behavioral responses, leading to restrictions in the amount of data available. This scarcity of data poses a significant challenge in developing accurate predictive models for thermal comfort in underground spaces. This study tackles this concern by establishing a comprehensive, 10-dimensional thermal comfort dataset. The dataset incorporates both environmental and personal data collected from underground spaces and involved 60 experimental participants (Table 2). This dataset can provide a solid database for training subsequent data generation models.

3. Methods

3.1. Data enhancement

Thermal comfort represents the citizens' thermal perception of the surrounding environment and results from a comprehensive evaluation of the influence of environmental factors and individual behavior. The research approaches in the thermal comfort field experienced a transformation from statistical models to AI algorithms, acquiring a more accurate analysis result. The complex data-driven AI algorithm needs a considerable amount of data, fitting to the best condition. However, due to some realistic limitations, it is difficult for researchers to collect sufficient human perception data. Therefore, the study employed a generative machine learning method to enhance thermal comfort data.

Several studies revealed that human thermal comfort is not a random perception result that strongly connects with environmental and individual characteristics and follows probability distributions in the data latent subspace [33]. Consequently, we constructed a thermal comfort data enhancement model based on a VAE framework to approximate the joint probability distribution of the thermal environment through variational inference [55]. VAE is a deep neural network framework from an AE. The AE differs from the usual progressively expanding deep neural network structures in that it has a bottleneck in the middle of the network structure, and the quantity of input and output dimensions is the same. As for the AE, the latent subspace is discrete. Thus, sampling from this subspace does not produce new results. The VAE attempts to learn the actual data probability distribution in the latent subspace, changing the discrete result to the continuous distribution.

Table 2
Data summary of the thermal comfort of underground space.

	Variable	Description	Data size	Reference
Independent Variables	Gender	Passenger gender distribution: male (53%), female (47%).	5161	[27,30,38]
	Age	Passengers' age ranges from 18 to 30 years.	5161	[19,28,39]
	BMI	Body Mass Index ranges from 14 to 36.	5161	[20,27,40]
	Clothing insulation	Clothing thermal insulation ranges from 0.3 to 0.7.	5161	[29,41,42]
	Temperature	The air temperature per space.	5161	[28,30,33]
	Temperature difference	Temperature variance from prior space reading.	5161	[42–44]
	Humidity	The relative humidity of per space during ride (±0.1% (0–90%); ±2%)	5161	[45–47]
	Humidity difference	The difference in space humidity from the previous measurement.	5161	[48,49]
	Wind velocity	The air velocity per space	5161	[50–52]
	Wind velocity difference	Wind speed change from prior space measurement.	5161	[51,53,54]
Dependent Variable	Thermal comfort	The thermal comfort of passengers.	5161	

The VAE framework contains three parts encoder, latent subspace, and decoder. The encoder is responsible for mapping the input data (x) to the latent subspace (z).

$$z \approx Enc(x) = q(z|x) \tag{1-1}$$

The decoder remaps z in the latent subspace back to $\hat{x}$, which is the process of decoding.

$$x \approx \hat{x} = Dnc(z) = p(x|z) \tag{1-2}$$

With the versatility and development of neural networks, the encoder or decoder part can utilize any algorithmic architecture. We capitalized MLP to model the encoder and decoder parts. Current research indicates that normal distribution can characterize latent subspace distributions and does not decrease the quality of generative results by VAE [56]. The latent subspace distribution z contains the information of x and is in line with a Gaussian distribution:

$$p(z) \approx Enc(x) = q(z|x) \sim N(\mu, \sigma^2) \tag{1-3}$$

$p(z)$ is the prior Gaussian distribution. The loss function of the VAE is different from that of the AE, which has two components. The first one is a general reconstruction loss, which is the difference between the input data set x and the VAE output dataset $\hat{x}$. This paper uses the mean square error (MSE) to construct this component:

$$L_{v1} = \frac{1}{n}\sum_{i=1}^{n}(x - \hat{x})^2 \tag{1-4}$$

The second component is the distribution loss, which is the difference between the Gaussian distribution $p(z)$ and the latent distribution $q(z|x)$. This study applied relative entropy, known as Kullback–Leibler divergence, to construct this loss function component. The information entropy between the two probability distributions uses the divergence to calculate the difference and fits the information loss of the actual distribution obtained using the latent subspace distribution:

$$L_{v2} = D_{kl}(q(z|x)\|p(z)) \tag{1-5}$$

$$L_v = L_{v1} + L_{v2} = \alpha\frac{1}{n}\sum_{i=1}^{n}(x - \hat{x})^2 + (1-\alpha)D_{kl}(q(z|x)\|p(z)) \tag{1-6}$$

where α is a hyperparameter of the loss function that adjusts the weights of the two parts of the loss and takes values in the range [0,1].

3.2. Thermal comfort modeling

Machin learning is a segmentation in the AI field. In recent years, the thermal comfort field has witnessed a high-speed improvement and outstanding advantage of machine learning. These algorithms can accurately dig the pattern between thermal comfort and environmental characteristics via iterating and adequate fitting data. For instance, Chai et al. employed the machine learning algorithm ANN to predict thermal comfort, and their findings demonstrated that the predictions obtained from the ANN model were significantly closer to the actual values when compared to traditional models [57]. Despite the subjective nature of people's perception of the thermal environment, basic machine learning models can still provide some support for modeling thermal comfort [28,58]. However, to further improve the accuracy and robustness of predictions, ensemble learning algorithms have emerged [59]. This is because ensemble learning algorithms combine numerous basic learners to construct a more remarkable learner and establish specific strategies to integrate the various fitting results from basic learners. In the field of thermal comfort research, several studies have focused on identifying the most suitable model by comparing different algorithms based on thermal comfort data [60]. For example, Feng et al. investigated using the ASHRAE RP-884 dataset, comparing the accuracy of individual models such as ELM, SCN, RF, SVR, and ensemble methods. The findings revealed that the integration of two or more algorithms achieved higher overall accuracy, with an RMSE as low as 0.157 [61]. These results demonstrate the potential of ensemble learning in improving the performance of thermal comfort models, underscoring the significance of adopting ensemble methods in this domain.

GBDT is a classical ensemble learning framework. In contrast with the random forest, GBDT utilizes the boosting method, i.e., training the model by transmitting the iterative residuals of decision trees, obtaining a more robust result. As the volume and dimensionality of data increased, researchers developed the LightGBM configuration based on GBDT. There are three key improvements in the LightGBM: GOSS, EFB, and Histogram [62]. In the training model process, the gradient can reflect the weight of the sample. GOSS uses this information to sample the samples, reducing a large number of samples with small gradients and focusing only on samples with high gradients in the following calculations. LightGBM resolves the problem of sparse high-dimensional features via EFB, which fuses and bounds some features, thus reducing the dimension of features. Meanwhile, the Histogram algorithm transforms continuous data split into discrete data, causing the regulation effect to increase the model's generalization. Due to these advantages, we selected the ensemble learning framework, LightGBM, for building a decoupling model for this thermal comfort research.

3.3. Thermal comfort model optimization

Although the MLM can spontaneously adjust parameters such as weights, split points, etc., through an iterative training procedure, the model requires pre-determined hyperparameters before the training. However, the LightGBM framework contains substantial hyperparameters while providing powerful capabilities, leading to a vast hyperparameter search space. In addition, due to the above thermal comfort data enhancement, each optimization search will consume considerable computing power. Accordingly, the study abandoned manual optimization as well as Grid Search and used BO to result in efficient global optimization of the expensive hyperparameter selection. BO estimates the hyperparameter objective function based on a finite number of observed experiments via a surrogate probability model and then employs the probability of improvement to determine the next observed experiment, optimizing the best combination of hyperparameters [63]. The study combinates the BO algorithm and the LightGBM framework to develop the BO-LightGBM model for challenges in thermal comfort modeling.

3.4. Collinearity testing of features

Before modeling thermal comfort, the variables in the model should be tested for the presence of collinearity. Collinearity means that the presence of a high correlation between the explanatory variables in the regression model distorts the model estimates. This study is a nonlinear fit of the thermal comfort patterns by the ensemble learning algorithm. The covariance matrix of variables is calculated using Spearman correlation coefficients. The formula is as follows:

$$\rho_{Spearman} = \frac{\sum_{i=1}^{N}(R_i - \overline{R})(S_i - \overline{S})}{\sqrt{\sum_{i=1}^{N}(R_i - \overline{R})^2 \sum_{i=1}^{N}(S_i - \overline{S})^2}} \tag{1-7}$$

where R_i and S_i are the ranks of sample i at features x and y, respectively; $\overline{R}$ and $\overline{S}$ are the average ranks of features x and y, respectively. The Spearman correlation coefficient can range from -1 to $+1$. The two features are highly correlated when the Spearman correlation coefficient exceeds 0.7.

3.5. Thermal comfort driver interpretation

The famous black-box problem in the AI field makes machine-learning models lose the same interpretability as linear statistical models. The ensemble learning model statistics the number of times features were used to estimate their importance. However, the method only provides global significance, which does not mention the nonlinear advantage of the MLM. Hence, We adopted an explainable approach called SHAP for an accurate nonlinear interpretation of thermal comfort drivers. The SHAP algorithm is based on the Shapley value of cooperative game theory, which estimates the feature's significance by calculating the marginal contribution of characteristic observations when they are added to the model. The calculation is as follows:

$$SHAP_j = \sum_{S \subseteq \{V_1 + V_2 + \cdots + V_p\} \setminus \{V_j\}} \frac{|S|!(p - |S| - 1)!}{p!} \left(f_x\left(S \cup \{V_j\}\right) - f_x(S)\right) \tag{1-8}$$

$$y_i = y_{base} + \sum_{j=1}^{k} SHAP\left(x_{i\_j}\right) \tag{1-9}$$

In the calculation, $SHAP_j$ is the SHAP value of the feature j; S is the subset of features used in the model; V_p is the feature of the model; p is the number of features; $f_x(S)$ is the prediction of the model at the subset; y_i is the predictive value of the model at sample i; y_{base} is the mean value of the predictive value at other samples; $SHAP(x_{i\_j})$ is the SHAP value of the feature j at sample i; k is the number of features.

4. Experiment and results

All code is written in Python with the help of the TensorFlow library. The experiments were run on a platform built with an Intel Core i9-13900KF CPU and an NVIDIA GeForce RTX 4090 GPU.

4.1. Correlation of thermal comfort variables

To study residents' thermal sentiment in the underground space, we collected individual and environmental characteristics as the thermal comfort model's variables, including nine independent variables and a dependent variable. In an attempt to make sure the validity of the model regressional result, the study calculated the correlation of these independent variables in female and male groups, respectively.

Fig. 1 demonstrates that there is no significant positive correlation between every two variables in the male group. RH and CI and RHD and TempD possess a negative correlation, but it is extremely slight. As for the female group, there is also a weak negative correlation between RHD and TempD. BMI and Age and WVD and WV have a comparatively significant positive correlation, but all of them are less than 0.7. In general, the absence of a strong correlation in these variables indicates that there is no severe collinearity problem in the independent variables, and all the above independent variables can participate in the thermal comfort model training.

4.2. Thermal comfort data enhancement

The thermal comfort dataset contains 5161 samples with ten dimensions, and these nine independent variables do not possess a significant correlation in the collinearity test. Due to the high-dimensional dataset having difficulty meeting the model's iterative needs, we adopted the VAE framework to generate new data following the distribution of the observed data. Fig. 2 exhibits the structure of the VAE model. Both the input and output layers contain ten neurons, corresponding to the 10-dimensional thermal comfort data. The study employed a 2-layer MLP to establish the encoder and decoder, and per layer includes 100 neurons to excavate the latent subspace distribution accurately. In the latent subspace, we set its dimensions to three, and the prior distribution is a normal distribution with a mean of 0 and a

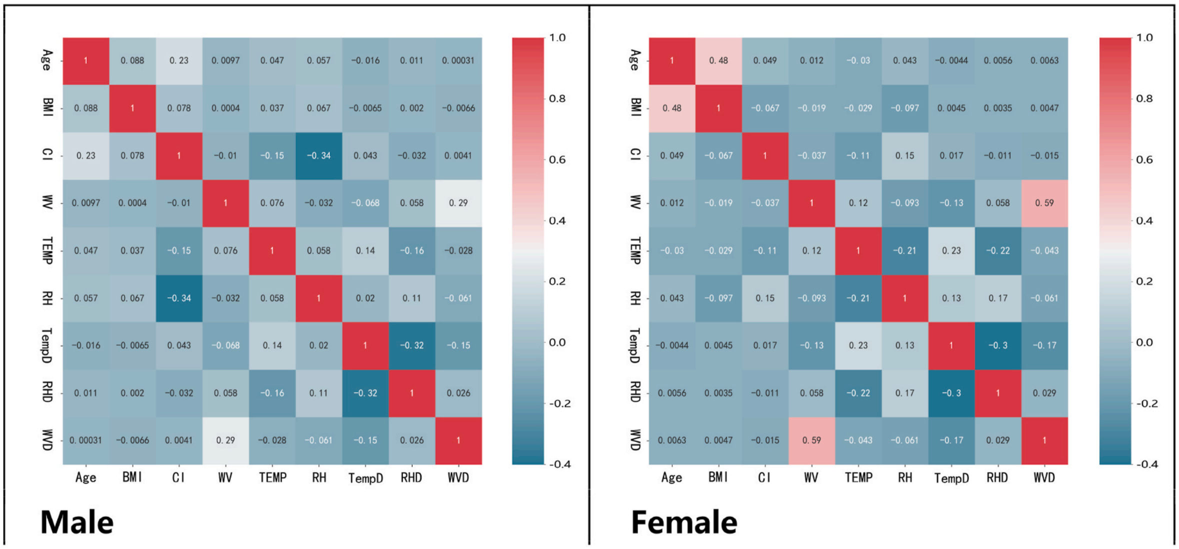

Fig. 1. Correlation of independent variables in female and male groups.

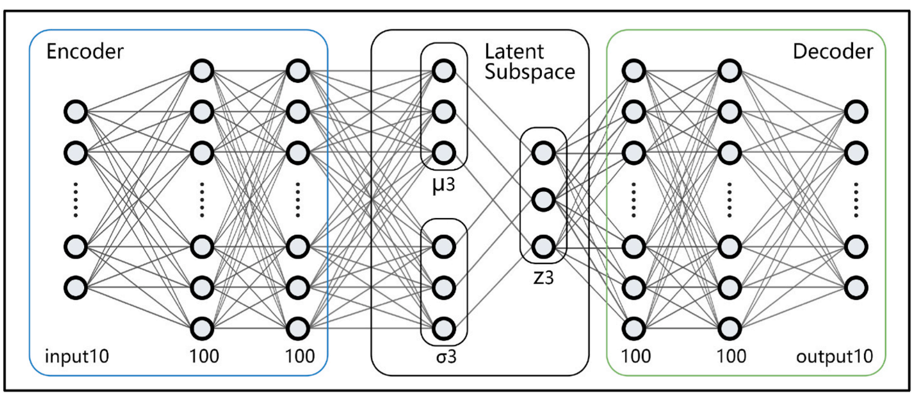

Fig. 2. The structure of the VAE model.

standard deviation of 0.1.

Before the VAE model training process, we normalize the data, transforming it to [0,1] to increase the stability of model iterations. In the encoder, latent subspace, and the first two layers of the decoder, the activation function is the ReLU function to avoid the issue of disappearing gradients. The Sigmoid function is used as the activation function of the output layer, making the model-generated result match the value range of the input data. We set the learning rate at 0.01, the batch size at 2000, and the training period at 5000. We chose Adam as the VAE model iteration algorithm and divided the thermal comfort dataset into 80% and 20% as the training and test sets. For more specific information on model construction, please refer to our previous research [1].

The study expanded the dataset via the VAE method. Qiao et al. demonstrated that 90% of the normal distribution sampling interval is the most suitable for the thermal comfort dataset [1]. Finally, we generated 100,000 thermal comfort data, including 50,000 female and 50,000 male perceptual data, to meet the data requirement of the thermal comfort decoupling model.

4.3. *Thermal comfort decoupling model*

For the purpose of decoupling the driver between environmental and individual characteristics and thermal comfort, we established the BO-LightGBM model through the generated thermal comfort data. The study built two BO-LightGBM models focused on the female and male groups, respectively, exploring the gender differences in thermal comfort. We fuse the data generated by VAE with the VAE training set, and the test set of VAE is used as the test set of BO-LightGBM models, thus ensuring the validity of the decoupling model evaluation. Consequently, the training and test sets in the male group include 52,623 data and 656 data, respectively. As for the female group, the training and test sets include 51,506 data and 376 data, respectively.

Evaluating the quality of non-image-generated data poses a challenge due to the difficulty in visual discernment. Consequently, scholars often assess the quality of generated data by examining the difference in model quality pre- and post-generation, a measure referred to as machine learning efficacy. Table 3 provides a comparative evaluation of the VAE and two prevalent oversampling methods, namely, SMOTE [64] and ADASYN [65] sampling approach, based on machine learning efficacy. Evidently, the VAE model demonstrates superior generation

Table 3
Machine learning efficacy of SMOTE, ADASYN, and VAE.

	Machine learning efficacy (R-square)			
	Male training set	Male test set	Female training set	Female test set
Baseline	0.82	0.62	0.81	0.58
SMOTE	0.96	0.83	0.97	0.82
ADASYN	0.94	0.84	0.94	0.85
VAE	**0.98**	**0.91**	**0.99**	**0.93**

quality. More specifically, for male thermal comfort, the R-square of the model based on the original data in training and test sets are 0.82 and 0.62, and based on the enhanced data by VAE, they are 0.98 and 0.91. As for the female group, the R-square of the model (original data) is 0.81 and 0.58, and that of the model (VAE enhanced data) is 0.99 and 0.93 from the training and test sets. On balance, the enhanced data contributes to a nearly 30% improvement in accuracy in the test set, which proves the effectiveness of the VAE data enhancement model we architected.

Fig. 3 demonstrates BO-LightGBM models' predicted capacity in test sets, showing that the generalization of the BO-LightGBM trained on the enhanced data significantly outperformed that trained on the original data.

From Fig. 3, we can obtain the distribution of predicted data. The forecast precision of per sample interval possesses a considerable variation, especially the predicted values that have significant deviations from the baseline ($y = x$) in 'Extremely uncomfortable' and 'Unbearable' samples. Apparently, our BO-LightGBM model consistently provides better predictive performance in arbitrary sample intervals, and the female group's model performance is slightly better than the male group's.

4.4. Thermal comfort model global interpretation

In the study, we adopted the BO-LightGBM to precisely establish the nonlinear relationship between the environmental and individual features and thermal comfort. Subsequently, the SHAP algorithm was implemented to essentially interpret the relationship found by the BO-LightGBM, which facilitates machine learning to acquire the same transparency as the classic statistical model, developing more effective environmental regulation strategies.

Fig. 4 shows the SHAP summary plots that describe the global effect and importance of environmental and individual variables on thermal comfort. The importance rank of environmental features is identical in the male and female groups. The air TEMP feature is the most critical variable in the two groups, which exceeds 20% impact on thermal comfort. The TempD feature is the variable second in importance to air TEMP in both groups of males and females, 16.32% and 13.60%, respectively. The influence of variables WV and WVD regarding the air flow rate is the weakest.

On the other hand, the sensitivity of thermal comfort to individual variables varies dramatically between males and females. In particular, The importance of the impact on thermal comfort is ranked from greatest to least in the male group: BMI, CI, and Age. However, as for females, the Age feature is the most crucial independent variable by up to 14.21%, surpassing the variable TempD.

4.5. Thermal comfort driver effect

The SHAP algorithm calculates the influence of the independent variable on thermal comfort at each observation sample. The attribution relationship between the independent variable and dependent variable, thermal comfort, includes the interaction influence of the other independent variables. As a result, the study must exclude the interaction among the independent variables, acquiring an unbiased effect. The calculation is as follows:

$$SHAP\left(X_{i\_jj}\right) = SHAP\left(X_{i\_j}\right) - \sum_{j \neq z}^{n} SHAP\left(X_{i\_jz}\right) \tag{2-1}$$

In this calculation, $SHAP(X_{i\_jj})$ is the eliminated-interactive-impact SHAP value of the feature j.

4.6. Individual variables affect

Fig. 5 demonstrates the influence of individual variables on the passenger's thermal perception in the male and female groups. X-axis display Age, BMI, and CI, and the Y-axis indicates their respective impact on thermal comfort, i.e., the SHAP values for each sample.

From Fig. 5, the Age feature looks like there are different impact trends, i.e., the male group's impact trend shows a decrease followed by an increase with the passenger's age, but the female group exhibits a downward trend in volatility. However, the value domains of these two groups are different. In the male group, when the Age variable is between 21 and 25, its SHAP value is mainly negative.

BMI is a commonly used international measure of how fat or thin a body is and whether it is healthy. In respect of males, the BMI's impact trend is like the Age feature. When their BMI is between 21 and 27, males feel more sensitive to the thermal environment. In particular, obese males with BMI over 27 dramatically increase the predicted value of thermal comfort value, indicating that obese men are prone to discomfort with the thermal environment. In the female group, we find that females with a BMI of less than 19.5 demonstrate better adaptability

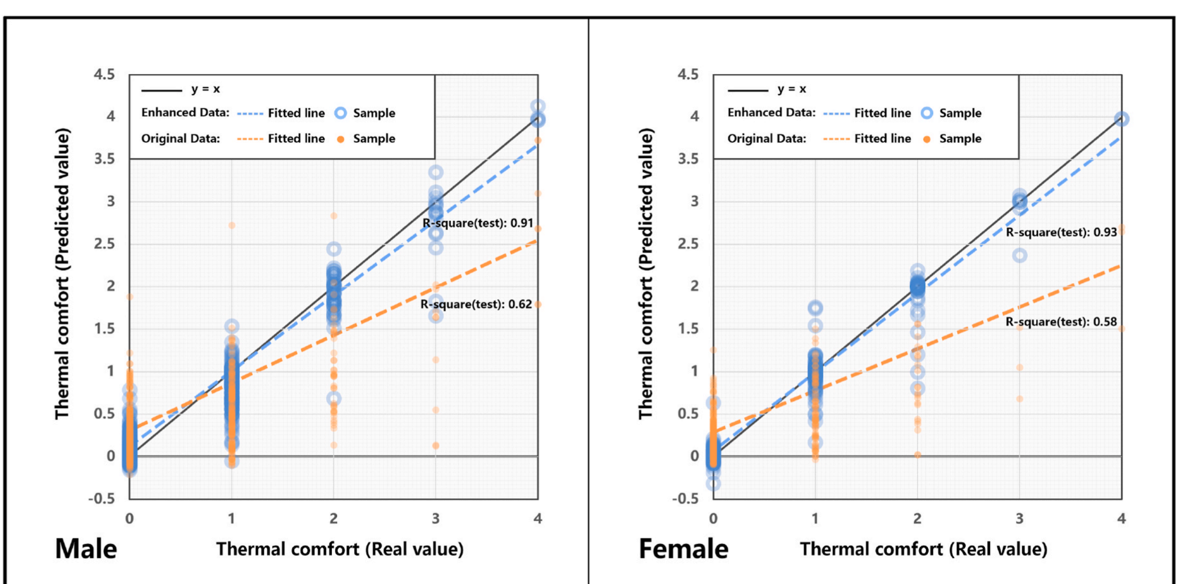

Fig. 3. Scatter plot comparison of BO-LightGBM performance based on VAE enhanced data versus original data.

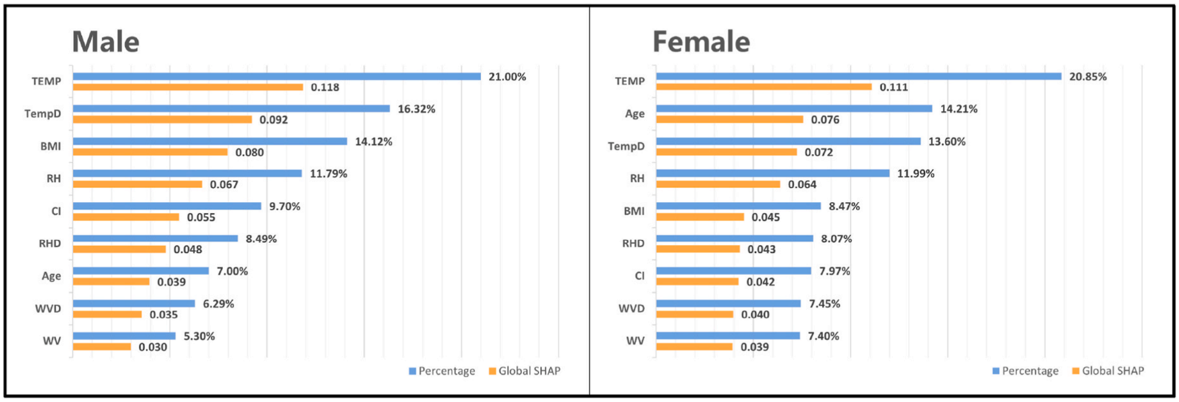

Fig. 4. Global SHAP of independent variables.

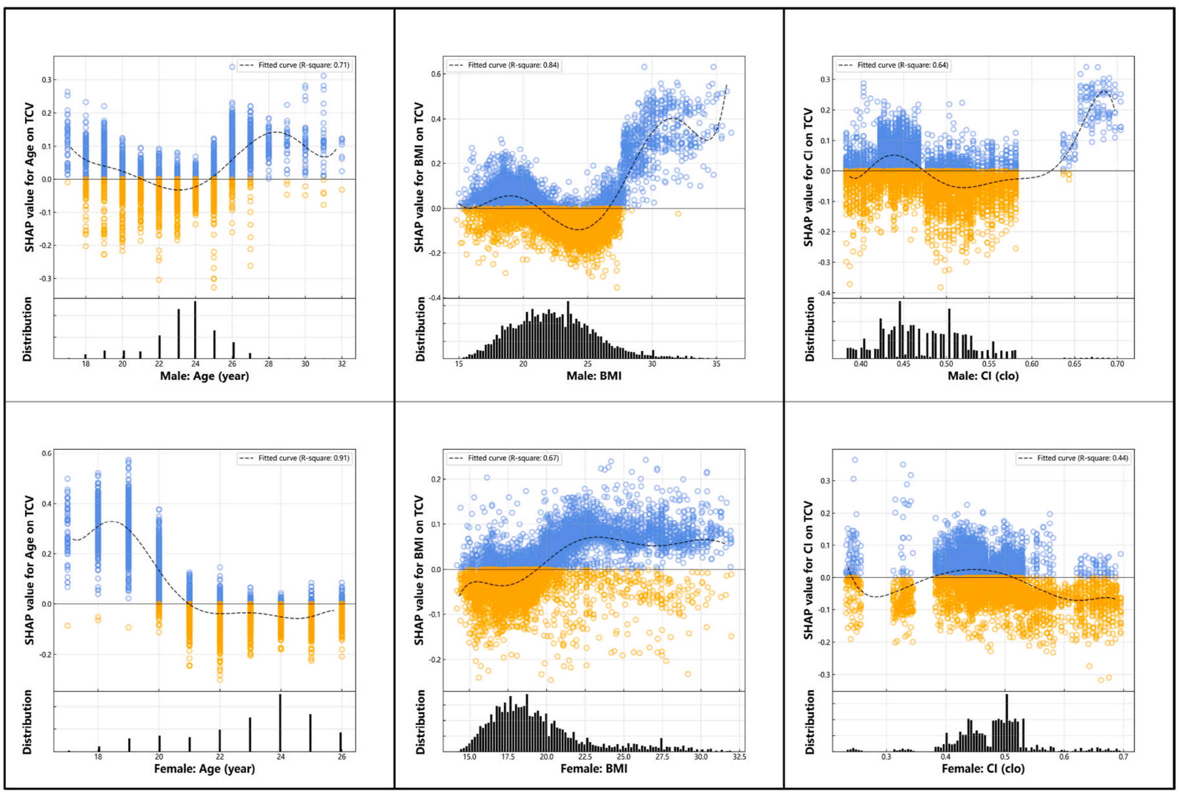

Fig. 5. SHAP dependence plots of individual variables.

and discomfort in females with a BMI over 22.5 did not increase with BMI.

According to our experimental results, the male CI has a marginal effect on thermal comfort when it surpasses a value of 0.6. Specifically, the mean CI value in summer has been found to be 0.6 [26]. Our experimental findings indicate that when both females and males exceed this seasonal mean CI value, females demonstrate better adaptation to the thermal environment while males experience discomfort. This suggests that males are more vulnerable to changes in the thermal environment when their CI surpasses the seasonal mean value.

4.7. Environmental variables affect

Fig. 6 demonstrates the influence of environmental variables. X-axis display air TEMP, RH, WV, TempD, RHD, and WVD, and the Y-axis indicates their respective impact on thermal comfort.

The influence domain of the TEMP feature is between −0.24 and 1.22 and between −0.26 and 0.91 in the male and female groups, respectively. It is the most vital variable in thermal comfort, with the same result as the global SHAP. The subgraphs about TEMP indicate 29 °C is a comfort threshold both in the male and female groups. The body's discomfort is rapidly elevated while the air TEMP exceeds this threshold. When the air TEMP is under 23.5° Celsius, females are more likely to feel discomfort than males. Meanwhile, when the air TEMP is higher than 29° Celsius, males are more likely to feel discomfort than females.

From the point of view of the TempD variable, commuters tend to a stable air TEMP environment, especially males. They generally accept a TEMP change of −2°C–1 °C. A rise of more than 1 °C in air TEMP can produce discomfort in both men and female, but females have a higher tolerance for cooling.

The subgraphs about RH in Fig. 6 reveal a similar trend in thermal

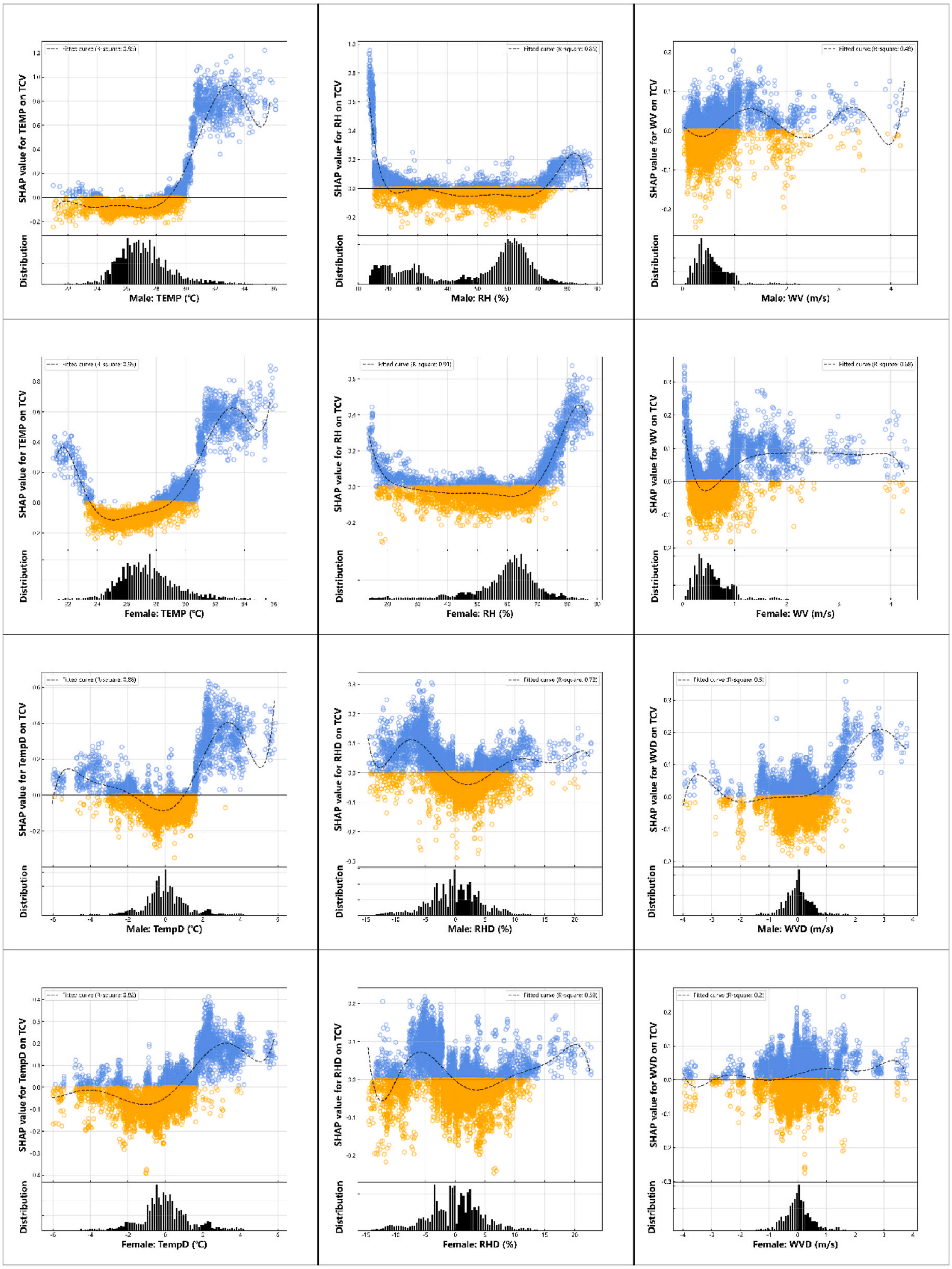

Fig. 6. SHAP dependence plots of environmental variables.

comfort by RH in the male and female groups. RH possesses a colossal comfort zone compared to air TEMP and is from 20% to 72% and 25%–69% in respect of males and females, respectively, which means that people are not sensitive to changes in humidity. Although there are no significant differences between males and females, males are more averse to extremely dry environments (less than 20%), and females are more averse to highly humid environments (more than 69%).

The effect trend of the RHD variable shows a double-peaked and double-valley pattern. The subgraphs of RHD present that passengers prefer progressively wetter spaces. In the male group, the samples' SHAP value is mainly negative when the RHD feature is from −2% to 6%. In the female group, the range is from 0% to 7.5%. Interestingly, a significant reduction in humidity can likewise enhance thermal comfort, a particularly noticeable phenomenon for females. This observation may be because increased humidity may negatively affect body evaporation and heat loss, which was also verified by Yan et al. [66].

With SHAP levels between 0 and 0.1, males and females are less sensitive to WV and WVD. However, there are minor variations in how susceptible WV males and females are. While variations in WV for males show peaks and valleys, indicating unstable qualities, changes in WV for females are more stable.

5. Discussion

5.1. Data augmentation

The extensive dimensions in the limited raw dataset cause the “curse of dimension.” In light of the results presented in Fig. 3, we observed that the raw data is characterized by “uneven data distribution and a small sample size of uncomfortable data,” resulting in a model based on the original data with low accuracy and poor generalization ability for the uncomfortable samples. To address this issue, we applied the VAE framework to construct implicit distributions of the data, enhancing the data volume. This augmented data effectively complemented the original data, resulting in BO-LightGBM models successfully fitting thermal comfort data at all stages. As a result, the accuracy and generalizability of the thermal comfort model trained using the augmented data were significantly enhanced. Notably, the generalizability increased by approximately 30% for both female and male models. To elaborate further, the R-square values of the training and test sets for the female thermal comfort model improved from 0.82 to 0.62 to 0.98 and 0.91, respectively. Similarly, the R-square values of the training and test sets for the male thermal comfort model improved from 0.81 to 0.58 to 0.99 and 0.93, respectively. Therefore, we argue that our proposed approach is an effective way to enhance the thermal comfort model in underground spaces.

5.2. Individual contributions

Significant gender differences in the perception of the thermal environment are attributable to the individual characteristics of females and males. Specifically, BMI plays a critical role in thermal comfort, with females exhibiting an increased degree of influence of thermal comfort value as BMI increases. Conversely, the male who is either underweight or overweight is more sensitive to the perception of thermal comfort value. An interesting observation is that when the BMI of men exceeds 27, the contribution of BIM to thermal comfort shows an increasing marginal effect. This can be attributed to the fact that in warmer conditions, obese individuals have a lower sweat rate per unit body surface area compared to lean individuals, which may not be sufficient to maintain the required evaporative heat loss to match metabolic heat production, leading to heat storage [48,67–69]. Therefore, males should be mindful of their BMI and maintain a healthy body shape to adapt to the changing external thermal environment.

5.3. Underground space environmental context

5.3.1. Dynamic environmental variables

The incorporation of dynamic environmental variables leads to the precise result of the thermal comfort model. The Shapley values reveal the impact of TempD, RHD, and WVD on the thermal environment. The findings indicate that individuals of both genders experience enhanced thermal comfort when there is consistency in the anterior and posterior environment. This outcome can be explained by physiological mechanisms. To be specific, the human body maintains a constant internal TEMP through various physiological mechanisms, such as sweating and shivering, to maintain homeostasis [70]. When the surrounding environment, including air TEMP, WV, and humidity, remains constant, the body can regulate its internal TEMP more easily and maintain its preferred thermal state. Therefore, any changes in these environmental factors can cause discomfort, requiring additional energy expenditure from the body to maintain its internal TEMP, leading to fatigue and stress [71]. Thus, a constant environment is preferred for optimal thermal comfort.

5.3.2. Environmental air temperature

Research findings indicate that the thermal comfort of individuals is heavily influenced by environmental TEMP. Interestingly, it has been observed that females are more sensitive to low TEMPs when the TEMP falls below 24 °C, while males are more susceptible to uncomfortable when the TEMP surpasses 29 °C, often referred to as the “females are afraid of cold, males are afraid of heat” phenomenon [72]. This gender-based difference in thermal comfort has been attributed to the anatomical and physiological differences between males and females. Specifically, females typically have smaller skeletons and faster skin heat dissipation rates [43], which leaves them relatively more vulnerable to cold TEMPs, resulting in greater fear of the cold than males. On the other hand, males have a higher metabolic rate than females, with an approximately 20% difference, and increased heat production, making them more sensitive to high TEMPs [46,73]. In the context of subway design, current Chinese design codes set the standard TEMP at 30 °C, which may not be the most comfortable range for all passengers. This study suggests that passengers feel most comfortable at TEMPs between 24 and 29 °C. As such, it is crucial to adjust the maximum subway air TEMP by 1–2 °C and optimize ventilation and air conditioning systems during summer to establish a comfortable thermal environment for both genders in underground spaces.

5.3.3. Relative humidity

This study proposes a more general solution for enhancing the thermal comfort of subway passengers by adjusting the RH in the subway to a specified range of 20%–70%. While the “Code for Design of Metro in China” (GB50157-2013) sets the standard RH range at 40%–70%, the study’s findings reveal that a broader range of 20%–70% is still considered comfortable for both genders. Passengers reported feeling comfortable even when the RH values in subway station halls and platforms exceeded the standard range. However, it is essential to note that while both genders exhibit lower sensitivity to RH, gender-specific variations in the perception of thermal comfort concerning RH still exist. Specifically, males tend to feel more uncomfortable in dry environments with RH below 20%, whereas females are more likely to feel uncomfortable in humid environments with RH above 70%. While previous studies have predominantly focused on descriptive comparisons of gender differences in thermal comfort perception with respect to RH [45,74], this study provides more precise insights into gender-based differences in thermal comfort perception.

5.3.4. Wind velocity

The findings of the present study suggest that while adjusting WV is a cost-effective and energy-efficient approach for enhancing thermal comfort in underground spaces [33], males and females are less responsive to changes in WV and WVD when it comes to their perception of thermal comfort. This is supported by the SHAP values, which indicate a range of −0.1 to 0.1. Recent research has shown that thermal comfort can be improved with increasing air velocity when the air TEMP exceeds a critical value of 33.7° Celsius [13]. However, in Beijing’s summer underground space, the air TEMP is regulated by the ventilation and air conditioning system between a relatively constant range of 24–30° Celsius, which does not surpass the critical TEMP value. Hence, subway stations in Beijing create a comfortable thermal environment for subway passengers without requiring additional energy expenditure or cost on ventilation systems.

5.3.5. Limitation

This study illuminates the impact of various individual and environmental factors on thermal comfort in underground spaces, with a particular focus on gender differences. Despite providing valuable insights, some limitations warrant further investigation. Firstly, the ISO 9920 and ASHRAE handbooks, which provide thermal insulation data for clothing, were published in 2007 and 2017 [75] and are based on experimental data collected several decades ago. Consequently, there is an urgent need to update and expand these existing databases to reflect advancements in materials engineering and the emergence of new clothing materials. Secondly, while 18-30-year-old passengers are the most frequent users of subway systems, it is crucial to consider passengers' thermal comfort across different age groups. Such an approach would enable the control and regulation of air conditioning systems according to the varying needs of different age cohorts.

Moreover, our future research will not only continue to examine gender differences but will also incorporate novel clothing material data and age groups. This will enable us to conduct more targeted research on thermal comfort in underground spaces and create a comfortable thermal environment for subway passengers across all demographics.

To foster further advancements in this field, our future research endeavors will encompass novel environmental conditions and unique environmental characteristics, aiming to enhance the diversity and comprehensiveness of thermal comfort studies. By incorporating these additional factors, we seek to broaden our understanding of the intricate interplay between thermal comfort and various environmental contexts.

6. Conclusion

With the city and society's development, the rapidly increasing demand for the efficiently sustainable underground public transportation system has prompted a greater focus on enhancing passengers' experience, especially for the thermal comfort in subway stations. To date, several studies have been conducted to understand the factors affecting thermal comfort in indoor environments. However, due to data limitations, limited studies have been conducted specifically for the subway station environment.

In response to these research gaps, this study proposed a thermal comfort data generated algorithm based on the VAE framework, which broke the data collection constraints, providing a sufficient, extensive database for predictive modeling. In addition, the study explored factors influencing thermal comfort in subway stations via BO-LightGBM MLMs through the generated extensive dataset, with a particular emphasis on the interplay between individual characteristics based on gender differences, such as Age, BMI, and CI, and environmental variables, including air TEMP, RH, WV, and environmental variables relative status ratio.

This study is summarized explicitly in the following points:

1. Data volume constraints underground public transportation space's thermal comfort effectively modeling. Mainly for the thermal discomfort interval, the model trained by the original dataset exhibits poor capability due to the absence of data. However, the VAE model established the data distributions in the latent space to enhance the thermal comfort dataset, increasing the accuracy of decoupling models at 29% and 35% in male and female test sets, respectively.
2. Individual variables, such as Age, BMI, and CI, influence thermal comfort in distinct ways for males and females. As for BMI, males demonstrate a heightened sensitivity to changes in this variable, notably exceeding 27 BMI, implying that their thermal comfort is more affected by differences in body composition than females. In addition, females are more adaptable to variations in CI levels. This suggests that their thermal comfort is less susceptible to changes in clothing and can more effectively cope with a broader range of insulation values without experiencing significant discomfort.
3. Environmental variables, including air TEMP and RH, significantly impact passengers' thermal comfort. However, WV has a limited impact on thermal comfort in the study area. The study suggests that people feel most comfortable at TEMPs between 24 and 29 °C and RH levels between 20% and 70%.
4. Gender differences in the environmental variables. Females are more sensitive to low TEMPs below 24 °C, while males are more sensitive to high TEMPs above 29 °C. Males tend to feel more uncomfortable in environments with RH below 20%, whereas females are likelier in settings with RH above 70%.
5. As for the subway stations' dynamic environmental assessment, the TempD, RHD, and WVD present minimal thermal comfort effects. However, these variables serve as important indicators that a stable environment can promote enhanced thermal comfort for passengers.

By investigating these complex relationships, the study aims to highlight the importance of considering gender differences and personal attributes in designing comfortable and inclusive public transportation environments. At the same time, The study's findings are expected to contribute valuable insights into human-centered transportation design and inform the development of evidence-based guidelines and recommendations for optimizing the thermal comfort of passengers in subway stations, ultimately creating more sustainable, impartial, and gender-friendly public transportation systems.

CRediT authorship contribution statement

Renlu Qiao: Writing – review & editing, Visualization, Supervision, Resources, Methodology, Conceptualization. **Zhiqaing Wu:** Conceptualization, Supervision. **Shuo Gao:** Formal analysis. **Qingrui Jiang:** Writing – original draft, Visualization, Supervision, Methodology, Data curation, Conceptualization. **Xiaochang Liu:** Writing – review & editing. **Chenyu Huang:** Software. **Li Xia:** Validation. **Mingze Chen:** Investigation.

Declaration of competing interest

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

Data availability

The authors do not have permission to share data.

References

[1] R. Qiao, T. Liu, Impact of building greening on building energy consumption: a quantitative computational approach, J. Clean. Prod. 246 (2020), 119020, https://doi.org/10.1016/j.jclepro.2019.119020.
[2] Q. Jiang, X. Liu, Z. Wu, Y. Wang, J. Dong, Decoupling of the municipal thermal environment using a spatial autoregressive model, Atmosphere 13 (2022) 2059, https://doi.org/10.3390/atmos13122059.
[3] D.V.L. Hunt, L.O. Makana, I. Jefferson, C.D.F. Rogers, Liveable cities and urban underground space, Tunn. Undergr. Space Technol. 55 (2016) 8–20, https://doi.org/10.1016/j.tust.2015.11.015.
[4] D. Lin, J.D. Nelson, M. Beecroft, J. Cui, An overview of recent developments in China's metro systems, Tunn. Undergr. Space Technol. 111 (2021), 103783, https://doi.org/10.1016/j.tust.2020.103783.
[5] 2020年城市轨道交通运营数据发布_部门政务_中国政府网, Available online: http://www.gov.cn/xinwen/2021-01/06/content_5577537.htm. (Accessed 31 December 2022).
[6] L. Xiong, Y. Yao, Study on an adaptive thermal comfort model with K-Nearest-Neighbors (KNN) algorithm, Build. Environ. 202 (2021), 108026, https://doi.org/10.1016/j.buildenv.2021.108026.
[7] S. Pan, Y. Liu, L. Xie, X. Wang, Y. Yuan, X. Jia, A thermal comfort field study on subway passengers during air-conditioning season in beijing, Sustain. Cities Soc. 61 (2020), 102218, https://doi.org/10.1016/j.scs.2020.102218.
[8] P.O. Fanger, Assessment of man's thermal comfort in practice, Occup. Environ. Med. 30 (1973) 313–324, https://doi.org/10.1136/oem.30.4.313.

[9] J. Yu, Y. Kang, Z. Zhai, John, Advances in research for underground buildings: energy, thermal comfort and indoor air quality, Energy Build. 215 (2020), 109916, https://doi.org/10.1016/j.enbuild.2020.109916.
[10] School of Housing, Building and planning, Universiti Aains Malaysia, 11800, Penang, Malaysia; Bahdad, A.A.S.; Department of Architectural Engineering & Environmental Planning, College of Engineering and Petroleum, Hadhramout University, Mukalla, Yemen; Fadzil, S.F.S.; School of Housing, Building and Planning, Universiti Sains Malaysia, 11800, Penang, Malaysia, An investigation-based optimization framework of thermal comfort analysis in underground enclosed spaces affected by multiple parameters for energy performance in tropics, J. Daylighting 9 (2022) 48–63, https://doi.org/10.15627/jd.2022.4.
[11] M.K. Singh, R. Ooka, H.B. Rijal, S. Kumar, A. Kumar, S. Mahapatra, Progress in thermal comfort studies in classrooms over last 50 Years and way forward, Energy Build. 188–189 (2019) 149–174, https://doi.org/10.1016/j.enbuild.2019.01.051.
[12] S. Lu, W. Wang, C. Lin, E.C. Hameen, Data-driven simulation of a thermal comfort-based temperature set-point control with ASHRAE RP884, Build. Environ. 156 (2019) 137–146, https://doi.org/10.1016/j.buildenv.2019.03.010.
[13] G. Katavoutas, M.N. Assimakopoulos, D.N. Asimakopoulos, On the determination of the thermal comfort conditions of a metropolitan city underground railway, Sci. Total Environ. 566–567 (2016) 877–887, https://doi.org/10.1016/j.scitotenv.2016.05.047.
[14] P. Roghanchi, K.C. Kocsis, M. Sunkpal, Sensitivity analysis of the effect of airflow velocity on the thermal comfort in underground mines, J. Sustain. Mining 15 (2016) 175–180, https://doi.org/10.1016/j.jsm.2017.03.005.
[15] L. Arakawa Martins, V. Soebarto, T. Williamson, A systematic review of personal thermal comfort models, Build. Environ. 207 (2022), 108502, https://doi.org/10.1016/j.buildenv.2021.108502.
[16] V. Cardoso, N.M.M. Ramos, R.M.S.F. Almeida, E. Barreira, J.P. Martins, M. Lurdes Simões, L. Sanhudo, B. Ribeiro, Thermal comfort evaluation in cruise terminals, Build. Environ. 126 (2017) 276–287, https://doi.org/10.1016/j.buildenv.2017.10.008.
[17] Y. Li, S. Geng, Y. Yuan, J. Wang, X. Zhang, Evaluation of climatic zones and field study on thermal comfort for underground engineering in China during summer, Sustain. Cities Soc. 43 (2018) 421–431, https://doi.org/10.1016/j.scs.2018.08.002.
[18] M. Favero, I. Sartori, S. Carlucci, Human thermal comfort under dynamic conditions: an experimental study, Build. Environ. 204 (2021), 108144, https://doi.org/10.1016/j.buildenv.2021.108144.
[19] S. Thapa, Insights into the thermal comfort of different naturally ventilated buildings of darjeeling, India – effect of gender, age and BMI, Energy Build. 193 (2019) 267–288, https://doi.org/10.1016/j.enbuild.2019.04.003.
[20] Z. Wang, H. Zhang, Y. He, M. Luo, Z. Li, T. Hong, B. Lin, Revisiting individual and group differences in thermal comfort based on ASHRAE database, Energy Build. 219 (2020), 110017, https://doi.org/10.1016/j.enbuild.2020.110017.
[21] H. Kim, D. Jung, H. Choi, T. Hong, Advanced prediction model for individual thermal comfort considering blood glucose and salivary cortisol, Build. Environ. 224 (2022), 109551, https://doi.org/10.1016/j.buildenv.2022.109551.
[22] N. Najafi, K. Movahed, Z. Barzegar, S. Samani, The effect of ventilation by window opening on stress, anxiety, and depression of female high school students, Int J School Health (2019), https://doi.org/10.5812/intjsh.87038. *In Press*.
[23] W. Yao, X. Li, W. Cao, G. Li, L. Ren, W. Gao, Research on the influence of indoor thermal environment and activity levels on thermal comfort in protective clothing, Energy Build. 279 (2023), 112681, https://doi.org/10.1016/j.enbuild.2022.112681.
[24] W. Luo, R. Kramer, Y. de Kort, W. van Marken Lichtenbelt, Effectiveness of personal comfort systems on whole-body thermal comfort – a systematic review on which body segments to target, Energy Build. 256 (2022), 111766, https://doi.org/10.1016/j.enbuild.2021.111766.
[25] H. Jin, S. Liu, J. Kang, Gender differences in thermal comfort on pedestrian streets in cold and transitional seasons in severe cold regions in China, Build. Environ. 168 (2020), 106488, https://doi.org/10.1016/j.buildenv.2019.106488.
[26] S. Karjalainen, Thermal comfort and gender: a literature review: thermal comfort and gender: a literature review, Indoor Air 22 (2012) 96–109, https://doi.org/10.1111/j.1600-0668.2011.00747.x.
[27] M. Indraganti, M.A. Humphreys, A comparative study of gender differences in thermal comfort and environmental satisfaction in air-conditioned offices in Qatar, India, and Japan, Build. Environ. 206 (2021), 108297, https://doi.org/10.1016/j.buildenv.2021.108297.
[28] A.A. Farhan, K. Pattipati, B. Wang, P. Luh, Predicting individual thermal comfort using machine learning algorithms, in: Proceedings of the 2015 IEEE International Conference on Automation Science and Engineering (CASE), IEEE, Gothenburg, Sweden, August 2015, pp. 708–713.
[29] B. Chegari, M. Tabaa, E. Simeu, F. Moutaouakkil, H. Medromi, Multi-objective optimization of building energy performance and indoor thermal comfort by combining artificial neural networks and metaheuristic algorithms, Energy Build. 239 (2021), 110839, https://doi.org/10.1016/j.enbuild.2021.110839.
[30] T. Chaudhuri, D. Zhai, Y.C. Soh, H. Li, L. Xie, Random forest based thermal comfort prediction from gender-specific physiological parameters using wearable sensing technology, Energy Build. 166 (2018) 391–406, https://doi.org/10.1016/j.enbuild.2018.02.035.
[31] J. Kim, Y. Zhou, S. Schiavon, P. Raftery, G. Brager, Personal comfort models: predicting individuals' thermal preference using occupant heating and cooling behavior and machine learning, Build. Environ. 129 (2018) 96–106, https://doi.org/10.1016/j.buildenv.2017.12.011.
[32] C. Huang, G. Zhang, J. Yao, X. Wang, J.K. Calautit, C. Zhao, N. An, X. Peng, Accelerated environmental performance-driven urban design with generative adversarial network, Build. Environ. 224 (2022), 109575, https://doi.org/10.1016/j.buildenv.2022.109575.
[33] R. Qiao, X. Li, S. Gao, X. Ma, Improvement of thermal comfort for underground space: data enhancement using variational autoencoder, Build. Environ. 207 (2022), 108457, https://doi.org/10.1016/j.buildenv.2021.108457.
[34] M. Quintana, S. Schiavon, K.W. Tham, C. Miller, Balancing thermal comfort datasets: we GAN, but should we?, in: Proceedings of the Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings Cities, and Transportation, November 18 2020, pp. 120–129.
[35] Beijing Subway, Wikipedia, 2023.
[36] X. Zhou, L. Xu, J. Zhang, B. Niu, M. Luo, G. Zhou, X. Zhang, Data-driven thermal comfort model via support vector machine algorithms: insights from ASHRAE RP-884 database, Energy Build. 211 (2020), 109795, https://doi.org/10.1016/j.enbuild.2020.109795.
[37] 14:00-17:00 ISO 10551. https://www.iso.org/standard/18636.html, 1995. (Accessed 22 June 2023).
[38] K.C. Parsons, The effects of gender, acclimation state, the opportunity to adjust clothing and physical disability on requirements for thermal comfort, Energy Build. 34 (2002) 593–599, https://doi.org/10.1016/S0378-7788(02)00009-9.
[39] K. Sinha, E. Rajasekar, Thermal comfort evaluation of an underground metro station in New Delhi using agent-based modelling, Build. Environ. 177 (2020), 106924, https://doi.org/10.1016/j.buildenv.2020.106924.
[40] M. Indraganti, R. Ooka, H.B. Rijal, Thermal comfort in offices in India: behavioral adaptation and the effect of age and gender, Energy Build. 103 (2015) 284–295, https://doi.org/10.1016/j.enbuild.2015.05.042.
[41] R.F. Rupp, O.B. Kazanci, J. Toftum, Investigating current trends in clothing insulation using a global thermal comfort database, Energy Build. 252 (2021), 111431, https://doi.org/10.1016/j.enbuild.2021.111431.
[42] E.J. Choi, Y.J. Choi, N.H. Kim, J.W. Moon, Seasonal effects of thermal comfort control considering real-time clothing insulation with vision-based model, Build. Environ. 235 (2023), 110255, https://doi.org/10.1016/j.buildenv.2023.110255.
[43] M. Fan, Z. Fu, J. Wang, Z. Wang, H. Suo, X. Kong, H. Li, A review of different ventilation modes on thermal comfort, air quality and virus spread control, Build. Environ. 212 (2022), 108831, https://doi.org/10.1016/j.buildenv.2022.108831.
[44] Q. Zhao, Z. Lian, D. Lai, Thermal comfort models and their developments: a review, Energy Built Environ. 2 (2021) 21–33, https://doi.org/10.1016/j.enbenv.2020.05.007.
[45] K. Parsons, Human Thermal Environments, ISBN 978-1-4200-2524-8, CRC Press, 2007.
[46] B. Griefahn, C. Künemund, The effects of gender, age, and fatigue on susceptibility to draft discomfort, J. Therm. Biol. 26 (2001) 395–400, https://doi.org/10.1016/S0306-4565(01)00050-X.
[47] L. Wu, H. Xia, X. Wang, Q. Dong, C. Lin, X. Liu, R. Liang, Indoor air quality and passenger thermal comfort in beijing metro transfer stations, Transport. Res. Transport Environ. 78 (2020), 102217, https://doi.org/10.1016/j.trd.2019.102217.
[48] F. Johnson, A. Mavrogianni, M. Ucci, A. Vidal-Puig, J. Wardle, Could increased time spent in a thermal comfort zone contribute to population increases in obesity?: indoor temperatures and obesity, Obes. Rev. 12 (2011) 543–551, https://doi.org/10.1111/j.1467-789X.2010.00851.x.
[49] H.C. Burridge, R.K. Bhagat, M.E.J. Stettler, P. Kumar, I. De Mel, P. Demis, A. Hart, Y. Johnson-Llambias, M.-F. King, O. Klymenko, et al., The ventilation of buildings and other mitigating measures for COVID-19: a focus on wintertime, rspa.2020.0855, 20200855, Proc. R. Soc. A A. 477 (2021), https://doi.org/10.1098/rspa.2020.0855.
[50] A.U. Weerasuriya, X. Zhang, B. Lu, K.T. Tse, C.-H. Liu, Optimizing lift-up design to maximize pedestrian wind and thermal comfort in 'hot-calm' and 'cold-windy' climates, Sustain. Cities Soc. 58 (2020), 102146, https://doi.org/10.1016/j.scs.2020.102146.
[51] K. Liu, Z. Lian, X. Dai, D. Lai, Comparing the effects of sun and wind on outdoor thermal comfort: a case study based on longitudinal subject tests in cold climate region, Sci. Total Environ. 825 (2022), 154009, https://doi.org/10.1016/j.scitotenv.2022.154009.
[52] Y. Du, C.M. Mak, Y. Li, A multi-stage optimization of pedestrian level wind environment and thermal comfort with lift-up design in ideal urban canyons, Sustain. Cities Soc. 46 (2019), 101424, https://doi.org/10.1016/j.scs.2019.101424.
[53] B.-J. He, J. Wang, H. Liu, G. Ulpiani, Localized synergies between heat waves and urban heat islands: implications on human thermal comfort and urban heat management, Environ. Res. 193 (2021), 110584, https://doi.org/10.1016/j.envres.2020.110584.
[54] R. Aghamolaei, M. Fallahpour, P.A. Mirzaei, Tempo-spatial thermal comfort analysis of urban heat island with coupling of CFD and building energy simulation, Energy Build. 251 (2021), 111317, https://doi.org/10.1016/j.enbuild.2021.111317.
[55] D.P. Kingma, M. Welling, Auto-Encoding Variational Bayes, 2013, https://doi.org/10.48550/ARXIV.1312.6114.
[56] B. Dai, D. Wipf, Diagnosing and Enhancing VAE Models, 2019, https://doi.org/10.48550/ARXIV.1903.05789.
[57] Q. Chai, H. Wang, Y. Zhai, L. Yang, Using machine learning algorithms to predict occupants' thermal comfort in naturally ventilated residential buildings, Energy Build. 217 (2020), 109937, https://doi.org/10.1016/j.enbuild.2020.109937.
[58] C. Fan, W. He, L. Liao, Real-time machine learning-based recognition of human thermal comfort-related activities using inertial measurement unit data, Energy Build. 294 (2023), 113216, https://doi.org/10.1016/j.enbuild.2023.113216.

[59] O. Sagi, L. Rokach, Ensemble learning: a survey, WIREs Data Mining Knowl. Discov. 8 (2018), https://doi.org/10.1002/widm.1249.

[60] B. Yang, X. Li, Y. Liu, L. Chen, R. Guo, F. Wang, K. Yan, Comparison of models for predicting winter individual thermal comfort based on machine learning algorithms, Build. Environ. 215 (2022), 108970, https://doi.org/10.1016/j.buildenv.2022.108970.

[61] X. Feng, E.B. Zainudin, H.W. Wong, K.J. Tseng, A hybrid ensemble learning approach for indoor thermal comfort predictions utilizing the ASHRAE RP-884 database, Energy Build. 290 (2023), 113083, https://doi.org/10.1016/j.enbuild.2023.113083.

[62] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, LightGBM: a highly efficient gradient boosting decision tree, in: Proceedings of the Advances in Neural Information Processing Systems vol. 30, Curran Associates, Inc., 2017.

[63] B. Shahriari, K. Swersky, Z. Wang, R.P. Adams, N. de Freitas, Taking the human out of the loop: a review of bayesian optimization, Proc. IEEE 104 (2016) 148–175, https://doi.org/10.1109/JPROC.2015.2494218.

[64] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling Technique, JAIR 16 (2002) 321–357, https://doi.org/10.1613/jair.953.

[65] Haibo He, Bai Yang, E.A. Garcia, Shutao Li Adasyn, Adaptive synthetic sampling approach for imbalanced learning, in: Proceedings of the 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), IEEE, Hong Kong, China, June 2008, pp. 1322–1328.

[66] H. Yan, Q. Liu, W. Zhao, C. Pang, M. Dong, H. Zhang, J. Gao, H. Wang, B. Hu, L. Yang, et al., The coupled effect of temperature, humidity, and air movement on human thermal response in hot–humid and hot–arid climates in summer in China, Build. Environ. 177 (2020), 106898, https://doi.org/10.1016/j.buildenv.2020.106898.

[67] R.F. Rupp, J. Kim, R. de Dear, E. Ghisi, Associations of occupant demographics, thermal history and obesity variables with their thermal comfort in air-conditioned and mixed-mode ventilation office buildings, Build. Environ. 135 (2018) 1–9, https://doi.org/10.1016/j.buildenv.2018.02.049.

[68] Z. Wang, R. de Dear, M. Luo, B. Lin, Y. He, A. Ghahramani, Y. Zhu, Individual difference in thermal comfort: a literature review, Build. Environ. 138 (2018) 181–193, https://doi.org/10.1016/j.buildenv.2018.04.040.

[69] K.A. Dougherty, M. Chow, W.L. Kenney, Responses of lean and obese boys to repeated summer exercise in the heat bouts, Med. Sci. Sports Exerc. 41 (2009) 279–289, https://doi.org/10.1249/MSS.0b013e318185d341.

[70] K. Chen, Q. Xu, B. Leow, A. Ghahramani, Personal thermal comfort models based on physiological measurements – a design of experiments based review, Build. Environ. 228 (2023), 109919, https://doi.org/10.1016/j.buildenv.2022.109919.

[71] J. Bligh, K.G. Johnson, Glossary of terms for thermal physiology, J. Appl. Physiol. 35 (1973) 941–961, https://doi.org/10.1152/jappl.1973.35.6.941.

[72] M. Racine, Y. Tousignant-Laflamme, L.A. Kloda, D. Dion, G. Dupuis, M. Choinière, A systematic literature review of 10 Years of research on sex/gender and experimental pain perception – Part 1: are there really differences between women and men? Pain 153 (2012) 602–618, https://doi.org/10.1016/j.pain.2011.11.025.

[73] W. Liu, Z. Lian, Q. Deng, Y. Liu, Evaluation of calculation methods of mean skin temperature for use in thermal comfort study, Build. Environ. 46 (2011) 478–488, https://doi.org/10.1016/j.buildenv.2010.08.011.

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

[75] S. Gao, R. Ooka, W. Oh, Experimental investigation of the effect of clothing insulation on thermal comfort indices, Build. Environ. 187 (2021), 107393, https://doi.org/10.1016/j.buildenv.2020.107393.