



Machine learning-based prediction of outdoor thermal comfort: Combining Bayesian optimization and the SHAP model

Ruiqi Guo^b, Bin Yang^{a,b,*}, Yuyao Guo^b, He Li^a, Zhe Li^b, Bin Zhou^b, Bo Hong^c, Faming Wang^d

^a School of Energy and Safety Engineering, Tianjin Chengjian University, Tianjin, China

^b School of Building Services Science and Engineering, Xi'an University of Architecture and Technology, Xi'an, China

^c College of Landscape Architecture & Arts, Northwest A&F University, Yangling, China

^d Division Animal and Human Health Engineering, Department of Biosystems (BIOSYST), KU Leuven, Kasteelpark Arenberg 30, BE-3001 Leuven, Belgium



ARTICLE INFO

Keywords:

Outdoor thermal comfort
Urban open space
Machine learning
Bayesian optimization
SHapley additive exPlanations

ABSTRACT

Rising global temperatures have resulted in urban heat waves in recent years, endangering residents' health and even their lives. As a result, accurate outdoor thermal comfort prediction is critical. The goal of this research is to develop highly accurate and interpretable machine learning models of outdoor thermal comfort. We created a summer outdoor thermal comfort dataset for cold regions using microclimate parameter measurements and questionnaires, and divided it into two datasets: without and with shading. The prediction performance of nine machine learning models was compared, as well as their prediction performance following Bayesian optimization. Finally, the SHAP model was used to explain the important features of the best machine learning models and their impact on thermal comfort. The results show that after partitioning the dataset by shading, the accuracy of the best machine learning prediction models for Thermal sensation vote (TSV), Thermal acceptable vote (TA), and Thermal comfort vote (TCV) in the unshaded space improves by 9.2%, 9.31%, and 6.16%, respectively, while the prediction accuracy in the shaded space remains basically unchanged. Extreme gradient boosting (XGBoost), Light Gradient Boosting Machine (LightGBM), and categorical boosting (CatBoost) outperform other methods. After Bayesian optimization (BO), the machine learning models' TSV, TA, and TCV prediction accuracies improved by 6.83%, 4.05%, and 2.55%, respectively. CatBoost model with Bayesian optimization (CatBoost + BO) was the best model in most cases. TSV, TA, and TCV predictions were heavily influenced by microclimate parameters, with mean radiant temperature being the most important in unshaded spaces and air temperature being critical in shaded spaces. Furthermore, age, body mass index, and emotional state all had a significant impact on TA and TCV. This research will contribute to improving the quality of life of urban residents by providing a scientific foundation for the design and planning of urban open space.

1. Introduction

The global urban population is projected to increase from 56% in 2021 to 68% in 2050 [1], indicating that the trend towards urbanization is irreversible. Urban open spaces provide citizens with material, environmental, social, and economic benefits, attracting a large population for outdoor activities and enhancing urban vitality [2–5]. As urban populations continue to expand, the importance of urban open spaces increases [6]. The importance of outdoor thermal comfort in attracting residents to urban open spaces [7,8] is widely recognized. Recent years, however, have seen a gradual increase in global temperatures, and the extreme heat generated by urban heatwaves (UHW) poses a grave threat to the health of residents and can even result in death [9–14]. The urban

heat island effect (UHI) causes a rise in temperature within cities, which exacerbates the occurrence of heat waves and creates a harsh outdoor thermal environment in urban areas [15,16].

To create a thermally comfortable urban open space, it is necessary to begin with people's needs; therefore, accurate assessment or prediction of outdoor human thermal comfort [17,18] is necessary. Thermal comfort in the outdoors is a multifaceted concept influenced by a multitude of variables. These factors include physical, physiological, and psychological elements that directly influence the thermal comfort of an individual while outdoors. Indirect contributors include factors such as behavior, individual characteristics, sociocultural factors, as well as thermal history and location. These interrelated factors influence an individual's perception of thermal comfort outdoors [19–22].

Currently, the predicted mean vote (PMV), physiologically

* Corresponding author. No. 26 Jinjing Road, Xiqing District, Tianjin, China.

E-mail address: binyang@tju.edu.cn (B. Yang).

Nomenclature	
<i>Abbreviation Description</i>	
ASHRAE	American Society of Heating, Refrigerating, and Air-Conditioning Engineers
BMI	Body Mass Index
BO	Bayesian Optimization
CatBoost	Categorical Boosting
Cl	Clothing Thermal Resistance (clo)
DT	Decision Tree
G	Global Radiation (W/m^2)
ISO	International Organization for Standardization
KNN	K-Nearest Neighbor
LightGBM	Light Gradient Boosting
LIME	Local Interpretable Model-Agnostic Explanations
Met	Metabolic Rate
MLR	Multinomial Logistics Regression
OPM	Ordered Probit Model
PDP	Partial Dependence Plot
PET	Physiologically Equivalent Temperature ($^{\circ}\text{C}$)
PMV	Predicted Mean Vote
PPD	Percentage of Predicted Dissatisfaction (%)
RF	Random Forest
RH	Relative Humidity (%)
SHAP	SHapley Additive exPlanations
SVF	Sky View Factor
SVM	Support Vector Machine
T_a	Air Temperature ($^{\circ}\text{C}$)
T_g	Globe Temperature ($^{\circ}\text{C}$)
T_{mrt}	Mean Radiant Temperature ($^{\circ}\text{C}$)
TA	Thermal Acceptability Vote
TCV	Thermal Comfort Vote
TSV	Thermal Sensation Vote
UTCI	Universal Thermal Climate Index ($^{\circ}\text{C}$)
V	Wind Speed (m/s)
XGBoost	Extreme Gradient Boosting

equivalent temperature (PET), and universal thermal climate index (UTCI) are the most commonly used models for thermal comfort [19]. The PMV model, with its condensed form and comprehensive consideration of environmental and individual factors, has found widespread application in studies of indoor thermal comfort [23,24]. Nonetheless, as a result of its development under steady-state conditions, the PMV model exhibits significant discrepancies between outdoor temperature predictions and actual human thermal perception [25,26]. PET and UTCI take into account more fully the characteristics of the outdoor environment and are more widely accepted in outdoor thermal comfort prediction [19,27–31]. However, PET and UTCI encounter challenges in addressing individual differences, struggling to fully account for variations in thermal history and other conditions, thereby leading to specific limitations in their predictive capabilities. Moreover, they lack the direct capability to forecast subjective thermal sensations [32]. In response to these shortcomings, researchers and technologists are considering the adoption of data-driven methodologies. In the initial stages, linear regression is employed to establish a model predicting human thermal sensation levels based on meteorological, physiological, and psychological parameters [33,34]. This approach aims to comprehensively integrate the influence of environmental, physiological, psychological, social, and cultural factors on outdoor thermal comfort. Consequently, the use of data-driven methods results in an improved accuracy of outdoor thermal comfort predictions, a simplified prediction process, and enhanced user-friendliness of the model [35]. However, it is noteworthy that, in the early stages of research, limitations in algorithms, computing power, and data volume constrained the overall prediction accuracy of these data-driven methodologies.

With the development of machine learning technologies and the increase in computational power, a growing number of researchers have turned to machine learning methods for directly predicting thermal comfort. Various machine learning algorithms, such as the ordered probit model (OPM), multinomial logistic regression (MLR), k-nearest neighbor (KNN), decision tree (DT), support vector machine (SVM), random forest (RF), and extreme gradient boosting (XGBoost), have been utilized by researchers to achieve promising predictive performance [36–43]. Different climate conditions, indoor and outdoor environments, age groups, and levels of physical activity have a significant impact on outdoor thermal comfort [31,44–58]. Researchers have conducted comparative analyses of numerous machine learning models [35, 59–68] in order to select the optimal machine learning prediction models for specific scenarios and populations. Prior comparative research on these machine learning models incorporated a variety of climate conditions, indoor and outdoor environments, age groups, and

levels of physical activity.

Despite the fact that significant progress has been made in the application research of machine learning for outdoor thermal comfort, there are still some obstacles: First, shading is a significant factor in outdoor thermal comfort [69–73], but little is known about its influence. Sky View Factor (SVF) is the most prevalent method for quantifying sun shading conditions. SVF is a constant calculated from fisheye photos; however, the position of the sun changes throughout the day, causing dynamic changes in the scene's shading conditions. As a result, SVF may not accurately reflect real shading conditions, affecting the precision of thermal comfort predictions. Second, With the continuous development in the field of machine learning, many new algorithms with superior performance have emerged, such as light gradient boosting (LightGBM) and categorical boosting (CatBoost). However, their performance in predicting outdoor thermal comfort has not yet been thoroughly investigated. Third, machine learning models for predicting outdoor thermal comfort are frequently perceived as "black boxes" with limited visibility into their operational principles. Understanding the significance of various features and their functions within the model is limited. This lack of transparency impedes model development.

To address these issues, the primary objective of this study is to compile a comprehensive data set, taking into account the influence of shading factors, by monitoring solar radiation levels and conducting on-site assistant observations. This dataset will be divided into two subsets, one for conditions with no shading and the other for conditions with shading. It will serve as the basis for our in-depth examination of the effect of outdoor shading conditions on thermal comfort. Using this dataset as a foundation, we investigated the performance differences between various machine learning algorithms for predicting outdoor thermal comfort. These algorithms encompass traditional methods such as MLR and KNN, as well as relatively newer approaches like XGBoost, LightGBM, and CatBoost. Using Bayesian optimization (BO) models, we chose the most appropriate machine learning models for various scenarios, including shaded and unshaded conditions, in an effort to improve the accuracy and dependability of outdoor thermal comfort forecasts. Finally, we applied the machine learning explainability model SHapley Additive xExplanations (SHAP) to obtain importance rankings of different features within the optimal machine learning model and evaluate their positive and negative effects on thermal comfort. This study contributes to a greater comprehension of the operational mechanisms of the model and provide support for its further optimization.

2. Methodology

The research methodology is divided into five main components: data collection, data preprocessing, model training, model optimization, and model interpretation. Fig. 1 is a flowchart illustrating the research framework used in this study. Initially, in accordance with the research

objectives and literature review, microclimate data and thermal comfort questionnaire data were collected from Xi'an's urban parks. Following data collection, a thorough data cleaning and normalization procedure was implemented. The processed data were then used to train nine different machine learning models and apply Bayesian Optimization to nine machine learning models, with the goal of obtaining optimal

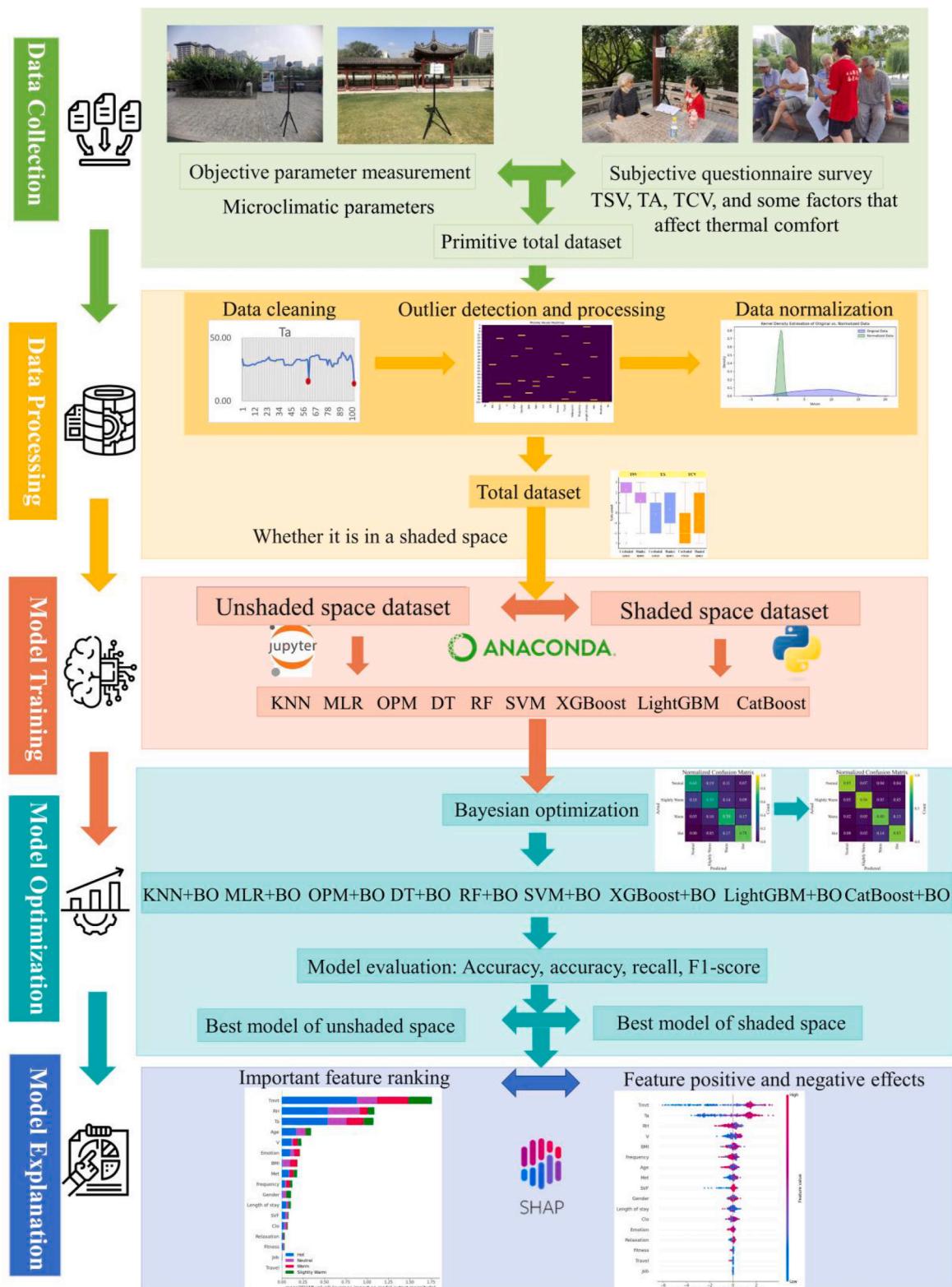


Fig. 1. Research framework.

performance models for both shaded and unshaded outdoor spaces. Finally, the SHAP method was used to interpret the best-performing models.

2.1. Outdoor data collection

2.1.1. Study area

Xi'an ($118^{\circ}5'18''$ E, $34^{\circ}15'44''$ N) is located in western China and serves as the capital city of Shaanxi Province. It is located in a cold climate zone in China and is influenced by the East Asian monsoon. According to the Köppen climate classification, it is located between a semi-arid climate (BSk) and a humid subtropical climate (Cwa). Summers are hot and humid, and winters are cold and dry [74]. Climate data from 2010 to 2019 show that the highest monthly average temperature (T_a) occurs in July (28.07°C), with a maximum temperature of 39.1°C . The lowest monthly average T_a is recorded in January (0.49°C), with a minimum temperature of -8.4°C . The annual average relative humidity ranges from 53.11% to 73.83% [75] (see Fig. 2).

The field experiment was carried out in the Xi'an Park around City, which is located in the heart of Xi'an (Fig. 3). Xi'an Park around City is a well-known urban park located just outside Xi'an's ancient city wall. It encircles the historic city wall and is known for its varied landscape elements, which include open-air plazas, traditional Chinese pavilions, Chinese-style corridors, Western-style promenades, and a protective moat. The park provides a relaxing setting for a variety of activities, allowing visitors to freely stroll, exercise, rest, and sightsee on the green expanses outside the city walls. It also provides an opportunity to learn about the history and culture of Xi'an, making it a popular social gathering place. To investigate the differences between thermal comfort prediction in unshaded spaces and shaded spaces, seven specific scenarios can be classified based on the approximate shading conditions of the specific research site. In unshaded spaces, there are primarily two landscape types: open ground (OG) and general pavement (GP), whereas shaded spaces have five: Western-style long corridor (WC), Chinese-style long corridor (CC), Chinese-style pavilion (CP), Lower tree (LT), and Tall tree (TT). However, depending on the solar azimuth, all scenarios have the potential to become unshaded or shaded. The surrounding environment influenced the local outdoor air temperature (T_a) within a central radius of 10–150 m [76]. As a result, the composition of each of the seven scene landscapes was measured from a point 10 m away from their physical centers (314 m^2). To compare the landscape characteristics and sky view factors of each space, quantitative assessments were performed (sky view factors were calculated using fisheye photos in Bmp format input into the Rayman software, see Table 1).

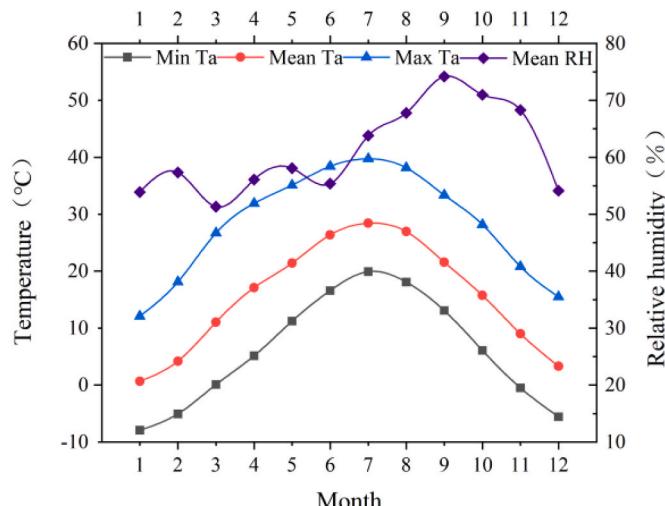


Fig. 2. Monthly mean/maximum/minimum T_a and mean RH in Xi'an from 2010 to 2019.

2.1.2. Procedure

This study included microclimate monitoring and questionnaire surveys from 9:30 a.m. to 6:00 p.m. An experimental assistant meticulously recorded the real-time shading conditions for each scene during this time. First, a one-day pre-experiment was conducted, and the objective parameters of that day were collected, as well as 20 subjective questionnaires distributed on site, to determine whether there were any flaws in the experimental procedure, whether the experimental instruments could be used normally, and whether the subjective questionnaire questions were straightforward and clear.

2.1.3. Physical measurements

Summer experiments were carried out in July and August of 2021 and 2022, coinciding with the hottest months in Xi'an during those years. In 2021, the experiment lasted 15 days, and in 2022, it lasted 11 days. Meteorological parameters were measured and recorded in various scenarios from 9:30 to 18:00 on each experimental day. The experimental assistant set up the instrument half an hour ahead of time at a height of 1.5 m above the ground. The temperature and humidity logger HOBO U12-011 (Onset), the Vientiane anemometer WFWWZY-1 (Tian Jian Hua Yi), the globe thermometer HQZY-1 (Tian Jian Hua Yi), and the solar radiation sensor JTR05 (Jiantong Technology) were among the instruments used. Measurements were taken every minute at 1-min intervals, capturing instantaneous values. Table 2 lists the specifications of the measuring instruments. All of the instruments used in this study met ISO 7726 standards, ensuring a quick response time and high accuracy [77]. Fig. 4 depicts the instrument used.

The mean radiant temperature T_{mrt} was calculated in accordance with the ISO 7726 standard, which is expressed by Equation (1) [77]:

$$T_{mrt} = \left[(T_g + 273)^4 + \frac{1.10 \times 10^8 V_a^{0.6}}{\varepsilon D^{0.4}} (T_g - T_a) \right]^{\frac{1}{4}} - 273 \quad (1)$$

where D is globe diameter (set to 0.15 m in this study); ε is emissivity (set to 0.95 for a black globe).

We manually fabricated radiation shields and conducted temperature and humidity calibration to account for the potential impact of outdoor solar radiation on temperature and humidity measurements during the summer season. This precautionary measure was implemented to reduce the impact of strong solar radiation on temperature and humidity measurements.

2.1.4. Questionnaire surveys

During the on-site experiments, seven research assistants were hired to help with questionnaire distribution. These research assistants were stationed near the measurement points and conducted the questionnaire surveys concurrently with the parameter measurements. When distributing the questionnaires, the research assistants first noted the questionnaire timestamp, whether the respondent was in a shaded area, and the specific location. The questionnaires were then filled out by the respondents. The outdoor thermal comfort survey included 2701 respondents ranging in age from 15 to 85 years old and taking place in both shaded and unshaded areas. All respondents were individuals who were freely engaged in activities within the park. Before completing the questionnaires or answering any questions, the majority of respondents had spent at least 20 min at the experimental site, acclimating to the outdoor thermal environment. A total of 2701 questionnaires were distributed, yielding 2652 valid responses. Fig. 5 depicts the simplified questionnaire.

The questionnaire design in this study adhered to ISO 10551 standards and consisted of two main sections: General Information and Thermal Sensation Information [78]. Respondents were asked to provide information such as their gender, age, height, weight, clothing, physical activities, purpose of park visit, visit frequency, duration of stay in the park, and emotional state in the General Information section. The Thermal Sensation questions were designed to assess respondents'

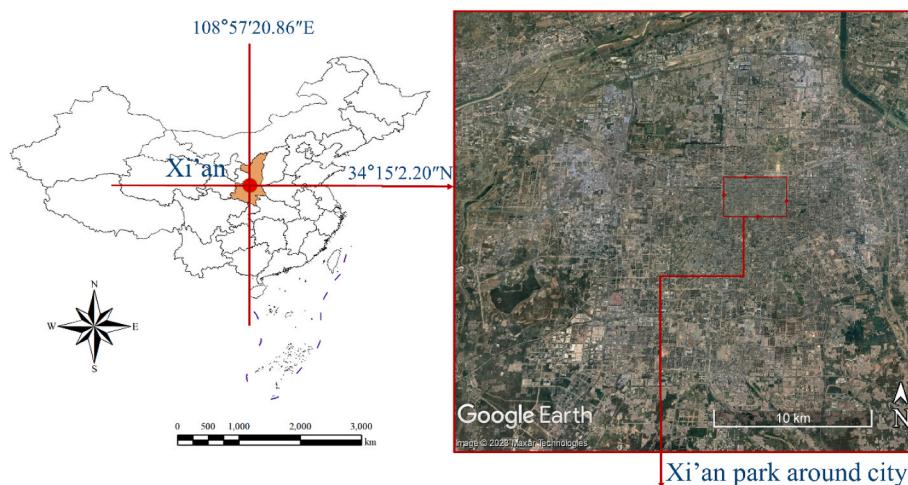


Fig. 3. Study site locations.

perceptions of thermal comfort in various environments. This section included the thermal sensation vote (TSV), thermal acceptability vote (TA), and thermal comfort vote (TCV). Following ASHRAE Standard 55–2020 definitions and accounting for local weather conditions, the TSV used a modified 7-point scale (−3: very cold; −2: cold; −1: cool; 0: neutral; 1: warm; 2: hot; 3: very hot). Meanwhile, the TA used a 5-point scale (−2: very unacceptable; −1: unacceptable; 0: neutral; 1: acceptable; 2: very acceptable). The TCV used a 7-point scale (−3: very uncomfortable; −2: uncomfortable; −1: slightly uncomfortable; 0: neutral; 1: slightly comfortable; 2: comfortable; 3: very comfortable).

The majority of the questions in the first section focused on factors that influence thermal sensation, excluding physiological factors. These included individual factors (gender, age, height, weight, clothing, and activity type), social factors (occupation and purpose of park visit), and psychological factors (city of residence, emotional state, visit frequency, and length of stay in the park). The Body Mass Index (BMI) is a common indicator of the relationship between a person's level of obesity and their thermal comfort. The formula for calculating BMI from a person's height and weight is Equation (2) [79]:

$$BMI = \frac{Weight(Kg)}{Height(m)^2} \quad (2)$$

In accordance with the ASHRAE Standard 55–2020 and ISO 7730 Standard, clothing insulation values and metabolic rates were determined based on respondents' clothing and activity types (Details are provided in Appendix 1–Tables A1 and A2) [80,81]. During the experimental process, research assistants attempted to balance the age and gender distribution of the participants in order to reduce sampling bias and variability.

2.2. Outdoor thermal comfort physical model

PET and UTCI stand as the primary physical models for predicting outdoor thermal comfort. However, despite their widespread use, these physical models do not fully account for the influence of individual differences, including the effects of diverse thermal histories on outdoor thermal comfort. Consequently, the accuracy of predictions is limited. However, as data science has advanced and computer capabilities have improved, data-driven machine learning has gradually improved the accuracy of TSV prediction, surpassing traditional physical models.

2.3. Machine learning model establishment and optimization

First, we preprocessed the dataset, removing outliers and normalizing features to ensure model stability and performance. The outdoor

shaded and unshaded datasets were then divided into training and testing sets. On the total dataset, unshaded space dataset, and shaded space dataset training sets, we used machine learning algorithms to build multi-class prediction models for human thermal sensation, thermal acceptability, and thermal comfort. On the testing set, we compared prediction results across different datasets to investigate the importance of distinguishing between unshaded and shaded outdoor thermal comfort datasets.

Second, we performed Bayesian optimization on the original machine learning models' hyperparameter space. We established parameter ranges for various algorithms, such as learning rate, maximum depth, and minimum samples per leaf node. We used Bayesian optimization to find the best parameter combinations within the hyperparameter space to maximize cross-validation average accuracy (using 5-fold cross-validation). We computed and analyzed the normalized confusion matrix data values for the original and Bayesian-optimized models.

Following that, we built nine machine learning multi-class models based on the best hyperparameters and used the SHAP model to interpret the models. We investigate the significant influencing characteristics of thermal sensation, thermal acceptability, and thermal comfort in unshaded and shaded spaces using the SHAP model, and determine the positive and negative effects of each feature, which helps to explain the model's prediction results and provides a foundation for further model optimization. The specifics are provided below.

2.3.1. Data preprocess

Data cleansing and normalization are examples of data preprocessing methods. When there were outliers in the dataset, such as data recording errors or questionnaire completion errors, we removed these samples. Following the initial cleaning, the dataset contained a total of 2651 data sets, with 1299 sets in the no-shade dataset and 1352 sets in the shade dataset. In this study, we randomly divided the dataset into a training set (80%) and a test set (20%).

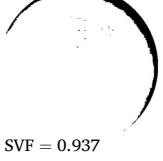
To deal with data in different units, we used the min-max scaling technique for data normalization. This method converts the data to a range between 0 and 1, allowing the input data to be scaled within this range. This process's specific formula is Equation (3) [62].

$$Data_{scaled} = \frac{Data - Data_{min}}{Data_{max} - Data_{min}} \quad (3)$$

where $Data_{scaled}$ is the normalized value of the input data; $Data$ is the input data; $Data_{max}$ is the maximum value of the input data; $Data_{min}$ is the minimum value of the input data.

Table 1

Descriptions of site configurations.

Space	Space characteristics	Plot photo	Fish-eye photo	Photo was processed by Rayman
Western corridor (WC)	Semi-open space with climbing plants at the top. (Shaded space)			 SVF = 0.077
Chinese corridor (CC)	The roof is wooden with unique arcades, carvings and decorative semi-open spaces. (Shaded space)			 SVF = 0.084
Chinese pavilion (Cp)	Semi-open space with traditional architectural style and exquisite decoration. (Shaded space)			 SVF = 0.159
Lower tree (LT)	Open space surrounded by shrubs and low trees. (Shaded space)			 SVF = 0.179
Tall tree (TT)	Open space surrounded by large trees. (Shaded space)			 SVF = 0.394
General pavement (GP)	General pavement with some trees around. (Unshaded space)			 SVF = 0.617
Open ground (OG)	Very empty open space around. (Unshaded space)			 SVF = 0.937

Note: SVF: Sky view factor.

Table 2

Instruments used for measurement of micrometeorological parameters.

Measured parameters	Sensor	Range	Accuracy	Height above the ground
Air temperature	Temperature and humidity recorder HOBO U12-011	-20~70 °C	±0.35 °C @0~50 °C	1.5 m
Relative humidity		5~95 % RH	±2.5%RH @10~90%RH	1.5 m
Globe temperature	Black ball thermometer HQZY-1	-20~80 °C	±0.3 °C	1.5 m
Wind speed	Vientiane anemometer WFWWZY-1	0.05~30 m/s	±0.05 m/s	1.5 m
Global radiation	Solar radiometer JTR05	0~2000 W/m ²	≤±2%	1.5 m



Fig. 4. Instruments for measuring.

2.3.2. Class reduction

Thermal feedback information from participants, including thermal sensation votes, thermal acceptability votes, and thermal comfort votes, was used as labels for training the models in this study. During the experimental phase, we gathered 7-point thermal sensation votes. Because the no-shade environment is generally hot during the summer, there were fewer selections for '-3' (very cold), '-2' (cold), and '-1'

(cool). Recognizing that data imbalance could lead to overfitting, we chose to remove data points corresponding to the '-3', '-2', and '-1' categories. This conversion reduced the original 7-class classification problem to a 4-class classification problem, a strategy supported by Yang et al. [56,79]. Similarly, in shading spaces, '-3' and '-2' options are minimal, so the 7 classification problem can be reduced to 5 classification.

Thermal acceptability was graded on a 5-point scale. Although there were fewer people who chose '2' (very acceptable) in the no-shade environment and '2' (very unacceptable) in the shade environment, the category imbalance was not significant. As a result, no category reduction was applied to the thermal acceptability data. In terms of thermal comfort, the 7-point data in the shade environment were relatively balanced, and category reduction was unnecessary. However, in the absence of shade, only a small number of participants chose '2' (comfortable) and '3' (very comfortable). To address the limited data

Outdoor thermal comfort questionnaire																																							
School of Construction Equipment Science and Engineering, Xi'an University of Architecture and Technology																																							
Data: _____ / _____	Time: _____ :			Location: _____																																			
1		2																																					
Unshaded space	Shaded space	Tall tree	Lower trees	Chinese corridor	Western corridor	Chinese pavilion	General pavement																																
Open ground																																							
1.Gender:Male/Female	2.Height(cm):	3.Weight(kg):																																					
4.Year: ①<10 ②11-20 ③21-30 ④31-40 ⑤41-50 ⑥51-60 ⑦61-70 ⑧71-80 ⑨>81																																							
5.The purpose of being here: ①fallow ②fitness ③journey ④work																																							
6.The frequency of coming here: _____																																							
7.Length of stay (minutes): ①30 ②60 ③90 ④120 ⑤150 ⑥180																																							
8.Ongoing activities: ①Chat (site) ②Chat (sit) ③rest ④Go for a walk ⑤Play chess ⑥badminton ⑦Table tennis ⑧running ⑨Square dance ⑩shadowboxing																																							
9.Mood: (1-Very negative 2-More negative 3-normal 4-More active 5-Very positive)																																							
10.Please describe how you feel now:	<table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse;"> <tr> <td style="text-align: center;">-3</td> <td style="text-align: center;">-2</td> <td style="text-align: center;">-1</td> <td style="text-align: center;">0</td> <td style="text-align: center;">+1</td> <td style="text-align: center;">+2</td> <td style="text-align: center;">+2</td> </tr> <tr> <td style="text-align: center;">Cold</td> <td style="text-align: center;">Cool</td> <td style="text-align: center;">Slightly cool</td> <td style="text-align: center;">Neutral</td> <td style="text-align: center;">Slightly warm</td> <td style="text-align: center;">Warm</td> <td style="text-align: center;">Warm</td> </tr> </table>							-3	-2	-1	0	+1	+2	+2	Cold	Cool	Slightly cool	Neutral	Slightly warm	Warm	Warm																		
-3	-2	-1	0	+1	+2	+2																																	
Cold	Cool	Slightly cool	Neutral	Slightly warm	Warm	Warm																																	
11.Please describe your current thermal acceptability:	<table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse;"> <tr> <td style="text-align: center;">-2</td> <td style="text-align: center;">-1</td> <td style="text-align: center;">0</td> <td style="text-align: center;">+1</td> <td style="text-align: center;">+2</td> <td colspan="2"></td> </tr> <tr> <td style="text-align: center;">Unacceptable</td> <td style="text-align: center;">Slightly Unacceptable</td> <td style="text-align: center;">Neutral</td> <td style="text-align: center;">Slightly Acceptable</td> <td style="text-align: center;">Acceptable</td> <td colspan="2"></td> </tr> </table>							-2	-1	0	+1	+2			Unacceptable	Slightly Unacceptable	Neutral	Slightly Acceptable	Acceptable																				
-2	-1	0	+1	+2																																			
Unacceptable	Slightly Unacceptable	Neutral	Slightly Acceptable	Acceptable																																			
12.Please describe your current thermal comfort:	<table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse;"> <tr> <td style="text-align: center;">-3</td> <td style="text-align: center;">-2</td> <td style="text-align: center;">-1</td> <td style="text-align: center;">0</td> <td style="text-align: center;">+1</td> <td style="text-align: center;">+2</td> <td style="text-align: center;">+3</td> </tr> <tr> <td style="text-align: center;">Very Uncomfortable</td> <td style="text-align: center;">Uncomfortable</td> <td style="text-align: center;">Slightly Uncomfortable</td> <td style="text-align: center;">Neutral</td> <td style="text-align: center;">Slightly Comfortable</td> <td style="text-align: center;">Comfortable</td> <td style="text-align: center;">Very Comfortable</td> </tr> </table>							-3	-2	-1	0	+1	+2	+3	Very Uncomfortable	Uncomfortable	Slightly Uncomfortable	Neutral	Slightly Comfortable	Comfortable	Very Comfortable																		
-3	-2	-1	0	+1	+2	+3																																	
Very Uncomfortable	Uncomfortable	Slightly Uncomfortable	Neutral	Slightly Comfortable	Comfortable	Very Comfortable																																	
13.How do you want the following meteorological environment parameters to change (-1 means need to be reduced, 0 means need not change, 1 means need to increase):	<table border="1" style="margin-left: auto; margin-right: auto; border-collapse: collapse;"> <tr> <td style="text-align: center;">Temperature</td> <td style="text-align: center;">-1</td> <td style="text-align: center;">0</td> <td style="text-align: center;">1</td> <td colspan="4"></td> </tr> <tr> <td style="text-align: center;">Humidness</td> <td style="text-align: center;">-1</td> <td style="text-align: center;">0</td> <td style="text-align: center;">1</td> <td colspan="4"></td> </tr> <tr> <td style="text-align: center;">Solar radiation</td> <td style="text-align: center;">-1</td> <td style="text-align: center;">0</td> <td style="text-align: center;">1</td> <td colspan="4"></td> </tr> <tr> <td style="text-align: center;">Wind</td> <td style="text-align: center;">-1</td> <td style="text-align: center;">0</td> <td style="text-align: center;">1</td> <td colspan="4"></td> </tr> </table>							Temperature	-1	0	1					Humidness	-1	0	1					Solar radiation	-1	0	1					Wind	-1	0	1				
Temperature	-1	0	1																																				
Humidness	-1	0	1																																				
Solar radiation	-1	0	1																																				
Wind	-1	0	1																																				
14.What you're wearing now:																																							
Default clothing: Top: Short-sleeved Long-sleeved Sun-protective clothing																																							
Bottoms: Shorts Skirts Trousers Long skirts																																							
Shoes: Low tops (Sandals or not) High tops																																							
Other clothing: Hat Sunglasses Mouth/Mask Ice sleeves																																							
More clothes: _____																																							
I wish you good health and all the best!																																							

Fig. 5. Simplified questionnaire.

volume, we transformed the original 7-class classification problem into a 5-class classification problem.

2.3.3. Machine learning algorithms

The machine learning algorithms selected in this study include OPM, DT, MLR, KNN, RF, SVM, XGBoost, LightGBM and CatBoost. The key characteristics of each algorithm are displayed in Table 3.

2.3.4. Hyperparameter optimization

To find the best hyperparameter configuration for each machine learning model, Bayesian optimization is used. Bayesian optimization, in contrast to the traditional GridSearch method, models the prior distribution of the objective function and gradually searches for the best solution in the candidate hyperparameter space [89].

In practice, each model's hyperparameter range and prior distribution are initially defined. Then, using Bayesian optimization, find the best hyperparameter configuration that maximizes model performance within the candidate hyperparameter space. This method allows for smarter hyperparameter searches, fewer unnecessary experiments, and more efficient discovery of the best hyperparameter combination [90].

2.3.5. Model evaluation

It is obvious that during the summer, there will be more people who perceive heat than those who perceive cold. As a result, it is critical to accurately identify those who are likely to experience heat discomfort. To address potential category imbalances, machine learning models that better predict summer thermal sensations are selected, and machine learning model performance metrics are evaluated using a weighted

average evaluation method. The basic idea behind this method is to compute the weighted average of performance metrics for each category (for example, accuracy, precision, recall, F1-score, and so on) based on the weights assigned to each category. The number of samples in the dataset that belong to each category determines the weights for that category. Because the distribution of samples among different categories may be uneven, with some categories having a large number of samples while others having only a few, the Weighted-average assessment method takes these category weights into account to provide a more equitable evaluation of performance, thus mitigating class imbalance issues. Appendix 2- Equations (A1)-(A4) show the calculation method for each performance metric in each category.

2.4. Model explanation

Despite complex machine learning models' remarkable predictive accuracy, they are frequently regarded as "black boxes," with no knowledge of the internal mechanisms, the impact of individual features on model outputs, or whether these features contribute positively or negatively. This lack of interpretability is a major impediment to further advancements and improvements in machine learning models. Researchers have proposed various methods to improve our understanding and interpretability of these models, with SHAP being one of the most widely used and highly regarded. The SHAP model is based on the cooperative game theory Shapley value concept, which treats each feature as a player in a game, the model's output as the game's outcome, and SHAP values to quantify the contributions of each feature to the game's outcome [91]. The SHAP model assigns SHAP values to each

Table 3

A brief overview of the 9 machine learning models and their hyperparameter spaces.

Algorithms	Features	Hyperparameters
KNN	KNN classifies based on the distance between samples, making classification decisions by comparing the nearest neighbor samples [82].	'leaf_size', 'N_neighbors', 'P'
MLR	MLR is a classification method that extends binary logistic regression to handle multiple classes, predicting the most probable class for a given input [64].	'C', 'max_iter'
OPM	OPM is a statistical approach that models the probabilities of ordered categorical outcomes, suitable for ranking or ordinal data analysis [64].	'Alpha'
DT	DT is a predictive modeling approach that uses a tree-like structure to make decisions based on features, facilitating the visualization of potential outcomes through hierarchical choices [83].	'Max_depth', 'min_samples_leaf', 'Min_samples_split'
RF	RF is an ensemble learning technique that aggregates multiple decision trees to enhance prediction accuracy, reduce overfitting, and capture complex relationships in data [84].	'min_samples_split', 'n_estimators'
SVM	SVM is a robust supervised learning method that identifies optimal decision boundaries by maximizing the margin between classes, effectively classifying data points and handling non-linear relationships through kernel functions [85].	'C', 'degree', 'kernel'
XGBoost	XGBoost is a gradient boosting algorithm that sequentially combines weak learners, such as decision trees, through optimization techniques to achieve high predictive accuracy and handle complex relationships within data [86].	'colsample_bytree', 'gamma', 'learning_rate', 'max_depth', 'min_child_weight', 'reg_alpha', 'reg_lambda', 'subsample'
LightGBM	LightGBM is a gradient boosting framework that employs histogram-based techniques for efficient decision tree construction, resulting in faster training, higher efficiency, and improved predictive accuracy for large-scale datasets [87].	'learning_rate', 'max_depth', 'min_child_samples'
CatBoost	CatBoost is an enhanced machine learning approach that incorporates boosting algorithms tailored to effectively handle categorical features, contributing to improved predictive accuracy in datasets containing such variables [88].	'depth', 'iterations', 'l2_leaf_reg', 'learning_rate'

feature using its built-in calculation methods, allowing for a comprehensive ranking of feature importance within the model. This methodology is currently being used extensively in a variety of domains, including thermal comfort prediction [92–94]. Equation (4) contains specific details on how to calculate SHAP values [95]:

$$f(\vec{X}) = J_0 + \sum_{i=1}^K J_i X_i \quad (4)$$

where f is an interpretation model; \vec{X} indicates whether input parameter i is present in the set of input parameters, and i is in the set of input parameters ($X'_i = 1$), i is not in the set of input parameters ($X_i' = 0$); K is the number of input parameters; J is the attribution of input parameters.

SHAP computes the contribution of each input feature to a sample by training two models: one that excludes the feature and one that includes it. The difference in the outputs of these two models represents the contribution of that input feature [96]. The overall feature importance for the entire dataset is calculated as the weighted average of each feature's contribution across all samples [97]. To conduct this analysis, we used Python 3.8.0's SHAP package, ranking the importance of various features affecting outdoor thermal comfort and determining whether their impact is positive or negative.

In terms of visualization, we used summary plots to demonstrate the importance ranking of features for each class. The horizontal axis in these plots represents the absolute values of SHAP, while the vertical axis shows the rankings of the sum of absolute SHAP values across all samples. We also used swarm plots to show the positive or negative impact of each feature on the model's output. In these plots, the horizontal axis represents SHAP values, with positive values indicating a positive impact of the feature on the model's output and negative values indicating a negative impact. Higher SHAP values indicate that the feature has a greater influence on the model's output. Furthermore, the color of each sample reflects the variation in that sample's feature value compared to the variation in feature values across all samples (in general, red represents a large variation, while blue represents a small variation) [98].

2.5. Software application

Based on Python 3.8, this study used data preprocessing, machine learning model development, model optimization, and model interpretation. The Jupyter Notebook platform was used to conduct and compile the entire process.

3. Dataset

There is currently a scarcity of comprehensive cross-sectional survey data for outdoor thermal comfort, as well as very few open-source databases for modeling and analysis. As a result, an independent outdoor thermal comfort dataset had to be created. Sections 3.1 and 3.2 contain specific results from a detailed analysis of the features of microclimate data and thermal comfort questionnaire data. As described in Section 3.3, the dataset has been divided into three categories: the total dataset, the dataset for unshaded spaces, and the dataset for shaded spaces.

3.1. Microclimatic data

Table 4 displays the statistical data for environmental parameters. The table shows that the microclimate parameters of the unshaded space differ significantly from those of the shaded space (see Table 5).

3.2. Thermal comfort questionnaire data

Male respondents in the shaded area have an average age of 46.32 years, a height of 172.36 cm, and a weight of 68.67 kg. Female

Table 4

Summary of Microclimate parameters of shaded space and unshaded space in the study site.

Variable	Shaded space dataset		Unshaded space dataset	
	Mean (std)	Min/Max	Mean (std)	Min/Max
T_a (°C)	32.09 (2.38)	21.23/39.12	33.06 (2.28)	24.88/43.25
RH (%)	60.5 (6.2)	38.5/85.8	57.8 (5.1)	30.3/81.0
T_g (°C)	33.3 (2.9)	22.9/45.8	39.5 (5.2)	23.9/55.7
T_{mrt} (°C)	34.90 (4.80)	21.31/60.71	49.24 (12.79)	25.90/93.68
V (m/s)	0.49 (0.53)	0.05/3.27	0.76 (0.82)	0.06/5.42
G (W/m ²)	54 (27)	6/150	264 (303)	8/933
Number of obs.	1352		1299	

Table 5

Volunteers attributes.

Space type	Gender	Number	Age	Height	Weight
			(years)	(cm)	(kg)
Shaded space	Male	802	46.32	172.36	68.67
	Female	550	42.42	161.75	56.24
Unshaded space	Male	674	46.57	171.19	67.93
	Female	605	44.90	161.97	56.48

respondents in the shaded area are 42.42 years old on average, stand 161.75 cm tall, and weigh 56.24 kg. Male respondents in the unshaded space have an average age of 46.57 years, a height of 171.19 cm, and a weight of 67.93 kg. In the unshaded space, female respondents have an average age of 44.9 years, a height of 161.97 cm, and a weight of 56.48 kg. Although there are minor demographic differences between the two space types, these differences are not statistically significant.

In the unshaded space, respondents have an average clothing thermal resistance of 0.43 clo and an average metabolic rate of 2.23 met, respectively, whereas in the shaded space, respondents have an average clothing thermal resistance of 0.42 clo and an average metabolic rate of 1.92 met. The frequency of respondents' park visits has been converted to the number of times per month. Respondents in the unshaded space visit the park 4.24 times per month on average, while respondents in the shaded space visit the park 4.33 times per month.

Fig. 6 depicts the information related to thermal indices. In terms of TSV, the average value in the unshaded space is 2.31, whereas it is only 1.21 in the shaded space. Similarly, for TA, the average value in the unshaded space is -0.16, while it is 0.38 in the shaded space, indicating that the unshaded space provides more thermal comfort. The average value of TCV in the unshaded space is -1.57, while it is -0.43 in the shaded space. This could imply a more positive assessment of thermal comfort in the shaded area. We can see from the analysis of these data that shading has a significant impact on thermal comfort. It should be noted that the data's standard deviation is large, and people's voting results are somewhat scattered, which could be attributed to individual differences, rapid changes in the outdoor environment, and other factors.

3.3. Split the dataset based on shading conditions

The outdoor thermal comfort dataset constructed in cold areas can be divided into two different datasets based on shading conditions, and the three datasets are introduced below.

Total Dataset: This dataset contains all data, excluding outliers. It consists of 18 features and three labels that cover various aspects of thermal comfort. Air temperature, humidity, wind speed, average radiant temperature, sky landscape factor, gender, BMI, age, clothing thermal resistance, purpose of visit (work, leisure, exercise, tourism), frequency of visits, length of stay, metabolic rate of the activity performed, and emotion are all characteristics of the scene. Labels include

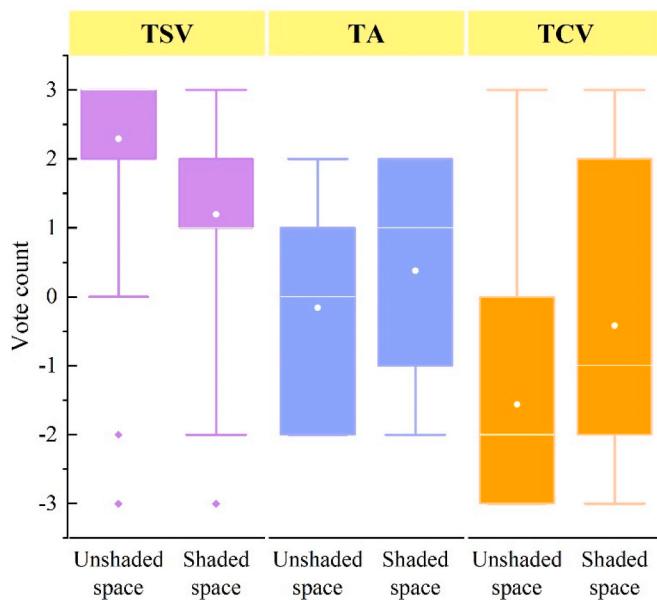


Fig. 6. Comparison of heat voting between unshaded spatial data set and shaded spatial dataset.

TSV, TA, and TCV.

Unshaded Space Dataset: This dataset is a subset of the total dataset that includes all data from unshaded spaces while excluding outliers. It has 17 features and 3 labels. The only difference between this dataset and the total dataset is the absence of the feature indicating whether it is a shaded space.

Shaded Space Dataset: Like the unshaded space dataset, this dataset is a subset of the total dataset and includes all data from shaded spaces while excluding outliers. It also has 17 features and 3 labels, with the same removal of the feature indicating whether it is a shaded space compared to the total dataset.

4. Results

4.1. Differences in machine learning models across different data sets

Appendix 3-Table A3 shows the performance of various machine learning algorithms for thermal sensation prediction on different datasets, with all accuracy, precision, recall, and F1-Score values weighted. Overall, there are some differences in performance across datasets. The default machine learning model performs fairly consistently in both the total dataset and the shaded space dataset, with average accuracy and F1-Scores around 50%. However, the overall model performance improves in the unshaded space dataset, with an average accuracy and F1-Score of around 65%.

CatBoost, with an accuracy of 65.49% and an F1-Score of 65.94%, is the best-performing model in the entire dataset. The CatBoost model outperforms the worst-performing model OPM in the entire dataset by 17.28% and 13.99%, respectively. The top-performing model in the shaded space dataset is XGBOOST, with an accuracy of 65.31% and an F1-Score of 66.95%, representing a 22.51% and 19.82% improvement over the worst-performing model OPM.

CatBoost is the best-performing model in the unshaded space dataset, with an accuracy of 73.08% and an F1-Score of 74.4%. This model outperforms the worst-performing model OPM in accuracy and F1-Score by 20.24% and 17.23%, respectively. Furthermore, when compared to the best-performing default CatBoost model in the total dataset, the top-performing default CatBoost model in the unshaded space dataset improved accuracy and F1-Score by 7.59% and 8.46%, respectively. In contrast, despite having half the training data, the best-performing

default CatBoost model in the shaded space dataset performs nearly equally to the best-performing default model in the total dataset.

Appendix 3-Tables A4 and A5 show the performance of various machine learning models for thermal acceptability and thermal comfort prediction on different datasets, where all accuracy, precision, recall, and F1-Score values are weighted. Similar to thermal sensation prediction models, machine learning models in the unshaded space dataset exhibit strong performance with relatively less data, while models in the shaded space dataset show similar performance to models in the total dataset with relatively less data. Overall, CatBoost models outperform XGBoost and LightGBM in TSV, TA, and TCV prediction.

4.2. Model optimization and selection

Nine different machine learning models were fine-tuned using Bayesian optimization techniques to improve their predictive performance. Several preliminary experiments were conducted to determine the hyperparameter space for each machine learning model before conducting formal hyperparameter space experiments. For example, for the XGBoost model, the defined hyperparameter space was as follows: {'learning_rate': (0.01, 0.5), 'max_depth': (3, 15), 'min_child_weight': (1, 10), 'gamma': (0, 0.5), 'subsample': (0.5, 1.0), 'colsample_bytree': (0.5, 1.0), 'reg_alpha': (0, 1.0), 'reg_lambda': (0, 1.0)}, with 500 iterations. To ensure model stability, each iteration used 5-fold cross-validation.

Tables A3–A5 in Appendix 3 compare the results of TSV, TA, and TCV prediction models before and after Bayesian optimization. The results show that Bayesian optimization improved the prediction accuracy of TSV, TA, and TCV significantly. Specifically, in TSV prediction, the CatBoost Classifier with Bayesian Optimization (CatBoost + BO) stood out as the optimal model in unshaded space, exhibiting an improvement of 9.61% in accuracy and 7.76% in F1-Score compared to the best model, CatBoost, before optimization. The optimal model in shaded space was XGBoost + BO, showcasing improvements in accuracy and F1-Score of 7.75% and 6.65%, respectively, compared to the best model, XGBoost, before optimization.

In TA prediction, LightGBM + BO emerged as the superior model in unshaded space, with improvements of 9.61% in accuracy and 9.88% in F1-Score compared to the best model, LightGBM, before optimization. CatBoost + BO claimed the top position in shaded space, showing enhancements of 3.33% in accuracy and 2.55% in precision compared to the best model, CatBoost, before optimization.

In TCV prediction, CatBoost + BO emerged as the optimal model in unshaded space, showcasing improvements of 6.54% in accuracy and 6.76% in F1-Score compared to the best model, CatBoost, before optimization. Similarly, in shaded space, CatBoost + BO claimed the top position, exhibiting enhancements of 6.65% in accuracy and 7.94% in F1-Score over the best model, CatBoost, before optimization. As before, machine learning models in unshaded spaces performed better when dealing with less data than models in the total dataset. Machine learning models in shaded spaces, on the other hand, showed no significant difference in performance when dealing with less data than models in the total dataset. However, there are still differences in performance among different machine learning models.

The normalized confusion matrices for TSV, TA, and TCV predictions on the test set for the best machine learning models in unshaded and shaded spaces before and after Bayesian optimization are shown in **Figs. 7–9**. The horizontal axis in these confusion matrices represents predicted values, while the vertical axis represents actual values. The diagonal values represent the prediction accuracy for each class label. You can see the model's predictions for each class label by comparing the color bars on the right side of the subplots. The results show that after Bayesian optimization, all models significantly improved prediction accuracy, particularly in cases where the models had difficulty predicting certain class labels. Prior to Bayesian optimization, it was difficult to predict the "Neutral" class in unshaded conditions and the

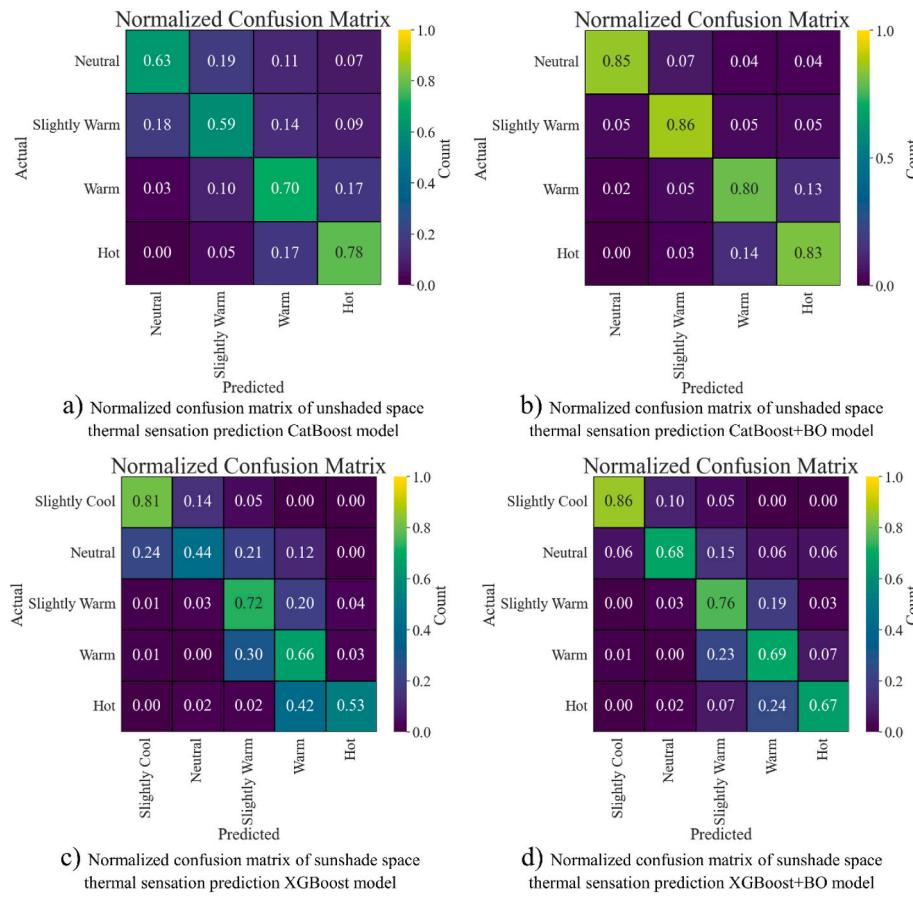


Fig. 7. Normalized confusion matrix of machine learning model for thermal sensation predictions.

“Neutral” and “Hot” classes in shaded conditions. However, after Bayesian optimization, accuracy improved significantly in these difficult scenarios. Despite this, some individual class labels exhibited low prediction accuracy, such as the “Slightly Comfortable” class in TCV in shaded conditions, which increased from 10% accuracy prior to optimization to only 24% after optimization. The CatBoost + BO model performed the best among the optimized models, followed by XGBoost + BO and LightGBM + BO, with an average improvement in accuracy of approximately 7%.

4.3. Model interpretation

4.3.1. Features importance ranking

As shown in Fig. 10, this study employs feature importance ranking based on SHAP values. The horizontal axis represents the average of the absolute values of SHAP values for all samples, indicating how input features contribute to the model’s output. A feature with a higher numerical value makes a greater contribution to the model. Different colors represent various TSV, TA, and TCV label categories.

The top ten features that have the most significant impact on the thermal sensation prediction model in unshaded outdoor spaces are shown in Fig. 10(a). Physical parameters such as mean radiant temperature, relative humidity, and air temperature have a significant influence on the overall model, according to the results. Furthermore, physiological and psychological factors such as age, emotion, and BMI have a significant impact. Although each category of thermal sensation prediction follows a consistent trend, relative humidity has the greatest impact when predicting “Neutral,” followed by mean radiant temperature and air temperature.

The influence of temperature on thermal sensation prediction in shaded conditions is shown in Fig. 10(b), with air temperature having a

significantly greater impact, though mean radiant temperature remains the most influential factor. Notably, for predictions of “Hot” and “Slightly Cool,” air temperature has a greater influence than mean radiant temperature.

Fig. 10(c) and (d) show the top ten features that have the greatest impact on thermal acceptability prediction models in unshaded and shaded outdoor spaces, respectively. According to the findings, mean radiant temperature has the greatest influence on thermal acceptability in unshaded spaces, while air temperature has the greatest influence in shaded spaces. It is important to note that emotion has a significant impact on thermal acceptability in both unshaded and shaded spaces.

Finally, in Fig. 10(e) and (f), we concentrate on predicting thermal comfort. The findings show that the primary factors influencing thermal comfort in unshaded spaces are mean radiant temperature, air temperature, emotion, relative humidity, and duration of stay in the park. The major influencing factors in shaded areas are air temperature, mean radiant temperature, relative humidity, emotion, wind speed, and sky view factor.

4.3.2. Features positive or negative effects

We converted these target labels into binary classifications in order to use the SHAP model to analyze the positive and negative effects of each feature on TSV, TA, and TCV. Larger values of TSV, TA, and TCV were labeled as 1, while smaller values were labeled as 0.

Fig. 11(a) and (b) show some of the key features’ positive and negative effects on thermal sensation prediction models in unshaded and shaded outdoor spaces, respectively. Blue represents smaller feature values, while red represents larger feature values, as indicated by the color bars on the right side of figures. A positive SHAP value indicates that the feature has a positive impact on thermal sensation, whereas a negative value indicates a negative impact. The magnitude of the SHAP

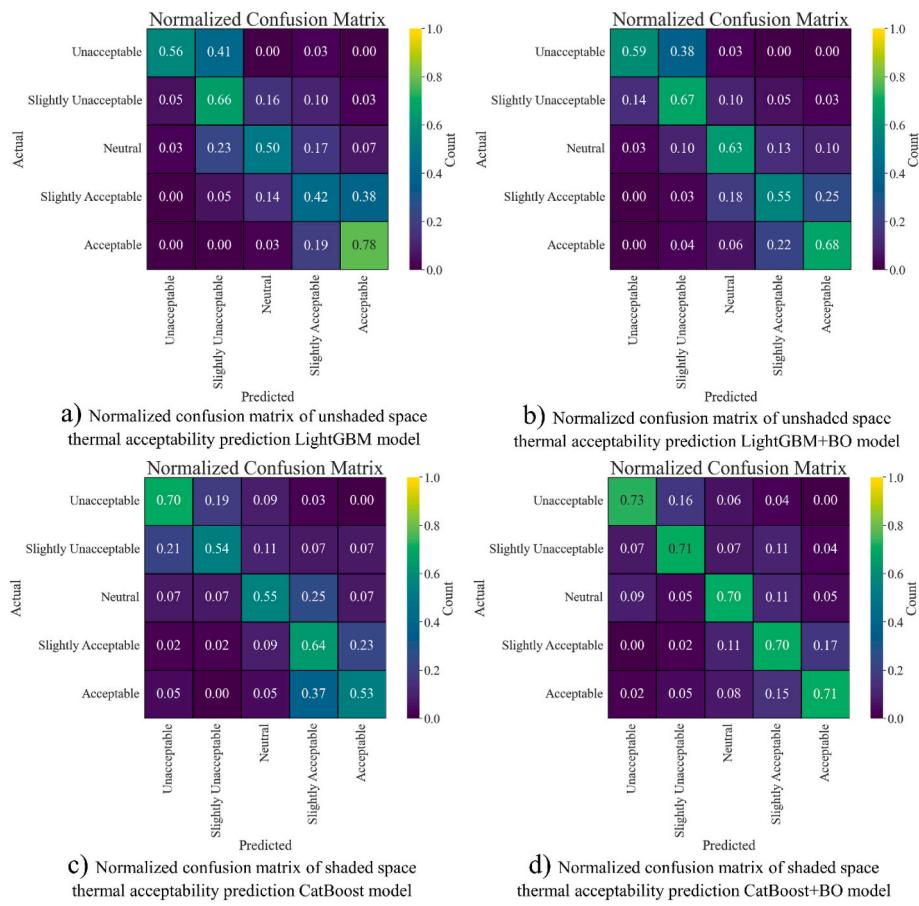


Fig. 8. Normalized confusion matrix of machine learning model for thermal acceptability predictions.

value reflects the influence of the feature on the model. Higher mean radiant temperature and air temperature have a significantly positive impact on thermal sensation, indicating that as mean radiant temperature and air temperature rise, people are more likely to perceive heat. Relative humidity has a negative impact on thermal sensation in unshaded outdoor spaces, while frequency of park visits has a pronounced negative impact, but wind speed has a positive impact. This suggests that higher relative humidity, more frequent park visits, and lower wind speeds contribute to a sense of coolness in unshaded outdoor spaces during the summer. Both relative humidity and clothing thermal resistance have a significant positive impact on thermal sensation in shaded outdoor spaces, while sky view factor and wind speed have a negative impact. This means that in summer, lower relative humidity, lower clothing thermal resistance, higher sky view factor, and higher wind speeds lead to a perception of coolness in shaded outdoor spaces.

The positive and negative impacts of partial important features on thermal acceptability prediction models in unshaded and shaded outdoor spaces are depicted in Fig. 11(c) and (d), respectively. Air temperature, mean radiant temperature, and wind speed all have a clear negative impact on thermal acceptability in unshaded outdoor spaces, indicating that people with better emotions have higher thermal acceptability under similar conditions. Increases in air temperature, relative humidity, and mean radiant temperature all have a negative impact on thermal acceptability in shaded outdoor spaces. Furthermore, age has a greater impact on thermal acceptability, with older people having higher thermal acceptability. Furthermore, we discovered that short-term heat history has a significant impact on thermal acceptability, implying that people who spend more time in the park have higher thermal acceptability.

The positive and negative impacts of partial important features on thermal comfort prediction models in unshaded and shaded outdoor

spaces are depicted in Fig. 11(e) and (f), respectively. Thermal comfort decreases in unshaded outdoor spaces as mean radiant temperature, air temperature, and BMI rise. Higher emotion scores and increased humidity are both associated with increased thermal comfort. Similarly, in shaded outdoor spaces, an increase in mean radiant temperature and air temperature is associated with decreased thermal comfort. Higher thermal comfort is also associated with higher emotion scores and higher relative humidity. Furthermore, age and length of stay in the park have a significant impact on thermal comfort. Higher thermal comfort is generally associated with improved emotion scores, and older people tend to have higher thermal comfort. Furthermore, we discovered that staying in the park for a longer period of time is typically associated with greater thermal comfort.

5. Discussion

5.1. Outdoor thermal comfort dataset segmentation

Previous research has found significant differences in thermal comfort prediction for different climatic conditions, indoor and outdoor environments, age groups, and physical activity levels [44–46,48, 50–57]. While research on these factors is relatively mature, studies on shading conditions are scarce and insufficiently comprehensive. Furthermore, differences in shading conditions are critical [69–73]. Currently, researchers use the SVF as a quantified input to represent shading conditions in models. However, SVF fails to fully capture a location's real-time shading situation because it remains a constant value while the position of the sun changes, resulting in dynamic shading conditions. In this study, a combination of solar radiometry and manual recording observations was used to dynamically assess shading conditions in various outdoor environments. These conditions were quantified

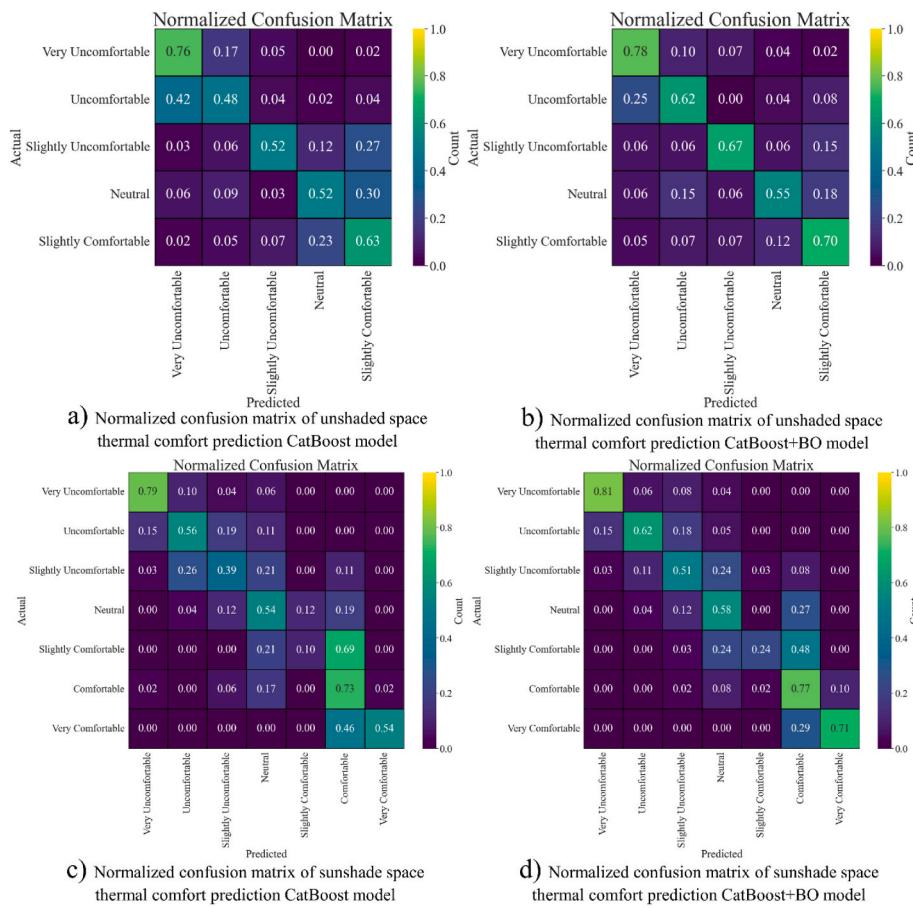


Fig. 9. Normalized confusion matrix of machine learning model for thermal comfort predictions.

using parameters such as average radiation temperature, and the outdoor datasets were then divided into unshaded and shaded space datasets. Statistical analysis reveals significant differences in environmental parameters and thermal comfort voting data for summer unshaded and shaded spaces. In particular, the average TSV in unshaded spaces reached 2.31, whereas it was only 1.21 in shaded spaces. [Tables A3–A5 in Appendix 3](#) compare model performance on various datasets. The results show that after dataset segmentation, the performance of TSV, TA, and overall TCV prediction models significantly improved. This may be attributed to the reduction of TSV categories from seven to four and five in unshaded and shaded spaces, respectively. It is worth noting that in unshaded areas, a significant proportion of people received a maximum thermal sensation score of 3. A 9-point thermal sensation scale could be investigated to more accurately assess outdoor human thermal sensation and mitigate data imbalances. To summarize, when the amount of data is insufficient and the characteristics are insufficiently comprehensive, it is necessary to divide the outdoor thermal comfort dataset into non-shaded space dataset and shading space dataset.

5.2. Differences in thermal sensation, thermal acceptance, and thermal comfort

To assess outdoor thermal comfort, a standardized 7-point thermal sensation scale, a 5-point thermal acceptance scale, and a 7-point thermal comfort scale were used in this study. In comparison to other metrics, the results show that thermal sensation has the highest predictive accuracy, reaching up to 82.69% in unshaded spaces, exceeding the highest predictive accuracy of 71.54% for thermal acceptance and 69.23% for thermal comfort. This could be because thermal sensation is

more strongly influenced by environmental factors and less so by other complex variables. Furthermore, the superior predictive performance of thermal sensation may be attributed to a reduction in the number of categories used during thermal sensation prediction [59,99]. Furthermore, a significant difference of approximately 8% was observed between the predictions for unshaded and shaded spaces, necessitating further investigation into the underlying causes. Average radiant temperature emerges as the most influential feature on thermal sensation in unshaded spaces, whereas air temperature takes precedence in shaded spaces. Humidity has a significant impact on thermal sensation in unshaded spaces, most likely due to the presence of interaction effects between high temperature and high humidity, amplifying the impact. Aside from average radiant temperature and temperature, emotion has a significant impact on thermal acceptance and thermal comfort. Positive emotions contribute to better environmental adaptation and an increased likelihood of achieving greater thermal comfort, which is consistent with previous research findings [100–102]. People with more positive emotions may show higher heat acceptance and thermal comfort in high heat-sensitive environments during the summer, according to partial sample data.

It is worth noting that the effect of metabolic rate was minor, in contrast to some previous research findings [100]. This disparity may be due to the relatively simple metabolic rate measurement methods used, which may not fully reflect actual human metabolic rates [93]. As a result, it is suggested that future research use more advanced measurement instruments to accurately assess human metabolic rates.

Finally, TSV, TA, and TCV can be used as feedback signals to guide thermal environment improvement device control. TCV aligns more closely with the ideal environmental state perceived by individuals than TA. Higher thermal comfort levels, on the other hand, may necessitate a

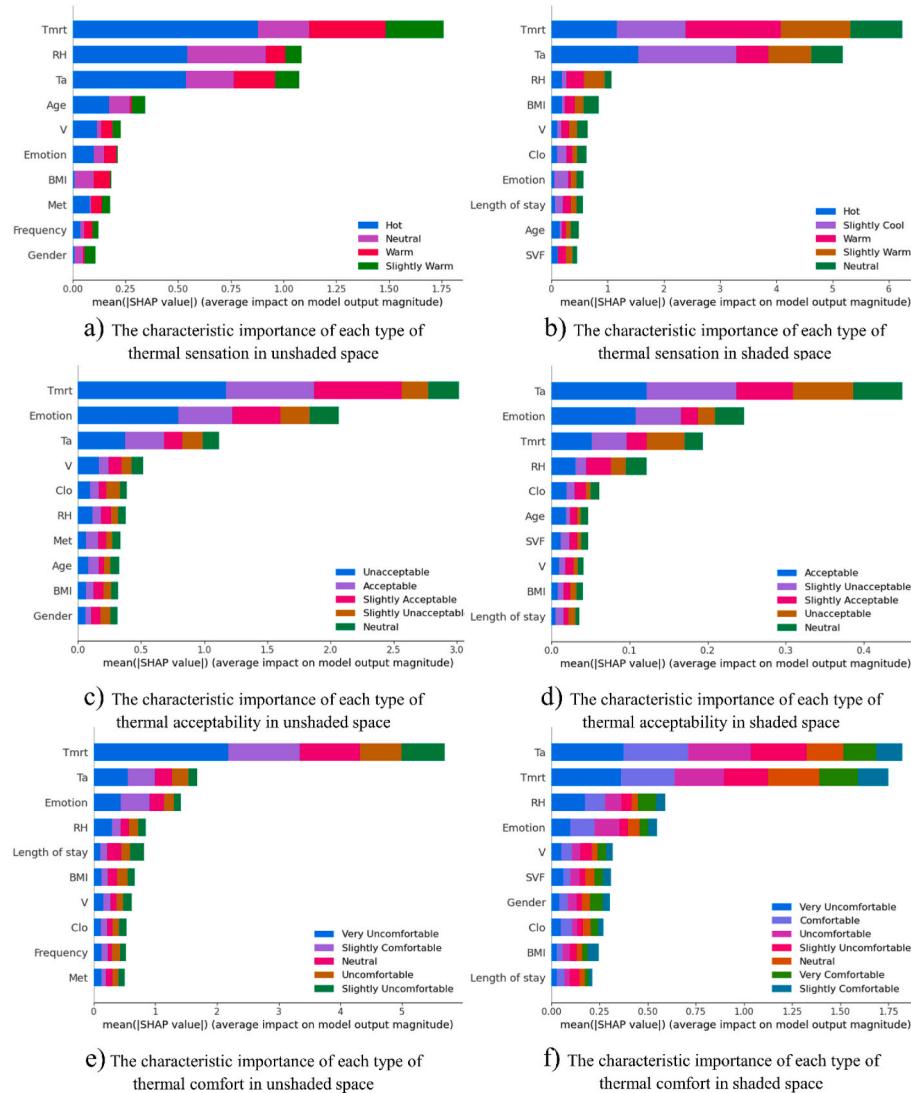


Fig. 10. SHAP summary of optimal TSV, TA, TCV prediction models for unshaded space and shaded space.

larger investment [59]. Because of the openness and complexity of outdoor environments, improving the thermal environment may necessitate more energy. As a result, it is best to start with TA as a guide for controlling outdoor thermal environment improvement devices and then consider incorporating thermal sensation to create new metrics for feedback signals.

5.3. Comparison of different machine learning models

Previous research has investigated various algorithms for thermal comfort prediction in terms of comparing different machine learning models. Researchers discovered that KNN performed best when predicting the 3-point indoor individual TSV after category reduction [59]. Other research has shown that MLR and SVM have promising predictive performance [36,41,42,62–64]. Recent research results, on the other hand, show that ensemble learning algorithms, such as RF and XGBoost, outperform traditional machine learning algorithms in thermal comfort prediction [35,92,93,103–105].

Appendix 3-Tables A3–A5 show that different machine learning models perform significantly differently on the same dataset. When compared to the OPM model, the CatBoost model achieves the highest accuracy improvement of up to 27.31%. Traditional classification models such as KNN, MLR, and OPM perform significantly worse in

thermal comfort prediction than ensemble learning models, which is consistent with previous research. XGBoost, LightGBM, and CatBoost, all of which have been proposed in recent years [86–88], outperform the other ensemble learning algorithms. Currently, XGBoost is widely used in thermal comfort prediction, whereas LightGBM and CatBoost are used to a lesser extent [92,93]. According to research, the CatBoost model outperforms the XGBoost model when predicting outdoor thermal comfort. These conclusions, however, are based solely on the dataset used in this study, and it is hoped that future research will investigate the performance of CatBoost and LightGBM models on other datasets.

5.4. Model optimization

Bayesian optimization is used to optimize machine learning model hyperparameters, simplifying the hyperparameter tuning process [89, 90]. When machine learning models were combined with Bayesian optimization, the results showed significant performance improvements in each predictive category. This optimized ensemble of machine learning models improves the predictive performance of outdoor thermal comfort models by up to 9.61%. It should be noted that different datasets may necessitate different hyperparameter space configurations. As a result, when dealing with multiple datasets, a reiteration of the hyperparameter tuning process is required to achieve optimal

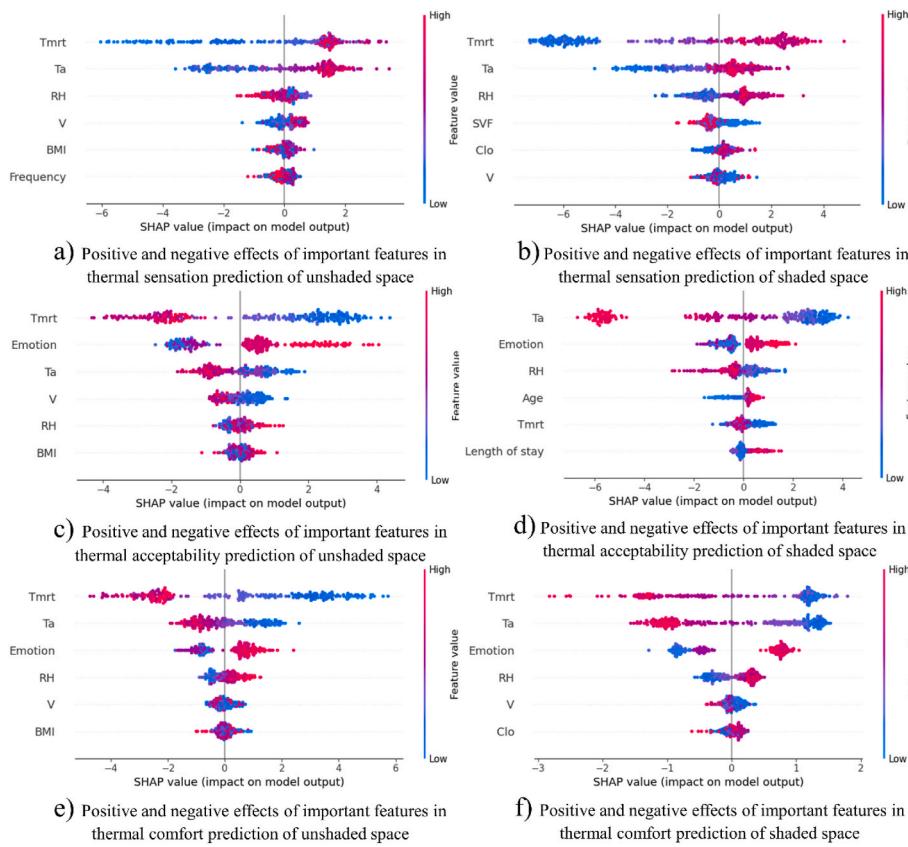


Fig. 11. The characteristics of TSV, TA and TCV prediction models without shade and shade space have positive and negative effects. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

performance. This iterative approach can improve the model's applicability and generalization capabilities even further.

5.5. Model interpretability

While machine learning models have made significant progress in terms of performance metrics such as accuracy when compared to traditional physical models, some complex machine learning models, such as ensemble learning, are frequently referred to as black-box models due to their complexity. These models' limited interpretability has limited their applications in certain domains. As a result, many researchers have dedicated themselves to improving the interpretability of complex black-box models, developing methods such as partial dependence plot (PDP), local interpretable Model-agnostic explanations (LIME), and SHAP to explain how these intricate machine learning models work [91,106–108]. The SHAP model has seen widespread use in the field of thermal comfort prediction [92,93]. In this study, we also introduced the SHAP model as a model interpretation tool to aid in the comprehension of complex ensemble learning models.

This study investigates the effects of environmental features on thermal sensation, thermal acceptability, and thermal comfort under various environmental conditions. Higher mean radiant temperature and air temperature are typically associated with higher thermal sensation, lower thermal acceptability, and lower thermal comfort during the summer. Higher relative humidity, increased park visit frequency, and lower wind speeds are associated with a perception of coolness in unshaded outdoor spaces. Higher relative humidity, higher thermal resistance of clothing, smaller sky view factor, and lower wind speed, on the other hand, result in people feeling hotter in shaded outdoor spaces. Individual physiological and psychological characteristics also have an important impact on thermal acceptability and comfort. Thermal acceptability and comfort are influenced by emotion, age, and

short-term heat history. Higher thermal acceptability and comfort are generally associated with increased emotion, while older people and longer park stays tend to exhibit higher thermal acceptability and comfort. These findings will help to improve the model further.

5.6. Limitations and future research

The following are the main limitations of this study and future research directions:

- (1) Data Scope: While data collection took place in 2021 and 2022, yielding 2651 valid data instances, it currently only covers data from one cold-weather city. To improve model robustness, future research could include outdoor data from more cities with cold climates.
- (2) Seasonality: The current models were developed using data from the summer and may not be applicable to other seasons. Subsequent studies could collect data from other seasons to make the models more complete.
- (3) Improved Interpretability: Evaluations of feature importance and positive/negative effects in machine learning models are not always accurate, and some judgments may deviate from logic, necessitating correction. Future research could look into methods like Abductive Learning, which combines machine learning and logical reasoning to improve model interpretability and robustness [109].

6. Conclusions

The following are the study's main findings:

- (1) After dividing the outdoor thermal comfort dataset into unshaded and shaded space datasets based on shading conditions, the optimal machine learning prediction models for TSV, TA, and TCV in unshaded spaces improved by 9.2%, 9.31%, and 6.16%, respectively, when the data volume was cut in half. The optimal machine learning prediction models for TSV, TA, and TCV achieved accuracy improvements of 1.58%, 0.4%, and 0.03%, respectively, in shaded spaces. Notably, ensemble learning models such as RF, XGBoost, LightGBM, and CatBoost performed relatively better in terms of prediction.
- (2) When compared to non-optimized models, Bayesian optimization-enhanced machine learning hybrid models improved average accuracy by 6.83%, 4.05%, and 2.55% for TSV, TA, and TCV predictions, respectively. The optimal machine learning prediction models for TSV, TA, and TCV showed accuracy improvements of 9.61%, 9.61%, and 6.54%, respectively, in unshaded spaces, and 7.75%, 3.33%, and 6.65%, respectively, in shaded spaces. CatBoost + BO was the optimal model of choice in the majority of cases. In certain scenarios, XGBoost + BO demonstrates superiority over CatBoost + BO. For instance, XGBoost + BO exhibited enhanced performance compared to CatBoost + BO in predicting TCV within shaded spaces. Additionally, in unshaded spaces, XGBoost + BO outperformed CatBoost + BO in TA prediction.
- (3) The four most influential features for TSV in the prediction of TSV, TA, and TCV in unshaded outdoor spaces during the summer were average radiant temperature, relative humidity, air temperature, and age. The key features for TA were average radiant temperature, emotion, air temperature, and wind speed, whereas the key features for TCV were average radiant temperature, air temperature, emotion, and relative humidity. Similarly, in shaded outdoor spaces, critical TSV features included average radiant temperature, air temperature, relative humidity, and BMI, while critical TA features included air temperature, emotion, average radiant temperature, and relative humidity, and critical TCV features included air temperature, average radiant temperature, relative humidity, and emotion. In summary, microclimate parameters dominated TSV, TA, and TCV predictions, with average radiant temperature in unshaded spaces and air temperature in shaded spaces being critical. Furthermore, age and BMI had an important impact on thermal sensation, while emotion had a high impact on thermal acceptance and comfort.
- (4) In addition to microclimate parameters, increased frequency of park visits in unshaded spaces was associated with lower thermal sensation. Higher relative humidity and clothing thermal resistance were generally associated with greater thermal sensation in shaded spaces, whereas an increase in the sky view factor was associated with decreased thermal sensation. Elevated emotion was generally associated with higher thermal acceptance and comfort, and older people tended to have higher thermal acceptance. Longer park stays were also found to be associated with higher thermal acceptance.

The findings of this study provide a scientific foundation for the design and planning of urban open spaces, thereby improving the quality of life for city dwellers, particularly in light of the thermal challenges posed by climate change. Furthermore, this study will advance the use of machine learning in urban planning and ecological and environmental sciences, resulting in more intelligent and sustainable urban development.

CRediT authorship contribution statement

Ruiqi Guo: Writing – original draft, Methodology, Investigation, Data curation. **Bin Yang:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **Yuyao Guo:** Investigation, Data curation. **He Li:** Investigation, Data curation. **Zhe Li:** Investigation, Data curation. **Bin Zhou:** Investigation, Data curation. **Bo Hong:** Supervision, Conceptualization. **Faming Wang:** Writing – review & editing, Supervision.

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

Data will be made available on request.

Acknowledgements

The project is financially supported by the National Natural Science Foundation of China (No. 52278119).

Appendix 1. Simplified garment checklist and Simplified list of activity types

Table A1
Simplified garment checklist

Body Parts	Clothing Types	clothing insulation (clo)
Upper Body	Hat	0.02
	Sunglasses	0.02
	Face Mask	0.05
	Cooling Sleeve	0.05
Lower Body	Short Sleeve	0.08
	Long Sleeve	0.25
	Sun Protective Clothing	0.07
	Underwear	0.07
Feet	Shorts	0.08
	Pants	0.24
	Mini Skirt	0.09
	Long Skirt	0.22
	Socks	0.02
	Sandals	0.02
	Low-Cut Shoes	0.05
	High-Cut Shoes	0.10

Table A2
Simplified list of activity types.

Activity Type	Metabolic Rate (Met)
Rest	1.0
Play Chess	1.5
Chat (Sitting)	1.5
Chat (Standing)	2.0
Walk	2.5
Sing	3.0
Practice Tai Chi	3.0
Dance in the Square	4.0
Play Badminton	4.0
Use Fitness Equipment	4.0
Play Table Tennis	4.0
Run	8.0

Appendix 2. Machine learning classification model evaluation metric calculation

Accuracy of each category refers to the percentage of correctly predicted samples within that specific category, calculated as shown in [Equation \(A1\)](#) [59].

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (\text{A1})$$

where TP is true positive; FP is false positive; FN is false negative; TN is true negative.

Precision of each category represents the percentage of correctly predicted samples within the subset of samples that receive positive predictions for that specific category, calculated as shown in [Equation \(A2\)](#) [59].

$$\text{Precision} = \frac{TP}{TP + FP} \quad (\text{A2})$$

Recall for each category indicates the percentage of correctly predicted samples for that category within the subset of samples that are actually positive for that category, calculated as shown in [Equation \(A3\)](#) [59].

$$\text{Recall} = \frac{TP}{TP + FN} \quad (\text{A3})$$

F1-score for each class provides a comprehensive assessment of accuracy and recall in the model. A higher F1-score, closer to 1, indicates better generalization ability of the model, calculated as shown in [Equation \(A4\)](#) [59].

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (\text{A4})$$

Appendix 3. Outdoor thermal comfort machine learning prediction model prediction accuracy statistics

Table A3

Performance comparative analysis of nine machine learning models (differentiated by Bayesian optimization) predicting human thermal sensation on a test set of different datasets (shaded spatial datasets, non-shaded spatial datasets, and global datasets) (note: Accuracy, accuracy, recall, and F1-Score are all weighted-average)

Data set	Algorithm	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
		Default	Model + BO						
Total	KNN	55.28	55.86	55.28	55.57	62.14	61.57	62.14	61.85
	MLR	57	58.75	57	57.86	64.25	62.75	64.25	63.49
	OPM	48.21	56.33	48.21	51.95	55.91	56.37	55.91	56.14
	DT	60.77	61.79	60.77	61.28	65.28	65.79	65.28	65.53
	RF	64.4	66.22	64.4	65.3	69.4	70.22	69.4	69.81
	SVM	63.82	64.44	63.82	64.13	69.11	71.95	69.11	70.50
	XGBoost	65.29	65.45	65.29	65.37	72.5	72.45	72.5	72.47
	LightGBM	65.08	65.26	65.08	65.17	72.08	72.26	72.08	72.17
	CatBoost	65.49	66.4	65.49	65.94	73.49	73.4	73.49	73.44
Unshaded	KNN	63.04	62.42	63.04	62.73	64.61	62.57	64.61	63.57
	MLR	62.18	63.19	62.18	62.68	71.53	68.24	71.53	69.85
	OPM	52.84	62.19	52.84	57.17	55.38	76.52	55.38	64.26
	DT	63.36	64.63	63.36	63.99	71.38	70.92	71.38	71.15
	RF	65.67	66.21	65.67	65.94	75.84	76.45	75.84	76.14
	SVM	65.03	64.33	65.03	64.68	71.38	72.73	71.38	72.05
	XGBoost	70.19	72.95	70.19	71.54	80.38	79.16	80.38	79.77
Shaded	LightGBM	71.96	72.79	71.96	72.37	79.08	80	79.08	79.54
	CatBoost	73.08	75.76	73.08	74.4	82.69	81.64	82.69	82.16
	KNN	48.2	47.64	48.2	47.92	53.46	51.88	53.46	52.66
	MLR	51.66	44.85	51.66	48.01	60.78	54.14	60.78	57.27

(continued on next page)

Table A3 (continued)

Data set	Algorithm	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
Default						Model + BO			
	OPM	42.8	51.69	42.8	46.83	48.75	55.8	48.75	52.04
	DT	56.83	56.84	56.83	56.83	63.74	63.46	63.74	63.60
	RF	62.89	63.17	62.89	63.03	68.63	70.66	68.63	69.63
	SVM	55.72	52.31	55.72	53.96	59.76	62.61	59.76	61.15
	XGBoost	65.31	68.8	65.31	66.95	73.06	73.8	73.06	73.50
	LightGBM	65.29	68.52	65.29	66.87	70.47	71.15	70.47	70.81
	CatBoost	63.52	67.86	63.52	65.62	71.59	72.05	71.59	71.82

Table A4

Performance comparative analysis of nine machine learning models (differentiated by Bayesian optimization) predicting human thermal acceptability on a test set of different datasets (shaded spatial datasets, non-shaded spatial datasets, and global datasets) (note: Accuracy, accuracy, recall, and F1-Score are all weighted-average)

Data set	Algorithm	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
Default						Model + BO			
Total	KNN	49.1	48.24	49.1	48.54	51.41	51.39	51.41	51.31
	MLR	52.12	49.91	52.12	50.18	52.54	49.16	52.54	49.54
	OPM	42.93	57.71	42.93	44.24	45.33	58.19	45.33	46.38
	DT	50.52	50.89	50.52	50.67	52.91	53.32	52.91	52.44
	RF	56.22	56.49	56.22	55.68	62.23	62.4	62.23	62.28
	SVM	50.47	48.34	50.47	49.21	56.7	55.65	56.7	55.81
	XGBoost	58.42	58.21	58.42	58.25	64.03	64.09	64.03	63.59
	LightGBM	58.62	58.42	58.62	58.43	62.23	62	62.23	61.86
	CatBoost	59.84	59.18	59.84	59.37	62.33	62.3	62.33	62.31
Unshaded	KNN	50.63	48.64	50.63	49.29	53.46	51.88	53.46	51.95
	MLR	53.9	50.33	53.9	51.69	55.76	53.24	55.76	53.88
	OPM	46.2	63.59	46.2	48.79	49.23	65.85	49.23	51.21
	DT	54.57	54.71	54.57	54.63	58.07	56.86	58.07	57.05
	RF	60.49	59.71	60.49	60.22	65.84	65.73	65.84	65.68
	SVM	55.38	54.06	55.38	54.31	59	56.88	59	57.35
	XGBoost	61.12	60.28	61.12	60.65	70.77	71.04	70.77	70.42
	LightGBM	61.93	60.77	61.93	61.25	71.54	71.68	71.54	71.13
	CatBoost	62.7	61.93	62.7	62.59	69.61	70.77	69.61	69.85
Shaded	KNN	47.92	46.32	47.92	46.64	49.44	49.63	49.44	48.74
	MLR	52.76	51.93	52.76	51.89	55.41	53.33	55.41	53.25
	OPM	48.94	54.24	48.94	48.67	50.55	54.33	50.55	50.47
	DT	49.4	49.63	49.4	49.49	54.61	56.46	54.61	54.76
	RF	59.05	58.68	59.05	58.56	62.36	62.33	64.21	61.51
	SVM	54.24	53.42	54.24	53.66	55.78	53.61	55.78	53.46
	XGBoost	54.61	55.1	54.61	54.71	60.15	59.4	60.15	59.11
	LightGBM	58.74	57.9	58.74	58.06	60.61	60.89	60.89	59.48
	CatBoost	59.4	59	59.4	59.15	62.73	62.63	62.73	61.7

Table A5

Performance comparative analysis of nine machine learning models (differentiated by Bayesian optimization) predicting human thermal comfort on a test set of different datasets (shaded spatial datasets, non-shaded spatial datasets, and global datasets) (note: Accuracy, accuracy, recall, and F1-Score are all weighted-average)

Data set	Algorithm	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
Default						Model + BO			
Total	KNN	43.88	42.33	43.88	42.79	47.03	45.98	47.03	46.17
	MLR	54.05	53.28	54.05	52.06	55.55	54.54	55.55	54.86
	OPM	33.7	43.43	33.7	34.29	38.11	45.23	38.11	38.23
	DT	53.4	53.53	53.4	53.45	55.55	56.15	55.55	55.92
	RF	58.87	58.62	58.87	58.51	61.03	61.08	61.03	61.21
	SVM	53.86	53.53	53.86	52.62	57.03	56.17	57.03	55.45
	XGBoost	59.88	58.9	59.88	59.24	61.56	61.17	61.56	61.32
	LightGBM	60.15	60.43	60.15	60.33	62.22	61.87	62.22	61.97
	CatBoost	60.22	60.89	60.22	60.51	63.07	62.59	63.07	62.72
Unshaded	KNN	52.07	48.84	52.07	50.01	52.3	47.03	52.3	48.55
	MLR	58.61	54.04	58.61	55.22	51.92	45.29	51.92	48.25
	OPM	40.04	63.63	40.04	41.89	40.76	61.17	40.76	42.56
	DT	56.69	57.26	56.69	56.95	54.23	50.93	54.23	51.72
	RF	59.54	59.85	59.54	59.44	62.41	62.54	62.41	64.66
	SVM	51.53	44.92	51.53	46.25	59.1	54.57	59.1	54.76
	XGBoost	61.83	62.5	61.83	61.95	66.15	64.79	66.15	65.21
	LightGBM	62.12	62.53	62.12	62.04	67.55	67.98	67.55	67.56
	CatBoost	62.69	62.9	62.69	62.29	69.23	69.91	69.23	69.45

(continued on next page)

Table A5 (continued)

Data set	Algorithm	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
		Default		Model + BO					
Shaded	KNN	43.29	43.29	43.29	42.83	44.65	48.68	44.65	44.13
	MLR	50.37	51.64	50.37	50.64	54.98	54.86	54.98	52.63
	OPM	53.87	62.13	53.87	54.88	53.93	61.9	53.93	55.58
	DT	51.99	52.52	51.99	52.22	53.51	53.71	53.51	53.08
	RF	56.85	57.42	56.85	56.75	60.87	60.28	60.87	60.15
	SVM	55.71	53.7	55.71	53.07	56.46	55.48	56.46	54.02
	XGBoost	56.45	57.34	56.45	56.23	57.93	59	57.93	57.12
	LightGBM	56.12	56.74	56.12	55.98	63.1	64.91	63.1	62.27
	CatBoost	56.45	54.82	56.45	54.85	63.1	66.67	63.1	62.79

References

- [1] U. Habitat, World Cities Report 2022: Envisaging the Future of Cities, United Nations Human Settlements Programme, Nairobi, Kenya, 2022, pp. 41–44.
- [2] C.W. Thompson, Urban open space in the 21st century, *Landsc. Urban Plann.* 60 (2) (2002) 59–72.
- [3] H. Woolley, Urban Open Spaces, Taylor & Francis, 2003.
- [4] T.-P. Lin, K.-T. Tsai, R.-L. Hwang, A. Matzarakis, Quantification of the effect of thermal indices and sky view factor on park attendance, *Landsc. Urban Plann.* 107 (2) (2012) 137–146.
- [5] D. Lai, B. Chen, K. Liu, Quantification of the Influence of Thermal Comfort and Life Patterns on Outdoor Space Activities, *Building Simulation*, Springer, 2020, pp. 113–125.
- [6] L.M. Brander, M.J. Koetse, The value of urban open space: meta-analyses of contingent valuation and hedonic pricing results, *J. Environ. Manag.* 92 (10) (2011) 2763–2773.
- [7] B. Yang, T. Olofsson, G. Nair, A. Kabanshi, Outdoor thermal comfort under subarctic climate of north Sweden—A pilot study in Umeå, *Sustain. Cities Soc.* 28 (2017) 387–397.
- [8] J. Niu, J. Liu, T.-c. Lee, Z.J. Lin, C. Mak, K.-T. Tse, B.-s. Tang, K.C. Kwok, A new method to assess spatial variations of outdoor thermal comfort: onsite monitoring results and implications for precinct planning, *Build. Environ.* 91 (2015) 263–270.
- [9] S. Luthi, C. Fairless, E.M. Fischer, N. Scovronick, A. Ben, M. Coelho, Y.L. Guo, Y. Guo, Y. Honda, V. Huber, J. Kysely, E. Lavigne, D. Roye, N. Ryti, S. Silva, A. Urban, A. Gasparrini, D.N. Bresch, A.M. Vicedo-Cabrera, Rapid increase in the risk of heat-related mortality, *Nat. Commun.* 14 (1) (2023) 4894.
- [10] A.M. Vicedo-Cabrera, N. Scovronick, F. Sera, D. Roye, R. Schneider, A. Tobias, C. Astrom, Y. Guo, Y. Honda, D.M. Hondula, R. Abrutzyk, S. Tong, M. de Sousa Zanotti Staglario Coelho, P.H.N. Saldiva, E. Lavigne, P.M. Correa, N.V. Ortega, H. Kan, S. Osorio, J. Kysely, A. Urban, H. Orru, E. Indermitte, J.J.K. Jaakkola, N. Ryti, M. Pascal, A. Schneider, K. Katsouyanni, E. Samoli, F. Mayvaneh, A. Entezari, P. Goodman, A. Zeka, P. Michelozzi, F. de' Donato, M. Hashizume, B. Alahmad, M.H. Diaz, C. De La Cruz Valencia, A. Overcenco, D. Houthuijs, C. Ameling, S. Rao, F.D. Ruscio, G. Carrasco-Escobar, X. Seposo, S. Silva, J. Madureira, I.H. Holobaca, S. Fratianni, F. Acquaotta, H. Kim, W. Lee, C. Iniguez, B. Forsberg, M.S. Ragettli, Y.L.L. Guo, B.Y. Chen, S. Li, B. Armstrong, A. Aleman, A. Zanobetti, J. Schwartz, T.N. Dang, D.V. Dung, N. Gillett, A. Haines, M. Mengel, V. Huber, A. Gasparrini, The burden of heat-related mortality attributable to recent human-induced climate change, *Nat. Clim. Change* 11 (6) (2021) 492–500.
- [11] L. Zhao, K. Oleson, E. Bou-Zeid, E.S. Krayenhoff, A. Bray, Q. Zhu, Z. Zheng, C. Chen, M. Oppenheimer, Global multi-model projections of local urban climates, *Nat. Clim. Change* 11 (2) (2021) 152–157.
- [12] A. Hsu, G. Sheriff, T. Chakraborty, D. Many, Disproportionate exposure to urban heat island intensity across major US cities, *Nat. Commun.* 12 (1) (2021) 2721.
- [13] K.M. Gabriel, W.R. Endlicher, Urban and rural mortality rates during heat waves in Berlin and Brandenburg, Germany, *Environ. Pollut.* 159 (8–9) (2011) 2044–2050.
- [14] P. Michelozzi, G. Accetta, M. De Sario, D. D'Ippoliti, C. Marino, M. Baccini, A. Biggeri, H.R. Anderson, K. Katsouyanni, F. Ballester, High temperature and hospitalizations for cardiovascular and respiratory causes in 12 European cities, *Am. J. Respir. Crit. Care Med.* 179 (5) (2009) 383–389.
- [15] L.W. Chew, X. Liu, X.-X. Li, L.K. Norford, Interaction between heat wave and urban heat island: a case study in a tropical coastal city, Singapore, *Atmos. Res.* 247 (2021) 105134.
- [16] L. Zhao, M. Oppenheimer, Q. Zhu, J.W. Baldwin, K.L. Ebi, E. Bou-Zeid, K. Guan, X. Liu, Interactions between urban heat islands and heat waves, *Environ. Res. Lett.* 13 (3) (2018) 034003.
- [17] P. Kumar, A. Sharma, Study on importance, procedure, and scope of outdoor thermal comfort—A review, *Sustain. Cities Soc.* 61 (2020) 102297.
- [18] L. Chen, E. Ng, Outdoor thermal comfort and outdoor activities: a review of research in the past decade, *Cities* 29 (2) (2012) 118–125.
- [19] D. Lai, Z. Lian, W. Liu, C. Guo, W. Liu, K. Liu, Q. Chen, A comprehensive review of thermal comfort studies in urban open spaces, *Sci. Total Environ.* 742 (2020) 140092.
- [20] S. Shooshtarian, C.K.C. Lam, I. Kenawy, Outdoor thermal comfort assessment: a review on thermal comfort research in Australia, *Build. Environ.* 177 (2020) 106917.
- [21] D. Lai, D. Guo, Y. Hou, C. Lin, Q. Chen, Studies of outdoor thermal comfort in northern China, *Build. Environ.* 77 (2014) 110–118.
- [22] E. Johansson, S. Thorsson, R. Emmanuel, E. Krüger, Instruments and methods in outdoor thermal comfort studies—The need for standardization, *Urban Clim.* 10 (2014) 346–366.
- [23] P.O. Fanger, Thermal Comfort. Analysis and Applications in Environmental Engineering, *Thermal Comfort. Analysis and Applications in Environmental Engineering*, 1970.
- [24] T. Cheung, S. Schiavon, T. Parkinson, P. Li, G. Brager, Analysis of the accuracy on PMV–PPD model using the ASHRAE global thermal comfort database II, *Build. Environ.* 153 (2019) 205–217.
- [25] O. Potchter, P. Cohen, T.-P. Lin, A. Matzarakis, Outdoor human thermal perception in various climates: a comprehensive review of approaches, methods and quantification, *Sci. Total Environ.* 631 (2018) 390–406.
- [26] J. Van Hoof, Forty years of Fanger's model of thermal comfort: comfort for all? *Indoor Air* 18 (3) (2008) 182–201.
- [27] A. Matzarakis, H. Mayer, M.G. Iziomron, Applications of a universal thermal index: physiological equivalent temperature, *Int. J. Biometeorol.* 43 (1999) 76–84.
- [28] K. Blazejczyk, Y. Epstein, G. Jendritzky, H. Staiger, B. Tinz, Comparison of UTCI to selected thermal indices, *Int. J. Biometeorol.* 56 (2012) 515–535.
- [29] G. Jendritzky, R. de Dear, G. Havenith, UTCI—why another thermal index? *Int. J. Biometeorol.* 56 (2012) 421–428.
- [30] Z. Fang, X. Feng, J. Liu, Z. Lin, C.M. Mak, J. Niu, K.-T. Tse, X. Xu, Investigation into the differences among several outdoor thermal comfort indices against field survey in subtropics, *Sustain. Cities Soc.* 44 (2019) 676–690.
- [31] B. Huang, B. Hong, Y. Tian, T. Yuan, M. Su, Outdoor thermal benchmarks and thermal safety for children: a study in China's cold region, *Sci. Total Environ.* 787 (2021).
- [32] K. Liu, T. Nie, W. Liu, Y. Liu, D. Lai, A machine learning approach to predict outdoor thermal comfort using local skin temperatures, *Sustain. Cities Soc.* 59 (2020) 102216.
- [33] M. Nikolopoulou, Designing Open Spaces in the Urban Environment: a Bioclimatic Approach, Centre for Renewable Energy Sources, EESD, FP52004.
- [34] R. Yao, B. Li, J. Liu, A theoretical adaptive model of thermal comfort—Adaptive Predicted Mean Vote (aPMV), *Build. Environ.* 44 (10) (2009) 2089–2096.
- [35] J. Jeong, J. Jeong, M. Lee, J. Lee, S. Chang, Data-driven approach to develop prediction model for outdoor thermal comfort using optimized tree-type algorithms, *Build. Environ.* 226 (2022).
- [36] S. Lee, I. Bilionis, P. Karava, A. Tzempelikos, A Bayesian approach for probabilistic classification and inference of occupant thermal preferences in office buildings, *Build. Environ.* 118 (2017) 323–343.
- [37] Y. Wu, H. Liu, B. Li, R. Kosonen, S. Wei, J. Jokisalo, Y. Cheng, Individual thermal comfort prediction using classification tree model based on physiological parameters and thermal history in winter, *Build. Simulat.* 14 (6) (2021) 1651–1665.
- [38] J. Zhang, F. Zhang, Z. Gou, J. Liu, Assessment of macroclimate and microclimate effects on outdoor thermal comfort via artificial neural network models, *Urban Clim.* 42 (2022).
- [39] D. Vučković, S. Jovic, R. Bozovic, V. Djamić, D. Kićović, Potential of neuro-fuzzy methodology for forecasting of outdoor thermal comfort index at urban open spaces, *Urban Clim.* 28 (2019).
- [40] E. Díz-Mellado, S. Rubino, S. Fernández-García, M. Gómez-Mármol, C. Rivera-Gómez, C. Galán-Marín, Applied machine learning algorithms for courtyards thermal patterns accurate prediction, *Mathematics* 9 (10) (2021).
- [41] 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).
- [42] L. Jiang, R. Yao, Modelling personal thermal sensations using C-Support Vector Classification (C-SVC) algorithm, *Build. Environ.* 99 (2016) 98–106.
- [43] J. Niu, J. Xiong, H. Qin, J. Hu, J. Deng, G. Han, J. Yan, Influence of thermal comfort of green spaces on physical activity: empirical study in an urban park in Chongqing, China, *Build. Environ.* 219 (2022).

- [44] L. An, B. Hong, X. Cui, Y. Geng, X. Ma, Outdoor thermal comfort during winter in China's cold regions: a comparative study, *Sci. Total Environ.* 768 (2021).
- [45] A. Bassoud, H. Khelafi, A.M. Mokhtari, A. Bada, Evaluation of summer thermal comfort in arid desert areas. Case study: old adobe building in Adrar (South of Algeria), *Build. Environ.* 205 (2021).
- [46] K. Labdaoui, S. Mazouz, S. Reiter, J. Teller, Thermal perception in outdoor urban spaces under the Mediterranean climate of Annaba, Algeria, *Urban Clim.* 39 (2021).
- [47] X. Ma, L. Zhang, J. Zhao, M. Wang, Z. Cheng, The outdoor pedestrian thermal comfort and behavior in a traditional residential settlement – a case study of the cave dwellings in cold winter of China, *Sol. Energy* 220 (2021) 130–143.
- [48] M. S, E. Rajasekar, Evaluating outdoor thermal comfort in urban open spaces in a humid subtropical climate: chandigarh, India, *Build. Environ.* 209 (2022).
- [49] T. Sharmin, K. Steemers, M. Humphreys, Outdoor thermal comfort and summer PET range: a field study in tropical city Dhaka, *Energy Build.* 198 (2019) 149–159.
- [50] Q. Yin, Y. Cao, C. Sun, Research on outdoor thermal comfort of high-density urban center in severe cold area, *Build. Environ.* 200 (2021).
- [51] M. Zhen, Q. Dong, P. Chen, W. Ding, D. Zhou, W. Feng, Urban outdoor thermal comfort in western China, *J. Asian Architect. Build. Eng.* 20 (2) (2020) 222–236.
- [52] N. Forcada, M. Gangolli, M. Casals, B. Tejedor, M. Macarulla, K. Gaspar, Field study on adaptive thermal comfort models for nursing homes in the Mediterranean climate, *Energy Build.* 252 (2021).
- [53] X. Ma, Y. Tian, M. Du, B. Hong, B. Lin, How to design comfortable open spaces for the elderly? Implications of their thermal perceptions in an urban park, *Sci. Total Environ.* 768 (2021).
- [54] E. Tarpani, I. Piglioutile, A.L. Pisello, On kids' environmental wellbeing and their access to nature in urban heat islands: hyperlocal microclimate analysis via surveys, modelling, and wearable sensing in urban playgrounds, *Urban Clim.* 49 (2023).
- [55] N. Wallenberg, D. Rayner, F. Lindberg, S. Thorsson, Present and future heat stress of preschoolers in five Swedish cities, *Clim. Risk Manag.* 40 (2023).
- [56] F. Yao, H. Fang, J. Han, Y. Zhang, Study on the outdoor thermal comfort evaluation of the elderly in the Tibetan plateau, *Sustain. Cities Soc.* 77 (2022).
- [57] P. Kumar, A. Sharma, Assessing the outdoor thermal comfort conditions of exercising people in the semi-arid region of India, *Sustain. Cities Soc.* 76 (2022).
- [58] Z. Fang, X. Xu, X. Zhou, S. Deng, H. Wu, J. Liu, Z. Lin, Investigation into the thermal comfort of university students conducting outdoor training, *Build. Environ.* 149 (2019) 26–38.
- [59] 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.
- [60] X. Shan, E.-H. Yang, Supervised machine learning of thermal comfort under different indoor temperatures using EEG measurements, *Energy Build.* 225 (2020).
- [61] K. Pantavou, K.K. Delibasis, G.K. Nikolopoulos, Machine learning and features for the prediction of thermal sensation and comfort using data from field surveys in Cyprus, *Int. J. Biometeorol.* 66 (10) (2022) 1973–1984.
- [62] M. Luo, J. Xie, Y. Yan, Z. Ke, P. Yu, Z. Wang, J. Zhang, Comparing machine learning algorithms in predicting thermal sensation using ASHRAE Comfort Database II, *Energy Build.* 210 (2020).
- [63] K. Liu, T. Nie, W. Liu, Y. Liu, D. Lai, A machine learning approach to predict outdoor thermal comfort using local skin temperatures, *Sustain. Cities Soc.* 59 (2020).
- [64] D. Lai, C. Chen, Comparison of the linear regression, multinomial logit, and ordered probability models for predicting the distribution of thermal sensation, *Energy Build.* 188–189 (2019) 269–277.
- [65] K. Katic, R. Li, W. Zeiler, Machine learning algorithms applied to a prediction of personal overall thermal comfort using skin temperatures and occupants' heating behavior, *Appl. Ergon.* 85 (2020) 103078.
- [66] A.C. Cosma, R. Simha, Machine learning method for real-time non-invasive prediction of individual thermal preference in transient conditions, *Build. Environ.* 148 (2019) 372–383.
- [67] A. Aryal, B. Becerik-Gerber, Thermal comfort modeling when personalized comfort systems are in use: comparison of sensing and learning methods, *Build. Environ.* 185 (2020).
- [68] M. Abdellatif, J. Chamoin, J.-M. Nianga, D. Defer, A thermal control methodology based on a machine learning forecasting model for indoor heating, *Energy Build.* 255 (2022).
- [69] M. Xu, B. Hong, R. Jiang, L. An, T. Zhang, Outdoor thermal comfort of shaded spaces in an urban park in the cold region of China, *Build. Environ.* 155 (2019) 408–420.
- [70] S. Watanabe, K. Nagano, J. Ishii, T. Horikoshi, Evaluation of outdoor thermal comfort in sunlight, building shade, and pergola shade during summer in a humid subtropical region, *Build. Environ.* 82 (2014) 556–565.
- [71] N. Nasrallah, Y. Namazi, M. Taleghani, The effect of urban shading and canyon geometry on outdoor thermal comfort in hot climates: a case study of Ahvaz, Iran, *Sustain. Cities Soc.* 65 (2021) 102638.
- [72] T.-P. Lin, A. Matzarakis, R.-L. Hwang, Shading effect on long-term outdoor thermal comfort, *Build. Environ.* 45 (1) (2010) 213–221.
- [73] R.-L. Hwang, T.-P. Lin, A. Matzarakis, Seasonal effects of urban street shading on long-term outdoor thermal comfort, *Build. Environ.* 46 (4) (2011) 863–870.
- [74] M.C. Peel, B.L. Finlayson, T.A. McMahon, Updated world map of the Köppen-Geiger climate classification, *Hydrol. Earth Syst. Sci.* 11 (5) (2007) 1633–1644.
- [75] C.M. Administration, Blue Book on Climate Change in China 2020, Science Press Beijing, China, 2020.
- [76] Y. Tian, B. Hong, Z. Zhang, S. Wu, T. Yuan, Factors influencing resident and tourist outdoor thermal comfort: a comparative study in China's cold region, *Sci. Total Environ.* 808 (2022) 152079.
- [77] ISO, Ergonomics of the Thermal Environment—Instruments for Measuring Physical Quantities, 1998 (ISO 7726: 1998).
- [78] I. 10551, Ergonomics of the Thermal Environment—Assessment of the Influence of the Thermal Environment Using Subjective Judgement Scales, ISO, Geneva, Switzerland, 1995.
- [79] F.Q. Nuttal, Body mass index: obesity, BMI, and health: a critical review, *Nutr. Today* 50 (3) (2015) 117.
- [80] ISO, International Standard, Ergonomics of the Thermal Environment—Analytical Determination and Interpretation of Thermal Comfort Using Calculation of the PMV and PPD Indices and Local Thermal Comfort Criteria, 2005.
- [81] A. Ashrae, S. Ashrae, Standard 55-2020, Thermal Environmental Conditions for Human Occupancy, American Society of Heating, Refrigerating, and Air-Conditioning Engineers, Inc. Atlanta, 2020.
- [82] G. Guo, H. Wang, D. Bell, Y. Bi, K. Greer, KNN Model-Based Approach in Classification, on the Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3–7, 2003. Proceedings, Springer, 2003, pp. 986–996.
- [83] C. Kingsford, S.L. Salzberg, What are decision trees? *Nat. Biotechnol.* 26 (9) (2008) 1011–1013.
- [84] L. Breiman, Random forests, *Mach. Learn.* 45 (2001) 5–32.
- [85] W.S. Noble, What is a support vector machine? *Nat. Biotechnol.* 24 (12) (2006) 1565–1567.
- [86] T. Chen, C. Guestrin, Xgboost: a scalable tree boosting system, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794.
- [87] 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, *Adv. Neural Inf. Process. Syst.* 30 (2017).
- [88] L. Prokhorenkova, G. Gusev, A. Vorobev, A.V. Dorogush, A. Gulin, CatBoost: unbiased boosting with categorical features, *Adv. Neural Inf. Process. Syst.* 31 (2018).
- [89] 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 (1) (2015) 148–175.
- [90] J. Snoek, H. Larochelle, R.P. Adams, Practical bayesian optimization of machine learning algorithms, *Adv. Neural Inf. Process. Syst.* 25 (2012).
- [91] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, *Adv. Neural Inf. Process. Syst.* 30 (2017).
- [92] G. Zheng, Y. Zhang, X. Yue, K. Li, Interpretable prediction of thermal sensation for elderly people based on data sampling, machine learning and SHapley Additive exPlanations (SHAP), *Build. Environ.* 242 (2023) 110602.
- [93] H. Lan, H. Hou, Z. Gou, A machine learning led investigation to understand individual difference and the human-environment interactive effect on classroom thermal comfort, *Build. Environ.* 236 (2023).
- [94] C. Molnar, Interpretable Machine Learning, Lulu. com, 2020.
- [95] I. Ekanayake, D. Meddage, U. Rathnayake, A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive explanations (SHAP), *Case Stud. Constr. Mater.* 16 (2022) e01059.
- [96] Y. Kim, Y. Kim, Explainable heat-related mortality with random forest and SHapley Additive exPlanations (SHAP) models, *Sustain. Cities Soc.* 79 (2022) 103677.
- [97] M. Mohammady, H.R. Pourghasemi, M. Amiri, Land subsidence susceptibility assessment using random forest machine learning algorithm, *Environ. Earth Sci.* 78 (2019) 1–12.
- [98] S.M. Lundberg, G.G. Erion, S.-I. Lee, Consistent Individualized Feature Attribution for Tree Ensembles, 2018 arXiv preprint arXiv:1802.03888.
- [99] A. Youssef, A. Youssef Ali Amer, N. Caballero, J.-M. Aerts, Towards online personalized-monitoring of human thermal sensation using machine learning approach, *Appl. Sci.* 9 (16) (2019) 3303.
- [100] T. Yan, H. Jin, Y. Jin, The mediating role of emotion in the effects of landscape elements on thermal comfort: a laboratory study, *Build. Environ.* 233 (2023) 110130.
- [101] G. Song, Z. Ai, G. Zhang, Y. Peng, W. Wang, Y. Yan, Using machine learning algorithms to multidimensional analysis of subjective thermal comfort in a library, *Build. Environ.* 212 (2022) 108790.
- [102] H. Wang, L. Liu, Experimental investigation about effect of emotion state on people's thermal comfort, *Energy Build.* 211 (2020) 109789.
- [103] Z. Wu, N. Li, J. Peng, H. Cui, P. Liu, H. Li, X. Li, Using an ensemble machine learning methodology-Bagging to predict occupants' thermal comfort in buildings, *Energy Build.* 173 (2018) 117–127.
- [104] Z. Wang, H. Yu, M. Luo, Z. Wang, H. Zhang, Y. Jiao, Predicting older people's thermal sensation in building environment through a machine learning approach: modelling, interpretation, and application, *Build. Environ.* 161 (2019).
- [105] 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.
- [106] H. Nori, S. Jenkins, P. Koch, R. Caruana, InterpretML: A Unified Framework for Machine Learning Interpretability, 2019 arXiv preprint arXiv:1909.09223.

- [107] B.M. Greenwell, pdp: an R package for constructing partial dependence plots, *R J.* 9 (1) (2017) 421.
- [108] M.T. Ribeiro, S. Singh, C. Guestrin, Model-agnostic Interpretability of Machine Learning, 2016 arXiv preprint arXiv:1606.05386.
- [109] Z.-H. Zhou, Abductive learning: towards bridging machine learning and logical reasoning, *Sci. China Inf. Sci.* 62 (2019) 1–3.