Energy and Built Environment 6 (2025) 66–79

Contents lists available at ScienceDirect

Energy and Built Environment

journal homepage: http://www.keaipublishing.com/en/journals/energy-and-built-environment/

An IMU dataset for human thermal comfort activities identification: Experimental designs and applications

Weilin He[a,c], Cheng Fan[a,b,c,*], Zebin Wu[a,c], Qiaoqiao Yong[a,c]

[a] Sino-Australia Joint Research Center in BIM and Smart Construction, Shenzhen University, Shenzhen, China
[b] Key Laboratory for Resilient Infrastructures of Coastal Cities, Ministry of Education, Shenzhen University, Shenzhen, China
[c] College of Civil and Transportation Engineering, Shenzhen University, Shenzhen, China

ARTICLE INFO

Keywords:
Thermal comfort
IMU dataset
Human activity recognition
Occupant behavior

ABSTRACT

Thermal comfort of occupants is key feedback information for improving indoor environment and managing building energy use. Through analyzing inertial measurement units (IMU) data from wearable devices with machine learning, thermal comfort of occupants can be detected in a non- intrusive method. This paper proposed a dataset consisted of IMU data collected from 30 participants (14 males and 16 females, aged 23.23 ± 1.70 years, height 168.67 ± 8.02 cm, and weight 59.55 ± 10.96 kg) who wore two IMUs on their hands while performing 30 thermal comfort activities (10 cold-related, 10 hot-related, and 10 neutral activities) according to their personal habits.

The database is divided into two parts: (1) Single activities data, which includes 4500 samples acquired from experiments where each participant was asked to perform 30 thermal comfort activities individually. (2) Continuous multi-activity data, which comprise 360 samples collected while participants performed a series of randomly assigned activities in a more natural and continuous manner. The combination of these two parts provides a comprehensive dataset for both the training and testing phases of machine learning models. By offering detailed labels, this database aims to serve as a foundation for research exploring machine learning applications in detecting occupant thermal comfort, ultimately contributing to improved indoor environments and more efficient building energy management.

1. Introduction

Traditional (heating, ventilation and air conditioning) HVAC systems generally operate based on fixed temperature setpoints. This mode of operation results in substantial energy wastage associated with HVAC systems while also frequently causing thermal discomfort for occupants due to insufficient consideration of individual differences and variations in thermal comfort [1,2]. Researchers in recent years have found that real-time thermal comfort information of occupants is a crucial reference for controlling HVAC systems to conserve energy and enhance indoor environmental comfort [3,4]. Presently, there are three methods of thermal comfort detection: the predicted mean vote (PMV) method [5], physiological detection method [6], and activities recognition-based method [7].

The PMV is a predictive method for occupant thermal comfort based on a statistical model. Fanger established the PMV model in 1970 using four environmental parameters (air temperature, air velocity, relative humidity, and mean radiant temperature) and two personal factors (clothing thermal resistance and metabolic rate) [8]. Many scholars have made different improvements based on Fanger's PMV model. Kim et al. developed a better-performing new PMV based on adaptive thermal comfort theory [9]. Yang et al. discovered that people living in hot regions for extended periods develop psychological adaptation to heat sensations, and the revised PMV model more accurately predicts heat sensations in humid and hot areas [10]. Gilani et al. developed an improved PMV model based on the strong correlation between average blood pressure and metabolic rate [11]. Due to its ease of implementation, various PMV models are most widely applied in buildings with HVAC systems. However, the PMV model is designed based on the average thermal sensation votes of a large group of people and cannot predict individual thermal comfort.

The physiological detection method establishes models for monitoring human thermal comfort by measuring physiological data related to thermal comfort sensation (such as skin temperature, heart rate, and blood oxygen capacity). Dai et al. developed a thermal comfort detection model using skin temperatures from different body parts as input variables and achieved an accuracy rate of 90 % using Support Vector Machine [12]. Yao et al. investigated the relationship between physiologi-

* Corresponding author at: Key Laboratory for Resilient Infrastructures of Coastal Cities, Ministry of Education, Shenzhen University, Shenzhen, China.
E-mail address: fancheng@szu.edu.cn (C. Fan).

https://doi.org/10.1016/j.enbenv.2023.09.001
Received 15 April 2023; Received in revised form 8 August 2023; Accepted 7 September 2023
Available online 9 September 2023

Table 1
Comparison of thermal comfort detection studies based on activities recognition.

Paper	Device(s)	Camera angle	Number of subjects	Number of activities	Data publicly available	Data type
[7]	–	Frontal Shot	–	12	No	skeleton data
[16]	1 camera	Frontal Shot	–	11	No	skeleton data
[17]	1 camera	Frontal Shot	16	5	No	skeleton data
[18]	3 Kinect cameras, 3 surveillance cameras	6 views at 6 heights (1.6 m–2.9 m)	52	16	Yes	RGB, skeleton data
[19]	1 camera	Frontal Shot	–	13	No	skeleton data
[20]	3 Kinect cameras	3 views at eye level	20	7	No	skeleton data
[21]	16 cameras	16 views at two heights (1.5 m and 2 m)	62	14	Yes	RGB

cal indicators such as heart rate variability and electroencephalograms with human thermal comfort, revealing a strong correlation between these indicators and thermal comfort state [13]. Park et al. measured physiological indicators (e.g., blood oxygen capacity, skin conductance, skin temperature, and heart rate) using smart bands and combined these data with transfer learning to achieve individual-level thermal comfort detection [14]. However, Chad et al. found that strenuous tasks could induce physiological disturbances [15]. Physiological detection methods may be affected by these disturbances, resulting in compromised detection accuracy, and are thus only suitable for constrained environments and limited populations.

To address these challenges, a thermal comfort detection method based on activities recognition has been proposed. Spontaneous activities performed by humans when feeling hot or cold are called thermal comfort activities.

Common thermal comfort activities include fanning with hands, taking off coat, and warming hands with breath. This method detects real-time human thermal comfort states by recognizing various thermal comfort activities. Table 1 summarizes current research on thermal comfort detection based on activities recognition, with a "–" denoting that the paper did not mention the information.

Yang et al. defined 12 thermal uncomfortable activities. They used computer vision technology to recognize those activities by importing videos into the open-source platform Openpose [16]. Duan et al. defined 16 thermal comfort activities and invited 52 participants for testing, releasing a public video dataset. Additionally, they developed a spatial-temporal graph convolutional network, achieving an accuracy rate of 78 % [18]. Liu et al. collected video data from 16 different angles and employed cutting-edge neural networks such as Two-Stream Inflated 3D ConvNet and SlowFast to achieve recognition accuracy exceeding 95 % for 14 thermal discomfort activities [21].

It is evident that existing research on thermal comfort activities recognition primarily focuses on analyzing image data using computer vision techniques. Recognizing human thermal comfort with image data has the advantages of not interfering with individuals' normal activities and being non-contact. However, this method has three drawbacks: first, the collection of video data may involve privacy issues; second, processing video data requires significant computational resources; and finally, recognition accuracy is easily affected by occlusion, lighting, and shooting angle [22,23].

With the continuous reduction in size and cost of inertial measurement units (IMUs), most mobile devices (such as smartwatches, smart bracelets, and smartphones) integrate inertial sensors like accelerometers and gyroscopes. Research on activities recognition based on IMU data has achieved substantial results in the fields of human basic activities recognition, motion posture recognition, and gesture recognition. Wannenburg et al. utilized smartphones placed in trouser pockets to measure acceleration during basic human activities such as sitting, standing, lying, walking, and jogging, and compared the performance of 10 classifiers. Experimental results showed that offline recognition accuracy of the K-nearest neighbors and kStar algorithms exceeded 99 % [24]. Vleugels et al. collected acceleration and angular velocity data from hockey sticks using IMUs, and employed convolutional neural network (CNN) to classify six different hockey activities, achieving an accuracy of 76 % [25]. Qi et al. applied support vector machine (SVM) and Hidden Markov models to construct a two-layer activities recognition framework using chest and arm acceleration data for various free-weight exercises (such as push-ups, squats, etc.) and aerobic exercises (such as jogging, rowing, etc.) [26]. Existing research indicates that activity recognition based on IMU data is accurate, stable, low-cost, and non-intrusive [27].

However, the use of IMU data for human activity recognition has not been extended to the identification of thermal comfort activities, leaving a void in the relevant datasets. The availability of relevant datasets is crucial. High-quality accessible datasets are the foundation of data-driven methods. As shown in Table 1, except for [18] and [21], existing studies have not made their datasets public. This not only leads to repetitive research work but also hinders scholars from establishing a unified benchmark to evaluate and compare the advantages of different methods. In summary, current thermal comfort activities datasets have the following three limitations: first, most datasets are not public; second, they contain few thermal comfort activities and do not consider neutral activities unrelated to human thermal comfort, creating a gap with real-world application scenarios; and finally, the collected data are image-based, making this thermal comfort detection method unsuitable for privacy-sensitive situations.

Therefore, this paper establishes an accessible IMU dataset for non-intrusive thermal comfort activities recognition. Specifically, the work includes the following:

(1) Through the implementation of a two-phase questionnaire survey, a total of 626 questionnaires were distributed, delineating 20 thermal comfort activities (i.e., 10 cold-related activities and 10 hot-related activities). From the results of the survey, a quantitative relationship was established between these 20 activities and various thermal sensations (i.e., cold, cool, slightly cool, slightly warm, warm, and hot).
(2) Recruiting 30 volunteers and collecting their IMU data during various thermal comfort activities using wearable devices, including 4500 single activities samples for training and 360 continuous multi-activity samples for testing.
(3) Developing a thermal comfort activities recognition model using 1D-CNN and long short-term memory (LSTM) networks, demonstrating the feasibility of thermal comfort detection based on IMU data.
(4) A strategy was designed to quantify indoor human thermal comfort based on detected thermal comfort activities, which can be utilized to guide the intelligent operation of HVAC systems.

2. Definition of thermal comfort activities

2.1. Thermal comfort activity questionnaire survey

To explore the activities pertaining to indoor occupants' thermal comfort, this research employed electronic questionnaires for two rounds of investigation (preliminary and in-depth survey). In the preliminary survey, we distributed 340 simple questionnaires wherein re-

Table 2
Details of the preliminary survey questionnaire.

Question 1: What activities would you take if you feel cold indoors?		
A. Buttoning up	B. Warming hands with breath	C. Rubbing hands
D. Rubbing ears	E. Rolling down sleeves	F. Tucking hands in sleeves
G. Rubbing arms	H. Putting hands in pockets	I. Wearing coat
J. Rubbing knees	K. All of the above	L. Other activities (please specify)
Question 2: What activities would you take if you feel hot indoors?		
A. Unbuttoning	B. Shaking T-shirt	C. Fanning with hands
D. Wiping sweat from face	E. Lifting bangs	F. Lifting hair on the back of the neck
G. Pushing up sleeves	H. Rolling up sleeves	I. Taking coat off
J. Wiping sweat from back	K. All of the above	L. Other activities (please specify)

Table 3
Details of the in-depth survey questionnaire (only three of the total 20 questions are listed).

Question 1: Indoors, you will **button up** when you feel___?			
A. Cold	B. Cool	C. Slightly cool	D. Neutral
E. Slightly warm	F. Warm	G. Hot	H. Never do it
Question 2: Indoors, you will **warm hands with breath** when you feel___?			
A. Cold	B. Cool	C. Slightly cool	D. Neutral
E. Slightly warm	F. Warm	G. Hot	H. Never do it
…			
Question 20: Indoors, you will **wipe sweat from back** when you feel___?			
A. Cold	B. Cool	C. Slightly cool	D. Neutral
E. Slightly warm	F. Warm	G. Hot	H. Never do it

spondents were requested to recall the activities they would engage in when experiencing sensations of cold or hot. The detailed content of the questionnaire is exhibited in Table 2.

The results of the preliminary questionnaire survey indicate a substantial correlation between the proposed 20 thermal comfort activities and human thermal comfort. These results align broadly with the findings of the surveys by Yang [16] and Duan [18]. However, our investigation identified a greater number of thermal comfort activities, such as "tucking hands in sleeves", "rolling down sleeves", "putting hands in pockets", "rubbing ears", and "rubbing knees" when feeling cold; and "unbuttoning", "lifting hair on the back of the neck", "rolling up sleeves", "lifting bangs" when feeling hot.

To further examine the relationships of these 20 thermal comfort activities, we conducted an additional questionnaire survey, distributing 284 questionnaires in total. The in-depth survey included 20 questions, for instance, "Indoors, you will button up when you feel__?" and "Indoors, you will shake T-shirt when you feel__?". Table 3 illustrates the details of the in-depth survey questionnaire.

Each question presented eight options, specifically, "cold", "cool", "slightly cool", "neutral", "slightly warm", "warm", "hot", or "never do it". These options derive from ASHRAE Standard-55 [28]: cold (−3), cool (−2), slightly cool (−1), neutral (0), slightly warm (+1), warm (+2), hot (+3). Fig. 1 depicts the results of the 284 questionnaires. Notably, the activities illustrated in Fig. 1 are organized in descending order based on the proportion of 'cold' or 'hot' responses.

Given that different thermal comfort activities may not uniformly reflect a person's degree of thermal comfort [7], the 20 thermal comfort activities, based on the two broad categories, "cold-related activities" and "hot-related activities", were further segmented into six thermal sensation levels, namely, "cold", "cool", "slightly cool", "slightly warm", "warm", "hot" using the *k*-means algorithm. The *k*-means algorithm is a commonly used unsupervised clustering algorithm that can divide data into k distinct clusters [29]. In this research, k was set at six. Table 4 presents the classification results of the 20 thermal comfort activities. In Table 4, P-comfort signifies the probability of a thermal comfort activity being motivated by thermal comfort state C. P-comfort can be calculated using Eq. (1) [18]:

$$P\left(C \mid A_i\right) = \sum_j P\left(S_j \mid A_i\right) = \sum_j \frac{R_{ij}}{1 - R_{(i-never)}} \tag{1}$$

C denotes cool-discomfort or hot-discomfort. A_i signifies the ith thermal comfort activity, where $i\in\{1,\ldots,20\}$. R_{ij} represents the statistical probability of the ith activity corresponding to the jth thermal sensation, where $j\in\{-3,-2,-1,1,2,3\}$, which correspond to "cold", "cool", "slightly cool", "slightly warm", "warm", "hot". $R_{(i-never)}$ indicates the statistical probability of the ith activity concerning the "never do it" option. $P(S_j|A_i)$ represents the probability of a specific activity A_i being motivated by a specific thermal sensation S_j.

Taking "wiping sweat from back" as an example, the questionnaire results revealed that the probability statistics R_{ij} for choosing "hot", "warm", "slightly warm", "neutral", "never do it" and other cases were 72.54 %, 10.92 %, 1.76 %, 1.76 %, and 11.27 %, respectively. Among these, 11.72 % of the respondents selected "never do it".

According to Eq. (1), $P(hot-discomfort|A)$ = (72.54 % + 10.92 % + 1.76 %) / (1–11.72 %) = 96.03 %. In other words, there is a 96.03 % correlation between "wiping sweat from back" and hot-discomfort. The P-comfort of other activities can be calculated in a similar manner. From Table 3, it is apparent that the P-comfort for the activities "putting hands in pockets", "rolling down sleeves", and "buttoning up" are 86.61 %, 88.69 %, and 83.58 %, respectively. The P-comfort for all other activities is above 90 %, further substantiating the strong correlation between the proposed activities and thermal comfort.

2.2. Thermal comfort activities library

In practice, people perform various other activities besides those related to thermal comfort. These activities are referred to as neutral activities, such as using a phone, looking at a watch, etc. To make this study more applicable to real-world scenarios, 10 neutral activities have been added based on the 10 cold-related and 10 hot-related activities. Table 5 displays the descriptions of the 30 proposed activities and their corresponding pathways for influencing human thermal balance.

3. Experimental data collection methods

3.1. Experimental setup

Single activities are individual movements that describe specific activities and can be part of more complex activities [30]. Daily human activities consist of multiple consecutive single activities. In this study, the activities in Table 5 are considered single activities, used to construct daily human activities. Based on the types of IMU data, the data collection experiment in this study is divided into two phases:

Phase 1 is the single thermal comfort activity experiment. In this phase, volunteers are asked to perform one activity from Table 5 at a time. The data collected in the phase 1 experiment is referred to as single activities data. Each single activity data sample only contains IMU data of a single activity. This data can be used to explore the performance of off-line activity recognition models or as training data.

Phase 2 is the continuous thermal comfort activities experiment. In this phase, volunteers are asked to perform several activities from Table 5 consecutively. The data collected in Phase 2 is referred to as continuous multi-activity data. Each continuous multi-activity data sample

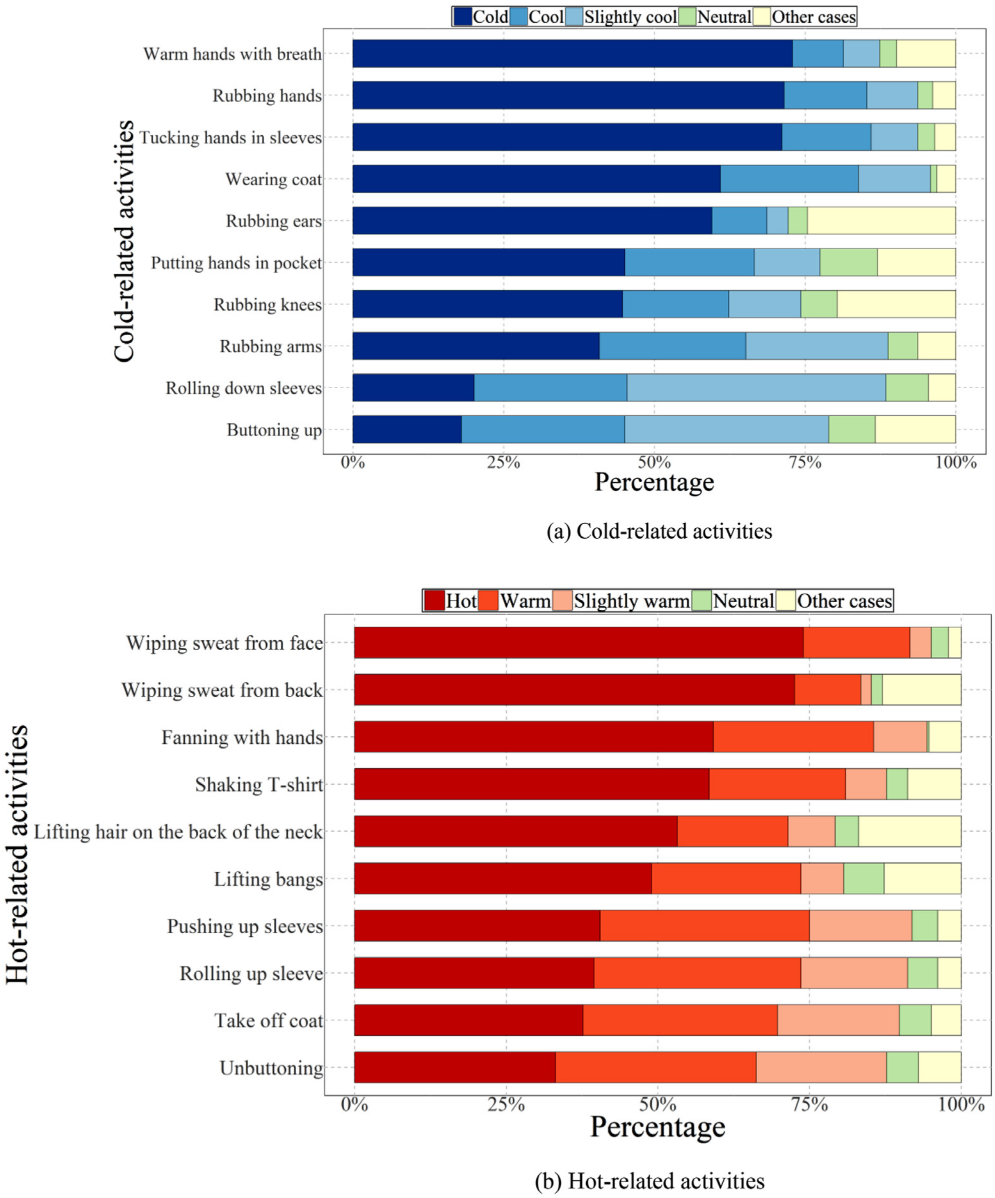


Fig. 1. Statistical results of the in-depth questionnaire.

includes IMU data of multiple single activities performed by the volunteer. Continuous multi-activity data can be used to evaluate the performance of real-time activity recognition algorithms or as testing data.

3.2. Description on experiment participants

Thirty volunteers (14 males and 16 females) were recruited for the experiment among the students of Shenzhen University. They were a cohort of right-handed and healthy individuals. Table 6 displays their anthropometric information, as well as their age, sex, and identifier (ID) in the database. All participants provided written informed consent before the beginning of the experiment. The names of the volunteers were replaced with ID in the dataset to ensure their privacy was protected.

3.3. Description on data acquisition tools

For tracking of fine motion of both hands, two IMUs named BWT61CL were used to collect various movement signals. This IMU is a multi-modal sensor, integrating a 3-axis accelerometer, 3-axis gyroscope, and 3-axis magnetometer. It can simultaneously collect 3-axis acceleration, 3-axis angular velocity, and 3-axis Euler angles at 100 Hz, and transmit these IMU data wirelessly in real-time to a computer. Fig. 2 displays the situation of volunteers wearing the IMUs and an overview of the collection and transmission of IMU data [31].

3.4. Experimental procedure

The data collection experiment for this study was conducted in the Human Factors Engineering Laboratory at Shenzhen University. Fig. 3 shows the experimental flowchart for collecting this dataset. After registering their personal information, each volunteer was required to wear the experimental equipment as instructed and sit at a table to participate in the experiment. The experiment was divided into two phases. Due to the wireless transmission and small size advantages of the IMU device used in this experiment, volunteers could perform various thermal comfort activities according to their habits.

3.4.1. Phase 1

In Phase 1, each volunteer was asked to complete all the activities listed in Table 5 in sequence. To ensure that the collected IMU data

Table 4
Analysis of thermal comfort activities based on in-depth questionnaire results.

Activities classification	Activities	Thermal comfort	P-comfort	Thermal sensation	P-sensation	Thermal sensation scores
Cool-related activities	Warming hands with breath	Cold-discomfort	93.58 %	Cold	78.11 %	−3
	Rubbing hands		96.03 %		73.29 %	
	Tucking hands in sleeves		95.34 %		72.40 %	
	Wearing coat		96.11 %		61.13 %	
	Rubbing ears		93.61 %		77.17 %	
	Putting hands in pockets		86.61 %	Cool	24.02 %	−2
	Rubbing knees		90.95 %		21.55 %	
	Rubbing arms		91.97 %		25.18 %	
	Rolling down sleeves		88.69 %	Slightly cool	43.11 %	−1
	Buttoning up		83.58 %		35.82 %	
Neutral Activities	–	Comfort	–	Neutral	–	0
Hot-related activities	Unbuttoning	Hot-discomfort	92.57 %	Slightly warm	22.68 %	+1
	Taking coat off		90.11 %		20.14 %	
	Rolling up sleeves		92.83 %		17.92 %	
	Pushing up sleeves		93.21 %		17.14 %	
	Lifting bangs		91.24 %	Warm	27.89 %	+2
	Lifting hair on the back of the neck		93.36 %		21.58 %	
	Shaking T-shirt		92.91 %		23.88 %	
	Fanning with hands		98.53 %		27.57 %	
	Wiping sweat from back		96.03 %	Hot	81.75 %	+3
	Wiping sweat from face		95.74 %		74.47 %	

Table 5
Description of 30 activities in the thermal comfort activities library.

Activities classification	No.	Activities	Description
Cold-related activities	1	Buttoning up	Buttoning up the buttons near the top of shirt to reduce the airflow between chest and shirt
	2	Warming hands with breath	Blowing warm air on the palms to warm them
	3	Rubbing hands	Rubbing the hands rapidly to generate heat through friction
	4	Rubbing ears	Rubbing the ears by hands to generate heat through friction
	5	Rolling down sleeves	Putting down the pulled sleeves of both sides to reduce the heat loss of the arms
	6	Tucking hands in sleeves	Tucking hands in the trouser pocket to warm them
	7	Rubbing arms	Crossing hands and rubbing arms to generate heat through friction
	8	Putting hands in pockets	Putting hands in the trouser pocket to warm them
	9	Wearing coat	Wearing the coat to reduce the heat loss of the body
	10	Rubbing knees	People with exposed knees rubbing their knees with their hands repeatedly to generate heat
Hot-related activities	11	Unbuttoning	Unbuttoning the buttons near the top of shirt to increase airflow between chest and shirt
	12	Shaking T-shirt	Shaking T-shirt repeatedly to increase airflow between chest and T-shirt
	13	Fanning with hands	Fanning the hands quickly on both sides of the face to increase the airflow near the face
	14	Wiping sweat from face	Wiping the sweat from face with a tissue
	15	Lifting bangs	Raising the hair in the front of forehead to accelerate the heat loss
	16	Lifting hair on the back of the neck	Raising hair at the back of the neck to prevent it sticking on the neck due to sweat and increase the airflow near the neck
	17	Pushing up sleeves	Pulling up the sleeves of both sides to accelerate the heat loss of the arms
	18	Rolling up sleeves	Folding up the sleeves of both sides carefully to accelerate the heat loss of the arms
	19	Taking coat off	Taking off the coat to accelerate the heat loss of the body
	20	Wiping sweat from back	Wiping the sweat on the back with a tissue
Neutral activities	21	Hands on chin	Putting the hands on chin
	22	Using phone	Picking up the phone on the table and using it with both hands for a while
	23	Typing	Typing by the keyboard on the desk
	24	Drinking water	Picking up and unscrewing the water bottle on the desk and drinking water
	25	Looking at watch	Turning the wrist and looking at the watch for a while
	26	Scratching head	Scratching head with hands
	27	Holding up glasses	Holding up the glasses on the nose
	28	Putting hands on legs	Putting hands on legs for relaxing
	29	Stretching	Raising hands up and stretching the body
	30	Picking something up off the ground	Bending down to pick something up off ground

could cover as many situations as possible, each volunteer was asked to repeat each activity five times (e.g., rubbing hands five times or pushing up sleeves five times). Consequently, each volunteer IMU data collection experiment consisted of 150 sub-experiments (30 activities × 5 times). To ensure the consistency of the experimental data, the IMU sensor was initialized before the start of each sub-experiment. A sub-experiment was considered complete after a volunteer performed a full thermal comfort activity. Once a sub-experiment was completed, the experimenter would immediately record the activity type as the label. Phase 1 took approximately 120 min per subject. Fig. 4 depicts the scenarios of a volunteer performing some cold-related activities, hot-related activities, and neutral activities.

Table 6
Anthropometry information of the participants.

ID	Gender	Age(years)	Height(cm)	Weight(kg)
1	Female	24	164	56
2	Female	24	161	56
3	Female	24	160	58
4	Male	26	178	80
5	Female	26	166	47.5
6	Female	23	155	51
7	Female	26	168	51.5
8	Female	24	164	53
9	Male	23	172	60
10	Female	23	164	50
11	Female	23	167	57.5
12	Female	23	168	47.5
13	Female	23	166	53
14	Female	19	170	55
15	Male	19	170	55
16	Male	25	177	72.5
17	Male	25	174	60
18	Female	24	163	50
19	Female	23	160	50
20	Female	23	156	52
21	Male	24	174	62
22	Male	24	180	65
23	Female	23	154	49
24	Male	22	169	58
25	Male	23	170	60
26	Male	23	173	62.5
27	Male	22	183	90
28	Male	21	185	86
29	Male	24	169	71
30	Male	21	180	67.5
Summary	14 Male, 16 Female	23.23(± 1.70)	168.67(± 8.02)	59.55(± 10.96)

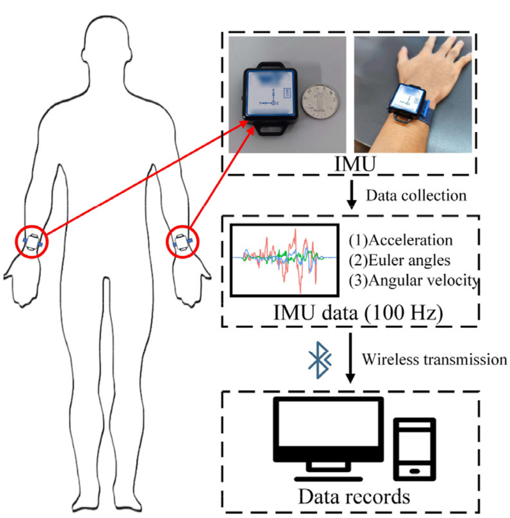


Fig. 2. An illustration of the IMU placement and data transmission.

3.4.2. Phase 2

In Phase 2, the 30 activities in Table 5 are randomly divided into six activity sets. Each activity set contains four to eight activities arranged in a random order. Volunteers are required to perform a series of activities according to the preset order of each activity set. When a volunteer completes all the activities within an activity set, it is considered as a sub-experiment completed. Each volunteer is required to repeat each activity set twice. Additionally, a camera is placed beside the volunteer to record the start and end times of each activity, serving as a basis for labeling the continuous multi-activity data in post-processing.

4. Experimental data

4.1. Data profile and visualization

As previously mentioned, since each volunteer is required to wear two IMUs, and each IMU can simultaneously collect data for nine variables (three-axis acceleration, three-axis angular velocity, and three-axis Euler angles), a total of 18 variables are included in the collected IMU signals. Table 7 displays the 18 variables contained in the dataset. The frequency of each variable is 100 Hz.

The single activities IMU data includes 4500 samples from 30 volunteers (30 subjects * 30 activities * 5 times). The total duration of all samples amounts to 642 min. As previously mentioned, each sample in the single activities data contains IMU data for only one specific activity. For example, a particular sample may include IMU data for a volunteer performing shaking T-shirt.

Table 8 summarizes the statistical characteristics of the duration of each thermal comfort activity sample in the dataset, including the minimum, average, maximum, and standard deviation values. Among them, buttoning up and rolling up sleeves have the highest average duration, at 13.35 s and 13.57 s, respectively; while rolling down sleeves and holding up glasses have the lowest average duration, at 5.14 s and 4.59 s, respectively.

Figure 5 displays the IMU data of multiple thermal comfort activities. Among them, (a) and (b) represent cold-related activities; (c) and (d) represent hot-related activities; (e) and (f) represent neutral activities. It is worth noting that the range of the *X*-axis in each figure is not uniform. This is because the duration of different thermal comfort activities often vary. As can be inferred from Fig. 5, the IMU data corresponding to different thermal comfort activities exhibits unique patterns. For instance, the angular velocity data related to “fanning with hands”, as depicted in Fig. 5(d), displays a certain degree of periodicity and large amplitude variations. Contrastingly, the “stretching” angular velocity data, shown in Fig. 5(f), lacks this periodicity and exhibits smaller amplitude changes than “fanning with hands”. Given its capacity to learn from data and encapsulate complex features, machine learning proves advantageous for discerning these differing patterns [32]. Thus, machine learning methods can effectively capture the distinct characteristics inherent in each activity, enabling accurate identification.

The continuous multi-activity data contains 360 samples (6 activities sets * 30 subjects * 2 times). The total duration of all continuous multi-activity samples is 225 min. As mentioned earlier, each continuous multi-activity data sample includes IMU data of volunteers performing multiple activities consecutively. Fig. 6 shows a continuous multi-activity sample, which records the IMU data of a volunteer performing drinking water, using phone, fanning with hands, buttoning up, and wiping sweat sequentially.

4.2. Dataset structures

The IMU Dataset is stored in NutCloud. Fig. 7 shows the organization of this dataset. Level 1 is the entire dataset. Level 2 contains two folders, one for storing single activities data and the other for storing continuous multi-activity data. Level 3 corresponds to different volunteers. Level 4 corresponds to different activities or activity sets. Level 5 stores csv files containing IMU data, with each file corresponding to different experiment times.

The specific file names are determined by the volunteer, thermal comfort activity, and activity count. For example, “s1_a1_t1.csv” refers to the csv file containing the IMU data of volunteer 1 performing activity 1 (Buttoning up) for the first time. “s1_c1_t2.csv” records the IMU data of volunteer 1 performing the activity set 1 for the second time. It is worth noting that the division of activity sets is random, and the activity sets corresponding to each volunteer are not the same. Additionally, the activities type, start times, and end times corresponding

Table 7
18 variables contained in the dataset.

Variables	Unit	IMUs placement					
		Left hand			Right hand		
		X-axis	*Y*-axis	*Z*-axis	*X*-axis	*Y*-axis	*Z*-axis
Acceleration	g	1	2	3	4	5	6
Angular velocity	°/s	7	8	9	10	11	12
Euler angles	°	13	14	15	16	17	18

Table 8
The statistical characteristics of activity duration (unit: seconds).

Num	Activities	Minimum	Mean	Maximum	Standard deviation
1	Buttoning up	7.34	13.35	21.87	3.01
2	Warming hands with breath	5.28	9.07	13.32	1.43
3	Rubbing hands	3.83	7.52	18.42	2.47
4	Rubbing ears	4.47	7.05	14.95	1.96
5	Rolling down sleeves	2.73	5.14	10.87	1.42
6	Tucking hands in sleeves	7.15	11.64	19.96	2.72
7	Rubbing arms	5.34	9.45	19.98	3.17
8	Putting hands in pockets	7.51	10.78	17.81	1.96
9	Wearing coat	6.90	12.51	20.88	3.40
10	Rubbing knees	4.47	7.61	12.86	1.96
11	Unbuttoning	5.43	9.01	16.88	2.12
12	Shaking T-shirt	2.91	6.08	15.54	2.28
13	Fanning with hands	3.67	6.60	12.47	1.62
14	Wiping sweat from face	4.78	9.00	19.87	2.97
15	Lifting bangs	4.49	8.83	17.73	2.24
16	Lifting hair on the back of the neck	3.61	7.41	15.40	2.73
17	Pushing up sleeves	2.86	5.62	12.70	1.93
18	Rolling up sleeves	6.35	13.57	25.73	4.43
19	Taking coat off	6.70	10.53	18.67	2.32
20	Wiping sweat from back	5.15	9.56	19.72	3.04
21	Hands on chin	5.21	9.21	15.76	1.94
22	Using phone	6.90	12.03	20.66	2.51
23	Typing	2.75	5.85	10.98	1.52
24	Drinking water	6.88	10.41	18.47	2.19
25	Looking at watch	2.64	5.49	12.13	1.95
26	Scratching head	3.22	6.61	15.66	2.06
27	Holding up glasses	2.46	4.59	11.33	1.47
28	Putting hands on legs	4.99	9.07	14.19	1.78
29	Stretching	3.98	7.21	13.44	1.65
30	Picking something up off the ground	3.36	5.30	7.37	0.92

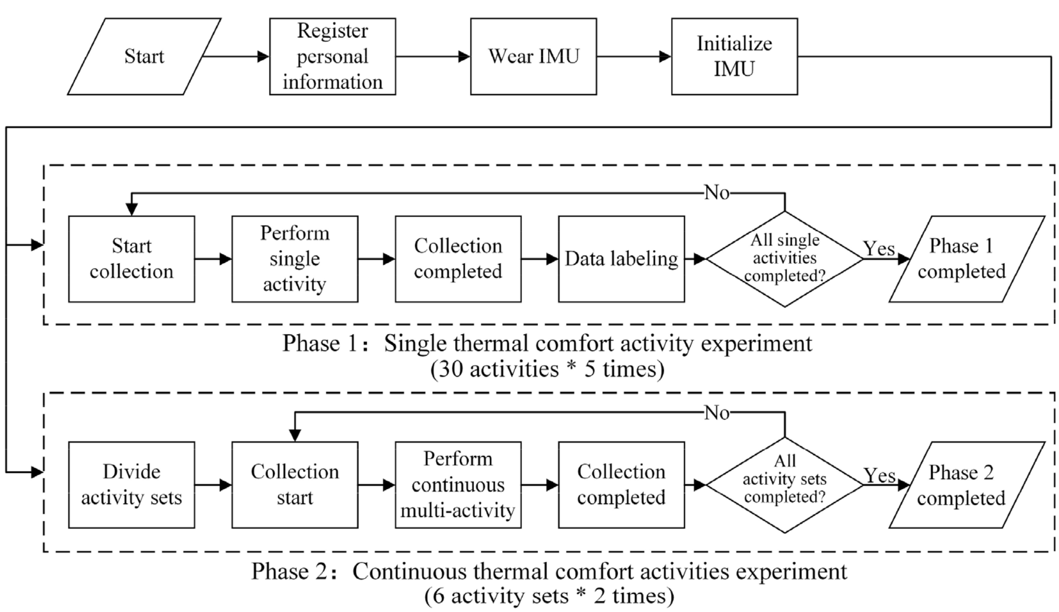


Fig. 3. Experimental flowchart.

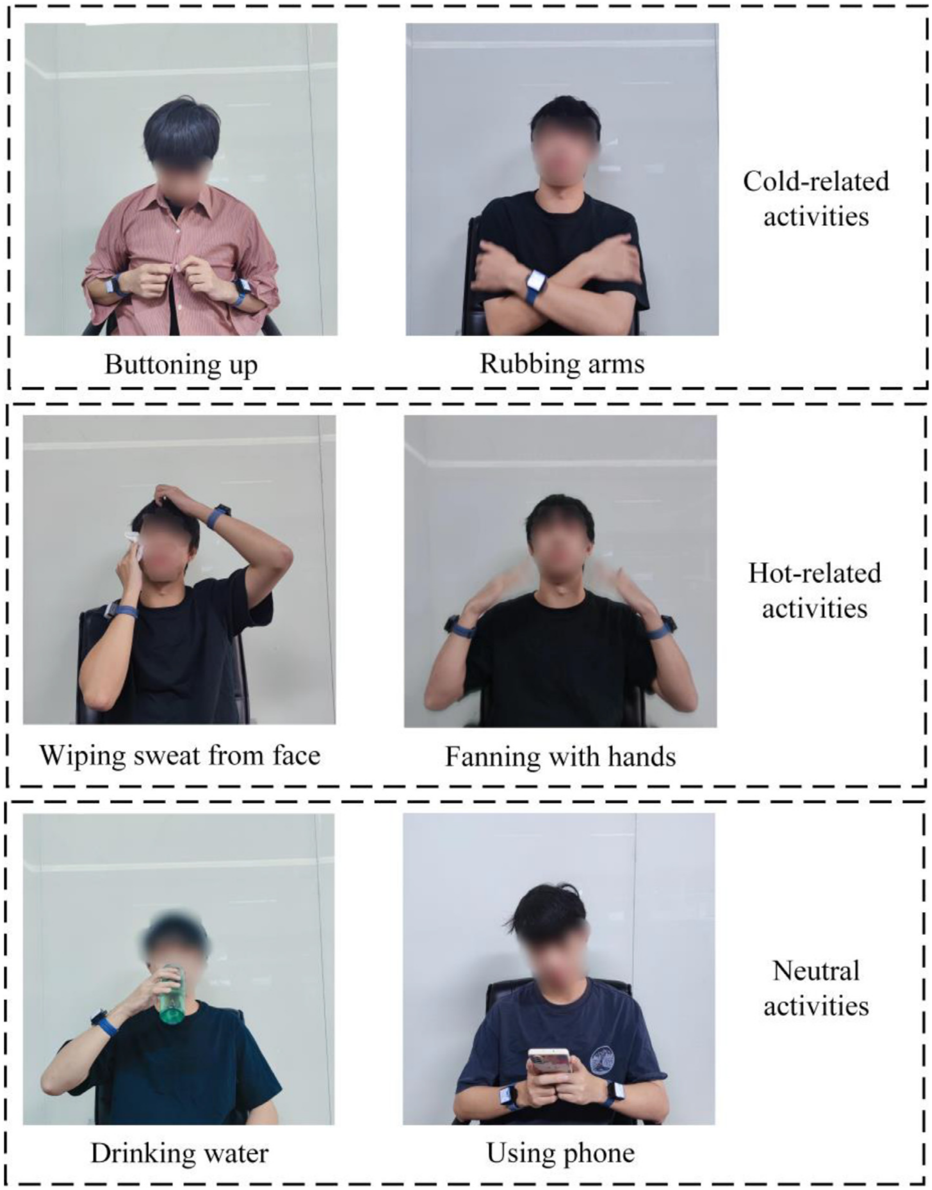


Fig. 4. Scenarios of a volunteer performing various thermal comfort activities.

to each continuous multi-activity sample are recorded in a file named "label.csv".

5. Case study on data-driven thermal comfort activity identification

5.1. Modeling development

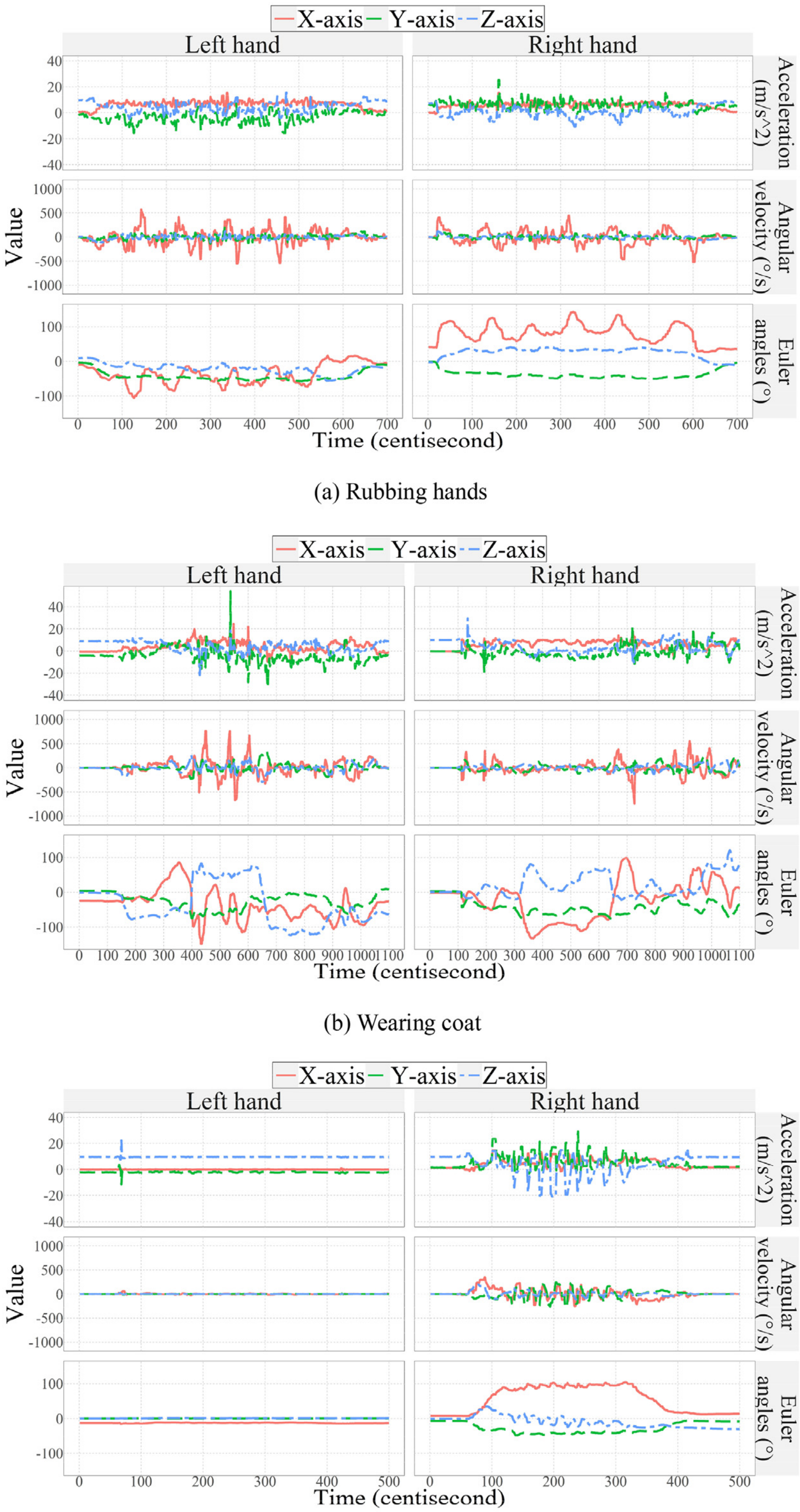
After creating the dataset, this study conducted a preliminary attempt based on the single activities data to classify 30 thermal comfort activities. 1D-CNN is commonly used to extract high-dimensional nonlinear features, while LSTM is adept at capturing the temporal information of input data [33]. Therefore, this study established an activity recognition model by combining 1D-CNN and LSTM.

Wearable devices on the market generally only come equipped with accelerometers and gyroscopes. Therefore, three-axis acceleration and three-axis angular velocity were used as input variables. To avoid the impact of dimension differences on model training, Z-score normalization was applied to all variables before inputting them into the model. Z-score normalization can transform the mean of each variable to 0 and the variance to 1. The neural network model was developed using the Keras package. All data analysis, data preprocessing, and modeling operations in this study were implemented using the R programming language [34].

Table 9
The grid-search settings for pre-trained model optimization.

Parameters	Candidate values
The number of 1D convolutional blocks	1, 2, 3, 4, 5, 6
The filter number in each 1D convolutional layer	100, 200, 300, 400, 500
Kernel size of 1D convolutions	4, 6, 8, 12, 16
The activation function in hidden layers	Relu, Tanh, Sigmoid
The number of recurrent units in LSTM layer	50, 100, 150, 200
Drop-out	0, 0.25, 0.5

To ascertain the optimal hyperparameters for the model, the single activities data was divided in a 6:2:2 ratio. Specifically, 60 % of the data served as the training set, while an additional 20 % functioned as the validation set, employed for verifying the model's performance during training; the remaining 20 % was utilized as a test set for assessing model performance. It is noteworthy that stratified sampling was conducted according to activity categories to ensure identical proportions of activity samples in the training, validation, and test sets. Table 9 summarizes the detailed grid search settings, which include dropout, a simple yet effective technique in the realm of deep learning for averting model



(a) Rubbing hands

(b) Wearing coat

(c) Shaking T-shirt

Fig. 5. IMU data of single thermal comfort activities.

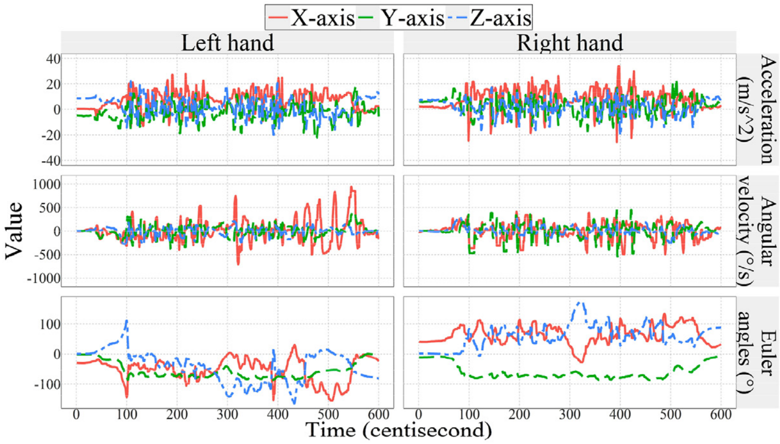
overfitting. Dropout entails randomly nullifying a certain proportion of nodes during model training [35].

After grid searching, the model adopted the following settings: 4 for the number of 1D convolutional blocks, 200 for the filter number in each 1D convolutional layer, 12 for the kernel size of 1D convolutions, Relu for the activation function, 100 for the number of recurrent units in the LSTM layer, and a 25 % dropout rate.

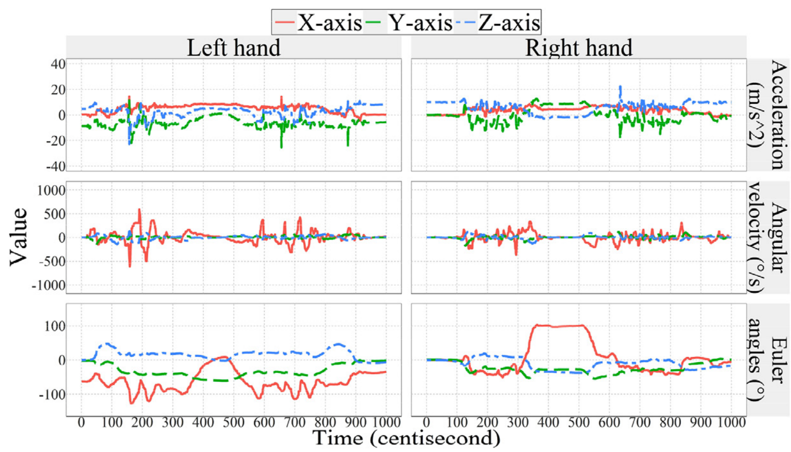
After grid searching, the model adopted the following settings: 4 for the number of 1D convolutional blocks, 200 for the filter number in each 1D convolutional layer, 12 for the kernel size of 1D convolutions, Relu for the activation function, 100 for the number of recurrent units in the LSTM layer, and a 25 % dropout rate.

5.2. Results of identification

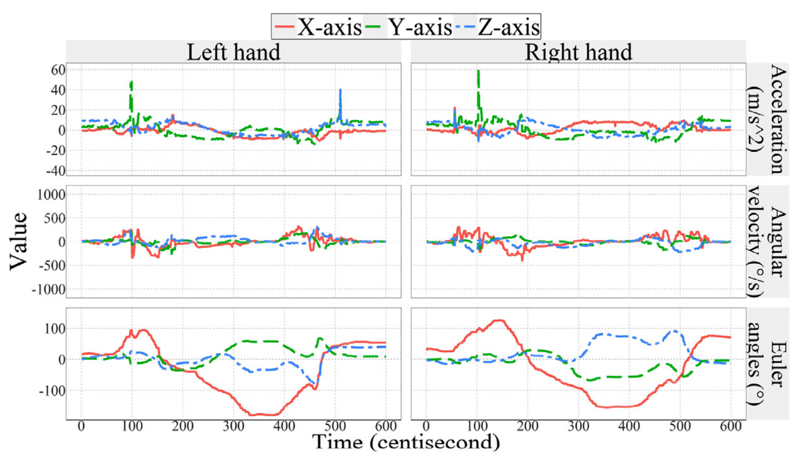
To mitigate the impact of randomness on results, this study employed 30-fold cross-validation to evaluate model performance. Notably, for



(d) Fanning with hands



(e) Drinking water



(f) Stretching

Fig. 5. Continued

the 30-fold cross-validation in this paper, each fold's dataset was partitioned based on volunteers. Specifically, during each fold validation, data from 29 volunteers was used for model training, and data from the remaining volunteer served as the test set for evaluating model performance.

The 1D-CNN-LSTM model demonstrated an overall accuracy of 82.51 % in distinguishing 30 thermal comfort activities. Fig. 8 exhibits the F1-scores for the 30 thermal comfort activities. It is evident that aside from a few activities (such as "scratching head" and "lifting hair on the back of the neck") which are similar and thus prone to misclassification, the F1-scores of most activities exceed 80 %. This underscores that the proposed model can accurately discern prevalent human thermal comfort activities. Constructing thermal comfort detection models with this dataset can produce commendable outcomes.

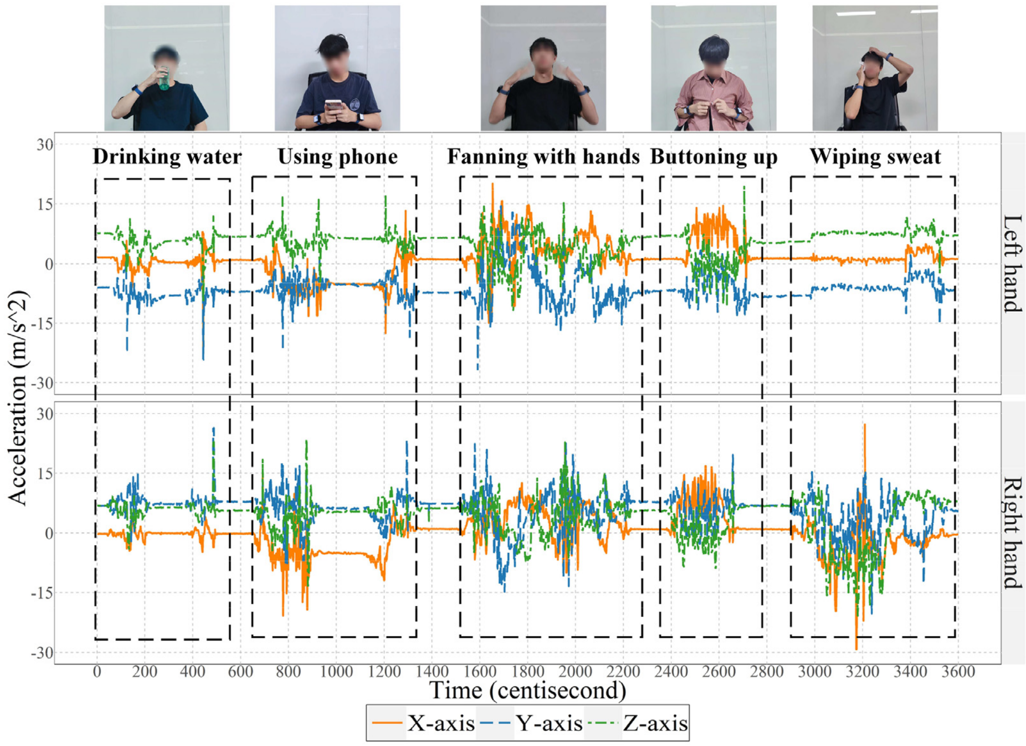


Fig. 6. Continuous multi-activity data (acceleration as an example).

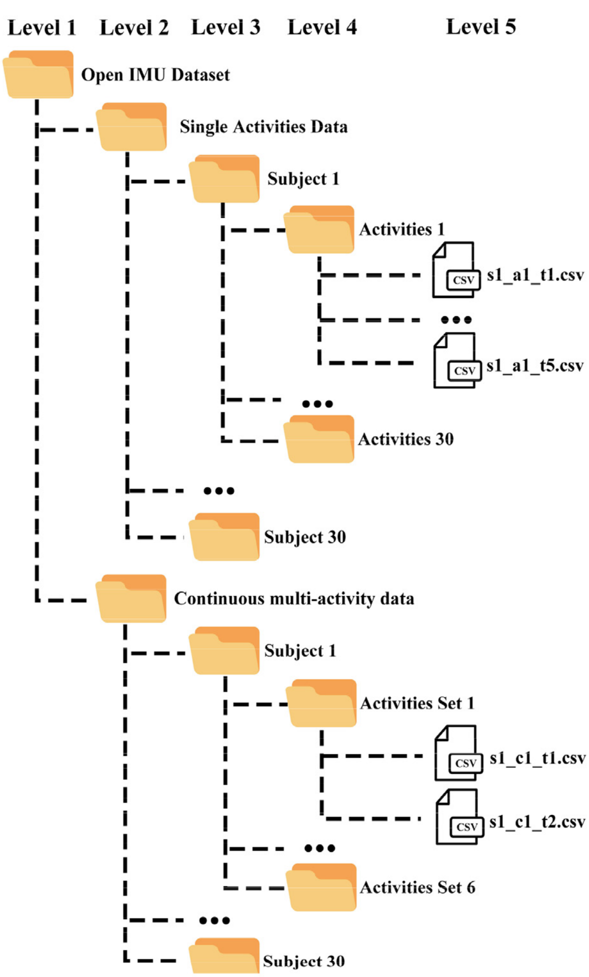


Fig. 7. Organization of the Open IMU dataset.

5.3. Quantification of thermal comfort for HVAC control

In order to guide the operation of HVAC systems through the detection of thermal comfort activities and to maximize the thermal comfort of indoor occupants, we proposed a strategy based on detected thermal comfort activities to quantify indoor human thermal comfort and control HVAC systems. The flowchart of this strategy can be seen in Fig. 9. Within this, the thermal comfort index (TCI) is a parameter defined in this paper to describe the human comfort level with the environmental temperature. When the TCI is zero, it signifies that the subject is in a state of thermal neutrality. If the TCI is greater than zero, it indicates that the subject feels hot, and the greater the TCI, the more intense the sensation of heat. Conversely, if the TCI is less than zero, it suggests that the subject feels cold.

The TCI varies based on the detected thermal comfort activities, with the degree of change contingent on the thermal sensation scores of those activities. For instance, when the activity “fanning with hands” is detected (with a thermal sensation score of +2), the TCI will increase by 2. When the activity “wearing coat” is detected (with a thermal sensation score of −3), the TCI will decrease by 3. The thermal sensation scores for the 20 thermal comfort activities proposed can be found in Table 4. The thermal sensation scores for neutral activities have been set at 0.

When the system detects multiple thermal comfort activities leading to the TCI exceeding a predetermined threshold, the system will signal the HVAC to adjust parameters such as temperature setpoints, thus accommodating the thermal comfort needs of indoor occupants. If no further thermal comfort activities are detected over a certain duration, the absolute value of the TCI is reset to 0.

Table 10 provides such an example: (1) At 9:00, the system begins operation, and the TCI is initialized to 0. (2) At 9:38, 9:39, and 9:40, by analyzing IMU data, the system detects that the occupant performed the activities of “fanning with hands”, “wiping sweat from the face”, “and shaking T-shirt” respectively. As a result, the TCI increased from 0 to 7 (7 = 0 + 2 + 3 + 2). (3) Since the TCI exceeded the pre-set threshold (assumed to be 6) at 9:40, the system sends a signal to the HVAC to cool

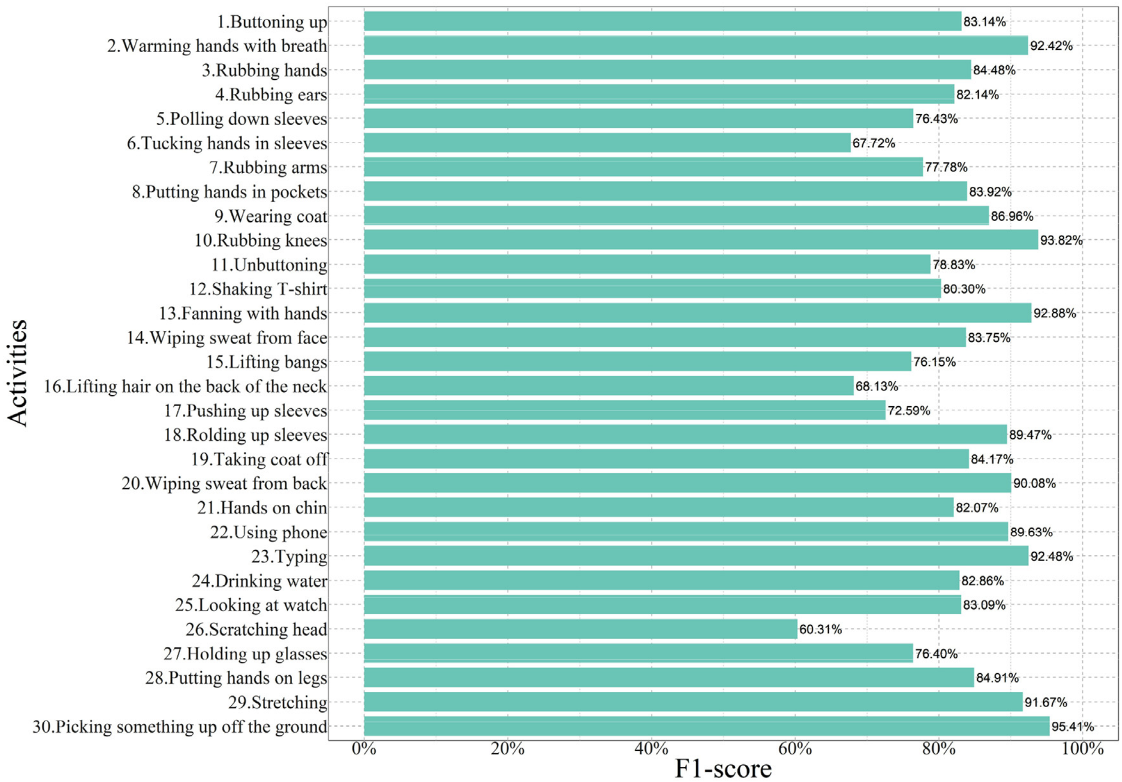


Fig. 8. F1-score for each thermal comfort activity.

Table 10
Example of triggering HVAC cooling upon detecting hot-related activities.

Time	Event number	Activities detected	Thermal sensation scores	TCI
9:00	0	Start		0
9:38	1	Fanning with hands	+2	2
9:39	2	Wiping sweat from face	+3	5
9:40	3	Shaking T-shirt	+2	7
9:40	4	TCI exceeds the threshold (+6), trigger HVAC to decrease temperature setpoint		7
9:50	5	No thermal comfort activities observed for 10 min, TCI return to 0		0

the room. (4) During the time period from 9:40 to 9:50, the system does not detect any additional thermal comfort activities, and the absolute value of the TCI is reset to 0.

6. Discussions

6.1. Insights into the data

The issues of data imbalance, classification challenges, and handling outliers are prominent concerns when utilizing data. In the following discussion, we will delve further into these typical problems, thereby illuminating our insights regarding this data.

6.1.1. Imbalanced data

Given the configuration of our experiment, the dataset is devoid of substantial bias. The study design ensured that every participant engaged in each activity five times, thereby fostering a balanced dataset. Consequently, via the contribution of 30 volunteers, we amassed a sum of 4500 samples, equating to 150 samples per activity.

6.1.2. Classification challenges

The process of activity classification can manifest hurdles, particularly when the activities under consideration exhibit similar characteristics. For example, the task of distinguishing between “lifting hair on the back of the neck” and “scratching head ” was arduous, as denoted by the respective F1 scores of 68.16 % and 60.13 %.

6.1.3. Outliers and noises

A multitude of factors can inject outliers and noise into the data, such as collisions of IMU or disruptions in data transmission. To alleviate the impacts of such anomalies, we suggest the implementation of customary IMU data preprocessing methodologies, such as mean filtration, median filtration, and Kalman filtration.

6.2. Limitations

Limitations do exist in this study which deserve further studies:

(1) The prediction of thermal comfort based on activity may not be consistent with the actual thermal comfort perception of an occupant [36,37]. For instance, an individual might display a certain behavior out of habit rather than an indicator of their thermal comfort level. In future work, we will explore combining IMU data with environmental factors (such as indoor temperature, humidity, wind speed, etc.) and individual physiological parameters

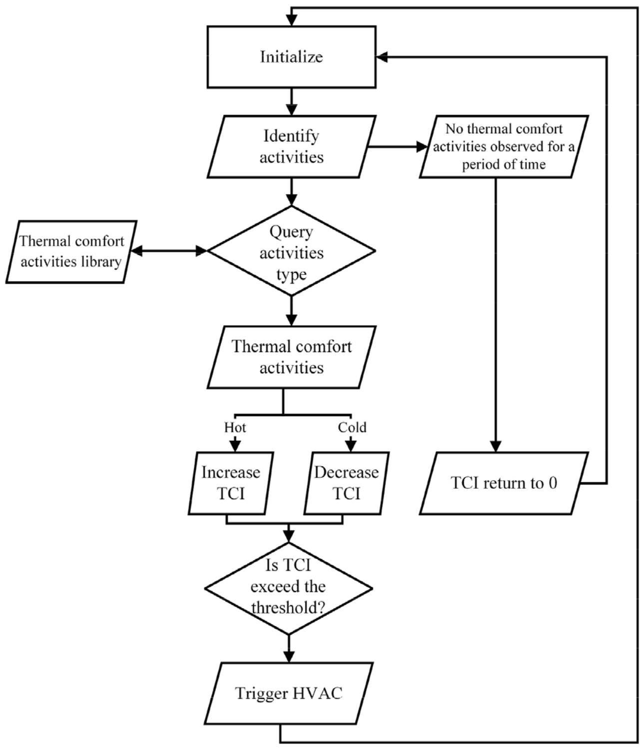


Fig. 9. Flow chart of HVAC operation based on thermal comfort activity recognition.

(like heart rate, skin temperature, etc.) to improve the accuracy of thermal comfort prediction.

(2) There is a limitation in the demographic diversity of the participants who took part in our IMU data collection experiment. Most of the participants were students, and different age groups and cultural backgrounds may display different habits when performing the same activity. This could challenge the accurate identification of thermal comfort activities in diverse populations based on our current dataset. Future efforts should aim to include volunteers with more diverse backgrounds to develop a more comprehensive thermal comfort activity IMU dataset.

7. Conclusions

Thermal comfort information of indoor occupants is a crucial factor in promoting HVAC system efficiency and maximizing indoor occupant comfort. Compared to image data, IMU-based activity recognition for thermal comfort detection boasts advantages such as high recognition accuracy, low cost, and non-intrusiveness. However, the lack of relevant IMU datasets has hindered further research. Consequently, this paper establishes an accessible IMU dataset, the first of its kind for human thermal comfort detection, successfully filling the research gap in IMU-based thermal comfort detection. Specifically, the paper encompasses the following work:

(1) Following two stages of questionnaire survey, 20 activities reflecting human thermal comfort were defined. Using the *k*-means algorithm to analyze the questionnaire results, these activities were divided into six categories (cold, cool, slightly cool, slightly warm, warm, hot). The probabilities, P-comfort, of each thermal comfort activity being stimulated by hot-discomfort or cool-discomfort were calculated.

(2) Thirty volunteers were recruited, and their 9-axis IMU data was collected using wearable devices while they performed the 30 thermal comfort activities. In this dataset, the sample size of single activities data amounts to 4500, suitable for training activity recognition models. The continuous multi-activity IMU data can be employed to assess the real-time activity recognition performance of the model.

(3) A thermal comfort activity recognition model was established using the 1D-CNN-LSTM network. Under 30-fold cross-validation, the recognition accuracy reached 82.51 %, indicating that favorable results can be achieved when applying this dataset for thermal comfort state detection.

Researchers can utilize this dataset and combine it with more advanced algorithms to develop real-time thermal comfort inference methods that balance privacy and recognition efficiency. The research outcomes can be integrated with HVAC systems to enable intelligent control and enhance indoor occupant thermal comfort levels.(4) A method for quantifying human thermal comfort based on the identification of thermal comfort activities has been proposed. This approach calculates the TCI through activity recognition, thus enabling the control of the HVAC system in accordance with the thermal comfort status of individuals.

Declaration of Competing Interests

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.

CRediT authorship contribution statement

Weilin He: Methodology, Investigation, Validation, Writing – original draft. **Cheng Fan:** Conceptualization, Funding acquisition, Project administration, Supervision, Writing – review & editing. **Zebin Wu:** Investigation. **Qiaoqiao Yong:** Investigation.

Acknowledgments

The authors gratefully acknowledge the support of this research by the National Natural Science Foundation of China (No. 52278117), the Philosophical and Social Science Program of Guangdong Province, China (GD22XGL20) and the Shenzhen Science and Technology Program (Nos. 20220531101800001 and 20220810160221001).

References

[1] Z. Wang, Y. Ji, J. Ren, Thermal adaptation in overheated residential buildings in severe cold area in China, Energy Build. 146 (2017) 322–332.
[2] J. Jiang, D. Wang, Y. Liu, Y. Di, J. Liu, A holistic approach to the evaluation of the indoor temperature based on thermal comfort and learning performance, Build. Environ. 196 (2021) 107803.
[3] T. Labeodan, W. Zeiler, G. Boxem, Y. Zhao, Occupancy measurement in commercial office buildings for demand-driven control applications—A survey and detection system evaluation, Energy Build. 93 (2015) 303–314.
[4] O. Ardakanian, A. Bhattacharya, D. Culler, Non-intrusive techniques for establishing occupancy related energy savings in commercial buildings, in: Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments, 2016, pp. 21–30.
[5] F.R. Ambrosio Alfano, B.W. Olesen, B.I. Palella, D. Pepe, G Riccio, Fifty years of PMV model: reliability, implementation and design of software for its calculation, Atmosphere 11 (1) (2019) 49.
[6] A. Dixit, U. Gade, A case study on human bio-heat transfer and thermal comfort within CFD, Build. Environ. 94 (2015) 122–130.
[7] A. Meier, W. Dyer, C. Graham, Using human gestures to control a building's heating and cooling System, in: Energy Efficiency in Domestic Appliances and Lighting (EEDAL'17), 2017, pp. 627–635.
[8] P.O. Fanger, Assessment of man's thermal comfort in practice, Occup. Environ. Med. 30 (4) (1973) 313–324.
[9] J.T. Kim, J.H. Lim, S.H. Cho, G.Y. Yun, Development of the adaptive PMV model for improving prediction performances, Energy Build. 98 (2015) 100–105.
[10] Y. Yang, B. Li, H. Liu, M. Tan, R. Yao, A study of adaptive thermal comfort in a well-controlled climate chamber, Appl. Therm. Eng. 76 (2015) 283–291.
[11] S.I.U.H. Gilani, M.H. Khan, M. Ali, Revisiting Fanger's thermal comfort model using mean blood pressure as a bio-marker: an experimental investigation, Appl. Therm. Eng. 109 (2016) 35–43.
[12] C. Dai, H. Zhang, E. Arens, Z. Lian, Machine learning approaches to predict thermal demands using skin temperatures: steady-state conditions, Build. Environ. 114 (2017) 1–10.

[13] Y. Yao, Z. Lian, W. Liu, C. Jiang, Y. Liu, H. Lu, Heart rate variation and electroencephalograph-the potential physiological factors for thermal comfort study, Indoor Air 19 (2) (2009) 93.
[14] H. Park, D.Y. Park, Prediction of individual thermal comfort based on ensemble transfer learning method using wearable and environmental sensors, Build. Environ. 207 (2022) 108492.
[15] K.E. Chad, J.M.M. Brown, Climatic stress in the workplace: its effect on thermoregulatory responses and muscle fatigue in female workers, Appl. Ergon. 26 (1) (1995) 29–34.
[16] B. Yang, X. Cheng, D. Dai, T. Olofsson, H. Li, A. Meier, Real-time and contactless measurements of thermal discomfort based on human poses for energy efficient control of buildings, Build. Environ. 162 (2019) 106284.
[17] J. Qian, X. Cheng, B. Yang, Z. Li, J. Ren, T. Olofsson, H. Li, Vision-based contactless pose estimation for human thermal discomfort, Atmosphere 11 (4) (2020) 376.
[18] W. Duan, Y. Wang, J. Li, Y. Zheng, C. Ning, P. Duan, Real-time surveillance-video-based personalized thermal comfort recognition, Energy Build. 244 (2021) 110989.
[19] Y. Wang, Y. Han, Y. Wu, E. Korkina, Z. Zhou, V. Gagarin, An occupant-centric adaptive façade based on real-time and contactless glare and thermal discomfort estimation using deep learning algorithm, Build. Environ. 214 (2022) 108907.
[20] K. Li, J. Wang, R. Yu, W. Xue, Non-invasive thermal sensation recognition based on human behavior postures in office environment, in: 2022 41st Chinese Control Conference (CCC, IEEE, 2022, pp. 6714–6719.
[21] Y. Liu, H. Li, P. Zheng, H. Wu, Z. Wang, Z. Li, B. Lin, Non-invasive measurements of thermal discomfort for thermal preference prediction based on occupants' adaptive behavior recognition, Build. Environ. 228 (2023) 109889.
[22] N. Gupta, S.K. Gupta, R.K. Pathak, V. Jain, P. Rashidi, J.S. Suri, Human activity recognition in artificial intelligence framework: a narrative review, Artif. Intell. Rev. 55 (6) (2022) 4755–4808.
[23] B. Yang, X. Li, Y. Hou, A. Meier, X. Cheng, J.H. Choi, Non-invasive (non–contact) measurements of human thermal physiology signals and thermal comfort/discomfort poses-a review, Energy Build. 224 (2020) 110261.
[24] J. Wannenburg, R. Malekian, Physical activity recognition from smartphone accelerometer data for user context awareness sensing, IEEE Trans. Syst., Man, Cybern. 47 (12) (2016) 3142–3149.
[25] R. Vleugels, B. Van Herbruggen, J. Fontaine, E. De Poorter, Ultra-wideband indoor positioning and IMU-based activity recognition for ice hockey analytics, Sensors 21 (14) (2021) 4650.
[26] J. Qi, P. Yang, M. Hanneghan, S. Tang, B. Zhou, A hybrid hierarchical framework for gym physical activity recognition and measurement using wearable sensors, IEEE Internet Things J. 6 (2) (2018) 1384–1393.
[27] S. Zhang, Y. Li, S. Zhang, F. Shahabi, S. Xia, Y. Deng, N. Alshurafa, Deep learning in human activity recognition with wearable sensors: a review on advances, Sensors 22 (4) (2022) 1476.
[28] R.J. De Dear, G.S. Brager, J. Reardon, F. Nicol, Developing an adaptive model of thermal comfort and preference/discussion, ASHRAE Trans. 104 (1998) 145.
[29] M. Ahmed, R. Seraj, S.M.S. Islam, The *k*-means algorithm: a comprehensive survey and performance evaluation, Electronics 9 (8) (2020) 1295.
[30] M. Vrigkas, C. Nikou, I.A. Kakadiaris, A review of human activity recognition methods, Front. Robot. AI 2 (2015) 28.
[31] C. Fan, W. He, L. Liao, Real-time machine learning-based recognition of human thermal comfort activities using inertial measurement unit data, Energy Build. 294 (2023) 113216.
[32] S. Qiu, H. Zhao, N. Jiang, Z. Wang, L. Liu, Y. An, Multi-sensor information fusion based on machine learning for real applications in human activity recognition: state-of-the-art and research challenges, Inf. Fusion 80 (2022) 241–265.
[33] C. Fan, J. Wang, W. Gang, S. Li, Assessment of deep recurrent neural network-based strategies for short-term building energy predictions, Appl. Energy 236 (2019) 700–710.
[34] R Development Core Team. (2013). R: a language and environment for statistical computing.
[35] C. Garbin, X. Zhu, O. Marques, Dropout vs. batch normalization: an empirical study of their impact to deep learning, Multimed. Tools Appl. 79 (2020) 12777–12815.
[36] B. Lala, A. Hagishima, A review of thermal comfort in primary schools and future challenges in machine learning based prediction for children, Buildings 12 (11) (2022) 2007.
[37] Z. Wang, J. Wang, Y. He, Y. Liu, B. Lin, T. Hong, Dimension analysis of subjective thermal comfort metrics based on ASHRAE global thermal comfort database using machine learning, J. Build. Eng. 29 (2020) 101120.