Personal thermal comfort models with wearable sensors

Shichao Liu[1,2*], Stefano Schiavon[1], Hari Prasanna Das[3], Ming Jin[3], Costas J. Spanos[3]

[1] Center for the Built Environment, University of California, Berkeley, CA, USA
[2] Department of Civil and Environmental Engineering, Worcester Polytechnic Institute, MA, USA
[3] Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA, USA

* Shichao Liu (sliu8@wpi.edu)

Citation:

Liu, Shichao, Stefano Schiavon, Hari Prasanna Das, Ming Jin, and Costas J. Spanos. "Personal thermal comfort models with wearable sensors." Building and Environment (2019), doi: 10.1016/j.buildenv.2019.106281

Highlights

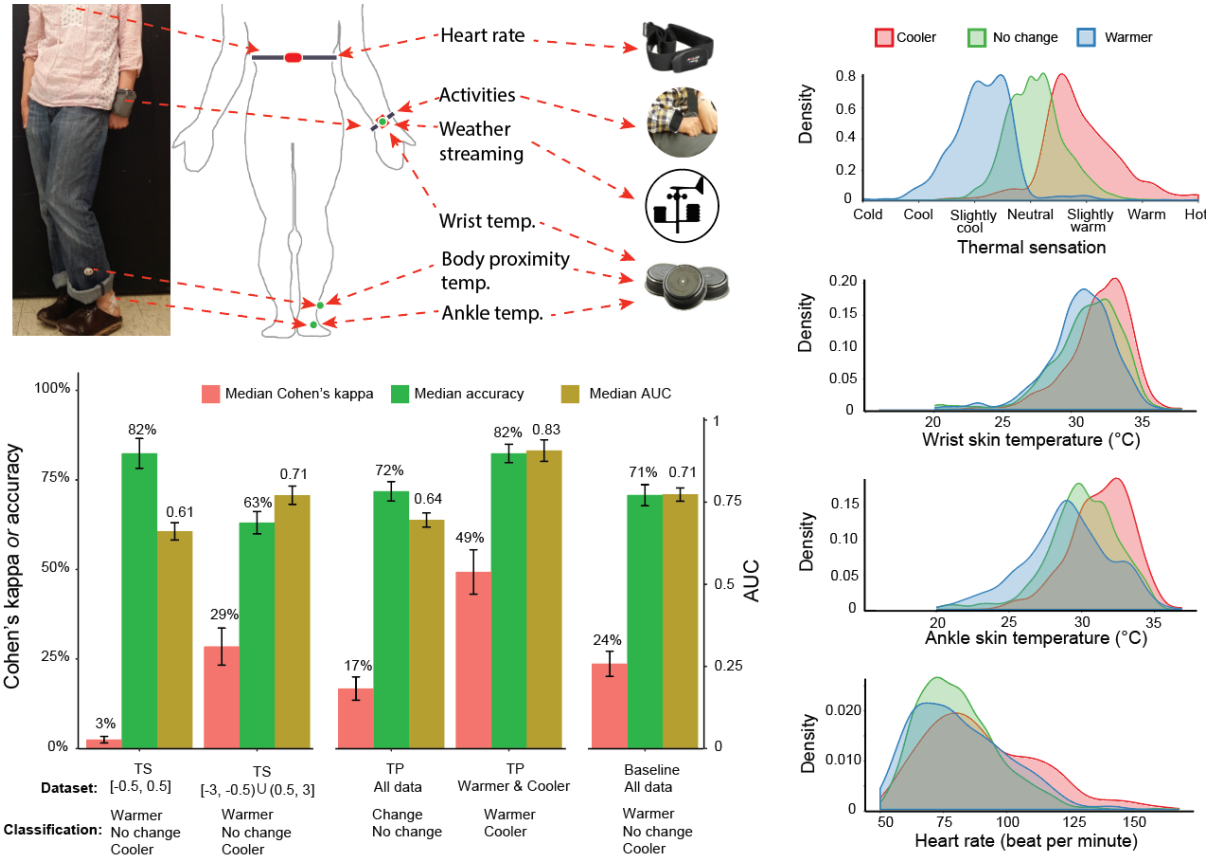
- We develop personal thermal comfort models using lab grade wearable sensors
- Median prediction power of the models is 24% /78% /0.79 (Cohen's kappa/accuracy/AUC)
- Developed models show strong performance outside thermal neutrality
- Ankle skin temperature is more predictive than wrist skin temperature

Abstract

A personal comfort model is an approach to thermal comfort modeling, for thermal environmental design and control, that predicts an individual's thermal comfort response, instead of the average response of a large population. We developed personal thermal comfort models using lab grade wearable in normal daily activities. We collected physiological signals (e.g., skin temperature, heart rate) of 14 subjects (6 female and 8 male adults) and environmental parameters (e.g., air temperature, relative humidity) for 2-4 weeks (at least 20 hours per day). Then we trained 14 models for each subject with different machine-learning algorithms to predict their thermal preference. The results show that the median prediction power could be up to 24% /78% /0.79 (Cohen's kappa/accuracy/AUC) with all features considered. The median prediction power reaches 21% /71% /0.7 after 200 subjective votes. We explored the importance of different features on the prediction performance by considering all subjects in one dataset. When all features included for the entire dataset, personal comfort models can generate the highest performance of 35% /76% /0.80 by the most predictive algorithm. Personal comfort models display the highest prediction power when occupants' thermal sensations is outside thermal neutrality. Skin temperature measured at the ankle is more predictive than measured at the wrist. We suggest that Cohen's kappa or AUC should be employed to assess the performance of personal thermal comfort models for imbalanced datasets due to the capacity to exclude random success.

Keywords: thermal preference; heart rate; skin temperature; machine learning; building-occupant interaction;

Graphical abstract



1. Introduction

Occupants' thermal comfort is associated with health [1,2], working productivity [3–6], learning performance [4,7] and well-being [8,9]. Indoor thermal environment design and thermostat settings in most buildings with mechanical systems rely on air temperature control values based on the existing predicted mean vote (PMV) model as described in thermal comfort standards as ASHRAE Standard 55 [10], EN 15251 [11] and ISO 7730 [12], while the adaptive model is used for free-running buildings[13].

Nevertheless, neither PMV nor the adaptive model incorporates individual differences and dynamics in thermal perceptions. Also, both models ignore aspects of human thermo-regulation and important personal psychophysics influencing the perception of thermal comfort [14,15]. The PMV predicts thermal sensation correctly only one out of three times and has a mean absolute error of one unit on the thermal sensation scale[16]. The main limitation of both PMV and adaptive models is that these two models were developed based on aggregated data from a large population. They were designed to predict the average thermal comfort of the entire population rather than an individual. It has been proven that their accuracy on predicting thermal comfort for a specific occupant is very low. Kim et al. [17] proposed a framework of personal comfort models that can predict an individual's thermal comfort responses by leveraging the Internet of Things (IoT) and machine learning, rather than the responses of an "average person." Such framework has been applied in a few recent studies that aimed to "customize" thermal comfort models for each occupant through users' feedback, IoT and machine learning [18–21]. The primary advantage of a personal thermal comfort model lies in its capacity of self-learning and updating to suit an individual with a data-driven approach, resulting in higher prediction power.

https://doi.org/10.1016/j.buildenv.2019.106281
https://escholarship.org/uc/item/3fb0p5gk

Numerous recent studies have developed personal thermal comfort models by feeding different variables into machine learning algorithms. The three primary categories of variables are 1) environmental information, 2) occupant behavior, and 3) physiological signals. Probability distributions of thermal comfort for each occupant were created over indoor temperature for HVAC controls [22]. A similar data-driven method with indoor environment was applied to classify occupants' personal thermal comfort with temperature and humidity sensors [23]. The second option is to track occupants' behaviors to infer thermal comfort and preference, such as adjusting thermostats [24] or changing the settings of personal heating/cooling devices [25]. A personal comfort model using only control behavior of a personal chair system can generate a prediction AUC of 69% compared to approximately 53% (almost random) for the PMV and adaptive model [26]. Along with the behavior-tracking, physiological signals, such as skin temperature [27–31], heart rate variability [32], electroencephalogram (EEG) [33], skin conductance [34], and accelerometry [35], show a strong relationship with human thermal sensation and comfort. Sim et al. [36] developed personal thermal sensation models based on wrist skin temperature measured by wearable sensors. In addition, studies using more than one category are also not uncommon. A "personalized" model can be developed by integrating environment, occupants' physiological and behavioral data [18]. Other recent attempts [37,38] applied commercial wearable sensors together with environmental sensors (e.g., temperature, air speed) to predict the comfort of each individual occupant.

Even though all the above-reviewed studies claimed an enhanced prediction accuracy over conventional PMV and adaptive models, we identify three major drawbacks or limitations in those studies. First, subjects involved in the studies were restrained in a climate-controlled laboratory environment for a short period of time, usually in hours [36,39,40]. The dynamics of thermal comfort among daily diverse activities (e.g., dining, commuting, working, shopping) and their interactions cannot be fully captured in steady-state short-term lab tests. Even in a relatively "static" office environment, occupants would be engaged in different tasks (e.g., attending meetings, working at computers, doing office chores). As such, studies at steady-state conditions could not capture human activity, circadian cycle and mobility. The feasibility and accuracy of personal thermal comfort models developed under real-life conditions are still unclear. From our literature review, the models developed directly from lab data [32,39] usually have a higher prediction power as compared to those from the real environment [18,26], resulting in ~90% vs ~70%.

Second, most studies evaluated the performance of personal models, which predict categorical responses (e.g., cooler, warmer, no change), using accuracy that is the number of correctly predicted instances divided by the total number of instances in the dataset. Previous studies using such metric reported the prediction accuracy 79% ± 32% (Mean ± SD) for personal comfort models developed from physiological data with wearable sensors [18,19,29,32,39,41]. However, this metric is problematic because it fails to exclude correct prediction purely due to randomness [42,43]. This issue will be discussed in detail in Section 2.8.

Third, previous studies with wearable sensors often employed commercialized low-cost sensors. The sensing accuracies of those sensors when they were worn were not known, even though manufacturers reported a high accuracy and strong reliance of the embedded sensors. In most situations, the manufacture specification was based on laboratory validation in a static environment, which could be quite different when sensors were used by end-users. For instance, Empatica E4 (Empatica Inc., USA) wristband or similar [44] might be only reliable with less movements, for example, during sleeping and sitting at the table.

In the present study, to address the prior identified limitations, we developed personal thermal comfort models by machine learning using lab grade wearable sensors that continuously monitor physiological signals (skin temperature, heart rate, accelerometry) for a long period in real settings. Since a personal comfort model applies individually relevant rather than group-averaged information for thermal comfort predictions, it can be better utilized to understand specific comfort needs and desires of individual occupants and satisfy their thermal comfort accordingly. With personal comfort models, a building system can provide optimal conditions for enhanced thermal satisfaction and energy efficiency

[17]. More practically, a personal model is able to evolve by adapting new data collected in future smart buildings. We aim to evaluate the prediction power of each personal comfort model using metrics that can compensate for randomness. The importance of physiological signals and environmental parameters for prediction were also assessed in this study.

2. Methodology

Unlike population-average models, a personal comfort model should be specifically developed for an individual occupant to account for great variations in personal factors. A personal model for an occupant might not be necessarily the same for another, even if its accuracy compared to a population-average model may be higher due to its flexibility. As such, personal models are inexplicitly determined using data-driven approaches such as continuous training of machine learning algorithms over streaming data [17]. In this study, we collected and formatted physiological responses from human subjects and then applied machine learning algorithms to train personal thermal comfort models for each subject. Thermal sensation and preference from surveys were utilized as ground truth for model training and evaluation. The following sub-sections describe our approaches in detail.

2.1 Subjects

Twenty subjects (half female and half male adults) living in Berkeley and San Francisco, CA, were initially recruited through posted announcements and snowball sampling method. All the subjects were working or studying in Berkeley. The recruitment procedure excluded people who were having medical treatment (e.g., medicine taking), smokers or heavy alcohol drinkers (more than 1000 ml of beer or equivalent alcohol per day). Six subjects quit the study before finishing the minimum two-week experiment. As such, we only analyzed the data from the fourteen subjects (6 female and 8 male adults) to develop personal thermal comfort models for each person. Table 1 describes the anthropometric data of the subjects. An "average" subject (27.4 years old, 21.2 kg/m^2 BMI) voted 275 times during the entire participation.

https://doi.org/10.1016/j.buildenv.2019.106281
https://escholarship.org/uc/item/3fb0p5gk

Table 1. Anthropometrics of subjects in this study

ID	Sex	Age	Height (m)	Weight (kg)	BMI* (kg/m^2)	Cold extremity experience§	Sensitivity to thermal environment†	Hours for working out per week	Cups (240 ml) of coffee intake per day	Months of living at Berkeley or San Francisco	Total number of survey taking	Participation period
1	Male	26	1.71	68	23.3	3.4	3.7	5	0.2	12	152	11/28 - 12/12/2016
2	Male	25	1.85	86	25.1	2	2.9	5	0.5	1	253	4/2 - 4/23/2017
3	Male	31	1.7	55	19	1.6	3.5	4	0	>12	323	5/1 - 5/19/2017
4	Female	38	1.63	54	20.3	0.4	2	14.6	1.1	3	261	5/23 - 6/6/2017
5	Male	24	1.73	52	17.4	3	3.5	7	0	12	271	10/17 - 11/10/2016
6	Female	28	1.73	86	28.7	1	3	4	1.3	>12	242	12/5 - 12/20/2016
7	Female	25	1.8	57	17.6	1.1	3.1	7.5	0.6	>12	393	4/5 - 4/23/2017
8	Male	23	1.75	57	18.6	3	4	6.2	0.1	7	353	4/30 - 5/17/2017
9	Male	21	1.81	73	22.3	0	3	12	1.5	2	261	5/19 - 6/8/2017
10	Female	48	1.63	57	21.5	2	3.7	12	2	>12	256	3/21 - 4/17/2017
11	Female	20	1.65	52	19.1	1.5	2.5	7	1	8	399	5/14 - 6/28/2017
12	Male	21	1.75	61	19.9	0	3	1.8	0	4	164	12/2 - 12/19/2016
13	Male	32	1.8	70	21.6	0	3	4	0	1	198	4/23 - 5/8/2017
14	Female	22	1.58	56	22.4	1	3	5	0	3	322	5/13 - 6/1/2017
Average		27.4	1.7	63.1	21.2	1.4	3.1	6.8	0.6	-	275	-
Standard deviation		7.8	0.1	11.7	3.1	1.1	0.5	3.7	0.7	-	76	-

*BMI: Body mass index = Weight/Height^2 (kg/m^2)

§ Cold extremity experience: continuous scale 0 (Never) to 5 (Always); *Question: Have you suffered from cold hands or feet during the past two months?*

† Sensitivity to the thermal environment: continuous scale 0 (Much lower sensitivity) to 5 (to Much higher sensitivity); *Question: please indicate how sensitive you think you are to thermal conditions.*

https://doi.org/10.1016/j.buildenv.2019.106281
https://escholarship.org/uc/item/3fb0p5gk

2.2 Surveys

We asked each subject to take an online survey at least once every hour during the day. They were required to take the survey at least 12 times per day to capture the dynamics of thermal conditions, especially when their thermal sensations changed, such as after working out or moving to a different thermal environment. In addition, we provided extra incentives when a subject took the survey more than 12 times a day. The Sub-section 3.4 describes compensations in detail.

Developed with Qualtrics (Qualtrics, LLC), the survey included three "right-now" questions: (1) *location (Indoor or Outdoor),* (2) *thermal sensation (continuous ASHRAE scale from -3 cold to 3 hot),* and (3) *thermal preference (Cooler, No change, and Warmer).* The questions were randomly displayed on the platform that can be accessed using a laptop or cellphone.

2.3 Wearable sensors

Previous studies found that heart rate, skin temperature, and physical activities (representing metabolism) could indicate or reflect occupants' thermal conditions [18,27,28,32,34,45–47]. Hence, this study considered skin temperature at wrist and ankle, heart rate, and wrist accelerometry. Another reason is that sensors for these parameters are already mature products available on the market. Although skin temperature measurements at more body locations might benefit the prediction, we focused on the skin temperature at wrist and ankle only because temperature monitoring at these two locations is most likely available in the daily life due to the least disruptiveness.

2.3.1 Sensor selection

We selected sensors for this study based on three criteria: (1) *accuracy*, (2) *raw data access* for research support, and (3) *convenience* to wear for 24/7. The commercial wrist-bands and smartwatches appeared as the most suitable due to commercial availability and infusion of multiple sensors, as what were employed in previous studies [19,38,48,49]. However, those devices might not measure parameters with acceptable accuracies as described in Appendix A1. Therefore, we selected the iButton (DS1923, Maxim Integrated Products, U.S.) [50,51] for skin temperature, the Polar H7 strap (Polar Electro, Ltd., Finland) for heart rate [52,53] and a small-size cell-phone (POSH Mobile, Ltd., U.S.) in a wrist pocket measured accelerometer data to represent activity levels. The built-in inertial sensors in regular smartphones were found to be reliable for the measurement of human body motion [54,55]. We calibrated the iButton against a refrigerated circulating bath (PD7LR-20, Polyscience, U.S.) in our lab beforehand.

2.3.2 Sensors in the experiments

Polar H7 monitored heart rate every second during participation. Skin temperatures were tracked at a wrist and ankle by two iButtons separately with the sampling frequency of 1 min. To capture the transition among different thermal environments (e.g., walking from one room to another), we monitored air temperature in the body proximity by attaching an extra iButton to a pin-badge with the sensing side facing outside. The badge was pinned at the lower pant (slightly above the ankle) to reduce the influence of body thermal plume (Figure 1). Subjects took off pants with the sensor badge before sleep. The data of the three iButtons was stored on the device memory and downloaded afterwards. The sampling frequency of accelerometry was 5 Hz or above, depending on the intensity of the movement. The cell-phone also served as a server to upload heart rate data via Bluetooth and 4G. Table 2 and Figure 1 describe the specification of the physiological sensors and wearing locations, respectively. All the measured data are publicly available on the website (https://doi.org/10.15146/R3S68S).

Table 2. The description of sensors for physiological data

Model	Uncertainty	Parameter measurement
iButton DS 1923 (Maxim Integrated Products, Inc., U.S.)	± 0.2 °C after calibration	Skin temperature and air temperature close to the body

https://doi.org/10.1016/j.buildenv.2019.106281
https://escholarship.org/uc/item/3fb0p5gk

Polar H7 Bluetooth Smart Heart Rate Sensor (Polar Electro, Ltd., Finland)	Concordance correlation coefficient, 0.99 [56]	Heart rate
Cell phone POSH built app Micro X S240 (POSH Mobile, Ltd., U.S.)	Not available	Accelerometer data to represent metabolic rates. Server to receive heart rate data

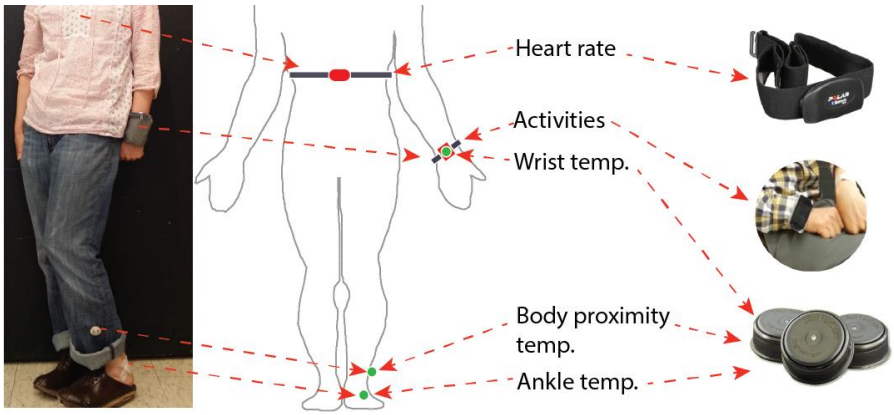


Figure 1. Physiological sensors and wearing locations

2.4 Procedure

To ensure the sensor reliability and comfortable wearing for almost 24/7, two of the authors participated in a preliminary study for approximately 4 weeks to refine the design of the experimental protocol and to ensure that the selected sensors met the criteria in the timeframe of participation. The institutional review board of University of California, Berkeley approved the experimental protocol (CPHS# 2016-09-9129). We trained all the subjects prior to their participation to ensure they wore all the sensors properly during the experiment. The duration of their participation was 14 days or longer. For each day, a subject needed to wear the sensors for at least 20 h and take the right-now survey for at least 12 times which were the minimum criteria to satisfy for compensation. The subjects had to compulsorily finish 14 days (not necessarily consecutive) with meeting the vote and wearing time requirements every day. Also, they were allowed to have a few days off during the participation and then make up the missing days within 30 days after the start of the participation.

The minimum compensation was $20 per day ($280 in total) contingent to the completion of 14-day participation. Some subjects participated up to 4 weeks in order to investigate if a long-term tracking improved the performance of personal comfort models. We encouraged the subjects to take the "right-now" survey as many times as possible, especially when they felt a change of thermal perception or preference. We provided $0.5 for each extra survey taking but the maximum payment was limited to $25 per day. Subjects received a text reminder with a survey link every hour.

2.5 Machine learning algorithms

We applied various machine learning algorithms to develop personal thermal comfort models over the collected dataset. The predicted response of the models was thermal preference ("Cooler", "No change" or "Warmer") because it is the most relevant parameter addressing thermal discomfort by specifying which action a heating, ventilation, and air conditioning (HVAC) system should take. The dataset consisted of numerical variables mainly measured by wearable sensors and subjective votes that were numerical (e.g., thermal sensation) or categorical (e.g., thermal preference).

The data cleaning and machine learning were conducted with the package of "caret" under R (version 1.1.383) [57]. The package considers over 200 algorithms. In this study, we applied four groups of

https://doi.org/10.1016/j.buildenv.2019.106281
https://escholarship.org/uc/item/3fb0p5gk

machine learning algorithms (1) linear methods, (2) non-linear methods, (3) trees and rules, and (4) ensembles of trees with each including several commonly used classification algorithms, with a total number of 14:

- Linear methods: *Linear Discriminant Analysis (abbreviation as "lda")* and *Logistic Regression ("regLogistic")*
- Non-Linear methods: *Neural Network ("nnet"), Support Vector Machine ("svmRadial"), K-Nearest Neighbors* ("*knn")* and *Naive Bayes("nb")*
- Trees and Rules: *Classification and Regression Trees ("rpart"), J48 Decision Tree ("J48"),* and *Rule-Based Classifier ("PART")*
- Ensembles of Trees: *C5.0 ("C5.0"), Bagged Classificaion and Regression Trees ("treebag"), Random Forest ("rf"), Random Forest by Randomization ("extraTrees")* and *Stochastic Gradient Boosting ("gbm")*

This algorithm selection ensured that prediction biases can be well balanced, preventing over- or under-prediction resulting from specific algorithms. Each algorithm can be applied to train a personal thermal comfort model based on the data-driven method, leading to 196 personal models in total. Some algorithms have been successfully applied previously to infer thermal comfort using environmental and/or physiological data, such as *Classification and Regression Trees* [19], *Bayesian network* [20,58], *Logistic regression* [23,26], *J48 decision tree* [59,60], and *Random forest* [26,38,61], and *SVM* [29]. In addition, the missing data in the total dataset were imputed using the *K-nearest neighbors ("knn")* algorithm.

2.6 Feature selection

The features for model training were extracted from raw data and consisted of skin temperatures (wrist and ankle), heart rate, body-proximity temperature and weather conditions (*wind, solar radiation, temperature, and humidity*). We downloaded weather data from the station *(https://www.wunderground.com/)* near the mostly stayed location of each subject during the participation. People spend most of the time indoors, so one may think that these parameters are relevant only when people are outdoors. We argue that weather conditions may affect people's clothing conditions, thermal expectation and, to a certain degree, the way how buildings are conditioned. These intermediate factors, which were not measured directly, may also cause variation of thermal perception and comfort. Nevertheless, we expected that wind and solar would have a low influence because of stronger spatial variations. Table 3 summaries the selected features in this study. The total number of features is 22.

For skin temperature and heart rate, we considered the average and gradient over the timeframes of 5 min and 60 min prior to a vote. The gradient was the slope of local linear regression (*time vs variable*) applied to the data within a timeframe window. A negative gradient of skin temperatures of the extremities possibly indicated a cool thermal sensation [27]. Likewise, an increased (positive gradient) heart rate and body acceleration might be associated with enhanced metabolism or energy expenditure [62,63]. The standard deviation of acceleration suggested the intensity of a physical activity (e.g., walking).

The selection of time frame window was based on two assumptions. First, in most real-life situations, occupants' thermal conditions change little within 5 min. Second, physiological signals 60 min ago or earlier have little reflection on the present thermal conditions. The Pearson correlation coefficients between averaged heart rate over *5 min vs 15 min, 5 min vs 30 min, 5 min vs 60 min, 15 min vs 30 min, 15 min vs 60 min,* and *30 min vs 60 min* are *0.95, 0.92, 0.91, 0.96, 0.94,* and *0.96,* respectively. The high correlations imply that finer or more timeframe windows might not be useful to improve prediction accuracy. Similar strong correlations can be also found for skin temperatures of both wrist and ankle: *0.94 (5 min vs 15 min), 0.84 (5 min vs 30 min), 0.67 (5 min vs 60 min), 0.95 (15 min vs 30 min), and*

https://doi.org/10.1016/j.buildenv.2019.106281
https://escholarship.org/uc/item/3fb0p5gk

0.92 (30 min vs 60 min). In addition, the timeframe windows for the average and gradient of meteorological parameters are *1 h* and *8 h*.

Table 3. Selected features for the development of personal thermal comfort models

Parameters	Features
Skin temperature at the ankle and wrist, body proximity temperature, and heart rate (*Numerical*)	Average over 5 min before a vote
	Gradient (slope of a linear regression) over 5 min before a vote
	Average over 60 min before a vote
	Gradient over 60 min before a vote
Wrist accelerometry (*Numerical*)	Standard deviation over 5 min before a vote
	Standard deviation over 60 min before a vote
Outdoor temperature, humidity, wind speed, and solar radiation retrieved from a weather station nearby the Berkeley campus (*Numerical*)	Average over 1 h before a vote
	Average over 8 h before a vote

2.7 Evaluation metrics

The performance of all the developed personal thermal comfort models from the machine learning algorithms was evaluated by three commonly used metrics: Cohen's kappa [64–66], accuracy [18,19,23,67], and Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) [26]. The accuracy (percentage of correct thermal preference prediction) is easy to understand, but it does not compensate for the successes that are due to mere chance [68]. Using the metric of accuracy could be an issue for an imbalanced dataset. For example, a naive model that always predicts "No change" will have a 90% accuracy if 90% of the votes in the dataset are "No change". The classification accuracy assumes equal misclassification costs and class distribution is known for the target environment [43]. However, these assumptions are probably not valid in real thermal comfort conditions. By contrast, Cohen's random successes. Cohen's kappa is defined as

$$k = \frac{P_0 - P_c}{1 - P_c}$$

Where P_0 is the accuracy and P_c is the probability which is due to change. In practice, Cohen's kappa ranges from 0 to 1 with zero being random prediction and 1 perfect prediction. AUC is the area under the ROC curve, and it varies between 0 and 1 and the larger the better. AUC of 0.5 is equal to random. In this study, we mainly applied Cohen's kappa to assess the performance of the personal thermal comfort models. Accuracy and AUC were also reported to facilitate comparison with previous studies.

K-fold cross-validation (CV) was employed and repeated multiple times to estimate average predictive performance [69]. For the entire dataset, the prediction was assessed with 5-fold CV repeated 20 times. In other words, a personal model was trained with 80% of the original data and cross-validated using the rest 20%. Nevertheless, lower folds (2-fold) CV repeated 150 times was employed to analyze the influence of data size on the prediction power because of possibly small data sizes. This approach follows the one described in [26].

3. Results

3.1 Dataset overview

We collected an extensive amount of high-quality data (*n = 3848*) comprised of physiological signals, meteorological information, and subjective votes on thermal comfort. Table 4 summarizes the statistics of the dataset of all the subjects during the participation. The outdoor temperature was 15.2 (12.7,

https://doi.org/10.1016/j.buildenv.2019.106281
https://escholarship.org/uc/item/3fb0p5gk

18.3) °C [median (Q1, Q3)] during the day (0600 - 1800), while 13.0 (11, 15.2) °C during the night time (1800 - 0600). Wrist skin temperature, mean (M) = 30.9 °C and standard deviation (SD) = 4.2 °C, was significantly (t = 72.7, p < 0.000) higher than but practically equal to ankle skin temperature (M = 30.3 °C, SD = 4.1 °C). Nevertheless, the two skin temperatures are only slightly correlated (Pearson ρ = 0.15). Figure 2 displays the relationship between wrist skin temperature and ankle skin temperature averaged over 5 min prior to a vote for 14 subjects. The majority of the data-points are in the range from 25 to 35 ˚C.

Additionally, the votes for "Cooler", "No change", and "Warmer" accounted for 15.0%, 68.5%, and 16.5% respectively, suggesting that people want a different thermal environment one-third of their time. Approximately 14% of the votes were conducted outdoors. The average voted right-now thermal sensation was 0.04 (SD = 0.7) indoors and 0.16 (SD = 1.0) outdoors. The results show that subjects' physiological data and self-reported thermal comfort varied during the whole day. Figure 3 presents the simultaneous skin temperature, heart rate, activity, location, thermal sensation and preference of a subject (ID = 11) for one day.

https://doi.org/10.1016/j.buildenv.2019.106281
https://escholarship.org/uc/item/3fb0p5gk

Table 4. Summary of monitored physiological, meteorological, and survey data for each subject during participation

	ID	1	2	3	4	5	6	7
Physiological signals	Skin T. at ankle (°C)	33.7 (32.3, 34.6)*	29.8 (28.0, 33.5)	30.3 (27.3, 33.0)	30.8 (28.9, 34.0)	29.6 (27.5, 33.8)	31.3 (29.0, 33.5)	32.2 (30.0, 34.3)
	Air T. at pant (°C)	19.7 (17.1, 22.8)	19.3 (17.5, 21.0)	21.2 (19.3, 22.9)	21.4 (19.9, 23.0)	21.4 (20.2, 22.8)	19.3 (17.1, 22.3)	17.9 (16.3, 21.4)
	Skin T. at wrist (°C)	32.6 (30.4, 33.7)	32.7 (30.5, 34.0)	33.5 (31.6, 34.6)	31.4 (29.7, 34.2)	31.8 (29.9, 33.5)	31.7 (30.2, 32.9)	31.9 (29.5, 34.6)
	Heart rate (bpm)	76 (68, 84)	61 (55, 70)	74 (63, 85)	68 (59, 77)	74 (61, 77)	89 (81, 101)	70 (58, 89)
Weather during the participation	Temperature (°C) - Day (6 AM- 6 PM)	11.2 (11.1, 13.6)	13.9 (11.7, 15.7)	15.2 (12.6, 18.2)	14.3 (12.1, 16.8)	16.6 (15.22, 19)	10.7 (8.8, 13.2)	13.7 (11.7, 15.5)
	Humidity (%) - Day (6 AM- 6 PM)	78 (55, 90)	71 (54.5, 84)	65 (52, 80)	81 (74, 89)	76 (66, 86)	79 (57, 93)	73 (59, 85)
	wind speed (m/s) - Day (6 AM- 6 PM)	0.9 (0.9, 1.8)	1.3 (0.9, 2.2)	1.3 (0.9, 1.8)	1.3 (0.9, 1.8)	0.9 (0.4, 1.8)	0.9 (0.4, 1.8)	1.3 (0.9, 2.2)
	Maximum solar (W/m^2) - Day (6 AM- 6 PM)	57 (69, 130)	314 (132, 594)	671 (331, 870)	526 (192, 826)	99 (42, 214)	49 (11.3, 119.8)	309 (132, 589)
	Temperature (°C) - Night (6 PM - 6 AM)	9.7 (9.8, 11.8)	11.8 (10.3, 13.7)	11.3 (10.3, 15.7)	11.3 (10.6, 12.4)	15.3 (14.4, 16.9)	8.9 (7.2, 11.8)	11.7 (9.9, 13.7)
	Humidity (%) - Night (6 PM - 6 AM)	80 (67, 92)	79 (65.25, 92)	77 (65, 89)	91 (86, 95)	81 (69, 89)	81 (68, 93)	81 (69, 93)
	wind speed (m/s) - Night (6 PM - 6 AM)	0.4 (0.4, 1.3)	0.9 (0, 1.8)	0.9 (0.4, 1.8)	0.9 (0.4, 1.8)	0.4 (0, 1.3)	0.9 (0.4, 1.8)	0.9 (0.4, 1.8)
Vote	Number of survey taking	152	253	323	261	271	242	393
	Thermal sensation	-0.3 (-0.5, 0.3)	0.1 (-0.3, 0.4)	-0.1 (-0.3, 0.4)	0.5 (-0.6, 0.8)	0.7 (-0.5, 1,2)	-0.6 (-1.1, 1)	0.4 (0.2, 0.6)
	% of votes for Cooler	11.2	28.5	18.6	5.9	9.6	28.5	21.6
	% of votes for No change	51.3	56.5	67.2	87.0	63.5	21.5	67.9
	% of votes for Warmer	37.5	15.0	14.2	7.0	26.9	50.0	10.4
	% of votes for Indoor	96.1	90.9	79.6	91.9	80.8	91.7	83.7
	% of votes for Outdoor	3.9	9.1	20.4	8.1	19.2	8.3	16.3

Note:* median (Q1, Q3)

https://doi.org/10.1016/j.buildenv.2019.106281
https://escholarship.org/uc/item/3fb0p5gk

Table 4. Summary of monitored physiological, weather, and survey data for each subject during participation (Cont.)

	ID	8	9	10	11	12	13	14
Physiological signals	**Skin T. at ankle (°C)**	31.7 (29.4, 33.2)*	30 (27.7, 33.4)	31.1 (29.1, 33.6)	30.8 (29.2, 33.0)	29.5 (25.2, 33.9)	31.9 (30.2, 34.0)	31.1 (22.4, 33.3)
	Air T. at pant (°C)	22.2 (19.5, 24.8)	23.2 (21.7, 24.8)	20.7 (19.0, 24.0)	22.5 (20.6, 24.3)	17.1 (15.6, 18.5)	22.3 (21.2, 23.5)	20.4 (19.4, 22.1)
	Skin T. at wrist (°C)	31.2 (29.0, 33.2)	30.2 (27.5, 32.2)	33.3 (31.9, 34.8)	32 (30.3, 33.5)	32.5 (30.8, 33.6)	32.3 (31.1, 33.3)	32.2 (21.4, 33.6)
	Heart rate (bpm)	81 (67, 92)	69 (55, 82)	70 (61, 83)	85 (64, 98)	91 (77, 103)	66 (56, 76)	65 (59, 78)
Weather during the participation	**Temperature (°C) - Day (6 AM- 6 PM)**	15.1 (12.6, 18.2)	15.3 (12.9, 18.5)	12.8 (11, 14.9)	16.3 (13.3, 19.3)	11 (9.2, 13.3)	16.4 (13.2, 21.6)	14.6 (12, 17.1)
	Humidity (%) - Day (6 AM- 6 PM)	65 (51, 80)	79 (65, 86)	72.5 (61, 87)	76 (62, 85)	76 (55, 91)	62 (45, 80)	77 (62.5, 87)
	wind speed (m/s) - Day (6 AM- 6 PM)	1.3 (0.9, 2.2)	1.3 (0.9, 1.8)	1.8 (0.9, 2.2)	0.9 (0.9, 1.8)	1.3 (0.4, 1.8)	1.3 (0.9, 1.8)	1.3 (0.9, 1.8)
	Maximum solar (W/m^2) - Day (6 AM- 6 PM)	679 (347, 876)	559.5 (205.3, 836)	254.5 (95.3, 469)	630 (240, 856)	50 (12, 125)	634.5 (270, 863.3)	585 (201.5, 843)
	Temperature (°C) - Night (6 PM - 6 AM)	11.2 (10.2, 15.2)	11.8 (10.7, 13.8)	11.3 (9.8, 13.1)	12.7 (11.1, 15.2)	9.6 (7.4, 12.6)	12.8 (10.8, 16.7)	11.2 (10.3, 13.1)
	Humidity (%) - Night (6 PM - 6 AM)	80 (68, 89)	90 (81, 95)	82 (70, 93)	88 (76, 93)	79 (67, 92)	73 (60, 86)	89 (75, 94)
	wind speed (m/s) - Night (6 PM - 6 AM)	0.9 (0.4, 1.8)	0.9 (0.4, 1.3)	0.9 (0.4, 1.8)	0.9 (0.4, 1.3)	0.4 (0.4, 1.8)	0.4 (0, 1.3)	0.9 (0.4, 1.8)
Vote	**Number of survey taking**	353	261	256	399	164	198	322
	Thermal sensation	0 (0, 0)	-0.3 (-0.6, 0.4)	0.25 (-0.4, 0.8)	0.2 (-0.2, 0.3)	-0.4 (-0.9, 0)	-0.3 (-0.5, 0.3)	-0.1 (-0.4, 0.4)
	% of votes for Cooler	7.1	8.4	23.0	12.0	3.0	31.3	3.7
	% of votes for No change	87.0	80.1	62.1	75.9	62.2	67.7	82.3
	% of votes for Warmer	5.9	11.5	14.8	12.0	34.8	1.0	14.0
	% of votes for Indoor	81.9	92.3	83.2	77.7	95.1	95.5	81.4
	% of votes for Outdoor	18.1	7.7	16.8	22.3	4.9	4.5	18.6

Note:* median (Q1, Q3)

https://doi.org/10.1016/j.buildenv.2019.106281
https://escholarship.org/uc/item/3fb0p5gk

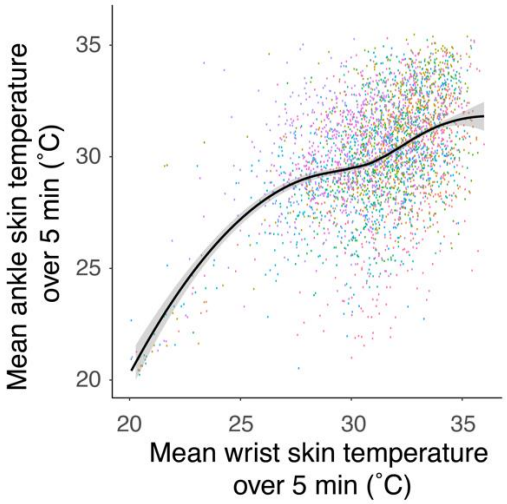


Figure 2. The relationship between wrist skin temperature and ankle skin temperature (averaged over 5 min prior to a vote) for 14 subjects with different colors; The Pearson correlation between the two temperatures is ρ = 0.15. The solid curve is local polynomial regression (LOESS) fit with 95% confidence interval (shaded area)

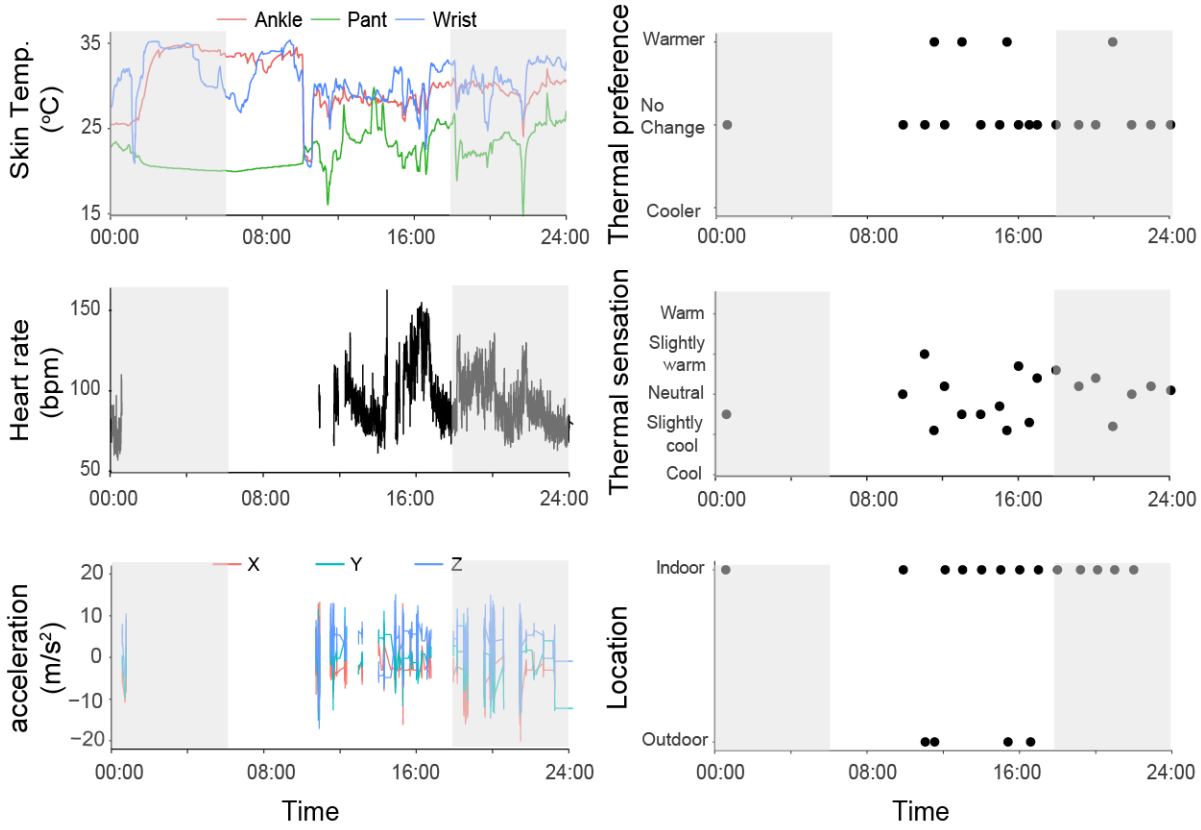


Figure 3. Monitored physiological signals and "right now" votes at a typical day for a subject (ID = 11)

https://doi.org/10.1016/j.buildenv.2019.106281
https://escholarship.org/uc/item/3fb0p5gk

3.2 Thermal sensation and preference

The plots in Figure 4 show the overall thermal sensation and preference (sorted by the percentage of "No change") for each subject. Consistent to a larger scale meta-data (the Comfort Database [70]), the median thermal sensations for all participants are within the thermal neutrality (-0.5< TS < 0.5) except for the subject (ID = 5, TS_{median} = 0.7) and the subject (ID = 6, TS_{median} = -0.6). However, the range of thermal sensations varies extensively among them. For instance, the participant (ID = 6) has a thermal sensation of -0.6 (-1.1, 1) [median (Q1, Q3)], compared to the participant (ID = 4) of 0.5 (-0.6, 0.8). We assessed the relationship between the variance of the thermal sensation deviation from the center of neutral sensation (TS= *0*), in the form of $P_{NC} = (\sum(TS - 0)^2)^{0.5}/N$ (N is the number of votes), and the percentage of "No change" preference. The calculated correlation is ρ = -0.53 (p = 0.05) implying that the subjects who had experienced smaller varieties of thermal sensation deviation from thermal neutrality tended to vote for "No change" more frequently. Moreover, the self-reported sensitivity to thermal environment collected from the background survey (Table 1) was not strongly correlated (ρ = -0.03, p = 0.91) to the variance of thermal sensation deviation from the center of neutrality.

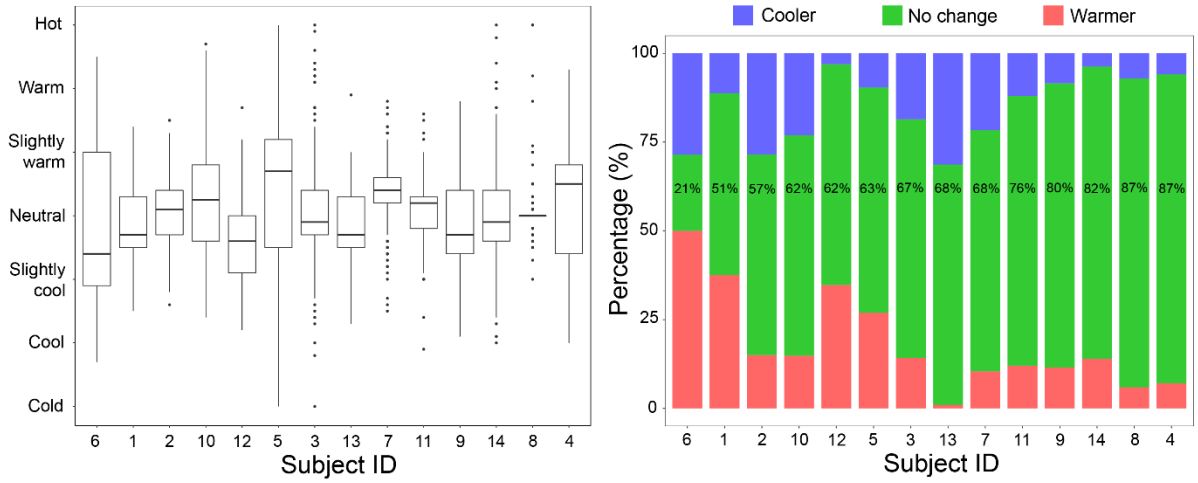


Figure 4. Thermal sensation and preference of each subject. Subject ID is sorted based on the increase of the percentage of "No change" thermal preference.

Thermal sensation is correlated with the skin temperature at wrist or ankle. Local polynomial regressions (LOESS) in Figure 5 denotes that thermal sensation does not linearly change with skin temperature at extremities, especially for wrist skin temperature. The data in Figure 5 is a subset with heart rate lower than 90 bpm (representing low activity levels) and when participants were indoors. Ankle skin temperature is more correlated with thermal sensation compared with wrist skin temperature. The results suggest that ankle skin temperature could be more predictive for thermal sensation for office workers or students. It is worth noting that the regression curves are developed based on the data of all subjects. Thus, the relationships using aggregate average data may not be generalized for an individual subject.

https://doi.org/10.1016/j.buildenv.2019.106281
https://escholarship.org/uc/item/3fb0p5gk

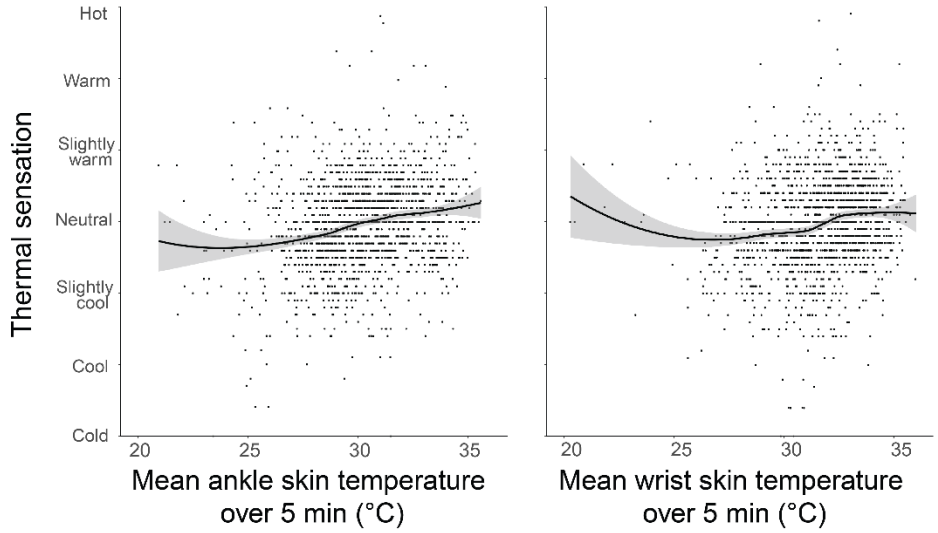


Figure 5. The relationship between skin temperatures and thermal sensation. Skin temperature is averaged over 5 min prior to a vote. The solid curves are local polynomial regression (LOESS) fit with 95% confidence interval (shaded areas)

3.3 Prediction power of personal thermal comfort models

Table 5 summarizes prediction performance with the metrics of Cohen's kappa/accuracy/AUC using 14 machine-learning algorithms for all participants. For all these subjects, the median prediction Cohen's kappa of personal comfort models is 20% with coincidence accuracy of 68% and AUC of 0.69. When only the best performing algorithm for each subject is considered, the median (based on kappa) prediction power is 24%/78%/0.79 (Cohen's kappa/accuracy/AUC). Kim et al. [26] reported the median prediction AUC of personal models, 0.73, by analyzing the heating and cooling behavior of 34 out of 38 occupants in an office building.

The results show that prediction power fluctuates among subjects and algorithms. The personal thermal comfort model of Subject 2 shows the highest median prediction power (44%/69%/0.73). By contrast, the model of Subject 4 displays the weakest performance (6%/85%/0.62), almost random "guessing" in terms of the low Cohen's kappa. Worthy to notice here that the accuracy would be misleading (Subject 4 has a higher accuracy than Subject 2) because of the imbalanced dataset. Subject 4's data were probably problematic, because we found that the subject always responded with "No change" for three consecutive days. In addition, preferring "Warmer" while feeling warm (TS = 1.4), and "Cooler" while cold (TS = -1.5), existed in the survey answers. As thermal comfort is subjective, we cannot conclude that those data are faulty. However, we can hypothesize that the subject did not answer carefully, and this could be a reason for the low prediction power. Another possibility is that some people might be less predictable than others.

https://doi.org/10.1016/j.buildenv.2019.106281
https://escholarship.org/uc/item/3fb0p5gk

Table 5. Prediction power (Cohen's kappa/accuracy/AUC) for each participant with 14 common algorithms

SubID /Data size	**1/152**	**2/253**	**3/323**	**4/261**	**5/271**	**6/242**	**7/393**
Lda	21%/56%/0.66	24%/61%/0.7	45%/74%/0.79	2%/83%/0.53	37%/69%/0.77	11%/48%/0.58	20%/68%/0.69
regLogistic	17%/56%/0.68	22%/62%/0.73	40%/75%/0.82	0%/87%/0.57	30%/69%/0.79	15%/53%/0.64	15%/70%/0.7
nnet	21%/56%/0.68	26%/62%/0.73	50%/77%/0.86	7%/78%/0.62	38%/70%/0.79	15%/51%/0.62	18%/68%/0.7
svmRadial	15%/54%/0.58	19%/60%/0.72	44%/76%/0.84	0%/87%/0.64	32%/70%/0.76	17%/55%/0.61	13%/69%/0.67
knn	11%/52%/0.61	24%/59%/0.71	43%/76%/0.83	1%/86%/0.65	29%/69%/0.76	15%/51%/0.59	17%/68%/0.62
nb	13%/49%/0.62	22%/55%/0.7	47%/72%/0.81	5%/75%/0.6	32%/63%/0.74	14%/45%/0.6	21%/65%/0.65
rpart	7%/49%/0.55	45%/71%/0.65	31%/71%/0.74	3%/85%/0.55	30%/68%/0.66	10%/50%/0.56	17%/66%/0.58
J48	7%/46%/0.55	52%/72%/0.61	40%/71%/0.7	9%/82%/0.53	31%/68%/0.66	10%/48%/0.55	13%/69%/0.54
PART	9%/47%/0.56	43%/68%/0.63	39%/71%/0.72	8%/82%/0.54	29%/64%/0.65	12%/46%/0.56	13%/61%/0.59
C5.0	10%/50%/0.61	65%/80%/0.73	47%/75%/0.85	6%/86%/0.63	32%/66%/0.78	21%/53%/0.58	19%/66%/0.65
treebag	11%/51%/0.61	49%/73%/0.73	46%/75%/0.84	6%/85%/0.65	34%/68%/0.76	16%/51%/0.6	17%/68%/0.66
gbm	19%/55%/0.67	54%/75%/0.78	50%/77%/0.85	7%/85%/0.68	4%/71%/0.8	16%/52%/0.62	20%/69%/0.68
extraTrees	19%/57%/0.67	51%/74%/0.78	50%/78%/0.88	7%/86%/0.73	37%/70%/0.8	17%/53%/0.63	21%/70%/0.7
rf	17%/55%/0.64	51%/74%/0.75	48%/76%/0.86	7%/87%/0.68	34%/68%/0.8	18%/54%/0.62	18%/69%/0.68
SubID /Data size	**8/353**	**9/261**	**10/256**	**11/399**	**12/164**	**13/198**	**14/322**
lda	40/87%/0.76	18%/77%/0.68	17%/62%/0.65	34%/79%/0.74	22%/64%/0.71	33%/71%/0.66	8%/80%/0.7
regLogistic	6%/87%/0.76	2%/80%/0.7	7%/62%/0.69	22%/79%/0.76	8%/62%/0.74	41%/75%/0.83	2%/82%/0.71
nnet	34%/85%/0.78	20%/74%/0.72	21%/64%/0.68	31%/79%/0.77	16%/62%/0.74	37%/73%/0.76	11%/78%/0.72
svmRadial	25%/88%/0.82	4%/79%/0.72	5%/62%/0.64	32%/80%/0.76	0%/60%/0.69	35%/75%/0.78	1%/82%/0.73
knn	9%/87%/0.75	15%/76%/0.73	17%/60%/0.66	18%/77%/0.7	6%/59%/0.63	40%/75%/0.7	0%/82%/0.68
nb	35%/84%/0.79	20%/70%/0.69	16%/48%/0.66	32%/71%/0.76	24%/62%/0.72	41%/73%/0.78	16%/74%/0.72
rpart	19%/84%/0.59	16%/75%/0.64	12%/61%/0.58	24%/76%/0.68	11%/60%/0.55	34%/72%/0.66	6%/75%/0.55
J48	27%/84%/0.58	22%/74%/0.61	11%/61%/0.56	20%/72%/0.63	19%/62%/0.64	37%/74%/0.65	6%/77%/0.57
PART	27%/85%/0.58	21%/75%/0.62	13%/55%/0.57	21%/72%/0.65	15%/62%/0.64	35%/72%/0.62	5%/77%/0.58
C5.0	27%/88%/0.8	23%/78%/0.75	16%/59%/0.65	33%/77%/0.76	19%/63%/0.67	38%/74%/0.69	6%/77%/0.66
treebag	30%/87%/0.78	21%/79%/0.77	18%/63%/0.65	30%/78%/0.76	18%/63%/0.71	42%/75%/0.78	3%/79%/0.66
gbm	31%/88%/0.79	24%/78%/0.79	15%/61%/0.67	37%/79%/0.79	19%/63%/0.74	47%/77%/0.81	6%/79%/0.71
extraTrees	33%/88%/0.84	21%/79%/0.81	20%/64%/0.7	32%/80%/0.8	18%/63%/0.76	46%/78%/0.81	4%/80%/0.75
rf	29%/87%/0.81	23%/79%/0.8	18%/63%/0.67	33%/79%/0.78	18%/64%/0.76	45%/76%/0.8	1%/80%/0.7

https://doi.org/10.1016/j.buildenv.2019.106281
https://escholarship.org/uc/item/3fb0p5gk

An algorithm that performs well for a subject might not necessarily the best for another. Figure 6 depicts that the most frequent algorithms with high prediction power are *Stochastic Gradient Boosting ("gbm")*, *Random Forest by Randomization ("extraTrees")*, *C5.0 ("C5.0")*, and *Random Forest ("rf")*. All of them belong to the category of "Ensembles of Trees". The *"gbm"* algorithm produces the highest median prediction power, approximately 25%/73%/74% (Cohen's kappa/accuracy/AUC).

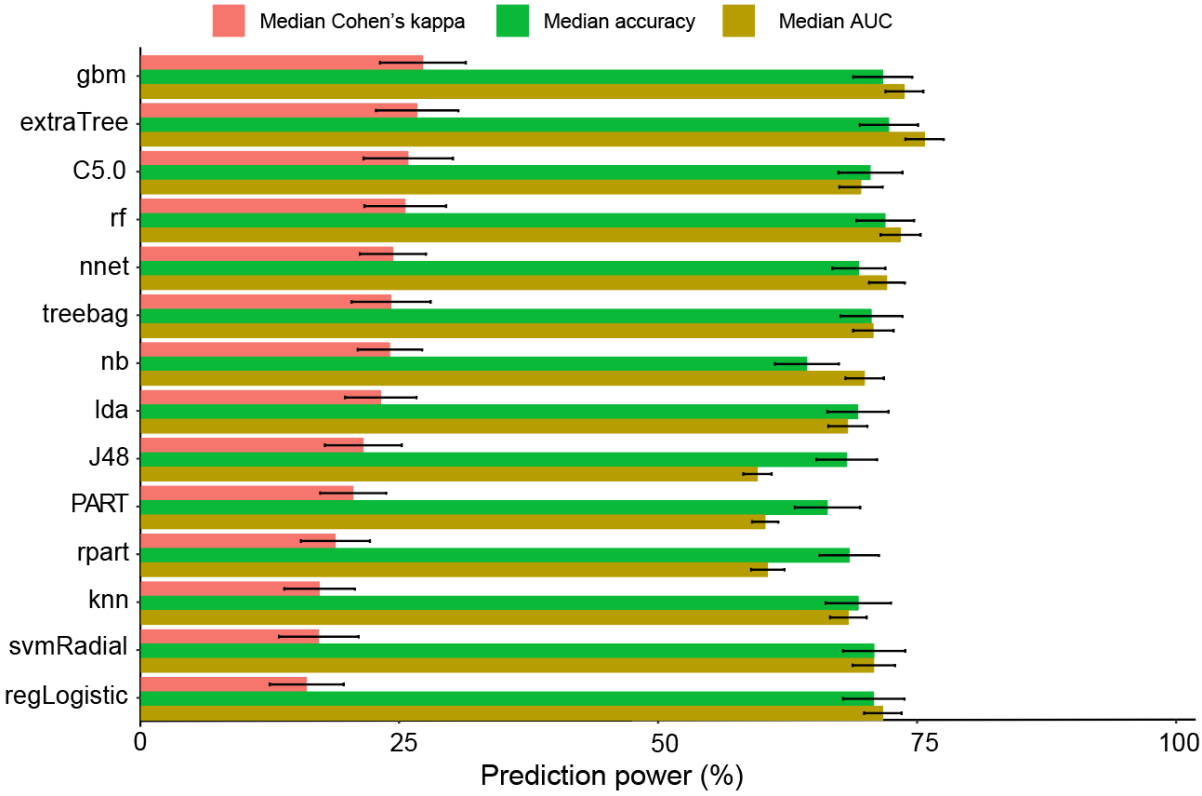


Figure 6. Prediction power of different algorithms. Bar plots and error bars represent the average and standard deviation of prediction power (Cohen's kappa/accuracy/AUC) across 14 participants respectively.

3.4 Importance of features

Many wearable sensors, such as wristbands, on the market have the capacity to measure skin temperature, heart rate, and activity, albeit the sampling accuracy is still a concern. Technically, incorporating additional functions such as extracting real-time weather data into wristbands is not difficult for thermal comfort model development. Nevertheless, understanding what features are the most contributing would help optimize the sampling efficiencies with wearable sensors.

We evaluated the predictive performance by starting with an individual parameter (e.g., wrist skin temperature) and then adding more until all features were included in terms of a step-wise approach. The added parameters were ordered based on the efforts involved in the data collection by the state-of-the-art wearable sensors on the market. For instance, sampling activity might need few efforts compared to body-proximity temperature because of the capacities of existing commercial wristbands.

Figure 7 shows the maximum prediction power (Cohen's kappa/accuracy/AUC) of the personal thermal comfort models using 14 algorithms, which represents the highest performance we could obtain by selecting the best algorithm. The models were trained on the entire dataset of all the 14 subjects instead of each subject and then averaged. In general, applying each individual feature or feature combinations

https://doi.org/10.1016/j.buildenv.2019.106281
https://escholarship.org/uc/item/3fb0p5gk

generates a prediction power greater than the conventional PMV/adaptive model that has a very low prediction power on thermal preference (AUC≈0.5) [16,26].

The baseline offers a maximum prediction power of 35% /76% /0.80. The features include time, weather (air temperature, wind speed, solar radiation, and relative humidity), body-proximity temperature and physiological signals (heart rate, wrist skin temperature, ankle skin temperature, and wrist movement acceleration). The time and weather data that can be extracted from the cloud using IoT can produce an accuracy of 18% /68% /0.66 that is only 51% of the baseline based on Cohen's kappa. In other words, the weather data only would not provide us compelling prediction performance from the limited dataset from this study.

Most wristbands on the market can measure wrist movement acceleration and heart rate. These features together with time predict thermal preference at the performance only 18% /71% /0.67, similar accuracy to that with weather and time. However, the wristbands could be easily updated with more functions, such as streaming weather data from the cloud and embedding a sensor to measure wrist skin temperature. By infusing these new functions, wristbands can produce prediction accuracy of up to 43% /77% /0.78, even slightly higher than the baseline based on Cohen's kappa. One might raise the question why the prediction power is higher with less features. We argue that the training error of unstructured and heterogeneous dataset using the data-driven approach accumulates especially when some features considered have a high level of sparsity. Still, the high accuracy of wristbands expanded with new functions might be a promising approach to achieve higher prediction power for personal thermal comfort models.

Another option is smart shoes. Sensors of feet movement acceleration, heart rate, and ankle skin temperature can be embedded into the shoes (i.e., shoe heel notch or top line). Based on the dataset of the study, the prediction accuracy of smart shoes can achieve 36% /78% /0.79, almost equivalent to the baseline performance. Note that we used wrist movement acceleration as the approximation of ankle acceleration and we used the ankle skin temperature. The temperature at the feet could carry even more information given that is an extremity. One advantage of smart shoes is that they are less intrusive than wristbands since most people wear shoes but not wristband. Considering all the factors regarding intrusiveness, possible market penetration, and prediction power, we think that smart shoes might be more suitable to implement personal thermal comfort models in the future smart personal environment control. Furthermore, Figure 7(d) shows the prediction performance of air temperature and skin temperature. The maximum power is 22% /69% /0.74, approximately 63% of the baseline's performance.

https://doi.org/10.1016/j.buildenv.2019.106281
https://escholarship.org/uc/item/3fb0p5gk

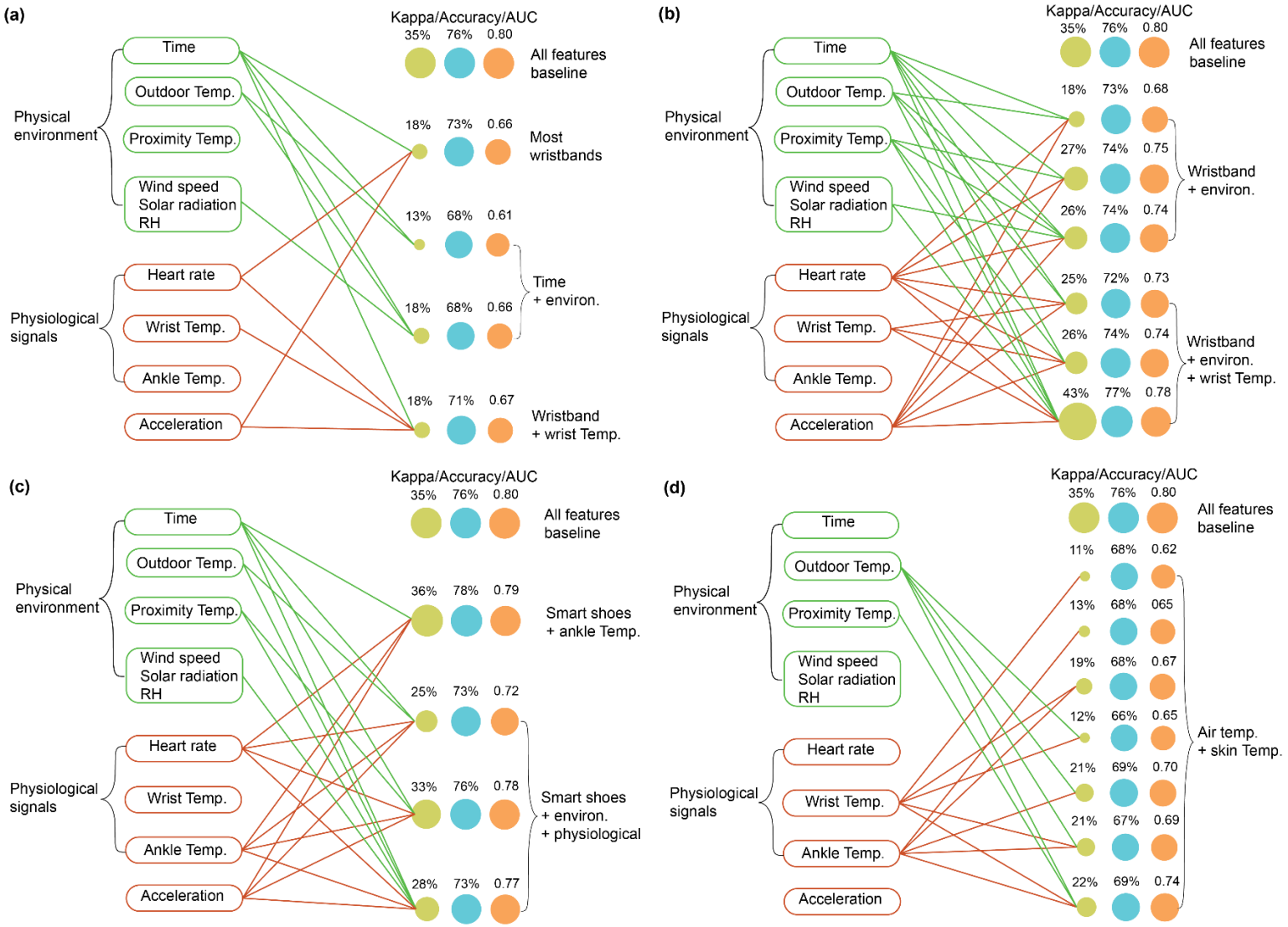


Figure 7. Maximum prediction power with variable feature combinations using 14 algorithms. The circle size denotes the prediction power relative to the baseline with all features considered. (a) and (b) prediction power of features (heart rate, time, acceleration) included in current wristbands and enhanced prediction performance when the wristbands are expanded with other functions; (c) Prediction power of smart shoes; (d) Prediction power of skin temperature and air temperature.

https://doi.org/10.1016/j.buildenv.2019.106281
https://escholarship.org/uc/item/3fb0p5gk

Figure A2 in the Appendix compares the normalized importance on a scale of 100 for each individual parameter and its derivations (e.g., gradient, standard deviation) determined by SVM classification algorithm. Please note that the order of importance might be slightly different for another algorithm. Overall, the results show that outdoor wind speed and relative humidity, as expected, are not very important. Furthermore, the standard deviation of the variables is a weak indicator of thermal preference prediction.

3.5 Comparison of prediction power for different thermal sensations and classification methods

We observed in the Comfort Database [70] that thermal preference is strongly correlated with thermal sensation only when people are outside the thermal neutrality. In specific, the Pearson correlation between thermal sensation and thermal preference was only -0.06 within the thermal neutrality (-0.5 to 0.5) as compared to -0.67 when thermal sensation was not neutral. We used a narrow definition of thermal neutrality to assess how the models work in the most challenging conditions. In practice, thermal sensation between -1.5 and 1.5 may be considered acceptable. Since physiological variables are strongly related to thermal sensation, thermal preference within thermal neutrality might be difficult to infer through physiological signals. As such, we hypothesize that wearable sensors may not be reliable for thermal preference inference within thermal neutrality. Rather, personal comfort models with wearable sensors might be most effective to bring people back to thermal neutrality when they are outside it. To test the hypothesis, we evaluated the prediction performance of personal models on the sub-datasets with thermal neutrality (-0.5 to 0.5) and non-neutrality.

Figure 8 shows that the median Cohen's kappa is only 3 % (slightly higher than random) on the dataset of thermal neutrality. The results suggest that the thermal preference prediction using wearable sensors could be very challenging when occupants' thermal sensation is neutral. The prediction accuracy is nevertheless as high as 82.4%, which is caused by the coincidently increased possibility of voting "No change." This is another example of why accuracy is not an appropriate metric to assess prediction power for an imbalanced dataset. When the thermal sensation is outside of neutrality, however, prediction power can be considerably enhanced (Cohen's kappa from 3% to 29%). The results also partially explain why prediction power is weaker for field studies than climate-controlled laboratory experiments, which dedicatedly exposed subjects to cold and/or hot environments more frequently than what was supposed to be in real life.

The PMV model has its highest prediction accuracy in the thermal neutral zone and it decays towards the extremes of thermal sensation scale [16]. We showed that personal thermal comfort models using wearable sensors do exactly the opposite. They have the highest prediction at the extremes. In addition, Figure 8 shows that the prediction power is significantly increased when only considering the subset data consisting of "Warmer" and "Cooler."

https://doi.org/10.1016/j.buildenv.2019.106281
https://escholarship.org/uc/item/3fb0p5gk

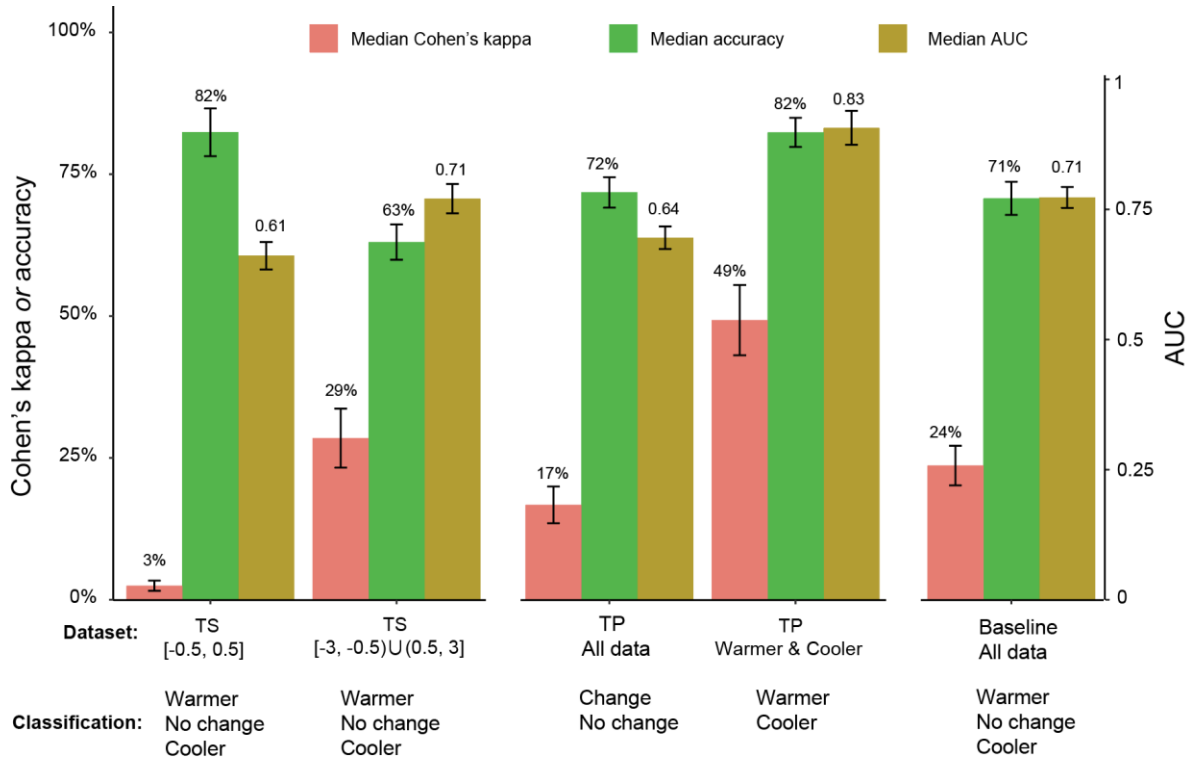


Figure 8. Prediction with different thermal sensations and classification methods

4. Discussions

Through this study, we identify many challenges when wearable sensors are applied to develop personal thermal comfort models. We will discuss them in this section as well as making suggestions for future studies.

4.1 The challenge of personal thermal comfort models within thermal neutrality

Figure 9 depicts the density plots of thermal preferences varying with thermal sensation, ankle skin temperature, wrist skin temperature, and heart rate. These plots show the overlapping regions for different thermal preferences. For the thermal sensation graph, it can be deduced that in the low prediction accuracy regions, even at the same self-reported thermal sensation, a subject likely prefers a different thermal environment. We believe that these overlapping regions near thermal neutrality are the fundamental reason for poor prediction presented in Section 4.5 using physiological signals and all other variables that could be related to thermal sensation.

According to the size of overlapping regions, ankle skin temperature appears more sensitive to thermal preference than wrist skin temperature or heart rate in intermediate ranges. The results imply that ankle skin temperature is more predictive in normal situations, while heart rate could be a strong predictor (of wanting cooler) at a high value. The preference of "No change" might be quite difficult to predict as the region is almost overlapped with "Cooler" or "Warmer" for both skin temperature and heart rate. An occupant is most likely to prefer "Warmer" when ankle skin temperature is lower than approximately 28 ˚C, wrist skin temperature lower than 31˚C, or heart rate blow 65 bpm. The cutoff value for "Cooler" preference is 31 ˚C for ankle skin temperature or wrist skin temperature, and 105 bpm for heart rate. In the middle ranges between the cutoff values (ankle skin temperature 29 - 31 ˚C or heart rate 65 -105 bpm), "No change" is preferred at the highest probability. However, the wrist skin temperature for "No change" preference might be indistinctive from those for other preferences.

https://doi.org/10.1016/j.buildenv.2019.106281
https://escholarship.org/uc/item/3fb0p5gk

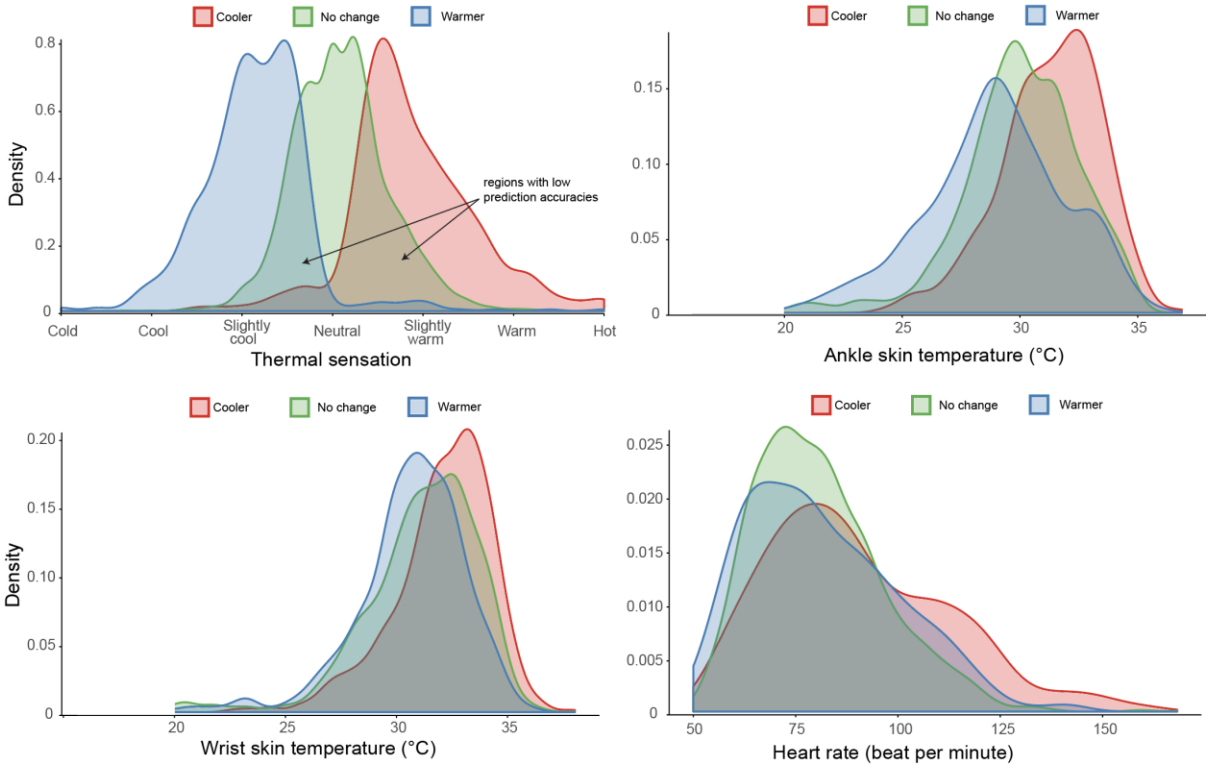


Figure 9. The distributions of thermal preference votes against thermal sensation, skin temperature and heart rate; overlapped areas of thermal preference are associated with low prediction accuracy

We further hypothesize hereof that using wearable sensors that monitor physiological data would not be an efficient and reliable way to predict thermal preference and control HVAC systems when a subject's thermal sensation is mainly neutral. Instead, it is much easier to predict thermal preference when occupants are cold or hot. Figure 10 illustrates the theoretical prediction power ("U" curve) of personal thermal comfort models varying over different thermal sensations.

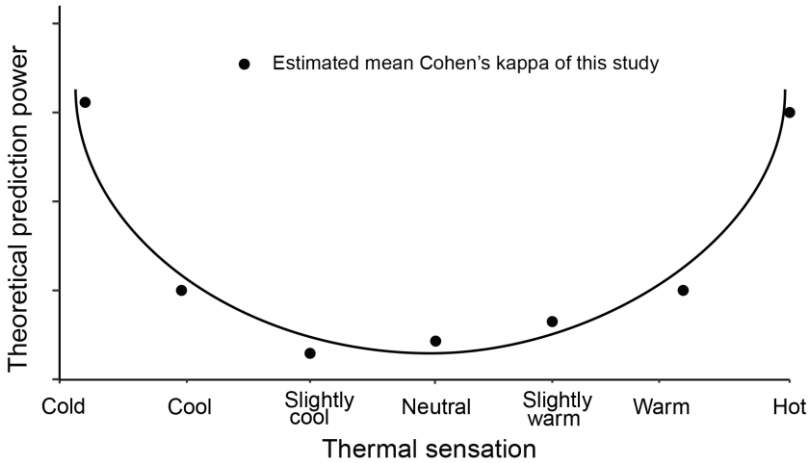


Figure 10. The diagram of the theoretical prediction performance of personal thermal comfort models using wearable sensors

https://doi.org/10.1016/j.buildenv.2019.106281
https://escholarship.org/uc/item/3fb0p5gk

4.2 Influence of data size on prediction power

Different from conventional thermal comfort models, such as PMV and adaptive model, personal thermal comfort models necessity model training and development based on a data-driven approach. Overall, a predictive model relies on a sufficient number of data points. As suggested by Kim et al. [26], dynamic machine learning models with cloud computing capacity would be updated when new data arrive. Figure 11 shows the median prediction power (Cohen's kappa/accuracy/AUC) of 14 algorithms varying with the training data size for each participant. The solid dark curves are the curve-fitted power of all subjects' models with 95% confidence intervals (shaded areas) using local polynomial regression (LOESS).

The aggregated curves show that prediction power can be improved by enlarging trained data. The aggregated prediction power shows that performance of 21% /71% /0.7 can be achieved after 200 votes. However, individual models display various sensitivities to the dataset size. For instance, Cohen's kappa curves of two subjects' models (ID = 4 and ID = 12) are almost flat over different data sizes, also possibly due to bad data. It is difficult to find a curve plateau to determine model convergence. In addition, a model might be only converged locally because new data could alter existing model patterns.

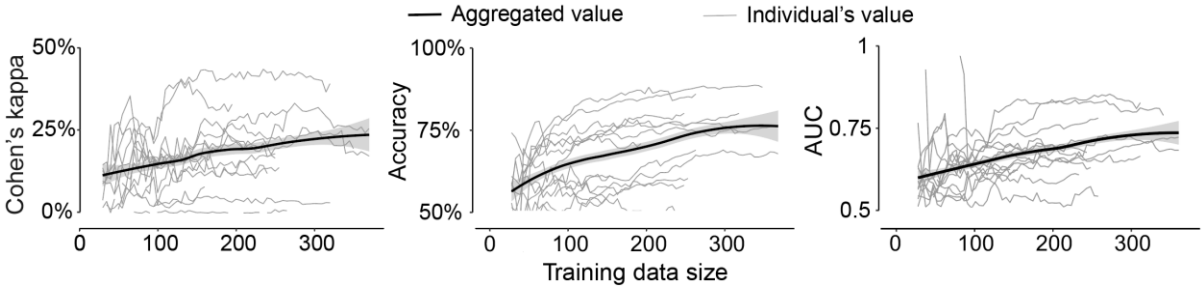


Figure 11. Median prediction accuracy varies with data size. The individual's values are the median Cohen's kappa/accuracy/AUC calculated with different personal comfort models. The solid curves are local polynomial regression (LOESS) fit with 95% confidence interval (shaded areas)

4.3 Exit survey, limitations, and future studies

Subjects took a short exit survey after they completed the participation. The average overall satisfaction (Likert scale from 1 to 7) with thermal environments of their home and offices is 5.3 ± 1.25 (Mean ± SD) and 4.8 ± 1.2, respectively. The Pearson correlation coefficient between their satisfaction and percentage of voting "No change" is 0.16 for offices and 0.14 for home. The weak correlations suggest that an occupant who is highly satisfied with the indoor environment in general (such as during post-occupancy survey) would still desire a probably mild "warmer" or "cooler" environment in daily life. The responses also consolidate that the ability to control micro-environment based on personal thermal comfort models is appreciated for occupants.

This study has a few limitations. First, the quality and quantity of the sampled data could be improved to gain highly predictive personal thermal comfort models. Although 14 subjects participated in the study for a few weeks, the average vote number of each one was only 275 because asking them to vote more frequently would interfere with their daily activities, even though incentives were provided for extra votes. In addition, the whole dataset has roughly 14% missing data resulting from the unstable sensing (e.g., loss of Internet or sensor batteries). Also, personal models can be vulnerable to "bad" data provided by randomly given votes. Continuous vote collection from occupants and model updating with steamed data would likely increase the prediction power after the model implementation into HVAC systems. Second, the accuracy of the models presented can be further improved more thorough feature engineering or hyperparameter tuning, as our feature engineering is just sufficient to justify the major goal, demonstrating the enhanced prediction power of personal models as opposed to the conventional PMV or adaptive models. Lastly, the models in this study share the common limitations (one-time batch learning and equal misclassification costs) as in previous works [26].

https://doi.org/10.1016/j.buildenv.2019.106281
https://escholarship.org/uc/item/3fb0p5gk

Future studies need to deal with the following issues. Using wearable sensors are still challenging to develop personal models due to sensor intrusiveness to occupants, data collection cost, and prediction power for thermal comfort. To our knowledge, many manufacturers do not allow users to access the raw data. Another important challenge is to detect bad data, especially subjective votes while model training, which would ruin in prediction accuracy.

5. Conclusions

Predicting thermal comfort/preference using physiological data could be potentially incorporated into HVAC system control for occupants' satisfaction and energy saving. The low-cost wearable sensors and cloud computing allow real-time thermal comfort/preference prediction using physiological and environmental data. We developed personal thermal comfort models for 14 participants using lab-grade wearable sensors. Based on physiological and meteorological data monitored for 2-4 weeks, we trained 14 personal comfort models using different machine learning algorithms for each participant. The results lead to the following conclusions:

- The developed personal thermal comfort models with long-term tracking of physiological and environmental data lead to a median prediction power of 24% /78% /79% (Cohen's kappa/accuracy/AUC) that is significantly greater than conventional PMV and adaptive models.
- The algorithm category of "Ensembles of Trees" such as Stochastic Gradient Boosting ("gbm"), Random Forest by Randomization ("extraTrees"), C5.0 ("C5.0") and Random Forest ("rf") showed the best performance to develop personal comfort models.
- The PMV model has its highest prediction accuracy in the thermal neutral zone and it decays towards the extremes of thermal sensation scale. We showed that personal thermal comfort models using wearable sensors do exactly the opposite. They have the highest prediction outside thermal neutrality. This is very useful in practice because we want to avoid people being over-cooled and over-heated.
- When the data from all subjects are merged together and all the features are included, the prediction power is 35% /76% /0.80 (baseline). Current smart wristbands with the data of time, heart rate, acceleration can generate a maximum prediction power of 18% /71% /0.67, only 51% of the baseline. However, the prediction performance can be enhanced to 43%/77%/0.78 when skin temperature sensing and weather data streaming are infused into the wristbands.
- Smart shoes might be suitable platforms to implement personal thermal comfort models. The maximum accuracy with the features of acceleration, ankle skin temperature, heart rate and weather is 36% /78% /0.79, equivalent the baseline.
- The prediction performance of personal comfort models with wearable sensors could reach 21% /71% /0.7 (Cohen's kappa/accuracy/AUC) after approximately 200 votes.
- Cohen's kappa and AUC are more appropriate metrics to evaluate the prediction performance of personal thermal comfort models because accuracy does not compensate for successes that are due to mere chance for an imbalanced dataset, especially when instance distribution is unknown beforehand. In addition, bad data should be detected and removed during model development with machine learning or other data-driven approaches.

Acknowledgment

This research is funded by the Republic of Singapore's National Research Foundation through a grant to the Berkeley Education Alliance for Research in Singapore (BEARS) for the Singapore - Berkeley Building Efficiency and Sustainability in the Tropics (SinBerBEST) Program. BEARS has been established by the University of California, Berkeley as a center for intellectual excellence in research and education in Singapore.

Competing interests

The authors declare no competing interests.

References

[1] D. Ormandy, V. Ezratty, Health and thermal comfort: From WHO guidance to housing strategies, Energy Policy. 49 (2012) 116–121. doi:10.1016/j.enpol.2011.09.003.

[2] K. Pantavou, G. Theoharatos, A. Mavrakis, M. Santamouris, Evaluating thermal comfort conditions and health responses during an extremely hot summer in Athens, Build. Environ. 46 (2011) 339–344. doi:10.1016/j.buildenv.2010.07.026.

[3] T. Akimoto, S. Tanabe, T. Yanai, M. Sasaki, Thermal comfort and productivity - Evaluation of workplace environment in a task conditioned office, Build. Environ. 45 (2010) 45–50. doi:10.1016/j.buildenv.2009.06.022.

[4] L. Lan, P. Wargocki, Z. Lian, Quantitative measurement of productivity loss due to thermal discomfort, Energy Build. 43 (2011) 1057–1062. doi:10.1016/j.enbuild.2010.09.001.

[5] O.A. Seppänen, W. Fisk, Some quantitative relations between indoor environmental quality and work performance or health, HVACR Res. 12 (2006) 957–973. doi:10.1080/10789669.2006.10391446.

[6] P. Wargocki, O. Seppänen, J. Andersson, D. Clements-Croome, K. Fitzner, S. Hanssen, Indoor climate and productivity in offices, REHVA, Federation of European Heating and Air-conditioning Associations, Brussels, Belgium, European Committee for Standardization, 2006.

[7] M.J. Mendell, G.A. Heath, Do indoor pollutants and thermal conditions in schools influence student performance? A critical review of the literature, Indoor Air. 15 (2005) 27–52. doi:10.1111/j.1600-0668.2004.00320.x.

[8] L. Lan, Z. Lian, L. Pan, The effects of air temperature on office workers' well-being, workload and productivity-evaluated with subjective ratings, Appl. Ergon. 42 (2010) 29–36. doi:10.1016/j.apergo.2010.04.003.

[9] H. Chappells, Comfort, well-being and the socio-technical dynamics of everyday life, Intell. Build. Int. 2 (2010) 286–298. doi:10.3763/inbi.2010.0003.

[10] ANSI/ASHRAE Standard 55, Thermal environmental conditions for human occupancy, Am. Soc. Heat. Refrig. Air-Cond. Eng. Atlanta GA. (2017).

[11] EN 15251:2007, Criteria for the indoor environment including thermal, indoor air quality, light and noise, Brussels, Belgium, European Committee for Standardization, 2007.

[12] ISO 7730: Ergonomics of the thermal environment - Analytical determination and interpretation of thermal comfort using calculation of the PMV and PPD indices and local thermal comfort criteria., Geneva, Switzerland, 2005.

[13] R. De Dear, G.S. Brager, Developing an adaptive model of thermal comfort and preference, ASHRAE Transactions. 104.(1998).

[14] K. Natsume, T. Ogawa, J. Sugenoya, N. Ohnishi, K. Imai, Preferred ambient temperature for old and young men in summer and winter, Int. J. Biometeorol. 36 (1992) 1–4.

[15] G. Havenith, I. Holmér, K. Parsons, Personal factors in thermal comfort assessment: clothing properties and metabolic heat production, Energy Build. 34 (2002) 581–591. doi:10.1016/S0378-7788(02)00008-7.

[16] 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. doi:10.1016/j.buildenv.2019.01.055.

[17] J. Kim, S. Schiavon, G. Brager, Personal comfort models – A new paradigm in thermal comfort for occupant-centric environmental control, Build. Environ. (2018). doi:10.1016/j.buildenv.2018.01.023.

[18] D. Li, C.C. Menassa, V.R. Kamat, Personalized human comfort in indoor building environments under diverse conditioning modes, Build. Environ. 126 (2017) 304–317. doi:10.1016/j.buildenv.2017.10.004.
[19] F. Salamone, L. Belussi, C. Currò, L. Danza, M. Ghellere, G. Guazzi, B. Lenzi, V. Megale, I. Meroni, Integrated method for personal thermal comfort assessment and optimization through users' feedback, IoT and machine learning: a case study, Sensors. 18(5) (2018) 1602. doi:10.3390/s18051602.
[20] F. Auffenberg, S. Stein, A. Rogers, A Personalised Thermal Comfort Model Using a Bayesian Network, in: Proc. 24th Int. Conf. Artif. Intell., AAAI Press, Buenos Aires, Argentina, 2015: pp. 2547–2553. http://dl.acm.org/citation.cfm?id=2832581.2832605 (accessed October 11, 2018).
[21] T.C.T. Cheung, S. Schiavon, E.T. Gall, M. Jin, W.W. Nazaroff, Longitudinal assessment of thermal and perceived air quality acceptability in relation to temperature, humidity, and CO2 exposure in Singapore, Build. Environ. 115 (2017) 80–90. doi:10.1016/j.buildenv.2017.01.014.
[22] D. Daum, F. Haldi, N. Morel, A personalized measure of thermal comfort for building controls, Build. Environ. 46 (2011) 3–11. doi:10.1016/j.buildenv.2010.06.011.
[23] A. Ghahramani, C. Tang, B. Becerik-Gerber, An online learning approach for quantifying personalized thermal comfort via adaptive stochastic modeling, Build. Environ. 92 (2015) 86–96. doi:10.1016/j.buildenv.2015.04.017.
[24] B. Huchuk, W. O'Brien, S. Sanner, A longitudinal study of thermostat behaviors based on climate, seasonal, and energy price considerations using connected thermostat data, Build. Environ. 139 (2018) 199–210. doi:10.1016/j.buildenv.2018.05.003.
[25] W. Pasut, H. Zhang, E. Arens, Y. Zhai, Energy-efficient comfort with a heated/cooled chair: Results from human subject tests, Build. Environ. 84 (2015) 10–21. doi:10.1016/j.buildenv.2014.10.026.
[26] J. Kim, Y. Zhou, S. Schiavon, P. Raftery, G. Brager, Personal comfort models: Predicting individuals' thermal preference using occupant heating and cooling behavior and machine learning, Build. Environ. 129 (2018) 96–106. doi:10.1016/j.buildenv.2017.12.011.
[27] D. Wang, H. Zhang, E. Arens, C. Huizenga, Observations of upper-extremity skin temperature and corresponding overall-body thermal sensations and comfort, Build. Environ. 42 (2007) 3933–3943. doi:10.1016/j.buildenv.2006.06.035.
[28] J.-H. Choi, V. Loftness, Investigation of human body skin temperatures as a bio-signal to indicate overall thermal sensations, Build. Environ. 58 (2012) 258–269. doi:10.1016/j.buildenv.2012.07.003.
[29] 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. doi:10.1016/j.buildenv.2016.12.005.
[30] 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. doi:10.1016/j.buildenv.2018.11.017.
[31] A.C. Cosma, R. Simha, Thermal comfort modeling in transient conditions using real-time local body temperature extraction with a thermographic camera, Build. Environ. 143 (2018) 36–47. doi:10.1016/j.buildenv.2018.06.052.
[32] K.N. Nkurikiyeyezu, Y. Suzuki, Y. Tobe, G.F. Lopez, K. Itao, Heart rate variability as an indicator of thermal comfort state, in: 2017 56th Annu. Conf. Soc. Instrum. Control Eng. Jpn. SICE, 2017: pp. 1510–1512. doi:10.23919/SICE.2017.8105506.
[33] F. Zhang, S. Haddad, B. Nakisa, M.N. Rastgoo, C. Candido, D. Tjondronegoro, R. de Dear, The effects of higher temperature setpoints during summer on office workers' cognitive load and thermal comfort, Build. Environ. 123 (2017) 176–188. doi:10.1016/j.buildenv.2017.06.048.
[34] A.P. Gagge, J.A.J. Stolwijk, J.D. Hardy, Comfort and thermal sensations and associated physiological responses at various ambient temperatures, Environ. Res. 1 (1967) 1–20. doi:10.1016/0013-9351(67)90002-3.

[35] M.P. Rothney, E.V. Schaefer, M.M. Neumann, L. Choi, K.Y. Chen, Validity of physical activity intensity predictions by ActiGraph, Actical, and RT3 Accelerometers, Obesity. 16 (2008) 1946–1952. doi:10.1038/oby.2008.279.
[36] S.Y. Sim, M.J. Koh, K.M. Joo, S. Noh, S. Park, Y.H. Kim, K.S. Park, Estimation of thermal sensation based on wrist skin temperatures, sensors. 16 (2016) 420. doi:10.3390/s16040420.
[37] M. Abdallah, C. Clevenger, T. Vu, A. Nguyen, Sensing occupant comfort using wearable technologies, in: Constr. Res. Congr. 2016, American Society of Civil Engineers, 2016: pp. 940–950. http://ascelibrary.org/doi/abs/10.1061/9780784479827.095 (accessed August 23, 2017).
[38] C.-C. (Jeff) Huang, R. Yang, M.W. Newman, The potential and challenges of inferring thermal comfort at home using commodity sensors, in: Proc. 2015 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput., ACM, New York, NY, USA, 2015: pp. 1089–1100. doi:10.1145/2750858.2805831.
[39] A. Ghahramani, G. Castro, B. Becerik-Gerber, X. Yu, Infrared thermography of human face for monitoring thermoregulation performance and estimating personal thermal comfort, Build. Environ. 109 (2016) 1–11. doi:10.1016/j.buildenv.2016.09.005.
[40] C. Sugimoto, Human sensing using wearable wireless sensors for smart environments, in: 2013 Seventh Int. Conf. Sens. Technol. ICST, 2013: pp. 188–192. doi:10.1109/ICSensT.2013.6727640.
[41] J.-H. Choi, D. Yeom, Study of data-driven thermal sensation prediction model as a function of local body skin temperatures in a built environment, Build. Environ. 121 (2017) 130–147. doi:10.1016/j.buildenv.2017.05.004.
[42] J. Demšar, Statistical comparisons of classifiers over multiple data sets, J. Mach. Learn. Res. 7 (2006) 1–30.
[43] F.J. Provost, T. Fawcett, R. Kohavi, The case against accuracy estimation for comparing induction algorithms., in: 1998: pp. 445–453.
[44] C. McCarthy, N. Pradhan, C. Redpath, A. Adler, Validation of the Empatica E4 wristband, in: 2016 IEEE EMBS Int. Stud. Conf. ISC, 2016: pp. 1–4. doi:10.1109/EMBSISC.2016.7508621.
[45] J.-H. Choi, V. Loftness, D.-W. Lee, Investigation of the possibility of the use of heart rate as a human factor for thermal sensation models, Build. Environ. 50 (2012) 165–175. doi:10.1016/j.buildenv.2011.10.009.
[46] W. Liu, Z. Lian, Y. Liu, Heart rate variability at different thermal comfort levels, Eur. J. Appl. Physiol. 103 (2008) 361–366. doi:10.1007/s00421-008-0718-6.
[47] A. Mishra, M. Loomans, R. Kosonen, Actimetry for Estimating Occupant Activity Levels in Buildings: A Step Toward Optimal and Energy-Efficient Indoor Conditioning, IEEE Consum. Electron. Mag. 8 (2019) 67–71. doi:10.1109/MCE.2018.2867983.
[48] M.H. Hasan, F. Alsaleem, M. Rafaie, Sensitivity study for the PMV thermal comfort model and the use of wearable devices biometric data for metabolic rate estimation, Build. Environ. 110 (2016) 173–183. doi:10.1016/j.buildenv.2016.10.007.
[49] E. Laftchiev, D. Nikovski, An IoT system to estimate personal thermal comfort, in: 2016 IEEE 3rd World Forum Internet Things WF-IoT, 2016: pp. 672–677. doi:10.1109/WF-IoT.2016.7845401.
[50] W.D. van Marken Lichtenbelt, H.A.M. Daanen, L. Wouters, R. Fronczek, R.J.E.M. Raymann, N.M.W. Severens, E.J.W. Van Someren, Evaluation of wireless determination of skin temperature using iButtons, Physiol. Behav. 88 (2006) 489–497. doi:10.1016/j.physbeh.2006.04.026.
[51] A.D.H. Smith, D.R. Crabtree, J.L.J. Bilzon, N.P. Walsh, The validity of wireless iButtons® and thermistors for human skin temperature measurement, Physiol. Meas. 31 (2010) 95. doi:10.1088/0967-3334/31/1/007.
[52] S.W. Cheatham, M.J. Kolber, M.P. Ernst, Concurrent validity of resting pulse-rate measurements: a comparison of 2 smartphone applications, the Polar H7 belt monitor, and a pulse oximeter with Bluetooth, J. Sport Rehabil. 24 (2015) 171–178. doi:10.1123/jsr.2013-0145.
[53] S. Gillinov, M. Etiwy, R. Wang, G. Blackburn, D. Phelan, A.M. Gillinov, P. Houghtaling, H. Javadikasgari, M.Y. Desai, Variable accuracy of wearable heart rate monitors during aerobic

exercise., Med. Sci. Sports Exerc. 49 (2017) 1697–1703. doi:10.1249/MSS.0000000000001284.
[54] D. Anguita, A. Ghio, L. Oneto, X. Parra, J.L. Reyes-Ortiz, A public domain dataset for human activity recognition using smartphones., in: 2013.
[55] Q. Mourcou, A. Fleury, C. Franco, F. Klopcic, N. Vuillerme, Performance evaluation of smartphone inertial sensors measurement for range of motion, Sensors. 15 (2015) 23168–23187. doi:10.3390/s150923168.
[56] R. Wang, G. Blackburn, M. Desai, D. Phelan, L. Gillinov, P. Houghtaling, M. Gillinov, Accuracy of wrist-worn heart rate monitors, JAMA Cardiol. 2 (2017) 104–106. doi:10.1001/jamacardio.2016.3340.
[57] M. Kuhn, Caret package, J. Stat. Softw. 28 (2008) 1–26.
[58] S. Aoki, E. Mukai, H. Tsuji, S. Inoue, E. Mimura, Bayesian networks for thermal comfort analysis, in: 2007 IEEE Int. Conf. Syst. Man Cybern., 2007: pp. 1919–1923. doi:10.1109/ICSMC.2007.4413772.
[59] Y. Gao, E. Tumwesigye, B. Cahill, K. Menzel, Using data mining in optimisation of building energy consumption and thermal comfort management, in: 2nd Int. Conf. Softw. Eng. Data Min., 2010: pp. 434–439.
[60] J.-H. Choi, D. Yeom, Investigation of the relationships between thermal sensations of local body areas and the whole body in an indoor built environment, Energy Build. 149 (2017) 204–215. doi:10.1016/j.enbuild.2017.05.062.
[61] A.A. Farhan, K. Pattipati, B. Wang, P. Luh, Predicting individual thermal comfort using machine learning algorithms, in: 2015 IEEE Int. Conf. Autom. Sci. Eng. CASE, 2015: pp. 708–713. doi:10.1109/CoASE.2015.7294164.
[62] G.A. Meijer, K.R. Westerterp, H. Koper, F.H. Ten, Assessment of energy expenditure by recording heart rate and body acceleration., Med. Sci. Sports Exerc. 21 (1989) 343–347.
[63] R.P. Wilson, C.R. White, F. Quintana, L.G. Halsey, N. Liebsch, G.R. Martin, P.J. Butler, Moving towards acceleration for estimates of activity-specific metabolic rate in free-living animals: the case of the cormorant, J. Anim. Ecol. 75 (2006) 1081–1090. doi:10.1111/j.1365-2656.2006.01127.x.
[64] A. Uzelac, N. Gligoric, S. Krco, A comprehensive study of parameters in physical environment that impact students' focus during lecture using Internet of Things, Comput. Hum. Behav. 53 (2015) 427–434. doi:10.1016/j.chb.2015.07.023.
[65] E. Ben-Joseph, J.S. Lee, E.K. Cromley, F. Laden, P.J. Troped, Virtual and actual: Relative accuracy of on-site and web-based instruments in auditing the environment for physical activity, Health Place. 19 (2013) 138–150. doi:10.1016/j.healthplace.2012.11.001.
[66] J. Cohen, A coefficient of agreement for nominal scales, Educ. Psychol. Meas. 20 (1960) 37–46. doi:10.1177/001316446002000104.
[67] J. Ranjan, J. Scott, ThermalSense: Determining dynamic thermal comfort preferences using thermographic imaging, (2016).
[68] A. Ben-David, About the relationship between ROC curves and Cohen's kappa, Eng. Appl. Artif. Intell. 21 (2008) 874–882. doi:10.1016/j.engappai.2007.09.009.
[69] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in: Montreal, Canada, 1995: pp. 1137–1145.
[70] V. Földváry Ličina, T. Cheung, H. Zhang, R. de Dear, T. Parkinson, E. Arens, C. Chun, S. Schiavon, M. Luo, G. Brager, P. Li, S. Kaam, M.A. Adebamowo, M.M. Andamon, F. Babich, C. Bouden, H. Bukovianska, C. Candido, B. Cao, S. Carlucci, D.K.W. Cheong, J.-H. Choi, M. Cook, P. Cropper, M. Deuble, S. Heidari, M. Indraganti, Q. Jin, H. Kim, J. Kim, K. Konis, M.K. Singh, A. Kwok, R. Lamberts, D. Loveday, J. Langevin, S. Manu, C. Moosmann, F. Nicol, R. Ooka, N.A. Oseland, L. Pagliano, D. Petráš, R. Rawal, R. Romero, H.B. Rijal, C. Sekhar, M. Schweiker, F. Tartarini, S. Tanabe, K.W. Tham, D. Teli, J. Toftum, L. Toledo, K. Tsuzuki, R. De Vecchi, A. Wagner, Z. Wang, H. Wallbaum, L. Webb, L. Yang, Y. Zhu, Y. Zhai, Y. Zhang, X. Zhou, Development of the ASHRAE global thermal comfort database II, Build. Environ. 142 (2018) 502–512. doi:10.1016/j.buildenv.2018.06.022.
[71] J. Pietilä, S. Mehrang, J. Tolonen, E. Helander, H. Jimison, M. Pavel, I. Korhonen, Evaluation of the accuracy and reliability for photoplethysmography based heart rate and beat-to-beat

detection during daily activities, in: H. Eskola, O. Väisänen, J. Viik, J. Hyttinen (Eds.), EMBEC NBC 2017, Springer Singapore, 2018: pp. 145–148.

https://doi.org/10.1016/j.buildenv.2019.106281
https://escholarship.org/uc/item/3fb0p5gk

Appendix

A1. Accuracy comparison among sensors

Previous studies reported that off-the-shelf wristbands or smartwatches might not measure physiological signals accurately. For instance, Basis Peak (Intel, Corp., U.S.) and Fitbit Charge HR (Fitbit, Inc., U.S.) inaccurately measure heart rate during exercise [56]. Moreover, wristband might have a higher sensitivity to wrist motion than Electrocardiography (ECG) chest strap to the motion of the torso. The percentage of correctly detected heart beats by Empatica E4 (Empatica Inc., U.S.) was 68% during sitting and only 9% during household work by tracking 25 male subjects [71]. Even if some parameters can be measured accurately by a wristband or smartwatch, to our knowledge, no comprehensive studies have been reported to validate all the embedded parameter sensors (e.g., skin temperature, heart rate) in one single wearable device. The inaccuracies of the measurements for some parameters could result in significant biases to personal thermal comfort models.

In this study, we evaluated the eligibility of two commercial wristbands, Empatica E4 and Polar V800 (Polar Electro, Ltd., Finland). The infusion of multiple sensors (e.g., heart rate, skin temperature, accelerometry) would cause minimum intrusiveness to occupants. The assessment of the two wristbands was conducted by comparing with already validated sensors described in Section 3.3.

Empatica E4 was claimed to measure skin temperature at a manufacturing accuracy of 0.2 °C within 36-39 °C, according to the specification. The accuracy of heart rate (derived from blood volume pulse) was not specified. For the Polar V800, the embedded temperature sensor was designed to measure air temperature since it was not completely contacted with skin. However, the sensor was positioned on the skin-contact side of the wristband and the air gap in between was only a few millimeters. Therefore, the measurement by the sensor could be strongly correlated to skin temperature.

Figure A1 shows the measured skin temperature and heart rate of four sensors (E4, V800, H7, and iButton) for approximately 10 h by one of the authors who was working in an office. Statistically significant deviations of the measurements between commercial wristbands (E4 and V800) and baseline sensors (iButton and H7) were observed. In addition, the *"spearman"* correlation of skin temperature is $r_s = 0.918$ between iButton and V800, and $r_s = 0.572$ between iButton and E4. The correlation of heart rate is even lower, $r_s = 0.242$, between H7 and E4. The inaccuracy of the commercial wristbands can be partially attributed to body movement that loosened the contact, while the iButton was fastened on the skin using a medical tape and H7 strap was damped and slightly tight on the chest. The measures ensured better contact with the skin during movement. As such, the results although based on one person for several hours imply that the current commercial wristbands might not be sufficiently accurate for physiological signal tracking during daily life, especially when extensive body movement occurs.

https://doi.org/10.1016/j.buildenv.2019.106281
https://escholarship.org/uc/item/3fb0p5gk

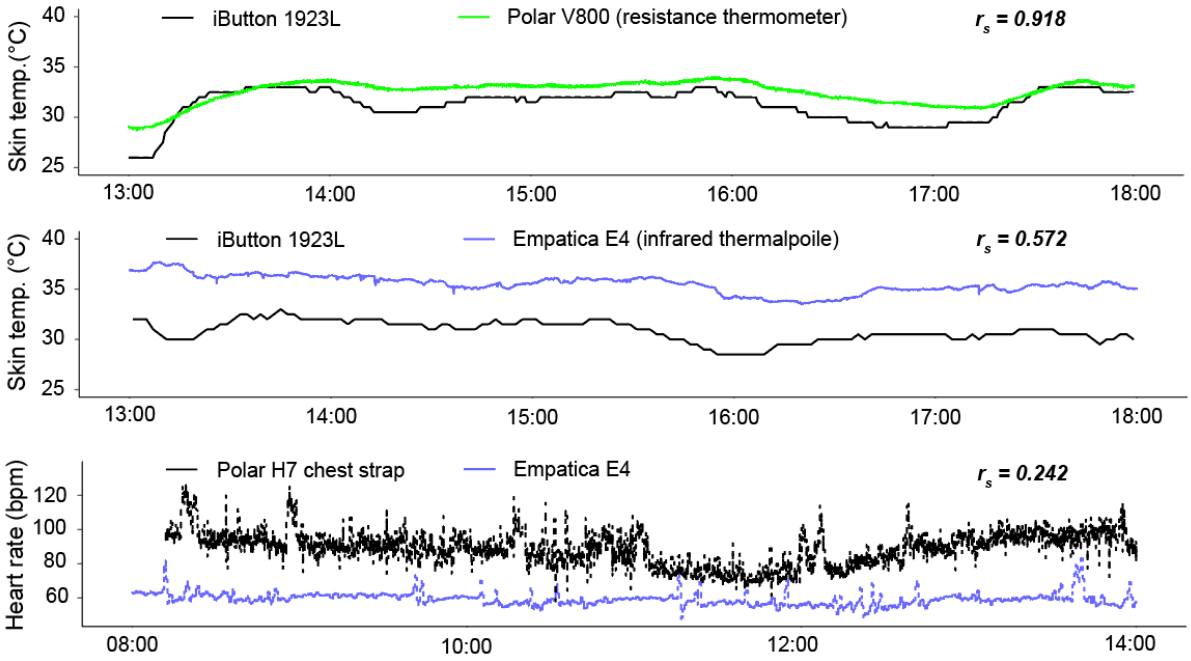


Figure A1. Comparison of measured daily skin temperature and heart rate by different sensors for one subject

A2. Correlation matrix among features

Figure A2 shows the correlation matrix of various parameters. The correlations between thermal sensation and other variables ordered from the highest to lowest are, 0.267 (outside air temperature), 0.226 (ankle skin temperature), 0.187 (close-proximity temperature), 0.184 (heart rate), 0.134 (wrist skin temperature), and 0.002 (activity). It is worth noting that these correlations are calculated using the entire dataset of the 14 participants. In addition, the correlations vary on the clusters of thermal preference ("Cooler", "No change", and "Warmer").

https://doi.org/10.1016/j.buildenv.2019.106281
https://escholarship.org/uc/item/3fb0p5gk

Figure A2. Correlation matrix among different variables. *TS:* thermal sensation; *TP:* thermal preference; *HR_5_avg (bpm):* average HR over 5 min; *T_60_avg_out (°C):* average outdoor air temperature over 1 h; *T_5_avg_wrist (°C):* average wrist skin temperature over 5 min; *T_5_avg_ankle (°C):* average ankle skin temperature over 5 min; *Acc_5_sd (m/s^2):* standard deviation of wrist acceleration over 5 min.

A3. Importance of features for prediction performance

Figure A3 displays the relative importance of each feature contributing to the performance of personal models using SVM classification. The order of the features would slightly vary when using another classification algorithm. The values were used as the first screen check for feature selections. It is observed from the chart that the standard deviation of variables is less important than the average or gradient. Also, unimportant features derived from some time frames can also be eliminated. The analysis yielded a number of 22 features for personal model development.

https://doi.org/10.1016/j.buildenv.2019.106281
https://escholarship.org/uc/item/3fb0p5gk

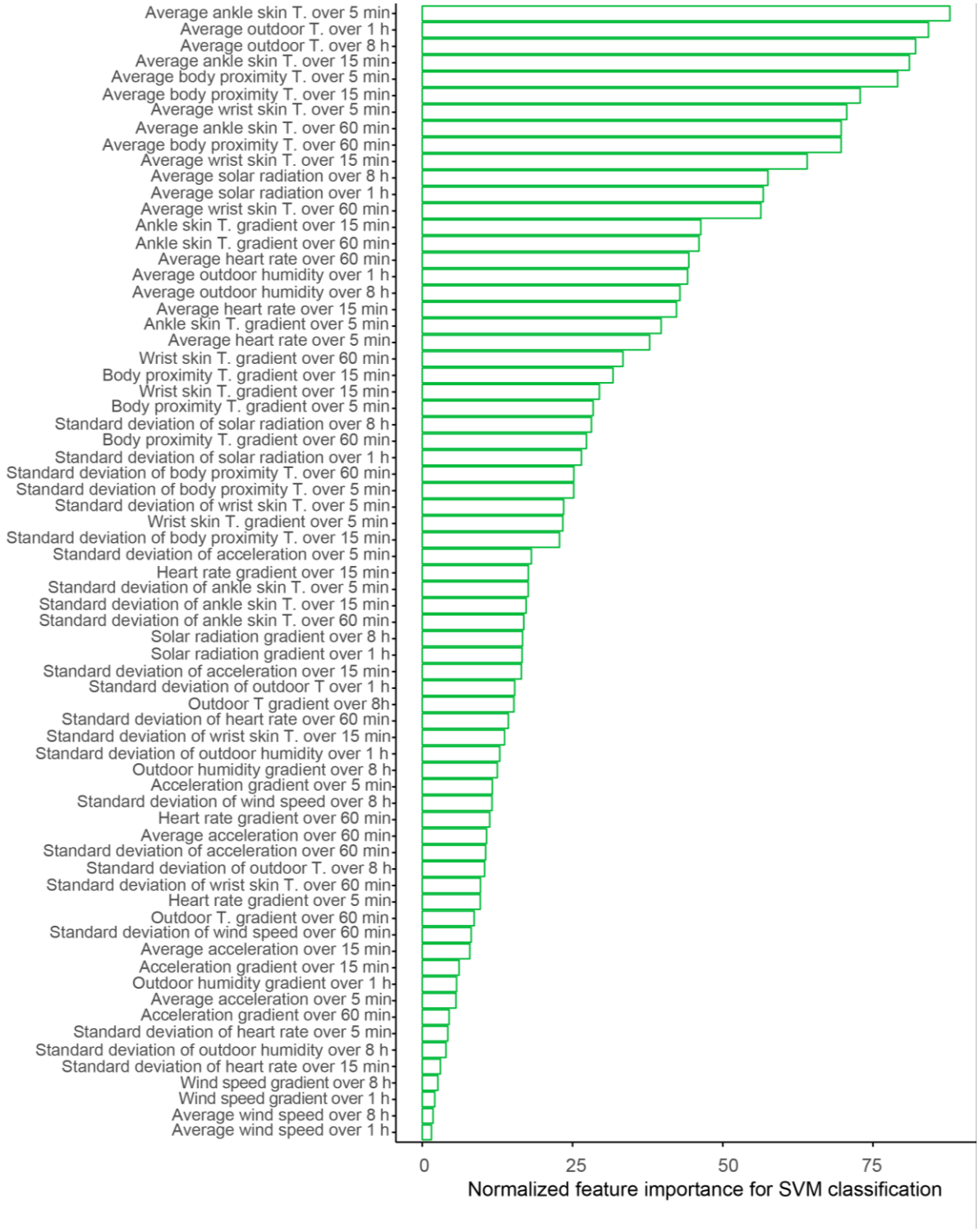


Figure A3. Importance of different features for prediction performance using SVM classification based on the whole dataset

https://doi.org/10.1016/j.buildenv.2019.106281
https://escholarship.org/uc/item/3fb0p5gk

View publication stats