Fall 12-3-2018

Data Driven Approach to Thermal Comfort Model Design

Mostafa Rafaiejokandan
University of Nebraska, mostafa@huskers.unl.edu

DATA DRIVEN APPROACH TO THERMAL COMFORT MODEL DESIGN

by

Mostafa Rafaiejokandan

A THESIS

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of the Requirements

For the Degree of Master of Science in Architectural Engineering

Major: Architectural Engineering

Under the Supervision of Professor Fadi Alsaleem

Lincoln, Nebraska

October 9, 2018

DATA DRIVEN APPROACH IN COMFORT MODEL DESIGN

Mostafa Rafaiejokandan, M.S.

University of Nebraska, 2018

Advisor: Fadi Alsaleem

Apart from the dominant environmental factors such as relative humidity, radiant, and ambient temperatures, studies have confirmed that thermal comfort significantly depends on internal personal parameters such as metabolic rate, age, and health status. This study reviews the sensitivity of the Predicted Mean Vote (PMV) thermal comfort model relative to its environmental and personal parameters of a group of people in a space. PMV model equations adapted in ASHRAE Standard 55–Thermal Environmental Conditions for Human Occupancy, are used in this investigation to conduct a parametric study by generating and analyzing multi-dimensional comfort zone plots. It has been found that personal parameters such as metabolic rate and clothing have the highest impact. Current and newly emerging advancements in state of the art wearable technology have made it possible to continuously acquired biometric information. This work proposes to access and exploit this data to build a new innovative thermal comfort model. Relying on various supervised machine-learning methods, a thermal comfort model has been produced and compared to a general model to show its superior performance. Finally, the study represents an

architecture to employ new thermal comfort model in inexpensive, responsive and extensible smart home service.

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Fadi Alsaleem for the continuous support of my master study and research. Also, I would like to thank the rest of my thesis committee: Prof. Prithviraj (Raj) Dasgupta and Dr. Iason Konstantzos for their encouragement, insightful comments and judgment.

My sincere thanks also go to Mr. Mohammad H Hasan and Mr. Mehari K. Tesfay for their invaluable assistance and cooperation in every step of operating the experiments, analyzing the result and publishing the work.

I also would like to thank my family who has always been supportive, especially my Dad and Mom. I would particularly like to appreciate my loving and supportive wife, Fariba who always provide infinite inspiration.

APPROVAL/SIGNATURE PAGE

We approve this Final Report of Click or tap here to enter your name.**:**

_____ _____

Fadi Alsaleem, PhD Date

Assistant Professor

The Durham School of Architectural Engineering and Construction

_____ _____

Prithviraj (Raj) Dasgupta, PhD Date

Professor

Computer Science Department, University of Nebraska, Omaha

_____ _____

Iason Konstantzos, PhD Date

Assistant Professor

The Durham School of Architectural Engineering and Construction

TABLE OF CONTENTS

Topic **Page**

LIST OF FIGURES

Figure Number **Topic** **Page**

LIST OF TABLES

Table Number **Title** **Page**

LIST OF SYMBOLS

Symbol	Physical meaning	Units
$\boldsymbol{a}$	Age	years
$\boldsymbol{BEE}$	Basal Energy Expenditure	$Kcal/day$
$\boldsymbol{BMR}$	Basal Metabolic Rate	$Kcal/day$
$\boldsymbol{D}$	Duration of Activity	$minutes$
$\boldsymbol{EER}$	Estimated Energy Requirement	$Kcal/day$
$\boldsymbol{f_{cl}}$	Clothing Factor	--
$\boldsymbol{h}$	Height	cm
$\boldsymbol{h_c}$	Convective Heat Transfer Coefficient	$W/m^2.K$
$\boldsymbol{I}$	Activity Intensity	--
$\boldsymbol{I_{cl}}$	Clothing Insulation	CLO
$\boldsymbol{L}$	Overall Heat Transfer around occupant	W/m^2
$\boldsymbol{M}$	Metabolic Rate	W/m^2
$\boldsymbol{MET}$	Metabolic Equivalence	MET
$\boldsymbol{m}$	mass	Kg
$\boldsymbol{P_a}$	Partial Pressure of Water	KPa
$\boldsymbol{PA}$	Physical Activity Level	--
$\boldsymbol{PAL}$	Physical Activity Level Factor	--
$\boldsymbol{PMV}$	Predicted Mean Vote	--

RH	Relative Humidity	--
$\boldsymbol{S}$	Sensitivity	$unit\hat{} - 1$
s	BMR Sex Factor	--
$\boldsymbol{t_a}$	Air Temperature	$°C$
$\boldsymbol{t_{cl}}$	Clothing temperature	$°C$
$\boldsymbol{\bar{t}_r}$	Mean Radiant Temperature	$°C$

CHAPTER 1

INTRODUCTION

1.1. BACKGROUND AND MOTIVATION

Currently people spend more than 90% of their time in indoor spaces (Frontczak and Wargocki 2011), (Höppe and Martinac 1998). Most of these indoor environments are conditioned with different types of HVAC systems that consume about 50% of primary energy in the building (Pérez-Lombard, Ortiz, and Pout 2008) to ensure the thermal satisfaction of the occupant (Wagner et al. 2007) and health (Allen et al. 2015). The impact of thermal satisfaction (Leaman and Bordass 1999; Salonen et al. 2012) on productivity in workplaces has made residential and commercial building designers to make comfort a design requirement of high priority. To ensure the criteria for the thermal comfort during the design and operation of the buildings, different types of standards and guidelines are usually used (EN 2007; Iso 2005). The American Society of Heating, Refrigerating, and Air-Conditioning Engineers (ASHRAE) Standard 55 is the most popular one that has been utilized extensively in the United States.

Predictive Mean Vote (PMV) and adaptive comfort models are the two primary models used in most of thermal comfort standards. In the PMV Model (Poul O Fanger and others 1970), personal and environmental information are fed to a mathematical equation to calculate the thermal comfort. This model is the default thermal comfort

model. On the other hand, the Adaptive comfort model (De Dear et al. 1998) uses a linear regression of acceptable indoor operative temperatures. While both models are designated to satisfy 80% of occupants, both models have some limitations. For example, none of these models supports variables such as gender or age. Moreover, the PMV model needs expensive sensors to capture data for airspeed or metabolic rate. Moreover, some studies (Van Hoof 2008; Auffenberg, Stein, and Rogers 2015) showed that the accuracy of these models for a small group of occupants is poor. Therefore, there is a need for a new comprehensive thermal comfort model.

1.2. LITRERTURE REVIEW

Thermal comfort is a subjective matter and may vary from person to person. There have been multiple attempts to develop a unified thermal comfort model that can widely be adopted. The most popular model is the PMV model, which was constructed by P. O. Fanger (1970) (Poul O Fanger and others 1970) and was later adapted into the ASHRAE Standard 55–Thermal Environmental Conditions for Human Occupancy. This model is meant to estimate the "average" thermal sensation that a group of people would report when occupying a given space. To this end, the model correlates multiple environmental parameters (air temperature, air velocity, relative humidity, and radiant temperature) and the average of personal parameters of group of people (metabolism and clothing) to different levels of comfort based on a rating between –3 and 3, where –3 means the body thermal sensation is very cold and 3 means the body thermal sensation is very hot. Typically, the goal is to control environmental

factors in order to keep the PMV value between –0.5 and 0.5, where the body is believed to be thermally satisfied or in comfort. Since the environmental parameters are relatively easy to measure, they have received a great deal of attention in this literature. Personal parameters, on the other hand, are more difficult to estimate or measure and, therefore, are usually assumed to have constant default values for a group of people (e.g. rest condition (Hang-yat and Wang 2013; Ku et al. 2015; Xue, Zhai, and Chen 2013; Gao and Keshav 2013)) and summer clothing (Nastase, Croitoru, and Lungu 2016; Hang-yat and Wang 2013; Ku et al. 2015), thus missing the opportunity to accommodate comfort variation due to clothing and metabolism. Multiple tables were created to map the metabolic rate and clothing conditions based on occupants activity level and clothing styles (McCullough, Jones, and Huck 1985; Hoovestol and Mikuls 2011). However, as most buildings don't include any feedback measurement loop from the occupants, one level of activity (metabolism) and one clothing insulation of the occupants is assumed for the sake of simplicity.

More recent works have adapted variable metabolic rate in the PMV calculation. For example, in (Luo et al. 2016) the metabolic rate was accurately measured using the medical Vmax encore™ station. It was shown that the metabolic rate of an occupant would change at different ambient temperatures without any change in activity. (Martellotta et al. 2016) studied the effect of thermal, visual, and acoustic factors on the overall comfort. In their case study, the metabolic rate and clothing values were considered by observation and in reference to metabolic rate and clothing tables. High

cost and its dependence on human observations and surveys may limit the use of these methods to estimate metabolism in real time.

Another limitation of the PMV model is that it fails to predict thermal comfort of a small group of people while accounting for their personal differences. Thus, the key for any new thermal comfort model is to account for the individual thermal choice in different conditions. In this regard, the study (Kim, Schiavon, and Brager 2018) reviewed the relevant papers to the new development in comfort modeling during the last ten years, and categorized the researches into two groups. The first one is a data-driven approach to model and predict thermal comfort of a general population(Dai et al. 2017; Chen, Wang, and Srebric 2015) and the second group is using the synthetic data to model personal comfort (Peng and Hsieh 2017; Ari et al. 2008). For the model output, most studies used the 3-point thermal preferences (warmer/no change/cooler) or ASHRAE 7-point thermal sensation scale. Also most these studies used indoor air temperature, mean radiant temperature, and relative humidity in their dataset while they tried to gather the individual information such as metabolism and rated skin temperature using wearable devices (Hasan, Alsaleem, and Rafaie 2016).

1.3. OBJECTIVES & ORGNIZATION

To overcome the current comfort modeling challenges, the goals of this study are to use wearable device data to: (1) propose a method to augment and enhance the PMV

model and (2) present a novel data-driven approach to define a new thermal comfort model. To achieve these goals, the following tasks were performed in this work:

- Obtain visualized multi-dimensional comfort zone plots using the PMV model
- Conduct parametric and sensitivity analysis for the PMV model
- Enhance the PMV model calculation using wearable devices data
- Perform new comfort modeling using wearable device biometric data
- Propose a new architecture to integrate the new comfort model with excising comfort systems

The organization of the thesis is as follows: In **chapter 2**, the methodologies employed and equipment used in this study have been discussed. The chapter begins with the definition of two different approaches that are used in this study. In a separate discussion, **chapter 2** covers the details about the devices, technologies, techniques and tools that were utilized to perform our investigation.

In **chapter 3**, we study the ASHRAE Standard 55 as the currently most agreed upon standard to review the effectiveness of different parameters on occupant thermal comfort. In addition to this review, we show how employing a smart-band to measure the metabolism rate could improve thermal comfort estimation as per the standard.

In **chapter 4**, a new data-driven approach has been introduced to create comfort models. Different machine learning types are used to create a general and personalized comfort models.

In Chapter 5 a new architecture to implement the new thermal comfort model in a smart home service has been introduced. An initial prototype is also conducted to review the challenges and measure the quality of the proposed architecture. The chapter also includes a new innovative idea for a virtual temperature sensor that is capable of measuring the local temperature without using a dedicated temperature sensor. Finally, **chapter 5** also summarizes our findings in this study and explores some details about future research plans.

CHAPTER 2

METHODOLOGY

Employing wearable devices to advance research that demands bio-information data is not a new approach. However, the availability of new high-quality equipment and devices in the market today has made conducting these research ideas and their implementations in real life much easier. Smart-bands and smart-watches are the most common wearable devices utilized for fitness, health monitoring, and communication. In this chapter, smart-bands are used, for the first time, to construct a new thermal comfort model.

2.1. SMART-BANDS

Smart bands, known with titles like activity tracker and fitness tracker, appeared in the market in the early 2000's. Invented for activity tracking, the first devices were equipped with accelerometers and altimeters sensors to calculate mileage, physical activity, and estimate calorie expenditure. Likewise, the early Fitbit, released at 2009, were worn at the waist. Such devices were capable of collecting bio-information and location data. However, they required additional interface to communicate with a primary computing device, such as a computer or a phone. More recently, these devices were designed as armbands and wristbands. These devices embed a large collection of sensors such as heart rate, skin resistance and GPS, which are used to

derive an appropriate summary of the user's fitness and health. The collected data in these devices can be transferred to the user's smartphone via Bluetooth or WiFi.

In this study, we used Fitbit HR and Microsoft Smart Band II in our investigation. The following sections include detailed technical review on these two devices.

2.1.1. Fitbit HR

Fitbit, a San Francisco based company, founded in 2007, released its first product in 2009. Until 2014, all Fitbit products were limited to track walked steps, climbed floors, sleep quality, and clocks. On January 2015, Fitbit achieved a marketing milestone by featuring their products with bio-information sensors. Fitbit Charge HR was released in early January 2015, to be the first product that is capable to continuously track heart rate. This new technology was advertised as the PurePulse® technology.

Fitbit Specifications

Current Fitbit bands include a variety of sensors such as: optical heart rate monitors, 3-axis accelerometers, altimeter sensors and skin temperature sensors. The band records the heart rate data at one-second intervals for workout tracking and at 5-second intervals in the normal conditions.

Figure 2-1 Fitbit Charge HR

Fitbit Charge HR could be connected to a smartphone via Bluetooth to transmit the collected data to the Fitbit application and, subsequently, to a cloud service. The cloud service is responsible for processing the data and generating different types of summary reports such as the weekly and monthly reports. To display the reports, the Fitbit application requires resynchronization with the cloud service. Finally, in addition to Bluetooth, the band can synchronize data with a computer via a dongle connection.

Data Accessibility

The Fitbit sensor data is stored in Fitbit cloud service after synchronizing the wearable device with a pair device (e.g. mobile or PC). The OAuth 2.0 protocol, supported by the Fitbit cloud service, was used to connect to the Fitbit cloud service. To access the Fitbit data, an application was built using the Flask micro framework. Flask is a python library that implements a lightweight micro framework based on Werkzeug and Jinja2.

This implementation is accessible through every internet browser. A user may initiate, using the application, a data request that is then forwarded to Fitbit cloud page for authentication. After a successful authentication, a callback service in flask will be triggered to request the data. Finally, a link to the extracted data as a comma-separated value file (CSV) would be shown on web page. These steps are summarized in Figure 2-2 and could be applied simultaneously for a group of occupants in a building to arrive to an average metabolism value that can be used in the PMV calculation.

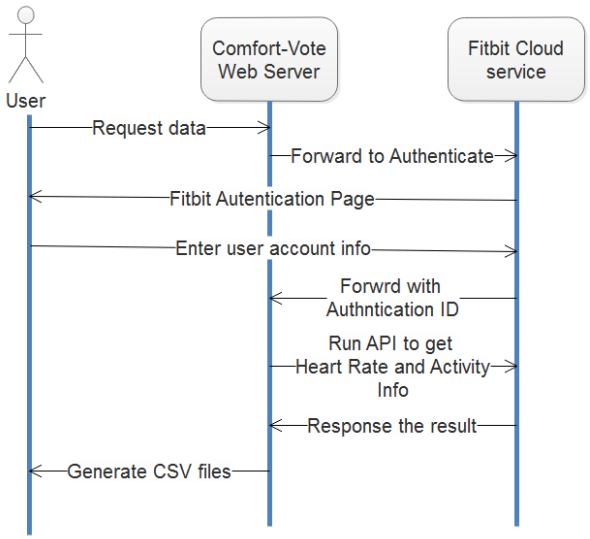


Figure 2-2 Application processes to extract Fitbit data

According to the Fitbit Web API reference, there are 11 groups of APIs. These APIs include the AUTH API, which is used to generate access tokens in the authentication process and the Time Series API, which is the most important API, used to get the time series of different data for the given data-time and resource type. Table 2-1 shows more details about these different APIs.

Table 2-1 Fitbit Web-API Summary

("Fitbit - Web API Reference" 2018)

API type	Details	Methods
Activity & Exercise Logs	The "Get Daily Activity Summary" endpoint retrieves a summary and list of a user's activities and activity log entries for a given day in the format requested using units corresponding to the Accept-Language header provided.	Get, Post, Delete
Auth	Authentication and authorization using OAuth 2.	Post
Body & Weight	The "Get Body Fat Logs" endpoint retrieves a list of all user's body fat log entries for a given day in the format requested. Body fat log entries are available only to authorized user. If the most recent entry is needed, "the Get Body Measurements" endpoint is used.	Get, Post, Delete
Devices	Fitbit allows the concurrent use of multiple activity tracker devices and scales. Furthermore, it automatically detects when a person switches from one tracker to another throughout their day or week. This unified data of this API is returned to users.	Get, Post, Delete
Food Logging	The "Get Food Locales" endpoint returns the food locales that the user may choose to search, log, and create food in.	Get, Post, Delete

Friends	The "Get Friends" endpoint returns data of a user's friends in the format requested using units which corresponds to the Accept-Language header provided.	Get
Heart Rate	The "Get Heart Rate Time Series" endpoint returns time series data in the specified range for a given resource in the format requested using units that corresponds to the Accept-Language header provided.	Get
Sleep Logs	The "Get Sleep Logs by Date" endpoint returns a summary and a list of a user's sleep log entries (including naps) as well as detailed sleep entry data for a given day.	Get, Post, Delete
Subscriptions	Because of the frequency of updates, Fitbit has developed the Subscriptions API that allows third-party applications to be notified when a Fitbit user's data changes. This allows apps to retain the user's latest data without having to implement a polling or scheduling system to retrieve user's data.	Get, Post, Delete
User Profile & Settings	The "Get Profile" endpoint returns the user's profile. The authenticated owner receives all values. Access to other user's profile is not available. If one wishes to retrieve the profile information of the authenticated owner's friends, Get Friends API is used. Numerical values are returned in the unit system specified in the Accept-Language header.	Get, Post

2.1.2. Microsoft Band 2

Microsoft is recognized as a software company with popular products like Microsoft Windows and Office. Around 2015, Microsoft announced its first product in activity tracker markets with a brand called Microsoft Band. The device combined features from a smart band, a smartwatch, and an activity tracker. Similar to the other smart-bands in the market, the band uses Bluetooth connection to pair with a phone and interact with the cloud service. Combining more health-oriented capabilities, Microsoft Band 2 was released in 2016.

Specifications

The second generation of the Microsoft smart band (Microsoft Band 2) is a wearable device that combines features from a smartwatch like communication and smart bands like activity tracking. It is also bundled with collection of sensors. Figure 2-2 shows a picture of the band including numbers referring to some sensors and features in the band. Other than the 12 sensors it contains, the band also includes a microphone (numbered 6 in the figure) to speak with the Microsoft Personal Digital Assistant (Cortana). This feature helps to operate some voice commands like sending texts.

Figure 2-2 Microsoft Band 2

("Get Started with Your Microsoft Band 2" 2018)

Unlike competitor devices, Microsoft band support real time data collection from all its inboard sensors. Table 2-2 shows sensors information; including the data generation frequency.

Table 2-2 Microsoft Band 2 Sensory details.

(Microsoft 2015)

Sensor Name	Details	Frequency	ID
Accelerometer	Provides X, Y, and Z acceleration in g units	62/31/8 Hz	-
Gyroscope	Provides X, Y, and Z angular velocity in degrees per second (°/sec) units.	62/31/8 Hz	-

Heart Rate	Provides the number of beats per minute; also indicates if the heart rate sensor is fully locked on to the wearer's heart rate.	1 Hz	2
Skin Temperature	Provides the current skin temperature of the wearer in degrees Celsius.	1 Hz	-
UV	Provides the current ultraviolet radiation exposure intensity.	1 HZ	3
Band Contact	Provides the current ultraviolet radiation exposure intensity.	Value change	-
Galvanic Skin Response	Provides the current skin resistance of the wearer in kohms.	0.2/5 Hz	6
RR Interval	Provides the interval in seconds between the last two continuous heart beats.	Value change	-
Ambient Light	Provides the current light intensity (illuminance) in lux (Lumes per sq. meter).	2 Hz	-
Barometer	Provides the current raw air pressure in hPa (hectopascals) and raw temperature in degrees Celsius.	1 Hz	1
Altimeter	Provides current elevation data like total gain/loss, steps ascended/descended, flights ascended/descended, and elevation rate.	1 Hz	-

Data Accessibility

The band supports two method of data connectivity. In the first method, applications can be installed directly onto the band as a new web tile. This tile is capable of accessing the band features and getting data from the internet.

The web tile is a custom application designed for the Band which includes a JSON file called "manifest.json". The json file defines the UI details like the form, font size and elements. In addition, it includes a channel that manages the dynamicity to get data from defined URL and show them on the Band's interface. To start the implementation, one needs to access the URL at: https://developer.microsoftband.com/webtile/ChooseLayout and proceed a wizard to generate a Web Tile application. Application deployment is facilitated through the Microsoft Health application. It also manages the connectivity the band to the internet. To install and activate the web tile in the band, one only needs to open the Web Tile application in their phone with the tile app running.

The second, and most popular method, is by developing a mobile application employing an SDK as an interface. This facilitates a direct connection to the band. This method is the only way to access sensory data directly from the band.

The Microsoft Band SDK supports Android, iOS and Windows mobile OS. The SDK includes a library for every mobile OS and can be added to the mobile application during the implementation. The SDK primarily utilize the Microsoft Health application to connect the band and manage the connectivity between the phone and the band. The SDK aids the direct access to the band's sensors and features and allows the creation of new Web Tiles on the band. In a regular applications, everything starts by initializing the connection to the band and configuring the sensors and features to

work as per needed. Beside the sensory data in Table 2-2, Table 2-3 presents additional features accessible via SDK.

Table 2-3 Additional Features accessible via Microsoft Band SDK

(Microsoft 2015)

Sensor Name	Details	Frequency
Pedometer	Provides the total number of steps the wearer has taken since the Band was last factory-reset.	Value change
Calories	Provides the total number of calories the wearer has burned since the Band was last factory-reset. This is a lifetime counter and not a daily or a 0-based counter. To determine the absolute number of calories burned between two readings, you must take the difference between the returned values.	Value change

2.2. EXPERIMENT DESIGN

Two distinct studies were conducted in this work to study the different aspects of using wearable device data in thermal comfort modeling. Each study used a different wearable devices (Fitbit or Microsoft Smart Band 2) and different biometric data set. In the first study, the metabolic rate was collected from a Fitbit wearable device to study the effects of the metabolic rate on the PMV comfort model; whereas in the

second study, as much data as possible was collected from the Microsoft Smart Band sensors to define a new thermal comfort model.

2.2.1. Metabolic rate estimation using Fitbit band (Study 1)

Metabolism is difficult to measure and is usually assumed to be a constant value for occupants in a building (e.g. rest condition (Ku et al. 2015)). However, due to the ever-increasing popularity and advancement of wearable fitness devices, the estimation of metabolism has becomes much easier and more convenient. In this study, we have used Fitbit Charge HR™ band data as one way to estimate metabolism as a case study. Recent Fitbit® wearable devices were shown to have an accuracy level up to 95%(Diaz et al. 2015). However, other wearable devices with more sensors and that are known to have higher accuracy could be used instead in this investigation. Fitbit® can be easily configured to share the metabolic rate, heart rate, and activity level of occupants, as discussed before, to a computing unit in real time. These pieces of information are updated every minute to enable, if needed, a fast response. However, to accommodate that the PMV model was designed assuming steady thermal conditions, a simple averaging was applied on these measurements.

Fitbit® metabolism calculation

The metabolism of a Fitbit® user is calculated as a multiple of the basal metabolic rate (BMR), which is defined as the minimum rate of energy expenditure per unit time by an endothermic human at rest (Mifflin et al. 1990):

$$BMR = (\frac{10m}{1\ Kg} + \frac{6.25h}{1\ cm} - \frac{0.5a}{1\ year} + s)\frac{Kcal}{day} \quad \textbf{(2-1)}$$

where *m* is the mass of the body in kilograms, *h* is the height of the body in cm, *a* is the age in years, and *s* is a factor relating to sex, as follows:

$$s = \begin{cases} +5 & for\ males \\ -161 & for\ females \end{cases} \quad \textbf{(2-2)}$$

Fitbit® also uses a built-in accelerometer to infer the activity level of the wearer(Hang-yat and Wang 2013). It uses this information to calculate the estimated the wearer's energy requirement (EER), which is related to the age, sex, weight, height, and physical activity of the user. For males, the EER is

$$EER = 864\ -\ 9.72\ a\,(years) +\ PA\left(14.2\,m(kg) +\ 503\ h\,(meters)\right) \quad \textbf{(2-3)}$$

and for females, the EER is

$$EER = 387\ -\ 7.31\ a\,(years) +\ PA\left(10.9\,m(kg) +\ 660.7\,h\,(meters)\right) \quad \textbf{(2-4)}$$

where PA is the physical activity level that is related to the motion of the person and is measured by Fitbit®'s built-in accelerometer as well as the physical properties of the wearer. The PA is calculated for men as (Gerrior, Juan, and Peter 2006)

$$PA = \begin{cases} 1\,, & 1.0 < PAL < 1.4\ (Sedentary) \\ 1.12, & 1.4 < PAL < 1.6\ (Low\ active) \\ 1.27, & 1.6 < PAL < 1.9\ (Active) \\ 1.54, & 1.9 < PAL < 2.5\ (Very\ active) \end{cases} \quad \textbf{(2-5)}$$

and for women as

$$PA = \begin{cases} 1\,, & 1.0 < PAL < 1.4\ (Sedentary) \\ 1.14, & 1.4 < PAL < 1.6\ (Low\ active) \\ 1.27, & 1.6 < PAL < 1.9\ (Active) \\ 1.45, & 1.9 < PAL < 2.5\ (Very\ active) \end{cases} \quad \textbf{(2-6)}$$

$$PAL = ((I - 1)\ [(1.15/0.9) \times D\ (minutes)]/1440))/\ (BEE/[0.0175 \times 1440 \times w\ (kg)]) \quad \textbf{(2-7)}$$

where I and D are the activity intensity (inferred from heart rate and accelerometer measurements) and duration, respectively, and BEE is the basal energy expenditure, given by

$$BEE = \begin{cases} 2933.8 \times a\ (years) + 456.4 \times h(meters) + 10.12 \times w\ (kg)\ \ Men \\ 2472.67 \times a\ (years) + 401.5 \times h(meters) + 8.6 \times w(kg), Women \end{cases} \quad \textbf{(2-8)}$$

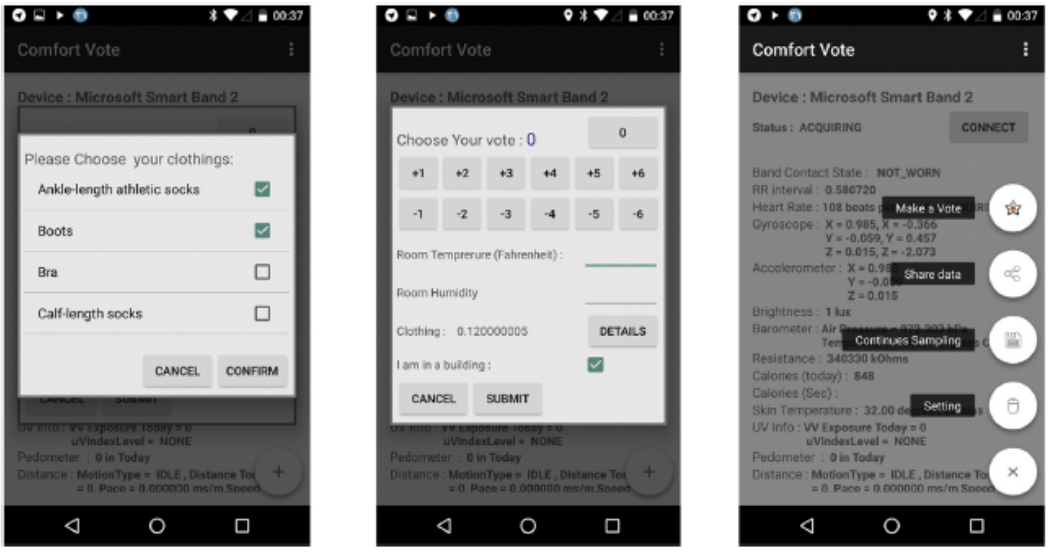
Once EER and BMR are calculated, the metabolic rate can be calculated as

$$MET = \frac{EER}{BMR} \quad \textbf{(2-9)}$$

From equation (2-9), MET is actually the ratio between the energy generated while performing some activity to the rest of the body's metabolic rate. This MET is used to estimate the number of calories burned by the human within thermal comfort equations such as the PMV model.

2.2.2. Thermal comfort modeling using the Microsoft band (Study 2)

In order to study the thermal comfort model using a data driven approach, we designed an experiment where participants carry the Microsoft band and a wireless temperature/humidity sensor. Five individuals were invited to take part in this study. These individuals were periodically prompted to vote on their thermal comfort throughout the day. This prompt came through their smart phones. The prompt they received can be seen in Figure 2-4. Voting was initially conducted with a scale from -6 to 6; with 6 being very hot, -6 being very cold, and 0 being comfortable. While this scale offers a large amount of variation, it was determined that some people may struggle to distinguish between minimal differences on this scale such as that between 5 and 6, therefore, introducing unnecessary human error. For this reason, the alternate scale of only three votes (1,0,-1) was introduced.



(a) (b) (c)

Figure 2-4 Several Screeshots from Mobile Application (Comfort Vote)

A mobile application implemented for this experiment is responsible for collecting the biometric data and vote. The provided data will be pre-processed and cleaned to be ready for a machine learning classifier application to create the new comfort model and perform the accuracy analysis. Figure 2- shows the flow of the data/tasks in this exercise, and in a similar order, the following sub-sections explain them in more details.

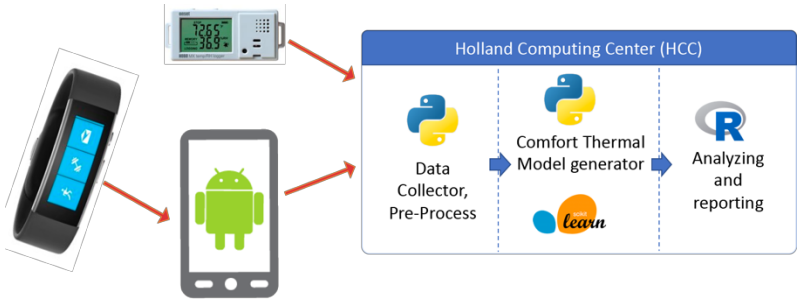


Figure 2-5 The flow of the data/tasks in Comprehensive sensory data experiment

2.2.2.1. ***Mobile Application***

The mobile application is an Android application capable of connecting to the Microsoft Smart band 2 and receives the sensor data in a customized fashion. In the design of this mobile application, three different requirements were addressed:

- **Collecting the comfort vote**: This functionality focuses on getting the feedback of the thermal comfort user as a number (called comfort vote here). The user vote and his bio-information data is the information needs to be stored and labeled with an accurate date-time. A notification in the mobile application

was another customizable functionality built to remind the user for entering his/her current thermal votes.

- **Automatic data sampling:** It is needed that the mobile application to automatically connect to the smart-band and to store its sensors measurements in an internal database. In addition, it is necessary to attach the user vote and temperature measurement to this data. Intuitively, performing these functionalities need to be within a separate thread from the main application running in the startup. This thread should work in the background and has an access to the shared resources like the database in the main mobile program. In addition to the main functionality, it is required to support customized scheduling via the mobile application.
- **Continuous high rate sampling:** the new application should allow storing high sampling data (16-64 sample per second)

The architecture of the developed mobile application to satisfy the above requirements has three major parts:

- **Main application**: It is the mobile application and includes the key functionalities interacting via the interface. It is also the container for other sections.
- **Background service**: As it is cleared, it covers all the process needs to be run as a background service.
- **Resources**: It includes the major libraries and available resources.

Figure 2-3 shows the mobile application architecture with the main components inside in every main category.

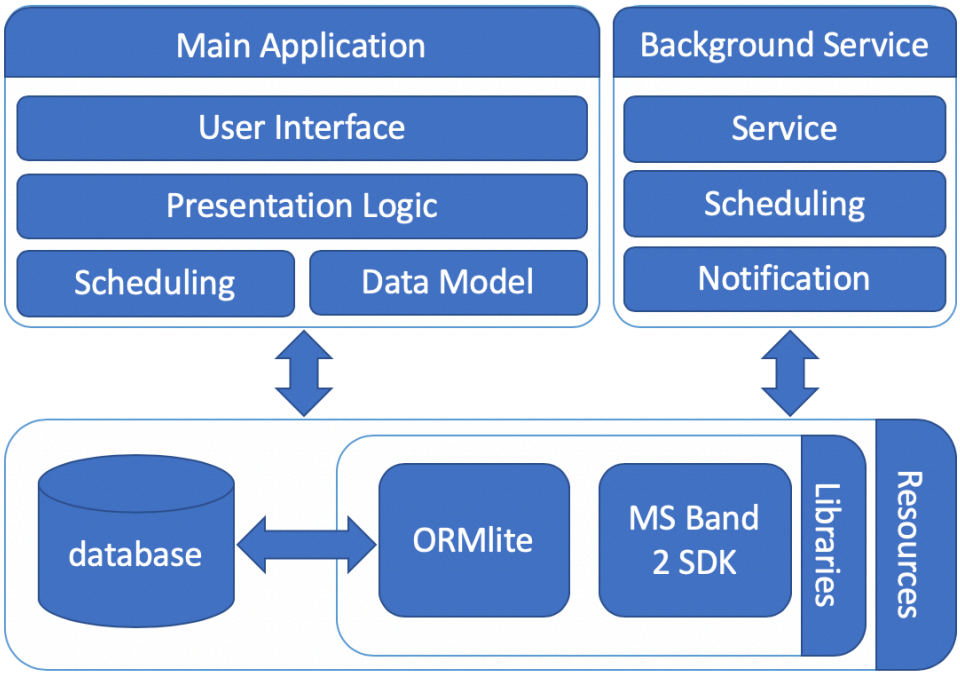


Figure 2-3 Mobile application architecture

2.2.2.1.1. *Main Application*

It's the primary container to include all the components and after installation causing to activate the other parts. As it is shows in Figure 2-3, it contains all the interface of displaying the data and getting the feedback from users. There is one Activity called MainActivity working as the container of the other interfaces, and accordingly, it loads the fragments inside. Details of the fragments are described below:

- **MainActivityFragment:** It's the first screen as it is shown in Error! Reference source not found.(c)**.** It initializes the connection to the Microsoft Band 2 and

shows the sensory data in the main view. Besides, it includes some functionalities like share database.

- **DataSamplingFragment:** It's the interface to run the continuous high rate sampling test. It includes the Start and Stop button to activate the process of storing sample data in the local database and stop it accordingly. Besides, there is a clear button helping the clear the database from sample data.
- **SettingDialogFragment:** As it is cleared from the Fragment name, it manages the setting section. It shows all the application configuration from scheduling and database management capable of updates or change as per setting modification.
- **VoteDialogFragment:** Error! Reference source not found.(a) and Error! Reference source not found. (b) show this Fragments activated. It includes two sections, Collecting the vote numbers and getting the clothing information via checklist and calculate the clothing score as per standards ASHRAE Standard 55(ASHRAE 2004).

As shown in Figure 2-1, Presentation logic is another essential component. Every Fragment loads the user interface and this component as an embedded functionality in every Fragment, responsible for controlling the flow and transferring the data to the interface component as per any changes or updates.

The data model describes the database definition and designes an interface to manage all the request toward the database. The definition details is mentioned in the mobile

resource section. The final section is the scheduling, it is responsible for the internal scheduling requirement capable of calling an interface or run a process or change an information.

2.2.2.1.2. *Background service*

This component is an extension to the Service class and is registered after the installation of the main application on the Android mobile phone. It is configured to run in phone startup and work live as a background job. This Service class works as the container responsible for reaching the resources and initiating the next two modules called Scheduling and Notification here.

Scheduling is an object inherited from TimerTask capable of activating a function in a configurable schedule. There are two scheduling services here. The first one controls the anonymous bio-information sampling functionality by connecting to the smart-band, collecting and storing the data. The next scheduling features the notification on the phone employed to inform the user sharing the comfort vote. The last part is the notification service extending the Android Notification features to raise a new customized one and trying to notify the user to share the new comfort vote. In addition, all the configuration needed to manage the element are stored in the internal database, and the service component facilitated the accessibility of this resource.

2.2.2.1.3. *Resource*

The resources are divided into two group libraries and shared resources. Besides the standard Android libraries, two new items are included in this project. The first one is called Microsoft Band SDK, which is the most critical one, implemented by Microsoft to facilitate the connection to the smart-band and moderate the sensory data requests. In addition to this SDK, ORMLite was imported to maintain the object-relational mapping for the internal database. Hence, to design a manageable database inside a mobile application, classes of object, annotations defining the variable type in the database, and a configuration determining the database type and location are needed. In this project, we chose the SQLite Android as the target database.

The database, as the shared repository of data and configuration, included four tables:

- **SensorData:** It is the base table to store the bio-information sampling information including all the sensory data supported in Microsoft Band SDK.
- **ComfData**: It includes all the fields in SensorData table plus votes, clothing details and environmental data (room temperature, room humidity, locationType). This table is used to keep the comfort vote data. The location type defined in the table is a marker showing the feedback recorded when the user was in a room or not.
- **SensorSampleData**: It is a class inherited from SensortData table to store continues high rate sampling test data.

- **SettingData**: As the name suggests, it simply stores the configuration data like the sampling interval and notification intervals.

2.2.2.2. ***Operations***

There are five notable operations in the application:

- **Connect to the Band**: The application automatically connects to the Microsoft Band. In case this operation fails, there is a button in the main interface that can reinitiate the process of connection to the band. The button is visible in the background of **Error! Reference source not found.**(c)
- **Make a vote**: It runs from the options button list and leads to a new interface. This new interface has the textboxes to get the room temperature, room humidity and a checkbox to control if the user is in a building or not. The clothing information is a checklist moderated by this operation, which leads to a number called clothing score. In the end, the user needs to click on a button named with a number varied from -6 to 6 to give his thermal comfort feedback. The digits on the button displays feeling hot if the sign is positive and feeling cold if the sign is negative. Also, the more decimal value indicates the severity of the condition for the user. The **Error! Reference source not found.**(a) and **Error! Reference source not found.**(b) show this operation
- **Configuration**: It includes all the setting and it is accessible via options button list in the first page.

- **Share data**: As the data is stored and collected in the phone, it is required to extract all the data and send it to the third-party software or services. In this operation, the process reads all the ComfData and SensoryData table, combines them in Comma Separated Value(CSV) file and share with Intent.ACTION_SEND feature implemented by Android. This Android facility will list all the application installed on the phone supporting this functionality like mail services and messaging services. As soon as the user would choose a program from the list, the CSV file will send it the application. In this experiment, we usually used the email services like Gmail or Google Inbox to export the data.
- **Continues Sampling**: accessible via options button list in the first page, it includes all the controls like stop and start the process, clear SensoreSampelingData table and share the stored data.

2.2.2.3. ***Date***

After the completion of the experiments, there were two groups of data:

- Data collected by Comfort Vote mobile application
- Environmental Data collected by Onset HOBO device.

All the data are formatted in the CSV format and stored in different files. In the beginning, it is needed to merge all the data files belonging to every group data for every participant. As a result, there were ten files, 5 of environmental data and the same amount for the Comfort Vote plus bio-information data. In the next step, data

of the occupants are required to merge as per timestamp to generate a dataset for everyone including the environmental information attached to the base data. To satisfy this requirement, a python code was implemented, which was ran on every group of data to generate the merged data.

2.2.2.3.1. *Data cleanup*

Now, that there are five datasets as per participant, the data needs to be filtered as per Table 2-4. As the experiment was defined for the work time, we excluded all the data before eight in the morning and after nine at night.

Table 2-4 Data cleanup rules

	Situation	Min Value	Max Value
1	Heart Rate	40	220
2	Skin Temperature	10	45
3	MET	0.4	0.8
4	Room Temperature	0	40
5	Room Humidity	0	100
6	Hour	8	21

The data includes voted data and the automated sampling info. In this experiment, we focused solely on the labeled data referred to voted data; hence, unlabeled data were filtered. In the final step, a new column was attached and marked with the participant's name, as it was needed for a specific model.

2.2.2.3.2. *General and Personal dataset*

The dataset for every user includes all the data related to him/her, and it is referred to as the personal dataset in this work. In an experiment, it is required to study the impact of data without considering the behavior of the user or individual features. So, all the dataset combined to create a new dataset titled as the general dataset here.

2.3. MACHINE LEARNING

Machine learning as an application of artificial intelligence (AI) is frequently employed to build analytical models capable of automated inference and prediction. Among the different varieties of categorization, it is commonly agreed to four groups: Supervised learning, Unsupervised Learning, Semi-supervised Learning, and Reinforcement Learning. Depending on the purpose of the study and the nature and size of data, mostly one of the categories is utilized. As the dataset size is tiny and the vote is the target label for a feature including bio-information, and environmental data lead to use the supervised machine learning. We picked five of the most prominent algorithms and tried to find the best potential models iterating on an individual collection of initializing variables.

2.3.1. Decision tree

As a non-parametric supervised learning method, it is one of the simplest and yet most successful forms of machine learning used for classification and regression. It is

represented in a tree-like graph or model regularly used as a classifier to decide from the possible choices (Ligeza 1995; Quinlan 2006). This algorithm also supports the regression while there are a limited number of labels in this study, it is utilized as the classifier here. The main parameters of this machine learning method are depths determining the maximum depth of the final tree.

2.3.2. Adaptive Boosting

It is a machine learning algorithm commonly knowns as AdaBoost and is an ensemble learning method designed to select a collection, or ensemble, of hypotheses from the hypothesis space and combine their predictions (Ligeza 1995). It's also a meta-estimator classifier creating a base fitted model from the dataset and using the different copies of the classifier with different weight to achieve a better outcome for inaccurately classified cases (Freund and Schapire 1997). It's possible to determine the estimator count, and for this study, we tried variables from five to a thousand with the step of five.

2.3.3. Gradient Boosting Classifier

It is another ensemble learning technique used for the classification and regression providing a prediction model in the form of a weak ensemble prediction models, typically decision trees (Mason et al. 1999). This classifier is known as a robust method for overfitting, and it assumes the higher estimator number will generate the better

performance(Vezhnevets and Barinova 2007). In this study, we employed Gradient Boosting Classifier with the exact number of estimators used for AdaBoost.

2.3.4. Random Forest Classifier

Utilizing the aggregation of decision trees built from various sub-samples of the datasets and averaging to improve the predictive accuracy is the base method in Random Forest to create the optimal model.(Ligeza 1995; Tin Kam Ho 1995; Ho 1998) Again, as an ensemble learning method, it can be used for classification and regression, and we employed it by the different estimator count to achieve the better accuracy.

2.3.5. Support Vector Machines

Also known as SVM, it's currently the most popular approach for "off-the-shelf" supervised learning (Ligeza 1995). Besides the linear classification, it uses the kernel trick to perform the non-linear classification. Linear, Poly, RBF, Sigmoid, and Precomputed are the main kernels, and we utilized the RBF for this experiment(Chang et al. 2010). SVM with RBF as the kernel can have the variable for the error term called C, and we investigated the models with different variables from 0.1 to 35.

2.3.6. Training and performance evaluation

The base dataset includes the Room Temperature, Room Humidity, Heart Rate, Skin Temperature, Clothing Score, and Metabolism rate and the new scaled Vote is the target class. Moreover, as per need to study the impact of a new sensory data like GSR, we combined the base data with new data to create a virtual or physical dataset. All the physical data is stored in CSV file, which is readable with the simple editor and Microsoft Excel.

We utilized the Python programming language to implement the Classifier application and import the Scikit-learn library to use various machine learning models. This library is open source implemented in Python programming language, and it features different classification, regression and clustering algorithms.(Pedregosa et al. 2012) To handle the model adaption dynamically, we used parameter passing to the python application. A unique ID translated to a unique combination of the dataset with a specific feature list was the only needed by the program. After that, the application could examine all the machine learning model with a different set of parameters customized as per machine learning and stored all the result in a file titled with runtime date-tima and the id.

We use cross-validation method which is a popular method to evaluate the accuracy of predictive methods. In this technique, the dataset is divided into N partitions and using the N-1 partition for the training and the remaining for the test, counting this

process with N possible combination, it will bring a proper judgment about the accuracy of the model.(Kohavi 1995; Ligeza 1995). We employed the 'cross_val_score' function from Scikit-learn library and ran it with four partitions. The result was a list of four values, and we used the mean value to measure the final accuracy of the model.

All the experiments were ran on a virtual server with 16 VCPUs and 30 GB Ram using the Anvil service from Holland Computing Center (HCC). Details of the results, including about 750 models, were stored in a CSV file. R programming language and RStudio environment were used to analyze and generate the charts and figures.

CHAPTER 3

ENHANCED PMV

Much recent work has adapted variable metabolic rate in the PMV calculation. For example, in (Luo et al. 2016) the metabolic rate was accurately measured using the medical Vmax encore™ station. It was shown that the metabolic rate of an occupant would change at different ambient temperatures without any change in activity. (Martellotta et al. 2016) studied the effect of thermal, visual, and acoustic factors on the overall comfort. In their case study, the metabolic rate and clothing values were considered by observation and in reference to metabolic rate and clothing tables. High cost and relying on human observations and surveys may limit the use of these methods to estimate metabolism in real time. In this chapter, we show the use of wearable devices data to provide a continuous feedback of occupants' metabolism value.

Evaluating PMV equations is computationally intensive and requires iterative processes. Hence, many approximations were made. For example, in (Ku et al. 2015), an artificial neural network (ANN) model was employed to capture the dynamics of the PMV model equations and then use them to evaluate the PMV value for any given thermal parameters set. Also, Zhang and You (Zhang and You 2014) introduced a sequential approximation to nonlinear equations, not only to simplify the calculation

but also to find the "air temperature, relative humidity" pair that leads to maximum energy savings (inverse problem solution).

Most human comfort research work has focused on studying the comfort effect of air temperature, which is widely accepted as the most important parameter in thermal comfort models, coupled with a few other environmental factors, such as air velocity and relative humidity (Hoovestol and Mikuls 2011). Less work has dealt with the effects of comfort and sensitivity to metabolism and clothing, which are personal parameters. This may reflect the fact that personal parameters are underestimated, or difficult to quantify and measure. In fact, metabolism and clothing thermal resistance play a vital role in defining the optimal thermal comfort conditions. While metabolism increases the rate of heat generation in the human body and decreases the desirable comfort temperature, clothing helps to tolerate colder conditions. Assuming clothing and metabolism to be constant values may lead to a false PMV calculation.

Focus in this paper will be given to metabolism and its direct effect on the PMV model. MET is the unit of metabolism in the PMV model. A single MET is equivalent to the heat a body produces while it inhales 3.5 ml of oxygen (O_2) per kg of weight each hour (H); ($\frac{ml\ O_2}{Kg.H}$) (Jetté, Sidney, and Blümchen 1990). Also, MET can be thought of as multiples of the resting metabolic rate for the occupant while he or she is engaged

in a physical (or mental) activity (Trumbo et al. 2002). Accurate measurement of metabolism requires knowing the amount of oxygen the body inhales or the amount of carbon dioxide and nitrogen waste were produced from the cellular breathing process (Ferrannini 1988). This task is not trivial, as it involves using devices such as mask calorimeters to measure the gas intake and outtake. Other devices can also be used to estimate the metabolic rate, such as pedometers, load transducers (also known as foot-contact monitors), accelerometers, and heart rate monitors. Those sensors individually provide an indirect estimate of the metabolic rate and often result in numerous errors. Recent advancement in smart wearable devices has made it possible to fit most of these sensors into a single smart band, thus allowing an accurate and continuous estimation of the metabolic rate. In this paper, in order to estimate the metabolic rate, we will use the Fitbit Charge HR™ smart wearable device that is equipped with a pedometer, an accelerometer, and a heart rate sensor.

In this chapter, after reviewing with the PMV definition as per ASHRAE Standard 55, we next conduct a parametric study for the various PMV environmental and personal parameters and highlight the PMV model sensitivity to these parameters. Next, we focus on the use of a wearable fitness device to acquire the metabolic rate for occupants during normal life activities.

3.1. PMV

The PMV value can be directly calculated using a system of highly nonlinear and iterative equations, which were later adapted in the ASHRAE Standard 55(Hoovestol and Mikuls 2011):

$$PMV = (0.028 + 0.3033e^{-0.036M}) \times L \quad \textbf{(3-1)}$$

$$\begin{aligned} L = (M - W) &- 3.05x10^{-3}(5733 - 6.99(M - W) - Pa) \\ &- 0.42(M - W - 58.15) \\ &- 1.7x10^{-5}(5867 - Pa) - 0.0014M(34 - t_a) \\ &- f_{cl}h_c(t_{cl} - t_a) \\ &- 3.96x10^{-8}f_{cl}[(t_{cl} + 273)^4 - (\overline{t}_r + 273)^4] \end{aligned} \quad \textbf{(3-2)}$$

where L defines the overall heat transfer around a single occupant in W/m^2, M is the metabolic rate in W/m^2, W is the work emitted by the occupant in W/m^2, Pa is the water vapor pressure, $\bar{t}_r$ is the mean radiant temperature in $°C$, t_a is the air temperature in $°C$, $\mathrm{f_{cl}}$ is the clothing insulation factor defined as the percentage of the total body surface area covered by clothing, I_{cl} is the clothing insulation in CLO and h_c is the convective rate heat transfer coefficient in $W/m^2.K$ given by:

$$h_c = \begin{cases} 2.38(t_{cl} - t_a)^{0.25}, if\ 2.38(t_{cl} - t_a)^{0.25} > 12.1\sqrt{V} \\ 12.1\sqrt{V}, \quad if\ 2.38(t_{cl} - t_a)^{0.25} > 12.1\sqrt{V} \end{cases} \quad \textbf{(3-3)}$$

Where V is the air velocity in ms^{-1}, and t_{cl} is the clothing temperature in $°C$, which can be calculated based on the conditions of the body using the following simple heat balance equation:

$$t_{cl} = 35.7 - 0.028(M - W) - 0.155I_{cl}\{3.96x10^{-8} \times f_{cl}[(t_{cl} + 273)^4 + (\bar{t}_r + 273)^4] + f_{cl} \times h_c(t_{cl} - t_a)\} \quad \textbf{(3-4)}$$

A heat balance approach is adapted in the PMV model to infer human thermal comfort(P. O. Fanger 1972). For example, equations (3-1), (3-2) were obtained by balancing the following heat modes: (1) heat generation due to metabolism $(M - W)$, (2) heat transfer by convection $0.0014M(34 - t_a)$, (3) heat transfer through the skin $3.05x10^{-3}(5733 - 6.99(M - W) - Pa)$, (4) heat transfer through latent respiration $1.7x10^{-5}(5867 - Pa)$, (5) heat transfer by dry respiration $0.0014M(34 - t_a)$, and (6) heat transfer by radiation $3.96x10^{-8}f_{cl}[(t_{cl} + 273)^4 - (\overline{t}_r + 273)^4]$. Also in equation (2), clothing temperature is estimated based on the heat generated by metabolism and heat transfer by way of convection and radiation. To solve for the clothing temperature t_{cl} in equation (3-4), ASHRAE incorporated an iterative process to continuously update the clothing temperature until the difference between the current and previous iteration is within a predefined margin.

3.2. PARAMETRIC STUDIES OF THE PMV PARAMETERS

There has been an increasing interest in studying the combined effect of temperature, humidity, and air velocity on PMV values, but less attention has been paid to the effect of metabolism and clothing. In this section, we follow a general approach in studying the interaction of these factors. First, we plot and discuss multiple areas of comfort under the 10% dissatisfaction criteria, i.e., PMV value is between –0.5 and 0.5, while varying the thermal comfort parameters (including the personal parameters) a pair at a time. Then, we construct plots showing comfort zones as a surface while varying three different PMV parameters. In the remainder of this paper, unless otherwise stated, the radiant temperature is assumed to be equal air temperature (Erickson and Cerpa 2012; Ku et al. 2015), and any parameter that does not vary in the simulation is assumed to be constant as follows: clothing = 0.65 CLO (clothing condition), relative humidity (RH) = 50%, MET = 1.0 (Xue, Zhai, and Chen 2013; Ku et al. 2015)(metabolism conditions), and air velocity = 0.5 ms^{-1}.

3.2.1. Parametric study and interactions of environmental thermal conditions

The comfort area (PMV value is between –0.5 and 0.5) as a function of temperature and humidity is depicted as a trapezoid in Figure 1(a). This plot can be used to explain the relative humidity effect on comfort. On a dry (low humidity) day, air has extra capacity to hold water compared to moist (high humidity) day. Hence, a body, through evaporative and latent respiration cooling, can lose heat faster, thus making it feel

colder at the same ambient temperature. This phenomenon explains, for example, why in Figure 1(a) the pair (t_a = 28ºC, RH = 30%) lies within the comfort zone, while the pair (t_a= 28ºC, RH = 60%) does not. For the same reason, high relative humidity means high vapor content in the air. Hence, the generated heat from a human body is trapped and cannot be rejected to the surrounding air by evaporative cooling. Figure 3-1(a) can be used as a guideline for a thermostat logic to maintain occupants' comfort based on temperature and humidity control.

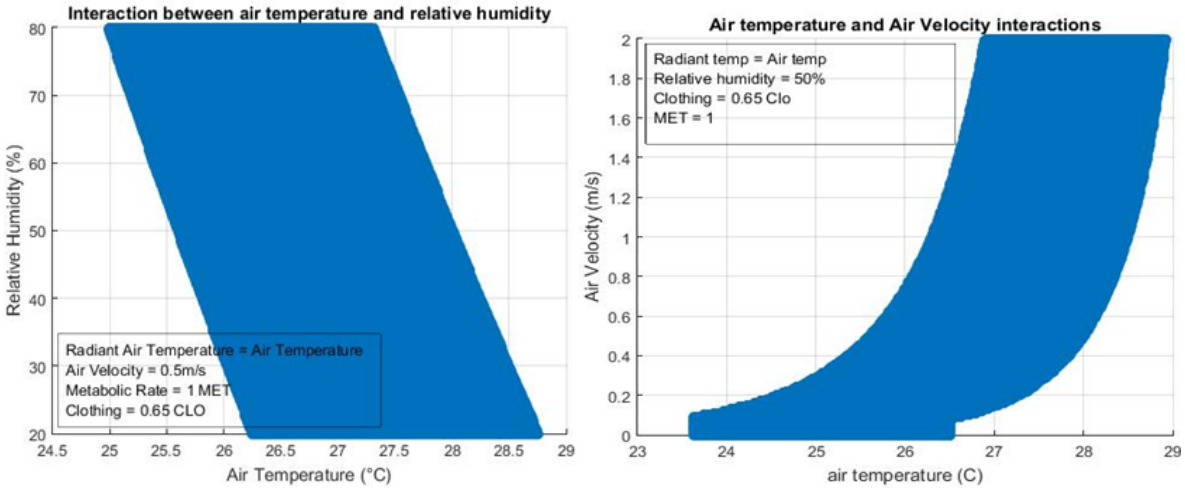


Figure 3-1 : Interaction of air temperature: (a) with relative humidity, (b) with air velocity, and resultant comfort zones

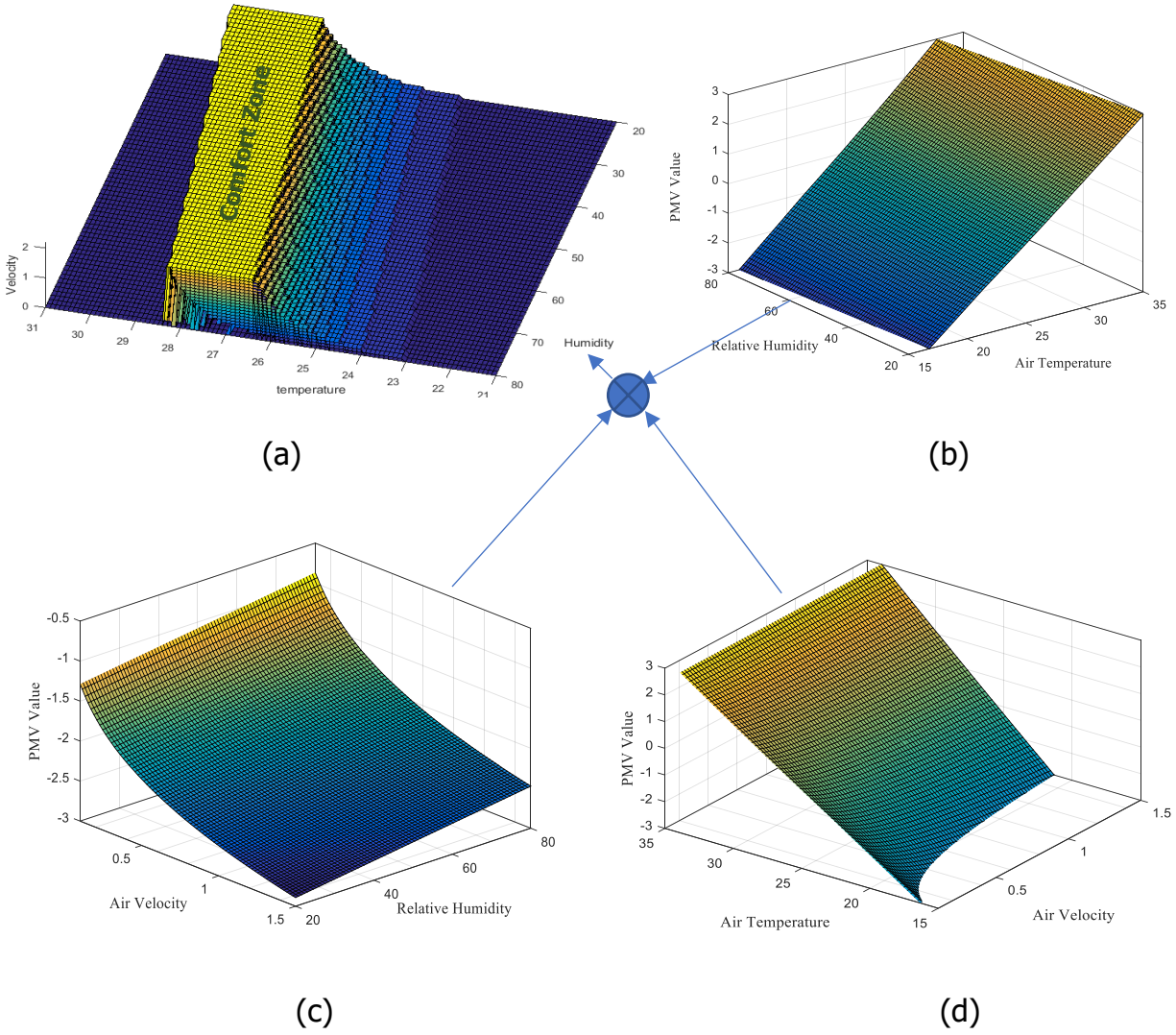


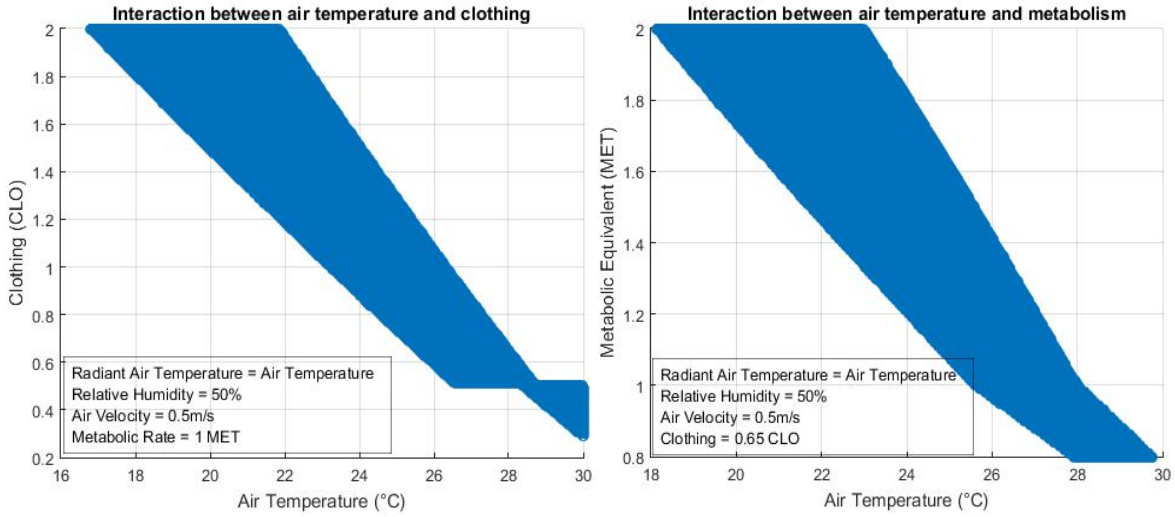
Figure 3-2 (a) 4D plot of interaction between air temperature, humidity, and air velocity, and resultant comfort zone. In this plot, the surface represents all points where the PMV value is between –0.5 and 0.5 (comfort zone) and its color is insignificant, (b), (c), (d) are 3D plots that were used to generate the 4D plot, simulating all three parameters pairs' combination on the PMV value

Air velocity helps in maintaining comfort at high temperatures by increasing the heat-rejection rate through force convection. Figure 3-1(b) shows the comfort zone as a function of air velocity and temperature. It is apparent from this figure that an occupant might tolerate higher temperatures as the air velocity increases. For example, even though the air temperature of 28.5^{o}C is considered to be uncomfortable for all possible relative humidity values above 30%, as shown in Figure 3-1(a), it is within the comfort zone once the air velocity exceeds 1.5 ms^{-1}. This argument necessitates the need for a 4D plot showing the comfort domain (fourth dimension) as a function of humidity, temperature, and air velocity, as shown in Figure 3-2 (a). This plot shows the humidity-temperature trapezoid comfort area shift to the left (higher temperature) as air velocity increases. Figure 3-2 (a) was obtained by merging at least three 3D plots, shown in Figure 3-2 (b), (c), (d), that simulate the multiple three parameters pairs' combination on the PMV value.

3.2.2. Parametric study and interactions of environmental thermal and personal parameters

Figure 3(a) shows the comfort zone area as a function of temperature and clothing, indicating that the comfort zone is very sensitive to clothing. An interesting extremely high sensitivity is noticed when CLO is around 0.5. This value lies between 0.36 CLO, the CLO value of wearing a short-sleeved shirt and shorts, and 0.57 CLO, the CLO value of wearing a short-sleeved shirt with trousers. This behavior occurs around a

clothing point where a comfortable human body changes from a hot to cold feeling. However, this behavior could be due to the use of two discrete equations in the PMV model to calculate the 'clothing area factor'. Figure 3(a) confirms the naive observation that two people under similar thermal conditions can feel differently because of their clothing. More specifically, higher clothing values make the body tolerant to lower temperatures. For example, an occupant at 17^{o}C air temperature and 2 CLO is predicted to be comfortable. This is true because clothing partially isolates the human body and helps to decrease the rate of heat rejection to the outside. On the other hand, high CLO values can quickly move a person out of his/her comfort zone at mild/high temperatures. In that case, body heat becomes trapped inside, creating an extreme feeling of warmth and thermal dissatisfaction. Figure 3(a) also shows that low CLO values result in people tolerating higher ambient temperatures and being very sensitive to mild ambient temperatures. For example, at 0.4 CLO, a human body can tolerate an ambient condition of over 30^{o}C air temperature and humidity ratio of 50%, but the same body can feel cold with ambient temperatures under 28^{o}C. From this discussion, it can be concluded that clothing is very critical to thermal comfort and should not assumed in a building to be a constant.



(a) (b)

Figure 3-3 Interaction of air temperature: (a) with clothing, (b) with air metabolism, and resultant comfort zone

Metabolism is related to heat generation in the human body. As the metabolism (MET) of an occupant increases, the heat generation rate of his/her body increases, leading to a warmness sensation. Metabolism is one of the personal factors in the PMV model that is difficult to estimate and is usually assumed to be constant in most proposed PMV comfort model applications. In a perfect world and if metabolism is measured, a heating, ventilation, and air-conditioning (HVAC) system needs to compensate for an increase in the metabolic rate by controlling the ambient environmental conditions such as decreasing the air temperature, increasing the air velocity, or reducing the relative humidly. Figure 3-3 (b) shows the comfort zone as a function of metabolic rate in METs and air temperature. As expected, the figure shows better comfort results

at low air temperature as the metabolic rate increases. This can be explained by the fact that lower ambient temperatures are needed to reject the internal heat through natural or forced convection. At a high metabolic rate, the body is highly intolerant to high and even mild temperatures. For example, as shown, at a metabolic rate of 1.5 MET, an occupant is thermally uncomfortable at 26°C, a temperature that is well within the comfort zone with a metabolic rate of 1.0 MET. This 1.0 MET, according to ASHRAE Standard 55 tables, is the metabolic equivalent of a person sitting at rest without engaging in any physical and mental activities, and is widely accepted to represent metabolism in the PMV equations when they are used to model comfort in buildings(Ku et al. 2015; Xue, Zhai, and Chen 2013). In this paper, however, we show the use of smart wearable devices to accommodate building occupants' metabolic rate variation in the PMV model calculation.

Figure 3-4 (a) shows the metabolism from a different angle, by displaying its interactions with air temperature and relative humidity. As the metabolic rate increases, the body requires a means to reject the heat/gain cooling in order to stay thermally comfortable. One option is to lower the relative humidity (i.e., air has the capacity to absorb water vapor) in order to provide more cooling through latent respiration and/or sweating. Another option is to increase the heat-rejection rate due to radiation, convection, and dry respiration by lowering the ambient temperature. The generated comfort surface shows a very high sensitivity to metabolism compared to relative humidity and temperature (i.e., surface gradient is higher in MET axis than

in direction of relative humidity axis). The sensitivity analysis is discussed in great details in the next section. The low-curve gradient in the direction of the relative humidity axis may reflect the fact that sweating and latent respiration alone are not fast enough to reject heat from the body at a higher metabolic rate.

We close our discussion by presenting Figure 3-5, which indicates the effects of varying air temperature, clothing, and metabolism on the comfort zone. In general, this figure shows that metabolism has a larger effect on an occupant's thermal sensitivity comfort compared to clothing. However, as shown in Figure 3-3(a), the effect of clothing is more dominate for lower CLO values.

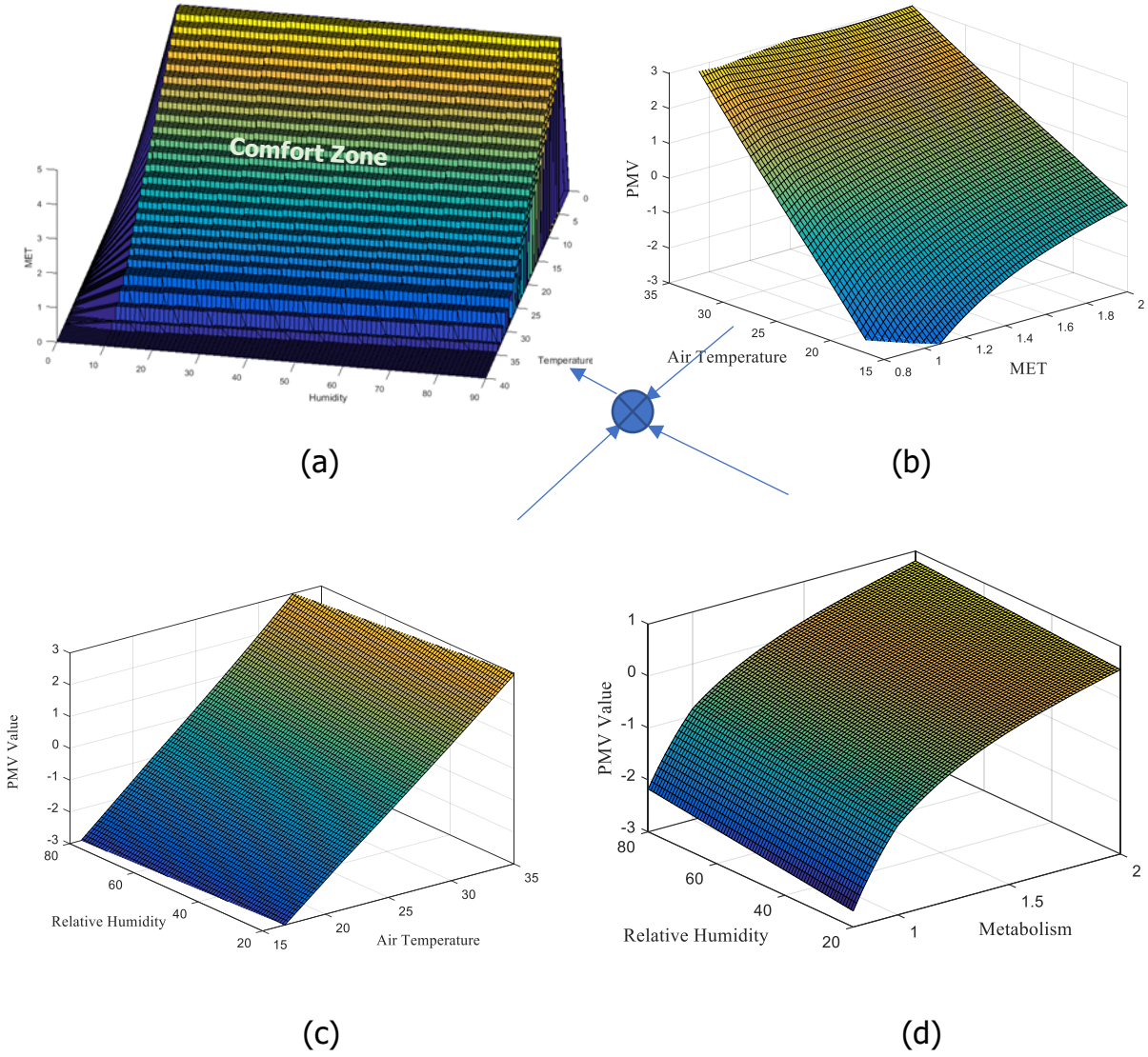


Figure 3-4 (a) 4D plot of interaction between air temperature, humidity, and metabolism, and resultant comfort zone. In this plot, the surface represents all points where the PMV value is between –0.5 and 0.5 and its color is insignificant, (b), (c), (d) are 3D plots that were used to generate the 4D plot, simulating all three parameters pair's combination on the PMV value

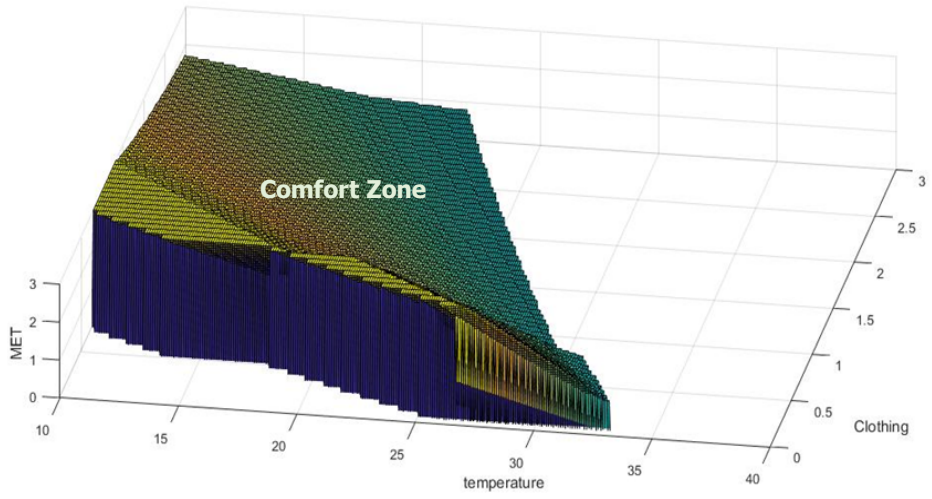


Figure 3-5 Comfort region surface relating to interaction between air temperature, clothing, and the metabolic rate

3.2.3. Sensitivity analysis of the PMV environmental thermal and personal parameters

In this section we study the PMV sensitivity to its thermal and personal parameters. Sensitivity is defined as the partial change in a dependent variable (PMV value) due to the change in one of its independent variables. Mathematically, it is defined as:

$$S_x[f(x,y,z)] = \frac{\partial f(x,y,z)}{\partial x} \qquad \textbf{(3-5)}$$

where S_x is the sensitivity with respect to parameter x, the independent variable. Forward finite difference equation is used to numerically evaluate the partial derivative as follows:

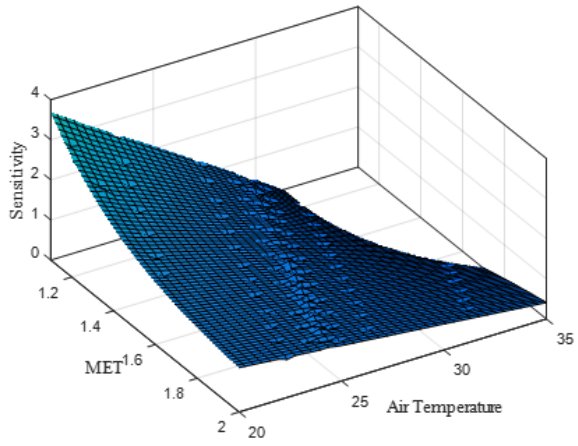
$$f'(x_i) = \frac{(f(x_i) - f(x_{i-1}))}{x_i - x_{i-1}} \quad \textbf{(3-6)}$$

Sensitivity gives a good indication about the magnitude of change in the PMV model value relative to one of its parameters. Some of the lowest-hanging fruit of this analysis, from an energy perspective, is the determination of the most effective control parameters to achieve comfort in a building.

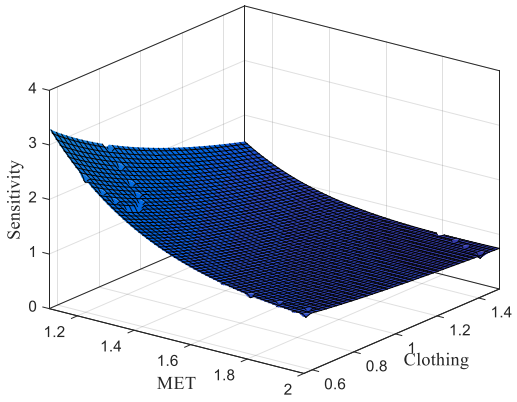
A complete sensitivity analysis for the PMV model is very complex due to the strong interaction between its parameters and is beyond the scope of this paper. Focus will be given to PMV sensitivity to metabolism. Figure 6 shows 3D plots for the PMV sensitivity to the metabolism value at different air temperature values (Figure 3-6(a)), different clothing values (Figure 3-6 (b)), different air velocity values (Figure 3-6 (c)), and different humidity values (Figure 3-6 (d)). Figure 3-6 (a), for example, was obtained by applying the partial derivative equation (6) over the data in Figure 4.b. The figures demonstrate that the PMV sensitivity to metabolism is not constant and it varies according to the following observations: (1) PMV value is more sensitive to metabolism at lower metabolism values, (2) PMV sensitivity to metabolism decreases at higher clothing and air temperature values, (3) humidity and air velocity has the least impact on the PMV sensitivity to metabolism. For more illustration on observations 1 and 2, we show multiple 2D slices of Figure 3-6 (a) at different air temperature values in Figure 3-6 (e). Figure 3-6 (e) shows that a sensitivity up to 3.5 MET^{-1} (i.e. one MET difference can cause up to 3.5 change in the PMV scale) can be

obtained at air temperature = 20C° at low MET values compared to 1.4 MET^{-1} at higher MET values. Moreover, the same figure shows that the sensitivity is around 1.0 MET^{-1} even for low MET values when air temperature =28 C°. It should be noted here that PMV calculations will yield exaggerated results at high metabolic rates due to heat compensation by sweating as shown by (Wang and Hu 2016).

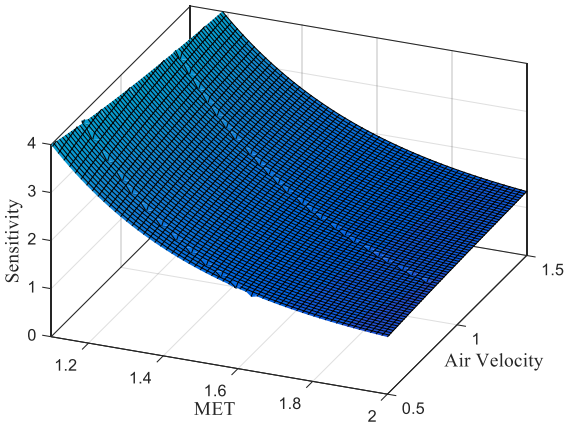
Figure 3-7 compares the PMV model sensitivity against five of its parameters (metabolism, air velocity, humidity, clothing, and air temperature). For each parameter, the sensitivity was calculated while the rest of the parameters were held at the constant default values that were listed at the beginning of the paper.


Sensitivity
MET
Air Temperature

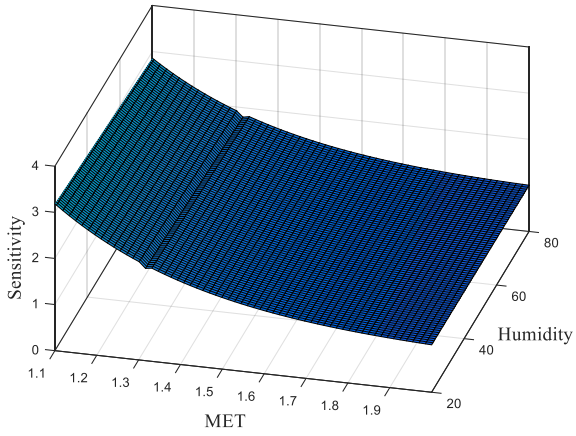
(a)


Sensitivity
MET
Clothing

(b)


Sensitivity
MET
Air Velocity

(c)


Sensitivity
MET
Humidity

(d)

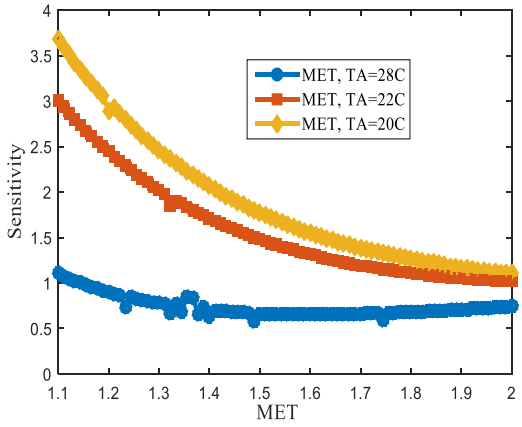
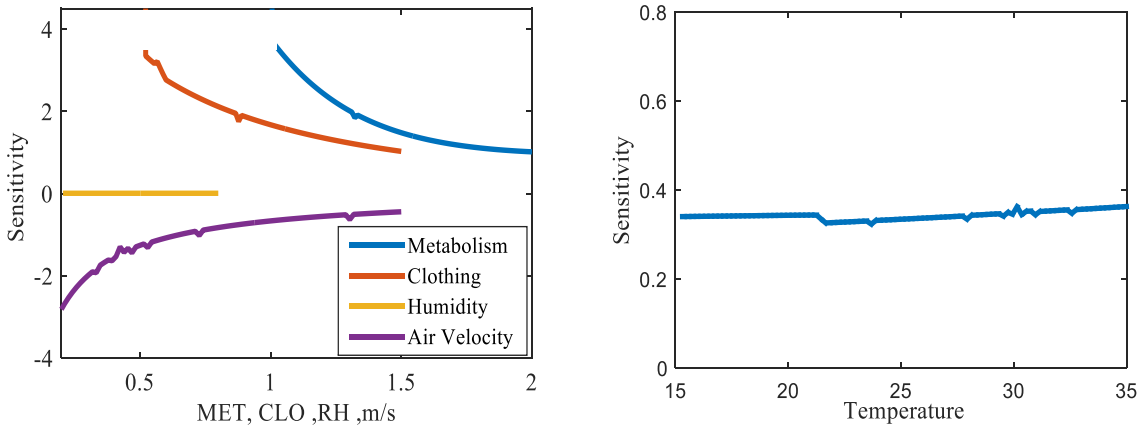

MET, TA=28C
MET, TA=22C
MET, TA=20C
Sensitivity
MET

Figure 3-6 3D plots for the PMV sensitivity to the metabolism value at different (a) air temperature, (b) clothing conditions, (c) air velocity, (d) humidity values, and (e) shows 2D slices of the PMV sensitivity to metabolism at air temperature (AT)=20, 22, and 28C. The plot shows higher PMV sensitivity to metabolism at lower metabolism values and low ambient temperatures

Due to clothing discontinuity issues around 0.5, the PMV sensitivity to clothing were plotted only for clothing values above 0.55. The figure confirms our early conclusion that PMV model is highly sensitive to the personal parameters. Moreover, the figure shows a very interesting behavior. The absolute PMV sensitivity value (PMV sensitivity to air velocity has the only negative sign as an air velocity increase (cooling effect) should reduce the PMV value) exponentially converges for all parameters, but not humidity and air temperature, to a low value as the PMV parameters increases. Air temperature and humidity are the only two parameters that the PMV sensitivity to them almost hold constant. Worth to mention that the sensitivity values in the figure represent the rate of change of the PMV value rather than the actual PMV value. Hence even the figure shows a very small constant sensitivity value for humidity ($\sim$.007 RH^{-1}), over the full expected humidity range (0%-100%) the PMV value changes by up to 0.7. This value is way much less than the expected change of the PMV value of 5.78 over air temperature range of (15 C^{o}- 32 C^{o}), assuming an average

PMV sensitivity to air temperature of $.34\,°C^{-1}$. 5.78 is almost near the full range of the PMV scale '6'. In other words, changes in air temperature by itself can drive the thermal sensation of an occupant from very cold to very hot. We close our PMV model sensitivity analysis with Table 3-1 PMV sensitivities to its parameters summary obtained from Figure 7, where summarized the PMV sensitivities to its parameters in Figure 3-7 are averaged and summarized.



(a) (b)

Figure 3-7 PMV model sensitivity to (a) metabolism, air velocity, humidity, and air temperature, (b) and to air temperature. The figure show that PMV model is very high sensitive to the personal parameters, and its sensitivity to metabolism, clothing, and air velocity, changes with these parameters values and hold constant for air temperature and humidity.

Table 3-1 PMV sensitivities to its parameters summary obtained from Figure 7

Parameter		**Sensitivity (mean)**	**Sensitivity (range)**
Air temperature (t_a)		$S_{AT} \cong .34\ °C^{-1}$	.04
Humidity (*RH*)		$S_{RH} \cong 0.007\ RH^{-1}$	~0
Clothing (CLO>0.5)		$S_{CLO} = 1.3\ CLO^{-1}$	1.22
Clothing (CLO<0.5)		$S_{CLO} = 5.53\ CLO^{-1}$	2.8
Air Velocity (*V>0.5*)		$S_{AV} = -.72\ m^{-1}s$	0.87
Air Velocity (*V<0.5*)		$S_{AV} = -2.2\ m^{-1}s$	2.9
Metabolism (*MET>1*)	t_a=20°C	$S_{MET} = 2.09\ MET^{-1}$	3.37
	t_a=22°C	$S_{MET} = 1.6\ MET^{-1}$	2.0
	t_a=28°C	$S_{MET} = 0.79\ MET^{-1}$	1.25

3.3. METABOLISM ESTIMATION USING SMART WEARABLE DEVICE

Metabolism is difficult to measure and is usually assumed to be a constant value for occupants in a building (e.g. rest condition(Ku et al. 2015)). However, due to the ever-increasing popularity and advancement of wearable fitness devices, the estimation of metabolism becomes much easier and more convenient. In this paper, we have used Fitbit Charge HR™ band data to estimate metabolism as a case study. Recent Fitbit® wearable devices were shown to have an accuracy level up to 95%(Diaz et al. 2015).

However, other wearable devices with more sensors and that are known to have higher accuracy could be used instead in this investigation. Fitbit® can be easily configured to share the metabolic rate, heart rate, and activity level of occupants to a computing unit in real time. These pieces of information are updated every minute to enable, if needed, a fast response. However, to accommodate that the PMV model was designed assuming steady thermal conditions, a simple averaging was applied on these measurements.

3.3.1. Monitoring metabolic rate during normal day activates

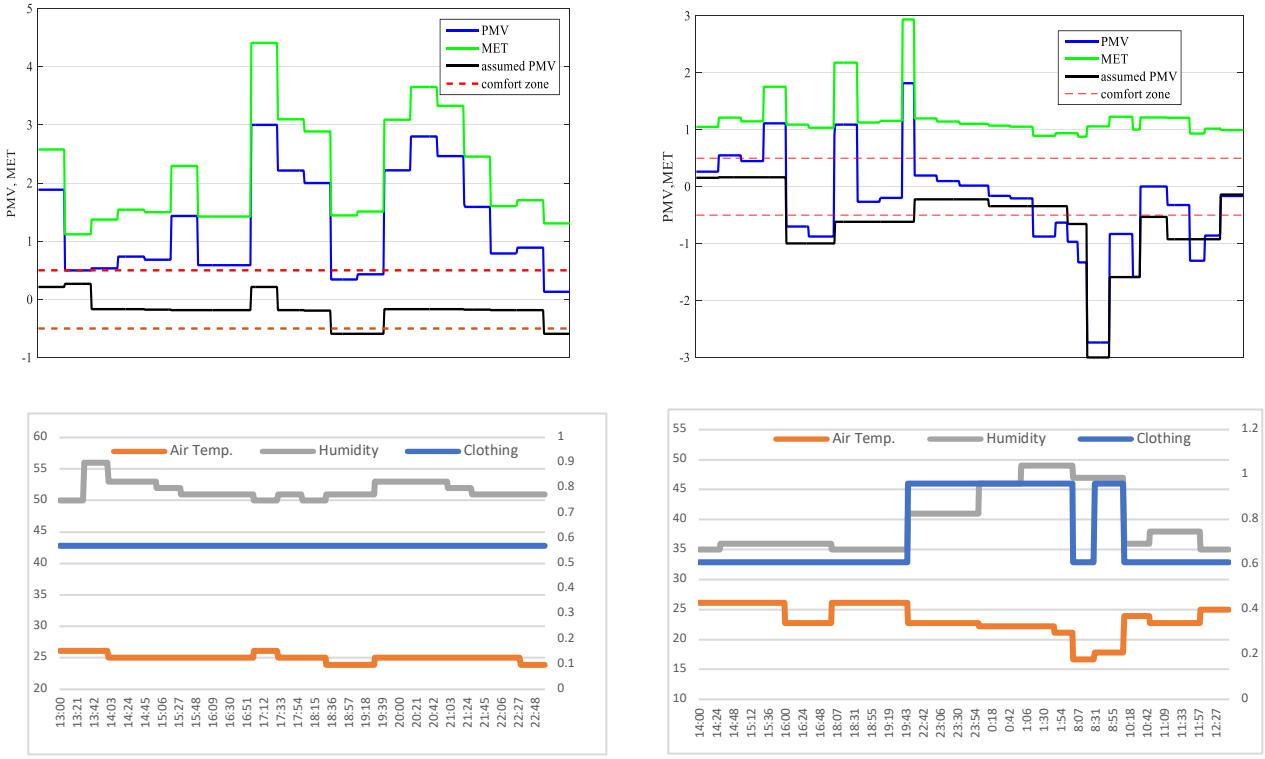
In this section, we investigate the effect of metabolism on the comfort level for building occupants while performing normal life activates. As a proof of concept, a simple experiment was conducted on a 22-year-old and 35-year-old male graduate students for more than a half day. These students were asked to carry a wireless HOBO MA1101 data logger to record their indoor environmental conditions (ambient temperature and humidity) while wearing a Fitbit® device to monitor their heart rate, activity level, and rate of caloric consumption per minute. Moreover, the students were asked to mark their clothing status through a smart phone application. The Fitbit wearable device data, the HOBO data, and the clothing status were all joined using a python-based application. These data were used to determine the students PMV values every minute and then averaged for each 30 minutes. In this experiment, the two students were asked to perform similar normal life activities while working at

office or at home. More experiments are planned to involve bigger human subject experiment size and to compare accuracy between different wearable devices.

3.3.2. Results

Figure 3-8 (a) and (b) show plots for the measured MET values using the Fitbit® device along with the corresponding PMV value and the assumed PMV value (i.e., using a constant metabolic rate of 1.0 MET (Hang-yat and Wang 2013; Ku et al. 2015; Xue, Zhai, and Chen 2013; Gao and Keshav 2013)). These plots show that the MET value keeps changing throughout the entire day. For example, in Figure 3-7 (a), for the younger student, the metabolic rate was consistently over 1.0 MET throughout the entire day, the lowest being 1.09 around 1:00 pm. Even at its lowest value, the metabolic rate was higher than the assumed value of 1.0 MET. The figure also shows a very large increase in the MET value, and consequently PMV value, during the student's study-hours at home between (3:00 pm – 5:00 pm) and (7:00 pm – 9:00pm), while it decreases during the student's relax hours between (5:00 pm and 7:00 pm). Throughout most of the studying period, the student was thermally dissatisfied, feeling hot, with an average PMV value of 2.5, while the assumption is that he should be comfortable with a PMV value of less than -0.2 (assuming a constant MET value of 1.0). Even though the student was not involved in physical labor, the plots show that mental work and simply standing and walking around seem to increase the MET value to an average of greater than 3.0. From our earlier findings about the PMV model's high sensitivity to metabolism, as shown previously in Figures 4 and 5,

this should explain the very large error between the assumed and actual PMV calculation.



(a) 22-year-old male student (b) 35-year-old male student

Figure 3-8 MET and PMV values, clothing, and indoor environmental conditions recorded for two students. The figure shows a big difference in the PMV value when the actual metabolism is used in the calculation instead of the assumed value of 1.0 MET. Also this difference is shown to be higher for the younger student. Data while students were not inside a building are ignored.

Figure 3-8 (b) shows similar higher metabolic rate during daily normal activities for the 34-year-old student. These activates include typing (office) and eating (home). The figure shows that the MET value was sometimes less than 1.0 at complete rest (sleeping). Figure 3-8 confirms that not only the MET value may vary over time for an occupant, but may also vary among occupants preforming similar daily activates. Hence a single constant value, such 1 MET or any other value, can't be used to represent the metabolic value in the PMV model.

3.4. CONCLUSION

The PMV value of an occupant is calculated based on multiple thermal environment and personal parameters. While it is not a perfect model, the best case should provide an 80% accuracy level, assuming that all input parameters are accurately measured. In this study, the quantitative sensitivity of the PMV value to its parameters is defined. We show that the effect of the personal parameters far exceeds that of the environmental factors within the normal parameter ranges. The PMV model simulation analysis performed in this work can help to prioritize measurements with the highest sensitivity. For example, the relative humidity has low thermal comfort sensitivity expect if it exceeds an extreme range (less than 30% and more than 60% (Balaras, Dascalaki, and Gaglia 2007; Wolkoff and Kjærgaard 2007). Hence, humidity is rarely in need of monitoring, whereas metabolism sensitivity is much higher and needs to be closely and continuously measured to assure reliability of the PMV comfort model. The metabolic rate is continuously changing over time, even without performing any

notable physical activities. For example, this paper shows that simple mental activities and doing regular tasks might lead to some increase in the MET value, which can lead to thermal discomfort.

The metabolic rate, which is assumed to be constant in real building throughout most of the literature, is now easy to estimate using wearable devices. We have shown, in our case study using a commercial wearable device, that a person who is not fully at rest will potentially have high MET value and may experience discomfort during the majority of working hours. The high sensitivity to the metabolic rate was apparent in our case study plots as the deviation between the estimated and actual PMV value is large. This shows the importance of measuring the metabolic rate in buildings to assure the reliability of the PMV model. Accurately measuring the metabolic rate of an occupant can extend the area of application of the PMV model to those who might be engaged in physical activities, such as waiters and waitresses in a restaurant, or people working out in gyms. And while it is practically impossible to make everyone comfortable, taking into account the individual differences in peoples' metabolism can help achieving the 80% true satisfaction rate that is expected from the PMV model.

CHAPTER 4

DATA DRIVEN THERMAL COMFORT MODEL

While a lot of work has been done on using wearable devices and smart phone sensor data to predict human activity type and level(Hasan, Alsaleem, and Rafaie 2016), very few have explored the use of these data to infer individual human comfort such as (Hang-yat and Wang 2013). The work of (Hang-yat and Wang 2013) tried to build a single generalized/global model to determine individual human comfort. In this model, the off-line sensing data (i.e., features such as humidity, temperature, and heart rate) and direct comfort vote/feedback from multiple users were used to build a single model that maps a new user sensing data to a prediction of his current comfort level. These works, did not address the sensitivity of each feature, the effect of the output class size and the machine learning tuning parameters on the thermal comfort model accuracy. In this work, however, questions such as what are the best features or the best machine learning method and parameter to model human comfort will be addressed.

In this study, we utilized dataset collected in the comprehensive sensory data collecting study. As it is described in chapter 2.2.2, a mobile application was developed to connect a Microsoft Band 2 and receive the bio-information data. An Onset HOBO device gathered the environmental information. In the final step, a python application

managed the merging the data to create the dataset. As a result, there are five datasets hold personal information and a dataset including all the data called General dataset here.

4.1. GENERAL THERMAL COMFORT MODEL

The thermal comfort models will be created using a machine learning method that will use several variables to predict whether or not an individual is comfortable. In order to do this accurately, several things must be considered. It must be determined which of the available variables have the largest impact on thermal comfort. Additionally, several machine learning systems can be used as the basis of the model. Some of these will be more accurate than others. Thus, it must also be determined which machine learning system will be incorporated. The goal of this experiment is to determine which combination of these (i.e. model type and feature list) gives the most accurate prediction of thermal comfort.

4.1.1. Data

Voting was initially conducted on the building occupants on a scale from -6 to 6 with 6 being very hot, -6 being very cold, and 0 being comfortable. While this scale offers a large amount of variation, it was determined that people struggle to distinguish between minimal differences on this scale such as that between 5 and 6 introducing unnecessary human error. For this reason, alternate scales were created. Table 4-1

and Table 4-2 define these alternate scales. The methodology for converting from the scale on which the individuals initially voted to the three and five group scales was simple. For the three-group scale, values from +/- 6 to +/- 2 were classified as +/- 1 respectively, while values from -1 to 1 were classified as 0. For the five group scale, +/- 6 to +/- 4 were classified as +/- 2 respectively, +/- 3 and +/- 2 were classified as +/- 1 respectively, and -1 through 1 were classified as 0.

Table 4-1 Three Group definition

	Situation	Scaled vote	Vote Range
1	Cold	-1	-6 to -2
2	Normal	0	-1 to 1
3	Hot	+1	2 to 6

Table 4-2 Five Group definition

	Situation	Scaled vote	Vote Range
1	Very Cold	-2	-6 to -5
2	Cold	-1	-4 to -2
3	Normal	0	-1 to 1
4	Hot	+1	2 to 4
5	Very Hot	+2	5 to 6

Initially, six variables were considered to be important for predicting thermal comfort of the participants. The occupant's skin temperature, metabolism, and heart rate were

all measured using the occupant's wearable Microsoft Smart Band 2™. The room temperature and room relative humidity were determined by the HOBO MA1101 data logger. Finally, the CLO value was determined using a smart phone prompt as shown. It was the goal to determine which of these variables can be used to most accurately predict thermal comfort.

4.1.2. Experiment

In order to determine which features have the largest impact on thermal comfort, these features were separated into 45 feature groups (lists) as shown in Figure 4-1. These groups consist of groupings of these features. When creating these groups, it was required that all groups have a minimum of one piece of external data (temperature and relative humidity), and one piece of wearable data (clothing score, heart rate, metabolism, and skin temperature). The significance of creating these groups was to allow individual variables to be separated from one another and for their individual effects to be studied. The features that were included in each list are represented by the circles. For example, feature 01 consists of both the clothing score and room humidity.

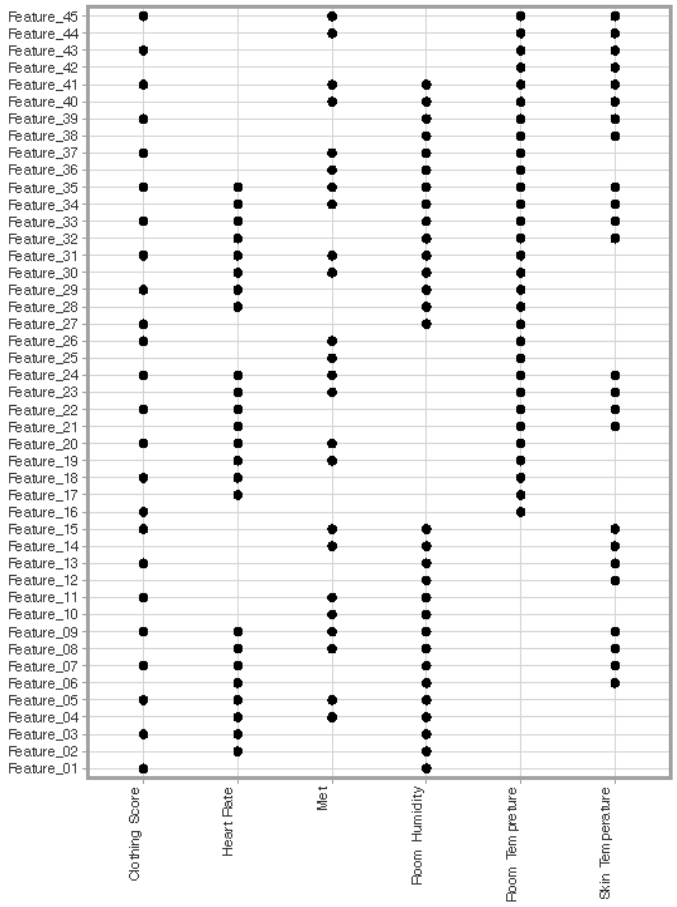


Figure 4-1 The definition of feature list based on different sensor type

Several machine-learning systems were also considered for creating the model to predict thermal comfort. In addition to the machine learning system, there are different ways in which these systems can be utilized. This is represented by using different C values and N numbers. The different combinations of the machine learning methods and their utilization are described in Figure 4-2. The scikit-learn package has been used to generate the models results in this work. The scikit-learn is a Python based program, built on top of SciPy and distributed under the 3-Clause BSD license

16. The Holland Super Computing Center at State University of Nebraska was used to carry out the heavy calculation needed in this investigation.

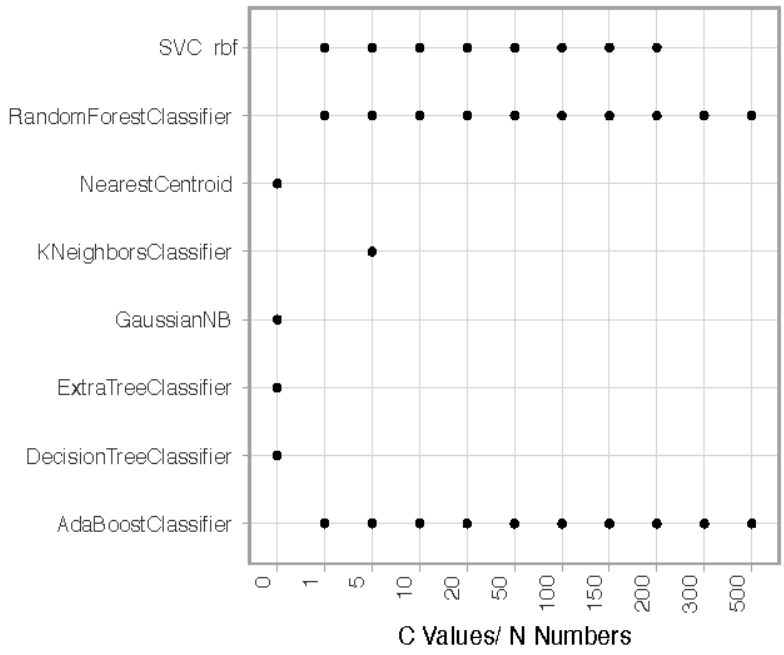


Figure 4-2 The different machine learning methods with their parameters used in this investigation

In order to determine the most accurate model, several models were created using every machine learning system and utilization shown in Figure 4-2. For every one of these models, 45 (the number of feature lists) thermal comfort predictions were made. Each of these predictions was also compared to both the three group and five group thermal comfort votes.

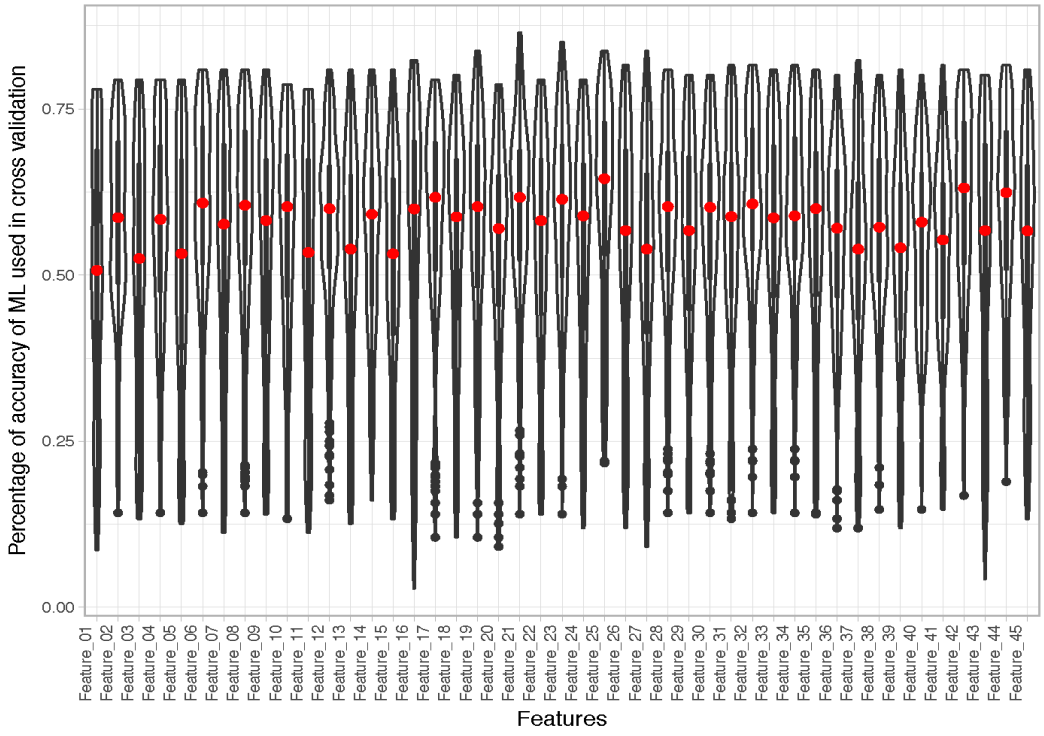


Figure 4-3 The different ML accuracy as per feature list while considering all other variations such as the different machine learning models and their parameters and the different output class size

Figure 4-3 shows the accuracy of the systems for each feature list. To generate one bar in in this figure (model accuracy at given feature list), 85 jobs were created. Each job consumes 6 Core CPU, each of 2GB Ram. The top of the black line represents the most accurate combination for a given feature, while the bottom represents the least accurate combination. The dots toward the middle represent the median accuracy for a given feature list.

Figure 4-4 represents a similar concept, however this time showing the accuracy of particular machine learning methods using only the best feature lists in Figure 5. The dots represent the median accuracy while the black lines extending upwards and downwards represent the maximum and minimum accuracy seen for a given machine learning system.

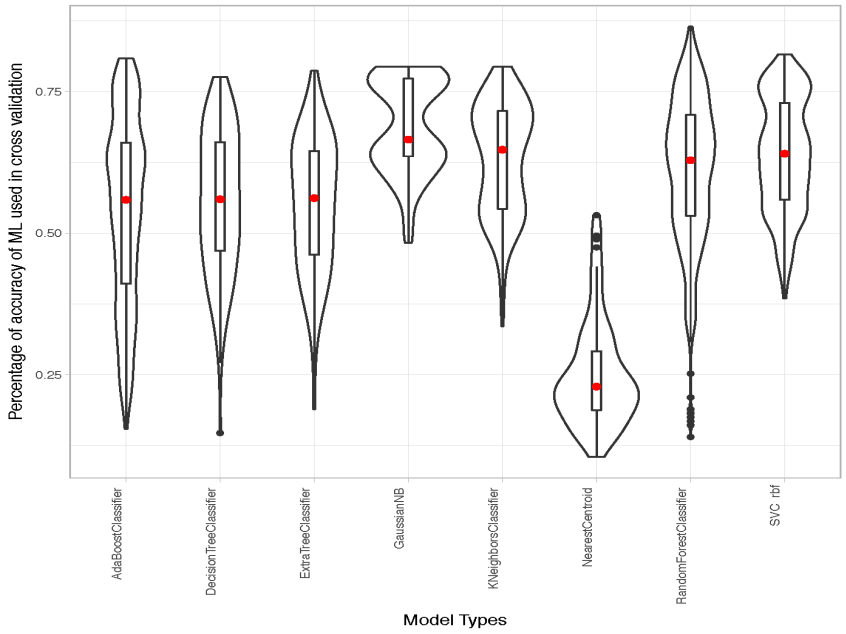


Figure 4-4 The Accuracy of the best ML model for the best feature lists

4.1.3. Results

Table 4-3 shows the feature lists with the highest median accuracy. 7 of the 8 most accurate feature lists include room temperature and 5 of the 8 most accurate feature lists include skin temperature. Meanwhile, 0 of the 14 most accurate feature lists include CLO and 0 of the 6 most accurate feature lists include relative humidity.

Therefore, it seems likely that using both room and skin temperatures to predict thermal comfort will give a relatively accurate thermal comfort prediction, while using the clothing insulation and relative humidity will result in a less accurate prediction. As we believe clothing is an important factor in human comfort, one can argue that skin temperature might have better representation for that factor compared to the individuals self-reporting of their clothing status. The results for the heart rate and metabolism are less conclusive. 10 of the 15 most accurate feature lists include the heart rate; however, this variable is not seen in any of the top 3 most accurate feature lists. Additionally, 2 of the 3 most accurate feature lists include metabolism, but only 8 of the top 15 most accurate feature lists include this variable. While less conclusive, it seems using metabolism and heart rate is relatively accurate at predicting thermal comfort.

Table 4-3 Feature lists with the highest median accuracy considering all other variations (class size, machine model type)

	Feature ID	Median ML Accuracy
1	Feature_25	0.65
2	Feature_42	0.63
3	Feature_44	0.62
4	Feature_17	0.62
5	Feature_21	0.62
6	Feature_23	0.61
7	Feature_06	0.61

8	Feature_32	0.61
9	Feature_08	0.60
10	Feature_10	0.60
11	Feature_19	0.60
12	Feature_28	0.60
13	Feature_30	0.60
14	Feature_12	0.60
15	Feature_35	0.60

Table 4-4 shows that the most accurate machine learning system is the GaussianNB. In addition, it can also be seen on Figure 6 that this system has the least variation in accuracy. Therefore, this suggests that GaussianNB is the machine learning system that can most accurately predict thermal comfort.

Table 4-4 Median of ML accuracy for the best features list considering all other variations

	ML Type	Median ML Accuracy
1	Gaussian NB	0.67
2	KNeighbors Classifier	0.65
3	SVC RBF	0.64
4	Random Forest Classifier	0.63
5	Extra Tree Classifier	0.56
6	Decision Tree Classifier	0.56
7	Ada Boost Classifier	0.56
8	Nearest Centroid	0.23

Table 4-5 and Table 4-6 show different machine learning method accuracy when the effects of the best features list and machine learning systems are considered. Table 4 is for the group of three voting scale while Table 5 is for the group of five voting scale. It is worth noting that the models are significantly more accurate when compared to the group of three scales than when compared to the group of five scales.

Table 4-5 Median of a specific ML accuracy for the best feature lists considering the three classes case

	ML Type	C/N	Features	Accuracy
1	SVC RBF	1	Feature_32	0.80
2	SVC RBF	1	Feature_35	0.79
3	SVC RBF	1	Feature_08	0.79
4	SVC RBF	1	Feature_06	0.78
5	SVC RBF	1	Feature_28	0.78
6	SVC RBF	1	Feature_30	0.78
7	Ada Boost Classifier	1	Feature_10	0.78
8	Gaussian NB	0	Feature_17	0.78
9	SVC RBF	1	Feature_42	0.78
10	Ada Boost Classifier	1	Feature_06	0.78
11	Ada Boost Classifier	1	Feature_08	0.78
12	Ada Boost Classifier	1	Feature_12	0.78
13	Ada Boost Classifier	5	Feature_06	0.78
14	Ada Boost Classifier	5	Feature_08	0.78
15	Random Forest Classifier	100	Feature_23	0.78

Table 4-6 Median of a specific ML accuracy for the best feature list considering the five classes case.

	ML Type	C/N	Features	Accuracy
1	SVC RBF	1	Feature_32	0.80
2	Gaussian NB	0	Feature_42	0.79
3	Gaussian NB	0	Feature_12	0.79
4	Ada Boost Classifier	5	Feature_12	0.78
5	SVC RBF	1	Feature_21	0.78
6	Ada Boost Classifier	1	Feature_06	0.78
7	Ada Boost Classifier	1	Feature_12	0.78
8	Gaussian NB	1	Feature_19	0.78
9	SVC RBF	1	Feature_44	0.78
10	Ada Boost Classifier	1	Feature_08	0.78
11	Ada Boost Classifier	1	Feature_10	0.78
12	Ada Boost Classifier	1	Feature_17	0.78
13	Ada Boost Classifier	5	Feature_19	0.78
14	Ada Boost Classifier	5	Feature_21	0.78
15	Ada Boost Classifier	1	Feature_23	0.78

Despite the observations above, it appears that it is not necessarily true that the combination of the GuassianNB machine learning system and the room temperature and skin temperature variables will give the most accurate prediction of thermal comfort. When the effects are combined, the SVC rbf machine learning method with

C/N=1 appears to be the most accurate, particularly with feature 32. This feature consists of the heart rate, room humidity, room temperature, and skin temperature variables.

4.1.4. Conclusion

The most accurate thermal comfort prediction made by any of the potential models was 80%. This is accurate enough and could potentially be a useful model for predicting thermal comfort in the future compared to the PMV model that requires many expensive measurements. This model was created using the SVC rbf machine learning system with a C/N value of 1, and considering the heart rate, room humidity, room temperature, and skin temperature. This combination was the most accurate using both the three group and five group voting scales. However, the 80% accuracy was for the more limited three-group scale. While this was still the most accurate combination, the accuracy using the five-group scale was only 66%.

We believe that the comfort model accuracy were capped by the data noise in comfort votes and sensing data. These data noises can occur as a voter may give different votes at the same measured features due to reporting error such as entering a wrong vote or because it is difficult to distinguish a comfort level 1 from 2 as an example, sensor noise, or illness. Filtering these data noise should improve the general model accuracy. Moreover, initial results show that an accuracy above 95% can be achieved for some of the individual, in this thermal comfort experiment, if a personalized model

is used rather than using a one single general model. Details of this analysis will be presented in a separate work.

4.2. PERSONAL THERMAL COMFORT MODEL

In this study, we just employed the personal dataset to design a new personal thermal comfort model and exercise the different machine learning methods to review the accuracy and the impact of the feature on the final models.

4.2.1. Data

In order to determine an experimental thermal comfort, five individuals were invited to take part in this study. These individuals were periodically prompted to vote on their thermal comfort throughout the day. This prompt came through their smart phones. Voting was initially conducted on the building occupants on a scale from -6 to 6 with 6 being very hot, -6 being very cold, and 0 being comfortable. While this scale offers a large amount of variation, it was determined that people struggle to distinguish between minimal differences on this scale such as that between 5 and 6 introducing unnecessary human error. For this reason, alternate scales were created.

Table 4-7 defines these alternate scales. In this scale, the values from +/- 6 to +/- 2 were classified as +/- 1 respectively, while values from -1 to 1 were classified as 0.

Table 4-7 Three Group Definition Vote

	Situation	Scaled vote	Vote Range
1	Hot	+1	-6 to -2
2	Normal	0	-1 to 1
3	Cold	-1	2 to 6

The outcome of the experiment is the collection of data files generated by the Comfort Vote application and Hobo. The export feature in the Comfort application ease the sharing the data to the server while for the Hobo, it's a manual process and we used the HOBO mobile® to extract information.

Automate data sampling feature in Comfort Vote generates a significant amount of unlabeled data - not including the user vote - and there is no need to use them in the modeling process. On the other hand, the data does not include the metabolism, and it needs to measure out from used-calories. These requirements are satisfied by employing a Python-based program called pre-process which is also merging the related data files for every person in a file. It also removes noisy data considering the possible values and finally scales the vote appended as a new field.

Three data set is chosen randomly, and a new dataset including whole data from these participants formed a new dataset called General dataset. Table 4-8 shows the

general information of the participants and datasets chosen for studying the Personal Comfort Modeling.

Table 4-8 Participant data

	Dataset	Gender	Age	Data size
1	Person 1	Male	20	54
2	Person 2	Male	24	91
3	Person 3	Female	21	143
4	General	-	-	286

4.2.2. Results

This section discusses the performances of different machine learning algorithms applied and the results obtained. Moreover, the relevance of some features in determining an accurate comfort model of an individual will be shown. It's has been found that human skin temperature of the person is the most salient feature that defines the thermal comfort level of the person. While doing this work, it has been observed that the galvanic skin resistance(GSR), also referred skin conductance, plays a vital role in accuracy of both the general and private models. However, it should also be noted that skin conductance change is an overall reaction in a subconscious level as a result of many human cognitive and emotional states, not only a manifestation of thermal comfort changes.

Figure 4-5 Decision Tree performance

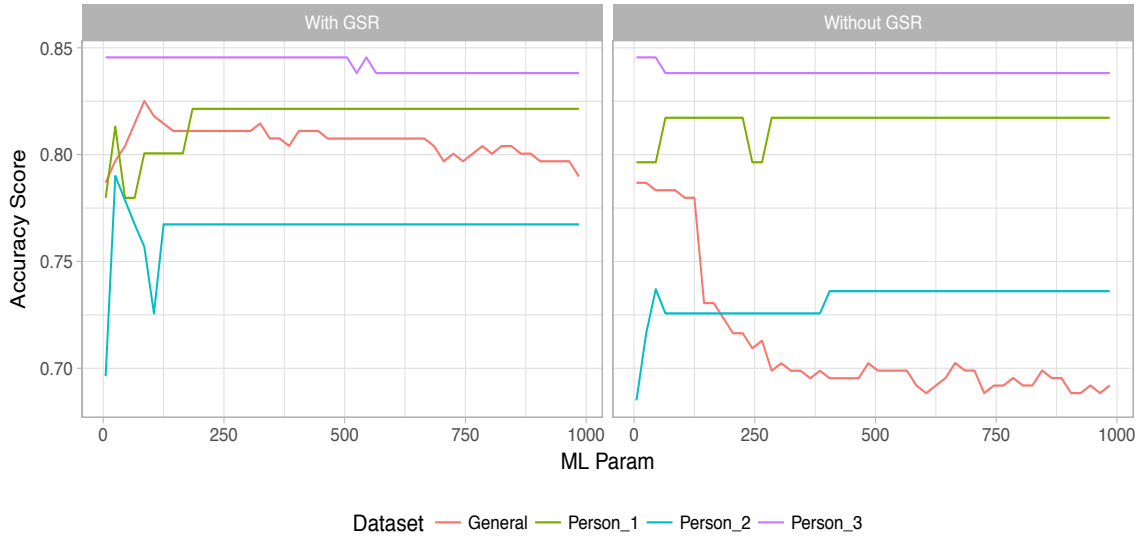


Figure 4-6 AdaBoost classifier performance

Feature data for skin temperature, skin conductance, room temperature, metabolism rate, clothing, heart rate and room humidity were collected and employed to build

general and private comfort models. To study the significance of the skin conductance, models with and without it have been developed and presented. In this work five different machine learning algorithms, namely Decision Tree, AdaBoost classifier, Gradient Boosting Classifier, Random Forest Classifier and Support Vector Machine-RBF, have been used to build the models.

Figure 4-7 Gradient Boosting classifier performance

As it could be observed from the figures above, in almost all cases, the models that were built by taking skin conductance into consideration are more accurate. In the same fashion it is evident that the private models happen to be a more representative of the comfort level of an individual compared to the general models. The random forest classifier exhibits the best accuracy of about 88% compared to other machine learning algorithms.

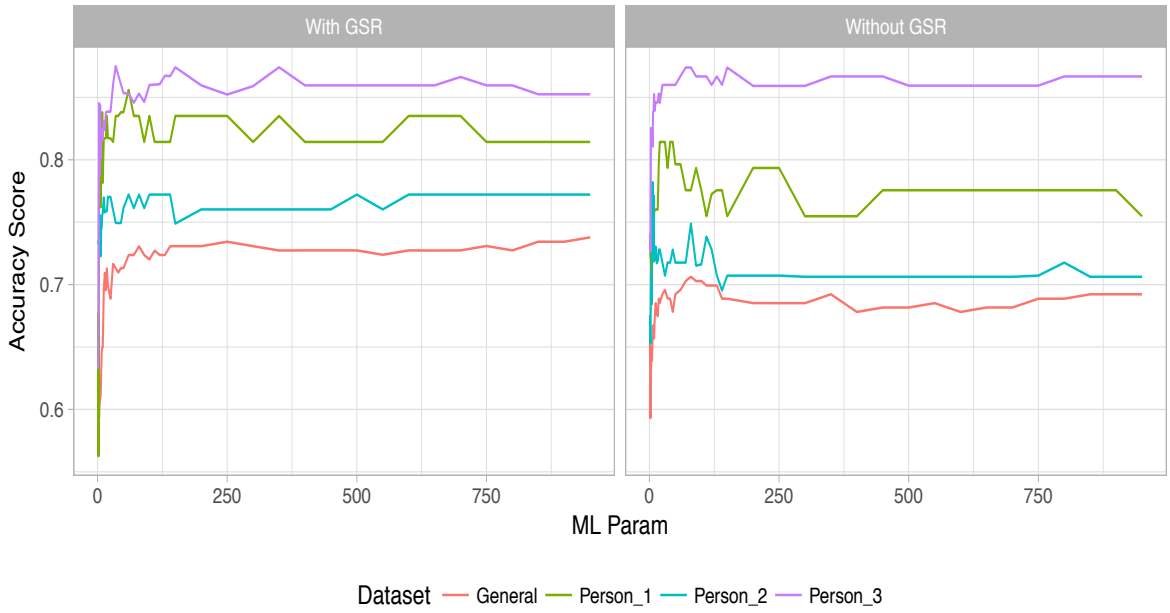


Figure 4-8 Random Forest classifier performance

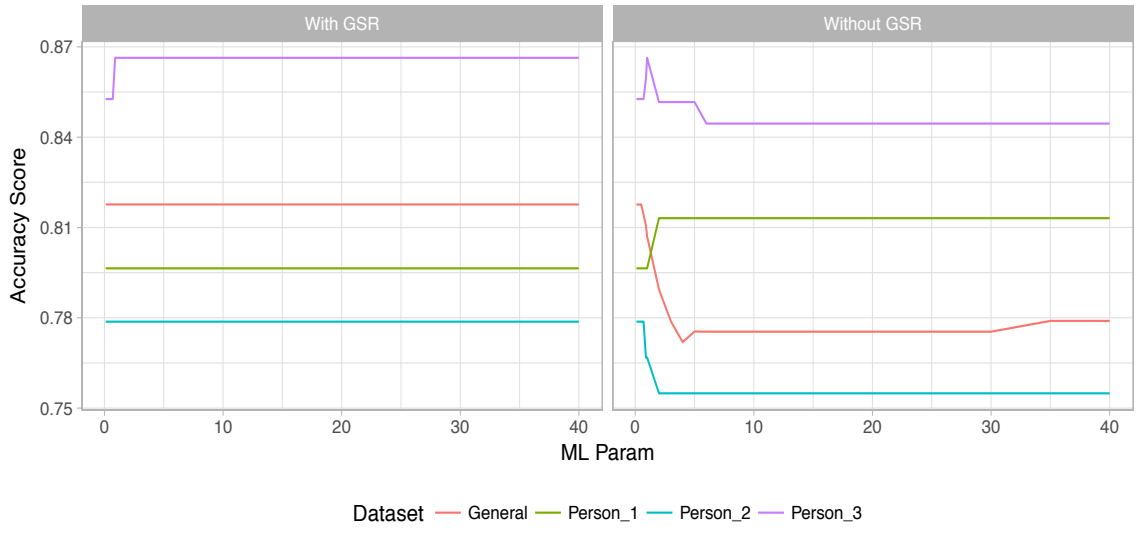


Figure 4-9 Support Vector Machine with RBF kernel performance

Table 4-9 Details of the performance review of Machine Learning for all the datasets

Dataset	Machine Learning Type	Include GSR	Max Accuracy	Min Accuracy	Mean Accuracy
General	AdaBoost classifier	Yes	0.8251	0.7868	0.8060
		No	0.7868	0.6883	0.7111
	Decision Tree	Yes	0.8212	0.5964	0.6550
		No	0.8176	0.5898	0.6113
	Gradient Boosting Classifier	Yes	0.7021	0.6280	0.6560
		No	0.6324	0.5334	0.5444
	Random Forest Classifier	Yes	0.7379	0.5650	0.6994
		No	0.7063	0.5931	0.6804
	Support Vector Machines - RBF	Yes	0.8176	0.8176	0.8176
		No	0.8176	0.7719	0.7838
Person 1	AdaBoost classifier	Yes	0.8214	0.7798	0.8167
		No	0.8173	0.7964	0.8152
	Decision Tree	Yes	0.7976	0.7202	0.7259
		No	0.7964	0.6643	0.6747
	Gradient Boosting Classifier	Yes	0.8393	0.8185	0.8385
		No	0.7768	0.7560	0.7685

	Random Forest Classifier	Yes	0.8560	0.5625	0.8130
		No	0.8143	0.6839	0.7683
	Support Vector Machines - RBF	Yes	0.7964	0.7964	0.7964
		No	0.8131	0.7964	0.8102
Person 2	AdaBoost classifier	Yes	0.7901	0.6963	0.7656
		No	0.7370	0.6850	0.7312
	Decision Tree	Yes	0.7475	0.6514	0.6561
		No	0.7475	0.6275	0.6416
	Gradient Boosting Classifier	Yes	0.7361	0.7044	0.7058
		No	0.6726	0.6606	0.6624
	Random Forest Classifier	Yes	0.7721	0.7227	0.7599
		No	0.7821	0.6528	0.7153
	Support Vector Machines - RBF	Yes	0.7787	0.7787	0.7787
		No	0.7787	0.7549	0.7583
Person 3	AdaBoost classifier	Yes	0.8455	0.8382	0.8421
		No	0.8455	0.8382	0.8386
	Decision Tree	Yes	0.8320	0.7804	0.8211
		No	0.8459	0.8244	0.8436
	Gradient Boosting Classifier	Yes	0.8586	0.7881	0.8295
		No	0.8447	0.7959	0.8230
		Yes	0.8751	0.6326	0.8410

Random Forest Classifier	No	0.8739	0.7269	0.8510
Support Vector Machines - RBF	Yes	0.8664	0.8527	0.8648
	No	0.8664	0.8445	0.8473

4.2.3. CONCLUSIONS

An improved private comfort model has been developed from biometric data gathered via wearable devices. Apart from skin temperature, skin conductance has been introduced and it has been observed that it is an important feature in creating a private comfort model. Moreover, the difference in the three private models of individuals presented in this work shows that it is hard to have a general model representative of everyone's comfort state, despite their common physical location. Limited data size contributes to the inaccuracies of the models.

CHAPTER 5

INTEGRATION

The work in previous chapters focused on the design of new thermal comfort models for the occupants of buildings. However, the practical implementation and feasibility of these new models, when deployed in a real world, remain unclear and demand a thorough investigation. This chapter focuses on the idea of a smart home services with human-in-the-loop (HITL) model where human interactions are leveraged ,along with machine intelligence, to create efficient machine learning models by proposing a new architecture. To demonstrate this architecture and evaluate the challenges and performance of the design, an experiment has been conducted on an efficient and inexpensive platform. Finally, an innovative idea is presented to reconstruct the temperature sensor, installed onsite, to obtain a precise value of local temperature.

5.1. ARCHITECTURE

The newly proposed home automation architecture, depicted in Figure 5-1, have the human-in-the-loop functionality and utilizes the new thermal model to control the HVAC system in homes. Apart from the typical sensing devices used to gather environmental information, this design requires the use of smart devices such as smart bands and smartwatches to collect the bio-info or feedback from the resident. In this design, it has been assumed that the core of the smart home is in the building. In

other words, even in the absence of internet service, the smart home would conveniently function as intended.

The service core includes two main modules. The first one defines the interface to all components such as sensing devices and home controls and is responsible for acquiring data and interacting with these devices. The second module utilizes the data collected from the first module and updates the model accordingly to improve the performance of the home control services.

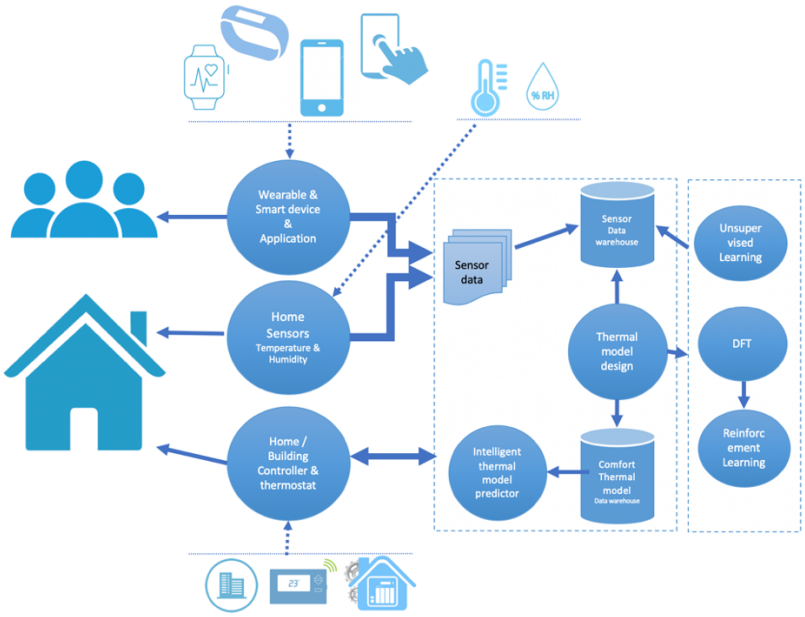


Figure 5-1 Home Automation architecture

5.1.1. Experimental smart home

The goal of the design of this platform is to incorporate open source software and customizable hardware devices to create an inexpensive and easy-to-use platform that can be extended to support new features and functionalities in the future. While doing so, licensing has been a concern, though it was resolved by adopting the software with the Apache License, Version 2.0 or MIT license.

5.1.1.1. ***Home Assistant***

Among different open source solutions available in the market, the Home Assistant platform has been chosen as it was found complementary with the requirements of our project design as an open-source home automation that puts local control and privacy first ("Home Assistant," n.d.). In version 0.82.1, it supports 1220 components varying from different devices like cameras, lights and smart-bands to cloud platforms like Alexa or Microsoft face identity. Most of the modules can be attached to the platform by easily updating a configuration file since it has capability that supports the extension for any new customization. Additionally, it also can be installed in any Linux distributions. **Error! Reference source not found.** illustrates the main interface, designed as the web page. The home assistant can be ran on desktops, tablets and a phones, as shown in the figure.

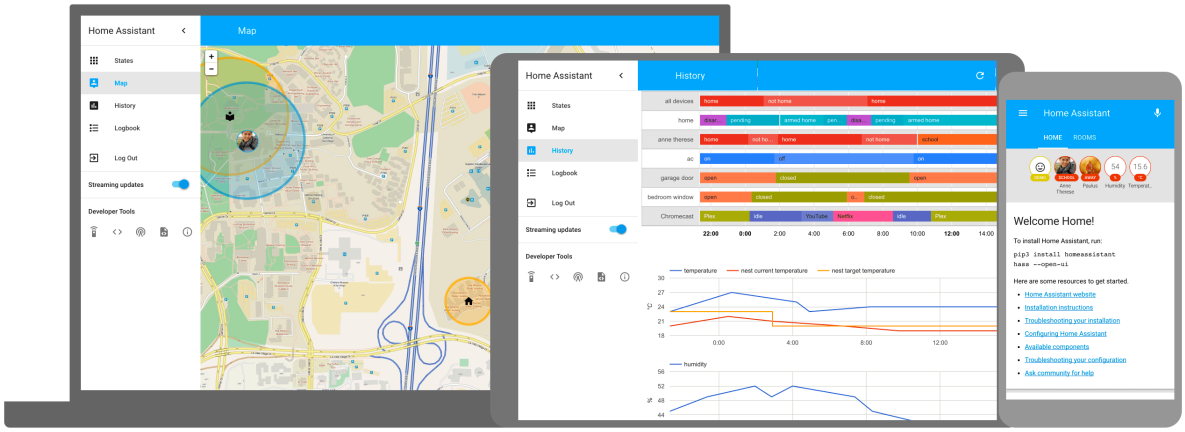
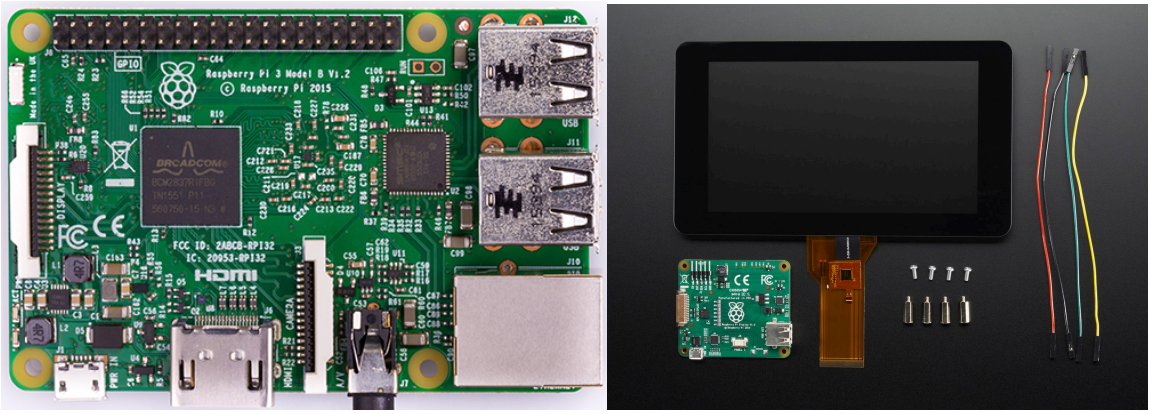


Figure 5-2 Home Assistant - screenshots from web interface

("Home Assistant," n.d.)

5.1.2. Hardware

As recommended by Home Assistant platform, for smoother operation purposes, we used a Raspberry Pi 3 Model B to install the service. In our experiment, we tried to make the platform standalone and user-friendly; hence, a touch screen monitor has been connected the Raspberry Pi for interface. **Error! Reference source not found.**(a) shows the Raspberry Pi 3 B and **Error! Reference source not found.**(b) present all the hardware related to the touchscreen.



(a) (b)

Figure 5-3 Home automation core hardwares

5.1.3. Sensory devices

Prior to the selection of the sensory devices, an investigation of the various protocols was conducted to identify those that support sensor integration and that are capable of working without any internet and Wi-Fi connection. Because of the specifications listed below, the Z-Wave protocol has been picked for this work.

- A source-routed mesh network architecture
- Primarily designed for home automation.
- Operate in power-save mode, so reducing power consumption

Based on the excellent reputation of the company that manufactures it and its ease of integration with Z-Wave technology, among other factors, the AeroTech sensory

devices were chosen for this experiment. The equipment and their descriptions are presented below.

Figure 5-4 AeroTech Sensory devices

- **Z-Stick Gen5**: A controller installed via a USB port to the Raspberry Pi, which connects all the devise in Z-Wave mesh to the raspberry Pi.
- **MultiSensor 6**: A device containing six different sensors (motion, humidity, temperature, light lux, UV, vibration sensor).
- **Smart Switch 6**: A switch that is controllable remotely to turn the connected devices off or on.
- **LED Bulb**: A controllable light that can generate about two million colors.

5.1.4. Evaluation

As a means of evaluating the performance and reliability the platform for this task, a simple installation and configuration setup was implemented in the lab to where environmental information was collected while running some minor control action like turning a device on or off. In order to interact with the platform, we picked the VoiceOver command methods by adopting Alexa service from Amazon and created a new skill set. The skill includes commands like turning on/off a device or requesting information about the temperature in the room. Using "ngrok" helped creating a backward proxy over Home Assistant Rest service. The REST Service was customized for a related command in Alexa skills. This skill sets Alexa to wait to be called from Amazon with a command, to which Alexa would respond to or execute accordingly. As soon as Alexa detects a command, a request is sent to Home Assistant and the configuration is updated to handle the request and run the expected request.

The experiment shows the considerable potential of the service to handle complicated architectures, which include local and remote devices in a secure environment. These architectures are required in the design of efficient smart home services. In a future work, we plan to add a new module to generate personal thermal comfort models and control the system to ensure occupant comfort.

While the outcome of integrating components from different device manufactures and open source firmware developers allows for remarkable for smart home automation

applications, the need for external modules increases the cost and inconvenience of these applications. Communication devices, such as computers and smartphones, have found their ways into consumers' houses and offices. Therefore, it is only intuitive to wonder if these devices can substitute the external modules and sensors that are currently employed to provide home automation.

Current day smart devices incorporate a variety of sensors to provide feedback and ensure sound operation. Due to their existence within ever-changing environments, these sensors inadvertently measure some environmental signals. This provides the possibility of eliminating the need for some external sensing devices by relying on the computing units internal sensors within a smart building. In this thesis, this concept is demonstrated by introducing a virtual temperature sensor, as shown in the next section.

5.2. VIRTUAL TEMPERATURE SENSOR

The advancement in technology miniaturization and the recent surge in electronic devices equipped with built-in sensors that are capable of capturing processes in real-time has become a valuable input for research. Temperature control and monitoring have been the center of most industrial automation processes in industries and buildings. This work investigates a mechanism in which indoor air temperature can be predicted using the battery temperature sensor of laptops or Android smartphones, where this data is easily accessible. This idea is different from the idea of using

crowdsourcing to predict the average outdoor city temperature for meteorological purposes, using smartphones battery sensor.(Overeem et al. 2013)

In recent works (Hasan et al., 2016) and (Rafaie et al., 2017) we have proposed an idea of blending traditional environmental sensing data streams (such as temperature) with newly available wearable sensing information to develop new models to predict thermal comfort of building occupants and to optimize building operations. However, a major challenge in that approach is the requirement for each occupant to carry a wireless temperature sensor. The idea in this thesis is to eliminate that requirement by employing the temperature sensors in smartphones or laptops' batteries (along with other parameters) without directly measuring the temperature in the room. Moreover, this new way of sensing indoor temperature can also be used in optimizing energy efficiency for applications of high-energy demand.

It is obvious that the estimation of indoor temperature from the battery temperature would be influenced by the insulation, intensity of load, quality, and location (pocket or indoor) of the smartphone. This work intends to tackle this bias and noises by working in a higher dimensional feature space. Apart from the temperature of the battery in the cellphone, there are dozens of more features of the laptop/smartphone where online real-time data can be collected and made useful in the model construction process. The indoor air temperature estimation could be more consolidated if features such as CPU internal temperature and other processor

activities are found to have a meaningful correlation with varying indoor air temperature values.

Data of more than 300 features from a running laptop along with the actual ambient air temperature were collected during the learning phase. The data collected is used to build non-linear models that best represents the relationship between the smartphone feature data and the indoor air temperature. In doing so, different machine learning algorithms of high-performance success are employed to develop a dependable generalized model that works fine not only in the same environment and cellphone but also for significantly different data and new user cellphones. It is quite intuitive to assume that different features will have different correlation degrees with the indoor air temperature. Therefore, it is imperative to give more relevance to features that show best correlations with the air temperature.

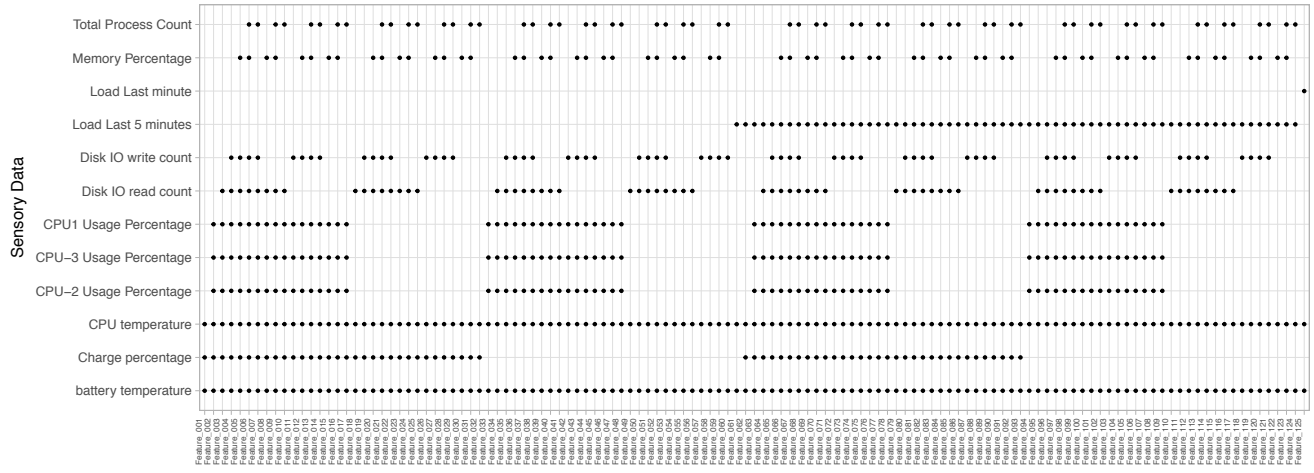
5.2.1. Experiment and Methodology

In this investigation, a new MacBook Pro with a built-in quad-core i7 processor and 16GB of RAM was used. Using tools such as Glances(Hennion, n.d.) and iStats(Naud-Dulude, n.d.), a dataset with about 300 features was collected online from this laptop with high sampling rate. Simultaneously, a Hobo data logger logs and stores the indoor temperature. Based on intuition and prior knowledge, preliminary filtering was made to reduce the number of features to eleven with a belief that these features

would capture the characteristics of the indoor air temperature. Battery temperature, CPU temperature, the percentage of processor usage of three CPUs, load intensity of the last few minutes and the total number of processes running in the system are among the features chosen for the analysis.

To investigate the combination of features that most capture the indoor air temperature, the features were grouped into 255 different possible combinations, **Error! Reference source not found.**. To avoid the omission of the features presumably considered as relevant such as battery temperature, and to minimize the number of possible combinations features that exhibit similar behaviors, we employed the following unification rule:

1) The battery temperature and CPU temperature are mandatory
2) The three processors information is considered as a variable meaning if they included in the feature, the data of all CPUs will be incorporated.



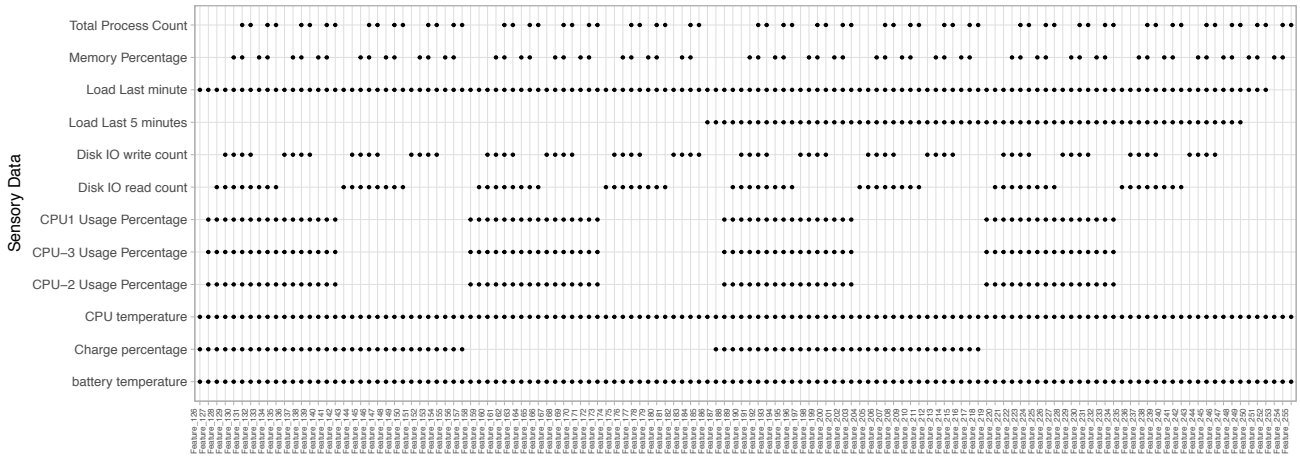


Figure 5-5 The different combinations of the feature list.

Utilizing the Python programming language, an application called trainer is implemented to create the models and evaluate the performance. The Scikit-learn package (Pedregosa et al. 2012) is the main library to facilitate the process of creating the models. The Scikit-Learn is a Python-based program, built on top of SciPy and distributed under the 3-Clause BSD license and includes a vast collection of machine learning methods. Running the trainer application for every dataset was managed by the Holland Supercomputing Center (HCC) at the University of Nebraska. The trainer application utilizes different regression machine learning methods including Epsilon-Support Vector Regression (SVR) with kernel RBF, PLS 2 blocks regression (PLS-Regression), Nu-Support Vector Regression (Nu-SVR) with kernel RBF, K-Nearest Neighbor Regression, and the Randomized Decision Trees regressor (Extra-Trees Regressor) (Ligeza 1995). The combining of the different machine learning methods and applying different initialization variable created 33 unique models. **Error! Reference source not found.** shows the details of the models used in this

experiment. To validate the accuracy of each model and to avoid the overfitting problem, we used the cross-validation with four-folds. This means that the data was partitioned randomly into four different groups. Three data groups are randomly selected as the training dataset and the last part use as the test dataset for evaluating the model accuracy and run this process for four times.

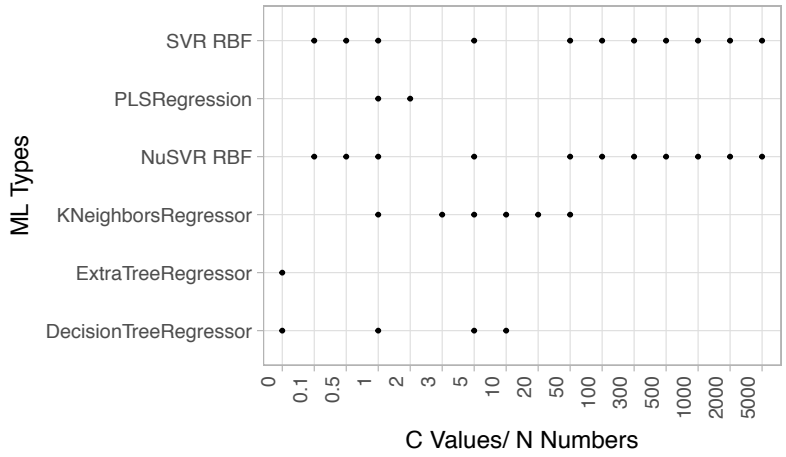


Figure 5-6 Machine learning methods with their parameters used in this investigation.

5.2.2. Results and discussion

Figure 5-7 shows the median accuracy of the ambient temperature prediction for each machine-learning algorithm. The far left of the black line represents the most accurate feature list for a given model, while the far right represents the least accurate feature list. The red dots toward the middle represent the median accuracy of a machine learning given all feature lists. In the figure, the y-axis is the machine learning type

identified with its kernel C value or N numbers, and the x-axis is average prediction error for the indoor temperature in C° (the lower the mean value, the better model accuracy).

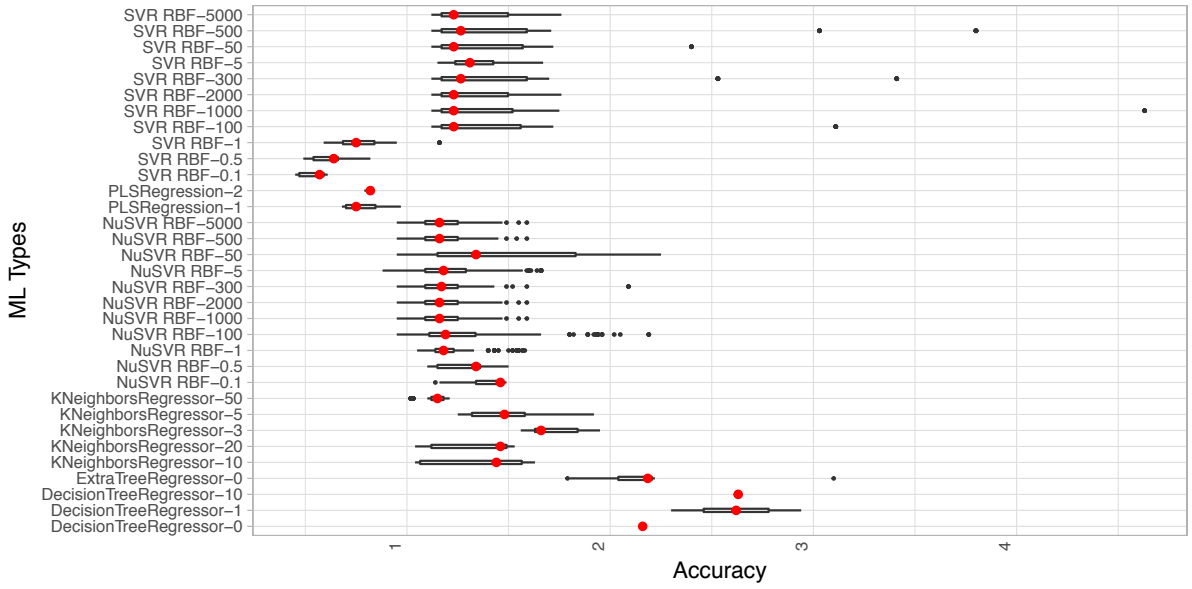


Figure 5-7 The median of accuracy for predicting indoor temperature using different machine learning methods.

Based on the results of **Error! Reference source not found.**, the best ten machine learning methods have been identified and are shown in **Error! Reference source not found.**. It can be seen that PLS regressor and Super Vector Regressor (SVR) with RBF kernel are the ones with superior performances.

Table 5-1 Median of ML accuracy for the best features list considering all other variations

	ML Types	Median ML Accuracy
1	SVR RBF-0.1	0.57
2	SVR RBF-0.5	0.64
3	PLSRegression-1	0.75
4	SVR RBF-1	0.75
5	PLSRegression-2	0.82
6	KNeigborsRegressor-50	1.15
7	NuSVR RBF-1000	1.16
8	NuSVR RBF-2000	1.16
9	NuSVR RBF-500	1.16
10	NuSVR RBF-5000	1.16

Error! Reference source not found., shows the best 40 feature lists that were obtained by averaging the median accuracy from the best machine learning methods listed in **Error! Reference source not found.**. **Error! Reference source not found.** lists the feature (sensor) name for these top feature lists. Finally, **Error! Reference source not found.** summarizes the details of the top 15 feature lists. This includes listing the name of the best machine learning model that provides the best accuracy for that feature, which surprisingly happened to be the same model for all these top feature lists. **Error! Reference source not found.**, visually confirms

the good accuracy of the best three feature lists using the best machine learning model (the first three rows in **Error! Reference source not found.**). The figure shows a great agreement between the actual indoor temperature and the model estimated temperature.

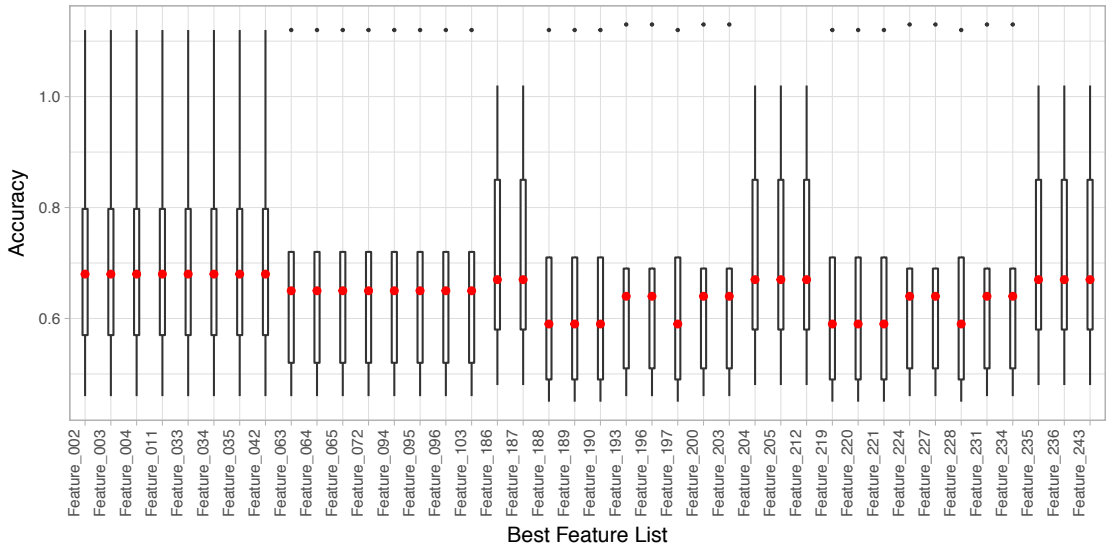


Figure 5-8 The top 40 feature lists obtained by averaging the median accuracy of the top machine learning methods in Table 1.

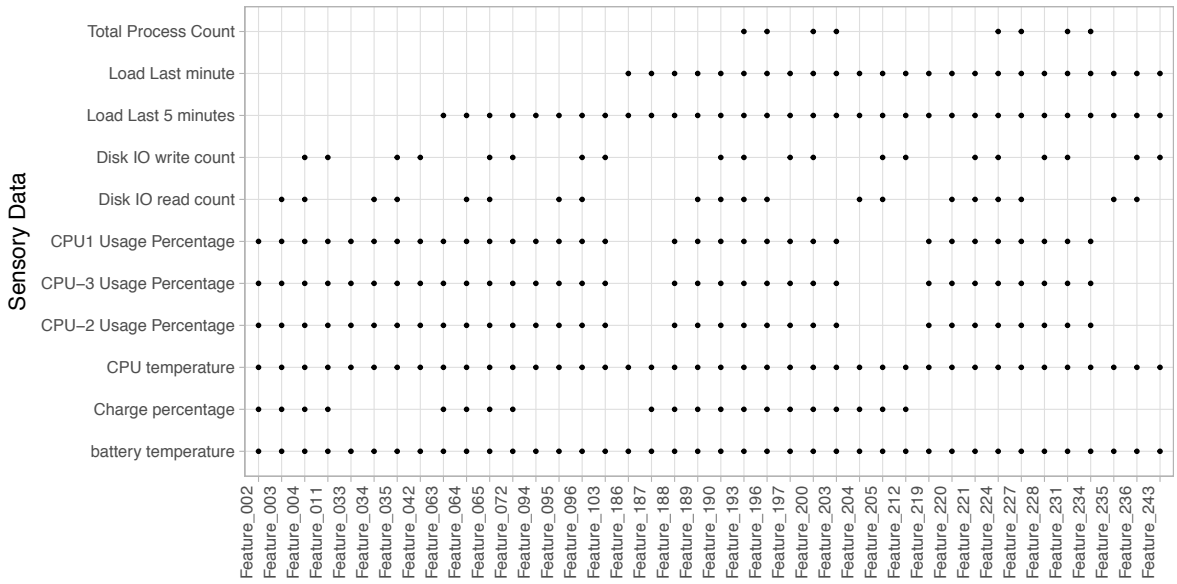


Figure 5-9 The actual sensor name for the top 40 feature lists.

The data in **Error! Reference source not found.** was used for counting the number of times a feature occurs in the top 40 feature lists. This count, shown in **Error! Reference source not found.**, helps to identify features with the highest likelihood of correlation with the actual indoor temperature. For example, one can conclude from **Error! Reference source not found.** that the laptop load intensity of the last 5 minute and the three CPU processing usage, even though they are not temperature measurements, are among the most reliable features that can be used to predict indoor temperature.

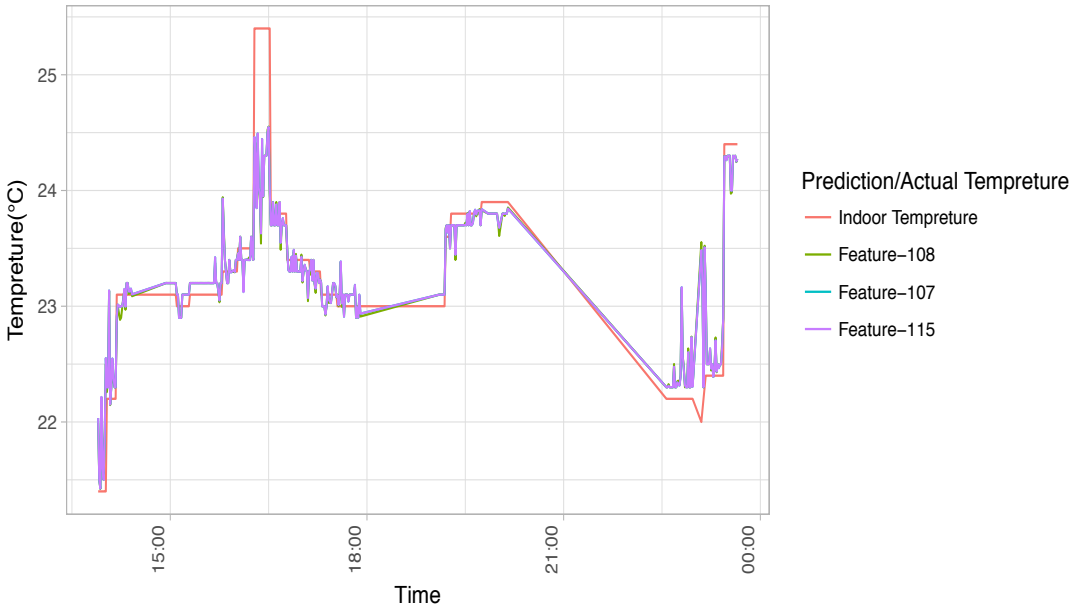


Figure 5-10 Model predicted Vs Actual indoor temperature

Table 5-2 The best machine learning method with the best feature list.

	ML Type	Features	Median ML Accuracy
1	SVR RBF-0.1	Feature_108	0.45
2	SVR RBF-0.1	Feature_107	0.45
3	SVR RBF-0.1	Feature_115	0.45
4	SVR RBF-0.1	Feature_106	0.45
5	SVR RBF-0.1	Feature_139	0.45
6	SVR RBF-0.1	Feature_138	0.45
7	SVR RBF-0.1	Feature_146	0.45

8	SVR RBF-0.1	Feature_137	0.45
9	SVR RBF-0.1	Feature_111	0.46
10	SVR RBF-0.1	Feature_114	0.46
11	SVR RBF-0.1	Feature_118	0.46
12	SVR RBF-0.1	Feature_121	0.46
13	SVR RBF-0.1	Feature_142	0.46
14	SVR RBF-0.1	Feature_145	0.46
15	SVR RBF-0.1	Feature_149	0.46

Table 5-3 Feature appearance frequency rank

	Feature	**Appearance in Feature List**
1	Bat_temp	40
2	Cpu_temp	40
3	Load_min5	32
4	Percpu_1_total	32
5	Percpu_2_total	32
6	Percpu_3_total	32
7	Load_min1	28
8	Charge_per	24
9	Diskio_disk0_read_count	20
10	Diskio_disk0_write_count	20
11	Processcount_total	8

5.2.3. Conclusion

This work identifies the best machine learning algorithms and features of MacBook Pro that could effectively model the indoor air temperature. As a result, features that are not temperature based such as CPU processing usage and the laptop load have been found as salient features in capturing and characterizing the surrounding temperature. This work can be extended for smartphones with the intention of controlling and individual's comfort in a room as opposed to the traditional holistic temperature control approach in buildings. We believe that a more reliable and accurate generalized model that works universally, across most modern electronic computing platforms, could be produced with better noise filtering approaches on raw data and internal sensors of high sensitivity.

REFERENCES

"Accessing the Fitbit API." 2018. Accessed October 1. https://dev.fitbit.com/build/reference/web-api/oauth2/#implicit-grant-flow.

Allen, Joseph G, Piers MacNaughton, Jose Guillermo Cedeno Laurent, Skye S Flanigan, Erika Sita Eitland, and John D Spengler. 2015. "Green Buildings and Health."

Current Environmental Health Reports 2 (3). Springer: 250–58.

Ari, S., P. Wilcoxen, H. E. Khalifa, J. F. Dannenhoffer, and C. Isik. 2008. "A Practical Approach to Individual Thermal Comfort and Energy Optimization Problem." In *Annual Conference of the North American Fuzzy Information Processing Society - NAFIPS*, 1–6. IEEE. doi:10.1109/NAFIPS.2008.4531261.

ASHRAE, ANSI. 2004. "Standard 55-2004, Thermal Environmental Conditions for Human Occupancy, Atlanta: American Society of Heating, Refrigerating, and Air-Conditioning Engineers." *Inc., USA*.

Auffenberg, Frederik, Sebastian Stein, and Alex Rogers. 2015. "A Personalised Thermal Comfort Model Using a Bayesian Network."

Balaras, Constantinos A., Elena Dascalaki, and Athina Gaglia. 2007. "HVAC and Indoor Thermal Conditions in Hospital Operating Rooms." *Energy and Buildings* 39 (4): 454–70. doi:10.1016/j.enbuild.2006.09.004.

Chang, Yin-Wen, Cho-Jui Hsieh, Kai-Wei Chang, Michael Ringgaard, and Chih-Jen Lin. 2010. "Training and Testing Low-Degree Polynomial Data Mappings via Linear SVM." *Journal of Machine Learning Research* 11: 1471–90.

Chen, Xiao, Qian Wang, and Jelena Srebric. 2015. "A Data-Driven State-Space Model of Indoor Thermal Sensation Using Occupant Feedback for Low-Energy Buildings." *Energy and Buildings* 91 (March). Elsevier: 187–98. doi:10.1016/j.enbuild.2015.01.038.

Dai, Changzhi, Hui Zhang, Edward Arens, and Zhiwei Lian. 2017. "Machine Learning Approaches to Predict Thermal Demands Using Skin Temperatures: Steady-State

Conditions." *Building and Environment* 114 (March). Pergamon: 1–10. doi:10.1016/j.buildenv.2016.12.005.

Dear, Richard J De, Gail Schiller Brager, James Reardon, Fergus Nicol, and others. 1998. "Developing an Adaptive Model of Thermal Comfort and Preference/Discussion." *ASHRAE Transactions* 104. American Society of Heating, Refrigeration and Air Conditioning Engineers, Inc.: 145.

Diaz, Keith M, David J Krupka, Melinda J Chang, James Peacock, Yao Ma, Jeff Goldsmith, Joseph E Schwartz, and Karina W Davidson. 2015. "Fitbit®: An Accurate and Reliable Device for Wireless Physical Activity Tracking." *International Journal of Cardiology*. NIH Public Access. doi:10.1016/j.ijcard.2015.03.038.

EN, C E N. 2007. "15251-2007, Criteria for the Indoor Environment Including Thermal, Indoor Air Quality, Light and Noise." *Brussels: European Committee for Standardization*.

Erickson, Varick L, and Alberto E Cerpa. 2012. "Thermovote: Participatory Sensing for Efficient Building Hvac Conditioning." In *Proceedings of the Fourth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, 9–16.

Fanger, P. O. 1972. *Thermal Comfort: Analysis and Applications in Environmental Engineering,*. New York: McGraw-Hill.

Fanger, Poul O, and others. 1970. "Thermal Comfort. Analysis and Applications in Environmental Engineering." *Thermal Comfort. Analysis and Applications in Environmental Engineering.* Copenhagen: Danish Technical Press.

Ferrannini, Eleuterio. 1988. "The Theoretical Bases of Indirect Calorimetry: A Review." *Metabolism*. W.B. Saunders. doi:10.1016/0026-0495(88)90110-2.

"Fitbit - Web API Reference." 2018. Accessed October 1. https://dev.fitbit.com/build/reference/web-api/.

Freund, Yoav, and Robert E Schapire. 1997. "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting." *Journal of Computer and System Sciences* 55 (1): 119–39. doi:10.1006/jcss.1997.1504.

Frontczak, Monika, and Pawel Wargocki. 2011. "Literature Survey on How Different Factors Influence Human Comfort in Indoor Environments." *Building and Environment* 46 (4). Elsevier: 922–37.

Gao, Peter Xiang, and S. Keshav. 2013. "Optimal Personal Comfort Management Using SPOT+." In *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings - BuildSys'13*, 1–8. doi:10.1145/2528282.2528297.

Gerrior, Shirley, Wenyen Juan, and Basiotis Peter. 2006. "An Easy Approach to Calculating Estimated Energy Requirements." *Preventing Chronic Disease* 3 (4). Centers for Disease Control and Prevention: A129.

"Get Started with Your Microsoft Band 2." 2018. Accessed November 1. https://support.microsoft.com/en-us/help/4000514/band-2-get-started.

Hang-yat, Lam Abraham, and Dan Wang. 2013. "Carrying My Environment with Me." In *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings - BuildSys'13*, 1–8. doi:10.1145/2528282.2528286.

Hasan, Mohammad H., Fadi Alsaleem, and Mostafa Rafaie. 2016. "Sensitivity Study

for the PMV Thermal Comfort Model and the Use of Wearable Devices Biometric Data for Metabolic Rate Estimation." *Building and Environment* 110: 173–83. doi:10.1016/j.buildenv.2016.10.007.

Hennion, Nicolas. n.d. "Glances an Eye on Your System."

Ho, Tin Kam. 1998. "The Random Subspace Method for Constructing Decision Forests." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20 (8): 832–44. doi:10.1109/34.709601.

"Home Assistant." n.d. https://www.home-assistant.io/.

Hoof, Joost Van. 2008. "Forty Years of Fanger's Model of Thermal Comfort: Comfort for All?" *Indoor Air* 18 (3). Wiley Online Library: 182–201.

Hoovestol, Ryan A., and Ted R. Mikuls. 2011. "Environmental Exposures and Rheumatoid Arthritis Risk." *Current Rheumatology Reports*. doi:10.1007/s11926-011-0203-9.

Höppe, Peter, and Ivo Martinac. 1998. "Indoor Climate and Air Quality." *International Journal of Biometeorology* 42 (1). Springer: 1–7.

Iso, En. 2005. "7730: 2005." *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*.

Jetté, M, K Sidney, and G Blümchen. 1990. "Metabolic Equivalents (METS) in Exercise Testing, Exercise Prescription, and Evaluation of Functional Capacity." *Clinical Cardiology* 13 (8): 555–65. doi:10.1002/clc.4960130809.

Kim, Joyce, Stefano Schiavon, and Gail Brager. 2018. "Personal Comfort Models – A

New Paradigm in Thermal Comfort for Occupant-Centric Environmental Control." *Building and Environment*. doi:10.1016/j.buildenv.2018.01.023.

Kohavi, Ron. 1995. "A Study of Cross Validation and Bootstrap for Accuracy Estimation and Model Selection." In *14th International Joint Conference on Artificial Intelligence (IJCAI)*, 2:1137–43.

Ku, K. L., J. S. Liaw, M. Y. Tsai, and T. S. Liu. 2015. "Automatic Control System for Thermal Comfort Based on Predicted Mean Vote and Energy Saving." *IEEE Transactions on Automation Science and Engineering* 12 (1): 378–83. doi:10.1109/TASE.2014.2366206.

Leaman, Adrian, and Bill Bordass. 1999. "Productivity in Buildings: The 'Killer'Variables." *Building Research & Information* 27 (1). Taylor & Francis: 4–19.

Ligeza, Antoni. 1995. "Artificial Intelligence: A Modern Approach." *Neurocomputing* 9 (2). Malaysia; Pearson Education Limited,: 215–18. doi:10.1016/0925-2312(95)90020-9.

Luo, Maohui, Wenjie Ji, Bin Cao, Qin Ouyang, and Yingxin Zhu. 2016. "Indoor Climate and Thermal Physiological Adaptation: Evidences from Migrants with Different Cold Indoor Exposures." *Building and Environment* 98: 30–38. doi:10.1016/j.buildenv.2015.12.015.

Martellotta, Francesco, Antonio Simone, Sabrina Della Crociata, and Michele D'Alba. 2016. "Global Comfort and Indoor Environment Quality Attributes for Workers of a Hypermarket in Southern Italy." *Building and Environment* 95: 355–64. doi:10.1016/j.buildenv.2015.09.029.

Mason, Llew, Jonathan Baxter, Peter Bartlett, and Marcus Frean. 1999. "Boosting Algorithms as Gradient Descent." *NIPS*, 512–18. doi:10.1109/5.58323.

McCullough, E a, B W Jones, and J Huck. 1985. "A Comprehensive Data Base for Estimating Clothing Insulation." *ASHRAE Transactions* 91 (2): 29–47.

Microsoft. 2015. "Microsoft Band SDK," 1–51.

Mifflin, Mark D, Sachiko T St Jeor, Lisa a Hill, Barbara J Scott, Sandra a Daugherty, and Young O Koh. 1990. "A New Predictive Equation for Resting Energy Expenditure in Healthy Individuals." *American Journal of Clinical Nutrition* 51 (2): 241–47. doi:10.1093/ajcn/51.2.241.

Nastase, Ilinca, Cristiana Croitoru, and CĂtĂlin Lungu. 2016. "A Questioning of the Thermal Sensation Vote Index Based on Questionnaire Survey for Real Working Environments." In *Energy Procedia*, 85:366–74. Elsevier. doi:10.1016/j.egypro.2015.12.263.

Naud-Dulude, Christophe. n.d. "Ruby Gem for Your Mac Stats."

Overeem, A., J. C. R. Robinson, H. Leijnse, G. J. Steeneveld, B. K. P. Horn, and R. Uijlenhoet. 2013. "Crowdsourcing Urban Air Temperatures from Smartphone Battery Temperatures." *Geophysical Research Letters* 40 (15). Wiley-Blackwell: 4081–85. doi:10.1002/grl.50786.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, et al. 2012. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12 (Oct): 2825–30. doi:10.1007/s13398-014-0173-7.2.

Peng, Bo, and Sheng-Jen Hsieh. 2017. "Data-Driven Thermal Comfort Prediction With Support Vector Machine." In *Volume 3: Manufacturing Equipment and Systems*, V003T04A044. ASME. doi:10.1115/MSEC2017-3003.

Quinlan, J R. 2006. "Simplifying Decision Trees, Int." *J* 27 (August 1986): 221–34. doi:10.1016/S0020-7373(87)80053-6.

Rafaie, M., F. Alsaleem, and A. Holthaus. 2017. "Data Fusion Application in Predicting Human Comfort." In *Structural Health Monitoring 2017: Real-Time Material State Awareness and Data-Driven Safety Assurance - Proceedings of the 11th International Workshop on Structural Health Monitoring, IWSHM 2017*. Vol. 2.

Salonen, Heidi, Jarek Kurnitski, Risto Kosonen, Ulla-Maija Hellgren, and Sanna Lappalainen. 2012. "The Effects of the Thermal Environment on Occupants ' Responses in Health Care Facilities : A Literature Review."

Tin Kam Ho. 1995. "Random Decision Forests." In *Proceedings of 3rd International Conference on Document Analysis and Recognition*, 1:278–82. doi:10.1109/ICDAR.1995.598994.

Trumbo, Paula, Sandra Schlicker, Allison A. Yates, and Mary Poos. 2002. "Dietary Reference Intakes for Energy, Carbohydrate, Fiber, Fat, Fatty Acids, Cholesterol, Protein and Amino Acids." *Journal of the American Dietetic Association*. doi:10.1016/S0002-8223(02)90346-9.

Vezhnevets, Alexander, and Olga Barinova. 2007. "Avoiding Boosting Overfitting by Removing Confusing Samples." *Work*. Berlin, Heidelberg: Springer Berlin Heidelberg, 430–41. doi:10.1007/978-3-540-74958-5_40.

Wagner, A, E Gossauer, C Moosmann, Th Gropp, and R Leonhart. 2007. "Thermal Comfort and Workplace Occupant Satisfaction—Results of Field Studies in German Low Energy Office Buildings." *Energy and Buildings* 39 (7). Elsevier: 758–69.

Wang, Haiying, and Songtao Hu. 2016. "Experimental Study on Thermal Sensation of People in Moderate Activities." *Building and Environment* 100 (February): 127–34. doi:10.1016/j.buildenv.2016.02.016.

Wolkoff, Peder, and Søren K. Kjærgaard. 2007. "The Dichotomy of Relative Humidity on Indoor Air Quality." *Environment International*. doi:10.1016/j.envint.2007.04.004.

Xue, Yu, Zhiqiang John Zhai, and Qingyan Chen. 2013. "Inverse Prediction and Optimization of Flow Control Conditions for Confined Spaces Using a CFD-Based Genetic Algorithm." *Building and Environment* 64: 77–84. doi:10.1016/j.buildenv.2013.02.017.

Zhang, Tian Hu, and Xue Yi You. 2014. "Applying Neural Networks to Solve the Inverse Problem of Indoor Environment." *Indoor and Built Environment* 23 (8): 1187–95. doi:10.1177/1420326X13499596.