Title	Incomplete Supervised Learning Method for Personal Thermal Comfort Model for Cyber-Physical Human Centric Systems
Author(s)	周, 晨冕
Citation	
Issue Date	2020-09
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/16857
Rights	
Description	Supervisor: リム 勇仁, 先端科学技術研究科, 修士(情報科学)

Master's Thesis

Incomplete Supervised Learning Method for Personal Thermal Comfort Model for Cyber-Physical Human Centric Systems

ZHOU, Chenmian

Supervisor Yuto Lim

Graduate School of Advanced Science and Technology
Japan Advanced Institute of Science and Technology
(Information Science)

August 2020

Abstract

A personal thermal comfort (PTC) model is a novel approach to predicting thermal sensation of individuals rather than groups of humans. The limitations of previous works are the difficulties of collecting feedback of thermal sensation, especially for feedback of neutral comfort, and the requirements of data size to achieve an acceptable performance of prediction. In this paper, we present a modified PTC model to predict real-time personal thermal sensation for cyber-physical human centric system (CPHCS), where psychological parameters are necessary for operating Heating, Ventilating, and Air Conditioning (HVAC) control system in order to offer satisfactory thermal comfort. The function of the proposed PTC model is given after data analysis of seven participants' experiments in smart home plant-iHouse. Then we presented a Personalized Predictive Classifier (PPC) specifically designed for CPHCS, which uses online learning and incomplete supervision to predict the 7-level thermal sensation of individuals. The results showed the appropriateness of using machine learning, Random Forest (RF) in particular, in the field of predicting personal thermal sensation with the performance of median accuracy of 0.86 using one RF classifier (RFC). Then we explored PPC using two cascaded RFCs, and it results in faster learning speed in most situations. We conclude that PTC model with PPC inside is able to offer psychological parameters(e.g., thermal sensation) inference in a timely manner to a continuous satisfactory control system in smart homes, for example, Energy Efficient Thermal Comfort Control (EETCC) system, so that satisfactory thermal comfort is available for individual living in smart home.

Keywords: Personal thermal comfort, Machine learning, Cyber-physical human centric system.

List of Abbreviations

AUC	Area Under the Receiver Operating Characteristic (ROC) Curve
CPHCS	Cyber-Physical Human Centric System
CPHS	Cyber-Physical Home System
CPS	Cyber-Physical System
EETCC	Energy Efficient Thermal Comfort Control
HVAC	Heating, Ventilating, and Air Conditioning
ICT	Information and Communication Technologies
PMV	Predictive Mean Vote
PPC	Personalized Predictive Classifier
PTC	Personal Thermal Comfort
PTSL	Personal Thermal Sensation Label
RF	Random Forest
RFC	Random Forest Classifier
SCL	Subjective Comfort Level

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Research Background

The comfort that the indoor environment provides is of great significance to people's satisfaction[1], productivity[2, 3], and health[4]. Thermal comfort is particularly important due that it is an important goal that Heating, Ventilating, and Air Conditioning (HVAC) systems, which are commonly installed to make people easeful, need to accomplish. Thermal comfort is an approach to evaluate people's satisfaction with the surrounding environment, which can be affected by both environmental and personal factors. The criteria for TC are described in standard as ASHRAE Standard 55[5] and ISO 7730[6], in which Predictive Mean Vote (PMV) and adaptive comfort models are mainly used. Fanger et al. gave mathematical expression in PMV model to describe the thermal sensation of occupants by subjective evaluation in [7].

Home is one of the most crucial indoor environments that can offer occupants with a safe and comfortable place satisfying both physical and psychological requirements. A smart home is one of the scenario applying ubiquitous computing and refers to the cooperation of Information and Communication Technologies (ICT) and smartness in home control, which includes appliance control and actuator automation, to provide services of comfort, healthcare, security, and power saving[8, 9]. Cyber-Physical Home System (CPHS), which is based on the concept of Cyber-Physical Systems (CPSs), combines many kinds of services and applications into smart home systems. Especially aiming for improving comfortability, providing healthcare monitoring, and reducing energy consumption, numerous pilot studies on CPHS have been conducted. One excellent research in CPHS domain is the work that Yuto et al. implemented Energy Efficient Thermal Comfort Control (EETCC) system in the real testbed to improve occupant's thermal comfort and optimize the energy consumption ate the same time.[10]

1.2 Problem Statement

However, there are limitations for PMV model to predict thermal comfort in the scenario of smart homes. First, Fanger et al. derived PMV model from chamber data during experiments, and PMV model is designed to satisfy the majority of a large occupant group. This leads to poor prediction power when PMV model is applied to predict individuals' thermal comfort[11]. Second, the input variables of PMV model are often assumed or simplified because specific variables that full implementation of PMV model needs, such as metabolic rate, are difficult to acquire in real-time in real environments. Third, the occupant's particular living conditions cannot accord with the experimental one perfectly hence it is hard to measure and quantify the input settings as what we do in a standardized experimental environment. Lastly, new variables cannot be taken into consideration, which has been proved relevant to thermal sensation (e.g., age, gender, and BMI [12]). In essence, these new variables are significant for predicting thermal comfort. Due to the above limitations, even though many proposed system models, which are based on PMV as an indicator of thermal comfort, are widely accepted and adopted, several studies [13, 14, 15] point out that these models are weak in predicting accurate thermal sensation of an occupant. Therefore, PMV-based EETCC system is inadequate in reflecting the actual thermal preference of an individual.

Also, There is no tight synchronization between occupants' preferences and systems in many research works in field of CPHS that consider human factors in the system design. Meanwhile, such system design ignores the variation of occupants preferences in a long term.

1.3 Research Motivation

To overcome the aforementioned shortcomings, Personal Thermal Comfort (PTC) model is proposed to satisfy the thermal comfort of an individual rather than the majority of a large population. There are several works[14, 16, 15, 13, 17] that have proposed PTC model, but only one of them considering using an online learning approach. What is more, None of them considered the difficulties of obtaining examples of neutral thermal sensation or thermal preference of no change. In this paper, a novel PTC model is presented to predict the real-time thermal sensation of an individual occupant in the field of smart home. This Proposed PTC model is able to distinguish whether an occupant is in a neutral comfort zone without gaining feedback on feeling comfort. The prediction of personal thermal sensation can help HVAC system with refined and precise control strategies as personalized profiles to offer not only a comfortable indoor environment but also happiness and well-being in smart home.

1.4 Research Objective

The objectives of this paper are to enhance the availability and accuracy of psychological parameters for PTC model by using an online learning and incomplete supervision method in Cyber-Physical Human Centric System (CPHCS), and to ensure the psychological parameters are always available for the continuous satisfactory control of a home control system (e.g., EETCC system) in a timely manner. In addition, the contributions that we offer in this papers are: (i) deciding necessary input variables for PTC model in order to make the model precise and simple; (ii) taking advantage of vast unlabeled data meanwhile reducing the dependence of thermal sensation feedback; (iii) proposing online learning method so that makes it possible for a quick start after implementation.

1.5 Thesis Organization

In chapter 2, background research information is listed in detail. We talk about Smart Homes, CPHCS, thermal comfort and sensation of humans. Then the concepts of PTC model and EETCC, which is implemented in the experiment base, are demonstrated. In chapter 3, a modified CPHCS framework is proposed. Based on this framework and the obtained data, we explain the reason that we use machine learning with an online learning method and incomplete supervision. In chapter 4, we introduce five metrics for evaluating performance. What is more, results are illustrated in detail. In chapter 5, we make a conclusion of our work.

Chapter 2

Background

2.1 Smart Homes

According to [18], Smart Home was first officially used in the form of term 'smart house' in 1984 by American Association of House Builders, which aimed to include new technologies in designing new homes. Smart home is defined in [19] as a home-like environment that adopts the combination of ambient intelligence and automatic control to react to the behaviors of occupants with a variety of facilities. Smart Home is one important branch ofCPS domains, which means that one key characteristic is the ability of interactions between the physical world and the cyber range. As one kind of these interactions, the interaction between the occupant and home appliances brings benefits of improving comfort, providing health care, ensuring security, and saving energy.[20] Smart homes are one of the key technologies facing the problem of an aging society. In the future, a smart home will integrate into daily life with dedicated artificial intelligence, computational power, communication skills, monitoring, and controlling abilities needed to improve everyday activities.

There are several implementations of smart home, and one of them is iHouse, in which our experiments are conducted in this paper. iHouse is a typical two-story building of Japanese style located in Nomi City, Ishikawa Prefecture, Japan. It is equipped with home appliances which are connected via ECHONET Lite version 1.1 and ECHONET version 3.6.2. Moreover, it incorporated more than 300 sensors and actuators to provide ample parameters for thermal comfort control and prediction.

2.2 Cyber-Physical Human Centric Systems (CPHCS)

Human and system are indivisible as systems are designed for satisfying people's needs. There are studies shown that it is of essence to interact

CPS with humans. Higashino and Uchiyama proposed the human centric cyber-physical system application in which the influence of human behaviors is taken into account to design CPS based societal system. Sowe et al. presented people in the loop of cyber-physical-human systems. Based on the consideration that human or occupant ought to be the first factor in CPS, we proposed the concept of Human Centric CPS shown in Figure 2.1.

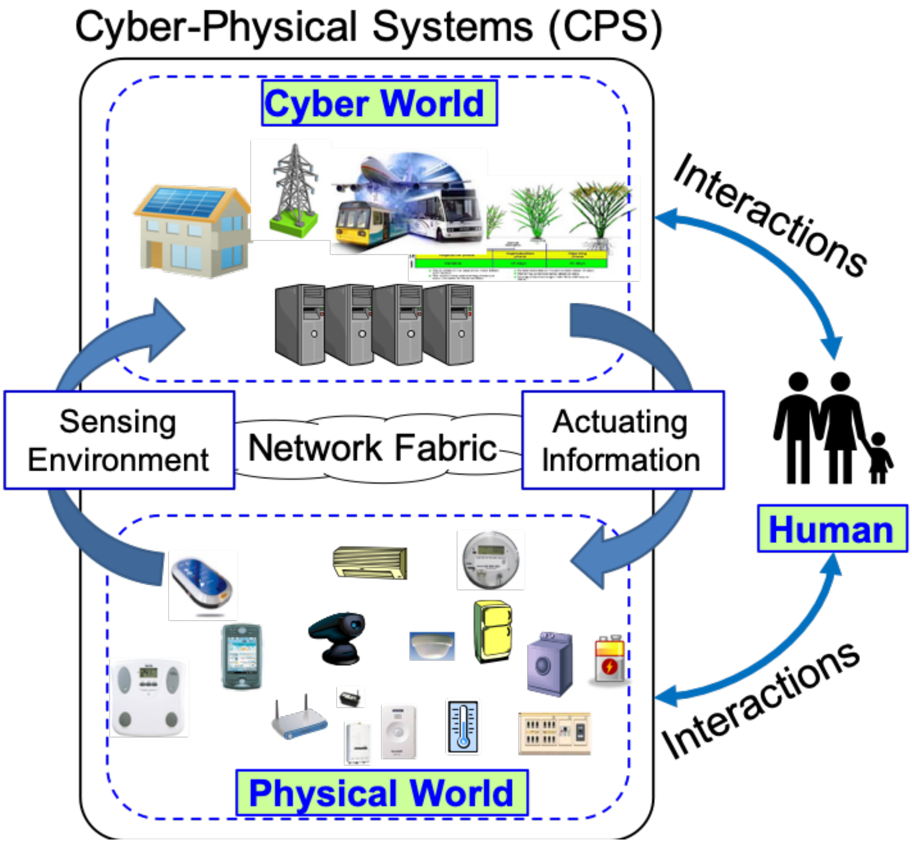


Figure 2.1: Concept of Human Centric CPS

Fang et al. proposed a CPHCS framework with the concept of Human Centric CPS. It centers human as the core element in the whole scheme so that human can interact with the surrounding physical environment and the user applications in the cyber world. Moreover, it provides energy saving and satisfies human's thermal comfort needs at the same time with sensing individual's thermal sensation. In order to make psychological parameters always available, we proposed a modified CPHCS framework based on the concept of CPHS, which is illustrated in Figure 2.2, for the sake of implementing PTC model into CPHCS.

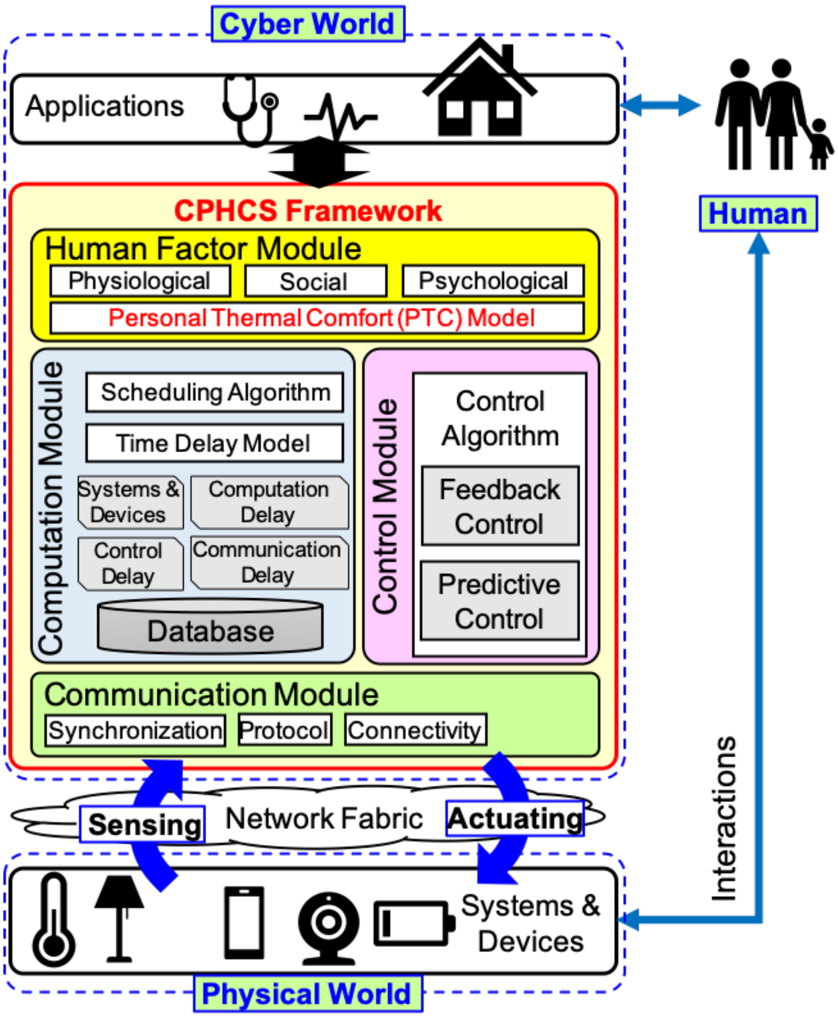


Figure 2.2: Cyber-physical Human Centric System Framework

2.3 Thermal Comfort of Human

Thermal comfort is in the description of the state that expresses whether ambient surroundings are satisfactory. Fanger et al. proposed the predicted mean vote/predicted percentage of dissatisfied (PMV/PPD) in the last century and has been included by ISO-7730 [6]. However, there are limitations for PMV/PPD model to predict thermal comfort in the scenario of smart home. For example, PMV/PPD derived model from chamber data during experiments, and PMV/PPD model is designed to satisfy the majority of large occupant groups other than an individual.

Even though the PMV/PPD model offers us a judging method for evaluating thermal comfort level, subjective evaluation is still needed. In this paper, make the subjective evaluation level into seven-level based on standard ASHRAE 55 [5], which is shown in Table 2.1. This subjective evaluation level has been modified into the Subjective Comfort Level (SCL) to collect direct human sensation by answering online questionnaires in [9].

Cold	Cool	Slightly Cool	Neutral	Slightly Warm	Warm	Hot
-3	-2	-1	0	+1	+2	+3

Table 2.1: Human's comfort degree on 7-point ASHRAE scale

PMV is calculated with a combination of air temperature, mean radiant temperature, relative humidity, airspeed, metabolic rate, and clothing insulation, which is defined as (2.1).

$$PMV = f_{pmv}(t_a, \bar{t}_r, v_{ar}, p_a, M, I_{cl}) \tag{2.1}$$

where t_a is air temperature, $\bar{t}_r$ is mean radiant temperature, v_{ar} is relative air velocity, p_a is water vapor partial pressure, M is the metabolic rate, and I_{cl} is the human's clothing insulation factor.

2.4 Thermal Sensation of Human

With the assumption that occupants tend to regain comfort when encountering discomfort, occupants play an active role in seeking comfortability by interacting with related elements such as using thermostats[23], openable windows[24], or fans[25]. Such interactions demonstrate thermal sensation or thermal preference of occupants. Compared to survey participation, these interactions can be tracked as the feedback of current thermal sensation requiring no additional effort except for normal behavior. In addition, obtaining thermal sensation in this way can be automated in Smart Homes with the help of CPHCS. However, when occupants feel comfortable with no preference of tending to be comfortable, it is of great difficulty to obtain such examples by inferring the thermal comfort of occupants from interaction with thermostats, etc.

2.5 Personal Thermal Comfort Model

A general personal thermal comfort aims to predict the thermal comfort sensation or preference of individuals other than the majority of a group of occupants and therefore improves the performance of HVAC control system. In[14], heating and cooling behaviors were adopted to predict personal thermal preference of 3 categories. In[13], several physiological signals (e.g., skin temperature, heart rate) were used to predict personal thermal preference of 3 categories. In[26], the authors presented a novel infrared thermography based technique to monitor an individual's thermoregulation performance and thermal comfort levels. Jung and Jazizadeh have investigated the effect of personal thermal comfort in selecting setpoints for thermostat and probabilities of thermal satisfaction. These works show the significance of considering personal thermal comfort instead of conventional predictors like PMV.

In the domain of CPHCS, PTC model diverts attention much more into human factors comparing with traditional control algorithms like model predictive control and proportional–integral–derivative controller. This proposed PTC model provides reliable psychological parameters output as the input for the PCA module shown inFigure 3.1. The PCA module outputs personal PMV for controllers, which are mainly driven by PMV index strategy. Therefore PTC model benefits the effect of these controllers(e.g., EETCC).

2.6 Energy Efficient and Thermal Comfort Control

EETCC algorithm is a rule-based control strategy that controls multiple actuators to maintain thermal comfort level dynamically. It is developed for smart homes, where PMV is adopted to evaluate the level of thermal comfort. [27] There are different control strategies (illustrated in Table 2.2) differing from the category of PMV range, which indicates thermal comfort circumstances.

Category	*PMV*
A	-0.2 < PMV < +0.2
B	-0.5 < PMV < +0.5
C	-0.7 < PMV < +0.7

Table 2.2: Categories of thermal comfort demands

EETCC system gives out adaptive PMV as a controlling index. However, as it is mentioned that PMV model is limited inherently in predicting the thermal comfort of an occupant in a real implementation, EETCC algorithm can not offer occupant highly acceptable thermal comfort and high energy efficiency simultaneously. In this paper, PMV value that we used is collected from EETCC system, and it is a little bit different from the original PMV.

Chapter 3

Machine Learning Method for CPHCS

In this chapter, a modified PTC model is first presented. Then we introduce the experiment setup and conduct the data analysis. Based on the result of data analysis, we present a novel supervised machine learning method using incomplete supervision and online learning for CPHCS.

3.1 Modified Personal Thermal Comfort Model for CPHCS

3.1.1 Architecture of Personal Thermal Comfort Model

iHouse, which is the experiment base, has evolved several times. Fang et al. have proposed CPHCS framework in [9], whose experimental base is iHouse as well. In the dissertation of Fang, which has not publicly published yet, she demonstrates a state-of-the-art work that combines time task, human factors, and control strategy in CPHCS. In order to fully utilize develop it, we proposed a modified system architecture for the implementation of PTC model in Figure 3.1.

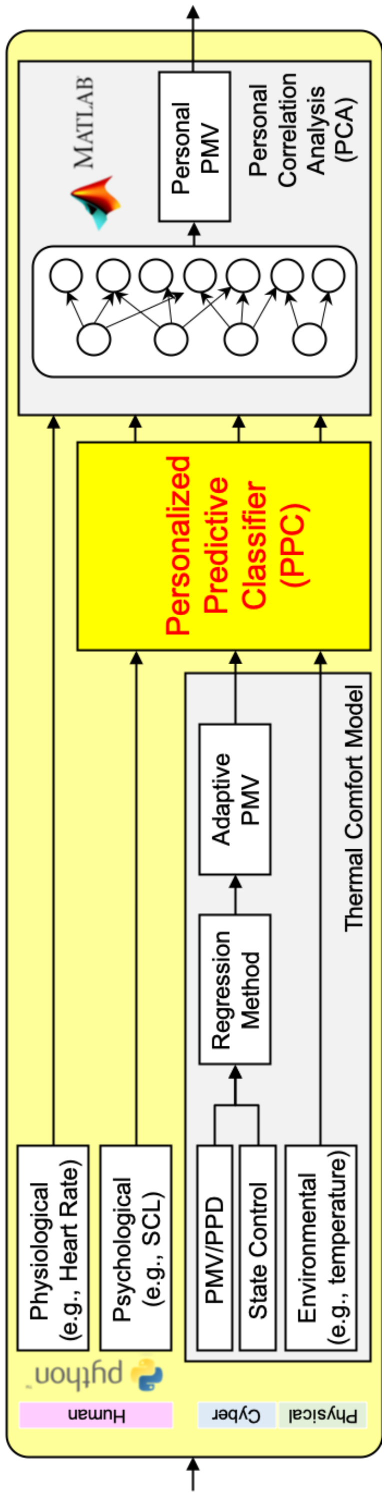


Figure 3.1: Architecture of Personal Thermal Comfort Model

This PTC model is specially designed for individuals rather than groups. First, we designed it in a smart home environment instead of experiment chambers. Second, it does not take the percentage of discomfort(PPD) into account. In other words, it aims to satisfy one individual in a specific environment.

3.1.2 Reasons for Adopting a 7-level Subjective Comfort Level (SCL)

In the design of PTC model, we adopt 7-level SCL as one psychological input variable. The main reason is that the energy consumption of HVAC systems accounts for half of building energy consumption in the USA[28]. In[29], five levels of thermal discomfort are set, and the results demonstrate the improvement of a 12% reduction in average airflow rate, which means lower power consumption. It is believed that more classes make it possible that we create elaborate profiles for HVAC control system in order to offer better thermal comfort and save energy meanwhile.

3.2 Experiment Setup

This experiment was conducted in corporation with Fang and the the data we used in this paper is extension of data set used in [9].

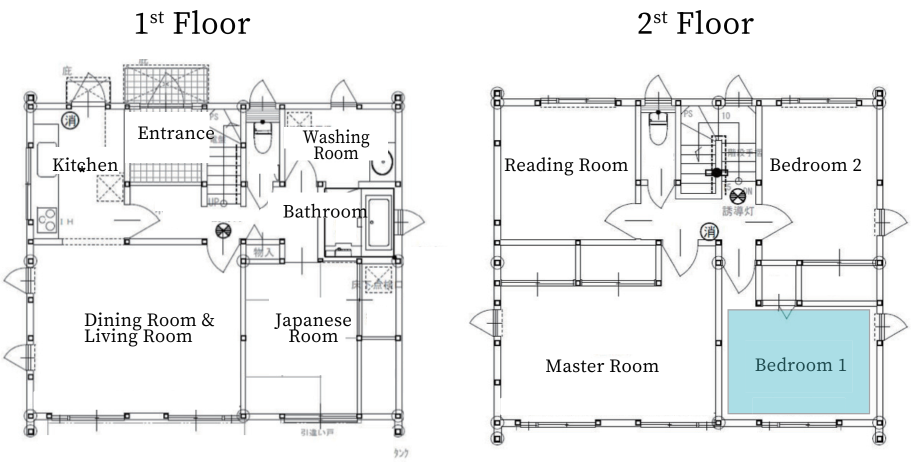
3.2.1 Plant: iHouse

Bedroom 1 on the second floor of iHouse was selected as the experiment environment this time. The layout of iHouse is shown in Figure 3.2. The bedroom 1 is 5.0 m in length, 4.1 m in width, and 2.4 m in height. Specific parameters of primary sensors installed in iHouse are shown in Table 3.1.

Type	Name	Range
Indoor temperature sensor	SHT75 digital sensor	[-40,125]°C ± 0.3 °C
Air velocity sensor	hot-wire anemometer sensor	[0.015,5]m/s ± 0.2%
Relative humidity sensor	SHT75 digital sensor	[0,100]% ± 1.8%

Table 3.1: Brief information of main sensors

(a) Exterior



(b) Floor plan

Figure 3.2: Overall appearance and internal structure of iHouse

3.2.2 Obtained Dataset

First, random submissions were adopted to record thermal sensation of participants when they felt a change in thermal comfort. The questionnaire link was sent to each participant, and they independently filled the questionnaire form via their personal computers or personal smartphones. The questionnaire form is made in Google Forms, and it consists of three parts: (a) date and time; (b) ID of participants; (c)thermal sensation. Because it takes time to fill a form, it will be automatically copied two times for ten seconds before and after the submission time. Examples of form are presented in Table 3.2. The questionnaire form will be named as SCL vote afterward.

Participant ID	Date and Time	Thermal Sensation	SCL
F1	2019/05/31 10:29:03	Hot	3
M1	2019/05/31 11:14:23	Warm	2
M2	2019/06/28 11:19:44	Slightly Cool	-1

Table 3.2: Examples of questionnaire form

Second, environment parameters and adaptive PMV value were recorded uninterruptedly per 10 seconds with the support of EETCC system. One example of records is presented in Table 3.3. Set $M = 1.0, I_{cl} = 0.5$ for summer and $M = 1.0, I_{cl} = 0.8$ for winter as default parameters for EETCC calculating adaptive PMV.

state window1	state curtain	sensor TemperatureIndoor	sensor TemperatureOutdoor	...	PMV	timestamp
49	49	27.5	23.1	...	0.653136	2019/05/31 10:59:41

Table 3.3: EETCC record example

3.3 Data Analysis

3.3.1 Information of Participants and Weather Conditions

There are seven participants of three female adults and four male adults. Consents of all participants are obtained before we experimented. All participants had no physical or mental illness, no alcohol intake before the experiment, and no lack of sleep. Detailed information on these participants is shown in Table 3.4.

ID	Gender	Age	Experiment date
F1	Female	38(39)	May 31, 2019 Feb. 02, 2020
F2	Female	26	Feb. 04, 2020
F3	Female	27	Feb. 01, 2020
M1	Male	26	May 31, 2019
M2	Male	23	June 28, 2019
M3	Male	24	June 28, 2019
M4	Male	27	Jan. 31, 2020

Table 3.4: Brief information of participants

Experiments conducted in Nomi City, Japan, were distributed over six days in 2019 and 2020. Detailed weather conditions information is shown in Table 3.5

	Mean temperature(SD) °C		Mean air velocity(SD) m/s		Mean relative humidity(SD) %	
Date	Indoor	outdoor	Indoor	outdoor	Indoor	outdoor
May 31, 2019	26.5(0.5)	17.9(2.4)	0.10(0.05)	1.77(1.48)	45.9(2.8)	92.0(15.6)
June 28, 2019	27.0(1.1)	22.9(2.1)	0.16(0.10)	2.56(1.41)	48.3(6.7)	85.4 (18.1)
Jan. 31, 2020	23.4(0.8)	2.7(0.4)	0.10(0.05)	1.9(0.8)	36.9(1.4)	100(0.0)
Feb. 01, 2020	21.0(3.3)	4.5(0.5)	0.13(0.07)	2.3(0.7)	38.8(5.2)	98.9(1.2)
Feb. 02, 2020	24.0(3.6)	8.4(1.7)	0.11(0.06)	1.2(0.6)	31.8(6.1)	68.2(8.1)
Feb. 04, 2020	21.4(2.5)	5.6(2.2)	0.11(0.05)	1.9(1.1)	39.0(4.2)	80.5(20.2)

Table 3.5: Weather conditions (SD: standard deviation)

3.3.2 SCL Distribution of Participants

We got eight sets of records of seven participants. The detailed information is as follows.

ID of participant	ID of dataset	data size	number of SCL vote
F1	ID1	1998	57
	ID8	744	115
F2	ID5	768	66
F3	ID7	744	89
M1	ID2	1998	60
M2	ID3	1133	114
M3	ID4	1133	114
M4	ID6	767	36

Table 3.6: Information of recorded data

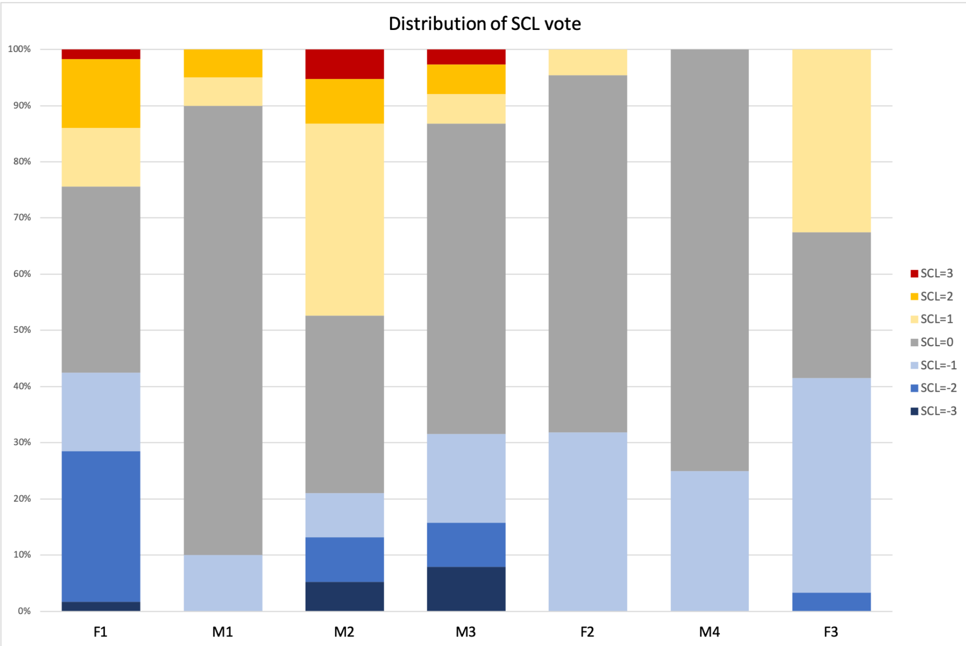


Figure 3.3: Distribution of SCL vote

3.3.3 Data Preprocessing

3.3.3.1 Personal Thermal Sensation Label

There are few samples with $SCL = -3\ or\ SCL = 3$ as a result of distribution analysis. Therefore we made a simple map from SCL, which is seven levels, to Personal Thermal Sensation Label (PTSL), which is five levels. Note that if the dataset from SCL is large and sufficient, PTSL can be extended to equivalent to the SCL.

<table>
<tr><th>SCL</th><th>PTSL</th></tr>
<tr><td>3</td><td rowspan="2">2</td></tr>
<tr><td>2</td></tr>
<tr><td>1</td><td>1</td></tr>
<tr><td>0</td><td>0</td></tr>
<tr><td>-1</td><td>-1</td></tr>
<tr><td>-2</td><td rowspan="2">-2</td></tr>
<tr><td>-3</td></tr>
</table>

Table 3.7: Correspondence between SCL and PTSL

3.3.3.2 Data preprocessing

Data preprocessing is conducted before analyzing the correlation among all recorded data. Average and difference values are calculated as the preprocessing method to temperature, air velocity, and relative humidity. Meanwhile, SCL will be transformed into PTSL pattern(e.g., $SCL = 3$ will be converted into $SCL = 2$), but the name will not be modified. The abbreviations and description of newly generated variables and original variables are shown in Table 3.8.

3.3.4 Correlation Analysis

In order to have an intuitive understanding of the relationship among environment parameters, adaptive PMV, and SCL, pairwise relationship figures are firstly plotted in Figure 3.4. Averaged variables are chosen rather than both of them. For example, average indoor temperature in 60 seconds will be selected to plot instead of using indoor temperature. Note that different color represents different SCL categories, and the legend is on the upper right of the figure. In order to orthogonalize these variables, the related variables will be removed. For example, $ControlSignal$ is highly related to air velocity indoor.

Variable name	description	abbreviations	mathematical representation	unit	type
sensor TemperatureIndoor	indoor temperature	Ti	T_{in}	°C	numerical
avg.TemperatureIndoor.60	average indoor temperature in 60 seconds	avg.Ti.60	$\bar{T}_{in}$	°C	numerical
avg.TemperatureIndoor.d.60	difference of average indoor temperature in 60 seconds	avg.Ti.d.60	$\Delta\bar{T}_{in}$	°C	numerical
sensor TemperatureOutdoor	outdoor temperature	To	T_{out}	°C	numerical
avg.TemperatureOutdoor.300	average outdoor temperature in 300 seconds	avg.To.300	$\bar{T}_{out}$	°C	numerical
avg.TemperatureIndoor.d.300	difference of average outdoor temperature in 300 seconds	avg.To.d.300	$\Delta\bar{T}_{out}$	°C	numerical
sensor AirSpeedIndoor	indoor air velocity	Ai	A_{in}	m/s	numerical
avg.AirSpeedIndoor.60	average indoor air velocity in 60 seconds	avg.Ai.60	$\bar{A}_{in}$	m/s	numerical
avg.AirSpeedIndoor.d.60	difference of average indoor air velocity in 60 seconds	avg.Ai.d.60	$\Delta\bar{A}_{in}$	m/s	numerical
sensor AirSpeedOutdoor	outdoor air velocity	Ao	A_{out}	m/s	numerical
avg.AirSpeedOutdoor.60	average outdoor air velocity in 60 seconds	avg.Ao.60	$\bar{A}_{out}$	m/s	numerical
avg.AirSpeedOutdoor.d.60	difference of average outdoor air velocity in 60 seconds	avg.Ao.d.60	$\Delta\bar{A}_{out}$	m/s	numerical
sensor HumidityIndoor	indoor relative humidity	Hi	H_{in}	%	numerical
avg.HumidityIndoor.60	average indoor relative humidity in 60 seconds	avg.Hi.60	$\bar{H}_{in}$	%	numerical
avg.HumidityIndoor.d.60	difference of average indoor relative humidity in 60 seconds	avg.Hi.d.60	$\Delta\bar{H}_{in}$	%	numerical
sensor HumidityOutdoor	outdoor relative humidity	Ho	H_{out}	%	numerical
avg.HumidityOutdoor.60	average outdoor relative humidity in 300 seconds	avg.Ho.300	$\bar{H}_{out}$	%	numerical
avg.HumidityOutdoor.d.60	difference of average outdoor relative humidity in 300 seconds	avg.Hi.d.300	$\Delta\bar{H}_{out}$	%	numerical
PMV	adaptive PMV value from EETCC	PMV	PMV	-	numerical
ControlSignal	state of actuators	-	-	-	category
TimePeriod	morning or afternoon	-	-	-	category
Season	Summer or Winter	-	-	-	category
SCL	subjective thermal sensation	SCL	SCL	-	category

Table 3.8: Abbreviation and description of variable

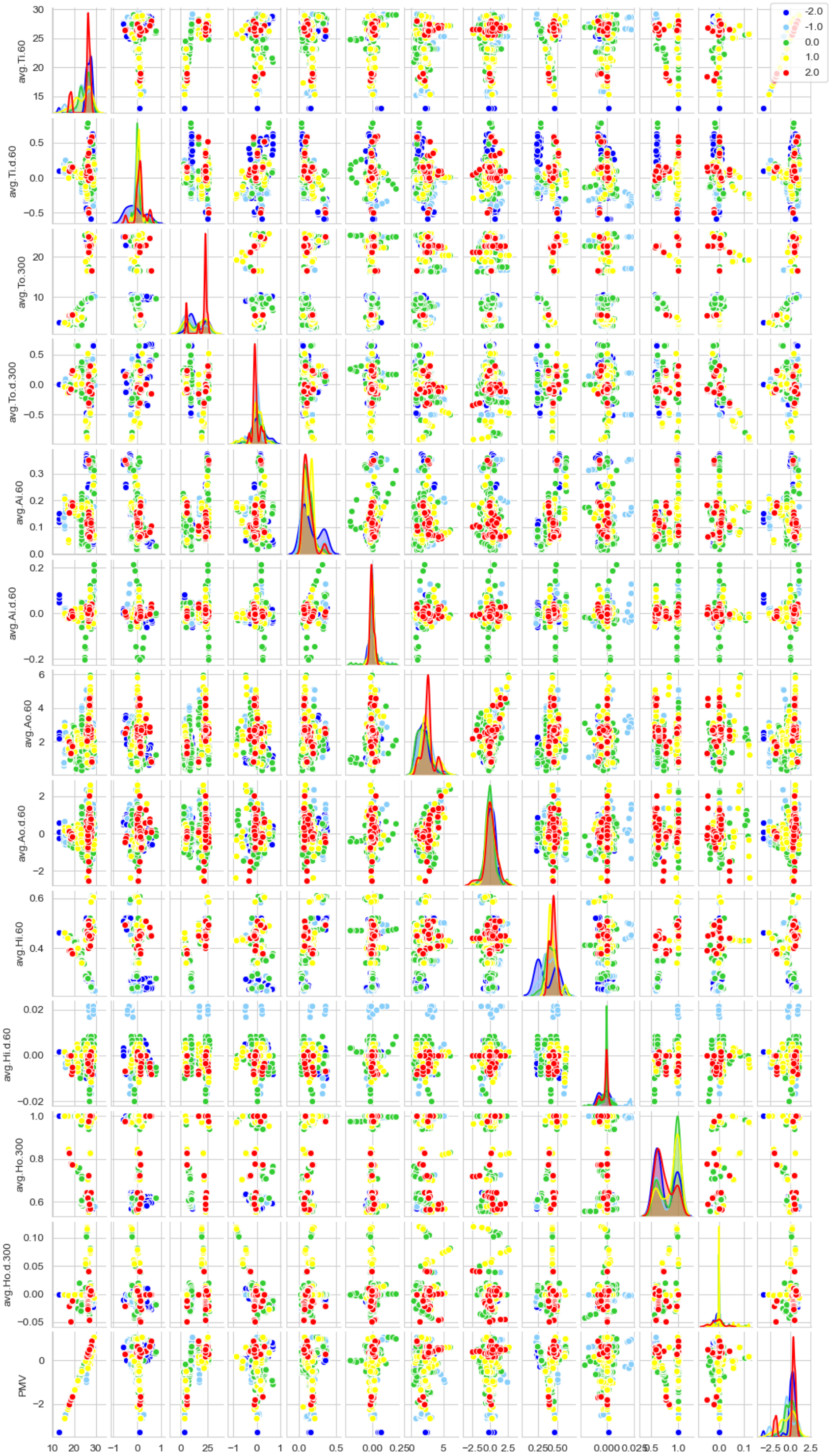


Figure 3.4: Pairwise plot of variables

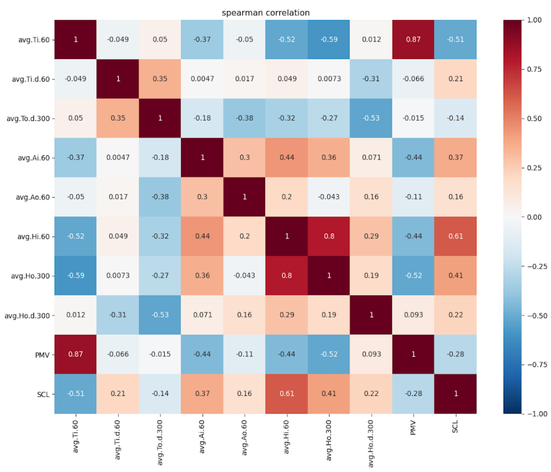
As SCL is in the form of category, we selected the Spearman's rank correlation coefficient as the index of correlation analysis, which is demonstrated in(3.1).

$$r_S = 1 - \frac{6\Sigma d_i^2}{n\left(n^2 - 1\right)} \tag{3.1}$$

where:

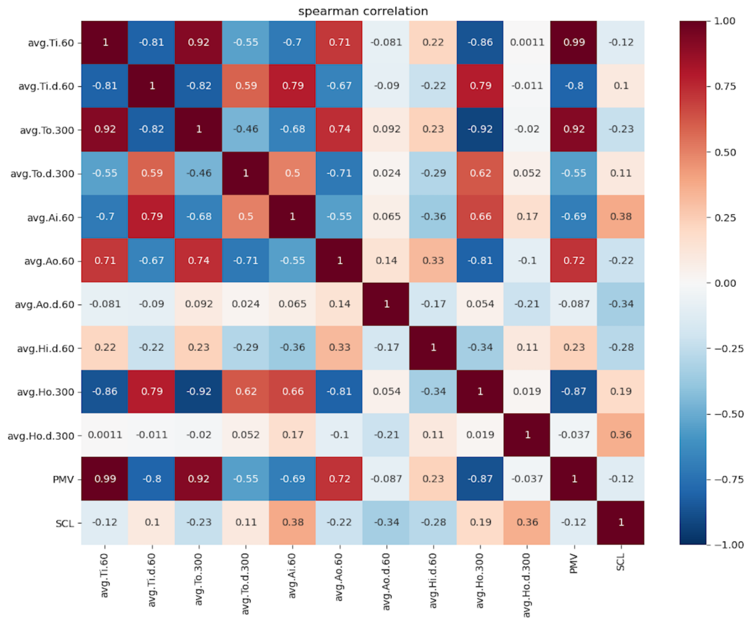
$d_i = rg\left(X_i\right) - rg\left(Y_i\right)$ is the difference between the two ranks of each observation,
n *is* the number of observations

After calculating Spearman's rank correlation coefficient of all variables for all participants, the variable which has a coefficient greater than 0.1 was kept. If the average one and original one (e.g., average indoor temperature in 60 seconds and indoor temperature) are both kept, the original one is removed. In addition, PMV value is always reserved for the sake of demonstrating a weak correlation between PMV and SCL. Correlation coefficient, which is in the form of heatmaps shown in Figure 3.5. The summary chart, which shows the difference between each participant, is shown in Figure 3.6. It clearly demonstrates that PTC model is necessary for the reason that the correlation coefficient is diverse among these participants.

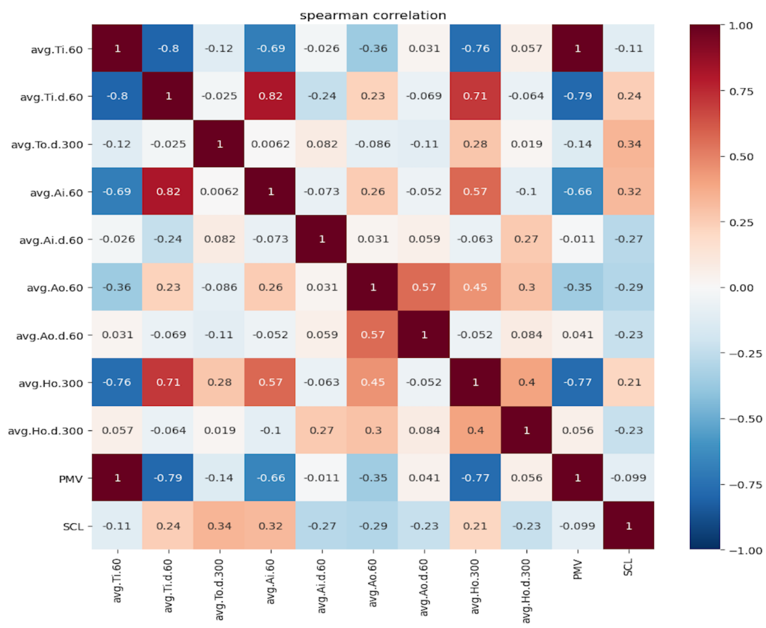


(a) F1

Figure 3.5: Heatmap of Spearman's Correlation coefficient

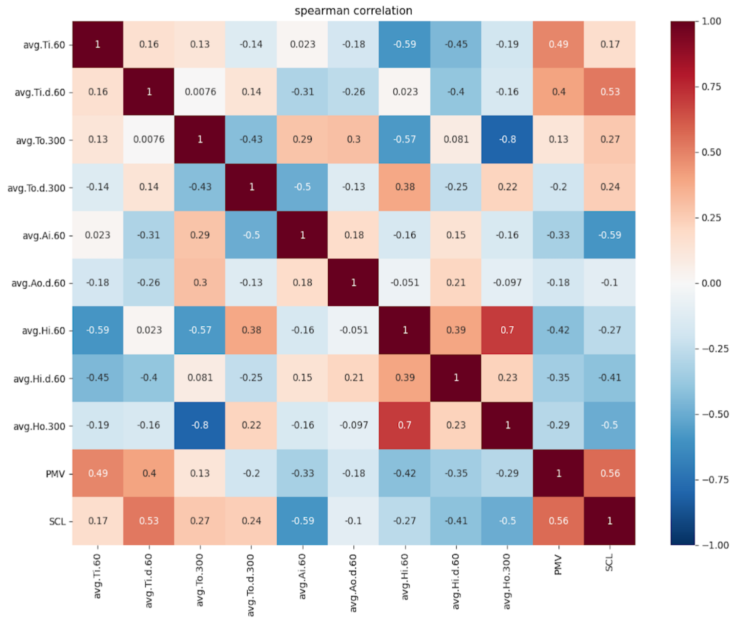


(b) F2

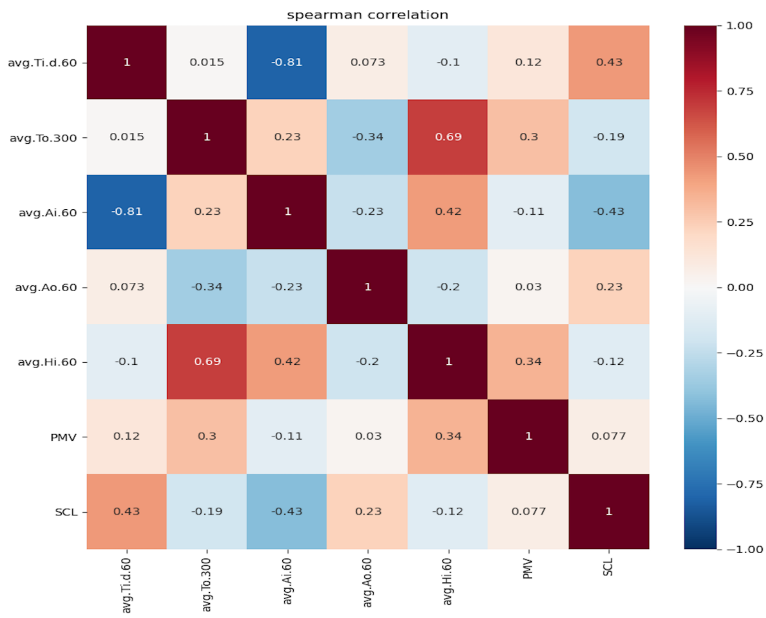


(c) F3

Figure 3.5: Heatmap of Spearman's Correlation coefficient

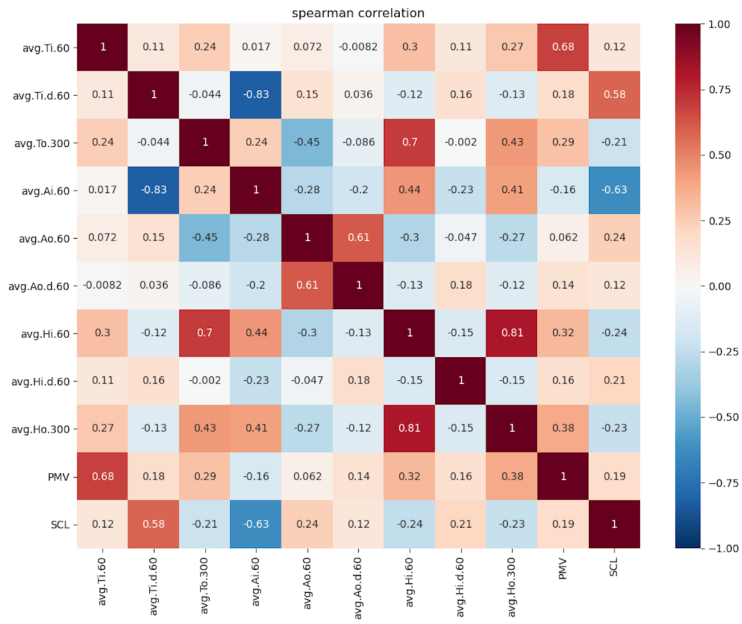


(f) M1

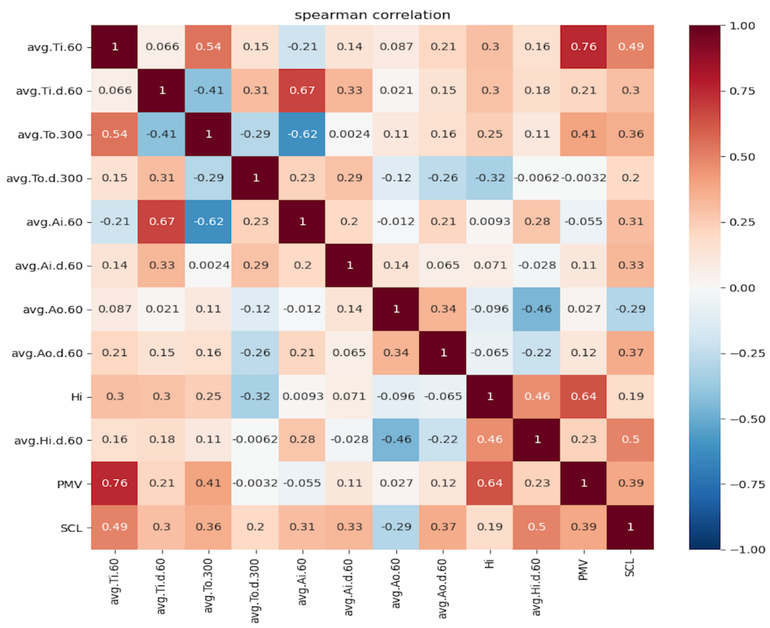


(g) M2

Figure 3.5: Heatmap of Spearman's Correlation coefficient



(j) M3



(k) M4

Figure 3.5: Heatmap of Spearman's Correlation coefficient

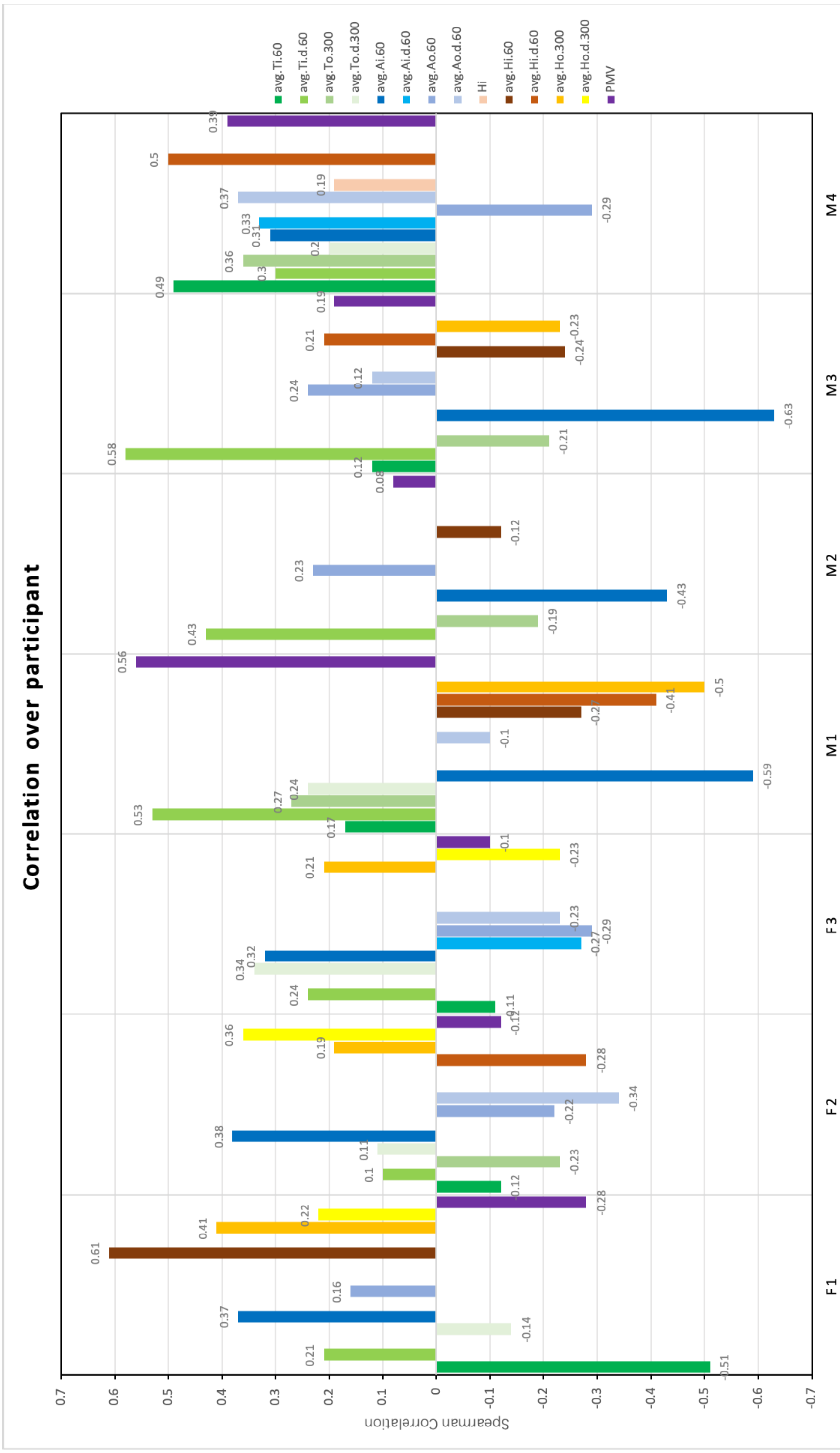


Figure 3.6: Correlation over each participant

As a result of correlation analysis, in order to make PTC model accurate and simplified, 11 variables are selected as the input for PPC. Hence we got the function of PTC model in (3.2).

$$\begin{aligned} F_{PTC}(t) = f_{PPC}(X(t)) = f_{PPC} \big(& \bar{T}_{in}(t), \Delta\bar{T}_{in}(t), \bar{T}_{out}(t), \Delta\bar{T}_{out}(t), \\ & \bar{A}_{in}(t), \bar{A}_{out}(t), \Delta\bar{A}_{out}(t), \\ & \bar{H}_{in}(t), \Delta\bar{H}_{in}(t), \bar{H}_{out}(t), PMV\big) \end{aligned} \tag{3.2}$$

3.4 Random Forest Classifier

In [14], authors compared several popular machine learning algorithms, which include Gaussian Process Classification, Gradient Boosting Method, Kernel Support Vector Machine, Random Forest (RF), etc. Random Forest displayed the best performance among the tested algorithms. Random forest is used for classification or regression using ensemble learning by constructing several decision tress. Therefor, it decrease the possibilities of overfitting compared to using one decision tree. In this paper, we choose Random Forest Classifier (RFC) as a primary classification algorithm in designing a specialized classifier for predicting thermal sensation. The algorithm of RFC is provided by Scikit-learn [30] of version 0.23.1.

3.5 Incomplete Supervised Learning

In binary classification, there are scenarios where negative samples consist of the majority, but they hard to be labeled. Retrieval [31, 32, 33], incomplete supervised learning is able to distinguish by using all unlabeled data as negative and then heuristically identifying reliable negative samples from unlabeled data. Shown inFigure 3.3, votes of $SCL = 0$ consist of the majority in the data set. However, we face the difficulties of obtain data with this label using occupant's thermal response introduced in the background. Therefore, incomplete supervised learning is suitable for resolving this issues. The primary process of incomplete supervised learning used for PTC model, which is in the framework of CPHCS, is illustrated in Figure 3.7.

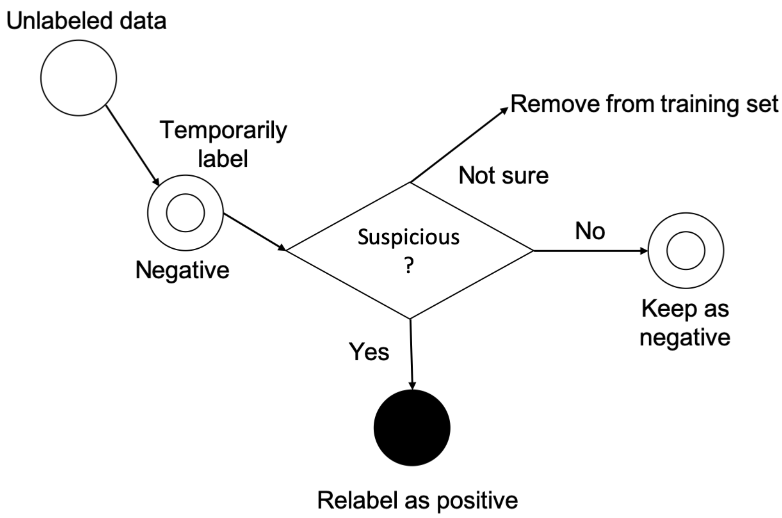


Figure 3.7: Process of incomplete supervised learning

3.6 Personalized Predictive Classifier

3.6.1 Architecture of Personalized Predictive Classifier

By illustrating where PTC model is in Figure 3.8, it is of great significance to offer accurate input so that controllers are capable of satisfying occupants with thermal comfort. We take variables from the filed of physical world, cyber world, and human. They have been examined in 3.3.

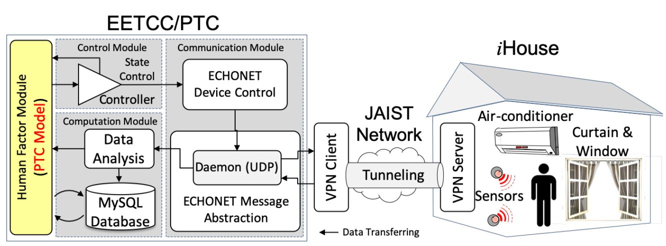


Figure 3.8: Proposed System Architecture

There are two main problems that we should take into account when designing the specific classifier: (i) observations of SCL are sparse along with time series so that proposed online classification algorithm ought to have the ability to learn quickly; (ii) proposed classifier ought to have the ability to predict PTSL of neutral comfort even no such data are given during training.

The core part of PTC is one classifier named Personalized Predictive Classifier (PPC), and the general design of PPC is illustrated in Figure 3.9. It is a classifier that is able to update in real-time and demarcate the comfortable and uncomfortable zone by just providing feedback of uncomfortableness only when occupant encounters thermal discomfort. It aims to achieve an acceptable level of performance using the method of online learning. In order to gain performance faster than traditional supervised machine learning algorithms, we take advantage of the vast numbers of unlabeled data.

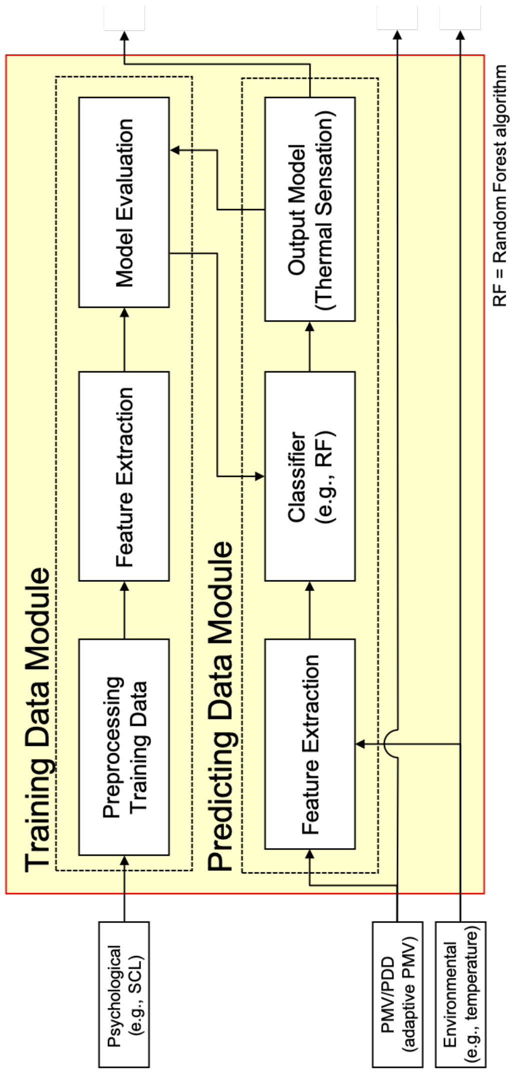


Figure 3.9: Personalized Predictive Classifier

Due that neutral comfort consists majority of SCL votes in the environment equipped with HVAC system, i.e., we get SCL information sparsely, we designed two cascaded classifiers to distinguish neutral comfort zone and which PTSL it belongs to. The function and output of the cascaded classifiers are illustrated in Figure 3.10.

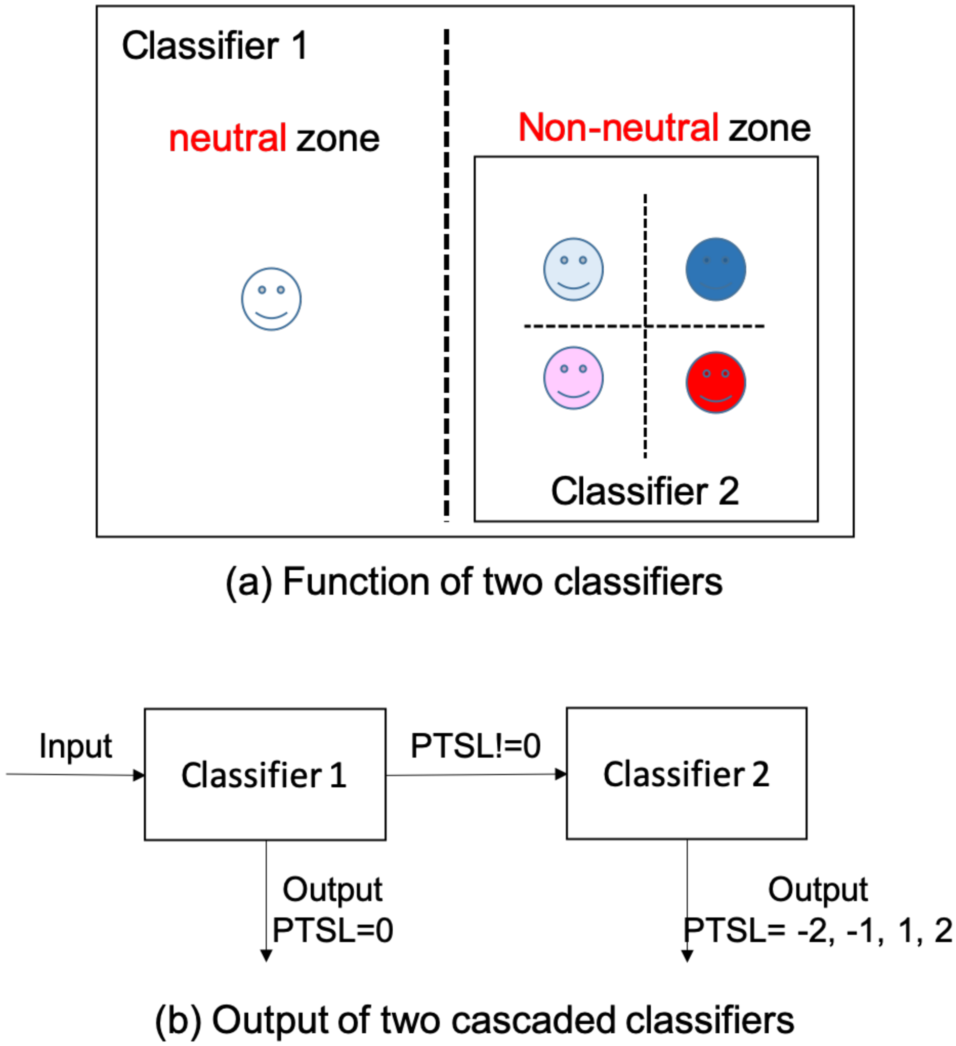


Figure 3.10: Cascaded Classifiers

3.6.2 Algorithm of Training Personalized Predictive Classifier

In this paper, we use a machine learning algorithm with incomplete supervision to resolve the multiclass classification of PTSL. This time questionnaire of SCL, other than thermal sensation response behavior, is adopted as the ground truth to verify the thermal sensation of an occupant at a specific time. Then, we expound the process of training these cascaded classifiers using incomplete supervised learning in the following algorithm tabulation.

Algorithm 3.1 Training process using incomplete supervised learning

```
Input: X(t), SCL
function SAMPLING()
    decide sampling interval based on size of training set
    sample buffer and put them into training set
end function
function UPDATE(X(t), SCL)
    put X(t), SCL into training set
    classifier2.fit()
    classifier1.fit()
    for x in training set do
        if probability(x | 1) is similar to probability(X(t) | 1) then
            set SCL of x = 1
        end if
    end for
end function
function REFINE()
    if size of training set is small then
        clean over fitting data
    else
        clean imprecise data
    end if
end function
while X(t) do
    if SCL is None then
        set SCL = 0
        push X(t), SCL into buffer
        continue
    else
        SAMPLING()
        UPDATE(X(t), SCL)
        REFINE()
        empty buffer
    end if
end while
```

Chapter 4

Simulation Results and Discussions

4.1 Performance Evaluation Metrics

In this paper, we adopted five commonly adopted metrics: accuracy, precision, recall, Area Under the Receiver Operating Characteristic (ROC) Curve (AUC), and Cohen's kappa coefficient.

4.1.1 Confusion Matrix

In order to clearly express selected metrics, confusion matrix is introduced at the very beginning.

		predicted class		
		p	**n**	**total**
actual class	$\mathbf{p}'$	TP	FN	P'
	$\mathbf{n}'$	FP	TN	N'
	total	P	N	

Table 4.1: Confusion Matrix

Where: p, P = positive; n, N = negative; TP = True Positive; FP = False Positive; TN = True Negative; FN = False Negative.

4.1.2 Accuracy

The accuracy represents the percentage of predicting the thermal sensation of occupants correctly, whose formula is shown in (4.1). However, it also causes an issue for imbalanced dataset. For instance, if a dataset consists of 80% data where $SCL = 0$, a naive predictive model is able to result in 80% accuracy.

$$Accuracy = \frac{TP + TN}{P + N} \tag{4.1}$$

4.1.3 Precision

Precision gives the result of the proportion of positive identification that is correct in fact. The formula of precision is shown in (4.2). Precision is also called positive predictive value (PPV).

$$Precision = \frac{TP}{TP + FP} \tag{4.2}$$

For multiclass situation, we use macro precision which is the average of precision of each class with same weight.

4.1.4 Recall

Recall demonstrates the proportion of actual positive example that is identified correctly. The formula of recall is shown in (4.3). Recall is also called true positive rate (TPR) or sensitivity.

$$Recall = \frac{TP}{TP + FN} \tag{4.3}$$

For multiclass situation, we use macro recall which is the average of recall of each class with same weight.

4.1.5 Area Under the Receiver Operating Characteristic Curve

ROC curve shows the performance of a classifier at all classification thresholds. The x-axis and y-axis of ROC curve are false positive rate (FPR) and true positive rate (TPR) respectively. The definitions of FPR and TPR are as follows.

$$TPR = \frac{TP}{TP + FN} \tag{4.4}$$

$$FPR = \frac{FP}{FP + TN} \tag{4.5}$$

AUC measures the two-dimensional area underneath the ROC curve from (0,0) to (1,1). The relationship between ROC and AUC is illustrated in Figure 4.1.

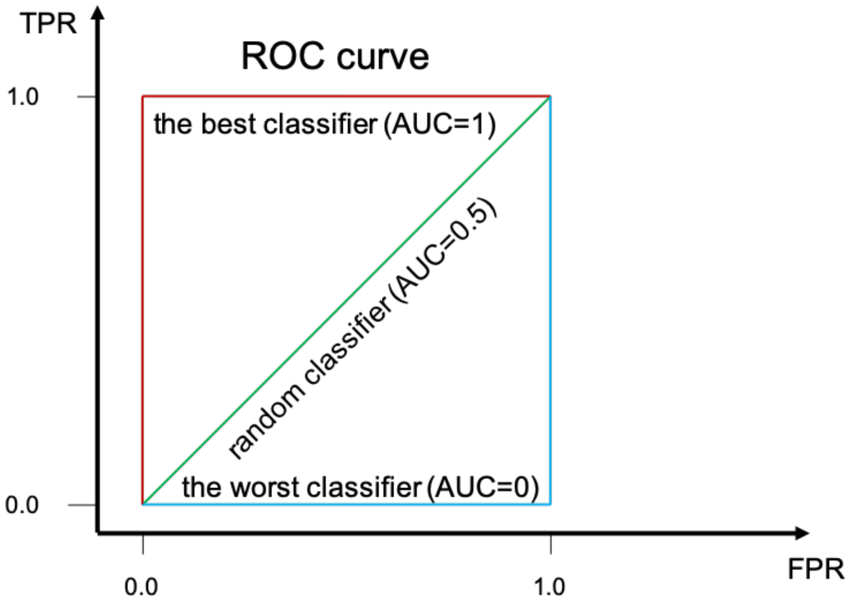


Figure 4.1: ROC and AUC

For multiclass situation, we use macro AUC which is the average of AUC of each class with same weight.

4.1.6 Cohen's Kappa Coefficient

Cohen's kappa measures the agreement between two raters who each classify N items into C mutually exclusive categories.[34] The definition of Cohen's kappa (κ) is shown in (4.6),

$$\kappa = \frac{p_o - p_e}{1 - p_e} \tag{4.6}$$

where p_o is accuracy and p_e is hypothetical probability of chance agreement. Cohen's Kappa has a range from 0-1 in general, with larger values indicating better reliability.

4.2 Performance of RFC

In order to evaluate the performance of RFC, we designed one RFC for PTC model and trained this RFC with the full dataset(i.e., including records of $SCL = 0$). The hyperparameters of RFC are as follows.

RFC: min sample leaf = 2, class weight = 'balanced'

The ratio of the training set and test set is 0.6:0.4 with stratification of labels, which means that examples of each class have the same ratio between the training set and test set. We ran this program 1000 times with randomly splitting training sets and test sets to get an average performance considering reducing contingency. The summary table is shown in Table 4.3.

Participant ID	Data Size	Average Accuracy(SD)	Average Precision(SD)	Average Recall(SD)	Average AUC(SD)	Average Cohen's kappa(SD)
F1	172	0.86(0.05)	0.88(0.06)	0.86(0.06)	0.98(0.02)	0.81(0.07)
F2	66	0.94(0.06)	0.88(0.13)	0.94(0.10)	0.99(0.02)	0.89(0.12)
F3	89	0.82(0.08)	0.81(0.09)	0.86(0.06)	0.82(0.08)	0.74(0.11)
M1	60	0.95(0.06)	0.94(0.11)	0.92(0.10)	0.99(0.01)	0.85(0.17)
M2	114	0.76(0.08)	0.82(0.08)	0.76(0.09)	0.95(0.04)	0.68(0.1)
M3	114	0.85(0.06)	0.85(0.09)	0.79(0.09)	0.96(0.04)	0.76(0.09)
M4	36	0.82(0.08)	0.81(0.09)	0.86(0.06)	0.97(0.02)	0.74(0.11)
Average	93	0.87(0.07)	0.88(0.09)	0.86(0.09)	0.97(0.03)	0.79(0.13)
Median	102	0.86(0.06)	0.87(0.09)	0.86(0.06)	0.98(0.02)	0.80(0.11)

Table 4.2: Predictive performance of RFC(SD:standard deviation)

Related work	Performance index	Result	Result of RFC
[14]	Median AUC	0.73	0.98
[13]	Median AUC	0.79	0.98
	Median accuracy	0.78	0.86
	Median Cohen's kappa	0.24	0.80
[15]	Average accuracy	0.7(SD:0.08)	0.87(SD:0.07)

Table 4.3: Performance of RFC vs. related work

4.3 Performance of PPC vs. RFC

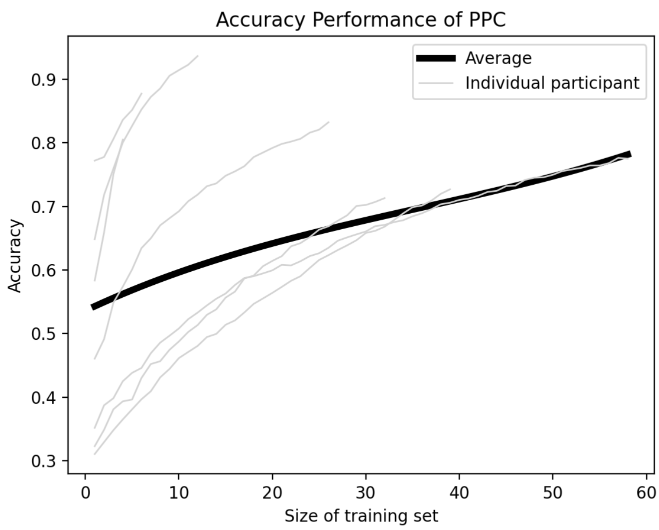
Hyperparameters for two classifier of PPC are set as:

RFC 1: min sample leaf = 2, class weight = 'balanced'
RFC 2: min sample leaf = 1, class weight = 'balanced'

For the sake of evaluating the performance with a small-size training set, we changed the ratio of the training set and test set to 0.5:0.5. To emphasize again, what needs special attention is that records of $SCL = 0$ are removed from training sets of PPC, but records of $SCL = 0$ are still kept in training sets of RFC and test sets of RFC and PPC. The result was the average of repeating 1000 times with no specified random seed.

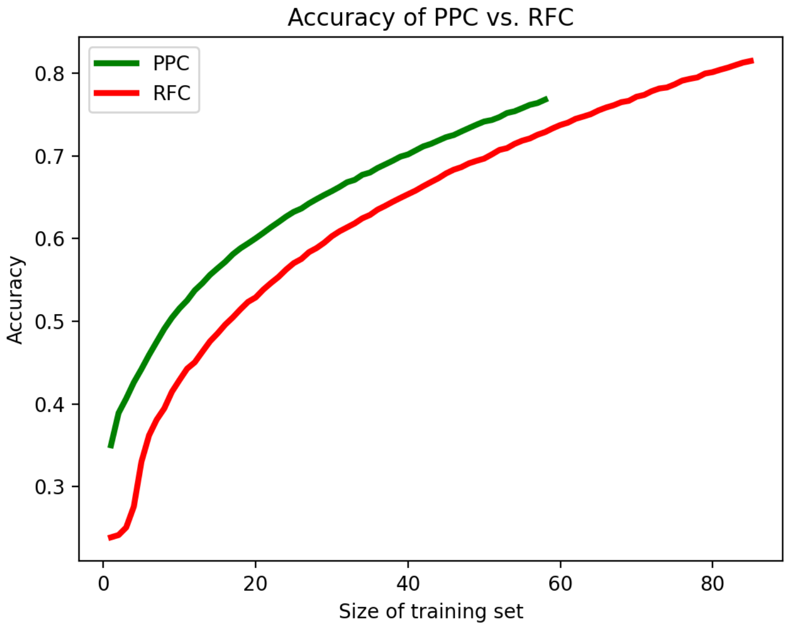
4.3.1 Accuracy

The summary figure is shown in Figure 4.2. First of all, except for F3, the results of other participants show that PPC, which is designed using an online learning method, improves performance quickly than RFC. For results of F2, M2, and M3, it is clear that the performance after training is equivalent to or beyond one RFC. For results of F2, M1, and M4, the size of the training data is small, and the samples of $SCL = 0$ consists of more than 60% of the training set. So that the performance does not get equivalent because update and refine function has not been executed for enough times. As to the result of F3, it is degraded.

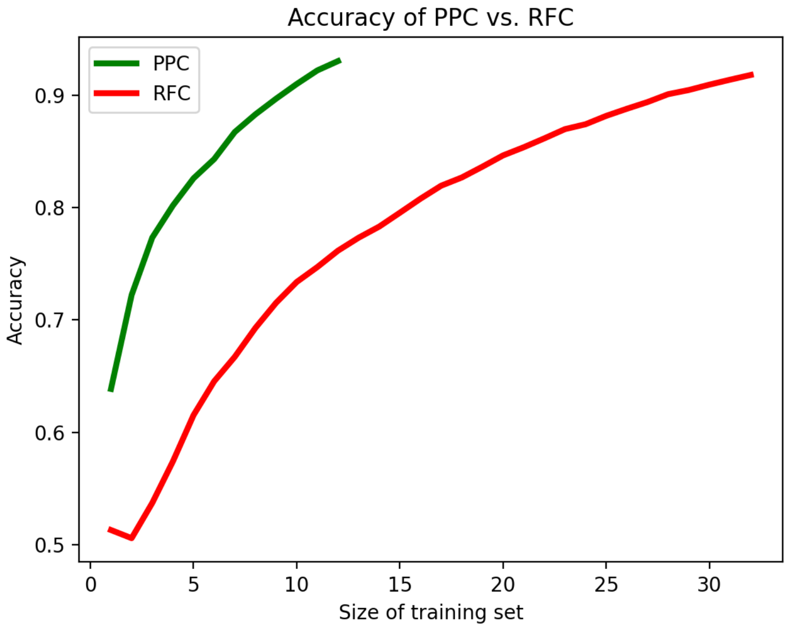


(a) Average

Figure 4.2: Accuracy result of PPC vs. RFC

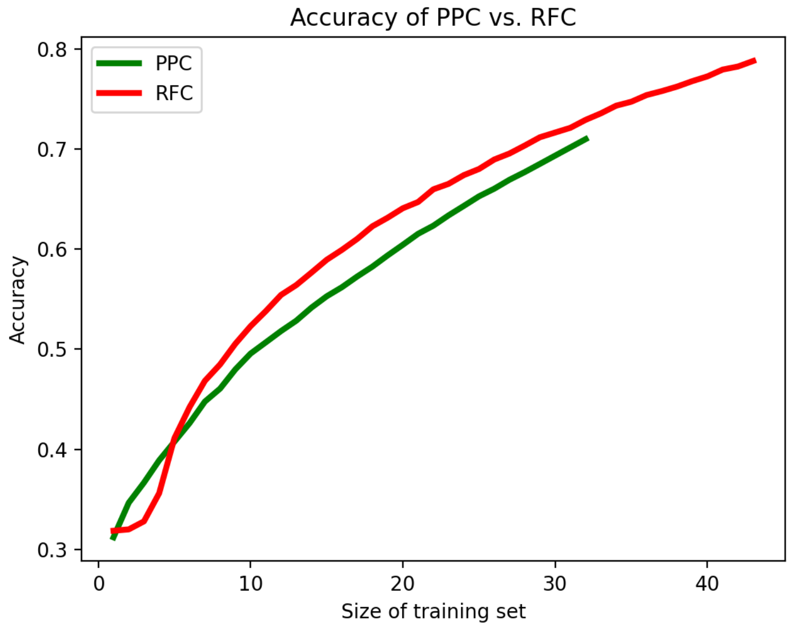


(b) F1

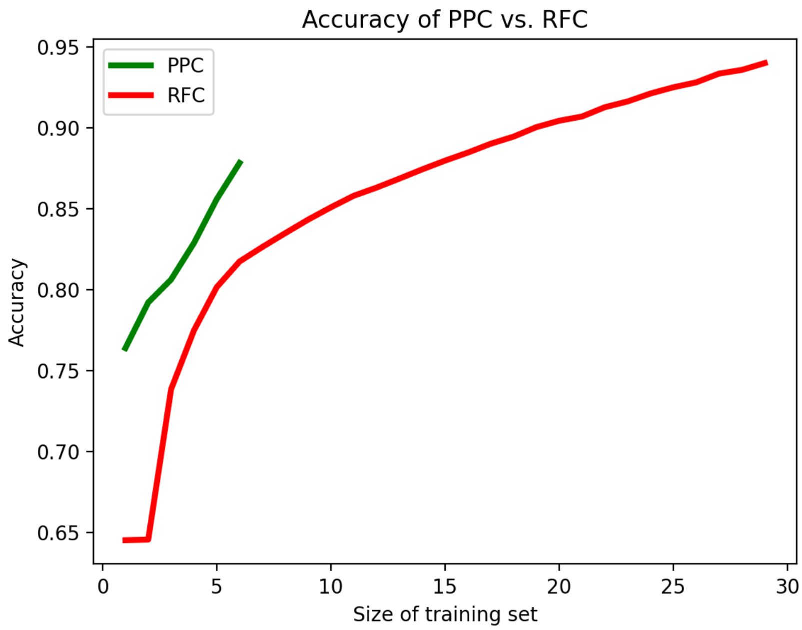


(c) F2

Figure 4.2: Accuracy result of PPC vs. RFC

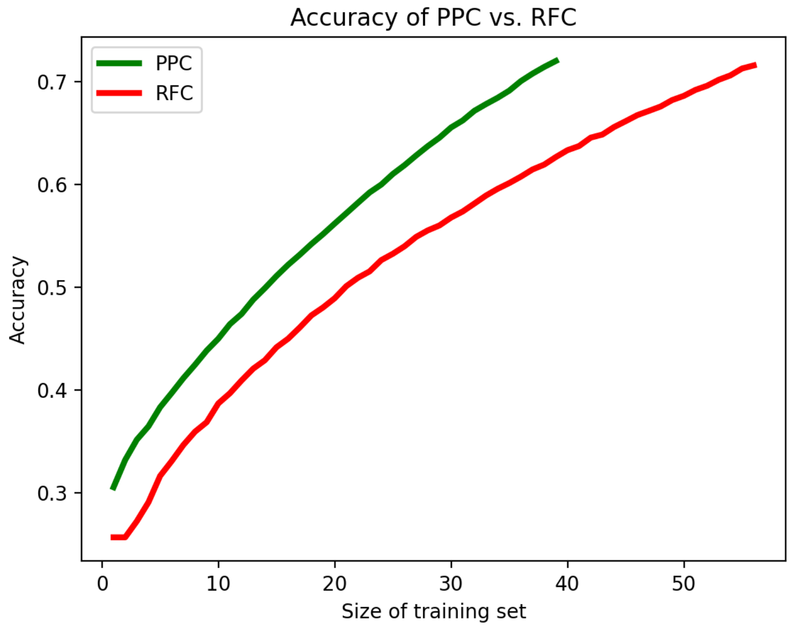


(f) F3

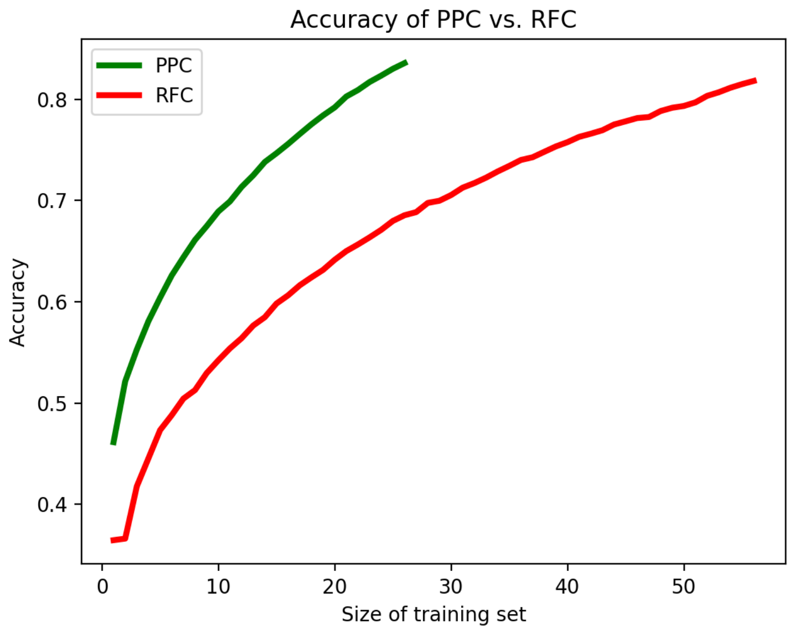


(g) M1

Figure 4.2: Accuracy result of PPC vs. RFC

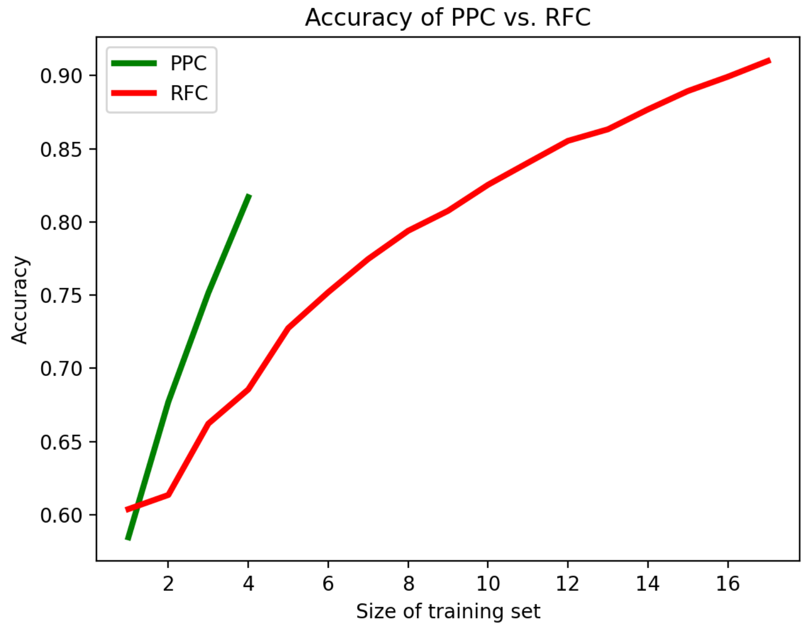


(j) M2



(k) M3

Figure 4.2: Accuracy result of PPC vs. RFC

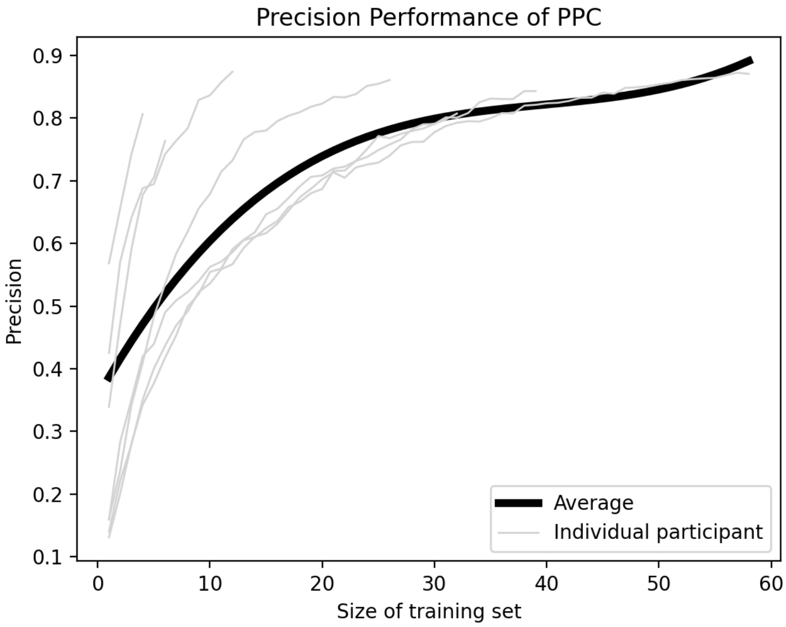


(n) M4

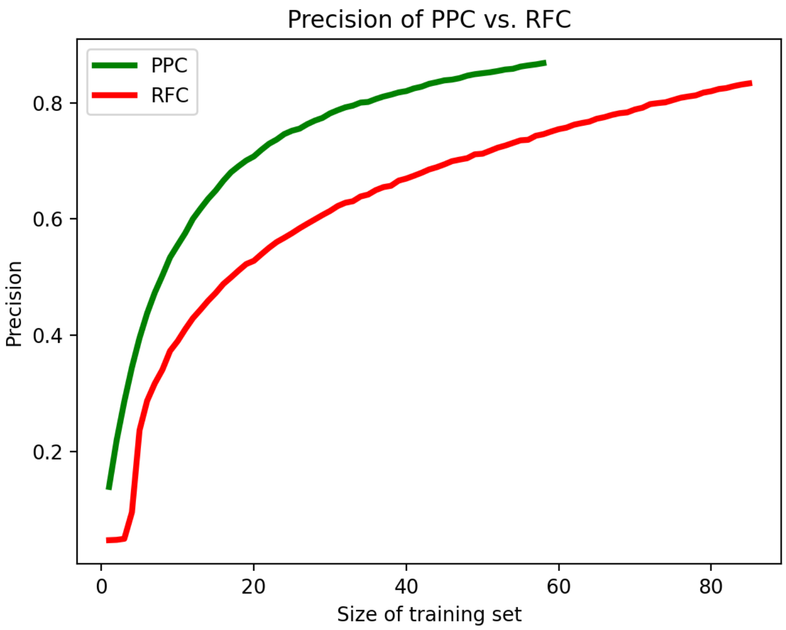
Figure 4.2: Accuracy result of PPC vs. RFC

4.3.2 Precision

The summary figure is shown in Figure 4.3. First of all, the results of all participants show that PPC improves its performance more quickly than RFC. For results of F1, F2, F3, M2, and M3, it is evident that the performance after training is equivalent to or beyond one RFC. For results of M1 and M4, the size of training data is small so that the performance does not get equivalent because update and refine function has not been executed enough times. Compared to accuracy, the result of F2 gets improved. The reason for it is, compared with M1 and M4, F2 has fewer samples of $SCL = 0$. As mentioned before, the neutral zone is ambiguous. Index of precision gives the result of proportion of the positive identification that is correct in fact. So it means PPC is good at recognizing scenes seen before.

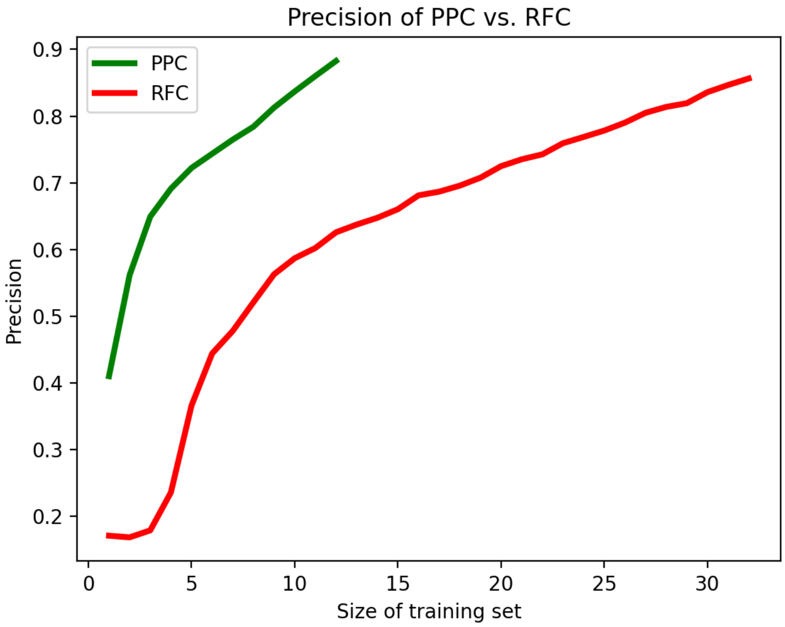


(a) Average

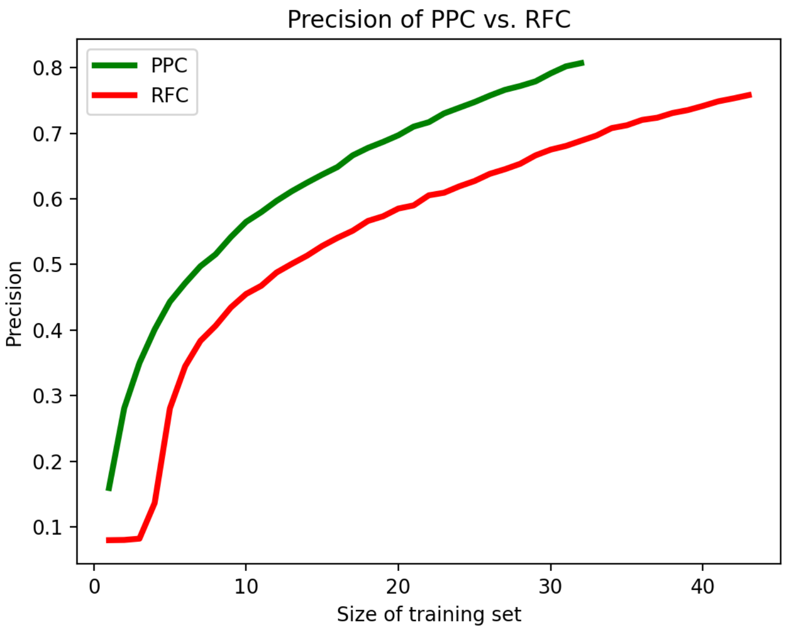


(b) F1

Figure 4.3: Precision result of PPC vs. RFC

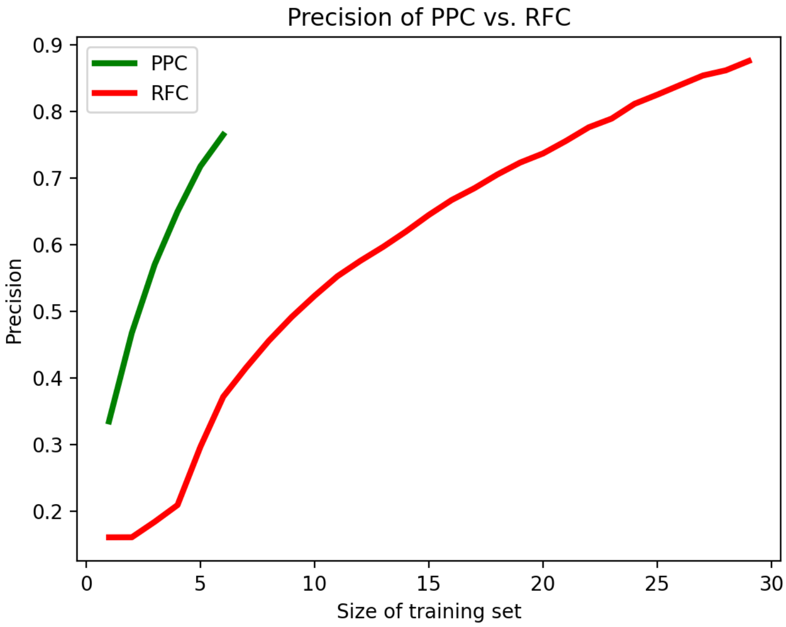


(c) F2

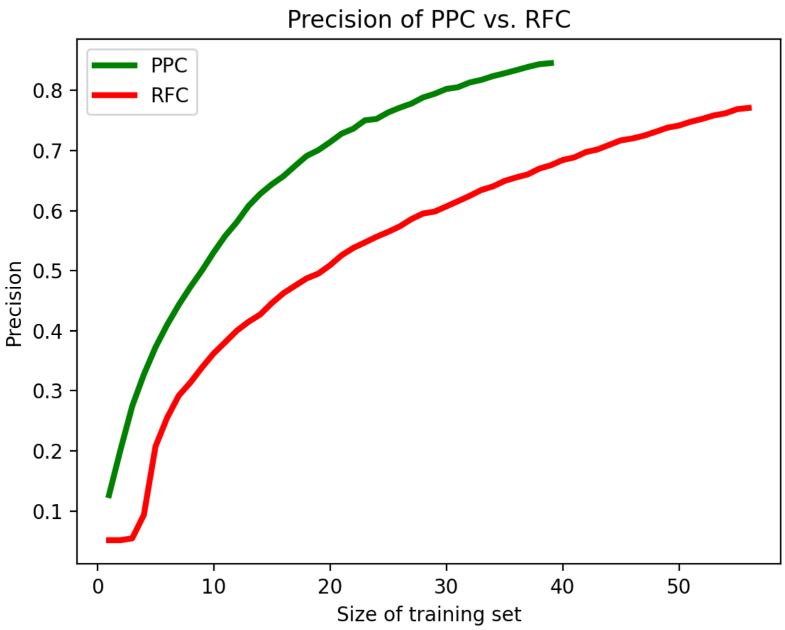


(d) F3

Figure 4.3: Precision result of PPC vs. RFC

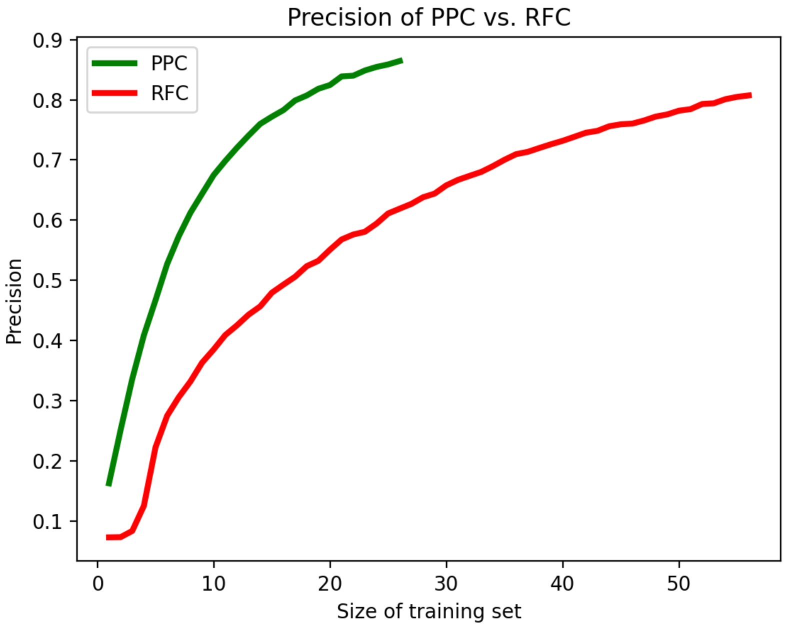


(g) M1

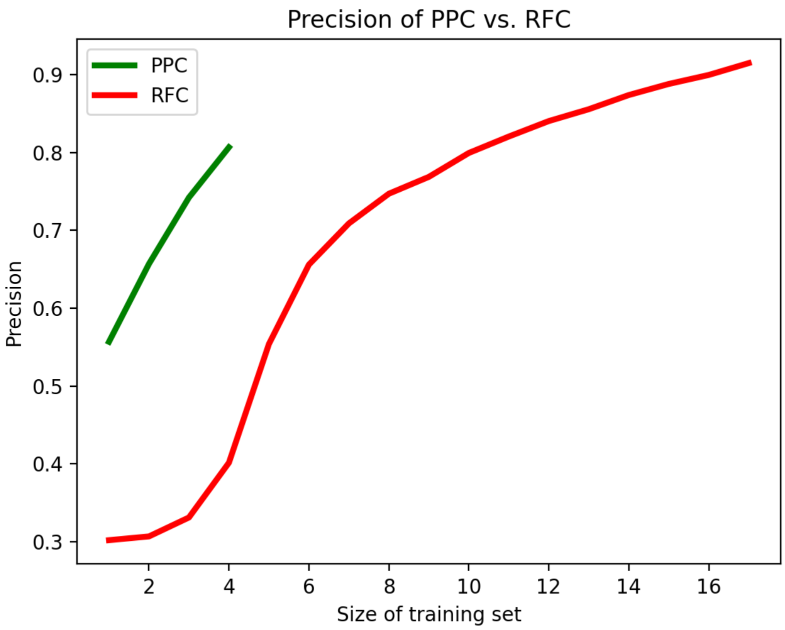


(h) M2

Figure 4.3: Precision result of PPC vs. RFC



(k) M3

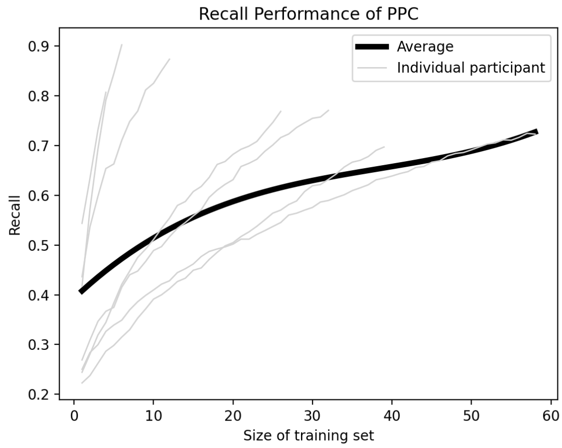


(l) M4

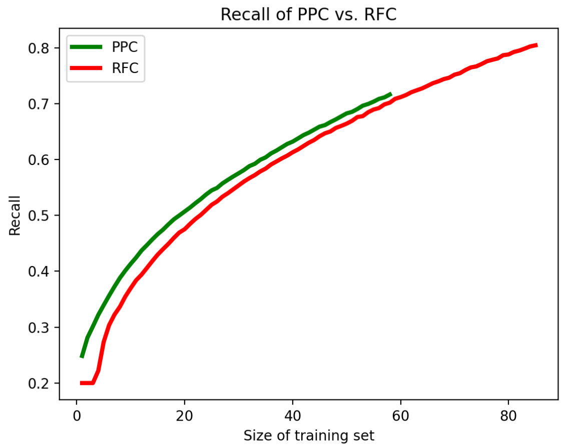
Figure 4.3: Precision result of PPC vs. RFC

4.3.3 Recall

The summary figure is shown in Figure 4.4. First of all, except for F3, the results of other participants show that PPC improves its performance more quickly than RFC. For results of F2, M1, M2, and M3, it shows that the performance after training is equivalent to or beyond one RFC. For results of F1 and M4, it shows that PPC is not skilled in recognizing all scenario of neutral comfort. We paid much more attention to enable PPC judge correctly. In other words, PPC makes fewer mistakes than radical predictions.

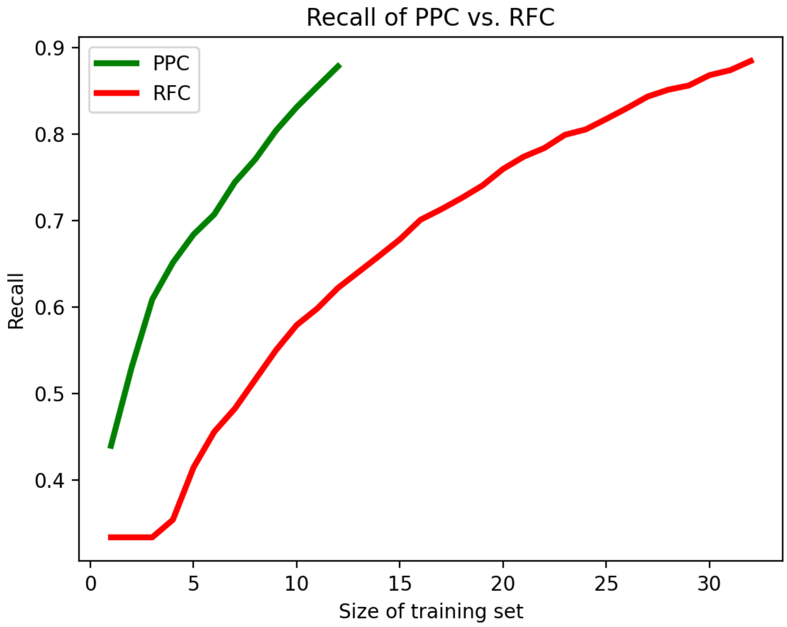


(a) Average

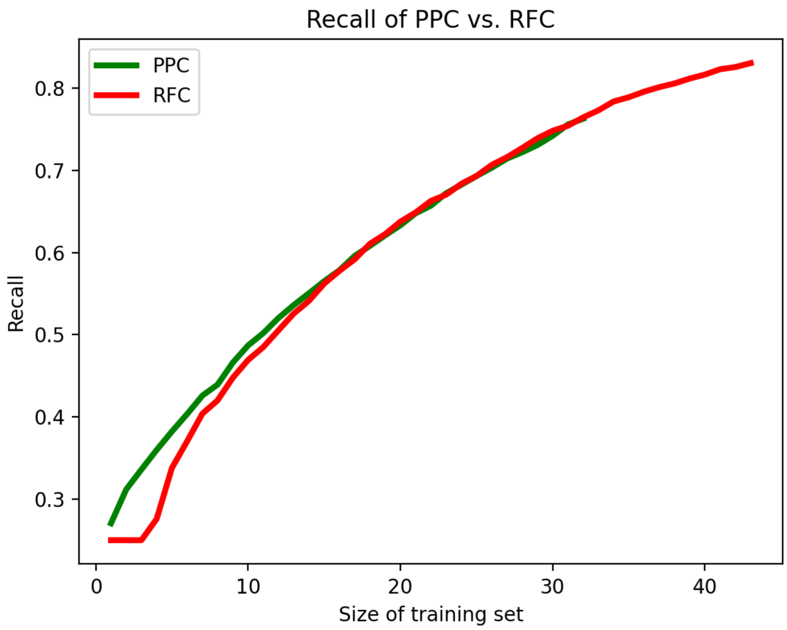


(b) F1

Figure 4.4: Recall result of PPC vs. RFC

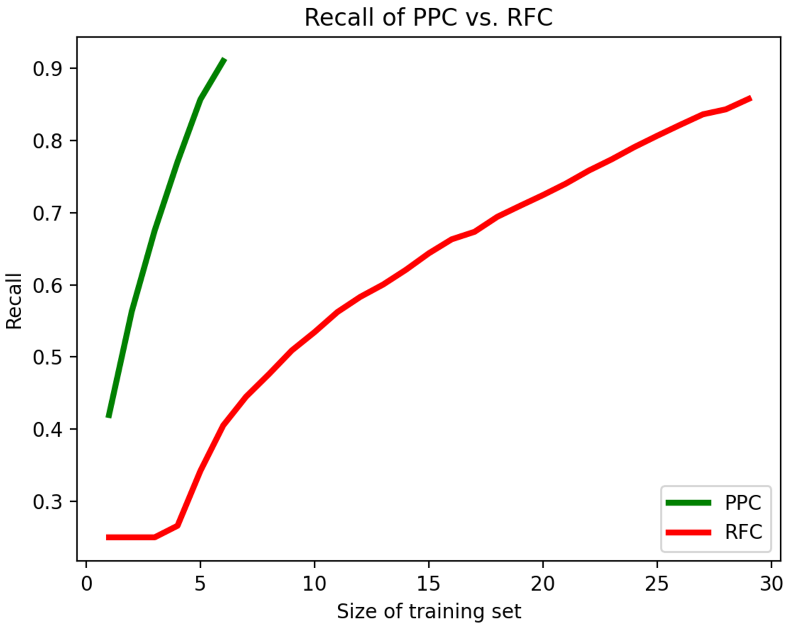


(c) F2

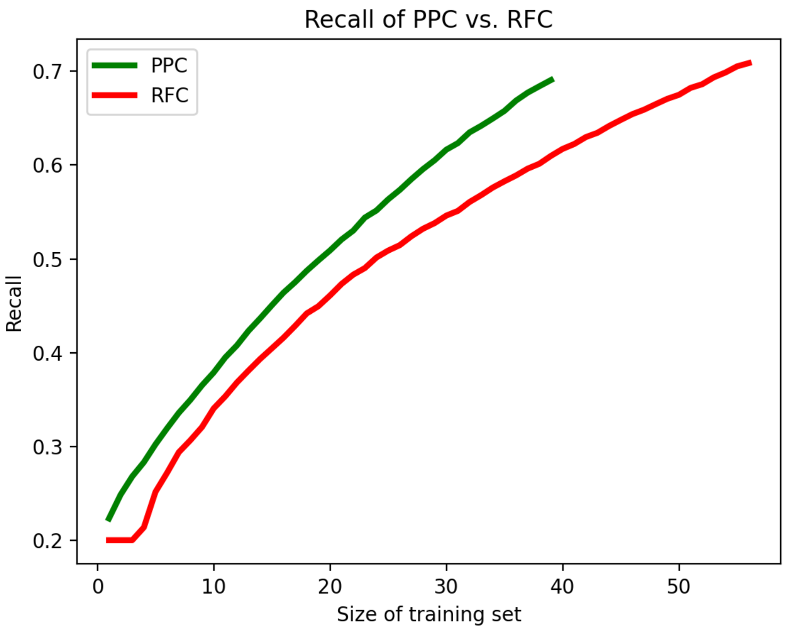


(d) F3

Figure 4.4: Recall result of PPC vs. RFC

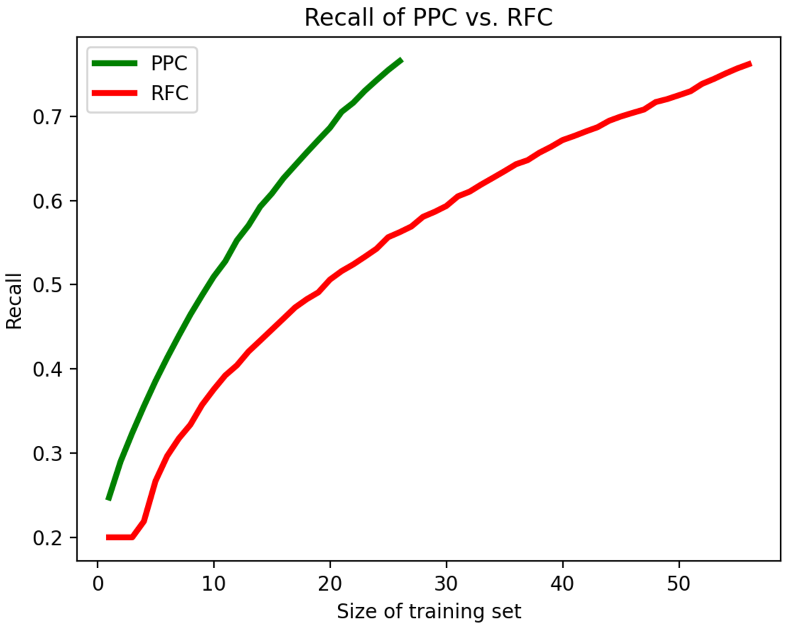


(g) M1

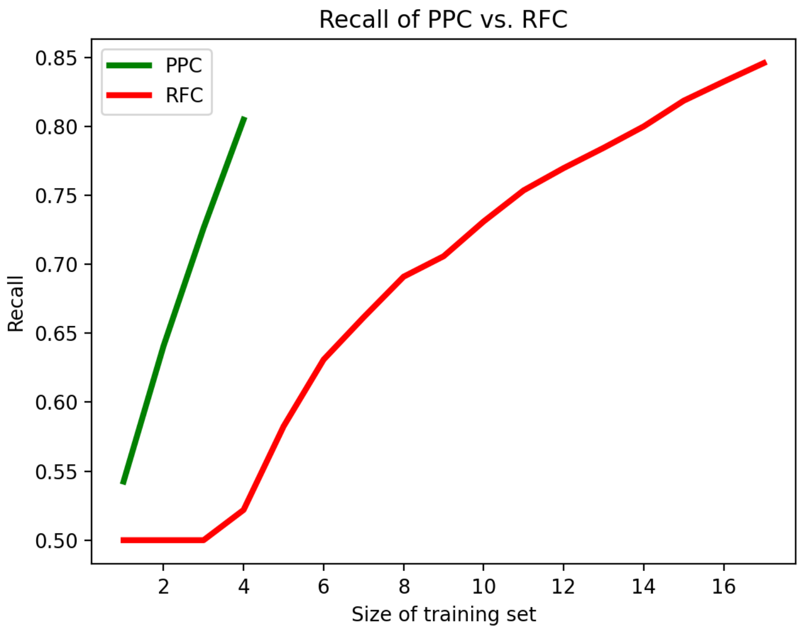


(h) M2

Figure 4.4: Recall result of PPC vs. RFC



(k) M3

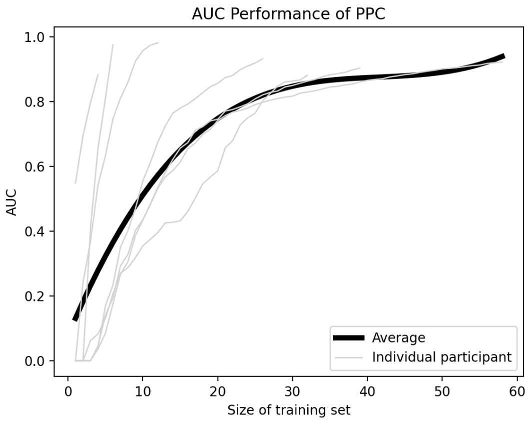


(l) M4

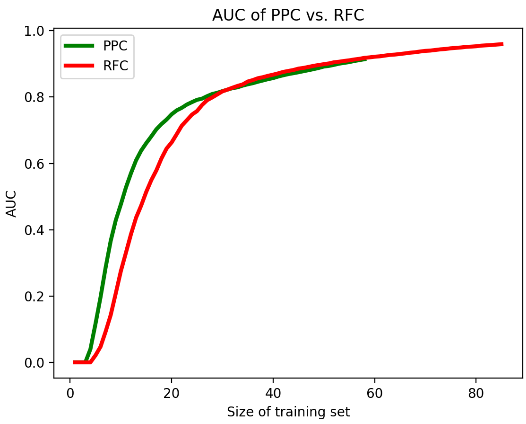
Figure 4.4: Recall result of PPC vs. RFC

4.3.4 AUC

Note that AUC was calculated after all classes appeared, which is different from others. The summary figure is shown in Figure 4.5. First of all, except for F3, the results of other participants show that PPC improves its performance more quickly than RFC. For results of F2, M1, M2, and M3, it shows that the performance after training is equivalent to or beyond one RFC. All results are in the acceptable level, and it means that PPC distinguishes different classes very clearly.

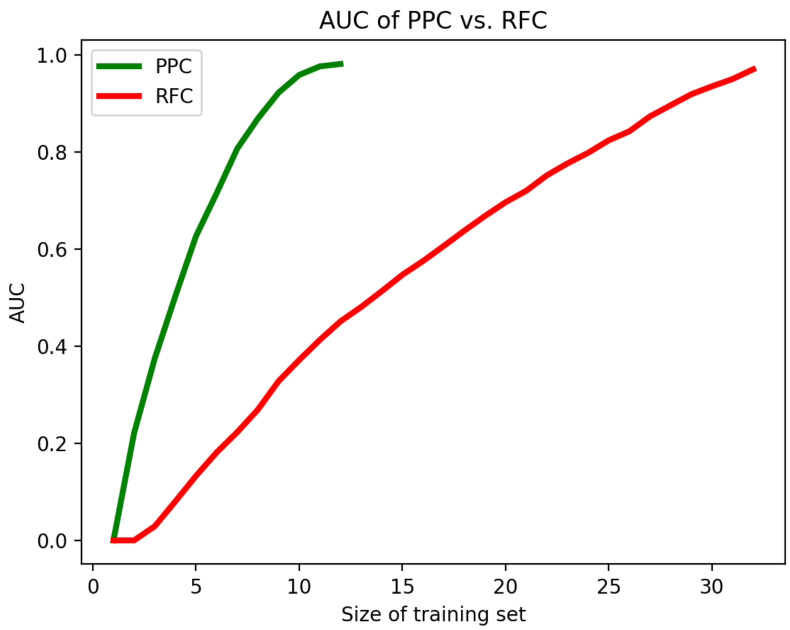


(a) Average

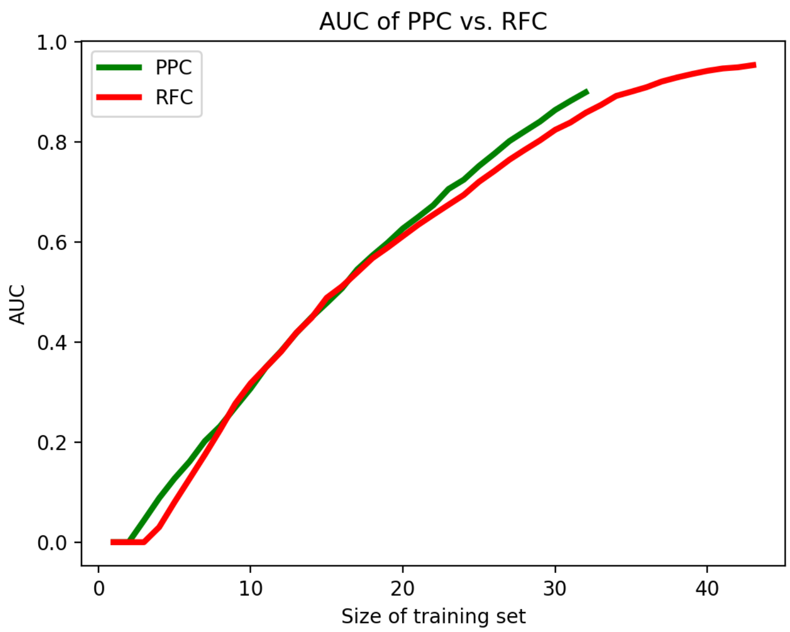


(b) F1

Figure 4.5: AUC result of PPC vs. RFC

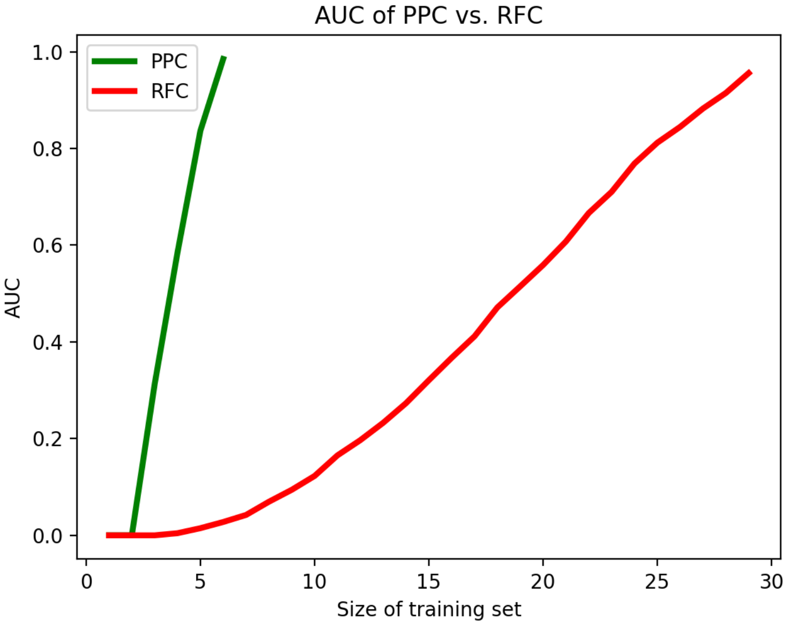


(c) F2

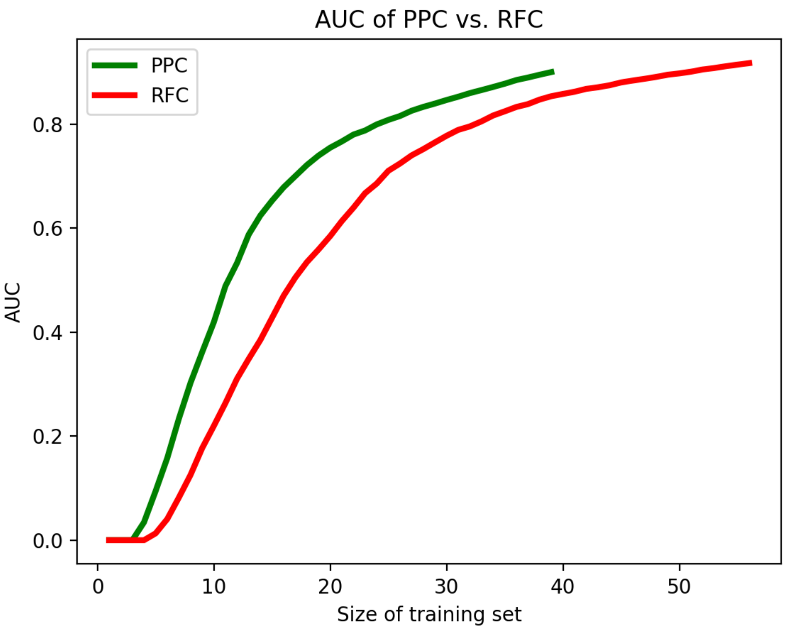


(d) F3

Figure 4.5: AUC result of PPC vs. RFC

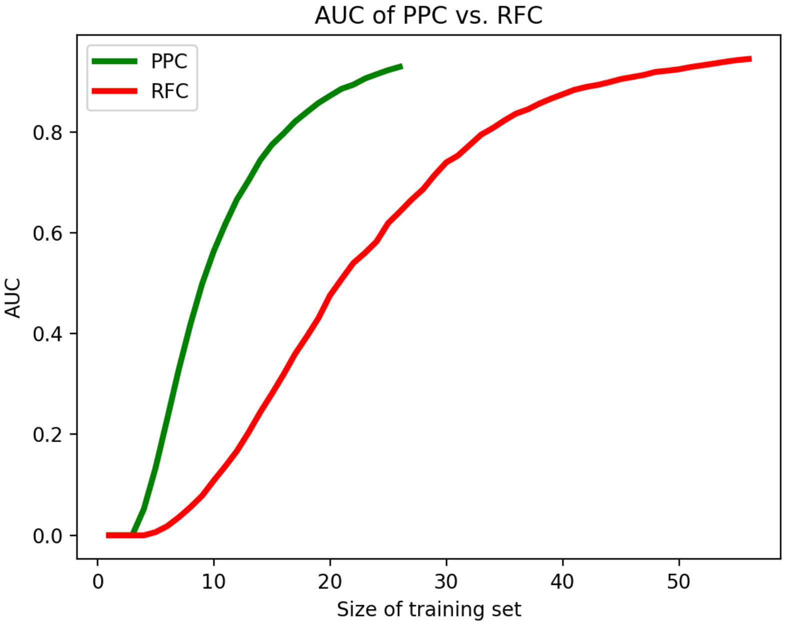


(g) M1

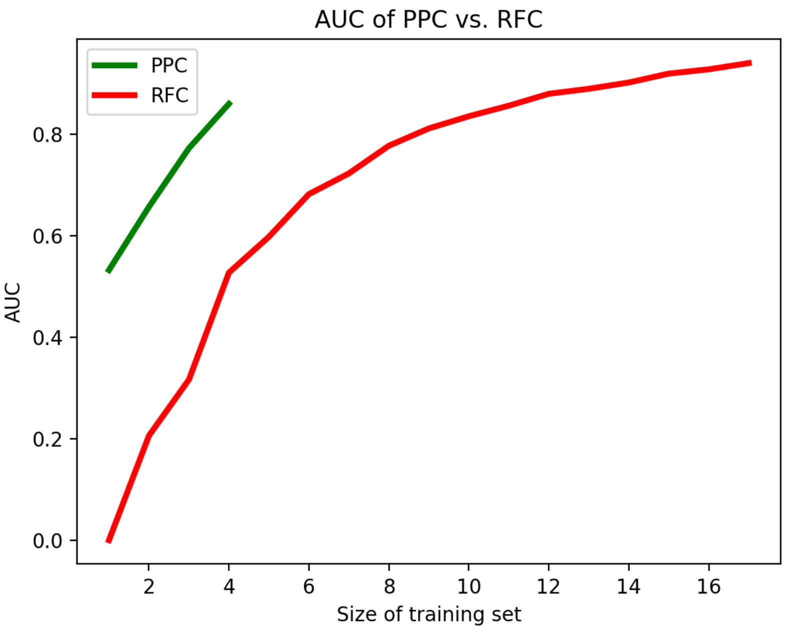


(h) M2

Figure 4.5: AUC result of PPC vs. RFC



(k) M3

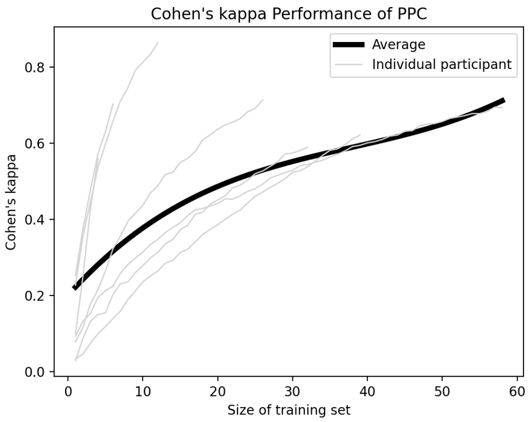


(l) M4

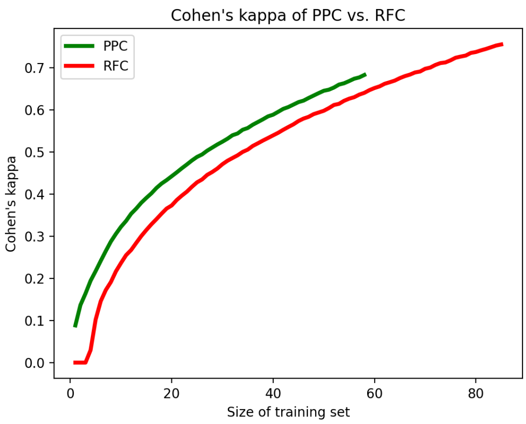
Figure 4.5: AUC result of PPC vs. RFC

4.3.5 Cohen's kappa Coefficient

The summary figure is shown in Figure 4.6. First of all, except for F3, the results of other participants show that PPC improves its performance more quickly than RFC. For results of F2, M2, and M3, it shows that the performance after training is equivalent to or beyond one RFC. Similar to recall, PPC lays emphasis on predicting positive samples correctly but not good at finding a boundary for the neutral zone. In other words, the classifier 1 in PPC is not well trained.

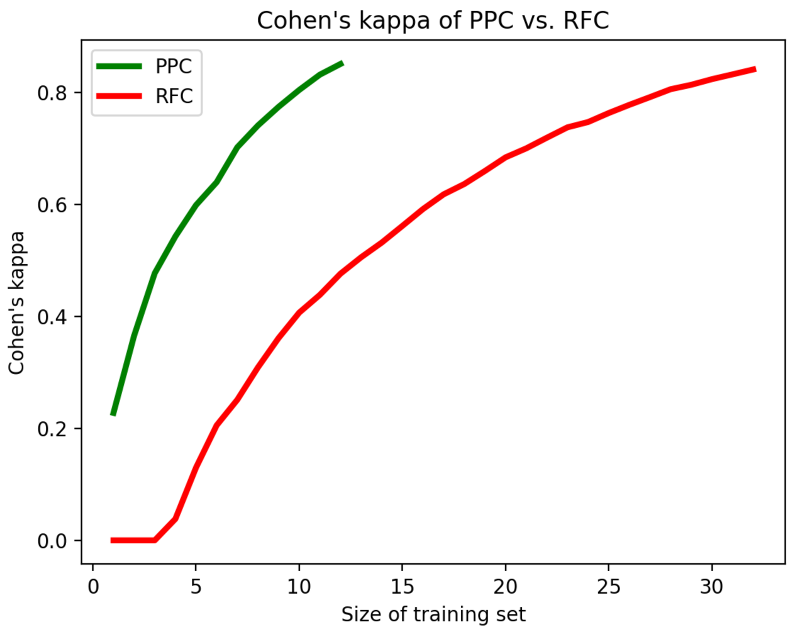


(a) Average

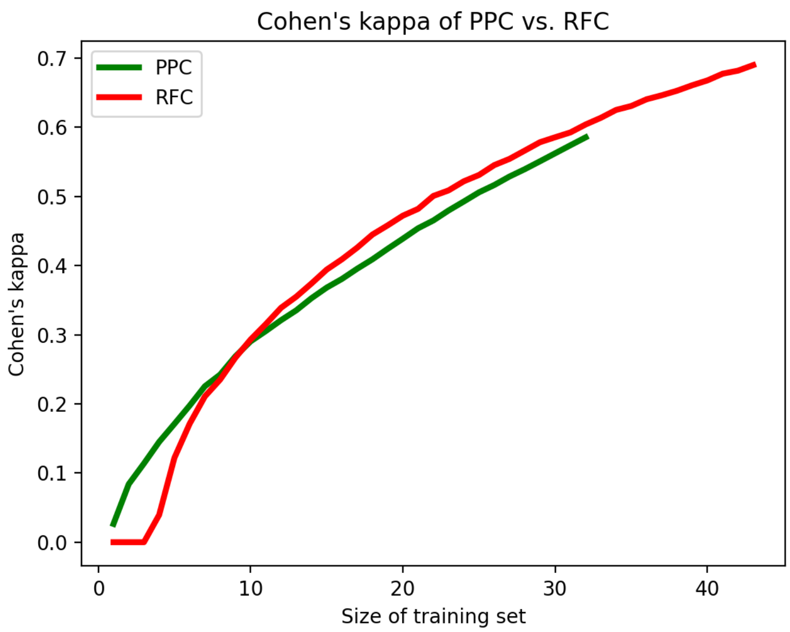


(b) F1

Figure 4.6: Cohen's kappa result of PPC vs. RFC

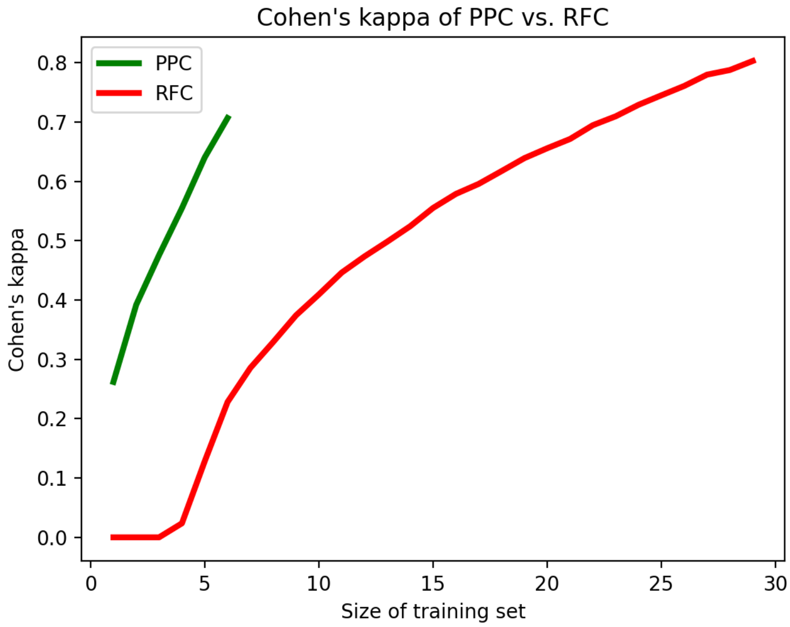


(c) F2

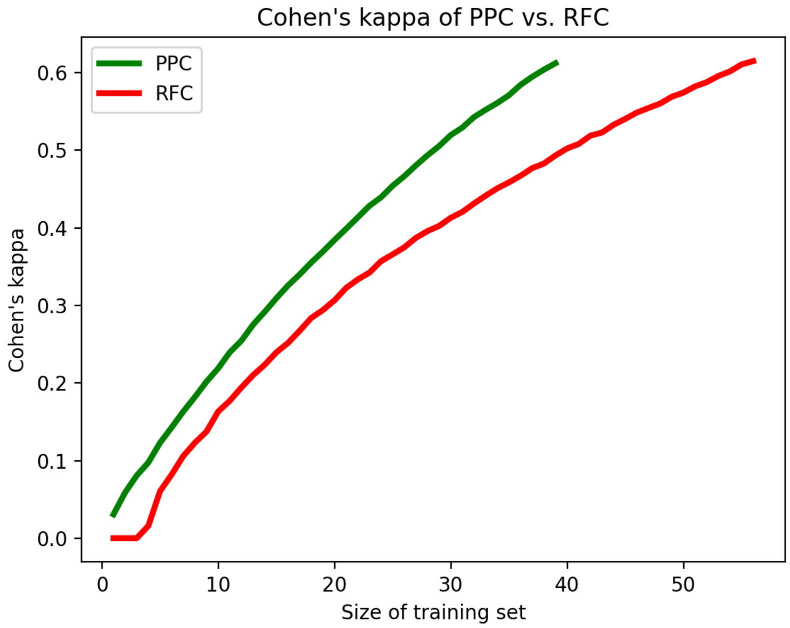


(d) F3

Figure 4.6: Cohen's kappa result of PPC vs. RFC

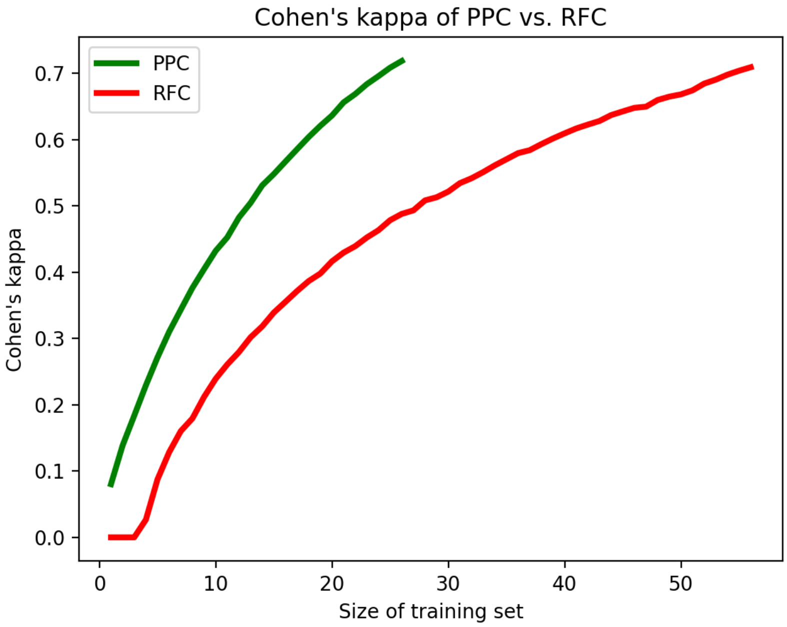


(g) M1

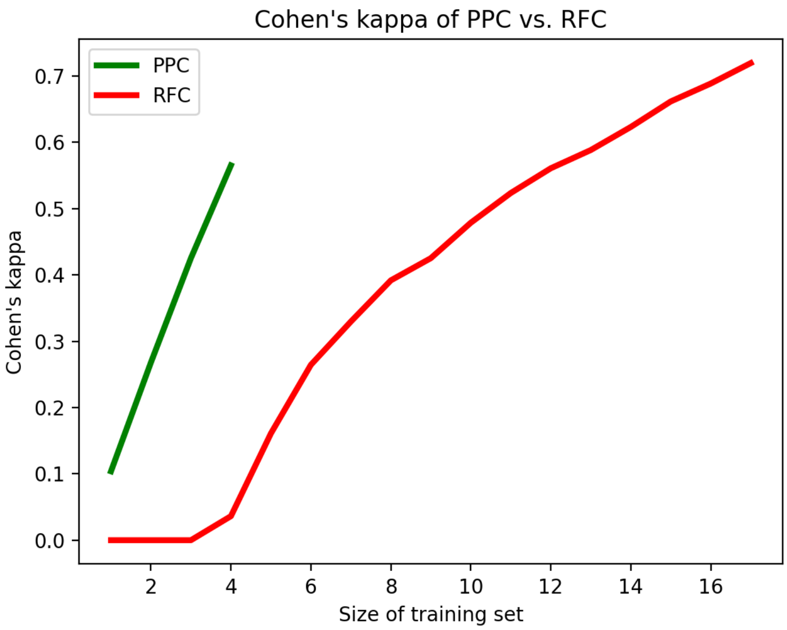


(h) M2

Figure 4.6: Cohen's kappa result of PPC vs. RFC



(k) M3



(l) M4

Figure 4.6: Cohen's kappa result of PPC vs. RFC

4.3.6 Summary of PPC vs. RF

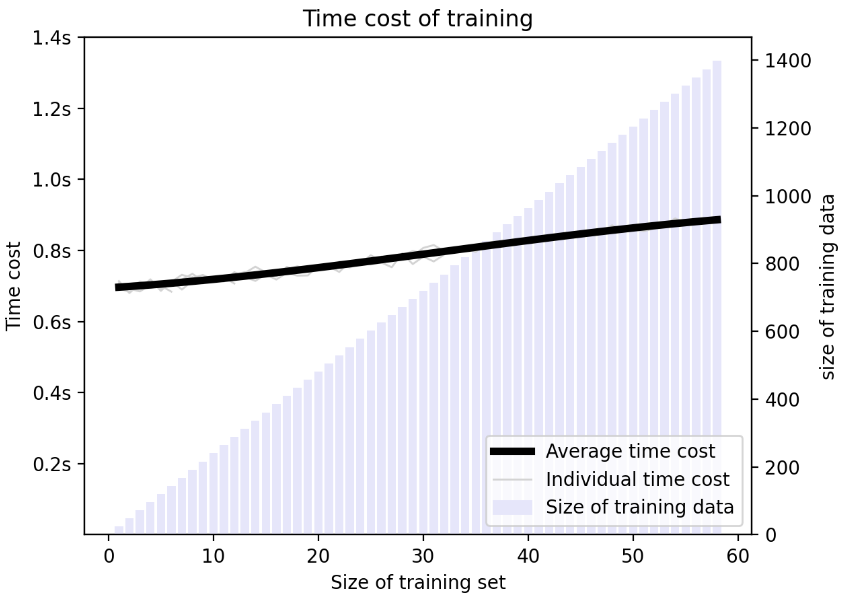
The performance improvements in percentage compared with RFC is shown in Table 4.4.

Participant ID	Average accuracy 5%/10%/25%/50%/100%	Average precision 5%/10%/25%/50%/100%	Average recall 5%/10%/25%/50%/100%	Average AUC 5%/10%/25%/50%/100%	Average Cohen's kappa 5%/10%/25%/50%/100%
F1	54.4/24.7/15.3/9.1/5.3	260.3/49.2/36.4/27.3/16.4	45/15.6/7.9/3.9/2.1	-/208/24/0.1/-0.4	557.1/53.8/22.4/11.4/6.7)
F2	42.8/43.9/39.6/30/22.1	233.1/262.6/193/60/41	59/82.6/84.1/54.3/41.2	-/1192.6/526.2/262/117.5	-/-/1319.3/179.6/78.7
F3	11.8/-0.8/-5.1/-6.2/-2.6	326.7/57.9/24.7/20.8/17.2	34.4/13.1/4.1/0.7/-0.2	-/59.4/-2.6/1.1/4.8	-/-40.3/0.8/-0.7-3.1
M1	22.7/22.7/9.2/7/7.4	189.9/189.9/209/209.9/105.6	125.4/125.4/170.1/189.7/124.7	-/-/-/1316.1/3473.9	-/-/-/2251.1/209.8
M2	29.1/21.1/16.9/14.8/14.9	403.9/79.6/46.4/40.3/25.2	34/20.1/11.7/10.4/13.3	-/622.1/85.3/29.1/5.4	-/103.4/35.3/25.4/24.1
M3	32.4/30.4/28.9/26.2/21.9	302.3/226.4/84.5/66.4/39.6	62/62.4/39.2/36.7/36.1	-/-/649.5/204.8/44.8	-/761.3/109.8/66.8/47.3/
M4	10.4/10.4/10.4/13.5/19.2	114.1/114.1/114.1/124.1/100.9	28.2/28.2/28.2/45.2/54.2	218.9/218.9/218.9/144.1/63.1	-/-/-/-/1459.6

Table 4.4: Performance improvement of PPC vs. RFC in percentage

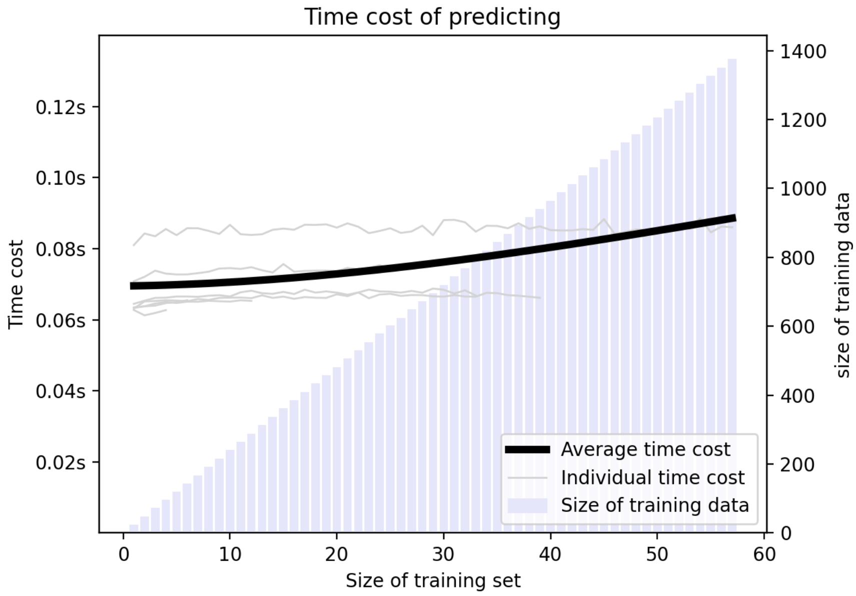
4.4 Time Cost of PPC

We conducted time cost evaluation to prove that PPC is able to be implemented as a real-time prediction into CPHCS. It is shown in Figure 4.7 that time consuming for predicting is low, but time-consuming for training is getting close to one second when the size of training sets increases. The refinement procedure for decreasing computational cost is future work.



(a) Training

Figure 4.7: Time Cost



(b) Predicting

Figure 4.7: Time Cost

4.5 Discussion

Concluded from the results above, PPC learns faster than one RFC in most situations. Meanwhile, it achieves equivalent or better prediction performance for half of the participants scaled by five metrics: accuracy, precision, recall, AUC, and Cohen's kappa. The reason for degradation for some situations is due to extremely unbalanced training sets, size of data, and the inherent weakness of distinguishing the neutral comfort zone without being trained with these data from the architecture design. However, this degradation can be solved by enlarging the training set with the conclusion in[14] of the requirement of over 60 responses.

Chapter 5

Conclusion

5.1 Concluding Remark

Thermal comfort is of high subjectivity, which is diverse among occupants. Hence, based on the concept of CPHCS, correct inference of personal comfort sensation makes HVAC controller 'understands' what human really needs. In this paper, we present a modified personal comfort model that ensures the availability of psychological parameters(e.g., thermal sensation) for CPHCS. From the results, we concluded:

1) A novel PTC model has been proposed. It is designed that which variables are related to personal thermal sensation and how these variables are adopted to work for CPHCS. The result of data analysis proves that PTC is necessary for that correlation coefficient differs over individuals.
2) Under the structure of PTC model, machine learning is proved to be capable of improving comfort predictions of actual occupants in smart homes. Random Forest algorithm has been evaluated, and it is one suitable supervised machine learning algorithm for predicting personal thermal sensation due that it produced satisfactory performance. The median performance of PTC model using an RFC is 0.86, 0.87, 0.86, 0.98, 0.80 for accuracy, precision, recall, AUC, and Cohen's kappa, respectively.
3) With the assumption that occupant's thermal sensation behavior is not obvious, we proposed PPC that is trained using only discomfort labels. Based on RF, the proposed PPC is able to judge the neutral comfort zone precisely without training date of neutral comfort, which is named as incomplete learning in this paper. Meanwhile, PPC adopts online learning methods to overcome the limitation of requiring a considerable amount of training data. The results show that PPC improved its performance more quickly in most cases and got equivalent or better performance in more than half of situations(except Cohen's kappa, which is less than a half).

PTC model enhanced by PPC provides the real-time prediction of personal thermal sensation, which represents occupants' comfort level. What's more, it offers Personal Correlation Analysis module reliable sensation inference, which is of necessity for PMV-driven HVAC control system(e.g., EETCC). However, there are many questions unsolved, such as how to divide a comfort zone and a discomfort zone more precisely. Meanwhile, the relationship between thermal sensation and thermal preference is still unclear. These topics are on the way to go along the development of CPHCS.

5.2 Contributions

In this paper, we explored the adoption of incomplete supervised learning with online learning method for CPHCS to predict personal thermal sensation. The predictions of personal thermal sensation help systems in CPHCS satisfy occupants with desired thermal comfort level, and generate specific user profiles for energy saving. In conclusion, we made the following contributions to PTC model:

1) We proposed a modified CPHCS framework and a PTC model. The CPHCS framework if an extension of CPHS, so it take human factors more into the balance between systems and environments.
2) We conducted the correlation analysis and presented the function of PTC model in the field of a smart home, whose name is iHouse.
3) we evaluated the appropriateness of RF for prediction of personal thermal sensation. Then we compared it with the proposed PPC, which showed a fast learning pace.

References

[1] M. Frontczak, S. Schiavon, J. Goins, E. Arens, H. Zhang, and P. Wargocki, "Quantitative relationships between occupant satisfaction and satisfaction aspects of indoor environmental quality and building design," *Indoor air*, vol. 22, no. 2, pp. 119–131, 2012.

[2] D. P. Wyon, "The effects of indoor air quality on performance and productivity," *Indoor air*, vol. 14, no. 1, pp. 92–101, 2004.

[3] K. W. Tham and H. C. Willem, "Room air temperature affects occupants' physiology, perceptions and mental alertness," *Building and Environment*, vol. 45, no. 1, pp. 40–44, 2010.

[4] J. G. Allen, P. MacNaughton, J. G. C. Laurent, S. S. Flanigan, E. S. Eitland, and J. D. Spengler, "Green buildings and health," *Current Environmental Health Reports*, vol. 2, no. 3, pp. 250–258, 2015.

[5] A. ANSI/ASHRAE Standard 55-2013, "Thermal environmental conditions for human occupancy," 2013.

[6] I. ISO, "7730: Ergonomics of the thermal environment—analytical determination and interpretation of thermal comfort using calculation of the pmv and ppd indices and local thermal comfort criteria," *Management*, vol. 3, no. 605, p. e615, 2005.

[7] P. O. Fanger *et al.*, "Thermal comfort. analysis and applications in environmental engineering." *Thermal comfort. Analysis and applications in environmental engineering.*, 1970.

[8] B. L. R. Stojkoska and K. V. Trivodaliev, "A review of internet of things for smart home: Challenges and solutions," *Journal of Cleaner Production*, vol. 140, pp. 1454–1464, 2017.

[9] Y. Fang, Y. Lim, S. E. Ooi, C. Zhou, and Y. Tan, "Study of human thermal comfort for cyber–physical human centric system in smart homes," *Sensors*, vol. 20, no. 2, p. 372, 2020.

[10] L. Yuto, O. S. En, M. Yoshiki, T. T. Kin, R. Alfred, and T. Yasuo, "Implementation of energy efficient thermal comfort control for cyber-physical home systems," *Advanced Science Letters*, vol. 23, no. 11, pp. 11 530–11 534, 2017.

[11] F. Auffenberg, S. Stein, and A. Rogers, "A personalised thermal comfort model using a bayesian network," in *Twenty-Fourth International Joint Conference on Artificial Intelligence*, 2015.

[12] P. Tuomaala, R. Holopainen, K. Piira, and M. Airaksinen, "Impact of individual characteristics such as age, gender, bmi and fitness on human thermal sensation," in *Proceedings of the thirteen Internatinal Building Performance Simulation Association conference*, 2013, pp. 2305–2311.

[13] S. Liu, S. Schiavon, H. P. Das, M. Jin, and C. J. Spanos, "Personal thermal comfort models with wearable sensors," *Building and Environment*, vol. 162, p. 106281, 2019.

[14] J. Kim, Y. Zhou, S. Schiavon, P. Raftery, and G. Brager, "Personal comfort models: predicting individuals' thermal preference using occupant heating and cooling behavior and machine learning," *Building and Environment*, vol. 129, pp. 96–106, 2018.

[15] A. Ghahramani, C. Tang, and B. Becerik-Gerber, "An online learning approach for quantifying personalized thermal comfort via adaptive stochastic modeling," *Building and Environment*, vol. 92, pp. 86–96, 2015.

[16] A. C. Cosma and R. Simha, "Machine learning method for real-time non-invasive prediction of individual thermal preference in transient conditions," *Building and Environment*, vol. 148, pp. 372–383, 2019.

[17] W. Jung and F. Jazizadeh, "Comparative assessment of hvac control strategies using personal thermal comfort and sensitivity models," *Building and Environment*, vol. 158, pp. 104–119, 2019.

[18] R. Harper, *Inside the smart home.* Springer Science & Business Media, 2006.

[19] L. C. De Silva, C. Morikawa, and I. M. Petra, "State of the art of smart homes," *Engineering Applications of Artificial Intelligence*, vol. 25, no. 7, pp. 1313–1321, 2012.

[20] M. Alaa, A. A. Zaidan, B. B. Zaidan, M. Talal, and M. L. M. Kiah, "A review of smart home applications based on internet of things," *Journal of Network and Computer Applications*, vol. 97, pp. 48–65, 2017.

[21] T. Higashino and A. Uchiyama, "A study for human centric cyber physical system based sensing–toward safe and secure urban life–," in *International Workshop on Information Search, Integration, and Personalization*. Springer, 2012, pp. 61–70.

[22] S. K. Sowe, E. Simmon, K. Zettsu, F. de Vaulx, and I. Bojanova, "Cyber-physical-human systems: Putting people in the loop," *IT professional*, vol. 18, no. 1, pp. 10–13, 2016.

[23] S. Karjalainen, "Thermal comfort and use of thermostats in finnish homes and offices," *Building and Environment*, vol. 44, no. 6, pp. 1237–1245, 2009.

[24] Y. Zhang and P. Barrett, "Factors influencing the occupants' window opening behaviour in a naturally ventilated office building," *Building and Environment*, vol. 50, pp. 125–134, 2012.

[25] I. A. Raja, J. F. Nicol, K. J. McCartney, and M. A. Humphreys, "Thermal comfort: use of controls in naturally ventilated buildings," *Energy and buildings*, vol. 33, no. 3, pp. 235–244, 2001.

[26] A. Ghahramani, G. Castro, B. Becerik-Gerber, and X. Yu, "Infrared thermography of human face for monitoring thermoregulation performance and estimating personal thermal comfort," *Building and Environment*, vol. 109, pp. 1–11, 2016.

[27] Z. Cheng, W. W. Shein, Y. Tan, and A. O. Lim, "Energy efficient thermal comfort control for cyber-physical home system," in *2013 IEEE International Conference on Smart Grid Communications (SmartGridComm)*. IEEE, 2013, pp. 797–802.

[28] L. Pérez-Lombard, J. Ortiz, and C. Pout, "A review on buildings energy consumption information," *Energy and buildings*, vol. 40, no. 3, pp. 394–398, 2008.

[29] A. Ghahramani, F. Jazizadeh, and B. Becerik-Gerber, "A knowledge based approach for selecting energy-aware and comfort-driven hvac temperature set points," *Energy and Buildings*, vol. 85, pp. 536–548, 2014.

[30] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[31] M. C. Du Plessis, G. Niu, and M. Sugiyama, "Analysis of learning from positive and unlabeled data," in *Advances in neural information processing systems*, 2014, pp. 703–711.

[32] Z.-Y. Zhang, P. Zhao, Y. Jiang, and Z.-H. Zhou, "Learning from incomplete and inaccurate supervision," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2019, pp. 1017–1025.

[33] E. Sansone, F. G. De Natale, and Z.-H. Zhou, "Efficient training for positive unlabeled learning," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 11, pp. 2584–2598, 2018.

[34] Wikipedia contributors, "Cohen's kappa — Wikipedia, the free encyclopedia," 2020, [Online; accessed 31-July-2020]. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Cohen%27s_kappa&oldid=969400637

List of Publications

[1] Fang, Y., Lim, Y., Ooi, S.E.; **Zhou, C.** and Tan, Y. "Study of Human Thermal Comfort for Cyber–Physical Human Centric System in Smart Homes," *Sensors* 20, 372. 2020 [SCI, IF: 3.031]